An Innovations Algorithm for the prediction of functional linear processes

Johannes Klepsch[*]   Claudia Klüppelberg[*]

July 19, 2016

Abstract

When observations are curves over some natural time interval, the field of functional data analysis comes into play. Functional linear processes account for temporal dependence in the data. The prediction problem for functional linear processes has been solved theoretically, but the focus for applications has been on functional autoregressive processes. We propose a new computationally tractable linear predictor for functional linear processes. It is based on an application of the Multivariate Innovations Algorithm to finite-dimensional subprocesses of increasing dimension of the infinite-dimensional functional linear process. We investigate the behavior of the predictor for increasing sample size. We show that, depending on the decay rate of the eigenvalues of the covariance and the spectral density operator, the resulting predictor converges with a certain rate to the theoretically best linear predictor.

AMS 2010 Subject Classifications: primary: 62M10, 62M15, 62M20   secondary: 62H25, 60G25
Keywords: functional linear process, functional principal components, functional time series, Hilbert space valued process, Innovations Algorithm, prediction, prediction error

1   Introduction

We consider observations which are consecutive curves over a fixed time interval within the field of functional data analysis (FDA). In this paper curves are representations of a functional linear process. The data generating process is a time series $X = (X_n)_{n \in \mathbb{Z}}$ where each X_n is a random element $X_n(t)$, $t \in [0,1]$, of a Hilbert space, often the space of square integrable functions on $[0,1]$.

Several books contain a mathematical or statistical treatment of dependent functional data as e.g. Bosq [4], Horvàth and Kokoszka [13], and Bosq and Blanke [7]. The main source of our paper is the book [4] on linear processes in function spaces, which gives the most general mathematical treatment of linear dependence in functional data, developing estimation, limit theorems and prediction for functional autoregressive processes. In [12] the authors develop limit theorems for the larger class of weakly dependent functional processes. More recently, [14] and [23] contribute to frequency domain methods of functional time series.

Solving the prediction equations in function spaces is problematic and research to-date has mainly considered first order autoregressive models. Contributions to functional prediction go

---

[*]Center for Mathematical Sciences, Technische Universität München, 85748 Garching, Boltzmannstrasse 3, Germany, e-mail: johannes.klepsch@tum.de , cklu@tum.de

hand in hand with an estimation method for the autoregressive parameter operator. The book [4] suggests a Yule-Walker type moment estimator, spline approximation is applied in [3], and [17] proposes a predictive factor method where the principal components are replaced by directions which may be more relevant for prediction.

When moving away from the autoregressive process, results on prediction of functional time series become sparse. An interesting theory for the prediction of general functional linear processes is developed in [6]. Necessary and sufficient conditions are derived for the best linear predictor to take the form $\phi_n(X_1, \ldots, X_n)$ with ϕ_n linear and bounded. However, due to the infinite dimensionality of function spaces boundedness of ϕ_n cannot be guaranteed. Consequently, most results, though interesting from a theoretical point of view, are not suitable for application.

More practical results are given for example in [1], where prediction is performed non-parametrically with a functional kernel regression technique, or in [2], [16] and [18], where the dimensionality of the prediction problem is reduced via functional principal component analysis. In a multivariate setting, the Innovations Algorithm proposed in [8] gives a robust prediction method for linear processes. However, as often in functional data analysis, the non-invertibility of covariance operators prevents an ad-hoc generalization of the Innovations Algorithm to functional linear processes.

We suggest a computationally feasible linear prediction method extending the Innovations Algorithm to the functional setting. For a functional linear process $(X_n)_{n\in\mathbb{Z}}$ with values in a Hilbert space H and with innovation process $(\varepsilon_n)_{n\in\mathbb{Z}}$ our goal is a linear predictor $\widehat{X}_{n+1}$ based on $X_1, \ldots, X_n$ such that $\widehat{X}_{n+1}$ is both computationally tractable and *consistent*. In other words, we want to find a bounded linear mapping ϕ_n with $\widehat{X}_{n+1} = \phi_n(X_1, \ldots, X_n)$ such that the statistical prediction error converges to 0 for increasing sample size; i.e.,

$$\lim_{n\to\infty} \mathbb{E}\|X_{n+1} - \widehat{X}_{n+1}\|^2 = \mathbb{E}\|\varepsilon_0\|^2. \tag{1.1}$$

To achieve convergence in (1.1) we work with finite dimensional projections of the functional process, similarly as in [2] and [18]. We start with a representation of the functional linear model in terms of an arbitrary orthonormal basis of the Hilbert space. We then focus on a representation of the model based on only finitely many basis functions. An intuitive choice for the orthonormal basis consists of the eigenfunctions of the covariance operator of the process. Taking the eigenfunctions corresponding to the D largest eigenvalues results in a truncated Karhunen-Loéve representation, and guarantees to capture most of the variance of the process (see [2]). Other applications may call for a different choice.

Though the idea of finite dimensional projections is not new, our approach differs significantly from existing ones. Previous approaches consider the innovations of the projected process as the projection of the innovation of the original functional process. Though this may be sufficient in practice, it is in general not theoretically accurate.

The Wold decomposition enables us to work with the exact dynamics of the projected process, which then allows us to derive precise asymptotic results. The task set for this paper is of a purely predictive nature: we assume knowing the dependence structure and do not perform model selection or covariance estimation. This will be the topic of a subsequent paper.

The truncated process $(X_{D,n})_{n\in\mathbb{Z}}$ based on D basis functions is called subprocess. We show that every subprocess of a stationary (and invertible) functional process is again stationary (and invertible). We then use an isometric isomorphy to a D-dimensional vector process to compute the best linear predictor of $(X_{D,n})_{n\in\mathbb{Z}}$ by the Multivariate Innovations Algorithm (see e.g. [8]).

As a special example we investigate the functional moving average process of finite order. We prove that every subprocess is again a functional moving average process of same order or less. Moreover, for this process the Innovations Algorithm simplifies. Invertibility is a natural assumption in the context of prediction (cf. [8], Section 5.5, and [22]), and we require it when proving limit results. The theoretical results on the structure of $(X_{D,n})_{n \in \mathbb{Z}}$ enable us to quantify the prediction error in (1.1). As expected, it can be decomposed in two terms, one due to the dimension reduction, and the other due to the statistical prediction error of the D-dimensional model. However, the goal of consistency as in (1.1) is not satisfied, as the error due to dimension reduction does not depend on the sample size.

Finally, in order to satisfy (1.1), we propose a modified version of the Innovations Algorithm. The idea is to increase D together with the sample size. Hence the iterations of our modified Innovations Algorithm are based on increasing subspaces. Here we focus on the eigenfunctions of the covariance operator of X as orthonormal basis of the function space.

Our main result states that the prediction error is a combination of two tail sums, one involving operators of the inverse representation of the process, and the other the eigenvalues of the covariance operator. We obtain a computationally tractable functional linear predictor for stationary invertible functional linear processes. As the sample size tends to infinity the predictor satisfies (1.1) with a rate depending on the eigenvalues of the covariance operator and of the spectral density operator.

Our paper is organized as follows. After summarizing prerequisites of functional time series in Section 2, we recall in Section 3 the framework of prediction in infinite dimensional Hilbert spaces, mostly based on the work of Bosq (see [4, 5, 6]). Here we also clarify the difficulties of linear prediction in infinite dimensional function spaces. In Section 4 we propose an Innovations Algorithm based on a finite dimensional subprocess of X. The predictor proposed in Section 4, though quite general, does not satisfy (1.1). Hence, in Section 5 we project the process on a finite-dimensional subspace spanned by the eigenfunctions of the covariance operator of X, and formulate the prediction problem in such a way that the dimension of the subprocess increases with the sample size. A modification of the Innovations Algorithm then yields a predictor, which satisfies (1.1) and remains computationally tractable. The proof of this result requires some work and is deferred to Section 6 along with some auxiliary results.

2   Methodology

Let $H = L^2([0,1])$ be the real Hilbert space of square integrable functions with norm $\|x\|_H = (\int_0^1 x^2(s)ds)^{1/2}$ generated by the inner product $\langle x, y \rangle = \int_0^1 x(s)y(s)ds$ for $x, y \in H$. We denote by $\mathcal{L}$ the space of bounded linear operators acting on H. If not stated differently, for $A \in \mathcal{L}$ we take the standard operator norm $\|A\|_{\mathcal{L}} = \sup_{\|x\| \leq 1} \|Ax\|$. Its adjoint A^* is defined by $\langle Ax, y \rangle = \langle x, A^*y \rangle$ for $x, y \in H$. The operator $A \in \mathcal{L}$ is called nuclear operator (denoted by $\mathcal{N}$), if $\|A\|_{\mathcal{N}} = \sum_{j=1}^{\infty} \langle Ae_j, e_j \rangle < \infty$ for some (and hence all) orthonormal basis (ONB) $(e_j)_{j \in \mathbb{N}}$ of H. If additionally A is self-adjoint, then $\|A\|_{\mathcal{N}} = \sum_{j=1}^{\infty} |\lambda_j| < \infty$, where $(\lambda_j)_{j \in \mathbb{N}}$ are the eigenvalues of A. We shall also use the estimate $\|AB\|_{\mathcal{N}} \leq \|A\|_{\mathcal{L}} \|B\|_{\mathcal{N}}$ for $A \in \mathcal{L}$ and $B \in \mathcal{N}$. For an introduction and more insight into Hilbert spaces we relied on Chapters 3.2 and 3.6 in [24].

Let $\mathcal{B}_H$ be the Borel σ-algebra of subsets of H. All random functions are defined on a probability space $(\Omega, \mathcal{A}, \mathcal{P})$ and are $\mathcal{A} - \mathcal{B}_H$-measurable. Then the space of square integrable random functions $L_H^2 := L^2(\Omega, \mathcal{A}, \mathcal{P})$ is a Hilbert space with inner product $\mathbb{E}\langle X, Y \rangle = \mathbb{E}\int_0^1 X(s)Y(s)ds$

for $X, Y \in L_H^2$. Furthermore, we say that X is integrable if $\mathbb{E}\|X\| = \mathbb{E}[(\int_0^1 X^2(t)dt)^{1/2}] < \infty$.

From Lemma 1.2 of [4] we know that X is a random function with values in H if and only if $\langle \mu, X \rangle$ is a real random variable for every $\mu \in H$. Hence, the following definitions are possible.

Definition 2.1. (i) If $X \in L_H^2$ is integrable, then there exists a unique $\mu \in H$ such that $\mathbb{E}\langle y, X \rangle = \langle y, \mu \rangle$ for $y \in H$. It follows that $\mathbb{E}X(t) = \mu(t)$ for almost all $t \in [0, 1]$, and $\mathbb{E}X \in H$ is called the *expectation* of X.

(ii) If $X \in L_H^2$ and $\mathbb{E}X = 0 \in H$, the *covariance operator* of X is defined as

$$C_X(y) = \mathbb{E}[\langle X, y \rangle X], \quad y \in H.$$

(iii) If $X, Y \in L_H^2$ and $\mathbb{E}X = \mathbb{E}Y = 0$, the *cross covariance operator* of X and Y is defined as

$$C_{X,Y}(y) = C_{Y,X}^*(y) = \mathbb{E}[\langle X, y \rangle Y], \quad y \in H.$$

The operators C_X and $C_{Y,X}$ belong to $\mathcal{N}$ (cf. [4], Section 1.5). Furthermore, C_X is a self-adjoint ($C_X = C_X^*$) and non-negative definite operator with spectral representation

$$C_X(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, \nu_j \rangle \nu_j, \quad x \in H,$$

for eigenpairs $(\lambda_j, \nu_j)_{j \in \mathbb{N}}$, where $(\nu_j)_{j \in \mathbb{N}}$ is an ONB of H and $(\lambda_j)_{j \in \mathbb{N}}$ is a sequence of positive real numbers such that $\sum_{j=1}^{\infty} \lambda_j < \infty$. When considering spectral representations, we assume that the λ_j are ordered decreasingly; i.e., $\lambda_i \geq \lambda_k$ for $i < k$.

For ease of notation we introduce the operator

$$x \otimes y(\cdot) = \langle x, \cdot \rangle y,$$

which allows us to write $C_X = \mathbb{E}[X \otimes X]$ and $C_{X,Y} = \mathbb{E}[X \otimes Y]$. Note also that

$$\mathbb{E}\|X\|^2 = \mathbb{E}\|X \otimes X\|_{\mathcal{N}} = \|C_X\|_{\mathcal{N}}. \tag{2.1}$$

Additionally, the following equalities are useful: for $A \in \mathcal{L}$ and $x_i, y_i \in H$ for $i = 1, 2$ we have

$$
\begin{aligned}
A(x_1 \otimes y_1) &= A(\langle x_1, \cdot \rangle y_1) = \langle x_1, \cdot \rangle A y_1 = x_1 \otimes A y_1, \\
(x_1 + x_2) \otimes (y_1 + y_2) &= x_1 \otimes y_1 + x_1 \otimes y_2 + x_2 \otimes y_1 + x_2 \otimes y_2.
\end{aligned}
\tag{2.2}
$$

We define now functional linear processes and state some of their properties, taken from [4], Section 1.5 and Section 3.1. We first define the driving noise sequence.

Definition 2.2. $(\varepsilon_n)_{n \in \mathbb{Z}}$ is *white noise (WN)* in L_H^2 if $\mathbb{E}\varepsilon_n = 0$, $0 < \mathbb{E}\|\varepsilon_n\|^2 = \sigma^2 < \infty$, $C_{\varepsilon_n} = C_\varepsilon$ is independent of n, and if $C_{\varepsilon_n, \varepsilon_m} = 0$ for all $n, m \in \mathbb{Z}$, $n \neq m$.

Definition 2.3. [[4], Definition 7.1] Let $(\varepsilon_n)_{n \in \mathbb{Z}}$ be WN and $(\psi_j)_{j \in \mathbb{N}}$ a sequence in $\mathcal{L}$. Define $\psi_0 = I_H$, the identity operator on H, and let $\mu \in H$. We call $(X_n)_{n \in \mathbb{Z}}$ satisfying

$$X_n = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{n-j}, \quad n \in \mathbb{Z}, \tag{2.3}$$

a *functional linear process* in L_H^2 with mean μ. The series in (2.3) converges in probability.

Note that by definition a functional linear process is causal. We now state assumptions to ensure stronger convergence of the above series.

4

Lemma 2.4 ([4], Lemma 7.1(2)). *Let $(\varepsilon_n)_{n\in\mathbb{Z}}$ be WN and $\sum_{j=0}^{\infty}\|\psi_j\|_{\mathcal{L}}^2 < \infty$. Then the series in (2.3) converges in L_H^2 and a.s., and $(X_n)_{n\in\mathbb{Z}}$ is (weakly) stationary.*

Strict stationarity of a functional linear process can be enforced by assuming that $(\varepsilon_n)_{n\in\mathbb{Z}}$ is additionally independent. In our setting weak stationarity will suffice. From here on, without loss of generality we set $\mu = 0$. For a stationary process $(X_n)_{n\in\mathbb{Z}}$, the covariance operator with lag h is denoted by

$$C_{X;h} = \mathbb{E}[X_0 \otimes X_h], \quad h \in \mathbb{Z}. \tag{2.4}$$

We now define the concept of invertibility of a functional linear process, which is a natural assumption in the context of prediction; cf. [8], Chapter 5.5 and [22].

