



Multivariate Wold decompositions: a Hilbert A -module approach

Simone Cerreia-Vioglio¹ · Fulvio Ortu² · Federico Severino³  · Claudio Tebaldi⁴

Received: 18 December 2021 / Accepted: 1 March 2023 / Published online: 4 April 2023

© The Author(s), under exclusive licence to Associazione per la Matematica Applicata alle Scienze Economiche e Sociali (AMASES) 2023

Abstract

Orthogonal decompositions are essential tools for the study of weakly stationary time series. Some examples are given by the classical Wold decomposition of Wold (A study in the analysis of stationary time series, Almqvist & Wiksells Boktryckeri, Uppsala, 1938) and the extended Wold decomposition of Ortu et al. (Quant Econ 11(1):203–230, 2020), which permits to disentangle shocks with heterogeneous degrees of persistence from a given weakly stationary process. The analysis becomes more involved when dealing with vector processes because of the presence of different simultaneous shocks. In this paper, we recast the standard treatment of multivariate time series in terms of Hilbert A -modules (where matrices replace the field of scalars) and we prove the abstract Wold theorem for self-dual pre-Hilbert A -modules with an isometric operator. This theorem allows us to easily retrieve the multivariate classical Wold

Simone Cerreia-Vioglio, Fulvio Ortu and Claudio Tebaldi have contributed equally to this work.

✉ Federico Severino
federico.severino@fsa.ulaval.ca

Simone Cerreia-Vioglio
simone.cerreia@unibocconi.it

Fulvio Ortu
fulvio.ortu@unibocconi.it

Claudio Tebaldi
claudio.tebaldi@unibocconi.it

¹ Department of Decision Sciences and IGIER, Università Bocconi, Via Roentgen 1, 20136 Milano, MI, Italy

² Department of Finance and IGIER, Università Bocconi, Via Roentgen 1, 20136 Milano, MI, Italy

³ Department of Finance, Insurance and Real Estate, Université Laval, 2325 Rue de la Terrasse, Quebec, QC G1V0A6, Canada

⁴ Department of Finance, IGIER and CAREFIN, Università Bocconi, Via Roentgen 1, 20136 Milano, MI, Italy

decomposition and the multivariate version of the extended Wold decomposition. The theory helps in handling matrix coefficients and computing orthogonal projections on closed submodules. The orthogonality notion is key to decompose the given vector process into uncorrelated subseries, and it implies a variance decomposition.

Keywords Multivariate stationary processes · Wold decomposition · Persistence heterogeneity · Hilbert A -modules

JEL Classification C18 · C32 · C50

Abbreviations

MCWD Multivariate classical Wold decomposition

MEWD Multivariate extended Wold decomposition

1 Introduction

A vector random 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 randomness [as ensured by the Wold decomposition Wold (1938)], each variable $x_{i,t}$ of a weakly stationary multivariate process can be affected by m possibly different shocks $\varepsilon_{1,t}, \varepsilon_{2,t}, \dots, \varepsilon_{m,t}$. The representation of vector processes involves matrix coefficients, and their study has proved to be fruitful in numerous macroeconomic and financial applications (Lütkepohl 2005).

In this paper, we recast the standard treatment of multivariate time series in terms of Hilbert A -modules and prove the Abstract Wold Theorem for Hilbert A -modules (Theorem 1). Our abstract A -module framework features a notion of orthogonality that, by means of Theorem 1, permits to easily retrieve two important orthogonal decompositions for weakly stationary vector processes that we illustrate in Sects. 3 and 4, respectively. One is the celebrated multivariate classical Wold decomposition (MCWD, henceforth), summarized in Theorem 2. The other is the multivariate extended Wold decomposition (MEWD, henceforth), which constitutes the multivariate version of the extended Wold decomposition¹ of Ortu et al. (2020a), and it is used by Bandi et al. (2019, 2021) in financial economics settings. See Theorem 4.

Both the MCWD and the MEWD rely on orthogonal innovations. However, only in the MEWD shocks are associated with different degrees of persistence. In econometrics, persistence is usually addressed by spectral analysis techniques to analyze the frequency domain. For instance, cross-spectrum and squared coherency may be used to quantify the linear association between single time series in a vector process (Brockwell and Davis 2006, Section 11.6). On the contrary, the MEWD permits to disentangle uncorrelated persistent components from a weakly stationary vector process by using exclusively the time domain. Each vector component is associated with

¹ An application of the extended Wold decomposition to market returns is provided in Di Virgilio et al. (2019).

a specific persistence level (or time scale), and it is sensible to a family of shocks with precise duration. With respect to the univariate extended Wold decomposition of Ortu et al. (2020a), the illustration of the MEWD is richer because of the persistence of different simultaneous shocks.

To derive the MCWD and the MEWD, we first embed multivariate time series in an abstract A -module framework from which orthogonal decompositions naturally arise. We use the algebra A of square matrices, define orthogonal projections on closed submodules and prove the Abstract Wold Theorem for Hilbert A -modules, which is key to decompose vector processes into the sum of uncorrelated components. We provide a self-contained compendium on Hilbert A -modules on a non-commutative and finite dimensional algebra (as the one of matrices) in “Appendix A”. In fact, the application of Hilbert A -modules in economic theory and statistics is not novel. Some examples are given by Hansen and Richard (1987), Gallant et al. (1990), Wiener and Masani (1957) and Cerreia-Vioglio et al. (2022).

To enter more the details of our construction, we consider the vector space of square-integrable m -dimensional random vectors and we substitute the field of scalars with the algebra of $m \times m$ matrices, obtaining an A -module H . We, then, endow H with an A -valued inner product which generalizes the inner product of L^2 and naturally conveys a notion of orthogonality. Such a structure is a Hilbert A -module. The properties of self-duality of H and complementability of closed submodules, that we discuss in Sect. 2.1, are crucial for the Abstract Wold Theorem for Hilbert A -modules (Theorem 1).

Such theorem generalizes the Abstract Wold Theorem for Hilbert spaces (Sz-Nagy et al. 2010, Theorem 1.1) that permits to orthogonally decompose a Hilbert space by an isometric operator. When this theorem is applied to the Hilbert space generated by the past realizations of a weakly stationary univariate time series with the *lag operator* as isometry, the classical Wold decomposition obtains (see, e.g., Wold 1938; Brockwell and Davis 2006, Section 5.7 or Severino 2016). The orthogonality induced by the theorem is responsible for the white noise of fundamental innovations. However, other choices for the isometry are possible. For instance, the univariate persistence-based decomposition of Ortu et al. (2020a) is obtained by using as isometry the *scaling operator* on the Hilbert space generated by the past fundamental innovations. In this decomposition, the orthogonality ensured by the Abstract Wold Theorem is retrieved in the absence of correlation between persistent components.

The MEWD for a weakly stationary vector process x_t comes from the application of Theorem 1 to the Hilbert A -module generated by the multivariate fundamental innovations given by the MCWD of x_t . The employed isometry is the scaling operator, adapted to A -modules. The A -module orthogonality given by the theorem guarantees that any entry of a multivariate persistent component at a given time scale is uncorrelated with any entry of any persistent component at a different scale (Theorem 4). This absence of correlation is crucial to associate each scale-specific response with the effect on the related time scale, without spurious correlation with the simultaneous effects at the other scales. In addition, the orthogonality of components induces a variance decomposition that permits to classify each entry of x_t as a short-, medium- or long-term process, according to the variance explained by the persistent components (Sect. 4.3).

The next section introduces the Hilbert A -module framework in which we embed weakly stationary vector processes and illustrates Theorem 1. Section 3 revisits the MCWD by emphasizing his connection with Theorem 1. Section 4 describes the MEWD and the persistence-based variance decomposition. As to appendices, “Appendix A” provides a comprehensive and self-contained treatment of Hilbert A -modules and Theorem 1, “Appendix B” contains the proofs about the MEWD and “Appendix C” provides some illustrations of the MEWD, including an application to the bivariate macroeconomic model of Blanchard and Quah (1989).

2 Hilbert A -modules for multivariate time series

In the first subsection, we condense the notions of Hilbert A -module theory that lead to the Abstract Wold Theorem for Hilbert A -modules (Theorem 1). “Appendix A” contains all the details (see also Cerreia-Vioglio et al. 2017, 2019). After that, we describe the Hilbert A -module $L^2(\mathbb{R}^m, \Omega, \mathcal{F}, \mathbb{P})$ and the submodules that we will need for the MCWD and the MEWD.

2.1 Hilbert A -modules

Hilbert A -modules are a generalization of Hilbert spaces where the scalar field \mathbb{R} is replaced by an abstract algebra A . Although these structures have been studied in the mathematical literature (Kaplansky 1953), few works deal with the case of our interest, where the algebra is real, non-commutative and finite dimensional (Goldstine and Horwitz 1966). Differently from Hilbert spaces, the Riesz Theorem about the representation of linear functionals (Theorem 5.5 in Brezis 2011) is not always valid and closed subspaces are not always complemented. Moreover, the self-duality property that we discuss is key in deriving Theorem 1.

We consider a real normed operator algebra A (the algebra of square matrices) with norm $\|\cdot\|_A$, an involution $*$: $A \rightarrow A$, an order \geq and a trace functional $\bar{\varphi} : A \rightarrow \mathbb{R}$. Then, we consider an A -module H with outer product $\cdot : A \times H \rightarrow H$ and we define an A -valued inner product $\langle \cdot, \cdot \rangle_H : H \times H \rightarrow A$, which satisfies the A -valued versions of the usual properties of inner products. This makes H a pre-Hilbert A -module.

A -valued operators $f : H \rightarrow A$ constitute the generalization of linear functionals, and the properties of A -linearity and boundedness can be defined accordingly. Importantly, H is *self-dual* when for each $f : H \rightarrow A$ which is A -linear and bounded there exists $y \in H$ such that $f(x) = \langle x, y \rangle_H$ for all $x \in H$. In other words, a version of the Riesz Theorem holds.

To put the theory at work, H can be endowed with two (real-valued) norms: $\|\cdot\|_{\bar{\varphi}}$ and $\|\cdot\|_H$. The first one is $\|\cdot\|_H : H \rightarrow [0, +\infty)$ defined by

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

The second one is induced by the (real-valued) inner product $\langle \cdot, \cdot \rangle_{\bar{\varphi}} : H \times H \rightarrow \mathbb{R}$ defined by $\langle x, y \rangle_{\bar{\varphi}} = \bar{\varphi}(\langle x, y \rangle_H) \quad \forall x, y \in H$, which makes H a pre-Hilbert space. The norm is $\|\cdot\|_{\bar{\varphi}} : H \rightarrow [0, +\infty)$ defined by

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

Proposition 12 in “Appendix A.3.1” shows that, if A is finite dimensional, then A admits a trace $\bar{\varphi}$ and the norms $\|\cdot\|_H$ and $\|\cdot\|_{\bar{\varphi}}$ are equivalent. In our construction, the use of $\|\cdot\|_{\bar{\varphi}}$ is convenient to establish the convergence of sequences in H , while $\langle \cdot, \cdot \rangle_H$ naturally induces the notion of orthogonality: $\langle x, y \rangle_H = 0$ with $x, y \in H$ and $0 \in A$.

We say that H is a Hilbert A -module when it is $\|\cdot\|_H$ complete. Theorem 15 in “Appendix A.5” establishes two characterizations of $\|\cdot\|_H$ completeness when A is finite dimensional: H is $\|\cdot\|_H$ complete if and only if it is $\|\cdot\|_{\bar{\varphi}}$ complete if and only if it is self-dual. The link between completeness and self-duality is, then, established.

The last ingredient for Theorem 1 regards the orthogonal complement of a given submodule $M \subseteq H$, i.e., $M^\perp = \{x \in H : \langle x, y \rangle_H = 0 \quad \forall y \in M\}$. By Proposition 17 in “Appendix A.5.1”, if A is finite dimensional and H is self-dual, M is $\|\cdot\|_H$ closed if and only if $H = M \oplus M^\perp$ (complementability). The projection map on closed submodules is, then, well-defined.

The decomposition in Theorem 1 is due to an isometry $\mathbf{T} : H \rightarrow H$, i.e., an A -linear operator such that $\langle \mathbf{T}x, \mathbf{T}y \rangle_H = \langle x, y \rangle_H$ for all $x, y \in H$. Moreover, the theorem prescribes the determination of a wandering submodule L such that $\mathbf{T}^n L \perp \mathbf{T}^m L$ for all $m, n \in \mathbb{N}_0$ with $m \neq n$. If H is self-dual, a wandering submodule is $L = (\mathbf{T}H)^\perp$.

Theorem 1 (Abstract Wold Theorem for Hilbert A -modules) *Let A be finite dimensional and H a pre-Hilbert A -module. If H is self-dual and $\mathbf{T} : H \rightarrow H$ is an isometry, then $H = \widehat{H} \oplus \widetilde{H}$ where*

$$\widehat{H} = \bigcap_{n=0}^{\infty} \mathbf{T}^n H, \quad \widetilde{H} = \bigoplus_{n=0}^{\infty} \mathbf{T}^n L, \quad L = (\mathbf{T}H)^\perp.$$

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

Proof See “Appendix A.6”. ■

Theorem 1 provides a (generalized) version for Hilbert A -modules of the Abstract Wold Theorem for Hilbert spaces (Sz-Nagy et al. 2010, Theorem 1.1). The properties of self-duality and complementability play a crucial role in this wider setting.

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

A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is given and, as usual, any two \mathcal{F} -measurable random vectors are defined to be equivalent when they coincide almost surely. We, then, consider the vector space $L^2(\mathbb{R}^m, \Omega, \mathcal{F}, \mathbb{P})$ of (equivalence classes of) measurable square-integrable random vectors x that take value in \mathbb{R}^m . We build on $L^2(\mathbb{R}^m, \Omega, \mathcal{F}, \mathbb{P})$ the structure of pre-Hilbert A -module, and we denote it by H . The whole discussion is summarized in Table 1.

Table 1 Relations between the general Hilbert A -module theory and the Hilbert A -module of square-integrable random vectors with zero mean

A-module theory	Random vectors	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 prod
Norm in A	$\ a\ _A = \sup_{x \in \mathbb{R}^m, \ x\ _2=1} \ ax\ _2$	Operator norm largest a' eigenvalue
Involution in A	$a^* = a'$	Matrix transposition
Order in A	$a - b \geq \mathbf{0}$	$a - b$ symmetric and positive semidefinite
Trace functional $\hat{\varphi} : A \rightarrow \mathbb{R}$	$\hat{\varphi} = \text{Tr}(a)$	Trace of matrix a
Hilbert A -module	$H = L^2(\mathbb{R}^m, \Omega, \mathcal{F}, \mathbb{P})$	Square-integrable random vectors
Outer prod. $\cdot : A \times H \rightarrow H$	$ax = [\dots, \sum_{k=1}^m a(i, k)x_k, \dots]'$	Matrix-by-vector prod
Inner prod. $\langle \cdot, \cdot \rangle_H : H \times H \rightarrow A$	$\langle x, y \rangle_H = \mathbb{E}[xy']$	Cross-covariance matrix
Squared norm $\ \cdot\ _H^2$	$\langle x, x \rangle_H = \mathbb{E}[xx']$	Covariance matrix
Orthogonality $\langle x, y \rangle_H = 0$	$\ x\ _H^2 = \lambda_{\max}$	Largest $\mathbb{E}[xx']$ eigenvalue
Inner prod. $\langle \cdot, \cdot \rangle_{\hat{\varphi}} : H \times H \rightarrow \mathbb{R}$	$\mathbb{E}[x_i y_j] = 0 \quad \forall i, j = 1, \dots, m$	Any x_i uncorrelated with any y_j
Squared norm $\ \cdot\ _{\hat{\varphi}}^2$	$\langle x, y \rangle_{\hat{\varphi}} = \text{Tr}(\mathbb{E}[xy']) = \sum_{i=1}^m \mathbb{E}[x_i y_i]$	$L^2(\mathbb{R}^m)$ inner prod
	$\ x\ _{\hat{\varphi}}^2 = \text{Tr}(\mathbb{E}[xx']) = \sum_{i=1}^m \mathbb{E}[x_i^2]$	$L^2(\mathbb{R}^m)$ squared norm sum of $\mathbb{E}[xx']$ eigenvalues

First of all, 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 $\| \cdot \|_A$ such that, for any $a = \{a_{i,j}\}_{i,j=1,\dots,m}$ in A ,

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

where $\| \cdot \|_2$ is the L^2 norm in \mathbb{R}^m . In particular, $\|a\|_A^2$ is the largest eigenvalue of the positive semidefinite matrix $a'a$ (Meyer 2000, Section 5.2), where a' is the transpose of a . The map that associates any matrix a with a' defines an involution in A . This map induces an order \geq such that, for any $a, b \in A$, $a \geq b$ when $a - b$ is a symmetric and positive semidefinite matrix (equivalently, $a - b \geq \mathbf{0}$).

