

Modeling random flocks through Generalized Factor Analysis

Giulio Bottegal and Giorgio Picci

Abstract—In this paper, we study modeling and identification of stochastic systems by Generalized Factor Analysis models. Although this class of models was originally introduced for econometric purposes, we present some possible applications of engineering interest. In particular, we show that there is a natural connection between Generalized Factor Analysis models and multi-agents systems. The common factor component of the model has an interpretation as a *flocking component* of the system behavior.

I. INTRODUCTION

Flocking is a commonly observed behavior in gregarious animals by which many equal individuals tend to group and follow, at least approximately a common path in space. The phenomenon has similarities with many scenarios observed in artificial/technological and biological environments and has been studied quite actively in recent years [6], [22], [17], [10]. A few examples are described below.

The mechanism of *formation* of flocks, is also called *convergence to consensus* and has been intensely studied in the literature, see e.g. [13], [18], [21], and there is now a quite articulated theory addressing the convergence to consensus under a variety of assumptions on the communication strategy among agents etc.

In this paper we want to address a different issue: given observations of the motion of a large set of equal agents and assuming statistical steady state, decide whether there is a flocking component in the collective motion and estimate its structural characteristics. The reason for doing this is that the very concept of flocking implies an *orderly motion* which must then admit a much simpler mathematical description than the whole ensemble. Once the flocking component (if present) has been separated, the motion of the ensemble splits naturally into flocking plus a random term which describes local random disagreements of the individual agents or the effect of external disturbances. Hence extracting a flocking structure is essentially a parsimonious modeling problem. Our approach to this problem is based on modeling the flock through *Generalized Factor Analysis* (GFA). This class of models, originally introduced in [8], [7], has been extended to a dynamic version (known as Generalized Dynamic Factor Models) in [15]. Some identification techniques for these models can be found in [14], [11].

G. Bottegal is with the Department of Information Engineering, University of Padova, Italy, and with the ACCESS Linnaeus Centre, School of Electrical Engineering, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden. bottegal@kth.se

G. Picci is with the Department of Information Engineering, University of Padova, Italy. picci@dei.unipd.it

This work has been partially supported by the European Research Council under the advanced grant LEARN, contract 267381

A. Detection of emitters

In this scenario we suppose there is an unknown number, say q , of emitters, each of them broadcasting radio impulse trains at a fixed common frequency. Such impulses are received by a large array of N antennas spread in space. The measurement of each antenna is corrupted by noise, generated by measurement errors or local disturbances, possibly correlated with that of neighbouring antennas. The set up can be described mathematically, by indexing with an integer number $i = 1, 2, \dots$ each antenna and denoting by $\mathbf{y}_i(t)$ the signal received at time t by antenna i . Then the following model can be used to describe the received signal

$$\mathbf{y}_i(t) = f_{i1}x_1(t) + \dots + f_{iq}x_q(t) + \tilde{\mathbf{y}}_i(t), \quad (1)$$

where $x_j(t)$ is the signal sent by the j -th emitter at time t ; f_{ij} is a coefficient related to the distance between j -th emitter and antenna i and $\tilde{\mathbf{y}}_i(t)$ is the disturbance affecting antenna i at time t . The disturbances are uncorrelated with the signal but those of neighboring antennas may be correlated. The goal is to detect the number of emitters q and possibly estimate the signal components $x_j(t)$ impinging on the antenna array. Let $\mathbf{y}(t)$, $\mathbf{x}(t)$, $\tilde{\mathbf{y}}(t)$ denote vector valued quantities in the model (1) of respective dimensions N , q and N . The model (1) can be compactly written as

$$\mathbf{y}(t) = F\mathbf{x}(t) + \tilde{\mathbf{y}}(t), \quad (2)$$

where \mathbf{y} is the N -dimensional random process of observables; $\mathbf{x}(t) = [\mathbf{x}_1(t) \dots \mathbf{x}_q(t)]^\top$ is an unknown hidden random signal; $F = \{f_{ij}\} \in \mathbb{R}^{N \times q}$ is an unknown matrix of coefficients and $\tilde{\mathbf{y}}$ is a N -dimensional random process of disturbances, uncorrelated with \mathbf{x} , describing the “local noise” features.

Note that in the model there are several hidden (non-measurable) variables, among which the dimension q . In general N is large; ideally one may take $N \rightarrow \infty$. We may identify $F\mathbf{x}(t)$ as the flocking component of $\mathbf{y}(t)$. The only statistical information on the system is an estimate of the covariance matrix of \mathbf{y} , say $\Sigma \in \mathbb{R}^{N \times N}$. A crude statistical approach assumes that all signals in the model are i.i.d. process, and the sample values $\{y(t)\}$ are interpreted as random samples generated by a underlying static model of the form

$$\mathbf{y} = F\mathbf{x} + \tilde{\mathbf{y}}. \quad (3)$$

The estimate of Σ may be computed assuming that a large enough collection of independent observations $\{\mathbf{y}(t)\}$ of \mathbf{y} are available.

A similar problem is automated speaker detection. This is the problem of detecting the speaking persons (emitters)

in a noisy environment at any particular time, from signals coming from a large array of N microphones distributed in a room. Here the number of emitters is generally small but could be varying with time. Robustly solving this problem is useful in areas such as surveillance systems, and human-machine interaction.

