# Modeling discrete time stochastic processes with block frames 

Maciej Klimek

# MODELING DISCRETE TIME STOCHASTIC PROCESSES WITH BLOCK FRAMES 

Maciej KLIMEK

March 10, 2011


#### Abstract

This paper develops the concept of block frames which was introduced in an earlier article to provide a geometric and yet computationally efficient approach to some problems of time series analysis. The concept is extended to deal with a greater range of issues, related in particular to weak stationarity and to causality relationships between stochastic flows. It is also shown how to use information generating function systems to force an essentially linear theory to account for non-linear phenomena.


Key words. Block frames in Hilbert spaces, discrete time stochastic processes, time series, matrix identities, weak stationarity, causality.

AMS Subject Classification:
Primary - 60G07, 62M10, 42C15, 46C05; Secondary - 15A24.

## 1 Introduction

Although modern time series analysis has grown into a very complex and diverse area of research, one task remains central to nearly all of its incarnations: how to use known history of evolution of a stochastic process to say something about its future development. In probabilistic terms, it translates into pursuit of different models of conditional expectation. Sometimes the abstract concept of the latter is practically invisible in simplified models. Often it is modeled by simple, computationally viable, algebraic or functional relationships like for instance in the constantly growing family of GARCH type models. In general, conditional expectation does not lend itself easily to computations, as in most non-trivial probabilistic models it involves projections of given random variables onto infinite dimensional spaces of other random variables. In practice, this
theoretical difficulty is compounded by practical considerations. For example, empirical data which by necessity is always finite - usually gives few clues as to realism of probabilistic assumptions imposed on it by researchers, sophisticated statistical tests notwithstanding.

Classic recursive linear representations of discrete time stochastic processes originating from the work of Levinson, Durbin, Whittle, Wiggins and Robinson (see e.g. [4], [9], [10], [11], [31], [30],[32]) have been expanded in many ways and widely used, usually with respect to stationary and nondegenerate stochastic sequences. A comprehensive methodology extending this type of approach has been proposed by the Japanese mathematician Yasunori Okabe, whose aim was to strike the right balance between the use of purely mathematical considerations and handling of empirical data. Okabe's theory of $\mathrm{KM}_{2} \mathrm{O}$-Langevin equations revolves around a computable approach to verification of different types of stationarity, causality and determinism, as well as handling of prediction and filtering problems. For more precise information see [20], [26], [17], [18] , [12], [7], [13], [8], [19] and references given there.

The goal of this paper is to build a unified, geometrically clear and yet computationally powerful set of tools well suited for handling of some problems of time series analysis. Initial steps in this direction were taken in the earlier paper [8], but here the ideas are developed much further, encompassing in particular practically all non-probabilistic mathematical notions needed in description of Okabe's theory of $\mathrm{KM}_{2} \mathrm{O}$-Langevin equations. By not referring to the probabilistic context, the frame properties themselves are often more general than the properties of stochastic flows explored in Okabe's work. Moreover, the block frame approach makes it possible to extract the geometric principles behind many aspects of that theory. As such, many of the statements generalize earlier results and can be potentially applied in many different settings.

The paper is organized as follows. The purpose of Section 2 is to establish notation. Section 3 introduces basic definitions and properties of block frames and improves on the approach to block frames proposed in [8]. In Section 4 we look at the relationship between block frame coefficients and matrix covariance functions. In particular, we obtain a new quick derivation of the Cholesky decomposition for block matrices. Section 5 takes up the subject of weak stationarity. In particular, we show that block frames can be conveniently used to characterize weak stationarity. In Section 6 we deal with different types of causality relationship between stochastic flows. We devote special attention to handling of non-linear information in fundamentally linear models. Essentially this is done by approximation of conditional expectation, which - in general - is based on an infinitedimensional model, by computationally viable finite dimensional projections combined with a nonlinear build-up of the given stochastic flow. We introduce a concept of information generating function systems to facilitate the discussion. We also define the concept of a moving composition of a stochastic sequence with a function to explain how the theory can be linked to practical applications.

## 2 Notation

In this and the next section we will modify slightly the concepts and notation from [8] in order to cater with more general applications of block frames.

If $L$ is a matrix (or a linear operator), then its transpose (resp. adjoint operator) is denoted by $L^{*}$. The vector space $\mathbb{R}^{d_{1} \times d_{2}}$ of real $\left(d_{1} \times d_{2}\right)$-matrices is furnished with the natural Frobenius inner product $\langle A, B\rangle=\operatorname{trace}\left(A B^{*}\right)$, where $A, B=\left[b_{i j}\right] \in \mathbb{R}^{d_{1} \times d_{2}}$. If $d$ is a positive integer and $H$ is a real Hilbert space with the inner product $(x, y) \mapsto\langle x, y\rangle$, then the Cartesian power $H^{d}$ is also a Hilbert space with the inner product $\left\langle\left(x_{1}, \ldots, x_{d}\right),\left(y_{1}, \ldots, y_{d}\right)\right\rangle=\sum_{i=1}^{d}\left\langle x_{i}, y_{i}\right\rangle$. It is convenient to refer to elements of $H^{d}$ as block vectors. For block vectors $\mathbf{x}=\left(x_{1}, \ldots, x_{d_{1}}\right) \in H^{d_{1}}$ and $\mathbf{y}=\left(y_{1}, \ldots, y_{d_{2}}\right) \in$ $H^{d_{2}}$, we define their Gram product $\langle\langle\mathbf{x}, \mathbf{y}\rangle\rangle=\left[\left\langle x_{i}, y_{j}\right\rangle\right] \in \mathbb{R}^{d_{1} \times d_{2}}$. The name is justified as $\langle\langle\mathbf{x}, \mathbf{x}\rangle\rangle$ is the Gram matrix corresponding to the vectors $x_{1}, \ldots, x_{d}$. Note that $\langle\langle\mathbf{x}, \mathbf{y}\rangle\rangle=0$ if and only if $x_{i}$ is orthogonal to $y_{j}$ for all choices of $i$ and $j$. Block vectors can be multiplied from the left by matrices of suitable size: if $A=\left[a_{i j}\right] \in \mathbb{R}^{d_{1} \times d_{2}}$ and $\mathbf{x}=\left(x_{1}, \ldots, x_{d_{2}}\right) \in H^{d_{2}}$, then the $i$-th entry of $A \mathbf{x} \in H^{d_{1}}$ is $\sum_{j=1}^{d_{2}} a_{i j} x_{j}$, for $i=1, \ldots, d_{1}$. In particular $A$ can be identified with a linear operator from $H^{d_{2}}$ to $H^{d_{1}}$.

Obviously, if $\mathbf{x}, \mathbf{y}$ are block vectors and $A, B$ are matrices - all of compatible sizes - then $\langle\langle\mathbf{x}, \mathbf{y}\rangle\rangle=$ $\langle\mathbf{y}, \mathbf{x}\rangle\rangle^{*}$ and $A\langle\langle\mathbf{x}, \mathbf{y}\rangle\rangle B=\left\langle\left\langle A \mathbf{x}, B^{*} \mathbf{y}\right\rangle\right\rangle$. In particular, if $\mathbf{x}=\mathbf{y}$ and $A=B^{*} \in \mathbb{R}^{1 \times d_{1}}$, the last property shows that $\langle\langle\mathbf{x}, \mathbf{x}\rangle\rangle$ is semi-positive definite. Moreover, with $A=I \in \mathbb{R}^{d_{1} \times d_{1}}$ and $B \in \mathbb{R}^{d_{1} \times 1}$, it shows that $\langle\langle\mathbf{x}, \mathbf{x}\rangle\rangle$ is invertible if and only if the components of $\mathbf{x}$ are linearly independent. If $\mathbf{x} \in H^{d}$ has linearly independent components, $C \in \mathbb{R}^{d \times d}$ and we define $\mathbf{y}=C(\sqrt{\langle\mathbf{x}, \mathbf{x}\rangle\rangle})^{-1} \mathbf{x}$, then $\langle\langle\mathbf{y}, \mathbf{y}\rangle\rangle=C C^{*}$.

By an integer interval with end-points $a, b \in \mathbb{Z} \cup\{-\infty, \infty\}$ we mean the set $\{a . . b\}=\mathbb{Z} \cap[a, b]$, if $a \leq b$ or the empty set otherwise.

Let $J$ be a non-empty index set and let $\mathbf{x}_{j}=\left(x_{1 j}, \ldots, x_{d_{1} j}\right) \in H^{d_{1}}$ where $j \in J$. The symbol $\operatorname{Span}\left(\mathrm{x}_{j}: j \in J\right)$ will denote the closed linear span of all of the components of the given block vectors, that is of the set $\left\{x_{i j} ; i=1, \ldots, d_{1}, j \in J\right\}$. If $J=\{p . . q\}$, with some $p, q \in \mathbb{Z}$, we also write $\operatorname{Span}\left(\mathbf{x}_{p}, \ldots, \mathbf{x}_{q}\right)$. With this notation in mind, $\mathbf{y} \in\left(\boldsymbol{S p a n}\left(\mathbf{x}_{1}, \ldots, \mathbf{x}_{n}\right)\right)^{d_{2}}$ if and only if there exist matrices $C_{1}, \ldots, C_{n} \in \mathbb{R}^{d_{2} \times d_{1}}$ such that $\mathbf{y}=\sum_{i=1}^{n} C_{i} \mathbf{x}_{i}$. The matrix coefficients $C_{i}$ are unique, if the vectors $x_{i j}$ are linearly independent.

## 3 Basic properties of block frames

Let $J$ be a linearly ordered finite or countable set. By $\ell^{2}(J)$ we will denote the Hilbert space of all square summable sequences $\left(\lambda_{j}\right)_{j \in J}$ of real numbers, with the usual inner product

$$
\langle\lambda, \eta\rangle=\sum_{j \in J}\left\langle\lambda_{j}, \eta_{j}\right\rangle
$$

Of course if $\# J=m$, then $\ell^{2}(J)=\mathbb{R}^{m}$. By $\ell^{2}(J)^{m \times n}$, we will mean the $m n$-th Cartesian power of $\ell^{2}(J)$ with the natural inner product, and with sequences of $m \times n$ matrices as elements.

We will begin with a short recap of the basic properties of frames. Let $H$ be a separable real Hilbert space. An ordered collection of vectors $F=\left(x_{j}\right)_{j \in J} \subset H$ is called a frame if for some constants $A, B>0$ (called frame bounds) we have

$$
\begin{equation*}
A\|x\|^{2} \leq \sum_{j \in J}\left\langle x, x_{j}\right\rangle^{2} \leq B\|x\|^{2}, \quad x \in H \tag{1}
\end{equation*}
$$

By necessity, any frame for $H$ is linearly dense in $H$. A frame is said to be tight if $A=B$. A tight frame is exact if $A=B=1$. Three linear mappings are associated in a natural way with each frame. The first is called the analysis operator (or the Bessel operator) and is defined as follows

$$
T_{F}: H \ni x \mapsto\left(\left\langle x, x_{j}\right\rangle\right)_{j \in J} \in \ell^{2}(J)
$$

Note that $T_{F}$ is injective because of the lower bound in (1). The second one is the adjoint of $T_{F}$ which is called the synthesis operator (or the pre-frame operator) and is given explicitly by the formula

$$
T_{F}^{*}: \ell^{2}(J) \ni\left(\lambda_{j}\right)_{j \in J} \mapsto \sum_{j \in J} \lambda_{1} x_{j} \in H
$$

It can be shown that $T_{F}^{*}$ is always surjective. The third one is the frame operator $S_{F}=T_{F}^{*} T_{F}$. Explicitly

$$
S_{F}: H \ni x \mapsto \sum_{j \in J}\left\langle x, x_{j}\right\rangle x_{j} \in H
$$

Clearly $\left\|T_{F}\right\| \leq \sqrt{B},\left\|T_{F}^{*}\right\| \leq \sqrt{B}$ and $\left\|S_{F}\right\| \leq B$. It is not difficult to check that $S_{F}$ is an isomorphism. The main reason why frames have inspired much interest is the so called resolution of identity:

$$
x=\sum_{j \in J}\left\langle x, S_{F}^{-1}\left(x_{j}\right)\right\rangle x_{j}, \quad x \in H
$$

Such representation of $x$ may potentially contain a large amount of redundant information about the vector, in comparison with an expansion with respect of an orthonormal basis, but this is actually
an asset in many applications of frames. An introduction to frame theory as well as list of further references can be found in e.g. [1].

Let $\mathbf{x}_{j}=\left(x_{1 j}, \ldots, x_{d j}\right) \in H^{d}$, where $j \in J$, be a family of vectors and let $M=\operatorname{Span}\left(\mathbf{x}_{j}: j \in J\right)$. In this case we say that $F=\left\{\mathbf{x}_{j}: j \in J\right\}$ is a block frame for $M^{d}$, if the family $\underline{F}=\left\{x_{i j}: i=\right.$ $1, \ldots, d, j \in J\}$ is a frame for $M$. We will be assuming that the index set $\{1, \ldots, d\} \times J$ is furnished with the reverse lexicographic order i.e. $(m, n) \leq(p, q)$ if either $n<q$ or $n=q$ and $m \leq p$. We will also say that $\underline{F}$ is the underlying frame for the block frame $F$. Obviously a block frame is not necessarily a frame for $M^{d}$. We define the block frame operator associated with $F$ as

$$
\begin{equation*}
S_{F}: M^{d} \ni \mathbf{z}=\left(z_{1}, \ldots, z_{d}\right) \mapsto\left(S_{\underline{F}}\left(z_{1}\right), \ldots, S_{\underline{E}}\left(z_{d}\right)\right) \in M^{d} . \tag{2}
\end{equation*}
$$

Obviously

$$
\begin{equation*}
S_{F}(\mathbf{z})=\sum_{j \in J}\left\langle\left\langle\mathbf{z}, \mathbf{x}_{j}\right\rangle\right\rangle \mathbf{x}_{j}, \quad \mathbf{z} \in M^{d}, \tag{3}
\end{equation*}
$$

and

$$
\begin{equation*}
S_{F}(C \mathbf{z})=C S_{F}(\mathbf{z}) \quad \mathbf{z} \in M^{d}, C \in \mathbb{R}^{d \times d} . \tag{4}
\end{equation*}
$$

We can say that $S_{F}$ is block homogeneous. Note also that

$$
S_{F}^{-1}: M^{d} \ni\left(w_{1}, \ldots, w_{d}\right) \mapsto\left(S_{\underline{F}}^{-1}\left(w_{1}\right), \ldots, S_{\underline{F}}^{-1}\left(w_{d}\right)\right) \in M^{d} .
$$

Since $S_{F}$ is bijective, $S_{F}^{-1}$ is also block homogeneous.
It will be also convenient to define block versions of the analysis and synthesis operators. The block analysis operator is defined as

$$
T_{F}: M^{d} \ni \mathbf{z} \mapsto\left(\left\langle\left\langle\mathbf{z}, \mathbf{x}_{j}\right\rangle\right\rangle\right)_{j \in J} \in \ell^{2}(J)^{d \times d}
$$

The block synthesis operator is defined as

$$
T_{F}^{*}: \ell^{2}(J)^{d \times d} \ni\left(C_{j}\right)_{j \in J} \mapsto \sum_{j \in J} C_{j} \mathbf{x}_{j} \in M^{d}
$$

