Isometric operators on Hilbert spaces and Wold Decomposition of stationary time series

Federico Severino[1]

September, 2016
This is a pre-print of an article published in *Decisions in Economics and Finance*.
The final authenticated version is available online at:
https://doi.org/10.1007/s10203-016-0181-5.

Abstract
Wold Theorem plays a fundamental role in the decomposition of weakly stationary time series. It provides a moving average representation of the process under consideration in terms of uncorrelated innovations, whatever the nature of the process is. From an empirical point of view, this result enables to identify orthogonal shocks, for instance in macroeconomic and financial time series. More theoretically, the decomposition of weakly stationary stochastic processes can be seen as a special case of the Abstract Wold Theorem, that allows to decompose Hilbert spaces by using isometric operators. In this work we explain this link in detail, employing the Hilbert space spanned by a weakly stationary time series and the lag operator as isometry. In particular, we characterize the innovation subspace by exploiting the adjoint operator. We also show that the isometry of the lag operator is equivalent to weak stationarity. Our methodology, fully based on operator theory, provides novel tools useful to discover new Wold-type decompositions of stochastic processes, in which the involved isometry is no more the lag operator. In such decompositions the orthogonality of innovations is ensured by construction since they are derived from the Abstract Wold Theorem.

Keywords: isometry, Hilbert space, Wold decomposition, stationary time series.

[1]Federico Severino is at the Department of Finance, Bocconi University, Milan, 20136, Italy.
Correspondence Information: Federico Severino federico.severino@unibocconi.it.
We wish to thank Fulvio Ortu, Andrea Tamoni and Claudio Tebaldi for fruitful discussions on the topic.

1 Introduction

Wold Decomposition Theorem, stated and proved for the first time by Wold (1938), is a milestone in theoretical econometrics. Generalizing the idea that autoregressive processes may have a moving average representation, this theorem shows how to decompose any weakly stationary time series into a *non-deterministic* component, driven by linear uncorrelated innovations, and a *deterministic* term, which is included in the so-called remote sigma-algebra of events. The non-deterministic part consists of a moving average which is extremely relevant for economic and financial time series. Indeed, the uncorrelation of innovations is useful to identify unexpected shocks which occur at calendar time. This way of reasoning underlies non-structural econometric modelling. In the macroeconomic literature, for example, Leeper, Sims, Zha, Hall, and Bernanke (1996) widely employed moving averages in order to evaluate the impact of independent monetary policy shocks to the economy.

The linearity of the non-deterministic term makes the most common econometric tools, e.g. ordinary least squares, suitable for the estimation of the so-called *impulse response functions*. Such coefficients capture the sensitivity of a time series with respect to a shock occurred several periods before. Or, from the opposite perspective, they quantify the future impact of today disturbances. For instance, through impulse response functions, Leeper et al. (1996) were able to relate money stock surprises to U.S. prices and output changes.

Wold decomposition has an abstract counterpart, i.e. the Abstract Wold Theorem, which is well-known in functional analysis and, in particular, in operator theory.[2] The Abstract Wold Theorem provides a way to decompose a Hilbert space into the direct sum of orthogonal subspaces, induced by the powers of an isometric operator. In this work we show that the Classical Wold Decomposition for weakly stationary time series, as stated, for instance, in Brockwell and Davis (2009) and in Bierens (2012), is a special case of the Abstract Wold Decomposition that involves isometries on Hilbert spaces. Briefly, the involved Hilbert space will be the span of the past realizations of a weakly stationary time series and the isometry will be the lag operator.

The applicability of our novel approach is not limited to the Wold decomposition showed in this paper, but it involves a broader class of Wold-type decompositions. Indeed, to deduce the Classical Wold Decomposition for stochastic processes from the abstract one, we develop

[2]See, for example, Nagy, Foias, Bercovici, and Kérchy (2010) as a reference about operators on Hilbert spaces.

a methodology that can be employed in more general contexts. Specifically, we exploit the properties of the adjoint operator, especially the features of its kernel, in order to find the so-called *innovation subspace*, or *wandering subspace*, which is the building block of the decomposition. Our procedure represents a simple path that can be followed to discover new Wold decompositions for weakly stationary time series, involving isometries different from the lag operator. For instance, a possibly isometric operator may be inspired by the discrete Haar transform that has been applied to economic time series by Ortu, Tamoni, and Tebaldi (2013) in order to capture heterogeneous layers of persistence.

The paper is organized as follows. Section 2 introduces weakly stationary time series and the Classical Wold Decomposition Theorem. Section 3 reviews operator theory in Hilbert spaces and states the Abstract Wold Theorem. Section 4 illustrates the main contribution of the paper, i.e. the derivation of the Classical Wold Decomposition for time series from the Abstract Wold Theorem. In particular, Proposition 12 establishes the equivalence between weak stationarity and isometry of the lag operator, promoting the isometry property to fundamental notion for our construction. Moreover, Proposition 19 provides a characterization of purely deterministic time series in terms of non-regularity. Finally, Section 5 outlines a possible generalization of the theory by exploiting different isometric operators. Appendix contains the proofs of lemmas.

2 Stationarity and Classical Wold Theorem

Let $\mathbf{x} = \{x_t\}_{t\in\mathbb{Z}}$ be a stochastic process in which any variable x_t is measurable with respect to a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We assume that the second moment of x_t is finite for any t, i.e. each x_t belongs to the vector space $L^2(\Omega, \mathcal{F}, \mathbb{P})$, in which two random variables are identified when they coincide almost surely.

Definition 1 *The process* $\mathbf{x} = \{x_t\}_{t\in\mathbb{Z}}$ *is* weakly stationary *when, for any* t*,*

i) $\mathbb{E}\left[x_t^2\right]$ *is finite,*

ii) $\mathbb{E}\left[x_t\right]$ *is independent of* t*,*

iii) *the cross moment* $\mathbb{E}\left[x_{t-h}x_{t-k}\right]$ *depends at most on the difference* $h-k$*, for any* $h, k \in \mathbb{Z}$*.*

When $\mathbf{x}$ is weakly stationary, the *autocovariance function* $\gamma : \mathbb{Z} \longrightarrow \mathbb{R}$ is well-defined. Indeed, by denoting μ the first moment of x_t, we have that, for any integer j,

$$\gamma(j) = \mathrm{Cov}\left(x_t, x_{t-j}\right) = \mathbb{E}\left[\left(x_t - \mu\right)\left(x_{t-j} - \mu\right)\right].$$

We will assume that the weakly stationary stochastic process $\mathbf{x}$ has zero mean, that is $\mu = 0$.[3] Relevant properties of autocovariance functions can be retrieved, for example, in Brockwell and Davis (2009). Here, we just recall that a weakly stationary process $\mathbf{x}$ is a *white noise* if $\mu = 0$ and $\gamma(j) = 0$ for all $j \neq 0$.

A key notion about time series is the so-called *regularity*, as defined, for instance, in Bierens (2012). Regularity concerns the orthogonal projection of the variable x_t on the closed subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$ spanned by the sequence $\{x_{t-k}\}_{k \in \mathbb{N}}$. See Section 4.1 for details about this subspace, that we denote by $\mathcal{H}_{t-1}(\mathbf{x})$.

Definition 2 *Given a weakly stationary stochastic process* $\mathbf{x}$*, let* $\mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t$ *denote the orthogonal projection of the random variable* x_t *on the vector space*

$$\mathcal{H}_{t-1}(\mathbf{x}) = \mathrm{cl}\left\{\sum_{k=1}^{+\infty} a_k x_{t-k} : \quad \sum_{k=1}^{+\infty}\sum_{h=1}^{+\infty} a_k a_h \gamma(k-h) < +\infty\right\}.$$

We say that $\mathbf{x}$ *is* regular *when*

$$\left\| x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t \right\| > 0 \qquad \forall t \in \mathbb{Z},$$

where $\|\cdot\|$ *is the* L^2*-norm.*

The regularity condition implies linear independence among variables x_t and allows us to identify non-trivial Wold decompositions (see Proposition 19 in Section 4.2).

We now state the Classical Wold Decomposition Theorem for zero-mean, regular, weakly stationary time series, that we will prove in the following sections.

Theorem 3 (Classical Wold Decomposition) *Let* $\mathbf{x} = \{x_t\}_{t \in \mathbb{Z}}$ *be a zero-mean, regular, weakly stationary stochastic process. Then, for any* $t \in \mathbb{Z}$*,* x_t *decomposes as*

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

where the equality is in the L^2*-norm and*

[3] Such assumption entails no loss of generality. Indeed, in case μ is different from zero, for our purposes it is enough to replace x_t with $x_t - \mu$.

i) $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t\in\mathbb{Z}}$ *is a white noise process with variance equal to* 1*;*

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

$$\alpha_k = \mathbb{E}\left[x_t \varepsilon_{t-k}\right] \qquad \textit{and} \qquad \sum_{k=0}^{+\infty} \alpha_k^2 < +\infty;$$

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

$$\nu_t \in \bigcap_{j=0}^{+\infty} \mathcal{H}_{t-j}(\mathbf{x}) \qquad \textit{and} \qquad \mathbb{E}\left[\nu_t \varepsilon_{t-k}\right] = 0 \quad \forall k \in \mathbb{N}_0;$$

iv)

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

The last property points out that the process $\boldsymbol{\nu}$ is predictable. Indeed, $\boldsymbol{\nu}$ is usually indicated as the *deterministic component* of the time series $\mathbf{x}$, while the term $\sum_{k=0}^{\infty} \alpha_k \varepsilon_{t-k}$ constitutes the *non-deterministic component*. We say that a weakly stationary time series is *purely non-deterministic* when the deterministic component is equal to zero. Similarly, we call *purely deterministic* those weakly stationary processes with null non-deterministic component.

