Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

Multivariate Wold Decompositions

Simone Cerreia-Vioglio, Fulvio Ortu,
Federico Severino, Claudio Tebaldi

Working Paper n. 606

This Version: September, 2017

IGIER – Università Bocconi, Via Guglielmo Röntgen 1, 20136 Milano –Italy
http://www.igier.unibocconi.it

Multivariate Wold Decompositions

Simone Cerreia-Vioglio Fulvio Ortu Federico Severino
Claudio Tebaldi*

Università Bocconi & IGIER

Preliminary version
September 2017

Abstract

The Wold decomposition of a weakly stationary time series extends to the multivariate case by allowing each entry of a weakly stationary vectorial process to linearly depend on the components of a vector of shocks. Since univariate coefficients are replaced by matrices, we propose a modelling approach based on Hilbert A-modules defined over the algebra of squared matrices. The Abstract Wold Theorem for Hilbert A-modules, that we prove, delivers two orthogonal decompositions of vectorial processes: the Multivariate Classical Wold Decomposition, which exploits the lag operator as isometry, and the Multivariate Extended Wold Decomposition, where a scaling operator is employed. The latter enables us to disentangle the heterogeneous levels of persistence of a weakly stationary vectorial process. Hence, the persistent components of the macro-financial variables into consideration are related to the overlapping of different sources of randomness with specific persistence. We finally provide a simple application to VAR models.

1 Introduction

A vectorial process $\mathbf{x} = \{x_t\}_{t\in\mathbb{Z}}$ is a collection of m univariate time series $x_{i,t}$. Although weakly stationary univariate processes generally depend on a unique source of innovations,[1] each variable $x_{t,i}$ of a weakly stationary multivariate process is possibly affected by m kinds of shocks $\varepsilon_{1,t}, \varepsilon_{2,t}, \dots, \varepsilon_{m,t}$. This peculiarity is captured by the

*We thank Giorgio Primiceri for valuable insights. Any errors or omissions are the sole responsibility of the authors.

[1]This is ensured by the Classical Wold Decomposition for weakly stationary time series.

use of matrix coefficients, which has been extremely fruitful for VAR processes in the macroeconomic and financial literature.[2] A big econometric issue is the shocks identification due to the large number of parameters involved, which makes necessary the imposition of several restrictions to make VAR models empirically tractable.[3]

The main purpose of this work is to study persistence in multivariate economic time series. Beyond the long-run risk literature (which focuses on the asymptotic properties of processes), this topic is usually addressed by spectral analysis techniques, developed in the frequency domain.[4] In this paper, we describe a methodology to disentangle uncorrelated persistent component from a weakly stationary vectorial process entirely in the time domain. Each vectorial component explains a specific layer of persistence and it is sensible to a family of shocks with determined half-life. In order to achieve this goal, we first renew the standard treatment of multivariate time series by using *Hilbert A-modules*. The Abstract Wold Theorem for Hilbert A-modules, indeed, allows us to easily retrieve the Multivariate Classical Wold Decomposition (MCWD henceforth) and to derive, in turns, the *Multivariate Extended Wold Decomposition* (MEWD), which is persistence-based.

The standard approach to multivariate time series modelling considers matrix coefficients as a mere collection of sensitivities of the variables $x_{i,t}$ with respect to each shock $\varepsilon_{j,t}$. Indeed, such matrices are not supposed to embody the projection meaning which is, actually, the distinctive feature of ordinary least squares in univariate modelling. Hence, we propose a new way to generalize one-dimensional time series to multidimensional ones, while keeping this meaning. Specifically, we replace the vector space $\mathbb{R}$ of the coefficients of univariate time series with the algebra A of $m \times m$ matrices. Accordingly, we substitute the vector space of square-integrable variables x_t with the A-module H, in which matrices play the role of coefficients. Finally, we endow H with an inner product, with values in A, which generalizes the inner product in L^2. Such a structure is a Hilbert A-module (see the recent Cerreia-Vioglio, Maccheroni and Marinacci [10]).[5]

In Hilbert A-modules, orthogonality and projections on closed submodules are defined. These notions allow us to describe two orthogonal decompositions of H. First, we provide a brief summary of the MCWD, then we focus on the MEWD, which dis-

[2]A comprehensive treatment of VAR processes is contained in Lütkepohl [19]. Financial applications can be found in Barberis [3] and Campbell and Viceira [9].

[3]Sims [23] provides a deep discussion about this topic. Relevant applications to monetary and fiscal policy can be found in Bernanke and Mihov [4] and Blanchard and Perotti [8] respectively.

[4]Cross-spectrum and squared coherency are, indeed, used to quantify the linear association between single time series in a vectorial process. See, for example, Brockwell and Davis [6].

[5]The application of Hilbert module in the economic theory is not a novelty. For instance, a pioneering use of Hilbert modules goes back to Hansen and Richard [16], who exploited this structure to formalize the effect of conditional information in intertemporal asset pricing models.

entangles heterogeneous layers of persistence from a vectorial process. The latter is a generalization of the univariate Extended Wold Decomposition of Ortu, Severino, Tamoni and Tebaldi [20]. The instrument to derive both the decompositions is the Abstract Wold Theorem for self-dual Hilbert A-modules, that we state and prove.[6]

The next subsection introduces the Hilbert A-module framework in which we embed multivariate weakly stationary processes and it provides a quick overview of the main results. Section 2 revisits the MCWD and provides a proof by employing the Abstract Wold Theorem for Hilbert A-modules. Section 3 states and proves the MEWD for weakly stationary vectorial processes. We describe some applications of the latter in Section 4. In particular we analyse Blanchard and Quah [7] model about demand and supply influence on GNP and unemployment from the perspective of persistence. Appendix A contains the main definitions and results about Hilbert A-modules (in particular, the Abstract Wold Theorem), while Appendices C and D include all the proofs.

1.1 Summary of main results

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we consider the vector space $L^2(\mathbb{R}^m, \Omega, \mathcal{F}, \mathbb{P})$ of measurable square-integrable random vectors x that take value in $\mathbb{R}^m$.[7] We build on $L^2(\mathbb{R}^m, \Omega, \mathcal{F}, \mathbb{P})$ the structure of Hilbert A-module and we denote it by H.[8] In particular, we consider the algebra $A = \mathbb{R}^{m\times m}$ of real $m \times m$ matrices. The outer product $A \times H \longrightarrow H$ is the standard matrix-by-vector product. This operation makes H an A-module.[9] Then, we define the A-valued inner product $\langle\ ,\ \rangle_H : H \times H \longrightarrow A$ that associates any $x = [x_1, \dots, x_m]'$, $y = [y_1, \dots, y_m]' \in H$ with the matrix

$$\langle x, y\rangle_H = \mathbb{E}\left[xy'\right] = \left\{\mathbb{E}\left[x_i y_j\right]\right\}_{i,j=1,\dots,m}.$$

$\langle\ ,\ \rangle_H$ satisfies the usual properties of inner products. In addition,

$$\langle x, x\rangle_H = \mathbb{E}\left[xx'\right] = \left\{\mathbb{E}\left[x_i x_j\right]\right\}_{i,j=1,\dots,m}$$

is the covariance matrix of x, which is symmetric and positive semidefinite.

It is useful to define the *trace* functional $\bar{\varphi} : A \longrightarrow \mathbb{R}$ by setting, for any matrix a,

$$\bar{\varphi}(a) = \mathrm{Tr}(a) = \sum_{i=1}^{m} a_{i,i}.$$

[6]See Theorem 7 in Appendix A.

[7]For any $i = 1, \dots, m$ the random variable x_i belongs to $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

[8]All details are collected in Appendix A.

[9]Note that the natural structure of real vector space of H is kept because of the relation

$$\lambda x = (\lambda I)x \qquad \forall x \in H, \quad \lambda \in \mathbb{R},$$

where I is the identity matrix.

Indeed, H is a Hilbert space with the inner product $\langle \ , \ \rangle_{\bar{\varphi}} : H \times H \longrightarrow \mathbb{R}$ defined by

$$\langle x, y \rangle_{\bar{\varphi}} = \bar{\varphi} \left(\langle x, y \rangle_H \right) = \mathrm{Tr} \left(\mathbb{E} \left[x y' \right] \right) = \sum_{i=1}^{m} \mathbb{E} \left[x_i y_i \right] \qquad \forall x, y \in H$$

since $\langle \ , \ \rangle_{\bar{\varphi}}$ coincides with the usual inner product of $L^2 \left(\mathbb{R}^m \right)$. The associated norm is $\| \ \|_{\bar{\varphi}} : H \longrightarrow [0, +\infty)$ such that

$$\|x\|_{\bar{\varphi}} = \sqrt{\langle x, x \rangle_{\bar{\varphi}}} = \sqrt{\mathrm{Tr} \left(\mathbb{E} \left[x x' \right] \right)} = \sqrt{\sum_{i=1}^{m} \mathbb{E} \left[x_i^2 \right]} \qquad \forall x \in H.$$

As A is finite dimensional, the norm $\| \ \|_H : H \longrightarrow [0, +\infty)$ defined by

$$\|x\|_H = \sqrt{\| \langle x, x \rangle_H \|_A} = \sqrt{\| \mathbb{E} \left[x x' \right] \|_A} \qquad \forall x \in H$$

is equivalent to $\| \ \|_{\bar{\varphi}}$.[10] In particular, $\|x\|_H = \sqrt{\lambda_{max}}$, where λ_{max} is the largest eigenvalue of the covariance matrix of x, i.e. the one associated with the Principal Component of $\mathbb{E} \left[x x' \right]$ that explains the most variance.[11] Proposition 13 in Appendix B shows that H is a Hilbert A-module, i.e. it is complete. Since A is finite dimensional, it follows that H is self-dual, as proved by Theorem 6 in Appendix A.

Now consider a multivariate process $\mathbf{x} = \{x_t\}_{t \in \mathbb{Z}}$ such that $x_t = [x_{1,t}, \ldots, x_{m,t}]' \in H$ for all $t \in \mathbb{Z}$. Assume that $\mathbf{x}$ is weakly stationary and, without loss of generality, that it has zero mean. The *autocovariance function* $\Gamma : \mathbb{Z} \longrightarrow A$ associates any integer n with the matrix $\Gamma_n = \left[\gamma_{i,j}(n) \right]_{i,j=1,\ldots,m}$ with

$$\gamma_{i,j}(n) = \mathrm{Cov} \left(x_{i,t}, x_{j,t+n} \right) = \mathbb{E} \left[x_{i,t} x_{j,t+n} \right].$$

If $\Gamma_n = \mathbf{0}$ for any $n \neq 0$, we are facing a *multivariate white noise*, which displays *unit variance* when Γ_0 is the identity matrix. In this case, the single time series of the multivariate white noise are uncorrelated. In general, the covariance matrix Γ_0 of $\mathbf{x}$ is symmetric and positive semidefinite. We will also suppose that Γ_0 is positive definite (hence it has a positive definite square root[12]), a requirement that parallels the *regularity* assumption in the univariate case.[13]

[10] Here $\| \|_A$ denotes the operator norm of matrices. As for the equivalence, see Proposition 6 in Appendix A.

[11] In case any x_i is uncorrelated with any x_j with $i \neq j$, the covariance matrix $\mathbb{E} \left[x x' \right]$ is diagonal and so $\|x\|_H = \sqrt{\max_{i=1,\ldots,m} \mathbb{E} \left[x_i^2 \right]}$.

[12] To avoid irrelevant complications in the theory sections, we will use square root matrices for factorizing covariance matrices. Anyway, our results do not depend on the way the covariance matrix is factorized. For instance, the Cholesky decomposition can be employed too, without affecting the conclusions.

[13] See Bierens [5].

Wold-type decompositions of one-dimensional processes follow from the Abstract Wold Theorem, a functional analytical result that allows to orthogonally decompose Hilbert spaces by using isometric operators.[14] Indeed, this theorem applies to the Hilbert space generated by the past realizations of a weakly stationary univariate time series $\mathbf{x} = \{x_t\}_t$ so that any x_t turns out to be the sum of uncorrelated variables (the so-called innovations). For example, the Classical Wold Decomposition[15] obtains when the isometry is the *lag operator*. On the other hand, other choices for the isometry are possible. For instance, Ortu, Severino, Tamoni and Tebaldi [20] derive a persistence-based decomposition (named Extended Wold Decomposition) by exploiting the *scaling operator*.

In order to address the decomposition of multidimensional processes, we provide a generalization of the Abstract Wold Theorem for self-dual Hilbert A-modules. Orthogonality in Hilbert A-modules mimics the same definition in Hilbert spaces, provided that the inner product $\langle\ ,\ \rangle_H$ is employed.[16] Hence, two elements $x, y \in H$ are orthogonal when any x_i is uncorrelated with any y_j for all $i, j = 1, \dots, m$. Similarly to the Hilbert space case, the theorem requires a Hilbert A-module H and an isometry[17] $T : H \to H$ and it delivers the orthogonal decomposition $H = \hat{H} \oplus \tilde{H}$, where

$$\hat{H} = \bigcap_{n=0}^{\infty} T^n (H), \qquad \tilde{H} = \bigoplus_{n=0}^{\infty} T^n (L).$$

$L = T(H)^{\perp}$, namely the orthogonal complement of $T(H)$, is called *wandering submodule* and it is uniquely determined by T. The submodule $\tilde{H}$ contains the orthogonal innovations obtained by iteratively applying the isometry T to L, while $\hat{H}$ is an invariant submodule.[18]

The MCWD obtains when we consider the Hilbert submodule $\mathcal{H}_t(\mathbf{x})$ of H spanned by the vectorial sequence $\{x_{t-n}\}_{n\in\mathbb{N}_0}$, i.e.

$$\mathcal{H}_t(\mathbf{x}) = \mathrm{cl}\left\{ \sum_{k=0}^{+\infty} a_k x_{t-k} : \quad a_k \in A, \quad \sum_{k=0}^{+\infty}\sum_{h=0}^{+\infty} \mathrm{Tr}\left(a_k \Gamma_{k-h} a_h'\right) < +\infty \right\},$$ [19]

and the *lag operator* $\mathbf{L}$, that maps any generator $\sum_{k=0}^{\infty} a_k x_{t-k}$ of $\mathcal{H}_t(\mathbf{x})$ into $\sum_{k=0}^{\infty} a_k x_{t-1-k}$. The submodule $\hat{H}$ delivers the purely deterministic term in the decomposition, while

[14] See Sz.-Nagy, Foias, Bercovici and Kérchy [24] as a reference.

[15] See, for instance, Brockwell and Davis [6] or the original work of Wold [26].

[16] Any $x, y \in H$ are orthogonal when $\langle x, y\rangle_H$ is the null matrix, i.e. $\mathbb{E}[x_i y_j] = 0$ for any $i, j = 1, \dots, m$.

[17] The operator $T : H \to H$ is an isometry in case it is A-linear and $\langle T(x), T(y)\rangle_H = \langle x, y\rangle_H$ for all $x, y \in H$.

[18] See Theorem 7 in Appendix A for the precise statement and the proof of the Abstract Wold Theorem for self-dual Hilbert A-modules.

[19] See Appendix B for details.

the wandering submodule is spanned by the vector $x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t$.[20] The normalization of such vector produces the multivariate classical Wold innovation (or *fundamental innovation*) ε_t.

The MEWD, instead, comes from the application of the Abstract Wold Theorem to a different submodule of H. Indeed, we consider the A-module $\mathcal{H}_t(\boldsymbol{\varepsilon})$ generated by the sequence of fundamental innovations $\{\varepsilon_{t-n}\}_{n\in\mathbb{N}_0}$, namely

$$\mathcal{H}_t(\boldsymbol{\varepsilon}) = \left\{ \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k} : \quad a_k \in A, \quad \sum_{k=0}^{+\infty} \operatorname{Tr}\left(a_k a_k'\right) < +\infty \right\}.$$

As isometry, we employ the *scaling operator* $\mathbf{R} : \mathcal{H}_t(\boldsymbol{\varepsilon}) \to \mathcal{H}_t(\boldsymbol{\varepsilon})$ such that

$$\mathbf{R} : \qquad \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k} \quad \longmapsto \quad \sum_{k=0}^{+\infty} \frac{a_k}{\sqrt{2}} \left(\varepsilon_{t-2k} + \varepsilon_{t-2k-1}\right).$$

The wandering submodule associated to $\mathbf{R}$ is spanned by the multivariate *details at scale* 1, namely $\varepsilon^{(1)}_{t-k2^j} = \left(\varepsilon_{t-2k} + \varepsilon_{t-2k-1}\right)/\sqrt{2}$, with $k \in \mathbb{N}_0$. Accordingly, the submodules $\mathbf{R}^j(L)$ are generated by the *details at scale* j

$$\varepsilon^{(j)}_{t-k2^j} = \frac{1}{\sqrt{2^j}} \left(\sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-k2^j-i} - \sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-k2^j-2^{j-1}-i} \right), \qquad k \in \mathbb{N}_0.$$

Each vector of shocks $\varepsilon^{(j)}_{t-k2^j}$ has half-life in the interval $[2^{j-1}, 2^j)$ and so its degree of persistence rises with the scale j. Moreover, the invariant submodule is null. As a result, any $x_t \in \mathcal{H}_t(\boldsymbol{\varepsilon})$ decomposes into an infinite sum of *(multivariate) persistent components* $g^{(j)}_t$ associated with different scales:

$$x_t = \sum_{j=1}^{+\infty} g^{(j)}_t, \qquad\qquad g^{(j)}_t = \sum_{k=0}^{+\infty} \beta^{(j)}_k \varepsilon_{t-k2^j} \quad \forall j \in \mathbb{N}.$$

A fundamental outcome of the Abstract Wold Theorem for Hilbert A-modules is that the components $g^{(j)}_t$ are orthogonal and so any spurious correlation within layers of persistence is ruled out. The matrices $\beta^{(j)}_k$, that we call *(multivariate) multiscale impulse responses* are, then, precisely associated with the scale j and the time shift $k2^j$. Moreover, each entry (p, q) of $\beta^{(j)}_k$ quantifies the sensitivity of the variable $x_{p,t}$ with respect to the q-th source of randomness in the vector of shocks, at the specific level of persistence j and time lag $k2^j$.

Different scales may capture diverse reactions with respect to shocks with specific persistence, that are not recognizable in the classical impulse responses. This is the case, for instance, of Blanchard and Quah bivariate model of GNP and unemployment, that we inspect in Section 4.2.

[20] $\mathcal{P}_M x_t$ denotes the orthogonal projection of x_t on the closed submodule M.

2 The Multivariate Classical Wold Decomposition

The Multivariate Wold Decomposition Theorem allows to decompose a zero-mean regular weakly stationary vectorial process $\mathbf{x}$ into the infinite sum of uncorrelated multivariate innovations that occur at different times. Although a proof of this result can be found in Rozanov [22], we show how to derive this decomposition by applying the Abstract Wold Theorem for self-dual Hilbert A-modules.

Given a zero-mean weakly stationary vectorial process $\mathbf{x} = \{x_t\}_{t\in\mathbb{Z}}$, we consider the Hilbert submodule of H spanned by the vectors x_{t-n} with $n \in \mathbb{N}_0$, namely

$$\mathcal{H}_t(\mathbf{x}) = \mathrm{cl}\left\{\sum_{k=0}^{+\infty} a_k x_{t-k} : \quad a_k \in A, \quad \sum_{k=0}^{+\infty}\sum_{h=0}^{+\infty} \mathrm{Tr}\left(a_k \Gamma_{k-h} a_h'\right) < +\infty\right\},$$ [21]

Definition 1 *We call* lag operator *the operator* $\mathbf{L} : \mathcal{H}_t(\mathbf{x}) \longrightarrow \mathcal{H}_t(\mathbf{x})$ *that acts on generators of* $\mathcal{H}_t(\mathbf{x})$ *as*

$$\mathbf{L}: \qquad \sum_{k=0}^{+\infty} a_k x_{t-k} \quad \longmapsto \quad \sum_{k=0}^{+\infty} a_k x_{t-1-k}.$$

$\mathbf{L}$ is A-linear and bounded, hence it can be extended to $\mathcal{H}_t(\mathbf{x})$ with continuity.[22] Moreover, $\mathbf{L}$ is isometric on $\mathcal{H}_t(\mathbf{x})$.[23]

In order to apply the Abstract Wold Theorem for Hilbert A-modules, we have to determine the images of $\mathcal{H}_t(\mathbf{x})$ through the powers of the operator $\mathbf{L}$, and the wandering submodule. Recall that, in a self-dual Hilbert A-module, the image of a closed submodule through an isometry is a closed submodule, too.[24] By exploiting this fact, we find that $\quad \mathbf{L}^j \mathcal{H}_t(\mathbf{x}) = \mathcal{H}_{t-j}(\mathbf{x}) \quad$ for any $j \in \mathbb{N}$.[25]

Then, we show that the A-module $\mathcal{H}_t(\mathbf{x})$ can be decomposed into the direct sum[26]

$$\mathcal{H}_t(\mathbf{x}) = \mathcal{H}_{t-1}(\mathbf{x}) \oplus \mathrm{span}\left\{x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t\right\}.$$

In other words, the wandering submodule associated with the lag operator is

$$\mathcal{L}_t^{\mathbf{L}} = \mathrm{span}\left\{x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t\right\}.$$

We say that $\mathbf{x}$ is *regular* when, for any $t \in \mathbb{Z}$, $\left\langle x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t, x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t \right\rangle_H$ is a symmetric positive definite matrix. Hence, there exists a symmetric positive definite square root matrix S such that

$$\left\langle x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t, x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t \right\rangle_H = SS.$$

[21] See Appendix B for details.
[22] See Proposition 15 in Appendix C.
[23] See Proposition 16 in Appendix C.
[24] See Lemma 2 in Appendix A.
[25] See Proposition 17 in Appendix C.
[26] See Proposition 18 in Appendix C.

As S is invertible, we define the *fundamental innovation process* $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t\in\mathbb{Z}}$ by

$$\varepsilon_t = S^{-1}\left(x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t\right), \qquad t \in \mathbb{Z}.$$

$\boldsymbol{\varepsilon}$ is a unit variance white noise, that is its components are uncorrelated.[27]

Lemma 4 in Appendix C shows that the lag and the projection operator commute: for any $k, j \in \mathbb{N}_0$,

$$\mathbf{L}^j\mathcal{P}_{\mathcal{H}_{t-k-1}(\mathbf{x})}x_{t-k} = \mathcal{P}_{\mathcal{H}_{t-k-j-1}(\mathbf{x})}x_{t-k-j}.$$

This result ensures that the covariance matrix of $x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t$ is actually not dependent on the time index $t \in \mathbb{Z}$ and that, for any $j \in \mathbb{N}$,

$$\mathbf{L}^j\mathcal{L}_t^{\mathbf{L}} = \operatorname{span}\left\{x_{t-j} - \mathcal{P}_{\mathcal{H}_{t-j-1}(\mathbf{x})}x_{t-j}\right\}.$$

We are now ready to apply the Abstract Wold Theorem to the Hilbert A-module $\mathcal{H}_t(\mathbf{x})$ with the isometry $\mathbf{L}$.

Theorem 1 *The Hilbert A-module $\mathcal{H}_t(\mathbf{x})$ decomposes into the orthogonal sum*

$$\mathcal{H}_t(\mathbf{x}) = \hat{\mathcal{H}}_t(\mathbf{x}) \oplus \tilde{\mathcal{H}}_t(\mathbf{x}),$$

where

$$\hat{\mathcal{H}}_t(\mathbf{x}) = \bigcap_{j=0}^{+\infty}\mathcal{H}_{t-j}(\mathbf{x}), \qquad \tilde{\mathcal{H}}_t(\mathbf{x}) = \bigoplus_{j=0}^{+\infty}\operatorname{span}\left\{x_{t-j} - \mathcal{P}_{\mathcal{H}_{t-j-1}(\mathbf{x})}x_{t-j}\right\}.$$

Proof. See Appendix C ■

The application to zero-mean regular weakly stationary vectorial processes is now straightforward.

Theorem 2 (Multivariate Classical Wold Decomposition) *Let $\mathbf{x} = \{x_t\}_{t\in\mathbb{Z}}$ be a zero-mean regular weakly stationary m-dimensional process. Then, for any $t \in \mathbb{Z}$, x_t decomposes as*

$$x_t = \sum_{k=0}^{+\infty}\alpha_k\varepsilon_{t-k} + \nu_t,$$

where the equality is in norm and

i) $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a unit variance m-dimensional white noise;

[27] Indeed,

$$\langle\varepsilon_t, \varepsilon_t\rangle_H = \mathbb{E}\left[\varepsilon_t\varepsilon_t'\right] = S^{-1}\left\langle x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t, x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t\right\rangle_H\left(S^{-1}\right)' = S^{-1}SSS^{-1} = I.$$

ii) for any $k \in \mathbb{N}_0$, the $m \times m$ matrices α_k do not depend on t,

$$\alpha_k = \mathbb{E}\left[x_t \varepsilon'_{t-k}\right] \qquad \textit{and} \qquad \sum_{k=0}^{+\infty} \mathrm{Tr}\left(\alpha_k \alpha'_k\right) < +\infty;$$

iii) $\boldsymbol{\nu} = \{\nu_t\}_{t \in \mathbb{Z}}$ is a zero-mean weakly stationary m-dimensional process,

$$\nu_t \in \bigcap_{j=0}^{+\infty} \mathcal{H}_{t-j}(\mathbf{x}) \qquad \textit{and} \qquad \mathbb{E}\left[\nu_t \varepsilon'_{t-k}\right] = \mathbf{0} \quad \forall k \in \mathbb{N}_0;$$

iv)

$$\nu_t \in \mathrm{cl}\left\{ \sum_{h=1}^{+\infty} a_h \nu_{t-h} \quad \in \bigcap_{j=1}^{+\infty} \mathcal{H}_{t-j}(\mathbf{x}) : \quad a_h \in A \right\}.$$

Proof. See Appendix C ■

The random vector $\sum_{k=0}^{\infty} \alpha_k \varepsilon_{t-k}$ is referred to as the *non-deterministic component*, while $\boldsymbol{\nu}$ constitutes the (predictable) *deterministic component* of $\mathbf{x}$. If $\boldsymbol{\nu}$ is the null vector, we call the process $\mathbf{x}$ *purely non-deterministic*. Similarly, we say that $\mathbf{x}$ is *purely deterministic* if the non-deterministic component is zero.

The main contribution of the approach that we followed so far is that the *multivariate impulse responses* α_h are fully characterized by the projection on Hilbert submodules. This feature generalizes the OLS methodology employed in the univariate case[28] and shows that the multivariate impulse responses are not only a collection of one-dimensional impulse responses, computed entry by entry. Indeed, each projection matrix α_h minimizes the distance of the outcome x_t from the submodule generated by the vectorial innovation ε_{t-h}.

This construction naturally delivers vectors of innovations composed by univariate sources of randomness that are uncorrelated within them. Indeed, $\boldsymbol{\varepsilon}$ is a unit variance m-dimensional white noise. This property opens the door to the big issue of identifying structural univariate shocks in vectorial processes.

3 The Multivariate Extended Wold Decomposition

The aim of this section is to generalize the Extended Wold Decomposition for weakly stationary time series (see Ortu, Severino, Tamoni and Tebaldi [20]) to multidimensional processes. Differently from the univariate case, in which Hilbert space techniques are employed, we embed multivariate processes in a Hilbert A-module framework.

[28]Indeed, we are actually treating multivariate multiple regressions.

Let $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t\in\mathbb{Z}}$ be a unit variance m-dimensional white noise and consider the Hilbert submodule of H generated by the sequence of ε_{t-n} with $n \in \mathbb{N}_0$, that is

$$\mathcal{H}_t(\boldsymbol{\varepsilon}) = \left\{ \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k} : \quad a_k \in A, \quad \sum_{k=0}^{+\infty} \mathrm{Tr}\left(a_k a_k'\right) < +\infty \right\}.$$

We define the *scaling operator* $\mathbf{R} : \mathcal{H}_t(\boldsymbol{\varepsilon}) \longrightarrow \mathcal{H}_t(\boldsymbol{\varepsilon})$ as follows[29]

$$\mathbf{R} : \qquad \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k} \quad \longmapsto \quad \sum_{k=0}^{+\infty} \frac{a_k}{\sqrt{2}} \left(\varepsilon_{t-2k} + \varepsilon_{t-2k-1}\right) = \sum_{k=0}^{+\infty} \frac{a_{\lfloor \frac{k}{2} \rfloor}}{\sqrt{2}} \varepsilon_{t-k}.$$

The scaling operator is well-defined, A-linear and isometric on $\mathcal{H}_t(\boldsymbol{\varepsilon})$.[30]

We first show the orthogonal decomposition of $\mathcal{H}_t(\boldsymbol{\varepsilon})$, obtained by applying the Abstract Wold Theorem for self-dual Hilbert A-modules with $\mathbf{R}$ as isometry. Then we derive the MEWD of a vectorial time series $\mathbf{x} = \{x_t\}_{t\in\mathbb{Z}}$ with fundamental innovations given by $\boldsymbol{\varepsilon}$.

3.1 The orthogonal decomposition of $\mathcal{H}_t(\boldsymbol{\varepsilon})$ induced by $\mathbf{R}$

Before entering the details of the decomposition induced by the scaling operator, we define the multivariate *detail process at scale* 1, denoted by $\boldsymbol{\varepsilon}^{(1)} = \left\{\varepsilon_t^{(1)}\right\}_{t\in\mathbb{Z}}$, where

$$\varepsilon_t^{(1)} = \frac{\varepsilon_t - \varepsilon_{t-1}}{\sqrt{2}}, \qquad t \in \mathbb{Z}.$$

Each $\varepsilon_t^{(1)}$ has zero mean and unit variance.[31] In general, we define the *detail process at scale j* in the following way.

Definition 2 *For any $j \in \mathbb{N}$, we call* detail process at scale j *the vectorial process $\boldsymbol{\varepsilon}^{(j)} = \left\{\varepsilon_t^{(j)}\right\}_{t\in\mathbb{Z}}$ where*

$$\varepsilon_t^{(j)} = \frac{1}{\sqrt{2^j}} \left(\sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-i} - \sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-2^{j-1}-i} \right).$$

[29] $\lfloor \cdot \rfloor$ denotes the floor function, that associates any $c \in \mathbb{R}$ with the integer $\lfloor c \rfloor = \max\{n \in \mathbb{Z} : n \leqslant c\}$.
[30] See Proposition 19 in Appendix D.
[31] Indeed,

$$\begin{aligned}
\mathbb{E}\left[\varepsilon_t^{(1)} \varepsilon_t^{(1)\prime}\right] &= \frac{1}{2}\mathbb{E}\left[\left(\varepsilon_t - \varepsilon_{t-1}\right)\left(\varepsilon_t' - \varepsilon_{t-1}'\right)\right] \\
&= \frac{1}{2}\left\{\mathbb{E}\left[\varepsilon_t \varepsilon_t'\right] - \mathbb{E}\left[\varepsilon_t \varepsilon_{t-1}'\right] - \mathbb{E}\left[\varepsilon_{t-1} \varepsilon_t'\right] + \mathbb{E}\left[\varepsilon_{t-1} \varepsilon_{t-1}'\right]\right\} \\
&= \frac{1}{2}\left\{I - \mathbf{0} - \mathbf{0} + I\right\} = I.
\end{aligned}$$

At any scale j, we consider the subseries of $\boldsymbol{\varepsilon}^{(j)}$ defined on the support $S_t^{(j)} = \{t - k2^j : k \in \mathbb{Z}\}$ in order to avoid overlap among the vectors $\varepsilon_t^{(j)}$. The vectors $\varepsilon_{t-k2^j}^{(j)}$, in fact, exhibit a dual nature depending on the support on which they are considered. Indeed, the process $\boldsymbol{\varepsilon}^{(j)}$ is an $MA(2^j - 1)$ with respect to the fundamental innovations of $\mathbf{x}$. Therefore, some spurious correlation is present between the vectors $\varepsilon_{t-k2^j}^{(j)}$ and $\varepsilon_{\tau-k2^j}^{(j)}$ with $|t-\tau| \leqslant 2^j - 1$. Nevertheless, each subseries $\left\{\varepsilon_{t-k2^j}^{(j)}\right\}_{k\in\mathbb{Z}}$ is a unit variance white noise on the support $S_t^{(j)}$.[32]

Now we want to determine the invariant submodule $\hat{\mathcal{H}}_t(\boldsymbol{\varepsilon})$ that arises when the Abstract Wold Theorem is applied on $\mathcal{H}_t(\boldsymbol{\varepsilon})$ with isometry $\mathbf{R}$. The definition of the scaling operator ensures that the submodule $\mathbf{R}\mathcal{H}_t(\boldsymbol{\varepsilon})$ is made of those linear combinations of the multivariate innovations ε_t that have the (matrix) coefficients equal to each others 2-by-2, that is

$$\mathbf{R}\mathcal{H}_t(\boldsymbol{\varepsilon}) = \left\{ \sum_{k=0}^{+\infty} c_k^{(1)} \left(\varepsilon_{t-2k} + \varepsilon_{t-2k-1}\right) \in \mathcal{H}_t(\boldsymbol{\varepsilon}) : \quad c_k^{(1)} \in A \right\}.$$

The same line of reasoning shows that, for any $j \in \mathbb{N}$, the submodules $\mathbf{R}^j\mathcal{H}_t(\boldsymbol{\varepsilon})$ consist of the linear combinations of the vectors ε_t with (matrix) coefficients equal to each others 2^j-by-2^j:

$$\mathbf{R}^j\mathcal{H}_t(\boldsymbol{\varepsilon}) = \left\{ \sum_{k=0}^{+\infty} c_k^{(j)} \left(\sum_{i=0}^{2^j-1} \varepsilon_{t-k2^j-i} \right) \in \mathcal{H}_t(\boldsymbol{\varepsilon}) : \quad c_k^{(j)} \in A \right\}.$$ [33]

It follows that the intersection of all submodules $\mathbf{R}^j\mathcal{H}_t(\boldsymbol{\varepsilon})$ contains only the zero element, that is $\hat{\mathcal{H}}_t(\boldsymbol{\varepsilon})$ is the null submodule: $\hat{\mathcal{H}}_t(\boldsymbol{\varepsilon}) = \{0\}$.