We use as outer product $A \times H \rightarrow H$ the standard matrix-by-vector product. This operation makes H an A -module. Then, we define the A -valued inner product $\langle \cdot \rangle_H : H \times H \rightarrow A$ that associates any $x = [x_1, \dots, x_m]'$, $y = [y_1, \dots, y_m]' \in H$ with the matrix

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

$\langle \cdot, \cdot \rangle_H$ satisfies the usual properties of inner products. In addition, if x has zero mean, $\langle x, x \rangle_H$ is the covariance matrix of x . Importantly, two random vectors $x, y \in H$ are orthogonal when $\langle x, y \rangle_H = \mathbf{0}$, that is $\mathbb{E} [x_i y_j] = 0$ for all $i, j = 1, \dots, m$. If x and y have zero mean, this means that any x_i is uncorrelated with any y_j .

We now define the two equivalent (real-valued) norms $\| \cdot \|_H$ and $\| \cdot \|_{\bar{\varphi}}$. Regarding $\| \cdot \|_H : H \rightarrow [0, +\infty)$, we have

$$\|x\|_H = \sqrt{\langle x, x \rangle_H \|A\|} = \sqrt{\| \mathbb{E} [xx'] \|_A} \quad \forall x \in H.$$

If x has zero mean, $\|x\|_H^2$ is the largest eigenvalue of the covariance matrix of x .

To construct $\| \cdot \|_{\bar{\varphi}}$, we first consider the trace functional $\bar{\varphi} : A \rightarrow \mathbb{R}$ defined, for any matrix a , by its trace $\bar{\varphi}(a) = \text{Tr}(a)$. Indeed, H is a pre-Hilbert space with the inner product $\langle \cdot \rangle_{\bar{\varphi}} : H \times H \rightarrow \mathbb{R}$ defined by

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

which coincides with the usual inner product of $L^2(\mathbb{R}^m)$. The associated norm $\| \cdot \|_{\bar{\varphi}} : H \rightarrow [0, +\infty)$ is

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

If x has zero mean, $\|x\|_{\bar{\varphi}}^2$ is the sum of the eigenvalues of the covariance matrix of x .

Proposition 20 in ‘‘Appendix B.1’’ shows that H is $\| \cdot \|_{\bar{\varphi}}$ complete, i.e., it is a Hilbert A -module. Thus, the self-duality property holds.

2.3 The submodules generated by weakly stationary vector processes

Now, consider a zero-mean weakly stationary multivariate process $\mathbf{x} = \{x_t\}_{t \in \mathbb{Z}}$ such that $x_t = [x_{1,t}, \dots, x_{m,t}]' \in H$ for all $t \in \mathbb{Z}$. The autocovariance function $\Gamma : \mathbb{Z} \rightarrow A$ associates any integer n with the matrix Γ_n with entries $\Gamma_n(p, q) = \mathbb{E}[x_{p,t}x_{q,t+n}]$ for $p, q = 1, \dots, m$. If $\Gamma_n = \mathbf{0}$ for any $n \neq 0$, we are facing a *multivariate white noise*, which features *unit variance* when Γ_0 is the identity matrix. In this case, the single time series of the multivariate white noise are uncorrelated also simultaneously.

The sequence $\{x_{t-n}\}_{n \in \mathbb{N}_0}$ spans the Hilbert submodule $\mathcal{H}_t(\mathbf{x}) \subseteq H$ defined by

$$\mathcal{H}_t(\mathbf{x}) = \text{cl} \left\{ \sum_{k=0}^{+\infty} a_k x_{t-k} : a_k \in A, \sum_{k=0}^{+\infty} \sum_{h=0}^{+\infty} \text{Tr}(a_k \Gamma_{k-h} a'_h) < +\infty \right\} \quad (1)$$

with

$$\left\| \sum_{k=0}^{+\infty} a_k x_{t-k} \right\|_{\bar{\varphi}}^2 = \sum_{k=0}^{+\infty} \sum_{h=0}^{+\infty} \text{Tr}(a_k \Gamma_{k-h} a'_h).$$

As we will see in Sect. 3, when \mathbf{x} is regular, the multivariate classical Wold decomposition (MCWD) can be obtained by applying Theorem 1 to $\mathcal{H}_t(\mathbf{x})$ with the lag operator as isometry.

An outcome of the MCWD is the unit variance multivariate white noise of *fundamental innovations* $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t \in \mathbb{Z}}$. The sequence $\{\varepsilon_{t-n}\}_{n \in \mathbb{N}_0}$ generates the submodule $\mathcal{H}_t(\boldsymbol{\varepsilon}) \subseteq H$ given by

$$\mathcal{H}_t(\boldsymbol{\varepsilon}) = \left\{ \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k} : a_k \in A, \sum_{k=0}^{+\infty} \text{Tr}(a_k a'_k) < +\infty \right\} \quad (2)$$

with

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

Proposition 20 in ‘‘Appendix B.1’’ shows that $\mathcal{H}_t(\boldsymbol{\varepsilon})$ is a closed submodule of H and so it is a Hilbert submodule. In Sect. 4, we apply Theorem 1 to $\mathcal{H}_t(\boldsymbol{\varepsilon})$ with the scaling operator as isometry in order to derive the multivariate extended Wold decomposition (MEWD) of x_t .

3 Multivariate classical Wold decomposition

In the MCWD, a zero-mean regular weakly stationary vector process $\mathbf{x} = \{x_t\}_{t \in \mathbb{Z}}$ is decomposed into the infinite sum of uncorrelated multivariate innovations that occur at different times, plus a deterministic component (Theorem 7.2 in Bierens 2005). Wiener and Masani (1957) and Rozanov (1967) provide a proof in the complex field. Here, we provide the roadmap to derive the MCWD via the Abstract Wold Theorem

for Hilbert A -modules in the real case. A similar derivation in the univariate case is illustrated in Severino (2016).

We consider the Hilbert submodule $\mathcal{H}_t(\mathbf{x}) \subseteq H$ of Eq. (1). The lag operator $\mathbf{L} : \mathcal{H}_t(\mathbf{x}) \rightarrow \mathcal{H}_t(\mathbf{x})$ acts on generators of $\mathcal{H}_t(\mathbf{x})$ as

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

\mathbf{L} is A -linear and bounded; hence, it can be extended to $\mathcal{H}_t(\mathbf{x})$ with continuity. Importantly, \mathbf{L} is isometric on $\mathcal{H}_t(\mathbf{x})$. Theorem 1 can, then, be applied.

Theorem 1 requires to determine the images of $\mathcal{H}_t(\mathbf{x})$ through the powers of \mathbf{L} and the wandering submodule $\mathcal{L}_t^{\mathbf{L}}$. Since, in a self-dual pre-Hilbert A -module, the image of a closed submodule through an isometry is a closed submodule (Lemma 18 in “Appendix A.6”), one can prove that $\mathbf{L}^j \mathcal{H}_t(\mathbf{x}) = \mathcal{H}_{t-j}(\mathbf{x})$ for any $j \in \mathbb{N}$.

Then, it is possible to show that $\mathcal{H}_t(\mathbf{x})$ can be decomposed into the direct sum

$$\mathcal{H}_t(\mathbf{x}) = \mathcal{H}_{t-1}(\mathbf{x}) \oplus \text{span} \{x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t\},$$

i.e., $\mathcal{L}_t^{\mathbf{L}}$ is the linear span of $x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t$, where \mathcal{P}_M denotes the orthogonal projection on the closed submodule M .

Since \mathbf{x} is regular, for any $t \in \mathbb{Z}$, $\langle x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t, x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t \rangle_H$ is a symmetric positive definite matrix (Bierens 2012, Section 6). Hence, by Theorem 7.2.6 in Horn and Johnson (1990), there exists a symmetric positive definite square root matrix σ such that

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

As σ is invertible, we define the *fundamental innovation process* $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t \in \mathbb{Z}}$ by setting, for any $t \in \mathbb{Z}$, $\varepsilon_t = \sigma^{-1}(x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t)$. In particular, $\boldsymbol{\varepsilon}$ is a unit variance white noise.

It is, then, possible to show that the lag and the projection operator commute: for any $k, j \in \mathbb{N}_0$, $\mathbf{L}^j \mathcal{P}_{\mathcal{H}_{t-k-1}(\mathbf{x})} x_{t-k} = \mathcal{P}_{\mathcal{H}_{t-k-j-1}(\mathbf{x})} x_{t-k-j}$. Therefore, the covariance matrix of $x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t$ is not dependent on the time index $t \in \mathbb{Z}$ and, for any $j \in \mathbb{N}$,

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

As a result, Theorem 1 implies that $\mathcal{H}_t(\mathbf{x}) = \widehat{\mathcal{H}}_t(\mathbf{x}) \oplus \widetilde{\mathcal{H}}_t(\mathbf{x})$ with

$$\widehat{\mathcal{H}}_t(\mathbf{x}) = \bigcap_{j=0}^{+\infty} \mathcal{H}_{t-j}(\mathbf{x}), \quad \widetilde{\mathcal{H}}_t(\mathbf{x}) = \bigoplus_{j=0}^{+\infty} \text{span} \{x_{t-j} - \mathcal{P}_{\mathcal{H}_{t-j-1}(\mathbf{x})} x_{t-j}\}.$$

The consequences for the process \mathbf{x} are, then, 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*

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

where the equality is in norm and

1. $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t \in \mathbb{Z}}$ is a unit variance m -dimensional white noise;
2. for any $h \in \mathbb{N}_0$, the $m \times m$ matrices α_h do not depend on t ,

$$\alpha_h = \mathbb{E} [x_t \varepsilon'_{t-h}] \quad \text{and} \quad \sum_{h=0}^{+\infty} \text{Tr} (\alpha_h \alpha'_h) < +\infty;$$

3. $\mathbf{v} = \{v_t\}_{t \in \mathbb{Z}}$ is a zero-mean weakly stationary m -dimensional process,

$$v_t \in \bigcap_{j=0}^{+\infty} \mathcal{H}_{t-j}(\mathbf{x}) \quad \text{and} \quad \mathbb{E} [v_t \varepsilon'_{t-h}] = \mathbf{0} \quad \forall h \in \mathbb{N}_0;$$

- 4.

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

Proof Wiener and Masani (1957, Theorem 6.11) and Rozanov (1967, Chapter II, Section 3) provide a proof in the complex field. A detailed proof using real Hilbert A -modules is available upon request. \blacksquare

The process \mathbf{v} constitutes the (predictable) deterministic component of \mathbf{x} . If each v_t is the null vector, \mathbf{x} is a *purely non-deterministic* process.

In our approach, the *multivariate impulse responses* α_h are fully characterized by the projection on Hilbert submodules via the inner product $\alpha_h = \langle x_t, \varepsilon_{t-h} \rangle_H$. This feature generalizes the OLS methodology employed in the univariate case by exploiting orthogonal projections in a more general sense. Indeed, each projection matrix α_h minimizes the distance of the outcome x_t from the submodule generated by the vector innovation ε_{t-h} .

4 Multivariate extended Wold decomposition

In this section, we generalize the extended Wold decomposition for weakly stationary time series of Ortu et al. (2020a) to multidimensional processes, by exploiting our Hilbert-module framework. We apply the Abstract Wold Theorem for Hilbert A -modules to $\mathcal{H}_t(\boldsymbol{\varepsilon})$, and, later on, we deduce the decomposition for a weakly stationary vector process \mathbf{x} with fundamental innovations given by $\boldsymbol{\varepsilon}$. We also provide a persistence-based variance decomposition.

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

Let $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t \in \mathbb{Z}}$ be a unit variance m -dimensional white noise and consider the Hilbert submodule $\mathcal{H}_t(\boldsymbol{\varepsilon}) \subseteq H$ of Eq. (2). In order to apply Theorem 1, we consider the scaling operator $\mathbf{R} : \mathcal{H}_t(\boldsymbol{\varepsilon}) \rightarrow \mathcal{H}_t(\boldsymbol{\varepsilon})$ that makes (normalized) two-by-two averages of subsequent innovations:

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

Here, the function $\lfloor \cdot \rfloor$ associates any $c \in \mathbb{R}$ with the integer $\lfloor c \rfloor = \max\{n \in \mathbb{Z} : n \leq c\}$. In the proof of Theorem 3, we show that \mathbf{R} is well-defined, A -linear and isometric on $\mathcal{H}_t(\boldsymbol{\varepsilon})$.

Following Ortu et al. (2020a), to illustrate the decomposition of $\mathcal{H}_t(\boldsymbol{\varepsilon})$ induced by \mathbf{R} , from the white noise $\boldsymbol{\varepsilon}$, we define the (multivariate) detail process at scale 1 $\boldsymbol{\varepsilon}^{(1)} = \{\varepsilon_t^{(1)}\}_{t \in \mathbb{Z}}$ by

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

Each $\varepsilon_t^{(1)}$ has zero mean and unit variance: $\mathbb{E}[\varepsilon_t^{(1)} \varepsilon_t^{(1)'}] = I$. In general, for any $j \in \mathbb{N}$, we define the (multivariate) detail process at scale j $\boldsymbol{\varepsilon}^{(j)} = \{\varepsilon_t^{(j)}\}_{t \in \mathbb{Z}}$ by

$$\varepsilon_t^{(j)} = \frac{1}{\sqrt{2^j}} \left(\sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-i} - \sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-2^{j-1}-i} \right), \quad t \in \mathbb{Z}. \tag{3}$$

This definition is the natural multivariate counterpart of eq. (6) in Ortu et al. (2020a), which is written in the same way. Equation (3) is reminiscent of the iterated application of the discrete Haar transform to the series of ε_t (Addison 2002, Chapter 3). High scales involve detail processes that have not been faded out by many applications of this transform. They are, therefore, associated with a high degree of persistence. For instance, if t evolves daily, one can interpret the details at scale 1 as 2-day shocks, those at scale 2 as 4-day shocks, those at scale 3 as 8-day shocks and so on. In few words, the scale j involves 2^j -day multivariate shocks and defines the degree of persistence j .

In order to avoid overlap among the vectors $\varepsilon_t^{(j)}$, at any scale j we consider the subseries of $\boldsymbol{\varepsilon}^{(j)}$ defined on the support $S_t^{(j)} = \{t - k2^j : k \in \mathbb{Z}\}$. Indeed, each detail at scale j is a vector $MA(2^j - 1)$ of the white noise $\boldsymbol{\varepsilon}$. Some spurious correlation is present between the details $\varepsilon_{t-k2^j}^{(j)}$ and $\varepsilon_{\tau-k2^j}^{(j)}$ with $|t - \tau| \leq 2^j - 1$, but each subseries $\{\varepsilon_{t-k2^j}^{(j)}\}_{k \in \mathbb{Z}}$ is a unit variance white noise. We formalize this fact in the first point of Theorem 4, and we will associate high scales with more persistent detail processes.

We now state the orthogonal decomposition of $\mathcal{H}_t(\boldsymbol{\varepsilon})$ implied by Theorem 1 when we use the scaling operator as isometry.

Theorem 3 *Let $\boldsymbol{\varepsilon}$ be a unit variance m -dimensional white noise. 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

$$\mathbf{R}^{j-1} \mathcal{L}_t^{\mathbf{R}} = \left\{ \sum_{k=0}^{+\infty} b_k^{(j)} \varepsilon_{t-2k}^{(j)} \in \mathcal{H}_t(\boldsymbol{\varepsilon}) : b_k^{(j)} \in A \right\}. \quad (4)$$

Proof See ‘‘Appendix B.2’’. ■

The main steps of the proof are the following.

We first determine the invariant submodule $\widehat{\mathcal{H}}_t(\boldsymbol{\varepsilon})$ prescribed by Theorem 1. From the definition of \mathbf{R} , the submodule $\mathbf{R}\mathcal{H}_t(\boldsymbol{\varepsilon})$ is composed by the linear combinations of innovations ε_t that have the (matrix) coefficients equal to each others 2-by-2. Similarly, for any $j \in \mathbb{N}$, the submodules $\mathbf{R}^j \mathcal{H}_t(\boldsymbol{\varepsilon})$ consist of the linear combinations of vectors ε_t with (matrix) coefficients equal to each others 2^j -by- 2^j :

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

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

We now turn to the submodule $\widetilde{\mathcal{H}}_t(\boldsymbol{\varepsilon})$. 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 $\mathcal{H}_t(\boldsymbol{\varepsilon})$. As \mathbf{R} is linear and bounded, such submodule coincides with the kernel of the adjoint operator \mathbf{R}^* (Proposition 19 in ‘‘Appendix A.7’’). Therefore,

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

Hence, $\mathcal{L}_t^{\mathbf{R}}$ contains the moving averages generated by the detail process at scale 1 on the support $S_t^{(1)}$. More generally, for each $j \in \mathbb{N}$, the image of $\mathcal{L}_t^{\mathbf{R}}$ through \mathbf{R}^{j-1} is the submodule in Eq. (4), which consists of the moving averages generated by the detail process at scale j on $S_t^{(j)}$.

