# Lessons 5-6: ARMA, regular and singular processes (Wold decomposition in 1D)

Marianna Bolla and Tamás Szabados, BME Math. Inst.

October 21, 2020

In this lesson we collect some basic facts about important classes of 1D (one-dimensional) stationary time series as a motivation for the multidimensional case. We do it inductively, while proceeding from the simplest 1D processes to more and more general ones.

Through the technique of linear filtering (i.e. applying a time-invariant linear filter TLF), parametric families, such as MA (moving average), AR (autoregressive), and ARMA (both AR and MA) processes are defined. For any 1D real, weakly stationary process with continuous spectral density $f$, it is possible to find both a causal AR and an invertible MA process with spectral density arbitrarily close to $f$. This is because the ARMA processes have rational spectral densities. Therefore, ARMA processes are vital in modelling 1D time series. Also, the linear structure of them is in close relation to the prediction theory of stationary processes.

Under certain conditions, a TLF, applied to a white noise process, results in a sliding summation (two-sided MA). We will see in the multidimensional case that these are the processes with spectral density matrix of constant rank. (In 1D, the spectral density is positive almost surely (a.s.) with respect to the Lebesgue measure on $[-\pi, \pi]$. The special class of them, the MA($\infty$) processes (one-sided MA) are the regular ones.

We also discuss the Wold decomposition of a weakly stationary time series into a regular and singular part. Again, the regular part is MA($\infty$), i.e. a causal (future-independent) TLF. We also consider the spectral form of the Wold decomposition and the types of singularities. It is important, that in the frame of non-singular processes, regular processes can coexist only with

Type (0) singular ones; while adding a regular part to Type (1) or Type (2) singularities, makes them regular.

# 1 Time invariant linear filtering

Time invariant linear filters can be used for smoothing the data, estimating the trend, and eliminating the seasonal and/or trend components of the data, making it stationary or more conveniently, regular. Latter notion is introduced in Section 7 of this lesson. The filters in practice are usually finite, though we first define them as an infinite summation as follows.

Linear filtering of a stationary time series $\{X_t\}_{t\in\mathbb{Z}}$ means applying a TLF (time-invariant linear filter) to it:

$$Y_t := \sum_{j=-\infty}^{\infty} c_{tj} X_j = \sum_{k=-\infty}^{\infty} b_k X_{t-k} \quad (t \in \mathbb{Z}).$$

Time-invariance means that the coefficient $c_{tj}$ depends only on $t - j$, i.e. $c_{tj} = b_{t-j}$, giving the final form of a TLF.

The so obtained $\{Y_t\}$ is also a weakly stationary sequence, with spectral measure

$$dF^Y(\omega) = |\hat{b}(\omega)|^2 \, dF^X(\omega) \tag{1}$$

and the pair $(X_t, Y_t)$ $(t \in \mathbb{Z})$ has a joint spectral measure (a complex measure in general)

$$dF^{Y,X}(\omega) = \hat{b}(\omega) \, dF^X(\omega), \tag{2}$$

where $\hat{b}(\omega) = \sum_{j=-\infty}^{\infty} b_j e^{-ij\omega}$.

It is important that the above formulas are valid when the sequence of weights $(b_j : j \in \mathbb{Z})$ is such that $\hat{b} \in L^2([-\pi, \pi], \mathcal{B}, dF)$. (When $b_j = 0$ for $|j| > N$, the square summability of $b_j$s automatically holds.) In another wording, the stationary time series $\{Y_t\}_{t\in\mathbb{Z}}$ is *subordinated* to the process $\{X_t\}_{t\in\mathbb{Z}}$, i.e., $Y_t \in H(\mathbf{X})$ for all $t$.

**Proposition 1** (Proposition 3.1.1 of Brockwell & Davis)**.** *If $\{X_t\}$ is any sequence of random variables such that $\sup_t \mathbb{E}|X_t| < \infty$, and if $\sum_{k=-\infty}^{\infty} |b_k| < \infty$, then the series*

$$b(L)X_t := \sum_{k=-\infty}^{\infty} b_k L^k X_t = \sum_{k=-\infty}^{\infty} b_k X_{t-k}$$

*converges absolutely with probability 1. If in addition* $\sup_t \mathbb{E}|X_t|^2 < \infty$*, then the series converges in mean square to the same limit.*

Here $L$ is the *backward shift (lag)* operator, so $L^k X_t = X_{t-k}$ for any integer $k$, and $b(z) = \sum_{j=-\infty}^{\infty} b_j z^j$. In case of a weakly stationary $\{X_t\}$, the sup conditions automatically hold.

The second order, weakly stationary 1D time series $\{\xi_t\}_{t\in\mathbb{Z}}$ is called *white noise sequence* if its autocovariances are $c(0) = \sigma^2$ and $c(h) = 0$ for $h = \pm 1, \pm 2, \dots$. In other words, $\xi_t$s are uncorrelated and have variance $\sigma^2$. We use the notation $\xi_t \sim \mathrm{WN}(\sigma^2)$. For example, if $\xi_t$s are i.i.d., with finite variance, they constitute a white noise sequence, and in the Gaussian case, the two notions are the same. When $\sigma^2 = 1$, we call the white noise sequence WN(1) *orthonormal sequence.*

Now the TLF with a white noise sequence is considered, and it is called *causal (future-independent, in other words, regular or purely non-deterministic)* if the so-called *sliding summation* from $-\infty$ to $\infty$ is, in fact, one-sided. It means that $X_t$ depends (randomly) only on the present and past values of $\xi_t$.

# 2 MA (moving average) processes

Let $\{\xi_t : t \in \mathbb{Z}\}$ be a WN(1) white noise sequence of complex valued random variables. Recall that its spectral density function is

$$f^{\xi}(\omega) = \frac{1}{2\pi} \quad (\omega \in [-\pi, \pi]). \tag{3}$$

Define a two-sided infinite complex valued *moving average* (MA) process (a so-called *sliding summation*) by

$$X_t = \sum_{k=-\infty}^{\infty} b_k \xi_{t-k}, \tag{4}$$

where the sequence of the non-random complex coefficients $\{b_k\}$ is in $\ell^2$, that is, $\sum_k |b_k|^2 < \infty$. Then by the Riesz–Fischer theorem, (4) is convergent in mean square. By Parseval's theorem, the covariance function is

$$c(h) = \mathbb{E}(X_{t+h}\bar{X}_t) = \sum_{k=-\infty}^{\infty} b_k \bar{b}_{k-h},$$

so that $\{X_t\}$ is a weakly stationary process. It is not difficult to show that $c(h) \to 0$ as $|h| \to \infty$. The spectral density of the sliding summation process (4) is

$$f^X(\omega) = \frac{1}{2\pi} \left| \sum_{k=-\infty}^{\infty} b_k e^{-ik\omega} \right|^2, \text{ where } \sum_{k=-\infty}^{\infty} |b_k|^2 < \infty. \tag{5}$$

If $b_k = 0$ whenever $k < 0$, then one obtains a *one-sided* (*causal, future-independent*) MA($\infty$) process:

$$X_t = \sum_{k=0}^{\infty} b_k \xi_{t-k}, \quad c(h) = \sum_{k=h}^{\infty} b_k \bar{b}_{k-h} \quad (h \geq 0), \quad c(-h) = \bar{c}(h). \tag{6}$$

This representation of the process is important as $\xi_t, \xi_{t-1}, \dots$ are considered as *random shocks* that influence $X_t$. In the following definition, finitely many random shocks suffice.