We now focus on the submodule $\tilde{\mathcal{H}}_t(\boldsymbol{\varepsilon})$. The wandering submodule $\mathcal{L}_t^{\mathbf{R}}$ associated with $\mathbf{R}$ is the orthogonal complement of $\mathbf{R}\mathcal{H}_t(\boldsymbol{\varepsilon})$ in $\mathcal{H}_t(\boldsymbol{\varepsilon})$, namely $\mathbf{R}\mathcal{H}_t(\boldsymbol{\varepsilon})^{\perp}$. As $\mathbf{R}$ is linear and bounded, such submodule coincides with the kernel of its adjoint operator, therefore

$$\mathcal{L}_t^{\mathbf{R}} = \left\{ \sum_{k=0}^{+\infty} b_k^{(1)} \varepsilon_{t-2k}^{(1)} \in \mathcal{H}_t(\boldsymbol{\varepsilon}) : \quad b_k^{(1)} \in A \right\}.$$ [34]

Hence, $\mathcal{L}_t^{\mathbf{R}}$ is the submodule of $\mathcal{H}_t(\boldsymbol{\varepsilon})$ that contains all infinite moving averages driven by the detail process at scale 1 on the support $S_t^{(1)}$. More generally, for each $j \in$

[32]See Proposition 20 in Appendix D.

[33]Note that the isometric operator $\mathbf{R}^j$ acts on the elements of $\mathcal{H}_t(\boldsymbol{\varepsilon})$ as

$$\mathbf{R}^j : \qquad \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k} \quad \longmapsto \quad \sum_{k=0}^{+\infty} \frac{a_k}{\sqrt{2^j}} \left(\sum_{i=0}^{2^j-1} \varepsilon_{t-k2^j-i} \right).$$

[34]See Proposition 10 in Appendix A and Proposition 21 in Appendix D.

$\mathbb{N}$, the image of $\mathcal{L}_t^{\mathbf{R}}$ through the powers of the scaling operator $\mathbf{R}^{j-1}$ is the submodule

$$\mathbf{R}^{j-1}\mathcal{L}_t^{\mathbf{R}} = \left\{ \sum_{k=0}^{+\infty} b_k^{(j)} \varepsilon_{t-k2^j}^{(j)} \in \mathcal{H}_t(\boldsymbol{\varepsilon}) : \quad b_k^{(j)} \in A \right\}.$$[35]

In sum, the submodules $\mathbf{R}^{j-1}\mathcal{L}_t^{\mathbf{R}}$ consist of all infinite moving averages with innovations given by the detail process at scale j on the support $S_t^{(j)}$.

We have now all the instruments to state the orthogonal decomposition of $\mathcal{H}_t(\boldsymbol{\varepsilon})$ induced by the scaling operator.

Theorem 3 *The Hilbert A-module $\mathcal{H}_t(\boldsymbol{\varepsilon})$ decomposes into the orthogonal sum*

$$\mathcal{H}_t(\boldsymbol{\varepsilon}) = \bigoplus_{j=1}^{+\infty} \mathbf{R}^{j-1}\mathcal{L}_t^{\mathbf{R}},$$

where

$$\mathbf{R}^{j-1}\mathcal{L}_t^{\mathbf{R}} = \left\{ \sum_{k=0}^{+\infty} b_k^{(j)} \varepsilon_{t-k2^j}^{(j)} \in \mathcal{H}_t(\boldsymbol{\varepsilon}) : \quad b_k^{(j)} \in A \right\}.$$

Proof. See Appendix D ■

3.2 The Multivariate Extended Wold Decomposition of x_t

Given a purely non-deterministic process $\mathbf{x}$, the MCWD ensures that x_t belongs to $\mathcal{H}_t(\boldsymbol{\varepsilon})$, where ε_t is the fundamental innovation of x_t. As a result, the orthogonal decomposition of the A-module $\mathcal{H}_t(\boldsymbol{\varepsilon})$ induces a decomposition of x_t. Indeed, there exists a sequence $\left\{ g_t^{(j)} \right\}_{j \in \mathbb{N}}$ of random vectors such that

$$x_t = \sum_{j=1}^{+\infty} g_t^{(j)}, \tag{1}$$

where each $g_t^{(j)}$ is the orthogonal projection of x_t on $\mathbf{R}^{j-1}\mathcal{L}_t^{\mathbf{R}}$, in the sense of A-modules.

Definition 3 *We call* persistent component at scale j *the orthogonal projection of x_t on the submodule $\mathbf{R}^{j-1}\mathcal{L}_t^{\mathbf{R}}$ of $\mathcal{H}_t(\boldsymbol{\varepsilon})$ and we denote it by $g_t^{(j)}$.*

Of course, given t, the components $g_t^{(j)}$ are orthogonal to each others. Moreover, each $g_t^{(j)}$ belongs to $\mathbf{R}^{j-1}\mathcal{L}_t^{\mathbf{R}}$ and so

$$g_t^{(j)} = \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)}$$

[35]See Proposition 22 in Appendix D.

where the $m \times m$ matrices $\beta_k^{(j)}$ satisfy

$$\left\| \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)} \right\|_{\tilde{\varphi}}^2 = \sum_{k=0}^{+\infty} \mathrm{Tr}\left(\beta_k^{(j)} \beta_k^{(j)\prime}\right) < +\infty.$$

Each $\beta_k^{(j)}$ is the matrix obtained by projecting x_t on the submodule generated by the detail $\varepsilon_{t-k2^j}^{(j)}$, that is

$$\beta_k^{(j)} = \langle x_t, \varepsilon_{t-k2^j}^{(j)} \rangle_H = \mathbb{E}\left[x_t \varepsilon_{t-k2^j}^{(j)\,\prime}\right].$$

By writing the explicit expression of $g_t^{(j)}$ into (1), we obtain the *Multivariate Extended Wold Decomposition of x_t.*

Definition 4 *We call* Multivariate Extended Wold Decomposition of x_t *the decomposition*

$$x_t = \sum_{j=1}^{+\infty} \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)}.$$

Moreover, we call the matrix $\beta_k^{(j)}$ the (multivariate) multiscale impulse response function *associated to the innovation at scale j and time translation $k2^j$.*

Since the details at different scales can be expressed in terms of the fundamental innovations ε_t, the MEWD and the MCWD exploit the same structure of shocks. Hence, we can retrieve the matrices $\beta_k^{(j)}$ from the matrices α_h of the MCWD. Furthermore, the matrices $\beta_k^{(j)}$ are independent of the time index t.

Proposition 1 *For any $j \in \mathbb{N}$, $k \in \mathbb{N}_0$,*

$$\beta_k^{(j)} = \frac{1}{\sqrt{2^j}} \left(\sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+2^{j-1}+i} \right),$$

hence $\beta_k^{(j)}$ does not depend on t. In addition, $\lim_{k\to+\infty} \beta_k^{(j)} = \mathbf{0}$ *for any $j \in \mathbb{N}$.*

Proof. See Appendix D ■

Note that an orthogonal decomposition of $\mathcal{H}_t(\boldsymbol{\varepsilon})$ into a finite number of submodules is also possible. Indeed, $\mathcal{H}_t(\boldsymbol{\varepsilon}) = \mathbf{R}\mathcal{H}_t(\boldsymbol{\varepsilon}) \oplus \mathcal{L}_t^{\mathbf{R}}$ and, by iteratively applying the scaling operator, we find:

$$\mathcal{H}_t(\boldsymbol{\varepsilon}) = \mathbf{R}^J \mathcal{H}_t(\boldsymbol{\varepsilon}) \oplus \bigoplus_{j=1}^{J} \mathbf{R}^{j-1} \mathcal{L}_t^{\mathbf{R}}.$$

We call *residual component* at scale j the orthogonal projection of x_t on the submodule $\mathbf{R}^j \mathcal{H}_t(\boldsymbol{\varepsilon})$ and we denote it by $\pi_t^{(j)}$. This random vector has the following expression[36]

$$\pi_t^{(j)} = \sum_{k=0}^{+\infty} \gamma_k^{(j)} \left(\sum_{i=0}^{2^j-1} \varepsilon_{t-k2^j-i} \right),$$

[36]See the proof of Proposition 1.

where each matrix $\gamma_k^{(j)}$ satisfies

$$\gamma_k^{(j)} = \frac{1}{2^j}\left(\sum_{i=0}^{2^j-1} \alpha_{k2^j+i}\right).$$

As a result, a MEWD of x_t holds both in the finite case, i.e. when a maximum scale J is chosen, and in the infinite one:

$$x_t = \pi_t^{(J)} + \sum_{j=1}^{J} g_t^{(j)} \qquad \text{or} \qquad x_t = \sum_{j=1}^{+\infty} g_t^{(j)}.$$

Specifically, the residual components $\pi_t^{(j)}$ and the persistent components $g_t^{(j)}$ have the following expressions:

$$\pi_t^{(j)} = \sum_{k=0}^{+\infty} \frac{1}{2^j}\left(\sum_{i=0}^{2^j-1} \alpha_{k2^j+i}\right)\left(\sum_{i=0}^{2^j-1} \varepsilon_{t-k2^j-i}\right),$$

$$g_t^{(j)} = \sum_{k=0}^{+\infty} \frac{1}{\sqrt{2^j}}\left(\sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+2^{j-1}+i}\right)\varepsilon_{t-k2^j}^{(j)}.$$

Theorem 4 (Multivariate Extended Wold Decomposition) *Let* $\mathbf{x}$ *be a zero-mean, weakly stationary purely non-deterministic m-dimensional process. Then x_t decomposes as*

$$x_t = \sum_{j=1}^{+\infty}\sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)},$$

where the equality is in norm and

i) for any fixed $j \in \mathbb{N}$, the m-dimensional process $\boldsymbol{\varepsilon}^{(j)} = \left\{\varepsilon_t^{(j)}\right\}_{t\in\mathbb{Z}}$ is an $MA(2^j-1)$ with respect to the classical Wold innovations of $\mathbf{x}$*:*

$$\varepsilon_t^{(j)} = \frac{1}{\sqrt{2^j}}\left(\sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-i} - \sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-2^{j-1}-i}\right)$$

and $\left\{\varepsilon_{t-k2^j}^{(j)}\right\}_{k\in\mathbb{Z}}$ is a unit variance white noise;

ii) for any $j \in \mathbb{N}$, $k \in \mathbb{N}_0$, the $m \times m$ matrices $\beta_k^{(j)}$ are unique and they satisfy

$$\beta_k^{(j)} = \frac{1}{\sqrt{2^j}}\left(\sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+2^{j-1}+i}\right),$$

hence they do not depend on t and $\sum_{k=0}^{\infty} \mathrm{Tr}\left(\beta_k^{(j)}\beta_k^{(j)\prime}\right) < +\infty$ for any $j \in \mathbb{N}$;

iii) *letting*

$$g_t^{(j)} = \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)},$$

then, for any $j, l \in \mathbb{N}, p, q, t \in \mathbb{Z}$, $\quad \mathbb{E}\left[g_{t-p}^{(j)} g_{t-q}^{(l)\,\prime}\right]$ *depends at most on* $j, l, p-q$. *Moreover,*

$$\mathbb{E}\left[g_{t-m2^j}^{(j)} g_{t-n2^l}^{(l)\,\prime}\right] = \mathbf{0} \qquad \forall j \neq l, \quad \forall m, n \in \mathbb{N}_0, \quad \forall t \in \mathbb{Z}.$$

Proof. See Appendix D ■

According to *iii)*, when t is fixed, the orthogonality among persistent components involves all the shifted vectors $g_{t-m2^j}^{(j)}$ and $g_{t-n2^l}^{(l)}$, for any $m, n \in \mathbb{Z}$, with time translation proportional to 2^j and 2^l respectively. What we can say in general is that the covariance matrix between $g_{t-p}^{(j)}$ and $g_{t-q}^{(l)}$ depends at most on the scales j, l and on the difference $p - q$.

By the MEWD we decompose a zero-mean, purely non-deterministic vectorial process into the sum of orthogonal components $g_t^{(j)}$ associated with the level of persistence j. Each vector $g_t^{(j)}$ has innovations on a grid $S_t^{(j)} = \{t - k2^j : \quad k \in \mathbb{Z}\}$ with time interval between two indices proportional to 2^j. When the scale j increases, the support $S_t^{(j)}$ becomes sparser and the degree of persistence of innovations rises. In case a multivariate multiscale impulse response $\beta_k^{(j)}$ is significantly different from the null matrix, with high j, we are facing a low-frequency component, that affects the process in the long run.

Although the innovations of the components have support $S_t^{(j)}$, the variables $g_t^{(j)}$ are defined for every $t \in \mathbb{Z}$. In particular, given two different time indices t and τ with $|t - \tau| \leqslant 2^j - 1$, the innovations of $g_t^{(j)}$ and $g_\tau^{(j)}$ belong to the different grids $S_t^{(j)}$ and $S_\tau^{(j)}$. Notwithstanding, $g_t^{(j)}$ and $g_\tau^{(j)}$ share the same matrix coefficients $\beta_k^{(j)}$, hence we are handling 2^j versions of the same process (in norm). According to the time index t we choose, we pick up one of these versions, namely the one with support $S_t^{(j)}$. Such a structure stems from the weak stationarity of the process $\mathbf{x}$.

Note that the MEWD properly generalizes the one-dimensional Extended Wold Decomposition of Ortu, Severino, Tamoni and Tebaldi [20]. Indeed, in case the matrix coefficients α_h are diagonal, any entry $x_{i,t}$ depends only on the innovations $\varepsilon_{i,t}$ and it satisfies the univariate Classical Wold Decomposition

$$x_{i,t} = \sum_{h=0}^{+\infty} \alpha_h(i,i) \varepsilon_{i,t-h}.$$

Accordingly, the multiscale impulse responses $\beta_k^{(j)}$ are diagonal matrices too and the MEWD delivers

$$x_{i,t} = \sum_{j=1}^{+\infty} \sum_{k=0}^{+\infty} \beta_k^{(j)}(i,i) \varepsilon_{i,t-k2^j}^{(j)},$$

where

$$\beta_k^{(j)}(i,i) = \frac{1}{\sqrt{2^j}} \left(\sum_{p=0}^{2^{j-1}-1} \alpha_{k2^j+p}(i,i) - \sum_{p=0}^{2^{j-1}-1} \alpha_{k2^j+2^{j-1}+p}(i,i) \right),$$

as prescribed by the univariate Extended Wold Decomposition.

A justification of the fact that the iterated application of $\mathbf{R}$ increases persistence is due to spectral analysis consideration and it is developed in detail in Ortu, Severino, Tamoni and Tebaldi [20] for the univariate case.

Finally, we built the MEWD of x_t as a refinement of the MCWD, where $\boldsymbol{\varepsilon}$ is precisely the process of fundamental innovations of $\mathbf{x}$. Nonetheless, such persistence-based decomposition holds also in case $\boldsymbol{\varepsilon}$ is any unit variance white noise that allows a moving average representation of x_t. In addition, in case $\boldsymbol{\varepsilon}$ has a positive definite covariance matrix Σ, then $\Sigma = SS$ for some symmetric positive definite $S \in A$. Then, $\eta_t = S^{-1}\varepsilon_t$ defines a unit variance white noise and the MCWD and the MEWD become respectively,

$$x_t = \sum_{h=0}^{+\infty} \widetilde{\alpha}_h \eta_{t-h}, \qquad\qquad x_t = \sum_{j=1}^{+\infty}\sum_{k=0}^{+\infty} \widetilde{\beta}_k^{(j)} \eta_{t-h},$$

where $\widetilde{\alpha}_h = \alpha_h S$ and $\widetilde{\beta}_k^{(j)} = \beta_k^{(j)} S$.

We now address the MEWD from the converse point of view. Suppose that the dynamics at all time scales are given. We are interested in rebuilding the vectorial process $\mathbf{x} = \{x_t\}_{t\in\mathbb{Z}}$ obtained by summing up such components. In order to make the sum feasible, we assume a common innovation process $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t\in\mathbb{Z}}$. This allows us to define at each scale $j \in \mathbb{N}$ the detail process $\boldsymbol{\varepsilon}^{(\mathbf{j})} = \left\{\varepsilon_t^{(j)}\right\}_{t\in\mathbb{Z}}$ as

$$\varepsilon_t^{(j)} = \frac{1}{\sqrt{2^j}} \left(\sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-i} - \sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-2^{j-1}-i} \right).$$

Then, at any scale j we consider the processes $\mathbf{g}^{(\mathbf{j})} = \left\{g_t^{(j)}\right\}_{t\in\mathbb{Z}}$ defined by

$$g_t^{(j)} = \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)},$$

where $\beta_k^{(j)}$ are matrices in A. Although each $\mathbf{g}^{(\mathbf{j})}$ is a moving average with respect to the innovations $\left\{\varepsilon_{t-k2^j}^{(j)}\right\}_{k\in\mathbb{Z}}$ the variables $g_t^{(j)}$ are defined for every $t \in \mathbb{Z}$. The process $\mathbf{x}$ obtained by the summation of all $g_t^{(j)}$ has the following properties.

Theorem 5 *Let $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t\in\mathbb{Z}}$ be a unit variance m-dimensional white noise process. For any $j \in \mathbb{N}$, define the detail process $\boldsymbol{\varepsilon}^{(\mathbf{j})} = \left\{\varepsilon_t^{(j)}\right\}_{t\in\mathbb{Z}}$ as*

$$\varepsilon_t^{(j)} = \frac{1}{\sqrt{2^j}} \left(\sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-i} - \sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-2^{j-1}-i} \right)$$

and consider a vectorial process $\mathbf{g}^{(\mathbf{j})} = \left\{g_t^{(j)}\right\}_{t\in\mathbb{Z}}$ such that

$$g_t^{(j)} = \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)}, \qquad \sum_{j=1}^{+\infty}\sum_{k=0}^{+\infty} \mathrm{Tr}\left(\beta_k^{(j)} \beta_k^{(j)\prime}\right) < +\infty.$$

Then, the vectorial process $\mathbf{x} = \{x_t\}_{t\in\mathbb{Z}}$ defined by

$$x_t = \sum_{j=1}^{+\infty} g_t^{(j)}$$

is zero-mean, weakly stationary purely non-deterministic and

$$x_t = \sum_{h=0}^{+\infty} \alpha_h \varepsilon_{t-h},$$

where, for any $h \in \mathbb{N}_0$,

$$\alpha_h = \sum_{j=1}^{+\infty} \frac{1}{\sqrt{2^j}} \beta^{(j)}_{\left\lfloor \frac{h}{2^j} \right\rfloor} \chi^{(j)}(h)$$

and

$$\chi^{(j)}(h) = \begin{cases} -1 & \text{if} \quad 2^j \left\lfloor \frac{h}{2^j} \right\rfloor \in \left\{h - 2^j + 1, \ldots, h - 2^{j-1}\right\}, \\ 1 & \text{if} \quad 2^j \left\lfloor \frac{h}{2^j} \right\rfloor \in \left\{h - 2^{j-1} + 1, \ldots, h\right\}. \end{cases}$$

Theorem 5 provides the moving average representation of the aggregated process $\mathbf{x}$ with respect to the underlying innovations $\boldsymbol{\varepsilon}$. If, in addition, the shocks ε_t coincide with the classical Wold innovations of $\mathbf{x}$, we exactly retrieve the MCWD of x_t form its MEWD.

4 Applications

To put the MEWD into practice we first compute the multiscale impulse responses of weakly stationary $VAR(1)$ and $VARMA(1,1)$ processes. Then, in Section 4.2, we analyse the persistent dynamics of Blanchard and Quah [7] bivariate process.

4.1 The MEWD of $VAR(1)$ and $VARMA(1,1)$ processes

Consider a weakly stationary purely non-deterministic vectorial $ARMA(1,1)$ process, or simply $VARMA(1,1)$, $\mathbf{x} = \{x_t\}_{t\in\mathbb{Z}}$ defined by

$$x_t = \rho x_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1},$$

where $\rho, \theta \in A$, $\rho + \theta \neq \mathbf{0}$ and $\boldsymbol{\varepsilon} = \{\varepsilon\}_{t\in\mathbb{Z}}$ is a multivariate unit variance white noise. The stationarity condition that we assume is $\|\rho\|_A < 1$.[37] Later, by setting $\theta = 0$ we will retrieve a $VAR(1)$ process as a special case.

By using the lag operator $\mathbf{L}$, we can rewrite the previous equation as

$$(I - \rho\mathbf{L})x_t = (I + \theta\mathbf{L})\varepsilon_t.$$

Since $\|\rho\|_A < 1$, the operator $\sum_{l=0}^{\infty} (\rho\mathbf{L})^l$ is well-defined.[38] Moreover,

$$(I - \rho\mathbf{L}) \sum_{l=0}^{+\infty} (\rho\mathbf{L})^l = I$$

and so the operator $(I - \rho\mathbf{L})$ is invertible with $\sum_{l=0}^{\infty} (\rho\mathbf{L})^l$ as inverse. This enables us to determine the moving average representation of x_t. In fact,

$$x_t = (I - \rho\mathbf{L})^{-1}(I + \theta\mathbf{L})\varepsilon_t = \varepsilon_t + \sum_{l=1}^{+\infty} \rho^{l-1}(\rho + \theta)\varepsilon_{t-l} = \sum_{h=0}^{+\infty} \alpha_h \varepsilon_{t-h},$$

where we define

$$\alpha_h = \begin{cases} 1 & \text{if } h = 0, \\ \rho^{h-1}(\rho + \theta) & \text{if } h \geqslant 1. \end{cases}$$

We now employ Proposition 1 for the computation of multiscale impulse responses. Fixed a scale $j \in \mathbb{N}$, we obtain

$$\beta_0^{(j)} = \frac{1}{\sqrt{2^j}} \left\{ I + (I - \rho)^{-1} \left(I - 2\rho^{2^{j-1}-1} + \rho^{2^j-1} \right) (\rho + \theta) \right\}$$

and, for any $k \in \mathbb{N}$,

$$\beta_k^{(j)} = \frac{1}{\sqrt{2^j}} (I - \rho)^{-1} \left(I - \rho^{2^{j-1}} \right)^2 \rho^{k2^j-1}(\rho + \theta).$$

[37] Recall that $\|\rho\|_A^2 = \lambda_{max}(\rho'\rho)$ is the largest eigenvalue of the positive semidefinite matrix $\rho'\rho$. In the literature, other assumptions are also considered, for instance *stability* (see Lütkepohl [19]).

[38] Indeed,

$$\left\| \sum_{l=0}^{+\infty} (\rho\mathbf{L})^l \right\| \leqslant \sum_{l=0}^{+\infty} \left\| (\rho\mathbf{L})^l \right\| \leqslant \sum_{l=0}^{+\infty} \|(\rho\mathbf{L})\|^l \leqslant \sum_{l=0}^{+\infty} \|\rho\|_A^l \|\mathbf{L}\|^l = \sum_{l=0}^{+\infty} \|\rho\|_A^l < +\infty.$$

By setting $\theta = 0$ we find the multiscale impulse responses for a $VAR(1)$. In particular, for any $k \in \mathbb{N}_0$, the matrix coefficients $\beta_k^{(j)}$ turn out to be

$$\beta_k^{(j)} = \frac{1}{\sqrt{2^j}}(I-\rho)^{-1}\left(I-\rho^{2^{j-1}}\right)^2 \rho^{k2^j}.$$

As an example, consider a weakly stationary bivariate $VAR(1)$ process with $x_t = [y_t, z_t]'$, $\varepsilon_t = [u_t, v_t]'$ as unit variance white noise and

$$\rho = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

that is

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} u_t \\ v_t \end{bmatrix}.$$

For any $j \in \mathbb{N}$ and $k \in \mathbb{N}_0$, the multiscale impulse responses $\beta_k^{(j)}$ turns out to be

$$\beta_k^{(j)} = \frac{1}{\sqrt{2^j}\left[(1-a)(1-d)-bc\right]} \begin{bmatrix} 1-d & -b \\ -c & 1-a \end{bmatrix} \left(I - \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{2^{j-1}}\right)^2 \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{k2^j}.$$

4.2 Blanchard and Quah model

As in Blanchard and Quah [7], we take into account a zero-mean weakly stationary purely non-deterministic bivariate time series $\mathbf{x} = \{x_t\}_{t\in\mathbb{Z}}$ such that

$$x_t = \sum_{h=0}^{+\infty} \alpha_h \varepsilon_{t-h}, \qquad \alpha_h \in A \tag{2}$$

where $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a unit variance bivariate with noise, $A = \mathbb{R}^2 \times \mathbb{R}^2$ and the matrix coefficients satisfy the *long-run restriction*

$$\sum_{h=0}^{+\infty} \alpha_h(1,1) = 0.$$

x_t is supposed to have also the MA representation

$$x_t = \sum_{h=0}^{+\infty} c_h \eta_{t-h}, \qquad c_h \in A, \tag{3}$$

where $\boldsymbol{\eta} = \{\eta_t\}_{t\in\mathbb{Z}}$ is a bivariate with noise with covariance matrix Ω. The latter is the usual formulation obtained by estimating the time series parameters from the data. Specifically, we first estimate an autoregressive form for x_t, that is

$$x_t = \sum_{k=1}^{N} b_k x_{t-k} + \eta_t, \qquad N \in \mathbb{N}.$$

The matrix Ω is obtained from the covariance matrix of the residuals in this multivariate regression. Then, the autoregressive form implies that

$$x_t = \eta_t + \sum_{k=1}^{N} b_k \sum_{h=0}^{+\infty} c_h \eta_{t-k-h} = \eta_t + \sum_{n=1}^{+\infty} \left(\sum_{h=\max\{n-N,0\}}^{n-1} b_{n-h} c_h \right) \eta_{t-n}.$$

Therefore,

$$c_0 = I, \qquad c_n = \sum_{h=\max\{n-N,0\}}^{n-1} b_{n-h} c_h \quad \forall n \in \mathbb{N}.$$

The MA representations (2) and (3) of x_t are related by

$$\eta_t = \alpha_0 \varepsilon_t, \qquad \alpha_h = c_h \alpha_0,$$

where the matrix α_0 is such that $\Omega = \alpha_0 \alpha_0'$. However, many choices for α_0 are possible since the factorization of Ω provides just three conditions for the identification of α_0. The long-run restriction is the fourth requirement that guarantees the identification (up to a sign restriction). Indeed, there exists a unique lower triangular matrix s, obtained by the Cholesky factorization, such that $\Omega = ss'$. Any α_0 such that $\Omega = \alpha_0 \alpha_0'$ is an orthonormal transformation of s, namely $\alpha_0 = sr'$ with $r \in A$ orthonormal. The long-run restriction and the sign restrictions $r(1,2) < 0$, $r(2,1) > 0$ imply that r is uniquely determined by

$$r = -\frac{1}{\sqrt{\vartheta^2 + 1}} \begin{bmatrix} \vartheta & 1 \\ -1 & \vartheta \end{bmatrix}$$

with

$$\vartheta = -\frac{s(2,2) \sum_{h=0}^{\infty} c_h(1,2)}{s(1,1) \sum_{h=0}^{\infty} c_h(1,1) \; + \; s(2,1) \sum_{h=0}^{\infty} c_h(1,2)}.$$

In Blanchard and Quah $x_t = [y_t, z_t]'$, where y_t is the first-difference process of log real GNP (or *output growth*) and z_t is the seasonally adjusted unemployment rate for males aged more than 20. Data are taken quarterly and they span from 1950 : $Q2$ to 1987 : $Q4$. The maximum autoregressive lag N is chosen equal to 8. To reduce non-stationarity, the unemployment rate is linearly detrended while the output growth is demeaned by splitting the sample in two parts: before and after 1973 : $Q4$. The multivariate innovation $\varepsilon_t = [u_t, v_t]'$ consists of the *demand shock* u_t and the *supply shock* v_t. The impulse responses of output are obtained by cumulating the impulse responses of output growth. They are plotted together with unemployment rate impulse responses in Figure 1, which reproduces Figures 1 and 2 in Blanchard and Quah [7]. The impulse responses of output with respect to u_t converge to zero in the long term as a consequence of the long-run restriction imposed from the beginning. This phenomenon, however, is not present in the impulse responses of GNP with respect to v_t. For this

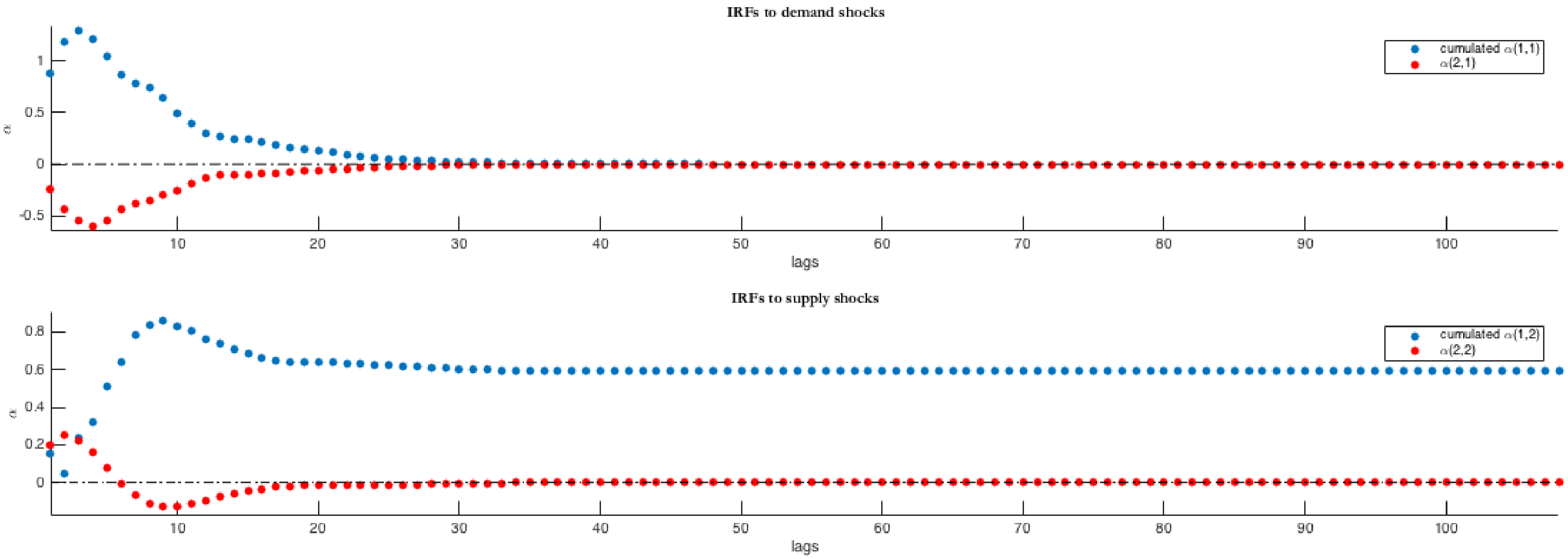

Figure 1: Impulse response functions of output (in blue) and unemployment rate (in red) with respect to demand or supply shocks. See Figures 1 and 2 in Blanchard and Quah [7].

reason, the demand shock u_t is associated with a transitory effect, while the supply disturbance v_t has a permanent impact on output.

In fact, Blanchard and Quah's result is mainly the following. Demand shocks have hump-shaped effects on output and unemployment, with a peak after two or four quarters. The main discrepancy between the two humps is given by their sign, that are opposite. Moreover, the impact of u_t vanishes after three or five years. The economic interpretation is that demand disturbances have similar relevant effects on GNP and employment but, definitively, the subsequent adjustment of prices and wages leads the economy back to the equilibrium. As for supply shocks, the influence of innovations v_t on output cumulates over time, reaching a peak after two years. Except for the first quarter, the evolution of GNP is increasing. Then, the output response declines and stabilizes on a steady level after five years from the initial shock. A different reaction, instead, characterizes the unemployment rate. Indeed, even if the supply disturbance is favourable (due for instance to a productivity increase), in the short term unemployment rises, plausibly because of wage rigidities. After several quarters unemployment drops and, later, it slowly reverts to the original value. No effect is present after five years.

Differently from Figure 1, Figure 2 displays the impulse responses of output growth, together with those of the unemployment rate. We notice a positive impact of demand innovations on output growth until the second quarter, followed by a negative oscillatory reaction up to roughly three years. Such behaviour of impulse responses reflects the hump of cumulated responses of GNP. The reactions of output growth to supply shocks are oscillatory too. Moreover, the response is positive except for the first quarter, captured by the coefficient $\alpha_1(1,2)$, and the third year. This is the counterpart of output increase of Figure 1, which is not always monotonic.

From the description above, it is apparent that impulse responses do not always follow definite dynamics. The unclear patterns of responses may hide the superposition of contemporary contrasting reactions. Therefore, we compute the multiscale impulse responses $\beta_k^{(j)}$ of x_t in order to disentangle the effects of demand or supply shocks with heterogeneous persistence. It comes out that the hump shape and the oscillations of responses are due to the overlapping of positive and negative reactions at different scales. The multiscale impulse responses of output growth and unemployment rate at the scales $j = 1, 2, 3, 4$ are displayed in Figures 3, 4, 5 and 6 respectively.