Definition 2.5. [[19], Definition 2] A functional linear process $(X_n)_{n\in\mathbb{Z}}$ is said to be *invertible* if it admits the representation

$$X_n = \varepsilon_n + \sum_{j=1}^{\infty} \pi_j X_{n-j}, \quad n \in \mathbb{Z}, \tag{2.5}$$

for $\pi_j \in \mathcal{L}$ and $\sum_{j=1}^{\infty}\|\pi_j\|_{\mathcal{L}} < \infty$.

Note that, as for univariate and multivariate time series models, every stationary causal functional autoregressive moving average (FARMA) process is a functional linear process (see [25], Theorem 2.3). Special cases include functional autoregressive processes of order $p \in \mathbb{N}$ (FAR(p)), which have been thoroughly investigated. Our focus is on functional linear models, with the functional moving average process of order $q \in \mathbb{N}$ (FMA(q)) as illustrating example, which we investigate in Section 4.2.

Definition 2.6. For $q \in \mathbb{N}$ a FMA(q) is a functional linear process $(X_n)_{n\in\mathbb{Z}}$ in L_H^2 such that

$$X_n = \varepsilon_n + \sum_{j=1}^{q} \gamma_j \varepsilon_{n-j}, \quad n \in \mathbb{Z}, \tag{2.6}$$

for WN $(\varepsilon_n)_{n\in\mathbb{Z}}$ and $\gamma_j \in \mathcal{L}$ for $j = 1,\ldots,q$.

A FMA(q) process can be characterized as follows:

Proposition 2.7 ([7], Prop. 10.2). *A stationary functional linear process $(X_n)_{n\in\mathbb{Z}}$ in L_H^2 is a FMA(q) for some $q \in \mathbb{N}$ if and only if $C_{X;q} \neq 0$ and $C_{X;h} = 0$ for $|h| > q$.*

3 Prediction in Hilbert spaces

In a finite dimensional setting, when the random elements take values in $\mathbb{R}^d$ equipped with the Euclidean norm, the concept of linear prediction of a random vector is well known (e.g. [8], Section 11.4). The best linear approximation of a random vector $\mathbf{X}$ based on vectors $(\mathbf{X}_1,\ldots,\mathbf{X}_n$ is the orthogonal projection of each component of $\mathbf{X}$ on the smallest closed linear subspace of $L_{\mathbb{R}}^2(\Omega, \mathcal{A}, \mathbb{P})$ generated by the components of $\mathbf{X}_i$. This results in

$$\widehat{\mathbf{X}} := \sum_{i=1}^{n} \mathbf{\Phi}_{n,i} \mathbf{X}_i$$

5

for $\boldsymbol{\Phi}_{n,i} \in \mathbb{R}^{d \times d}$. In infinite dimensional Hilbert spaces we proceed similarly, but need a rich enough subspace on which to project. The concept of linear prediction in infinite dimensional Hilbert spaces was introduced by Bosq; see Section 1.6 in [4]. We start by recalling the notion of $\mathcal{L}$-closed subspaces (LCS), introduced in [10].

Definition 3.1. $\mathcal{G}$ is said to be an $\mathcal{L}$-*closed subspace* (LCS) of L_H^2 if $\mathcal{G}$ is a Hilbertian subspace of L_H^2, and if $X \in \mathcal{G}$ and $l \in \mathcal{L}$ imply $lX \in \mathcal{G}$.

We now give a characterization of a LCS generated by a subset of L_H^2.

Proposition 3.2 ([4], Theorem 1.8)**.** *Let $F \subseteq L_H^2$. Then the* LCS *generated by F, denoted by* LCS(F)*, is the closure with respect to $\|\cdot\|$ of*

$$F' = \Big\{ \sum_{i=1}^k l_i X_i : \ l_i \in \mathcal{L}, \ X_i \in F, \ k \geq 1 \Big\}.$$

We are now ready to define the best linear predictor in an infinite dimensional Hilbert space analogous to the finite dimensional setting.

Definition 3.3. Let $X_1, \ldots, X_n$ be zero mean random elements in L_H^2. Define

$$F_n = \{X_1, \ldots, X_n\} \quad \text{and} \quad \widehat{X}_{n+1} = P_{\mathrm{LCS}(F_n)}(X_{n+1}), \tag{3.1}$$

i.e., $\widehat{X}_{n+1}$ is the orthogonal projection of X_{n+1} on LCS(F_n). Then $\widehat{X}_{n+1}$ is called *best linear functional predictor* of X_{n+1} based on LCS(F_n).

Note however that, since F' is not closed, $\widehat{X}_{n+1}$ as in (3.1) has in general not the form $\widehat{X}_{n+1} = \sum_{i=1}^n l_i X_i$ for $l_i \in \mathcal{L}$ (e.g. [6], Proposition 2.2). For functional linear processes the above representation is purely theoretical. In the following we develop an alternative approach based on finite dimensional projections of the functional process.

4 Prediction based on a finite dimensional projection

For a stationary functional linear process $(X_n)_{n \in \mathbb{Z}}$ the infinite dimensional setting makes the computation of $\widehat{X}_{n+1}$ as in (3.1) basically impossible. A natural solution lies in finite dimensional projections of the functional process $(X_n)_{n \in \mathbb{Z}}$. For fixed $D \in \mathbb{N}$ we define

$$(\nu_i)_{i \in \mathbb{N}} \quad \text{and} \quad A_D = \overline{\mathrm{sp}}\{\nu_1, \ldots, \nu_D\}, \tag{4.1}$$

where $(\nu_i)_{i \in \mathbb{N}}$ is some ONB of H, and consider the projection of a functional random element on A_D. In [2] and [18] the authors consider the projection of a FAR process $(X_n)_{n \in \mathbb{Z}}$ on A_D, where $\nu_1, \ldots, \nu_D$ are the eigenfunctions corresponding to the largest eigenvalues of C_X. However, instead of considering the true dynamics of the subprocess, they work with an approximation which lies in the same model class as the original functional process; e.g. projections of functional AR(p) models are approximated by multivariate AR(p) models. The following examples clarifies this concept.

Example 4.1. Consider a FAR(1) process $(X_n)_{n \in \mathbb{Z}}$ as defined in Section 3.2 of [4] by

$$X_n = \Phi X_{n-1} + \varepsilon_n, \qquad n \in \mathbb{Z}, \tag{4.2}$$

for some $\Phi \in \mathcal{L}$ and WN $(\varepsilon_n)_{n\in\mathbb{Z}}$. Let furthermore $(\nu_i)_{i\in\mathbb{N}}$ be an arbitrary ONB of H. Then (4.2) can be rewritten in terms of $(\nu_i)_{i\in\mathbb{N}}$ as

$$
\begin{pmatrix} \langle X_n,\nu_1\rangle \\ \vdots \\ \langle X_n,\nu_D\rangle \\ \hline \langle X_n,\nu_{D+1}\rangle \\ \vdots \end{pmatrix} = \left(\begin{array}{ccc|cc} \langle \phi\nu_1,\nu_1\rangle & \cdots & \langle \phi\nu_D,\nu_1\rangle & \langle \phi\nu_{D+1},\nu_1\rangle & \cdots \\ \vdots & \ddots & \vdots & \vdots & \ddots \\ \langle \phi\nu_1,\nu_D\rangle & \cdots & \langle \phi\nu_D,\nu_D\rangle & \langle \phi\nu_{D+1},\nu_D\rangle & \cdots \\ \hline \langle \phi\nu_1,\nu_{D+1}\rangle & \cdots & \langle \phi\nu_D,\nu_{D+1}\rangle & \langle \phi\nu_{D+1},\nu_{D+1}\rangle & \cdots \\ \vdots & \ddots & \vdots & \vdots & \ddots \end{array}\right) \begin{pmatrix} \langle X_{n-1},\nu_1\rangle \\ \vdots \\ \langle X_{n-1},\nu_D\rangle \\ \hline \langle X_{n-1},\nu_{D+1}\rangle \\ \vdots \end{pmatrix} + \begin{pmatrix} \langle \varepsilon_n,\nu_1\rangle \\ \vdots \\ \langle \varepsilon_n,\nu_D\rangle \\ \hline \langle \varepsilon_n,\nu_{D+1}\rangle \\ \vdots \end{pmatrix},
$$

which we abbreviate as

$$
\begin{pmatrix} \mathbf{X}_{D,n} \\ \hline \mathbf{X}_n^\infty \end{pmatrix} = \left[\begin{array}{c|c} \boldsymbol{\Phi}_D & \boldsymbol{\Phi}_D^\infty \\ \hline \vdots & \vdots \end{array}\right] \begin{pmatrix} \mathbf{X}_{D,n-1} \\ \hline \mathbf{X}_{n-1}^\infty \end{pmatrix} + \begin{pmatrix} \mathbf{E}_{D,n} \\ \hline \mathbf{E}_n^\infty \end{pmatrix}. \tag{4.3}
$$

We are interested in the dynamics of the D-dimensional subprocess $(\mathbf{X}_{D,n})_{n\in\mathbb{Z}}$. From (4.3) we find that $(\mathbf{X}_{D,n})_{n\in\mathbb{Z}}$ satisfies

$$
\mathbf{X}_{D,n} = \boldsymbol{\Phi}_D \mathbf{X}_{D,n-1} + \boldsymbol{\Phi}_D^\infty \mathbf{X}_{n-1}^\infty + \mathbf{E}_{D,n}, \quad n \in \mathbb{Z}, \tag{4.4}
$$

which does in general not define a FAR(1) process. This can be seen from the following example, similar to Example 3.7 in [4]. For some $a \in \mathbb{R}$ with $0 < a < 1$ let

$$
\Phi(x) = a\sum_{j=1}^\infty \langle x,\nu_j\rangle \nu_1 + a\sum_{i=1}^\infty \langle x,\nu_i\rangle \nu_{i+1}, \qquad x \in H.
$$

Furthermore, assume that $\mathbb{E}\langle \varepsilon_n,\nu_1\rangle^2 > 0$ but $\mathbb{E}\langle \varepsilon_n,\nu_j\rangle^2 = 0$ for all $j > 1$. Since $\|\Phi\|_{\mathcal{L}} = a < 1$, $(X_n)_{n\in\mathbb{Z}}$ is a stationary FAR(1) process. However, with (4.4) for $D = 1$,

$$
\mathbf{X}_{1,n} = \langle X_n,\nu_1\rangle = a\sum_{j=1}^\infty \langle X_{n-1},\nu_j\rangle + \langle \varepsilon_n,\nu_1\rangle
$$

$$
= a\langle X_{n-1},\nu_1\rangle + a\sum_{j=2}^\infty \Big\langle \Big(a\sum_{j'=1}^\infty \langle X_{n-2},e_{j'}\rangle \nu_1 + a\sum_{i=1}^\infty \langle X_{n-2},\nu_i\rangle e_{i+1} + \varepsilon_{n-1}\Big),\nu_j\Big\rangle + \langle \varepsilon_n,\nu_1\rangle
$$

$$
= a\langle X_{n-1},\nu_1\rangle + a^2\langle X_{n-2},\nu_1\rangle + a^2\sum_{j=2}^\infty \langle X_{n-2},\nu_j\rangle + \langle \varepsilon_n,\nu_1\rangle
$$

$$
= \sum_{j=1}^\infty a^j \mathbf{X}_{1,n-j} + \mathbf{E}_{n,1}.
$$

Hence, $(\mathbf{X}_{1,n})_{n\in\mathbb{Z}}$ follows an AR(∞) model and $(\mathbf{X}_{1,n}\nu_1)_{n\in\mathbb{Z}}$ a FAR(∞) model.

In [2] and [18], $(\mathbf{X}_{D,n})_{n\in\mathbb{Z}}$ is approximated by $(\tilde{\mathbf{X}}_{D,n})_{n\in\mathbb{Z}}$ satisfying

$$
\tilde{\mathbf{X}}_{D,n} = \boldsymbol{\Phi}_D \tilde{\mathbf{X}}_{D,n-1} + \mathbf{E}_{D,n}, \quad n \in \mathbb{Z},
$$

such that $(\tilde{\mathbf{X}}_{D,n})_{n\in\mathbb{Z}}$ follows a vector AR(1) process.

We pursue the idea of Example 4.1 for functional linear processes and work with the true dynamics of a finite-dimensional subprocess.

4.1   Prediction of functional linear processes

For a functional linear process $(X_n)_{n\in\mathbb{Z}}$ we focus on the orthogonal projection

$$
X_{D,n} = P_{A_D}(X_n) = \sum_{j=1}^D \langle X_n,\nu_j\rangle \nu_j, \quad n \in \mathbb{Z}, \tag{4.5}
$$

for $(\nu_i)_{i\in\mathbb{N}}$ and A_D as in (4.1). We will often use the following isometric isomorphism between two Hilbert spaces of the same dimension.

Lemma 4.2. *Define A_D as in (4.1). The map $T : A_D \to \mathbb{R}^D$ defined by $Tx = (\langle x, \nu_i \rangle)_{i=1,\dots,D}$ is a bijective linear mapping with $\langle Tx, Ty \rangle_{\mathbb{R}^D} = \langle x, y \rangle$ for all $x, y \in A_D$. Hence, $\mathrm{LCS}(F_{D,n})$ is isometrically isomorphic to $\overline{\mathrm{sp}}\{\mathbf{X}_{D,1}, \dots, \mathbf{X}_{D,n}\}$. Moreover, $(X_{D,n})_{n \in \mathbb{Z}}$ as defined in (4.5) is isometrically isomorphic to the D-dimensional vector process*

$$\mathbf{X}_{D,n} := (\langle X_n, \nu_1 \rangle, \dots, \langle X_n, \nu_D \rangle)^\top, \quad n \in \mathbb{Z}. \tag{4.6}$$

When choosing $(\nu_i)_{i \in \mathbb{N}}$ as the eigenfunctions of the covariance operator C_X of $(X_n)_{n \in \mathbb{Z}}$, the representation (4.5) is a truncated version of the Karhunen-Loéve decomposition (see e.g. [4], Theorem 1.5).

As known from Example 4.1, the structure of $(X_n)_{n \in \mathbb{Z}}$ does in general not immediately reveal the dynamics of $(X_{D,n})_{n \in \mathbb{Z}}$. Starting with the representation of $(X_{D,n})_{n \in \mathbb{Z}}$ as in (2.3) with $\psi_0 = I_H$ and using similar notation as in (4.4), the D-dimensional vector process $(\mathbf{X}_{D,n})_{n \in \mathbb{Z}}$ can be written as

$$\mathbf{X}_{D,n} = \mathbf{E}_{D,n} + \sum_{j=1}^{\infty} \Big(\boldsymbol{\Psi}_{D,j} \mathbf{E}_{D,n-j} + \boldsymbol{\Psi}_{D,j}^{\infty} \mathbf{E}_{n-j}^{\infty} \Big), \quad n \in \mathbb{Z}, \tag{4.7}$$

where the blocks $\boldsymbol{\Psi}_{D,j}$, $\boldsymbol{\Psi}_{D,j}^{\infty}$, $\mathbf{E}_{D,n} = (\langle \varepsilon_n, \nu_1 \rangle, \dots, \langle \varepsilon_n, \nu_D \rangle)^\top$, and $\mathbf{E}_n^{\infty} = (\langle \varepsilon_n, \nu_{D+1} \rangle, \langle \varepsilon_n, \nu_{D+2} \rangle, \dots)^\top$ are defined analogously to the blocks in (4.3). Note that this is in general not a vector $\mathrm{MA}(\infty)$ representation of a process with innovation $(\mathbf{E}_{D,n})_{n \in \mathbb{Z}}$.

The following proposition summarizes general results on the structure of $(X_{D,n})_{n \in \mathbb{Z}}$. Its proof is given in Section 6.

