Lessons 7-8: VARMA, regular and singular processes in multi-dimension (multi-dimensional Wold decomposition)

Marianna Bolla and Tamás Szabados, BME Math. Inst.

October 28, 2020

In this lesson we investigate properties of multidimensional, weakly stationary time series, similarly to the 1D case. In contrast to the 1D situation, here we proceed deductively: from the most general constant rank processes, via rational (causal) ones, to the state space models.

Not surprisingly, here the classification is more complicated; for example, the spectral density matrix can be of reduced rank, but not zero. Linear filtering and constant rank processes are considered, and we prove that the spectral density matrix of a d-dimensional process has a constant rank $r \leq d$ if and only if the process is a two-sided moving average, obtained as a TLF with an r-dimensional white noise process; in this way, the spectral density matrix is also factorized. Special cases are the regular processes, when the TLF is a one-sided MA process; i.e., a regular process has a causal MA representation.

A non-singular process has a Wold decomposition in multi-dimension too. Here only the so-called innovation subspaces are unique, whereas their dimension is equal to the constant rank of the process. Further subclass of regular processes are the ones with a rational spectral density matrix. These can be finitely parametrized, and have either a state space or stable VARMA (Vector AutoRegressive Moving Average) representation, or an MFD (Matrix Fractional Description). Here, in the factorization of the spectral density matrix, the so-called spectral factors are also rational matrices. This fact has important consequences, for example, in the dynamic factor analysis and relations in the time domain.

1 Linear transformations, subordinated and causally subordinated processes

Here the extension of notions of *linear filter* and *subordinated process* from 1D to MD (multi-dimension) follows.

Let $\{\mathbf{X}_t\}_{t\in\mathbb{Z}}$ be an d-dimensional stationary time series with spectral representation

$$\mathbf{X}_t = \int_{-\pi}^{\pi} e^{it\omega} d\mathbf{Z}_\omega$$

and spectral measure matrix $d\mathbf{F} = d\mathbf{F}^X$. Assume that we are given a matrix function

$$\boldsymbol{T}(\omega) = [t_{jk}(\omega)]_{m\times d}, \quad t_{jk} \in L^2([-\pi,\pi], \mathcal{B}, \mathrm{tr}(d\mathbf{F})). \tag{1}$$

By definition, the m-dimensional process $\{\mathbf{Y}_t\}_{t\in\mathbb{Z}}$ is a *linear transform* of or obtained by a *time invariant linear filter (TLF)* from $\{\mathbf{X}_t\}$ if

$$\mathbf{Y}_t = \int_{-\pi}^{\pi} e^{it\omega} \boldsymbol{T}(\omega)\, d\mathbf{Z}_\omega \quad (t\in\mathbb{Z}). \tag{2}$$

It means that with the random measure

$$d\mathbf{Z}_\omega^Y := \boldsymbol{T}(\omega)\, d\mathbf{Z}_\omega$$

we have a representation of the process $\{\mathbf{Y}_t\}$:

$$\mathbf{Y}_t = \int_{-\pi}^{\pi} e^{it\omega} d\mathbf{Z}_\omega^Y \quad (t\in\mathbb{Z}).$$

Likrwise,

$$\begin{aligned}\mathrm{Cov}(\mathbf{Y}_{t+h}, \mathbf{Y}_t) &= \left[\int_{-\pi}^{\pi} e^{ih\omega} \sum_{r,s=1}^{n} t_{pr}(\omega)\overline{t_{qs}(\omega)} dF^{rs}(\omega)\right]_{p,q=1}^{m} \\ &= \int_{-\pi}^{\pi} e^{ih\omega} \boldsymbol{T}(\omega) d\mathbf{F}(\omega) \boldsymbol{T}^*(\omega) \quad (h\in\mathbb{Z}).\end{aligned} \tag{3}$$

Thus $\{\mathbf{Y}_t\}$ is also a stationary time series with spectral measure matrix

$$d\mathbf{F}^Y = \boldsymbol{T}\, d\mathbf{F}^X\, \boldsymbol{T}^*. \tag{4}$$

Considering $\mathrm{Cov}(\mathbf{Y}_{t+h}\mathbf{X}_t) = \mathbb{E}(\mathbf{Y}_{t+h}\mathbf{X}_t^*)$, one can similarly obtain that $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ are jointly stationary and their joint spectral density matrix is

$$d\mathbf{F}^{Y,X} = \boldsymbol{T}\, d\mathbf{F}^X. \tag{5}$$

It is not difficult to show that the last formula is not only necessary, but sufficient for obtaining $\{\mathbf{Y}_t\}$ from $\{\mathbf{X}_t\}$ by linear transformation. Because of (5) we may call $\boldsymbol{T}$ the *conditional spectral density* of $\{\mathbf{Y}_t\}$ w.r.t. $\{\mathbf{X}_t\}$ and denote $\boldsymbol{T} = \boldsymbol{f}^{Y|X}$.

By means of Fourier series, we can rewrite the definition of linear transform in the time domain, assuming that condition (1) holds:

$$\boldsymbol{T}(\omega) = \sum_{j=-\infty}^{\infty} \boldsymbol{\tau}(j)e^{-ij\omega}, \quad \boldsymbol{\tau}(j) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{ij\omega}\boldsymbol{T}(\omega)d\omega, \quad \sum_{j=-\infty}^{\infty} \|\boldsymbol{\tau}(j)\|_F^2 < \infty. \tag{6}$$

Because of the isometry between $H_{\mathbf{X}}$ and $L^2([-\pi,\pi],\mathcal{B},\mathrm{tr}(d\mathbf{F}))$, definition (2) of linear transform is equivalent to

$$\begin{aligned}\mathbf{Y}_t &= \int_{-\pi}^{\pi} e^{it\omega} \sum_{j=-\infty}^{\infty} \boldsymbol{\tau}(j)e^{-ij\omega}\, d\mathbf{Z}_\omega = \sum_{j=-\infty}^{\infty} \boldsymbol{\tau}(j) \int_{-\pi}^{\pi} e^{i(t-j)\omega}\, d\mathbf{Z}_\omega \\ &= \sum_{j=-\infty}^{\infty} \boldsymbol{\tau}(j)\mathbf{X}_{t-j}, \qquad t \in \mathbb{Z}.\end{aligned} \tag{7}$$

This shows that the linear transform $\{\mathbf{Y}_t\}$ is, in fact, obtained by linear filtering from $\{\mathbf{X}_t\}$, that is, by a sliding summation with given $m \times d$ matrix weights $\boldsymbol{\tau}(j)$.

We call an m-dimensional stationary time series $\{\mathbf{Y}_t\}_{t\in\mathbb{Z}}$ *causally subordinated* to an d-dimensional time series $\{\mathbf{X}_t\}_{t\in\mathbb{Z}}$ if $\{\mathbf{Y}_t\}$ is obtained from $\{\mathbf{X}_t\}$ by a linear transform (2) or (7), and, also,

$$H_k^-(\mathbf{Y}) \subset H_k^-(\mathbf{X}) \text{ for all } k \in \mathbb{Z}. \tag{8}$$

By stationarity, it suffices if this last relationship holds for a specific value of k, for example, for $k = 0$.

In the causally subordinated case (8) if $\boldsymbol{T}(\omega)$ has entries $t_{jk} \in L^2([-\pi,\pi],\mathcal{B},d\omega)$, the filtering equation (7) modifies as a causal one-sided moving average process:

$$\mathbf{Y}_t = \sum_{j=0}^{\infty} \boldsymbol{\tau}(j)\mathbf{X}_{t-j}, \quad \boldsymbol{\tau}(j) = 0 \quad \text{if} \quad j < 0. \tag{9}$$

In other words, the Fourier series of $\boldsymbol{T}$ becomes one-sided:

$$\boldsymbol{T}(\omega) = \sum_{j=0}^{\infty} \boldsymbol{\tau}(j) e^{-ij\omega}, \quad \sum_{j=0}^{\infty} \|\boldsymbol{\tau}(j)\|_F^2 < \infty. \tag{10}$$

2 Stationary time series of constant rank

Assume that $\mathbf{X}_t = (X_t^1, \ldots, X_t^d)$ $(t \in \mathbb{Z})$ is a d-dimensional complex valued weakly stationary time series with absolutely continuous spectral measure with density matrix $\boldsymbol{f}$ on $[-\pi, \pi]$. We say that $\{\mathbf{X}_t\}$ has *constant rank* r if the matrix $\boldsymbol{f}(\omega)$ has rank r for almost all $\omega \in [-\pi, \pi]$.