B. Inference on gene regulatory networks

In systems biology, an important task is the inference on gene regulatory networks to understand cell physiology and pathology. Genes are known to interact among each other forming a network, and their expression is directly regulated by few proteins called transcription factors (TFs). Understanding this action of regulation is of paramount importance. Methods for measuring the gene expressions using microarray data are extremely popular, however there are still issues in observing the action of TFs. Thus, several strategies were proposed to understand the action of the TFs from available measurements of gene expressions. To this end, a reasonable assumption on the interaction between genes and TFs is commonly accepted and consists in considering the TFs as elements of the “gene space” [5]. In this way, one should be able to reconstruct the TFs actions from the gene expressions using proper estimation techniques.

Denoting the measured expression profile of the i -th gene of the network by a random variable y_i , \mathbf{y} representing the N -dimensional vector of all the gene expressions, Equation (3) can be used as model also for this scenario. In this case, each component of the vector \mathbf{x} is associated with a TF. Since in this model $\mathbb{E}\mathbf{x}_i\mathbf{x}_j = 0$, $i \neq j$, we assume that each TF action is uncorrelated with the others. Furthermore, the $N \times q$ matrix F is the strength of the TFs effect on each gene and the vector $\tilde{\mathbf{y}}$ is used to describe local interactions of the genes. We shall see in the following that, in our modeling paradigm, such a vector has “weakly correlated components”.

In the model specification it is customary to assume that the noise vector $\tilde{\mathbf{y}}$ has uncorrelated components. In this case the model (3), is a (*static*) *Factor Analysis Model* (FA). Statistical inference on these models leads in general to ill-posed problems and to resolve the issue it is often imposed that the variances of the scalar components of $\tilde{\mathbf{y}}$ should be equal. Then a well-known bag of tricks applies and the inference problem can be solved essentially by computing the smallest eigenvalue of the covariance matrix Σ [20]. The assumption of equal variance and uncorrelated noises is however rather unrealistic in many instances.

II. PROBLEM STATEMENT AND SCOPE OF THE PAPER

Let $\mathbf{y}(k, t)$ be a second order (finite variance) random field depending on a space variable k and on a time variable t . The variable k is indexing an ensemble of equal “agents” which at each time t produce measurements, $\mathbf{y}(k, t)$, of a scalar quantity, say the received voltage signal of the k -th antenna or the expression level of the k -th gene in a microdata array. We shall assume that k varies on some ordered index set of N elements and let $t \in \mathbb{Z}$ or \mathbb{Z}_+ , depending on the

context. Eventually we shall be interested in problems where $N = \infty$. We shall denote by $\mathbf{y}(t)$ the random (column) vector with components $\{\mathbf{y}(k, t); k = 1, 2, \dots, N\}$. Suitable mathematical assumptions on this process will be specified in due time.

A (**random**) **flock** is a random field having the multiplicative structure $\hat{\mathbf{y}}(k, t) = \sum_{i=1}^q f_i(k)\mathbf{x}_i(t)$, or equivalently,

$$\hat{\mathbf{y}}(t) = \sum_{i=1}^q f_i\mathbf{x}_i(t)$$

where $f_i = [f_i(1) \ f_i(2) \ \dots \ f_i(N)]^\top$, $i = 1, 2, \dots, q$ are nonrandom N -vectors and $\mathbf{x}_i(t)$ are random processes depending on the time variable only. In this paper we shall investigate when a second order random field has a flocking component and study the problem of extracting it from sample measurements of $\mathbf{y}(k, t)$. This means that one should be searching for decompositions of the type:

$$\mathbf{y}(t) = \sum_{i=1}^q f_i\mathbf{x}_i(t) + \tilde{\mathbf{y}}(t)$$

where $q \geq 1$ and $\tilde{\mathbf{y}}(t)$ is a “random noise” field which should not contain flocking components. Naturally for the problem to be well-defined one has to specify conditions making this decomposition *unique*. The issue is a crucial one. Uniqueness has been addressed in the early literature on static GFA models [8] but in our opinion the reasons why this property may hold remain rather elusive. In this paper we shall propose an explanation which clarifies the question completely, at least in the stationary case. See also our earlier paper [3].

III. GENERALIZED FACTOR ANALYSIS MODELS

In this section we shall review the main ideas of the papers [8], [15], restricting for now to the *static* case. Consider a zero-mean finite variance stochastic process $\mathbf{y} := \{\mathbf{y}(k), k \in \mathbb{N}\}$, which we shall normally represent as a random column vector with an infinite number of components. The index k will have the interpretation of a space variable. We want to describe the process as a linear combination of a finite number of common random components plus “noise”, i.e.