Finally, we recall that the *impulse response functions* α_k are the least squares coefficients concerning the orthogonal projection of x_t on the linear subspace generated by the innovation ε_{t-k}. Note, indeed, that the variance of each ε_{t-k} is equal to 1.

3 Operators and Abstract Wold Theorem

In this section we recall some basic notions about operators on Hilbert spaces. Our main reference is Nagy, Foias, Bercovici, and Kérchy (2010). The Hilbert spaces framework is rather flexible and it allows us to state an abstract version of the Wold decomposition. The main notions of Hilbert space theory that we employ are isometry and orthogonal projection on a closed subspace.

Abstract Hilbert spaces are suitable when dealing with stochastic processes because they constitute the generalization of the space $L^2(\Omega, \mathcal{F}, \mathbb{P})$ to which the realizations x_t of a weakly stationary process $\mathbf{x}$ belong.

Let $\mathcal{H}$ be a Hilbert space and consider an operator $\mathbf{V} : \mathcal{H} \longrightarrow \mathcal{H}$. We will deal with bounded linear operators and we will occasionally compute their *adjoint operator*.

Definition 4 *Let* $\mathbf{V} : \mathcal{H} \longrightarrow \mathcal{H}$ *be a bounded linear operator. The* adjoint*, or* transposed*, of* $\mathbf{V}$ *is the bounded linear operator* $\mathbf{V}^* : \mathcal{H} \longrightarrow \mathcal{H}$ *that satisfies the relation*

$$\langle \mathbf{V}x, y \rangle = \langle x, \mathbf{V}^* y \rangle \qquad \forall x, y \in \mathcal{H}.$$

In particular, $\|\mathbf{V}\| = \|\mathbf{V}^*\|$*.*

In the following lemma we state a fundamental property of the kernel of the adjoint of a bounded linear operator, that is the orthogonal complement of the image $\mathbf{V}\mathcal{H}$ coincides with the kernel of $\mathbf{V}^*$.

Lemma 1 *Let* $\mathbf{V} : \mathcal{H} \longrightarrow \mathcal{H}$ *be a bounded linear operator and* $\mathbf{V}^*$ *its adjoint. Then,*

$$\ker(\mathbf{V}^*) = \mathcal{H} \ominus \mathbf{V}\mathcal{H}.$$

Proof. See Luenberger (1968). ■

As a result, we can compute the kernel of $\mathbf{V}^*$ whenever we need to characterize the elements that are orthogonal to any of the elements in the image of $\mathcal{H}$ through $\mathbf{V}$.

The crucial feature of the linear operator $\mathbf{V}$ that induces a Wold decomposition is the *isometry*. According to this property, inner products are preserved from the Hilbert space $\mathcal{H}$ to its image $\mathbf{V}\mathcal{H}$. Moreover, the norm of isometric operators is equal to 1.

Definition 5 *Let* $\mathbf{V} : \mathcal{H} \longrightarrow \mathcal{H}$ *be a linear operator.* $\mathbf{V}$ *is an* isometry *when*

$$\langle \mathbf{V}x, \mathbf{V}y \rangle = \langle x, y \rangle \qquad \forall x, y \in \mathcal{H},$$

or, equivalently,

$$\mathbf{V}^*\mathbf{V} = I,$$

where I *denotes the* identity map *on* $\mathcal{H}$*.*

Clearly, if $\mathbf{V}$ is an isometric operator, also its powers $\mathbf{V}^j$, with $j \in \mathbb{N}$, are isometries. In addition, the following lemma will be helpful when dealing with the image of $\mathcal{H}$ through $\mathbf{V}$.

Lemma 2 *Let* $\mathcal{I}$ *be a closed subspace of a Hilbert space* $\mathcal{H}$ *and* $\mathbf{V} : \mathcal{I} \longrightarrow \mathcal{I}$ *an isometry. Then, the image of* $\mathcal{I}$ *through* $\mathbf{V}$ *is closed in* $\mathcal{I}$*.*

Proof. See Appendix. ■

Note that isometric operators are not required to be onto, hence $\mathbf{V}\mathcal{H}$ may be a proper subspace of $\mathcal{H}$. The following definition addresses this point.

Definition 6 *Let* $\mathbf{V} : \mathcal{H} \longrightarrow \mathcal{H}$ *be a linear operator.* $\mathbf{V}$ *is* unitary *when it is isometric and onto, that is*

$$\mathbf{V}^*\mathbf{V} = I, \qquad \mathbf{V}\mathcal{H} = \mathcal{H}.$$

The last relations are equivalent to $\mathbf{V}^* = \mathbf{V}^{-1}$*.*

Isometries allow us to orthogonally decompose Hilbert spaces. The decomposition is based on the so-called *wandering subspace*, which is also indicated as *innovation subspace* or *detail subspace*.

Definition 7 *Let* $\mathbf{V} : \mathcal{H} \longrightarrow \mathcal{H}$ *be an isometry. We call* wandering subspace for $\mathbf{V}$ *a subspace* $\mathcal{L}^{\mathbf{V}}$ *of* $\mathcal{H}$ *such that* $\mathbf{V}^h\mathcal{L}^{\mathbf{V}}$ *and* $\mathbf{V}^k\mathcal{L}^{\mathbf{V}}$ *are orthogonal for every* $h, k \in \mathbb{N}_0$*, with* $h \neq k$*. Since* $\mathbf{V}$ *is an isometry, this definition is equivalent to require that* $\mathbf{V}^n\mathcal{L}^{\mathbf{V}}$ *is orthogonal to* $\mathcal{L}^{\mathbf{V}}$ *for every* $n \in \mathbb{N}$*.*

Given a wandering subspace $\mathcal{L}^{\mathbf{V}}$, it is possible to define the orthogonal sum

$$\bigoplus_{j=0}^{+\infty} \mathbf{V}^j\mathcal{L}^{\mathbf{V}},$$

in which the convergence of the infinite direct sum is in the norm induced in the Hilbert space $\mathcal{H}$ by the inner product. It may happen that such an orthogonal sum covers the whole Hilbert space $\mathcal{H}$. This leads to the definition of *unilateral shift.*

Definition 8 *Let* $\mathbf{V} : \mathcal{H} \longrightarrow \mathcal{H}$ *be an isometry. We say that* $\mathbf{V}$ *is a* unilateral shift *when there exists a subspace* $\mathcal{L}^{\mathbf{V}}$ *of* $\mathcal{H}$ *which is wandering for* $\mathbf{V}$ *and such that*

$$\bigoplus_{j=0}^{+\infty} \mathbf{V}^j\mathcal{L}^{\mathbf{V}} = \mathcal{H}.$$

Given a unilateral shift, the innovation subspace is uniquely determined, as we claim in the following lemma.

Lemma 3 *Let* $\mathbf{V} : \mathcal{H} \longrightarrow \mathcal{H}$ *be an isometry. If* $\mathbf{V}$ *is a unilateral shift, then the wandering subspace* $\mathcal{L}^{\mathbf{V}}$ *is uniquely determined by* $\mathbf{V}$ *and it satisfies the relation*

$$\mathcal{L}^{\mathbf{V}} = \mathcal{H} \ominus \mathbf{V}\mathcal{H}.$$

Proof. See Appendix. ■

We refer to Nagy et al. (2010) for further details, as well as for the proof of the Wold decomposition of Hilbert spaces. Such decomposition is based solely on isometric operators defined on Hilbert spaces. Therefore, we name this result Abstract Wold Decomposition Theorem.

Theorem 9 (Abstract Wold Decomposition) *Let* $\mathcal{H}$ *be a Hilbert space and* $\mathbf{V} : \mathcal{H} \longrightarrow \mathcal{H}$ *be an isometry. Then,* $\mathcal{H}$ *decomposes uniquely into an orthogonal sum*

$$\mathcal{H} = \hat{\mathcal{H}} \oplus \tilde{\mathcal{H}},$$

such that

$$\mathbf{V}\hat{\mathcal{H}} = \hat{\mathcal{H}}, \qquad \mathbf{V}\tilde{\mathcal{H}} \subset \tilde{\mathcal{H}},$$

the restriction of $\mathbf{V}$ *on* $\hat{\mathcal{H}}$ *is unitary and the restriction of* $\mathbf{V}$ *on* $\tilde{\mathcal{H}}$ *is a unilateral shift. In particular,*

$$\hat{\mathcal{H}} = \bigcap_{j=0}^{+\infty} \mathbf{V}^j \mathcal{H}, \qquad \tilde{\mathcal{H}} = \bigoplus_{j=0}^{+\infty} \mathbf{V}^j \mathcal{L}^{\mathbf{V}},$$

with $\mathcal{L}^{\mathbf{V}} = \mathcal{H} \ominus \mathbf{V}\mathcal{H}$*.*

Proof. See Nagy et al. (2010). ■

Observe that one of the subspaces $\hat{\mathcal{H}}$ and $\tilde{\mathcal{H}}$ may be null (not both, unless $\mathcal{H}$ is trivial). Note also that $\hat{\mathcal{H}}$ and $\tilde{\mathcal{H}}$ play, respectively, the role of the deterministic and the non-deterministic components that come up in the Classical Wold Decomposition for time series. Finally, remember that the convergence of the infinite direct sum in $\tilde{\mathcal{H}}$ is in the norm of the Hilbert space.

The aim of the next section is to explain in detail how the Classical Wold Decomposition for time series may be derived by applying Theorem 9 to a particular Hilbert space on which the lag operator is isometric.

4 The Classical Wold Decomposition from the abstract one

In this chapter we build, step by step, all the constituents of the theoretical framework that makes the Classical Wold Decomposition arise as a special case of the Wold Theorem for isometric operators. We begin with the definition of a Hilbert space, $\mathcal{H}_t(\mathbf{x})$, and we show that the *lag operator* is isometric on it. This is what we need to apply Theorem 9. Afterwards, we devote the second subsection to the characterization of the innovation subspace, by taking advantage of the properties of the adjoint of the lag operator. Finally, we retrieve the Classical Wold Decomposition for time series and we prove the usual properties of such representation by exploiting the isometry of the lag operator.