To begin with, consider the output growth reaction to demand disturbances. Multiscale impulse responses at scales 2 and 3 reflect the behaviour of classical impulse responses, positive in the short term and negative later. However, the negative reaction is negligible at scale 1 and 4, which reveal a favourable feedback from demand innovations. As a result, the decline of the hump in the responses of GNP is mainly

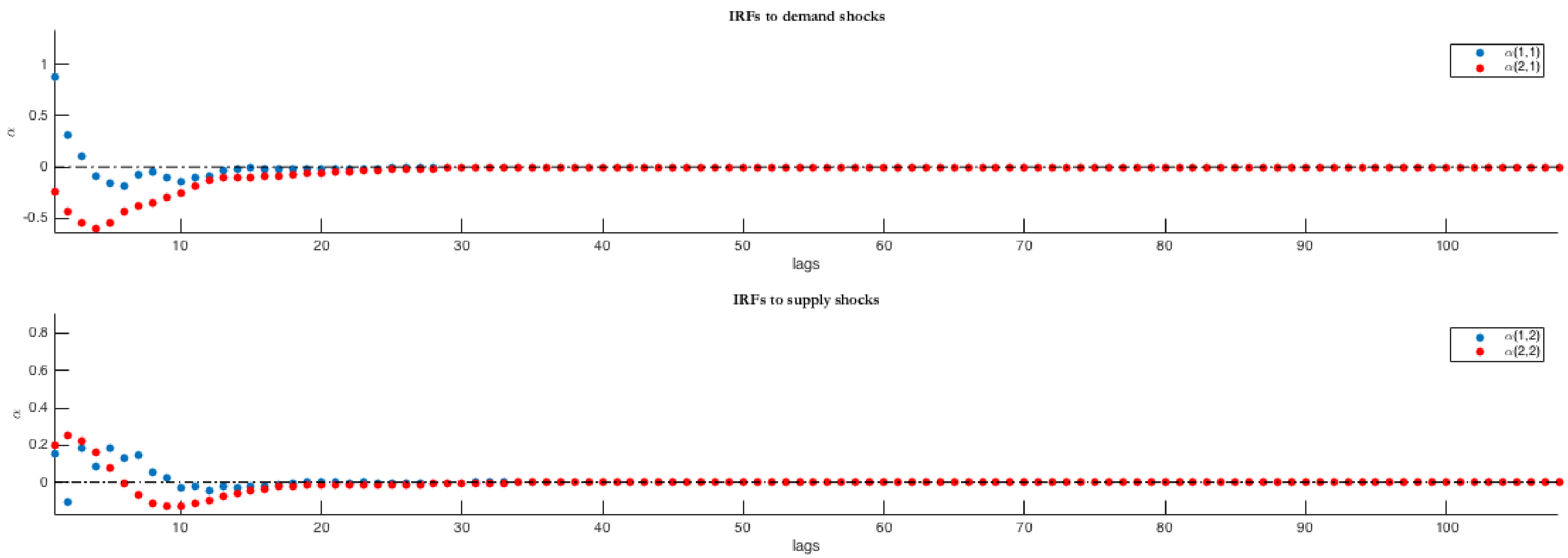

Figure 2: Impulse response functions of output growth (in blue) and unemployment rate (in red) with respect to demand or supply shocks.

due to yearly or biennial shocks occurring at scales 2 and 3 respectively. As for supply shocks, the influence on output growth at scales 1 and 4 is generally positive, while at scale 3 it is negative. This means that the reaction on a biennial basis is counterproductive. In addition, a brake to GNP growth is captured by $\beta_0^{(2)}(1,2)$ and $\beta_0^{(3)}(1,2)$. Hence, the negative reaction explained by $\alpha_1(1,2)$ actually starts at the previous quarter and involves annual and biennial innovations. Such effect is actually concealed by the contemporary positive response quantified by $\beta_0^{(1)}(1,2)$ and $\beta_0^{(4)}(1,2)$.

Now we focus on the unemployment rate. Multiscale responses to demand disturbances are generally negative with the exception of $\beta_0^{(1)}(2,1)$ and $\beta_0^{(2)}(2,1)$, which provide evidence for an immediate and temporary positive reaction to biannual and yearly innovations. On the other hand, the negative coefficient $\beta_0^{(4)}(2,1)$ is prevailing and all the classical impulse responses of unemployment are negative. However, the short-term positive impact of demand shocks delays the large drop in unemployment and makes the hump shape arise. As for the impact of supply disturbances, the behaviour of multiscale responses reflects that of classical responses, except for $\beta_0^{(1)}(2,2)$ which reveals a temporary mean-reversion.

As a result, the MEWD allows us to disaggregate demand/supply calendar-time shocks and to quantify the impact of innovations with different persistent levels. The rigidities advocated by Blanchard and Quah to justify the dynamics of output and unemployment act differently across scales. Moreover, the shocks $\varepsilon_t^{(j)}$ may be due to policies of diverse nature according to the scale, from temporary tax-benefits to long-lasting monetary policy interventions, for instance. Therefore, the fact that the vector process $\mathbf{x}$ comes from the superposition of persistent components with scale-specific behaviours is useful from the policy maker perspective, too. Indeed, the impact of the introduction of short, medium or long-term innovations in the economy is easily quantified. This is the approach formally described by Theorem 5.

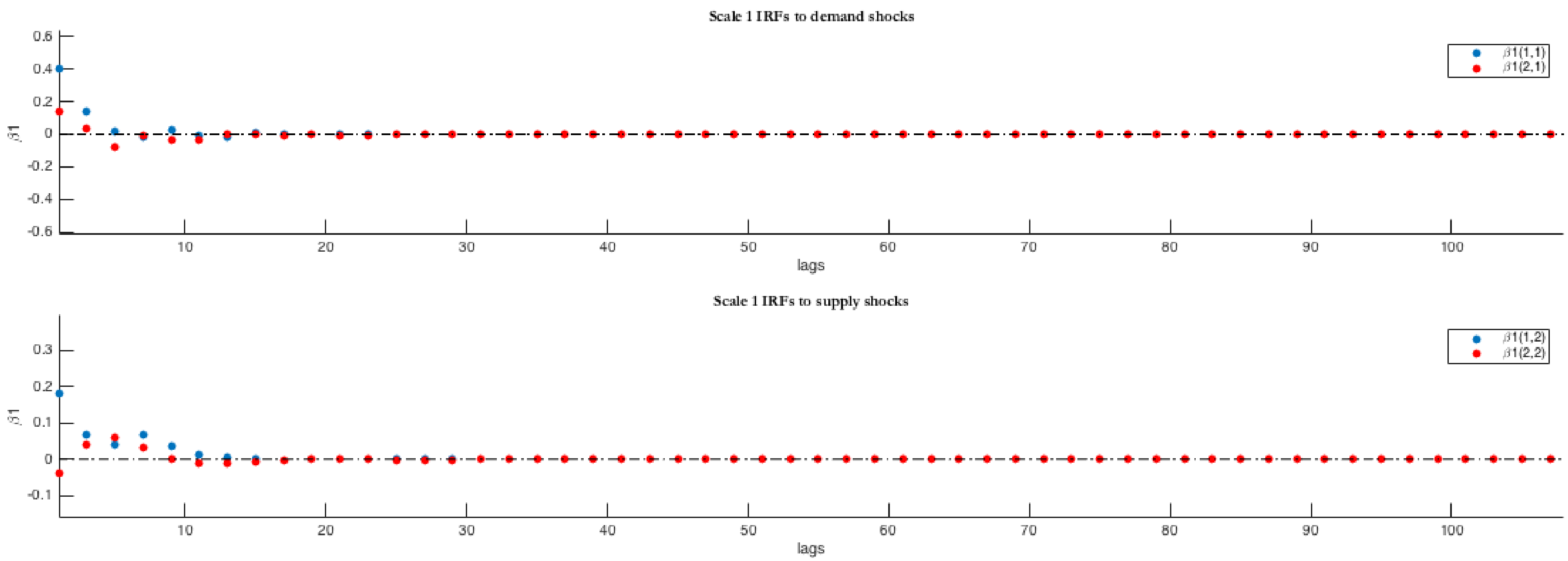

Figure 3: Multiscale impulse response functions of output growth (in blue) and unemployment rate (in red) with respect to demand or supply shocks at scale 1.

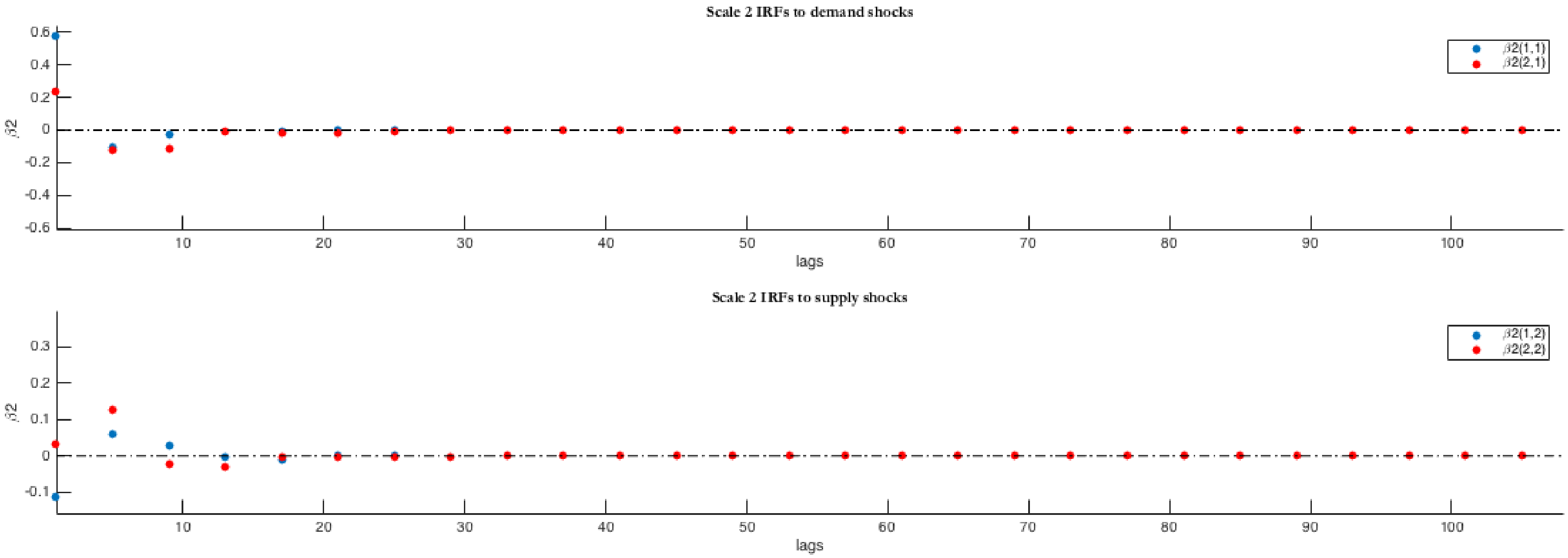

Figure 4: Multiscale impulse response functions of output growth (in blue) and unemployment rate (in red) with respect to demand or supply shocks at scale 2.

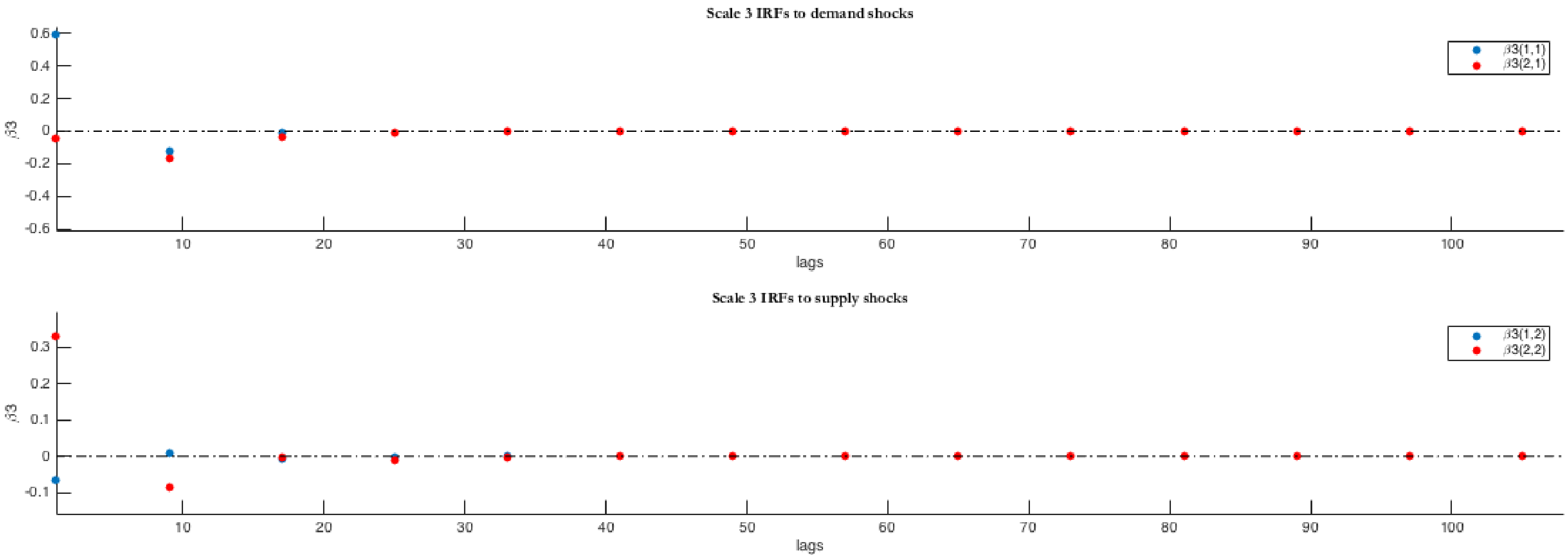

Figure 5: Multiscale impulse response functions of output growth (in blue) and unemployment rate (in red) with respect to demand or supply shocks at scale 3.

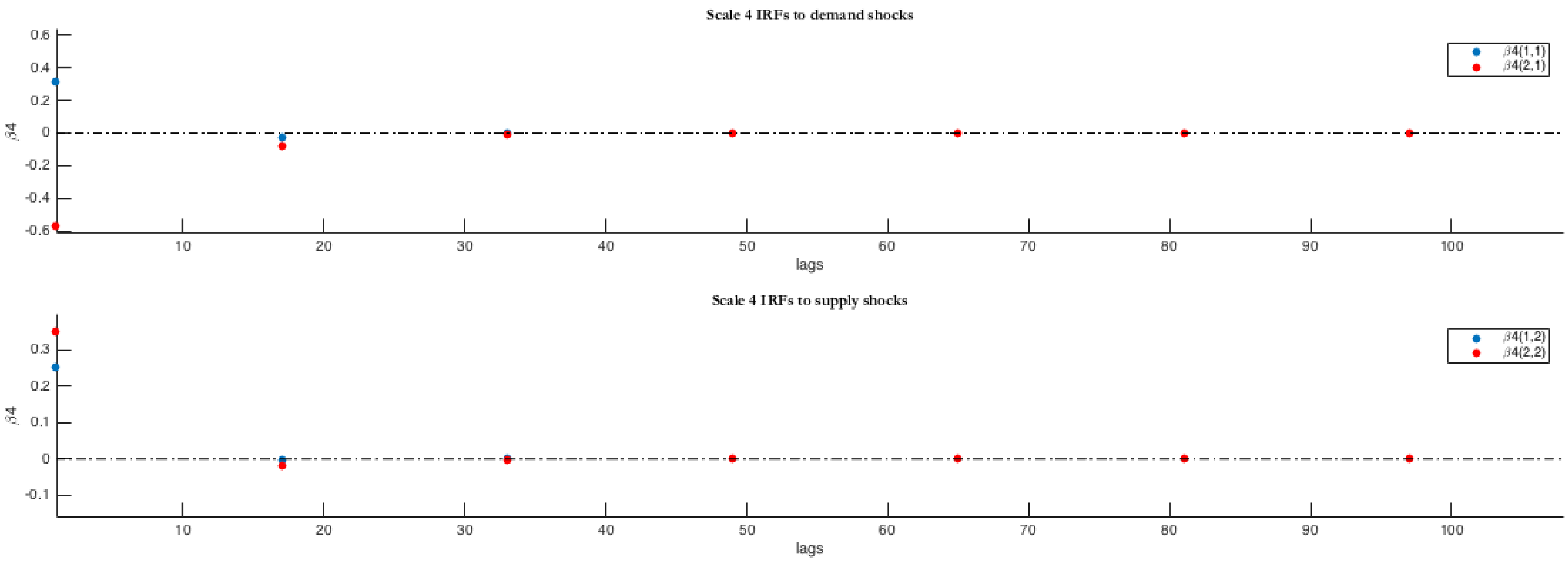

Figure 6: Multiscale impulse response functions of output growth (in blue) and unemployment rate (in red) with respect to demand or supply shocks at scale 4.

5 Conclusion

A wide literature supports the idea that the realizations of economic time series are actually the outcome of the reaction to contemporary phenomena with heterogeneous persistence. Daily news and demographic trends are examples at opposite sides of the spectrum. The situation is even more involved when multivariate processes are taken into account. Indeed, the diversity of persistence commingles with a collection of, possibly correlated, sources of randomness. As things stand, we reach our purpose of eliciting persistent components from vectorial processes by the following plan.

We first revisit the standard treatment of multivariate time series in a Hilbert module framework, where the role of matrix coefficient is clarified. We, then, prove the Abstract Wold Theorem for Hilbert modules that allows us to derive two orthogonal decompositions of the original process: the well-known Multiscale Classical Wold Decomposition and the persistence-based Multivariate Extended Wold Decomposition. The latter provides a decomposition into uncorrelated vectorial components that explain idiosyncratic layers of persistence. Multivariate multiscale impulse responses quantify the dependence on persistent shocks.

As we saw in the analysis of Blanchard and Quah's model, the MEWD provides useful information about the dynamics of multivariate processes. Such information is, indeed, often unrecognisable when the aggregate process is observed by the lenses of classical impulse response functions. Hence, we expect our methodology to be fruitful applied to macroeconomic and financial variables usually modelled by VAR processes. Further potential applications may involve DSGE models, too.

References

[1] Y. A. Abramovich and C. D. Aliprantis, *An Invitation to Operator Theory*, American Mathematical Society, Providence, 2002.

[2] C. D. Aliprantis and K. C. Border, *Infinite Dimensional Analysis*, 3rd ed., Springer Verlag, Berlin, 2006.

[3] N. Barberis, *Investing for the long run when returns are predictable*, The Journal of Finance, volume 55, 1, 225–264, Wiley Online Library, 2000.

[4] B. S. Bernanke and I. Mihov, *Measuring monetary policy*, National Bureau of Economic Research, 1995

[5] H. J. Bierens, *The Wold Decomposition*, 2012.

[6] P. J. Brockwell and R. A. Davis, *Time series: theory and methods*, Springer, 2009.

[7] O. Blanchard and D. Quah, *The dynamic effects of aggregate demand and supply disturbances*, The American Economic Review, Vol. 79, n. 4, 655–673, 1989.

[8] O. Blanchard and R. Perotti, *An empirical characterization of the dynamic effects of changes in government spending and taxes on output*, The Quarterly Journal of Economics, 2002.

[9] J. Y. Campbell and L. M. Viceira, *Strategic asset allocation: portfolio choice for long-term investors*, Oxford University Press, USA, 2002.

[10] S. Cerreia-Vioglio, F. A. Maccheroni, and M. Marinacci, *Hilbert A-modules*, Journal of Mathematical Analysis and Applications, 2017.

[11] M. Frank, *Hilbert C*-modules and related subjects – a guided reference overview*, arXiv, 1996.

[12] H. H. Goldstine and L. P. Horwitz, *Hilbert space with non-associative scalars II*, Mathematische Annalen, 164, 291–316, 1966.

[13] T. Guo, *On Some Basic Theorems of Continuous Module Homomorphisms between Random Normed Modules*, Journal of Function Spaces and Applications, 1–13, 2013.

[14] J. D. Hamilton, *Time series analysis*, Vol. 2, Princeton university press, Princeton, 1994.

[15] R. A. Horn and C. R. Johnson, *Matrix analysis*, Cambridge university press, 2012.

[16] L. P. Hansen and S. F. Richard, *The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models*, Econometrica, 55, 587–613, 1987.

[17] I. Kaplansky, *Modules Over Operator Algebras*, American Journal of Mathematics, 75, 839–858, 1953.

[18] E. C. Lance, *Hilbert C*-modules: A toolkit for operator algebraists*, Cambridge University Press, Cambridge, 1995.

[19] H. Lütkepohl, *New introduction to multiple time series analysis*, Springer Verlag, Berlin, 2005.

[20] F. Ortu, F. Severino, A. Tamoni, and C. Tebaldi *A persistence-based Wold-type decomposition for stationary time series*, Available at SSRN 1973049, 2017.

[21] I. Raeburn and D. P. Williams, *Morita Equivalence and Continuous-trace C*-algebras*, American Mathematical Society, Providence, 1998.

[22] Y. A. Rozanov, *Stationary random processes*, Holden-Day 1967.

[23] C. A. Sims, *Macroeconomics and reality*, Econometrica, 1–48, 1980.

[24] B. Sz.-Nagy, C. Foias, H. Bercovici, and L. Kérchy, *Harmonic Analysis of Operators in Hilbert Spaces*, Springer, New York, 2010.

[25] L. Trefethen and D. Bau III, *Numerical linear algebra*, Vol. 50, Siam, 1970.

[26] H. Wold, *A study in the analysis of stationary time series*, 1938.

A Hilbert A-modules

In this section of the appendix we present a short primer on Hilbert A-modules. The purpose is twofold: a) to present a uniform and self-contained treatment of the topic, b) to present results that are key for our theory and we could not find in the literature. We will mostly focus our attention to the case of A being the algebra of squared matrices, but we will keep our setting abstract in order to avoid getting lost in useless details.

Hilbert A-modules are nothing else, but a generalization of Hilbert spaces. In particular, one starts from the observation that the scalar field $\mathbb{R}$ in a Hilbert space can be replaced by an abstract algebra A: for example, the algebra of matrices. All definitions[39] are then kept identical to the ones of the scalar case. Since the seminal paper of Kaplansky [17], Hilbert modules have been widely studied in Mathematics. In Economics instead, Hilbert A-modules were studied and fruitfully used by Hansen and Richard [16] to prove a conditional version of the fundamental theorem of asset pricing. Mostly, the mathematical literature focused on complex C^*-algebras and developed very rapidly and in a non systematic/scattered way.[40] On the other hand, the real case received little attention. Particularly, most of the results available have been developed for algebras that are commutative, which is not our case here. This includes the work of Hansen and Richard [16], Guo [13], and Cerreia-Vioglio, Maccheroni, Marinacci [10]. One notable exception to this is the paper of Goldstine and Horwitz [12] which deals with the case we have at hand here: the algebra of squared real matrices.

We conclude by observing that the reader might be tempted to think that Hilbert A-modules behave exactly like Hilbert spaces. For example, one might think that, as it is the case for Hilbert spaces, each linear and bounded functional can be represented by using the (generalized) inner product: the famous Riesz Theorem. Similarly, one could also think that any closed subspace is automatically complemented. Unfortunately, this is not the case and much of the truth of these statements depends on both the properties of A and H.[41] In what follows, we will derive both results and, in so doing, we will highlight what are the connections and differences with the existing literature. These two results will be instrumental in proving the Abstract Wold Theorem.

A.1 Introduction

Let A be a *real* C^*-algebra with (multiplicative) unit e which is isomorphic to the *real* C^*-algebra of bounded operators over a *real* Hilbert space. We order A with the partial order induced by the cone of its positive and self-adjoint elements. It follows that A

[39]See, e.g., Definition 5.

[40]See also Frank [11].

[41]This applies even to the complex case (see Lance [18, p. 7]).

has a (multiplicative) unit e. We next proceed by defining the objects we study in this paper.

Definition 5 *An abelian group* $(H,+)$ *is an* A-module *if and only if an outer product* $\cdot : A\times H\to H$ *is well defined with the following properties, for each* $a,b\in A$ *and for each* $x,y\in H$*:*

(1) $a\cdot(x+y)=a\cdot x+a\cdot y$*;*

(2) $(a+b)\cdot x=a\cdot x+b\cdot x$*;*

(3) $a\cdot(b\cdot x)=(ab)\cdot x$*;*

(4) $e\cdot x=x$*.*

An A*-module is a* pre-Hilbert A-module *if and only if an inner product* $\langle\ ,\ \rangle_H : H\times H\to A$ *is well defined with the following properties, for each* $a\in A$ *and for each* $x,y,z\in H$*:*

(5) $\langle x,x\rangle_H\geq 0$*, with equality if and only if* $x=0$*;*

(6) $\langle x,y\rangle_H=\langle y,x\rangle_H^*$*;*

(7) $\langle x+y,z\rangle_H=\langle x,z\rangle_H+\langle y,z\rangle_H$*;*

(8) $\langle a\cdot x,y\rangle_H=a\langle x,y\rangle_H$*.*

For $A=\mathbb{R}$ conditions (1)-(4) define vector spaces, while (5)-(8) define pre-Hilbert spaces.[42]

Given a pre-Hilbert A-module, we will show that[43]

$$\langle x,y\rangle_H^*\langle x,y\rangle_H\leq\|\langle x,x\rangle_H\|_A\langle y,y\rangle_H\qquad\forall x,y\in H$$

where $\|\ \|_A$ is the norm of A.

Given an element $y\in H$, note that $\langle\ ,\ \rangle_H$ induces an operator $f:H\to A$ defined as $f(x)=\langle x,y\rangle_H$ with the following properties:

- **A-linearity** $f(a\cdot x+b\cdot y)=af(x)+bf(y)$ for all $a,b\in A$ and for all $x,y\in H$;
- **Boundedness** There exists $M>0$ such that $\|f(x)\|_A^2\leq M\|\langle x,x\rangle_H\|_A$ for all $x\in H$.

[42] We will use Latin letters a,b,c to denote elements of A, Latin letters x,y,z to denote elements of H, and Greek letters α,β to denote elements of $\mathbb{R}$.

[43] We will adapt the techniques of Raeburn and Williams [21, Lemma 2.5] to the real case.

In light of this fact, we give the following definition:

Definition 6 *Let H be a pre-Hilbert A-module. We say that H is* self-dual *if and only if for each $f : H \to A$ which is A-linear and bounded there exists $y \in H$ such that*

$$f(x) = \langle x, y \rangle_H \qquad \forall x \in H.$$

A.2 C^*-algebras

In our case, we will consider A to be isomorphic to the algebra of bounded operators on a real Hilbert space H'. In particular, A is a *real normed* algebra with multiplicative unit e, we denote by $\| \|_A$ the norm of A. We denote the norm dual of A by A^*. Recall that A is also a C^*-algebra with unit, that is, there exists an involution $^* : A \to A$ such that for each $a, b \in A$ and $\alpha \in \mathbb{R}$

$$(a+b)^* = a^* + b^*,\ (ab)^* = b^* a^*,\ (\alpha a)^* = \alpha a^*, \text{ and } a^{**} = (a^*)^* = a.$$

The involution also well behaves with the norm, that is,

$$\|a\|_A^2 = \|a^* a\|_A \qquad \forall a \in A.$$

The algebra A is also naturally ordered by the order $\geq$ induced by the closed convex cone of positive elements that are such that $a = a^*$.[44] We denote by $A_+ = \{a \in A : a \geq 0\}$. The following useful properties will be very useful in what follows:

1. $\|a\|_A = \|a^*\|_A$;
2. If $a \in A$, then we have that $a^* a \in A_+$;
3. If $a \geq 0$, then $bab^* \leq \|a\|_A bb^*$;
4. If $a \geq b \geq 0$, then $\|a\|_A \geq \|b\|_A$;
5. If A is finite dimensional, then there exists a continuous linear functional $\bar{\varphi} : A \to \mathbb{R}$ such that

$$\begin{aligned} a \geq 0 &\implies \bar{\varphi}(a) \geq 0 \\ a \geq 0 \text{ and } \bar{\varphi}(a) = 0 &\iff a = 0 \\ \bar{\varphi}(a) = \bar{\varphi}(a^*) &\qquad \forall a \in A \\ \exists K > 0 \text{ such that } \|a\|_A \leq \bar{\varphi}(a) \leq K \|a\|_A &\qquad \forall a \geq 0. \end{aligned}$$

We will call a continuous and linear functional $\bar{\varphi}$ that satisfies the first three properties of point 5. *strictly positive*. We will call a functional as in point 5 a trace. Since A_+ is a closed convex cone, there exists a closed and convex set $C \subseteq A^*$ such that

$$a \geq b \iff \varphi(a) \geq \varphi(b) \qquad \forall \varphi \in C. \tag{4}$$

[44] In the real case, the extra requirement $a = a^*$ is not redundant.

A.3 The vector space structure of H

In this section, we will first show that a pre-Hilbert A-module has a natural structure of vector space. Next, we will show that the Avalued inner product $\langle\ ,\ \rangle_H$ shares some of the properties of standard real-valued inner products. In particular, under the assumption that A admits a strictly positive functional $\bar{\varphi}$, we will show that it also induces a real valued inner product on H, thus making H into a pre-Hilbert space.

We use the outer product $\cdot$ to define a scalar product:

$$\begin{array}{rccc} \cdot^e : & \mathbb{R}\times H & \rightarrow & H \\ & (\alpha, x) & \mapsto & (\alpha e)\cdot x \end{array}.$$

We next show that $\cdot^e$ makes the abelian group H into a real vector space.

Proposition 2 *Let H be an A-module. $(H, +, \cdot^e)$ is a real vector space.*

Proof. By assumption, H is an abelian group. For each $\alpha, \beta \in \mathbb{R}$ and each $x, y \in H$, we have that

(1) $\alpha \cdot^e (x + y) = \alpha e \cdot (x + y) = (\alpha e)\cdot x + (\alpha e)\cdot y = \alpha \cdot^e x + \alpha \cdot^e y;$

(2) $(\alpha + \beta) \cdot^e x = ((\alpha + \beta) e) \cdot x = (\alpha e + \beta e)\cdot x = (\alpha e)\cdot x + (\beta e)\cdot x = \alpha \cdot^e x + \beta \cdot^e x;$

(3) $\alpha \cdot^e (\beta \cdot^e x) = (\alpha e)\cdot ((\beta e)\cdot x) = ((\alpha e)(\beta e))\cdot x = ((\alpha\beta) e)\cdot x = (\alpha\beta)\cdot^e x;$

(4) $1 \cdot^e x = (1e)\cdot x = e \cdot x = x.$

∎

From now on, we will often write αx in place of $\alpha \cdot^e x$.

