2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

A note on Generalized Factor Analysis models

Giulio Bottegal and Giorgio Picci

***Abstract*— An interesting generalization of dynamic factor analysis models has been proposed recently by Forni, Lippi and collaborators. These models, called *generalized dynamic factor analysis models* describe observations of infinite cross-sectional dimension. Quite surprisingly the inherent non-uniqueness of factor analysis models does not occur in this generalized context. We attempt an explanation of this fact by restricting the analysis to static generalized factor models. We show that there is a natural interpretation of generalized factor analysis models in terms of Wold decomposition of stationary sequences. A stationary sequence admits a (unique) generalized factor analysis decomposition if and only if two rather natural conditions are satisfied.**

I. Introduction

Factor analysis models have a long history; they have been first introduced by psychologists [28], [5] and have successively been studied and applied in various branches of statistics and econometrics [20], [21], [3], [19], [6], [7]. With a few notable exceptions however, [18], [29], [25], [26], [10], little attention has been payed to these models in the control engineering community.

Dynamic versions of factor models have also been introduced in the econometric literature, see e.g. [15], [23], [24], [17] and references therein.

Recently, we have been witnessing a revival of interest on these models, motivated on one hand by the need of modeling very large aggregates or very large dimensional time series. Vector AR or ARMA models are inadequate for large-dimensional data sets, because they involve a huge number of parameters to estimate which may sometime turn out to be larger than the sample size. On the other hand, an interesting generalization of dynamic factor analysis models allowing the cross-sectional dimension of the observed time series to go to infinity, has been proposed recently by Forni Lippi and collaborators in a series of widely quoted papers [13], [14]. This new modeling paradigm is attracting a considerable attention also in the engineering system identification community [1], [10], [24]. The models proposed by Forni and Lippi, called *Generalized Dynamic Factor Analysis Models* (GDFM)(see e.g. [14], [13] and references therein) are motivated by economic applications. However large-dimensional time series occur often in engineering and signal processing applications, and typically occur, for example, in computer vision and dynamic image processing. The role of identification in image processing and computer vision has been addressed by several authors. We may refer the reader to the recent survey [8] for more details and references. For instance, if we are interested in modeling a "dynamic texture" (see [12] for an example) we may end dealing with a signal $\mathbf{y}(t) := \mathrm{vec}(\mathbf{I}(\cdot, t))$, obtained by vectorizing at a given time t, the signals extracted from the image intensities $\mathbf{I}(\cdot, t)$ at each pixel, forming a vector, say $\mathbf{y}(t) \in \mathbb{R}^m$, with a "large" number (typically tens of thousands) of components. One is interested, therefore, in classes of models (and identification methodologies thereof) which are suited for high dimensional data. Note also that the number N of samples (i.e. $t = 1, .., N$) is very often of the same order (and sometimes smaller) than the data dimensionality ($N < m$). For instance, in dynamic textures modeling, the number N of images in the sequences is of the order of a few hundreds while m (which is equal to the number of pixels of the image) is certainly of the order of a few hundreds or thousands [12], [4]. It is therefore apparent that some sort of dimensionality reduction is absolutely necessary in this context.

In this paper boldface symbols will denote random variables or random arrays, either finite or infinite. Due to page limitations some of the proofs will not be given. A more complete version with all proofs will appear elsewhere and can be obtained from the authors upon request.

G. Bottegal and G. Picci are with the Department of Information Engineering, University of Padova, Italy bottegal@dei.unipd.it, picci@dei.unipd.it

This work has been partially supported by the PRIN Project "New Methods and Algorithms for Identification and Adaptive Control of Technological Systems", funded by the Italian ministery for higher education (MIUR).

II. Static Factor Analysis models

A (static) *Factor Analysis* model is a representation

$$\mathbf{y} = A\mathbf{x} + \mathbf{e}, \tag{1}$$

of m observable variables $\mathbf{y} = [\,\mathbf{y}_1 \dots \mathbf{y}_m\,]^\top$, assumed zero-mean and with finite variance, as linear combinations of n *common factors* $\mathbf{x} = [\,\mathbf{x}_1 \dots \mathbf{x}_n\,]^\top$, plus uncorrelated "noise" or "error" terms $\mathbf{e} = [\,\mathbf{e}_1 \dots \mathbf{e}_m\,]^\top$. An essential part of the model specification is that the m components of the error e should be (zero-mean and) mutually uncorrelated random variables, i.e.

$$\Sigma_{xe} := \mathbb{E}\,\mathbf{x}\mathbf{e}^\top = 0, \tag{2}$$

$$\Sigma_e := \mathbb{E}\,\mathbf{e}\mathbf{e}^\top = \mathrm{diag}\{\sigma_1^2, \dots, \sigma_m^2\}\,. \tag{3}$$

The aim of these models is to provide an "explanation" of the mutual interrelation between the observable variables $\mathbf{y}$ in terms of a small number of common factors, in the sense that, setting

$$\hat{\mathbf{y}}_i := a_i^\top \mathbf{x}, \tag{4}$$

where $a_i^\top$ is the i-th row of the matrix A, one has exactly

$$\mathbb{E}\,\mathbf{y}_i\mathbf{y}_j = \mathbb{E}\,\hat{\mathbf{y}}_i\hat{\mathbf{y}}_j, \tag{5}$$

978-1-61284-799-3/11/$26.00 ©2011 IEEE

for all $i \neq j$. This property is just *conditional orthogonality* (or conditional independence in the Gaussian case) of the family of random variables $\{\mathbf{y}_1, \ldots, \mathbf{y}_m\}$ *given* $\mathbf{x}$ and is a characteristic property of the factors. It is in fact not difficult to see that $\mathbf{y}$ admits a representation of the type (1) if and only if $\mathbf{x}$ renders $\{\mathbf{y}_1, \ldots, \mathbf{y}_m\}$ conditionally orthogonal given $\mathbf{x}$ [25], [2]. We stress that conditional orthogonality given $\mathbf{x}$ is actually equivalent to the orthogonality (uncorrelation) of the components of the noise vector $\mathbf{e}$.