Proposition 4.3. *Let $(X_n)_{n \in \mathbb{Z}}$ be a stationary (and invertible) functional linear process with WN $(\varepsilon_n)_{n \in \mathbb{Z}}$, such that all eigenvalues of the covariance operator C_ε of $(\varepsilon_n)_{n \in \mathbb{Z}}$ are positive. Then $(X_{D,n})_{n \in \mathbb{Z}}$ is also a stationary (and invertible) functional linear process with some WN $(\tilde{\varepsilon}_{D,n})_{n \in \mathbb{Z}}$. $(\tilde{\varepsilon}_{D,n})_{n \in \mathbb{Z}}$ is isometrically isomorphic to the D-dimensional vector process $(\tilde{\mathbf{E}}_{D,n})_{n \in \mathbb{Z}}$, defined by $\tilde{\mathbf{E}}_{D,n} := (\langle \tilde{\varepsilon}_{D,n}, \nu_1 \rangle, \dots, \langle \tilde{\varepsilon}_{D,n}, \nu_D \rangle)^\top$. Furthermore define $\mathcal{M}_{D,n} = \overline{\mathrm{sp}}\{\mathbf{X}_{D,t}, -\infty < t \le n\}$. Then*

$$\tilde{\mathbf{E}}_{D,n} = \mathbf{E}_{D,n} + \boldsymbol{\Psi}_{D,1}^{\infty}(\mathbf{E}_{n-1}^{\infty} - P_{\mathcal{M}_{D,n-1}}(\mathbf{E}_{n-1}^{\infty})) := \mathbf{E}_{D,n} + \boldsymbol{\Delta}_{D,n-1}, \quad n \in \mathbb{Z}. \tag{4.8}$$

The lagged covariance operator $C_{X_D;h}$ of $(X_{D,n})_{n \in \mathbb{Z}}$ is given by

$$C_{X_D;h} = \mathbb{E}[P_{A_D} X_0 \otimes P_{A_D} X_h] = P_{A_D} \mathbb{E}[X_0 \otimes X_h] P_{A_D} = P_{A_D} C_{X;h} P_{A_D}, \quad h \in \mathbb{Z}. \tag{4.9}$$

By Lemma 4.2, $(X_{D,n})_{n \in \mathbb{Z}}$ is isomorphic to the D-dimensional vector process $(\mathbf{X}_{D,n})_{n \in \mathbb{Z}}$ as defined in (4.6). The prediction problem can therefore be solved by methods from multivariate time series analysis. More precisely, we define for fixed $D \in \mathbb{N}$

$$F_{D,n} = \{X_{D,1}, \dots, X_{D,n}\} \quad \text{and} \quad \widehat{X}_{D,n+1} = P_{\mathrm{LCS}(F_{D,n})}(X_{n+1}),$$

i.e., $\widehat{X}_{D,n+1}$ is the best linear functional predictor based on $F_{D,n}$ for $n \in \mathbb{N}$.

We formulate the Multivariate Innovations Algorithm for this setting.

Proposition 4.4 ([8], Proposition 11.4.2)**.** *Let $(X_n)_{n \in \mathbb{Z}}$ be a stationary functional linear process and $(X_{D,n})_{n \in \mathbb{Z}} = (P_{A_D} X_n)_{n \in \mathbb{Z}}$ as in (4.5). If C_{X_D} is invertible on A_D, then the best linear*

functional predictor $\widehat{X}_{D,n+1}$ *of* X_{n+1} *based on* $\mathrm{LCS}(F_{D,n})$ *can be computed by the following set of recursions:*

$$\widehat{X}_{D,1} = 0 \quad and \quad V_{D,0} = C_{X_D;0},$$

$$\widehat{X}_{D,n+1} = \sum_{i=1}^{n} \theta_{D,n,i}(X_{D,n+1-i} - \widehat{X}_{D,n+1-i}), \tag{4.10}$$

$$\theta_{D,n,n-i} = \Big(C_{X_D;n-i} - \sum_{j=0}^{i-1} \theta_{D,n,n-j}\, V_{D,j}\, \theta^*_{D,i,i-j}\Big)V_{D,i}^{-1}, \quad i = 1,\dots,n-1, \tag{4.11}$$

$$V_{D,n} = C_{X_{D,n+1}-\widehat{X}_{D,n+1}} = C_{X_D;0} - \sum_{j=0}^{n-1} \theta_{D,n,n-j}V_{D,j}\theta^*_{D,n,n-j}. \tag{4.12}$$

The recursions can be solved explicitly in the following order: $V_{D,0}, \theta_{D,1,1}, V_{D,1}, \theta_{D,2,2}, \theta_{D,2,1}\dots$. Thus we found a predictor, which is in contrast to $\widehat{X}_{n+1}$ from (3.1) easy to compute. However, since we are not using all available information, we loose predictive power. To evaluate this loss we bound the prediction error. We show that the error bound can be decomposed in two terms. One is due to the dimension reduction, and the other to the statistical prediction error of the finite dimensional model.

Theorem 4.5. *Let* $(X_n)_{n\in\mathbb{Z}}$ *be a stationary functional linear process with WN* $(\varepsilon_n)_{n\in\mathbb{Z}}$ *such that all eigenvalues of* C_ε *are positive. Assume furthermore that* C_X *is invertible on* A_D. *Recall the best linear functional predictor* $\widehat{X}_{n+1}$ *from Definition 3.3.*
(i) Then for all $n \in \mathbb{N}$ *the prediction error is bounded:*

$$E\|X_{n+1} - \widehat{X}_{n+1}\|^2 \le E\|X_{n+1} - \widehat{X}_{D,n+1}\|^2 = \sum_{i>D}\langle C_X\nu_i, \nu_i\rangle + \|V_{D,n}\|^2_{\mathcal{N}}. \tag{4.13}$$

(ii) If additionally $(X_n)_{n\in\mathbb{Z}}$ *is invertible, then*

$$\lim_{n\to\infty} E\|X_{n+1} - \widehat{X}_{D,n+1}\|^2 = \sum_{i>D}\langle C_X\nu_i, \nu_i\rangle + \|C_{\tilde{\varepsilon}_D}\|^2_{\mathcal{N}}.$$

Proof. (i) Since $\widehat{X}_{D,n+1} = P_{\mathrm{LCS}(F_{D,n})}(X_{n+1})$ and $\widehat{X}_{n+1} = P_{\mathrm{LCS}(F_n)}(X_{n+1})$, and since $\mathrm{LCS}(F_{D,n}) \subseteq \mathrm{LCS}(F_n)$, the first inequality follows immediately from the projection theorem. Furthermore, since $X_{n+1} - X_{D,n+1} \in A_D^{\perp}$ (the orthogonal complement of A_D) and $X_{D,n+1}, X_{D,n+1} - \widehat{X}_{D,n+1} \in A_D$, we have $\langle X_{n+1} - X_{D,n+1}, X_{D,n+1} - \widehat{X}_{D,n+1}\rangle = 0$. Therefore,

$$\mathbb{E}\|X_{n+1} - \widehat{X}_{D,n+1}\|^2 = \mathbb{E}\|X_{n+1} - X_{D,n+1} + X_{D,n+1} - \widehat{X}_{D,n+1}\|^2$$
$$= \mathbb{E}\|X_{n+1} - X_{D,n+1}\|^2 + \mathbb{E}\|X_{D,n+1} - \widehat{X}_{D,n+1}\|^2.$$

By (2.1) we have $\mathbb{E}\|X_{D,n+1} - \widehat{X}_{D,n+1}\|^2 = \|\mathbb{E}[(X_{D,n+1} - \widehat{X}_{D,n+1}) \otimes (X_{D,n+1} - \widehat{X}_{D,n+1})]\|_{\mathcal{N}}$, which is equal to $\|V_{D,n}\|_{\mathcal{N}}$ by (4.12). Furthermore,

$$\mathbb{E}\|X_{n+1} - X_{D,n+1}\|^2 = \mathbb{E}\Big\langle \sum_{i>D}\langle X_{n+1}, \nu_i\rangle\nu_i, \sum_{j>D}\langle X_{n+1}, \nu_j\rangle\nu_j\Big\rangle$$
$$= \sum_{i,j>D}\mathbb{E}\big\langle X_{n+1}\langle X_{n+1}, \nu_i\rangle, \nu_j\big\rangle\langle \nu_i, \nu_j\rangle$$
$$= \sum_{i>D}\langle C_X\nu_i, \nu_i\rangle.$$

(ii) By (i), what is left to show is that $\|V_{D,n}\|^2_{\mathcal{N}} \to \|C_{\tilde{\varepsilon}_D}\|^2_{\mathcal{N}}$ for $n \to \infty$. However, this is an immediate consequence of the Multivariate Innovations Algorithm under the assumption that $(X_{D,n})_{n\in\mathbb{Z}}$ is invertible (see Remark 4 in Chapter 11 of [8]). However, invertibility of $(X_{D,n})_{n\in\mathbb{Z}}$ is given by Proposition 4.3, which finishes the proof. $\qquad\square$

The above theorem states that for a stationary, invertible functional linear process, for increasing sample size the predictor restricted to the D-dimensional space performs arbitrarily well in the sense that in the limit only the statistical prediction error remains. However, our goal in (1.1) is not satisfied. The dimension reduction induces the additional error term $\sum_{i>D}\langle C_X(\nu_i), \nu_i\rangle$ independently of the sample size. If A_D is spanned by eigenfunctions of the covariance operator C_X with eigenvalues λ_i, the prediction error due to dimension reduction is $\sum_{i>D}\lambda_i$.

We now investigate the special case of functional moving average processes with finite order.

4.2 Prediction of $\mathrm{FMA}(q)$

$\mathrm{FMA}(q)$ processes for $q \in \mathbb{N}$ as in Definition 2.6 are an interesting and not very well studied class of functional linear processes. We start with the $\mathrm{FMA}(1)$ process as example.

Example 4.6. Consider a $\mathrm{FMA}(1)$ process $(X_n)_{n\in\mathbb{Z}}$ defined by

$$X_n = \psi\varepsilon_{n-1} + \varepsilon_n, \quad n \in \mathbb{Z},$$

for some $\psi \in \mathcal{L}$ and WN $(\varepsilon_n)_{n\in\mathbb{Z}}$. The representation of (4.7) reduces to

$$\mathbf{X}_{D,n} = \mathbf{\Psi}_D\mathbf{E}_{D,n-1} + \mathbf{\Psi}_D^\infty\mathbf{E}_{n-1}^\infty + \mathbf{E}_{D,n}, \quad n \in \mathbb{Z}.$$

As $\mathbf{X}_{D,n}$ depends on $\mathbf{E}_{D,n-1}$, $\mathbf{E}_{n-1}^\infty$ and $\mathbf{E}_{D,n}$, it is in general not a vector MA(1) process with WN $(\mathbf{E}_{D,n})_{n\in\mathbb{Z}}$. □

However, we can state the dynamics of a finite dimensional subprocess of a $\mathrm{FMA}(q)$ process.

Theorem 4.7. *Let $(X_n)_{n\in\mathbb{Z}}$ be a stationary $\mathrm{FMA}(q)$ process for $q \in \mathbb{N}$ and A_D be as in (4.1). Then $(X_{D,n})_{n\in\mathbb{Z}} = (P_{A_D}X_n)_{n\in\mathbb{Z}}$ as defined in (4.5) is a stationary $\mathrm{FMA}(q^*)$ process for $q^* \le q$ satisfying*

$$X_{D,n} = \sum_{j=1}^{q^*} \tilde{\psi}_{D,j}\tilde{\varepsilon}_{D,n-j} + \tilde{\varepsilon}_{D,n}, \qquad n \in \mathbb{Z}, \tag{4.14}$$

where $\tilde{\psi}_{D,j} \in \mathcal{L}$ for $j = 1,\ldots,q^*$ and $(\tilde{\varepsilon}_{D,n})_{n\in\mathbb{Z}}$ is WN. Moreover, $(\tilde{\varepsilon}_{D,n})_{n\in\mathbb{Z}}$ is isometrically isomorphic to $(\tilde{\mathbf{E}}_{D,n})_{n\in\mathbb{Z}}$ as defined in (4.8). If $q^* = 0$, then $(X_{D,n})_{n\in\mathbb{Z}}$ is WN.

Proof. By Proposition 4.3 $(X_{D,n})_{n\in\mathbb{Z}}$ is stationary. Furthermore, by (4.9) and Proposition 2.7 $C_{X_D;h} = P_{A_D}C_{X;h}P_{A_D} = 0$ for $h > q$, since $C_{X;h} = 0$ for $h > q$. Hence, again by Proposition 2.7 $(X_{D,n})_{n\in\mathbb{Z}}$ is a $\mathrm{FMA}(q^*)$ process, where q^* is the largest lag $j \le q$ such that $C_{X_D;j} = P_{A_D}C_{X;j}P_{A_D} \ne 0$. Thus, (4.14) holds for some linear operators $\tilde{\psi}_{D,j} \in \mathcal{L}$ and WN $(\tilde{\varepsilon}_{D,n})_{n\in\mathbb{Z}}$. The fact that $(\tilde{\varepsilon}_{D,n})_{n\in\mathbb{Z}}$ is isometrically isomorphic to $(\tilde{\mathbf{E}}_{D,n})_{n\in\mathbb{Z}}$ as in (4.8) is again a consequence of the Wold decomposition of $(X_{D,n})_{n\in\mathbb{Z}}$ and follows from the proof of Proposition 4.3. □

The fact that every subprocess of a $\mathrm{FMA}(q)$ is a $\mathrm{FMA}(q^*)$ with $q^* \le q$ simplifies the algorithm of Proposition 4.4. Since $C_{X_D;h} = 0$ for $|h| > q$ modifies (4.10)-(4.12) as follows: for $n > q^*$,

$$\widehat{X}_{D,n+1} = \sum_{i=1}^{q^*} \theta_{D,n,i}(X_{D,n+1-i} - \widehat{X}_{D,n+1-i})$$

$$\theta_{D,n,k} = \Big(C_{X_D;k} - \sum_{j=0}^{n-k-1} \theta_{D,n,n-j}\, V_{D,j}\, \theta_{D,n-k,n-k-j}^*\Big)V_{D,n-k}^{-1}, \quad k = 1,\ldots,q^*,$$

$$V_{D,n} = C_{X_{D,n+1}-\widehat{X}_{D,n+1}} = C_{X_d;0} - \sum_{j=1}^{q^*} \theta_{D,n,j}V_{D,n-j}\theta_{D,n,j}^*.$$

We now investigate the prediction error $E\|X_{n+1} - \widehat{X}_{D,n+1}\|^2$ of Theorem 4.5 for a functional linear process. For $D \to \infty$, obviously, $\sum_{i>D}\langle C_{X_0}(\nu_i), \nu_i\rangle \to 0$. However, the second term $\|V_{D,n}\|_{\mathcal{N}}$ on the rhs of (4.13) is not defined in the limit, since the inverse of $V_{D,j}$ in (4.11) is no longer bounded when $D \to \infty$. To see this, take e.g. $V_{D,0}$. By (4.12), since $\widehat{X}_{D,1} = 0$ and since $(X_{D,n})_{n\in\mathbb{Z}}$ is stationary,

$$V_{D,0} = C_{X_{D,1} - \widehat{X}_{D,1}} = C_{X_{D,1}} = C_{X_D}.$$

By (4.9) for $h = 0$ we find $C_{X_D} = P_{A_D} C_X P_{A_D}$, hence for all $x \in H$, $\|(C_X - C_{X_D})(x)\| \to 0$ for $D \to \infty$. But, since C_X is not invertible on the entire of H, neither is $\lim_{D\to\infty} C_{X_D}$. Therefore, $\lim_{D\to\infty} \widehat{X}_{D,n+1}$ is not defined.