**Definition 1.** *The* $q$th order moving average process*, denoted by MA(q), is the following finite moving average:*

$$X_t = \beta(L)\xi_t = \sum_{k=0}^{q} \beta_k \; \xi_{t-k}, \tag{7}$$

*where L is the lag operator and it is substituted for the complex variable z of the following MA(q) polynomial:*

$$\beta(z) = \sum_{k=0}^{q} \beta_k z^k. \tag{8}$$

The covariance function of the MA($q$) process is

$$c(h) = \sum_{k=h}^{q} \beta_k \bar{\beta}_{k-h} \quad (0 \leq h \leq q), \quad c(-h) = \bar{c}(h), \quad c(h) = 0 \text{ if } |h| > q.$$

Conversely, the following is true.

**Remark 1** (Proposition 3.2.1 of Brockwell & Davis)**.** *If the autocovariance function of the zero mean stationary process is such that $c(h) = 0$ for $|h| > q$ and $c(q) \neq 0$, then it is a MA(q) process.*

The spectral density of an MA($q$) process is

$$f^X(\omega) = \frac{1}{2\pi}|\hat{\beta}(\omega)|^2, \qquad \hat{\beta}(\omega) = f^{X|\xi}(\omega) = \sum_{k=0}^{q} \beta_k e^{-ik\omega}. \tag{9}$$

Observe that $f^X(\omega)$ is a $2q$th degree polynomial of $e^{-i\omega}$. A bit more general, so-called rational spectral densities come into existence in the subsequent sections.

# 3 AR (autoregressive) processes

Next let us consider the *first order autoregressive process* AR(1), which is, in fact, a first order stochastic linear difference equation:

$$X_t = \alpha X_{t-1} + \beta \xi_t, \quad t \in \mathbb{Z},$$

where $\alpha$ and $\beta \neq 0$ are complex constants. We assume that $\{\xi_t\}$ is a WN(1) (orthonormal) sequence and each $X_t$ depends only on the present and past values $(\xi_t, \xi_{t-1}, \dots)$ of the driving white noise process. Iterating the equation $k$ times we get

$$X_t = \beta\xi_t + \alpha\beta\xi_{t-1} + \alpha^2\beta\xi_{t-2} + \dots + \alpha^k\beta\xi_{t-k} + \alpha^{k+1}X_{t-k-1}.$$

In order that the right side have bounded norm as $k \to \infty$, it is necessary that $|\alpha| < 1$ (this will be called stability). If that is so, the series $\sum_{k=0}^{\infty} \alpha^k \xi_{t-k}$ converges and $\alpha^{k+1}X_{t-k-1} \to 0$ as $k \to \infty$, since we are looking for a stationary process, where $\|X_t\| = (\mathbb{E}(|X_t|^2))^{1/2}$ is constant. Then

$$X_t = \beta \sum_{k=0}^{\infty} \alpha^k \xi_{t-k}, \quad t \in \mathbb{Z},$$

converges in mean square, and it is the only stationary solution of the AR(1) process. From the above form, we can as well see that it is a causal MA($\infty$) process. Its covariance function is

$$c(h) = |\beta|^2 \frac{\alpha^h}{1-|\alpha|^2} \quad (h \in \mathbb{Z}),$$

converging to 0 exponentially fast.

Then we generalize the above recursion to the situation, when $X_t$ depends on its $p$ past values, above the previous white noise term.

**Definition 2.** *A* $p$th order *autoregressive process AR(p), or a pth order stochastic linear difference equation, is*

$$X_t = \sum_{j=1}^{p} \alpha_j X_{t-j} + \beta \xi_t, \quad \beta \neq 0, \quad t \in \mathbb{Z}, \tag{10}$$

*or concisely,*

$$\alpha(L) X_t = \beta \xi_t,$$

*where* $L$ *is the lag operator and*

$$\alpha(z) = 1 - \alpha_1 z - \cdots - \alpha_p z^p \tag{11}$$

*is the AR(p) polynomial; further,* $\{\xi_t\} \sim \mathrm{WN}(1)$ *is white noise sequence.*

This means that $X_t$ is determined by finitely many preceding past values of itself plus a random shock. If the shocks are accumulated in a recursion, we can write $X_t$ as the infinite summation of those.

To find a stationary solution of this equation, let us look for a causal MA($\infty$) solution of form

$$X_t = \sum_{k=0}^{\infty} b_k \xi_{t-k}, \quad t \in \mathbb{Z}, \quad b_k \in \mathbb{C}. \tag{12}$$

Substitute this in Equation (10) and equate the coefficients of $\xi_\ell$ on both sides, starting with $\xi_t$ and working toward the past:

$$\begin{aligned} b_0 &= \beta, \\ b_1 - \alpha_1 b_0 &= 0, \\ b_2 - \alpha_1 b_1 - \alpha_2 b_0 &= 0, \\ &\vdots \\ b_p - \alpha_1 b_{p-1} - \cdots - \alpha_p b_0 &= 0, \\ b_{p+k} - \alpha_1 b_{p+k-1} - \cdots - \alpha_p b_k &= 0 \qquad (k \geq 1). \end{aligned} \tag{13}$$

If $\{\alpha_j : j = 1, \ldots, p\}$ and $\beta$ are known, these equations uniquely determine the coefficients $\{b_k : k \geq 0\}$ by recursion. The question is whether or not this sequence will be square-summable, so the proposed solution will converge in mean square.

If each (complex) root of the AR($p$) polynomial is greater than 1 in absolute value, then the sequence $\{b_0, b_1, b_2, \dots\}$ will be square summable. Equivalently, if *there are no roots of the* AR($p$) *polynomial on the closed unit disc* ($|z| \le 1$), then (12) and (13) give a causal MA($\infty$) stationary solution of the AR($p$) process (10). In this case we call the AR($p$) process *stable*. Let us check it for $p = 1$. The AR(1) polynomial with the only $\alpha$ is $1 - \alpha z$. For its root $z$, $\alpha z = 1$, so $|z| = 1/|\alpha| > 1$ holds if and only if $|\alpha| < 1$ that was necessary and sufficient for stability.

The covariance function of AR($p$) can be obtained in terms of the coefficients $\{b_k\}$ by (12):

$$c(h) = \mathbb{E}(X_{t+h}\bar{X}_t) = \sum_{k=h}^{\infty} b_k \bar{b}_{k-h} \quad (h \ge 0), \quad c(-h) = \bar{c}(h), \tag{14}$$

and $c(h)$ converges to 0 exponentially fast as $|h| \to \infty$.