Unfortunately these models although providing a quite natural and useful data compression scheme in many circumstances, suffer from a serious non-uniqueness problem. Note that the property of making $\{\mathbf{y}_1, \ldots, \mathbf{y}_m\}$ conditionally orthogonal is really a property of the subspace of random variables linearly generated by the components of the vector $\hat{\mathbf{y}} := A\mathbf{x}$, denoted[1] $X := H(\hat{\mathbf{y}})$ and it will hold for any set of generators of X. Any set of generating variables for X can serve as a common factors vector and there is no loss of generality to choose the generating vector $\mathbf{x}$ for X of minimal cardinality (a basis) and normalized, i.e. $\mathbb{E}\,\mathbf{x}\mathbf{x}^\top = I$, which we shall always do in the following. A subspace X making the components of $\mathbf{y}$ conditionally independent is called a *splitting subspace* for $\{\mathbf{y}_1, \ldots, \mathbf{y}_m\}$. The so-called "true" variables $\hat{\mathbf{y}}_i$ are then just the orthogonal projections $\hat{\mathbf{y}}_i = \mathbb{E}\,[\mathbf{y}_i \mid X]$.

We may then call $n = \dim \mathbf{x} = \dim X$ the dimension of the model. Obviously a model of dimension n will automatically have $\mathrm{rank} A = n$ as well. Two F.A. models for the same observable $\mathbf{y}$, whose factors span the same splitting subspace X are regarded as *equivalent*. This is a trivial kind of non-uniqueness since two equivalent F.A. models will have factor vectors related by a real orthogonal transformation matrix.

The serious non-uniqueness comes from the fact that there are in general many (possibly infinitely many) minimal splitting subspaces for a given family of observables $\{\mathbf{y}_1, \ldots, \mathbf{y}_m\}$. This is by now well known [25], [22]. Hence there are in general many nonequivalent minimal F.A. models (with normalized factors) representing a fixed m-tuple of random variables $\mathbf{y}$. For example, by choosing for each $k \in \{1, \ldots, m\}$, $\mathbf{x} := [\,\mathbf{y}_1 \ldots \mathbf{y}_{k-1}\; \mathbf{y}_{k+1} \ldots \mathbf{y}_m\,]^\top$, one obtains m "extremal" F.A. models called *elementary regressions*, of the form

$$\begin{cases} \mathbf{y}_1 = [\,1 \ldots 0\,]\,\mathbf{x} + 0 \\ \ldots \\ \mathbf{y}_k = \hat{a}_k^\top \mathbf{x} + \mathbf{e}_k \\ \ldots \\ \mathbf{y}_m = [\,0 \ldots 1\,]\,\mathbf{x} + 0 \end{cases} \tag{6}$$

where $\hat{a}_k^\top = \mathbb{E}\,\mathbf{y}_k\mathbf{x}^\top(\mathbb{E}\,\mathbf{x}\mathbf{x}^\top)^{-1}$. The inherent nonuniqueness of F.A. models is called "factor indeterminacy" (or unindentifiability) in the literature and the term is usually referred to parameter unidentifiability as it may appear that there are always "too many" parameters to be estimated. It may be argued that once a model (in essence, a splitting subspace) is selected, it can always be parametrized in a one-to-one (and hence identifiable) way. Unfortunately, the classification of all possible (minimal) F.A. representations and an explicit characterization of minimality are, to a large extent, still an open problem. The difficulty is indeed a serious one.

Since, as we have argued, in essence non-uniqueness is just a consequence of uncorrelation of the noise components, one may try to get uniqueness by mitigating the requirement of uncorrelation of the components of e. This however turns out to be an ill-defined problem as the basic goal of uniquely splitting the external signal into a noiseless component plus "additive noise" is made vacuous, unless some extra assumptions are made on the model and on the very notion of "noise". As we shall see, for models describing an *infinite* number of observables a meaningful weakening of the uncorrelatedness property can be made, which can guarantee the uniqueness of the decomposition.

[1] In the following we shall denote by the symbol $H(\mathbf{v})$ the inner-product space of random variables linearly generated by the scalar components $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ of a generic m-dimensional random vector $\mathbf{v}$.

III. Generalized Factor Analysis models

In this section we shall review the main points of the construction of [14] particularized to the *static* case. Our point in this review is the observation that the dynamics does not seem to add anything to the structure of the underlying model and tends instead to obscure certain important points. Although we shall not attempt to do so, one could possibly recapture the original dynamic picture by assuming that all real random variables are substituted by random elements taking values in sequence spaces of time series. The object of our study are called *(static) generalized factor analysis models*.