Corollary 1 *Let H be an A-module. If $f : H \rightarrow A$ is an A-linear operator, then f is linear.*

Proof. Consider $x, y \in H$ and $\alpha, \beta \in \mathbb{R}$. We have that

$$\begin{aligned} f(\alpha x + \beta y) &= f((\alpha e)\cdot x + (\beta e)\cdot y) = (\alpha e) f(x) + (\beta e) f(y) \\ &= \alpha f(x) + \beta f(y), \end{aligned}$$

proving the statement. ∎

Assume A admits a strictly positive functional $\bar{\varphi}$. Define $\langle\ ,\ \rangle_{\bar{\varphi}} : H \times H \rightarrow \mathbb{R}$ by

$$\langle x, y\rangle_{\bar{\varphi}} = \bar{\varphi}(\langle x, y\rangle_H) \quad \forall x, y \in H.$$

Proposition 3 *Let H be a pre-Hilbert A-module. If A admits a strictly positive functional $\bar{\varphi}$, then $\langle\ ,\ \rangle_{\bar{\varphi}}$ is an inner product.*

Proof. We prove four properties:

a. Consider $x \in H$. By assumption, we have that $\langle x, x\rangle_H \geq 0$. Since $\bar{\varphi}$ is positive, it follows that

$$\langle x, x\rangle_{\bar{\varphi}} = \bar{\varphi}\left(\langle x, x\rangle_H\right) \geq 0.$$

Since $\bar{\varphi}$ is strictly positive and $\langle x, x\rangle_H \geq 0$, note also that

$$\bar{\varphi}\left(\langle x, x\rangle_H\right) = 0 \iff \langle x, x\rangle_H = 0 \iff x = 0.$$

b. Consider $x, y \in H$. Since $\bar{\varphi}(a) = \bar{\varphi}(a^*)$ for all $a \in A$, we have that

$$\langle y, x\rangle_{\bar{\varphi}} = \bar{\varphi}\left(\langle y, x\rangle_H\right) = \bar{\varphi}\left(\langle y, x\rangle_H^*\right) = \bar{\varphi}\left(\langle x, y\rangle_H\right) = \langle x, y\rangle_{\bar{\varphi}}.$$

c. Consider $x, y, z \in H$. Since $\bar{\varphi}$ is linear, we obtain that

$$\begin{aligned}\langle x+y, z\rangle_{\bar{\varphi}} &= \bar{\varphi}\left(\langle x+y, z\rangle_H\right) = \bar{\varphi}\left(\langle x, z\rangle_H + \langle y, z\rangle_H\right)\\ &= \bar{\varphi}\left(\langle x, z\rangle_H\right) + \bar{\varphi}\left(\langle y, z\rangle_H\right) = \langle x, z\rangle_{\bar{\varphi}} + \langle y, z\rangle_{\bar{\varphi}}.\end{aligned}$$

d. Consider $x, y \in H$ and $\alpha \in \mathbb{R}$. Since $\bar{\varphi}$ is linear, we obtain that

$$\begin{aligned}\langle \alpha x, y\rangle_{\bar{\varphi}} &= \bar{\varphi}\left(\langle (\alpha e)\cdot x, y\rangle_H\right) = \bar{\varphi}\left((\alpha e)\langle x, y\rangle_H\right)\\ &= \bar{\varphi}\left(\alpha\langle x, y\rangle_H\right) = \alpha\langle x, y\rangle_{\bar{\varphi}}.\end{aligned}$$

Properties a-d yield the statement. ■

Corollary 2 *Let H be a pre-Hilbert A-module. If A admits a strictly positive functional $\bar{\varphi}$, then $\left(H, +, \cdot^e, \langle\ ,\ \rangle_{\bar{\varphi}}\right)$ is a pre-Hilbert space.*

Proposition 4 *Let H be a pre-Hilbert A-module. The following statements are true:*

1. $\langle x, y\rangle_H^* \langle x, y\rangle_H \leq \|\langle x, x\rangle_H\|_A \langle y, y\rangle_H$ *for all $x, y \in H$;*

2. $\|\langle x, y\rangle_H\|_A^2 \leq \|\langle x, x\rangle_H\|_A \|\langle y, y\rangle_H\|_A$ *for all $x, y \in H$;*

3. $\|\langle x, y\rangle_H\|_A \leq \|\langle x, x\rangle_H\|_A^{\frac{1}{2}} \|\langle y, y\rangle_H\|_A^{\frac{1}{2}}$ *for all $x, y \in H$.*

Proof. Consider $w, z \in H$ and assume that $\langle w, z\rangle_H = \langle w, z\rangle_H^*$. It follows that for each $t \geq 0$

$$\begin{aligned}0 &\leq \langle w+tz, w+tz\rangle_H = \langle w, w+tz\rangle_H + \langle tz, w+tz\rangle_H\\ &= \langle w, w\rangle_H + \langle w, tz\rangle_H + \langle tz, w\rangle_H + \langle tz, tz\rangle_H\\ &= \langle w, w\rangle_H + t\langle w, z\rangle_H + t\langle z, w\rangle_H + t^2\langle z, z\rangle_H\\ &= \langle w, w\rangle_H + t\langle w, z\rangle_H + t\langle w, z\rangle_H^* + t^2\langle z, z\rangle_H\\ &= \langle w, w\rangle_H + 2t\langle w, z\rangle_H + t^2\langle z, z\rangle_H.\end{aligned}$$

Consider $\varphi \in C$. It follows that

$$\begin{aligned}0 &\leq \varphi\left(\langle w+tz, w+tz\rangle_H\right) = \varphi\left(\langle w,w\rangle_H + 2t\langle w,z\rangle_H + t^2\langle z,z\rangle_H\right)\\ &= \varphi\left(\langle w,w\rangle_H\right) + 2t\varphi\left(\langle w,z\rangle_H\right) + t^2\varphi\left(\langle z,z\rangle_H\right),\end{aligned}$$

yielding that

$$\varphi\left(\langle w,z\rangle_H\right)^2 \leq \varphi\left(\langle w,w\rangle_H\right)\varphi\left(\langle z,z\rangle_H\right). \tag{5}$$

Choose $\bar{x}, \bar{y} \in H$. Define $w = \langle \bar{x}, \bar{y}\rangle_H^* \bar{x}$ and $\bar{y} = y$. It follows that

$$\langle w,z\rangle_H = \langle \bar{x},\bar{y}\rangle_H^* \langle \bar{x},\bar{y}\rangle_H,$$

yielding that $\langle w,z\rangle_H = \langle w,z\rangle_H^*$ and (5) holds. In particular, we have that

$$\begin{aligned}\varphi\left(\langle \bar{x},\bar{y}\rangle_H^*\langle \bar{x},\bar{y}\rangle_H\right)^2 &= \varphi\left(\langle w,z\rangle_H\right)^2 \leq \varphi\left(\langle w,w\rangle_H\right)\varphi\left(\langle z,z\rangle_H\right)\\ &= \varphi\left(\langle \bar{x},\bar{y}\rangle_H^*\langle \bar{x},\bar{x}\rangle_H\langle \bar{x},\bar{y}\rangle_H\right)\varphi\left(\langle \bar{y},\bar{y}\rangle_H\right).\end{aligned}$$

Define $a = \langle \bar{x},\bar{x}\rangle_H$ and $b = \langle \bar{x},\bar{y}\rangle_H^*$. Recall that

$$bab^* \leq \|a\|_A\, bb^* \text{and } bb^* \geq 0.$$

Thus, we have that

$$\begin{aligned}\varphi\left(\langle \bar{x},\bar{y}\rangle_H^*\langle \bar{x},\bar{y}\rangle_H\right)^2 &\leq \varphi\left(\langle \bar{x},\bar{y}\rangle_H^*\langle \bar{x},\bar{x}\rangle_H\langle \bar{x},\bar{y}\rangle_H\right)\varphi\left(\langle \bar{y},\bar{y}\rangle_H\right)\\ &\leq \varphi\left(\|\langle \bar{x},\bar{x}\rangle_H\|_A\langle \bar{x},\bar{y}\rangle_H^*\langle \bar{x},\bar{y}\rangle_H\right)\varphi\left(\langle \bar{y},\bar{y}\rangle_H\right)\\ &\leq \|\langle \bar{x},\bar{x}\rangle_H\|_A\,\varphi\left(\langle \bar{x},\bar{y}\rangle_H^*\langle \bar{x},\bar{y}\rangle_H\right)\varphi\left(\langle \bar{y},\bar{y}\rangle_H\right)\end{aligned}$$

and

$$\varphi\left(\langle \bar{x},\bar{y}\rangle_H^*\langle \bar{x},\bar{y}\rangle_H\right) \geq 0.$$

We thus have that

$$\varphi\left(\langle \bar{x},\bar{y}\rangle_H^*\langle \bar{x},\bar{y}\rangle_H\right) \leq \varphi\left(\|\langle \bar{x},\bar{x}\rangle_H\|_A\langle \bar{y},\bar{y}\rangle_H\right). \tag{6}$$

Since φ was arbitrarily chosen, we have that (6) holds for all $\varphi \in C$, that is, by (4)

$$\langle \bar{x},\bar{y}\rangle_H^*\langle \bar{x},\bar{y}\rangle_H \leq \|\langle \bar{x},\bar{x}\rangle_H\|_A\langle \bar{y},\bar{y}\rangle_H.$$

Since $\bar{x}$ and $\bar{y}$ were arbitrarily chosen, the statement follows.

2. Consider $x, y \in H$. Call $a = \langle x,y\rangle_H$ and $b = \|\langle x,x\rangle_H\|_A\langle y,y\rangle_H$. By point 1, we have that

$$0 \leq a^*a \leq b.$$

It follows that

$$\begin{aligned}\|\langle x,y\rangle_H\|_A^2 = \|a\|_A^2 = \|a^*a\|_A \leq \|b\|_A &= \left\|\|\langle x,x\rangle_H\|_A\langle y,y\rangle_H\right\|_A\\ &= \|\langle x,x\rangle_H\|_A\|\langle y,y\rangle_H\|_A,\end{aligned}$$

proving the point.

3. It trivially follows from point 2. ■

A.3.1 Topological structure

The $\| \ \|_H$ norm

Define $\| \ \|_H : H \to [0, \infty)$ by

$$\|x\|_H = \sqrt{\|\langle x, x\rangle_H\|_A} \qquad \forall x \in H.$$

Proposition 5 *Let H be pre-Hilbert A-module. The following statements are true:*

1. *$\| \ \|_H$ is a norm;*
2. *$\|a \cdot x\|_H \leq \|a\|_A \|x\|_H$ for all $a \in A$ and all $x \in H$.*

Proof. 1. Note that

$$\|x\|_H = 0 \iff \|\langle x, x\rangle_H\|_A = 0 \iff \langle x, x\rangle_H = 0 \iff x = 0.$$

Note also that for each $\alpha \in \mathbb{R}$ and $x \in H$

$$\begin{aligned} \|\alpha x\|_H &= \sqrt{\|\langle \alpha x, \alpha x\rangle_H\|_A} = \sqrt{\|\alpha^2 \langle x, x\rangle_H\|_A} \\ &= |\alpha| \sqrt{\|\langle x, x\rangle_H\|_A} = |\alpha| \|x\|_H . \end{aligned}$$

Finally, we have that for each $x, y \in H$

$$\begin{aligned} \|x + y\|_H^2 &= \|\langle x + y, x + y\rangle_H\|_A \\ &= \|\langle x, x\rangle_H + \langle x, y\rangle_H + \langle y, x\rangle_H + \langle y, y\rangle_H\|_A \\ &\leq \|\langle x, x\rangle_H\|_A + \|\langle x, y\rangle_H\|_A + \|\langle y, x\rangle_H\|_A + \|\langle y, y\rangle_H\|_A \\ &\leq \|\langle x, x\rangle_H\|_A + \sqrt{\|\langle x, x\rangle_H\|_A \|\langle y, y\rangle_H\|_A} + \sqrt{\|\langle y, y\rangle_H\|_A \|\langle x, x\rangle_H\|_A} + \|\langle y, y\rangle_H\|_A \\ &= \|\langle x, x\rangle_H\|_A + 2\sqrt{\|\langle x, x\rangle_H\|_A \|\langle y, y\rangle_H\|_A} + \|\langle y, y\rangle_H\|_A \\ &= \left(\|\langle x, x\rangle_H\|_A^{\frac{1}{2}} + \|\langle y, y\rangle_H\|_A^{\frac{1}{2}}\right)^2 . \end{aligned}$$

We can thus conclude that

$$\begin{aligned} \|x + y\|_H &\leq \left(\|\langle x, x\rangle_H\|_A^{\frac{1}{2}} + \|\langle y, y\rangle_H\|_A^{\frac{1}{2}}\right) \\ &= \|x\|_H + \|y\|_H , \end{aligned}$$

proving that $\| \ \|_H$ is a norm.

2. Given any $a \in A$ and $x \in H$, define $b = \langle x, x\rangle_H \geq 0$.

$$\begin{aligned} \|a \cdot x\|_H^2 &= \|\langle a \cdot x, a \cdot x\rangle_H\|_A = \|a \langle x, x\rangle_H a^*\|_A = \|aba^*\|_A \leq \|b\|_A \|aa^*\|_A \\ &\leq \|\langle x, x\rangle_H\|_A \|a\|_A^2 = \|a\|_A^2 \|x\|_H^2 , \end{aligned}$$

proving the statement. ■

By Proposition 4, it readily follows that

$$\|\langle x,y\rangle_H\|_A \leq \|x\|_H \|y\|_H \qquad \forall x,y \in H. \tag{7}$$

Corollary 3 *Let H be a pre-Hilbert A-module. For each $y \in H$, the functional $\langle \cdot, y\rangle_H : H \to A$ is A-linear, $\|\ \|_H - \|\ \|_A$ continuous, and has norm $\|y\|_H$.*

Proof. Fix $y \in H$. It is immediate to see that the operator induced by y is A-linear, thus, linear. Continuity easily follows from (7). Since the norm of the linear operator is given by

$$\sup\left\{\|\langle x,y\rangle_H\|_A / \|x\|_H : x \neq 0\right\},$$

the statement easily follows from (7) and the definition of $\|\ \|_H$. ■

The $\|\ \|_{\bar{\varphi}}$ norm

Assume A admits a strictly positive functional $\bar{\varphi}$. Define $\|\ \|_{\bar{\varphi}} : H \to [0,\infty)$ by

$$\|x\|_{\bar{\varphi}} = \sqrt{\langle x,x\rangle_{\bar{\varphi}}} \qquad \forall x \in H. \tag{8}$$

By Corollary 2, $\langle\ ,\ \rangle_{\bar{\varphi}}$ is an inner product on H and it is immediate to see that $\|\ \|_{\bar{\varphi}}$ is a norm and

$$\|x\|_{\bar{\varphi}} = \sqrt{\bar{\varphi}\left(\langle x,x\rangle_H\right)} \qquad \forall x \in H. \tag{9}$$

Relations among norms

Assume A admits a strictly positive functional $\bar{\varphi}$. Since $\bar{\varphi}$ is a continuous linear functional, it follows that there exists $K > 0$ such that

$$\bar{\varphi}(a) \leq K \|a\|_A \qquad \forall a \in A_+.$$

This implies that

$$\|x\|_{\bar{\varphi}}^2 = \bar{\varphi}\left(\langle x,x\rangle_H\right) \leq K \|\langle x,x\rangle_H\|_A = \|x\|_H^2 \qquad \forall x \in H,$$

that is,

$$\|x\|_{\bar{\varphi}} \leq \sqrt{K}\, \|x\|_H \qquad \forall x \in H.$$

We can conclude that

$$x_n \overset{\|\ \|_H}{\to} 0 \implies x_n \overset{\|\ \|_{\bar{\varphi}}}{\to} 0.$$

Proposition 6 *Let H be a pre-Hilbert A-module. If A is finite dimensional, then A admits a trace $\bar{\varphi}$ and the norms $\|\ \|_{\bar{\varphi}}$ and $\|\ \|_H$ are equivalent.*

Proof. Since A is finite dimensional, there exists $K > 0$ such that $\|a\|_A \leq \bar{\varphi}(a) \leq K \|a\|_A$ for all $a \geq 0$. It follows that

$$\begin{aligned} \|x\|_H &= \sqrt{\|\langle x, x\rangle_H\|_A} \leq \sqrt{\bar{\varphi}(\langle x, x\rangle_H)} = \|x\|_{\bar{\varphi}} \\ &= \sqrt{\bar{\varphi}(\langle x, x\rangle_H)} \leq \sqrt{K}\sqrt{\|\langle x, x\rangle_H\|_A} \\ &= \sqrt{K}\,\|x\|_H \qquad \forall x \in H, \end{aligned}$$

proving the statement. ■

A.4 Dual module

Given a pre-Hilbert A-module H, we define

$$H^{\sim} = \left\{ f \in A^H : f \text{ is } A\text{-linear and bounded} \right\}.$$

By definition of boundedness and $\| \ \|_H$, we have that f is bounded if and only if there exists $M > 0$ such that

$$\|f(x)\|_A \leq M \|x\|_H \qquad \forall x \in H.$$

Recall that if $f \in H^{\sim}$, then f is linear. Thus, in this case, we have that $H^{\sim} \subseteq B(H, A)$, where the latter is the set of all bounded linear operators from H to A when H is endowed with $\| \ \|_H$ and A is endowed with $\| \ \|_A$.

Proposition 7 *If H is a pre-Hilbert A-module, then $H^{\sim}$ is an A-module.*

Proof. Define $+ : H^{\sim} \times H^{\sim} \to H^{\sim}$ to be such that for each $f, g \in H^{\sim}$

$$(f + g)(x) = f(x) + g(x) \qquad \forall x \in H.$$

In other words, $+$ is the usual pointwise sum of operators. Define $\cdot : A \times H^{\sim} \to H^{\sim}$ to be such that for each $a \in A$ and for each $f \in H^{\sim}$

$$(a \cdot f)(x) = f(x)\, a^* \qquad \forall x \in H.$$

It is immediate to verify that $H^{\sim}$ is closed under $+$ and $\cdot$. In particular, $(H, +)$ is an abelian group. Note that for each $a, b \in A$ and each $f, g \in H^{\sim}$:

1. $(a \cdot (f + g))(x) = ((f + g)(x))\, a^* = (f(x) + g(x))\, a^* = f(x)\, a^* + g(x)\, a^* = (a \cdot f)(x) + (a \cdot g)(x) = (a \cdot f + a \cdot g)(x)$ for all $x \in H$, that is, $a \cdot (f + g) = a \cdot f + a \cdot g$.

2. $((a + b) \cdot f)(x) = f(x)\,(a + b)^* = f(x)\,(a^* + b^*) = f(x)\, a^* + f(x)\, b^* = (a \cdot f)(x) + (b \cdot f)(x) = (a \cdot f + b \cdot f)(x)$ for all $x \in H$, that is, $(a + b) \cdot f = a \cdot f + b \cdot f$.

3. $(a \cdot (b \cdot f))(x) = ((b \cdot f)(x))\, a^* = (f(x)\, b^*)\, a^* = f(x)(b^* a^*) = f(x)(ab)^* = ((ab) \cdot f)(x)$ for all $x \in H$, that is, $a \cdot (b \cdot f) = (ab) \cdot f$.

4. $(e \cdot f)(x) = f(x)\, e^* = f(x)\, e = f(x)$ for all $x \in H$, that is, $e \cdot f = f$. ■

Since $H^\sim$ is an A-module, it is also a vector space. Note that the scalar product $\cdot^e$ coincides with the usual scalar product defined on $B(H, A)$ once restricted to $H^\sim$. Thus, we can also define a norm on $H^\sim$ defined as $\| \ \|_{H^\sim} : H^\sim \to [0, \infty)$ such that

$$\|f\|_{H^\sim} = \sup_{\|x\|_H = 1} \|f(x)\|_A \qquad \forall f \in H^\sim.$$

Define $S^\sim : H \to H^\sim$ by

$$S^\sim(y) = \langle \ , y\rangle_H \qquad \forall y \in H.$$

Given Corollary 3 and the properties of $\langle \ , \ \rangle_H$, the map $S^\sim$ is well defined and linear. In fact, for each $\alpha, \beta \in \mathbb{R}$ and each $y, z \in H$

$$\begin{aligned} S^\sim(\alpha y + \beta z)(x) &= \langle x, \alpha y + \beta z\rangle_H = \langle x, \alpha y\rangle_H + \langle x, \beta z\rangle_H \\ &= \langle x, (\alpha e) \cdot y\rangle_H + \langle x, (\beta e) \cdot z\rangle_H \\ &= \langle x, y\rangle_H (\alpha e)^* + \langle x, z\rangle_H (\beta e)^* \\ &= (S^\sim(y)(x))(\alpha e)^* + (S^\sim(z)(x))(\beta e)^* \\ &= ((\alpha e) \cdot S^\sim(y))(x) + (\beta e \cdot S^\sim(z))(x) \\ &= ((\alpha e) \cdot S^\sim(y) + (\beta e) \cdot S^\sim(z))(x) \qquad \forall x \in H, \end{aligned}$$

proving that

$$S^\sim(\alpha y + \beta z) = (\alpha e) \cdot S^\sim(y) + (\beta e) \cdot S^\sim(z) = \alpha S^\sim(y) + \beta S^\sim(z).$$

Proposition 8 *Let H be a pre-Hilbert A-module. The following statements are true:*

1. *$H^\sim$ is $\| \ \|_{H^\sim}$ complete.*

2. *$S^\sim$ is an isometry, that is, $\|S^\sim(y)\|_{H^\sim} = \|y\|_H$ for all $y \in H$.*

3. *If H is self-dual, then $S^\sim$ is onto and H is $\| \ \|_H$ complete.*

Proof. 1. By Proposition 7, $H^\sim$ is an A-module. In particular, $H^\sim$ is a vector subspace of $B(H, A)$. Consider a $\| \ \|_{H^\sim}$ Cauchy sequence $\{f_n\}_{n \in \mathbb{N}} \subseteq H^\sim \subseteq B(H, A)$. By Aliprantis and Border [2, Theorem 6.6], we have that there exists $f \in B(H, A)$ such that $f_n \overset{\| \ \|_{H^\sim}}{\to} f$. We are left to show that f is A-linear. First, observe that $f : H \to A$ is such that

$$f(x) = \lim_n f_n(x) \qquad \forall x \in H$$

where the limit is in $\| \|_A$ norm. We can conclude that for each $a, b \in A$ and $x, y \in H$

$$f_n(x) \overset{\| \|_A}{\to} f(x), \; f_n(y) \overset{\| \|_A}{\to} f(y) \implies af_n(x) \overset{\| \|_A}{\to} af(x), \; bf_n(y) \overset{\| \|_A}{\to} bf(y)$$
$$\implies af_n(x) + bf_n(y) \overset{\| \|_A}{\to} af(x) + bf(y).$$

At the same time, $af_n(x)+bf_n(y) = f_n(a\cdot x + b\cdot y) \overset{\| \|_A}{\to} f(a\cdot x + b\cdot y)$ for all $a, b \in A$ and $x, y \in H$. By the uniqueness of the limit, we can conclude that $f(a\cdot x + b\cdot y) = af(x) + bf(y)$ for all $a, b \in A$ and $x, y \in H$, proving the statement.

2. Define $S^{\sim} : H \to H^{\sim}$ by

$$S^{\sim}(y)(x) = \langle x, y\rangle_H \qquad \forall x \in H.$$

By Corollary 3, it follows that $\|S^{\sim}(y)\|_{H^{\sim}} = \|y\|_H$ for all $y \in H$.

3. If H is self-dual, it is immediate to see that $S^{\sim}$ is onto. Consider a $\| \|_H$ Cauchy sequence $\{x_n\}_{n\in\mathbb{N}} \subseteq H$. Since $S^{\sim}$ is an isometry, it follows that $\{S^{\sim}(x_n)\}_{n\in\mathbb{N}}$ is a $\| \|_{H^{\sim}}$ Cauchy sequence in $H^{\sim}$. Since $H^{\sim}$ is $\| \|_{H^{\sim}}$ complete and $S^{\sim}$ is onto, it follows that there exists $f \in H^{\sim}$ such that $S^{\sim}(x_n) \overset{\| \|_{H^{\sim}}}{\to} f = S^{\sim}(x)$ for some $x \in H^{\sim}$. Since $S^{\sim}$ is an isometry, we have that $x_n \overset{\| \|_H}{\to} x$, proving that H is $\| \|_H$ complete. ∎

A.5 Self-duality

Theorem 6 *Let A be finite dimensional and H a pre-Hilbert A-module. The following statements are equivalent:*

(i) H is $\| \|_H$ complete, that is, H is a Hilbert A-module;

(ii) H is $\| \|_{\bar{\varphi}}$ complete;

(iii) H is self-dual.

Proof. Since A is finite dimensional, it admits a trace $\bar{\varphi}$.

(i) *implies* (ii). By Proposition 6 and since A is finite dimensional, $\| \|_{\bar{\varphi}}$ and $\| \|_H$ are equivalent. It follows that H is $\| \|_{\bar{\varphi}}$ complete.

(ii) *implies* (iii). By Corollary 2 and since H is $\| \|_{\bar{\varphi}}$ complete, it follows that H is a Hilbert space with inner product $\langle \; , \; \rangle_{\bar{\varphi}}$. Consider $f : H \to A$ which is A-linear and bounded. In particular, by the proof of Proposition 6, we have that there exists $M > 0$ such that

$$\|f(x)\|_A \leq M\|x\|_H \leq M\|x\|_{\bar{\varphi}} \qquad \forall x \in H.$$

We can conclude that $f : H \to A$ is linear and $\| \|_{\bar{\varphi}} - \| \|_A$ continuous. Consider the linear functional $l = \bar{\varphi} \circ f$. Since $\bar{\varphi}$ is $\| \|_A$ continuous and f is $\| \|_{\bar{\varphi}} - \| \|_A$ continuous, we have that l is $\| \|_{\bar{\varphi}}$ continuous. By the standard Riesz representation theorem, there exists (a unique) $y \in H$ such that $l(x) = \langle x, y \rangle_{\bar{\varphi}}$ for all $x \in H$. It follows that

$$\bar{\varphi}(f(x) - \langle x, y \rangle_H) = \bar{\varphi}(f(x)) - \bar{\varphi}(\langle x, y \rangle_H) = l(x) - \langle x, y \rangle_{\bar{\varphi}} = 0 \qquad \forall x \in H. \quad (10)$$

Fix $\bar{x} \in H$. Define $a = (f(\bar{x}) - \langle \bar{x}, y \rangle_H)^* \in A$. By (10), we have that

$$\begin{aligned} 0 &= \bar{\varphi}(f(a \cdot \bar{x}) - \langle a \cdot \bar{x}, y \rangle_H) = \bar{\varphi}(af(\bar{x}) - a \langle \bar{x}, y \rangle_H) \\ &= \bar{\varphi}(a(f(\bar{x}) - \langle \bar{x}, y \rangle_H)) = \bar{\varphi}(aa^*). \end{aligned}$$

Since $\bar{\varphi}$ is a trace and $aa^* \geq 0$, this implies that $aa^* = 0$, that is, $\|a^*\|_A^2 = \|a\|_A^2 = \|aa^*\|_A = 0$. We can conclude that $f(\bar{x}) - \langle \bar{x}, y \rangle_H = a^* = 0$. Since $\bar{x}$ was arbitrarily chosen, it follows that $f(x) = \langle x, y \rangle_H$ for all $x \in H$, proving that H is self-dual.

(iii) implies (i). By point 3. of Proposition 8, it follows that H is $\| \|_H$ complete. ■

The implication *(ii) implies (iii)* can be found in Goldstine and Horwitz [12] although few mathematical differences are present. Namely, Goldstine and Horwitz use a different norm over A. The characterization of self-duality, that is, the remaining implications, to the best of our knowledge is novel. A similar observation holds for the implication *(i) implies (ii)* of Proposition 9.

A.5.1 Orthogonal decompositions

Pre-Hilbert modules behave very much like Hilbert spaces also in terms of orthogonal decompositions. Consider a pre-Hilbert A-module H and let $M \subseteq H$. Define

$$M^\perp = \{x \in H : \langle x, y \rangle_H = 0 \quad \forall y \in M\}.$$

It is immediate to prove that $M^\perp$ is a submodule.[45] It is also immediate to show that $M \cap M^\perp = \{0\}$ and that $M^{\perp\perp} \supseteq M$ where

$$M^{\perp\perp} = \left(M^\perp\right)^\perp = \left\{y \in H : \langle x, y \rangle_H = 0 \quad \forall x \in M^\perp\right\}.$$

Before stating our result on orthogonal decompositions, we need an ancillary fact.

Lemma 1 *Let H be a pre-Hilbert A-module. If $M \subseteq H$, then $M^\perp$ is $\| \|_H$ closed.*

[45] A subset N of H is a submodule if and only if for each $a, b \in A$ and $x, y \in N$

$$a \cdot x + b \cdot y \in N.$$

Proof. Fix $z \in H$ and define $\ker\{z\} = \{x \in H : \langle x, z \rangle_H = 0\}$. Consider a sequence $\{x_n\}_{n\in\mathbb{N}} \subseteq \ker\{z\}$ such that $x_n \overset{\| \|_H}{\to} x$. Since $S^{\sim}(z)$ is $\| \|_H - \| \|_A$ continuous, it follows that $S^{\sim}(z)(x) = 0$, that is, $\ker\{z\}$ is closed. Since $M^{\perp} = \bigcap_{y\in M} \ker\{y\}$, the statement follows. ■

Proposition 9 *Let A be finite dimensional and H a pre-Hilbert A-module. If H is self-dual and M is a submodule of H, then the following statements are equivalent:*

(i) M is $\| \|_H$ closed;

(ii) $H = M \oplus M^{\perp}$;

(iii) $M = M^{\perp\perp}$.

Proof. (i) *implies* (ii). Clearly, $M \oplus M^{\perp} \subseteq H$. We next prove the opposite inclusion. Since M is a submodule of H, if we define $\langle \ , \ \rangle_M$ as the restriction of $\langle \ , \ \rangle_H$ to $M \times M$, then $(M, +, \cdot, \langle \ , \ \rangle_M)$ is a pre-Hilbert A-module. It is immediate to see that $\| \|_M = \| \|_H$ once the latter is restricted to M. Since M is $\| \|_H$ closed, it follows that M is $\| \|_M$ complete and is itself self-dual. Fix $y \in H$. The map defined on M by $x \mapsto \langle x, y \rangle_H$ is A-linear and bounded. Since M is self dual, it follows that there exists a unique $y_1 \in M$ such that

$$\langle x, y_1 \rangle_H = \langle x, y_1 \rangle_M = \langle x, y \rangle_H \qquad \forall x \in M.$$

Define $y_2 = y - y_1$. It follow that

$$\langle x, y - y_1 \rangle_H = 0 \qquad \forall x \in M,$$

that is, $y_2 \in M^{\perp}$. It is also immediate to see that $y_1 + y_2 = y$. Since y was arbitrarily chosen, we can conclude that $H \subseteq M \oplus M^{\perp}$.

(ii) *implies* (iii). Since $M \subseteq M^{\perp\perp}$, we only need to prove the opposite inclusion. By assumption, if $x \in M^{\perp\perp}$, then there exists $x_M \in M$ and $x_{M^{\perp}} \in M^{\perp}$ such that $x = x_M + x_{M^{\perp}}$. Since $M \subseteq M^{\perp\perp}$, we have that $M^{\perp} \ni x_{M^{\perp}} = x - x_M \in M^{\perp\perp}$. Since $M^{\perp} \cap M^{\perp\perp} = \{0\}$, this implies that $x - x_M = 0$, that is, $x = x_M \in M$, proving the opposite inclusion and the statement.

(iii) *implies* (i). By Lemma 1 and since $M = M^{\perp\perp} = \left(M^{\perp}\right)^{\perp}$, it follows that M is $\| \|_H$ closed. ■

We conclude with a last piece of notation given $M, N \subseteq H$ we write $M \perp N$ if and only if $\langle x, y \rangle_H = 0$ for all $x \in M$ and $y \in N$. Clearly, we have that $M \perp M^{\perp}$ for all $M \subseteq H$.

Proposition 9 allows us to define the (orthogonal) projection of an element $x \in H$ on a $\| \|_H$ closed submodule M.

Definition 7 *Let A be finite dimensional, H a Hilbert A-module and $M \subseteq H$ a $\| \|_H$ closed submodule. We call* projection on M *the linear map $\mathcal{P}_M : H \to M$ such that, for any $x \in H$,*

$$\mathcal{P}_M x = x_M,$$

where $x_M \in M$ and $x_{M^\perp} \in M^\perp$ are the unique elements that satisfy $x = x_M + x_{M^\perp}$.

Given $x \in H$ and $y \in M$, $y = \mathcal{P}_M x$ if and only if $\langle x - y, z \rangle_H = 0$ for all $z \in M$. Moreover, $\mathcal{P}_M x$ minimizes the distance between x and the submodule M because, for any $z \in M$,

$$\|x - z\|_H^2 = \|(x - \mathcal{P}_M x) + (\mathcal{P}_M x - z)\|_H^2 = \|x - \mathcal{P}_M x\|_H^2 + \|\mathcal{P}_M x - z\|_H^2 .$$

A.6 The Abstract Wold Theorem for Hilbert A-modules

In this section we prove an abstract version for Hilbert modules of the Wold Decomposition Theorem.[46] It is important to observe that the properties of self-duality and complementability (see Theorem 6 and Proposition 9) are fundamental in allowing us to follow the proof strategy used for Hilbert spaces. We say that $T : H \to H$ is an isometry if and only if T is A-linear and such that

$$\langle T(x), T(y) \rangle_H = \langle x, y \rangle_H \qquad \forall x, y \in H. \tag{11}$$

Note that an isometry in this sense satisfies the usual property

$$\|T(x)\|_H = \|x\|_H \qquad \forall x \in H. \tag{12}$$

It is immediate to prove that for each $n \in \mathbb{N}_0$ the iterate T^n satisfies (11) and (12).[47] In particular, by Abramovich and Aliprantis [1, Theorem 2.5], if H is $\| \|_H$ complete, $T^n(H)$ is a $\| \|_H$ closed submodule of H.

Definition 8 *Let $T : H \to H$ be an isometry. We say that a submodule L is wandering if and only if for all $m, n \in \mathbb{N}_0$ such that $m \neq n$*

$$T^n(L) \perp T^m(L).$$

Lemma 2 *Let $T : H \to H$ be an isometry. If H is self-dual, then the following statements are true:*

1. *If M is $\| \|_H$ closed, so is $T(M)$.*

[46] See Sz.-Nagy, Foias, Bercovici and Kérchy [24].

[47] Recall that $T^0 = I$.

2. *If $L = T(H)^{\perp}$, then L is wandering.*

3. *If $L = T(H)^{\perp}$, then for each $n \in \mathbb{N}_0$*

$$T^n(H) = T^n(L) \oplus T^{n+1}(H) \ \textit{ and } T^n(L) \perp T^{n+1}(H).$$

4. *If $L = T(H)^{\perp}$, then for each $k \in \mathbb{N}_0$*

$$\bigoplus_{n=0}^{k} T^n(L) = T^{k+1}(H)^{\perp}.$$

Proof. 1. Since T is A-linear, that is, for each $a, b \in A$ and each $x, y \in H$

$$T(a \cdot x + b \cdot y) = a \cdot T(x) + b \cdot T(y),$$

we have that T is linear. By the proof of Abramovich and Aliprantis [1, Theorem 2.5] and since T satisfies (11), we have that $T(M)$ is closed.

2. Observe that $T^n(H) \subseteq H$ for all $n \in \mathbb{N}_0$. It follows that $T^n(H) \subseteq T(H)$ for all $n \in \mathbb{N}$. Since $L \subseteq H$, it also follows that $T^n(L) \subseteq T^n(H) \subseteq T(H)$ for all $n \in \mathbb{N}$. Since $T(H) \perp L$, this implies that $T^n(L) \perp L$ for all $n \in \mathbb{N}$. Next, consider $m, n \in \mathbb{N}_0$ such that $m \neq n$. Wlog, assume that $n > m$. By the previous part of the proof, we have that $T^{n-m}(L) \perp L$. By (11), we can conclude that $T^n(L) \perp T^m(L)$.

3. We proceed by induction.

Initial Step. $n = 0$. By definition of L, point 1, and Proposition 9 and since H is self dual, we have that L is a $\| \, \|_H$ closed submodule and

$$T^n(H) = H = L \oplus L^{\perp} = L \oplus T(H) = T^n(L) \oplus T^{n+1}(H)$$

proving the Step.

Inductive Step. Assume the statement is true for n. By assumption, it follows that $T^n(H) = T^n(L) \oplus T^{n+1}(H)$ and $T^n(L) \perp T^{n+1}(H)$. By (11) we have that

$$T^{n+1}(L) \perp T^{n+2}(H) \tag{13}$$

as well as

$$\begin{aligned} T^{n+1}(H) &= T(T^n(H)) = T\left(T^n(L) \oplus T^{n+1}(H)\right) \\ &= T(T^n(L)) + T\left(T^{n+1}(H)\right) \\ &= T(T^n(L)) \oplus T\left(T^{n+1}(H)\right) \end{aligned}$$

where the last equality follows from (13).