$$\mathbf{y}(k) = f(k)^\top \mathbf{x} + \tilde{\mathbf{y}}(k), \quad k = 1, 2, \dots \quad (4)$$

where \mathbf{x} is a q -dimensional random vector which we shall take, without loss of generality, with orthonormal components; i.e. $\text{Var}[\mathbf{x}] = I_q$. The components \mathbf{x}_i , $i = 1, \dots, q$ are the *common factors* and the deterministic vectors f_i are sometimes called *factor loadings*. The linear combination $f(k)^\top \mathbf{x}$ is denoted $\hat{\mathbf{y}}(k)$ for short and as $\hat{\mathbf{y}}$ in vector notation. The $\tilde{\mathbf{y}}(k)$'s are zero mean random variables orthogonal to \mathbf{x} . The specific character of the process $\tilde{\mathbf{y}}$ will be discussed later (see the definition of idiosyncratic noise below). The infinite covariance matrix of the vector \mathbf{y} is formally written as $\Sigma := \mathbb{E}\mathbf{y}\mathbf{y}^\top$. We let Σ_n indicate 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 $\Sigma > 0$ means that all submatrices Σ_n

of Σ are positive definite, which we shall always assume in the following. Letting $\hat{\Sigma} := \mathbb{E}\hat{\mathbf{y}}\hat{\mathbf{y}}^\top$ and $\tilde{\Sigma} := \mathbb{E}\tilde{\mathbf{y}}\tilde{\mathbf{y}}^\top$, the orthogonality of the noise term and the factor components implies that

$$\Sigma = \hat{\Sigma} + \tilde{\Sigma}, \quad (5)$$

that is, $\Sigma_n = \hat{\Sigma}_n + \tilde{\Sigma}_n$, $\forall n \in \mathbb{N}$. Even imposing $\hat{\Sigma}$ of low rank, this is a priori a highly non unique decomposition.

There are situations/examples in which the $\tilde{\Sigma}$ is diagonal as in the static case, but these situations are exceptional.

Example 1 (Exchangeable processes): Consider the case of a (weakly) exchangeable process \mathbf{y} ; i.e. a process whose second order statistics are invariant with respect to all finite index permutations. Clearly the covariances $\sigma(k, j) = \mathbb{E}\mathbf{y}(k)\mathbf{y}(j)$ must be independent of k, j for $k \neq j$ and $\sigma(k, k) = \sigma^2 > 0$ independent of k [2]. Letting $\rho := \sigma(k, j)$, $k \neq j$, one has

$$\Sigma = \begin{bmatrix} \sigma^2 & \rho & \rho & \rho & \dots \\ \rho & \sigma^2 & \rho & \rho & \dots \\ \dots & & \ddots & & \dots \end{bmatrix} \quad (6)$$

where we must have $\sigma^2 > |\rho|$ for positive definiteness. Letting f denote an infinite column vector with components all equal to ρ , one can decompose Σ as $\Sigma = ff^\top + (\sigma^2 - \rho)I$, where here I denotes an infinite identity matrix. This is a Factor Analysis decomposition of rank $q = 1$ of Σ with $\tilde{\Sigma}$ a diagonal matrix.

Let $\ell^2(\Sigma)$ denote the Hilbert space of infinite sequences $a := \{a(k), k \in \mathbb{N}\}$ such that $\|a\|_\Sigma^2 := a^\top \Sigma a < \infty$. When $\Sigma = I$ we simply use the symbol ℓ^2 , denoting the corresponding norm with the symbol $\|\cdot\|$.

The sequence whose first n elements are the same as in a , while the others are set equal to zero is denoted $a_{[n]}$.

Definition 1: Let $\{a_n, n \in \mathbb{N}\}$ be a sequence of elements of the set $\ell^2 \cap \ell^2(\Sigma)$. We say that $\{a_n, n \in \mathbb{N}\}$ is an *averaging sequence* (AS) for \mathbf{y} , if $\lim_{n \rightarrow \infty} \|a_n\| = 0$.

Example 2: The sequence of elements in ℓ^2

$$a_n = \frac{1}{n} \underbrace{[1 \dots 1]}_n 0 \dots]^\top \quad (7)$$

is an averaging sequence.

An AS can be seen just as a sequence of linear functionals in $\ell^2 \cap \ell^2(\Sigma)$ converging strongly to zero. The definition is instrumental to the concept of *idiosyncratic sequence* of random variables.

Definition 2 ([15]): We say that \mathbf{y} is *idiosyncratic* if $\lim_{n \rightarrow \infty} a_n^\top \mathbf{y} = 0$ for any averaging sequence $a_n \in \ell^2 \cap \ell^2(\Sigma)$.

Example 3: Let $\mathbf{1}$ be an infinite column vector of 1's and let \mathbf{x} be a scalar random variable uncorrelated with $\tilde{\mathbf{y}}$, a zero-mean weakly stationary ergodic sequence. Consider the process

$$\mathbf{y} = \mathbf{1}\mathbf{x} + \tilde{\mathbf{y}}$$

Since $L^2 - \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \tilde{\mathbf{y}}(k) = \mathbb{E}\tilde{\mathbf{y}}(k) = 0$, we have $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbf{y}(k) = \mathbf{x}$ hence we can recover the latent

factor by averaging. In fact, it will be shown below that $\lim_{n \rightarrow \infty} a_n^\top \tilde{\mathbf{y}} = 0$ for any averaging sequence.

The following definition is meant to capture the phenomenon described in the example. Let $H(\mathbf{y})$ be the Hilbert space spanned by the sequence $\{\mathbf{y}(k), k \in \mathbb{N}\}$.