Consider an infinite collection of zero-mean finite variance random variables $\mathbf{y} := \{\mathbf{y}_k,\, k \in \mathbb{N}\}$, which we shall occasionally represent as a random vector with infinite components. We want to describe every element of such a sequence as a linear combination of a finite number of common components plus " noise", i.e.

$$\mathbf{y}_k = f_k^\top \mathbf{x} + \tilde{\mathbf{y}}_k\,, \qquad k = 1, 2, \ldots \tag{7}$$

where $\mathbf{x}$ is a q-dimensional fixed random vector which can be taken with orthonormal components ($\mathrm{Var}\,[\mathbf{x}] = I_q$) and $\tilde{\mathbf{y}}_k$ is a random variable which is orthogonal to $\mathbf{x}$, whose specific character (see the definition of idiosyncratic noise below) will be discussed later. The linear combination $f_k^\top \mathbf{x}$ is also denoted by $\hat{\mathbf{y}}_k$ ($\hat{\mathbf{y}}$ in vector notation).

The infinite covariance matrix of the vector $\mathbf{y}$ is denoted by Σ, while Σ_n indicates the top-left $n \times n$ block of Σ, equal to the covariance matrix of the first n components of $\mathbf{y}$, the corresponding n-dimensional vector being denoted by $\mathbf{y}^n$. The inequality $\Sigma \geq 0$ means that all submatrices Σ_n of Σ are at least positive semidefinite.

The orthogonality of the noise term and the common components implies that

$$\Sigma_n = \hat{\Sigma}_n + \tilde{\Sigma}_n\,, \quad \forall n \in \mathbb{N}\,, \tag{8}$$

where $\hat{\Sigma}_n := \mathbb{E}\hat{\mathbf{y}}^n\hat{\mathbf{y}}^{n\top}$ and $\tilde{\Sigma}_n := \mathbb{E}\tilde{\mathbf{y}}^n\tilde{\mathbf{y}}^{n\top}$.

Let $\ell_2(\Sigma)$, $\Sigma \geq 0$ denote the Hilbert space of infinite sequences $\mathfrak{a} := \{a_k,\, k \in \mathbb{N}\}$ such that $\|\mathfrak{a}\|_\Sigma^2 := \mathfrak{a}^\top \Sigma \mathfrak{a} < \infty$. When $\Sigma = I$ we simply use the symbol ℓ_2, indicating the corresponding norm with the symbol $\|\cdot\|$.
The sequence whose first n elements are the same as in $\mathfrak{a}$, while the others are set equal to zero is denoted $\mathfrak{a}^{[n]}$.

Definition 1: Let $\{\mathfrak{a}_n,\, n \in \mathbb{N}\}$ be a sequence of elements of the set $\ell_2 \cap \ell_2(\Sigma)$. We say that $\{\mathfrak{a}_n,\ n \in \mathbb{N}\}$ is an averaging sequence *(AS) if* $\lim_{n\to\infty} \|\mathfrak{a}_n\| = 0$.

Example 1: The sequence of elements in ℓ_2

$$\mathfrak{a}_n = \frac{1}{n}[\underbrace{1 \dots 1}_{n}\, 0 \dots]^\top \tag{9}$$

is an averaging sequence.
The definition of AS allows us to introduce the concept of *idiosyncratic sequence* of random variables.

Definition 2: We say that $\mathbf{y}$ *is* idyosincratic *if* $\lim_{n\to\infty} \mathfrak{a}_n^\top \mathbf{y} = 0$ *for any averaging sequence* $\mathfrak{a}_n \in \ell_2 \cap \ell_2(\Sigma)$.
Another useful defintion is the following. Let $H(\mathbf{y})$ be the Hilbert space spanned by the sequence $\{\mathbf{y}_k,\, k \in \mathbb{N}\}$.

Definition 3: Let $z \in H(\mathbf{y})$. We say that the random variable $\mathbf{z}$ is an aggregate (of $\mathbf{y}$) *if there exists an AS $\mathfrak{a}_n$ such that $\lim_{n\to\infty} \mathfrak{a}_n^\top \mathbf{y} = \mathbf{z}$. The set of all aggregate random variables in $H(\mathbf{y})$ is denoted by $\mathcal{G}(\mathbf{y})$ and called the* aggregation subspace *of $H(\mathbf{y})$.*
It is straightforward to check that $\mathcal{G}(\mathbf{y})$ is a closed subspace. Clearly, if $\mathbf{y}$ is an idiosyncratic sequence then $\mathcal{G}(\mathbf{y}) = \{0\}$. Furthermore, it is possible to define an orthogonal decomposition of the type

$$\mathbf{y} = \mathbb{E}[\mathbf{y} \mid \mathcal{G}(\mathbf{y})] + \mathbf{u}\,, \tag{10}$$

where all components u_k are uncorrelated with $\mathcal{G}(\mathbf{y})$. The idea behind this decomposition is that, in case $\mathcal{G}(\mathbf{y})$ is finite dimensional, say generated by a q-dimensional random vector $\mathbf{x}$, one may naturally capture a unique decomposition of $\mathbf{y}$ as in (7). Unfortunately, in general $\mathcal{G}(\mathbf{y}) = \{0\}$ does not imply that $\mathbf{y}$ is idiosyncratic, as it can be seen in the following example.