However, it is more important to describe the covariance function from the defining formula (10). Multiply the complex conjugate of (10) by $X_{t-k}$, $k \ge 0$, and take expectation. This way we obtain the important *Yule–Walker equations*:

$$c(-k) - \sum_{j=1}^{p} c(j-k)\bar{\alpha}_j = \delta_{0k}\bar{\beta}, \qquad k \ge 0. \tag{15}$$

The significance of the Yule–Walker equations is that taking them for $0 \le k \le p$, one obtains a system of linear equations for the unknowns $\alpha_1, \dots, \alpha_p$ ($\alpha_0 = 1$) and $\beta$ if the covariances $c(0), c(1), \dots, c(p)$ are known:

$$\begin{bmatrix} c(0) & c(1) & \dots & c(p-1) \\ \bar{c}(1) & c(0) & \dots & c(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \bar{c}(p-1) & \bar{c}(p-2) & \dots & c(0) \end{bmatrix} \begin{bmatrix} \bar{\alpha}_1 \\ \bar{\alpha}_2 \\ \vdots \\ \bar{\alpha}_p \end{bmatrix} = \begin{bmatrix} \bar{c}(1) \\ \bar{c}(2) \\ \vdots \\ \bar{c}(p) \end{bmatrix},$$
$$\bar{\beta} = c(0) - c(1)\bar{\alpha}_1 - c(2)\bar{\alpha}_2 - \dots - c(p)\bar{\alpha}_p. \tag{16}$$

The matrix of the system is the $p \times p$ Toeplitz matrix $\boldsymbol{C}_p$, defined in Lesson 1, self-adjoint and non-negative definite. Introducing $\boldsymbol{\alpha} := [\alpha_1, \dots, \alpha_p]^T$, $\mathbf{c} := [c(1), \dots, c(p)]^T$, we can write the above system of linear equations as

$$\boldsymbol{C}_p\, \bar{\boldsymbol{\alpha}} = \bar{\mathbf{c}}. \tag{17}$$

If $\boldsymbol{C}_p$ is positive definite, then there exists a unique solution: $\bar{\boldsymbol{\alpha}} = \boldsymbol{C}_p^{-1}\bar{\mathbf{c}}$. Otherwise, one may use the Moore–Penrose inverse $\boldsymbol{C}_p^{-}$ instead. The system (17) is always consistent, always has a solution, because it is a Gauss normal equation, we will further consider in the context of predictions. Indeed, taking $\mathbf{X} := [X_{p-1}X_{p-2}\dots X_0]^T$, we have

$$\boldsymbol{C}_p = \mathbb{E}(\mathbf{X}\mathbf{X}^*), \quad \bar{\mathbf{c}} = \mathbb{E}(\mathbf{X}\bar{X}_p),$$

and this shows that the right hand side of (17) is in the column space of the left hand side.

If the AR($p$) process is stable, then $\text{rank}(\boldsymbol{C}_p) = p$, so the system (17) has a unique solution. In case of a real time series no conjugation is needed, and $\boldsymbol{\alpha} = (\boldsymbol{C}_p^T)^{-1}\mathbf{c}$, which is exactly the same as the solution of the Gauss normal equation discussed in the subsequent lessons.

Since the autocovariances can be easily estimated from a random sample, the resulting estimated version of (16) gives a practical method for estimating the coefficients $\alpha_1, \dots, \alpha_p$ and $\beta$ of an AR($p$) process. This issue will be further discussed in the prediction lesson.

On the other hand, if the coefficients $\alpha_1, \dots, \alpha_p$ and $\beta$ are known, then one can determine the covariances $c(0), c(1), \dots, c(p)$ from the first $(p+1)$ Yule–Walker equations (15), and then one can determine $c(p+1), c(p+2), \dots$ recursively.

Let us find the spectral density of an AR($p$) process $\{X_t\}$, assuming that no roots of the AR($p$) polynomial lie within the closed unit disc. This is

$$f^X(\omega) = \frac{1}{2\pi}\frac{1}{|\alpha(e^{-i\omega})|^2}. \tag{18}$$

Observe that $f^X(\omega)$ is the reciprocal of a $2p$th degree polynomial of $e^{-i\omega}$.

# 4 ARMA (both autoregressive and moving average) processes

The ARMA($p, q$) processes ($p \geq 0$, $q \geq 0$) are generalizations of both AR($p$) and MA($q$) processes.

**Definition 3.** *With $p, q \geq 0$ integers, $\{X_t\}$ is a $p$th order autoregressive and* *qth order moving average process (at the same time) if*

$$X_t = \sum_{j=1}^{p} \alpha_j X_{t-j} + \sum_{j=0}^{q} \beta_j \xi_{t-j}, \quad t \in \mathbb{Z}, \tag{19}$$

*or concisely,*

$$\alpha(L)X_t = \beta(L)\xi_t,$$

*where $L$ is the lag operator and $\alpha(z)$ and $\beta(z)$ are the previously defined AR($p$) and MA($q$) polynomials, respectively; further, $\{\xi_t\} \sim \mathrm{WN}(1)$ is white noise sequence.*

Again, we want to find a causal stationary solution $\{X_t\}$ of this equation, that is, a MA solution of form (12). The following can be stated.

**Theorem 1** (Theorem 3.1.1 of Brockwell & Davis)**.** *Let $\{X_t\}$ be an ARMA($p,q$) process for which the polynomials $\alpha(z)$ and $\beta(z)$ have no common zeros. Then $\{X_t\}$ is causal if and only if $\alpha(z) \neq 0$ for all $|z| \leq 1$ (stability). The coefficients of the MA($\infty$) process $X_t = \sum_{j=0}^{\infty} b_j \xi_{t-j}$ are obtainable by the power series expansion*

$$b(z) = \sum_{j=0}^{\infty} b_j z^j = \alpha^{-1}(z)\beta(z), \quad |z| \leq 1. \tag{20}$$

The coefficients $b_j$s of the so-called *transfer function* $b(z)$ are called *impulse responses.*

In the ARMA model $X_t$ is determined by finitely many preceding past values of itself plus finitely many random shocks. Here $X_t$ can also be written as the infinite sum of impulse responses multiplied by random shocks as follows.