Both operators can be expressed more directly in terms of the underlying frame $\underline{F}$. To this end define

$$
\operatorname{Row}_{k}\left(C_{j}\right)_{j \in J}=\left\{C_{j}(k, m): j \in J, m=1, \ldots, d\right\}
$$

for any family $C=\left(C_{j}\right)_{j \in J}$ of $d \times d$-matrices, where $C_{j}=\left[C_{j}(k, m)\right]_{k, m=1, \ldots, d}$. Then

$$
\operatorname{Row}_{k}\left[T_{F}\left(z_{1}, \ldots, z_{d}\right)\right]=T_{\underline{F}}\left(z_{k}\right), \quad k=1, \ldots, d,
$$

and for $C=\left(C_{j}\right)_{j \in J}$

$$
T_{F}^{*}(C)=\left(T_{\underline{F}}^{*}\left(\operatorname{Row}_{1}(C)\right), \ldots, T_{\underline{F}}^{*}\left(\operatorname{Row}_{d}(C)\right)\right) .
$$

Clearly the operators $T_{F}$ and $T_{F}^{*}$ are bounded. Moreover, $T_{F}$ is injective, $T_{F}^{*}$ is surjective and $S_{F}=T_{F}^{*} T_{F}$, just as in the case of ordinary frames. Furthermore $T_{F}^{*}$ is indeed the adjoint of $T_{F}$, because for $C=\left(C_{j}\right)_{j \in J} \in \ell^{2}(J)^{d \times d}$ we have

$$
\begin{aligned}
\left\langle T_{F}(\mathbf{z}), C\right\rangle & =\sum_{j \in J} \operatorname{trace}\left[\left\langle\left\langle\mathbf{z}, \mathbf{x}_{j}\right\rangle\right\rangle C_{j}^{*}\right]=\sum_{j \in J} \operatorname{trace}\left\langle\left\langle\mathbf{z}, C_{j} \mathbf{x}_{j}\right\rangle\right\rangle \\
& =\operatorname{trace}\left\langle\left\langle\mathbf{z}, \sum_{j \in J} C_{j} \mathbf{x}_{j}\right\rangle\right\rangle=\left\langle\mathbf{z}, T_{F}^{*}(C)\right\rangle .
\end{aligned}
$$

In particular it follows that the block frame operator $S_{F}$ is self-adjoint.
If $L: H^{d} \longrightarrow H^{d}$ is a bounded linear operator, we will say that it is block self-adjoint if

$$
\langle\langle L(\mathbf{x}), \mathbf{y}\rangle\rangle=\langle\langle\mathbf{x}, L(\mathbf{y})\rangle\rangle
$$

for all $\mathbf{x}, \mathbf{y} \in H^{d}$.
If $M$ is a closed subspace of $H$, then by $\mathbf{P}_{M}$ we will denote the orthogonal projection of $H$ onto $M$. In particular, for the orthogonal projection $\mathbf{P}_{M^{d}}: H^{d} \longrightarrow H^{d}$ we have

$$
\mathbf{P}_{M^{d}}(\mathbf{x})=\left(\mathbf{P}_{M}\left(x_{1}\right), \ldots, \mathbf{P}_{M}\left(x_{n}\right)\right) \in M^{d}, \quad \mathbf{x}=\left(x_{1}, \ldots, x_{d}\right) \in H^{d}
$$

For later use we will also define the block projection error by the formula

$$
\begin{equation*}
\operatorname{Err}_{M^{d}}(\mathbf{x})=\left\langle\left\langle\mathbf{P}_{\left(M^{\perp}\right)^{d}}(\mathbf{x}), \mathbf{P}_{\left(M^{\perp}\right)^{d}}(\mathbf{x})\right\rangle\right\rangle=\left\langle\left\langle\mathbf{P}_{\left(M^{\perp}\right)^{d}}(\mathbf{x}), \mathbf{x}\right\rangle\right\rangle . \tag{5}
\end{equation*}
$$

If $M, N \subset H$ are two non-trivial closed subspaces and $M \perp N$, then clearly

$$
\begin{equation*}
\mathbf{P}_{(M \oplus N)^{d}}(\mathbf{x})=\mathbf{P}_{M^{d}}(\mathbf{x})+\mathbf{P}_{N^{d}}(\mathbf{x}), \quad \mathbf{x} \in H^{d} . \tag{6}
\end{equation*}
$$

If in addition $F, G$ are block frames for $M, N$ respectively, then $F \cup G$ is a block fame for $M \oplus N$ and

$$
\begin{equation*}
S_{F \cup G}(\mathbf{x}+\mathbf{y})=S_{F}(\mathbf{x})+S_{G}(\mathbf{y}) \quad \mathbf{x} \in M^{d}, \mathbf{x} \in N^{d} . \tag{7}
\end{equation*}
$$

The following theorem lists the basic properties of block projections and block frame operators.

Theorem 3.1 Let $F=\left(\mathbf{x}_{j}\right)_{j \in J}$ be a block frame for $M^{d}$.
(i) The block frame operator $S_{F}: M^{d} \longrightarrow M^{d}$ is a block self-adjoint isomorphism.
(ii) Let $d^{\prime}$ be positive integer. The orthogonal projection onto $M^{d^{\prime}}$ is given by the formula

$$
\begin{equation*}
\mathbf{P}_{M^{d^{\prime}}}(\mathbf{y})=\sum_{j=1}^{n}\left\langle\left\langle\mathbf{y}, S_{F}^{-1}\left(\mathbf{x}_{j}\right)\right\rangle\right\rangle \mathbf{x}_{j}=\sum_{j=1}^{n}\left\langle\left\langle\mathbf{y}, \mathbf{x}_{j}\right\rangle\right\rangle S_{F}^{-1}\left(\mathbf{x}_{j}\right), \quad \mathbf{y} \in H^{d^{\prime}} \tag{8}
\end{equation*}
$$

Moreover,

$$
\begin{equation*}
\left\langle\left\langle\mathbf{P}_{M^{d^{\prime}}}(\mathbf{y}), \mathbf{z}\right\rangle\right\rangle=\left\langle\left\langle\mathbf{y}, \mathbf{P}_{M^{d}}(\mathbf{z})\right\rangle\right\rangle, \quad \mathbf{y} \in H^{d^{\prime}}, \mathbf{z} \in H^{d} \tag{9}
\end{equation*}
$$

In particular, the operator $\mathbf{P}_{M^{d^{\prime}}}$ is block self-adjoint.
(iii) Given $\mathbf{y} \in H^{d^{\prime}}$, define the affine subspace $N(\mathbf{y})$ of $\ell^{2}(J)^{d^{\prime} \times d}$ by the formula

$$
N(\mathbf{y})=\left\{Y=\left(Y_{j}\right)_{j \in J} \in \ell^{2}(J)^{d^{\prime} \times d}: \mathbf{P}_{M^{d^{\prime}}}(\mathbf{y})=\sum_{j \in J} Y_{j} \mathbf{x}_{j}\right\}
$$

Then the orthogonal projection of the origin in $\ell^{2}(J)^{d^{\prime} \times d}$ onto $N(\mathbf{y})$ is given by

$$
\begin{equation*}
Y=\left(Y_{j}\right)_{j \in J}=\left(\left\langle\left\langle\mathbf{y}, S_{F}^{-1}\left(\mathbf{x}_{j}\right)\right\rangle\right\rangle\right)_{j \in J} \tag{10}
\end{equation*}
$$

In other words, the vector (10), consisting of the block frame coefficients for $\mathbf{y}$, furnishes the minimum norm solution to the equation

$$
\mathbf{P}_{M^{d^{\prime}}}(\mathbf{y})=\sum_{j \in J} Y_{j} \mathbf{x}_{j}
$$

Proof: Because of (2) it is clear that $S_{F}$ is an isomorphism. In view of (3) $S_{F}$ is block self-adjoint. In order to show (ii), suppose first that $d^{\prime}=1$. Then

$$
\begin{align*}
\mathbf{P}_{M}(\mathbf{y}) & =\sum_{i=1}^{d} \sum_{j=1}^{n}\left\langle\mathbf{y}, S_{\underline{F}}^{-1}\left(x_{i j}\right)\right\rangle x_{i j}=\sum_{j=1}^{n}\left(\sum_{i=1}^{d}\left\langle\mathbf{y}, S_{\underline{F}}^{-1}\left(x_{i j}\right)\right\rangle x_{i j}\right)  \tag{11}\\
& =\sum_{j=1}^{n}\left\langle\left\langle\mathbf{y}, S_{F}^{-1}\left(\mathbf{x}_{j}\right)\right\rangle\right\rangle \mathbf{x}_{j} \tag{12}
\end{align*}
$$

as needed. The second equality in (8) follows similarly from the basic properties of $S_{\underline{F}}$.

For an arbitrary $d^{\prime}$, we observe that for $\mathbf{y}=\left(y_{1}, \ldots, y_{d^{\prime}}\right) \in H^{d^{\prime}}$ we have

$$
\begin{aligned}
\mathbf{P}_{M^{d^{\prime}}}(\mathbf{y}) & =\left(\mathbf{P}_{M}\left(y_{1}\right), \ldots, \mathbf{P}_{M}\left(y_{d^{\prime}}\right)\right) \\
& =\left(\sum_{j=1}^{n}\left\langle\left\langle y_{1}, S_{F}^{-1}\left(\mathbf{x}_{j}\right)\right\rangle\right\rangle \mathbf{x}_{j}, \ldots, \sum_{j=1}^{n}\left\langle\left\langle y_{d^{\prime}}, S_{F}^{-1}\left(\mathbf{x}_{j}\right)\right\rangle\right\rangle \mathbf{x}_{j}\right) \\
& =\sum_{j=1}^{n}\left\langle\left\langle\mathbf{y}, S_{F}^{-1}\left(\mathbf{x}_{j}\right)\right\rangle\right\rangle \mathbf{x}_{j}
\end{aligned}
$$

The other equality in (8) can be obtained in the same way.
The two representations of the projection in (8) imply that the left-hand side of (9) is the same as the right-hand side because

$$
\sum_{i=1}^{n}\left\langle\left\langle\mathbf{y}, \mathbf{x}_{i}\right\rangle\right\rangle\left\langle\left\langle S_{F}^{-1}\left(\mathbf{x}_{i}\right), \mathbf{z}\right\rangle\right\rangle=\sum_{i=1}^{n}\left\langle\left\langle\mathbf{y},\left\langle\left\langle\mathbf{z}, S_{F}^{-1}\left(\mathbf{x}_{i}\right)\right\rangle\right\rangle \mathbf{x}_{i}\right\rangle\right\rangle=\left\langle\left\langle\mathbf{y}, \sum_{i=1}^{n}\left\langle\left\langle\mathbf{z}, S_{F}^{-1}\left(\mathbf{x}_{i}\right)\right\rangle\right\rangle \mathbf{x}_{i}\right\rangle\right\rangle
$$

To prove (iii) we can use an argument from [8]. It suffices to prove the statement for $\mathbf{y} \in M^{d^{\prime}}$. We want to solve the equation $T_{F}^{*}(Y)=\mathbf{y}$ with respect to $Y$. Any solution has the form $Y=Y^{\prime}+Y^{\prime \prime}$, where $Y^{\prime}$ is a uniquely determined element of the range of $T_{F}$ and $Y^{\prime \prime}$ is an arbitrary element of the kernel of $T_{F}^{*}$. Thus the minimum norm solution must be by necessity $Y=Y^{\prime}$. Now

$$
\left(\left\langle\left\langle\mathbf{y}, S_{F}^{-1}\left(\mathbf{x}_{j}\right)\right\rangle\right\rangle\right)_{j \in J}=\left(\left\langle\left\langle S_{F}^{-1}(\mathbf{y}), \mathbf{x}_{j}\right\rangle\right\rangle\right)_{j \in J}=T_{F}\left(S_{F}^{-1}(\mathbf{y})\right)
$$

is the required solution because $T_{F}^{*} T_{F}=S_{F}$.

The vectors $S_{F}^{-1}\left(\mathbf{x}_{1}\right), \ldots, S_{F}^{-1}\left(\mathbf{x}_{n}\right)$ form another block frame for $M^{d}$, which can be referred to as the dual block frame for the block frame $\mathbf{x}_{1}, \ldots, \mathbf{x}_{n}$.

A number of block frame properties which were shown in [8] in the case of a finite $J$ and with $d^{\prime}=d$, remain true in the more general settings adopted here. The notation in this paper is slightly different to allow the proofs and statements of those properties to be applied nearly verbatim in the present case. More precisely, the target spaces in symbols for orthogonal projections and error terms are now clearly specified e.g. like in $\mathbf{P}_{M^{d}}$ or $\mathbf{E r r}_{M^{d}}$. Below we will only state explicitly the formulas needed in this paper.

We will use the ampersand symbol \& to denote concatenation of linearly ordered sets. So if $F$ and $G$ are such sets, then $F \& G$ is also linearly ordered with any element of $F$ regarded as preceding any element of $G$. If $G=\{g\}$ is a singleton, we will write $F \& g$ rather that $F \& G$.

For any matrix $A$, let $A^{\dagger}$ denote the Moore-Penrose pseudoinverse of $A$.

Theorem 3.2 Let $F=\left(\mathbf{x}_{j}\right)_{j \in J}$ be a block frame for $M^{d}$ and let $\mathbf{y} \in H^{d}$. Furthermore, let $M_{\mathbf{y}}=$ $M+\operatorname{Span}(\mathbf{y}), F_{\mathbf{y}}=F \& \boldsymbol{y}$ and let $d^{\prime} \in \mathbb{N}$. Then

$$
\begin{equation*}
\mathbf{P}_{M_{\mathbf{y}}^{d^{\prime}}}(\mathbf{z})=\mathbf{P}_{M^{d^{\prime}}}(\mathbf{z})+\left\langle\left\langle\mathbf{z}, S_{F_{\mathbf{y}}}^{-1}(\mathbf{y})\right\rangle\right\rangle \mathbf{P}_{\left(M^{\perp}\right)^{d}}(\mathbf{y}), \quad \mathbf{z} \in H^{d^{\prime}} \tag{13}
\end{equation*}
$$

Moreover,

$$
\begin{equation*}
\operatorname{Err}_{M_{\mathbf{y}}^{d^{\prime}}}(\mathbf{z})=\left(I-\left\langle\left\langle\mathbf{z}, S_{F_{\mathbf{y}}}^{-1}(\mathbf{y})\right\rangle\right\rangle\left\langle\left\langle\mathbf{y}, S_{F_{\mathbf{z}}}^{-1}(\mathbf{z})\right\rangle\right\rangle\right) \operatorname{Err}_{M^{d^{\prime}}}(\mathbf{z}), \quad \mathbf{z} \in H^{d^{\prime}} \tag{14}
\end{equation*}
$$

where $F_{\mathbf{z}}=F \& \boldsymbol{z}, M_{\mathbf{z}}=M+\mathbf{S p a n}(\mathbf{z})$ and $I$ denotes here the $d^{\prime} \times d^{\prime}$ identity matrix. The frame coefficients and error terms satisfy the identities

$$
\begin{align*}
\left\langle\left\langle\mathbf{z}, S_{F_{\mathbf{y}}}^{-1}\left(\mathbf{x}_{j}\right)\right\rangle\right\rangle & =\left\langle\left\langle\mathbf{z}, S_{F}^{-1}\left(\mathbf{x}_{j}\right)\right\rangle\right\rangle-\left\langle\left\langle\mathbf{z}, S_{F_{\mathbf{y}}}^{-1}(\mathbf{y})\right\rangle\right\rangle\left\langle\left\langle\mathbf{y}, S_{F}^{-1}\left(\mathbf{x}_{j}\right)\right\rangle\right\rangle, \quad j \in J .  \tag{15}\\
\left\langle\left\langle\mathbf{z}, S_{F_{\mathbf{y}}}^{-1}(\mathbf{y})\right\rangle\right\rangle \mathbf{E r r}_{M^{d}}(\mathbf{y}) & =\mathbf{E r r}_{M^{d^{\prime}}}(\mathbf{z})\left\langle\left\langle S_{F_{\mathbf{z}}}^{-1}(\mathbf{z}), \mathbf{y}\right\rangle\right\rangle . \tag{16}
\end{align*}
$$

In particular, if $\mathbf{y} \perp M^{d}$, then

$$
\begin{equation*}
\mathbf{P}_{\left(M_{\mathbf{y}}\right)^{d^{\prime}}}(\mathbf{z})=\mathbf{P}_{M^{d^{\prime}}}(\mathbf{z})+\langle\langle\mathbf{z}, \mathbf{y}\rangle\rangle\langle\langle\mathbf{y}, \mathbf{y}\rangle\rangle^{\dagger} \mathbf{y}, \quad \mathbf{z} \in H^{d^{\prime}} \tag{17}
\end{equation*}
$$

Proof: The first four properties above can be checked by straightforward modifications of the arguments used in [8] to prove their less general counterparts. Formula (17) follows from (6) combined with Corollary 4 in [8].