The statement follows by induction.

4. We proceed by induction.

Initial Step. $k = 0$. By definition of L,

$$\bigoplus_{n=0}^{k} T^n(L) = T^0(L) = L = T(H)^{\perp} = T^{k+1}(H)^{\perp}.$$

Inductive Step. Assume the statement is true for k. By assumption, it follows that $\bigoplus_{n=0}^{k} T^n(L) = T^{k+1}(H)^{\perp}$. By Proposition 9 and since H is self-dual and since $T^{k+1}(H)$ is a $\| \|_H$ closed submodule, this implies that

$$H = T^{k+1}(H) \oplus T^{k+1}(H)^{\perp} = \bigoplus_{n=0}^{k} T^n(L) \oplus T^{k+1}(H).$$

At the same time, by point 3., we also have that $T^{k+1}(H) = T^{k+1}(L) \oplus T^{k+2}(H)$ and $T^{k+1}(L) \perp T^{k+2}(H)$. We can conclude that

$$\bigoplus_{n=0}^{k+1} T^n(L) \oplus T^{k+2}(H) = H \text{ and } \bigoplus_{n=0}^{k+1} T^n(L) = T^{k+2}(H)^{\perp}.$$

The statement follows by induction. ∎

Theorem 7 (Abstract Wold Theorem for Hilbert A-modules) *Let $T : H \to H$ be an isometry. If H is self-dual, then $H = \hat{H} \oplus \tilde{H}$ where*

$$\hat{H} = \bigcap_{n=0}^{\infty} T^n(H), \qquad \tilde{H} = \bigoplus_{n=0}^{\infty} T^n(L), \qquad L = T(H)^{\perp}.$$

Moreover, the submodules orthogonal decomposition, $\left(\hat{H}, \tilde{H}\right)$, of H is the unique submodules decomposition such that $T\left(\hat{H}\right) = \hat{H}$ and $\tilde{H} = \bigoplus_{n=0}^{\infty} T^n(L)$ given a wandering set L.

Proof. Define $L = T(H)^{\perp}$. Define also $M_k = \bigoplus_{n=0}^{k} T^n(L)$ for all $k \in \mathbb{N}_0$, $\tilde{H} = \bigoplus_{n=0}^{\infty} T^n(L)$,[48] and $\hat{H} = \tilde{H}^{\perp}$. It is immediate to see that $\hat{H}$ and $\tilde{H}$ are two $\| \|_H$ closed submodules. Note that $M_k \subseteq M_{k+1}$ for all $k \in \mathbb{N}_0$ and $\tilde{H} = cl_{\| \|_H}\left(\bigcup_{k \in \mathbb{N}_0} M_k\right)$. By construction, we have that $\hat{H} \perp M_k$ for all $k \in \mathbb{N}_0$. By Lemma 2, it follows that $M_k = T^{k+1}(H)^{\perp}$ for all $k \in \mathbb{N}$. By Proposition 9, this implies that if $x \in \hat{H}$, then $x \in M_k^{\perp} = T^{k+1}(H)$ for all $k \in \mathbb{N}_0$. We can conclude that $x \in \bigcap_{n=1}^{\infty} T^n(H) \cap H = \bigcap_{n=0}^{\infty} T^n(H)$. Vice versa, since $M_k = T^{k+1}(H)^{\perp}$ for all $k \in \mathbb{N}$, if $x \in \bigcap_{n=0}^{\infty} T^n(H)$, then

$$\langle x, y \rangle_H = 0 \qquad \forall y \in \bigcup_{n=0}^{\infty} M_n.$$

[48] With the notation $\bigoplus_{n=0}^{\infty} T^n(L)$, we mean the $\| \|_H$ closure of the set $\bigcup_{k \in \mathbb{N}_0} M_k$.

Since $cl_{\| \|_H}\left(\bigcup_{n=0}^{\infty} M_n\right) = \tilde{H}$, this implies that $x \in \tilde{H}^{\perp} = \hat{H}$. In other words, we proved that $\hat{H} = \bigcap_{n=0}^{\infty} T^n(H)$.

We next prove uniqueness. Since $T^{n+1}(H) \subseteq T^n(H) \subseteq H$ for all $n \in \mathbb{N}_0$, it follows that

$$T\left(\hat{H}\right) = T\left(\bigcap_{n=0}^{\infty} T^n(H)\right) = \bigcap_{n=0}^{\infty} T\left(T^n(H)\right) = \bigcap_{n=1}^{\infty} T^n(H) = \bigcap_{n=1}^{\infty} T^n(H) \cap H = \hat{H}.$$

Assume that $\left(\hat{H}', \tilde{H}'\right)$ is another decomposition. Consider the wandering set L' generating $\tilde{H}'$. By construction and since L' is wandering, we have that $L' \perp T\left(\tilde{H}'\right)$ and $L' \oplus T\left(\tilde{H}'\right) = \tilde{H}'$. By construction and (11), this implies that

$$L = T(H)^{\perp} = T\left(\hat{H}' \oplus \tilde{H}'\right)^{\perp} = \left(\hat{H}' \oplus T\left(\tilde{H}'\right)\right)^{\perp} = L',$$

proving the statement. ■

A.7 Adjoints

Given a pre-Hilbert A-module, we define by $B^{\sim}(H)$ the collection of all bounded A-linear operators. In other words, $T \in B^{\sim}(H)$ if and only if

$$T(a \cdot x + b \cdot y) = a \cdot T(x) + b \cdot T(y) \qquad \forall a, b \in A, \forall x, y \in H$$

and there exists $M > 0$ such that

$$\|T(x)\|_H \leq M \|x\|_H \qquad \forall x \in H.$$

Since any A-linear operator is linear, we have that $B^{\sim}(H) \subseteq B(H)$ where the latter is the set of all bounded and linear operators from H to H.

Given $T \in B^{\sim}(H)$, we define the adjoint of T, denoted by T^*, to be such that

$$\langle T(x), y \rangle_H = \langle x, T^*(y) \rangle_H \qquad \forall x, y \in H. \tag{14}$$

The next result shows that adjoints are well defined and all the properties that hold for Hilbert spaces are satisfied once suitably adjusted to Hilbert modules. It is immediate to see that $B^{\sim}(H)$ is a vector subspace of $B(H)$. Note also that if $S, T \in B^{\sim}(H)$, then the composition of S with T is also in $B^{\sim}(H)$.[49]

[49]Note that $S, T \in B^{\sim}(H) \subseteq B(H)$, thus $ST \in B(H)$. We only need to prove A-linearity. We have that for each $a, b \in A$ and $x, y \in H$

$$\begin{aligned}(ST)(a \cdot x + b \cdot y) &= S(T(a \cdot x + b \cdot y)) = S(a \cdot T(x) + b \cdot T(y)) \\ &= a \cdot S(T(x)) + b \cdot S(T(y)) = a \cdot (ST)(x) + b \cdot (ST)(y),\end{aligned}$$

proving A-linearity.

Proposition 10 *Let H be a self-dual pre-Hilbert A-module. The following statements are true:*

1. $^* : B^{\sim}(H) \to B^{\sim}(H)$ *is a well defined, injective, and linear;*
2. $T^{**} = T$ *for all* $T \in B^{\sim}(H)$;
3. $^* : B^{\sim}(H) \to B^{\sim}(H)$ *is a surjective;*
4. $\|T\| = \|T^*\|$ *for all* $T \in B^{\sim}(H)$;
5. $\|ST\| \leq \|S\| \|T\|$ *for all* $S, T \in B^{\sim}(H)$;
6. $\|T^*T\| = \|TT^*\| = \|T\|^2$ *for all* $T \in B^{\sim}(H)$;
7. $(ST)^* = T^*S^*$ *for all* $S, T \in T \in B^{\sim}(H)$;
8. *For all* $T \in B^{\sim}(H)$, $\ker(T^*) = T(H)^{\perp}$, *where*

$$\ker(T^*) = \{x \in H : \quad T^*(x) = 0\}.$$

Proof. 1. Consider $T \in B^{\sim}(H)$. Fix $y \in H$. Since T is A-linear and bounded, note that the element y induces a bounded A-linear operator on H to A via the map

$$x \mapsto \langle T(x), y \rangle_H \qquad \forall x \in H.$$

Since H is self-dual, there exists a unique $z_y \in H$ such that

$$\langle T(x), y \rangle_H = \langle x, z_y \rangle_H \qquad \forall x \in H.$$

We define $T^* : H \to H$ to be such that $T^*(y) = z_y$. It follows that T^* is well defined and satisfies (14). Next, observe that for each $y_1, y_2 \in H$ and $\alpha_1, \alpha_2 \in \mathbb{R}$

$$\begin{aligned}
\langle x, T^*(\alpha_1 y_1 + \alpha_2 y_2) \rangle_H &= \langle x, z_{\alpha_1 y_1 + \alpha_2 y_2} \rangle_H = \langle T(x), \alpha_1 y_1 + \alpha_2 y_2 \rangle_H \\
&= \alpha_1 \langle T(x), y_1 \rangle_H + \alpha_2 \langle T(x), y_2 \rangle_H \\
&= \alpha_1 \langle x, z_{y_1} \rangle_H + \alpha_2 \langle x, z_{y_2} \rangle_H = \langle x, \alpha_1 z_{y_1} + \alpha_2 z_{y_2} \rangle_H \\
&= \langle x, \alpha_1 T^*(y_1) + \alpha_2 T^*(y_2) \rangle_H \qquad \forall x \in H,
\end{aligned}$$

yielding that T^* is linear. Finally, note that

$$\begin{aligned}
\|T^*\| &= \sup_{\|y\|_H = 1} \|T^*(y)\|_H = \sup_{\|y\|_H = 1} \left(\sup_{\|x\|_H = 1} \|\langle x, T^*(y) \rangle_H\|_A \right) \\
&= \sup_{\|y\|_H = 1} \left(\sup_{\|x\|_H = 1} \|\langle T(x), y \rangle_H\|_A \right) \leq \sup_{\|y\|_H = 1} \left(\sup_{\|x\|_H = 1} \|T(x)\|_H \|y\|_H \right) \\
&\leq \sup_{\|y\|_H = 1} \left(\sup_{\|x\|_H = 1} \|T\| \|x\|_H \|y\|_H \right) \leq \|T\|,
\end{aligned}$$

proving that $T^* \in B^\sim(H)$ and * is well defined. Next, fix $y \in H$. Consider $S, T \in B^\sim(H)$ and $\alpha, \beta \in \mathbb{R}$. Observe that

$$\begin{aligned}\langle x, (\alpha S + \beta T)^* (y)\rangle_H &= \langle (\alpha S + \beta T)(x), y\rangle_H = \langle \alpha S(x) + \beta T(x), y\rangle_H \\ &= \alpha \langle S(x), y\rangle_H + \beta \langle T(x), y\rangle_H = \alpha \langle x, S^*(y)\rangle_H + \beta \langle x, T^*(y)\rangle_H \\ &= \langle x, \alpha S^*(y) + \beta T^*(y)\rangle_H = \langle x, (\alpha S^* + \beta T^*)(y)\rangle_H \qquad \forall x \in H.\end{aligned}$$

We can conclude that $(\alpha S + \beta T)^*(y) = (\alpha S^* + \beta T^*)(y)$. Since y was arbitrarily chosen, we can conclude that $(\alpha S + \beta T)^* = (\alpha S^* + \beta T^*)$, that is, * is linear. Next, fix $x \in H$ and assume that $T^* = S^*$. It follows that

$$\langle T(x), y\rangle_H = \langle x, T^*(y)\rangle_H = \langle x, S^*(y)\rangle_H = \langle S(x), y\rangle_H \qquad \forall y \in H.$$

We can conclude that $T(x) = S(x)$. Since x was arbitrarily chosen, we can conclude that $T = S$, that is, * is injective.

2. Fix $x \in H$. By definition of T^* and T^{**}, we have that

$$\langle T(x), y\rangle_H = \langle x, T^*(y)\rangle_H = \langle T^*(y), x\rangle_H^* = \langle y, T^{**}(x)\rangle_H^* = \langle T^{**}(x), y\rangle_H \qquad \forall y \in H.$$

We can conclude that $T(x) = T^{**}(x)$. Since x was arbitrarily chosen, we can conclude that $T = T^{**}$.

3. Consider $S \in B^\sim(H)$ and consider $T = S^*$. By point 2, it follows that $T^* = S^{**} = S$, that is, * is surjective.

4. By the proof of point 1, we have that

$$\|T^*\| \leq \|T\| \qquad \forall T \in B^\sim(H).$$

In particular, we have that $\|T^{**}\| \leq \|T^*\| \leq \|T\|$ for all $T \in B^\sim(H)$. By point 2, we can conclude that $\|T\| \leq \|T^*\| \leq \|T\|$ for all $T \in B^\sim(H)$, proving the statement.

5. Consider $S, T \in B^\sim(H)$. We have that

$$\|ST\| = \sup_{\|y\|_H = 1} \|S(T(y))\|_H \leq \sup_{\|y\|_H = 1} (\|S\| \|T(y)\|_H) \leq \|S\| \sup_{\|y\|_H = 1} (\|T(y)\|_H) \leq \|S\| \|T\|,$$

proving the statement.

6. Consider $S \in B^\sim(H)$. By points 4 and 5, observe that

$$\begin{aligned}\|S\|^2 &= \sup_{\|x\|_H = 1} \|S(x)\|_H^2 = \sup_{\|x\|_H = 1} \|\langle S(x), S(x)\rangle_H\|_A \\ &= \sup_{\|x\|_H = 1} \|\langle x, S^*(S(x))\rangle_H\|_A \leq \sup_{\|x\|_H = 1} \|x\|_H \|S^*(S(x))\|_H \\ &= \sup_{\|x\|_H = 1} \|(S^*S)(x)\|_H = \|S^*S\| \leq \|S^*\| \|S\| = \|S\|^2,\end{aligned}$$

yielding that $\|S^*S\| = \|S\|^2$. If we choose $S = T$, then $\|T^*T\| = \|T\|^2$. If we choose $S = T^*$, then $\|TT^*\| = \|T^*\|^2 = \|T\|^2$.

7. Consider $S, T \in B^\sim(H)$. Fix $y \in H$. We have that, for any $x \in H$,

$$\begin{aligned}\langle x, (ST)^*(y)\rangle_H &= \langle (ST)(x), y\rangle_H = \langle S(T(x)), y\rangle_H \\ &= \langle T(x), S^*(y)\rangle_H = \langle x, T^*(S^*(y))\rangle_H .\end{aligned}$$

It follows that $(ST)^*(y) = T^*(S^*(y))$. Since y was arbitrarily chosen, it follows that $(ST)^*(y) = T^*(S^*(y))$ for all $y \in H$, that is, $(ST)^* = T^*S^*$.

8. First we show that $\ker(T^*)$ is included in $T(H)^\perp$. Equivalently, we prove that each $y^* \in \ker(T^*)$ is orthogonal to any $y \in T(H)$. Note that $T^*(y^*) = 0$ and that $y = T(x)$ for some $x \in H$. By the definition of adjoint operator,

$$\langle y, y^*\rangle_H = \langle T(x), y^*\rangle_H = \langle x, T^*(y^*)\rangle_H = 0,$$

proving the orthogonality of y^* and y. Conversely, consider any $y^* \in T(H)^\perp$, that is $\langle T(x), y^*\rangle_H = 0$ for all $x \in H$. Since T^* is the adjoint operator, $\langle x, T^*(y^*)\rangle_H = 0$ for all $x \in H$ and this ensures that $T^*(y^*) = 0$. ■

Point 1 can also be found in Goldstine and Horwitz [12]. Also in this case, there is a technical difference in terms of norm used over A.

A.8 Ancillary results

Proposition 11 *Let H be a pre-Hilbert A-module. The following statements are true:*

1. $\langle z, x + y\rangle_H = \langle z, x\rangle_H + \langle z, y\rangle_H$ *for all* $x, y, z \in H$;
2. $\langle x, a \cdot y\rangle_H = \langle x, y\rangle_H a^*$ *for all* $a \in A$ *and for all* $x, y \in H$;
3. $\langle x, \alpha y\rangle_H = \alpha \langle x, y\rangle_H$ *for all* $\alpha \in \mathbb{R}$ *for all* $x, y \in H$.

Proof. 1. Consider $x, y, z \in H$. We have that

$$\begin{aligned}\langle z, x + y\rangle_H &= \langle x + y, z\rangle_H^* = (\langle x, z\rangle_H + \langle y, z\rangle_H)^* \\ &= \langle x, z\rangle_H^* + \langle y, z\rangle_H^* = \langle z, x\rangle_H + \langle z, y\rangle_H ,\end{aligned}$$

proving the point.

2. Consider $x, y \in H$ and $a \in A$. We have that

$$\langle x, a \cdot y\rangle_H = \langle a \cdot y, x\rangle_H^* = (a \langle y, x\rangle_H)^* = \langle y, x\rangle_H^* a^* = \langle x, y\rangle_H a^*,$$

proving the point.

3. Consider $x, y \in H$ and $\alpha \in \mathbb{R}$. We have that

$$\langle x, \alpha y \rangle_H = \langle x, (\alpha e) \cdot y \rangle_H = \langle x, y \rangle_H (\alpha e)^* = \alpha \langle x, y \rangle_H ,$$

proving the point. ■

Proposition 12 *Let H be a self-dual pre-Hilbert A-module. If M, N, P, Q are four $\| \|_H$ closed submodules such that*

$$H = M \oplus N, \ N = P \oplus Q, \ N = M^\perp, \text{ and } P \perp Q,$$

then $Q^\perp = M \oplus P$.

Proof. Note that

$$H = M \oplus P \oplus Q. \tag{15}$$

Consider $x \in M$, $y \in P$, and $z \in Q$. Since $x \in M$ and $z \in Q \subseteq N$ and $N = M^\perp$, we have that $\langle x, z \rangle_H = 0$. Since $y \in P$, $z \in Q$ and $P \perp Q$, we have that $\langle y, z \rangle_H = 0$. It follows that

$$\langle x + y, z \rangle_H = \langle x, z \rangle_H + \langle y, z \rangle_H = 0.$$

Since z was arbitrarily chosen, it follows that $\langle x + y, z \rangle_H = 0$ for all $z \in Q$, that is, $x + y \in Q^\perp$. Since $x \in M$ and $y \in P$ were arbitrarily chosen, we have that $M \oplus P \subseteq Q^\perp$. Next, consider $y \in Q^\perp$. By (15), we have that there exist $x_1 \in M \oplus P$ and $x_2 \in Q$ such that $y = x_1 + x_2$. Since $M \oplus P \subseteq Q^\perp$, it follows that $Q^\perp \ni y - x_1 = x_2 \in Q$, proving that $y - x_1 = 0$, that is, $y = x_1 \in M \oplus P$. Since $y \in Q^\perp$ was arbitrarily chosen, we can conclude the opposite inclusion $Q^\perp \subseteq M \oplus P$. ■

B Hilbert A-modules for multivariate time series

We describe the special Hilbert A-modules that we employ for the treatment of random vectors. We first introduce the Hilbert A-module $H = L^2(\mathbb{R}^m, \Omega, \mathcal{F}, \mathbb{P})$ that generalizes the space of square-integrable random vectors, allowing for matrix coefficients. Then, we review the main definitions of weakly stationary vector processes and we describe the submodules of H induced by them.

B.1 The Hilbert A-module $L^2(\mathbb{R}^m, \Omega, \mathcal{F}, \mathbb{P})$

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we consider the vector space $L^2(\mathbb{R}^m, \Omega, \mathcal{F}, \mathbb{P})$ of measurable square-integrable random vectors x that take value in $\mathbb{R}^m$.[50] We build on

[50] For any $i = 1, \dots, m$ the component x_i belongs to $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

$L^2(\mathbb{R}^m, \Omega, \mathcal{F}, \mathbb{P})$ the structure of Hilbert A-module and we denote it by H. We consider the algebra $A = \mathbb{R}^{m\times m}$ of real $m \times m$ matrices. The unit in A is the identity matrix and the product in A is the usual row-by-column product. A is normed by the operator norm $\| \ \|_A$ such that, for any $a = \{a_{i,j}\}_{i,j=1,\ldots,m}$ in A,

$$\|a\|_A = \sup_{x\in\mathbb{R}^m, \|x\|_2=1} \|ax\|_2,$$

where $\| \ \|_2$ is the L^2 norm in $\mathbb{R}^m$. In the algebra A, the involution that associates any matrix a with its transposed a' is defined. Such operation induces an order $\geq$ on the convex cone of symmetric positive semidefinite matrices. For any $a, b \in A$, $a \geq b$ when the matrix $a - b$ is symmetric and positive semidefinite (equivalently, $a - b \geq \mathbf{0}$).

Moreover, we define the *trace* functional $\bar{\varphi} : A \longrightarrow \mathbb{R}$ by setting, for any matrix a,

$$\bar{\varphi}(a) = \mathrm{Tr}(a) = \sum_{i=1}^{m} a_{i,i}.$$

$\bar{\varphi}$ satisfies the properties:[51]

- $a \geq \mathbf{0} \quad \Longrightarrow \quad \bar{\varphi}(a) \geq \mathbf{0}$,
- $a \geq \mathbf{0}, \quad \bar{\varphi}(a) = 0 \quad \Longleftrightarrow \quad a = \mathbf{0}$,
- $\bar{\varphi}(a) = \bar{\varphi}(a') \quad \forall a \in A$,
- $\|a\|_A \leq \bar{\varphi}(a) \leq m\|a\|_A \quad \forall a \geq \mathbf{0}$.

The outer product $A \times H \longrightarrow H$ is the standard matrix-by-vector product. This operation makes H an A-module. Note that the natural structure of real vector space of H is kept because of the relation

$$\lambda x = (\lambda I)x \qquad \forall x \in H, \quad \lambda \in \mathbb{R},$$

where I is the identity matrix.

[51]Indeed, the trace of a symmetric positive semidefinite matrix a is nonnegative and it equals zero just when a is null. Moreover, $\mathrm{Tr}(a) = \mathrm{Tr}(a')$ and, by denoting $\{z_i\}_{i=1}^m$ an orthonormal basis of eigenvectors of a (any z_i is an eigenvector of the nonnegative eigenvalue λ_i), we have

$$\bar{\varphi}(a) = \mathrm{Tr}(a) = \sum_{i=1}^{m} \lambda_i \|z_i\|_2 = \sum_{i=1}^{m} \|az_i\|_2 \leqslant m\|a\|_A.$$

In addition, for any vector $x = \sum_i \mu_i z_i$ with $\|x\|_2^2 = \sum_i \mu_i^2 = 1$,

$$\|ax\|_2 = \left\| \sum_{i=1}^{m} \mu_i a z_i \right\|_2 \leq \sum_{i=1}^{m} |\mu_i| \, \|az_i\|_2 \leq \sum_{i=1}^{m} \|az_i\|_2 = \sum_{i=1}^{m} \lambda_i \|z_i\|_2 = \mathrm{Tr}(a).$$

Consequently, $\|a\|_A \leq \bar{\varphi}(a)$.

We define the A-valued inner product $\langle\ ,\ \rangle_H : H \times H \longrightarrow A$ that associates any $x = [x_1, \ldots, x_m]'$, $y = [y_1, \ldots, y_m]' \in H$ with the matrix

$$\langle x, y\rangle_H = \mathbb{E}[xy'] = \{\mathbb{E}[x_i y_j]\}_{i,j=1,\ldots,m}.$$

$\langle\ ,\ \rangle_H$ satisfies, for any $x, y, z \in H$ and $M \in A$,

1. $\langle x, x\rangle_H \geq \mathbf{0}$ with equality if and only if $x = 0$;[52]
2. $\langle x, y\rangle_H = \langle y, x\rangle'_H$;
3. $\langle x + y, z\rangle_H = \langle x, z\rangle_H + \langle y, z\rangle_H$;
4. $\langle ax, y\rangle_H = a\langle x, y\rangle_H$.

As a result, H is a pre-Hilbert A-module with the above operations. A useful consequence of the previous properties is that

$$\langle ax, by\rangle_H = a\langle x, y\rangle_H b' \qquad \forall a, b \in A.$$

In addition,

$$\langle x, x\rangle_H = \mathbb{E}[xx'] = \{\mathbb{E}[x_i x_j]\}_{i,j=1,\ldots,m}$$

is the covariance matrix of x which is symmetric and positive semidefinite. Hence, $\mathbb{E}[xx']$ has a unique symmetric positive semidefinite square root matrix S such that $\langle x, x\rangle_H = SS$. Such matrix is positive definite in case $\mathbb{E}[xx']$ is.[53]

As $\bar{\varphi}$ is a trace functional, it is strictly positive. Therefore, H is also a pre-Hilbert space with the inner product $\langle\ ,\ \rangle_{\bar{\varphi}} : H \times H \longrightarrow \mathbb{R}$ defined by

$$\langle x, y\rangle_{\bar{\varphi}} = \bar{\varphi}(\langle x, y\rangle_H) = \mathrm{Tr}(\langle x, y\rangle_H) = \mathrm{Tr}(\mathbb{E}[xy']) = \sum_{i=1}^{m} \mathbb{E}[x_i y_i] \qquad \forall x, y \in H.$$

$\langle\ ,\ \rangle_{\bar{\varphi}}$ actually coincides with the usual inner product of $L^2(\mathbb{R}^m)$ and the associated norm $\|\ \|_{\bar{\varphi}} : H \longrightarrow [0, +\infty)$ is

$$\|x\|_{\bar{\varphi}} = \sqrt{\langle x, x\rangle_{\bar{\varphi}}} = \sqrt{\mathrm{Tr}(\mathbb{E}[xx'])} = \sqrt{\sum_{i=1}^{m} \mathbb{E}[x_i^2]} \qquad \forall x \in H.$$

[52] $\langle x, x\rangle_H \geq \mathbf{0}$ means that $\langle x, x\rangle_H$ is a symmetric semipositive definite matrix.

[53] See Horn and Johnson [15, Theorem 7.2.6] as reference. Note that an alternative factorization is provided by the Cholesky decomposition (see Trefethen and Bau III [25]). If $\mathbb{E}[xx']$ is positive definite, its Cholesky decomposition $\mathbb{E}[xx'] = LL'$ is unique and L is a lower triangular matrix which has positive entries on the diagonal. However, if $\mathbb{E}[xx']$ is only positive semidefinite, the decomposition is still possible (with zero entries on the diagonal), but it is not unique.

As A is finite dimensional, the norm $\| \ \|_H : H \longrightarrow [0, +\infty)$ defined by

$$\|x\|_H = \sqrt{\|\langle x, x\rangle_H\|_A} = \sqrt{\|\mathbb{E}\left[xx'\right]\|_A} \qquad \forall x \in H$$

is equivalent to $\| \ \|_{\bar{\varphi}}$.[54] In particular, observe that $\|x\|_H = \sqrt{\lambda_{max}}$, where λ_{max} is the largest eigenvalue of the covariance matrix of x, i.e. the one associated with the Principal Component of $\mathbb{E}\left[xx'\right]$ that explains the most variance.[55] In case any x_i is uncorrelated with any x_j with $i \neq j$, the covariance matrix $\mathbb{E}\left[xx'\right]$ is diagonal and so $\|x\|_H = \sqrt{\max_{i=1,\ldots,m} \mathbb{E}\left[x_i^2\right]}$.

In addition, an equivalent formulation of $\|x\|_H$ exploits the Rayleigh-Ritz ratio:[56]

$$\|x\|_H^2 = \max_{\|y\|_2=1} y'\mathbb{E}\left[xx'\right]y.$$

Proposition 13 *H is a Hilbert A-module.*

Proof. We already described that H is a pre-Hilbert A-module. We are just left to show that H is $\| \ \|_{\bar{\varphi}}$ complete. Thus, we consider a Cauchy sequence $\left\{x^{(n)}\right\}_n \subset H$, i.e. for any $\varepsilon > 0$ there exists $N > 0$ such that

$$\left\|x^{(n)} - x^{(m)}\right\|_{\bar{\varphi}}^2 = \sum_{i=1}^m \mathbb{E}\left[\left(x_i^{(n)} - x_i^{(m)}\right)^2\right] < \varepsilon^2 \qquad \forall n, m > N.$$

Therefore, for any component $i = 1, \ldots, m$, the sequence $\left\{x_i^{(n)}\right\}_n \subset L^2(\Omega, \mathcal{F}, \mathbb{P})$ satisfies the Cauchy condition. As $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is complete, there exists $x_i \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ such that

$$\mathbb{E}\left[\left(x_i^{(n)} - x_i\right)^2\right] < \frac{\varepsilon^2}{m} \qquad \forall n > N_i.$$

As a result, by defining $x = [x_1, \ldots, x_m]' \in H$, we have

$$\left\|x^{(n)} - x\right\|_{\bar{\varphi}}^2 = \sum_{i=1}^m \mathbb{E}\left[\left(x_i^{(n)} - x_i\right)^2\right] < \varepsilon^2 \qquad \forall n > \max_{i=1,\ldots,m} N_i$$

---

[54] See Proposition 6 in Appendix A.

[55] Indeed, let $\{z_i\}_{i=1}^m$ be an orthonormal basis of eigenvectors of $\mathbb{E}\left[xx'\right]$, where each z_i is an eigenvector of the nonnegative eigenvalue λ_i. Then, for any vector $y = \sum_i \mu_i z_i$ such that $\|y\|_2^2 = \sum_i \mu_i^2 = 1$,

$$\|\mathbb{E}\left[xx'\right]y\|_2 = \left\|\sum_{i=1}^m \mu_i \mathbb{E}\left[xx'\right]z_i\right\|_2 = \left\|\sum_{i=1}^m \mu_i \lambda_i z_i\right\|_2 = \sqrt{\sum_{i=1}^m \mu_i^2 \lambda_i^2}.$$

As a result,

$$\|x\|_H^2 = \|\mathbb{E}\left[xx'\right]\|_A = \sup_{\sum_i \mu_i^2 = 1} \sqrt{\sum_{i=1}^m \mu_i^2 \lambda_i^2} = \sqrt{\lambda_{max}^2} = \lambda_{max}$$

because the supremum is attained by associating the highest weight 1 with the largest eigenvalue λ_{max}. Hence, $\|x\|_H = \sqrt{\lambda_{max}}$.

[56] See Horn and Johnson [15].

and so H is $\| \|_{\bar{\varphi}}$ complete. ■

Hence, H is self-dual by Theorem 6 in Appendix A.

We summarize in Table 1 the connection between general Hilbert modules and the special case of $H = L^2(\mathbb{R}^m, \Omega, \mathcal{F}, \mathbb{P})$.

B.2 Weakly stationary multivariate time series

Let $\mathbf{x} = \{x_t\}_{t\in\mathbb{Z}}$ be a stochastic process that takes value in $\mathbb{R}^m$, that is $x_t = [x_{1,t}, \dots, x_{m,t}]'$ with $x_{i,t}$ measurable on $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 9 *The vectorial process* $\mathbf{x} = \{x_t\}_{t\in\mathbb{Z}}$ *is* weakly stationary *when, for any* t*,*

i) $\mathbb{E}\left[x_{i,t}^2\right]$ *is finite for all* $i = 1, \dots, m$*, i.e.* $x_t \in L^2(\mathbb{R}^m, \Omega, \mathcal{F}, \mathbb{P})$*,*

ii) $\mathbb{E}\left[x_t\right]$ *does not depend on* t*,*

iii) the cross moments matrix $\mathbb{E}\left[x_t x'_{t+k}\right]$ *depends at most on* k*, for any* $k \in \mathbb{Z}$*.*

With no loss of generality we assume that x_t has zero mean. When $\mathbf{x}$ is weakly stationary, the *autocovariance function* $\Gamma : \mathbb{Z} \longrightarrow A$ is well-defined: for any integer n, Γ_n is the $m \times m$ matrix such that $\Gamma_n = \left[\gamma_{i,j}(n)\right]_{i,j=1,\dots,m}$ with

$$\gamma_{i,j}(n) = \mathrm{Cov}\left(x_{i,t}, x_{j,t+n}\right) = \mathbb{E}\left[x_{i,t} x_{j,t+n}\right].$$

Note that $\gamma_{i,j}(n) = \gamma_{j,i}(-n)$, that is $\Gamma_n = \Gamma'_{-n}$.[57]

Definition 10 *A weakly stationary vectorial process* $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t\in\mathbb{Z}}$ *is a* multivariate white noise *if* $\mathbb{E}\left[\varepsilon_t\right] = 0$ *and* $\Gamma_n = \mathbf{0}$ *for any* $n \neq 0$*.*

The covariance matrix Γ_0 of $\boldsymbol{\varepsilon}$ is symmetric and positive semidefinite. We will also suppose that Γ_0 is positive definite (hence it has a positive definite square root), a requirement that parallels the *regularity* assumption in the univariate case.[58] In addition, we say that $\boldsymbol{\varepsilon}$ is a *unit variance* multivariate white noise when Γ_0 is the identity matrix. In this case, the components of ε_t are uncorrelated.

Given a white noise $\boldsymbol{\varepsilon}$, we consider the submodule $\mathcal{H}_t(\boldsymbol{\varepsilon})$ of H generated by the innovations ε_{t-k} with $k \in \mathbb{N}_0$:

$$\mathcal{H}_t(\boldsymbol{\varepsilon}) = \left\{ \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k} : \quad a_k \in A, \quad \left\| \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k} \right\|_{\bar{\varphi}} < +\infty \right\}.$$

[57] See Hamilton [14] as a reference.