Theorem 1. *We have the following back-and-forth statements.*

(a) Assume that the stationary time series $\mathbf{X}_t = (X_t^1, \ldots, X_t^d)$ $(t \in \mathbb{Z})$ *has an absolutely continuous spectral measure with density matrix* $\boldsymbol{f}$ *of constant rank* r*. Then* $\boldsymbol{f}$ *can be factored as*

$$\boldsymbol{f}(\omega) = \frac{1}{2\pi} \boldsymbol{\phi}(\omega) \boldsymbol{\phi}^*(\omega) \quad \text{for a.e. } \omega \in [-\pi, \pi],$$

where $\boldsymbol{\phi}(\omega) \in \mathbb{C}^{d \times r}$*,* $r \leq d$*, called a* spectral factor*. Also,* $\mathbf{X}$ *can be represented as a two-sided infinite MA process (a sliding summation)*

$$\mathbf{X}_t = \sum_{j=-\infty}^{\infty} \boldsymbol{b}(j) \boldsymbol{\xi}_{t-j}, \tag{11}$$

where $\{\boldsymbol{\xi}_t\}$ *is a* $\mathrm{WN}(\boldsymbol{I}_r)$ *(orthonormal) sequence and* $\boldsymbol{b}(j) = [b_{k\ell}(j)] \in \mathbb{C}^{d \times r}$ $(j \in \mathbb{Z})$ *is a non-random matrix-valued sequence, the Fourier coefficients of the spectral factor* $\boldsymbol{\phi}(\omega)$*:*

$$\boldsymbol{\phi}(\omega) = \sum_{j=-\infty}^{\infty} \boldsymbol{b}(j) e^{-ij\omega}, \quad \boldsymbol{b}(j) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \boldsymbol{\phi}(\omega) d\omega, \quad \sum_{j=-\infty}^{\infty} \|\boldsymbol{b}(j)\|_F^2 < \infty. \tag{12}$$

The Fourier series converges to $\boldsymbol{\phi}$ *in mean square.*

(b) Conversely, any stationary time series $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ *represented as a two-sided infinite MA process* (11) *has an absolutely continuous spectral measure with density matrix* $\boldsymbol{f}$ *of constant rank* r*, where* r *is the dimension of the white noise process* $\{\boldsymbol{\xi}_t\}$*.*

We do not prove the theorem, just remark that the proof of statement (a) depends only on the fact that one has a factorization of the spectral density of the form

$$\boldsymbol{f}(\omega) = \frac{1}{2\pi}\boldsymbol{\phi}(\omega)\boldsymbol{\phi}^*(\omega), \quad \boldsymbol{\phi}(\omega) \in \mathbb{C}^{d\times r}, \quad \text{for a.e.} \quad \omega \in [-\pi, \pi].$$

3 Multidimensional Wold decomposition

The multidimensional version of Wold decomposition goes similarly as its 1D case The notations, definitions of remote past and *singular, regular, non-singular processes* can be extended to the MD case with no essential change, therefore will not be repeated here.

Theorem 2. *Assume that* $\{\mathbf{X}_t\}_{t\in\mathbb{Z}}$ *is an d-dimensional non-singular stationary time series. Then it can be represented in the form*

$$\mathbf{X}_t = \mathbf{R}_t + \mathbf{Y}_t = \sum_{j=0}^{\infty} \boldsymbol{b}(j)\boldsymbol{\xi}_{t-j} + \mathbf{Y}_t \quad (t \in \mathbb{Z}), \tag{13}$$

where

1. $\{\mathbf{R}_t\}$ *is a d-dimensional regular time series causally subordinated to* $\{\mathbf{X}_t\}$*;*

2. $\{\mathbf{Y}_t\}$ *is an d-dimensional singular time series causally subordinated to* $\{\mathbf{X}_t\}$*;*

3. $\{\boldsymbol{\xi}_t\}$ *is an r-dimensional* $(r \leq d)$ $\mathrm{WN}(\boldsymbol{I}_r)$ *(orthonormal) sequence causally subordinated to* $\{\mathbf{X}_t\}$*;*

4. $\{\mathbf{R}_t\}$ *and* $\{\mathbf{Y}_t\}$ *are orthogonal to each other:* $\mathbb{E}(\mathbf{R}_t\mathbf{Y}_s^*) = \boldsymbol{O}$ *for* $t, s \in \mathbb{Z}$*;*

5. $\boldsymbol{b}(j) = [b_{k\ell}(j)] \in \mathbb{C}^{d\times r}$ *for* $j \geq 0$ *and* $\sum_{j=0}^{\infty} \|\boldsymbol{b}(j)\|_F^2 < \infty$.

As we are going to see in the next section, the orthonormal series $\{\boldsymbol{\xi}_t\}$ can be chosen the same as the orthonormal series in Theorem 1. Also, by the definition $\{\boldsymbol{\xi}_t\}$ in the present theorem, $H_k^-(\boldsymbol{\xi}) \subset H_k^-(\mathbf{X})$, so $\{\boldsymbol{\xi}_t\}$ is also subordinated to $\{\mathbf{X}_t\}$. This comes out from the proof of the theorem. It also follows from the proof that $\boldsymbol{\xi}_0$ is unique up to pre-multiplication by an arbitrary $r \times r$ unitary matrix $\boldsymbol{U}$. We do not write the complete proof here, but discuss some considerations related to predictions (to be introduced in the next lesson), and what were the base of the original proof of H. Wold in the 1D case.

For simplicity, let us fix the present time as time 0. Then the *best linear t-step ahead prediction* $\hat{\mathbf{X}}_t$ of $\mathbf{X}_t$ is by definition the projection of $\mathbf{X}_t$ to the past until 0 that is, to H_0^-:

$$\hat{\mathbf{X}}_t = \sum_{j=t}^{\infty} \boldsymbol{b}(j)\boldsymbol{\xi}_{t-j} + \mathbf{Y}_t \quad (t \in \mathbb{Z}), \tag{14}$$

since the right hand side of (14) is in H_0^-, whereas the error term of the t-step ahead prediction,

$$\mathbf{X}_t - \hat{\mathbf{X}}_t = \sum_{j=0}^{t-1} \boldsymbol{b}(j)\boldsymbol{\xi}_{t-j}$$

is orthogonal to H_0^- (recall the projection theorem in Hilbert spaces). Hence, the mean square error of the prediction is given by

$$\sigma_t^2 := \|\mathbf{X}_t - \hat{\mathbf{X}}_t\|^2 = \sum_{j=0}^{t-1} \|\boldsymbol{b}(j)\|_F^2. \tag{15}$$

Let H_k^- denote again the Hilbert space spanned by the past of $\{\mathbf{X}_t\}$ until time k and let $\mathrm{Proj}_{H_{t-1}^-}\mathbf{X}_t$ denote the orthogonal projection of $\mathbf{X}_t$ to H_{t-1}^- (1-step ahead prediction), which exists uniquely by the projection theorem. Define the *innovation process*

$$\boldsymbol{\eta}_t := \mathbf{X}_t - \mathrm{Proj}_{H_{t-1}^-}\mathbf{X}_t, \quad t \in \mathbb{Z}. \tag{16}$$

If the covariance matrix of $\{\boldsymbol{\eta}_t\}$ is

$$\boldsymbol{\Sigma} := \mathbb{E}(\boldsymbol{\eta}_0\boldsymbol{\eta}_0^*), \tag{17}$$

then $\{\boldsymbol{\eta}_t\}$ is a $\mathrm{WN}(\boldsymbol{\Sigma})$ process,

$$\mathbb{E}(\boldsymbol{\eta}_t\boldsymbol{\eta}_s^*) = \delta_{ts}\boldsymbol{\Sigma} \quad \forall t, s \in \mathbb{Z}.$$

The one-step ahead prediction of $\mathbf{X}_t$ based on H_{t-1}^- is exactly $\mathrm{Proj}_{H_{t-1}^-}\mathbf{X}_t$, the prediction error is $\boldsymbol{\eta}_t$, and the covariance matrix of this error is $\boldsymbol{\Sigma}$.

It is important that the rank $r \le d$ of $\mathbf{\Sigma}$ is *the dimension of the innovation subspace* (Hilbert sace approach), and it is the *rank of the process* $\{\mathbf{X}_t\}$. If $\{\mathbf{X}_t\}$ is non-singular, then clearly we must have $r \ge 1$. In the special case when $r = d$, the process is called *full rank*. If the process has full rank, then $\mathbf{\Sigma}$ is invertible, and we can transform the innovation process into an orthonormal process with the transformation

$$\boldsymbol{\xi}_t := \mathbf{\Sigma}^{-\frac{1}{2}} \boldsymbol{\eta}_t, \quad t \in \mathbb{Z}.$$

Obviously, $\boldsymbol{\xi}_t$ is an orthonormal $\mathrm{WN}(\boldsymbol{I}_d)$ process.

In generic, not necessarily full rank case we proceed as follows. If the rank of $\mathbf{\Sigma}$ is $r \le d$, then it has the Gram decomposition $\mathbf{\Sigma} = \boldsymbol{A}\boldsymbol{A}^T$, where $\boldsymbol{A}$ is $d \times r$ matrix, and it is unique up to post-multiplication by an arbitrary $r \times r$ unitary matrix $\boldsymbol{U}$. Then $\boldsymbol{\eta}_t = \boldsymbol{A}\boldsymbol{\xi}_t$, where $\boldsymbol{\xi}_t \sim \mathrm{WN}(\boldsymbol{I}_r)$. The post-multiplication of $\boldsymbol{A}$ with $\boldsymbol{U}$ causes pre-multiplication of $\boldsymbol{\xi}_t$ with $\boldsymbol{U}^*$, but $\boldsymbol{U}^*\boldsymbol{\xi} \sim \mathrm{WN}(\boldsymbol{I}_r)$ too.

The Wold decomposition with the innovation process has the following form (see also the forthcoming lesson about predictons):

$$\mathbf{X}_t = \mathbf{R}_t + \mathbf{Y}_t = \sum_{j=0}^{\infty} \boldsymbol{a}(j)\boldsymbol{\eta}_{t-j} + \mathbf{Y}_t \quad (t \in \mathbb{Z}), \tag{18}$$

where $\boldsymbol{a}(j)$s are $d \times d$ matrices, $\{\mathbf{Y}_t\}$ is a singular process, $\mathbf{Y}_t \in H_{-\infty}$, $\mathbf{R}_s \perp \mathbf{Y}_t$ for any $s, t \in \mathbb{Z}$. On the other hand, $\{\mathbf{R}_t\}$ is a regular process of rank r that can equivalently be written as

$$\mathbf{R}_t = \sum_{j=0}^{\infty} \boldsymbol{a}(j)\boldsymbol{A}\boldsymbol{\xi}_{t-j} = \sum_{j=0}^{\infty} \boldsymbol{b}(j)\boldsymbol{\xi}_{t-j},$$

where $\boldsymbol{b}(j) = \boldsymbol{a}(j)\boldsymbol{A}$ is $d \times r$ matrix and unique up to post-multiplication by an arbitrary $r \times r$ unitary matrix $\boldsymbol{U}$, $j = 0, 1, \dots$. This is the required Wold decomposition.