To resolve this problem, we propose a tool used before in functional data analysis, for instance in [4] for the estimation of FAR(1). We increase the dimension D together with the sample size n by choosing $d_n := D(n)$ and $d_n \to \infty$ with $n \to \infty$. However, since the Innovations Algorithm is based on a recursion, it will always start with $V_{d_n,0} = C_{X_{d_n}}$, which again is not invertible on H for $d_n \to \infty$. For the Innovations Algorithm we increase D iteratively such that $V_{d_1,0}$ is inverted on say A_1, $V_{d_2,1}$ is inverted on $A_2, \ldots$ and so on. To quantify a convergence rate in Theorem 5.3 below we restrict ourselves to projections on eigenspaces of the covariance operator C_X of the underlying process.

5 Prediction based on projections of increasing subspaces of H

In this section we propose a functional version of the Innovations Algorithm. Starting with the same idea as in Section 4, we project the functional data on a finite dimensional space. However, we now let the dimension of the space on which we project depend on the sample size. More precisely, let $(X_n)_{n\in\mathbb{Z}}$ be a stationary functional linear process with covariance operator C_X. For some positive, increasing sequence $(d_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$ such that $d_n \to \infty$ with $n \to \infty$, we define

$$(\nu_i)_{i\in\mathbb{N}} \quad \text{and} \quad A_{d_n} = \overline{\mathrm{sp}}\{\nu_1, \ldots, \nu_{d_n}\}, \quad n \in \mathbb{N}, \tag{5.1}$$

where $(\nu_i)_{i\in\mathbb{N}}$ are now chosen as the eigenfunctions of C_X, and $(A_{d_n})_{n\in\mathbb{N}}$ is an increasing sequence of subspaces of H. Instead of applying the Innovations Algorithm to $(P_{A_d}X_1, \ldots, P_{A_d}X_n)$ as in Proposition 4.4, we apply it now to $(P_{A_{d_1}}X_1, \ldots, P_{A_{d_n}}X_n)$.

Proposition 5.1. *Let $(X_n)_{n\in\mathbb{Z}}$ be a stationary functional linear process with covariance operator C_X with eigenpairs $(\lambda_j, \nu_j)_{j\in\mathbb{N}}$, where $\lambda_j > 0$ for all $j \in \mathbb{N}$. Let $(d_n)_{n\in\mathbb{N}}$ be a positive sequence in $\mathbb{N}$ such that $d_n \uparrow \infty$ as $n \to \infty$. Define furthermore the best linear predictor of X_{n+1} based on $\mathrm{LCS}(F'_{d_n,n})$ for $n \in \mathbb{N}$ as*

$$F'_{d_n,n} = \{X_{d_1,1}, X_{d_2,2}, \ldots, X_{d_n,n}\} \quad \text{and} \quad \widehat{X}_{d_{n+1},n+1} = P_{\mathrm{LCS}(F'_{d_n,n})}(X_{n+1}). \tag{5.2}$$

Then $\widehat{X}_{d_{n+1},n+1}$ is given by the following set of recursions:

$$\widehat{X}_{d_1,1} = 0 \quad \text{and} \quad V_{d_1,0} = C_{X_{d_1}},$$

$$\widehat{X}_{d_{n+1},n+1} = \sum_{i=1}^{n} \theta_{d_{n-i+1},n,i}(X_{d_{n+1-i},n+1-i} - \widehat{X}_{d_{n+1-i},n+1-i}), \tag{5.3}$$

$$\theta_{d_{i+1},n,n-i} = \left(P_{A_{d_{n+1}}} C_{X;n-i} P_{A_{d_{i+1}}} - \sum_{j=0}^{i-1} \theta_{d_{j+1},n,n-j} \, V_{d_{j+1},j} \, \theta^*_{d_{j+1},i,i-j}\right) V^{-1}_{d_{i+1},i}, \quad i = 1, \ldots, n-1, \tag{5.4}$$

$$V_{d_{n+1},n} = C_{X_{d_{n+1},n+1} - \widehat{X}_{d_{n+1},n+1}} = C_{X_{d_{n+1}}} - \sum_{j=0}^{n-1} \theta_{d_{j+1},n,n-j} V_{d_{j+1},j} \theta^*_{d_{j+1},n,n-j}. \tag{5.5}$$

11

Proof. The proof is based on the proof of Proposition 11.4.2 in [8]. First notice that the representation

$$\widehat{X}_{d_{n+1},n+1} = \sum_{i=1}^{n} \theta_{d_{n-i+1},n,i}(X_{d_{n+1-i},n+1-i} - \widehat{X}_{d_{n+1-i},n+1-i}), \quad n \in \mathbb{N},$$

results from the definition of $\widehat{X}_{d_{n+1},n+1} = P_{\mathrm{LCS}(F'_{d_n,n})}(X_{n+1})$. Multiplying both sides of (5.3) with $\langle X_{d_{k+1},k+1} - \widehat{X}_{d_{k+1},k+1}, \cdot \rangle$ for $0 \le k \le n$ and taking expectations, we get

$$E\big[(X_{d_{k+1},k+1} - \widehat{X}_{d_{k+1},k+1}) \otimes \widehat{X}_{d_{n+1},n+1}\big]$$
$$= \sum_{i=1}^{n} \theta_{d_{n-i+1},n,i} E\big[(X_{d_{k+1},k+1} - \widehat{X}_{d_{k+1},k+1}) \otimes (X_{d_{n+1-i},n+1-i} - \widehat{X}_{d_{n+1-i},n+1-i})\big]$$
$$= \theta_{d_{k+1},n,n-k} E\big[(X_{d_{k+1},k+1} - \widehat{X}_{d_{k+1},k+1}) \otimes (X_{d_{k+1},k+1} - \widehat{X}_{d_{k+1},k+1})\big],$$

where we used that $E\langle X_{d_{n+1},n+1} - \widehat{X}_{d_{n+1},n+1}, X_{d_{k+1},k+1} - \widehat{X}_{d_{k+1},k+1}\rangle = 0$ for $k \ne n$. Now with the definition of $V_{d_{k+1},k}$ in (5.5),

$$E\big[(X_{d_{k+1},k+1} - \widehat{X}_{d_{k+1},k+1}) \otimes X_{d_{n+1},n+1}\big] = \theta_{d_{k+1},n,n-k} V_{d_{k+1},k}. \tag{5.6}$$

By representation (5.3) for $n = k$ and the fact that $V_{d_{k+1},k}$ is finite dimensional and therefore invertible, since all eigenvalues of C_X are positive,

$$\theta_{d_{k+1},n,n-k} = \Big(P_{A_{d_{n+1}}} C_{X;n-k} P_{A_{d_{k+1}}} - \sum_{i=1}^{k} E\big[(X_{d_i,i} - \widehat{X}_{d_i,i}) \otimes X_{d_{n+1},n+1}\big]\theta^*_{d_i,k,k-i-1}\Big) V^{-1}_{d_{k+1},k}.$$

However, with (5.6) the expectation on the right-hand-side can be replaced by $\theta_{d_i,n,n+1-i} V_{d_i,i-1}$, for $i = 1, \dots, k$, which leads to

$$\theta_{d_{k+1},n,n-k} = \Big(P_{A_{d_{n+1}}} C_{X;n-k} P_{A_{d_{k+1}}} - \sum_{i=1}^{k} \theta_{d_i,n,n+1-i} V_{d_i,i-1} \theta^*_{d_i,k,k-i-1}\Big) V^{-1}_{d_{k+1},k}.$$

Finally, the projection theorem gives

$$V_{d_{n+1},n} = C_{X_{d_{n+1},n+1} - \widehat{X}_{d_{n+1},n+1}} = C_{X_{d_{n+1}}} - C_{\widehat{X}_{d_{n+1},n+1}} = C_{X_{d_{n+1}}} - \sum_{j=0}^{n-1} \theta_{d_{j+1},n,n-j} V_{d_{j+1},j} \theta^*_{d_{j+1},n,n-j}.$$

$\square$

Remark 5.2. Notice that $X_{d_1,1}, X_{d_2,2}, \dots, X_{d_n,n}$ is not necessarily stationary. However, the recursions above can still be applied, since stationarity is not required for the application of the Innovations Algorithm in finite dimensions, see Proposition 11.4.2 in [8].

If $(X_n)_{n\in\mathbb{Z}}$ is invertible, we can derive asymptotics for $\widehat{X}_{d_{n+1},n+1}$ as $d_n \to \infty$ and $n \to \infty$.

Theorem 5.3. *Let $(X_n)_{n\in\mathbb{Z}}$ be a stationary, invertible functional linear process with WN $(\varepsilon_n)_{n\in\mathbb{Z}}$ such that all eigenvalues of C_ε are positive. Assume furthermore that all eigenvalues λ_j, $j \in \mathbb{N}$, of C_X are positive.*
(i) Let $m_n \to \infty$, $m_n < n$ and $m_n/n \to 0$ for $n \to \infty$ and $d_n \to \infty$ for $n \to \infty$ be two positive increasing sequences in $\mathbb{N}$. Then

$$E\|X_{n+1} - \widehat{X}_{d_{n+1},n+1} - \varepsilon_{n+1}\|^2 = O\Big(\sum_{j>m_n} \|\pi_j\|_{\mathcal{L}} + \sum_{j>d_{n-m_n}} \lambda_j\Big) \to 0, \quad n \to \infty. \tag{5.7}$$

12

(ii) Denote by $C_{\mathbf{X}_{d_n};h}$ the covariance matrix of the subprocess $(\mathbf{X}_{d_n})_{n\in\mathbb{Z}}$ as defined in Lemma 4.2. Then all eigenvalues of the spectral density matrix $f_{\mathbf{X}_{d_n}}[\omega] := \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} e^{-ih\omega} C_{\mathbf{X}_{d_n};h}$ for $-\pi < \omega \le \pi$ are positive. Denote by $\alpha_{d_n} > 0$ the infimum of these eigenvalues. If

$$\frac{1}{\alpha_{d_n}}\Big(\sum_{j>m_n}\|\pi_j\|_{\mathcal{L}} + \sum_{j>d_n-m_n}\lambda_j\Big) \to 0, \quad n \to \infty. \tag{5.8}$$

then for $i = 1, \ldots, n$ and for all $x \in H$,

$$\|(\theta_{d_n,n,i} - \gamma_i)(x)\| \to 0, \quad n \to \infty.$$

The proof of this Theorem is given in the next Section 6.

Remark 5.4. (a) Part (i) of Theorem 5.3 requires only that $d_n \to \infty$ as $n \to \infty$. No rate is required, and we do not impose any coupling condition of d_n with m_n. The theory would suggest to let d_n increase as fast as possible. In practice, when quantities such as the lagged covariance operators of the underlying process have to be estimated, the variance of the estimators of $P_{d_n}C_{X;h}P_{d_n}$ increases with d_n. In fact, for instance, for the estimation of $\theta_{d_1,1,1}$ the statistician is faced with the inversion of $P_{d_1}C_X P_{d_1}$. Small errors in the estimation of small empirical eigenvalues of $P_{d_1}C_X P_{d_1}$ may have severe consequences for the estimation of $\theta_{d_1,1,1}$. This suggests a conservative choice for d_n. The problem is similar to the choice of k_n in Chapter 9.2 of [4] concerned with the estimation of the autoregressive parameter operator in a FAR(1). The authors propose to choose k_n based on validation of the empirical prediction error.

(b) The choice of m_n in (5.7) allows us to calibrate two error terms: under the restriction that $m_n/n \to 0$, choosing a larger m_n increases $\sum_{j>d_n-m_n}\lambda_j$, the error caused by dimension reduction. Choosing a smaller m_n will on the other hand increase $\sum_{j>m_n}\|\pi_j\|$.

6 Proofs

Before presenting a proof of Theorem 5.3 we give some notation and auxiliary results. Recall that throughout I_H denotes the identity operator on H. We also recall the notation and results provided in Section 2, which we shall use below without specific referencing.

Let $(X_n)_{n\in\mathbb{Z}}$ be a stationary functional linear process. Then for $n \in \mathbb{N}$ define the covariance operator of the vector $(X_n, \ldots, X_1)$ by

$$\Gamma_n := \begin{pmatrix} \mathbb{E}[X_n \otimes X_n] & \mathbb{E}[X_n \otimes X_{n-1}] & \ldots & \mathbb{E}[X_n \otimes X_1] \\ \mathbb{E}[X_{n-1} \otimes X_1] & \mathbb{E}[X_{n-1} \otimes X_{n-1}] & \ldots & \vdots \\ \vdots & & \ddots & \\ \mathbb{E}[X_1 \otimes X_n] & \ldots & & \mathbb{E}[X_1 \otimes X_1] \end{pmatrix} = \begin{pmatrix} C_X & C_{X;1} & \ldots & C_{X;n-1} \\ C_{X;-1} & C_X & \ldots & \vdots \\ \vdots & & \ddots & \\ C_{X;-(n-1)} & \ldots & & C_X \end{pmatrix}, \tag{6.1}$$

i.e., Γ_n is an operator acting on H^n, where H^n is the Cartesian product of n copies of H. Recall that H^n is again a Hilbert space, when equipped with the scalar product

$$\langle x, y \rangle_n = \sum_{i=1}^{n} \langle x_i, y_i \rangle$$

(see [4], Section 5 for details). As the covariance operator of $(X_n, X_{n-1}, \ldots, X_1)$, Γ_n is self-adjoint, nuclear, and has the spectral representation (cf. Theorem 5.1 in [11])

$$\Gamma_n = \sum_{j=1}^{\infty} \lambda_j^{(n)} \nu_j^{(n)} \otimes \nu_j^{(n)}, \quad n \in \mathbb{N},$$

with eigenpairs $(\lambda_j^{(n)}, \nu_j^{(n)})_{j \in \mathbb{N}}$.

Furthermore, define the operators $P_{(d_n)}$ and P_D acting on H^n by

$$P_{(d_n)} = \mathrm{diag}\left(P_{A_{d_n}}, P_{A_{d_{n-1}}}, \ldots, P_{A_{d_1}}\right) \quad \text{and} \quad P_D = \mathrm{diag}\left(P_{A_D}, P_{A_D}, \ldots, P_{A_D}\right), \qquad (6.2)$$

then

$$\Gamma_{(d_n),n} := P_{(d_n)} \Gamma_n P_{(d_n)} \quad \text{and} \quad \Gamma_{D,n} := P_D \Gamma_n P_D.$$

Note that $\Gamma_{(d_n),n}$ is in fact the covariance operator of $(X_{d_n,n}, \ldots, X_{d_1,1})$ and has rank $k_n := \sum_{i=1}^n d_i$, whereas $\Gamma_{D,n}$ is the covariance operator of $(X_{D,n}, \ldots, X_{D,1})$ and has rank $D \cdot n$. The operators $\Gamma_{(d_n),n}$ and $\Gamma_{d_n,n}$ are therefore self-adjoint nuclear operators with spectral representations

$$\Gamma_{(d_n),n} = \sum_{j=1}^{k_n} \lambda_{(d_n),j}^{(n)} e_{(d_n),j}^{(n)} \otimes e_{(d_n),j}^{(n)} \quad \text{and} \quad \Gamma_{d_n,n} = \sum_{j=1}^{d_n \cdot n} \lambda_{d_n,j}^{(n)} e_{d_n,j}^{(n)} \otimes e_{d_n,j}^{(n)}. \qquad (6.3)$$

We need the following auxiliary results.