After multiplying both sides of Equation (20) with $\alpha(z)$ (we can do it as it is not zero for $|z| \leq 1$) and equating the coefficients of the $z$-powers on both sides, gives the following recursion for $b_j$s:

$$\begin{aligned}
b_0 &= \beta_0, \\
b_1 &= \beta_1 + \alpha_1 b_0, \\
b_2 &= \beta_2 + \alpha_1 b_1 + \alpha_2 b_0, \\
&\vdots \\
b_q &= \beta_q + \alpha_1 b_{q-1} + \cdots + \alpha_q b_0, \\
&\vdots \\
b_p &= \alpha_1 b_{p-1} + \cdots + \alpha_p b_0, \\
b_{p+k} &= \alpha_1 b_{p+k-1} + \cdots + \alpha_p b_k \qquad (k \geq 1).
\end{aligned} \tag{21}$$

Here we assumed that $q < p$ (in the opposite case, similar formulas are valid). If $\{\alpha_j : j = 1, \ldots, p\}$ and $\{\beta_j : j = 1, \ldots, q\}$ are known, these equations uniquely determine the coefficients $\{b_k : k \geq 0\}$ by the above recursion. Also, the previous Equations (13) are special cases of the present Equations (21) when $q = 0$.

The spectral density of $\{X_t\}$ is

$$f^X(\omega) = \frac{1}{2\pi} \frac{|\beta(e^{-i\omega})|^2}{|\alpha(e^{-i\omega})|^2}. \tag{22}$$

Observe that this is a rational function (ratio of polynomials) of $e^{-i\omega}$. Rational spectral densities play an important role in state space models and those belong, in fact, to ARMA models as we will see in the next lesson. It can also be shown that any continuous spectral density can be approximated with rational ones with any small error, albeit with possibly large $p$ and/or $q$.

**Example 1** (Examples 4.2.2 and 4.4.1 from Brockwell and Davis: Introduction to Time Series and Forecastig, Springer)**.** *Yearly sunspot numbers are observed spanning 1770-1869. Based on experiences (the first two autocovariances are significant), an AR(2) model was fitted. The spectral density with the estimated parameters on* $[0, \pi]$ *had a peak at* $\omega = 0.556$ *radians per year. Therefore, the corresponding period is* $2\pi/0.556 = 11.3$ *years. (This is the period of the* $\sin(0.556t)$ *function, which is a term of finite sum of simusoids, see Example 1 of Lesson 2). The model thus reflects the approximating cyclic behavior of the data with around 11 years of cycles. Also, if we consider* *the 120 Fourier frequencies, the peak happens at about the 11th Fourier frequency. It means that the frequency domain* $[0, 0.556]$ *roughly explains the process; extrapolating, about 11 years contain the relevant information.*

*In another context, based on Equation (2) of Example 1 of Lesson 2: this finite time series can be approximated with the only sinusoid of form* $A\sin(\omega t)$*, where* $\omega = 0.556$*. The period of this function is* $2\pi/0.556 = 11.3$ *in years, approximately. Since* $F(\omega) = \int_0^\omega f(s)\, ds$*, and* $F(\omega)$ *is the variance of* $Z_\omega$*, the majority of the variance of the orthogonal increment process is explained until the frequency* $0.556$ *in the frequency domain. We may also think that the principal component* $Z_{0.556}$ *explains best the process, and it is around at the first Fourier-frequency of the time series until the time period 11.3. About principal components (PC) see Remark 2 of Lesson 2.*

# 5 Wold decomposition in 1D

Recall that we use the notation $\overline{\text{span}}\{X_t : t \in A\}$ for the closed linear span of the random variables $\{X_t : t \in A\} \subset L^2(\Omega, \mathcal{F}, \mathbb{P})$. Then for a weakly stationary time series $\{X_t\}_{t\in\mathbb{Z}}$, we define

$$H_n^- = \overline{\text{span}}\{X_t : t \leq n\} \quad (n \in \mathbb{Z}), \quad H_{-\infty} = \bigcap_{n\in\mathbb{Z}} H_n^-.$$

$H_n^-$ is called the *past of* $\{X_t\}$ *until* $n$ and $H_{-\infty}$ is the *remote past of* $\{X_t\}$. Clearly, $H(X) = \overline{\text{span}}_{t\in\mathbb{Z}} H_n^-$ and by the weak stationarity of $\{X_t\}_{t\in\mathbb{Z}}$, $H_n^- = S^n H_0^-$ $(t \in \mathbb{Z})$, where $S$ is the unitary operator of right time-shift. The time series is called *singular (deterministic)* if $H_{-\infty} = H(X)$, or equivalently, if $H_k^- = H_{k+1}^-$ for some $k \in \mathbb{Z}$, that is, if $H_k^- = H_{k+m}^-$ for any $k, m \in \mathbb{Z}$; otherwise, it is non-singular. The time series is *regular* if $H_{-\infty} = \{0\}$. We are going to show that in general, a weakly stationary time series can be written as an orthogonal sum of a regular and a singular time series. So regularity means no remote past at all (such sequences are also called *purely non-deterministic*), whereas singularity means an overwhelming remote past (such sequences are also called *deterministic*). Between these two extremes there are the *non-singular* processes that are not purely deterministic. They can be purely non-deterministic (regular) or must have a regular part. In the latter case, we will see that the regular part can coexist only with a special type of singularity.

Assume that $\{X_t\}_{t\in\mathbb{Z}}$ is non-singular, so $H_{-1}^- \neq H_0^- = \overline{\text{span}}\{H_{-1}^-, X_0\}$. Let

$$X_0 = X_0^- + X_0^+, \quad X_0^- \in H_{-1}^-, \quad X_0^+ \perp H_{-1}^-.$$

Define the random variable $\xi_0 := X_0^+/\|X_0^+\|$. Then $\xi_0 \in H_0^-$, $\|\xi_0\| = 1$, $\xi_0 \perp H_{-1}^-$. Define $\xi_n := S^n \xi_0$ $(t \in \mathbb{Z})$. Clearly, $\xi_n \in H_n^-$, $\xi_n \perp H_{n-1}^-$, and $H_n^- = \overline{\text{span}}\{H_{n-1}^-, \xi_n\}$. Thus $\{\xi_n\}_{n\in\mathbb{Z}}$ is an orthonormal sequence and $\xi_n \perp H_{-\infty}$ for each $n$. This procedure resembles the Gram–Schmidt orthogonalization. The orthogonal components are sometimes called *innovations*, and we will use matrix decompositions to obtain them, in the prediction lesson.

Now let us expand $X_0$ into its orthogonal series w.r.t. $\{\xi_n\}_{t\in\mathbb{Z}}$:

$$X_0 = \sum_{k=0}^{\infty} b_k \xi_{-k} + Y_0. \tag{23}$$

Here $b_k = \langle X_0, \xi_{-k}\rangle$, $\sum_k |b_k|^2 < \infty$, and $b_k = 0$ for $k < 0$ because $\xi_k \perp H_0^-$ when $k \geq 1$. In particular,

$$b_0 = \langle X_0, \xi_0\rangle = \langle X_0^+, X_0^+/\|X_0^+\|\rangle = \|X_0^+\| > 0. \tag{24}$$

The vector $Y_0$ is simply the remainder term, which of course is 0 if $\{\xi_n\}_{t\in\mathbb{Z}}$ span $H(X)$, but not in general. It is not hard to see that $Y_0 \in H_{-\infty}$ and $\xi_{-k} \perp H_{-\infty}$ for any $k \geq 0$.