4.2 The multivariate extended Wold decomposition of \mathbf{x}_t

As in Sect. 3, we consider a zero-mean regular weakly stationary m -dimensional process $\mathbf{x} = \{x_t\}_{t \in \mathbb{Z}}$. We also require \mathbf{x} to be purely non-deterministic in order to focus on the persistence generated by the shocks aggregation.

Theorem 2 (the MCWD) ensures that x_t belongs to $\mathcal{H}_t(\boldsymbol{\varepsilon})$, where $\boldsymbol{\varepsilon}$ is the process of fundamental innovations of \mathbf{x} . As a result, the orthogonal decomposition of $\mathcal{H}_t(\boldsymbol{\varepsilon})$ of Theorem 3 induces a decomposition of x_t . Indeed, there exists a sequence $\{g_t^{(j)}\}_{j \in \mathbb{N}}$ of random processes such that

$$x_t = \sum_{j=1}^{+\infty} g_t^{(j)}, \tag{5}$$

where each $g_t^{(j)}$ is the orthogonal projection of x_t on the submodule $\mathbf{R}^{j-1} \mathcal{L}_t^{\mathbf{R}}$ of $\mathcal{H}_t(\boldsymbol{\varepsilon})$. We refer to $g_t^{(j)}$ as the (multivariate) persistent component at scale j . Clearly, given t , the components $g_t^{(j)}$ are orthogonal to each others. In addition, since each $g_t^{(j)}$ belongs to $\mathbf{R}^{j-1} \mathcal{L}_t^{\mathbf{R}}$,

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

where $\|\sum_{k=0}^{\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)}\|_{\varphi}^2 = \sum_{k=0}^{\infty} \text{Tr}(\beta_k^{(j)} \beta_k^{(j)'})$ is finite. Each $\beta_k^{(j)}$ is obtained by projecting x_t on the submodule generated by the detail $\varepsilon_{t-k2^j}^{(j)}$ and so

$$\beta_k^{(j)} = \left\langle x_t, \varepsilon_{t-k2^j}^{(j)} \right\rangle_H = \mathbb{E} \left[x_t \varepsilon_{t-k2^j}^{(j)'} \right].$$

By replacing the expression of $g_t^{(j)}$ into Eq. (5), we obtain the MEWD of x_t stated in Eq. (6). The explicit expression of matrices $\beta_k^{(j)}$ in Eq. (7) below turns out to be the multivariate version of eq. (7) in Ortu et al. (2020a), which features the same writing.

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

$$x_t = \sum_{j=1}^{+\infty} \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)}, \tag{6}$$

where the equality is in norm and

1. for any fixed $j \in \mathbb{N}$, the m -dimensional process $\boldsymbol{\varepsilon}^{(j)} = \{\varepsilon_t^{(j)}\}_{t \in \mathbb{Z}}$ with

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

is a $MA(2^j - 1)$ with respect to the classical Wold innovations of \mathbf{x} and $\{\varepsilon_{t-k2^j}^{(j)}\}_{k \in \mathbb{Z}}$ is a unit variance white noise;

2. for any $j \in \mathbb{N}, k \in \mathbb{N}_0$, the $m \times m$ matrices $\beta_k^{(j)}$ are uniquely determined via

$$\beta_k^{(j)} = \frac{1}{\sqrt{2^j}} \left(\sum_{i=0}^{2^j-1} \alpha_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+2^{j-1}+i} \right); \tag{7}$$

hence, they are independent of t and $\sum_{k=0}^{\infty} \text{Tr}(\beta_k^{(j)} \beta_k^{(j)'}) < +\infty$ for any $j \in \mathbb{N}$;

3. letting

$$g_t^{(j)} = \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}, \tag{8}$$

for any $j, l \in \mathbb{N}, p, q, t \in \mathbb{Z}, \mathbb{E}[g_{t-p}^{(j)} g_{t-q}^{(l)'}]$ depends at most on $j, l, p - q$.

Moreover, $\mathbb{E}[g_{t-m2^j}^{(j)} g_{t-n2^l}^{(l)'}] = \mathbf{0}$ for all $j \neq l, m, n \in \mathbb{N}_0$ and $t \in \mathbb{Z}$.

Proof See ‘‘Appendix B.3.’’ ■

The matrix $\beta_k^{(j)}$ is the (multivariate) scale-specific response associated with 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 share the same shocks. For this reason, we can retrieve the matrices $\beta_k^{(j)}$ from the matrices α_n of the MCWD through Eq. (7).

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,

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

The (multivariate) residual component at scale j is the orthogonal projection of x_t on the submodule $\mathbf{R}^j \mathcal{H}_t(\boldsymbol{\varepsilon})$, and we denote it by $\pi_t^{(j)}$. As can be seen in the proof of Theorem 4, $\pi_t^{(j)}$ has the expression:

$$\pi_t^{(j)} = \sum_{k=0}^{+\infty} \gamma_k^{(j)} \left(\frac{1}{\sqrt{2^j}} \sum_{i=0}^{2^j-1} \varepsilon_{t-k2^j-i} \right), \quad \gamma_k^{(j)} = \frac{1}{\sqrt{2^j}} \sum_{i=0}^{2^j-1} \alpha_{k2^j+i}. \tag{9}$$

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:

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

According to the third point of Theorem 4, 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. In general, the cross-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, x_t is decomposed into the sum of orthogonal components $g_t^{(j)}$ associated with different persistence levels j . Each vector $g_t^{(j)}$ has innovations on a grid $S_t^{(j)} = \{t - k2^j : k \in \mathbb{Z}\}$ 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 details rises. For instance, if j is high and a scale-specific response $\beta_k^{(j)}$ is remarkably different from the null matrix, x_t is affected by an important low-frequency component. On the contrary, if j is low and some $\beta_k^{(j)}$ differs from zero, a high-frequency component is not negligible and impinges on x_t in the short term.

A justification of the fact that the iterated application of \mathbf{R} increases persistence is due to spectral analysis considerations, and it is developed in detail in Ortu et al. (2020a) for the univariate case: see Section 2.1 therein and the supplementary material (Ortu et al. 2020b). The scaling operator defines, in fact, an approximate low-pass filter. In addition, as suggested in the same paper, bases different from 2 can be used to derive similar persistence-based decompositions.

The MEWD properly generalizes the univariate Extended Wold Decomposition of Ortu et al. (2020a). Indeed, in case the matrix coefficients α_h are diagonal, for any $i = 1, \dots, m$ the entry $x_{i,t}$ depends only on the innovations $\varepsilon_{i,t}$ and the scale-specific responses $\beta_k^{(j)}$ are diagonal matrices, too. Each $x_{i,t}$ satisfies the decompositions

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

with

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

as in the univariate Extended Wold Decomposition.

Finally, the MEWD of x_t turned out to be a refinement of the MCWD, where ε is the process of fundamental innovations of \mathbf{x} . Nonetheless, such persistence-based decomposition holds also in case ε is any unit variance white noise that allows a moving average representation of x_t . Furthermore, in case ε has a positive definite covariance matrix ξ , then $\xi = \zeta \zeta$ for some symmetric positive definite $\zeta \in A$. Then, $\eta_t = \zeta^{-1} \varepsilon_t$ defines a unit variance white noise and the MCWD and the MEWD of x_t become, respectively,

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

with $\tilde{\alpha}_h = \alpha_h \zeta$ and $\tilde{\beta}_k^{(j)} = \beta_k^{(j)} \zeta$. Alternatively, other tools can be used to factorize ξ , as the Cholesky decomposition (see, for instance, the application in ‘‘Appendix C.3’’).

4.3 Persistence-based variance decomposition

One of the strength of the MEWD is that it allows us to define a variance decomposition across the different persistence layers. The notion of orthogonality permits to

decompose the variance of each entry $x_{i,t}$ of the vector x_t into the sum of the variances of the corresponding entries of each persistent component $g_t^{(j)}$. This permits to disentangle the exposure towards shocks with heterogeneous persistence without distinguishing among the different univariate disturbances. As a further step, the null correlation between the univariate innovations in the (multivariate) details $\varepsilon_t^{(j)}$ allows us to decompose the variance of each entry of $g_t^{(j)}$ in order to quantify the contribution of each source of randomness. In so doing, one can individuate the main time scales at which each univariate shock impacts the aggregate process x_t .

To illustrate the variance decompositions, we consider a weakly stationary bivariate process $x_t = [y_t, z_t]'$ with unit variance white noise $\varepsilon_t = [u_t, v_t]'$. We focus on y_t and on the first entry of the persistent components $g_t^{(j)}$. The portion of variance of y_t associated with the latter is

$$\text{var}^{(j)}(y_t) = \sum_{k=0}^{+\infty} \left[\left(\beta_k^{(j)}(1, 1) \right)^2 + \left(\beta_k^{(j)}(1, 2) \right)^2 \right] \quad (10)$$

and

$$\text{var}(y_t) = \sum_{j=1}^{+\infty} \text{var}^{(j)}(y_t).$$

Operationally, in order to assess the overall importance of each persistence level in explaining the total variance of y_t , we can compute, for each scale j , the ratio $\text{var}^{(j)}(y_t)/\text{var}(y_t)$. In addition, to capture the effect of the persistence of each single shock, we can compute, at any scale j , the ratios

$$\sum_{k=0}^{+\infty} \left(\beta_k^{(j)}(1, 1) \right)^2 / \text{var}(y_t), \quad \sum_{k=0}^{+\infty} \left(\beta_k^{(j)}(1, 2) \right)^2 / \text{var}(y_t). \quad (11)$$

Although a persistence-based variance decomposition is already present in Ortu et al. (2020a), the orthogonality notion employed in the multivariate case makes it possible to quantify both the persistence of each time series in the vector process and the persistence of single shocks impacting each of them.

5 Conclusions

In this paper, we recast the standard treatment of multivariate time series in a Hilbert A -module framework and prove the Abstract Wold Theorem for Hilbert A -modules. This result allows us to revisit the MCWD and to derive multivariate version of the Extended Wold Decomposition of Ortu et al. (2020a), by using two different isometric operators. Interestingly, the MEWD provides a decomposition of the given vector process into uncorrelated persistent components driven by shocks with longer and longer duration. The orthogonality ensured by the theorem induces a variance decomposition that

permits to establish the relative importance of each persistence component. Moreover, scale-specific responses allow us to isolate, on different time scales, dynamics that are not recognizable from the impulse responses of the aggregated process (some illustrations are in “Appendix C”).

The statistical inference about scale-specific responses is an important issue that goes beyond the scope of the paper. As to the estimators of the parameters in moving average models, one can refer to the literature summarized in the introduction of Ghysels et al. (2003), especially to the seminal work of Durbin (1959). However, a simple way to statistically test whether scale-specific responses are zero is using the bootstrap procedure (Efron and Tibshirani 1986) in the pipeline described in “Appendix C.2”. The bootstrap empirical distribution of scale-specific responses can be easily employed to obtain the test p value.

Acknowledgements We thank Giorgio Primiceri for valuable insights and the participants at the 14th International Conference on Computational and Financial Econometrics (CFE), University of London (virtual, 2020), at the 2021 Annual Congress of the Swiss Society of Economics and Statistics (SSES), Universität Zürich (virtual, 2021) and at the 6th Canadian Conference in Applied Statistics, Concordia University (virtual, 2021).

Funding S. Cerreia-Vioglio acknowledges the financial support of ERC Grant SDDM-TEA.

Data Availability Data are publicly available.

Code Availability Codes are available upon request.

Declarations

Conflict of interest The authors declare that they do not have any financial or non-financial conflicts of interest.

Appendices

“Appendix A” focuses on the Hilbert A -module theory. “Appendix B” illustrates the derivation of the multivariate extended Wold decomposition (MEWD). “Appendix C” provides some illustrations of the MEWD, including an application to Blanchard and Quah (1989).

Appendix A Hilbert A -modules

In this appendix, we present a primer on Hilbert A -modules. The purpose is twofold: (a) to present a uniform and self-contained treatment of the topic, (b) to present tools and 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 square matrices with *real values*, but we will keep our setting abstract in order to avoid getting lost in useless details.

Hilbert A -modules are to algebras as Hilbert spaces are to the real/complex field. 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 (e.g., Definition 1) are then kept identical to the ones of the scalar case. Since the seminal paper of Kaplansky (1953), Hilbert A -modules have been widely studied in mathematics. In applications, their use seems to be scarcer even though there are notable exceptions. In economics, a particular Hilbert A -module was studied and used by Hansen and Richard (1987) to prove a conditional version of the fundamental theorem of asset pricing, while in statistics, Wiener and Masani (1957) studied the complex version of the Hilbert A -module we use in our application, to provide a proof for the multivariate classical Wold decomposition (MCWD). This is consistent with the mathematical literature. This literature, starting with Kaplansky (1953), focused on complex algebras and developed very rapidly and in a non-systematic/scattered way (see, e.g., Frank 1996 for an account). On the other hand, the real case received little attention. One notable exception to this is the paper of Goldstine and Horwitz (1966) which deals with the case we have at hand here: the algebra of square real matrices. As the appendix progresses, we will highlight what are the overlaps between our work and theirs. Most notably, they prove the Riesz representation theorem and the projection theorem, in a reverse order. Both results are instrumental in proving the main result of this appendix which is the Abstract Wold Theorem for Hilbert A -modules (Theorem 1).

The reader might be tempted to think that Hilbert A -modules behave exactly like Hilbert spaces. One key feature which makes them appealing for applications is that most of the statements valid for Hilbert spaces, seemingly carry over to this more general structure: the caveat is that the proofs not always generalize in a similar fashion.

A.1 C^* -algebras: the new scalars

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. In particular, A is a *real normed* algebra with multiplicative unit e , we denote by $\|\cdot\|_A$ 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 $*$: $A \rightarrow A$ 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,

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

The algebra A is also naturally ordered by the order \geq induced by the closed convex cone of positive elements that are such that $a = a^*$ (in the real case, the extra requirement $a = a^*$ is not redundant). We denote by $A_+ = \{a \in A : a \geq 0\}$. The following 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 $\bar{\varphi} : A \rightarrow \mathbb{R}$ such that

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

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

$$a \geq b \iff \varphi(a) \geq \varphi(b) \quad \forall \varphi \in C. \tag{A1}$$

A.2 Pre-Hilbert A -modules

Consider A as above. We next proceed by defining the objects we study in this paper.

Definition 1 An abelian group $(H, +)$ is an A -module if and only if an outer product $\cdot : A \times H \rightarrow H$ 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 $\langle \cdot, \cdot \rangle_H : H \times H \rightarrow A$ 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.^{2,3}

² 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} .

³ It is routine to show that the following statements, which we will use later on, 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 $\alpha \in \mathbb{R}$ and for all $x, y \in H$.

Given a pre-Hilbert A -module, by adapting the techniques of Raeburn and Williams (1998, Lemma 2.5) to the real case, we will show that

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

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

Given an element $y \in H$, note that $\langle \cdot, y \rangle_H$ induces an operator $f : H \rightarrow A$ 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 $x \in H$.

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

Definition 2 Let H be a pre-Hilbert A -module. We say that H is *self-dual* if and only if for each $f : H \rightarrow A$ which is A -linear and bounded there exists $y \in H$ such that

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

It is rather easy to see that if H is self-dual, then each A -linear and bounded $f : H \rightarrow A$ is represented by a *unique* vector y .

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 A -valued inner product $\langle \cdot, \cdot \rangle_H$ 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:

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

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

Proposition 5 *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 6 *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

$$f(\alpha x + \beta y) = f((\alpha e) \cdot x + (\beta e) \cdot y) = (\alpha e) f(x) + (\beta e) f(y) = \alpha f(x) + \beta f(y).$$

■

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

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

Proposition 7 *Let H be a pre-Hilbert A -module. If A admits a strictly positive functional $\bar{\varphi}$, then $\langle \cdot, \cdot \rangle_{\bar{\varphi}}$ 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 $\langle x, x \rangle_{\bar{\varphi}} = \bar{\varphi}(\langle x, x \rangle_H) \geq 0$. Since $\bar{\varphi}$ is strictly positive and $\langle x, x \rangle_H \geq 0$, note also that

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

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

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

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

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

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

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

Properties a–d yield the statement.