Lemma 6.1 (Theorem 1.2 in [20])**.** *Let $(\mathbf{X}_{D,n})_{n \in \mathbb{Z}}$ be a D-variate stationary, invertible linear process satisfying*

$$\mathbf{X}_{D,n} = \sum_{i=1}^{\infty} \mathbf{\Psi}_i \mathbf{E}_{n-i} + \mathbf{E}_n, \quad n \in \mathbb{Z},$$

with $\sum_{i=1}^{\infty} \|\mathbf{\Psi}_i\|_2 < \infty$ ($\|\cdot\|_2$ denotes the Euclidean matrix norm) and WN $(\mathbf{E}_{D,n})_{n \in \mathbb{Z}}$ in $L_{\mathbb{R}^D}^2$ with non-singular covariance matrix $C_{\mathbf{E}_D}$. Let $C_{\mathbf{X}_D}$ be the covariance matrix of $(\mathbf{X}_{D,n})_{n \in \mathbb{Z}}$. Then the spectral density matrix $f_{\mathbf{X}_D}[\omega] := \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} e^{-ih\omega} C_{\mathbf{X}_{D;h}}$ for $-\pi < \omega \leq \pi$ has only positive eigenvalues. Let α_D be their infimum. Then the eigenvalues $(\lambda_i^{(n)})_{i=1,\ldots,D \cdot n}$ of $\Gamma_{D,n}$ as in (6.1) are bounded below as follows:

$$0 < 2\pi \alpha_D \leq \lambda_{D \cdot n}^{(n)} \leq \cdots \leq \lambda_1^{(n)}.$$

The following is a consequence of the Cauchy-Schwarz inequality.

Lemma 6.2. *For $j, l \in \mathbb{N}$ let (λ_j, ν_j) and (λ_l, ν_l) be eigenpairs of C_X. Then for $i, k \in \mathbb{Z}$,*

$$\langle C_{X;i-k}\nu_j, \nu_l \rangle \leq \lambda_j^{1/2} \lambda_l^{1/2}. \qquad (6.4)$$

Proof. With the definition of the lagged covariance operators in (2.4) and then the Cauchy-Schwarz inequality, we get by stationarity of $(X_n)_{n \in \mathbb{Z}}$

$$\langle C_{X;i-k}\nu_j, \nu_l \rangle = \langle \mathbb{E}\langle X_i, \nu_j \rangle X_k, \nu_l \rangle = \mathbb{E}[\langle X_i, \nu_j \rangle \langle X_k, \nu_l \rangle]$$

$$\leq \left(\mathbb{E}\langle X_i, \nu_j \rangle^2\right)^{1/2} \left(\mathbb{E}\langle X_k, \nu_l \rangle^2\right)^{1/2}.$$

We find $\mathbb{E}\langle X_i, \nu_j \rangle^2 = \mathbb{E}\langle \langle X_i, \nu_j \rangle X_i, \nu_j \rangle = \langle C_X \nu_j, \nu_j \rangle = \lambda_j$, which implies (6.4). $\qquad \square$

So far we only considered the real Hilbert space $H = L^2([0,1])$. There is a natural extension to the complex Hilbert space by defining the scalar product $\langle x, y \rangle = \int_0^1 x(t)\bar{y}(t)dt$ for complex valued functions $x, y : [0,1] \to \mathbb{C}$. As in Section 7.2 of [4], for $(\psi_j)_{j \in \mathbb{N}} \subset \mathcal{L}$ we define the *complex operators*

$$A[z] := \sum_{j=0}^{\infty} z^j \psi_j, \quad z \in \mathbb{C}, \qquad (6.5)$$

such that the series converges in the operator norm. We need some methodology on frequency analysis of functional time series, recently studied in [23]. The functional discrete Fourier transform of $(X_1, \ldots, X_n)$ is defined by

$$S_n(\omega) = \sum_{j=1}^{n} X_j e^{-ij\omega}, \quad \omega \in (-\pi, \pi].$$

By Theorem 4 of [9], for all $\omega \in (-\pi, \pi]$, if $(X_n)_{n \in \mathbb{Z}}$ is a linear process with $\sum_{i=1}^{\infty} \|\psi_j\|_{\mathcal{L}} < \infty$, then $\frac{1}{\sqrt{n}} S_n(\omega)$ converges in distribution as $n \to \infty$ to a complex Gaussian random element with covariance operator

$$2\pi \mathcal{F}_X[\omega] := \sum_{h \in \mathbb{Z}} C_{X;h} e^{-ih\omega}.$$

The *spectral density operator* $\mathcal{F}_X[\omega]$ of $(X_n)_{n \in \mathbb{Z}}$ is non-negative, self-adjoint and nuclear (see Proposition 2.1 in [23]).

Theorem 1 and 4 of [9] infer the following duality between $C_{X;h}$ and $\mathcal{F}_X[\omega]$, with $A[z]$ as in (6.5) and adjoint $A[z]^*$:

$$C_{X;h} = \int_{-\pi}^{\pi} \mathcal{F}_X[\omega] e^{ih\omega} d\omega, \quad h \in \mathbb{Z} \quad \text{and} \quad \mathcal{F}_X[\omega] = \frac{1}{2\pi} A[e^{-i\omega}] C_\varepsilon A[e^{-i\omega}]^*, \quad \omega \in (-\pi, \pi]. \quad (6.6)$$

The following Lemma is needed for the subsequent proofs, but may also be of interest by itself.

Lemma 6.3. *Let $(X_n)_{n \in \mathbb{Z}}$ be a stationary, invertible functional linear process with WN $(\varepsilon_n)_{n \in \mathbb{Z}}$, such that all eigenvalues of C_ε are positive. Then for all $\omega \in (-\pi, \pi]$ the spectral density operator $\mathcal{F}_X[\omega]$ has only positive eigenvalues.*

Proof. The proof is an extension of the proof of Theorem 3.1 in [22] to the infinite dimensional setting. Define for $A[z]$ as in (6.5) and $(\pi_i)_{i \in \mathbb{N}}$ as in (2.5)

$$P[z] := \sum_{j=0}^{\infty} z^j \pi_j \quad \text{and} \quad D[z] := P[z] A[z], \quad z \in \mathbb{C}.$$

Since $A[z]$ and $P[z]$ are power series, also $D[z]$ can be represented by

$$D[z] = \sum_{j=0}^{\infty} z^j \delta_j, \quad z \in \mathbb{C},$$

for $\delta_j \in \mathcal{L}$. Let B be the backshift operator. Then $X_n = A[B]\varepsilon_n$ and $\varepsilon_n = P[B]X_n$; in particular,

$$\varepsilon_n = P[B]X_n = P[B]A[B]\varepsilon_n = D[B]\varepsilon_n, \quad n \in \mathbb{Z}. \quad (6.7)$$

Since all eigenvalues of C_ε are positive, by equating the coefficients in (6.7), $D[z] = I_H$ for all $z \in \mathbb{C}$.

Assume that there exists some non-zero $v \in H$ such that $\mathcal{F}_X[\omega](v) = 0$. Then by (6.6),

$$\frac{1}{2\pi} A[e^{-i\omega}] C_\varepsilon A[e^{-i\omega}]^*(v) = 0.$$

But since all eigenvalues of C_ε are positive, there exists some non-zero $u \in H$ such that $A[e^{i\omega}](u) = 0$. However, since $D[z] = P[z]A[z] = I_H$ for all $z \in \mathbb{C}$, this is a contradiction, and $\mathcal{F}_X[\omega]$ can only have positive eigenvalues for all $\omega \in (-\pi, \pi]$. $\quad\square$

6.1 Proof of Proposition 4.3

Stationarity of $(X_{D,n})_{n\in\mathbb{Z}}$ follows immediately from stationarity of $(X_n)_{n\in\mathbb{Z}}$, since P_{A_D} is a linear shift-invariant transformation. The functional Wold decomposition (e.g. Definition 3.1 in [5]) gives a representation of $(X_{D,n})_{n\in\mathbb{Z}}$ as a linear process with WN, say $(\tilde{\varepsilon}_n)_{n\in\mathbb{Z}}$ in L_H^2. By Lemma 4.2, $(X_{D,n})_{n\in\mathbb{Z}}$ is isometrically isomorphic to the vector process $(\mathbf{X}_{D,n})_{n\in\mathbb{Z}}$ as in (4.6). Analogously, $(\tilde{\mathbf{E}}_{D,n})_{n\in\mathbb{Z}}$ defined by $\tilde{\mathbf{E}}_{D,n} := (\langle\tilde{\varepsilon}_{D,n},\nu_1\rangle,\dots,\langle\tilde{\varepsilon}_{D,n},\nu_D\rangle)^\top$ is isometrically isomorphic to $(\tilde{\varepsilon}_{D,n})_{n\in\mathbb{Z}}$. We give a representation of $(\tilde{\mathbf{E}}_{D,n})_{n\in\mathbb{Z}}$.

Define $\mathcal{M}_{D,n} = \overline{\mathrm{sp}}\{\mathbf{X}_{D,t}, -\infty < t \le n\}$. Then from the multivariate Wold decomposition the WN of $(\mathbf{X}_{D,n})_{n\in\mathbb{Z}}$ in $L_{\mathbb{R}^D}^2$ is defined by

$$\tilde{\mathbf{E}}_{D,n} = \mathbf{X}_{D,n} - P_{\mathcal{M}_{D,n-1}}(\mathbf{X}_{D,n}), \quad n \in \mathbb{Z}. \tag{6.8}$$

Now recall (4.7) in the following form

$$\mathbf{X}_{D,n} = \mathbf{E}_{D,n} + \sum_{j=1}^{\infty} \mathbf{\Psi}_{D,j}\mathbf{E}_{D,n-j} + \mathbf{\Psi}_{D,1}^{\infty}\mathbf{E}_{n-1}^{\infty} + \sum_{j=2}^{\infty} \mathbf{\Psi}_{D,j}^{\infty}\mathbf{E}_{n-j}^{\infty}, \quad n \in \mathbb{Z},$$

and we apply the projection operator to all terms. Firstly, $P_{\mathcal{M}_{D,n-1}}(\mathbf{E}_{D,n}) = 0$, and $\mathbf{E}_{D,n-j}$ and $\mathbf{E}_{n-j-1}^{\infty}$ belong to $\mathcal{M}_{D,n-1}$ for all $j \ge 1$. Hence,

$$P_{\mathcal{M}_{D,n-1}}(\mathbf{X}_{D,n}) = \sum_{i=1}^{\infty} \mathbf{\Psi}_{D,j}\mathbf{E}_{D,n-j} + \sum_{i=2}^{\infty} \mathbf{\Psi}_{D,j}^{\infty}\mathbf{E}_{n-j}^{\infty} + \mathbf{\Psi}_{D,1}P_{\mathcal{M}_{D,n-1}}(\mathbf{E}_{n-1}^{\infty}), \quad n \in \mathbb{Z},$$

which together with (6.8) implies (4.8).

We now show that $(\mathbf{X}_{D,n})_{n\in\mathbb{Z}}$ is invertible. The Wold decomposition gives the following representation

$$\mathbf{X}_{D,n} = \sum_{i=1}^{\infty} \tilde{\mathbf{\Psi}}_{D,i}(\tilde{\mathbf{E}}_{D,n-i}) + \tilde{\mathbf{E}}_{D,n}, \quad n \in \mathbb{Z} \tag{6.9}$$

for appropriate $\tilde{\mathbf{\Psi}}_{D,i}$ and innovation process as in (6.8). Theorem 1 of [22] gives conditions for the invertibility of the stationary D-variate linear process $(\mathbf{X}_{D,n})_{n\in\mathbb{Z}}$ satisfying (6.9).

We verify these conditions one by one.

(1) We start by showing that for all $\omega \in (-\pi,\pi]$ the matrix $\mathcal{F}_{\mathbf{X}_D}[\omega]$ is invertible, equivalently, $\langle\mathcal{F}_{\mathbf{X}_D}[\omega]\mathbf{x},\mathbf{x}\rangle_{\mathbb{R}^D} > 0$ for all non-zero $\mathbf{x} \in \mathbb{R}^D$. By the isometric isomorphy between $\mathbb{R}^D$ and A_D from Lemma 4.2 we have

$$\langle\mathcal{F}_{\mathbf{X}_D}[\omega]\mathbf{x},\mathbf{x}\rangle_{\mathbb{R}^D} = \langle\mathcal{F}_{X_D}[\omega]x,x\rangle.$$

By (4.9) the spectral density operator $\mathcal{F}_{X_D}[\omega]$ of $(X_{D,n})_{n\in\mathbb{Z}}$ satisfies

$$\mathcal{F}_{X_D}[\omega] = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}}C_{X_D;h}e^{-ih\omega} = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}}P_{A_D}C_{X;h}P_{A_D}e^{-ih\omega}$$

$$= P_{A_D}\Big(\frac{1}{2\pi}\sum_{h\in\mathbb{Z}}C_{X;h}e^{-ih\omega}\Big)P_{A_D} = P_{A_D}\mathcal{F}_X[\omega]P_{A_D}. \tag{6.10}$$

However, since $(X_n)_{n\in\mathbb{Z}}$ is invertible, by Lemma 6.3 all eigenvalues of $\mathcal{F}_X[\omega]$ are positive for all $\omega \in (-\pi,\pi]$. Using first (6.10), then that $x \in A_D$ and finally that all eigenvalues of $\mathcal{F}_X[\omega]$ are positive, we get

$$\langle\mathcal{F}_{X_D}[\omega]x,x\rangle = \langle P_{A_D}\mathcal{F}_X[\omega]P_{A_D}x,x\rangle = \langle\mathcal{F}_X[\omega]x,x\rangle > 0.$$

16

Hence, $\langle \mathcal{F}_{\mathbf{X}_D}[\omega]\mathbf{x}, \mathbf{x}\rangle_{\mathbb{R}^D} > 0$ and thus $\mathcal{F}_{X_D}[\omega]$ is invertible.

(2)   We next show that the covariance matrix $C_{\tilde{\mathbf{E}}_D}$ of $(\tilde{\mathbf{E}}_{D,n})_{n\in\mathbb{Z}}$ as in (4.8) is invertible. Since $\mathbf{E}_{D,n}$ and $\boldsymbol{\Delta}_{D,n-1}$ from (4.8) are uncorrelated, $C_{\tilde{\mathbf{E}}_D} = C_{\mathbf{E}_D} + C_{\boldsymbol{\Delta}_D}$. All eigenvalues of C_ε are positive by assumption. For all $x \in A_D$ we get $\langle x, C_\varepsilon x\rangle = \langle \mathbf{x}, C_{\mathbf{E}_D}\mathbf{x}\rangle_{\mathbb{R}^d}$ where x and $\mathbf{x}$ are related by the isometric isomorphism T of Lemma 4.2. With the characterization of the eigenvalues of a self-adjoint operator via the Rayleigh quotient as in Theorem 4.2.7 in [15], all eigenvalues of $C_{\mathbf{E}_D}$ are positive. Therefore, all eigenvalues of $C_{\tilde{\mathbf{E}}_D} = C_{\mathbf{E}_D} + C_{\boldsymbol{\Delta}_D}$ are positive, and $C_{\tilde{\mathbf{E}}_D}$ is invertible.