Definition 3: Let $\mathbf{z} \in H(\mathbf{y})$. The random variable \mathbf{z} is an *aggregate* (of \mathbf{y}) if there exists an AS $\{a_n\}$ such that $\lim_{n \rightarrow \infty} a_n^\top \mathbf{y} = \mathbf{z}$. 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

$$\mathbf{y} = \mathbb{E}[\mathbf{y} | \mathcal{G}(\mathbf{y})] + \mathbf{u}, \quad (8)$$

where all components $\mathbf{u}(k)$ 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 (4). 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 4: ([15]) Consider a (possibly non-stationary) white noise sequence \mathbf{y} with $\mathbf{y}(j) \perp \mathbf{y}(h) \forall j \neq h$, and let \mathbf{z} be an aggregate random variable, so that there must exist an AS a_n such that

$$\mathbf{z} = \lim_{n \rightarrow \infty} a_n^\top \mathbf{y} = \lim_{n \rightarrow \infty} \sum_{j=1}^{\infty} a_n(j) \mathbf{y}_j. \quad (9)$$

Note that, being $\mathbf{z} \in H(\mathbf{y})$ and \mathbf{y} an orthogonal basis of this space, we can uniquely express \mathbf{z} as $\mathbf{z} = \sum_{j=1}^{\infty} b(j) \mathbf{y}(j)$ and, by uniqueness of the representation, it follows that $\lim_{n \rightarrow \infty} a_n(j) = b(j) \forall j$. On the other hand, being a_n an AS, the limits of $a_n(j)$ must be zero, so that $b(j) = 0$. Hence $\mathbf{z} = 0$. Thus for a white noise process one always has $\mathcal{G}(\mathbf{y}) = \{0\}$. However if $\{\mathbf{y}(k)\}$ has unbounded variance, the sequence may not be idiosyncratic, say if $\|\mathbf{y}(k)\|^2 = k$, given the AS

$$d_n = \frac{1}{\sqrt{n}} \underbrace{[0 \dots 0]}_n 1 0 \dots]^\top, \quad (10)$$

we have $\|d_n^\top \mathbf{y}\| = 1 \forall n$. Later it will be shown that idiosyncratic sequences must have a bounded covariance matrix.

The nature of an idiosyncratic sequence is strictly related to the behavior of the eigenvalues of its covariance matrix. To explain this point, we need to introduce some notations and facts about the eigenvalues of infinite covariance matrices. Denote by $\lambda_{n,k}(\Sigma)$ the k -th eigenvalue of the $n \times n$ upper left submatrix Σ_n of Σ . The $\lambda_{n,k}(\Sigma)$ are real nonnegative and can be ordered in decreasing magnitude. Forni and Lippi [15, Fact M], show that the k -th eigenvalue of Σ_n is a non decreasing function of n and hence has a limit, $\lambda_k(\Sigma)$, 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

the infinite matrix Σ is bounded as a linear operator in ℓ^2 [4]. The limits however are in general not true eigenvalues of this operator, as it is well-known that Σ may not have eigenvalues. For example, a bounded symmetric Toeplitz matrix has a purely continuous spectrum [16]. Anyway since Σ is symmetric and positive, its spectrum lies on the positive half line and its elements can also be ordered. Henceforth we shall denote by $\lambda_1(\Sigma)$ the maximal eigenvalue of Σ , with the convention that $\lambda_1(\Sigma) = +\infty$ when there are infinite eigenvalues as defined above.

A strong characterization of idiosyncratic sequences is rephrased in the following theorem, stated after [15] with some obvious simplifications.

Theorem 1: The sequence \mathbf{y} is idiosyncratic if and only if $\lambda_1(\Sigma)$ is finite. Equivalently, \mathbf{y} is idiosyncratic if and only if its covariance matrix defines a bounded operator in ℓ^2 .

The following definition characterizes a particular type of sequences.

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 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 are q infinite column vectors (real sequences) $f_1(\cdot), f_2(\cdot), \dots, f_q(\cdot)$ such that

$$\mathbf{y}(k) = \sum_{i=1}^q f_i(k) \mathbf{x}_i, \quad k \in \mathbb{N}, \quad (11)$$

The functions $f_1(\cdot), f_2(\cdot), \dots, f_q(\cdot)$ must be linearly independent, for otherwise the rank of \mathbf{y} would be smaller than q . Notation: we shall write $F := [f_1 \ f_2 \ \dots \ f_q]$ and let $F(k)$ denote the k -th row of this infinite matrix.

We want to relate the concept of q -PD to the idea of aggregation subspace of \mathbf{y} , as defined earlier. In particular we would like to identify \mathbf{x} as an orthonormal basis in $\mathcal{G}(\mathbf{y})$. Quite unfortunately however, there are nontrivial sequences representable in the form (11) which are idiosyncratic (or contain idiosyncratic sequences). See the Example below.