Corollary 8 *Let H be a pre-Hilbert A -module. If A admits a strictly positive functional $\bar{\varphi}$, then $(H, +, \cdot^e, \langle \cdot, \cdot \rangle_{\bar{\varphi}})$ is a pre-Hilbert space.*

Proposition 9 *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 $\langle w, z \rangle_H = \langle w, z \rangle_H^*$. Then, 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

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

yielding that

$$\varphi(\langle w, z \rangle_H)^2 \leq \varphi(\langle w, w \rangle_H) \varphi(\langle z, z \rangle_H). \quad (\text{A2})$$

Choose $\bar{x}, \bar{y} \in H$. Define $w = \langle \bar{x}, \bar{y} \rangle_H^* \cdot \bar{x}$ and $z = \bar{y}$. 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 (A2) holds. In particular, we have that

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

Define $a = \langle \bar{x}, \bar{x} \rangle_H$ and $b = \langle \bar{x}, \bar{y} \rangle_H^*$. Recall that $bab^* \leq \|a\|_A bb^*$ and $bb^* \geq 0$. Thus, we have that

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

and $\varphi(\langle \bar{x}, \bar{y} \rangle_H^* \langle \bar{x}, \bar{y} \rangle_H) \geq 0$. We thus have that

$$\varphi(\langle \bar{x}, \bar{y} \rangle_H^* \langle \bar{x}, \bar{y} \rangle_H) \leq \varphi(\| \langle \bar{x}, \bar{x} \rangle_H \|_A \langle \bar{y}, \bar{y} \rangle_H). \quad (\text{A3})$$

Since φ was arbitrarily chosen, we have that (A3) holds for all $\varphi \in C$, that is, by (A1)

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

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

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

3. It trivially follows from point 2. ■

A.3.1 Topological structure

The $\| \cdot \|_H$ norm

Define $\| \cdot \|_H : H \rightarrow [0, +\infty)$ by

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

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

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

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

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

$$\begin{aligned} \| x + y \|_H^2 &= \| \langle x + y, x + y \rangle_H \|_A = \| \langle x, x \rangle_H + \langle x, y \rangle_H + \langle y, x \rangle_H + \langle y, y \rangle_H \|_A \\ &\leq \| \langle x, x \rangle_H \|_A + \| \langle x, y \rangle_H \|_A + \| \langle y, x \rangle_H \|_A + \| \langle y, y \rangle_H \|_A \\ &\leq \| \langle x, x \rangle_H \|_A + \sqrt{ \| \langle x, x \rangle_H \|_A \| \langle y, y \rangle_H \|_A } \\ &\quad + \sqrt{ \| \langle y, y \rangle_H \|_A \| \langle x, x \rangle_H \|_A } + \| \langle y, y \rangle_H \|_A \\ &= \| \langle x, x \rangle_H \|_A + 2\sqrt{ \| \langle x, x \rangle_H \|_A \| \langle y, y \rangle_H \|_A } + \| \langle y, y \rangle_H \|_A \end{aligned}$$

$$= \left(\|\langle x, x \rangle_H\|_A^{\frac{1}{2}} + \|\langle y, y \rangle_H\|_A^{\frac{1}{2}} \right)^2.$$

We can thus conclude that

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

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

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

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

■

By Proposition 9, it readily follows that

$$\|\langle x, y \rangle_H\|_A \leq \|x\|_H \|y\|_H \quad \forall x, y \in H. \quad (\text{A4})$$

Corollary 11 *Let H be a pre-Hilbert A -module. For each $y \in H$, the functional $\langle \cdot, y \rangle_H : H \rightarrow A$ is A -linear, $\|\cdot\|_H - \|\cdot\|_A$ 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 (A4). Since the norm of the linear operator is given by

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

the statement easily follows from (A4) and the definition of $\|\cdot\|_H$. ■

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

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

$$\|x\|_{\bar{\varphi}} = \sqrt{\langle x, x \rangle_{\bar{\varphi}}} \quad \forall x \in H. \quad (\text{A5})$$

By Corollary 8, $\langle \cdot, \cdot \rangle_{\bar{\varphi}}$ is an inner product on H and it is immediate to see that $\|\cdot\|_{\bar{\varphi}}$ is a norm and

$$\|x\|_{\bar{\varphi}} = \sqrt{\bar{\varphi}(\langle x, x \rangle_H)} \quad \forall x \in H. \quad (\text{A6})$$

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

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

This implies that

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

that is,

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

We can conclude that

$$x_n \xrightarrow{\|\cdot\|_H} 0 \implies x_n \xrightarrow{\|\cdot\|_{\bar{\varphi}}} 0.$$

Proposition 12 *Let H be a pre-Hilbert A -module. If A is finite dimensional, then A admits a trace $\bar{\varphi}$ and the norms $\|\cdot\|_{\bar{\varphi}}$ and $\|\cdot\|_H$ 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

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

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

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

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 $\|\cdot\|_H$ and A is endowed with $\|\cdot\|_A$.

Proposition 13 *If H is a pre-Hilbert A -module, then H^\sim is an A -module.*

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

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

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

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

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, $e \cdot f = f$. \blacksquare

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

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

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

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

Given Corollary 11 and the properties of $\langle \cdot, \cdot \rangle_H$, the map S^\sim is well-defined and linear. In fact, for each $\alpha, \beta \in \mathbb{R}$ and for each $x, y, z \in H$

$$S^\sim(\alpha y + \beta z)(x) = \langle x, \alpha y + \beta z \rangle_H = \langle x, \alpha y \rangle_H + \langle x, \beta z \rangle_H = \alpha \langle x, y \rangle_H + \beta \langle x, z \rangle_H,$$

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

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

Proof 1. By Proposition 13, H^\sim is an A -module. In particular, H^\sim is a vector subspace of $B(H, A)$. Consider a $\| \cdot \|_{H^\sim}$ Cauchy sequence $\{f_n\}_{n \in \mathbb{N}} \subseteq H^\sim \subseteq B(H, A)$. By Aliprantis and Border (2006, Theorem 6.6) and since A is $\| \cdot \|_A$ complete, we have that there exists $f \in B(H, A)$ such that $f_n \xrightarrow{\| \cdot \|_{H^\sim}} f$. We are left to show that f is A -linear. First, observe that $f : H \rightarrow A$ is such that

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

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

$$\begin{aligned} f_n(x) \xrightarrow{\| \cdot \|_A} f(x), f_n(y) \xrightarrow{\| \cdot \|_A} f(y) &\implies af_n(x) \xrightarrow{\| \cdot \|_A} af(x), bf_n(y) \xrightarrow{\| \cdot \|_A} bf(y) \\ &\implies af_n(x) + bf_n(y) \xrightarrow{\| \cdot \|_A} af(x) + bf(y). \end{aligned}$$

At the same time, $af_n(x) + bf_n(y) = f_n(a \cdot x + b \cdot y) \xrightarrow{\| \cdot \|_A} f(a \cdot x + b \cdot y)$ 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. Recall that $S^\sim : H \rightarrow H^\sim$ is defined by

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

By Corollary 11, 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 $\| \cdot \|_H$ 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 $\| \cdot \|_{H^\sim}$ Cauchy sequence in H^\sim . Since H^\sim is $\| \cdot \|_{H^\sim}$ complete and S^\sim is onto, it follows that there exists $f \in H^\sim$ such that $S^\sim(x_n) \xrightarrow{\| \cdot \|_{H^\sim}} f = S^\sim(x)$ for some $x \in H^\sim$. Since S^\sim is an isometry, we have that $x_n \xrightarrow{\| \cdot \|_H} x$, proving that H is $\| \cdot \|_H$ complete. ■

A.5 Self-duality

Theorem 15 *Let A be finite dimensional and H a pre-Hilbert A -module. The following statements are equivalent:*

- (i) H is $\| \cdot \|_H$ complete, that is, H is a Hilbert A -module;
- (ii) H is $\| \cdot \|_{\bar{\varphi}}$ complete;
- (iii) H is self-dual.

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

(i) implies (ii). By Proposition 12 and since A is finite dimensional, $\| \cdot \|_{\bar{\varphi}}$ and $\| \cdot \|_H$ are equivalent. It follows that H is $\| \cdot \|_{\bar{\varphi}}$ complete.

(ii) implies (iii). By Corollary 8 and since H is $\| \cdot \|_{\bar{\varphi}}$ complete, it follows that H is a Hilbert space with inner product $\langle \cdot, \cdot \rangle_{\bar{\varphi}}$. Consider $f : H \rightarrow A$ which is A -linear and bounded. In particular, by the proof of Proposition 12, we have that there exists $M > 0$ such that

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

We can conclude that $f : H \rightarrow A$ is linear and $\| \cdot \|_{\bar{\varphi}} - \| \cdot \|_A$ continuous. Consider the linear functional $l = \bar{\varphi} \circ f$. Since $\bar{\varphi}$ is $\| \cdot \|_A$ continuous and f is $\| \cdot \|_{\bar{\varphi}} - \| \cdot \|_A$ continuous, we have that l is $\| \cdot \|_{\bar{\varphi}}$ continuous. By the standard Riesz representation theorem (Brezis 2011, Theorem 5.5), 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, for all $x \in H$,

$$\bar{\varphi}(f(x) - \langle x, y \rangle_H) = \bar{\varphi}(f(x)) - \bar{\varphi}(\langle x, y \rangle_H) = l(x) - \langle x, y \rangle_{\bar{\varphi}} = 0. \tag{A7}$$

Fix $\bar{x} \in H$. Define $a = (f(\bar{x}) - \langle \bar{x}, y \rangle_H)^* \in A$. By Eq. (A7), 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, $\|a^*\|_A^2 = \|aa^*\|_A = 0$. 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 14, it follows that H is $\|\cdot\|_H$ complete. ■

The implication (ii) implies (iii) can be found in Goldstine and Horwitz (1966) 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 17.

A.5.1 Orthogonal decompositions

Pre-Hilbert A -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

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

If M is nonempty, then it is immediate to prove that M^\perp is a submodule.⁴ It is also immediate to show that $M \cap M^\perp = \{0\}$ and that $M^{\perp\perp} \supseteq M$ where

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

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

Lemma 16 *Let H be a pre-Hilbert A -module. If $M \subseteq H$, then M^\perp is $\|\cdot\|_H$ closed.*

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 $x_n \xrightarrow{\|\cdot\|_H} x$. Since $S^\sim(z)$ is $\|\cdot\|_H - \|\cdot\|_A$ 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 17 *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:*

⁴ A nonempty subset N of H is a submodule if and only if, for each $a, b \in A$ and $x, y \in N, a \cdot x + b \cdot y \in N$.

- (i) M is $\| \cdot \|_H$ 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 $\langle \cdot, \cdot \rangle_M$ as the restriction of $\langle \cdot, \cdot \rangle_H$ to $M \times M$, then $(M, +, \cdot, \langle \cdot, \cdot \rangle_M)$ is a pre-Hilbert A -module. It is immediate to see that $\| \cdot \|_M = \| \cdot \|_H$ once the latter is restricted to M . By Theorem 15 and since M is $\| \cdot \|_H$ closed, it follows that M is $\| \cdot \|_M$ 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 $\langle x, y_1 \rangle_H = \langle x, y_1 \rangle_M = \langle x, y \rangle_H$ for all $x \in M$. Define $y_2 = y - y_1$. It follows that $\langle x, y - y_1 \rangle_H = 0$ for all $x \in M$, 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 16 and since $M = M^{\perp\perp} = (M^\perp)^\perp$, it follows that M is $\| \cdot \|_H$ 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 17 allows us to define the (orthogonal) projection of an element $x \in H$ on a $\| \cdot \|_H$ closed submodule M .

Definition 3 Let A be finite dimensional, H a Hilbert A -module, and $M \subseteq H$ a $\| \cdot \|_H$ closed submodule. We call *projection on M* the linear map $\mathcal{P}_M : H \rightarrow M$ 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 $y \in M$, we have that $y = \mathcal{P}_M x$ if and only if $\langle x - y, z \rangle_H = 0$ for all $z \in M$. Moreover, since $\mathcal{P}_M x \in M$ and $x - \mathcal{P}_M x \in M^\perp$, $\mathcal{P}_M x$ minimizes the distance between x and the submodule M since, for all $z \in M$,

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

A.6 The abstract Wold theorem for Hilbert A -modules

In this section, we prove a (generalized) version for Hilbert A -modules of the Abstract Wold Theorem (see, e.g., Sz-Nagy et al. 2010, Theorem 1.1). It is important to observe that the properties of self-duality and complementability (see Theorem 15 and Proposition 17) are fundamental in allowing us to follow the proof strategy used for Hilbert spaces.

Definition 4 We say that $\mathbf{T} : H \rightarrow H$ is an isometry if and only if \mathbf{T} is A -linear and such that

$$\langle \mathbf{T}x, \mathbf{T}y \rangle_H = \langle x, y \rangle_H \quad \forall x, y \in H. \quad (\text{A8})$$

Note that an isometry in this sense satisfies the usual property

$$\|\mathbf{T}x\|_H = \|x\|_H \quad \forall x \in H. \quad (\text{A9})$$

It is immediate to prove by induction that for each $n \in \mathbb{N}_0$ the iterate \mathbf{T}^n satisfies Eqs. (A8) and (A9). In particular, by Abramovich and Aliprantis (2002, Theorem 2.5), if H is $\|\cdot\|_H$ complete, $\mathbf{T}^n H$ is a $\|\cdot\|_H$ closed submodule of H .

Definition 5 Let $\mathbf{T} : H \rightarrow H$ be an isometry. We say that a submodule L is wandering if and only if $\mathbf{T}^n L \perp \mathbf{T}^m L$ for all $m, n \in \mathbb{N}_0$ such that $m \neq n$.

Lemma 18 Let A be finite dimensional and H a pre-Hilbert A -module. If H is self-dual and $\mathbf{T} : H \rightarrow H$ an isometry, then the following statements are true:

1. If M is $\|\cdot\|_H$ closed, so is $\mathbf{T}M$.
2. If $L = (\mathbf{T}H)^\perp$, then L is wandering.
3. If $L = (\mathbf{T}H)^\perp$, then for each $n \in \mathbb{N}_0$

$$\mathbf{T}^n H = \mathbf{T}^n L \oplus \mathbf{T}^{n+1} H \quad \text{and} \quad \mathbf{T}^n L \perp \mathbf{T}^{n+1} H.$$

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

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

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

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

we have that \mathbf{T} is linear. By the proof of Abramovich and Aliprantis (2002, Theorem 2.5) and since \mathbf{T} satisfies Eq. (A9), we have that $\mathbf{T}M$ is closed.

2. Observe that $\mathbf{T}^n H \subseteq H$ for all $n \in \mathbb{N}_0$. It follows that $\mathbf{T}^n H \subseteq \mathbf{T}H$ for all $n \in \mathbb{N}$. Since $L \subseteq H$, it also follows that $\mathbf{T}^n L \subseteq \mathbf{T}^n H \subseteq \mathbf{T}H$ for all $n \in \mathbb{N}$. Since $\mathbf{T}H \perp L$, this implies that $\mathbf{T}^n L \perp L$ for all $n \in \mathbb{N}$. Next, consider $m, n \in \mathbb{N}_0$ such that $m \neq n$. Without loss of generality, assume that $n > m$. By the previous part of the proof, we have that $\mathbf{T}^{n-m} L \perp L$. By Eq. (A8), we can conclude that $\mathbf{T}^n L \perp \mathbf{T}^m L$.

3. We proceed by induction.

Initial Step. $n = 0$. By definition of L , point 1, and Proposition 17 and since H is self-dual, we have that L is a $\|\cdot\|_H$ closed submodule and

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

proving the step.

Inductive Step. Assume the statement is true for n . By assumption, it follows that $\mathbf{T}^n H = \mathbf{T}^n L \oplus \mathbf{T}^{n+1} H$ and $\mathbf{T}^n L \perp \mathbf{T}^{n+1} H$. By Eq. (A8), we have that

$$\mathbf{T}^{n+1} L \perp \mathbf{T}^{n+2} H \tag{A10}$$

as well as

$$\begin{aligned} \mathbf{T}^{n+1} H &= \mathbf{T}\mathbf{T}^n H = \mathbf{T} \left(\mathbf{T}^n L \oplus \mathbf{T}^{n+1} H \right) \\ &= \mathbf{T}\mathbf{T}^n L + \mathbf{T}\mathbf{T}^{n+1} H = \mathbf{T}^{n+1} L \oplus \mathbf{T}^{n+2} H, \end{aligned}$$

where the last equality follows from (A10). The statement follows by induction.

4. We proceed by induction.

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

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

Inductive Step. Assume the statement is true for k . By assumption, it follows that $\bigoplus_{n=0}^k \mathbf{T}^n L = \left(\mathbf{T}^{k+1} H \right)^\perp$. By Proposition 17 and since H is self-dual and since $\mathbf{T}^{k+1} H$ is a $\| \cdot \|_H$ closed submodule, this implies that

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

At the same time, by point 3, we also have that $\mathbf{T}^{k+1} H = \mathbf{T}^{k+1} L \oplus \mathbf{T}^{k+2} H$ and $\mathbf{T}^{k+1} L \perp \mathbf{T}^{k+2} H$. We can conclude that⁵

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

The statement follows by induction. ▀

Theorem 1 Let A be finite dimensional and H a pre-Hilbert A -module. If H is self-dual and $\mathbf{T} : H \rightarrow H$ is an isometry, then $H = \widehat{H} \oplus \widetilde{H}$ where

$$\widehat{H} = \bigcap_{n=0}^{+\infty} \mathbf{T}^n H, \quad \widetilde{H} = \bigoplus_{n=0}^{+\infty} \mathbf{T}^n L, \quad L = (\mathbf{T}H)^\perp.$$

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

⁵ We are relying on the following fact whose proof is routine. Given a H self-dual pre-Hilbert A -module, if M, N, P, Q are four $\| \cdot \|_H$ closed submodules such that $H = M \oplus N$, $N = P \oplus Q$, $N = M^\perp$; and $P \perp Q$, then $Q^\perp = M \oplus P$.

