# Modeling Complex Systems by Generalized Factor Analysis

Giulio Bottegal and Giorgio Picci, Life Fellow, IEEE

*Abstract*—We propose a new modeling paradigm for large dimensional aggregates of stochastic systems by Generalized Factor Analysis (GFA) models. These models describe the data as the sum of a *flocking* plus an uncorrelated *idiosyncratic* component. The flocking component describes a sort of collective orderly motion which admits a much simpler mathematical description than the whole ensemble while the idiosyncratic component describes weakly correlated noise. We first discuss static GFA representations and characterize in a rigorous way the properties of the two components. The extraction of the dynamic flocking component is discussed for time-stationary linear systems and for a simple classes of separable random fields.

*Index Terms*—Collective behavior, complex systems, flocking, generalized factor analysis, multi-agent systems, stochastic systems.

#### I. INTRODUCTION

T has been observed in several circumstances [3], [16]-**[**18] that modeling and identification of complex stochastic systems by traditional AR or ARMA models may lead to problems where the number of parameters can be of the same order of magnitude or larger than the sample size. The only way out of this problem seems to be to change our ideas on modeling. In this paper we propose a new paradigm on stochastic modeling of complex systems based on the theory of Generalized Factor Analysis (GFA) and the idea of stochastic flocking. Although the two terminologies belong to different cultures which seem to have little in common, our point in this paper will be to show that dynamic GFA modeling of a large ensemble of interacting random units hinges on splitting the overall motion into a component which deserves the name of *flocking* plus a weakly correlated kind of noise. The latter is called the *idiosyncratic* component. The first component describes the average random motion of the system by a rather simple statistical model while the second aims at describing the stochastic dynamics which pertains exclusively to individual fluctuations about the average.

The word *Flocking* is used to describe a commonly observed behavior in gregarious animals by which many equal individu-

G. Bottegal is with the ACCESS Linnaeus Centre, School of Electrical Engineering, KTH Royal Institute of Technology, 10044 Stockholm, Sweden (e-mail: bottegal@kth.se).

G. Picci is with the Department of Information Engineering, University of Padova, 35131 Padova, Italy (e-mail: picci@dei.unipd.it).

Digital Object Identifier 10.1109/TAC.2014.2357913

als tend to group and follow, at least approximately, a common path in space. The phenomenon has been studied very actively in recent years; see e.g., [9], [48], [55], [57] and the literature on this subject is now huge, consisting of hundreds of papers which would be impossible to discuss here. Our interest in flocking derives from the fact that the phenomenon has similarities with many scenarios observed in artificial/technological environments a few examples of which will be described below.

The mechanism of formation of flocks is sometimes also called *convergence to consensus* and has been intensely studied in the literature. There is now a quite articulated theory addressing the convergence to consensus under a variety of assumptions on the communication strategy among agents, specific nonlinearities of the dynamics, the kind of permissible local control actions etc. see e.g., [15], [21], [29], [41], [42], [51], [53], [54] and references therein.

In this paper, we want to address a different and possibly more basic issue: given observations of the motion of a large set of interacting agents and assuming statistical steady state, find out whether there is a flocking component in the collective motion and estimate its characteristics. The rationale for this search is that the very concept of flocking implies an *orderly* motion which must then admit a much simpler mathematical description than that of 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. Prediction of the future behavior and control of a complex ensemble of random agents could then reasonably be restricted to the flocking component and be based on the simple model thereof.

#### A. Problem Statement and Scope of the Paper

We start by setting notations: In this paper boldface symbols will normally denote random arrays, either finite or infinite. All random variables will be real, zero-mean and with finite variance. In the following we shall denote by the symbol  $H(\mathbf{v})$ the standard Hilbert space of random variables linearly generated by the scalar components  $\{\mathbf{v}_1, \ldots, \mathbf{v}_n, \ldots\}$  of a (possibly infinite) family of random variables which we generically denote  $\mathbf{v}$ . For  $\boldsymbol{\xi}, \boldsymbol{\eta} \in H(\mathbf{v})$ , the inner product is the mathematical expectation  $\langle \boldsymbol{\xi}, \boldsymbol{\eta} \rangle := \mathbb{E} \boldsymbol{\xi} \boldsymbol{\eta}$  which induces the (variance) norm of random variables by setting  $\|\boldsymbol{\xi}\|^2 = \mathbb{E} \boldsymbol{\xi}^2$ . Convergence of random sequences will always be understood with respect to this norm.