[58] See Bierens [5].

General theory	Random vectors theory	Description
Algebra	$A = \mathbb{R}^{m\times m}$	real $m \times m$ matrices
multiplicative unit in A	$e = I$	identity matrix
product in A	$ab = \{\sum_{k=1}^{m} a_{i,k} b_{k,j}\}_{i,j}$	row-by-vector product
norm in A	$\\|a\\|_A = \sup_{x\in\mathbb{R}^m, \\|x\\|_2=1} \\|ax\\|_2$	operator norm
involution in A	$a^* = a'$	transposition of matrices
order in A	$a - b \geq \mathbf{0}$	$a - b$ symmetric and positive semidefinite
trace functional $\bar{\varphi} : A \to \mathbb{R}$	$\bar{\varphi} = \mathrm{Tr}(a) = \sum_{i=1}^{m} a_{i,i}$	trace of the matrix a
Hilbert A-module	$H = L^2(\mathbb{R}^m, \Omega, \mathcal{F}, \mathbb{P})$	square-integrable random vectors
outer product $\cdot : A \times H \to H$	$ax = [\ldots, \sum_{k=1}^{m} a_{i,k} x_k, \ldots]'$	matrix-by-vector product
inner product $\langle\ ,\ \rangle_H : H \times H \to A$	$\langle x, y\rangle_H = \mathbb{E}[xy'] = \{\mathbb{E}[x_i y_j]\}_{i,j}$	matrix of cross-covariances
inner product $\langle x, x\rangle_H$	$\langle x, x\rangle_H = \mathbb{E}[xx'] = \{\mathbb{E}[x_i x_j]\}_{i,j}$	covariance matrix of x
norm $\\|\ \\|_H$	$\\|x\\|_H = \sqrt{\lambda_{max}}$	square root of the largest eigenvalue of $\mathbb{E}[xx']$
orthogonality $\langle x, y\rangle_H = 0$	$\mathbb{E}[x_i y_j] = 0 \quad \forall i, j = 1, \ldots, m$	any x_i uncorrelated with any y_j
inner product $\langle\ ,\ \rangle_{\bar{\varphi}} : H \times H \to \mathbb{R}$	$\langle x, y\rangle_{\bar{\varphi}} = \mathrm{Tr}(\mathbb{E}[xy']) = \sum_{i=1}^{m} \mathbb{E}[x_i y_i]$	inner product of $L^2(\mathbb{R}^m)$
norm $\\|\ \\|_{\bar{\varphi}}$	$\\|x\\|_{\bar{\varphi}} = \sqrt{\mathrm{Tr}(\mathbb{E}[xx'])} = \sqrt{\sum_{i=1}^{m} \mathbb{E}[x_i^2]}$	norm of $L^2(\mathbb{R}^m)$ square root of the sum of eigenvalues of $\mathbb{E}[xx']$

Table 1: Relations between the general Hilbert module theory and the Hilbert A-module used to model random vectors.

Lemma 3 *The following equality holds:*

$$\left\| \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k} \right\|_{\bar{\varphi}}^2 = \sum_{k=0}^{+\infty} \mathrm{Tr}\left(a_k \Gamma_0 a_k'\right).$$

Proof.

$$\begin{aligned}\left\| \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k} \right\|_{\bar{\varphi}}^2 &= \mathrm{Tr}\left(\left\langle \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k}, \sum_{h=0}^{+\infty} a_h \varepsilon_{t-h} \right\rangle_H \right) = \mathrm{Tr}\left(\sum_{k=0}^{+\infty} \sum_{h=0}^{+\infty} a_k \left\langle \varepsilon_{t-k}, \varepsilon_{t-h} \right\rangle_H a_h' \right) \\ &= \mathrm{Tr}\left(\sum_{k=0}^{+\infty} a_k \Gamma_0 a_k' \right) = \sum_{k=0}^{+\infty} \mathrm{Tr}\left(a_k \Gamma_0 a_k'\right).\end{aligned}$$

■

We call *generators* of $\mathcal{H}_t(\boldsymbol{\varepsilon})$ the random vectors $\sum_k a_k \varepsilon_{t-k}$ that satisfy the previous summability requirement. From now on, $\boldsymbol{\varepsilon}$ is supposed to be a unit variance white noise, so that Γ_0 is the identity matrix and we rewrite $\mathcal{H}_t(\boldsymbol{\varepsilon})$ as

$$\mathcal{H}_t(\boldsymbol{\varepsilon}) = \left\{ \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k} : \quad a_k \in A, \quad \sum_{k=0}^{+\infty} \mathrm{Tr}\left(a_k a_k'\right) < +\infty \right\}.$$

In case $m = 1$ we retrieve the usual square-summability requirement for univariate time series.

Proposition 14 *$\mathcal{H}_t(\boldsymbol{\varepsilon})$ is a closed submodule of H.*

Proof. Consider $x \in H$ such that there exists a sequence $\left\{x^{(n)}\right\}_n \subset \mathcal{H}_t(\boldsymbol{\varepsilon})$ such that

$$\left\| x^{(n)} - x \right\|_{\bar{\varphi}} \longrightarrow 0.$$

We show that $x \in \mathcal{H}_t(\boldsymbol{\varepsilon})$, too. Observe that any $x^{(n)}$ can be written as

$$x^{(n)} = \sum_{k=0}^{+\infty} \left\langle x^{(n)}, \varepsilon_{t-k} \right\rangle_H \varepsilon_{t-k}$$

because, if $x^{(n)} = \sum_{k=0}^{\infty} a_k^{(n)} \varepsilon_{t-k}$ with $a_k^{(n)} \in A$, we have

$$\left\langle x^{(n)}, \varepsilon_{t-k} \right\rangle_H = \left\langle \sum_{h=0}^{+\infty} a_h^{(n)} \varepsilon_{t-h}, \varepsilon_{t-k} \right\rangle_H = \sum_{h=0}^{+\infty} a_h^{(n)} \left\langle \varepsilon_{t-h}, \varepsilon_{t-k} \right\rangle_H = a_k^{(n)}.$$

In addition, the limit x can be decomposed as

$$x = \sum_{k=0}^{+\infty} \left\langle x, \varepsilon_{t-k} \right\rangle_H \varepsilon_{t-k} + \nu$$

with $\nu \in H$ such that $\langle \nu, \varepsilon_{t-k} \rangle_H = \mathbf{0}$ for all $k \in \mathbb{N}_0$. This implies that

$$\langle \nu, \varepsilon_{t-k} \rangle_{\bar{\varphi}} = \mathrm{Tr}\left(\langle \nu, \varepsilon_{t-k} \rangle_H\right) = 0.$$

In consequence,

$$\left\| x^{(n)} - x \right\|_{\bar{\varphi}}^2 = \left\| \sum_{k=0}^{+\infty} \left\langle x^{(n)} - x, \varepsilon_{t-k} \right\rangle_H \varepsilon_{t-k} - \nu \right\|_{\bar{\varphi}}^2 = \left\| \sum_{k=0}^{+\infty} \left\langle x^{(n)} - x, \varepsilon_{t-k} \right\rangle_H \varepsilon_{t-k} \right\|_{\bar{\varphi}}^2 + \|\nu\|_{\bar{\varphi}}^2.$$

As $\left\| x^{(n)} - x \right\|_{\bar{\varphi}}$ is arbitrary small, $\|\nu\|_{\bar{\varphi}} = 0$ and so $\nu = 0$. Thus,

$$x = \sum_{k=0}^{+\infty} \langle x, \varepsilon_{t-k} \rangle_H \varepsilon_{t-k} \quad \in \quad \mathcal{H}_t(\boldsymbol{\varepsilon}).$$

■

Similarly, if $\mathbf{x}$ is a weakly stationary vectorial process, we define the closed submodule $\mathcal{H}_t(\mathbf{x})$ of H by

$$\mathcal{H}_t(\mathbf{x}) = \mathrm{cl} \left\{ \sum_{k=0}^{+\infty} a_k x_{t-k} : \quad a_k \in A, \quad \left\| \sum_{k=0}^{+\infty} a_k x_{t-k} \right\|_{\bar{\varphi}} < +\infty \right\},$$

where *cl* denotes the closure in the $\| \ \|_{\bar{\varphi}}$ or $\| \ \|_{\bar{\varphi}}$.

Equivalently,

$$\mathcal{H}_t(\mathbf{x}) = \mathrm{cl} \left\{ \sum_{k=0}^{+\infty} a_k x_{t-k} : \quad a_k \in A, \quad \sum_{k=0}^{+\infty} \sum_{h=0}^{+\infty} \mathrm{Tr}\left(a_k \Gamma_{k-h} a_h'\right) < +\infty \right\}.$$

C Proofs about the Multivariate Classical Wold Decomposition

Statement and proof of Proposition 15

Proposition 15 *The operator* $\mathbf{L}$ *is well-defined and it is* A*-linear and bounded on the span of generators of* $\mathcal{H}_t(\mathbf{x})$*. Hence, it can be extended to* $\mathcal{H}_t(\mathbf{x})$ *with continuity.*

Proof. Consider any generator $X = \sum_{k=0}^{\infty} a_k x_{t-k}$ in $\mathcal{H}_t(\mathbf{x})$, that is

$$\|X\|_{\bar{\varphi}}^2 = \sum_{k=0}^{+\infty} \sum_{h=0}^{+\infty} \mathrm{Tr}\left(a_k \Gamma_{k-h} a_h'\right) < +\infty.$$

By the definition of $\mathbf{L}X$ and the weak stationarity of $\mathbf{x}$,

$$\|\mathbf{L}X\|_{\bar{\varphi}}^2 = \sum_{k=0}^{+\infty} \sum_{h=0}^{+\infty} \mathrm{Tr}\left(a_k \Gamma_{(k+1)-(h+1)} a_h'\right) = \sum_{k=0}^{+\infty} \sum_{h=0}^{+\infty} \mathrm{Tr}\left(a_k \Gamma_{k-h} a_h'\right) = \|X\|_{\bar{\varphi}}^2 < +\infty.$$

Therefore, $\mathbf{L}X$ belongs to $\mathcal{H}_t(\mathbf{x})$ and so $\mathbf{L}$ is well-defined.

To show A-linearity, we take two arbitrary generators of $\mathcal{H}_t(\mathbf{x})$, that is $X = \sum_{k=0}^{\infty} a_k x_{t-k}$, $Y = \sum_{k=0}^{\infty} b_k x_{t-k}$ and a matrix $m \in A$. The element $X + mY$ is defined by

$$X + mY = \sum_{k=0}^{+\infty} (a_k + mb_k)\, x_{t-k}.$$

The lag operator maps $X + mY$ to the vector

$$\mathbf{L}(X + mY) = \sum_{k=0}^{+\infty} (a_k + mb_k)\, x_{t-1-k} = \sum_{k=0}^{+\infty} a_k x_{t-1-k} + m \sum_{k=0}^{+\infty} b_k x_{t-1-k} = \mathbf{L}X + m\mathbf{L}Y$$

and so $\mathbf{L}$ is A-linear.

As for boundedness, we already proved that $\|\mathbf{L}X\|_{\bar{\varphi}} = \|X\|_{\bar{\varphi}}$ for any generator $X \in \mathcal{H}_t(\mathbf{x})$. In consequence, $\mathbf{L}$ is a bounded operator and it can be extended to the closed submodule $\mathcal{H}_t(\mathbf{x})$ with continuity. Indeed, consider the limit $Z \in \mathcal{H}_t(\mathbf{x})$ of a sequence of generators $\{Z_n\}_n$, namely

$$\|Z_n - Z\|_{\bar{\varphi}} \longrightarrow 0.$$

$\{Z_n\}_n$ is a Cauchy sequence and so is the sequence $\{\mathbf{L}Z_n\}_n$ because

$$\begin{aligned} \|\mathbf{L}Z_n - \mathbf{L}Z_m\|_{\bar{\varphi}}^2 &= \mathrm{Tr}\left(\langle \mathbf{L}Z_n - \mathbf{L}Z_m, \mathbf{L}Z_n - \mathbf{L}Z_m\rangle_H\right) \\ &= \mathrm{Tr}\left(\langle Z_n - Z_m, Z_n - Z_m\rangle_H\right) = \|Z_n - Z_m\|_{\bar{\varphi}}^2 \end{aligned}$$

is arbitrarily small when n, m are big enough. Since $\mathcal{H}_t(\mathbf{x})$ is $\| \ \|_{\bar{\varphi}}$ complete, there exists the limit of $\{\mathbf{L}Z_n\}_n$, that we denote ℓ_Z. Such vector does not depend on the sequence of generators that we choose. Indeed, consider the sequence of generators $\{Y_n\}_n$ convergent to Z. Similarly, $\mathbf{L}Y_n$ tends to a limit ℓ_Y. Then, take the sequence $\{X_n\}_n$, where X_n equals Z_n when n is odd and Y_n when n is even. Accordingly, $\mathbf{L}X_n$ converges to a limit $\ell_X = \ell_Z = \ell_Y$. Hence, the extension of $\mathbf{L}$ on $Z \in \mathcal{H}_t(\mathbf{x})$ is unique and we can define $\mathbf{L}Z = \ell_Z$. In addition, the continuity of $\| \ \|_{\bar{\varphi}}$ ensures that the extended $\mathbf{L}$ is A-linear and bounded, too. ■

Statement and proof of Proposition 16

Proposition 16 *If* $\mathbf{x}$ *is a weakly stationary vectorial process,* $\mathbf{L}$ *is an isometry on the Hilbert* A*-module* $\mathcal{H}_t(\mathbf{x})$ *for any* $t \in \mathbb{Z}$.

Proof. We prove the isometry property just for the generators of $\mathcal{H}_t(\mathbf{x})$. Indeed, the continuity of the extension of $\mathbf{L}$ on the closure ensures that the property is satisfied on the whole $\mathcal{H}_t(\mathbf{x})$. Consider, then, any $X = \sum_{k=0}^{\infty} a_k x_{t-k}, \quad Y = \sum_{h=0}^{\infty} b_h x_{t-h}$ in $\mathcal{H}_t(\mathbf{x})$. By the weak stationarity of $\mathbf{x}$, we have

$$\begin{aligned}\langle \mathbf{L}X, \mathbf{L}Y\rangle_H &= \sum_{k=0}^{+\infty}\sum_{h=0}^{+\infty} a_k \langle x_{t-k-1}, x_{t-h-1}\rangle_H b_h' = \sum_{k=0}^{+\infty}\sum_{h=0}^{+\infty} a_k \Gamma_{(k+1)-(h+1)} b_h' \\ &= \sum_{k=0}^{+\infty}\sum_{h=0}^{+\infty} a_k \Gamma_{k-h} b_h' = \sum_{k=0}^{+\infty}\sum_{h=0}^{+\infty} a_k \langle x_{t-k}, x_{t-h}\rangle_H b_h' = \langle X, Y\rangle_H .\end{aligned}$$

Therefore, $\mathbf{L}$ is an isometry on $\mathcal{H}_t(\mathbf{x})$. ■

Statement and proof of Proposition 17

Proposition 17 *For any $j \in \mathbb{N}$,* $\quad \mathbf{L}^j \mathcal{H}_t(\mathbf{x}) = \mathcal{H}_{t-j}(\mathbf{x})$.

Proof. We start with showing that $\mathbf{L}\mathcal{H}_t(\mathbf{x}) = \mathcal{H}_{t-1}(\mathbf{x})$.

Consider any generator $X = \sum_{k=0}^{\infty} a_k x_{t-k}$ of $\mathcal{H}_t(\mathbf{x})$. Its image $\mathbf{L}X$ belongs to $\mathcal{H}_{t-1}(\mathbf{x})$ by definition. As for the generic elements of $\mathcal{H}_t(\mathbf{x})$, the continuity of the extension of $\mathbf{L}$ and the closure of $\mathcal{H}_{t-1}(\mathbf{x})$ ensure that the whole $\mathbf{L}\mathcal{H}_{t-1}(\mathbf{x})$ is included in $\mathcal{H}_{t-1}(\mathbf{x})$.

Conversely, take any generator $Y = \sum_{k=1}^{\infty} b_k x_{t-k}$ of $\mathcal{H}_{t-1}(\mathbf{x})$. Y is the image of the element $X = \sum_{k=0}^{\infty} b_{k+1} x_{t-k}$ belonging to $\mathcal{H}_t(\mathbf{x})$ because

$$\mathbf{L}X = \sum_{k=0}^{+\infty} b_{k+1} x_{t-1-k} = \sum_{k=1}^{+\infty} b_k x_{t-k} = Y.$$

Consequently, the generators of $\mathcal{H}_{t-1}(\mathbf{x})$ are contained in $\mathbf{L}\mathcal{H}_t(\mathbf{x})$. Since $\mathbf{L}$ is an isometry and $\mathcal{H}_t(\mathbf{x})$ is a self-dual Hilbert A-module, by Lemma 2 $\mathbf{L}\mathcal{H}_t(\mathbf{x})$ is a closed submodule of $\mathcal{H}_t(\mathbf{x})$ and so, by taking the closures we get that the whole $\mathcal{H}_{t-1}(\mathbf{x})$ is included in $\mathbf{L}\mathcal{H}_t(\mathbf{x})$.

Following the same steps with the isometric operator $\mathbf{L}^j : \mathcal{H}_t(\mathbf{x}) \longrightarrow \mathcal{H}_t(\mathbf{x})$ that acts on generators of $\mathcal{H}_t(\mathbf{x})$ as

$$\mathbf{L}^j : \qquad \sum_{k=0}^{+\infty} a_k x_{t-k} \quad \longmapsto \quad \sum_{k=0}^{+\infty} a_k x_{t-j-k},$$

it can be proved that $\mathbf{L}^j \mathcal{H}_t(\mathbf{x}) = \mathcal{H}_{t-j}(\mathbf{x})$ for any $j \in \mathbb{N}$. ■

Statement and proof of Proposition 18

Proposition 18 *The wandering submodule associated with the isometric operator* $\mathbf{L}$ *on the Hilbert A-module $\mathcal{H}_t(\mathbf{x})$ is*

$$\mathcal{L}_t^{\mathbf{L}} = \operatorname{span}\left\{ x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t \right\}.$$

Proof. We show that

$$\mathcal{H}_t(\mathbf{x}) = \mathcal{H}_{t-1}(\mathbf{x}) \oplus \operatorname{span}\left\{x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t\right\}.$$

Trivially, $\mathcal{H}_{t-1}(\mathbf{x}) \oplus \operatorname{span}\left\{x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t\right\} \subset \mathcal{H}_t(\mathbf{x})$.

Conversely, if $Y \in \mathcal{H}_t(\mathbf{x})$, there exists a sequence $\{Y_n\}_n \subset \mathcal{H}_t(\mathbf{x})$ of generators such that $\|Y_n - Y\|_{\bar{\varphi}} \longrightarrow 0$ as n goes to infinity. In particular, for any n

$$Y_n = \sum_{k=0}^{+\infty} a_k^{(n)} x_{t-k}.$$

Hence, we can write $Y_n = W_n + Z_n$ with

$$W_n = a_0^{(n)}\mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t + \sum_{k=1}^{+\infty} a_k^{(n)} x_{t-k}, \qquad Z_n = a_0^{(n)}\left(x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t\right).$$

W_n belongs to $\mathcal{H}_{t-1}(\mathbf{x})$, while Z_n is included in the span of the vector $x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t$. The two addends are orthogonal, thanks to the definition of orthogonal projection on the closed submodule $\mathcal{H}_{t-1}(\mathbf{x})$.

Observe that Y can be decomposed into its orthogonal projections on the closed submodules $\mathcal{H}_{t-1}(\mathbf{x})$ and $\mathcal{H}_t(\mathbf{x}) \ominus \mathcal{H}_{t-1}(\mathbf{x})$, that is

$$Y = \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}Y + \mathcal{P}_{\mathcal{H}_t(\mathbf{x})\ominus\mathcal{H}_{t-1}(\mathbf{x})}Y.$$

The orthogonality ensures that, for any $n \in \mathbb{N}_0$,

$$\begin{aligned}
\|Y_n - Y\|_{\bar{\varphi}}^2 &= \|W_n + Z_n - Y\|_{\bar{\varphi}}^2 \\
&= \left\|\left(W_n - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}Y\right) + \left(Z_n - \mathcal{P}_{\mathcal{H}_t(\mathbf{x})\ominus\mathcal{H}_{t-1}(\mathbf{x})}Y\right)\right\|_{\bar{\varphi}}^2 \\
&= \left\|W_n - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}Y\right\|_{\bar{\varphi}}^2 + \left\|Z_n - \mathcal{P}_{\mathcal{H}_t(\mathbf{x})\ominus\mathcal{H}_{t-1}(\mathbf{x})}Y\right\|_{\bar{\varphi}}^2
\end{aligned}$$

and this quantity converges to zero as n increases. As a result,

$$\left\|Z_n - \mathcal{P}_{\mathcal{H}_t(\mathbf{x})\ominus\mathcal{H}_{t-1}(\mathbf{x})}Y\right\|_{\bar{\varphi}}^2 \longrightarrow 0.$$

Since each Z_n belongs to the (closed) span of $x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t$, we deduce that

$$\mathcal{P}_{\mathcal{H}_t(\mathbf{x})\ominus\mathcal{H}_{t-1}(\mathbf{x})}Y \in \operatorname{span}\left\{x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t\right\}.$$

Hence Y is contained in $\mathcal{H}_{t-1}(\mathbf{x}) \oplus \operatorname{span}\left\{x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t\right\}$ and this shows that $\mathcal{H}_t(\mathbf{x}) \subset \mathcal{H}_{t-1}(\mathbf{x}) \oplus \operatorname{span}\left\{x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t\right\}$. ■

Statement and proof of Lemma 4

Lemma 4 *For any $k, j \in \mathbb{N}_0$,*

$$\mathbf{L}^j \mathcal{P}_{\mathcal{H}_{t-k-1}(\mathbf{x})} x_{t-k} = \mathcal{P}_{\mathcal{H}_{t-k-j-1}(\mathbf{x})} x_{t-k-j}.$$

Proof. We prove the result for $k = 0$ without losing generality. In addition, we show the property for $j = 1$ because the general case follows by induction. Precisely, we prove that

$$\mathbf{L}\mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t = \mathcal{P}_{\mathcal{H}_{t-2}(\mathbf{x})} x_{t-1}.$$

Orthogonality of projections on the submodules $\mathcal{H}_{t-2}(\mathbf{x})$ and $\mathcal{H}_{t-1}(\mathbf{x})$ leads to the following normal equations: for any $l \in \mathbb{N}_0$,

$$\left\langle x_{t-1} - \mathcal{P}_{\mathcal{H}_{t-2}(\mathbf{x})} x_{t-1}, x_{t-2-l} \right\rangle_H = \mathbf{0}, \qquad \left\langle x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t, x_{t-1-l} \right\rangle_H = \mathbf{0}.$$

The first one may be rewritten as

$$\langle x_{t-1}, x_{t-2-l} \rangle_H = \left\langle \mathcal{P}_{\mathcal{H}_{t-2}(\mathbf{x})} x_{t-1}, x_{t-2-l} \right\rangle_H \qquad \forall l \in \mathbb{N}_0,$$

while the second one becomes

$$\langle x_t, x_{t-1-l} \rangle_H = \left\langle \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t, x_{t-1-l} \right\rangle_H \qquad \forall l \in \mathbb{N}_0.$$

By using the isometry of the operator $\mathbf{L}$ in the last equation,

$$\langle x_{t-1}, x_{t-2-l} \rangle_H = \left\langle \mathbf{L}\mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t, x_{t-2-l} \right\rangle_H \qquad \forall l \in \mathbb{N}_0,$$

where the left-hand side coincides with the one of the first equation. By matching the expressions above we obtain

$$\left\langle \mathbf{L}\mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t, x_{t-2-l} \right\rangle_H = \left\langle \mathcal{P}_{\mathcal{H}_{t-2}(\mathbf{x})} x_{t-1}, x_{t-2-l} \right\rangle_H \qquad \forall l \in \mathbb{N}_0,$$

namely

$$\left\langle \mathbf{L}\mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t - \mathcal{P}_{\mathcal{H}_{t-2}(\mathbf{x})} x_{t-1}, x_{t-2-l} \right\rangle_H = \mathbf{0} \qquad \forall l \in \mathbb{N}_0.$$

The Hilbert A-module $\mathcal{H}_{t-2}(\mathbf{x})$ is also a Hilbert space with the inner product $\langle \ , \ \rangle_{\bar{\varphi}}$. In addition, $\mathcal{H}_{t-2}(\mathbf{x})$ is countably generated and so it has a countable complete orthonormal system, that we denote by $\mathcal{E}_{t-2}$. As the last normal equation holds for all vectors $\{x_{t-2-l}\}_{l \in \mathbb{N}_0}$, that generate $\mathcal{H}_{t-2}(\mathbf{x})$, the condition is also satisfied by $\mathcal{E}_{t-2}$, namely:

$$\left\langle \mathbf{L}\mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t - \mathcal{P}_{\mathcal{H}_{t-2}(\mathbf{x})} x_{t-1}, e \right\rangle_H = \mathbf{0} \qquad \forall e \in \mathcal{E}_{t-2}.$$

As $\mathcal{E}_{t-2}$ is a complete orthonormal system, we conclude that

$$\mathbf{L}\mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t - \mathcal{P}_{\mathcal{H}_{t-2}(\mathbf{x})} x_{t-1} = 0,$$

as we wanted to show. ∎

Proof of Theorem 1

Since $\mathcal{H}_t(\mathbf{x})$ is a Hilbert A-module, by Theorem 6 in Appendix A, it is self-dual. Since, in addition, $\mathbf{L}$ is an isometry on $\mathcal{H}_t(\mathbf{x})$, we can apply the Abstract Wold Theorem for self-dual Hilbert A-modules.[59]

As for the submodule $\hat{\mathcal{H}}_t(\mathbf{x})$, since each $\mathbf{L}^j\mathcal{H}_t(\mathbf{x})$ coincides with $\mathcal{H}_{t-j}(\mathbf{x})$ by Proposition 17, we find

$$\hat{\mathcal{H}}_t(\mathbf{x}) = \bigcap_{j=0}^{+\infty} \mathbf{L}^j\mathcal{H}_t(\mathbf{x}) = \bigcap_{j=0}^{+\infty} \mathcal{H}_{t-j}(\mathbf{x}).$$

Now consider the submodule $\tilde{\mathcal{H}}_t(\mathbf{x})$. By the Abstract Wold Theorem for Hilbert A-modules,

$$\tilde{\mathcal{H}}_t(\mathbf{x}) = \bigoplus_{j=0}^{+\infty} \mathbf{L}^j\mathcal{L}_t^{\mathbf{L}},$$

where $\mathcal{L}_t^{\mathbf{L}}$ is the innovation submodule defined by $\mathcal{L}_t^{\mathbf{L}} = \mathcal{H}_t(\mathbf{x})\ominus\mathbf{L}\mathcal{H}_t(\mathbf{x})$. By Proposition 18,

$$\mathcal{L}_t^{\mathbf{L}} = \text{span}\left\{x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t\right\}$$

and Lemma 4 ensures that, for any $j \in \mathbb{N}_0$,

$$\mathbf{L}^j\mathcal{L}_t^{\mathbf{L}} = \text{span}\left\{x_{t-j} - \mathcal{P}_{\mathcal{H}_{t-j-1}(\mathbf{x})}x_{t-j}\right\}.$$

■

Proof of Theorem 2

For any $t \in \mathbb{Z}$, the vector x_t belongs to $\mathcal{H}_t(\mathbf{x})$ and so we apply Theorem 1. We denote ν_t the orthogonal projection of x_t on the submodule $\hat{\mathcal{H}}_t(\mathbf{x})$ and we define the process $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t\in\mathbb{Z}}$ by setting, for any $t \in \mathbb{Z}$,

$$\varepsilon_t = S^{-1}\left(x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t\right),$$

where S is the square root matrix of $\left\langle x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t, x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t\right\rangle_H$, i.e.

$$\left\langle x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t, x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t\right\rangle_H = SS.$$

Specifically, S is positive definite because $\mathbf{x}$ is regular and so $\left\langle x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t, x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t\right\rangle_H$ is a symmetric positive definite matrix for any $t \in \mathbb{Z}$. Moreover, as discussed in Lemma 4, the isometry of $\mathbf{L}$ guarantees that S is not dependent on t.

[59] See Theorem 7 in Appendix A.

We call α_k the matrix in A that constitutes the projection coefficient of x_t on the submodule generated by ε_{t-k} and we find the decomposition

$$x_t = \sum_{k=0}^{+\infty} \alpha_k \varepsilon_{t-k} + \nu_t,$$

in which the equality is in norm. We still need to show that α_k are not time-dependent.

i) $\boldsymbol{\varepsilon}$ has unit variance because, for any $t \in \mathbb{Z}$,

$$\mathbb{E}\left[\varepsilon_t \varepsilon_t'\right] = S^{-1} \left\langle x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t, x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t \right\rangle_H \left(S^{-1}\right)' = S^{-1} S S S^{-1} = I.$$

In addition, for any $h \neq k$, it is true that $\mathbb{E}\left[\varepsilon_{t-h} \varepsilon_{t-k}'\right] = \mathbf{0}$ because the vectors ε_t belong to orthogonal submodules. Hence, in order to prove that the process $\boldsymbol{\varepsilon}$ is an m-dimensional white noise, we are just left to show that $\mathbb{E}\left[\varepsilon_t\right] = 0$ for any t. To begin with, we show that $\mathbb{E}\left[\mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t\right] = 0$.

In case the projection of x_t on $\mathcal{H}_{t-1}(\mathbf{x})$ coincides with a generator of $\mathcal{H}_{t-1}(\mathbf{x})$, that is

$$\mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t = \sum_{h=0}^{+\infty} \beta_h x_{t-1-h}, \qquad \beta_h \in A,$$

we immediately see that

$$\mathbb{E}\left[\mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t\right] = \sum_{h=0}^{+\infty} \beta_h \mathbb{E}\left[x_{t-1-h}\right] = 0$$

as $\mathbf{x}$ is a zero-mean process.

In general, we can find a sequence $\{X^{(n)}\}_n$ of random vectors $X^{(n)} = \sum_{h=0}^{\infty} \beta_h^{(n)} x_{t-1-h}$ that converges to $\mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t$ in norm:

$$\left\| X^{(n)} - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t \right\|_{\bar{\varphi}} \longrightarrow 0 \qquad \text{as} \qquad n \to +\infty.$$

For any $j = 1, \ldots, m$ let u_j be the j-th vector of the canonical basis of $\mathbb{R}^m$. $\left\| u_j \right\|_{\bar{\varphi}} = 1$ and the Cauchy-Schwartz' inequality ensures that

$$\begin{aligned}
\left| \mathbb{E}\left[\left(X^{(n)} - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t\right)_j\right] \right| &= \operatorname{Tr}\left(\mathbb{E}\left[\left(X^{(n)} - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t\right) u_j'\right]\right) \\
&= \left\langle X^{(n)} - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t, u_j \right\rangle_{\bar{\varphi}}^2 \\
&\leqslant \left\| X_n - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t \right\|_{\bar{\varphi}}^2 \left\| u_j \right\|_{\bar{\varphi}}^2 \\
&\leqslant \left\| X_n - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t \right\|_{\bar{\varphi}}^2 .
\end{aligned}$$

Therefore, when n goes to infinity,

$$\left| \mathbb{E}\left[X^{(n)}\right] - \mathbb{E}\left[\mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t\right] \right| \longrightarrow 0.$$

On the other hand, for any $n \in \mathbb{N}$,

$$\mathbb{E}\left[X^{(n)}\right] = \sum_{h=0}^{+\infty} \beta_h^{(n)} \mathbb{E}\left[x_{t-1-h}\right] = 0.$$

Since the limit is unique, we deduce that $\mathbb{E}\left[\mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t\right] = 0$.