Remark 3.3 If the notation is extended to allow $F=\emptyset, F_{\mathbf{y}}=\{\mathbf{y}\}$ and $M=\{0\}$, then and the formulas (13), (14), (16) and (17) remain true.

## 4 Matrix covariance function and the blueprint algorithm

With any indexed collection of vectors $F=\left(\mathbf{x}_{j}\right)_{j \in J} \subset H^{d}$ one can associate its matrix covariance function $R_{F}$ defined by the formula:

$$
R_{F}: J \times J \ni(i, j) \mapsto\left\langle\left\langle\mathbf{x}_{i}, \mathbf{x}_{j}\right\rangle\right\rangle \in \mathbb{R}^{d \times d}
$$

If $J$ is finite, say $J=\left\{j_{1}, \ldots, j_{n}\right\}$, then the then we can regard $R_{F}$ as a block matrix in $\mathbb{R}^{n d \times n d}$, with the building blocks $R_{F}\left(j_{p}, j_{q}\right) \in \mathbb{R}^{d \times d}$, where $p, q=1, \ldots, n$. Moreover, if

$$
\underline{F}=\left\{x_{1, j_{1}}, \ldots, x_{d, j_{1}}, x_{1, j_{2}}, \ldots, x_{d, j_{2}}, \ldots ., x_{1, j_{n}}, \ldots, x_{d, j_{n}}\right\}
$$

then

$$
\begin{equation*}
R_{F}=R_{\underline{F}} . \tag{18}
\end{equation*}
$$

This is so, as the ( $k, l$ )-entry of the block matrix $R_{F}$ is the same as the $(r, s)$-entry of the $(p, q)$-block $\left\langle\left\langle\mathbf{x}_{j_{p}}, \mathbf{x}_{j_{q}}\right\rangle\right.$, where

$$
p=\left\lfloor\frac{k}{d}\right\rfloor+1, \quad q=\left\lfloor\frac{l}{d}\right\rfloor+1
$$

and

$$
r=k-(p-1) d, \quad s=l-(q-1) d .
$$

Consequently, the matrix $R_{F}$ is symmetric and semi-positive definite.
Any symmetric $(n d \times n d)$-matrix divided into $(d \times d)$-blocks is necessarily self-adjoint with respect to its building blocks, in the sense that the transpose of the $(p, q)$-block is the same as the ( $q, p$ ) block. In particular the matrix $R_{F}$ has this property, but it is generally not symmetric with respect the the $(d \times d)$-blocks forming it.

It is easy to check that any semi-positive definite symmetric matrix furnishes a covariance function for some vectors ${ }^{1}$ or block-vectors in the case of block matrices. For example, let $d=1$ and let $R \in \mathbb{R}^{n \times n}$ be symmetric and semi-positive definite, with a decomposition $R=L L^{*}$, with some $L \in \mathbb{R}^{n \times n}$. (The classic Cholesky decomposition is one of the possible choices here, or alternatively the square root of $R$ can be taken as $L$.) If $\mathbf{e} \in H^{n}$ has orthonormal components, then the components of $\mathbf{x}=L \mathbf{e}$ have $R$ as the covariance matrix. The case $d>1$ follows from (18). For other constructions of this type see also [18] or [13].

In [8] (see Corollary 7 there), it was shown that basic properties of block frames lead to a quick derivation of an algorithm calculating frame coefficients and error terms on the basis of known matrix covariance function. More precisely, given a finite block frame $F=\left[\mathbf{x}_{1}, \ldots, \mathbf{x}_{N}\right]$ for $M^{d}$ and $\mathbf{z} \in H^{d^{\prime}}$, the algorithm is a recipe for calculating the matrix coefficients of $\mathbf{P}_{M^{d^{\prime}}}(\mathbf{z})$ with respect to $F$ and $\operatorname{Err}_{M^{d^{\prime}}}(\mathbf{z})$. In [8] it was assumed that, $d^{\prime}=d$, but in view of the comments made in the previous section $d^{\prime}$ can be arbitrary. What is remarkable about this algorithm is that the only information about the vectors $\mathbf{x}_{1}, \ldots, \mathbf{x}_{N}, \mathbf{z}$ it requires, are the matrices $\langle\langle\mathbf{u}, \mathbf{v}\rangle\rangle$, where $\mathbf{u}, \mathbf{v} \in\left\{\mathbf{x}_{1}, \ldots, \mathbf{x}_{N}, \mathbf{z}\right\}$. As such the algorithm can be viewed as a generalization of earlier versions of Levinson's type algorithms like for instance those presented in [9] (with an extra assumption of non-degeneracy) and in [5], [6], [2] (with an extra assumption of stationarity). Formally, it is also equivalent to a generalization of older results proposed by Matsuura's [15] who dispensed with the additional assumptions of non-degeneracy or stationarity. Matsuura obtained a minimum norm version of the algorithm motivated by the need to lower sensitivity to observation noise in practical applications. In what follows the name the blueprint algorithm will be used when referring to the algorithm from $[8]$ and its generalization to the case $d^{\prime} \neq d$. Calling the algorithm a block frame

[^0]extension algorithm would have also been accurate, but a different name seems more appropriate as block frames and block vectors feature on the algorithm only indirectly through matrices associated with them. The algorithm exploits interrelationships between these matrices, never using directly the underlying vector structure. One could say that it creates an matrix algebra blueprint for relationships between block frames and other block vectors.

In the case of $d=d^{\prime}$, we can express more succinctly what the blueprint algorithm allows us to do.

Theorem 4.1 If $F=\left\{\mathbf{x}_{j}: m \leq j \leq n\right\}$ is a finite block frame for $M^{d} \subset H^{d}, \mathbf{x}_{n+1} \in H^{d}$ and all entries of $R=R_{F \& \mathbf{x}_{n+1}}$ are known, then one can calculate matrices

$$
V_{R} \in \mathbb{R}^{d \times d} \text { and } \gamma_{R}(j) \in \mathbb{R}^{d \times d}(\text { for } j \in\{m, \ldots, n\}),
$$

such that

$$
V_{R}=\operatorname{Err}_{M^{d}}\left(\mathbf{x}_{n+1}\right) \text { and } \gamma_{R}(j)=-\left\langle\left\langle\mathbf{x}_{n+1}, S_{F}^{-1}\left(\mathbf{x}_{j}\right)\right\rangle\right\rangle(\text { for } j \in\{m, \ldots, n\}) .
$$

The calculation of the matrices $\gamma_{R}(j)$ and $V_{R}$ uses the values of the matrix covariance function $R$ as the only input data. Furthermore, the matrix $V_{R}$ is semi-positive definite and

$$
\begin{align*}
R(n+1, k) & =-\sum_{j=m}^{n} \gamma_{R}(j) R(j, k), \quad k=m, \ldots, n,  \tag{19}\\
R(n+1, n+1) & =-\sum_{j=m}^{n} \gamma_{R}(j) R(j, n+1)+V_{R} . \tag{20}
\end{align*}
$$

Moreover, if the matrix covariance function $R$ leads to the matrices $\gamma_{R}(j)$ and $V_{R}$, then the matrix covariance function $R^{*}$ leads to the matrices $\gamma_{R}(j)^{*}$ and $V_{R}^{*}$.

The above theorem yields the following useful property.

Corollary 4.2 Let $F^{(i)}=\left\{\mathbf{x}_{j}^{(i)}: m \leq j \leq n\right\}$, where $i=1,2$, be finite block frames for $M_{(i)}^{d} \subset H^{d}$. Let $\mathbf{x}_{n+1}^{(i)} \in H^{d}$, for $i=1,2$. If

$$
R_{F^{(1)} \& \mathbf{x}_{n+1}^{(1)}}=R_{F^{(2)} \& x_{n+1}^{(2)}},
$$

then

$$
\begin{equation*}
\left\|\mathbf{P}_{M_{(1)}^{d}}\left(\mathbf{x}_{n+1}^{(1)}\right)\right\|=\left\|\mathbf{P}_{M_{(2)}^{d}}\left(\mathbf{x}_{n+1}^{(2)}\right)\right\| . \tag{21}
\end{equation*}
$$

Proof: Since the expression $\left\langle\left\langle\mathbf{P}_{M_{(i)}^{d}}\left(\mathbf{x}_{n+1}^{(i)}\right), \mathbf{P}_{M_{(i)}^{d}}\left(\mathbf{x}_{n+1}^{(i)}\right)\right\rangle\right.$ depends only on the matrix covariance function and

$$
\left\|\mathbf{P}_{M_{(i)}^{d}}\left(\mathbf{x}_{n+1}^{(i)}\right)\right\|=\operatorname{trace}\left\langle\left\langle\mathbf{P}_{M_{(i)}^{d}}\left(\mathbf{x}_{n+1}^{(i)}\right), \mathbf{P}_{M_{(i)}^{d}}\left(\mathbf{x}_{n+1}^{(i)}\right)\right\rangle\right.
$$

the result follows.
Basic properties of frames lead to an alternative derivation of the block matrix Cholesky decomposition shown in [13]. More precisely we have the following.

Theorem 4.3 Let $\mathbf{x}_{1}, \ldots, \mathbf{x}_{n} \in H^{d}$, where $n>1$. Let $M_{k}=\boldsymbol{\operatorname { S p a n }}\left(\mathbf{x}_{1}, \ldots, \mathbf{x}_{k}\right)$ and $F_{k}=\left\{\mathbf{x}_{1}, \ldots, \mathbf{x}_{k}\right\}$ for $k=1, \ldots, n-1$. Let $\mathbf{I}_{d}$ and $\mathbf{O}_{d}$ denote the $(d \times d)$-identity matrix and $(d \times d)$-zero matrix respectively. Define three block matrices with $(d \times d)$-blocks as entries:

$$
\begin{aligned}
& \mathbf{G}=\left[\mathbf{g}_{i j}\right]_{i, j=1, \ldots, n} \text { where } \mathbf{g}_{i j}= \begin{cases}-\left\langle\left\langle\mathbf{x}_{i}, S_{F_{i-1}}^{-1}\left(\mathbf{x}_{j}\right)\right\rangle\right\rangle & \text { if } i>j ; \\
\mathbf{I}_{d} & \text { if } i=j ; \\
\mathbf{O}_{d} & \text { if } i<j,\end{cases} \\
& \mathbf{R}=\left[\left\langle\left\langle\mathbf{x}_{i}, \mathbf{x}_{j}\right\rangle\right\rangle\right]_{i, j=1, \ldots, n}, \\
& \mathbf{D}=\left[\mathbf{d}_{i j}\right]_{i, j=1, \ldots, n} \text { where } \mathbf{d}_{i j}= \begin{cases}\mathbf{x}_{1} & \text { if } i=j=1 ; \\
\mathbf{E r r}_{M_{i-1}^{d}}\left(\mathbf{x}_{i}\right) & \text { if } i=j>1 ; \\
\mathbf{O}_{d} & \text { if } i \neq j .\end{cases}
\end{aligned}
$$

Then the minimum norm solution of the matrix equation

$$
\mathbf{X R X}^{*}=\mathbf{D} .
$$

is given by the block matrix $\mathbf{X}=\mathbf{G}$. Moreover, $\mathbf{L}=\mathbf{G}^{-1}$ is block lower-triangular - with matrices $\mathbf{I}_{d}$ on the diagonal - and the Cholesky decomposition holds:

$$
\mathbf{L D L}^{*}=\mathbf{R} .
$$

Proof: Basically all we have to do is to calculate the $(i, j)$-th block entry of the block matrix $\widetilde{\mathbf{D}}=\left[\widetilde{\mathbf{d}}_{i j}\right]_{i, j=1, \ldots, n}=$ GRG $^{*}$. With the convention that an empty sum is equal to zero we have:

$$
\begin{aligned}
\widetilde{\mathbf{d}}_{i j}= & \sum_{k=1}^{j-1}\left(\sum_{l=1}^{i-1}\left\langle\left\langle\mathbf{x}_{i}, S_{F_{i-1}}^{-1}\left(\mathbf{x}_{l}\right)\right\rangle\right\rangle\left\langle\left\langle\mathbf{x}_{l}, \mathbf{x}_{k}\right\rangle\right\rangle-\left\langle\left\langle\mathbf{x}_{i}, \mathbf{x}_{k}\right\rangle\right\rangle\right)\left\langle\left\langle S_{F_{j-1}}^{-1}\left(\mathbf{x}_{k}\right), \mathbf{x}_{j}\right\rangle\right\rangle \\
& -\left(\sum_{l=1}^{i-1}\left\langle\left\langle\mathbf{x}_{i}, S_{F_{i-1}}^{-1}\left(\mathbf{x}_{l}\right)\right\rangle\right\rangle\left\langle\left\langle\mathbf{x}_{l}, \mathbf{x}_{j}\right\rangle\right\rangle-\left\langle\left\langle\mathbf{x}_{i}, \mathbf{x}_{j}\right\rangle\right\rangle\right),
\end{aligned}
$$

where the expressions in the two sets of round parentheses are of course the the entries of the $i$-th row of GR, with a changed sign. Putting $M_{0}=\{0\}$, we can write that

$$
\begin{aligned}
\tilde{\mathbf{d}}_{i j} & =\sum_{k=1}^{j-1}\left(\left\langle\left\langle\mathbf{P}_{M_{i-1}^{d}}\left(\mathbf{x}_{i}\right), \mathbf{x}_{k}\right\rangle\right\rangle-\left\langle\left\langle\mathbf{x}_{i}, \mathbf{x}_{k}\right\rangle\right\rangle\right)\left\langle\left\langle S_{F_{j-1}}^{-1}\left(\mathbf{x}_{k}\right), \mathbf{x}_{j}\right\rangle\right\rangle-\left\langle\left\langle\mathbf{P}_{M_{i-1}^{d}}\left(\mathbf{x}_{i}\right), \mathbf{x}_{j}\right\rangle\right\rangle+\left\langle\left\langle\mathbf{x}_{i}, \mathbf{x}_{j}\right\rangle\right\rangle \\
& =-\sum_{k=1}^{j-1}\left\langle\left\langle\mathbf{P}_{\left(M_{i-1}^{\perp}\right)^{d}}\left(\mathbf{x}_{i}\right), \mathbf{x}_{k}\right\rangle\right\rangle\left\langle\left\langle S_{F_{j-1}}^{-1}\left(\mathbf{x}_{k}\right), \mathbf{x}_{j}\right\rangle\right\rangle+\left\langle\left\langle\mathbf{P}_{\left(M_{i-1}^{\perp}\right)^{d}}\left(\mathbf{x}_{i}\right), \mathbf{x}_{j}\right\rangle\right\rangle \\
& =\left\langle\left\langle\mathbf{P}_{\left(M_{i-1}^{\perp}\right)^{d}}\left(\mathbf{x}_{i}\right), \mathbf{P}_{\left(M_{j-1}^{\perp}\right)^{d}}\left(\mathbf{x}_{j}\right)\right\rangle\right\rangle .
\end{aligned}
$$

Obviously $\widetilde{\mathbf{d}}_{i i}=\mathbf{d}_{i i}$. On the other hand, if $i<j$, then $M_{i-1} \subset M_{j-1}$ and $\mathbf{x}_{i} \in M_{j-1}$. Thus $\mathbf{P}_{\left(M_{i-1}^{\perp}\right)^{\perp}}\left(\mathbf{x}_{i}\right) \in M_{j-1}$, which implies that $\widetilde{\mathbf{d}}_{i j}=\mathbf{O}_{d}=\mathbf{d}_{i j}$. The case $i>j$ is symmetric.