Example 2: ([14]) Consider a sequence $\mathbf{y}$ with $\mathbf{y}_j \perp \mathbf{y}_h \,\forall j \neq h$ (i.e. $\mathbf{y}$ is a white noise), such that $\|\mathbf{y}_j\|^2 = j$. This sequence is not idiosyncratic, since, given the AS

$$\mathfrak{d}_n = \frac{1}{\sqrt{n}}[\underbrace{0 \dots 0\, 1}_{n}\, 0 \dots]^\top\,, \tag{11}$$

we obtain that $\|\mathfrak{d}_n^\top \mathbf{y}\| = 1\, \forall n$. Let then $\mathbf{z}$ be an aggregate random variable, so that there must exist an AS $\mathfrak{a}_n$ such that

$$\mathbf{z} = \lim_{n\to\infty} \mathfrak{a}_n^\top \mathbf{y} = \lim_{n\to\infty} \sum_{j=1}^{\infty} a_{n,j} \mathbf{y}_j\,. \tag{12}$$

Note that, being $\mathbf{z} \in H(\mathbf{y})$ and $\mathbf{y}$ an orthogonal basis of such space, we can uniquely express $\mathbf{z}$ as

$$\mathbf{z} = \sum_{j=1}^{\infty} b_j \mathbf{y}_j\,, \tag{13}$$

and, by uniqueness of the representation, it follows that $\lim_{n\to\infty} a_{n,j} = b_j \,\forall j$. On the other hand, being $\mathfrak{a}_n$ an AS, the limits of $a_{n,j}$ must be zero, so that $b_j = 0$. Hence $z = 0$. Thus $\mathcal{G}(\mathbf{y}) = \{0\}$ but $\mathbf{y}$ is not idiosyncratic.
Note that the sequence $\mathbf{y}$, interpreted as a stochastic process with respect to the cross-sectional index k, is *non-stationary.*
The nature of an idiosyncratic sequence is strictly related to the behaviour of the eigenvalues of its covariance matrix. To explain this point, it is useful to introduce some notations and facts about the eigenvalues of infinite covariance matrices. Denote by $\lambda_{n,k}^y$ the k–th eigenvalue of the $n \times n$ upper left submatrix Σ_n of Σ. The $\lambda_{n,k}^y$ are real nonnegative and can be ordered in decreasing magnitude. Forni and Lippi [14, Fact M], show that the k–th eigenvalue of Σ_n is a non decreasing function of n and hence has a limit, λ_k^y, which may possibly be $+\infty$. Each such limit is called an eigenvalue of Σ. In case all the limits are finite one can show that they are bona–fide eigenvalues of the infinite matrix Σ (considered as a linear operator on ℓ_2). Clearly these eigenvalues can also be ordered. Henceforth we shall denote by λ_1^y the maximal eigenvalue of Σ, with the convention that $\lambda_1^y = +\infty$ when there are infinite eigenvalues as defined above.
A strong characterization of idiosyncratic sequences is given by the following theorem, stated after [14] after some obvious simplifications.

Theorem 1: The sequence $\mathbf{y}$ is idiosyncratic if and only if λ_1^y is finite.

Proof: Assume first that $\lim_{n\to\infty} \lambda_{n,1}^y = +\infty$. Since $\Sigma_n \geq 0$, for every n one has the diagonalization

$$U_n^\top \Sigma_n U_n = D_n\,, \tag{14}$$

where U_n is orthonormal and

$$D_n = \operatorname{diag}\{\, \lambda_{n,1}^y, \dots, \lambda_{n,n}^y \,\} \tag{15}$$

For every n, consider the first column of U_n, say u_1^n, which is the eigenvector of the eigenvalue $\lambda_{n,1}^y$ and define the sequence of elements in the set $\ell_2 \cap \ell_2(\Sigma)$

$$\mathfrak{a}_n := \frac{1}{\sqrt{\lambda_{n,1}^y}} \begin{bmatrix} u_1^{n\top} & 0 & \dots \end{bmatrix}^\top. \tag{16}$$

Note that this is an AS, whose application to $\mathbf{y}$ gives $\|\mathfrak{a}_n^\top \mathbf{y}\| = 1$ for every n, thus the sequence $\mathbf{y}$ cannot be idiosyncratic.
Conversely, suppose now that $\lambda_1^y < +\infty$ and again apply the diagonalization $\Sigma_n = U_n D_n U_n^\top$. Let $\mathfrak{a}_n$ be an arbitrary AS and consider the random variable

$$\mathbf{z} = \lim_{n\to\infty} \mathfrak{a}_n^\top \mathbf{y} = \lim_{n\to\infty} \mathfrak{a}_n^{n\top} \mathbf{y}^n\,, \tag{17}$$

which has variance

$$\operatorname{var}[\mathbf{z}] = \lim_{n\to\infty} \mathfrak{a}_n^{n\top} U_n D_n U_n^\top \mathfrak{a}_n^n := \mathfrak{d}_n^{n\top} D_n \mathfrak{d}_n^n\,, \tag{18}$$

where $\mathfrak{d}_n$ is an AS whose first n elements form a vector equal to $U_n^\top \mathfrak{a}_n^n$, while the remaining can be taken equal to

those of $\mathfrak{a}_n$.
Since $\mathfrak{d}_n^{n\top} D_n \mathfrak{d}_n^n = \sum_{i=1}^n \lambda_{n,i}^y d_{n,i}^2$ one can write

$$\begin{aligned}\operatorname{var}[\mathbf{z}] &= \lim_{n\to\infty} \sum_{i=1}^{n} \lambda_{n,i}^y d_{n,i}^2 \leq \lim_{n\to\infty} \lambda_1^y \sum_{i=1}^{n} d_{n,i}^2 \\ &= \lim_{n\to\infty} \lambda_1^y \|\mathfrak{d}_n^n\|^2 = 0\end{aligned}$$