Summing up, since S is a matrix of real numbers,

$$\mathbb{E}\left[\varepsilon_t\right] = \mathbb{E}\left[S^{-1}\left(x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t\right)\right] = S^{-1}\left(\mathbb{E}\left[x_t\right] - \mathbb{E}\left[\mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t\right]\right) = 0$$

and we conclude that $\boldsymbol{\varepsilon}$ is a multivariate white noise.

ii) Each α_k comes from the projection of the random vector x_t on the submodule generated by ε_{t-k}. For any $k \in \mathbb{N}_0$,

$$\begin{aligned}\langle x_t, \varepsilon_{t-k} \rangle_H = \mathbb{E}\left[x_t \varepsilon'_{t-k}\right] &= \mathbb{E}\left[\left(\sum_{h=0}^{+\infty} \alpha_h \varepsilon_{t-h} + \nu_t\right) \varepsilon'_{t-k}\right] \\ &= \sum_{h=0}^{+\infty} \alpha_h \mathbb{E}\left[\varepsilon_{t-h} \varepsilon'_{t-k}\right] + \mathbb{E}\left[\nu_t \varepsilon'_{t-k}\right] = \alpha_k,\end{aligned}$$

because $\boldsymbol{\varepsilon}$ is a unit variance white noise and, in addition, ν_t and ε_{t-k} belong to orthogonal submodules. Moreover, $\sum_k \alpha_k \varepsilon_{t-k}$ belongs to $\mathcal{H}_t(\mathbf{x})$ and so

$$\left\| \sum_{k=0}^{+\infty} \alpha_k \varepsilon_{t-k} \right\|_{\bar{\varphi}}^2 = \sum_{k=0}^{+\infty} \mathrm{Tr}\left(\alpha_k \alpha'_k\right) < +\infty.$$

We are left to prove that each matrix α_k does not depend on t. By Lemma 4, for any $j, k \in \mathbb{N}_0$,

$$\mathbf{L}^j \varepsilon_{t-k} = \mathbf{L}^j S^{-1}\left(x_{t-k} - \mathcal{P}_{\mathcal{H}_{t-k-1}(\mathbf{x})} x_{t-k}\right) = S^{-1}\left(x_{t-k-j} - \mathcal{P}_{\mathcal{H}_{t-k-j-1}(\mathbf{x})} x_{t-k-j}\right) = \varepsilon_{t-k-j}.$$

Given the decomposition of

$$x_t = \sum_{k=0}^{+\infty} \alpha_k \varepsilon_{t-k} + \nu_t,$$

we apply the lag operator on both sides to obtain

$$x_{t-1} = \mathbf{L}\left(\sum_{k=0}^{+\infty} \alpha_k \varepsilon_{t-k}\right) + \mathbf{L}\nu_t = \sum_{k=0}^{+\infty} \alpha_k \mathbf{L} \varepsilon_{t-k} + \mathbf{L}\nu_t = \sum_{k=0}^{+\infty} \alpha_k \varepsilon_{t-k-1} + \mathbf{L}\nu_t,$$

where we exploited the A-linearity and the continuity of $\mathbf{L}$. Hence, we note that the projection matrices of x_{t-1} on the submodules generated by ε_{t-k-1} are the same as the projection matrices of x_t on the submodules generated by ε_{t-k}, for all $k \in \mathbb{N}_0$. In other words, α_k is also equal to

$$\alpha_k = \mathbb{E}\left[x_{t-1}\varepsilon'_{t-k-1}\right].$$

More generally, by using the operator $\mathbf{L}^j$, we have

$$\alpha_k = \mathbb{E}\left[x_{t-j}\varepsilon'_{t-k-j}\right] \qquad \forall j \in \mathbb{N}_0$$

and so α_k is independent of the time index t, for any k.

iii) By the Abstract Wold Theorem for Hilbert A-modules, ν_t is the projection of x_t on the submodule $\hat{\mathcal{H}}_t(\mathbf{x})$ and $\langle \nu_t, \varepsilon_{t-k}\rangle_H = \mathbb{E}\left[\nu_t\varepsilon'_{t-k}\right] = \mathbf{0}$ for any $k \in \mathbb{N}_0$ because ν_t and ε_{t-k} are in orthogonal submodules of $\mathcal{H}_t(\mathbf{x})$. Moreover, $\boldsymbol{\nu}$ is zero-mean because, for any $t \in \mathbb{Z}$,

$$\mathbb{E}[\nu_t] = \mathbb{E}\left[x_t - \sum_{k=0}^{+\infty}\alpha_k\varepsilon_{t-k}\right] = \mathbb{E}[x_t] - \sum_{k=0}^{+\infty}\alpha_k\mathbb{E}[\varepsilon_{t-k}] = 0$$

because both $\mathbf{x}$ and $\boldsymbol{\varepsilon}$ are zero-mean processes.

Before showing the weak stationarity of $\boldsymbol{\nu}$, we establish that

$$\mathbb{E}\left[x_{t-k}\varepsilon'_{t-l}\right] = \mathbf{0} \qquad \forall l \in \{0, \ldots, k-1\}.$$

Indeed, the vector x_{t-k} decomposes as

$$x_{t-k} = \sum_{h=0}^{+\infty}\alpha_h\varepsilon_{t-k-h} + \nu_{t-k} = \sum_{l=0}^{+\infty}\beta_l\varepsilon_{t-l} + \nu_{t-k},$$

where we defined the matrices

$$\beta_l = \begin{cases} \alpha_h & \text{if } l = k+h \quad \text{for some } h \in \mathbb{N}_0, \\ \mathbf{0} & \text{if } l \in \{0, \ldots, k-1\}. \end{cases}$$

The above expression enables us to embed the submodule $\mathcal{H}_{t-k}(\mathbf{x})$ in $\mathcal{H}_t(\mathbf{x})$. Since the decomposition of x_{t-k} is unique in $\mathcal{H}_t(\mathbf{x})$, it follows that $\mathbb{E}\left[x_{t-k}\varepsilon'_{t-l}\right] = \beta_l = \mathbf{0}$ for all $l \in \{0, \ldots, k-1\}$.

Now, the last step in order to show that $\boldsymbol{\nu}$ is weakly stationary is to prove that $\mathbb{E}\left[\nu_{t-p}\nu'_{t-q}\right]$ depends at most on the difference $p-q$, for any $p, q \in \mathbb{N}_0$. Indeed, suppose that $q \geqslant p+1$:

$$
\begin{aligned}
\mathbb{E}\left[\nu_{t-p}\nu'_{t-q}\right] &= \mathbb{E}\left[\left(x_{t-p} - \sum_{k=0}^{+\infty}\alpha_k\varepsilon_{t-p-k}\right)\left(x_{t-q} - \sum_{h=0}^{+\infty}\alpha_h\varepsilon_{t-q-h}\right)'\right] \\
&= \mathbb{E}\left[x_{t-p}x'_{t-q}\right] - \mathbb{E}\left[x_{t-p}\left(\sum_{h=0}^{+\infty}\alpha_h\varepsilon_{t-q-h}\right)'\right] \\
&\quad - \mathbb{E}\left[x_{t-q}\left(\sum_{k=0}^{+\infty}\alpha_k\varepsilon_{t-p-k}\right)'\right] + \mathbb{E}\left[\left(\sum_{k=0}^{+\infty}\alpha_k\varepsilon_{t-p-k}\right)\left(\sum_{h=0}^{+\infty}\alpha_h\varepsilon_{t-q-h}\right)'\right] \\
&= \Gamma_{p-q} - \sum_{h=0}^{+\infty}\mathbb{E}\left[x_{t-p}\varepsilon'_{t-q-h}\right]\alpha'_h \\
&\quad - \sum_{k=0}^{+\infty}\mathbb{E}\left[x_{t-q}\varepsilon'_{t-p-k}\right]\alpha'_k + \sum_{k=0}^{+\infty}\sum_{h=0}^{+\infty}\alpha_k\mathbb{E}\left[\varepsilon_{t-p-k}\varepsilon'_{t-q-h}\right]\alpha'_h \\
&= \Gamma_{p-q} - \sum_{h=0}^{+\infty}\alpha_{q-p+h}\alpha'_h - \sum_{k=0}^{q-p-1}\mathbb{E}\left[x_{t-q}\varepsilon'_{t-p-k}\right]\alpha_k \\
&\quad - \sum_{k=q-p}^{+\infty}\mathbb{E}\left[x_{t-q}\varepsilon'_{t-p-k}\right]\alpha'_k + \sum_{k=q-p}^{+\infty}\alpha_k\alpha'_{p-q+k} \\
&= \Gamma_{p-q} - \sum_{h=0}^{+\infty}\alpha_{q-p+h}\alpha'_h - \mathbf{0} - \sum_{k=q-p}^{+\infty}\alpha_{p-q+k}\alpha'_k + \sum_{k=q-p}^{+\infty}\alpha_k\alpha'_{p-q+k} \\
&= \Gamma_{p-q} - \sum_{h=0}^{+\infty}\alpha_{q-p+h}\alpha'_h + \sum_{k=q-p}^{+\infty}\left(\alpha_k\alpha'_{p-q+k} - \alpha_{p-q+k}\alpha'_k\right).
\end{aligned}
$$

In consequence, $\mathbb{E}\left[\nu_{t-p}\nu'_{t-q}\right]$ depends at most on $p-q$ and so $\boldsymbol{\nu}$ is weakly stationary.

iv) Since

$$\nu_t \in \hat{\mathcal{H}}_t(\mathbf{x}) = \bigcap_{j=0}^{+\infty}\mathcal{H}_{t-j}(\mathbf{x}),$$

ν_t is also an element of the closed submodule $\mathcal{H}_{t-1}(\mathbf{x})$ and so we can find a sequence of vectors $\left\{X^{(n)}\right\}_n \subset \mathcal{H}_{t-1}(\mathbf{x})$ that converges to ν_t in norm. For instance, we can set

$$X^{(n)} = \sum_{k=0}^{+\infty}\beta_k^{(n)}x_{t-1-k}$$

with

$$\left\|\nu_t - X^{(n)}\right\|_{\bar{\varphi}} = \left\|\nu_t - \sum_{k=0}^{+\infty}\beta_k^{(n)}x_{t-1-k}\right\|_{\bar{\varphi}} \longrightarrow 0 \qquad \text{as} \qquad n \to +\infty.$$

Any of the variables x_{t-1-k} has, in turns, a Multivariate Wold Decomposition

$$x_{t-1-k} = \sum_{h=0}^{+\infty} \alpha_h \varepsilon_{t-1-k-h} + \nu_{t-1-k},$$

in which the equality is in norm. By combining these facts together, we get

$$\begin{aligned}
&\left\| \nu_t - \sum_{k=0}^{+\infty} \beta_k^{(n)} \left(\sum_{h=0}^{+\infty} \alpha_h \varepsilon_{t-1-k-h} + \nu_{t-1-k} \right) \right\|_{\bar{\varphi}} \\
&\qquad\qquad \leqslant \left\| \nu_t - \sum_{k=0}^{+\infty} \beta_k^{(n)} x_{t-1-k} \right\|_{\bar{\varphi}} \\
&\qquad\qquad\quad + \left\| \sum_{k=0}^{+\infty} \beta_k^{(n)} x_{t-1-k} - \sum_{k=0}^{+\infty} \beta_k^{(n)} \left(\sum_{h=0}^{+\infty} \alpha_h \varepsilon_{t-1-k-h} + \nu_{t-1-k} \right) \right\|_{\bar{\varphi}} \\
&\qquad\qquad \leqslant \left\| \nu_t - \sum_{k=0}^{+\infty} \beta_k^{(n)} x_{t-1-k} \right\|_{\bar{\varphi}} \\
&\qquad\qquad\quad + \sum_{k=0}^{+\infty} \left| \beta_k^{(n)} \right| \left\| x_{t-1-k} - \left(\sum_{h=0}^{+\infty} \alpha_h \varepsilon_{t-1-k-h} + \nu_{t-1-k} \right) \right\|_{\bar{\varphi}} \\
&\qquad\qquad = \left\| \nu_t - \sum_{k=0}^{+\infty} \beta_k^{(n)} x_{t-1-k} \right\|_{\bar{\varphi}}
\end{aligned}$$

When n goes to infinity, the right-hand side converges to zero. Therefore, also

$$\left\| \nu_t - \sum_{k=0}^{+\infty} \beta_k^{(n)} \left(\sum_{h=0}^{+\infty} \alpha_h \varepsilon_{t-1-k-h} + \nu_{t-1-k} \right) \right\|_{\bar{\varphi}} \quad \longrightarrow \quad 0.$$

The last convergence may be rewritten as

$$\left\| \nu_t - \sum_{k=0}^{+\infty} \beta_k^{(n)} \nu_{t-1-k} - \sum_{l=0}^{+\infty} \left(\sum_{k=0}^{l} \beta_k^{(n)} \alpha_{l-k} \right) \varepsilon_{t-1-l} \right\|_{\bar{\varphi}} \quad \longrightarrow \quad 0.$$

Observe that the element

$$\sum_{l=0}^{+\infty} \left(\sum_{k=0}^{l} \beta_k^{(n)} \alpha_{l-k} \right) \varepsilon_{t-1-l}$$

belongs to $\tilde{\mathcal{H}}_t(\mathbf{x})$, while

$$\nu_t - \sum_{k=0}^{+\infty} \beta_k^{(n)} \nu_{t-1-k}$$

is contained in $\hat{\mathcal{H}}_t(\mathbf{x})$. Since $\tilde{\mathcal{H}}_t(\mathbf{x})$ and $\hat{\mathcal{H}}_t(\mathbf{x})$ are orthogonal submodules, in the limit it must hold that

$$\left\| \nu_t - \sum_{k=0}^{+\infty} \beta_k^{(n)} \nu_{t-1-k} \right\|_{\bar{\varphi}} \quad \longrightarrow \quad 0.$$

As a result, we can claim that

$$\nu_t \in \mathrm{cl}\left\{ \sum_{h=1}^{+\infty} a_h \nu_{t-h} \quad \in \bigcap_{j=1}^{+\infty} \mathcal{H}_{t-j}(\mathbf{x}) : \quad a_h \in A \right\}.$$

D Proofs about the Multivariate Extended Wold Decomposition

Statement and proof of Proposition 19

Proposition 19 *The operator* $\mathbf{R} : \mathcal{H}_t(\boldsymbol{\varepsilon}) \longrightarrow \mathcal{H}_t(\boldsymbol{\varepsilon})$ *is well-defined, A-linear and isometric.*

Proof. In order to show that the scaling operator $\mathbf{R}$ is well-defined on $\mathcal{H}_t(\boldsymbol{\varepsilon})$, consider any element $X = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k} \in \mathcal{H}_t(\boldsymbol{\varepsilon})$, i.e.

$$\|X\|_{\tilde{\varphi}}^2 = \sum_{k=0}^{+\infty} \mathrm{Tr}\left(a_k a_k'\right) < +\infty.$$

Our purpose is to prove that $\mathbf{R}X$ belongs to $\mathcal{H}_t(\boldsymbol{\varepsilon})$ too, namely

$$\|\mathbf{R}X\|_{\tilde{\varphi}}^2 = \sum_{k=0}^{+\infty} \mathrm{Tr}\left(\frac{a_{\lfloor \frac{k}{2} \rfloor}}{\sqrt{2}} \frac{a'_{\lfloor \frac{k}{2} \rfloor}}{\sqrt{2}} \right) < +\infty.$$

Since index k is either even or odd, the sum actually is

$$\|\mathbf{R}X\|_{\tilde{\varphi}}^2 = \sum_{k=0}^{+\infty} \mathrm{Tr}\left(\frac{a_{\lfloor \frac{k}{2} \rfloor}}{\sqrt{2}} \frac{a'_{\lfloor \frac{k}{2} \rfloor}}{\sqrt{2}} \right) = \frac{1}{2} \sum_{k=0}^{+\infty} \mathrm{Tr}\left(a_{\lfloor \frac{k}{2} \rfloor} a'_{\lfloor \frac{k}{2} \rfloor} \right) = \sum_{p=0}^{+\infty} \mathrm{Tr}\left(a_p a_p' \right) = \|X\|_{\tilde{\varphi}}^2$$

and this quantity is finite. As a result, $\mathbf{R}$ is well-defined. In addition $\mathbf{R}$ is a bounded operator because $\|\mathbf{R}X\|_{\tilde{\varphi}} = \|X\|_{\tilde{\varphi}}$.

About the A-linearity of the scaling operator, consider any matrix $m \in A$ and two arbitrary elements $X = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}, \quad Y = \sum_{k=0}^{\infty} b_k \varepsilon_{t-k}$ in $\mathcal{H}_t(\boldsymbol{\varepsilon})$. The element $X + mY = \sum_k c_k \varepsilon_{t-k}$ has for coefficients the matrices $c_k = a_k + m b_k$ for every k in $\mathbb{N}_0$. The operator $\mathbf{R}$ maps $X + mY$ to the element

$$\mathbf{R}(X + mY) = \sum_{k=0}^{+\infty} \frac{c_{\lfloor \frac{k}{2} \rfloor}}{\sqrt{2}} \varepsilon_{t-k} \quad \in \mathcal{H}_t(\boldsymbol{\varepsilon}), \qquad \text{with} \qquad c_{\lfloor \frac{k}{2} \rfloor} = a_{\lfloor \frac{k}{2} \rfloor} + m b_{\lfloor \frac{k}{2} \rfloor} \quad \forall k \in \mathbb{N}_0.$$

As a result, $\mathbf{R}(X + mY) = \mathbf{R}X + m\mathbf{R}Y$, that is the scaling operator is A-linear.

Finally, we prove that the scaling operator is isometric on $\mathcal{H}_t(\boldsymbol{\varepsilon})$. Consider any two elements $X = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}, \quad Y = \sum_{h=0}^{\infty} b_h \varepsilon_{t-h}$ in $\mathcal{H}_t(\boldsymbol{\varepsilon})$. By exploiting the properties of the multivariate white noise $\boldsymbol{\varepsilon}$, we find that

$$\begin{aligned}\langle \mathbf{R}X, \mathbf{R}Y\rangle_H &= \left\langle \sum_{k=0}^{+\infty} \frac{a_{\lfloor \frac{k}{2} \rfloor}}{\sqrt{2}} \varepsilon_{t-k}, \sum_{h=0}^{+\infty} \frac{b_{\lfloor \frac{h}{2} \rfloor}}{\sqrt{2}} \varepsilon_{t-k} \right\rangle_H = \frac{1}{2} \sum_{k=0}^{+\infty} \sum_{h=0}^{+\infty} a_{\lfloor \frac{k}{2} \rfloor} \langle \varepsilon_{t-k}, \varepsilon_{t-h} \rangle\, b'_{\lfloor \frac{h}{2} \rfloor} \\ &= \frac{1}{2} \sum_{k=0}^{+\infty} a_{\lfloor \frac{k}{2} \rfloor} b'_{\lfloor \frac{k}{2} \rfloor} = \sum_{k=0}^{+\infty} a_k b'_k = \langle X, Y \rangle_H .\end{aligned}$$

Hence, $\mathbf{R}$ is an isometry. ■

Statement and proof of Proposition 20

Proposition 20 *For any fixed $j \in \mathbb{N}$, the process $\left\{ \varepsilon^{(j)}_{t-k2^j} \right\}_{k \in \mathbb{Z}}$ is a unit variance white noise.*

Proof. First of all, we show that $\left\{ \varepsilon^{(j)}_{t-k2^j} \right\}_{k \in \mathbb{Z}}$ is weakly stationary.

i) The vectors ε_t are the classical Wold innovations of the process $\mathbf{x}$, hence $\mathbb{E}\left[\varepsilon_{t-p}\varepsilon'_{t-q}\right] = 0$ for all $p \neq q$ and $\mathbb{E}\left[\varepsilon_t \varepsilon'_t\right] = I$ for any t. Therefore, for any $k \in \mathbb{Z}$,

$$\begin{aligned}\mathbb{E}\left[\varepsilon^{(j)}_{t-k2^j} {\varepsilon^{(j)}_{t-k2^j}}'\right] &= \frac{1}{2^j} \mathbb{E}\Bigg[\left(\sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-k2^j-i} - \sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-k2^j-2^{j-1}-i} \right) \\ &\qquad \cdot \left(\sum_{i=0}^{2^{j-1}-1} \varepsilon'_{t-k2^j-i} - \sum_{i=0}^{2^{j-1}-1} \varepsilon'_{t-k2^j-2^{j-1}-i} \right) \Bigg] \\ &= \frac{1}{2^j} \sum_{i=0}^{2^j-1} \mathbb{E}\left[\varepsilon_t \varepsilon'_t\right] = \frac{2^j}{2^j} \mathbb{E}\left[\varepsilon_t \varepsilon'_t\right] = I.\end{aligned}$$

Hence, $\mathbb{E}\left[\varepsilon^{(j)}_{t-k2^j} {\varepsilon^{(j)}_{t-k2^j}}'\right]$ is finite and it does not depend on k.

ii) Since $\mathbb{E}\left[\varepsilon_t\right] = 0$ for any t, by A-linearity we find that $\mathbb{E}\left[\varepsilon^{(j)}_{t-k2^j}\right] = 0$ for any $k \in \mathbb{Z}$ and so the expectation does not depend on k.

$iii)$ Consider the cross moments matrix in the support $S_t^{(j)}$. By taking $h \neq k$,

$$\mathbb{E}\left[\varepsilon_{t-h2^j}^{(j)}\varepsilon_{t-k2^j}^{(j)}{}'\right] = \frac{1}{2^j}\mathbb{E}\Bigg[\left(\sum_{i=0}^{2^{j-1}-1}\varepsilon_{t-h2^j-i} - \sum_{i=0}^{2^{j-1}-1}\varepsilon_{t-h2^j-2^{j-1}-i}\right)$$
$$\cdot\left(\sum_{l=0}^{2^{j-1}-1}\varepsilon'_{t-k2^j-l} - \sum_{l=0}^{2^{j-1}-1}\varepsilon'_{t-k2^j-2^{j-1}-l}\right)\Bigg]$$
$$= \frac{1}{2^j}\Bigg\{\sum_{i=0}^{2^{j-1}-1}\sum_{l=0}^{2^{j-1}-1}\mathbb{E}\left[\varepsilon_{t-h2^j-i}\varepsilon'_{t-k2^j-l}\right]$$
$$-\sum_{i=0}^{2^{j-1}-1}\sum_{l=0}^{2^{j-1}-1}\mathbb{E}\left[\varepsilon_{t-h2^j-i}\varepsilon'_{t-k2^j-2^{j-1}-l}\right]$$
$$-\sum_{i=0}^{2^{j-1}-1}\sum_{l=0}^{2^{j-1}-1}\mathbb{E}\left[\varepsilon_{t-h2^j-2^{j-1}-i}\varepsilon'_{t-k2^j-l}\right]$$
$$+\sum_{i=0}^{2^{j-1}-1}\sum_{l=0}^{2^{j-1}-1}\mathbb{E}\left[\varepsilon_{t-h2^j-2^{j-1}-i}\varepsilon'_{t-k2^j-2^{j-1}-l}\right]\Bigg\}.$$

Since $h \neq k$, the sets of indices $\{h2^j, \dots, h2^j + 2^j - 1\}$ and $\{k2^j, \dots, k2^j + 2^j - 1\}$ are disjoint and so all the last sums are null. In consequence,

$$\mathbb{E}\left[\varepsilon_{t-h2^j}^{(j)}\varepsilon_{t-k2^j}^{(j)}{}'\right] = \mathbf{0} \quad \forall h \neq k.$$

To recap, $\left\{\varepsilon_{t-k2^j}^{(j)}\right\}_{k\in\mathbb{Z}}$ turns out to be weakly stationary on its support $S_t^{(j)}$. In particular, it is a unit variance white noise. ■

Statement and proof of Proposition 21

Proposition 21 *The adjoint of* $\mathbf{R}$ *is the operator* $\mathbf{R}^* : \mathcal{H}_t(\boldsymbol{\varepsilon}) \longrightarrow \mathcal{H}_t(\boldsymbol{\varepsilon})$ *such that*

$$\mathbf{R}^* : \quad \sum_{k=0}^{+\infty} a_k\varepsilon_{t-k} \quad \longmapsto \quad \sum_{k=0}^{+\infty}\frac{a_{2k}+a_{2k+1}}{\sqrt{2}}\varepsilon_{t-k}.$$

In addition, the kernel of $\mathbf{R}^*$ *is*

$$\ker(\mathbf{R}^*) = \left\{\sum_{k=0}^{+\infty} b_k^{(1)}\varepsilon_{t-2k}^{(1)} \in \mathcal{H}_t(\boldsymbol{\varepsilon}) : \quad b_k^{(1)} \in A\right\}.$$

Proof. To prove that $\mathbf{R}^*$ is well-defined, we take any $Y = \sum_{k=0}^{\infty} a_k\varepsilon_{t-k}$ in $\mathcal{H}_t(\boldsymbol{\varepsilon})$, i.e. $\|Y\|_{\tilde{\varphi}}^2 = \sum_{h=0}^{\infty}\mathrm{Tr}\left(a_h a_h'\right) < +\infty$. Then,

$$\|\mathbf{R}^*Y\|_{\tilde{\varphi}}^2 = \frac{1}{2}\sum_{k=0}^{+\infty}\mathrm{Tr}\left(\left(a_{2k}+a_{2k+1}\right)\left(a'_{2k}+a'_{2k+1}\right)\right).$$

For any $i = 1, \ldots, m$, the (i,i)-entry of the matrix $a_h a_h'$ is $\sum_{j=1}^{m} \left(a_h^{(i,j)} \right)^2$, while the (i,i)-entry of the matrix $(a_{2k} + a_{2k+1}) \left(a_{2k}' + a_{2k+1}' \right)$ is $\sum_{j=1}^{m} \left(a_{2k}^{(i,j)} + a_{2k+1}^{(i,j)} \right)^2$. Hence

$$\operatorname{Tr}\left(a_h a_h' \right) = \sum_{i=1}^{m} \sum_{j=1}^{m} \left(a_h^{(i,j)} \right)^2$$

and

$$\operatorname{Tr}\left((a_{2k} + a_{2k+1}) \left(a_{2k}' + a_{2k+1}' \right) \right) = \sum_{i=1}^{m} \sum_{j=1}^{m} \left(a_{2k}^{(i,j)} + a_{2k+1}^{(i,j)} \right)^2.$$

Therefore,

$$\begin{aligned} \|\mathbf{R}^* Y\|_{\bar{\varphi}}^2 &= \frac{1}{2} \sum_{k=0}^{+\infty} \sum_{i=1}^{m} \sum_{j=1}^{m} \left(a_{2k}^{(i,j)} + a_{2k+1}^{(i,j)} \right)^2 \\ &\leqslant \sum_{k=0}^{+\infty} \sum_{i=1}^{m} \sum_{j=1}^{m} \left\{ \left(a_{2k}^{(i,j)} \right)^2 + \left(a_{2k+1}^{(i,j)} \right)^2 \right\} \\ &= \sum_{h=0}^{+\infty} \sum_{i=1}^{m} \sum_{j=1}^{m} \left(a_h^{(i,j)} \right)^2 \\ &= \|Y\|_{\bar{\varphi}}^2. \end{aligned}$$

We deduce that $\|\mathbf{R}^* Y\|^2$ is finite.

Now we establish the validity of the relation $\langle \mathbf{R} X, Y \rangle_H = \langle X, \mathbf{R}^* Y \rangle_H$ for any vectors in $\mathcal{H}_t(\boldsymbol{\varepsilon})$ as $X = \sum_{h=0}^{\infty} b_h \varepsilon_{t-h}$ and $Y = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}$. By the unit variance white noise properties, we have

$$\begin{aligned} \langle \mathbf{R} X, Y \rangle_H &= \sum_{h=0}^{+\infty} \sum_{k=0}^{+\infty} \frac{b_{\lfloor \frac{h}{2} \rfloor}}{\sqrt{2}} \langle \varepsilon_{t-h}, \varepsilon_{t-k} \rangle_H a_k' = \sum_{k=0}^{+\infty} b_{\lfloor \frac{k}{2} \rfloor} \frac{a_k'}{\sqrt{2}} = \sum_{k=0}^{+\infty} b_k \frac{a_{2k}' + a_{2k+1}'}{\sqrt{2}} \\ &= \sum_{h=0}^{+\infty} \sum_{k=0}^{+\infty} b_h \langle \varepsilon_{t-h}, \varepsilon_{t-k} \rangle_H \frac{a_{2k}' + a_{2k+1}'}{\sqrt{2}} = \langle X, \mathbf{R}^* Y \rangle_H. \end{aligned}$$

Therefore, $\mathbf{R}^*$ is the adjoint of the scaling operator.

Finally we prove that

$$\ker(\mathbf{R}^*) = \left\{ \sum_{k=0}^{+\infty} d_k^{(1)} \left(\varepsilon_{t-2k} - \varepsilon_{t-2k-1} \right) \in \mathcal{H}_t(\boldsymbol{\varepsilon}) : \quad d_k^{(1)} \in A \right\}.$$

Any element of $\mathcal{H}_t(\boldsymbol{\varepsilon})$ as, for instance,

$$X = \sum_{k=0}^{+\infty} d_k^{(1)} \left(\varepsilon_{t-2k} - \varepsilon_{t-2k-1} \right)$$

can be rewritten as $X = \sum_{h=0}^{\infty} a_h \varepsilon_{t-h}$ with $a_{2k+1} = -a_{2k}$ for every $k \in \mathbb{N}_0$, i.e. $a_{2k} + a_{2k+1} = \mathbf{0}$. Consequently, $\mathbf{R}^* X = 0$ and so

$$\left\{ \sum_{k=0}^{+\infty} d_k^{(1)} \left(\varepsilon_{t-2k} - \varepsilon_{t-2k-1} \right) \in \mathcal{H}_t(\boldsymbol{\varepsilon}) : \quad d_k^{(1)} \in A \right\} \subset \ker(\mathbf{R}^*).$$

Conversely, let $X = \sum_{h=0}^{\infty} a_h \varepsilon_{t-h}$ belong to $\ker(\mathbf{R}^*)$. Since $\|\mathbf{R}^* X\|_{\bar{\varphi}} = 0$, by Lemma 3, we have

$$\frac{1}{2} \sum_{k=0}^{+\infty} \operatorname{Tr}\left((a_{2k} + a_{2k+1})(a'_{2k} + a'_{2k+1}) \right) = 0.$$

As we observed before,

$$\operatorname{Tr}\left((a_{2k} + a_{2k+1}) \left(a'_{2k} + a'_{2k+1} \right) \right) = \sum_{i=1}^{m} \sum_{j=1}^{m} \left(a_{2k}^{(i,j)} + a_{2k+1}^{(i,j)} \right)^2$$

and so

$$\sum_{k=0}^{+\infty} \sum_{i=1}^{m} \sum_{j=1}^{m} \left(a_{2k}^{(i,j)} + a_{2k+1}^{(i,j)} \right)^2 = 0.$$

It follows that $a_{2k+1}^{(i,j)} = -a_{2k}^{(i,j)}$ for any $k \in \mathbb{N}_0$, $i, j = 1, \ldots, m$. Therefore, $a_{2k+1} = -a_{2k}$ for any $k \in \mathbb{N}_0$. As a result,

$$X = \sum_{k=0}^{+\infty} d_k^{(1)} \left(\varepsilon_{t-2k} - \varepsilon_{t-2k-1} \right),$$

with $d_k^{(1)} = a_{2k}$ and so

$$\ker(\mathbf{R}^*) \subset \left\{ \sum_{k=0}^{+\infty} d_k^{(1)} \left(\varepsilon_{t-2k} - \varepsilon_{t-2k-1} \right) \in \mathcal{H}_t(\boldsymbol{\varepsilon}) : \quad d_k^{(1)} \in A \right\}.$$

■

Statement and proof of Proposition 22

Proposition 22 *For any $j \in \mathbb{N}$,*

$$\mathbf{R}^{j-1} \mathcal{L}_t^{\mathbf{R}} = \left\{ \sum_{k=0}^{+\infty} b_k^{(j)} \varepsilon_{t-k2^j}^{(j)} \in \mathcal{H}_t(\boldsymbol{\varepsilon}) : \quad b_k^{(j)} \in A \right\}.$$

Proof. As the general case follows by induction, we prove that

$$\mathbf{R} \mathcal{L}_t^{\mathbf{R}} = \left\{ \sum_{k=0}^{+\infty} b_k^{(2)} \varepsilon_{t-4k}^{(2)} \in \mathcal{H}_t(\boldsymbol{\varepsilon}) : \quad b_k^{(2)} \in A \right\}.$$

In particular, we show that

$$\mathbf{R}\mathcal{L}_t^{\mathbf{R}} = \left\{ \sum_{k=0}^{+\infty} d_k^{(2)} \left(\varepsilon_{t-4k} + \varepsilon_{t-4k-1} - \varepsilon_{t-4k-2} - \varepsilon_{t-4k-3} \right) \in \mathcal{H}_t(\boldsymbol{\varepsilon}) : \quad d_k^{(2)} \in A \right\}.$$

Consider any element $Y \in \mathbf{R}\mathcal{L}_t^{\mathbf{R}}$. As Y is the image of some element $X \in \mathcal{L}_t^{\mathbf{R}}$, there exists a sequence of matrices $\left\{ d_k^{(1)} \right\}_k$ such that

$$X = \sum_{k=0}^{+\infty} d_k^{(1)} \left(\varepsilon_{t-2k} - \varepsilon_{t-2k-1} \right)$$

and

$$Y = \sum_{k=0}^{+\infty} \frac{d_k^{(1)}}{\sqrt{2}} \left(\varepsilon_{t-4k} + \varepsilon_{t-4k-1} - \varepsilon_{t-4k-2} - \varepsilon_{t-4k-3} \right).$$

As a result,

$$\mathbf{R}\mathcal{L}_t^{\mathbf{R}} \subset \left\{ \sum_{k=0}^{+\infty} d_k^{(2)} \left(\varepsilon_{t-4k} + \varepsilon_{t-4k-1} - \varepsilon_{t-4k-2} - \varepsilon_{t-4k-3} \right) \in \mathcal{H}_t(\boldsymbol{\varepsilon}) : \quad d_k^{(2)} \in A \right\}.$$

Conversely, consider any element $Y \in \mathcal{H}_t(\boldsymbol{\varepsilon})$ of the kind

$$Y = \sum_{k=0}^{+\infty} d_k^{(2)} \left(\varepsilon_{t-4k} + \varepsilon_{t-4k-1} - \varepsilon_{t-4k-2} - \varepsilon_{t-4k-3} \right).$$

Then Y belongs to $\mathbf{R}\mathcal{L}_t^{\mathbf{R}}$ too, because it is the image of the element

$$X = \sum_{k=0}^{+\infty} \sqrt{2} d_k^{(2)} \left(\varepsilon_{t-2k} - \varepsilon_{t-2k-1} \right) \quad \in \mathcal{L}_t^{\mathbf{R}}.$$

Consequently,

$$\left\{ \sum_{k=0}^{+\infty} d_k^{(2)} \left(\varepsilon_{t-4k} + \varepsilon_{t-4k-1} - \varepsilon_{t-4k-2} - \varepsilon_{t-4k-3} \right) \in \mathcal{H}_t(\boldsymbol{\varepsilon}) : \quad d_k^{(2)} \in A \right\} \subset \mathbf{R}\mathcal{L}_t^{\mathbf{R}}.$$

■

Proof of Theorem 3

$\mathcal{H}_t(\boldsymbol{\varepsilon})$ is a Hilbert A-module and so, by Theorem 6 in Appendix A, it is self-dual. Moreover, on $\mathcal{H}_t(\boldsymbol{\varepsilon})$ the operator $\mathbf{R}$ is isometric by Proposition 19. Thus, we apply the Abstract Wold Decomposition for Hilbert A-modules, which provides the orthogonal decomposition $\mathcal{H}_t(\boldsymbol{\varepsilon}) = \hat{\mathcal{H}}_t(\boldsymbol{\varepsilon}) \oplus \tilde{\mathcal{H}}_t(\boldsymbol{\varepsilon})$, where

$$\hat{\mathcal{H}}_t(\boldsymbol{\varepsilon}) = \bigcap_{j=0}^{+\infty} \mathbf{R}^j \mathcal{H}_t(\boldsymbol{\varepsilon}),$$

$$\tilde{\mathcal{H}}_t(\boldsymbol{\varepsilon}) = \bigoplus_{j=0}^{+\infty} \mathbf{R}^j \mathcal{L}_t^{\mathbf{R}} \qquad \text{with} \qquad \mathcal{L}_t^{\mathbf{R}} = \mathcal{H}_t(\boldsymbol{\varepsilon}) \ominus \mathbf{R}\mathcal{H}_t(\boldsymbol{\varepsilon}).$$

First, we show that $\hat{\mathcal{H}}_t(\boldsymbol{\varepsilon})$ is the null submodule. Indeed, the submodules $\mathbf{R}^j\mathcal{H}_t(\boldsymbol{\varepsilon})$ consist of linear combinations of innovations with matrix coefficients equal to each others 2^j-by-2^j:

$$\mathbf{R}^j\mathcal{H}_t(\boldsymbol{\varepsilon}) = \left\{ \sum_{k=0}^{+\infty} c_k^{(j)} \left(\sum_{i=0}^{2^j-1} \varepsilon_{t-k2^j-i} \right) \in \mathcal{H}_t(\boldsymbol{\varepsilon}) : \quad c_k^{(j)} \in A \right\}.$$

Therefore, $\hat{\mathcal{H}}_t(\boldsymbol{\varepsilon})$, being the intersection of all $\mathbf{R}^j\mathcal{H}_t(\boldsymbol{\varepsilon})$, can just include vectors as $\sum_{h=0}^{\infty} c\varepsilon_{t-h}$ with $c \in A$. Such vectors must belong to $\mathcal{H}_t(\boldsymbol{\varepsilon})$, hence

$$\sum_{k=0}^{\infty} \mathrm{Tr}\,(cc') = \sum_{k=0}^{+\infty} \sum_{i=1}^{m} \sum_{j=1}^{m} \left(c^{(i,j)}\right)^2$$

is finite. Since the addends do not depend on k, it follows that $c^{(i,j)} = 0$ for all $i, j = 1, \dots, m$ and so c is the null matrix. Consequently, $\hat{\mathcal{H}}_t(\boldsymbol{\varepsilon}) = \{0\}$ and so the orthogonal decomposition of $\mathcal{H}_t(\boldsymbol{\varepsilon})$ simplifies to $\mathcal{H}_t(\boldsymbol{\varepsilon}) = \tilde{\mathcal{H}}_t(\boldsymbol{\varepsilon})$.