Example 5: Consider a sequence \mathbf{y} whose k -th element is

$$\mathbf{y}(k) = \lambda^k \mathbf{x}, \quad |\lambda| < 1, \quad (12)$$

where \mathbf{x} is a zero-mean random variable of positive variance σ^2 . Clearly, \mathbf{y} is 1-PD, its spanned subspace $H(\mathbf{y})$ being the one-dimensional space $H(\mathbf{x})$. It can be easily seen that $\text{rank}(\Sigma) = 1$ and $\lambda_1(\Sigma) = \frac{\sigma^2 \lambda^2}{1 - \lambda^2}$, 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 (4) has to be unique. In the following, we shall call a q -PD sequence q -**aggregate** if its covariance matrix has q nonzero eigenvalues (i.e. $\text{rank} \hat{\Sigma}_n = q, \forall n$), which are *all infinite*. The question is which properties need to be satisfied by the functions f_1, f_2, \dots, f_q for \mathbf{y} to be a q -aggregate sequence. One necessary condition is easily

found: the f_i cannot be in ℓ^2 since otherwise any sequence of functionals $\{a_n\}$ in ℓ^2 converging strongly (and hence also weakly) to zero would lead to $\lim_{n \rightarrow \infty} a_n^\top f_i = 0$, so that $\lim_{n \rightarrow \infty} a_n^\top \mathbf{y} = 0$ as well. This is clearly the problem with Example 5. To this end, [8], [15] impose a condition on the eigenvalues of the covariance matrix of a q -PD sequence. We instead introduce the following definition.

Definition 5: The vectors $f_i, i = 1, \dots, q$ in \mathbb{R}^∞ are **strongly linearly independent** if

$$\lim_{n \rightarrow \infty} \|f_i^n(\cdot) - \Pi[f_i^n(\cdot) | \mathcal{F}_i^n]\|_2 = +\infty. \quad (13)$$

where Π is the orthogonal projection onto the Euclidean space $\mathcal{F}_i^n = \text{span}\{f_j^n(\cdot), j = 1, \dots, q, j \neq i\}$.

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

$$\mathbf{y}(k) = F(k) \mathbf{x} = \sum_{i=1}^q f_i(k) \mathbf{x}_i, \quad k \in \mathbb{N}; \quad (14)$$

then \mathbf{y} is q -aggregate if and only if, the vectors $f_i, i = 1, \dots, q$ are strongly linearly independent.

We eventually come to a precise definition of the basic object of our study. The following definition is the static version of a similar one of [15] for the dynamic setting.

Definition 6: Let F be an $\infty \times q$ real matrix with rows $F(k) k = 1, 2, \dots$. The sequence \mathbf{y} is a q -**factor sequence** (q -FS) if it can be written as an orthogonal sum

$$\mathbf{y}(k) = F(k) \mathbf{x} + \tilde{\mathbf{y}}(k), \quad k = 1, 2, \dots \quad (15)$$

where $\hat{\mathbf{y}}(k) := F(k) \mathbf{x}$ is a q -aggregate sequence and $\tilde{\mathbf{y}}$ is idiosyncratic and orthogonal to \mathbf{x} . The representation (15) is called a **Generalized Factor Analysis model with q factors**. The crucial question is now which random sequences are q -FS. Chamberlain and Rothschild [8] (see also [15]) provide the following criterion.

Theorem 3 (Chamberlain-Rothschild): If and only if for $n \rightarrow \infty$, Σ_n has q **unbounded eigenvalues** and $\lambda_{q+1}(\Sigma_n)$ stays bounded, then Σ has a GFA decomposition with q latent factors; i.e.

$$\Sigma = FF^\top + \tilde{\Sigma}, \quad \text{with } F = [f_1 \ \dots \ f_q], \quad f_i \in \mathbb{R}^\infty \quad (16)$$

where $\lambda_1(\tilde{\Sigma}) < \infty$. This decomposition is unique.

Corollary 1: The covariance Σ has a GFA decomposition with q latent factors if and only if it can be decomposed as the sum of a matrix operator $\tilde{\Sigma}$ which is bounded in ℓ^2 and a rank q perturbation $\hat{\Sigma} = FF^\top$ where $F \in \mathbb{R}^{\infty \times q}$ has strongly linearly independent columns.

Let $f_{i,n} \in \mathbb{R}^n$ be the eigenvectors of Σ_n corresponding to the q largest eigenvalues of Σ_n and let $F_n := [f_{1,n} \ \dots \ f_{q,n}]$. Then we have $\lim_{n \rightarrow \infty} F_n F_n^\top = FF^\top$ and hence we can get q and F by a limit procedure of PCA on Σ_n .

Unfortunately, as shown by Example 5, there are sequences which do not admit a q -FS (with q finite). The covariance matrix of the scalar sequence \mathbf{y} of Example 5 is diagonal with infinitely many eigenvalues of Σ equal to $+\infty$, according to the definition given above. Hence the sequence may be called a ∞ -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 y is a purely-non-deterministic sequence. On the other hand \mathbf{y} is not idiosyncratic either since we have shown that $\mathcal{G}(\mathbf{y}) = \{0\}$.

A. 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. $\mathbb{E}\mathbf{y}(k)\mathbf{y}(j) = \sigma(k-j)$ for $k, j \geq 0$. It is well known, see e.g. [12], [19] that, introducing the *remote future subspace* of \mathbf{y} , namely $H_\infty(\mathbf{y}) := \bigcap_{t \geq 0} H_t(\mathbf{y})$, the sequence of orthogonal wandering subspaces $\tilde{E}_t := H_t(\mathbf{y}) \ominus H_{t+1}(\mathbf{y})$ and their orthogonal direct sum $\tilde{H}(\mathbf{y}) := \bigoplus_{t \geq 0} \tilde{E}_t$, one has a unique orthogonal decomposition

$$\mathbf{y} = \hat{\mathbf{y}} + \tilde{\mathbf{y}}, \quad \hat{\mathbf{y}}(k) \in H_\infty(\mathbf{y}), \quad \tilde{\mathbf{y}}(k) \in \tilde{H}(\mathbf{y}) \quad (17)$$

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 following theorem enlightens the nature of GFA decompositions.