which shows that $\mathbf{y}$ is idiosyncratic. ∎

Definition 4: Let q be a finite integer. A sequence $\mathbf{y}$ is *purely deterministic of rank* q (in short q-PD) if $H(\mathbf{y})$ has dimension q.
Clearly for a q-PD sequence $\mathbf{y}$ can be seen as a (in general non-stationary) purely deterministic process in the classical sense of the term, see [9]. Let $\mathbf{x}$ be an orthonormal basis in $H(\mathbf{y})$. Obviously $\mathbf{y}$ is a q-PD random sequence if and only if there is a $\mathbb{R}^q$-valued function $f(k)$, $k \in \mathbb{N}$, such that

$$\mathbf{y}_k = f^\top(k)\mathbf{x} = \sum_{i=1}^{q} f_i(k)\,\mathbf{x}_i\,, \qquad k \in \mathbb{N}, \tag{19}$$

where the functions $f_1(\cdot), f_2(\cdot), \ldots f_q(\cdot)$ must be linearly independent, for otherwise the rank of $\mathbf{y}$ would be smaller than q.
We want to relate this concept with the idea of aggregation subspace of $\mathbf{y}$, as defined earlier. Let now $\mathbf{x}$ be an orthonormal basis in $\mathcal{G}(\mathbf{y})$; quite unfortunately, there are nontrivial sequences representable in the form (19) which are idiosyncratic (or contain idiosyncratic sequences). See Example 3 below.

Example 3: Consider a sequence $\mathbf{y}$ whose $k-$th element is

$$\mathbf{y}_k = \lambda^k \mathbf{x} \quad , |\lambda| < 1, \tag{20}$$

where $\mathbf{x}$ is a zero–mean random variable of variance σ^2. Clearly, $\mathbf{y}$ is non–stationary and 1-PD, its spanning subspace $H(\mathbf{y})$ being the one-dimensional space $H(\mathbf{x})$. The covariance matrix of the first n components of $\mathbf{y}$ is

$$\Sigma_n = \mathbb{E}\mathbf{y}^n\mathbf{y}^{n\top} = \sigma^2 \begin{bmatrix} \lambda^2 & \lambda^3 & \ldots & \lambda^{n+1} \\ \lambda^3 & \lambda^4 & \ldots & \lambda^{n+2} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda^{n+1} & \lambda^{n+2} & \ldots & \lambda^{2n} \end{bmatrix} \tag{21}$$

Being $\operatorname{rank}(\Sigma_n) = 1$ for every n, we have

$$\lambda_1^y = \lim_{n\to\infty} \operatorname{tr}(\Sigma_n) = \lim_{n\to\infty} \sigma^2 \sum_{k=1}^{n} \lambda^{2k} = \frac{\sigma^2\lambda^2}{1-\lambda^2}\,, \tag{22}$$

thus, in force of Theorem 1, $\mathbf{y}$ is idiosyncratic. Hence there are non-stationary $q-$PD sequences which are idiosyncratic. This is a possibility which we clearly must exclude if the decomposition (7) has to be unique. To this end Forni and Lippi seem to impose a condition on the eigenvalues of the covariance matrix of a $q-$PD sequence. We introduce the following definition.

Definition 5: Let $\mathbf{y}$ *be a* $q-$*PD sequence; then* $\mathbf{y}$ *is called* q-aggregate *if the* q *nonzero eigenvalues of its covariance matrix are all infinite.*

The question now is which properties need to be satisfied by the functions $f_1, f_2, \ldots f_q$ for $\mathbf{y}$ to be a q-aggregate sequence. The answer is in the following theorem.

Theorem 2: Let $\mathbf{y}$ *be a* $q-$*PD sequence, i.e. let*

$$\mathbf{y}_k = f^\top(k)\mathbf{x} = \sum_{i=1}^{q} f_i(k)\,\mathbf{x}_i\,, \qquad k \in \mathbb{N}\,; \tag{23}$$

then $\mathbf{y}$ *is* $q-$*aggregate if and only if, for each* $i = 1, \ldots, q$, *it holds that*

$$\lim_{n\to\infty} \|f_i^n(\cdot) - \Pi[f_i^n(\cdot)\,|\,\mathcal{F}_i^n]\|_2 = +\infty\,. \tag{24}$$

where Π *is the orthogonal projection onto the Euclidean space*

$$\mathcal{F}_i^n = \operatorname{span}\{f_j^n(\cdot),\, j = 1, \ldots, q,\, j \neq i\} \tag{25}$$

The proof of this theorem is rather long and will be given elsewhere.

Example 4: Consider the $2-$PD sequence

$$\mathbf{y}_k := \sum_{i=1}^{2} f_i(k)\,\mathbf{x}_i \tag{26}$$

with

$$f_1(k) = 1 \quad \forall k\,, \qquad f_2(k) = 1 - \left(\frac{1}{2}\right)^k.$$

It is not difficult to check that this sequence does not satisfy condition (24). We shall show that this sequence is not 2-aggreggate. The Gramian matrix of the functions f_1, f_2 restricted to $[1,\, n]$ is

$$f^{n\top} f^n = \begin{bmatrix} \|f_1^n\|_2^2 & < f_1^n,\, f_2^n >_2 \\ < f_1^n,\, f_2^n >_2 & \|f_2^n\|_2^2 \end{bmatrix} \tag{27}$$

and it is easy to see that as $n \to \infty$, the second eigenvalue converges to $\frac{5}{3}$. Hence one eigenvalue of the covariance matrix of $\mathbf{y}$ is finite and the sequence is not 2-aggregate.