4 Regular and singular time series

Assume that $\{\mathbf{X}_t\}_{t\in\mathbb{Z}}$ is a regular stationary time series. By Wold's decomposition (13),

$$\mathbf{X}_t = \sum_{j=0}^{\infty} \boldsymbol{b}(j)\boldsymbol{\xi}_{t-j} \quad (t \in \mathbb{Z}), \quad \boldsymbol{b}(j) = [b_{k\ell}(j)]_{d\times r}, \tag{19}$$
$$\sum_{j=0}^{\infty} \|\boldsymbol{b}(j)\|_F^2 < \infty, \qquad \{\boldsymbol{\xi}_t\}_{t\in\mathbb{Z}} \sim \mathrm{WN}(I_r).$$

By Theorem 1, this representation of $\{\mathbf{X}_t\}$ implies that $\{\mathbf{X}_t\}$ has an absolutely continuous spectral measure with density matrix $\boldsymbol{f}(\omega)$ and with constant rank $\tilde{r}$ for a.e. $\omega \in [-\pi, \pi]$,

$$\boldsymbol{f}(\omega) = \frac{1}{2\pi}\boldsymbol{\phi}(\omega)\boldsymbol{\phi}^*(\omega), \quad \boldsymbol{\phi}(\omega) = [\phi_{k\ell}(\omega)]_{d\times\tilde{r}}, \quad \text{for a.e. } \omega \in [-\pi, \pi] \tag{20}$$

and

$$\mathbf{X}_t = \sum_{j=-\infty}^{\infty} \tilde{\boldsymbol{b}}(j)\tilde{\boldsymbol{\xi}}_{t-j} \quad (t \in \mathbb{Z}), \quad \tilde{\boldsymbol{b}}(j) = [\tilde{b}_{k\ell}(j)]_{d\times\tilde{r}}, \tag{21}$$
$$\sum_{j=-\infty}^{\infty} \|\tilde{\boldsymbol{b}}(j)\|_F^2 < \infty, \qquad \{\tilde{\boldsymbol{\xi}}_t\}_{t\in\mathbb{Z}} \sim \mathrm{WN}(I_r).,$$

Compare now (19) and (21). Since the orthonormal process $\{\boldsymbol{\xi}_t\}$ in (19) is unique up to pre-multiplication by an arbitrary $r \times r$ unitary matrix $\boldsymbol{U}$, it follows that

1. $\tilde{r} = r$,
2. $\tilde{\boldsymbol{b}}(j) = 0$ if $j < 0$,
3. $\tilde{\boldsymbol{\xi}}_0 = \boldsymbol{U}\boldsymbol{\xi}_0$, $\tilde{\boldsymbol{\xi}}_k = S^k\boldsymbol{U}\boldsymbol{\xi}_0$ $(k \in \mathbb{Z})$.

Corollary 1. *The dimension r of the WN($\boldsymbol{I}_r$) (orthonormal) innovation process $\{\boldsymbol{\xi}_t\}$ in* (19) *and of the WN($\boldsymbol{\Sigma}$) innovation process $\{\boldsymbol{\eta}_t\}$ in* (16) *are equal to the a.e. constant rank of the spectral density matrix $\boldsymbol{f}$ of the regular time series $\{\mathbf{X}_t\}$.*

From now on we assume that $\boldsymbol{\xi}_0 = \tilde{\boldsymbol{\xi}}_0$ is chosen as in (19), so we may omit all 'tildes'.

Corollary 2. *A stationary time series* $\{\mathbf{X}_t\}_{t\in\mathbb{Z}}$ *is regular if and only it has an absolutely continuous spectral measure with spectral density that can be factored in the form*

$$\boldsymbol{f}(\omega) = \frac{1}{2\pi}\boldsymbol{\phi}(\omega)\boldsymbol{\phi}^*(\omega), \quad \boldsymbol{\phi}(\omega) = [\phi_{k\ell}(\omega)]_{d\times r}, \quad \textit{for a.e. } \omega \in [-\pi, \pi],$$

where

$$\boldsymbol{\phi}(\omega) = \sum_{j=0}^{\infty} \boldsymbol{b}(j)e^{-ij\omega}, \quad \|\boldsymbol{\phi}\|_2^2 = \sum_{j=0}^{\infty} \|\boldsymbol{b}(j)\|_F^2 < \infty.$$

Then

$$\boldsymbol{\phi}(\omega) = \boldsymbol{\Phi}(e^{-i\omega}), \quad \boldsymbol{\Phi}(z) = \sum_{j=0}^{\infty} \boldsymbol{b}(j)z^j, \quad z \in D, \tag{22}$$

where D is the open unit disc. Thus the entries of the spectral factor $\boldsymbol{\Phi}(z) = [\Phi_{jk}(z)]_{d\times r}$ *are analytic functions in the open unit disc D and belong to the class $L^2(T)$, consequently, they belong to H^2 (Hardy space). Briefly, we can write that $\boldsymbol{\Phi} \in H^2$.*

Take the spectral factor $\boldsymbol{\Phi}(z) = [\Phi_{k\ell}(z)]_{d\times r} \in H^2$ defined in (22). The spectral factor $\boldsymbol{\Phi}(z)$ contains all information needed for finding the orthonormal innovation process $\{\boldsymbol{\xi}_t\}$ through its boundary values on the unit circle. Also, the coefficients $\boldsymbol{b}(j)$ can be obtained from $\boldsymbol{\Phi}$ by power series expansion. As soon as we have these information, we may get the optimal linear prediction and its mean square error by formulas (14) and (15).

4.1 Full rank processes

Assume that $\{\mathbf{X}_t\}$ is a d-dimensional stationary time series, with spectral measure matrix

$$d\boldsymbol{F} = d\boldsymbol{F}_a + d\boldsymbol{F}_s, \quad d\boldsymbol{F}_a \ll d\omega, \quad d\boldsymbol{F}_s \perp d\omega, \tag{23}$$

where $d\omega$ denotes Lebesgue measure, $d\boldsymbol{F}_a(\omega) = \boldsymbol{f}(\omega)d\omega$, $\boldsymbol{f}$ is the spectral density matrix of $\{\mathbf{X}_t\}$, $d\boldsymbol{F}_s$ is supported on a zero Lebesgue measure subset of $[-\pi, \pi]$. We say that $\{\mathbf{X}_t\}$ has *full rank* if $\text{rank}(\boldsymbol{f}(\omega)) = d$ for a.e. $\omega \in$

$[-\pi, \pi]$. It means that $\boldsymbol{f}(\omega)$ is a non-singular matrix a.e. Note that a 1D spectral density is of full rank 1 if $f > 0$ a.s. on $[0, 2\pi]$.

The next theorem is an extension of the Kolmogorov–Szegő formula to the MD full rank case.

Theorem 3. *A d-dimensional stationary time series $\{\mathbf{X}_t\}$ is of full rank non-singular process if and only if* $\log\det \boldsymbol{f} \in L^1$*, that is,*

$$\int_{-\pi}^{\pi} \log\det \boldsymbol{f}(\omega) d\omega > -\infty.$$

In this case, if $\boldsymbol{\Sigma}$ denotes the covariance matrix of the innovation process $\{\boldsymbol{\eta}_t\}$, that is, of the one-step ahead prediction error process defined in (16) *and* (17)*, then*

$$\det \boldsymbol{\Sigma} = (2\pi)^d \exp \int_{-\pi}^{\pi} \log\det \boldsymbol{f}(\omega) \frac{d\omega}{2\pi}. \tag{24}$$

Corollary 3. *If $\{\mathbf{X}_t\}$ is of full rank non-singular time series, then the singular process $\{\mathbf{Y}_t\}$ in its Wold decomposition $\mathbf{X}_t = \mathbf{R}_t + \mathbf{Y}_t$ has singular spectral measure. More exactly, using the notations introduced above,*

$$d\boldsymbol{F}^R = d\boldsymbol{F}_a, \quad d\boldsymbol{F}^Y = d\boldsymbol{F}_s.$$

Corollary 4. *A stationary time series $\{\mathbf{X}_t\}$ is regular and of full rank if and only if*

1. *it has an absolutely continuous spectral measure matrix $d\boldsymbol{F}$ with density matrix $\boldsymbol{f}$;*
2. $\log\det \boldsymbol{f} \in L^1$.

Then the Kolmogorov–Szegő formula (24) *also holds.*

4.2 Generic regular processes

The next theorem is an extension of Corollary 4 to the general, not necessarily full rank, case. Let $\{\mathbf{X}_t\}$ be a d-dimensional stationary time series. Assume that its spectral measure matrix $d\boldsymbol{F}$ is absolutely continuous with density matrix $\boldsymbol{f}(\omega)$ which has rank r, $1 \le r \le d$, for a.e. $\omega \in [-\pi, \pi]$. Take the parsimonious spectral decomposition of the self-adjoint, non-negative definite matrix $\boldsymbol{f}(\omega)$:

$$\boldsymbol{f}(\omega) = \sum_{j=1}^{r} \lambda_j(\omega)\mathbf{u}_j(\omega)\mathbf{u}_j^*(\omega) = \tilde{\boldsymbol{U}}(\omega)\boldsymbol{\Lambda}_r(\omega)\tilde{\boldsymbol{U}}^*(\omega), \tag{25}$$

where

$$\boldsymbol{\Lambda}_r(\omega) = \mathrm{diag}[\lambda_1(\omega), \dots, \lambda_r(\omega)], \quad \lambda_1(\omega) \geq \cdots \geq \lambda_r(\omega) > 0, \tag{26}$$

and $\tilde{\boldsymbol{U}}(\omega) \in \mathbb{C}^{d\times r}$ is a sub-unitary matrix, containing the corresponding orthonormal eigenvectors columnwise.