Since $\boldsymbol{\Delta}=\mathbf{G}-\mathbf{I}$ is strictly lower triangular, it is nilpotent with $\boldsymbol{\Delta}^{n d}=0$. Also, consecutive powers of $\boldsymbol{\Delta}$ have shorter-and-shorter non-zero parts of their rows. Thus the formula

$$
\mathbf{G}^{-1}=(\mathbf{I}+\boldsymbol{\Delta})^{-1}=\sum_{k=0}^{n d-1}(-1)^{k} \boldsymbol{\Delta}^{k},
$$

justifies the last conclusion of the theorem.

## 5 Weak stationarity and block frames

Let $J$ be an integer interval. We say that an indexed collection of vectors $F=\left(\mathbf{x}_{j}\right)_{j \in J} \subset H^{d}$ is weakly stationary if its matrix covariance function satisfies the condition

$$
R_{F}(i, j)=\rho(i-j), \quad i, j \in J,
$$

for some function $\rho:\{i-j: i, j \in J\} \longrightarrow \mathbb{R}^{d \times d}$.
Before focusing our attention on characterization of stationary sequences, it will be convenient to introduce some notation and terminology. Some of the formulas we will have to deal with are not particularly intuitive, so clearer terminology may be quite useful. Suppose that a family $\left\{\mathbf{x}_{j}: j \in J\right\} \subset H^{d}$ is given, where the index set $J$ is an integer interval. If $\{p . . q\} \subset J$, we define

$$
\begin{align*}
& F[p, q]=\left\{\mathbf{x}_{j}: j \in\{p . . q\}\right\}, \text { if }\{p . . q\} \neq \emptyset  \tag{22}\\
& M[p, q]= \begin{cases}\boldsymbol{\operatorname { S p a n }}\{F[p, q]\} & \text { if }\{p . . q\} \neq \emptyset, \\
\{0\} & \text { if }\{p . . q\}=\emptyset\end{cases} \tag{23}
\end{align*}
$$

By $\mathbb{L}$ we will denote the lag operator shifting backward an indexed expression by one unit. We will use positive and negative powers of the lag operator in the usual manner. More specifically, if $k \in \mathbb{Z}$ and an expression to which $\mathbb{L}^{k}$ is applied involves any of the following: $\mathbf{x}_{m}, F[p, q], M[p, q]$, then the resulting expression will contain respectively $\mathbf{x}_{m-k}, F[p-k, q-k], M[p-k, q-k]$. Applications of non-zero integer powers of the lag operator are also referred to as time shifts. A time shift is admissible if it can be applied within $J$. An application of either $\mathbb{L}$ or $\mathbb{L}^{-1}$, provided that it is admissible, is referred to as a minimal time shift. A time shift $\mathbb{L}^{k}$ is called maximal if $\mathbb{L}^{k}$ is admissible, but $\mathbb{L}^{k+\operatorname{sign}(k)}$ is not admissible. We will say that a block vector $\mathbf{x}$ and a block frame $F[p, q]$ form a projection pair, if either $\mathbf{x}=\mathbf{x}_{p-1}$ or $\mathbf{x}=\mathbf{x}_{q+1}$. In such a case the parameters of this projection pair are the matrices

$$
\left\langle\left\langle\mathbf{x}, S_{F[p, q]}^{-1}\left(\mathbf{x}_{p}\right)\right\rangle\right\rangle,\left\langle\left\langle\mathbf{x}, S_{F[p, q]}^{-1}\left(\mathbf{x}_{p+1}\right)\right\rangle\right\rangle, \ldots,\left\langle\left\langle\mathbf{x}, S_{F[p, q]}^{-1}\left(\mathbf{x}_{q}\right)\right\rangle\right\rangle, \operatorname{Err}_{M[p, q]^{d}}(\mathbf{x})
$$

A projection pair is docked within the range $J$, if either $p=\min J$ or $q=\max J$.
The following characterization of stationarity is a block frame version of Theorem 4.2 from [17] and Theorem 6.5 from [13]. Our formulation of this Characterization Theorem, includes additional equivalent conditions of stationarity formulated in terms of the lag operator. It should be noted that a more elaborate variation of the theorem - for double stochastic sequences or for periodic stationarity (see [17] and [13]) - can also be shown in a similar manner as a statement about block frames.

Theorem 5.1 (Characterization Theorem) Let $N$ be a positive integer and let

$$
\mathbf{X}=\left\{\mathbf{x}_{j}: j \in\{0 . . N\}\right\} \subset H^{d}
$$

such that

$$
\begin{equation*}
\left\langle\left\langle\mathbf{x}_{0}, \mathbf{x}_{0}\right\rangle\right\rangle=\left\langle\left\langle\mathbf{x}_{N}, \mathbf{x}_{N}\right\rangle\right\rangle ; \tag{24}
\end{equation*}
$$

Then the following conditions are equivalent:
(a) $\mathbf{X}$ is weakly stationary;
(b) parameters of all projection pairs are invariant with respect to all admissible time shifts;
(c) parameters of all docked projection pairs are invariant with respect to all minimal and maximal time shifts;
(d) for all $n, i$ such that $1 \leq i<n \leq N$ :

$$
\begin{align*}
& \left\langle\left\langle\mathbf{x}_{n}, S_{F[0, n-1]}^{-1}\left(\mathbf{x}_{i}\right)\right\rangle\right\rangle= \\
& \quad=\left\langle\left\langle\mathbf{x}_{n-1}, S_{F[0, n-2]}^{-1}\left(\mathbf{x}_{i-1}\right)\right\rangle\right\rangle-\left\langle\left\langle\mathbf{x}_{n}, S_{F[0, n-1]}^{-1}\left(\mathbf{x}_{0}\right)\right\rangle\right\rangle\left\langle\left\langle\mathbf{x}_{N-n+1}, S_{F[N-n+2, N]}^{-1}\left(\mathbf{x}_{N-n+1+i}\right)\right\rangle\right\rangle \tag{25}
\end{align*}
$$

$$
\begin{align*}
& \left\langle\left\langle\mathbf{x}_{N-n}, S_{F[N-n+1, N]}^{-1}\left(\mathbf{x}_{N-i}\right)\right\rangle\right\rangle= \\
& =\left\langle\left\langle\mathbf{x}_{N-n+1}, S_{F[N-n+2, N]}^{-1}\left(\mathbf{x}_{N-i+1}\right)\right\rangle\right\rangle-\left\langle\left\langle\mathbf{x}_{N-n}, S_{F[N-n+1, N]}^{-1}\left(\mathbf{x}_{N}\right)\right\rangle\right\rangle\left\langle\left\langle\mathbf{x}_{n-1}, S_{F[0, n-2]}^{-1}\left(\mathbf{x}_{n-1-i}\right)\right\rangle\right\rangle, \tag{26}
\end{align*}
$$

and for all $n$ such that $1 \leq n \leq N$ :

$$
\begin{equation*}
\operatorname{Err}_{M[0, n-1]^{d}}\left(\mathbf{x}_{n}\right)=\left(I-\left\langle\left\langle\mathbf{x}_{n}, S_{F[0, n-1]}^{-1}\left(\mathbf{x}_{0}\right)\right\rangle\right\rangle\left\langle\left\langle\mathbf{x}_{N-n}, S_{F[N-n+1, N]}^{-1}\left(\mathbf{x}_{N}\right)\right\rangle\right\rangle\right) \operatorname{Err}_{M[0, n-2]^{d}}\left(\mathbf{x}_{n-1}\right), \tag{27}
\end{equation*}
$$

$$
\begin{align*}
& \mathbf{E r r}_{M[N-n+1, N]^{d}}\left(\mathbf{x}_{N-n}\right)= \\
& \quad=\left(I-\left\langle\left\langle\mathbf{x}_{N-n}, S_{F[N-n+1, N]}^{-1}\left(\mathbf{x}_{N}\right)\right\rangle\right\rangle\left\langle\left\langle\mathbf{x}_{n}, S_{F[0, n-1]}^{-1}\left(\mathbf{x}_{0}\right)\right\rangle\right\rangle\right) \mathbf{E r r}_{M[N-n+2, N]^{d}}\left(\mathbf{x}_{N-n+1}\right), \tag{28}
\end{align*}
$$

$$
\begin{equation*}
\left\langle\left\langle\mathbf{x}_{n}, S_{F[0, n-1]}^{-1}\left(\mathbf{x}_{0}\right)\right\rangle\right\rangle \operatorname{Err}_{M[N-n+2, N]^{d}}\left(\mathbf{x}_{N-n+1}\right)=\operatorname{Err}_{M[0, n-2]^{d}}\left(\mathbf{x}_{n-1}\right)\left\langle\left\langle S_{F[N-n+1, N]}^{-1}\left(\mathbf{x}_{N}\right), \mathbf{x}_{N-n}\right\rangle\right\rangle . \tag{29}
\end{equation*}
$$

Before giving a proof of the above theorem a few comments are in order. Although the above condition (c) may look formidable, it simply a list of slightly modified general properties (15), (14) and $(16)^{2}$. The only difference is that some frame coefficients are shifted forward or backward. Using the lag operator the formulas $(25-29)$ can be re-written equivalently as follows:

$$
\begin{align*}
& \left\langle\left\langle\mathbf{x}_{n}, S_{F[0, n-1]}^{-1}\left(\mathbf{x}_{i}\right)\right\rangle\right\rangle= \\
& =\mathbb{L}\left[\left\langle\left\langle\mathbf{x}_{n}, S_{F[1, n-1]}^{-1}\left(\mathbf{x}_{i}\right)\right\rangle\right\rangle\right]-\left\langle\left\langle\mathbf{x}_{n}, S_{F[0, n-1]}^{-1}\left(\mathbf{x}_{0}\right)\right\rangle\right\rangle \mathbb{L}^{-(N-n+1)}\left[\left\langle\left\langle\mathbf{x}_{0}, S_{F[1, n-1]}^{-1}\left(\mathbf{x}_{i}\right)\right\rangle\right\rangle\right]  \tag{30}\\
& \begin{array}{l}
\left\langle\left\langle\mathbf{x}_{N-n}, S_{F[N-n+1, N]}^{-1}\left(\mathbf{x}_{N-i}\right)\right\rangle\right\rangle= \\
=
\end{array} \\
& \quad \mathbb{L}^{-1}\left[\left\langle\left\langle\mathbf{x}_{N-n}, S_{F[N-n+1, N-1]}^{-1}\left(\mathbf{x}_{N-i}\right)\right\rangle\right\rangle\right]- \\
& \quad-\left\langle\left\langle\mathbf{x}_{N-n}, S_{F[N-n+1, N]}^{-1}\left(\mathbf{x}_{N}\right)\right\rangle\right\rangle \mathbb{L}^{N-n+1}\left[\left\langle\left\langle\mathbf{x}_{N}, S_{F[N-n+1, N-1]}^{-1}\left(\mathbf{x}_{N-i}\right)\right\rangle\right\rangle\right] \tag{31}
\end{align*}
$$

[^1]\[

$$
\begin{align*}
& \operatorname{Err}_{M[0, n-1]^{d}}\left(\mathbf{x}_{n}\right)= \\
& \left(I-\left\langle\left\langle\mathbf{x}_{n}, S_{F[0, n-1]}^{-1}\left(\mathbf{x}_{0}\right)\right\rangle\right\rangle \mathbb{L}^{-(N-n)}\left[\left\langle\left\langle\mathbf{x}_{0}, S_{F[1, n]}^{-1}\left(\mathbf{x}_{n}\right)\right\rangle\right\rangle\right]\right) \mathbb{L}\left[\operatorname{Err}_{M[1, n-1]^{d}}\left(\mathbf{x}_{n}\right)\right],  \tag{32}\\
& \operatorname{Err}_{M[N-n+1, N]^{d}}\left(\mathbf{x}_{N-n}\right)= \\
& =\left(I-\left\langle\left\langle\mathbf{x}_{N-n}, S_{F[N-n+1, N]}^{-1}\left(\mathbf{x}_{N}\right)\right\rangle\right\rangle \mathbb{L}^{N-n}\left[\left\langle\left\langle\mathbf{x}_{N}, S_{F[N-n, N-1]}^{-1}\left(\mathbf{x}_{N-n}\right)\right\rangle\right\rangle\right]\right) .  \tag{33}\\
& \cdot \mathbb{L}^{-1}\left[\operatorname{Err}_{M[N-n+1, N-1]^{d}}\left(\mathbf{x}_{N-n}\right)\right], \\
& \left\langle\left\langle\mathbf{x}_{n}, S_{F[0, n-1]}^{-1}\left(\mathbf{x}_{0}\right)\right\rangle\right\rangle \mathbb{L}^{-(N-n+1)}\left[\operatorname{Err}_{M[1, n-1]^{d}}\left(\mathbf{x}_{0}\right)\right]= \\
& =\mathbb{L}\left[\operatorname{Err}_{M[1, n-1]^{d}}\left(\mathbf{x}_{n}\right)\right] \mathbb{L}^{-(N-n)}\left[\left\langle\left\langle S_{F[1, n]}^{-1}\left(\mathbf{x}_{n}\right), \mathbf{x}_{0}\right\rangle\right\rangle .\right] \tag{34}
\end{align*}
$$
\]

Proof: The implication $(\mathrm{a}) \Rightarrow(\mathrm{b})$ follows directly from the blueprint algorithm. The implication $(\mathrm{b}) \Rightarrow(\mathrm{c})$ is obvious, whereas $(\mathrm{c}) \Rightarrow(\mathrm{d})$ because the properties $(25-29)$ are identical with the properties $(30-34)$. Let us assume that (d) holds true.