4.1 The Hilbert space $\mathcal{H}_t(\mathbf{x})$ and the lag operator as isometry

We consider the vector space $L^2(\Omega, \mathcal{F}, \mathbb{P})$, which is a Hilbert space when it is equipped with the inner product:

$$\langle A, B \rangle = \mathbb{E}\left[AB\right] \qquad \forall A, B \in L^2(\Omega, \mathcal{F}, \mathbb{P}).$$

We introduce a zero-mean weakly stationary stochastic process $\mathbf{x}$ such that each random variable x_t is measurable. In particular, we are interested in the closed subspace $\mathcal{H}_t(\mathbf{x})$ of $L^2(\Omega, \mathcal{F}, \mathbb{P})$ spanned by the sequence $\{x_{t-k}\}_{k \in \mathbb{N}_0}$. This is the space of random variables $y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ such that there exists a sequence of real numbers $\{a_k\}_{k \in \mathbb{N}_0}$ so that

$$\left\| y - \sum_{k=0}^{n} a_k x_{t-k} \right\| \longrightarrow 0, \qquad \text{as} \qquad n \longrightarrow +\infty.$$

In an equivalent way, we can claim that the subspace $\mathcal{H}_t(\mathbf{x})$ is defined by

$$\mathcal{H}_t(\mathbf{x}) = \mathrm{cl}\left\{ \sum_{k=0}^{+\infty} a_k x_{t-k} : \quad \sum_{k=0}^{+\infty} \sum_{h=0}^{+\infty} a_k a_h \gamma(k-h) < +\infty \right\}, \tag{1}$$

where *cl* denotes the closure in the L^2-norm and $\gamma : \mathbb{Z} \longrightarrow \mathbb{R}$ is the autocovariance function.

$\mathcal{H}_t(\mathbf{x})$ is a Hilbert subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$; see Bierens (2012) for the proof. We recall that the inner product of $L^2(\Omega, \mathcal{F}, \mathbb{P})$ acts on square-summable linear combinations, for instance

$$A = \sum_{k=0}^{+\infty} a_k x_{t-k}, \qquad B = \sum_{h=0}^{+\infty} b_h x_{t-h},$$

as

$$\langle A, B\rangle = \mathbb{E}\left[\left(\sum_{k=0}^{+\infty} a_k x_{t-k}\right)\left(\sum_{h=0}^{+\infty} b_h x_{t-h}\right)\right] = \sum_{k=0}^{+\infty}\sum_{h=0}^{+\infty} a_k b_h \gamma(k-h).$$

In addition, one can show that the subspace $\mathcal{H}_t(\mathbf{x})$ may be built from arbitrary finite linear combinations of variables x_{t-k}, according to the relation

$$\mathcal{H}_t(\mathbf{x}) = \text{cl}\left\{\bigcup_{n=0}^{+\infty} \text{span}\,(x_t, x_{t-1}, \ldots, x_{t-n})\right\}.$$

Nevertheless, when dealing with $\mathcal{H}_t(\mathbf{x})$, we will refer to the characterization (1) and, in particular, we will call *generators* of $\mathcal{H}_t(\mathbf{x})$ the random variables $\sum_{k=0}^{\infty} a_k x_{t-k}$ that satisfy the square-summability requirement

$$\left\|\sum_{k=0}^{+\infty} a_k x_{t-k}\right\|^2 = \mathbb{E}\left[\left(\sum_{k=0}^{+\infty} a_k x_{t-k}\right)^2\right] = \sum_{k=0}^{+\infty}\sum_{h=0}^{+\infty} a_k a_h \gamma(k-h) < +\infty.$$

Now we provide the definition of *lag operator* on the space $\mathcal{H}_t(\mathbf{x})$.

Definition 10 *We call* lag operator *the operator* $\mathbf{L} : \mathcal{H}_t(\mathbf{x}) \longrightarrow \mathcal{H}_t(\mathbf{x})$ *that acts on generators of* $\mathcal{H}_t(\mathbf{x})$ *as*

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

The first problem we face is the definition of $\mathbf{L}$ on the whole closed space $\mathcal{H}_t(\mathbf{x})$. The following proposition helps us in extending $\mathbf{L}$ continuously by means of an operator that we still call $\mathbf{L}$, with a little abuse, which has the same norm of the original $\mathbf{L}$.

Proposition 11 *The operator* $\mathbf{L}$ *is well-defined and it is linear and bounded on the span of generators of* $\mathcal{H}_t(\mathbf{x})$*. Hence, it can be extended to* $\mathcal{H}_t(\mathbf{x})$ *with continuity.*

Proof. In order to show that $\mathbf{L}$ is well-defined, take any generator $X = \sum_{k=0}^{\infty} a_k x_{t-k}$ in $\mathcal{H}_t(\mathbf{x})$, meaning that

$$\|X\|^2 = \sum_{k=0}^{+\infty}\sum_{h=0}^{+\infty} a_k a_h \gamma(k-h) < +\infty.$$

Recall that $\mathbf{L}X = \sum_{k=0}^{\infty} a_k x_{t-1-k}$ so that, by the weak stationarity of $\mathbf{x}$,

$$\|\mathbf{L}X\|^2 = \sum_{k=0}^{+\infty}\sum_{h=0}^{+\infty} a_k a_h \gamma\left((k+1)-(h+1)\right) = \sum_{k=0}^{+\infty}\sum_{h=0}^{+\infty} a_k a_h \gamma(k-h) = \|X\|^2,$$

which is finite. As a result, $\mathbf{L}X$ belongs to $\mathcal{H}_t(\mathbf{x})$ too, so $\mathbf{L}$ is well-defined.

Linearity of $\mathbf{L}$ is immediate to prove.

As for boundedness, we already showed that $\|\mathbf{L}X\| = \|X\|$ for any generator $X \in \mathcal{H}_t(\mathbf{x})$. This implies that $\|\mathbf{L}\| = 1$. Therefore, $\mathbf{L}$ is a bounded operator and so it can be extended to the closed space $\mathcal{H}_t(\mathbf{x})$ with continuity. ■

We stress the fact that the weak stationarity of $\mathbf{x}$ is the key-property to ensure that the lag operator is an isometry on the Hilbert space $\mathcal{H}_t(\mathbf{x})$. Indeed, the following proposition asserts the equivalence between the isometry of $\mathbf{L}$ and the weak stationarity of $\mathbf{x}$.

Proposition 12 *Let $\mathbf{x}$ be a zero-mean stochastic process such that $x_t \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ for all $t \in \mathbb{Z}$. The operator $\mathbf{L}$ is an isometry on $\mathcal{H}_t(\mathbf{x})$ for any $t \in \mathbb{Z}$ if and only if $\mathbf{x}$ is weakly stationary.*

Proof. Assume that $\mathbf{x}$ is a weakly stationary time series. We just prove the isometry property for generators of $\mathcal{H}_t(\mathbf{x})$. Indeed, by continuity of the extension of $\mathbf{L}$ on the closure, it follows that the property is satisfied on the whole closed space $\mathcal{H}_t(\mathbf{x})$. Take, then, any $X = \sum_{k=0}^{\infty} a_k x_{t-k}$, $Y = \sum_{h=0}^{\infty} b_h x_{t-h}$ in $\mathcal{H}_t(\mathbf{x})$. By the weak stationarity of $\mathbf{x}$, we have

$$\langle \mathbf{L}X, \mathbf{L}Y \rangle = \sum_{k=0}^{+\infty}\sum_{h=0}^{+\infty} a_k b_h \gamma\left((k+1)-(h+1)\right) = \sum_{k=0}^{+\infty}\sum_{h=0}^{+\infty} a_k b_h \gamma(k-h) = \langle X, Y \rangle.$$

As a result, $\mathbf{L}$ is an isometry on $\mathcal{H}_t(\mathbf{x})$.

Conversely, suppose that $\mathbf{L}$ is an isometric operator on $\mathcal{H}_t(\mathbf{x})$. Hence, for any $X = \sum_{k=0}^{\infty} a_k x_{t-k}$ in $\mathcal{H}_t(\mathbf{x})$, it holds $\langle X, X \rangle = \langle \mathbf{L}X, \mathbf{L}X \rangle$, namely

$$\sum_{k=0}^{+\infty}\sum_{h=0}^{+\infty} a_k a_h \mathbb{E}\left[x_{t-k} x_{t-h}\right] = \sum_{k=0}^{+\infty}\sum_{h=0}^{+\infty} a_k a_h \mathbb{E}\left[x_{t-1-k} x_{t-1-h}\right].$$

By choosing $X = a_n x_{t-n}$, with $a_n \neq 0$, $n \in \mathbb{N}_0$, the equality becomes

$$a_n^2 \mathbb{E}\left[x_{t-n} x_{t-n}\right] = a_n^2 \mathbb{E}\left[x_{t-1-n} x_{t-1-n}\right].$$

As a result, $\mathbb{E}\left[x_{t-n}^2\right] = \mathbb{E}\left[x_{t-1-n}^2\right]$ for any $n \in \mathbb{N}_0$.