Then, still, we have

$$\boldsymbol{\Lambda}_r(\omega) = \tilde{\boldsymbol{U}}^*(\omega)\boldsymbol{f}(\omega)\tilde{\boldsymbol{U}}(\omega). \tag{27}$$

Take care that here we use the word ‘spectral’ in two different meanings. On one hand, we use the spectral density of a time series in terms of a Fourier spectrum, on the other hand we take the spectral decomposition of a matrix in terms of eigenvalues and eigenvectors.

Theorem 4. *A d-dimensional stationary time series $\{\mathbf{X}_t\}$ is regular and of rank r, $1 \leq r \leq d$, if and only if each of the following conditions hold:*

1. *It has an absolutely continuous spectral measure matrix $d\boldsymbol{F}$ with density matrix $\boldsymbol{f}(\omega)$ which has rank r for a.e. $\omega \in [-\pi, \pi]$.*
2. *For $\boldsymbol{\Lambda}_r(\omega)$ defined by (26) one has $\log\det\boldsymbol{\Lambda}_r \in L^1 = L^1([-\pi,\pi], \mathcal{B}, d\omega)$, which is equivalent to*

$$\int_{-\pi}^{\pi} \sum_{j=1}^{r} \log\lambda_j(\omega)\, d\omega > -\infty.$$

3. *The sub-unitary matrix function $\tilde{\boldsymbol{U}}(\omega)$ appearing in the spectral decomposition of $\boldsymbol{f}(\omega)$ in (25) belongs to the Hardy space $H^\infty \subset H^2$, so*

$$\tilde{\boldsymbol{U}}(\omega) = \sum_{j=0}^{\infty} \boldsymbol{\psi}(j)e^{-ij\omega}, \quad \boldsymbol{\psi}(j) \in \mathbb{C}^{d\times r}, \quad \sum_{j=0}^{\infty} \|\boldsymbol{\psi}(j)\|_F^2 < \infty.$$

The quintessence of the proof is that $\mathbf{\Lambda}_r(\omega)$ can be considered as the spectral density function of an r-dimensional stationary time series $\{\mathbf{V}_t\}_{t\in\mathbf{Z}}$ of full rank r; the auto-covariance function of $\{\mathbf{V}_t\}$ is

$$\boldsymbol{C}_{\mathbf{V}}(h) = \boldsymbol{E}(\mathbf{V}_{t+h}\mathbf{V}_t^*) = \int_{-\pi}^{\pi} e^{ih\omega}\mathbf{\Lambda}_r(\omega)d\omega, \quad h \in \mathbb{Z}. \tag{28}$$

Then $\{\mathbf{V}_t\}$ is a regular time series.

Remark 1. *Assume that $\{\mathbf{X}_t\}$ is a d-dimensional regular time series of rank r and it has the spectral representation*

$$\mathbf{X}_t = \int_{-\pi}^{\pi} e^{it\omega} d\mathbf{Z}_\omega, \quad t \in \mathbf{Z}.$$

Assume as well that its spectral density matrix $\boldsymbol{f}$ has the spectral decomposition (25). Then the r-dimensional time series $\{\mathbf{V}_t\}$ corresponding to (28) can be written as a linear filtering of $\{\mathbf{X}_t\}$:

$$\mathbf{V}_t = \int_{-\pi}^{\pi} e^{it\omega}\tilde{\boldsymbol{U}}^*(\omega)d\mathbf{Z}_\omega, \quad t \in \mathbf{Z}, \tag{29}$$

see (2).

By (4) and (27), its spectral density is

$$\boldsymbol{f}_{\mathbf{V}}(\omega) = \tilde{\boldsymbol{U}}^*(\omega)\boldsymbol{f}(\omega)\tilde{\boldsymbol{U}}(\omega) = \mathbf{\Lambda}_r(\omega), \tag{30}$$

and

$$\begin{aligned}
\tilde{\boldsymbol{U}}^*(\omega) &= \sum_{j=0}^{\infty} \boldsymbol{\psi}^*(j)e^{ij\omega}, \qquad \sum_{j=0}^{\infty} \|\boldsymbol{\psi}^*(j)\|_F^2 < \infty, \\
\mathbf{V}_t &= \int_{-\pi}^{\pi} e^{it\omega} \sum_{j=0}^{\infty} \boldsymbol{\psi}^*(j)e^{ij\omega} d\mathbf{Z}_\omega = \sum_{j=0}^{\infty} \boldsymbol{\psi}^*(j) \int_{-\pi}^{\pi} e^{i(t+j)\omega} d\mathbf{Z}_\omega \\
&= \sum_{j=0}^{\infty} \boldsymbol{\psi}^*(j)\mathbf{X}_{t+j}.
\end{aligned} \tag{31}$$

Note that the above $\mathbf{V}_t$ will be later called Principal Component (PC) process.

Remark 2. *Comparing Corollary 4 and Theorem 4 shows that in the full rank case, condition (3) in Theorem 4 follows from conditions (1) and (2).*

Corollary 5. *Assume that $\{\mathbf{X}_t\}$ is a d-dimensional regular stationary time series of rank r, $1 \le r \le d$. Then a Kolmogorov–Szegő formula holds:*

$$\det \mathbf{\Sigma}_r = (2\pi)^r \exp \int_{-\pi}^{\pi} \log\det \mathbf{\Lambda}_r(\omega)\frac{d\omega}{2\pi} = (2\pi)^r \exp \int_{-\pi}^{\pi} \sum_{j=1}^{r} \log \lambda_j(\omega)\frac{d\omega}{2\pi},$$

where $\mathbf{\Lambda}_r$ is defined by (26) and $\mathbf{\Sigma}_r$ is the covariance matrix of the innovation process of an r-dimensional subprocess $\{\mathbf{X}_t^{(r)}\}$ of rank r.

4.3 Classification of non-regular multidimensional time series

The classification of multidimensional time series is – not surprisingly – more complex than the one-dimensional ones, see the 1D classes in the preceding lesson. We call a time series *non-regular* if either it is singular or its Wold decomposition contains two orthogonal, non-vanishing processes: a regular and a singular one. The classification below follows from Theorem 4.

In dimension $d > 1$ a non-singular process beyond its regular part may have a singular part with non-vanishing spectral density. For example, if $d = 3$ and the components $\{(X_t^1, X_t^2, X_t^3)\}$ are orthogonal to each other, it is possible that $\{X_t^1\}$ is regular of rank 1, $\{X_t^2\}$ is Type (1) singular, and $\{X_t^3\}$ is Type (2) singular.

Below we are considering a d-dimensional stationary time series $\{\mathbf{X}_t\}$ with spectral measure $d\boldsymbol{F}$.

- *Type (0) non-regular processes.* In this case the spectral measure $d\boldsymbol{F}$ of the time series $\{\mathbf{X}_t\}$ is singular w.r.t. the Lebesgue measure in $[-\pi, \pi]$. Clearly, type (0) non-regular processes are simply singular ones. Like in the 1D case, we may further divide this class into processes with a discrete spectrum or processes with a continuous singular spectrum or processes with both.

- *Type (1) non-regular processes.* The time series has an absolutely continuous spectral measure with density $\boldsymbol{f}$, but rank$(\boldsymbol{f})$ is not constant. It means that there exist measurable subsets $A, B \subset [-\pi, \pi]$ such that

$d\omega(A) > 0$ and $d\omega(B) > 0$, $\text{rank}(\boldsymbol{f}(\omega)) = r_1$ if $\omega \in A$, $\text{rank}(\boldsymbol{f}(\omega)) = r_2$ if $\omega \in B$, and $r_1 \neq r_2$. Here $d\omega$ denotes the Lebesgue measure in $[-\pi, \pi]$.

- *Type (2) non-regular processes.* The time series has an absolutely continuous spectral measure with density $\boldsymbol{f}$ which has constant rank r a.e., $1 \leq r \leq d$, but

$$\int_{-\pi}^{\pi} \log \det \boldsymbol{\Lambda}_r(\omega) d\omega = \int_{-\pi}^{\pi} \sum_{j=1}^{r} \log \lambda_j(\omega)\, d\omega = -\infty,$$

  where $\boldsymbol{\Lambda}_r$ is defined by (26).

- *Type (3) non-regular processes.* The time series has an absolutely continuous spectral measure with density $\boldsymbol{f}$ which has constant rank r a.e., $1 \leq r < d$,

$$\int_{-\pi}^{\pi} \log \det \boldsymbol{\Lambda}_r(\omega) d\omega = \int_{-\pi}^{\pi} \sum_{j=1}^{r} \log \lambda_j(\omega)\, d\omega > -\infty,$$

  but the unitary matrix function $\tilde{\boldsymbol{U}}(\omega)$ appearing in the spectral decomposition of $\boldsymbol{f}(\omega)$ in (27) does not belong to the Hardy space H^2.

By Corollary 3, if $\{\mathbf{X}_t\}$ has full rank $r = d$ and it is non-singular, then it can have only a Type (0) singular part.

5 Low rank approximation

The aim of this section is to approximate a time series of constant rank r with one of smaller rank k. This problem was treated by Brillinger, where it was called Principal Component Analysis (PCA) in the Frequency Domain. We show the important fact that when the process is regular, the low rank approximation can also be chosen regular.