Now apply the operator $S^t$ to (23):

$$X_t = \sum_{k=0}^{\infty} b_k \xi_{t-k} + Y_t =: R_t + Y_t \quad (t \in \mathbb{Z}), \tag{25}$$

where we have defined $Y_t := S^t Y_0$ $(t \in \mathbb{Z})$. This way we have proved the important *Wold decomposition* of $\{X_t\}_{t\in\mathbb{Z}}$.

**Theorem 2.** *Assume that $\{X_t\}_{t\in\mathbb{Z}}$ is a* non-singular *weakly stationary time series. Then we can decompose $\{X_t\}$ in the form* (25)*, where $\{R_t\}_{t\in\mathbb{Z}}$ is a regular time series (that is, a causal* MA($\infty$) *process) and $\{Y_t\}_{t\in\mathbb{Z}}$ is a singular time series, $Y_t \in H_{-\infty}$ for all t; the two processes are orthogonal to each other.*

The *best linear t-step ahead prediction* $\hat{X}_t$ of $X_t$ by definition the projection of $X_t$ to the past until 0, that is, to $H_0^-$. (Time 0 is considered the present and $\hat{X}_t$ is called a $t$-step ahead prediction.) By the Wold decomposition (25) and the projection theorem, the best prediction is

$$\hat{X}_t = \sum_{k=t}^{\infty} b_k \xi_{t-k} + Y_t \quad (t \in \mathbb{Z}), \tag{26}$$

since the right hand side of (26) is in $H_0^-$ and the difference $(X_t - \hat{X}_t)$ is orthogonal to $H_0^-$. Hence, the prediction error (the mean-square error) of the $t$-step ahead prediction is given by

$$\sigma_t^2 := \|X_t - \hat{X}_t\|^2 = \sum_{k=0}^{t-1} |b_k|^2. \tag{27}$$

This implies that

$$\lim_{t\to\infty} \sigma_t^2 = \sum_{k=0}^{\infty} |b_k|^2 = \left\| \sum_{k=0}^{\infty} b_k \xi_{\ell-k} \right\|^2 \quad (\ell \in \mathbb{Z}), \qquad \lim_{t\to\infty} \sum_{k=t}^{\infty} b_k \xi_{t-k} = 0.$$

Clearly, the MA part $\{R_t\}_{t\in\mathbb{Z}}$ of $\{X_t\}_{t\in\mathbb{Z}}$ is a regular time series. For, the past $H_n^-(R) := \overline{\text{span}}\{R_t : t \le n\}$ of $\{R_t\}$ is spanned by the vectors $\{\xi_t : t \le n\}$, so

$$H_{-\infty}(R) := \bigcap_{n\in\mathbb{Z}} H_n^-(R) = \lim_{n\to-\infty} H_n^-(R) = \{0\}.$$

If the process $\{X_t\}_{t\in\mathbb{Z}}$ itself is regular, then $Y_t = 0$ for each $t \in \mathbb{Z}$,

$$\lim_{t\to\infty} \sigma_t^2 \to \|X_\ell\|^2 \quad (\ell \in \mathbb{Z}), \text{ and } \lim_{t\to\infty} \hat{X}_t = 0.$$

It is also easy to see that $\{Y_t\}_{t\in\mathbb{Z}}$ is a singular process and spans $H_{-\infty}$. Clearly, if a process $\{X_t\}_{t\in\mathbb{Z}}$ is singular, then $X_t \in H_0^-$ for any $t \in \mathbb{Z}$, so perfect linear prediction is possible: $\hat{X}_t = X_t$ for any $t \in \mathbb{Z}$.

## 6 Spectral form of the Wold decomposition

The regular part of a non-singular time series $\{X_t\}$ is a causal MA($\infty$) process, so by (5) it has an absolutely continuous spectral measure with density

$$f(\omega) = \frac{1}{2\pi} \left| \sum_{j=0}^{\infty} b_j e^{-ij\omega} \right|^2 =: \frac{1}{2\pi} |B(e^{-i\omega})|^2, \tag{28}$$

$$B(z) = \sum_{j=0}^{\infty} b_j z^j \quad (|z| \leq 1), \quad \sum_{j=0}^{\infty} |b_j|^2 < \infty. \tag{29}$$

Therefore $B(z)$ is analytic in the open unit disc $D = \{z : |z| < 1\}$ and is in $L^2$ on the unit circle $T$. Clearly, these constitute a necessary and sufficient condition for the regularity of a stationary time series.

By a theorem of Frigyes and Marcel Riesz, an $L^2(T)$ function $f$ of "power series type", that is, with Fourier coefficients $b_j = 0$ when $j < 0$ and not identically 0, vanishes only on a set of measure 0, so it is positive a.e. on $T$. It means that $f \in H^2$ (Hardy space). This applies to the spectral density $f$ of the regular part of any non-singular process as well; see Theorems 4 and 5 below.

The next lemma shows that the Wold decomposition (25) is a spectral decomposition in the sense too that the support of the spectral measure of the singular process $\{Y_t\}$ must be disjoint from the set $\{\omega \in [-\pi, \pi] : f(\omega) \neq 0\}$. Consequently, the spectral measure of $\{Y_t\}$ is a singular measure w.r.t. Lebesgue measure. This way we see that the Wold decomposition (25) of $\{X_t\}$ is equivalent to a decomposition of a non-singular process into an absolutely continuous and a singular part of its spectral measure w.r.t. Lebesgue measure on $[-\pi, \pi]$.

**Lemma 1.** *Assume that $\{X_t\}$ is a stationary time series, $X_t = U_t + V_t$, where $U_t, V_t \in H(X)$ for all $t \in \mathbb{Z}$, $\{U_t\}$ and $\{V_t\}$ are stationary processes, and $\langle U_t, V_t \rangle = 0$ for all $t, s$. Then there is a decomposition $A \cup B = [-\pi, \pi]$, $A \cap B = \emptyset$, $A$ and $B$ are measurable, and*

$$X_t = \int_{-\pi}^{\pi} e^{it\omega} dZ_\omega = \int_{-\pi}^{\pi} e^{it\omega} \mathbf{1}_A dZ_\omega + \int_{-\pi}^{\pi} e^{it\omega} \mathbf{1}_B dZ_\omega = U_t + V_t \quad (t \in \mathbb{Z}).$$

**Theorem 3.** *Assume that the spectral density $f$ of a regular stationary time series $\{X_t\}$ has a factorization $f = \frac{1}{2\pi} \phi \cdot \overline{\phi}$, where $\overline{\phi} \in H^2$ and $1/\overline{\phi} \in H^2$.*

*Then the orthonormal sequence appearing in the Wold decomposition is given by the random variables*

$$\xi_t = \int_{-\pi}^{\pi} \frac{e^{it\omega}}{\phi(\omega)} dZ_\omega \quad (t \in \mathbb{Z}). \tag{30}$$

Later, $\phi$ will be called *spectral factor*, and will be dealt in more details in the multidimensional case.