Theorem 4: Every stationary sequence with a bounded spectral density and remote future space of dimension q is a q -factor sequence and admits a unique generalized factor analysis decomposition (15) where $\hat{\mathbf{y}}$ is the PD and $\tilde{\mathbf{y}}$ the PND component of \mathbf{y} .

It is not hard to show that the assumption of stationarity here is crucial. In fact, Example 5 discussed before shows that in the non-stationary case a PD process \mathbf{y} whose remote future is the one-dimensional space $H(\mathbf{x})$ may be idiosyncratic.

IV. USE AND INTERPRETATION OF GFA MODELS

We now suggest an interpretation of GFA models in the framework of an ensemble of infinitely many agents distributed in space and interacting in a random fashion.

1) *Short distance interaction:* The idiosyncratic covariances $\{\tilde{\sigma}(k, j)\}$ describe the mutual influence of neighbouring units noises $\tilde{\mathbf{y}}(k), \tilde{\mathbf{y}}(j)$. Since $\tilde{\Sigma}$ is a bounded operator in ℓ^2 , it is a known fact [1, Section 26] that $\tilde{\sigma}(k, j) \rightarrow 0$ as $|k-j| \rightarrow \infty$ so in a sense the idiosyncratic component $\tilde{\mathbf{y}}$ models only *short range interactions* among the agents, which are decaying with distance. Agents which are far away from each other do not resent of mutual influence.

2) *Factor loadings and long range influence:* Since $\mathbb{E}\hat{\mathbf{y}}(k)\hat{\mathbf{y}}(j) = \sum_i f_i(k)f_i(j)$ and the elements of the column vectors $f_i \in \mathbb{R}^\infty$ do not decay with distance, the products $f_i(k)f_i(j)$ do not vanish when $|k-j| \rightarrow \infty$. Hence the factor loadings describe “long range” correlation between the factor components and the $\hat{\mathbf{y}}$ component of \mathbf{y} can be interpreted as variables produced by *long range interaction* among agents.

A. Stochastic realization

Assume we are given a GFA decomposition $\hat{\Sigma} + \tilde{\Sigma}$ of the covariance Σ . How do we find the hidden variables in the representation $\mathbf{y} = F\mathbf{x} + \tilde{\mathbf{y}}$?

We can answer this question under the constraint is that both \mathbf{x} and $\tilde{\mathbf{y}}$ belong to $H(\mathbf{y})$. Models of this kind are called

internal in stochastic realization. Assume that $\hat{\Sigma} = FF^\top$ and that $\tilde{\Sigma}$ is invertible, then a classical orthogonal projection formula yields

$$\hat{\mathbf{x}} := \hat{\mathbb{E}}(\mathbf{x} | \mathbf{y}) = (I + F^\top \tilde{\Sigma}^{-1} F)^{-1} F^\top \tilde{\Sigma}^{-1} \mathbf{y}$$

since the variance matrix of \mathbf{x} is the $q \times q$ identity. Now since \mathbf{x} belong already to $H(\mathbf{y})$ we have $\mathbf{x} \equiv \hat{\mathbf{x}}$ modulo orthogonal transformations. Then $\tilde{\mathbf{y}} = \mathbf{y} - F\hat{\mathbf{x}}$ by the orthogonality principle.

V. GFA MODELS OF RANDOM FIELDS

We come back to the question raised in section II namely when does a second order random field have a flocking component and how to extract it from sample measurements of $\mathbf{y}(k, t)$. A simple class of random fields for which this question can be answered positively is the class of multiplicative processes

$$\mathbf{y}(k, t) = \mathbf{v}(k)\mathbf{u}(t) \quad (18)$$

where the time component \mathbf{u} has finite variance and \mathbf{v} is second-order weakly stationary satisfying the conditions of Theorem 4. This model can be generalized, for example making both $\mathbf{v}(k)$ and $\mathbf{u}(t)$ vector-valued but for the sake of clarity we shall restrain to the scalar case. According to the model (18) all agents evolve in time with essentially the same temporal dynamics described by the process $\{\mathbf{u}(t)\}$ which is amplitude-modulated by a random space-dependent multiplicative coefficient $\mathbf{v}(k)$. Let

$$\mathbf{v}(k) = \sum_{i=1}^q f_i(k)\mathbf{z}_i + \tilde{\mathbf{v}}(k) \quad (19)$$

be the unique generalized factor analysis decomposition of \mathbf{v} where $\hat{\mathbf{v}}(k) := \sum_i f_i(k)\mathbf{z}_i$ is the PD and $\tilde{\mathbf{v}}(k)$ the PND components of $\mathbf{v}(k)$. Then setting $\mathbf{x}_i(t) = \mathbf{z}_i\mathbf{u}(t)$ and $\tilde{\mathbf{y}}(k, t) := \tilde{\mathbf{v}}(k)\mathbf{u}(t)$ one can rewrite (18) as a dynamic GFA,

$$\mathbf{y}(k, t) = \sum_{i=1}^q f_i(k)\mathbf{x}_i(t) + \tilde{\mathbf{y}}(k, t) := \hat{\mathbf{y}}(k, t) + \tilde{\mathbf{y}}(k, t)$$

The following assumption specifies in probabilistic terms the multiplicative structure (18) of the random field $\mathbf{y}(k, t)$.