In order to show that $\mathbf{X}$ is weakly stationary we will use an induction argument to prove the following assertion for $n \in\{0 . . N\}$ :

Auxiliary Claim: The sequence $\left\{\mathbf{x}_{0}, \ldots, \mathbf{x}_{n}\right\}$ is weakly stationary and $\left.\left\langle\left\langle\mathbf{x}_{i}, \mathbf{x}_{j}\right\rangle\right\rangle=\mathbb{L}^{N-n}\left[\left\langle\mathbf{x}_{i}, \mathbf{x}_{j}\right\rangle\right\rangle\right]$ for any $i, j \in\{N-n . . N\}$.

We will consider first the case of $n=1$. We have:

$$
\begin{aligned}
\left\langle\left\langle\mathbf{x}_{0}, \mathbf{x}_{1}\right\rangle\right\rangle & =\left\langle\left\langle\mathbf{x}_{0}, \mathbf{x}_{0}\right\rangle\right\rangle\left\langle\left\langle S_{F[0,0]}^{-1}\left(\mathbf{x}_{0}\right), \mathbf{x}_{1}\right\rangle\right\rangle+\left\langle\left\langle\mathbf{x}_{0}, \mathbf{P}_{(M[0,0] \perp)^{d}}\left(\mathbf{x}_{1}\right)\right\rangle\right\rangle \quad(\text { by } \quad(8)) \\
& \left.=\left\langle\left\langle\mathbf{x}_{0}, \mathbf{x}_{0}\right\rangle\right\rangle\left\langle\left\langle S_{F[0,0]}^{-1}\left(\mathbf{x}_{0}\right), \mathbf{x}_{1}\right\rangle\right\rangle=\left\langle\left\langle\mathbf{x}_{N}, \mathbf{x}_{N}\right\rangle\right\rangle\left\langle\left\langle S_{F[0,0]}^{-1}\left(\mathbf{x}_{0}\right), \mathbf{x}_{1}\right\rangle\right\rangle \quad \text { by }(24)\right) \\
& =\left\langle\left\langle\mathbf{x}_{N-1}, S_{F[N, N]}^{-1}\left(\mathbf{x}_{N}\right)\right\rangle\right\rangle\left\langle\left\langle\mathbf{x}_{N}, \mathbf{x}_{N}\right\rangle\right\rangle \quad(\text { by }(29) \text { and }(24)) \\
& =\left\langle\left\langle\mathbf{x}_{N-1}, S_{F[N, N]}^{-1}\left(\mathbf{x}_{N}\right)\right\rangle\right\rangle\left\langle\left\langle\mathbf{x}_{N}, \mathbf{x}_{N}\right\rangle\right\rangle+\left\langle\left\langle\mathbf{P}_{\left(M[N, N]^{\perp}\right)^{d}}\left(\mathbf{x}_{N-1}\right), \mathbf{x}_{N}\right\rangle\right\rangle \\
& =\left\langle\left\langle\mathbf{x}_{N-1}, \mathbf{x}_{N}\right\rangle\right\rangle \quad(\text { by }(8)) .
\end{aligned}
$$

Because of (8) and the above we have

$$
\begin{aligned}
\left.\left\langle\mathbf{x}_{1}, \mathbf{x}_{1}\right\rangle\right\rangle & =\left\langle\left\langle\mathbf{x}_{1}, \mathbf{x}_{0}\right\rangle\right\rangle\left\langle\left\langle S_{F[0,0]}^{-1}\left(\mathbf{x}_{0}\right), \mathbf{x}_{1}\right\rangle\right\rangle+\left\langle\left\langle\mathbf{x}_{1}, \mathbf{P}_{\left(M[0,0]^{\perp}\right)^{d}}\left(\mathbf{x}_{1}\right)\right\rangle\right\rangle \\
& \left.=\left\langle\left\langle\mathbf{x}_{N}, \mathbf{x}_{N-1}\right\rangle\right\rangle\left\langle\left\langle S_{F[0,0]}^{-1}\left(\mathbf{x}_{0}\right), \mathbf{x}_{1}\right\rangle\right\rangle+\operatorname{Err}_{M[0,0]^{d}} \mathbf{x}_{1}\right) \\
& =\left\langle\left\langle\mathbf{x}_{N}, \mathbf{x}_{N}\right\rangle\right\rangle\left\langle\left\langle S_{F[N, N]}^{-1}\left(\mathbf{x}_{N}\right), \mathbf{x}_{N-1}\right\rangle\right\rangle\left\langle\left\langle S_{F[0,0]}^{-1}\left(\mathbf{x}_{0}\right), \mathbf{x}_{1}\right\rangle\right\rangle+\operatorname{Err}_{M[0,0]^{d}}\left(\mathbf{x}_{1}\right) \\
& =\left\{\left\langle\left\langle\mathbf{x}_{0}, \mathbf{x}_{0}\right\rangle\right\rangle-\operatorname{Err}_{M[0,0]^{d}}\left(\mathbf{x}_{1}\right)\right\}^{*}+\operatorname{Err}_{M[0,0]^{d}}\left(\mathbf{x}_{1}\right) \\
& =\left\langle\left\langle\mathbf{x}_{0}, \mathbf{x}_{0}\right\rangle\right\rangle,
\end{aligned}
$$

in view of (27). Similarly, but using (28) we can check that

$$
\left\langle\left\langle\mathbf{x}_{N-1}, \mathbf{x}_{N-1}\right\rangle\right\rangle=\left\langle\left\langle\mathbf{x}_{N}, \mathbf{x}_{N}\right\rangle\right\rangle,
$$

which completes the proof of the case $n=1$.
Now assume that, the auxiliary claim is true for some $(n-1) \geq 1$. We will deduce its validity for $n$. For $j \in\{0 . . n\}$ we have

$$
\begin{align*}
\left\langle\left\langle\mathbf{x}_{n}, \mathbf{x}_{j}\right\rangle\right\rangle & =\left\langle\left\langle\mathbf{x}_{n}, S_{F[0, n-1]}^{-1}\left(\mathbf{x}_{0}\right)\right\rangle\right\rangle\left\langle\left\langle\mathbf{x}_{0}, \mathbf{x}_{j}\right\rangle\right\rangle+\sum_{i=1}^{n-1}\left\langle\left\langle\mathbf{x}_{n}, S_{F[0, n-1]}^{-1}\left(\mathbf{x}_{i}\right)\right\rangle\right\rangle\left\langle\left\langle\mathbf{x}_{i}, \mathbf{x}_{j}\right\rangle\right\rangle+ \\
& +\operatorname{Err}_{M[0, n-1]^{d}}\left(\mathbf{x}_{j}\right) \\
& =\sum_{i=0}^{n-2}\left\langle\left\langle\mathbf{x}_{n-1}, S_{F[0, n-2]}^{-1}\left(\mathbf{x}_{i}\right)\right\rangle\right\rangle\left\langle\left\langle\mathbf{x}_{i+1}, \mathbf{x}_{j}\right\rangle\right\rangle+\operatorname{Err}_{M[0, n-1]^{d}}\left(\mathbf{x}_{j}\right)+ \\
& \left.+\left\langle\left\langle\mathbf{x}_{n}, S_{F[0, n-1]}^{-1}\left(\mathbf{x}_{0}\right)\right\rangle\right\rangle\left[\left\langle\left\langle\mathbf{x}_{0}, \mathbf{x}_{j}\right\rangle\right\rangle-\sum_{i=1}^{n-1}\left\langle\left\langle\mathbf{x}_{N-n+1}, S_{F[N-n+2, N]}^{-1}\left(\mathbf{x}_{N-n+1+i}\right)\right\rangle\right\rangle\left\langle\mathbf{x}_{i}, \mathbf{x}_{j}\right\rangle\right\rangle\right] \tag{35}
\end{align*}
$$

The first of the above equalities is a consequence of (8), the second one follows from (25).
If $j \in\{1 . . n-1\}$, then by the induction hypothesis and because $\mathbf{x}_{j} \in F[0, n-1]$, we can see that

$$
\begin{aligned}
& \sum_{i=0}^{n-2}\left\langle\left\langle\mathbf{x}_{n-1}, S_{F[0, n-2]}^{-1}\left(\mathbf{x}_{i}\right)\right\rangle\right\rangle\left\langle\left\langle\mathbf{x}_{i+1}, \mathbf{x}_{j}\right\rangle\right\rangle+\operatorname{Err}_{M[0, n-1]^{d}}\left(\mathbf{x}_{j}\right)= \\
& \sum_{i=0}^{n-2}\left\langle\left\langle\mathbf{x}_{n-1}, S_{F[0, n-2]}^{-1}\left(\mathbf{x}_{i}\right)\right\rangle\right\rangle\left\langle\left\langle\mathbf{x}_{i}, \mathbf{x}_{j-1}\right\rangle\right\rangle=\left\langle\left\langle\mathbf{x}_{n-1}, \mathbf{x}_{j-1}\right\rangle\right\rangle .
\end{aligned}
$$

Moreover, by the induction hypothesis and because $\mathbf{x}_{N-n+1+j} \in F[N-n+2, N]$, we have

$$
\begin{aligned}
\sum_{i=1}^{n-1}\left\langle\left\langle\mathbf{x}_{N-n+1}, S_{F[N-n+2, N]}^{-1}\left(\mathbf{x}_{N-n+1+i}\right)\right\rangle\right\rangle\left\langle\left\langle\mathbf{x}_{i}, \mathbf{x}_{j}\right\rangle\right\rangle & = \\
=\sum_{i=1}^{n-1}\left\langle\left\langle\mathbf{x}_{N-n+1}, S_{F[N-n+2, N]}^{-1}\left(\mathbf{x}_{N-n+1+i}\right)\right\rangle\right\rangle\left\langle\left\langle\mathbf{x}_{N-n+1+i}, \mathbf{x}_{N-n+1+j}\right\rangle\right\rangle & =\left\langle\left\langle\mathbf{x}_{N-n+1}, \mathbf{x}_{N-n+1+j}\right\rangle\right\rangle .
\end{aligned}
$$

Hence, once again by the induction hypothesis, for $j \in\{1 . . n-1\}$ the formula (35) reduces to

$$
\begin{equation*}
\left\langle\left\langle\mathbf{x}_{n}, \mathbf{x}_{j}\right\rangle\right\rangle=\left\langle\left\langle\mathbf{x}_{n-1}, \mathbf{x}_{j-1}\right\rangle\right\rangle . \tag{36}
\end{equation*}
$$

A similar argument using (26) shows that

$$
\begin{equation*}
\left\langle\left\langle\mathbf{x}_{N-n}, \mathbf{x}_{N-j}\right\rangle\right\rangle=\left\langle\left\langle\mathbf{x}_{N-n+1}, \mathbf{x}_{N-j+1}\right\rangle\right\rangle . \tag{37}
\end{equation*}
$$

If $j=0$, then $\operatorname{Err}_{M[0, n-1]^{d}}\left(\mathbf{x}_{0}\right)=0$ and

$$
\begin{aligned}
& \sum_{i=1}^{n-1}\left\langle\left\langle\mathbf{x}_{N-n+1}, S_{F[N-n+2, N]}^{-1}\left(\mathbf{x}_{N-n+1+i}\right)\right\rangle\right\rangle\left\langle\left\langle\mathbf{x}_{i}, \mathbf{x}_{0}\right\rangle\right\rangle= \\
= & \sum_{i=1}^{n-1}\left\langle\left\langle\mathbf{x}_{N-n+1}, S_{F[N-n+2, N]}^{-1}\left(\mathbf{x}_{N-n+1+i}\right)\right\rangle\right\rangle\left\langle\left\langle\mathbf{x}_{N-n+1+i}, \mathbf{x}_{N-n+1}\right\rangle\right\rangle= \\
= & \left\langle\left\langle\mathbf{P}_{M[N-n+2, N]^{d}}\left(\mathbf{x}_{N-n+1}\right), \mathbf{x}_{N-n+1}\right\rangle\right\rangle= \\
= & \left\langle\left\langle\mathbf{x}_{N-n+1}, \mathbf{x}_{N-n+1}\right\rangle\right\rangle-\operatorname{Err}_{M[N-n+2, N]^{d}}\left(\mathbf{x}_{N-n+1}\right) .
\end{aligned}
$$

Thus - in view of the above combined with the induction hypothesis - the formula (35) becomes:

$$
\begin{equation*}
\left\langle\left\langle\mathbf{x}_{n}, \mathbf{x}_{0}\right\rangle\right\rangle=\sum_{i=0}^{n-2}\left\langle\left\langle\mathbf{x}_{n-1}, S_{F[0, n-2]}^{-1}\left(\mathbf{x}_{i}\right)\right\rangle\right\rangle\left\langle\left\langle\mathbf{x}_{i+1}, \mathbf{x}_{0}\right\rangle\right\rangle+\left\langle\left\langle\mathbf{x}_{n}, S_{F[0, n-1]}^{-1}\left(\mathbf{x}_{0}\right)\right\rangle\right\rangle \operatorname{Err}_{M[N-n+2, N]^{d}}\left(\mathbf{x}_{N-n+1}\right) \tag{38}
\end{equation*}
$$

Similarly

$$
\begin{align*}
\left\langle\left\langle\mathbf{x}_{N-n}, \mathbf{x}_{N}\right\rangle\right\rangle & =\sum_{i=0}^{n-2}\left\langle\left\langle\mathbf{x}_{N-n+1}, S_{F[N-n+2, N]}^{-1}\left(\mathbf{x}_{N-i}\right)\right\rangle\right\rangle\left\langle\left\langle\mathbf{x}_{N-i-1}, \mathbf{x}_{N}\right\rangle\right\rangle  \tag{39}\\
& +\left\langle\left\langle\mathbf{x}_{N-n}, S_{F[N-n+1, N]}^{-1}\left(\mathbf{x}_{N}\right)\right\rangle\right\rangle \mathbf{E r r}_{M[0, n-2]^{d}}\left(\mathbf{x}_{n-1}\right)
\end{align*}
$$

The first terms on the right-hand sides of the formulas (38) and (39) are closely related, namely one is the adjoint of the other. This can be seen as follows. In view of the Theorem 4.1, (36) and (9),
we see that the term of interest in (38) is simply $\left\langle\left\langle\mathbf{x}_{n}, \mathbf{P}_{M[1, n-1]^{d}}\left(\mathbf{x}_{0}\right)\right\rangle\right\rangle$. Expanding the projection, using (36), the induction hypothesis and (37), we get the adjoint of the respective term in (37), modulo a change of the summation variable.