5.1 Approximation of time series of constant rank

Assume that $\{\mathbf{X}_t\}$ is a d-dimensional stationary time series of constant rank r, $1 \leq r \leq d$. By Theorem 1, it is equivalent to the assumption that $\{\mathbf{X}_t\}$

can be written as a sliding summation of form (11). The spectral density $\boldsymbol{f}$ of the process has rank r a.e., and so we may write its eigenvalues as

$$\lambda_1(\omega) \geq \cdots \geq \lambda_r(\omega) > 0, \quad \lambda_{r+1}(\omega) = \cdots = \lambda_d(\omega) = 0. \tag{32}$$

Also, the spectral decomposition of $\boldsymbol{f}$ is

$$\boldsymbol{f}(\omega) = \sum_{j=1}^{r} \lambda_j(\omega)\mathbf{u}_j(\omega)\mathbf{u}_j^*(\omega) = \tilde{\boldsymbol{U}}_r(\omega)\tilde{\boldsymbol{\Lambda}}_r(\omega)\tilde{\boldsymbol{U}}_r^*(\omega), \qquad \text{a.e.} \quad \omega \in [-\pi, \pi], \tag{33}$$

where $\tilde{\boldsymbol{\Lambda}}_r(\omega) := \operatorname{diag}[\lambda_1(\omega), \ldots, \lambda_r(\omega)]$, $\mathbf{u}_j(\omega) \in \mathbb{C}^d$ $(j = 1, \ldots, r)$ are the corresponding orthonormal eigenvectors, and $\tilde{\boldsymbol{U}}_r(\omega) \in \mathbb{C}^{d\times r}$ is the matrix of these column vectors.

Now the problem we are treating can be described as follows. Given an integer k, $1 \leq k \leq r$, find a process $\{\mathbf{X}_t^{(k)}\}$ of constant rank k which is a linear transform of $\{\mathbf{X}_t\}$ and which minimizes the distance

$$\begin{aligned} \|\mathbf{X}_t - \mathbf{X}_t^{(k)}\|^2 &= \mathbb{E}\left\{(\mathbf{X}_t - \mathbf{X}_t^{(k)})^* \,(\mathbf{X}_t - \mathbf{X}_t^{(k)})\right\} \\ &= \operatorname{tr} \operatorname{Cov}\left\{(\mathbf{X}_t - \mathbf{X}_t^{(k)}), (\mathbf{X}_t - \mathbf{X}_t^{(k)})\right\}. \end{aligned} \tag{34}$$

Consider the spectral representations of $\{\mathbf{X}_t\}$ and $\{\mathbf{X}_t^{(k)}\}$, see (2):

$$\mathbf{X}_t = \int_{-\pi}^{\pi} e^{it\omega} d\mathbf{Z}_\omega, \quad \mathbf{X}_t^{(k)} = \int_{-\pi}^{\pi} e^{it\omega}\boldsymbol{T}(\omega) d\mathbf{Z}_\omega, \quad t \in \mathbb{Z}.$$

Then we can rewrite (34) as

$$\begin{aligned} \|\mathbf{X}_t - \mathbf{X}_t^{(k)}\|^2 &= \operatorname{tr} \int_{-\pi}^{\pi} (I_d - \boldsymbol{T}(\omega))\boldsymbol{f}(\omega)(I_d - \boldsymbol{T}^*(\omega))\, d\omega \\ &= \operatorname{tr} \int_{-\pi}^{\pi} (I_d - \boldsymbol{T}(\omega))\tilde{\boldsymbol{U}}_r(\omega)\tilde{\boldsymbol{\Lambda}}_r(\omega)\tilde{\boldsymbol{U}}_r^*(\omega)(I_d - \boldsymbol{T}^*(\omega))\, d\omega, \end{aligned} \tag{35}$$

which clearly does not depend on $t \in \mathbb{Z}$.

To find the minimizing linear transformation $\boldsymbol{T}(\omega)$, we have to study the non-negative definite quadratic form

$$\mathbf{v}^*\boldsymbol{f}(\omega)\mathbf{v} = \mathbf{v}^*\tilde{\boldsymbol{U}}_r(\omega)\tilde{\boldsymbol{\Lambda}}_r(\omega)\tilde{\boldsymbol{U}}_r^*(\omega)\mathbf{v}, \quad \mathbf{v} \in \mathbb{C}^d, \quad |\mathbf{v}| = 1.$$

By (32), there is a monotonicity: taking the orthogonal projections $\mathbf{u}_j\mathbf{u}_j^*$ $(j = 1, \dots, r)$ in the space $\mathbb{C}^d$ one-by-one, the sequence

$$\mathbf{v}_j^* \boldsymbol{f}(\omega)\mathbf{v}_j, \qquad \mathbf{v}_j \in \mathbf{u}_j(\omega)\mathbf{u}_j^*(\omega)\,\mathbb{C}^d \qquad (j = 1, \dots, r),$$

is non-increasing. Since $I_d = \sum_{j=1}^d \mathbf{u}_j(\omega)\mathbf{u}_j^*(\omega)$ and $\boldsymbol{T}(\omega)$ must have rank k a.e., (35) implies that the minimizing linear transformation must be

$$\boldsymbol{T}(\omega) = \sum_{j=1}^{k} \mathbf{u}_j(\omega)\mathbf{u}_j^*(\omega) = \tilde{\boldsymbol{U}}_k(\omega)\tilde{\boldsymbol{U}}_k^*(\omega). \tag{36}$$

This is in accord with the theory of law rank approximations, where $\boldsymbol{T}(\omega)$ is the projection onto the subspace spanned by the k leading eigenvectors of $\boldsymbol{f}(\omega)$ that correspond to the k largest eigenvalues of this matrix.

Thus we have proved that

$$\mathbf{X}_t^{(k)} = \int_{-\pi}^{\pi} e^{it\omega}\tilde{\boldsymbol{U}}_k(\omega)\tilde{\boldsymbol{U}}_k^*(\omega)d\mathbf{Z}_\omega, \quad t \in \mathbb{Z}. \tag{37}$$

Then by (4), the spectral density of $\{\mathbf{X}_t^{(k)}\}$ is

$$\begin{aligned}\boldsymbol{f}_k(\omega) &= \tilde{\boldsymbol{U}}_k(\omega)\tilde{\boldsymbol{U}}_k^*(\omega)\tilde{\boldsymbol{U}}_r(\omega)\tilde{\boldsymbol{\Lambda}}_r(\omega)\tilde{\boldsymbol{U}}_r^*(\omega)\tilde{\boldsymbol{U}}_k(\omega)\tilde{\boldsymbol{U}}_k^*(\omega)\\ &= \tilde{\boldsymbol{U}}_k(\omega)\begin{bmatrix} I_k & 0_{k\times(r-k)}\end{bmatrix}\tilde{\boldsymbol{\Lambda}}_r(\omega)\begin{bmatrix} I_k \\ 0_{(r-k)\times k}\end{bmatrix}\tilde{\boldsymbol{U}}_k^*(\omega)\\ &= \tilde{\boldsymbol{U}}_k(\omega)\tilde{\boldsymbol{\Lambda}}_k(\omega)\tilde{\boldsymbol{U}}_k^*(\omega), \quad \omega \in [-\pi, \pi]. \end{aligned} \tag{38}$$

Further, the covariance function of $\{\mathbf{X}_t^{(k)}\}$ is

$$\boldsymbol{C}_k(h) := \int_{-\pi}^{\pi} e^{ih\omega}\boldsymbol{f}_k(\omega)d\omega, \qquad h \in \mathbb{Z}. \tag{39}$$

The next theorem summarizes the results above.

Theorem 5. *Assume that $\{\mathbf{X}_t\}$ is a d-dimensional stationary time series of constant rank r, $1 \le r \le d$, with spectral density $\boldsymbol{f}$. Let (32) and (33) be the spectral decomposition of $\boldsymbol{f}$.*

(a) Then

$$\mathbf{X}_t^{(k)} = \int_{-\pi}^{\pi} e^{it\omega}\tilde{\boldsymbol{U}}_k(\omega)\tilde{\boldsymbol{U}}_k^*(\omega)d\mathbf{Z}_\omega, \quad t \in \mathbb{Z},$$

is the approximating process of rank k, $1 \le k \le r$, which minimizes the mean square error of the approximation.

(b) For the mean square error we have

$$\|\mathbf{X}_t - \mathbf{X}_t^{(k)}\|^2 = \int_{-\pi}^{\pi} \sum_{j=k+1}^{r} \lambda_j(\omega)\, d\omega, \qquad t \in \mathbb{Z}, \tag{40}$$

and

$$\frac{\|\mathbf{X}_t - \mathbf{X}_t^{(k)}\|^2}{\|\mathbf{X}_t\|^2} = \frac{\int_{-\pi}^{\pi} \sum_{j=k+1}^{r} \lambda_j(\omega)\, d\omega}{\int_{-\pi}^{\pi} \sum_{j=1}^{r} \lambda_j(\omega)\, d\omega}, \quad t \in \mathbb{Z}. \tag{41}$$

(c) If condition

$$\lambda_k(\omega) \geq \Delta > \epsilon \geq \lambda_{k+1}(\omega) \qquad \forall \omega \in [-\pi, \pi], \tag{42}$$

holds then we also have

$$\|\mathbf{X}_t - \mathbf{X}_t^{(k)}\| \leq (2\pi(r-k)\epsilon)^{1/2}, \quad t \in \mathbb{Z}$$

and

$$\frac{\|\mathbf{X}_t - \mathbf{X}_t^{(k)}\|}{\|\mathbf{X}_t\|} \leq \left(\frac{(r-k)\epsilon}{r\Delta} \right)^{\frac{1}{2}}, \qquad t \in \mathbb{Z}. \tag{43}$$

Equation (37) can be factored. As in Theorem 4, one can take the Fourier series of the sub-unitary matrix function $\tilde{\boldsymbol{U}}_k(\omega) \in L^2$:

$$\tilde{\boldsymbol{U}}_k(\omega) = \sum_{j=-\infty}^{\infty} \boldsymbol{\psi}(j) e^{-ij\omega}, \quad \boldsymbol{\psi}(j) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ij\omega} \tilde{\boldsymbol{U}}_k(\omega) d\omega \in \mathbb{C}^{d \times k},$$

where $\sum_{j=-\infty}^{\infty} \|\boldsymbol{\psi}(j)\|_F^2 < \infty$. Consequently,

$$\tilde{\boldsymbol{U}}_k^*(\omega) = \sum_{j=-\infty}^{\infty} \boldsymbol{\psi}^*(j) e^{ij\omega}, \quad \omega \in [-\pi, \pi].$$