Manuscript received September 5, 2013; revised March 26, 2014, March 27, 2014, April 25, 2014, August 4, 2014, and August 18, 2014; accepted September 11, 2014. Date of publication September 15, 2014; date of current version February 19, 2015. This work was supported in part by the Swedish Research Council under contract 621-2009-4017. Recommended by Associate Editor L. H. Lee.

Let  $\mathbf{y}(k,t)$  be a finite variance random field depending on a discrete space variable k and on a time variable t. We shall denote by 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. The variable k is indexing (space) locations of a large ensemble of "agents" each of which produces at time t a scalar measurement,  $\mathbf{y}(k, t)$ , of an observable quantity.<sup>1</sup> 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  $N = \infty$ . The following is a first attempt to define in precise terms a random flock. The definition is given for a finite ensemble and, as it stands, may lead to non-unique modeling of the same system, which is instead of paramount importance in statistical identification theory. The notion will have to be refined later for an infinite ensemble. See Sections IV and VI.

A q-dimensional **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

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

where  $f_i = [f_i(1) \ f_i(2) \ \dots \ f_i(N)]^\top$ ,  $i = 1, 2, \dots, q$  are nonrandom N-vectors which do not depend on time and  $\mathbf{x}(t) := [\mathbf{x}_1(t) \ \dots \ \mathbf{x}_q(t)]^\top$  is a random processes depending on the time variable only, which can be chosen with orthonormal components; i.e.,  $\mathbb{E}\mathbf{x}(t)\mathbf{x}(t)^\top = I_q, \ t \in \mathbb{Z}$ .

The idea is that a random flock can essentially be regarded as a rigid deterministic geometric configuration of N objects (or points) in space moving synchronously in a random fashion. A very simple intuitive picture can be imagined extending for a moment our model to allow for three dimensional (vector valued) outputs  $\mathbf{y}(k,t) \in \mathbb{R}^3$ ,  $k = 1, 2, \ldots$  The k-th agent can then be visualized as a point moving in 3-D space. Let q be also equal to three and think of the 3-dimensional random motion with coordinates  $\mathbf{x}(t)$  as the motion of, say, the barycenter of the ensemble. This implies that all different agents follow the same trajectory in 3-dimensional space, modulo a constant offset depending on their relative location. In general however the agent's output does not need to be of the same dimension of the common state  $\mathbf{x}(t)$ . As already said, here for simplicity we restrict to one dimensional output components.

The above may look like a rather crude mathematical idealization of animal flocking behavior especially in that the spatial pattern of flocks of birds or herds of animals etc. may also deform its shape in time [27]. Note however that the deformations may be interpreted as random fluctuations about an average common trajectory that all agents of the flock aim at and that these fluctuations could in principle be embodied in the "noisy component"  $\tilde{y}$  of our modeling scheme. The denomination of random flock above may be reasonable as a description of the average behavior of a realistic flock. It should however be said very clearly that our objective in this paper is not to address animal behavior, but rather to discuss dynamical modeling of complex technological systems.

The main goal of this paper is to 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

$$\mathbf{y}(t) = \sum_{i=1}^{q} f_i \mathbf{x}_i(t) + \tilde{\mathbf{y}}(t)$$
(2)

where  $q \ge 1$  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*.