**Corollary 1.** *Under the assumptions of Theorem 3, we have the Fourier series*

$$\phi(\omega) = \sum_{k=0}^{\infty} b_k e^{-ik\omega}, \ \textit{where} \ \sum_{k=0}^{\infty} |b_k|^2 < \infty, \tag{31}$$

*and this defines the coefficients of the Wold representation:*

$$X_t = \sum_{k=0}^{\infty} b_k \xi_{t-k}. \tag{32}$$

Theorem 3 and Corollary 1 together give an explicit solution to causal MA($\infty$) representation of regular stationary time series under conditions that, according to equations (28) and (29) and Lemma 2 below, are almost necessary. If $\{X_t\}$ is an arbitrary non-singular time series, then Theorem 3 and Corollary 1 are valid for its regular part.

# 7 Classification of stationary time series in 1D

The theorems in this section are based on the seminal paper "Stationary sequences in Hilbert space" by A.N. Kolmogorov from 1941.

Wold decomposition in Section 5 showed that a stationary time series $\{X_t\}_{t\in\mathbb{Z}}$ is regular if and only if it can be represented as a causal MA process

$$X_t = \sum_{j=0}^{\infty} b_j \xi_{t-j}, \tag{33}$$

with an orthonormal sequence $\{\xi_t\}_{t\in\mathbb{Z}}$. In turn, in (28) and (29) we saw that these are equivalent to the fact that the spectral density function $f^X$ of $X$

can be written as $f^X(\omega) = \frac{1}{2\pi}|B(e^{-i\omega})|^2$, where $B(z)$ is analytic in the open unit disc $D$ and is in $L^2$, more accurately, is in $H^2$, on the unit circle $T$. The next lemma gives an even more precise and interesting description of $B$.

**Lemma 2.** *For any regular stationary time series $\{X_t\}_{t\in\mathbb{Z}}$, the analytic function $B(z)$ has no zeros in the open unit disc $D$.*

**Theorem 4.** *A stationary time series $\{X_t\}_{t\in\mathbb{Z}}$ is regular if and only if the following three conditions hold on $[-\pi, \pi]$:*

1. *its spectral measure $dF$ is absolutely continuous w.r.t. Lebesgue measure;*
2. *its spectral density $f$ is positive almost everywhere;*
3. (Kolmogorov's condition) $\log f$ *is integrable.*

**Remark 2.** *[Kolmogorov–Szegő formula] If $\{X_t\}_{t\in\mathbb{Z}}$ is regular, then*

$$b_0 = B(0) = \sqrt{2\pi}e^{Q(0)} = \sqrt{2\pi}e^{\alpha_0} = \sqrt{2\pi}\exp\left(\frac{1}{4\pi}\int_{-\pi}^{\pi}\log f(\omega)d\omega\right),$$

*so*

$$\sigma_1^2 = b_0^2 = 2\pi\exp\int_{-\pi}^{\pi}\log f(\omega)\frac{d\omega}{2\pi}.$$

The next theorem gives the different classes of singular time series.

**Theorem 5.** *Assume that $\{X_t\}_{t\in\mathbb{Z}}$ is a stationary time series with spectral measure $dF$ on $[-\pi, \pi]$. Then there exists a unique Lebesgue decomposition*

$$dF = dF_a + dF_s, \quad dF_a \ll d\omega, \quad dF_s \perp d\omega, \tag{34}$$

*where $d\omega$ denotes the Lebesgue measure, $dF_a(\omega) = f_a(\omega)d\omega$ is the absolutely continuous part of $dF$ with density $f_a$ and $dF_s$ is the singular part of $dF$, which is concentrated on a zero Lebesgue measure subset of $[-\pi, \pi]$.*

*The following three cases are distinguished:*

*(1) $f_a(\omega) = 0$ on a set of positive Lebesgue measure on $[-\pi, \pi]$;*

*(2) $f_a(\omega) > 0$ a.e. , but $\int_{-\pi}^{\pi}\log f_a(\omega)d\omega = -\infty$;*

*(3) $f_a(\omega) > 0$ a.e. and $\int_{-\pi}^{\pi}\log f_a(\omega)d\omega > -\infty$.*

*Then in cases (1) and (2), the time series X is singular. In case (3), the time series $\{X_t\}$ is non-singular, so we may apply the Wold decomposition to write*

$$X_t = R_t + Y_t = \sum_{j=0}^{\infty} b_j \xi_{t-j} + Y_t \quad (t \in \mathbb{Z}), \tag{35}$$

*where $\{R_t\}$ is a regular time series with absolutely continuous spectral measure $dF_a(\omega) = f_a(\omega)d\omega$ and $\{Y_t\}$ is a singular time series with singular spectral measure $dF_s$, as described by* (34).

**Corollary 2.** *In the proof of the previous theorem we saw that for any non-singular time series $\{X_t\}$ the Wold decomposition $X_t = R_t + Y_t$ gives a regular process $\{R_t\}$ whose spectral density function $f^R$ is a.e. the same as the spectral density function $f_a$ of $\{X_t\}$. Thus the Kolmogorov-Szegő formula (Remark 2) is valid for any non-singular process $\{X_t\}$ and its spectral density function $f_a$.*

# 8 Examples for singular time series

## 8.1 Type (0) singular time series

In the Lebesgue decomposition (34) one can further decompose the singular spectral measure:

$$dF_s = dF_d + dF_c,$$

where $dF_d$ is the discrete spectrum corresponding to at most countable many jumps of the spectral distribution function $F$, while $dF_c$ is the continuous singular spectrum.
(a) An example for a process with discrete spectrum:

$$X_t = \sum_{j=1}^{n} A_j e^{it\omega_j} \quad (t \in \mathbb{Z}),$$

where $-\pi < \omega_1 < \cdots < \omega_n \leq \pi$; $A_1, \dots, A_n$ are uncorrelated random variables with mean 0 and variance $\sigma_j^2$ ($j = 1, \dots, n$). (The $A_j$'s can be e.g. Gaussian random variables.)