(3)   Finally, summability in Euclidean matrix norm of the matrices $\tilde{\boldsymbol{\Psi}}_{D,i}$ over $i \in \mathbb{N}$ follows from the properties of the Wold decomposition (see e.g. Theorem 5.7.1 in [8]) and from the summability of $\|\psi_i\|_{\mathcal{L}}$ over $i \in \mathbb{N}$.

Therefore, all conditions of Theorem 1 of [22] are satisfied and $(\mathbf{X}_{D,n})_{n\in\mathbb{Z}}$ is invertible. $\square$

6.2   Proof of Theorem 5.3 (i)

First note that by the projection theorem (e.g. Theorem 2.3.1 in [8]),

$$\mathbb{E}\|X_{n+1} - \widehat{X}_{d_{n+1},n+1}\|^2 \le \mathbb{E}\|X_{n+1} - \sum_{i=1}^n \eta_i X_{d_{n+1-i},n+1-i}\|^2, \quad n \in \mathbb{N}, \tag{6.11}$$

for all $\eta_i \in \mathcal{L}$, $i = 1, \ldots, n$. Hence, (6.11) holds in particular for $\eta_i = \pi_i$ for $i = 1, \ldots, n$, where π_i are the operators in the inverse representation of $(X_n)_{n\in\mathbb{Z}}$ of (2.5). Furthermore, by the orthogonality of ε_{n+1} and X_k for $k < n + 1$ and $n \in \mathbb{N}$,

$$\mathbb{E}\|X_{n+1} - \widehat{X}_{d_{n+1},n+1}\|^2 = \mathbb{E}\|X_{n+1} - \widehat{X}_{d_{n+1},n+1} - \varepsilon_{n+1}\|^2 + \mathbb{E}\|\varepsilon_{n+1}\|^2$$

Now (6.11) with $\eta_i = \pi_i$ and then the invertibility of $(X_n)_{n\in\mathbb{Z}}$ yield

$$\mathbb{E}\|X_{n+1} - \widehat{X}_{d_{n+1},n+1} - \varepsilon_{n+1}\|^2 \le \mathbb{E}\|X_{n+1} - \sum_{i=1}^n \pi_i X_{d_{n+1-i},n+1-i}\|^2 - \mathbb{E}\|\varepsilon_{n+1}\|^2$$

$$= \mathbb{E}\|\sum_{i=1}^\infty \pi_i X_{n+1-i} + \varepsilon_{n+1} - \sum_{i=1}^n \pi_i X_{d_{n+1-i},n+1-i}\|^2 - \mathbb{E}\|\varepsilon_{n+1}\|^2$$

$$= \mathbb{E}\|\sum_{i=1}^n \pi_i(X_{n+1-i} - X_{d_{n+1-i},n+1-i}) + \varepsilon_{n+1} + \sum_{i>n} \pi_i X_{n+1-i}\|^2 - \mathbb{E}\|\varepsilon_{n+1}\|^2.$$

Again by the orthogonality of ε_{n+1} and X_k, for $k < n + 1$, since $X_{d_n,n} = P_{A_{d_n}} X_n$, and then using that for $X, Y \in L_H^2$, $\mathbb{E}\|X + Y\|^2 \le 2\mathbb{E}\|X\|^2 + 2\mathbb{E}\|Y\|^2$, we get

$$\mathbb{E}\|X_{n+1} - \widehat{X}_{d_{n+1},n+1} - \varepsilon_{n+1}\|^2 \le \mathbb{E}\|\sum_{i=1}^n \pi_i(I_H - P_{A_{d_{n+1-i}}})X_{n+1-i} + \sum_{i>n} \pi_i X_{n+1-i}\|^2$$

$$\le 2\mathbb{E}\|\sum_{i=1}^n \pi_i(I_H - P_{A_{d_{n+1-i}}})X_{n+1-i}\|^2 + 2\mathbb{E}\|\sum_{i>n} \pi_i X_{n+1-i}\|^2 \tag{6.12}$$

$$=: 2J_1 + 2J_2.$$

We consider the two terms in (6.12) separately. From (2.1) we get for the first term in (6.12)

$$J_1 = \left\|\mathbb{E}\Big[\sum_{i=1}^n \pi_i(I_H - P_{A_{d_{n+1-i}}})X_{n+1-i} \otimes \sum_{i=1}^n \pi_i(I_H - P_{A_{d_{n+1-i}}})X_{n+1-i}\Big]\right\|_{\mathcal{N}}.$$

Using the triangle inequality together with properties of the nuclear operator norm given in Section 2, and then the definition of $C_{X;h}$ in (2.4),

$$J_1 \le \sum_{i,j=1}^n \|\pi_i\|_{\mathcal{L}} \|\pi_j\|_{\mathcal{L}} \big\| \mathbb{E}\big[(I_H - P_{A_{d_{n+1-i}}}) X_{n+1-i} \otimes (I_H - P_{A_{d_{n+1-j}}}) X_{n+1-j} \big] \big\|_{\mathcal{N}}$$

$$= \sum_{i,j=1}^n \|\pi_i\|_{\mathcal{L}} \|\pi_j\|_{\mathcal{L}} \big\| (I_H - P_{A_{d_{n+1-i}}}) C_{X;i-j} (I_H - P_{A_{d_{n+1-j}}}) \big\|_{\mathcal{N}} := \sum_{i,j=1}^n \|\pi_i\|_{\mathcal{L}} \|\pi_j\|_{\mathcal{L}} K(i,j). \quad (6.13)$$

By the definition of A_d in (5.1) and, since by (4.5) we have $(I_H - P_{A_{d_i}}) = \sum_{l>d_i} \nu_l \otimes \nu_l$,

$$K(i,j) = \Big\| \Big(\sum_{l'>d_{n+1-i}} \nu_{l'} \otimes \nu_{l'} \Big) C_{X;i-j} \Big(\sum_{l>d_{n+1-j}} \nu_l \otimes \nu_l \Big) \Big\|_{\mathcal{N}} = \Big\| \sum_{l'>d_{n+1-i}} \sum_{l>d_{n+1-j}} \langle C_{X;i-j}(\nu_l), \nu_{l'} \rangle \nu_l \otimes \nu_{l'} \Big\|_{\mathcal{N}}.$$

With Lemma 6.2, the definition of the nuclear norm given in Section 2 and the orthogonality of the $(\nu_i)_{i\in\mathbb{N}}$, we get

$$K(i,j) \le \Big\| \sum_{l'>d_{n+1-i}} \sum_{l>d_{n+1-j}} \lambda_l^{1/2} \lambda_{l'}^{1/2} \nu_l \otimes \nu_{l'} \Big\|_{\mathcal{N}}$$

$$= \sum_{k=1}^{\infty} \Big\langle \sum_{l'>d_{n+1-i}} \sum_{l>d_{n+1-j}} \lambda_l^{1/2} \lambda_{l'}^{1/2} \nu_l \otimes \nu_{l'}(\nu_k), \nu_k \Big\rangle$$

$$= \sum_{k>\max(d_{n+1-j},d_{n+1-i})} \lambda_k \le \sum_{k>d_{n+1-j}} \lambda_k. \quad (6.14)$$

Plugging (6.14) into (6.13), and recalling that $\sum_{i=1}^{\infty} \|\pi_i\|_{\mathcal{L}} =: M_1 < \infty$, we conclude

$$J_2 \le M_1 \sum_{j=1}^n \|\pi_j\|_{\mathcal{L}} \sum_{l>d_{n+1-j}} \lambda_l. \quad (6.15)$$

Now for some $m_n < n$,

$$\sum_{j=1}^n \|\pi_j\|_{\mathcal{L}} \sum_{l>d_{n+1-j}} \lambda_l = \sum_{j=1}^{m_n} \|\pi_j\|_{\mathcal{L}} \sum_{l>d_{n+1-j}} \lambda_l + \sum_{j=m_n}^n \|\pi_j\|_{\mathcal{L}} \sum_{l>d_{n+1-j}} \lambda_l. \quad (6.16)$$

Since $\sum_{j=1}^{m_n} \|\pi_j\|_{\mathcal{L}} \le \sum_{j=1}^{\infty} \|\pi_j\|_{\mathcal{L}} = M_1 < \infty$, the first term on the rhs of (6.16) can be bounded by

$$\sum_{j=1}^{m_n} \|\pi_j\|_{\mathcal{L}} \sum_{l>d_{n+1-j}} \lambda_l \le M_1 \sum_{l>d_{n+1-m_n}} \lambda_l. \quad (6.17)$$

Furthermore, since $\sum_{l>d_{n+1-j}} \lambda_l \le \sum_{l=1}^{\infty} \lambda_l = \|C_X\|_{\mathcal{N}} < \infty$, the second term of the rhs in (6.16) can be bounded by

$$\sum_{j=m_n}^n \|\pi_j\|_{\mathcal{L}} \sum_{l>d_{n+1-j}} \lambda_l \le \|C_X\|_{\mathcal{N}} \sum_{j=m_n}^n \|\pi_j\|_{\mathcal{L}}. \quad (6.18)$$

Hence, from (6.15) together with (6.16), (6.17) and (6.18) we obtain

$$J_1 = O\Big(\sum_{j=m_n}^n \|\pi_j\|_{\mathcal{L}} + \sum_{l>d_{n+1-m_n}} \lambda_l \Big). \quad (6.19)$$

Concerning J_2, the second term of (6.12) with (2.2), and then the definition of $C_{X;h}$ in (2.4) yield

$$J_2 = \mathbb{E}\Big\| \sum_{i>n} \pi_i X_{n+1-i} \Big\|^2 = \Big\| \mathbb{E}\Big[\sum_{i>n} \pi_i X_{n+1-i} \otimes \sum_{j>n} \pi_j X_{n+1-j} \Big] \Big\|_{\mathcal{N}}$$

$$= \Big\| \sum_{i,j>n} \pi_i C_{X;i-j} \pi_j^* \Big\|_{\mathcal{N}} \le \sum_{i,j>n} \|\pi_i\|_{\mathcal{L}} \|\pi_j\|_{\mathcal{L}} \|C_{X;i-j}\|_{\mathcal{N}}.$$

Since $C_{X;i-j} \in \mathcal{N}$ for all $i, j \in \mathbb{N}$, $\|C_{X;i-j}\|_{\mathcal{N}} =: M_2 < \infty$, and for some $m_n < n$,

$$J_2 \leq M_2 \Big(\sum_{i>n} \|\pi_i\|_{\mathcal{L}} \Big)^2 = O\Big(\sum_{i>m_n} \|\pi_i\|_{\mathcal{L}} \Big). \tag{6.20}$$

Finally the combination of (6.12), (6.19) and (6.20) yields assertion (i). $\qquad\square$

6.3 Proof of Theorem 5.3 (ii)

Note first that by the projection theorem there is an equivalent representation of $\widehat{X}_{d_{n+1},n+1}$ to (5.3) given by

$$\widehat{X}_{d_{n+1},n+1} = P_{\mathrm{LCS}(F'_{d_n,n})}(X_{n+1}) = \sum_{i=1}^{n} \beta_{d_{n+1-i},n,i} X_{d_{n+1-i},n+1-i} \tag{6.21}$$

for $F'_{d_n,n}$ as in (5.2) and $\beta_{d_{n+1-i},n,i} \in \mathcal{L}$ for $i = 1, \ldots, n$. Furthermore, for $k = 1, \ldots, n$, we define the best linear predictor of X_{n+1} based on $F'_{d_n,n}(k) = \{X_{d_{n+1-k},n+1-k}, X_{d_{n-k+2},n+2-k}, \ldots, X_{d_n,n}\}$ by

$$\widehat{X}_{d_{n+1},n+1}(k) = P_{\mathrm{LCS}(F'_{d_n,n}(k))}(X_{n+1}) = \sum_{i=1}^{k} \beta_{d_{n+1-i},k,i} X_{d_{n+1-i},n+1-i}. \tag{6.22}$$

We start with the following Proposition, which is an infinite-dimensional extension to Proposition 2.2 in [21].

Proposition 6.4. *Under the assumptions of Theorem 5.3 the following assertions hold:*
(i) The operators $\beta_{d_{n+1-i},n,i}$ from (6.21) and $\theta_{d_{n+1-i},n,i}$ from (5.3) are for $n \in \mathbb{N}$ related by

$$\theta_{d_{n+1-i},n,i} = \sum_{j=1}^{i} \beta_{d_{n+1-j},n,j} \theta_{d_{n+1-i},n-j,i-j}, \quad i = 1, \ldots, n. \tag{6.23}$$

Furthermore, for every $i, j \in \mathbb{N}$ and $x \in H$, as $n \to \infty$,
(ii) $\big\|(\beta_{d_{n+1-i},n,i} - \pi_i)(x)\big\| \to 0$,
(iii) $\big\|(\beta_{d_{n+1-i},n,i} - \beta_{d_{n+1-i-j},n-j,i})(x)\big\| \to 0$,
(iv) $\big\|(\theta_{d_{n+1-i},n,i} - \theta_{d_{n+1-i-j},n-j,i})(x)\big\| \to 0$.

Proof. **(i)** Set $\theta_{d_{n+1},n,0} := I_H$. By adding the term $\theta_{d_{n+1},n,0}(X_{d_{n+1},n+1} - \widehat{X}_{d_{n+1},n+1})$ to both sides of (5.3), we get

$$X_{d_{n+1},n+1} = \sum_{j=0}^{n} \theta_{d_{n+1-j},n,j}(X_{d_{n+1-j},n+1-j} - \widehat{X}_{d_{n+1-j},n+1-j}), \quad n \in \mathbb{N}.$$

Plugging this representation of $X_{d_{n+1-i},n+1-i}$ into (6.21) for $i = 1, \ldots, n$ yields

$$\widehat{X}_{d_{n+1},n+1} = \sum_{i=1}^{n} \beta_{d_{n+1-i},n,i} \Big(\sum_{j=0}^{n-i} \theta_{d_{n+1-i-j},n-i,j}(X_{d_{n+1-i-j},n+1-i-j} - \widehat{X}_{d_{n+1-i-j},n+1-i-j}) \Big).$$

Equating the coefficients of the innovations $(X_{d_{n+1-i},n+1-i} - \widehat{X}_{d_{n+1-i},n+1-i})$ with the innovation representation (5.3), the identity

$$\widehat{X}_{d_{n+1},n+1} = \sum_{i=1}^{n} \theta_{d_{n-i+1},n,i}(X_{d_{n+1-i},n+1-i} - \widehat{X}_{d_{n+1-i},n+1-i})$$

19

leads by linearity of the operators to (6.23).

(ii) Let

$$B_{(d_n),n} = (\beta_{d_n,n,1}, \ldots, \beta_{d_1,n,n}) \quad \text{and} \quad \Pi_n = (\pi_1, \ldots, \pi_n), \tag{6.24}$$

which are both operators from H^n to H defined as follows: let $x = (x_1, \ldots, x_n) \in H^n$ with $x_i \in H$ for $i = 1, \ldots, n$. Then $B_{(d_n),n}\, x = \sum_{i=1}^{n} \beta_{d_{n+1-i},n,i} x_i \in H$. By definition of the norm in H^n we have for all $x \in H$

$$\left\|(B_{(d_n),n} - \Pi_n)(x)\right\| = \sum_{i=1}^{n} \left\|(\beta_{d_{n+1-i},n,i} - \pi_i)(x)\right\|.$$

We show that this tends to 0 as $n \to \infty$, which immediately gives $\left\|(\beta_{d_{n+1-i},n,i} - \pi_i)(x)\right\| \to 0$ for all $i \in \mathbb{N}$.