## B. Examples

1) 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 neighboring antennas. The setup can be described mathematically, by indexing each antenna by an integer i = 1, 2, ..., N and denoting by  $\mathbf{y}_i(t)$  the signal received at time t by antenna i. Then, model (2) can be used to describe the received signal, with  $\mathbf{x}(t)$  being the signals sent by the emitters at time t,  $f_i$  coefficients related to the distance between the emitters and antenna i and  $\tilde{\mathbf{y}}_i(t)$  some disturbance affecting antenna i at time t. Hence, we may identify  $\sum_{i=1}^{q} f_i \mathbf{x}_i(t)$  as the flocking component of  $\mathbf{y}(t)$ . The goal is to detect the number of emitters q and possibly estimate the signal components impinging on the antenna array. Note that in the model there are several hidden (non-measurable) variables, including the dimension q. In our setting N is assumed to be very large; ideally we shall assume  $N \to \infty$ . One should note that estimation of this model from observations of  $\mathbf{y}(t)$ consists first of estimating the model parameters, say  $\{f_i\}$  and the covariance matrix of  $\tilde{\mathbf{y}}(t)$  but also in constructing the hidden random quantities  $\mathbf{x}(t)$  and  $\tilde{\mathbf{y}}(t)$ . The covariance matrix of  $\mathbf{y}(t)$ , say  $\Sigma \in \mathbb{R}^{N \times N}$  may be obtained from the data by standard procedures.

A problem leading to models of similar structure 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.

2) Inference on Gene Regulatory Networks: Transcription Factors (TFs) are proteins which regulate gene expression binding to specific sequences of the promoter region of a gene. This regulation brings to the transcription of genes into mRNAs, which are in turn translated into proteins giving rise to a complex

<sup>&</sup>lt;sup>1</sup>An extension of the theory presented in this paper to a more general setting where each component  $\mathbf{y}(k,t)$  may take vector values, say in  $\mathbb{R}^m$ , is conceptually straightforward, although of course at the price of more complicated notations. For the sake of clarity we shall here restrict to scalar-valued processes.

gene regulatory network. In a schematic representation of this network, TF regulation of genes is represented by directed links with a weight proportional to the strength of the regulation on each gene; possible mutual interactions among genes are also accounted for [36], [47]. Usually, measurements of the activity of a large number of genes can be collected, but no information is available about their regulators (i.e., the TFs). Hence, retrieving the TF activity from gene expression data is a challenging problem in gene regulatory networks.

Referring to the model (2), we may use the vector  $\mathbf{y}(t)$  to represent the measured expression profile of the genes of the network in the *t*-th experiment. Describe the TF activity by the variable  $\mathbf{x}(t)$ , the strength of the effect of the TFs on the *i*-th gene of the network by the loading vector  $f_i$ 's and the gene mutual influences by the random vector  $\tilde{\mathbf{y}}(t)$ , we obtain a description of the network as a "flocking component plus local interactions" model. Estimating the flocking component  $\mathbf{x}(t)$ due to the action of the TFs may constitute a preliminary step towards understanding the TF activity in the network.

3) Modeling Energy Consumption: In this example, we may want to model the energy consumption (or production) of a network of N users distributed geographically in a certain area, say a city or a region. The energy consumption  $\mathbf{y}_i(t)$  of user i is a random variable which can be seen as the sum of the two contributions in (2), where the term  $f_i^{\dagger} \mathbf{x}(t)$ , the flocking component of the model, represents a linear combination of qhidden variables  $\mathbf{x}_i(t)$  which model different factors affecting the energy consumption (or production) of the whole ensemble; say heating or air conditioning consumption related to seasonal climatic variations, energy production related to the current status of the economy etc. The factor vector  $\mathbf{x}(t)$  determines the average time pattern of energy consumption/production of each unit, the importance of each scalar factor being determined by a q-ple of constant weight coefficients  $f_{i,k}$ . The terms  $\tilde{\mathbf{y}}_i(t)$ , represent local random fluctuations which model the consumption due to appliances or devices that are usually activated randomly, for short periods of time. They are assumed uncorrelated with the process x. The covariance  $\mathbb{E}\tilde{\mathbf{y}}_i(t)\tilde{\mathbf{y}}_i(t)$  could be non-zero for neighboring users but is reasonable to expect that it decays to zero when |i - j| is large.