Also the second terms on the right-hand sides of (38) and (39) are mutually adjoint, because of (29). Consequently

$$
\begin{equation*}
\left\langle\left\langle\mathbf{x}_{0}, \mathbf{x}_{n}\right\rangle\right\rangle=\left\langle\left\langle\mathbf{x}_{N-n}, \mathbf{x}_{N}\right\rangle\right\rangle . \tag{40}
\end{equation*}
$$

Finally, if $j=n$, then by (36) and the induction hypothesis

$$
\begin{align*}
\sum_{i=0}^{n-2}\left\langle\left\langle\mathbf{x}_{n-1}, S_{F[0, n-2]}^{-1}\left(\mathbf{x}_{i}\right)\right\rangle\right\rangle\left\langle\left\langle\mathbf{x}_{i+1}, \mathbf{x}_{n}\right\rangle\right\rangle & =\sum_{i=0}^{n-2}\left\langle\left\langle\mathbf{x}_{n-1}, S_{F[0, n-2]}^{-1}\left(\mathbf{x}_{i}\right)\right\rangle\right\rangle\left\langle\left\langle\mathbf{x}_{i}, \mathbf{x}_{n-1}\right\rangle\right\rangle \\
& =\left\langle\left\langle\mathbf{x}_{n-1}, \mathbf{x}_{n-1}\right\rangle\right\rangle-\mathbf{E r r}_{M[0, n-2]^{d}}\left(\mathbf{x}_{n-1}\right)  \tag{41}\\
& =\left\langle\left\langle\mathbf{x}_{0}, \mathbf{x}_{0}\right\rangle\right\rangle-\operatorname{Err}_{M[0, n-2] d}\left(\mathbf{x}_{n-1}\right) .
\end{align*}
$$

Furthermore, by (40), the induction hypothesis and a change of summation variable

$$
\begin{align*}
& \left\langle\left\langle\mathbf{x}_{0}, \mathbf{x}_{n}\right\rangle\right\rangle-\sum_{i=1}^{n-1}\left\langle\left\langle\mathbf{x}_{N-n+1}, S_{F[N-n+2, N]}^{-1}\left(\mathbf{x}_{N-n+1+i}\right)\right\rangle\right\rangle\left\langle\left\langle\mathbf{x}_{i}, \mathbf{x}_{n}\right\rangle\right\rangle \\
& \left\langle\left\langle\mathbf{x}_{N-n}, \mathbf{x}_{N}\right\rangle\right\rangle-\sum_{i=0}^{n-2}\left\langle\left\langle\mathbf{x}_{N-n+1}, S_{F[N-n+2, N]}^{-1}\left(\mathbf{x}_{N-i}\right)\right\rangle\right\rangle\left\langle\left\langle\mathbf{x}_{N-i-1}, \mathbf{x}_{N}\right\rangle\right\rangle=  \tag{42}\\
& \left\langle\left\langle\mathbf{x}_{N-n}, S_{F[N-n+1, N]}^{-1}\left(\mathbf{x}_{N}\right)\right\rangle\right\rangle \operatorname{Err}_{M[0, n-2]^{d}}\left(\mathbf{x}_{n-1}\right),
\end{align*}
$$

where the last identity follows from (39). Now, substituting (41) and (42) into (35) (with $j=n$ ), and then using (27) we get the identity

$$
\begin{equation*}
\left\langle\left\langle\mathbf{x}_{n}, \mathbf{x}_{n}\right\rangle\right\rangle=\left\langle\left\langle\mathbf{x}_{0}, \mathbf{x}_{0}\right\rangle\right\rangle . \tag{43}
\end{equation*}
$$

Similarly, but using (28), we derive the identity

$$
\begin{equation*}
\left\langle\left\langle\mathbf{x}_{N-n}, \mathbf{x}_{N-n}\right\rangle\right\rangle=\left\langle\left\langle\mathbf{x}_{N}, \mathbf{x}_{N}\right\rangle\right\rangle . \tag{44}
\end{equation*}
$$

The principle of mathematical induction implies now weak stationarity of $\mathbf{X}$.

Corollary 5.2 Let $n, d \in \mathbb{N}$ and let $\operatorname{dim} H \geq d(n+1)$. If a sequence $\left\{\mathbf{x}_{1}, \ldots, \mathbf{x}_{n}\right\} \subset H^{d}$ is weakly stationary, then it can be extended to a weakly stationary sequence $\left\{\mathbf{x}_{1}, \ldots, \mathbf{x}_{n}, \mathbf{x}_{n+1}\right\} \subset H^{d}$, in such $a$ way that the components of $\mathbf{x}_{n+1}$ are not in $\boldsymbol{\operatorname { S p a n }}\left(\mathbf{x}_{1}, \ldots, \mathbf{x}_{n}\right)$.

Proof: Let $\mathbf{x}_{n+1} \in H^{d}$. Obviously the target sequence $\left\{\mathbf{x}_{m}, \ldots, \mathbf{x}_{n+1}\right\}$ is weakly stationary if and only if the sequences $\left\{\mathbf{x}_{j}: j=1, \ldots, n\right\}$ and $\left\{\mathbf{x}_{j}: j=2, \ldots, n+1\right\}$ have the same matrix covariance function. Define

$$
\mathbf{x}=\sum_{j=2}^{n}\left\langle\left\langle\mathbf{x}_{n}, S_{F}^{-1}\left(\mathbf{x}_{j-1}\right)\right\rangle\right\rangle \mathbf{x}_{j}
$$

where $F=\left\{\mathbf{x}_{1}, \ldots, \mathbf{x}_{n-1}\right\}$. Clearly for $k \in\{2, \ldots, n\}$,

$$
\left\langle\left\langle\mathbf{x}, \mathbf{x}_{k}\right\rangle\right\rangle=\left\langle\left\langle\mathbf{x}_{n}, \mathbf{x}_{k-1}\right\rangle\right\rangle
$$

and

$$
\langle\langle\mathbf{x}, \mathbf{x}\rangle\rangle=\left\langle\left\langle\mathbf{P}_{M^{d}}\left(\mathbf{x}_{n}\right), \mathbf{P}_{M^{d}}\left(\mathbf{x}_{n}\right)\right\rangle\right\rangle,
$$

where $M=\boldsymbol{\operatorname { S p a n }}\left(\mathbf{x}_{1}, \ldots, \mathbf{x}_{n-1}\right)$. It is now enough to define:

$$
\mathbf{x}_{n+1}=\mathbf{x}+\sqrt{\operatorname{Err}_{M^{d}}\left(\mathbf{x}_{n}\right)}(\sqrt{\langle\langle\mathbf{y}, \mathbf{y}\rangle\rangle})^{-1} \mathbf{y}
$$

where $\mathbf{y} \in\left(\left(M \& \mathbf{x}_{n}\right)^{\perp}\right)^{d}$ is an arbitrarily chosen block vector with linearly independent components.

Using induction we can easily deduce a more general statement:

Corollary 5.3 (Extension Property) Let $p, m, n, q \in \mathbb{Z}$, where $p \leq m<n \leq q$ and let $\operatorname{dim} H \geq$ $d(q-p+1)$. If a sequence $\left\{\mathbf{x}_{m}, \ldots, \mathbf{x}_{n}\right\} \subset H^{d}$ is weakly stationary, then it can be extended to $a$ weakly stationary sequence $\left\{\mathbf{x}_{p}, \ldots, \mathbf{x}_{q}\right\} \subset H^{d}$, in such a way that the components of the added block vectors are not in $\boldsymbol{\operatorname { S p a n }}\left(\mathbf{x}_{m}, \ldots, \mathbf{x}_{n}\right)$.

For earlier versions of the extension property see [13] and [22].

## 6 Causality

We will be working now in the Hilbert space $H=L^{2}(\Omega, \mathcal{F}, \mathbb{P})$, where $(\Omega, \mathcal{F}, \mathbb{P})$ is a given probability space. If $\mathcal{X} \subset H$ is a non-empty family of random variables, then by $\sigma(\mathcal{X})$ we denote the smallest $\sigma$-algebra with respect to which all members of $\mathcal{X}$ are measurable. The subspace $L^{2}(\Omega, \sigma(\mathcal{X}), \mathbb{P})$ will be called the information space generated by $\mathcal{X}$. If $\mathcal{X}$ if finite, say $\mathcal{X}=\left\{x_{1}, \ldots, x_{n}\right\}$, then it is well known that $\sigma(\mathcal{X})=\left(x_{1}, \ldots, x_{n}\right)^{-1}\left(\mathcal{B}_{n}\right) \subset \mathcal{F}$, where $\mathcal{B}_{n}$ denotes the $\sigma$-algebra of Borel subsets of $\mathbb{R}^{n}$.

Following [27], we say that a subspace $S \subset \mathcal{L}^{2}(\Omega, \mathcal{F}, \mathbb{P})$ is probabilistic if $S=\bar{S}, \mathbf{1}_{\Omega} \in S$, and $S$ is closed with respect to the operations $(x, y) \mapsto x \wedge y=\min (x, y)$ and $(x, y) \mapsto x \vee y=\max (x, y)$.

If $\emptyset \neq A \subset \mathcal{L}^{2}(\Omega, \mathcal{F}, \mathbb{P})$, then its lattice envelope $\operatorname{Latt}(A)$ is the smallest probabilistic subspace of $\mathcal{L}^{2}(\Omega, \mathcal{F}, \mathbb{P})$ containing $A$. Since a direct proof of the following property is not readily available in literature, we enclose it here.

Lemma 6.1 If $x_{1}, \ldots, x_{n} \in L^{2}(\Omega, \mathcal{F}, \mathbb{P})$, then

$$
\operatorname{Latt}\left(x_{1}, \ldots, x_{n}\right)=L^{2}\left(\Omega, \mathbf{x}^{-1}\left(\mathcal{B}_{n}\right), \mathbb{P}\right)
$$

where $\mathbf{x}=\left(x_{1}, \ldots, x_{n}\right)$ and $\mathcal{B}_{n}$ denotes the $\sigma$-algebra of Borel subsets of $\mathbb{R}^{n}$.

Proof: Let $\mathcal{G}=\mathbf{x}^{-1}\left(\mathcal{B}_{n}\right)$. Obviously $\operatorname{Latt}\left(x_{1}, \ldots, x_{n}\right) \subset \mathcal{L}^{2}(\Omega, \mathcal{G}, \mathbb{P})$. We want to show the opposite inclusion. We have:

$$
\begin{aligned}
\mathcal{L}^{2}(\Omega, \mathcal{G}, \mathbb{P}) & =\operatorname{Latt}\left(\left(\mathbf{1}_{G}\right)_{G \in \mathcal{G}}\right)=\mathbf{L a t t}\left(\left(\mathbf{1}_{G}\right)_{G \in \text { arbitrary } \pi \text {-system generating } \mathcal{G})}\right. \\
& =\boldsymbol{\operatorname { L a t t }}\left(\left\{\mathbf{1}_{\mathbf{x}^{-1}\left(\left(-\infty, c_{1}\right] \times \ldots \times\left(-\infty, c_{n}\right]\right)}:\left(c_{1}, \ldots, c_{n}\right) \in \mathbb{R}^{n}\right\}\right) \\
& =\boldsymbol{\operatorname { L a t t }}\left(\left\{\mathbf{1}_{y^{-1}((-\infty, c])}: c \in \mathbb{R}, y \in\left\{x_{1}, \ldots, x_{n}\right\}\right\}\right) .
\end{aligned}
$$

(Recall that a $\pi$-system is a nonempty family of sets closed with respect to finite intersections.) We want to show that the last set is a subspace of $\boldsymbol{\operatorname { L a t }}\left(x_{1}, \ldots, x_{n}\right)$. Take $\epsilon>0$ and $c \in \mathbb{R}$. Let

$$
A=y^{-1}((-\infty, c]), \quad B_{m}=y^{-1}((c+1 / m, \infty)) \quad C_{m}=y^{-1}((c, c+1 / m])
$$

where $m$ is chosen so that $\mathbb{P}\left(C_{m}\right)<\epsilon$. Define

$$
y_{m}=\mathbf{1}_{\Omega}-m\left[(y \vee c-c) \wedge \frac{1}{m}\right]
$$

Then

$$
\left.y_{m}\right|_{A} \equiv 1,\left.\quad y_{m}\right|_{B_{m}} \equiv 0, \quad y_{m}\left(C_{m}\right) \subset[0,1]
$$

Therefore

$$
\int_{\Omega}\left(\mathbf{1}_{A}-y_{m}\right)^{2} d \mathbb{P}=\int_{C_{m}} y_{m}^{2} d \mathbb{P} \leq \mathbb{P}\left(C_{m}\right)<\epsilon
$$

Since $\boldsymbol{\operatorname { L a t t }}\left(x_{1}, \ldots, x_{n}\right)$ is a closed subset, $\mathbf{1}_{A} \in \operatorname{Latt}\left(x_{1}, \ldots, x_{n}\right)$.

If $\mathcal{X}=\left(x_{n}\right)_{n \in \mathbb{N}}$, then

$$
\begin{equation*}
L^{2}(\Omega, \sigma(\mathcal{X}), \mathbb{P})=\overline{\bigcup_{n \in \mathbb{N}} L^{2}\left(\Omega,\left(x_{1}, \ldots, x_{n}\right)^{-1}\left(\mathcal{B}_{n}\right), \mathbb{P}\right)} \tag{45}
\end{equation*}
$$

Clearly

$$
\mathbf{P}_{L^{2}(\Omega, \sigma(\mathcal{X}), \mathbb{P})}(Z)=\lim _{n \rightarrow \infty} \mathbf{P}_{L^{2}\left(\Omega,\left(x_{1}, \ldots, x_{n}\right)^{-1}\left(\mathcal{B}_{n}\right), \mathbb{P}\right)}(Z), \quad Z \in H
$$

We say that there exists a causality relation between $\mathcal{X}$ (the cause) and a random variable $y$ (the effect), and we write

$$
\mathcal{X} \xrightarrow{C} y,
$$

if $y \in L^{2}(\Omega, \sigma(\mathcal{X}), \mathbb{P})$. If $\mathcal{X}$ is finite, say $\mathcal{X}=\left\{x_{1}, \ldots, x_{n}\right\}$, then such causality relation means simply that for some Borel function $F$ of $n$-real variables $y=F\left(x_{1}, \ldots, x_{n}\right)$.