After that, choose the element $X = a_p x_{t-p} + a_q x_{t-q}$, where $a_p, a_q \neq 0$ and $p \neq q$, with $p, q \in \mathbb{N}_0$. Then,

$$\langle X, X\rangle = a_p^2 \mathbb{E}\left[x_{t-p}^2\right] + a_q^2 \mathbb{E}\left[x_{t-q}^2\right] + 2a_p a_q \mathbb{E}\left[x_{t-p}x_{t-q}\right] = \left(a_p^2 + a_q^2\right)\mathbb{E}\left[x_t^2\right] + 2a_p a_q \mathbb{E}\left[x_{t-p}x_{t-q}\right].$$

and, similarly,

$$\begin{aligned}\langle \mathbf{L}X, \mathbf{L}X\rangle &= a_p^2 \mathbb{E}\left[x_{t-1-p}^2\right] + a_q^2 \mathbb{E}\left[x_{t-1-q}^2\right] + 2a_p a_q \mathbb{E}\left[x_{t-1-p}x_{t-1-q}\right] \\ &= \left(a_p^2 + a_q^2\right)\mathbb{E}\left[x_t^2\right] + 2a_p a_q \mathbb{E}\left[x_{t-1-p}x_{t-1-q}\right].\end{aligned}$$

Since $\langle X, X\rangle = \langle \mathbf{L}X, \mathbf{L}X\rangle$, we find that

$$\mathbb{E}\left[x_{t-p}x_{t-q}\right] = \mathbb{E}\left[x_{t-1-p}x_{t-1-q}\right].$$

As a consequence, $\mathbb{E}\left[x_{t-p}x_{t-q}\right]$ depends at most on $p - q$. Since the process $\mathbf{x}$ is also zero-mean and $x_t \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ for all $t \in \mathbb{Z}$, it follows that $\mathbf{x}$ is weakly stationary. ■

Once the isometry of the lag operator is ensured, we focus on the image of $\mathbf{L}$ and of its powers $\mathbf{L}^j$, with $j \in \mathbb{N}$. Indeed, these subspaces will be present in the Wold decomposition of $\mathcal{H}_t(\mathbf{x})$ when the operator $\mathbf{L}$ is employed.

Proposition 13 *For any $j \in \mathbb{N}$,*

$$\mathbf{L}^j \mathcal{H}_t(\mathbf{x}) = \mathcal{H}_{t-j}(\mathbf{x}).$$

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

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

Conversely, consider any generator $Y = \sum_{k=1}^{\infty} a_k x_{t-k}$ in $\mathcal{H}_{t-1}(\mathbf{x})$. It is easy to see that Y is the image of the element $X = \sum_{k=0}^{\infty} a_{k+1} x_{t-k}$ belonging to $\mathcal{H}_t(\mathbf{x})$ because

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

As a result, generators of $\mathcal{H}_{t-1}(\mathbf{x})$ are included in $\mathbf{L}\mathcal{H}_t(\mathbf{x})$. By Lemma 2, $\mathbf{L}\mathcal{H}_t(\mathbf{x})$ is a closed subspace of $\mathcal{H}_t(\mathbf{x})$ and so, by taking the closures we get that the whole $\mathcal{H}_{t-1}(\mathbf{x})$ is contained in $\mathbf{L}\mathcal{H}_t(\mathbf{x})$.

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

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

it is easy to show that $\mathbf{L}^j\mathcal{H}_t(\mathbf{x}) = \mathcal{H}_{t-j}(\mathbf{x})$ for any $j \in \mathbb{N}$. Indeed, $\mathbf{L}^j$ is an isometry too and the proof we described above still works. ■

As the lag operator turned out to be an isometry on $\mathcal{H}_t(\mathbf{x})$, we can apply the Abstract Wold Theorem to the isometry $\mathbf{L}$ on the Hilbert space $\mathcal{H}_t(\mathbf{x})$. However, in order to complete the decomposition, we need to describe explicitly the innovation subspace $\mathcal{L}_t^{\mathbf{L}}$.

4.2 The wandering subspace $\mathcal{L}_t^{\mathbf{L}}$

In this subsection we determine $\mathcal{L}_t^{\mathbf{L}}$, that is the wandering subspace associated with the lag operator, in two ways. The first one employs Gram-Schmidt's orthonormalization procedure, it is rather straightforward and it is particularly suitable for the lag operator on $\mathcal{H}_t(\mathbf{x})$. Differently, the second method exploits Lemma 1. Indeed, $\mathcal{L}_t^{\mathbf{L}}$ is the orthogonal complement of the subspace $\mathbf{L}\mathcal{H}_t(\mathbf{x})$ and, accordingly, it coincides with the kernel of the adjoint of $\mathbf{L}$. Hence, we will focus on the determination of $\mathbf{L}^*$. The latter approach has wider applicability than the first one because it may be used when isometries different from the lag operator are involved.

A convenient way to proceed is to find a countable complete orthonormal system of generators for $\mathcal{H}_t(\mathbf{x})$. Indeed, such a system enables us to write any element of $\mathcal{H}_t(\mathbf{x})$ as a convergent series of orthogonal variables. See, for instance, Luenberger (1968) or Rudin (1987) as references. Thus, let us consider the set

$$\mathcal{X}_t = \{x_{t-k} : \quad k \in \mathbb{N}_0\}.$$

$\mathcal{X}_t$ is a countable complete linearly independent system of generators for $\mathcal{H}_t(\mathbf{x})$; note that linearly independence is guaranteed by the regularity of the process $\mathbf{x}$. We can exploit Gram-Schmidt's procedure in order to build a countable complete system $\mathcal{E}_t$ of generators for $\mathcal{H}_t(\mathbf{x})$ that is also orthonormal:

$$\mathcal{E}_t = \{e_{t-k} : \quad k \in \mathbb{N}_0\},$$

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

$$\langle e_{t-k}, e_{t-h} \rangle = \begin{cases} 1 & \text{if } k = h, \\ 0 & \text{if } k \neq h. \end{cases}$$

As a consequence, we can write any element X of $\mathcal{H}_t(\mathbf{x})$ as a convergent series in norm:

$$X = \sum_{k=0}^{+\infty} c_k e_{t-k} \qquad \text{with} \qquad \sum_{k=0}^{+\infty} c_k^2 < +\infty.$$

Therefore, we can rewrite the Hilbert space $\mathcal{H}_t(\mathbf{x})$ as

$$\begin{aligned} \mathcal{H}_t(\mathbf{x}) &= \mathrm{cl}\left\{ \sum_{k=0}^{+\infty} a_k x_{t-k} : \quad \sum_{k=0}^{+\infty}\sum_{h=0}^{+\infty} a_k a_h \gamma(k-h) < +\infty \right\} \\ &= \left\{ \sum_{k=0}^{+\infty} c_k e_{t-k} : \quad \sum_{k=0}^{+\infty} c_k^2 < +\infty \right\}. \end{aligned}$$

The closure is not needed when we use the complete orthonormal system $\mathcal{E}_t$.

We now focus on the details of Gram-Schmidt's construction. The method proceeds inductively. In particular, consider $\mathcal{E}_{t-1}$ as a countable complete orthonormal system of generators for $\mathcal{H}_{t-1}(\mathbf{x})$. Then, the space $\mathcal{H}_t(\mathbf{x})$ is generated by the system $\{x_t, \mathcal{E}_{t-1}\}$. Following Gram-Schmidt's procedure, we discover that a complete orthonormal system for $\mathcal{H}_t(\mathbf{x})$ is given by $\{e_t, \mathcal{E}_{t-1}\}$, where the variable e_t is built as

$$e_t = \frac{x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t}{\left\| x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t \right\|}.$$

We recap this observation in the following proposition.

Proposition 14 *Let $\mathcal{E}_{t-1}$ be a countable complete orthonormal system of generators for $\mathcal{H}_{t-1}(\mathbf{x})$. Then, a countable complete orthonormal system of generators for $\mathcal{H}_t(\mathbf{x})$ is given by $\{e_t, \mathcal{E}_{t-1}\}$, where*

$$e_t = \frac{x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t}{\left\| x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t \right\|}.$$

Proof. Clearly, a countable complete system of generators for $\mathcal{H}_t(\mathbf{x})$ is given by $\{x_t, \mathcal{E}_{t-1}\}$. Let y be an element of $\mathcal{H}_t(\mathbf{x})$. Then, there exist a real number α and a variable $w \in \mathcal{H}_{t-1}(\mathbf{x})$ such that

$$y = \alpha x_t + w,$$

where the equality is in norm. Therefore,

$$y = \alpha \left\| x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t \right\| \frac{x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t}{\left\| x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t \right\|} + \alpha \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t + w = \beta e_t + z,$$

where $\beta = \alpha \left\| x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t \right\|$ is a real number and $z = \alpha \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t + w$ is a random variable belonging to $\mathcal{H}_{t-1}(\mathbf{x})$. Thus, z can be written in terms of the orthonormal system $\mathcal{E}_{t-1}$. Moreover, all the involved variables in the decomposition of y are orthonormal and so $\{e_t, \mathcal{E}_{t-1}\}$ makes up a countable complete orthonormal system of generators for $\mathcal{H}_t(\mathbf{x})$. ■

An immediate consequence of Proposition 14 is that the space $\mathcal{H}_t(\mathbf{x})$ can be decomposed into the direct sum

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

Equivalently, the linear space generated by the variable $x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t$ is the wandering subspace associated with the isometric operator $\mathbf{L}$ on $\mathcal{H}_t(\mathbf{x})$:

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

As a consequence, the Abstract Wold Decomposition, stated in Theorem 9, can be applied.

Now we turn to the alternative approach. Nevertheless, before concentrating on the adjoint of the lag operator, we investigate the way in which $\mathbf{L}$ acts on elements of $\mathcal{E}_t$. In particular, we study what happens when we apply the lag operator to the orthogonal projection of x_t on the subspace $\mathcal{H}_{t-1}(\mathbf{x})$. Unsurprisingly, the isometry of $\mathbf{L}$ ensures that

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

Roughly speaking, the lag operator and the projection map commute, thanks to the weak stationarity of the process $\mathbf{x}$. Consequently, as for the elements of the system $\mathcal{E}_t$, we find that

$$\mathbf{L} e_t = e_{t-1}.$$

A more general case is described in the following lemma.