To identify such a model one should start from real data of energy consumption collected from a large amount of units. A possible application for such a model is the forecasting of the average requirement of energy in a certain geographical area.

4) Dynamic Modeling in Computer Vision: Largedimensional time series occur often in signal processing applications, typically 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 survey [13] for more details and references. One starts from a signal  $\mathbf{y}(t) := \operatorname{vec}(\mathbf{I}(\cdot, t))$ , obtained by vectorizing at each time t, the intensities  $\mathbf{I}(\cdot, t)$  at each pixel of an image, into a vector, say  $\mathbf{y}(t) \in \mathbb{R}^N$ , with a "large" number (typically tens of thousands) of components. We may for instance be interested in modeling (and in identification methodologies thereof) of "dynamic textures" (see [20]), by linear state space models or in extracting classes of models describing rigid motions of objects of a scene. Most of these models involve hidden variables, say the state of linear models of textures, or the displacement-angular velocity coordinates of the rigid motions of objects in the scene. The purpose is of course to compress high dimensional data into simple mathematical structures. Note that the number of samples that can be used for identification is very often of the same order (and sometimes smaller) than the data dimensionality. For instance, in dynamic textures modeling, the number of images in the sequences is of the order of a few hundreds while N (which is equal to the number of pixels of the image) is certainly of the order of a few hundreds or thousands [6], [16].

#### C. Structure of the Paper

The organization of the paper is as follows: In Section II we quickly review static finite-dimensional Factor Analysis; in Section III following the basic definition of [23], we provide a neat mathematical characterization of idiosyncratic sequences (Theorem 3.1) which is believed to be new. In the following Section IV, based on the characterization of idiosyncratic sequences, the notion of strong linear independence is introduced and shown to be a crucial ingredient to provide a unique representations by GFA models. New results characterizing the factor loadings are presented in this section and related to the condition of diverging eigenvalues from the literature on GFA. The problem of representation by GFA models is discussed in Section V. Here the proposed notion of q-aggregate sequence from [23] is made concretely operational and a procedure to compute asymptotically the factor and the idiosyncratic components is presented for the first time. The literature on GFA does not seem to distinguish between GFA representations of a covariance matrix and of a random sequence. However while the first may be unique there may be quite diverse random components  $\hat{\mathbf{y}}, \tilde{\mathbf{y}}$  yielding a GFA representation of the same string y. To guarantee uniqueness one needs for example to impose that  $\hat{\mathbf{y}}, \tilde{\mathbf{y}}$  have components in  $H(\mathbf{y})$ . An interpretations of the two GFA components in terms of short and long range interaction of a large ensemble of stochastic agents is discussed in Section V-A. Also, new necessary and sufficient conditions for a (weakly) stationary sequence to have a GFA representation are presented at the end of the section in V-B. In Section VI, which is believed to be completely original, dynamic GFA representations of two classes of random fields of interest in applications are discussed. Time-stationary random fields provide in particular a wide class of linear stochastic models which can describe complex systems arising in a variety of applications in the system and control area. The extraction of the flocking component for these systems reduces to the study of an infinite dimensional Lyapunov equation.

Some of the material of this paper has been presented in a preliminary form at conferences [7], [8].

# II. A SHORT REVIEW OF STATIC FACTOR ANALYSIS MODELS

Factor Analysis (FA) has a long history; it has apparently first been introduced by psychologists and successively been studied and applied in various branches of Statistics and Econometrics [5], [32]–[34]. Dynamic versions of factor models have also been introduced in the econometric literature, see e.g., [24], [28], [43] and references therein. With a few exceptions however, [30], [31], [39], [40], [45], [46], little attention has been paid to these models in the system and control engineering community. Recently, we have been witnessing a revival of interest in FA, due to the generalization proposed by Chamberlain, Rothschild, Forni, Lippi and collaborators in a series of widely quoted papers [10], [11], [22], [23]. This new modeling paradigm is attracting a considerable attention also in the system identification community [3], [16]–[18], [44]. The new models, called *Generalized Factor Analysis (GFA)* models, although initially motivated by financial econometrics seem to have a potential to be useful also in engineering applications.