Proof Define $L = (\mathbf{T}H)^\perp$. Define also $M_k = \bigoplus_{n=0}^k \mathbf{T}^n L$ for all $k \in \mathbb{N}_0$, $\tilde{H} = \bigoplus_{n=0}^\infty \mathbf{T}^n L$,⁶ and $\widehat{H} = \tilde{H}^\perp$. It is immediate to see that \widehat{H} and \tilde{H} are two $\|\cdot\|_H$ closed submodules. Note that $M_k \subseteq M_{k+1}$ for all $k \in \mathbb{N}_0$ and $\tilde{H} = \text{cl}_{\|\cdot\|_H}(\bigcup_{k \in \mathbb{N}_0} M_k)$. By construction, we have that $\widehat{H} \perp M_k$ for all $k \in \mathbb{N}_0$. By Lemma 18, it follows that $M_k = (\mathbf{T}^{k+1}H)^\perp$ for all $k \in \mathbb{N}$. By Proposition 17, this implies that if $x \in \widehat{H}$, then $x \in M_k^\perp = \mathbf{T}^{k+1}H$ for all $k \in \mathbb{N}_0$. We can conclude that $x \in \bigcap_{n=1}^\infty \mathbf{T}^n H \cap H = \bigcap_{n=0}^\infty \mathbf{T}^n H$. Vice versa, since $M_k = (\mathbf{T}^{k+1}H)^\perp$ for all $k \in \mathbb{N}$, if $x \in \bigcap_{n=0}^\infty \mathbf{T}^n H$, then

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

Since $\text{cl}_{\|\cdot\|_H}(\bigcup_{n=0}^\infty M_n) = \tilde{H}$, this implies that $x \in \tilde{H}^\perp = \widehat{H}$. In other words, we proved that $\widehat{H} = \bigcap_{n=0}^\infty \mathbf{T}^n H$.

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

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

Assume that $(\widehat{H}', \tilde{H}')$ is another decomposition. Consider the wandering set L' generating \tilde{H}' . By construction and since L' is wandering, we have that $L' \perp \mathbf{T}\tilde{H}'$ and $L' \oplus \mathbf{T}\tilde{H}' = \tilde{H}'$. By construction and Eq. (A8), this implies that

$$L = (\mathbf{T}H)^\perp = (\mathbf{T}(\widehat{H}' \oplus \tilde{H}'))^\perp = (\widehat{H}' \oplus \mathbf{T}\tilde{H}')^\perp = L',$$

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, $\mathbf{T} \in B^\sim(H)$ if and only if

$$\mathbf{T}(a \cdot x + b \cdot y) = a \cdot \mathbf{T}x + b \cdot \mathbf{T}y \quad \forall a, b \in A, \forall x, y \in H$$

and there exists $M > 0$ such that

$$\|\mathbf{T}x\|_H \leq M \|x\|_H \quad \forall x \in H.$$

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

$$\langle \mathbf{T}x, y \rangle_H = \langle x, \mathbf{T}^*y \rangle_H \quad \forall x, y \in H. \tag{A11}$$

⁶ With the notation $\bigoplus_{n=0}^\infty \mathbf{T}^n L$, we mean the $\|\cdot\|_H$ closure of the set $\bigcup_{k \in \mathbb{N}_0} M_k$.

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 $\mathbf{S}, \mathbf{T} \in B^\sim(H)$, then the composition of \mathbf{S} with \mathbf{T} is also in $B^\sim(H)$,⁷

Proposition 19 *Let H be a self-dual pre-Hilbert A -module. The following statements are true:*

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

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

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

$$x \mapsto \langle \mathbf{T}x, y \rangle_H \quad \forall x \in H.$$

Since H is self-dual, there exists a unique $z_y \in H$ such that $\langle \mathbf{T}x, y \rangle_H = \langle x, z_y \rangle_H$ for all $x \in H$. We define $\mathbf{T}^* : H \rightarrow H$ to be such that $\mathbf{T}^*y = z_y$. It follows that \mathbf{T}^* is well-defined and satisfies Eq. (A11). Next, observe that for each $y_1, y_2 \in H$ and $a_1, a_2 \in A$

$$\begin{aligned} \langle x, \mathbf{T}^*(a_1 \cdot y_1 + a_2 \cdot y_2) \rangle_H &= \langle x, z_{a_1 \cdot y_1 + a_2 \cdot y_2} \rangle_H = \langle \mathbf{T}x, a_1 \cdot y_1 + a_2 \cdot y_2 \rangle_H \\ &= \langle \mathbf{T}x, y_1 \rangle_H a_1^* + \langle \mathbf{T}x, y_2 \rangle_H a_2^* = \langle x, z_{y_1} \rangle_H a_1^* + \langle x, z_{y_2} \rangle_H a_2^* \\ &= \langle x, a_1 \cdot z_{y_1} + a_2 \cdot z_{y_2} \rangle_H = \langle x, a_1 \cdot \mathbf{T}^*y_1 + a_2 \cdot \mathbf{T}^*y_2 \rangle_H \quad \forall x \in H, \end{aligned}$$

yielding that \mathbf{T}^* is A -linear and, in particular, linear. Finally, note that

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

⁷ Note that $\mathbf{ST} \in B^\sim(H) \subseteq B(H)$, thus $\mathbf{ST} \in B(H)$. We only need to prove A -linearity. Indeed, for each $a, b \in A$ and $x, y \in H$

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

$$\leq \sup_{\|y\|_H=1} \left(\sup_{\|x\|_H=1} \|\mathbf{T}\| \|x\|_H \|y\|_H \right) \leq \|\mathbf{T}\|,$$

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

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

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

$$\langle \mathbf{T}x, y \rangle_H = \langle x, \mathbf{T}^*y \rangle_H = \langle x, \mathbf{S}^*y \rangle_H = \langle \mathbf{S}x, y \rangle_H \quad \forall y \in H.$$

We can conclude that $\mathbf{T}x = \mathbf{S}x$. Since x was arbitrarily chosen, we can conclude that $\mathbf{T} = \mathbf{S}$, that is, $*$ is injective.

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

$$\langle \mathbf{T}x, y \rangle_H = \langle x, \mathbf{T}^*y \rangle_H = \langle \mathbf{T}^*y, x \rangle_H^* = \langle y, \mathbf{T}^{**}x \rangle_H^* = \langle \mathbf{T}^{**}x, y \rangle_H \quad \forall y \in H.$$

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

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

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

$$\|\mathbf{T}^*\| \leq \|\mathbf{T}\| \quad \forall \mathbf{T} \in B^\sim(H).$$

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

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

$$\begin{aligned} \|\mathbf{S}\mathbf{T}\| &= \sup_{\|y\|_H=1} \|\mathbf{S}(\mathbf{T}y)\|_H \leq \sup_{\|y\|_H=1} (\|\mathbf{S}\| \|\mathbf{T}y\|_H) \\ &\leq \|\mathbf{S}\| \sup_{\|y\|_H=1} (\|\mathbf{T}y\|_H) \leq \|\mathbf{S}\| \|\mathbf{T}\|. \end{aligned}$$

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

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

$$= \sup_{\|x\|_H=1} \|(\mathbf{S}^*\mathbf{S})x\|_H = \|\mathbf{S}^*\mathbf{S}\| \leq \|\mathbf{S}^*\| \|\mathbf{S}\| = \|\mathbf{S}\|^2,$$

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

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

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

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

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

$$\langle y, \bar{y} \rangle_H = \langle \mathbf{T}x, \bar{y} \rangle_H = \langle x, \mathbf{T}^*\bar{y} \rangle_H = 0,$$

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

Point 1 can also be found in Goldstine and Horwitz (1966). Also in this case, there is a technical difference in terms of norm used over A .

Appendix B Proofs about the MEWD

B.1 Properties of $L^2(\mathbb{R}^m, \Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{H}_t(\boldsymbol{\varepsilon})$

Proposition 20 $H = L^2(\mathbb{R}^m, \Omega, \mathcal{F}, \mathbb{P})$ is a Hilbert A -module. Moreover, if $\boldsymbol{\varepsilon}$ is a m -dimensional white noise, the submodule $\mathcal{H}_t(\boldsymbol{\varepsilon})$ of H defined in Eq. (2) is closed.

Proof We already described in Sect. 2.2 that H is a pre-Hilbert A -module. Here, we show that H is also $\|\cdot\|_{\bar{\varphi}}$ complete. We consider a Cauchy sequence $\{x^{(n)}\}_n \subset H$, i.e., for any $\varepsilon > 0$ there exists $N > 0$ such that

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

For any entry $i = 1, \dots, m$, the sequence $\{x_i^{(n)}\}_n \subset L^2(\Omega, \mathcal{F}, \mathbb{P})$ satisfies the Cauchy condition. Since $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is complete, there exists $x_i \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{E}[(x_i^{(n)} - x_i)^2] < \varepsilon^2/m$ for all $n > N_i$. As a result, by defining $x = [x_1, \dots, x_m]' \in H$, we have

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

and so H is $\|\cdot\|_{\bar{\varphi}}$ complete, i.e., it is a Hilbert A -module by Theorem 15.

As to the submodule $\mathcal{H}_t(\boldsymbol{\varepsilon})$, take $x \in H$ such that there exists a sequence $\{x^{(n)}\}_n \subset \mathcal{H}_t(\boldsymbol{\varepsilon})$ with $\|x^{(n)} - x\|_{\bar{\varphi}} \rightarrow 0$. We show that $x \in \mathcal{H}_t(\boldsymbol{\varepsilon})$, too. Any $x^{(n)}$ can be written as $x^{(n)} = \sum_{k=0}^{\infty} \langle x^{(n)}, \varepsilon_{t-k} \rangle_H \varepsilon_{t-k}$ because, if $x^{(n)} = \sum_{k=0}^{\infty} a_k^{(n)} \varepsilon_{t-k}$ with $a_k^{(n)} \in A$,

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

In addition, the limit x can be decomposed as $x = \sum_{k=0}^{\infty} \langle x, \varepsilon_{t-k} \rangle_H \varepsilon_{t-k} + v$ with $v \in H$ such that $\langle v, \varepsilon_{t-k} \rangle_H = \mathbf{0}$ for all $k \in \mathbb{N}_0$. This implies that $\langle v, \varepsilon_{t-k} \rangle_{\bar{\varphi}} = \text{Tr}(\langle v, \varepsilon_{t-k} \rangle_H) = 0$. In consequence,

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

As $\|x^{(n)} - x\|_{\bar{\varphi}}$ is arbitrary small, $\|v\|_{\bar{\varphi}} = 0$ and so $v = 0$. Thus, $x = \sum_{k=0}^{\infty} \langle x, \varepsilon_{t-k} \rangle_H \varepsilon_{t-k}$ belongs to $\mathcal{H}_t(\boldsymbol{\varepsilon})$. ■

B.2 Proof of Theorem 3

By Proposition 20, $\mathcal{H}_t(\boldsymbol{\varepsilon})$ is a closed submodule of $H = L^2(\mathbb{R}^m, \Omega, \mathcal{F}, \mathbb{P})$, which is a Hilbert A -module. Hence, $\mathcal{H}_t(\boldsymbol{\varepsilon})$ is a Hilbert A -module too and so, by Theorem 15, it is self-dual. Before applying the Abstract Wold Theorem for Hilbert A -modules (Theorem 1), we prove that the scaling operator \mathbf{R} is well-defined, A -linear and isometric on $\mathcal{H}_t(\boldsymbol{\varepsilon})$.

To show that \mathbf{R} is well-defined on $\mathcal{H}_t(\boldsymbol{\varepsilon})$, consider any $X = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k} \in \mathcal{H}_t(\boldsymbol{\varepsilon})$, i.e., $\|X\|_{\bar{\varphi}}^2 = \sum_{k=0}^{\infty} \text{Tr}(a_k a_k') < +\infty$. Then,

$$\begin{aligned} \|\mathbf{R}X\|_{\bar{\varphi}}^2 &= \sum_{k=0}^{+\infty} \text{Tr} \left(\frac{a_{\lfloor \frac{k}{2} \rfloor}}{\sqrt{2}} \frac{a'_{\lfloor \frac{k}{2} \rfloor}}{\sqrt{2}} \right) = \frac{1}{2} \sum_{k=0}^{+\infty} \text{Tr} \left(a_{\lfloor \frac{k}{2} \rfloor} a'_{\lfloor \frac{k}{2} \rfloor} \right) \\ &= \sum_{p=0}^{+\infty} \text{Tr} (a_p a_p') = \|X\|_{\bar{\varphi}}^2 \end{aligned}$$

and this quantity is finite. Thus, \mathbf{R} is well-defined and it is a bounded operator.

About A -linearity, consider any matrix $m \in A$ and $X = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}$, $Y = \sum_{k=0}^{\infty} b_k \varepsilon_{t-k}$ in $\mathcal{H}_t(\boldsymbol{\varepsilon})$. The element $X + mY = \sum_{k=0}^{\infty} c_k \varepsilon_{t-k}$ has for coefficients the matrices $c_k = a_k + mb_k$ for any k in \mathbb{N}_0 . Then, \mathbf{R} maps $X + mY$ to the element

$$\mathbf{R}(X + mY) = \sum_{k=0}^{+\infty} \frac{c_{\lfloor \frac{k}{2} \rfloor}}{\sqrt{2}} \varepsilon_{t-k} \in \mathcal{H}_t(\boldsymbol{\varepsilon}), \quad \text{with } c_{\lfloor \frac{k}{2} \rfloor} = a_{\lfloor \frac{k}{2} \rfloor} + mb_{\lfloor \frac{k}{2} \rfloor} \quad \forall k \in \mathbb{N}_0.$$

As a result, $\mathbf{R}(X + mY) = \mathbf{R}X + m\mathbf{R}Y$, i.e., \mathbf{R} is A -linear.

To prove that \mathbf{R} is isometric on $\mathcal{H}_t(\boldsymbol{\varepsilon})$, consider again any X and Y as before. Since $\boldsymbol{\varepsilon}$ is a multivariate white noise,

$$\begin{aligned} \langle \mathbf{R}X, \mathbf{R}Y \rangle_H &= \left\langle \sum_{k=0}^{+\infty} \frac{a_{\lfloor \frac{k}{2} \rfloor}}{\sqrt{2}} \varepsilon_{t-k}, \sum_{h=0}^{+\infty} \frac{b_{\lfloor \frac{h}{2} \rfloor}}{\sqrt{2}} \varepsilon_{t-h} \right\rangle_H \\ &= \frac{1}{2} \sum_{k=0}^{+\infty} \sum_{h=0}^{+\infty} a_{\lfloor \frac{k}{2} \rfloor} \langle \varepsilon_{t-k}, \varepsilon_{t-h} \rangle b'_{\lfloor \frac{h}{2} \rfloor} \\ &= \frac{1}{2} \sum_{k=0}^{+\infty} a_{\lfloor \frac{k}{2} \rfloor} b'_{\lfloor \frac{k}{2} \rfloor} = \sum_{k=0}^{+\infty} a_k b'_k = \langle X, Y \rangle_H. \end{aligned}$$

Hence, \mathbf{R} is an isometry on $\mathcal{H}_t(\boldsymbol{\varepsilon})$. Theorem 1 provides the orthogonal decomposition $\mathcal{H}_t(\boldsymbol{\varepsilon}) = \widehat{\mathcal{H}}_t(\boldsymbol{\varepsilon}) \oplus \widetilde{\mathcal{H}}_t(\boldsymbol{\varepsilon})$, where

$$\widehat{\mathcal{H}}_t(\boldsymbol{\varepsilon}) = \bigcap_{j=0}^{+\infty} \mathbf{R}^j \mathcal{H}_t(\boldsymbol{\varepsilon}), \quad \widetilde{\mathcal{H}}_t(\boldsymbol{\varepsilon}) = \bigoplus_{j=1}^{+\infty} \mathbf{R}^{j-1} \mathcal{L}_t^{\mathbf{R}}$$

and $\mathcal{L}_t^{\mathbf{R}} = \mathcal{H}_t(\boldsymbol{\varepsilon}) \ominus \mathbf{R}\mathcal{H}_t(\boldsymbol{\varepsilon})$ is the wandering submodule.