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

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

and so any element e_{t-k} of the orthonormal system $\mathcal{E}_t$ satisfies the relation

$$\mathbf{L}^j e_{t-k} = e_{t-k-j}.$$

Proof. See Appendix. ■

The proof exploits the normal equations that define orthogonal projections on closed subspaces. The isometry of the lag operator is crucial in combining together such equations in order to derive the desired result. In addition, an immediate outcome of Lemma 4 is that

$$\left\| x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t \right\|$$

is not dependent on the time index $t \in \mathbb{Z}$.

At this point, we have all the instruments to determine explicitly the adjoint of the lag operator. In order to do so, we represent the elements of $\mathcal{H}_t(\mathbf{x})$ in the complete orthonormal system $\mathcal{E}_t$.

Proposition 15 *The adjoint of* $\mathbf{L}$ *is the operator* $\mathbf{L}^* : \mathcal{H}_t(\mathbf{x}) \longrightarrow \mathcal{H}_t(\mathbf{x})$ *that acts on elements of* $\mathcal{H}_t(\mathbf{x})$*, written with respect to the orthonormal system* $\mathcal{E}_t$*, as*

$$\mathbf{L}^* : \qquad \sum_{k=0}^{+\infty} b_k e_{t-k} \quad \longmapsto \quad \sum_{k=0}^{+\infty} b_{k+1} e_{t-k}.$$

Proof. First of all, note that $\mathbf{L}^*$ is well-defined. Indeed, the square-summability requirement is satisfied because the element $\sum_{k=0}^{\infty} b_k e_{t-k}$ belongs to $\mathcal{H}_t(\mathbf{x})$:

$$\sum_{k=0}^{+\infty} b_{k+1}^2 \leqslant \sum_{k=0}^{+\infty} b_k^2 < +\infty.$$

We are just left to establish the relation

$$\langle \mathbf{L}X, Y \rangle = \langle X, \mathbf{L}^* Y \rangle,$$

for any elements of $\mathcal{H}_t(\mathbf{x})$ as $X = \sum_{h=0}^{\infty} c_h e_{t-h}$, $Y = \sum_{k=0}^{\infty} b_k e_{t-k}$.

Observe that, by Lemma 4,

$$\mathbf{L}X = \mathbf{L}\left(\sum_{h=0}^{+\infty} c_h e_{t-h} \right) = \sum_{h=0}^{+\infty} c_h \mathbf{L} e_{t-h} = \sum_{h=0}^{+\infty} c_h e_{t-1-h}$$

and so, by exploiting orthonormality,

$$\langle \mathbf{L}X, Y \rangle = \sum_{h=0}^{+\infty} \sum_{k=0}^{+\infty} c_h b_k \langle e_{t-1-h}, e_{t-k} \rangle = \sum_{h=0}^{+\infty} c_h b_{h+1} = \sum_{h=0}^{+\infty} \sum_{k=0}^{+\infty} c_h b_{k+1} \langle e_{t-h}, e_{t-k} \rangle = \langle X, \mathbf{L}^* Y \rangle,$$

as we wanted to check. ■

We are now able to compute the kernel of the adjoint operator $\mathbf{L}^*$.

Proposition 16 *The kernel of the operator* $\mathbf{L}^*$ *is*

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

Proof. We write any element of $\mathcal{H}_t(\mathbf{x})$ as a convergent series of elements in the complete orthonormal system $\mathcal{E}_t$. The kernel of $\mathbf{L}^*$ is defined by

$$\begin{aligned}\ker(\mathbf{L}^*) &= \{X \in \mathcal{H}_t(\mathbf{x}) : \quad \mathbf{L}^* X = 0\} \\ &= \left\{\sum_{k=0}^{+\infty} b_k e_{t-k} \in \mathcal{H}_t(\mathbf{x}) : \quad \mathbf{L}^*\left(\sum_{k=0}^{+\infty} b_k e_{t-k}\right) = 0\right\},\end{aligned}$$

where the equality $\mathbf{L}^* X = 0$ is in norm. Recall, by Proposition 15, that

$$\mathbf{L}^* X = \mathbf{L}^*\left(\sum_{k=0}^{+\infty} b_k e_{t-k}\right) = \sum_{k=0}^{+\infty} b_{k+1} e_{t-k}.$$

By the orthonormality of elements e_{t-k}, the last quantity equals zero, in norm, if and only if

$$\sum_{k=0}^{+\infty} (b_{k+1})^2 = 0.$$

This happens if and only if $b_k = 0$ for all $k \geqslant 1$, that is, $X = b_0 e_t$. As a result, we find out that the kernel of $\mathbf{L}^*$ coincides with the linear space generated by e_t. Finally, Gram-Schmidt's construction immediately implies that

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

■

As Lemma 1 ensures, we retrieve that the wandering subspace $\mathcal{L}_t^{\mathbf{L}}$ is the linear space generated by the variable $x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t$.

Finally, in order to fill in the Wold decomposition of $\mathcal{H}_t(\mathbf{x})$, we need to analyse the image of $\mathcal{L}_t^{\mathbf{L}}$ through the powers of $\mathbf{L}$.

Corollary 17 *Let* $\mathcal{L}_t^{\mathbf{L}} = \text{span}\left\{x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t\right\}$. *Then, for any* $j \in \mathbb{N}$,

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

Proof. By Lemma 4, the elements e_{t-j} of the orthonormal system $\mathcal{E}_t$ coincide with $\mathbf{L}^j e_t$, for any $j \in \mathbb{N}$. Therefore,

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

■

The last corollary is actually based on the interaction between the lag operator and the projection map highlighted by Lemma 4. We are now ready to show precisely how the orthogonal decomposition of the space $\mathcal{H}_t(\mathbf{x})$ is deduced from the Abstract Wold Theorem when the lag operator is the involved isometry.

Theorem 18 *Consider the Hilbert space $\mathcal{H}_t(\mathbf{x})$ and the lag operator $\mathbf{L}$. Then, $\mathcal{H}_t(\mathbf{x})$ decomposes uniquely into an orthogonal sum*

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

such that

$$\mathbf{L}\hat{\mathcal{H}}_t(\mathbf{x}) = \hat{\mathcal{H}}_t(\mathbf{x}), \qquad \mathbf{L}\tilde{\mathcal{H}}_t(\mathbf{x}) \subset \tilde{\mathcal{H}}_t(\mathbf{x}),$$

the restriction of $\mathbf{L}$ on $\hat{\mathcal{H}}_t(\mathbf{x})$ is unitary and the restriction of $\mathbf{L}$ on $\tilde{\mathcal{H}}_t(\mathbf{x})$ is a unilateral shift. In particular,

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

Proof. Since $\mathbf{L}$ is an isometry we apply the Abstract Wold Decomposition, i.e. Theorem 9.

As for the subspace $\hat{\mathcal{H}}_t(\mathbf{x})$, since each $\mathbf{L}^j\mathcal{H}_t(\mathbf{x})$ coincides with $\mathcal{H}_{t-j}(\mathbf{x})$ by Proposition 13, we have that

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

Now, turn to the subspace $\tilde{\mathcal{H}}_t(\mathbf{x})$. By Theorem 9, we know that

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

where $\mathcal{L}_t^{\mathbf{L}}$ is the innovation subspace defined by

$$\mathcal{L}_t^{\mathbf{L}} = \mathcal{H}_t(\mathbf{x}) \ominus \mathbf{L}\mathcal{H}_t(\mathbf{x}) = \mathcal{H}_t(\mathbf{x}) \ominus \mathcal{H}_{t-1}(\mathbf{x}).$$

Since the orthogonal complement of $\mathbf{L}\mathcal{H}_t(\mathbf{x})$ is the kernel of the adjoint operator $\mathbf{L}^*$ (see Lemma 1), by Proposition 16,

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

In addition, by Corollary 17, we know that subspaces $\mathbf{L}^j \mathcal{L}_t^{\mathbf{L}}$, with $j \in \mathbb{N}_0$, are

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

and this completes the decomposition. ■

The theorem essentially collects all the results we mentioned so far. The decomposition is still abstract but the application to time series is now attainable. Additional details are discussed in Section 4.3.

We conclude this subsection by analysing the interaction between non-regularity and the deterministic component of a weakly stationary time series. The question arises from the requirement, in the Classical Wold Decomposition, of a regular time series $\mathbf{x}$. In fact, when regularity lacks, the process $\mathbf{x}$ turns out to be purely deterministic. Furthermore, the converse implication is also true: purely deterministic processes are not regular. As before, the isometry of the lag operator is the fundamental property that is employed. To summarize, the regularity requirement ensures a non-trivial Wold decomposition, in which the non-deterministic component is non-null.

Proposition 19 *Let $\mathbf{x} = \{x_t\}_{t\in\mathbb{Z}}$ be a weakly stationary stochastic process. Then, $\mathbf{x}$ is purely deterministic if and only if it is not regular.*

Proof. Suppose that the process $\mathbf{x}$ is purely deterministic. Then, for any $t \in \mathbb{Z}$,

$$x_t \in \bigcap_{j=0}^{+\infty} \mathcal{H}_{t-j}(\mathbf{x});$$

in particular, x_t belongs both to $\mathcal{H}_t(\mathbf{x})$ and $\mathcal{H}_{t-1}(\mathbf{x})$. Hence, the projection of x_t on the subspace $\mathcal{H}_{t-1}(\mathbf{x})$ is x_t itself:

$$\left\|x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t\right\| = 0.$$

For this reason, $\mathbf{x}$ is not regular.