If the time series $\{\mathbf{V}_t\}$ is defined by the linear filter

$$\mathbf{V}_t := \int_{-\pi}^{\pi} e^{it\omega} \tilde{\boldsymbol{U}}_k^*(\omega) d\mathbf{Z}_\omega \in \mathbb{C}^k, \quad t \in \mathbb{Z},$$

then similarly to (31) it follows that $\{\mathbf{V}_t\}$ can be obtained from the original time series $\{\mathbf{X}_t\}$ by a sliding summation:

$$\mathbf{V}_t = \sum_{j=-\infty}^{\infty} \boldsymbol{\psi}^*(j) \mathbf{X}_{t+j}, \quad t \in \mathbb{Z},$$

and similarly to (30), its spectral density is a diagonal matrix:

$$f_{\mathbf{V}}(\omega) = \mathbf{\Lambda}_k(\omega) = \mathrm{diag}[\lambda_1(\omega), \dots, \lambda_k(\omega)].$$

It means that the covariance matrix function of $\{\mathbf{V}_t\}$ is also diagonal:

$$\boldsymbol{C}_{\mathbf{V}}(h) = \mathrm{diag}[c_{11}(h), \dots, c_{kk}(h)], \quad c_{jj}(h) = \int_{-\pi}^{\pi} e^{ih\omega} \lambda_j(\omega) d\omega, \quad h \in \mathbb{Z},$$

that is, the components of the process $\{\mathbf{V}_t\}$ are orthogonal to each other.

Using a second linear filtration, equivalently, a second sliding summation, one can obtain the k-rank approximation $\{\mathbf{X}_t^{(k)}\}$ from $\{\mathbf{V}_t\}$:

$$\begin{aligned}\mathbf{X}_t^{(k)} &= \int_{-\pi}^{\pi} e^{it\omega} \tilde{\boldsymbol{U}}_k(\omega) \tilde{\boldsymbol{U}}_k^*(\omega) d\mathbf{Z}_\omega = \int_{-\pi}^{\pi} e^{it\omega} \tilde{\boldsymbol{U}}_k(\omega) d\mathbf{Z}_\omega^{\mathbf{V}} \\ &= \sum_{j=-\infty}^{\infty} \boldsymbol{\psi}(j) \mathbf{V}_{t-j}, \quad t \in \mathbb{Z}.\end{aligned}$$

Notice the *dimension reduction* in this approximation. Dimension d of the original process $\{\mathbf{X}_t\}$ can be reduced to dimension $k < d$ with the cross-sectionally orthogonal process $\{\mathbf{V}_t\}$, obtained by linear filtration, from which the low-rank approximation $\{\mathbf{X}_t^{(k)}\}$ can be reconstructed also by linear filtration. Of course, this is useful only if the error of the approximation given by Theorem 5 is small enough.

Since $\tilde{\boldsymbol{U}}_k \tilde{\boldsymbol{U}}_k^* \in L^2$ as well, one can take the L^2-convergent Fourier series

$$\tilde{\boldsymbol{U}}_k(\omega) \tilde{\boldsymbol{U}}_k^*(\omega) = \sum_{j,\ell=-\infty}^{\infty} \boldsymbol{\psi}(j) e^{-ij\omega} \boldsymbol{\psi}^*(\ell) e^{i\ell\omega} = \sum_{m=-\infty}^{\infty} \boldsymbol{w}(m) e^{-im\omega},$$

where $\omega \in [-\pi, \pi]$ and

$$\boldsymbol{w}(m) = \sum_{j=-\infty}^{\infty} \boldsymbol{\psi}(j)\, \boldsymbol{\psi}^*(j-m) \in \mathbb{C}^{d \times d}, \quad \sum_{m=-\infty}^{\infty} \|\boldsymbol{w}(m)\|_F^2 < \infty. \tag{44}$$

By (7) it implies that the filtered process $\{\mathbf{X}_t^{(k)}\}$ can be obtained directly from $\{\mathbf{X}_t\}$ by a two-sided sliding summation:

$$\mathbf{X}_t^{(k)} = \sum_{m=-\infty}^{\infty} \boldsymbol{w}(m) \mathbf{X}_{t-m}, \quad t \in \mathbb{Z}.$$

5.2 Approximation of regular time series

In the special case when $\{\mathbf{X}_t\}$ is a d-dimensional regular time series of rank r, it follows by Theorem 4 that $\tilde{\boldsymbol{U}}_r(\omega)$ belongs to the Hardy space $H^\infty \subset H^2$, so the same holds for $\tilde{\boldsymbol{U}}_k(\omega)$ as well. Then its Fourier series is one-sided:

$$\tilde{\boldsymbol{U}}_k(\omega) = \sum_{j=0}^{\infty} \boldsymbol{\psi}(j) e^{-ij\omega}, \quad \boldsymbol{\psi}(j) \in \mathbb{C}^{d\times k}, \quad \sum_{j=0}^{\infty} \|\boldsymbol{\psi}(j)\|_F^2 < \infty.$$

It is clear that for the approximating process $\{\mathbf{X}_t^{(k)}\}$ each of the conditions (1), (2) and (3) of Theorem 4 hold, thus it is also a regular time series.

Theorem 5 and Corollary **??** are valid for regular processes without change. However, the factorization of the approximation discussed above is different in the regular case, because several of the summations become one-sided. Thus we have

$$\tilde{\boldsymbol{U}}_k^*(\omega) = \sum_{j=0}^{\infty} \boldsymbol{\psi}^*(j) e^{ij\omega}, \quad \omega \in [-\pi, \pi].$$

Consequently, the k-dimensional, cross-sectionally orthogonal process $\{\mathbf{V}_t\}$ becomes

$$\mathbf{V}_t = \sum_{j=0}^{\infty} \boldsymbol{\psi}^*(j) \mathbf{X}_{t+j}, \quad t \in \mathbb{Z}.$$

Further, the reconstruction of the k-rank approximation $\{\mathbf{X}_t^{(k)}\}$ from $\{\mathbf{V}_t\}$ is

$$\mathbf{X}_t^{(k)} = \sum_{j=0}^{\infty} \boldsymbol{\psi}(j) \mathbf{V}_{t-j}, \quad t \in \mathbb{Z}.$$

The direct evaluation of $\{\mathbf{X}_t^{(k)}\}$ from $\{\mathbf{X}_t\}$ takes now the following form:

$$\tilde{\boldsymbol{U}}_k(\omega)\tilde{\boldsymbol{U}}_k^*(\omega) = \sum_{j,\ell=0}^{\infty} \boldsymbol{\psi}(j) e^{-ij\omega} \boldsymbol{\psi}^*(\ell) e^{i\ell\omega} = \sum_{m=-\infty}^{\infty} \boldsymbol{w}(m) e^{-im\omega}, \quad \omega \in [-\pi, \pi],$$

where $\boldsymbol{w}(m) = \sum_{j=\max(0,m)}^{\infty} \boldsymbol{\psi}(j)\, \boldsymbol{\psi}^*(j-m)$. It implies that the filtered process $\{\mathbf{X}_t^{(k)}\}$ is *not* causally subordinated to the original regular process $\{\mathbf{X}_t\}$ in general, since it can be obtained from $\{\mathbf{X}_t\}$ by a two-sided sliding summation:

$$\mathbf{X}_t^{(k)} = \sum_{m=-\infty}^{\infty} \boldsymbol{w}(m) \mathbf{X}_{t-m}, \quad t \in \mathbb{Z},$$

see (7). On the other hand, it is clear that if $\|\boldsymbol{\psi}(j)\|_F$ goes to 0 fast enough as $j \to \infty$, one does not have to use too many 'future' terms of $\{\mathbf{X}_t\}$ to get a good enough approximation of $\{\mathbf{X}_t^{(k)}\}$. In practice one can also replace the future values of $\{\mathbf{X}_t\}$ by $\mathbf{0}$ to get a causal approximation of $\{\mathbf{X}_t^{(k)}\}$.

6 Rational spectral densities

An important subclass of the class of regular stationary time series, which in turn is a subclass of time series with constant rank, is such that each entry $f^{k\ell}(\omega)$ in the spectral density matrix $\boldsymbol{f}$ is a rational complex function in $z = e^{-i\omega}$. This subclass is the same as that of the stable VARMA(p, q) processes. Also, this is the subclass of stable stochastic linear systems with finite dimensional state space representation. Moreover, this is the subclass of stable stochastic linear systems with rational transfer function. In sum, this is the subclass of weakly stationary time series that can be described by finitely many complex valued parameters.

Remark 3. *Every minor (that is, the determinant of a sub-matrix) of a rational matrix is a rational function which is either identically zero or has zeros and poles at only finitely many points. Hence a weakly stationary time series with a spectral density* $\boldsymbol{f}$ *which is a rational matrix in* $z = e^{-i\omega}$ *must be of constant rank* r*. By Theorem 1 it implies that such a process can be represented as a two-sided infinite MA process. More accurately, Theorem 7 below shows that such a* stable *process is regular with a one-sided causal MA(*∞*) representation.*

6.1 Smith–McMillan form

The Smith–McMillan form is a useful tool by which one can diagonalize a non-negative definite rational matrix so that both the obtained diagonal matrix and the transformation matrix used for the diagonalization are rational matrices. The usual technique of linear algebra which uses eigenvalues and eigenvectors does not have this important property, since the eigenvalues and the entries of eigenvectors of a rational matrix are not rational functions in general.