Assumption : *The space and time evolutions of $\mathbf{y}(k, t)$ are multiplicatively uncorrelated in the sense that*

$$\mathbb{E}_{\mathbf{v}}\{\mathbf{v}(k_1)\mathbf{v}(k_2) | \mathbf{u}(t_1)\mathbf{u}(t_2)\} = \mathbb{E}_{\mathbf{v}}\{\mathbf{v}(k_1)\mathbf{v}(k_2)\} \quad (20)$$

where the first conditional expectation is made with respect to the conditional probability distribution of \mathbf{v} given the random variables $\mathbf{u}(t_1), \mathbf{u}(t_2)$, while the second expectation is with respect to the marginal distribution of \mathbf{v} .

From (20) we get

$$\begin{aligned} \mathbb{E}\{\mathbf{v}(k_1)\mathbf{v}(k_2)\mathbf{u}(t_1)\mathbf{u}(t_2)\} &= \\ \mathbb{E}\{\mathbf{v}(k_1)\mathbf{v}(k_2)\} \mathbb{E}\{\mathbf{u}(t_1)\mathbf{u}(t_2)\} &= \sigma_{\mathbf{v}}(k_1 - k_2) \sigma_{\mathbf{u}}(t_1, t_2) \end{aligned} \quad (21)$$

where $\sigma_{\mathbf{v}}$ and $\sigma_{\mathbf{u}}$ are the covariance functions of the two processes. Hence the covariance function of the random field inherits the multiplicative structure. If \mathbf{v} and \mathbf{u} are jointly

Gaussian, the multiplicative uncorrelation property follows from the multiplicative structure of the joint covariance.

Proposition 1: If the processes \mathbf{v} and \mathbf{u} are multiplicatively uncorrelated then the two terms $\hat{\mathbf{y}}(k_1, t_1)$ and $\tilde{\mathbf{y}}(k_2, t_2)$ in the GFA model (V) are uncorrelated for any $k_1, k_2 \in \mathbb{Z}_+$, $t_1, t_2 \in \mathbb{Z}$.

Proof: We have

$$\mathbb{E}\{\hat{\mathbf{y}}(k_1, t_1)\tilde{\mathbf{y}}(k_2, t_2)\} = \sum_{i=1}^q f_i(k_1)\mathbb{E}\{\mathbf{z}_i\mathbf{u}(t_1)\tilde{\mathbf{v}}(k_2)\mathbf{u}(t_2)\}$$

where the last term can be written as

$$\begin{aligned} & \mathbb{E}\{\mathbf{z}_i\tilde{\mathbf{v}}(k_2)\mathbf{u}(t_1)\mathbf{u}(t_2)\} = \\ & \mathbb{E}\{\mathbb{E}_{\mathbf{v}}[\mathbf{z}_i\tilde{\mathbf{v}}(k_2) \mid \mathbf{u}(t_1)\mathbf{u}(t_2)]\mathbf{u}(t_1)\mathbf{u}(t_2)\} = \\ & \mathbb{E}\{\mathbb{E}_{\mathbf{v}}[\mathbf{z}_i\tilde{\mathbf{v}}(k_2)]\mathbf{u}(t_1)\mathbf{u}(t_2)\} = 0 \end{aligned} \quad (22)$$

since the \mathbf{z}_i 's are random variables in $H(\hat{\mathbf{v}})$ and $\tilde{\mathbf{v}}(k)$ is $(\mathbb{E}_{\mathbf{v}})$ -orthogonal to this space. ■

A. Statistical estimation

Assume that we have a snapshot of the system at certain time t_0 ; that is we have observations of a “very large” portion of the process $\{\mathbf{y}(k, t_0), k = 1, 2, \dots, N\}$ at some fixed time t_0 . With these sample data we may form the sample covariance estimates

$$\begin{aligned} \hat{\sigma}_N(h, t_0) & := \frac{1}{N} \sum_{k=1}^N y(k+h, t_0)y(k, t_0) \\ & = \frac{1}{N} \sum_{k=1}^N v(k+h)v(k)u(t_0)^2, \quad h = 0, 1, \dots \end{aligned} \quad (23)$$

which also have the multiplicative structure,

$$\hat{\sigma}_N(h, t_0) = \hat{\sigma}_{\mathbf{v}, N}(h)u(t_0)^2 \quad (24)$$

where $\hat{\sigma}_{\mathbf{v}, N}(h)$ is the sample covariance estimate of the \mathbf{v} process based on N data. Now by the assumptions made on the space-process \mathbf{v} the limit $\lim_{N \rightarrow \infty} \hat{\sigma}_N(h, t_0)$ exists (although it may be sample dependent for the PD part), so the sample matrix covariance estimate, which has the form

$$\begin{aligned} \hat{\Sigma}_N(t_0) & = \text{Toeplitz}\{\hat{\sigma}_N(0, t_0), \dots, \hat{\sigma}_N(N-1, t_0)\} \\ & = u(t_0)^2 \hat{\Sigma}_{\mathbf{v}, N}. \end{aligned} \quad (25)$$