A classical (static) *Factor Analysis* model is a representation of N observable random variables  $\mathbf{y} = [\mathbf{y}(1) \dots \mathbf{y}(N)]^{\top}$ , as linear combinations of *q* common factors  $\mathbf{x} = [\mathbf{x}_1 \dots \mathbf{x}_q]^{\top}$ , plus uncorrelated "noise" or "error" terms  $\mathbf{e} = [\mathbf{e}(1) \dots \mathbf{e}(N)]^{\top}$  of the type

$$\mathbf{y} = F\mathbf{x} + \mathbf{e}.\tag{3}$$

The columns  $\{f_1, f_2, \ldots, f_q\}$  of the matrix F, called the *factor loadings* can be chosen linearly independent and the common factors can be normalized in such a way that  $\mathbb{E}\mathbf{x}\mathbf{x}^{\top} = I$ , which we shall always assume in the following. An essential part of the model specification is that the N components of the error  $\mathbf{e}$  should be (zero-mean and) mutually uncorrelated random variables, i.e.,

$$\mathbb{E}\mathbf{x}\mathbf{e}^{\top} = 0, \qquad \mathbb{E}\mathbf{e}\mathbf{e}^{\top} = \operatorname{diag}\left\{\sigma_{1}^{2}, \dots, \sigma_{N}^{2}\right\}.$$
(4)

The aim of these models is to provide an "explanation" of the mutual correlations of the observable variables  $\mathbf{y}(i)$  in terms of a small number of common factors, in the sense that, setting:  $\hat{\mathbf{y}}(k) := \sum f_i(k)\mathbf{x}_i$ , where  $f_i(k)$  is the k-th component of  $f_i$ , one has exactly  $\mathbb{E}\mathbf{y}(i)\mathbf{y}(j) = \mathbb{E}\hat{\mathbf{y}}(i)\hat{\mathbf{y}}(j)$ , for all  $i \neq j$ . Note that a FA representation then induces a decomposition of the covariance matrix  $\Sigma$  of  $\mathbf{y}$  as

$$\Sigma = FF^{\top} + \operatorname{diag}\left\{\sigma_{\mathbf{e}_{1}}^{2}, \dots, \sigma_{\mathbf{e}_{N}}^{2}\right\}$$
(5)

which can be seen as a special kind of *low rank plus sparse* decomposition of a covariance matrix [12], [50], a diagonal matrix being, in intuitive terms, as sparse as one could possibly ask for.

Unfortunately, these models, although providing in many circumstances a quite natural and useful data compression scheme, suffer from a serious non-uniqueness problem coming from the fact that, even for a fixed dimension q there are in general many (generally infinitely many) statistically non-equivalent FA models describing the same family of observables  $\{\mathbf{y}(1), \ldots, \mathbf{y}(N)\}$ . In addition, determining the minimal integer q for which a FA decomposition holds for a given symmetric positive definite matrix  $\Sigma$  has been an open problem since the beginning of the last century. It is by now a well-known fact that estimation of FA models (say minimal models with F's of rank q and normalized factors) is an ill-posed problem.

This inherent non-uniqueness is called "factor indeterminacy," or unindentifiability in the literature and the often acritical use of FA models has been vehemently criticized by Kalman in a series of papers, see e.g., [30], [31]. Non-uniqueness of the factors is an intrinsic difficulty common to stochastic models with latent variables, whose role is to enforce some conditional independence relation among the observables. As a rule the choice of the latent variables is non-unique. It is known, see [4], [37], [45], that a factor structure is also equivalent to a relation of conditional independence of the observables given the factors and this is in turn *equivalent to the uncorrelation of the noise components*.