We say that there is a linear causality relation between $\mathcal{X}$ (the cause) and a random variable $y$ (the effect), and we write

$$
\mathcal{X} \xrightarrow{L C} y,
$$

if $y \in \mathbf{S p a n}(\mathcal{X})$.
These definitions can be extended in an obvious way to the case of vector valued families of random variables. Suppose that $\mathcal{X}=\left(\mathbf{x}_{j}\right)_{j \in J} \subset H^{d}$ and $\mathbf{y} \in H^{d^{\prime}}$, where $\mathbf{x}_{j}=\left(x_{1 j}, \ldots, x_{d j}\right)$ and $\mathbf{y}=$ $\left(y_{1}, \ldots, y_{d^{\prime}}\right)$. Let $\underline{\mathcal{X}}=\left\{x_{i j}: i=1, \ldots, d, j \in J\right\}$. Then by definition

$$
\begin{aligned}
\mathcal{X} \xrightarrow{L C} \mathbf{y} & \Longleftrightarrow \underline{\mathcal{X}} \xrightarrow{L C} y_{k} \text { for each } k=1, \ldots, d^{\prime} \\
\mathcal{X} \xrightarrow{C} \mathbf{y} & \Longleftrightarrow \underline{\mathcal{X}} \xrightarrow{C} y_{k} \text { for each } k=1, \ldots, d^{\prime} .
\end{aligned}
$$

Let $\mathcal{X}=\left\{\mathbf{x}_{j}: j \in J\right\}$ be a family of $\mathbb{R}^{d}$-valued random variables, where $J$ is finite or not. A family of real-valued Borel functions $\mathbb{F}$ of several real variables is called an information generating function system (or IGFS) for $\mathcal{X}$, if it is at most countable and for any finite subset $\mathcal{Y}=\left\{\mathbf{y}_{1}, \ldots, \mathbf{y}_{n}\right\}$ of $\mathcal{X}$, there exists a subset $\mathbb{G} \subset \mathbb{F}$ such that the random variables

$$
\left\{G\left(\mathbf{y}_{1}, \ldots, \mathbf{y}_{n}\right): G \in \mathbb{G}\right\}
$$

are linearly dense in the information space generated by $\underline{\mathcal{Y}}$. Just as before $\underline{\mathcal{Y}}=\left\{y_{i j}: i=\right.$ $1, \ldots, d, j=1, \ldots, n\}$, where $\mathbf{y}_{j}=\left(y_{1 j}, \ldots, y_{d j}\right)$. In other words we have then

$$
\operatorname{Span}\left\{G\left(\mathbf{y}_{1}, \ldots, \mathbf{y}_{n}\right): G \in \mathbb{G}\right\}=L^{2}\left(\Omega,\left(\mathbf{y}_{1}, \ldots, \mathbf{y}_{n}\right)^{-1}\left(\mathcal{B}_{n d}\right), \mathbb{P}\right) .
$$

Example 6.2 A result of Dobrushin and Minlos from 1977 (see [3]) yields a very important example of an IGFS. Suppose that $x_{1}, \ldots, x_{n}$ are random variables such that

$$
\begin{equation*}
e^{\left|x_{j}\right|} \in \mathcal{L}^{\alpha}(\Omega, \mathcal{F}, \mathbb{P}) \tag{46}
\end{equation*}
$$

for some $\alpha>0$ and for $j=1, \ldots, n$. Suppose also that $\mathcal{P}\left(\mathbb{R}^{n}\right)$ denotes the family of all real polynomials of $n$-real variables. Then any sequence $\left(P_{j}\right)_{j \in \mathbb{N}} \subset \mathcal{P}\left(\mathbb{R}^{n}\right)$, with the property that every polynomial in $\mathcal{P}\left(\mathbb{R}^{n}\right)$ is a linear combination of elements of this sequence, is an IGFS for $x_{1}, \ldots, x_{n}$. In fact, since we are dealing with powers of random variables it can be shown that the vector space $V\left(x_{1}, \ldots, x_{n}\right)=\left\{P\left(x_{1}, \ldots, x_{n}\right): P \in \mathcal{P}\left(\mathbb{R}^{n}\right)\right\}$ is not only a subset of $L^{p}(\Omega, \mathcal{F}, \mathbb{P})$ (for each $p \geq 1$ ), but is also dense there. Interestingly, the property that $V\left(x_{1}, \ldots, x_{n}\right) \subset L^{p}(\Omega, \mathcal{F}, \mathbb{P})$ for all $p \geq 1$ and for some random variables $x_{1}, \ldots, x_{n}$, does not imply that (46) is satisfied. To see this consider $\Omega=\mathbb{R}, \mathcal{F}=\mathcal{B}_{1}$ and the probability measure

$$
\mathbb{P}(A)=\frac{\sum_{m=1}^{\infty} m^{-\ln m} \mathbf{1}_{A}(m)}{\sum_{m=1}^{\infty} m^{-\ln m}}, \quad A \in \mathcal{F} .
$$

We have

$$
\frac{m^{k}}{m^{\ln m}}<\frac{1}{m^{2}}
$$

if $k \geq 1$, provided that $m$ is large enough. Also, for any $\alpha>0$

$$
\lim _{m \rightarrow \infty} \sqrt[m]{\frac{e^{\alpha m}}{m^{\ln m}}}=e^{\alpha}>1
$$

Thus every polynomial $P(x)$ is in all the spaces $L^{p}$ for $p \geq 1$, but there is no positive $\alpha$ for which $e^{x}$ is in $L^{\alpha}$.

Example 6.3 Suppose that the range $\mathbf{x}(\Omega)$ of the random vector $\mathbf{x}=\left(x_{1}, \ldots, x_{n}\right): \Omega \longrightarrow \mathbb{R}^{n}$ is countable and $\mathrm{E}|\mathbf{x}|<\infty$. Then $\left\{\mathbf{1}_{\{a\}}: a \in \mathbf{x}(\Omega)\right\}$ is an IGFS for $x_{1}, \ldots, x_{n}$. Note that in this case the information space generated by the random variables $x_{1}, \ldots, x_{n}$ is given by

$$
\left\{\sum_{a \in \mathbf{x}(\Omega)} \lambda(a) \mathbf{1}_{\mathbf{x}^{-1}(a)}: \lambda \in L^{2}\left(\mathbf{x}(\Omega), 2^{\mathbf{x}(\Omega)}, \mathbb{P} \circ \mathbf{x}^{-1}\right)\right\} .
$$

The concept of information generating function systems is useful if one wants to approximate the conditional expectation $E\left[y \mid x_{1}, \ldots, x_{n}\right]$, where $y, x_{1}, \ldots, x_{n}$ are random variables with finite variance. The conditional expectation is in this situation simply the orthogonal projection of $Y$ onto the information space generated by $x_{1}, \ldots, x_{n}$. Unfortunately the information space is usually infinite-dimensional making the problem computationally difficult. In applications, a common way of avoiding this difficulty is to assume that the random variables $y, x_{1}, \ldots, x_{n}$ have a joint Gaussian distribution in which case $E\left[y \mid x_{1}, \ldots, x_{n}\right]$ is a linear combination of the constant function $\mathbf{1}$ and the random variables $x_{1}, \ldots, x_{n}$. However, this additional assumption is often quite unrealistic. The general problem is quite pertinent to the distinction between linear and non-linear causality which we we are going to address now.

In the proof of the next theorem we will need the following easy to check property of orthogonal projections.

Lemma 6.4 Suppose $E_{1} \subset E_{2} \subset \ldots E_{n} \subset \ldots H$ is a sequence of a closed subspaces of a Hilbert space H. Let

$$
E=\overline{\bigcup_{j \in \mathbb{N}} E_{j}}
$$

Then for any $\mathbf{x} \in H$ the sequence $\left\|\mathbf{P}_{E_{j}}(\mathbf{x})\right\|$ is non-decreasing and $\mathbf{P}_{E_{j}}(\mathbf{x}) \rightarrow \mathbf{P}_{E}(\mathbf{x})$ as $j \rightarrow \infty$.

Linear causality can be characterized in several ways (see [20] and [26] for earlier versions):

Theorem 6.5 Let $\left(\mathbf{x}_{j}\right)_{j \in \mathbb{Z}} \subset H^{d}$ and $\left(\mathbf{y}_{j}\right)_{j \in \mathbb{Z}} \subset H^{d^{\prime}}$ be such that the sequence $\left(\mathbf{x}_{j}, \mathbf{y}_{j}\right)_{j \in \mathbb{Z}} \subset H^{d+d^{\prime}}$ is weakly stationary. The following conditions are equivalent:
(a) $\left\{\mathbf{x}_{j}\right\}_{-\infty<j \leq 0} \xrightarrow{L C} \mathbf{y}_{0}$;
(b) $\left\{\mathbf{x}_{j}\right\}_{-\infty<j \leq n} \xrightarrow{L C} \mathbf{y}_{n}$ for some $n \in \mathbb{Z}$;
(c) $\left\{\mathbf{x}_{j}\right\}_{-\infty<j \leq m} \xrightarrow{L C} \mathbf{y}_{m}$ for each $m \in \mathbb{Z}$;
(d) If $n \rightarrow \infty$, then

$$
\left\|\mathbf{P}_{M[0, n] d^{\prime}}\left(\mathbf{y}_{n}\right)\right\| \nearrow\left\|\mathbf{y}_{0}\right\|
$$

where $M[0, n]$ is defined as in (23).

Proof: Obviously $(\mathrm{a}) \Rightarrow(\mathrm{b})$. Assume that (b) is satisfied with some value of $n$ and pick $m \in \mathbb{Z}$. Since $\left(\mathbf{y}_{j}\right)_{j \in \mathbb{Z}}$ is weakly stationary,

$$
\left\|\mathbf{y}_{n}\right\|=\sqrt{\operatorname{trace}\left\langle\left\langle\mathbf{y}_{n}, \mathbf{y}_{n}\right\rangle\right\rangle}=\sqrt{\operatorname{trace}\left\langle\left\langle\mathbf{y}_{m}, \mathbf{y}_{m}\right\rangle\right\rangle}=\left\|\mathbf{y}_{m}\right\| .
$$

But in view of Lemma 6.4 and Corollary 4.2

$$
\begin{aligned}
\left\|\mathbf{y}_{n}\right\| & =\left\|\mathbf{P}_{M[-\infty, n] d^{\prime}}\left(\mathbf{y}_{n}\right)\right\|=\lim _{k \rightarrow \infty}\left\|\mathbf{P}_{M[n-k, n]^{d^{\prime}}}\left(\mathbf{y}_{n}\right)\right\| \\
& =\lim _{k \rightarrow \infty}\left\|\mathbf{P}_{M[m-k, m]^{d^{\prime}}}\left(\mathbf{y}_{m}\right)\right\|=\left\|\mathbf{P}_{M[-\infty, m]^{d^{\prime}}}\left(\mathbf{y}_{m}\right)\right\| .
\end{aligned}
$$

Hence

$$
\left\|\mathbf{P}_{M[-\infty, m]^{d^{\prime}}}\left(\mathbf{y}_{m}\right)\right\|=\left\|\mathbf{y}_{m}\right\|
$$

which means that $\left\{\mathbf{x}_{j}\right\}_{-\infty<j \leq m} \xrightarrow{C} \mathbf{y}_{m}$. Since $m$ was arbitrarily chosen we get (c) Now if (c) holds, then in particular

$$
\begin{equation*}
\left\|\mathbf{P}_{M[-n, 0]^{d^{\prime}}}\left(\mathbf{y}_{0}\right)\right\| \nearrow\left\|\mathbf{y}_{0}\right\| \text { as } n \rightarrow \infty \tag{47}
\end{equation*}
$$

but then by Corollary 4.2 we see that (d) is equivalent to (47). The latter yields (a).
If any of these equivalent conditions are satisfied, we will say that there is a linear causality relationship between $\left(\mathbf{x}_{j}\right)_{j \in \mathbb{Z}}$ (the cause) and $\left(\mathbf{y}_{j}\right)_{j \in \mathbb{Z}}$ (the effect). In this case we write

$$
\left(\mathbf{x}_{j}\right)_{j \in \mathbb{Z}} \xrightarrow{L C}\left(\mathbf{y}_{j}\right)_{j \in \mathbb{Z}} .
$$

By analogy to linear causality, we will say that there is a causality relationship between $\left(\mathbf{x}_{j}\right)_{j \in \mathbb{Z}}$ (the cause) and $\left(\mathbf{y}_{j}\right)_{j \in \mathbb{Z}}$ (the effect), if $\left(\mathbf{x}_{j}\right)_{-\infty j \leq n}$ is the cause of $\mathbf{y}_{n}$ for each $n \in \mathbb{Z}$. In symbols

$$
\left[\left(\mathbf{x}_{j}\right)_{j \in \mathbb{Z}} \xrightarrow{C}\left(\mathbf{y}_{j}\right)_{j \in \mathbb{Z}}\right] \Longleftrightarrow\left[\left(\mathbf{x}_{j}\right)_{-\infty j \leq n} \xrightarrow{C} \mathbf{y}_{n} \text { for each } n \in \mathbb{Z} .\right]
$$

A natural questions can be asked here. How to detect and measure linear causality when dealing with empirical data? The answer can be obtained with the help of property (d) in the above theorem combined with the computational recipe provided in the following statement.

Theorem 6.6 Let $F=\left\{\mathbf{x}_{m}, \ldots, \mathbf{x}_{n}\right\} \subset H^{d}$ and $\mathbf{y} \in H^{d^{\prime}}$, where $m<n$. Define

$$
\nu(j)= \begin{cases}\mathbf{x}_{m}, & \text { if } j=0  \tag{48}\\ \mathbf{P}_{\left(M[m, m+j-1]^{\perp}\right)^{d}}\left(\mathbf{x}_{m+j}\right), & \text { if } j=1, \ldots, n-m\end{cases}
$$

where $M[\cdot, \cdot]$ is defined as in (23). Define also

$$
\alpha(j)= \begin{cases}\left\langle\left\langle\mathbf{x}_{m}, \mathbf{y}\right\rangle\right\rangle, & \text { if } j=0  \tag{49}\\ \left\langle\left\langle\mathbf{x}_{m+j}, \mathbf{y}\right\rangle\right\rangle-\sum_{k=0}^{j-1}\left\langle\left\langle\mathbf{x}_{m+j}, S_{F_{j-1}}^{-1}\left(\mathbf{x}_{m+k}\right)\right\rangle\right\rangle\left\langle\left\langle\mathbf{x}_{m+k}, \mathbf{y}\right\rangle\right\rangle, & \text { if } j=1, \ldots, n-m,\end{cases}
$$

where $F_{j}=\left\{\mathbf{x}_{m}, \ldots, \mathbf{x}_{m+j}\right\}$. Then

$$
\begin{equation*}
\left\|\mathbf{P}_{M[m, n]^{d^{\prime}}}(\mathbf{y})\right\|=\sqrt{\operatorname{trace}\left\{\sum_{j=0}^{n-m} \alpha(j)^{*}\langle\langle\nu(j), \nu(j)\rangle\rangle^{\dagger} \alpha(j)\right\}} \tag{50}
\end{equation*}
$$

Proof: As a consequence of (17) and Remark 3.3

$$
\mathbf{P}_{M[m, n]^{d^{\prime}}}(\mathbf{y})=\sum_{j=0}^{n-m}\langle\langle\mathbf{y}, \nu(j)\rangle\rangle\langle\langle\nu(j), \nu(j)\rangle\rangle^{\dagger} \nu(j),
$$

and hence

$$
\left.\left.\left\|\mathbf{P}_{M[m, n]^{d^{\prime}}}(\mathbf{y})\right\|^{2}=\operatorname{trace}\left\{\sum_{j=0}^{n-m}\langle\mathbf{y}, \nu(j)\rangle\right\rangle\langle\nu(j), \nu(j)\rangle\right\rangle^{\dagger}\langle\langle\nu(j), \mathbf{y}\rangle\rangle\right\},
$$

because if $A$ is a matrix then $A^{\dagger} A A^{\dagger}=A^{\dagger}$ and $\nu(i) \perp \nu(j)$ for $i \neq j$. Now it is enough to observe that

$$
\nu(j)=\mathbf{x}_{m+j}-\sum_{k=0}^{j-1}\left\langle\left\langle\mathbf{x}_{m+j}, S_{F_{j-1}}^{-1}\left(\mathbf{x}_{m+k}\right)\right\rangle\right\rangle \mathbf{x}_{m+k},
$$

and $\alpha(j)=\langle\langle\nu(j), \mathbf{y}\rangle\rangle$.
The theorem generalizes a slightly different formula obtained in [20] under the assumption that $d^{\prime}=1$ and $\langle\langle\nu(j), \nu(j)\rangle$ is invertible. Note that the blueprint algorithm allows calculation of all necessary quantities if the matrix covariance function for $\left\{\mathbf{x}_{m}, \ldots, \mathbf{x}_{n}, \mathbf{y}\right\}$ is known.

It was shown in [20] (see also [26] and [13]) that tests for linear causality can be adapted to deal with the non-linear case. We will show that these earlier results can be generalized within the framework adopted in this paper. Before proving a "non-linear" counterpart of Theorem 6.5 we need a lattice version of Corollary 4.2.