Conversely, let $\mathbf{x}$ be non-regular. Thus, we can find some integer τ such that

$$\left\|x_\tau - \mathcal{P}_{\mathcal{H}_{\tau-1}(\mathbf{x})} x_\tau\right\| = 0.$$

Since the lag operator is isometric on $\mathcal{H}_t(\mathbf{x})$, see Proposition 12, from Lemma 4 it follows that, for any $j \in \mathbb{N}_0$,

$$\left\|x_{\tau-j} - \mathcal{P}_{\mathcal{H}_{\tau-j-1}(\mathbf{x})} x_{\tau-j}\right\| = \left\|\mathbf{L}^j \left(x_\tau - \mathcal{P}_{\mathcal{H}_{\tau-1}(\mathbf{x})} x_\tau\right)\right\| = \left\|x_\tau - \mathcal{P}_{\mathcal{H}_{\tau-1}(\mathbf{x})} x_\tau\right\| = 0.$$

Similarly, it also holds that, for any $j \in \mathbb{N}$,

$$\left\|x_{\tau+j} - \mathcal{P}_{\mathcal{H}_{\tau+j-1}(\mathbf{x})} x_{\tau+j}\right\| = \left\|\mathbf{L}^j \left(x_{\tau+j} - \mathcal{P}_{\mathcal{H}_{\tau+j-1}(\mathbf{x})} x_{\tau+j}\right)\right\| = \left\|x_\tau - \mathcal{P}_{\mathcal{H}_{\tau-1}(\mathbf{x})} x_\tau\right\| = 0.$$

Therefore,

$$\left\|x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t\right\| = 0 \qquad \forall t \in \mathbb{Z}.$$

The last equality implies that, fixed any $t \in \mathbb{Z}$, the variable $x_t \in \mathcal{H}_t(\mathbf{x})$ belongs also to $\mathcal{H}_{t-1}(\mathbf{x})$. Consequently, we are able to find a sequence $\{X_n\}_n$ of generators

$$X_n = \sum_{k=0}^{+\infty} (\beta_n)_k \, x_{t-1-k} \quad \in \mathcal{H}_{t-1}(\mathbf{x}),$$

such that

$$\|x_t - X_n\| = \left\|x_t - \sum_{k=0}^{+\infty} (\beta_n)_k \, x_{t-1-k}\right\| \longrightarrow 0, \qquad \text{as} \qquad n \longrightarrow +\infty.$$

Since also

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

we have

$$\begin{aligned}
&\left\|x_t - \sum_{k=1}^{+\infty} (\beta_n)_k \, x_{t-1-k} - (\beta_n)_0 \, \mathcal{P}_{\mathcal{H}_{t-2}(\mathbf{x})} x_{t-1}\right\| \\
&\qquad = \left\|x_t - \sum_{k=0}^{+\infty} (\beta_n)_k \, x_{t-1-k} + (\beta_n)_0 \, x_{t-1} - (\beta_n)_0 \, \mathcal{P}_{\mathcal{H}_{t-2}(\mathbf{x})} x_{t-1}\right\| \\
&\qquad \leqslant \left\|x_t - \sum_{k=0}^{+\infty} (\beta_n)_k \, x_{t-1-k}\right\| + \left|(\beta_n)_0\right| \left\|x_{t-1} - \mathcal{P}_{\mathcal{H}_{t-2}(\mathbf{x})} x_{t-1}\right\| \\
&\qquad = \left\|x_t - \sum_{k=0}^{+\infty} (\beta_n)_k \, x_{t-1-k}\right\|.
\end{aligned}$$

As the right-hand side converges to zero, when n increases, we deduce that

$$\left\| x_t - \sum_{k=1}^{+\infty} (\beta_n)_k \, x_{t-1-k} - (\beta_n)_0 \, \mathcal{P}_{\mathcal{H}_{t-2}(\mathbf{x})} x_{t-1} \right\| \quad \longrightarrow \quad 0.$$

From the last expression, we see that x_t belongs also to the subspace $\mathcal{H}_{t-2}(\mathbf{x})$. By induction, it is possible to prove that, for any $t \in \mathbb{Z}$,

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

and this fact ensures that the process $\mathbf{x}$ is purely deterministic. ■

In case regularity lacks at some time index τ, the isometry of $\mathbf{L}$ ensures that regularity is missing at any time $t \in \mathbb{Z}$. This fact forces the process $\mathbf{x}$ to be purely deterministic.

Hence, a weakly stationary time series $\mathbf{x}$ is regular if and only if it is not purely deterministic. This happens when the non-deterministic component is not trivial. In other words, a weakly stationary stochastic process is regular if and only if its Wold decomposition

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

has at least one non-zero impulse response function α_k.

4.3 Classical Wold Decomposition for time series

The final step consists in applying the Wold decomposition of Hilbert spaces stated in Theorem 18 to zero-mean, regular, weakly stationary processes, in order to recover the Classical Wold Decomposition for time series. Our proof is based on the Hilbert space framework and it exploits the isometry of the lag operator. This approach is quite general and can be useful in several applications besides the one presented here.

Theorem 20 (Classical Wold Decomposition) *Let $\mathbf{x} = \{x_t\}_{t \in \mathbb{Z}}$ be a zero-mean, regular, weakly stationary stochastic process. Then, for any $t \in \mathbb{Z}$, x_t decomposes as*

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

where the equality is in the L^2-norm and *i)* $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t\in\mathbb{Z}}$ *is a white noise process with variance equal to* 1*;*

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

$$\alpha_k = \mathbb{E}\left[x_t \varepsilon_{t-k}\right] \qquad \textit{and} \qquad \sum_{k=0}^{+\infty} \alpha_k^2 < +\infty;$$

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

$$\nu_t \in \bigcap_{j=0}^{+\infty} \mathcal{H}_{t-j}(\mathbf{x}) \qquad \textit{and} \qquad \mathbb{E}\left[\nu_t \varepsilon_{t-k}\right] = 0 \quad \forall k \in \mathbb{N}_0;$$

iv)

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

Proof. We apply Theorem 18 to the random variable x_t which belongs to $\mathcal{H}_t(\mathbf{x})$, for any fixed $t \in \mathbb{Z}$. We denote ν_t the orthogonal projection of x_t on the subspace $\hat{\mathcal{H}}_t(\mathbf{x})$ and we define the process $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t\in\mathbb{Z}}$ by setting, for any $t \in \mathbb{Z}$,

$$\varepsilon_t = \frac{x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t}{\left\| x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t \right\|}.$$

By calling α_k the projection coefficient of x_t on the linear subspace generated by ε_{t-k}, we find the decomposition

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

in which the equality is in norm. However, we still need to show that the coefficients α_k do not depend on time index t.

i) By definition we have that $\mathbb{E}\left[\varepsilon_t^2\right] = 1$ for each t. Moreover, as discussed in Lemma 4, the isometry of $\mathbf{L}$ guarantees that

$$\left\| x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t \right\|$$

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

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

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

for some sequence of square-summable coefficients $\{\beta_h\}_h$, we immediately check that

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

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

In general, we can find a sequence $\{X_n\}_n$ of random variables

$$X_n = \sum_{h=0}^{+\infty} (\beta_n)_h \, x_{t-1-h}$$

that converges to $\mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t$ in norm:

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

By exploiting Cauchy-Schwartz' inequality, we can claim that

$$\begin{aligned}\left|\mathbb{E}\left[X_n\right] - \mathbb{E}\left[\mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t\right]\right| &= \left|\mathbb{E}\left[\left(X_n - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t\right) \cdot 1\right]\right| \\ &\leqslant \left\|X_n - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t\right\|.\end{aligned}$$

Therefore, when n goes to infinity,

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

However, we know that, for any $n \in \mathbb{N}$,

$$\mathbb{E}\left[X_n\right] = \sum_{h=0}^{+\infty} (\beta_n)_h \, \mathbb{E}\left[x_{t-1-h}\right] = 0.$$

As a result, by uniqueness of the limit, we infer that $\mathbb{E}\left[\mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t\right] = 0$.

Summing up,

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

and so $\boldsymbol{\varepsilon}$ is a white noise process.

ii) Each α_k is the usual least squares coefficient coming from the projection of the random variable x_t on the linear subspace generated by ε_{t-k}. More explicitly, for any $k \in \mathbb{N}_0$,

$$\begin{aligned}\mathbb{E}\left[x_t \varepsilon_{t-k}\right] &= \mathbb{E}\left[\left(\sum_{h=0}^{+\infty} \alpha_h \varepsilon_{t-h} + \nu_t\right) \varepsilon_{t-k}\right] \\ &= \sum_{h=0}^{+\infty} \alpha_h \mathbb{E}\left[\varepsilon_{t-h} \varepsilon_{t-k}\right] + \mathbb{E}\left[\nu_t \varepsilon_{t-k}\right] \\ &= \alpha_k + \mathbb{E}\left[\nu_t \varepsilon_{t-k}\right] = \alpha_k,\end{aligned}$$

because $\boldsymbol{\varepsilon}$ is white noise and, in addition, ν_t and ε_{t-k} belong to orthogonal subspaces (see Theorem 18). Moreover,

$$\sum_{k=0}^{+\infty} \alpha_k^2 < +\infty$$

because the element $\sum_{k=0}^{\infty} \alpha_k \varepsilon_{t-k}$ belongs to $\tilde{\mathcal{H}}_t(\mathbf{x})$, which is a subspace of $\mathcal{H}_t(\mathbf{x})$.