One may then try to obtain uniqueness by giving up or by mitigating the requirement of uncorrelation of the components of e. Obviously this tends to make the problem ill-defined as the basic goal of uniquely splitting the external observable 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." Quite surprisingly, as we shall see, for models describing an *infinite* number of observables a meaningful weakening of the uncorrelation property can be introduced, so as to guarantee the uniqueness of the decomposition.

# III. STATIC GENERALIZED FACTOR ANALYSIS AND IDIOSYNCRATIC SEQUENCES

In this section, we shall review Generalized Factor Analysis restricting for now to the *static case*.

Consider a zero-mean finite variance stochastic process  $\mathbf{y} := {\mathbf{y}(k), k \in \mathbb{Z}_+}$  represented as a column vector with an infinite number of random components. We want to represent  $\mathbf{y}$  by an infinite dimensional FA model of the form

$$\mathbf{y}(k) = \sum_{i=1}^{q} f_i(k)\mathbf{x}_i + \tilde{\mathbf{y}}(k), \qquad k = 1, 2, \dots$$
(6)

where, in analogy to finite-dimensional FA, the random variables  $\mathbf{x}_i$ , i = 1, ..., q are the common factors and the deterministic vectors  $f_i \in \mathbb{R}^{\infty}$  the factor loadings. The  $\mathbf{x}_i$  form a q-dimensional random vector  $\mathbf{x}$  with orthonormal components; i.e.,  $\mathbb{E}\mathbf{x}\mathbf{x}^{\top} = I_q$ . The  $\tilde{\mathbf{y}}(k)$ 's are zero mean random variables orthogonal to (uncorrelated with)  $\mathbf{x}$ . The linear combinations  $\hat{\mathbf{y}}(k) := \sum f_i(k)\mathbf{x}_i; k = 1, 2, ...$  are the components of an infinite random vector denoted  $\hat{\mathbf{y}}$  which, together with the noise terms  $\tilde{\mathbf{y}}(k)$ , give the representation  $\mathbf{y} = \hat{\mathbf{y}} + \tilde{\mathbf{y}}$ , a compact way to write the model (6) in vector form.

Which specific characteristics qualify the process  $\tilde{y}$  as "noise" is a nontrivial issue which will be the main theme of this section and will lead to the concept of *idiosyncratic random sequence* below. The underlying idea can be extracted from the following example.

*Example 3.1:* Let  $\mathbb{1}$  be an infinite column vector of 1's, let x be a zero-mean scalar random variable and  $\tilde{y}$  a zero-mean weakly stationary ergodic sequence uncorrelated with x. Consider the FA model

and the sequence of vectors in  $\mathbb{R}^{\infty}$ 

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

Since  $\lim_{n\to\infty} a_n^{\top} \mathbb{1} = 1$  and  $\lim_{n\to\infty} a_n^{\top} \tilde{\mathbf{y}} = \lim_{n\to\infty} (1/n) \sum_{k=1}^n \tilde{\mathbf{y}}(k) = \mathbb{E} \tilde{\mathbf{y}}(k) = 0$  (limit in  $L^2$ ) we have  $\lim_{n\to\infty} a_n^{\top} \mathbf{y} = \mathbf{x}$ ; hence we can recover the latent factor by an operation of averaging. There are in fact more general sequences  $a_n$  of infinite vectors, such that  $\lim_{n\to\infty} a_n^{\top} \mathbb{1}$  exists and is nonzero and in addition  $\lim_{n\to\infty} a_n^{\top} \tilde{\mathbf{y}} = 0$  for processes like  $\tilde{\mathbf{y}}$ . These sequences recover  $\mathbf{x}$  from the FA model uniquely.