The following proposition, which follows trivially from Theorem 1, guarantees uniqueness of the decomposition (7) when $\hat{\mathbf{y}}$ is q-aggregate and $\tilde{\mathbf{y}}$ is idiosyncratic.

Proposition 1: A q*-aggregate sequence* $\hat{\mathbf{y}}$ *can be idiosyncratic only if it is the zero sequence.*
The next definition is the static version of a similar one of [14] for the dynamic setting.

Definition 6: The sequence $\mathbf{y}$ *is a* $q-$*factor sequence (*$q-$*FS) if it can be written as an orthogonal sum*

$$\mathbf{y}_k = f_k^\top \mathbf{x} + \tilde{\mathbf{y}}_k\,, \tag{28}$$

where $\hat{\mathbf{y}}_k := f_k^\top \mathbf{x}$ *is a* q*-aggregate sequence and* $\tilde{\mathbf{y}}$ *is idiosyncratic (and orthogonal to* $\mathbf{x}$*). The representation* (28) *is called a* generalized factor model with q factors.
Hence $\mathbf{y}$ is a $q-$FS if and only if it admits a representation by a generalized factor model with q factors.
The crucial question is now to give a criterion telling us which random sequences are $q-$FS. Forni and Lippi [14] provide a criterion based on the unboundedness of the eigenvalues of the covariance matrix. The criterion is rephrased below for the static setting which concerns us here.

Theorem 3: If the sequence $\mathbf{y}$ *is a* $q-$*FS then* $\lambda_q^y = +\infty$ *but* λ_{q+1}^y *is bounded.*

The condition is clearly necessary since the sequence $\tilde{\mathbf{y}}$ is idiosyncratic iff the first (i.e. maximal) eigenvalue of $\tilde{\Sigma}$ is finite and the covariance matrix $\hat{\Sigma}$ of the sequence $\hat{\mathbf{y}} := \{\hat{\mathbf{y}}_k = f_k^\top \mathbf{x},\, n \in \mathbb{N}\}$ has q unbounded eigenvalues, i.e. $\lambda_q^{\hat{y}} = +\infty$. The proof of sufficiency in [14] is very involved and we shall not discuss it.

Unfortunately, as shown by Example 2, there are sequences which do not admit a $q-$FS (with q finite). The covariance matrix of the scalar sequence $\mathbf{y}$ of Example 2 is diagonal with infinitely many eigenvalues of Σ equal to $+\infty$, according to the definition given above. Hence the sequence may be called a $\infty-$aggregate sequence. However $\mathbf{y}$ cannot be $q-$aggregate since it cannot be $q-$PD for any finite q; in fact, it can be shown that $\mathbf{y}$ is a purely-non-deterministic sequence. On the other hand $\mathbf{y}$ is not idiosyncratic either since we have shown that $\mathcal{G}(\mathbf{y}) = \{0\}$.

IV. Stationary sequences and the Wold decomposition

As we have just seen, non-stationarity can bring in many pathologies which seem to be difficult to rule out. We consider now the special case in which the sequence $\mathbf{y}$, defined on $\mathbb{Z}_+$, is (weakly) stationary; i.e. $\mathbb{E}\mathbf{y}_t\mathbf{y}_s = r(t-s)$ for $t, s \geq 0$. It is well known, see e.g. [11], [27] that, introducing the *remote future subspace* of $\mathbf{y}$:

$$H_\infty(\mathbf{y}) = \bigcap_{t\geq 0} H_t(\mathbf{y}) \tag{29}$$

the sequence of orthogonal wandering subspaces $E_t := H_t(\mathbf{y}) \ominus H_{t+1}(\mathbf{y})$ and their orthogonal direct sum

$$\tilde{H}(\mathbf{y}) = \bigoplus_{t\geq 0} E_t$$

one has a unique orthogonal decomposition

$$\mathbf{y} = \hat{\mathbf{y}} + \tilde{\mathbf{y}}\,, \qquad \hat{\mathbf{y}}_k \in H_\infty(\mathbf{y}) \quad \tilde{\mathbf{y}}_k \in \tilde{H}(\mathbf{y}) \tag{30}$$

for all $k \in \mathbb{Z}_+$, the component $\hat{\mathbf{y}}$ being the purely deterministic (PD) component while $\tilde{\mathbf{y}}$ the purely non deterministic (PND) one. The two sequences are orthogonal and uniquely determined. Furthermore, it is well known that $\tilde{\mathbf{y}}$ has an absolutely continuous spectrum with a spectral density function, say $S_y(\omega)$, while $\hat{\mathbf{y}}$ has a singular spectral distribution (for example consisting only of jumps) possibly together with a singular spectral density such that

$$\int \log S_y(\omega)\, d\omega = -\infty\,. \tag{31}$$

In this section we want to give an interpretation of the decomposition (7) in the light of the the Wold decomposition. First we prove the following lemma.