will converge to a limit for $N \rightarrow \infty$. Following [8], [15] the idea is now to do PCA on the covariance estimate for increasing N and isolate q eigenvalues which tend to grow without bound as $N \rightarrow \infty$ while the others stay bounded. The q corresponding eigenvectors will tend as $N \rightarrow \infty$ to the q factor loadings f_1, \dots, f_q and therefore provide asymptotically the F.A. decomposition of the $\Sigma_{\mathbf{v}}$ matrix

$$\Sigma_{\mathbf{v}} = FF^{\top} + \tilde{\Sigma}_{\mathbf{v}} \quad (26)$$

After F and $\tilde{\Sigma}_{\mathbf{v}}$ are estimated, the stochastic realization procedure of Section IV-A permits to construct the factor vector \mathbf{z} and the idiosyncratic component $\tilde{\mathbf{v}}$ of the GFA representation of \mathbf{v} as in (19). The reconstruction of the time varying factor variables $\mathbf{x}_i(t) = \mathbf{z}_i\mathbf{u}(t)$ of \mathbf{y} from the observations $\mathbf{y}(k, t) = \mathbf{v}(k)\mathbf{u}(t)$ can be done, in several equivalent ways, by averaging on the space variable.

VI. CONCLUSIONS

In this paper we have presented a rather complete analysis of space-stationary Generalized Factor Analysis models. We have discussed the interpretation of the factor component as a flocking configuration for an ensemble of equal agents and the modeling of simple multiplicative random fields as the sum of a flocking component plus a random term which describes local random disagreements of the individual agents or the effect of external disturbances.

REFERENCES

- [1] N. Akhiezer and I. M. Glazman, *Theory of Linear Operators in Hilbert Space Vol I*. New York: Fredrik Ungar Pub. Co., 1961.
- [2] D. Aldous, “Exchangeability and related topics,” in *École d’Été de Probabilités de Saint-Flour XIII*, 1985, vol. Volume 1117, pp. 1–198.
- [3] G. Bottegal and G. Picci, “A note on generalized factor analysis models,” in *Proc. 50th Decision and Control Conference (CDC)*, Orlando FLA, USA, 2011, pp. 1485–1490.
- [4] —, “Flocking and generalized factor analysis,” Available on arXiv.org, Tech. Rep., 2013.
- [5] P. Brazhnik, A. de la Fuente, and P. Mendes, “Gene networks: how to put the function in genomics,” *Trends in Biotechnology*, vol. 20, no. 11, pp. 467–472, 2002.
- [6] R. W. Brockett, “On the control of a flock by a leader,” in *Differential equations and topology. I*, ser. Tr. Mat. Inst. Steklova. Moscow: MAIK Nauka/Interperiodica, 2010, vol. 268, pp. 56–63.
- [7] G. Chamberlain, “Funds, factors and diversification in arbitrage pricing models,” *Econometrica*, vol. 51, no. 5, pp. 1305–1324, 1983.
- [8] G. Chamberlain and M. Rothschild, “Arbitrage, factor structure and meanvariance analysis on large asset markets,” *Econometrica*, vol. 51, no. 5, pp. 1281–1304, 1983.
- [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] F. Cucker and S. Smale, “Emergent behavior in flocks,” *IEEE Transactions on Automatic Control*, pp. 852–862, 2007.
- [11] M. Deistler, B. D. O. Anderson, A. Filler, C. Zinner, and W. Chen, “Generalized linear dynamic factor models: An approach via singular autoregressions,” *European Journal of Control*, vol. 3, pp. 211–224, 2010.
- [12] 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.
- [13] F. Fagnani and S. Zampieri, “Randomized consensus algorithms over large scale networks,” *IEEE J. Selected Areas Commun.*, vol. 26, p. 634649, 2008.
- [14] 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.
- [15] M. Forni and M. Lippi, “The generalized dynamic factor model: representation theory,” *Econometric Theory*, vol. 17, pp. 1113–1141, 2001.
- [16] P. Hartman and A. Wintner, “The spectra of Toeplitz matrices,” *American Journal of Mathematics*, vol. 76, pp. 867–882, 1954.
- [17] R. Olfati-Saber, “Flocking for multi-agent dynamic systems: Algorithms and theory,” *IEEE Transactions on Automatic Control*, vol. 51, p. 401420, 2006.
- [18] R. Olfati-Saber, J. Fax, and R. Murray, “Consensus and cooperation in networked multi-agent systems,” *Proc. of the IEEE*, vol. 95, pp. 215–233, 2007.
- [19] Y. A. Rozanov, *Stationary Random Processes*, Holden-Day, Ed. Holden Days, 1967.
- [20] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Prentice Hall, 1997.
- [21] A. Tahbaz-Salehi and A. Jadbabaie, “Consensus over ergodic stationary graph processes,” *IEEE Transactions on Automatic Control*, vol. 55, pp. 225–230, 2010.
- [22] J. Veerman, J. Caughman, G. Lafferriere, and A. Williams, “Flocks and formations,” *J. Stat.Phys.*, vol. 121, Vol 5-6, pp. 901–936, 2005.