The infinite covariance matrix of the vector **y** is formally written as  $\Sigma := \mathbb{E}\mathbf{y}\mathbf{y}^{\top}$ . Let  $\ell^2(\Sigma)$  denote the Hilbert space of infinite sequences  $a := \{a(k), k \in \mathbb{Z}_+\}$  such that  $||a||_{\Sigma}^2 := a^{\top}\Sigma a < \infty$ . When  $\Sigma = I$ , we use the standard symbol  $\ell^2$ , denoting the corresponding norm by  $|| \cdot ||_2$ . The following definition was introduced in [23]

Definition 3.1: A sequence  $\{a_n\}_{n\in\mathbb{Z}_+}\subset \ell^2\cap\ell^2(\Sigma)$  is an averaging sequence (AS) for y, if  $\lim_{n\to\infty} ||a_n||_2 = 0$ .

We say that a sequence of random variables  $\mathbf{y}$  is **idiosyncratic** if  $\lim_{n\to\infty} a_n^\top \mathbf{y} = 0$  for any averaging sequence  $a_n \in \ell^2 \cap \ell^2(\Sigma)$ .

Whenever the covariance  $\Sigma$  is a bounded operator on  $\ell^2$ one has  $\ell^2(\Sigma) \subset \ell^2$ ; in this case an AS can be seen just as a sequence of linear functionals in  $\ell^2$  converging strongly to zero.

The elements of the sequence (7) together with their limit are clearly in  $\ell^2$  and form an averaging sequence. For a more general class of AS's, let  $P_n$  denote the compression of the *n*-th power of the left shift operator to the space  $\ell^2$ ; i.e.,  $[P_n a](k) =$ a(k-n) for  $k \ge n$  and zero otherwise. Then  $\lim_{n\to\infty} P_n a = 0$ for all  $a \in \ell^2$  [25] so that  $\{P_n a\}_{n \in \mathbb{Z}_+}$  is an AS for any  $a \in \ell^2$ .

The nature of an idiosyncratic sequence is related to certain properties of its covariance matrix. To explain this point, we need to introduce some notations and facts about the eigenvalues of sequences of covariance matrices. We let  $\Sigma_n$  indicate the top-left  $n \times n$  block of  $\Sigma$ , 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  $\Sigma_n$  of  $\Sigma$  are positive definite, which we shall always assume in the following. Letting  $\hat{\Sigma} := \mathbb{E}\hat{\mathbf{y}}\hat{\mathbf{y}}^{\top} = FF^{\top}$ and  $\tilde{\Sigma} := \mathbb{E}\tilde{\mathbf{y}}\tilde{\mathbf{y}}^{\top}$ , the orthogonality of the noise term and the factor components implies that

$$\Sigma = \hat{\Sigma} + \tilde{\Sigma} \tag{8}$$

that is,  $\Sigma_n = \hat{\Sigma}_n + \tilde{\Sigma}_n$ ,  $\forall n \in \mathbb{Z}_+$ . Even imposing that  $\hat{\Sigma}$  should be of low rank, this is *a priori* a highly non-unique decomposition. There are situations/examples in which  $\tilde{\Sigma}$  is diagonal as in the finite-dimensional FA case, but these situations are exceptional. Denote by  $\lambda_k(\Sigma_n)$  the k-th eigenvalue of the  $n \times n$  upper left submatrix  $\Sigma_n$  of  $\Sigma$ . The  $\lambda_k(\Sigma_n)$ 's are real nonnegative and will always be ordered by decreasing magnitude. By Weyl's theorem [52, p. 203], see also [23, Fact M], the k-th eigenvalue of  $\Sigma_n$  is a non decreasing function of nand hence has a limit,  $\lambda_k(\Sigma)$ , which may possibly be equal to  $+\infty$ . Each such limit is called an *eigenvalue of*  $\Sigma$ . These limits however are in general not true eigenvalues, as it is well-known that  $\Sigma$  may not have eigenvalues. For example, a bounded symmetric Toeplitz matrix has a purely continuous spectrum [26]. Anyway since  $\Sigma$  is symmetric and positive, its spectrum lies on the positive half line and its elements can be ordered. Henceforth we shall denote by  $\lambda_1(\Sigma)$  the maximal eigenvalue of  $\Sigma$ , as defined above, with the convention that  $\lambda_1(\Sigma) = +\infty$ when there are infinite eigenvalues. The following Lemma will be instrumental in understanding the structure of idiosyncratic processes.