As for the submodule $\tilde{\mathcal{H}}_t(\boldsymbol{\varepsilon})$,

$$\tilde{\mathcal{H}}_t(\boldsymbol{\varepsilon}) = \bigoplus_{j=0}^{+\infty} \mathbf{R}^j \mathcal{L}_t^{\mathbf{R}} = \bigoplus_{j=1}^{+\infty} \mathbf{R}^{j-1} \mathcal{L}_t^{\mathbf{R}},$$

where $\mathcal{L}_t^{\mathbf{R}}$ is the wandering submodule. As the orthogonal complement of $\mathbf{R}\mathcal{H}_t(\mathbf{x})$ is the kernel of the adjoint operator $\mathbf{R}^*$ (see Proposition 10), by Proposition 21 the wandering submodule is

$$\mathcal{L}_t^{\mathbf{R}} = \mathcal{H}_t(\boldsymbol{\varepsilon}) \ominus \mathbf{R}\mathcal{H}_t(\boldsymbol{\varepsilon}) = \ker(\mathbf{R}^*) = \left\{ \sum_{k=0}^{+\infty} b_k^{(1)} \varepsilon_{t-2k}^{(1)} \in \mathcal{H}_t(\boldsymbol{\varepsilon}) : \quad b_k^{(1)} \in A \right\}.$$

Moreover, by Proposition 22, for any $j \in \mathbb{N}$,

$$\mathbf{R}^{j-1}\mathcal{L}_t^{\mathbf{R}} = \left\{ \sum_{k=0}^{+\infty} b_k^{(j)} \varepsilon_{t-k2^j}^{(j)} \in \mathcal{H}_t(\boldsymbol{\varepsilon}) : \quad b_k^{(j)} \in A \right\}$$

and so the decomposition of the Hilbert A-module $\mathcal{H}_t(\boldsymbol{\varepsilon})$ is proved. ■

Proof of Proposition 1

To begin with, we show that, for any fixed scale $j \in \mathbb{N}$,

$$g_t^{(j)} = \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)}$$

belongs to $\mathbf{R}^{j-1}\mathcal{L}_t^{\mathbf{R}}$, that is we assess the convergence of $\left\|\sum_{k=0}^{\infty}\beta_k^{(j)}\varepsilon_{t-k2^j}^{(j)}\right\|_{\bar{\varphi}}$. By making the variables $\varepsilon_{t-k2^j}^{(j)}$ explicit with respect to the classical Wold innovations of x_t, we derive that

$$g_t^{(j)} = \sum_{k=0}^{+\infty}\left(\sum_{l=0}^{2^{j-1}-1}\frac{\beta_k^{(j)}}{\sqrt{2^j}}\varepsilon_{t-k2^j-l} - \sum_{l=0}^{2^{j-1}-1}\frac{\beta_k^{(j)}}{\sqrt{2^j}}\varepsilon_{t-k2^j-2^{j-1}-l}\right).$$

For any $h \in \mathbb{N}_0$, we uniquely find $k \in \mathbb{N}_0$ and $l \in \{0,1,\ldots,2^j-1\}$ such that $h = k2^j + l$. Consequently, we can express the component $g_t^{(j)}$ as

$$g_t^{(j)} = \sum_{h=0}^{+\infty}\eta_h^{(j)}\varepsilon_{t-h},$$

where the matrices $\eta_h^{(j)}$ are defined by

$$\eta_h^{(j)} = \begin{cases} \frac{\beta_k^{(j)}}{\sqrt{2^j}} & \text{for } k \in \mathbb{N}_0, \quad l \in \{0,\ldots,2^{j-1}-1\} \\ -\frac{\beta_k^{(j)}}{\sqrt{2^j}} & \text{for } k \in \mathbb{N}_0, \quad l \in \{2^{j-1},\ldots,2^j-1\}. \end{cases}$$

Therefore, we have to check the convergence of the series

$$\sum_{h=0}^{+\infty}\mathrm{Tr}\left(\eta_h^{(j)}\eta_h^{(j)\prime}\right) = \sum_{k=0}^{+\infty}\sum_{l=0}^{2^j-1}\mathrm{Tr}\left(\frac{\beta_k^{(j)}\beta_k^{(j)\prime}}{2^j}\right) = \sum_{k=0}^{+\infty}\mathrm{Tr}\left(\beta_k^{(j)}\beta_k^{(j)\prime}\right).$$

As we observed in the proof of Proposition 21,

$$\mathrm{Tr}\left(\alpha_h\alpha_h'\right) = \sum_{p=1}^{m}\sum_{q=1}^{m}\alpha_h^2(p,q), \qquad \mathrm{Tr}\left(\beta_k^{(j)}\beta_k^{(j)\prime}\right) = \sum_{p=1}^{m}\sum_{q=1}^{m}\left(\beta_k^{(j)}(p,q)\right)^2.$$

Note that

$$
\begin{aligned}
\sum_{k=0}^{+\infty} \operatorname{Tr}\left(\beta_k^{(j)} \beta_k^{(j)\prime}\right) &= \sum_{k=0}^{+\infty} \sum_{p=1}^{m} \sum_{q=1}^{m} \left(\beta_k^{(j)}(p,q)\right)^2 \\
&= \frac{1}{2^j} \sum_{k=0}^{+\infty} \sum_{p=1}^{m} \sum_{q=1}^{m} \left(\sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+i}(p,q) - \sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+2^{j-1}+i}(p,q) \right)^2 \\
&\leqslant \frac{2}{2^j} \sum_{k=0}^{+\infty} \sum_{p=1}^{m} \sum_{q=1}^{m} \left(\sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+i}(p,q) \right)^2 + \left(\sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+2^{j-1}+i}(p,q) \right)^2 \\
&\leqslant \frac{2}{2^j} \sum_{k=0}^{+\infty} \sum_{p=1}^{m} \sum_{q=1}^{m} \left(2^{j-1} \left(\sum_{i=0}^{2^{j-1}-1} \alpha^2_{k2^j+i}(p,q) \right) + 2^{j-1} \left(\sum_{i=0}^{2^{j-1}-1} \alpha^2_{k2^j+2^{j-1}+i}(p,q) \right) \right) \\
&\leqslant \frac{2^j}{2^j} \sum_{k=0}^{+\infty} \sum_{p=1}^{m} \sum_{q=1}^{m} \sum_{l=0}^{2^j-1} \alpha^2_{k2^j+l}(p,q) \\
&= \sum_{h=0}^{+\infty} \sum_{p=1}^{m} \sum_{q=1}^{m} \alpha_h^2(p,q) \\
&= \sum_{h=0}^{+\infty} \operatorname{Tr}\left(\alpha_h \alpha_h'\right),
\end{aligned}
$$

where $\sum_{h=0}^{\infty} \operatorname{Tr}\left(\alpha_h \alpha_h'\right) = \|x_t\|^2_{\bar{\varphi}}$ is finite because x_t belongs to $\mathcal{H}_t(\boldsymbol{\varepsilon})$. Observe that we used twice the inequality

$$
\left(\sum_{i=0}^{n} a_i\right)^2 \leqslant (n+1) \sum_{i=0}^{n} a_i^2, \qquad n \in \mathbb{N}.
$$

As a result $g_t^{(j)}$ belongs to $\mathbf{R}^{j-1}\mathcal{L}_t^{\mathbf{R}}$. Furthermore, since $\sum_{k=0}^{\infty} \operatorname{Tr}\left(\beta_k^{(j)} \beta_k^{(j)\prime}\right)$ is finite, for any fixed scale $j \in \mathbb{N}$ we have that $\operatorname{Tr}\left(\beta_k^{(j)} \beta_k^{(j)\prime}\right)$ tends to zero as k increases. Hence, any (p,q)-entry of the matrix $\beta_k^{(j)}$ must converge to zero and so $\lim_{k\to\infty} \beta_k^{(j)} = \mathbf{0}$ entry by entry.

In order to find the exact expression of the matrices $\beta_k^{(j)}$, we exploit the orthogonal decompositions of the Hilbert A-module $\mathcal{H}_t(\boldsymbol{\varepsilon})$ at different scales $J \in \mathbb{N}$:

$$
\mathcal{H}_t(\boldsymbol{\varepsilon}) = \mathbf{R}^J \mathcal{H}_t(\boldsymbol{\varepsilon}) \oplus \bigoplus_{j=1}^{J} \mathbf{R}^{j-1} \mathcal{L}_t^{\mathbf{R}}.
$$

We call $\pi_t^{(j)}$ the orthogonal projection of x_t on the submodule $\mathbf{R}^j \mathcal{H}_t(\boldsymbol{\varepsilon})$ and we proceed inductively.

Let us start by the first decomposition $x_t = \pi_t^{(1)} + g_t^{(1)}$ coming from scale $J = 1$:

$$
\mathcal{H}_t(\boldsymbol{\varepsilon}) = \mathbf{R} \mathcal{H}_t(\boldsymbol{\varepsilon}) \oplus \mathcal{L}_t^{\mathbf{R}}.
$$

By using the characterization of submodules $\mathbf{R}\mathcal{H}_t(\boldsymbol{\varepsilon})$ and $\mathcal{L}_t^{\mathbf{R}}$, as remarked in Section 3.1 and stated in Proposition 21 respectively, we set

$$\pi_t^{(1)} = \sum_{k=0}^{+\infty} \gamma_k^{(1)} \left(\varepsilon_{t-2k} + \varepsilon_{t-(2k+1)}\right),$$

$$g_t^{(1)} = \sum_{k=0}^{+\infty} \beta_k^{(1)} \varepsilon_{t-2k}^{(1)} = \sum_{k=0}^{+\infty} d_k^{(1)} \left(\varepsilon_{t-2k} - \varepsilon_{t-2k-1}\right)$$

for some sequences of matrices $\left\{\gamma_k^{(1)}\right\}_k$ and $\left\{d_k^{(1)}\right\}_k$, or equivalently $\left\{\beta_k^{(1)}\right\}_k$, to determine in order to have $x_t = \pi_t^{(1)} + g_t^{(1)}$, where we set $\sqrt{2}d_k^{(1)} = \beta_k^{(1)}$. The expressions above may be rewritten as

$$x_t = \sum_{k=0}^{+\infty} \left\{ \left(\gamma_k^{(1)} + d_k^{(1)}\right) \varepsilon_{t-2k} + \left(\gamma_k^{(1)} - d_k^{(1)}\right) \varepsilon_{t-2k-1} \right\}.$$

However, from the Multivariate Classical Wold Decomposition of $\mathbf{x}$, we know that

$$x_t = \sum_{k=0}^{+\infty} \left\{ \alpha_{2k}\varepsilon_{t-2k} + \alpha_{2k+1}\varepsilon_{t-2k-1} \right\},$$

where we use the same fundamental innovations as before. By exploiting the uniqueness of writing deriving from the Classical Wold Decomposition, the two expressions for x_t must coincide. As a result, $\gamma_k^{(1)}$ and $d_k^{(1)}$ are the solutions of the linear system

$$\begin{cases} \gamma_k^{(1)} + d_k^{(1)} & = \alpha_{2k} \\ \gamma_k^{(1)} - d_k^{(1)} & = \alpha_{2k+1}, \end{cases}$$

that is,

$$\gamma_k^{(1)} = \frac{\alpha_{2k} + \alpha_{2k+1}}{2}, \qquad d_k^{(1)} = \frac{\alpha_{2k} - \alpha_{2k+1}}{2}.$$

In particular, we find that

$$\beta_k^{(1)} = \frac{\alpha_{2k} - \alpha_{2k+1}}{\sqrt{2}}.$$

Hence,

$$\pi_t^{(1)} = \sum_{k=0}^{+\infty} \frac{\alpha_{2k} + \alpha_{2k+1}}{2} \left(\varepsilon_{t-2k} + \varepsilon_{t-2k-1}\right), \qquad g_t^{(1)} = \sum_{k=0}^{+\infty} \frac{\alpha_{2k} - \alpha_{2k+1}}{\sqrt{2}} \varepsilon_{t-2k}^{(1)}.$$

Now, focus on the scale $J = 2$. We exploit the decomposition of the submodule

$$\mathbf{R}\mathcal{H}_t(\boldsymbol{\varepsilon}) = \mathbf{R}^2\mathcal{H}_t(\boldsymbol{\varepsilon}) \oplus \mathbf{R}\mathcal{L}_t^{\mathbf{R}},$$

that implies the relation $\pi_t^{(1)} = \pi_t^{(2)} + g_t^{(2)}$. We follow the same track as in the previous case, by using the features of the elements in $\mathbf{R}^2\mathcal{H}_t(\boldsymbol{\varepsilon})$ and in $\mathbf{R}\mathcal{L}_t^{\mathbf{R}}$ and, finally, by comparing the expression of $\pi_t^{(2)} + g_t^{(2)}$ with the (unique) writing of $\pi_t^{(1)}$ that we found before. We easily discover that

$$\gamma_k^{(2)} = \frac{\alpha_{4k} + \alpha_{4k+1} + \alpha_{4k+2} + \alpha_{4k+3}}{4}, \qquad d_k^{(2)} = \frac{\alpha_{4k} + \alpha_{4k+1} - \alpha_{4k+2} - \alpha_{4k+3}}{4}$$

and, in particular,

$$\beta_k^{(2)} = \frac{\alpha_{4k} + \alpha_{4k+1} - \alpha_{4k+2} - \alpha_{4k+3}}{2}.$$

Consequently,

$$\pi_t^{(2)} = \sum_{k=0}^{+\infty} \frac{\alpha_{4k} + \alpha_{4k+1} + \alpha_{4k+2} + \alpha_{4k+3}}{4} \left(\varepsilon_{t-4k} + \varepsilon_{t-(4k+1)} + \varepsilon_{t-(4k+2)} + \varepsilon_{t-(4k+3)}\right),$$

$$g_t^{(2)} = \sum_{k=0}^{+\infty} \frac{\alpha_{4k} + \alpha_{4k+1} - \alpha_{4k+2} - \alpha_{4k+3}}{2} \varepsilon_{t-4k}^{(2)}.$$

As for the generic scale $J = j$, we conjecture that

$$\gamma_k^{(j)} = \frac{1}{2^j} \left(\sum_{i=0}^{2^j-1} \alpha_{k2^j+i} \right), \qquad \beta_k^{(j)} = \frac{1}{\sqrt{2^j}} \left(\sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+2^{j-1}+i} \right).$$

In addition, it may be helpful to highlight the expressions of $\pi_t^{(j)}$ and $g_t^{(j)}$:

$$\pi_t^{(j)} = \sum_{k=0}^{+\infty} \frac{1}{2^j} \left(\sum_{i=0}^{2^j-1} \alpha_{k2^j+i} \right) \left(\sum_{i=0}^{2^j-1} \varepsilon_{t-k2^j-i} \right),$$

$$g_t^{(j)} = \sum_{k=0}^{+\infty} \frac{1}{\sqrt{2^j}} \left(\sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+2^{j-1}+i} \right) \varepsilon_{t-k2^j}^{(j)}.$$

Proof of Theorem 4

The representation of x_t comes from the Wold decomposition of the space $\mathcal{H}_t(\boldsymbol{\varepsilon})$, claimed in Theorem 3. Indeed, by applying the Multivariate Classical Wold Decomposition to the zero-mean, weakly stationary purely non-deterministic process $\mathbf{x}$, we find that x_t belongs to the Hilbert A-module $\mathcal{H}_t(\boldsymbol{\varepsilon})$, where $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_t$ is the unit variance white noise of classical Wold innovations of $\mathbf{x}$. Afterwards, by exploiting the orthogonal decomposition of $\mathcal{H}_t(\boldsymbol{\varepsilon})$ given by Theorem 3, justified by the fact that the scaling operator $\mathbf{R}$ is isometric on $\mathcal{H}_t(\boldsymbol{\varepsilon})$, we know that

$$\mathcal{H}_t(\boldsymbol{\varepsilon}) = \bigoplus_{j=1}^{+\infty} \mathbf{R}^{j-1} \mathcal{L}_t^{\mathbf{R}},$$

where

$$\mathbf{R}^{j-1}\mathcal{L}_t^{\mathbf{R}} = \left\{ \sum_{k=0}^{+\infty} b_k^{(j)} \varepsilon_{t-k2^j}^{(j)} \in \mathcal{H}_t(\boldsymbol{\varepsilon}) : \quad b_k^{(j)} \in A \right\},$$

as stated in Proposition 22. Recall that the random vectors $\varepsilon_t^{(j)}$ are defined by

$$\varepsilon_t^{(j)} = \frac{1}{\sqrt{2^j}} \left(\sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-i} - \sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-2^{j-1}-i} \right).$$

Hence, by denoting $g_t^{(j)}$ the orthogonal projections of the vector x_t on the submodules $\mathbf{R}^{j-1}\mathcal{L}_t^{\mathbf{R}}$, we find that

$$x_t = \sum_{j=1}^{+\infty} g_t^{(j)},$$

where the equality is in norm. Then, by using the characterizations of submodules $\mathbf{R}^{j-1}\mathcal{L}_t^{\mathbf{R}}$, for any scale $j \in \mathbb{N}$ we can find a sequence of matrices $\left\{ \beta_k^{(j)} \right\}_k$ such that

$$g_t^{(j)} = \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)}$$

with $\sum_{k=0}^{\infty} \operatorname{Tr}\left(\beta_k^{(j)} \beta_k^{(j)\prime} \right) < +\infty$. As a consequence, we can decompose the vector x_t as

$$x_t = \sum_{j=1}^{+\infty} \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)},$$

where the equality is in norm.

i) As we can see in the definition of vectors $\varepsilon_t^{(j)}$, the process $\boldsymbol{\varepsilon_t^{(j)}}$ is an $MA(2^j - 1)$ with respect to the fundamental innovations $\boldsymbol{\varepsilon}$ of the process $\mathbf{x}$. In addition, as claimed in Proposition 20, the subprocess $\left\{ \varepsilon_{t-k2^j}^{(j)} \right\}_{k \in \mathbb{Z}}$ is a unit variance white noise.

ii) For any fixed scale $j \in \mathbb{N}$, once the detail process $\boldsymbol{\varepsilon^{(j)}}$ is defined, since the vectors $\varepsilon_{t-k2^j}^{(j)}$ are orthonormal when k varies, the component $g_t^{(j)}$ has a unique representation of the kind

$$g_t^{(j)} = \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)}.$$

Thus, the matrices $\beta_k^{(j)}$ are uniquely defined.

By Proposition 1, $\beta_k^{(j)}$ do not depend on t and they can be expressed in terms of the matrices α_h of the Classical Wold Decomposition of x_t:

$$\beta_k^{(j)} = \frac{1}{\sqrt{2^j}} \left(\sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+2^{j-1}+i} \right).$$

Moreover $\sum_{k=0}^{\infty} \operatorname{Tr}\left(\beta_k^{(j)} \beta_k^{(j)\prime}\right) < +\infty$ for any $j \in \mathbb{N}$, as explicitly shown in the proof of Proposition 1. Indeed, it holds that

$$\sum_{k=0}^{+\infty} \operatorname{Tr}\left(\beta_k^{(j)} \beta_k^{(j)\prime}\right) \leqslant \sum_{h=0}^{+\infty} \operatorname{Tr}\left(\alpha_h \alpha_h'\right) < +\infty.$$

iii) First of all, when t is fixed, $\langle g_t^{(j)}, g_t^{(l)} \rangle_H = \mathbb{E}\left[g_t^{(j)} g_t^{(l)\prime}\right] = \mathbf{0}$ for all $j \neq l$ because $g_t^{(j)}$ and $g_t^{(l)}$ are, respectively, the projections of x_t on the submodules $\mathbf{R}^{j-1}\mathcal{L}_t^{\mathbf{R}}$ and $\mathbf{R}^{l-1}\mathcal{L}_t^{\mathbf{R}}$ which are orthogonal by construction. Now, consider any $g_{t-m2^j}^{(j)}$ with $m \in \mathbb{N}_0$. Clearly, $g_{t-m2^j}^{(j)}$ belongs to $\mathbf{R}^{j-1}\mathcal{L}_{t-m2^j}^{\mathbf{R}}$ but, by the definition of $g_t^{(j)}$, we can write

$$g_{t-m2^j}^{(j)} = \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-(m+k)2^j}^{(j)} = \sum_{K=0}^{+\infty} \beta_K^{(j)} \varepsilon_{t-K2^j}^{(j)},$$

where

$$\beta_K^{(j)} = \begin{cases} \mathbf{0} & \text{if } K \in \{0, \dots, m-1\}, \\ \beta_k^{(j)} & \text{if } K = m+k \quad \text{for some } k \in \mathbb{N}_0. \end{cases}$$

As a result, $g_{t-m2^j}^{(j)}$ belongs to $\mathbf{R}^{j-1}\mathcal{L}_t^{\mathbf{R}}$, too. Similarly, at scale l, taken any $n \in \mathbb{N}_0$, it is easy to see that $g_{t-n2^l}^{(l)}$ belongs to $\mathbf{R}^{l-1}\mathcal{L}_t^{\mathbf{R}}$. Hence, the orthogonality of such submodules guarantees that

$$\langle g_{t-m2^j}^{(j)}, g_{t-n2^l}^{(l)} \rangle_H = \mathbb{E}\left[g_{t-m2^j}^{(j)} g_{t-n2^l}^{(l)\prime}\right] = \mathbf{0} \qquad \forall j \neq l, \quad \forall m, n \in \mathbb{N}_0.$$

As for the more general requirement concerning $\mathbb{E}\left[g_{t-p}^{(j)} g_{t-q}^{(l)\prime}\right]$ for any $j, l \in \mathbb{N}$ and $p, q, t \in \mathbb{Z}$, we have that

$$\begin{aligned}
\mathbb{E}\left[g_{t-p}^{(j)} g_{t-q}^{(l)\prime}\right] &= \sum_{k=0}^{+\infty} \sum_{h=0}^{+\infty} \beta_k^{(j)} \mathbb{E}\left[\varepsilon_{t-p-k2^j}^{(j)} \varepsilon_{t-q-h2^l}^{(l)\prime}\right] \beta_h^{(l)\prime} \\
&= \frac{1}{\sqrt{2^{j+l}}} \sum_{k=0}^{+\infty} \sum_{h=0}^{+\infty} \beta_k^{(j)} \sum_{u=0}^{2^{j-1}-1} \sum_{v=0}^{2^{l-1}-1} \Big\{ \mathbb{E}\left[\varepsilon_{t-p-k2^j-u} \varepsilon'_{t-q-h2^l-v}\right] \\
&\quad - \mathbb{E}\left[\varepsilon_{t-p-k2^j-u} \varepsilon'_{t-q-h2^l-2^{l-1}-v}\right] - \mathbb{E}\left[\varepsilon_{t-p-k2^j-2^{j-1}-u} \varepsilon'_{t-q-h2^l-v}\right] \\
&\quad + \mathbb{E}\left[\varepsilon_{t-p-k2^j-2^{j-1}-u} \varepsilon'_{t-q-h2^l-2^{l-1}-v}\right] \Big\} \beta_h^{(l)\prime}
\end{aligned}$$

and so

$$\mathbb{E}\left[g_{t-p}^{(j)} g_{t-q}^{(l)\,\prime}\right] = \frac{1}{\sqrt{2^{j+l}}} \sum_{k=0}^{+\infty} \sum_{h=0}^{+\infty} \beta_k^{(j)} \sum_{u=0}^{2^{j-1}-1} \sum_{v=0}^{2^{l-1}-1} \left\{ \Gamma_{p-q+k2^j+u-h2^l-v} \right.$$
$$- \Gamma_{p-q+k2^j+u-h2^l-2^{l-1}-v}$$
$$- \Gamma_{p-q+k2^j+2^{j-1}+u-h2^l-v}$$
$$\left. + \Gamma_{p-q+k2^j+2^{j-1}+u-h2^l-2^{l-1}-v} \right\} \beta_h^{(l)\prime},$$

where the matrices $\beta_k^{(j)}$, $\beta_h^{(l)}$ do not depend on t, as remarked in ii). Hence, after the summations over u, v and k, h, the one remaining variables are $j, l, p-q$. In other words, $\mathbb{E}\left[g_{t-p}^{(j)} g_{t-q}^{(l)\,\prime}\right]$ depends at most on $j, l, p-q$.

■

Proof of Theorem 5

First, observe that processes $\mathbf{g}^{(\mathbf{j})}$ are well-defined. Indeed,

$$g_t^{(j)} = \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)} = \sum_{h=0}^{+\infty} \frac{1}{\sqrt{2^j}} \beta_{\lfloor \frac{h}{2^j} \rfloor}^{(j)} \chi^{(j)}(h) \varepsilon_{t-h},$$

where[60]

$$\chi^{(j)}(h) = \begin{cases} -1 & \text{if} \quad 2^j \left\lfloor \frac{h}{2^j} \right\rfloor \in \{h - 2^j + 1, \dots, h - 2^{j-1}\}, \\ 1 & \text{if} \quad 2^j \left\lfloor \frac{h}{2^j} \right\rfloor \in \{h - 2^{j-1} + 1, \dots, h\}. \end{cases}$$

Hence,

$$\sum_{h=0}^{+\infty} \operatorname{Tr}\left(\frac{1}{\sqrt{2^j}} \beta_{\lfloor \frac{h}{2^j} \rfloor}^{(j)} \chi^{(j)}(h) \frac{1}{\sqrt{2^j}} \beta_{\lfloor \frac{h}{2^j} \rfloor}^{(j)\,\prime} \chi^{(j)}(h) \right) = \sum_{h=0}^{+\infty} \frac{1}{2^j} \operatorname{Tr}\left(\beta_{\lfloor \frac{h}{2^j} \rfloor}^{(j)} \beta_{\lfloor \frac{h}{2^j} \rfloor}^{(j)\,\prime} \right),$$

where the last quantity equals $\sum_{k=0}^{+\infty} \operatorname{Tr}\left(\beta_k^{(j)} \beta_k^{(j)\prime}\right)$, which is finite by assumption. As a result, each $\mathbf{g}^{(\mathbf{j})}$ is well-defined.

For any $j \neq l$, the components $g_t^{(j)}$ and $g_t^{(l)}$ belong to orthogonal submodules of $\mathcal{H}_t(\boldsymbol{\varepsilon})$. This ensures the well-definition of x_t:

$$x_t = \sum_{j=1}^{+\infty} g_t^{(j)} = \sum_{j=1}^{+\infty} \sum_{h=0}^{+\infty} \frac{1}{\sqrt{2^j}} \beta_{\lfloor \frac{h}{2^j} \rfloor}^{(j)} \chi^{(j)}(h) \varepsilon_{t-h} = \sum_{h=0}^{+\infty} \left(\sum_{j=1}^{+\infty} \frac{1}{\sqrt{2^j}} \beta_{\lfloor \frac{h}{2^j} \rfloor}^{(j)} \chi^{(j)}(h) \right) \varepsilon_{t-h}$$

[60]We used the identity $h = k2^j + i$, with $k = \left\lfloor \frac{h}{2^j} \right\rfloor$ and $i = h - 2^j \left\lfloor \frac{h}{2^j} \right\rfloor$.

because

$$\sum_{h=0}^{+\infty} \operatorname{Tr}\left(\sum_{j=1}^{+\infty} \frac{1}{\sqrt{2^j}} \beta^{(j)}_{\left\lfloor \frac{h}{2^j} \right\rfloor} \chi^{(j)}(h) \sum_{l=1}^{+\infty} \frac{1}{\sqrt{2^l}} \beta^{(l)}_{\left\lfloor \frac{h}{2^l} \right\rfloor}{}' \chi^{(l)}(h)\right)$$

$$= \sum_{h=0}^{+\infty} \sum_{j=1}^{+\infty} \operatorname{Tr}\left(\frac{1}{\sqrt{2^j}} \beta^{(j)}_{\left\lfloor \frac{h}{2^j} \right\rfloor} \chi^{(j)}(h) \frac{1}{\sqrt{2^j}} \beta^{(j)}_{\left\lfloor \frac{h}{2^j} \right\rfloor}{}' \chi^{(j)}(h)\right)$$

$$= \sum_{j=1}^{+\infty} \sum_{h=0}^{+\infty} \frac{1}{2^j} \operatorname{Tr}\left(\beta^{(j)}_{\left\lfloor \frac{h}{2^j} \right\rfloor} \beta^{(j)}_{\left\lfloor \frac{h}{2^j} \right\rfloor}{}'\right)$$

$$= \sum_{j=1}^{+\infty} \sum_{k=0}^{+\infty} \operatorname{Tr}\left(\beta^{(j)}_k \beta^{(j)\prime}_k\right),$$

which is finite by assumption.

Now we show that $\mathbf{x}$ is a zero-mean weakly stationary multivariate process.

i) The well-definition of x_t already ensures that the second moments of each variable $x_{t,i}$ are finite and not dependent on t.

ii) Since the processes $\mathbf{g}^{(\mathbf{j})}$ have zero mean, also x_t has zero mean for all $t \in \mathbb{Z}$.

iii) For any $k \in \mathbb{Z}$ consider the cross moments matrix

$$\mathbb{E}\left[x_t x_{t+k}'\right] = \mathbb{E}\left[\left(\sum_{j=1}^{+\infty} g_t^{(j)}\right)\left(\sum_{l=1}^{+\infty} g_{t+k}^{(l)}{}'\right)\right] = \sum_{j=1}^{+\infty} \sum_{l=1}^{+\infty} \mathbb{E}\left[g_t^{(j)} g_{t+k}^{(l)}{}'\right].$$

Following the same steps of Theorem 4 *iii*), we have that $\mathbb{E}\left[g_t^{(j)} g_{t+k}^{(l)}{}'\right]$ depends at most on j, l, k. Hence, $\mathbb{E}\left[x_t x_{t+k}'\right]$ depends at most on k.

It follows that $\mathbf{x}$ is weakly stationary, with zero mean.

As we remarked in the previous lines, each variable x_t decomposes as

$$x_t = \sum_{h=0}^{+\infty}\left(\sum_{j=1}^{+\infty} \frac{1}{\sqrt{2^j}} \beta^{(j)}_{\left\lfloor \frac{h}{2^j} \right\rfloor} \chi^{(j)}(h)\right) \varepsilon_{t-h}.$$

As a result, the decomposition of x_t with respect to the process $\boldsymbol{\varepsilon}$ turns out to be

$$x_t = \sum_{h=0}^{+\infty} \alpha_h \varepsilon_{t-h} \qquad \text{with} \qquad \alpha_h = \sum_{j=1}^{+\infty} \frac{1}{\sqrt{2^j}} \beta^{(j)}_{\left\lfloor \frac{h}{2^j} \right\rfloor} \chi^{(j)}(h) \quad \forall h \in \mathbb{N}_0.$$

The process $\mathbf{x}$ has null purely deterministic part, therefore it is purely non-deterministic.
■