First, we show that $\widetilde{\mathcal{H}}_t(\boldsymbol{\varepsilon})$ is the null submodule. Indeed, the submodules $\mathbf{R}^j \mathcal{H}_t(\boldsymbol{\varepsilon})$ consist of linear combinations of innovations ε_t with matrix coefficients equal to each others 2^j -by- 2^j :

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

Therefore, $\widehat{\mathcal{H}}_t(\boldsymbol{\varepsilon})$ 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,

$$\sum_{k=0}^{+\infty} \text{Tr}(cc') = \sum_{k=0}^{+\infty} \sum_{p=1}^m \sum_{q=1}^m c^2(p, q)$$

is finite. Since the addends do not depend on k , $c(p, q) = 0$ for all $p, q = 1, \dots, m$ and so c is the null matrix. Consequently, $\widehat{\mathcal{H}}_t(\boldsymbol{\varepsilon}) = \{0\}$ and $\mathcal{H}_t(\boldsymbol{\varepsilon}) = \widetilde{\mathcal{H}}_t(\boldsymbol{\varepsilon})$.

We now turn to the submodule $\widetilde{\mathcal{H}}_t(\boldsymbol{\varepsilon})$. As the orthogonal complement of $\mathbf{R}\mathcal{H}_t(\boldsymbol{\varepsilon})$ is the kernel of the adjoint operator \mathbf{R}^* (Proposition 19), we determine \mathbf{R}^* . In particular, $\mathbf{R}^* : \mathcal{H}_t(\boldsymbol{\varepsilon}) \rightarrow \mathcal{H}_t(\boldsymbol{\varepsilon})$ is defined by

$$\mathbf{R}^* : \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k} \mapsto \sum_{k=0}^{+\infty} \frac{a_{2k} + a_{2k+1}}{\sqrt{2}} \varepsilon_{t-k}.$$

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.,

$$\|Y\|_{\varphi}^2 = \sum_{h=0}^{+\infty} \text{Tr}(a_h a'_h) = \sum_{h=0}^{+\infty} \sum_{p=1}^m \sum_{q=1}^m a_h^2(p, q) < +\infty.$$

Similarly,

$$\begin{aligned} \|\mathbf{R}^* Y\|_{\varphi}^2 &= \frac{1}{2} \sum_{k=0}^{+\infty} \text{Tr}((a_{2k} + a_{2k+1})(a'_{2k} + a'_{2k+1})) \\ &= \frac{1}{2} \sum_{k=0}^{+\infty} \sum_{p=1}^m \sum_{q=1}^m (a_{2k}(p, q) + a_{2k+1}(p, q))^2 \end{aligned}$$

and so

$$\begin{aligned} \|\mathbf{R}^* Y\|_{\varphi}^2 &\leq \sum_{k=0}^{+\infty} \sum_{p=1}^m \sum_{q=1}^m \left((a_{2k}(p, q))^2 + (a_{2k+1}(p, q))^2 \right) \\ &= \sum_{h=0}^{+\infty} \sum_{p=1}^m \sum_{q=1}^m a_h^2(p, q) = \|Y\|_{\varphi}^2. \end{aligned}$$

We deduce that $\|\mathbf{R}^* Y\|^2$ is finite and \mathbf{R}^* is well-defined.

We now establish that $\langle \mathbf{R}X, Y \rangle_H = \langle X, \mathbf{R}^* Y \rangle_H$ for any $X = \sum_{h=0}^{\infty} b_h \varepsilon_{t-h}$ and $Y = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}$ in $\mathcal{H}_t(\boldsymbol{\varepsilon})$. By the unit variance white noise properties of $\boldsymbol{\varepsilon}$,

$$\begin{aligned} \langle \mathbf{R}X, Y \rangle_H &= \sum_{h=0}^{+\infty} \sum_{k=0}^{+\infty} \frac{b_{\lfloor \frac{h}{2} \rfloor}}{\sqrt{2}} \langle \varepsilon_{t-h}, \varepsilon_{t-k} \rangle_H a'_k = \sum_{k=0}^{+\infty} b_{\lfloor \frac{k}{2} \rfloor} \frac{a'_k}{\sqrt{2}} = \sum_{k=0}^{+\infty} b_k \frac{a'_{2k} + a'_{2k+1}}{\sqrt{2}} \\ &= \sum_{h=0}^{+\infty} \sum_{k=0}^{+\infty} b_h \langle \varepsilon_{t-h}, \varepsilon_{t-k} \rangle_H \frac{a'_{2k} + a'_{2k+1}}{\sqrt{2}} = \langle X, \mathbf{R}^* Y \rangle_H. \end{aligned}$$

Therefore, \mathbf{R}^* is the adjoint of the scaling operator.

Regarding the kernel of \mathbf{R}^* , we show that

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

Any element $X = \sum_{k=0}^{\infty} d_k^{(1)} (\varepsilon_{t-2k} - \varepsilon_{t-2k-1})$ of $\mathcal{H}_t(\boldsymbol{\varepsilon})$ 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., $a_{2k} + a_{2k+1} = \mathbf{0}$. Conse-

quently, $\mathbf{R}^*X = 0$ and so

$$\left\{ \sum_{k=0}^{+\infty} d_k^{(1)} (\varepsilon_{t-2k} - \varepsilon_{t-2k-1}) \in \mathcal{H}_t(\mathfrak{e}) : d_k^{(1)} \in A \right\} \subset \ker(\mathbf{R}^*). \tag{B12}$$

Conversely, take any $X = \sum_{h=0}^{+\infty} a_h \varepsilon_{t-h}$ in $\ker(\mathbf{R}^*)$. Since $\|\mathbf{R}^*X\|_{\bar{\varphi}} = 0$,

$$\begin{aligned} & \frac{1}{2} \sum_{k=0}^{+\infty} \text{Tr}((a_{2k} + a_{2k+1})(a'_{2k} + a'_{2k+1})) \\ &= \frac{1}{2} \sum_{k=0}^{+\infty} \sum_{p=1}^m \sum_{q=1}^m (a_{2k}(p, q) + a_{2k+1}(p, q))^2 = 0. \end{aligned}$$

It follows that $a_{2k+1}(p, q) = -a_{2k}(p, q)$ for any $k \in \mathbb{N}_0$ and $p, q = 1, \dots, m$. Therefore, $a_{2k+1} = -a_{2k}$ for any $k \in \mathbb{N}_0$. As a result, $X = \sum_{k=0}^{+\infty} d_k^{(1)} (\varepsilon_{t-2k} - \varepsilon_{t-2k-1})$ with $d_k^{(1)} = a_{2k}$ and so the converse inclusion in (B12) holds. As a result,

$$\mathcal{L}_t^{\mathbf{R}} = \ker(\mathbf{R}^*) = \left\{ \sum_{k=0}^{+\infty} b_k^{(1)} \varepsilon_{t-2k} \in \mathcal{H}_t(\mathfrak{e}) : b_k^{(1)} \in A \right\}.$$

Moreover,

$$\mathbf{R}\mathcal{L}_t^{\mathbf{R}} = \left\{ \sum_{k=0}^{+\infty} b_k^{(2)} \varepsilon_{t-4k} \in \mathcal{H}_t(\mathfrak{e}) : b_k^{(2)} \in A \right\}$$

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

$$\mathbf{R}^{j-1}\mathcal{L}_t^{\mathbf{R}} = \left\{ \sum_{k=0}^{+\infty} b_k^{(j)} \varepsilon_{t-2kj} \in \mathcal{H}_t(\mathfrak{e}) : b_k^{(j)} \in A \right\}.$$

As the case with $j \in \mathbb{N}$ follows by induction, we focus on $\mathbf{R}\mathcal{L}_t^{\mathbf{R}}$ and prove that

$$\mathbf{R}\mathcal{L}_t^{\mathbf{R}} = \left\{ \sum_{k=0}^{+\infty} d_k^{(2)} (\varepsilon_{t-4k} + \varepsilon_{t-4k-1} - \varepsilon_{t-4k-2} - \varepsilon_{t-4k-3}) \in \mathcal{H}_t(\mathfrak{e}) : d_k^{(2)} \in A \right\}. \tag{B13}$$

Consider any $Y \in \mathbf{R}\mathcal{L}_t^{\mathbf{R}}$. As Y is the image of some $X \in \mathcal{L}_t^{\mathbf{R}}$, there exists a sequence of matrices $\{d_k^{(1)}\}_k$ such that $X = \sum_{k=0}^{+\infty} d_k^{(1)} (\varepsilon_{t-2k} - \varepsilon_{t-2k-1})$ and

$$Y = \sum_{k=0}^{+\infty} \frac{d_k^{(1)}}{\sqrt{2}} (\varepsilon_{t-4k} + \varepsilon_{t-4k-1} - \varepsilon_{t-4k-2} - \varepsilon_{t-4k-3}).$$

As a result, $\mathbf{R}\mathcal{L}_t^{\mathbf{R}}$ is included in the module in (B13).

Conversely, consider any $\sum_{k=0}^{\infty} d_k^{(2)} (\varepsilon_{t-4k} + \varepsilon_{t-4k-1} - \varepsilon_{t-4k-2} - \varepsilon_{t-4k-3})$ in $\mathcal{H}_t(\boldsymbol{\varepsilon})$. Then, Y belongs to $\mathbf{R}\mathcal{L}_t^{\mathbf{R}}$ too, because it is the image of $X = \sum_{k=0}^{\infty} \sqrt{2} d_k^{(2)} (\varepsilon_{t-2k} - \varepsilon_{t-2k-1})$ in $\mathcal{L}_t^{\mathbf{R}}$. Consequently, the relation in (B13) holds.

The decomposition of the Hilbert A -module $\mathcal{H}_t(\boldsymbol{\varepsilon})$ is, then, achieved. ■

B.3 Proof of Theorem 4

By applying the MCWD (Theorem 2) 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 $\boldsymbol{\varepsilon}$ is the unit variance white noise of classical Wold innovations of \mathbf{x} . Notably, $\mathcal{H}_t(\boldsymbol{\varepsilon})$ orthogonally decomposes as in Theorem 3. By denoting $g_t^{(j)}$ the orthogonal projections of 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 (recall that $\|\cdot\|_H$ and $\|\cdot\|_{\bar{\varphi}}$ are equivalent by Proposition 12). Then, by using the characterizations of submodules $\mathbf{R}^{j-1}\mathcal{L}_t^{\mathbf{R}}$, for any scale $j \in \mathbb{N}$ we find a sequence of matrices $\{\beta_k^{(j)}\}_k$ such that Eq. (8) holds with $\sum_{k=0}^{\infty} \text{Tr}(\beta_k^{(j)} \beta_k^{(j)'}) < +\infty$. As a consequence, we can decompose x_t as in Eq. (6).

1. As we can see in Eq. (3), the process $\boldsymbol{\varepsilon}_t^{(j)}$ is a $MA(2^j - 1)$ with respect to the fundamental innovations $\boldsymbol{\varepsilon}$. In addition, the subprocess $\{\boldsymbol{\varepsilon}_{t-k2^j}^{(j)}\}_{k \in \mathbb{Z}}$ is weakly stationary. Indeed, since $\boldsymbol{\varepsilon}$ is a multivariate white noise, $\mathbb{E}[\boldsymbol{\varepsilon}_{t-k2^j}^{(j)} \boldsymbol{\varepsilon}_{t-k2^j}^{(j)'}]$ is finite and it does not depend on k : for any $k \in \mathbb{Z}$,

$$\begin{aligned} \mathbb{E} \left[\boldsymbol{\varepsilon}_{t-k2^j}^{(j)} \boldsymbol{\varepsilon}_{t-k2^j}^{(j)'} \right] &= \frac{1}{2^j} \mathbb{E} \left[\left(\sum_{i=0}^{2^j-1} \varepsilon_{t-k2^j-i} - \sum_{i=0}^{2^j-1} \varepsilon_{t-k2^j-2^{j-1}-i} \right) \right. \\ &\quad \cdot \left. \left(\sum_{i=0}^{2^j-1} \varepsilon'_{t-k2^j-i} - \sum_{i=0}^{2^j-1} \varepsilon'_{t-k2^j-2^{j-1}-i} \right) \right] \\ &= \frac{1}{2^j} \sum_{i=0}^{2^j-1} \mathbb{E} [\varepsilon_t \varepsilon_t'] = \frac{2^j}{2^j} \mathbb{E} [\varepsilon_t \varepsilon_t'] = I. \end{aligned}$$

In addition, $\mathbb{E}[\boldsymbol{\varepsilon}_{t-k2^j}^{(j)}] = 0$ for any $k \in \mathbb{Z}$ and the expectation does not depend on k . Regarding the cross-moment matrix on the support $S_t^{(j)} = \{t - k2^j : k \in \mathbb{N}_0\}$, for any $h \neq k$,

$$\begin{aligned} \mathbb{E} \left[\boldsymbol{\varepsilon}_{t-h2^j}^{(j)} \boldsymbol{\varepsilon}_{t-k2^j}^{(j)'} \right] &= \frac{1}{2^j} \mathbb{E} \left[\left(\sum_{i=0}^{2^j-1} \varepsilon_{t-h2^j-i} - \sum_{i=0}^{2^j-1} \varepsilon_{t-h2^j-2^{j-1}-i} \right) \right. \\ &\quad \cdot \left. \left(\sum_{l=0}^{2^j-1} \varepsilon'_{t-k2^j-l} - \sum_{l=0}^{2^j-1} \varepsilon'_{t-k2^j-2^{j-1}-l} \right) \right] \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{2^j} \left\{ \sum_{i=0}^{2^{j-1}-1} \sum_{l=0}^{2^{j-1}-1} \mathbb{E} \left[\varepsilon_{t-h2^j-i} \varepsilon'_{t-k2^j-l} \right] \right. \\
 &\quad - \sum_{i=0}^{2^{j-1}-1} \sum_{l=0}^{2^{j-1}-1} \mathbb{E} \left[\varepsilon_{t-h2^j-i} \varepsilon'_{t-k2^j-2^{j-1}-l} \right] \\
 &\quad - \sum_{i=0}^{2^{j-1}-1} \sum_{l=0}^{2^{j-1}-1} \mathbb{E} \left[\varepsilon_{t-h2^j-2^{j-1}-i} \varepsilon'_{t-k2^j-l} \right] \\
 &\quad \left. + \sum_{i=0}^{2^{j-1}-1} \sum_{l=0}^{2^{j-1}-1} \mathbb{E} \left[\varepsilon_{t-h2^j-2^{j-1}-i} \varepsilon'_{t-k2^j-2^{j-1}-l} \right] \right\}.
 \end{aligned}$$

The sets of indices $\{h2^j, \dots, h2^j + 2^j - 1\}$ and $\{k2^j, \dots, k2^j + 2^j - 1\}$ are disjoint because of $h \neq k$ and so the last sums are null. In consequence, $\mathbb{E}[\varepsilon_{t-h2^j}^{(j)} \varepsilon_{t-k2^j}^{(j)'}] = \mathbf{0}$ for all $h \neq k$. Thus, $\{\varepsilon_{t-k2^j}^{(j)}\}_{k \in \mathbb{Z}}$ turns out to be weakly stationary on $S_t^{(j)}$. In particular, it is a unit variance white noise.

2. 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 $\mathbf{R}^j \mathcal{H}_t(\boldsymbol{\varepsilon})$, and we proceed inductively.