Lemma 6.7 Let $\left\{\mathbf{x}_{m}^{(i)}, \ldots, \mathbf{x}_{n}^{(i)}\right\} \subset H^{d}$ and $\mathbf{y}^{(i)} \in H^{d^{\prime}}$, where $i=1,2$. Let $N_{(i)}=\boldsymbol{\operatorname { L a t t }}\left(\mathbf{x}_{m}^{(i)}, \ldots, \mathbf{x}_{n}^{(i)}\right)$. If $\left(\mathbf{x}_{m}^{(1)}, \ldots, \mathbf{x}_{n}^{(1)}, \mathbf{y}^{(1)}\right)$ and $\left(\mathbf{x}_{m}^{(2)}, \ldots, \mathbf{x}_{n}^{(2)}, \mathbf{y}^{(2)}\right)$ have the same probability distribution, then

$$
\left\|\mathbf{P}_{N_{(1)}^{d^{\prime}}}\left(\mathbf{y}^{(1)}\right)\right\|=\left\|\mathbf{P}_{N_{(2)}^{d^{\prime}}}\left(\mathbf{y}^{(2)}\right)\right\| .
$$

Proof: Let $z_{1}, \ldots, z_{k}$ denote the scalar components of the block vectors $\mathbf{x}_{m}^{(i)}, \ldots, \mathbf{x}_{n}^{(i)}$ (where $i$ is fixed). Let $y$ denote a chosen component of $\mathbf{y}^{(i)}$. It is enough to show that the norm of $\mathrm{E}\left[y \mid z_{1}, \ldots, z_{k}\right]=\mathbf{P}_{\mathbf{L a t t}\left(z_{1}, \ldots, z_{k}\right)}(y)$ does not depend on the choice of $i$.

Without loss of generality we may assume that $z_{1}, \ldots, z_{k} \geq 0$ (because if $z$ is in a probabilistic subspace, then so are $z \vee 0$ and $-(z \wedge 0)$ ). Note that for any $j$, we have the $L^{2}$-convergence $z_{j} \wedge t \longrightarrow z_{j}$ as $t \rightarrow \infty$, where $t \in \mathbb{N}$. Thus

$$
\overline{\bigcup_{t \in \mathbb{N}} \operatorname{Latt}\left(z_{1} \wedge t, \ldots, z_{k} \wedge t\right)}=\boldsymbol{\operatorname { L a t t }}\left(z_{1}, \ldots, z_{k}\right)
$$

and consequently - in view of Lemma 6.4 - we may suppose that $z_{1}, \ldots, z_{k}$ are bounded. Since under this assumption (46) is obviously satisfied, we can use the result of Dobrushin and Minlos
described in Example 6.2 and approximate $E=\boldsymbol{\operatorname { L a t t }}\left(z_{1}, \ldots, z_{k}\right)$ with a non-decreasing family of subspaces $E_{j}=\operatorname{Span}\left\{\pi\left(z_{1}, \ldots, z_{k}\right): \pi \in \mathcal{P}^{(j)}\right\}$, where $\mathcal{P}^{(j)}$ is a finite family of polynomials. Now it is enough to use Corollary 4.2 and Lemma 6.4.

Let $J=\{p . . q\}$ be an integer interval for some $p, q \in \mathbb{Z} \cup\{ \pm \infty\}$. Recall that a sequence $\left(\mathbf{z}_{j}\right)_{j \in J} \subset H^{d}$ is said to be strictly stationary, if for any choice of integers $m, s \in \mathbb{N}$ and $n_{1}, \ldots, n_{m} \in J$ such that $n_{1}+s, \ldots, n_{m}+s \in J$, the random vectors $\left(\mathbf{z}_{n_{1}}, \ldots, \mathbf{z}_{n_{m}}\right)$ and ( $\mathbf{z}_{n_{1}+s}, \ldots, \mathbf{z}_{n_{m}+s}$ ) have the same probability distribution. Obviously strict stationarity implies weak stationarity. Moreover if the random vectors are chosen like in the definition and $h: \mathbb{R}^{d \times m} \longrightarrow \mathbb{R}^{n}$ is a Borel function (for some $n)$, then the random vectors $h\left(\mathbf{z}_{n_{1}}, \ldots, \mathbf{z}_{n_{m}}\right)$ and $h\left(\mathbf{z}_{n_{1}+s}, \ldots, \mathbf{z}_{n_{m}+s}\right)$ have the same probability distribution.

Theorem 6.8 Let $\left(\mathbf{x}_{j}\right)_{j \in \mathbb{Z}} \subset H^{d}$ and $\left(\mathbf{y}_{j}\right)_{j \in \mathbb{Z}} \subset H^{d^{\prime}}$ be such that the sequence $\left(\mathbf{x}_{j}, \mathbf{y}_{j}\right)_{j \in \mathbb{Z}} \subset H^{d+d^{\prime}}$ is strictly stationary. Let $N[p, q]$ denote the information space generated by all the components of $\mathbf{x}_{j}$, where $j \in \mathbb{Z} \cap[p, q]$ and $-\infty \leq p \leq q \leq \infty$. The following conditions are equivalent:
(a) $\left\{\mathbf{x}_{j}\right\}_{-\infty<j \leq 0} \xrightarrow{C} \mathbf{y}_{0}$;
(b) $\left\{\mathbf{x}_{j}\right\}_{-\infty<j \leq n} \xrightarrow{C} \mathbf{y}_{n}$ for some $n \in \mathbb{Z}$;
(c) $\left\{\mathbf{x}_{j}\right\}_{-\infty<j \leq m} \xrightarrow{C} \mathbf{y}_{m}$ for each $m \in \mathbb{Z}$;
(d) If $n \rightarrow \infty$, then

$$
\left\|\mathbf{P}_{N[0, n]^{\prime}}\left(\mathbf{y}_{n}\right)\right\| \nearrow\left\|\mathbf{y}_{0}\right\|
$$

Proof: The proof is similar to that of Theorem 6.5 except that we have to use $N[p, q]$ and Lemma 6.7 instead of $M[p, q]$ and Corollary 4.2, respectively.

Note that the $k$-th component of $\mathbf{P}_{N[0, n]{ }^{d^{\prime}}}\left(\mathbf{y}_{n}\right)$ is simply the conditional expectation of the $k$-th component of $\mathbf{y}_{n}$, conditioned on the $\sigma$-algebra generated by the components of $\mathbf{x}_{0}, \ldots, \mathbf{x}_{n}$.

For earlier versions of the above property see [20] and [26].
We are going to close the paper with a few comments explaining how information generating function systems can enter practical time series analysis or, in other words, how the above linear theory can account for non-linear phenomena. As an example we will consider a relatively recent procedure called the abnormality test or $\operatorname{Test}(\mathrm{ABN})$ developed in [24] in connection with analysis of stock market data. Apart from possible use for financial risk analysis (see also [28]), the abnormality test is useful in detection of early signs of deep low-frequency earthquakes (see [29]).

Definition of moving compositions: Let $A, B \neq \emptyset$ and let $k \in \mathbb{Z}_{+}$. Let $J$ be an integer interval. If a (finite or infinite) sequence $\mathbf{X}: J \longrightarrow A$ of elements of the set $A$ is given and $f: A^{k+1} \longrightarrow B$ is a given function, then the moving composition of $f$ with $\mathbf{X}$ is the sequence of elements of the set $B$ defined by the formula

$$
(f \circ \mathbf{X})(n)=f(\mathbf{X}(n-k), \mathbf{X}(n-k+1), \ldots, \mathbf{X}(n-1), \mathbf{X}(n))
$$

for all values of $n$ for which the right-hand side of the above formula makes sense.
Suppose that a time series $\mathbf{X}$ represents some empirical data. Test(ABN) is meant to screen the given time series for statistical anomalies or - to be more descriptive - for breakdown of stationarity. A finite family of Borel functions $\mathcal{P}$ is chosen. For $p, q \in \mathcal{P}, p \neq q$, we form

$$
\binom{\# \mathcal{P}}{2} \text { time series } \mathbf{X}_{p q}=\binom{p \circ \mathbf{X}}{q \circ \mathbf{X}}
$$

We apply a weak stationarity test to all $\mathbf{X}_{p q}$ and count the number $\mathbf{n}_{\mathbf{X}}(t)$ of positive results at time $t$. We say that a period of anomalies starts at time $t$ if $\mathbf{n}_{\mathbf{X}}(t) \leq M$ but $\mathbf{n}_{\mathbf{X}}(t-1)>M$, where $M$ is a chosen threshold value. Test(ABN) can also be applied simultaneously to several concurrent time series, say $\mathbf{X}_{1}, \ldots, \mathbf{X}_{n}$, in which case the behavior of $\min \left\{\mathbf{n}_{\mathbf{X}_{j}}(t): j=1, \ldots, n\right\}$ may serve as an indicator of breakdown of stationarity.

## References

[1] P. Casazza, The Art of Frame Theory, Taiwanese J. Math., 4(2) (2000), pp. 129-201.
[2] S. DÉgerine, Canonical partial autocorrelation function of a multivariate time series, Ann. Statist., 18 (1990), pp. 961-971.
[3] R.L. Dobrushin and R.A. Minlos, Polynomials in linear random functions, Russian Math. Surveys, 32(2) (1977), pp. 71-127.
[4] J. Durbin, The fitting of time series models, Rev. Int. Inst. Stat., 28 (1960), pp. 233-244.
[5] Y. Inouye, Modeling of multichannel time series and extrapolation of matrix-valued autocorrelation sequences, IEEE Trans. Acoust., Speech, Signal Process., 31 (1983), pp. 45-55.
[6] Y. Inouye, Autoregressive model fitting for multichannel time series of degenerate rank: Limit properties, IEEE Trans. Circuits and Systems, 32 (1985), pp. 252-259.
[7] M. Klimek, E. Karlsson, M. Matsuura and Y. Okabe, A geometric proof of the fluctuation-dissipation theorem for $\mathrm{KM}_{2} \mathrm{O}$-Langevin equation, Hokkaido Math. J., 31(3) (2002), pp. 615-628.
[8] M. Klimek, M. Matsuura and Y. Okabe, Stochastic flows and finite block frames, Journal of Mathematical Analysis and Applications, 342(2) (2008), pp. 816-829.
[9] H. Lev-Ari and T. Kailath, Schur and Levinson algorithms for nonstationary processes, Proceedings of 1981 IEEE Int. Conf. on Acoust., Speech, Signal Process., (1981), pp. 860-864.
[10] H. Lev-Ari and T. Kailath, Lattice Filter Parametrization and Modelling of Nonstationary processes, IEEE Trans. Inf. Theory, IT-30(1) (1984), pp. 2-16.
[11] N. Levinson, The Wiener rms error criterion in filter design and prediction, J. Math. Phys., 25 (1947), pp. 261-278.
[12] M. Matsuura and Y. Okabe, On a non-linear prediction problem for one-dimensional stochastic processes, Japanese Journal of Mathematics, $27(1)$ (2001), pp. 51-112.
[13] M. Matsuura and Y. Okabe, On the theory of $K_{2} O$-Langevin equations for non-stationary and degenerate flows, J. Math. Soc. Japan, Vol., 55(2) (2003), pp. 523-563.
[14] M. Matsuura, On a New Fluctuation-Dissipation Theorem for Degenerate Stationary Flows, Methodology and Computing in Applied Probability, 5 (2003), pp. 369-387.
[15] M. Matsuura, Relations among the minimum norm coefficients for degenerate non-stationary flows, SIAM J. Matrix Anal. Appl., 27(3) (2005), pp. 654-664.
[16] Y. Okabe, Langevin equations and causal analysis. In: Selected papers on analysis, probability, and statistics, pp. 19-50, Amer. Math. Soc. Transl. Ser. 2, 161, Amer. Math. Soc., Providence, RI, 1994.
[17] Y. Okabe, On the theory of $K M_{2} O$-Langevin equations for stationary flows (1): characterization theorem, J. Math. Soc. Japan, 51 (1999), pp. 817-841.
[18] Y. Okabe, On the theory of $\mathrm{KM}_{2} \mathrm{O}$-Langevin equations for stationary flows (2): construction theorem, Acta. Appl. Math., 63(1-3) (2000), pp. 307-322.
[19] Y. Okabe, Experimental mathematics - analysis of time series, earthquakes, northern lights, brain waves and voices, Asakura Shoten, Tokyo 2005 (in Japanese).
[20] Y. Okabe and A. Inoue, The theory of $\mathrm{KM}_{2} \mathrm{O}$-Langevin equations and its applications to data analysis (II): Causal analysis (1), Nagoya Math J., 134 (1994), pp. 1-28.
[21] Y. Okabe and A. Kaneko, On a non-linear prediction analysis for multi-dimensional stochastic processes with its applications to data analysis, Hokkaido Math. J., 29 (2000), pp. 601-657.
[22] Y. Okabe and M. Matsuura, On the theory of $\mathrm{KM}_{2} \mathrm{O}$-Langevin equations for stationary flows (3): extension theorem, Hokkaido Math. J., 29 (2000), pp. 369-382.
[23] Y. Okabe and M. Matsuura, On a non-linear filtering problems for discrete time stochastic processes, J. Math. Soc. Japan, Vol. 57 (2005), pp. 1067-1076.
[24] Y. Okabe, M. Matsuura and M. Klimek, On a method for detecting certain signs of stock market crashes by non-linear stationarity tests, Int. J. Pure Appl Math., 3(4) (2002), pp. 443-484.
[25] Y. Okabe and Y. Nakano, The theory of $\mathrm{KM}_{2} \mathrm{O}$-Langevin equations and its applications to data analysis (I): Stationary analysis, Hokkaido Math. J., 20 (1991), pp. 45-90.
[26] Y. Okabe and T. Yamane, The theory of $\mathrm{KM}_{2} \mathrm{O}$-Langevin equations and its applications to data analysis (III): Deterministic analysis, Nagoya Math. J., 152 (1998), pp. 175-201.
[27] C.G. Small and D.L. McLeish, Hilbert Space Methods in Probability and Statistical Inference, John Wiley, New York 1994.
[28] K. Suzuki, Y. Okabe and T. Fujir, On a Non-linear Risk Analysis for Stock Market Indexes, Asia-Pacific Financial Markets, 13 (2006), pp. 235-258.
[29] M. Takeo, H. Ueda, Y. Okabe and M. Matsuura, Waveform characteristics of deep low-frequency earthquakes: time-series evolution based on the theory of the $\mathrm{KM}_{2} \mathrm{O}$-Langevin equation, Geophys. J. Int., 165 (2006), pp. 87-107.
[30] P. Whittle, On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix, Biometrika, 50 (1963), pp. 129-134.
[31] R.A. Wiggins and E.A. Robinson, Recursive solution to the multichannel filtering problem. J. Geophys. Res., 70 (1965), pp. 1885-1891.
[32] A.E. Yagle, Fast algorithms for structured matrices in signal processing. In: Handbook of Statistics, Vol 10, Ed. by N.K.Bose and C.R. Rao, North-Holland, Amsterdam 1993, pp. 933972.

Maciej Klimek
Department of Mathematics, Uppsala University, P.O.Box 480, 75106 Uppsala, SWEDEN.
e-mail: maciej.klimek@maths.uu.se


[^0]:    ${ }^{1}$ In fact much more general results can be found in literature also for infinite $J$.

[^1]:    ${ }^{2}$ See also Remark 3.3.