Lemma 1: Let $\mathbf{y}$ *be stationary and PND and assume that its spectral density is bounded; i.e.*

$$S_y(\omega) \in L^\infty([-\pi,\, \pi])\,. \tag{32}$$

Then $\mathbf{y}$ *is idiosyncratic.*

Proof: Consider an AS $\mathfrak{a}_n$; then

$$\|\mathfrak{a}_n^\top \mathbf{y}\|^2 = \|\mathfrak{a}_n\|_\Sigma^2 = \mathfrak{a}_n^\top \Sigma \mathfrak{a}_n \leq \lambda_1^y \|\mathfrak{a}_n\|^2\,. \tag{33}$$

Since $\mathbf{y}$ is PND and its spectral density is bounded, for a well known theorem of Szegö [16, p.65], Σ has bounded eigenvalues, thus $\|\mathfrak{a}_n^\top \mathbf{y}\|^2 \to 0$, i.e. $\mathbf{y}$ is idiosyncratic. ■

Lemma 1 has an important consequence, namely

Lemma 2: Let $\mathbf{y}$ *be a stationary sequence with a bounded spectral density, then* $\mathcal{G}(\mathbf{y}) \subseteq H_\infty(\mathbf{y})$.

Note that the statement holds in particular for PD processes with a singular spectrum, as in this case $S_y(\omega) \equiv 0$. The converse inclusion, i.e. $H_\infty(\mathbf{y}) \subseteq \mathcal{G}(\mathbf{y})$, is in general not true. However, for stationary sequences with a *finite dimensional remote future*, we can state the following.

Theorem 4: Assume that $\mathbf{y}$ *is a stationary sequence with a bounded spectral density and that* $\dim H_\infty(\mathbf{y}) < \infty$. *Then* $H_\infty(\mathbf{y}) \equiv \mathcal{G}(\mathbf{y})$.

Hence,

Theorem 5: Every stationary sequence with bounded spectral density and remote future space of dimension q *is a* $q-$*factor sequence and admits a unique generalized factor analysis decomposition* (28) *where* $\hat{\mathbf{y}}$ *is the PD and* $\tilde{\mathbf{y}}$ *the PND components of* $\mathbf{y}$.

It is not hard to show that the assumption of stationarity here is crucial. In fact, Example 3 discussed before shows that in the non-stationary case a PD process $\mathbf{y}$ whose remote future is the one-dimensional space $H(\mathbf{z})$ may be idiosyncratic.

In the following example, we show how to build an AS that generates a basis in a finite-dimensional remote future space.

Example 5: Consider a PD process $\mathbf{y}$, with a remote future of finite dimension 2ν. It is well-known that any such process can be expressed as a sum of elementary oscillations of the form $\mathbf{y}_k = \sum_{i=1}^{\nu} \mathbf{v}_i \cos\omega_i k + \mathbf{w}_i \sin\omega_i k\,,$ where $\mathbf{v}_i$ and $\mathbf{w}_i$ are mutually uncorrelated zero–mean random variables with $\mathrm{var}[\mathbf{v}_i] = \mathrm{var}[\mathbf{w}_i]$. It can be seen that $H_\infty(\mathbf{y}) = \mathrm{span}\{\mathbf{v}_i,\, \mathbf{w}_i,\, i = 1, \ldots, \nu\}$. Consider the AS $\mathfrak{a}_n$ whose elements have components

$$a_{n,k} = \begin{cases} \dfrac{1}{n}\sum_{i=1}^{\nu} \cos\omega_i k + \sin\omega_i k & k \leq n \\ 0 & k > n \end{cases} \tag{34}$$

Applying $\mathfrak{a}_n$ to $\mathbf{y}$ we obtain the random variable

$$\mathbf{z}_n = \sum_{k=1}^{n} a_{n,k}\mathbf{y}_k = \frac{1}{n}\sum_{k=1}^{n}\left[\sum_{i=1}^{\nu} z_{n,k}^i + \sum_{i,j=1, i\neq j}^{\nu} z_{n,k}^{i,j}\right],$$

where

$$\begin{aligned} z_{n,k}^i &= \mathbf{v}_i \cos^2\omega_i k + \mathbf{w}_i \sin^2\omega_i k + \frac{\mathbf{v}_i + \mathbf{w}_i}{2}\sin 2\omega_i k \\ z_{n,k}^{i,j} &= \frac{\mathbf{v}_i - \mathbf{w}_i}{2}\cos(\omega_i + \omega_j)k + \frac{\mathbf{v}_i + \mathbf{w}_i}{2}\times \\ &\quad \ldots [\cos(\omega_i - \omega_j)k + \sin(\omega_i + \omega_j)k + \sin(\omega_i - \omega_j)k] \end{aligned}$$

As n tends to infinity, all the non quadratic terms vanish, giving

$$\mathbf{z} = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{\nu} \mathbf{v}_i \cos^2\omega_i k + \mathbf{w}_i \sin^2\omega_i k \tag{35}$$

and so

$$\mathbf{z} = \sum_{i=1}^{\nu} c_i \mathbf{v}_i + d_i \mathbf{w}_i \,, \tag{36}$$

where c_i, d_i are constants in the interval $[0, 1]$. Thus the random variable $\mathbf{z}$ is an aggregate of $\mathbf{y}$, so $\mathcal{G}(\mathbf{y}) \neq \{0\}$ and $\mathbf{y}$ is not idiosyncratic. This example suggests a method to obtain a basis of $H_\infty(\mathbf{y})$. In order to obtain only one of the random variables spanning $H_\infty(\mathbf{y})$, say for example $\mathbf{v}_p$, $p \leq \nu$, it is sufficient to apply to $\mathbf{y}$ the AS $\mathfrak{a}_n$ of components $a_{n,k} = \frac{1}{n} \cos \omega_p k$ for $k \leq n$ and equal to zero for $k > n$, obtaining

$$\mathbf{z} = \lim_{n\to\infty} \mathfrak{a}_n^\top \mathbf{y} = \frac{1}{n} \sum_{k=1}^{n} \mathbf{v}_p \cos^2 \omega_p k = c_p \mathbf{v}_p \,, \tag{37}$$

with $c_p > 0$. Analogously, one can obtain the random variables $\mathbf{w}_i$ using a sine instead of the cosine.