We begin with the decomposition $x_t = \pi_t^{(1)} + g_t^{(1)}$ coming from scale $J = 1$, i.e., $\mathcal{H}_t(\boldsymbol{\varepsilon}) = \mathbf{R} \mathcal{H}_t(\boldsymbol{\varepsilon}) \oplus \mathcal{L}_t^{\mathbf{R}}$. By using the characterization of submodules $\mathbf{R} \mathcal{H}_t(\boldsymbol{\varepsilon})$ and $\mathcal{L}_t^{\mathbf{R}}$ described in the proof of Theorem 3, we set

$$\begin{aligned}
 \pi_t^{(1)} &= \sum_{k=0}^{+\infty} \gamma_k^{(1)} \frac{\varepsilon_{t-2k} + \varepsilon_{t-(2k+1)}}{\sqrt{2}} = \sum_{k=0}^{+\infty} c_k^{(1)} (\varepsilon_{t-2k} + \varepsilon_{t-(2k+1)}), \\
 g_t^{(1)} &= \sum_{k=0}^{+\infty} \beta_k^{(1)} \varepsilon_{t-2k} = \sum_{k=0}^{+\infty} d_k^{(1)} (\varepsilon_{t-2k} - \varepsilon_{t-2k-1})
 \end{aligned}$$

for some sequences of matrices $\{c_k^{(1)}\}_k$ and $\{d_k^{(1)}\}_k$, or equivalently $\{\gamma_k^{(1)}\}_k$ and $\{\beta_k^{(1)}\}_k$, to determine in order to have $x_t = \pi_t^{(1)} + g_t^{(1)}$, where we set $\sqrt{2}c_k^{(1)} = \gamma_k^{(1)}$ and $\sqrt{2}d_k^{(1)} = \beta_k^{(1)}$. The expressions above may be rewritten as

$$x_t = \sum_{k=0}^{+\infty} \left((c_k^{(1)} + d_k^{(1)}) \varepsilon_{t-2k} + (c_k^{(1)} - d_k^{(1)}) \varepsilon_{t-2k-1} \right).$$

However, from Theorem 2 we know that

$$x_t = \sum_{k=0}^{+\infty} (\alpha_{2k} \varepsilon_{t-2k} + \alpha_{2k+1} \varepsilon_{t-2k-1}),$$

where the same fundamental innovations ε_t are present. By the uniqueness of writing of the MCWD, the two expressions for x_t must coincide. As a result, $c_k^{(1)}$ and $d_k^{(1)}$ are the solutions of the linear system

$$\begin{cases} c_k^{(1)} + d_k^{(1)} = \alpha_{2k} \\ c_k^{(1)} - d_k^{(1)} = \alpha_{2k+1}, \end{cases}$$

that is,

$$c_k^{(1)} = \frac{\alpha_{2k} + \alpha_{2k+1}}{2}, \quad d_k^{(1)} = \frac{\alpha_{2k} - \alpha_{2k+1}}{2}.$$

In particular, we find

$$\gamma_k^{(1)} = \frac{\alpha_{2k} + \alpha_{2k+1}}{\sqrt{2}}, \quad \beta_k^{(1)} = \frac{\alpha_{2k} - \alpha_{2k+1}}{\sqrt{2}}.$$

Now, we focus on the scale $J = 2$. We exploit the decomposition of the submodule $\mathbf{R}\mathcal{H}_t(\boldsymbol{\varepsilon}) = \mathbf{R}^2\mathcal{H}_t(\boldsymbol{\varepsilon}) \oplus \mathbf{R}\mathcal{L}_t^{\mathbf{R}}$, which 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. Since

$$\pi_t^{(2)} = \sum_{k=0}^{+\infty} \gamma_k^{(2)} \frac{\varepsilon_{t-4k} + \varepsilon_{t-(4k+1)} + \varepsilon_{t-(4k+2)} + \varepsilon_{t-(4k+3)}}{2}, \quad g_t^{(2)} = \sum_{k=0}^{+\infty} \beta_k^{(2)} \varepsilon_{t-4k},$$

by solving a linear system, we get

$$\gamma_k^{(2)} = \frac{\alpha_{4k} + \alpha_{4k+1} + \alpha_{4k+2} + \alpha_{4k+3}}{2}, \quad \beta_k^{(2)} = \frac{\alpha_{4k} + \alpha_{4k+1} - \alpha_{4k+2} - \alpha_{4k+3}}{2}.$$

At the generic scale $J = j$, we retrieve the expressions of $\beta_k^{(j)}$ and $\gamma_k^{(j)}$ of Eqs. (7) and (9), where $\pi_t^{(j)}$ is also defined.

3. First of all, when t is fixed, $\langle g_t^{(j)}, g_t^{(l)} \rangle_H = \mathbb{E}[g_t^{(j)} g_t^{(l)'}] = \mathbf{0}$ 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 $g_{t-2j}^{(j)}$ with $m \in \mathbb{N}_0$. Clearly, $g_{t-2j}^{(j)}$ belongs to $\mathbf{R}^{j-1}\mathcal{L}_{t-2j}^{\mathbf{R}}$ but, by the definition of $g_t^{(j)}$, we can write

$$g_{t-2j}^{(j)} = \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-(m+k)2j} = \sum_{K=0}^{+\infty} \beta_K^{(j)} \varepsilon_{t-K2j},$$

with $\beta_K^{(j)} = \mathbf{0}$ if $K = 0, \dots, m - 1$ and $\beta_K^{(j)} = \beta_k^{(j)}$ if $K = m + k$ for some $k \in \mathbb{N}_0$. 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 $\mathbb{E}[g_{t-m2^j}^{(j)} g_{t-n2^l}^{(l)'}] = \mathbf{0}$ for all $j \neq l$ and $m, n \in \mathbb{N}_0$.

As for the general requirement about $\mathbb{E}[g_{t-p}^{(j)} g_{t-q}^{(l)'}]$ for any $j, l \in \mathbb{N}$ and $p, q, t \in \mathbb{Z}$, we have

$$\begin{aligned} \mathbb{E} \left[g_{t-p}^{(j)} g_{t-q}^{(l)'} \right] &= \sum_{k=0}^{+\infty} \sum_{h=0}^{+\infty} \beta_k^{(j)} \mathbb{E} \left[\varepsilon_{t-p-k2^j}^{(j)} \varepsilon_{t-q-h2^l}^{(l)'} \right] \beta_h^{(l)'} \\ &= \frac{1}{\sqrt{2^{j+l}}} \sum_{k=0}^{+\infty} \sum_{h=0}^{+\infty} \beta_k^{(j)} \sum_{u=0}^{2^{j-1}-1} \sum_{v=0}^{2^{l-1}-1} \left\{ \mathbb{E} \left[\varepsilon_{t-p-k2^j-u} \varepsilon_{t-q-h2^l-v}' \right] \right. \\ &\quad - \mathbb{E} \left[\varepsilon_{t-p-k2^j-u} \varepsilon_{t-q-h2^l-2^{l-1}-v}' \right] - \mathbb{E} \left[\varepsilon_{t-p-k2^j-2^{j-1}-u} \varepsilon_{t-q-h2^l-v}' \right] \\ &\quad \left. + \mathbb{E} \left[\varepsilon_{t-p-k2^j-2^{j-1}-u} \varepsilon_{t-q-h2^l-2^{l-1}-v}' \right] \right\} \beta_h^{(l)'} \end{aligned}$$

and so

$$\begin{aligned} \mathbb{E} \left[g_{t-p}^{(j)} g_{t-q}^{(l)'} \right] &= \frac{1}{\sqrt{2^{j+l}}} \sum_{k=0}^{+\infty} \sum_{h=0}^{+\infty} \beta_k^{(j)} \sum_{u=0}^{2^{j-1}-1} \sum_{v=0}^{2^{l-1}-1} \left\{ \Gamma_{p-q+k2^j+u-h2^l-v} \right. \\ &\quad - \Gamma_{p-q+k2^j+u-h2^l-2^{l-1}-v} - \Gamma_{p-q+k2^j+2^{j-1}+u-h2^l-v} \\ &\quad \left. + \Gamma_{p-q+k2^j+2^{j-1}+u-h2^l-2^{l-1}-v} \right\} \beta_h^{(l)'} \end{aligned}$$

where the matrices $\beta_k^{(j)}, \beta_h^{(l)}$ do not depend on t and Γ_n denotes the autocovariance matrix of ε at lag $n \in \mathbb{Z}$. Hence, after the summations over u, v and k, h , the one remaining variables are $j, l, p - q$. Therefore, $\mathbb{E}[g_{t-p}^{(j)} g_{t-q}^{(l)'}]$ depends at most on $j, l, p - q$. ■

Appendix C Illustrations of the MEWD

To put the MEWD into practice, we first compute the scale-specific responses of weakly stationary $VAR(1)$ and $VARMA(1, 1)$ processes in closed form. Afterwards, we briefly describe how to estimate the MEWD of weakly stationary processes in general. Finally, we analyze the persistent dynamics of Blanchard and Quah (1989) bivariate process through the MEWD.

C.1 The MEWD of VAR(1) and VARMA(1,1)

Consider a weakly stationary purely non-deterministic vector $ARMA(1, 1)$ process, or $VARMA(1, 1)$, $\mathbf{x} = \{x_t\}_{t \in \mathbb{Z}}$ 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 $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t \in \mathbb{Z}}$ is a multivariate unit variance white noise. We assume the stationarity condition $\|\rho\|_A < 1$.⁸

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. Moreover, the operator $(I - \rho\mathbf{L})$ is invertible with $\sum_{l=0}^{\infty} (\rho\mathbf{L})^l$ as inverse. Therefore, the moving average representation of x_t is

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

with $\alpha_0 = 1$ and $\alpha_h = \rho^{h-1}(\rho + \theta)$ for all $h \in \mathbb{N}$.

We compute the scale-specific responses by Eq. (7). Fixed a scale $j \in \mathbb{N}$, we obtain

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

$$\beta_k^{(j)} = \frac{1}{\sqrt{2^j}} (I - \rho)^{-1} \left(I - \rho^{2^{j-1}} \right)^2 \rho^{k2^j-1} (\rho + \theta) \quad \forall k \in \mathbb{N}.$$

By setting $\theta = 0$, we find the scale-specific responses of a $VAR(1)$:

$$\beta_k^{(j)} = \frac{1}{\sqrt{2^j}} (I - \rho)^{-1} \left(I - \rho^{2^{j-1}} \right)^2 \rho^{k2^j} \quad \forall k \in \mathbb{N}_0.$$

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 $\rho = [a, b; c, d]$, 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 scale-specific responses $\beta_k^{(j)}$ turn out to be

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

C.2 Estimation of the MEWD

When the moving average representation of the weakly stationary process $\mathbf{x} = \{x_t\}_{t \in \mathbb{Z}}$ is not known, to obtain the MEWD, one can follow the similar procedure of Subsections 3.1.1 and 3.2.1 in Ortu et al. (2020a) and Section 5 in Di Virgilio et al. (2019). The

⁸ In the literature, other assumptions are also considered, e.g., stability (Lütkepohl 2005, Chapter 2).

first step is to estimate a vector autoregressive form for x_t . Then, a moving average representation can be retrieved. Finally, detail processes and scale-specific responses can be obtained from Eqs. (3) and (7), respectively. More details can be found in the next subsection, in the special case of Blanchard and Quah (1989) model.

C.3 Blanchard and Quah’s model

The macroeconomic model of Blanchard and Quah (1989) studies the impulse responses of GNP and unemployment to demand and supply shocks. We reproduce the model and analyze the responses through the MEWD in order to quantify the persistence of different disturbances.

We consider the algebra A of 2×2 matrices, and 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

$$x_t = \sum_{h=0}^{+\infty} \alpha_h \varepsilon_{t-h}, \quad \alpha_h \in A, \tag{C14}$$

where $\varepsilon = \{\varepsilon_t\}_{t \in \mathbb{Z}}$ is a unit variance bivariate white noise and $\sum_{h=0}^{+\infty} \text{Tr}(\alpha_h \alpha_h')$ is finite. In Blanchard and Quah (1989), $x_t = [y_t, z_t]'$, 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. The impulse responses satisfy the *long-run restriction* $\sum_{h=0}^{+\infty} \alpha_h (1, 1) = 0$, which is crucial for the identification. Indeed, x_t is supposed to have also the *MA* representation

$$x_t = \sum_{h=0}^{+\infty} c_h \eta_{t-h}, \quad c_h \in A, \tag{C15}$$

where $\eta = \{\eta_t\}_{t \in \mathbb{Z}}$ is a bivariate with noise with covariance matrix ω , generally different from the identity. Here, η_t are the reduced-form residuals, while ε_t are the structural shocks. Equation (C15) provides the formulation obtained by estimating the time series parameters from the data. Specifically, we first estimate a vector autoregressive form for x_t , that is

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

The matrix ω is estimated by the covariance matrix of the residuals in the multivariate regression. Then, the autoregressive form implies that

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

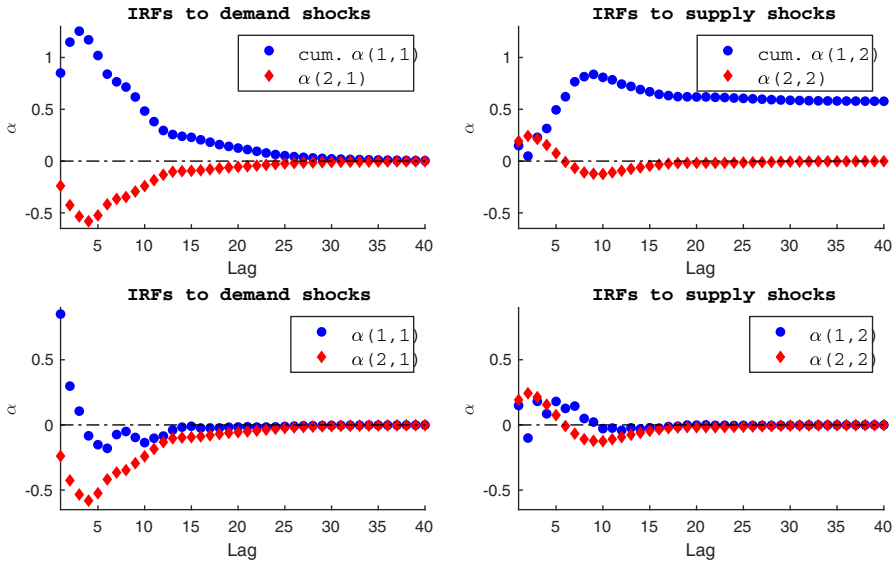


Fig. 1 Bottom: impulse responses of output growth (blue) and unemployment rate (red) with respect to demand (left panel) and supply (right panel) shocks. The sole difference of upper graphs is that they contain the impulse responses of output (the cumulated responses of output growth) (Color figure online)

From the last expression, we can find the *MA* matrix coefficients of Eq. (C15):

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

The *MA* representations (C14) and (C15) of x_t are related by $\eta_t = \alpha_0 \varepsilon_t$ and $\alpha_h = c_h \alpha_0$, 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 only three conditions for the identification of α_0 . The long-run restriction (together with the signs restrictions) is an additional requirement that ensures the identification of structural shocks (Lütkepohl 2005, Section 9.1.4).

By the Cholesky factorization, there exists a unique lower triangular matrix s , such that $\omega = s s'$ (Trefethen and Bau III 1997, Lecture 23). Any α_0 such that $\omega = \alpha_0 \alpha_0'$ is an orthonormal transformation of s , namely $\alpha_0 = s r'$ 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

$$r = -\frac{1}{\sqrt{\vartheta^2 + 1}} \begin{bmatrix} \vartheta & 1 \\ -1 & \vartheta \end{bmatrix}, \quad \vartheta = -\frac{s(2, 2) \sum_{h=0}^{\infty} c_h(1, 2)}{s(1, 1) \sum_{h=0}^{\infty} c_h(1, 1) + s(2, 1) \sum_{h=0}^{\infty} c_h(1, 2)}.$$

By using these parameters, we get the *MA* of Eq. (C14), where the univariate shocks are simultaneously uncorrelated.

Blanchard and Quah (1989) data, which are freely released, are quarterly and span from 1950 : $Q2$ to 1987 : $Q4$. The maximum autoregressive lag N is chosen equal to 8. To reduce non-stationarity, z_t is linearly detrended, while y_t is demeaned by splitting the sample into two parts: before and after 1973 : $Q4$. The bivariate structural 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 y_t . They are plotted (together with the impulse responses of z_t) in the top panels of Fig. 1, which reproduce Figures 1 and 2 in Blanchard and Quah (1989). The impulse responses of output with respect to u_t converge to zero in the long term by the long-run restriction. This phenomenon is not present in the impulse responses of output with respect to v_t : demand shocks have transitory effects, while supply disturbances have a permanent impact on output.

Demand shocks have opposite hump-shaped effects on output and unemployment, with a peak after two or four quarters. Moreover, the impact of u_t vanishes after 3 or 5 years: 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 2 years. Except for the first quarter, output is increasing. Then, the output response declines and stabilizes on a steady level after 5 years from the initial shock. A different reaction, however, characterizes the unemployment rate. Indeed, even if the supply disturbance is favorable (due, e.g., to a productivity increase), in the short term z_t rises, plausibly because of wage rigidities. After several quarters z_t drops and, later, it slowly reverts to the original value. No effect is present after 5 years.

The lower graphs of Fig. 1 display the (non-cumulated) impulse responses of y_t , together with those of z_t . The non-monotonic responses of output in top panels correspond to oscillatory and sign changing responses of y_t in bottom panels. The latter reveal the presence of contrasting reactions that, overall, generate the hump-shaped responses. We now use the MEWD to shed light on the persistence of the shocks causing opposite reactions on different time scales.

From a sample of 159 data points, we estimate the first 7 scales in the MEWD and we plot in Fig. 2 the variance decompositions of y_t and z_t illustrated in Sect. 4.3. The two left panels, inspired by Eq. (10), regard the decomposition across scales without distinguishing between the two sources of randomness. The middle and right panels further disentangle the contribution across time scales of demand and supply shocks, following Eq. (11). From the figure, it is apparent that the unemployment rate is a longer term phenomenon than output growth. Indeed, most of the variance of z_t is explained by persistent components at scales 4, 5 and 6, involving shocks from 4 to 16 years. On the contrary, the variance of y_t comes from scales 1–3, i.e., from disturbances lasting from 6 months to 2 years, with an additional sizable contribution given by scale 4. Moreover, the analysis of individual shocks reveals that demand shocks provide the most variability of both output growth and unemployment. In agreement with the long-run restriction, the response of y_t to supply shocks operates mainly at scale 4 (involving 4-year innovations), while the response of the same variable to demand shocks is more concentrated on lower scales (1–3). However, demand shocks explain relatively more variance.