Lemma 1. *Let* $\boldsymbol{A}(z) = [a_{jk}(z)]_{d\times d}$ *be a rational matrix which is self-adjoint and non-negative definite for* $z \in T$*, having no poles on* T*, and whose rank* *is r, $1 \le r \le d$, for $z \in T \setminus Z$, where Z is a finite set. Then we can write*

$$\boldsymbol{A}(z) = \boldsymbol{E}(z)\boldsymbol{\Lambda}(z)\boldsymbol{E}^*(z) \quad (z \in T \setminus Z), \tag{45}$$

where $\boldsymbol{\Lambda}(z)$ and $\boldsymbol{E}(z)$ are rational matrices,

$$\boldsymbol{\Lambda}(z) = \mathrm{diag}[\lambda_1(z), \dots, \lambda_r(z)], \quad \lambda_j(z) > 0 \quad (j = 1, \dots, r;\ z \in T \setminus Z).$$

Here $\mathrm{diag}[\lambda_1, \dots, \lambda_r]$ denotes an $r \times r$ diagonal matrix with entries λ_j ($j = 1, \dots, r$) in its main diagonal. Also,

$$\boldsymbol{E}(z) = [e_{jk}(z)]_{d \times r} \quad (z \in T \setminus Z),$$

is a lower unit trapezoidal matrix:

1. *$e_{jk}(z) = 0$ if $k > j$,*
2. *$e_{jk}(z) = 1$ if $k = j$.*

6.2 Spectral factors of a rational spectral density matrix

Theorem 6. *Let $\boldsymbol{A}(z) = [a_{jk}(z)]_{d \times d}$ be a rational matrix which is self-adjoint and non-negative definite for $z \in T$, having no poles on T, and whose rank is r for all $z \in T \setminus Z$, where Z is a finite set. Then we can write*

$$\boldsymbol{A}(z) = \frac{1}{2\pi}\boldsymbol{\Phi}(z)\boldsymbol{\Phi}^*(z) \quad (z \in T \setminus Z),$$

where $\boldsymbol{\Phi}(z) = [\Phi_{jk}(z)]_{d \times r}$ is a rational matrix, analytic in the open unit disc $D = \{z : |z| < 1\}$, and has rank r for any $z \in T \setminus Z$.

Theorem 7. *Let $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ be a d-dimensional weakly stationary time series with spectral density $\boldsymbol{f}(\omega)$ which is a rational function in $z = e^{-i\omega}$. By Remark 3, $\boldsymbol{f}(\omega)$ has constant rank r for all $\omega \in [-\pi, \pi] \setminus Z$, where Z is a finite set. Then $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ can be represented as a regular process, a causal MA process*

$$\mathbf{X}_t = \sum_{j=0}^{\infty} \boldsymbol{b}(j)\boldsymbol{\xi}_{t-j} \quad (t \in \mathbb{Z}), \quad \sum_{j=0}^{\infty} \|\boldsymbol{b}(j)\|_F^2 < \infty,$$

where $\boldsymbol{b}(j) \in \mathbb{C}^{d \times r}$ ($j = 0, 1, \dots$) is a non-random matrix-valued function and $\{\boldsymbol{\xi}_t\}_{t \in \mathbb{Z}}$ is a $\mathrm{WN}(I_r)$ (orthonormal) process.

7 Multidimensional ARMA (VARMA) processes

More special regular processes are the ones that can be finitely parametrized. Those are, in fact, the causal VARMA (Vector AutoRegressive Moving Average) processes that also have an MFD (matrix fractional description) or state space representation.

The d-dimensional VARMA(p, q) process of $\mathbf{0}$ mean is defined as follows:

$$\mathbf{X}_t = \boldsymbol{\alpha}_1 \mathbf{X}_{t-1} + \cdots + \boldsymbol{\alpha}_p \mathbf{X}_{t-p} + \mathbf{U}_t + \boldsymbol{\beta}_1 \mathbf{U}_{t-1} + \cdots + \boldsymbol{\beta}_q \mathbf{U}_{t-q},$$

where $\{\mathbf{U}_t\} \sim WN(\mathbf{0}, \boldsymbol{\Sigma})$ is d-dimensional white noise and $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_p, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_q$ are $d \times d$ complex matrices, for the time being, $\boldsymbol{\alpha}_0 = \boldsymbol{I}_d$. This defining equation can concisely be written as

$$\boldsymbol{\alpha}(L)\, \mathbf{X}_t = \boldsymbol{\beta}(L)\, \mathbf{U}_t,$$

where $\boldsymbol{\alpha}(z) = \boldsymbol{I} - \boldsymbol{\alpha}_1 z - \cdots - \boldsymbol{\alpha}_p z^p$ and $\boldsymbol{\beta}(z) = \boldsymbol{I} + \boldsymbol{\beta}_1 z + \cdots + \boldsymbol{\beta}_q z^q$ are matrix-valued complex polynomials, namely, the AR and MA polynomials; whereas, L is the backward shift operator. In particular, in the $q = 0$ case we have a VAR(p), whereas, in the $p = 0$ case we have a VMA(q) process.

If the condition $|\boldsymbol{\alpha}(z)| \neq 0$ for $|z| \leq 1$ for the VAR polynomial is satisfied, it is called *stability*, then we have a causal representation of the process:

$$\mathbf{X}_t = \sum_{j=0}^{\infty} \boldsymbol{H}_j \mathbf{U}_{t-j}, \tag{46}$$

with $\{\mathbf{U}_t\} \sim \mathrm{WN}(\mathbf{0}, \boldsymbol{\Sigma})$ and the coefficient matrices $\boldsymbol{H}_j$s come from the power series expansion of the *transfer function*, as in the 1D case:

$$\boldsymbol{H}(z) = \sum_{j=0}^{\infty} \boldsymbol{H}_j z^j, \quad |z| \leq 1,$$

where $\boldsymbol{H}(z) = \boldsymbol{\alpha}^{-1}(z)\, \boldsymbol{\beta}(z)$ and $\boldsymbol{H}_j$s are called *impulse responses*. This is the MFD. So we can write the original process as

$$\mathbf{X}_t = \boldsymbol{\alpha}^{-1}(z)\ \boldsymbol{\beta}(z)\, \mathbf{U}_t = \boldsymbol{H}(z)\, \mathbf{U}_t.$$

If, in addition to the stability condition, the *inverse stability or strict miniphase condition*, i.e., $|\boldsymbol{\beta}(z)| \neq 0$ for $|z| \leq 1$ also holds (concerning the

MA polynomial), then $\mathbf{U}_t$ can also be expanded in terms of $\mathbf{X}_k$s ($k \le t$). Also, under stability and inverse stability, Equation (47) is the multidimensional Wold decomposition of the VARMA process with the innovations $\mathbf{U}_t$s (there is no singular part).

Note that the innovations can be transformed into an orthonormal process. Indeed, if the white noise covariance matrix is non-singular, it can be decomposed as $\mathbf{\Sigma} = \boldsymbol{B}\boldsymbol{B}^T$ with the $d \times d$ non-singular $\boldsymbol{B}$, and Equation (47) can be written like

$$\mathbf{X}_t = \sum_{j=0}^{\infty} \boldsymbol{H}_j \mathbf{U}_{t-j} = \sum_{j=0}^{\infty} \boldsymbol{H}_j \boldsymbol{B}_j \boldsymbol{B}_j^{-1} \mathbf{U}_{t-j} = \sum_{j=0}^{\infty} (\boldsymbol{H}_j \boldsymbol{B}_j) \boldsymbol{\xi}_{t-j},$$

where $\{\boldsymbol{\xi}_t\} \sim \mathrm{WN}(\boldsymbol{I}_d)$ is an orthonormal process (both longitudinally and cross-sectionally).

Going further, even if $\mathbf{\Sigma}$ is singular (it has rank $r < d$), it has the Gram-decomposition like $\mathbf{\Sigma} = \boldsymbol{B}\boldsymbol{B}^T$ with the $d \times r$ matrix $\boldsymbol{B}$ of full rank, and the above equation is also valid with $\{\boldsymbol{\xi}_t\} \sim \mathrm{WN}(\boldsymbol{I}_r)$.

It is important, then a VAR(p) process makes rise of a finite prediction of $\mathbf{X}_t$ based on its p-length long past:

$$\mathbf{X}_t = \mathbf{a}_1 \mathbf{X}_{t-1} + \cdots + \mathbf{a}_p \mathbf{X}_{t-p} + \mathbf{U}_t.$$

Consequently,

$$\hat{\mathbf{X}}_t = \mathbf{a}_1 \mathbf{X}_{t-1} + \cdots + \mathbf{a}_p \mathbf{X}_{t-p}$$

is the same as the prediction of $\mathbf{X}_t$ with its p-length long past that extends to the infinite past prediction. The multidimensional Yule-Walker equations also work in this situation, see the next lesson for details.

VARMA processes also have a state spece representation, we will tuch upon this setup when discussing Kálmán filtering.

The importance and wide applicability of VARMA processes is further emphasized by the next proposition.

Proposition 1. *The stable VARMA processes are dense among the regular time series. More exactly, for any regular stationary time series* $\{\mathbf{X}_t\}$ *and for any* $\epsilon > 0$ *there exists a positive integer* N *and a VMA(*N*) process* $\{\mathbf{Y}_t\}$ *such that*

$$\|\mathbf{X}_t - \mathbf{Y}_t\| < \epsilon, \quad \forall t \in \mathbb{Z}.$$

This shows that stable VARMA processes are dense among the regular time series.