V. Discussion

In the paper [14] stationarity with respect to the cross-sectional index is not required. However without stationarity, there may be sequences which satisfy the eigenvalue conditions of Theorem 3 but do not admit a generalized factor analysis decomposition. Example 2 shows one such sequence. Understanding which class of non-stationary sequences admits a generalized factor analysis decomposition seems still to be an open problem.

References

[1] B. Anderson and M. Deistler, "Generalized linear dynamic factor models - a structure theory," in *Proc. of the IEEE Decision and Control conference (CDC)*, 2008.

[2] D. J. Bartholomew, "The foundations of factor analysis," *Biometrika*, vol. 71(2), pp. 221–232, 1984.

[3] P. A. Bekker and J. de Leeuw, "The rank of reduced dispersion matrices," *Psychometrika*, vol. 52, no. 1, pp. 125–135, 1987.

[4] A. Bissacco, A. Chiuso, and S. Soatto, "Classification and recognition of dynamical models: The role of phase, independent components, kernels and optimal transport," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, pp. 1958–1972, 2007.

[5] C. Burt, "Experimental tests of general intelligence," *British Journal of Psychology*, vol. 3, pp. 94–177, 1909.

[6] G. Chamberlain, "Funds, factors and diversification in arbitrage pricing models," *Econometrica*, vol. 51, no. 5, pp. 1305–1324, 1983.

[7] G. Chamberlain and M. Rothschild, "Arbitrage, factor structure and meanvariance analysis on large asset markets," *Econometrica*, vol. 51, no. 5, pp. 1281–1304, 1983.

[8] A. Chiuso and G. Picci, "Some identification techniques in computer vision," in *Proc. of CDC*, 2008.

[9] H. Cramèr, "On some classes of non-stationary stochastic processes," in *Proc. IV Berkeley Symposium on Math. Statistics and Probability*, 1961, vol. II, pp. 57–77.

[10] M. Deistler and C. Zinner, "Modelling high-dimensional time series by generalized dynamic factor models: an introductory survey," *Communications on Information and Systems*, vol. 7, no. 2, pp. 153–166, 2007.

[11] J. L. Doob, *Stochastic processes*, ser. Wiley Classics Library. New York: John Wiley & Sons Inc., 1990, reprint of the 1953 original, A Wiley-Interscience Publication.

[12] G. Doretto, A. Chiuso, S. Soatto, and Y. Wu, "Dynamic textures," *International Journal of Computer Vision*, vol. 51, no. 2, pp. 91–109, February 2003.

[13] M. Forni, M. Hallin, M. Lippi, and L. Reichlin, "The generalized dynamic factor model: identification and estimation," *The review of Economic and Statistics*, vol. 65, pp. 453–473, 2000.

[14] M. Forni and M. Lippi, "The generalized dynamic factor model: representation theory," *Econometric Theory*, vol. 17, pp. 1113–1141, 2001.

[15] J. Geweke, "The dynamic factor analysis of economic time series," in *Latent Variables in Socio-Economic Models*, D. A. . A. Goldberger, Ed. North-Holland, 1977, pp. 365–383.

[16] U. Grenander and G. Szegö, *Toeplitz forms and their applications, second ed.* New York: Chelsea, 1984.

[17] Y. Hu and R. Chou, "On the Peña-Box model," *Journal of Time Series Analysis*, vol. 25, pp. 811–830, 2004.

[18] R. Kalman, "Identifability and problems of model selction in econometrics," in *Advances in econometrics*, W. Hildebrandt, Ed. Cambridge: Cambridge University Press, 1983.

[19] D. N. Lawley and A. E. Maxwell, *Factor Analysis as a Statistical Method, Second ed.* London: Butterworths, 1971.

[20] W. Ledermann, "On the rank of the reduced correlation matrix in multiple factor analysis," *Psychometrika*, vol. 2, pp. 85–93, 1937.

[21] ——, "On a problem concerning matrices with variable diagonal elements," *Proc. Royal Soc. Edinburgh*, vol. XL, pp. 1–17, 1939.

[22] A. Lindquist and G. Picci, *Linear stochastic systems: a geometric approach*. In preparation, 2011.

[23] D. Peña and G. Box, "Identifying a simplifying structure in time series," *J. Amer. Stat. Ass.*, vol. 82, pp. 836–843, 1987.

[24] D. Peña and P. Poncela, "Nonstationary dynamic factor analysis," *Journal of Statistical Planning and Inference*, vol. 136, pp. 1237–1257, 2006.

[25] G. Picci, "Parametrization of factor analysis models," *Journal of Econometrics*, vol. 41, pp. 17–38, 1987.

[26] G. Picci and S. Pinzoni, "Dynamic factor-analysis models for stationary processes." *IMA Journal of Math. Control and Information*, vol. 3, pp. 185–210, 1986.

[27] Y. A. Rozanov, *Stationary Random Processes*, Holden-Day, Ed. Holden Days, 1967.

[28] C. Spearman, "General intelligence, objectively determined and measured," *American Journal of Psychology*, vol. 15, pp. 201–203, 1904.

[29] J. van Schuppen, "Stochastic realization problems motivated by econometric modeling," in *Modeling Identification and robust control*, C. Byrnes and A. Lindquist, Eds. Amsterdam: North-Holland, 1986.