First notice that for $x \in H^n$ and with $P_{(d_n)}$ defined in (6.2), the triangular inequality yields

$$\left\|(B_{(d_n),n} - \Pi_n)(x)\right\| \leq \left\|(B_{(d_n),n} - \Pi_n P_{(d_n)})(x)\right\| + \left\|\Pi_n(I_{H^n} - P_{(d_n)})(x)\right\|$$
$$=: J_1(d_n, n)(x) + J_2(d_n, n)(x),$$

with identity operator I_{H^n} on H^n. We find bounds for $J_1(d_n, n)(x)$ and $J_2(d_n, n)(x)$. Since uniform convergence implies pointwise convergence, we consider the operator norm of $J_1(d_n, n)(x)$

$$J_1(d_n, n) := \left\|B_{(d_n),n} - \Pi_n P_{(d_n)}\right\|_{\mathcal{L}}$$

and show that $J_1(d_n, n) \to 0$ as $n \to \infty$. From Theorem 2.1.8 in [24] we find

$$\left\|B_{(d_n),n} - \Pi_n P_{(d_n)}\right\|_{\mathcal{L}}^2 = \left\|(B_{(d_n),n} - \Pi_n P_{(d_n)})(B_{(d_n),n} - \Pi_n P_{(d_n)})^*\right\|_{\mathcal{L}}. \tag{6.25}$$

Recall the spectral representation of $\Gamma_{(d_n),n}$ as in (6.3). By the definition of $B_{(d_n),n}$ and $\Pi_n P_{(d_n)}$, note that $(B_{(d_n),n} - \Pi_n P_{(d_n)}) P_{(d_n)} = B_{(d_n),n} - \Pi_n P_{(d_n)}$. Extracting the smallest positive eigenvalue $\lambda_{(d_n),k_n}^{(n)}$ of $\Gamma_{(d_n),n}$, we get

$$\left\|(B_{(d_n),n} - \Pi_n P_{(d_n)}) \Gamma_{(d_n),n} (B_{(d_n),n} - \Pi_n P_{(d_n)})^*\right\|_{\mathcal{L}}$$
$$= \left\|(B_{(d_n),n} - \Pi_n P_{(d_n)}) \sum_{j=1}^{k_n} \lambda_{(d_n),j}^{(n)} (e_{(d_n),j}^{(n)} \otimes e_{(d_n),j}^{(n)}) (B_{(d_n),n} - \Pi_n P_{(d_n)})^*\right\|_{\mathcal{L}}$$
$$\geq \lambda_{(d_n),k_n}^{(n)} \left\|(B_{(d_n),n} - \Pi_n P_{(d_n)})(B_{(d_n),n} - \Pi_n P_{(d_n)})^*\right\|_{\mathcal{L}}. \tag{6.26}$$

Since $A_{d_i} \subseteq A_{d_n}$ for all $i \leq n$ we obtain $\mathbf{A}_{(d_n)} := (A_{d_n}, A_{d_{n-1}}, \ldots, A_{d_1}) \subseteq \mathbf{A}_{d_n} := (A_{d_n}, A_{d_n}, \ldots, A_{d_n})$ and, therefore, $P_{d_n} P_{(d_n)} = P_{(d_n)}$. Together with the definition of $\Gamma_{(d_n),n}$ this implies

$$\Gamma_{(d_n),n} = P_{(d_n)} \Gamma_n P_{(d_n)} = P_{(d_n)} P_{d_n} \Gamma_n P_{d_n} P_{(d_n)} = P_{(d_n)} \Gamma_{d_n,n} P_{(d_n)}.$$

Since $\langle x, \Gamma_{(d_n),n} x \rangle = \langle x, \Gamma_{d_n,n} x \rangle$ for all $x \in \mathbf{A}_{d_n}$, and $\mathbf{A}_{(d_n)} \subseteq \mathbf{A}_{d_n}$, we get

$$\lambda_{(d_n),k_n}^{(n)} = \min_{x \in \mathbf{A}_{(d_n)}} \frac{\langle x, \Gamma_{(d_n),n} x \rangle}{\|x\|^2} = \min_{x \in \mathbf{A}_{(d_n)}} \frac{\langle x, \Gamma_{d_n,n} x \rangle}{\|x\|^2} \geq \min_{x \in \mathbf{A}_{d_n}} \frac{\langle x, \Gamma_{d_n,n} x \rangle}{\|x\|^2} = \lambda_{d_n,d_n\cdot n}^{(n)},$$

where the first and last equality hold by application of Theorem 4.2.7 in [15]. Furthermore, by Lemma 6.1, $\lambda^{(n)}_{d_n,d_n\cdot n} \geq 2\pi\alpha_{d_n}$. Therefore,

$$\lambda^{(n)}_{d_n,k_n} \geq \lambda^{(n)}_{d_n,d_n\cdot n} \geq 2\pi\alpha_{d_n}. \tag{6.27}$$

With (6.26) and (6.27), we get

$$\|B_{(d_n),n} - \Pi_n P_{(d_n)}\|^2 \leq \frac{1}{2\pi\alpha_{d_n}}\left\|(B_{(d_n),n} - \Pi_n P_{(d_n)})\Gamma_{(d_n),n}(B_{(d_n),n} - \Pi_n P_{(d_n)})^*\right\|_{\mathcal{L}} =: J_1'(d_n,n). \tag{6.28}$$

Furthermore, since $\langle Ax, y\rangle = \langle x, A^*y\rangle$ for $A \in \mathcal{L}$ and $x, y \in H$, and by (6.24) and the structure of $\Gamma_{(d_n),n}$,

$$\left\|(B_{(d_n),n} - \Pi_n P_{(d_n)})\Gamma_{(d_n),n}(B_{(d_n),n} - \Pi_n P_{(d_n)})^*\right\|_{\mathcal{L}}$$
$$\leq \left\|\mathbb{E}\left[\sum_{i=1}^n(\beta_{d_{n+1-i},n,i} - \pi_i P_{A_{d_{n+1-i}}})X_{d_{n+1-i},n+1-i} \otimes \sum_{j=1}^n(\beta_{d_{n+1-j},n,j} - \pi_j P_{A_{d_{n+1-j}}})X_{d_{n-j+1},n-j+1}\right]\right\|_{\mathcal{L}}.$$

Now with (2.5) and (6.21) we get

$$\left\|\mathbb{E}\left[\sum_{i=1}^n(\beta_{d_{n+1-i},n,i} - \pi_i P_{A_{d_{n+1-i}}})X_{d_{n+1-i},n+1-i} \otimes \sum_{j=1}^n(\beta_{d_{n+1-j},n,j} - \pi_j P_{A_{d_{n+1-j}}})X_{d_{n-j+1},n-j+1}\right]\right\|_{\mathcal{L}}$$
$$= \left\|\mathbb{E}\left[\left(\widehat{X}_{d_{n+1},n+1} - X_{n+1} + \varepsilon_{n+1} + \sum_{i>n}\pi_i X_{n+1-i} + \sum_{i=1}^n\pi_i(I - P_{A_{d_{n+1-i}}})X_{n+1-i}\right)\right.\right.$$
$$\left.\left.\otimes \left(\widehat{X}_{d_{n+1},n+1} - X_{n+1} + \varepsilon_{n+1} + \sum_{j>n}\pi_j X_{n+1-j} + \sum_{j=1}^n\pi_j(I - P_{A_{d_{n+1-j}}})X_{n+1-j}\right)\right]\right\|_{\mathcal{L}} \tag{6.29}$$

With the trianglular inequality, (6.29) decomposes in the following four terms giving with (6.28):

$$2\pi\alpha_{d_n}\|B_{(d_n),n} - \Pi_n P_{(d_n)}\|^2$$
$$\leq \left\|\mathbb{E}\left[\left(\widehat{X}_{d_{n+1},n+1} - X_{n+1} + \varepsilon_{n+1}\right) \otimes \left(\widehat{X}_{d_{n+1},n+1} - X_{n+1} + \varepsilon_{n+1}\right)\right]\right\|_{\mathcal{L}}$$
$$+ \left\|\mathbb{E}\left[\left(\sum_{i>n}\pi_i X_{n+1-i} + \sum_{i=1}^n\pi_i(I - P_{A_{d_{n+1-i}}})X_{n+1-i}\right) \otimes \left(\sum_{j>n}\pi_j X_{n+1-j} + \sum_{j=1}^n\pi_j(I - P_{A_{d_{n+1-j}}})X_{n+1-j}\right)\right]\right\|_{\mathcal{L}}$$
$$+ \left\|\mathbb{E}\left[\left(\widehat{X}_{d_{n+1},n+1} - X_{n+1} + \varepsilon_{n+1}\right) \otimes \left(\sum_{j>n}\pi_j X_{n+1-j} + \sum_{j=1}^n\pi_j(I - P_{A_{d_{n+1-j}}})X_{n+1-j}\right)\right]\right\|_{\mathcal{L}}$$
$$+ \left\|\mathbb{E}\left[\left(\sum_{i>n}\pi_i X_{n+1-i} + \sum_{i=1}^n\pi_i(I - P_{A_{d_{n+1-i}}})X_{n+1-i}\right) \otimes \left(\widehat{X}_{d_{n+1},n+1} - X_{n+1} + \varepsilon_{n+1}\right)\right]\right\|_{\mathcal{L}}.$$

Define

$$f(n, d_n, m_n) := \left(\sum_{j>m_n}\|\pi_j\|_{\mathcal{L}} + \sum_{j>d_{n-m_n}}\lambda_j\right).$$

By Theorem 5.3 the first term is of the order $f(n, d_n, m_n)$. The second term is of the same order by the calculations following (6.12). Concerning the remaining two terms, using first that

21

$\|C_{X,Y}\|_{\mathcal{L}} \le \mathbb{E}\|X\|\|Y\|$, and then applying the Cauchy-Schwarz inequality gives

$$\left\|\mathbb{E}\left[\left(\widehat{X}_{d_{n+1},n+1} - X_{n+1} + \varepsilon_{n+1}\right) \otimes \left(\sum_{j>n} \pi_j X_{n+1-j} + \sum_{j=1}^n \pi_j(I - P_{A_{d_{n+1-j}}})X_{n+1-j}\right)\right]\right\|_{\mathcal{L}}^2$$

$$\le \left(\mathbb{E}\left\|\widehat{X}_{d_{n+1},n+1} - X_{n+1} + \varepsilon_{n+1}\right\|_{\mathcal{L}}\left\|\sum_{j>n}\pi_j X_{n+1-j} + \sum_{j=1}^n \pi_j(I - P_{A_{d_{n+1-j}}})X_{n+1-j}\right\|\right)^2 \quad (6.30)$$

$$\le \mathbb{E}\left\|\widehat{X}_{d_{n+1},n+1} - X_{n+1} + \varepsilon_{n+1}\right\|_{\mathcal{L}}^2 \mathbb{E}\left\|\sum_{j>n}\pi_j X_{n+1-j} + \sum_{j=1}^n \pi_j(I - P_{A_{d_{n+1-j}}})X_{n+1-j}\right\|_{\mathcal{L}}^2.$$

Both terms are of the order $f(n, d_n, m_n)$ by Theorem 5.3(i).

Hence, $\|B_{(d_n),n} - \Pi_n P_{(d_n)}\|^2$ is of the order $f(n, d_n, m_n)/\alpha_{d_n}$, and with the assumption (5.8),

$$J_1(d_n, n)^2 \to 0, \quad n \to \infty. \quad (6.31)$$

We now estimate $J_2(d_n, n)(x)$, which we have to consider pointwise. For every $x = (x_1, \ldots, x_n) \in H^n$ with $x_i \in H$ for $1 \le i \le n$ and $\|x\| \le 1$,

$$J_2(d_n, n) = \|\Pi_n(I - P_{(d_n)})(x)\|$$
$$= \left\|\left(\pi_1(I_H - P_{A_{d_n}}), \pi_2(I_H - P_{A_{d_{n-1}}}), \ldots, \pi_n(I_H - P_{A_{d_1}})\right)(x)\right\|$$
$$= \sum_{i=1}^n \left\|\pi_i(I_H - P_{A_{d_{n+1-i}}})(x_i)\right\|.$$

Let $m \in \mathbb{N}$ such that $m < n$. Then,

$$\sum_{i=1}^n \left\|\pi_i(I_H - P_{A_{d_{n+1-i}}})(x_i)\right\| = \sum_{i=1}^m \left\|\pi_i(I_H - P_{A_{d_{n+1-i}}})(x_i)\right\| + \sum_{i=m+1}^n \left\|\pi_i(I_H - P_{A_{d_{n+1-i}}})(x_i)\right\|. \quad (6.32)$$

Note that $I_H - P_{A_{d_n}}$ is a projection operator on the orthogonal complement of A_{d_n}. Hence for all $n \in \mathbb{N}$, we have $\|I_H - P_{A_{d_n}}\| = 1$ (see e.g. Theorem 2.1.9 in [24]). Furthermore, for $A, B \in \mathcal{L}$ and $x \in H$, $\|ABx\| \le \|A\|\|B\|\|x\|$, and since $\|x_i\| \le 1$,

$$\sum_{i=m+1}^n \left\|\pi_i(I_H - P_{A_{d_{n+1-i}}})(x_i)\right\| \le \sum_{i>m} \|\pi_i\|_{\mathcal{L}}. \quad (6.33)$$

Furthermore, since $A_{d_j} \subseteq A_{d_i}$ for $j \le i$,

$$\sum_{i=1}^m \left\|\pi_i(I_H - P_{A_{d_{n+1-i}}})(x_i)\right\| \le \sum_{i=1}^m \|\pi_i\|_{\mathcal{L}}\left\|(I_H - P_{A_{d_{n+1-m}}})(x_i)\right\|. \quad (6.34)$$

Since $\sum_{i=1}^\infty \|\pi_i\|_{\mathcal{L}} < \infty$, for every $\delta > 0$ there exists some $m_\delta \in \mathbb{N}$, such that $\sum_{i>m_\delta} \|\pi_i\|_{\mathcal{L}} < \delta/2$. Hence, with (6.32), (6.33) and (6.34) we estimate

$$\sum_{i=1}^n \left\|\pi_i(I_H - P_{A_{d_{n+1-i}}})(x_i)\right\| \le \sum_{i=1}^{m_\delta} \|\pi_i\|_{\mathcal{L}}\left\|(I_H - P_{A_{d_{n+1-m_\delta}}})(x_i)\right\| + \delta/2. \quad (6.35)$$

Furthermore, for the first term of the rhs of (6.35),

$$\sum_{i=1}^{m_\delta} \|\pi_i\|_{\mathcal{L}}\left\|(I_H - P_{A_{d_{n+1-m_\delta}}})(x_i)\right\| \le \max_{1 \le j \le m_\delta}\left\|(I_H - P_{A_{d_{n+1-m_\delta}}})(x_j)\right\|\sum_{i=1}^{m_\delta} \|\pi_i\|_{\mathcal{L}}.$$

Now note that $\|(I_H - P_{A_{d_{n+1-m_\delta}}})(x)\| \to 0$ for $n \to \infty$ for all $x \in H$. Hence, there exists some $n_\delta \in \mathbb{N}$ such that

$$\max_{1 \le j \le m_\delta} \|(I_H - P_{A_{d_{n_\delta+1-m_\delta}}})(x_j)\| \sum_{i=1}^{m_\delta} \|\pi_i\|_{\mathcal{L}} < \delta/2.$$

Hence, $J_2(d_n, n)(x) < \delta$ for all $n \ge n_\delta$ and all $x \in H$.

Together with (6.31), this proves (ii).