8 Summary

The d-dimensional, weakly stationary time series $\{\mathbf{X}_t\}$ has a spectral density matrix $\boldsymbol{f}$ with constant rank $r \le d$ (almost everywhere on $[-\pi, \pi]$) if and only if $\boldsymbol{f}$ can be factored as

$$\boldsymbol{f}(\omega) = \frac{1}{2\pi}\boldsymbol{\phi}(\omega)\boldsymbol{\phi}^*(\omega) \quad \text{for a.e. } \omega \in [-\pi, \pi],$$

where $\boldsymbol{\phi}(\omega) \in \mathbb{C}^{d \times r}$:

$$\boldsymbol{\phi}(\omega) = \sum_{j=-\infty}^{\infty} \boldsymbol{b}(j) e^{-ij\omega}, \quad \sum_{j=-\infty}^{\infty} |b_{qs}(j)|^2 < \infty \quad (q = 1, \dots, d;\ s = 1, \dots, r).$$

Equivalently, $\mathbf{X_t}$ can be represented as a two-sided TLF

$$\mathbf{X}_t = \sum_{j=-\infty}^{\infty} \boldsymbol{b}(j) \boldsymbol{\xi}_{t-j},$$

where $\boldsymbol{b}(j) \in \mathbb{C}^{d \times r}$ $(j \in \mathbb{Z})$ is a non-random matrix-valued sequence, the Fourier coefficients of the function $\boldsymbol{\phi}(\omega)$; further, $\{\boldsymbol{\xi}_s\}$ is a WN$(\boldsymbol{I}_r)$ sequence.

Regular time series are subclasses of the constant rank ones in that they have a one-sided MA representation:

$$\mathbf{X}_t = \sum_{j=0}^{\infty} \boldsymbol{b}(j) \boldsymbol{\xi}_{t-j}, \quad t \in \mathbb{Z}.$$

This representation of $\{\mathbf{X}_t\}$, as a process of constant rank spectral density matrix $\boldsymbol{f}(\omega)$, is equivalent that its spectral density can be factored in the form

$$\boldsymbol{f}(\omega) = \frac{1}{2\pi}\boldsymbol{\phi}(\omega)\boldsymbol{\phi}^*(\omega), \quad \boldsymbol{\phi}(\omega) = [\phi_{j\ell}(\omega)]_{d \times r}, \quad \text{for a.e. } \omega \in [-\pi, \pi],$$

where

$$\boldsymbol{\phi}(\omega) = \sum_{k=0}^{\infty} \boldsymbol{b}(k) e^{-ik\omega}, \quad \|\phi_{j\ell}\|_2^2 = \sum_{k=0}^{\infty} |b_{j\ell}(k)|^2 < \infty.$$

The multi-dimensional Wold decomposition also works as follows. Assume that $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ is an d-dimensional non-singular stationary time series. Then it can be represented as

$$\mathbf{X}_t = \mathbf{R}_t + \mathbf{Y}_t = \sum_{j=0}^{\infty} \boldsymbol{b}(j) \boldsymbol{\xi}_{t-j} + \mathbf{Y}_t \quad (t \in \mathbb{Z}),$$

where $\{\mathbf{R}_t\}$ is a d-dimensional regular time series subordinated to $\{\mathbf{X}_t\}$; $\{\mathbf{Y}_t\}$ is an d-dimensional singular time series subordinated to $\{\mathbf{X}_t\}$; $\{\boldsymbol{\xi}_t\}$ is an r-dimensional ($r \leq d$) WN($\boldsymbol{I}_r$) sequence subordinated to $\{\mathbf{X}_t\}$. Further, $\{\mathbf{R}_t\}$ and $\{\mathbf{Y}_t\}$ are orthogonal to each other, i.e.,

$$\mathbb{E}(\mathbf{R}_t \mathbf{Y}_s^*) = \boldsymbol{O}_d \quad (t, s \in \mathbb{Z});$$

$\boldsymbol{b}(j) = [b_{qs}(j)] \in \mathbb{C}^{d \times r}$ for each $j \geq 0$ and

$$\sum_{j=0}^{\infty} |b_{qs}(j)|^2 < \infty \quad (q = 1, \dots, d;\ s = 1, \dots, r).$$

It is important that the *orthonormal innovation process* $\{\boldsymbol{\xi}_t\}$ is not unique, but $\boldsymbol{\xi}_t$s are within the pairwise orthogonal *innovation subspaces* that are unique and their dimension is equal to the constant rank r of the spectral density matrix $\boldsymbol{f}$ of the process $\{\mathbf{X}_t\}$.

More special regular processes are the ones that can be finitely parametrized. Those are, in fact, the causal VARMA (Vector AutoRegressive Moving Average) processes that also have an MFD or state space representation.

The d-dimensional VARMA(p, q) process of $\mathbf{0}$ mean is defined as follows:

$$\mathbf{X}_t + \boldsymbol{\alpha}_1 \mathbf{X}_{t-1} + \cdots + \boldsymbol{\alpha}_p \mathbf{X}_{t-p} = \mathbf{U}_t + \boldsymbol{\beta}_1 \mathbf{U}_{t-1} + \cdots + \boldsymbol{\beta}_q \mathbf{U}_{t-q},$$

where $\{\mathbf{U}_t\} \sim WN(\mathbf{0}, \boldsymbol{\Sigma})$ is d-dimensional white noise and $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_p, \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_q$ are $d \times d$ complex matrices, for the time being, $\boldsymbol{\alpha}_0 = \boldsymbol{I}_d$. This defining equation can concisely be written as

$$\boldsymbol{\alpha}(L)\, \mathbf{X}_t = \boldsymbol{\beta}(L)\, \mathbf{U}_t,$$

where $\boldsymbol{\alpha}(z) = \boldsymbol{I} + \boldsymbol{\alpha}_1 z + \cdots + \boldsymbol{\alpha}_p z^p$ and $\boldsymbol{\beta}(z) = \boldsymbol{I} + \boldsymbol{\beta}_1 z + \cdots + \boldsymbol{\beta}_q z^q$ are matrix-valued complex polynomials, namely, the AR and MA polynomials; whereas, L is the backward shift operator. In particular, in the $q = 0$ case we have a VAR(p), whereas, in the $p = 0$ case we have a VMA(q) process.

If the condition $|\boldsymbol{\alpha}(z)| \neq 0$ for $|z| \leq 1$ for the VAR polynomial is satisfied (it is called stability), then we have a causal representation of the process:

$$\mathbf{X}_t = \sum_{j=0}^{\infty} \boldsymbol{H}_j \mathbf{U}_{t-j}, \tag{47}$$

with $\{\mathbf{U}_t\} \sim \mathrm{WN}(\mathbf{0}, \mathbf{\Sigma})$ and the coefficient matrices $\boldsymbol{H}_j$s come from the power series expansion of the *transfer function*:

$$\boldsymbol{H}(z) = \sum_{j=0}^{\infty} \boldsymbol{H}_j z^j, \quad |z| \le 1,$$

where $\boldsymbol{H}(z) = \boldsymbol{\alpha}^{-1}(z)\,\boldsymbol{\beta}(z)$ and $\boldsymbol{H}_j$s are called *impulse responses*. So we can write the original process as

$$\mathbf{X}_t = \boldsymbol{\alpha}^{-1}(z)\ \boldsymbol{\beta}(z)\,\mathbf{U}_t = \boldsymbol{H}(z)\,\mathbf{U}_t.$$

If, in addition to the stability condition, the *inverse stability or strict miniphase condition*, i.e., $|\boldsymbol{\beta}(z)| \ne 0$ for $|z| \le 1$ also holds (concerning the MA polynomial), then $\mathbf{U}_t$ can also be expanded in terms of $\mathbf{X}_k$s $(k \le t)$. Also, under stability and inverse stability, Equation (47) is the multidimensional Wold decomposition of the VARMA process with the innovations $\mathbf{Z}_t$s (there is no singular part).

Note that the innovations can be transformed into an orthonormal process. Indeed, if the white noise covariance matrix is non-singular, it can be decomposed as $\mathbf{\Sigma} = \boldsymbol{B}\boldsymbol{B}^T$ with the $d \times d$ non-singular $\boldsymbol{B}$, and Equation (47) can be written like

$$\mathbf{X}_t = \sum_{j=0}^{\infty} \boldsymbol{H}_j \mathbf{U}_{t-j} = \sum_{j=0}^{\infty} \boldsymbol{H}_j \boldsymbol{B}_j \boldsymbol{B}_j^{-1} \mathbf{U}_{t-j} = \sum_{j=0}^{\infty} (\boldsymbol{H}_j \boldsymbol{B}_j) \boldsymbol{\xi}_{t-j},$$

where $\{\boldsymbol{\xi}_t\} \sim \mathrm{WN}(\boldsymbol{I}_d)$ is an orthonormal process (both longitudinally and cross-sectionally).

Going further, even if $\mathbf{\Sigma}$ is singular (it has rank $r < d$), it has the Gram-decomposition like $\mathbf{\Sigma} = \boldsymbol{B}\boldsymbol{B}^T$ with the $d \times r$ matrix $\boldsymbol{B}$ of full rank, and the above equation is also valid with $\{\boldsymbol{\xi}_t\} \sim \mathrm{WN}(\boldsymbol{I}_r)$.

It is important, then a VAR(p) process makes rise of a finite prediction of $\mathbf{X}_t$ based on its p-length long past:

$$\mathbf{X}_t = \mathbf{a}_1 \mathbf{X}_{t-1} + \cdots + \mathbf{a}_p \mathbf{X}_{t-p} + \mathbf{U}_t.$$

Consequently,

$$\hat{\mathbf{X}}_t = \mathbf{a}_1 \mathbf{X}_{t-1} + \cdots + \mathbf{a}_p \mathbf{X}_{t-p}$$

is the same as the prediction of $\mathbf{X}_t$ with its p-length long past that extends to the infinite past prediction, see the next lesson. The multidimensional Yule-Walker equations also work in this situation. It is also proved that the stable VARMA processes are dense among the regular time series.