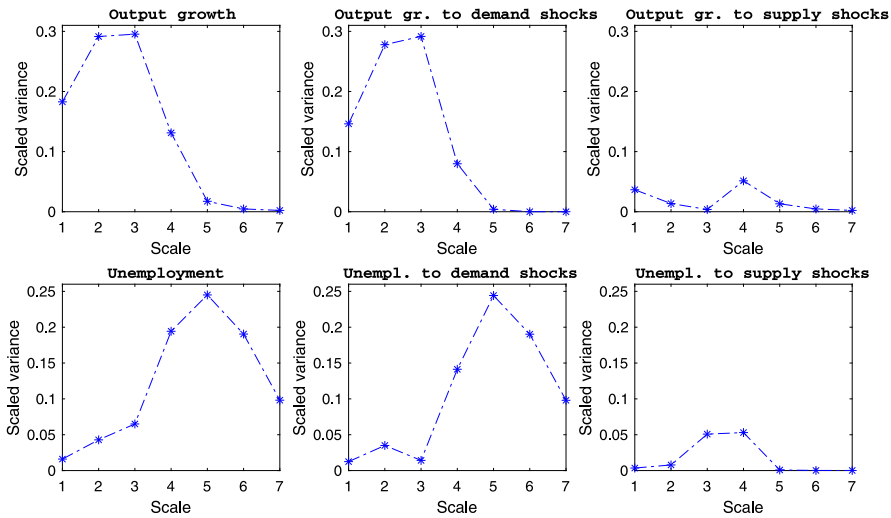


Fig. 2 Relative variance of output growth and unemployment rate explained by each scale

We can compare the results of Fig. 2 with the analysis of the spectral density matrix of x_t . Following Example 11.8.1 in Brockwell and Davis (2006), we rewrite Eq. (C14) by using a time-invariant linear filter U :

$$x_t = U(\mathbf{L})\varepsilon_t \quad \text{with} \quad U(\mathbf{L}) = \sum_{h=0}^{+\infty} \alpha_h \mathbf{L}^h,$$

where we assume that the entries of $U(\mathbf{L})$ are absolutely summable. Then, the spectral density matrix of x_t is, for any $\pi < \lambda \leq \pi$,

$$f_x(\lambda) = \frac{1}{2\pi} U \left(e^{-i\lambda} \right) U' \left(e^{i\lambda} \right) = \begin{bmatrix} f_y(\lambda) & f_{yz}(\lambda) \\ f_{zy}(\lambda) & f_z(\lambda) \end{bmatrix},$$

where the spectral density of output growth satisfies

$$\begin{aligned} 2\pi f_y(\lambda) &= \left| \sum_{h=0}^{+\infty} \alpha_h(1, 1)e^{-ih\lambda} \right|^2 + \left| \sum_{h=0}^{+\infty} \alpha_h(1, 2)e^{-ih\lambda} \right|^2 \\ &= \sum_{h=0}^{+\infty} \sum_{k=0}^{+\infty} (\alpha_h(1, 1)\alpha_k(1, 1) + \alpha_h(1, 2)\alpha_k(1, 2)) \cos(\lambda(k - h)), \end{aligned} \quad (C16)$$

the spectral density of unemployment is similar, i.e.,

$$2\pi f_z(\lambda) = \left| \sum_{h=0}^{+\infty} \alpha_h(2, 1)e^{-ih\lambda} \right|^2 + \left| \sum_{h=0}^{+\infty} \alpha_h(2, 2)e^{-ih\lambda} \right|^2, \quad (C17)$$

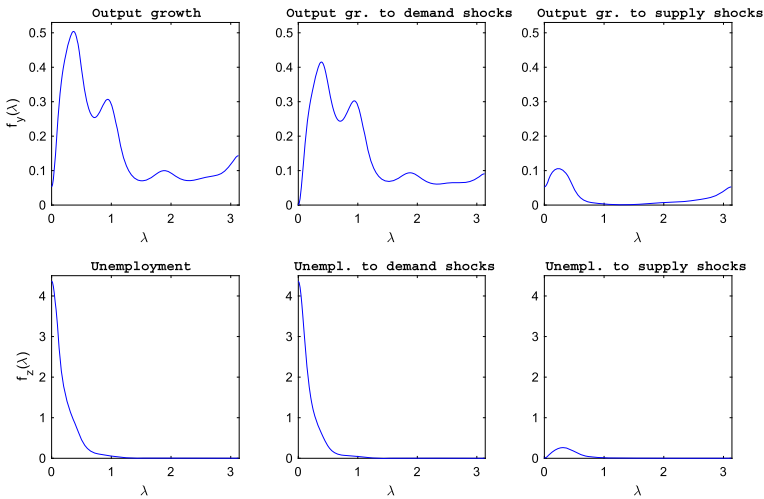


Fig. 3 On the left, spectral densities of output growth and unemployment rate. In the middle, contributions of demand shocks to such densities. On the right, contributions of supply shocks to the same densities

the cross-spectrum satisfies

$$2\pi f_{yz}(\lambda) = \sum_{h=0}^{+\infty} \sum_{k=0}^{+\infty} (\alpha_h(1, 1)\alpha_k(2, 1) + \alpha_h(1, 2)\alpha_k(2, 2)) e^{i(k-h)\lambda}$$

and $f_{zy}(\lambda)$ is the complex conjugate of $f_{yz}(\lambda)$.

The left panels of Fig. 3 display the spectral densities of output growth and unemployment. Middle panels represent the contributions of demand shocks to such densities, i.e., the first summation terms in Eqs. (C16) and (C17). Similarly, right panels display the contribution of supply shocks to f_y and f_z . Demand shocks keep a prominent role with respect to supply disturbances, as it was already clear from Fig. 2. The peaks of f_y are consistent with the variance of output growth explained by scales 1–4 (top-left panels in both figures). Moreover, the low density at 0 is consistent with the negligible variance of output explained by high scales. Supply shocks feature a peak of frequencies at roughly 0.23, in line with the peak of variance at scale 4 (top-right panels in the same figures). The spectral analysis confirms the persistent nature of unemployment, which features a concentration of frequencies around 0 (bottom-left panels). Hence, Figs. 2 and 3 are consistent within them. However, the variance decomposition of the MEWD provides simple interpretations directly in the time domain.

Figure 4 displays the scale-specific responses $\beta_k^{(j)}$ of x_t . As to supply shocks, we confirm the positive contemporaneous reaction of output growth with respect to 4-year innovations, captured by $\beta_0^{(4)}(1, 2)$. Regarding the reaction of output growth to demand shocks, the positive reaction causing the surge of the hump in cumulated responses is principally due to the coefficients $\beta_0^{(j)}(1, 1)$ with $j = 1, \dots, 4$. Such responses are simultaneous with the shock and positive. The responses at other lags and scales

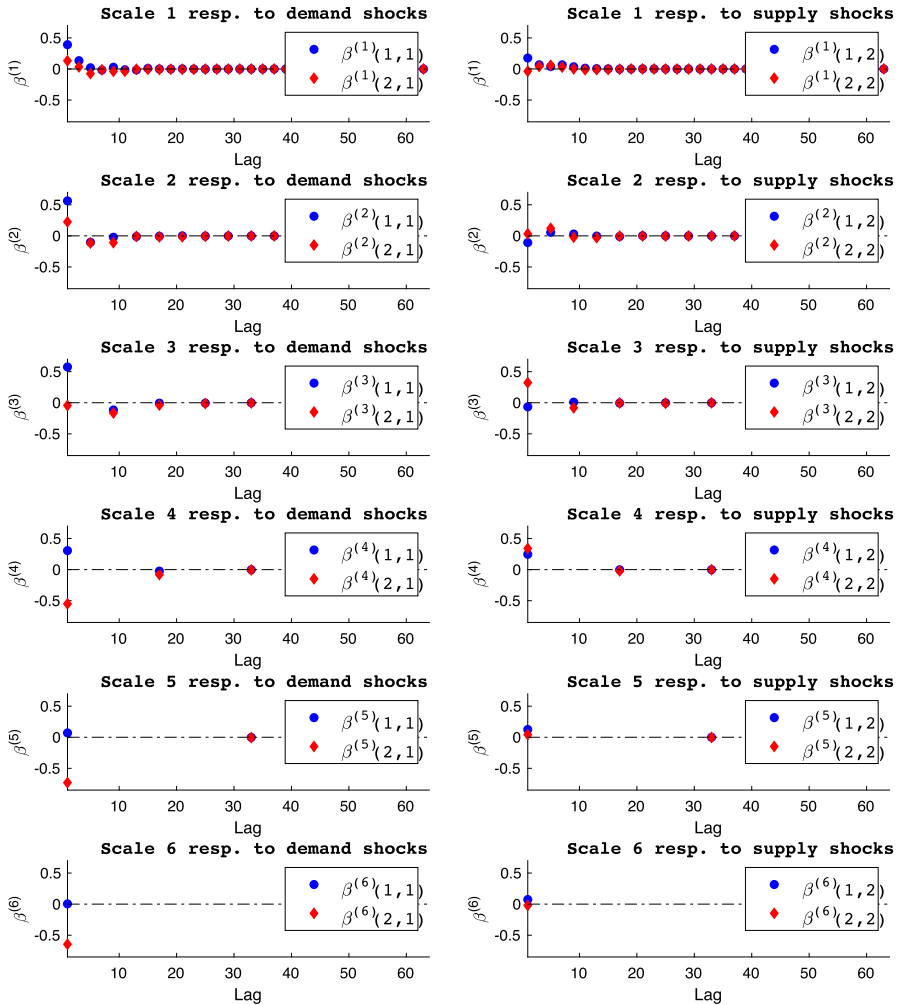


Fig. 4 Scale-specific responses of output growth (in blue) and unemployment rate (in red) with respect to demand (left panels) and supply (right panels) shocks at scales 1–6 (Color figure online)

mainly contribute to stabilize, and definitely fade out, this positive effect over time. In line with the variance decompositions, the responses of unemployment rate to demand shocks are mainly dictated by the coefficients $\beta_0^{(j)}(2, 1)$ with $j = 4, 5, 6$, which are negative and occur instantaneously (lag zero). These scale-specific responses generate the fall in the hump of the responses of z_t to demand shocks. Interestingly, other scale-specific coefficients, as for example $\beta_0^{(2)}(2, 1)$, are positive, proving the coexistence of contrasting reactions at different time scales. The simultaneous positive impact of demand shocks at scale 2 concurs to delay the large drop in unemployment and to make the hump arise.

Hence, by the MEWD we can disaggregate demand/supply calendar-time shocks and quantify the impact of innovations with heterogeneous duration in Blanchard and Quah's model. First, the permanent effect of supply shocks on GNP growth (assumed by the long-run restriction) turns out to be linked to 4-year innovations. Overall, demand shocks are more important than supply shocks in the variance explanation. In particular, output growth positively responds to demand shocks lasting from 6 months to 2 years, while the unemployment rate evolves primarily on a 4, 8 or 16-year basis. In addition, the analysis of scale-specific responses can help in disentangling the positive and negative reactions at different time scales that, aggregated, generate the hump-shaped behavior of cumulated responses. The economic rigidities advocated by Blanchard and Quah (1989) can be individuated at some scale and lag, providing useful tools to the policy maker.

References

- Abramovich, Y.A., Aliprantis, C.D.: *An Invitation to Operator Theory*. American Mathematical Society, Providence (2002)
- Addison, P.S.: *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance*. IOP Publishing, London (2002)
- Aliprantis, C.D., Border, K.C.: *Infinite Dimensional Analysis: A Hitchhiker's Guide*, 3rd edn. Springer, Berlin (2006)
- Bandi, F.M., Perron, B., Tamoni, A., Tebaldi, C.: The scale of predictability. *J. Econom.* **208**(1), 120–140 (2019)
- Bandi, F.M., Chaudhuri, S., Lo, A.W., Tamoni, A.: Spectral factor models. *J. Financ. Econ.* **142**(1), 214–238 (2021)
- Bierens, H.J.: *Introduction to the Mathematical and Statistical Foundations of Econometrics*. Cambridge University Press, Cambridge (2005)
- Bierens, H.J.: The Wold decomposition (2012). <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.231.2308>
- Blanchard, O.J., Quah, D.: The dynamic effects of aggregate demand and supply disturbances. *Am. Econ. Rev.* **79**(4), 655–673 (1989)
- Brezis, H.: *Functional Analysis. Sobolev Spaces and Partial Differential Equations*. Springer, New York (2011)
- Brockwell, P.J., Davis, R.A.: *Time Series: Theory and Methods*, 2nd edn. Springer, New York (2006)
- Cerreia-Vioglio, S., Maccheroni, F.A., Marinacci, M.: Hilbert A -modules. *J. Math. Anal. Appl.* **446**(1), 970–1017 (2017)
- Cerreia-Vioglio, S., Maccheroni, F.A., Marinacci, M.: Orthogonal decompositions in Hilbert A -modules. *J. Math. Anal. Appl.* **470**(2), 846–875 (2019)
- Cerreia-Vioglio, S., Ortu, F., Rotondi, F., Severino, F.: On horizon-consistent mean-variance portfolio allocation. *Ann. Oper. Res.* (2022). <https://doi.org/10.1007/s10479-022-04798-x>
- Di Virgilio, D., Ortu, F., Severino, F., Tebaldi, C.: Optimal asset allocation with heterogeneous persistent shocks and myopic and intertemporal hedging demand. In: Venezia, I. (ed.) *Behavioral Finance: The Coming of Age*, pp. 57–108. World Scientific, Singapore (2019)
- Durbin, J.: Efficient estimation of parameters in moving-average models. *Biometrika* **46**(3/4), 306–316 (1959)
- Efron, B., Tibshirani, R.: Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.* **1**, 54–75 (1986)
- Frank, M.: Hilbert C^* -modules and related subjects—a guided reference overview (1996). [arXiv:funct-an/9605003](https://arxiv.org/abs/1906.05003)
- Gallant, A.R., Hansen, L.P., Tauchen, G.: Using conditional moments of asset payoffs to infer the volatility of intertemporal marginal rates of substitution. *J. Econom.* **45**(1–2), 141–179 (1990)
- Ghysels, E., Khalaf, L., Vodounou, C.: Simulation based inference in moving average models. *Ann. Econ. Stat.* **69**, 85–99 (2003)

- Goldstine, H.H., Horwitz, L.P.: Hilbert space with non-associative scalars II. *Math. Ann.* **164**(4), 291–316 (1966)
- Hansen, L.P., Richard, S.F.: The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica* **55**(3), 587–613 (1987)
- Horn, R.A., Johnson, C.R.: *Matrix Analysis*. Cambridge University Press, Cambridge (1990)
- Kaplansky, I.: Modules over operator algebras. *Am. J. Math.* **75**(4), 839–858 (1953)
- Lütkepohl, H.: *New Introduction to Multiple Time Series Analysis*. Springer, Berlin (2005)
- Meyer, C.D.: *Matrix Analysis and Applied Linear Algebra*. SIAM, Philadelphia (2000)
- Ortu, F., Severino, F., Tamoni, A., Tebaldi, C.: A persistence-based Wold-type decomposition for stationary time series. *Quant. Econ.* **11**(1), 203–230 (2020a)
- Ortu, F., Severino, F., Tamoni, A., Tebaldi, C.: Supplement to A persistence-based Wold-type decomposition for stationary time series. *Quant. Econ.* (2020b). <https://doi.org/10.3982/QE994>
- Raeburn, I., Williams, D.P.: *Morita Equivalence and Continuous-Trace C^* -Algebras*. American Mathematical Society, Providence (1998)
- Rozanov, Y.A.: *Stationary Random Processes*. Holden-Day, San Francisco (1967)
- Severino, F.: Isometric operators on Hilbert spaces and Wold decomposition of stationary time series. *Decis. Econ. Finance* **39**(2), 203–234 (2016)
- Sz-Nagy, B., Foias, C., Bercovici, H., Ké rchy, L.: *Harmonic Analysis of Operators on Hilbert Spaces*. Springer, New York (2010)
- Trefethen, L.N., Bau, D., III.: *Numerical Linear Algebra*. SIAM, Philadelphia (1997)
- Wiener, N., Masani, P.: The prediction theory of multivariate stochastic processes. *Acta Math.* **98**(1–4), 111–150 (1957)
- Wold, H.: *A Study in the Analysis of Stationary Time Series*. Almqvist & Wiksells Boktryckeri, Uppsala (1938)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.