This process is weakly stationary with

$$c(k) = \mathbb{E}(X_{t+k}\overline{X_t}) = \sum_{j=1}^{n} \mathbb{E}(|A_j|^2) e^{ik\omega_j} = \sum_{j=1}^{n} \sigma_j^2 e^{ik\omega_j} = \int_{\pi}^{\pi} e^{ik\omega}\, dF(\omega),$$

where

$$F(\omega) = \sum_{\omega_j \le \omega} \sigma_j^2 \quad (\omega \in [-\pi, \pi]).$$

Observe that the covariance function does not tend to 0 as $k \to \infty$.
(b) The standard example for a continuous singular function on $[0, 1]$ is the Cantor function $\gamma$, "the devil's ladder". Suppose $\mathcal{C}$ is the Cantor set in $[0, 1]$, that is, $x \in \mathcal{C}$ if and only if in base 3 expansion,

$$x = \sum_{n=1}^{\infty} a_n 3^{-n}, \quad a_n = 0 \text{ or } 2.$$

Then the Cantor function $\gamma : [0, 1] \to [0, 1]$ can be defined as

$$\gamma(x) = \begin{cases} \sum_{n=1}^{\infty} \frac{1}{2} a_n 2^{-n}, & x = \sum_{n=1}^{\infty} a_n 3^{-n} \in \mathcal{C}; \\ \sup\{\gamma(y) : y \le x,\ y \in \mathcal{C}\}, & x \in [0, 1] \setminus \mathcal{C}. \end{cases}$$

Then $\gamma(0) = 0$, $\gamma(1) = 1$, $\gamma$ is non-decreasing on $[0, 1]$, and $\gamma'(x) = 0$ for a.e. $x \in [0, 1]$.

The definition

$$F(\omega) = \gamma\left(\frac{\omega + \pi}{2\pi}\right), \quad \omega \in [-\pi, \pi],$$

gives the spectral distribution function of a singular stationary time series $X$. Heuristically, the spectrum of $X$ consists of the points of an uncountable but zero Lebesgue measure Cantor set, with infinitesimally small amplitudes.

## 8.2 Type (1) singular time series

A simple example for a singular time series corresponding to case (1) of Theorem 5 is the one with spectral density function

$$f(\omega) = \begin{cases} \frac{1}{2}, & |\omega| \le 1; \\ 0, & 1 < |\omega| \le \pi. \end{cases}$$

With the construction of Lesson 3-4, one can construct a singular stationary time series with this spectral density.

Its covariance function is

$$c(k) = \int_{-1}^{1} e^{ik\omega} \frac{1}{2} d\omega = \begin{cases} 1, & k = 0; \\ \frac{\sin k}{k}, & k \ne 0. \end{cases}$$

It is an example for an absolutely non-summable covariance function which still corresponds to an absolutely continuous spectral measure. On the other hand, $\sum_k c(k)$ converges conditionally; also, $\sum_k |c(k)|^2 < \infty$.

Generalizing the previous example, observe the following interesting phenomenon. For any $\delta > 0$ fixed, a time series with spectral density

$$f(\omega) = \begin{cases} \frac{1}{2(\pi-\delta)}, & |\omega| \le \pi - \delta; \\ 0, & \pi - \delta < |\omega| \le \pi. \end{cases}$$

is still singular, like the one above. On the other hand, if we take $\delta = 0$, that is, $f(\omega) = \frac{1}{2\pi}$ for any $\omega \in [-\pi, \pi]$, then the time series becomes a regular, orthonormal series.

## 8.3 Type (2) singular time series

An example for Case (2) of Theorem 5 is the following spectral density:

$$f(\omega) = e^{-\frac{1}{|\omega|}}, \quad \omega \in [-\pi, \pi] \setminus \{0\}, \quad f(0) = 0.$$

Then $f(\omega) > 0$ almost everywhere and $f$ is continuous everywhere on $[-\pi, \pi]$,

$$\int_{-\pi}^{\pi} f(\omega) d\omega < \infty, \quad \int_{\pi}^{\pi} \log f(\omega) d\omega = \int_{-\pi}^{\pi} -\frac{1}{|\omega|} = -\infty.$$

By the construction of Lesson 3-4, one can construct a singular stationary time series $\{X_t\}$ with this spectral density. A theorem later shows that a time series can be represented as a two-sided infinite MA (a sliding summation) if and only if it has constant rank, that is, in 1D its spectral density is positive a.e., like in the case of $\{X_t\}$. However, since this $\{X_t\}$ is singular, it cannot be represented as a causal (one-sided) infinite MA. In general, in 1D the same is true for any singular time series of Type (2) and only for these ones.

**Corollary 3.** *Theorem 5 has the following interesting corollary. Any non-singular process may contain a Type (0) singular component, but cannot contain a Type (1) or (2) singular part. For, if case (1) or (2) of Theorem 5 holds for a time series then that process must be singular. Adding an orthogonal regular part to a Type (1) or (2) singular process results in a regular process.*

# 9 Summary

Linear filtering of a stationary time series $\{X_t\}_{t\in\mathbb{Z}}$ means applying a TLF (time-invariant linear filter) to it:

$$Y_t := \sum_{j=-\infty}^{\infty} c_{tj} X_j = \sum_{k=-\infty}^{\infty} b_k X_{t-k} \quad (t \in \mathbb{Z}).$$

Time-invariance means that the coefficient $c_{tj}$ depends only on $t-j$, i.e. $c_{tj} = b_{t-j}$, giving the final form of a TLF. The so obtained $\{Y_t\}$ is also a weakly stationary sequence, with spectral measure

$$dF^Y(\omega) = |\hat{b}(\omega)|^2 \, dF^X(\omega) \tag{36}$$

and the pair $(X_t, Y_t)$ $(t \in \mathbb{Z})$ has a joint spectral measure (a complex measure in general)

$$dF^{Y,X}(\omega) = \hat{b}(\omega) \, dF^X(\omega), \tag{37}$$

where $\hat{b}(\omega) := \sum_{j=-\infty}^{\infty} b_j e^{-ij\omega}$.

It is important that the above formulas are valid when the sequence of weights $(b_j : j \in \mathbb{Z})$ is such that $\hat{b} \in L^2([-\pi, \pi], \mathcal{B}, dF)$. A sufficient condition is that $\sum_{k=-\infty}^{\infty} |b_k| < \infty$. (When $b_j = 0$ for $|j| > N$, the square and absolute summability of $b_j$s automatically holds.) In another wording, the stationary time series $\{Y_t\}_{t\in\mathbb{Z}}$ is *subordinated* to the process $\{X_t\}_{t\in\mathbb{Z}}$.

The second order, weakly stationary time series $\{\xi_t\}_{t\in\mathbb{Z}}$ is called *white noise sequence* if its autocovariances are $c(0) = \sigma^2$ and $c(h) = 0$ for $h = \pm 1, \pm 2, \dots$. In other words, $\xi_t$s are uncorrelated and have variance $\sigma^2$. For example, if $\xi_t$s are i.i.d., with finite variance, they constitute a white noise sequence, and in the Gaussian case, the two notions are the same. When $\sigma^2 = 1$, we call the white noise sequence *orthonormal sequence.*

The *qth order moving average process*, MA($q$) process, is defined by

$$X_t = \sum_{k=0}^{q} \beta_k \xi_{t-k}, \quad c(h) = \sum_{k=h}^{q} \beta_k \bar{\beta}_{k-h} \quad (0 \le h \le q), \quad c(-h) = \bar{c}(h),$$

$c(h) = 0$ if $|h| > q$, where $\{\xi_n\}$ is an orthonormal sequence. The spectral density of an MA($q$) process is

$$f^X(\omega) = \frac{1}{2\pi} |\hat{\beta}(\omega)|^2, \qquad \hat{\beta}(\omega) = f^{X|\xi}(\omega) = \sum_{k=0}^{q} b_k e^{-ik\omega}.$$

Likewise, the spectral density of the MA($\infty$) process is

$$f^X(\omega) = \frac{1}{2\pi} \left| \sum_{k=-\infty}^{\infty} b_k e^{-ik\omega} \right|^2, \text{ where } \sum_{k=-\infty}^{\infty} |b_k|^2 < \infty.$$

The AR($p$), i.e. the *pth order autoregressive process* $\{X_t\}_{t\in\mathbb{Z}}$ is defined by

$$X_t = \sum_{j=1}^{p} \alpha_j X_{t-j} + \beta\xi_t = \sum_{j=1}^{p} a_j X_{t-j} + \beta\xi_t,$$

where $\{\xi_t\} \sim \text{WN}(1)$ is an orthonormal process. This form is well comparable to the prediction theory of Chapter 5.