(iii) Similarly to the proof of (ii), we start by defining for every $n \in \mathbb{N}$,

$$\widetilde{B}_{(d_n),n-j} := (\beta_{d_n,n-j,1}, \beta_{d_{n-1},n-j,2}, \ldots, \beta_{d_{j+1},n-j,n-j}, 0_H, \ldots, 0_H), \quad j = 1, \ldots, n,$$

where the last j entries are 0_H, the null operator on H. Then $\widetilde{B}_{(d_n),n-j}$ is a bounded linear operator from H^n to H. Analogously to the beginning of the proof of (ii), we show that $\|\widetilde{B}_{(d_n),n} - \widetilde{B}_{(d_n),n-j}\|_{\mathcal{L}} \to 0$ for $n \to \infty$. With the same calculation as deriving (6.28) from (6.25), we obtain

$$\|\widetilde{B}_{(d_n),n} - \widetilde{B}_{(d_n),n-j}\|_{\mathcal{L}}^2 \le \frac{1}{2\pi\alpha_{d_n}}\|(\widetilde{B}_{(d_n),n} - \widetilde{B}_{(d_n),n-j})\Gamma_{(d_n),n}(\widetilde{B}_{(d_n),n} - \widetilde{B}_{(d_n),n-j})^*\|_{\mathcal{L}}$$

$$=: \frac{1}{2\pi\alpha_{d_n}}\widetilde{J}_1'(d_n, n).$$

Applying the same steps as when bounding $J_1(d_n, n)$ in the proof of (ii), and setting $\beta_{d_{n+j},n,m} = 0$ for $m > n$, we obtain

$$\widetilde{J}_1'(d_n, n) = \Big\|\mathbb{E}\Big[\Big(\sum_{i=1}^n (\beta_{d_{n-i+1},n,i} - \beta_{d_{n-i+1},n-j,i})X_{d_{n+1-i},n+1-i}\Big)$$

$$\otimes \Big(\sum_{l=1}^n (\beta_{d_{n-l+1},n,l} - \beta_{d_{n-l+1},n-j,l})X_{d_{n+1-l},n+1-l}\Big)\Big]\Big\|_{\mathcal{L}}$$

$$= \Big\|\mathbb{E}\Big[\big(\widehat{X}_{d_{n+1},n+1} - \widehat{X}_{d_{n+1},n+1}(n-j)\big) \otimes \big(\widehat{X}_{d_{n+1},n+1} - \widehat{X}_{d_{n+1},n+1}(n-j)\big)\Big]\Big\|_{\mathcal{L}},$$

where $\widehat{X}_{d_{n+1},n+1}(k) = \sum_{l=1}^k \beta_{d_{n-l+1},k,l}X_{d_{n+1-l},n+1-l}$ is defined as in (6.22). By adding and subtracting $X_{d_{n+1},n+1} + \varepsilon_{n+1}$ and then using the linearity of the scalar product we get

$$\widetilde{J}_1'(K, n) = \Big\|\mathbb{E}\Big[\big((\widehat{X}_{d_{n+1},n+1} - X_{d_{n+1},n+1} - \varepsilon_{n+1}) - (\widehat{X}_{d_{n+1},n+1}(n-j) - X_{d_{n+1},n+1} - \varepsilon_{n+1})\big)$$

$$\otimes \big((\widehat{X}_{d_{n+1},n+1} - X_{d_{n+1},n+1} - \varepsilon_{n+1}) - (\widehat{X}_{d_{n+1},n+1}(n-j) - X_{d_{n+1},n+1} - \varepsilon_{n+1})\big)\Big]\Big\|_{\mathcal{L}}$$

$$\le \Big\|\mathbb{E}\Big[(\widehat{X}_{d_{n+1},n+1} - X_{d_{n+1},n+1} - \varepsilon_{n+1}) \otimes (\widehat{X}_{d_{n+1},n+1} - X_{d_{n+1},n+1} - \varepsilon_{n+1})\Big]\Big\|_{\mathcal{L}}$$

$$+ \Big\|\mathbb{E}\Big[(\widehat{X}_{d_{n+1},n+1}(n-j) - X_{d_{n+1},n+1} - \varepsilon_{n+1}) \otimes (\widehat{X}_{d_{n+1},n+1}(n-j) - X_{d_{n+1},n+1} - \varepsilon_{n+1})\Big]\Big\|_{\mathcal{L}}$$

$$+ \Big\|\mathbb{E}\Big[(\widehat{X}_{d_{n+1},n+1} - X_{d_{n+1},n+1} - \varepsilon_{n+1}) \otimes (\widehat{X}_{d_{n+1},n+1}(n-j) - X_{d_{n+1},n+1} - \varepsilon_{n+1})\Big]\Big\|_{\mathcal{L}}$$

$$+ \Big\|\mathbb{E}\Big[(\widehat{X}_{d_{n+1},n+1}(n-j) - X_{d_{n+1},n+1} - \varepsilon_{n+1}) \otimes (\widehat{X}_{d_{n+1},n+1} - X_{d_{n+1},n+1} - \varepsilon_{n+1})\Big]\Big\|_{\mathcal{L}}.$$

For $n \to \infty$ the first term converges to 0 by Theorem 5.3 (i). For every fixed $j \in \{1, \ldots, n\}$ the second term converges to 0 by exactly the same arguments. Similar arguments as in the proof of (ii) show that the third and fourth terms also converge to 0. Indeed, applying the Cauchy-Schwarz inequality, we find as in (6.30),

$$\Big\|\mathbb{E}\Big[(\widehat{X}_{d_{n+1},n+1} - X_{d_{n+1},n+1} - \varepsilon_{n+1}) \otimes (\widehat{X}_{d_{n+1},n+1}(n-j) - X_{d_{n+1},n+1} - \varepsilon_{n+1})\Big]\Big\|_{\mathcal{L}}^2$$

$$\le \mathbb{E}\|\widehat{X}_{d_{n+1},n+1} - X_{d_{n+1},n+1} - \varepsilon_{n+1}\|_{\mathcal{L}}^2 \, \mathbb{E}\|\widehat{X}_{d_{n+1},n+1}(n-j) - X_{d_{n+1},n+1} - \varepsilon_{n+1}\|_{\mathcal{L}}^2.$$

Since both these terms tend to 0 for $n \to \infty$, $\widetilde{J}_1'(d_n, n) \to 0$ for $n \to \infty$, which finishes the proof of (iii).

(iv) By (6.23)

$$\theta_{d_{n+1-k},n,k} = \sum_{l=1}^{k} \beta_{d_{n+1-l},n,l} \theta_{d_{n+1-k},n-l,k-l}, \quad k = 1, \ldots, n,$$

and we get $\theta_{d_n,n,1} = \beta_{d_n,n,1}$. Hence, for $n \to \infty$ and fixed $j \in \mathbb{N}$,

$$\|(\theta_{d_n,n,1} - \theta_{d_n,n-j,1})(x)\| = \|(\beta_{d_n,n,1} - \beta_{d_n,n-j,1})(x)\| \to 0. \tag{6.36}$$

For some fixed $j \in \mathbb{N}$ by a shift of (6.23), we obtain

$$\theta_{d_{n+1-k},n-j,k} = \sum_{l=1}^{k} \beta_{d_{n+1-l},n-j,l} \theta_{d_{n+1-l},n-j-l,k-l}. \tag{6.37}$$

With (6.37) and then the triangular equality after adding and subtracting $\beta_{d_{n+1-l},n,l} \theta_{d_{n+1-l},n-j-l,k-l}(x)$ for $l = 1, \ldots, k$,

$$\left\|(\theta_{d_{n+1-k},n,k} - \theta_{d_{n+1-k},n-j,k})(x)\right\| = \left\|\left(\sum_{l=1}^{k} \beta_{d_{n+1-l},n,l} \theta_{d_{n+1-k},n-l,k-l} - \beta_{d_{n+1-l},n-j,l} \theta_{d_{n+1-l},n-j-l,k-l}\right)(x)\right\|$$

$$\leq \left\|\sum_{l=1}^{k} \beta_{d_{n+1-l},n,l} (\theta_{d_{n+1-k},n-l,k-l} - \theta_{d_{n+1-l},n-j-l,k-l})(x)\right\|$$

$$+ \left\|(\beta_{d_{n+1-l},n,l} - \beta_{d_{n+1-l},n-j,l}) \theta_{d_{n+1-l},n-j-l,k-l}(x)\right\|$$

By (iii) $\|(\beta_{d_{n+1-l},n,l} - \beta_{d_{n+1-l},n-j,l})(x)\| \to 0$ as $n \to \infty$. Furthermore, if for all $l = 1, \ldots, i-1$, $\|(\theta_{d_{n+1-l},n,l} - \theta_{d_{n+1-l},n-j,l}(x)\| \to 0$, then $\|(\theta_{d_{n+1-i},n,i} - \theta_{d_{n+1-i},n-j,i}(x)\| \to 0$. The proof then follows by induction with the initial step given in (6.36). $\square$

We are now ready to prove Theorem 5.3(ii).

Proof of Theorem 5.3(ii). Set $\pi_0 := -I_H$. By (2.5) and the definition of a linear process (2.3)

$$-\varepsilon_n = \sum_{i=0}^{\infty} \pi_i(X_{n-i}) = \sum_{i=0}^{\infty} \pi_i\left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{n-i-j}\right), \quad n \in \mathbb{Z}.$$

Setting $k = i + j$, this can be rewritten as

$$-\varepsilon_n = \sum_{i=0}^{\infty} \pi_i\left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{n-i-j}\right) = \sum_{k=0}^{\infty}\left(\sum_{i+j=k} \pi_j \psi_i\right)\varepsilon_{n-k} = \sum_{k=0}^{\infty}\sum_{j=0}^{k} \pi_j \psi_{k-j} \varepsilon_{n-k}.$$

Equating the coefficients we get $\sum_{j=0}^{k} \pi_j \psi_{k-j} = 0$ for $k > 0$. Since $-\pi_0 = I_H$, extracting the first term of the series, $\sum_{j=1}^{k} \pi_j \psi_{k-j} - I_H \psi_k = 0$, hence,

$$\sum_{j=1}^{k} \pi_j \psi_{k-j} = \psi_k.$$

24

Furthermore, by (6.23) we get for all $x \in H$,

$$\left\|\left(\theta_{d_{n+1-i},n,i} - \psi_i\right)(x)\right\| = \left\|\left(\sum_{j=1}^{i} \beta_{d_{n+1-j},n,j}\theta_{d_{n+1-i},n-j,i-j} - \sum_{j=1}^{i} \pi_j\psi_{i-j}\right)(x)\right\|$$

$$= \left\|\sum_{j=1}^{i}(\beta_{d_{n+1-j},n,j} - \pi_j)\theta_{d_{n+1-i},n-j,i-j}(x) - \sum_{j=1}^{i}\pi_j(\psi_{i-j} - \theta_{d_{n+1-i},n-j,i-j})(x)\right\|$$

$$\leq \left\|\sum_{j=1}^{i}(\beta_{d_{n+1-j},n,j} - \pi_j)\theta_{d_{n+1-i},n-j,i-j}(x)\right\| + \left\|\sum_{j=1}^{i}\pi_j(\psi_{i-j} - \theta_{d_{n+1-i},n,i-j})(x)\right\|$$

$$+ \left\|\left(\sum_{j=1}^{i}\pi_j(\theta_{d_{n+1-i},n,i-j} - \theta_{d_{n+1-i},n-j,i-j})\right)(x)\right\|,$$

where we have added and subtracted $\theta_{d_{n+1-i},n,i-j}$ and applied the triangular inequality for the last equality. Now, for $n \to \infty$, the last term tends to 0 by Proposition 6.4 (iv). The first term tends to 0 by Proposition 6.4 (ii). The second term tends to 0 by induction, where the initial step is clear, since $\psi_1 = -\pi_1$ and $\theta_{d_n,n,1} = \beta_{d_n,n,1}$. $\qquad\square$

References

[1] A. Antoniadis, E. Paparoditis, and T. Sapatinas. A functional wavelet-kernel approach for time series prediction. *Journal of the Royal Statistical Society, Series B*, 68:837–857, 2006.

[2] A. Aue, D. Norinho, and S. Hoermann. On the prediction of stationary functional time series. *Journal of the American Statistical Association*, 110(509):378–392, 2015.

[3] P. Besse and H. Cardot. Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1. *Canadian Journal of Statistics*, 24:467–487, 1996.

[4] D. Bosq. *Linear Processes in Function Spaces: Theory and Applications.* Springer, New York, 2000.

[5] D. Bosq. General linear processes in Hilbert spaces and prediction. *Journal of Statistical Planning and Inference*, 137:879–894, 2007.

[6] D. Bosq. Computing the best linear predictor in a Hilbert space. Applications to general ARMAH processes. *Journal of Multivariate Analysis*, 124:436–450, 2014.

[7] D. Bosq and D. Blanke. *Inference and Prediction in Large Dimensions.* Wiley-Dunod, Chichester, 2007.

[8] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods (2nd Ed.).* Springer, New York, 1991.

[9] C. Cerovecki and S. Hörmann. On the CLT for discrete Fourier transforms of functional time series. *arXiv:1506.00970v2 [math.PR]*, 2015.

[10] R. Fortet. *Vecteurs, fonctions et distributions aléatoires dans les espaces de Hilbert.* Hermes, Paris, 1995.

[11] I. Gohberg, S. Goldberg, and M.A. Kaashoek. *Basic Classes of Linear Operators.* Birkhäuser, Basel, 2003.

[12] S. Hörmann and P. Kokoszka. Weakly dependent functional data. *Annals of Statistics*, 38(3):1845–1884, 2010.

[13] L. Horvàth and P. Kokoszka. *Inference for Functional Data with Applications*. Springer, New York, 2012.

[14] S. Hörmann, L. Kidzinski, and M. Hallin. Dynamic functional principal components. *Journal of the Royal Statistical Society: Series B*, 77(2):319–348, 2015.

[15] T. Hsing and R. Eubank. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley, West Sussex, UK, 2015.

[16] R.J. Hyndman and H.L. Shang. Forecasting functional time series. *Journal of the Korean Statistical Society*, 38:199–221, 2009.

[17] V. Kargin and A. Onatski. Curve forecasting by functional autoregression. *Journal of Multivariate Analysis*, 99:2508–2526, 2008.

[18] J. Klepsch, C. Klüppelberg, and T. Wei. Prediction of functional ARMA processes with an application to traffic data. *arXiv:1603.02049 [math.PR]*, 2016.

[19] F. Merlevède. Sur l'inversibilité des processus linéaires à valeurs dans un espace de Hilbert. *C. R. Acad. Sci. Paris, Serie I*, 321:477–480, 1995.

[20] H. Mitchell. *Topics in Multiple Time Series*. PhD thesis, Royal Melbourne Institute of Technology, 1996.

[21] H. Mitchell and P.J. Brockwell. Estimation of the coefficients of a multivariate linear filter using the Innovations Algorithm. *Journal of Time Series Analysis*, 18(2):157–179, 1997.

[22] S. Nsiri and R. Roy. On the invertibility of multivariate linear processes. *Journal of Time Series Analysis*, 14(3):305–316, 1993.

[23] V.M. Panaretros and S. Tavakoli. Fourier analysis of stationary time series in function space. *Annals of Statistics*, 41(2):568–603, 2013.

[24] B. Simon. *Operator Theory - A Comprehensive Course in Analysis, Part 4*. AMS, 2015.

[25] F. Spangenberg. Strictly stationary solutions of ARMA equations in Banach spaces. *Journal of Multivariate Analysis*, 121:127–138, 2013.