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

$$\mathbf{L}^j \varepsilon_{t-k} = \varepsilon_{t-k-j}.$$

Given the decomposition of x_t

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

we apply the lag operator on both sides to obtain

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

where we used the linearity and continuity of $\mathbf{L}$. Thus, we observe that the projection coefficients of x_{t-1} on the subspaces generated by ε_{t-k-1} are the same as the projection coefficients of x_t on the subspaces generated by ε_{t-k}, for all $k \in \mathbb{N}_0$. Hence α_k is also equal to

$$\alpha_k = \mathbb{E}\left[x_{t-1} \varepsilon_{t-k-1}\right].$$

More generally, by using the operator $\mathbf{L}^j$, it can be shown that

$$\alpha_k = \mathbb{E}\left[x_{t-j} \varepsilon_{t-k-j}\right] \qquad \forall j \in \mathbb{N}_0$$

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

iii) By following Theorem 18, ν_t is the projection of x_t on the subspace $\hat{\mathcal{H}}_t(\mathbf{x})$ and $\mathbb{E}\left[\nu_t \varepsilon_{t-k}\right] = 0$ for any $k \in \mathbb{N}_0$ because ν_t and ε_{t-k} belong to orthogonal subspaces of $\mathcal{H}_t(\mathbf{x})$.

In addition to this, $\boldsymbol{\nu}$ is zero-mean because, for any $t \in \mathbb{Z}$,

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

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

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

$$\mathbb{E}\left[x_{t-k}\varepsilon_{t-l}\right] = 0 \qquad \forall l \in \{0, \dots, k-1\}.$$

In fact, the variable x_{t-k} has the following decomposition:

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

where we defined

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

Note that the above expression enables us to embed the subspace $\mathcal{H}_{t-k}(\mathbf{x})$ in $\mathcal{H}_t(\mathbf{x})$. Now, since the decomposition of the variable x_{t-k} is unique in $\mathcal{H}_t(\mathbf{x})$, it follows that

$$\mathbb{E}\left[x_{t-k}\varepsilon_{t-l}\right] = \beta_l = 0 \qquad \forall l \in \{0, \dots, k-1\}.$$

Now, the last step in order to show that $\boldsymbol{\nu}$ is weakly stationary is to prove that $\mathbb{E}\left[\nu_{t-p}\nu_{t-q}\right]$ depends at most on the difference $p - q$, for any $p, q \in \mathbb{N}_0$. Indeed, suppose that $q \geqslant p + 1$:

$$
\begin{aligned}
\mathbb{E}\left[\nu_{t-p}\nu_{t-q}\right] &= \mathbb{E}\left[\left(x_{t-p} - \sum_{k=0}^{+\infty} \alpha_k \varepsilon_{t-p-k}\right)\left(x_{t-q} - \sum_{h=0}^{+\infty} \alpha_h \varepsilon_{t-q-h}\right)\right] \\
&= \mathbb{E}\left[x_{t-p}x_{t-q}\right] - \mathbb{E}\left[x_{t-p}\sum_{h=0}^{+\infty} \alpha_h \varepsilon_{t-q-h}\right] \\
&\quad - \mathbb{E}\left[x_{t-q}\sum_{k=0}^{+\infty} \alpha_k \varepsilon_{t-p-k}\right] + \mathbb{E}\left[\left(\sum_{k=0}^{+\infty} \alpha_k \varepsilon_{t-p-k}\right)\left(\sum_{h=0}^{+\infty} \alpha_h \varepsilon_{t-q-h}\right)\right] \\
&= \gamma(p-q) - \sum_{h=0}^{+\infty} \alpha_h \mathbb{E}\left[x_{t-p}\varepsilon_{t-q-h}\right] \\
&\quad - \sum_{k=0}^{+\infty} \alpha_k \mathbb{E}\left[x_{t-q}\varepsilon_{t-p-k}\right] + \sum_{k=0}^{+\infty}\sum_{h=0}^{+\infty} \alpha_k\alpha_h \mathbb{E}\left[\varepsilon_{t-p-k}\varepsilon_{t-q-h}\right] \\
&= \gamma(p-q) - \sum_{h=0}^{+\infty} \alpha_h\alpha_{q-p+h} - \sum_{k=0}^{q-p-1} \alpha_k \mathbb{E}\left[x_{t-q}\varepsilon_{t-p-k}\right] \\
&\quad - \sum_{k=q-p}^{+\infty} \alpha_k \mathbb{E}\left[x_{t-q}\varepsilon_{t-p-k}\right] + \sum_{k=q-p}^{+\infty} \alpha_k\alpha_{p-q+k} \\
&= \gamma(p-q) - \sum_{h=0}^{+\infty} \alpha_h\alpha_{q-p+h} - 0 - \sum_{k=q-p}^{+\infty} \alpha_k\alpha_{p-q+k} + \sum_{k=q-p}^{+\infty} \alpha_k\alpha_{p-q+k} \\
&= \gamma(p-q) - \sum_{h=0}^{+\infty} \alpha_h\alpha_{q-p+h}.
\end{aligned}
$$

As a result, $\mathbb{E}\left[\nu_{t-p}\nu_{t-q}\right]$ depends at most on $p-q$ and so $\boldsymbol{\nu}$ is weakly stationary.

iv) Given that

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

we can see ν_t as an element of the closed subspace $\mathcal{H}_{t-1}(\mathbf{x})$ and so we can find a sequence of variables $\{X_n\}_n \subset \mathcal{H}_{t-1}(\mathbf{x})$ that converges to ν_t in norm. For example, we can set

$$
X_n = \sum_{k=0}^{+\infty} \left(\beta_n\right)_k x_{t-1-k},
$$

with

$$
\left\|\nu_t - X_n\right\| = \left\|\nu_t - \sum_{k=0}^{+\infty} \left(\beta_n\right)_k x_{t-1-k}\right\| \longrightarrow 0, \qquad \text{as} \qquad n \to +\infty.
$$

Note that any of the variables x_{t-1-k} has a Classical Wold Decomposition, that is

$$x_{t-1-k} = \sum_{h=0}^{+\infty} \alpha_h \varepsilon_{t-1-k-h} + \nu_{t-1-k},$$

in which the equality is in norm. By combining these facts together, we get

$$\begin{aligned}
&\left\| \nu_t - \sum_{k=0}^{+\infty} (\beta_n)_k \left(\sum_{h=0}^{+\infty} \alpha_h \varepsilon_{t-1-k-h} + \nu_{t-1-k} \right) \right\| \\
&\qquad \leqslant \left\| \nu_t - \sum_{k=0}^{+\infty} (\beta_n)_k \, x_{t-1-k} \right\| \\
&\qquad\quad + \left\| \sum_{k=0}^{+\infty} (\beta_n)_k \, x_{t-1-k} - \sum_{k=0}^{+\infty} (\beta_n)_k \left(\sum_{h=0}^{+\infty} \alpha_h \varepsilon_{t-1-k-h} + \nu_{t-1-k} \right) \right\| \\
&\qquad \leqslant \left\| \nu_t - \sum_{k=0}^{+\infty} (\beta_n)_k \, x_{t-1-k} \right\| \\
&\qquad\quad + \sum_{k=0}^{+\infty} \left| (\beta_n)_k \right| \left\| x_{t-1-k} - \left(\sum_{h=0}^{+\infty} \alpha_h \varepsilon_{t-1-k-h} + \nu_{t-1-k} \right) \right\| \\
&\qquad = \left\| \nu_t - \sum_{k=0}^{+\infty} (\beta_n)_k \, x_{t-1-k} \right\|
\end{aligned}$$

When n goes to infinity, the right-hand side converges to zero. Therefore, also

$$\left\| \nu_t - \sum_{k=0}^{+\infty} (\beta_n)_k \left(\sum_{h=0}^{+\infty} \alpha_h \varepsilon_{t-1-k-h} + \nu_{t-1-k} \right) \right\| \quad \longrightarrow \quad 0.$$

The last convergence may be rewritten as

$$\left\| \nu_t - \sum_{k=0}^{+\infty} (\beta_n)_k \, \nu_{t-1-k} - \sum_{l=0}^{+\infty} \left(\sum_{k=0}^{l} (\beta_n)_k \, \alpha_{l-k} \right) \varepsilon_{t-1-l} \right\| \quad \longrightarrow \quad 0.$$

Note that the element

$$\sum_{l=0}^{+\infty} \left(\sum_{k=0}^{l} (\beta_n)_k \, \alpha_{l-k} \right) \varepsilon_{t-1-l}$$

belongs to $\tilde{\mathcal{H}}_t(\mathbf{x})$, while

$$\nu_t - \sum_{k=0}^{+\infty} (\beta_n)_k \, \nu_{t-1-k}$$

is contained in $\hat{\mathcal{H}}_t(\mathbf{x})$. Since $\tilde{\mathcal{H}}_t(\mathbf{x})$ and $\hat{\mathcal{H}}_t(\mathbf{x})$ are orthogonal subspaces, in the limit it must hold that

$$\left\| \nu_t - \sum_{k=0}^{+\infty} (\beta_n)_k \, \nu_{t-1-k} \right\| \quad \longrightarrow \quad 0.$$

As a consequence, we can claim that

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

■

Beyond the isometry of **L** and the use of Lemma 4, the key notion exploited in the proof is the orthogonality between the subspaces induced by the Abstract Wold Decomposition of the Hilbert space $\mathcal{H}_t(\mathbf{x})$. The methodology is general and it has wide applicability. In particular, most of the characteristics of the Wold coefficients α_k and of the process $\boldsymbol{\nu}$ are due either to the orthogonality between $\tilde{\mathcal{H}}_t(\mathbf{x})$ and $\hat{\mathcal{H}}_t(\mathbf{x})$ or to the orthogonality among the images of the wandering subspace through the powers of **L**. Properties *i)* and *ii)* help us in characterizing the non-deterministic component of **x**, while the other two fully describe the nature of the deterministic component, especially its predictability.

5 A rescaling operator as isometry

In this final section we suggest an alternative operator that, in case it is isometric, can be used instead of **L** to obtain a novel Wold-type decomposition of a weakly stationary process. In particular, we start from the problem of extracting persistent components from an economic time series.