Under the stability conditions, the above equation has a causal MA solution of the form

$$X_t = \sum_{k=0}^{\infty} b_k \xi_{t-k} \quad (b_k \in \mathbb{C})$$

satisfying the following relations:

$$\begin{aligned} b_0 &= \beta, \\ b_1 + \alpha_1 b_0 &= 0, \\ b_2 + \alpha_1 b_1 + \alpha_2 b_0 &= 0, \\ &\vdots \\ b_p + \alpha_1 b_{p-1} + \cdots + \alpha_p b_0 &= 0, \\ b_{p+k} + \alpha_1 b_{p+k-1} + \cdots + \alpha_p b_k &= 0 \quad (k \geq 1). \end{aligned}$$

The question is whether the resulting sequence $\{b_0, b_1, b_2, \dots\}$ will be square summable. It holds if no roots of the *AR polynomial* $\alpha(z) = 1 - \alpha_1 z - \cdots - \alpha_p z^p$ are on the open unit disc. Knowing the covariances, we get the Yule-Walker equations for $\alpha_j$s that coincide with the one-step ahead prediction of Chapter 5, based on the $p$-length long past. If $\boldsymbol{C}_p$ of Chapter 1 is positive definite, this condition holds.

The ARMA($p, q$) processes ($p \geq 0$, $q \geq 0$) are generalizations of both AR($p$) and MA($q$) processes. The model equation is:

$$X_t = \sum_{j=1}^{p} \alpha_j X_{t-j} + \sum_{k=0}^{q} \beta_k \xi_{t-k}, \qquad \beta_0 \neq 0.$$

With the *AR polynomial* $\alpha(z) = 1 - \alpha_1 z - \cdots - \alpha_p z^p$ and the *MA polynomial* $\beta(z) = \sum_{\ell=0}^{q} \beta_\ell z^\ell$, the model equation of the ARMA process can be written in the simple form

$$\alpha(L)X_t = \beta(L)\xi_t, \quad t \in \mathbb{Z},$$

where $L$ is the left (backward) shift, or in other words, the *lag operator*.

We want to find a causal stationary solution $\{X_t\}$ of this equation, that is a MA($\infty$) process. Again, no roots of the AR polynomial can be on the closed unit disc, called *stability*.

We can state the following. Let $\{X_t\}$ be an ARMA($p, q$) process for which the polynomials $\alpha(z)$ and $\beta(z)$ have no common zeros. Then $\{X_t\}$ is causal if and only if $\alpha(z) \neq 0$ for all $|z| \leq 1$ (stability). The coefficients of the MA($\infty$) process $X_t = \sum_{j=0}^{\infty} b_j \xi_{t-j}$ are obtainable by the power series expansion

$$b(z) = \sum_{j=0}^{\infty} b_j z^j = \alpha^{-1}(z)\beta(z), \quad |z| \leq 1.$$

The coefficients $b_j$s of the so-called *transfer function* $b(z)$ are called *impulse responses.*

The spectral density of the ARMA($p, q$) process is

$$f^X(\omega) = \frac{1}{2\pi} \frac{|\hat{\beta}(\omega)|^2}{|\hat{\alpha}(\omega)|^2}, \quad \hat{\beta}(\omega) = \sum_{\ell=0}^{q} \beta_\ell e^{-i\ell\omega}, \quad \hat{\alpha}(\omega) = \sum_{j=0}^{p} \alpha_j e^{-ij\omega}.$$

This is a rational spectral density.

The Wold decomposition in 1D guarantees that any non-singular weakly stationary time series (not completely determined by the remote past) can be decomposed into a regular (MA($\infty$), purely non-deterministic, future independent) and a singular (purely deterministic, completely determined by a remote past) process that are orthogonal to each other. The original proof of Wold is based on the one-step ahead predictions with longer and longer past, where the prediction errors converge to an optimum value (that is not zero in the presence of a regular part), see the subsequent lesson about predictions.

The spectral form of the Wold decomposition is

$$f(\omega) = \frac{1}{2\pi} \left| \sum_{j=0}^{\infty} b_j e^{-ij\omega} \right|^2 = \frac{1}{2\pi} |B(e^{-i\omega})|^2,$$

$$B(z) = \sum_{j=0}^{\infty} b_j z^j \quad (|z| \leq 1), \quad \sum_{j=0}^{\infty} |b_j|^2 < \infty.$$

By a theorem of A.N. Kolmogorov, a stationary time series is regular if and only if the following three conditions hold on $[-\pi, \pi]$:

1. its spectral measure $dF$ is absolutely continuous w.r.t. Lebesgue measure;

2. its spectral density $f$ is positive almost everywhere;

3. *(Kolmogorov's condition)* $\log f$ is integrable.

Note that we have different classes of singular time series. The spectral measure of a stationary time series has a unique Lebesgue decomposition

$$dF = dF_a + dF_s, \quad dF_a \ll d\omega, \quad dF_s \perp d\omega,$$

where $d\omega$ denotes Lebesgue measure, $dF_a(\omega) = f_a(\omega)d\omega$ is the absolutely continuous part of $dF$ with density $f_a$ and $dF_s$ is the singular part of $dF$, which is concentrated on a zero Lebesgue measure subset of $[-\pi, \pi]$. We call it Type (0) singularity. Apart from this, we distinguish between the following three cases:

(1) $f_a(\omega) = 0$ on a set of positive Lebesgue measure on $[-\pi, \pi]$;

(2) $f_a(\omega) > 0$ a.e. , but $\int_{-\pi}^{\pi} \log f_a(\omega) d\omega = -\infty$;

(3) $f_a(\omega) > 0$ a.e. and $\int_{-\pi}^{\pi} \log f_a(\omega) d\omega > -\infty$.

In cases (1) and (2), the time series is singular, whereas in case (3), the time series is non-singular and has the Wold decomposition. So a regular (Type (3)) process can coexist only with a Type (0) singular one, while adding an orthogonal regular process to a Type (1) or (2) singularity gives a regular process.