A convenient way of achieving this goal is given by the application of the discrete Haar transform (or other multiresolution methods) as it is done, for instance, in Ortu, Tamoni, and Tebaldi (2013). This methodology relays on a rescaling operator $\tilde{\mathbf{R}}$, that maps the realizations x_t of a time series $\mathbf{x} = \{x_t\}_{t\in\mathbb{Z}}$ into the sequence of two-by-two averages:

$$\breve{\mathbf{R}}_{\mathbf{x}} : \quad \{x_{t-k}\}_{k\in\mathbb{N}_0} \quad \longmapsto \quad \left\{ \frac{x_{t-2k} + x_{t-2k-1}}{\sqrt{2}} \right\}_{k\in\mathbb{N}_0}.$$

This operator comes from the composition of two operators. Indeed, $\breve{\mathbf{R}}_{\mathbf{x}} = \breve{\mathbf{D}}_{\mathbf{x}}\breve{\mathbf{M}}_{\mathbf{x}}$, where

$$\breve{\mathbf{M}}_{\mathbf{x}} : \quad \{x_{t-k}\}_{k\in\mathbb{N}_0} \quad \longmapsto \quad \left\{ \frac{x_{t-k} + x_{t-k-1}}{2} \right\}_{k\in\mathbb{N}_0}$$

is the dyadic mean operator and

$$\breve{\mathbf{D}}_{\mathbf{x}} : \quad \{x_{t-k}\}_{k\in\mathbb{N}_0} \quad \longmapsto \quad \left\{ \sqrt{2}\, x_{t-2k} \right\}_{k\in\mathbb{N}_0}$$

is the dyadic dilation operator.

Hence, given a zero-mean weakly stationary process $\mathbf{x}$, we define a proper operator $\mathbf{R_x}$ on the Hilbert space $\mathcal{H}_t(\mathbf{x})$ in a way that it mimics the original $\breve{\mathbf{R}}_\mathbf{x}$ on linear combinations of random variables.

Definition 21 *We call* $\mathbf{R_x}$ *the operator* $\mathbf{R_x} : \mathcal{H}_t(\mathbf{x}) \longrightarrow \mathcal{H}_t(\mathbf{x})$ *that acts on generators of* $\mathcal{H}_t(\mathbf{x})$ *as*

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

The usual requirements that $\mathbf{x}$ is zero-mean, regular and weakly stationary are not sufficient to ensure that $\mathbf{R_x}$ is well-defined and isometric on $\mathcal{H}_t(\mathbf{x})$. Other conditions need to be satisfied by the process $\mathbf{x}$. When the operator $\mathbf{R_x}$ is isometric, the Abstract Wold Theorem supplies a novel Wold-type decomposition. In this decomposition, the non-deterministic and the deterministic components have a new interpretation because they are now associated with a rescaling operator and not with $\mathbf{L}$. Similarly to what we did in Subsection 4.2, a convenient way to determine the wandering subspace is to exploit the kernel of the adjoint operator, as suggested by Lemma 1. The main outcome of the application of the Abstract Wold Theorem consists of the orthogonality of the arising components in the new decomposition, which is guaranteed by construction. An in-depth discussion on this topic is present in Ortu, Severino, Tamoni, and Tebaldi.

6 Conclusion

In this work we showed how the Classical Wold Decomposition for time series can be derived from the Abstract Wold Theorem, that involves isometries on Hilbert spaces in an abstract framework. In particular, we dealt with the space $\mathcal{H}_t(\mathbf{x})$ spanned by a weakly stationary stochastic process $\mathbf{x}$ and we employed the lag operator as an isometry. The orthogonality of the abstract decomposition allowed us to retrieve the properties of the deterministic and the non-deterministic components into which a weakly stationary time series can be decomposed.

The fundamental step in the construction is the characterization of the wandering subspace. In order to determine this subspace, we analyzed the kernel of the adjoint of the lag operator.

This procedure suggests a more general methodology that can be followed in order to find Wold-type decompositions with isometries different from the lag operator. Any of these decompositions inherit the orthogonality properties of the Abstract Wold Theorem. Therefore, they are suitable for econometric applications because they provide a feasible shock identification.

Appendix

Proof of Lemma 2

Take into consideration an element $y \in \mathcal{I}$ such that there exists a sequence $\{y_n\}_n \subset \mathbf{V}\mathcal{I}$ so that

$$\|y_n - y\| \longrightarrow 0, \qquad \text{as} \qquad n \to +\infty.$$

Our aim is to show that actually $y \in \mathbf{V}\mathcal{I}$.

Since each y_n belongs to the image $\mathbf{V}\mathcal{I}$, we can find an element $x_n \in \mathcal{I}$ such that $y_n = \mathbf{V}x_n$. As a result, we come up with a sequence $\{x_n\}_n \subset \mathcal{I}$. In addition, since $\{y_n\}_n$ is convergent in $\mathcal{I}$, it is also a Cauchy sequence and so, for any $\varepsilon > 0$, there exists a natural number N_ε such that, for any $n, m \in \mathbb{N}$, with $n, m > N_\varepsilon$,

$$\varepsilon > \|y_n - y_m\| = \|\mathbf{V}(x_n - x_m)\| = \|x_n - x_m\|.$$

The last equality follows from the fact that $\mathbf{V}$ is an isometry. We can infer that $\{x_n\}_n$ is a Cauchy sequence in $\mathcal{I}$, which is a complete subspace. Thus, this sequence converges to some element $x \in \mathcal{I}$. By continuity of the operator $\mathbf{V}$, it follows that $\mathbf{V}x_n$ converges to $\mathbf{V}x$ in $\mathbf{V}\mathcal{I}$, namely

$$\|y_n - \mathbf{V}x\| \longrightarrow 0, \qquad \text{as} \qquad n \to +\infty.$$

By uniqueness of the limit, $y = \mathbf{V}x$ in the norm of $\mathcal{H}$, so y belongs to the image $\mathbf{V}\mathcal{I}$. Hence, we conclude that $\mathbf{V}\mathcal{I}$ is closed in $\mathcal{I}$.

Proof of Lemma 3

As $\mathbf{V}$ is a unilateral shift, there exists a wandering subspace $\mathcal{L}^{\mathbf{V}}$ that allows us to decompose the Hilbert space $\mathcal{H}$ as

$$\mathcal{H} = \bigoplus_{j=0}^{+\infty} \mathbf{V}^j \mathcal{L}^{\mathbf{V}}$$

and so

$$\mathbf{V}\mathcal{H} = \bigoplus_{j=1}^{+\infty} \mathbf{V}^j \mathcal{L}^{\mathbf{V}}.$$

As a consequence, we find

$$\mathcal{H} = \mathcal{L}^{\mathbf{V}} \oplus \left(\bigoplus_{j=1}^{+\infty} \mathbf{V}^j \mathcal{L}^{\mathbf{V}} \right) = \mathcal{L}^{\mathbf{V}} \oplus \mathbf{V}\mathcal{H},$$

namely $\mathcal{L}^{\mathbf{V}} = \mathcal{H} \ominus \mathbf{V}\mathcal{H}$. The last relation ensures that the innovation subspace $\mathcal{L}^{\mathbf{V}}$ is uniquely determined by $\mathbf{V}$.

Proof of Lemma 4

Without loss of generality, we prove the result when $k = 0$. Moreover, we show the property when $j = 1$, as the general case easily follows by induction. So, we provide evidence of the fact that

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

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

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

The first one may be rewritten as

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

while the second one becomes

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

By exploiting the isometry of the operator $\mathbf{L}$ in the last equation, we deduce that

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

but now we recognize that the left-hand side is exactly the same as the one of the first equation. By matching expressions we get that

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

that is,

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

Since the last normal equation holds for all variables $\{x_{t-2-l}\}_{l \in \mathbb{N}_0}$, that generate the subspace $\mathcal{H}_{t-2}(\mathbf{x})$, the condition is also satisfied by the orthonormal system $\mathcal{E}_{t-2}$, namely:

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

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

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

and so,

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

What we showed so far enables us to prove that $\mathbf{L}e_t = e_{t-1}$. Indeed, by taking advantage again of the isometry of $\mathbf{L}$, we infer that

$$\begin{aligned}\mathbf{L}e_t &= \frac{\mathbf{L}\left(x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t\right)}{\left\|x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t\right\|} = \frac{\mathbf{L}x_t - \mathbf{L}\mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t}{\left\|\mathbf{L}x_t - \mathbf{L}\mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})}x_t\right\|} \\ &= \frac{x_{t-1} - \mathcal{P}_{\mathcal{H}_{t-2}(\mathbf{x})}x_{t-1}}{\left\|x_{t-1} - \mathcal{P}_{\mathcal{H}_{t-2}(\mathbf{x})}x_{t-1}\right\|} = e_{t-1}.\end{aligned}$$

References

Herman J Bierens. The wold decomposition. 2012.

Peter J Brockwell and Richard A Davis. *Time series: theory and methods*. Springer, 2009.

Eric M Leeper, Christopher A Sims, Tao Zha, Robert E Hall, and Ben S Bernanke. What does monetary policy do? *Brookings papers on economic activity*, 1996(2):1–78, 1996.

David G Luenberger. *Optimization by vector space methods*. John Wiley & Sons, 1968.

Béla Sz Nagy, Ciprian Foias, Hari Bercovici, and László Kérchy. *Harmonic analysis of operators on Hilbert space*. Springer, 2010.

Fulvio Ortu, Andrea Tamoni, and Claudio Tebaldi. Long-run risk and the persistence of consumption shocks. *Review of financial studies*, 26(11):2876–2915, 2013.

Fulvio Ortu, Federico Severino, Andrea Tamoni, and Claudio Tebaldi. A persistence-based wold-type decomposition for stationary time series. *Available at SSRN 1973049*, 2015.

Walter Rudin. *Real and complex analysis*. Tata McGraw-Hill Education, 1987.

Herman Wold. A study in the analysis of stationary time series. 1938.