*Lemma 3.1:* A symmetric matrix  $\Sigma$  defines a bounded operator on  $\ell^2$  if and only if  $\lambda_1(\Sigma)$  is finite.

Proof: See the Appendix.

A characterization of idiosyncratic sequences is stated in the following theorem. The proof will also be given in the Appendix.

*Theorem 3.1:* The sequence y is idiosyncratic if and only if its covariance matrix defines a bounded operator on  $\ell^2$ .

In particular, a white noise process with uniformly bounded variance is idiosyncratic. This follows since the covariance of a white noise process is a diagonal matrix with uniformly bounded entries and therefore is a bounded operator. However the notion of idiosyncratic process is much more general than that of white noise. For example any (weakly) stationary purely non deterministic process with a bounded spectral density is idiosyncratic. See Section V-B.

A test for idiosyncracy of a random sequence can be based on Lemma 3.1, whereby y is idiosyncratic if and only if  $\lambda_1(\Sigma)$  is finite; this last characterization is due to [23] (where however the characterization in terms of boundedness of  $\Sigma$  was not noticed).

#### **IV. PURELY DETERMINISTIC SEQUENCES**

The notion of a (discrete-time) *purely deterministic* random sequence, or process, is well-known, see e.g., [14] which originated the terminology for not necessarily stationary processes.

*Definition 4.1:* Let q be a finite natural number. A sequence y is *purely deterministic of rank* q (in short q-PD) if  $H(\mathbf{y})$  has dimension q.

Clearly a q-PD sequence y is a (in general non-stationary) purely deterministic process in the classical sense of the term. More specifically, y is a q-PD random sequence if and only if there are q deterministic infinite column vectors  $f_1, f_2, \ldots f_q$ or, for short, an  $\infty \times q$  matrix  $F = [f_1 \ f_2 \ \ldots \ f_q]$ , such that

$$\mathbf{y}(k) = \sum_{i=1}^{q} f_i(k) \mathbf{x}_i, \quad k \in \mathbb{Z}_+, \quad \text{or} \quad \mathbf{y} = F \mathbf{x}$$
(9)

for some random variables  $\mathbf{x}_1, \ldots, \mathbf{x}_q$ . Without loss of generality, the columns  $f_1, f_2, \ldots, f_q$  can be assumed to be linearly independent, for otherwise one of them could be expressed as a linear combination of the others and eliminated. In this case  $\{\mathbf{x}_1, \ldots, \mathbf{x}_q\}$  can be taken to be an orthonormal basis in  $H(\mathbf{y})$ .

We want to relate this concept to the factor component of y, as defined earlier. The factors will later on be shown to originate the flocking component of y(t) in the time varying case. As it stands the *q*-PD condition is however insufficient to guarantee

 $\square$ 

uniqueness. Unfortunately it turns out that there are nontrivial sequences representable in the form (9) which are idiosyncratic (or contain idiosyncratic sequences). See the example below.

*Example 4.1:* Consider a sequence y whose k-th element is

$$\mathbf{y}(k) = \lambda^k \mathbf{x} \quad , |\lambda| < 1 \tag{10}$$

where x is a zero-mean random variable of positive variance  $\sigma^2$ . Clearly, y is 1-PD, its spanned subspace H(y) being the one-dimensional space H(x). The covariance matrix of the first n components of y is

$$\Sigma_n = \mathbb{E} \mathbf{y}_n \mathbf{y}_n^{\top} = \sigma^2 \begin{bmatrix} \lambda^2 & \lambda^3 & \dots & \lambda^{n+1} \\ \lambda^3 & \lambda^4 & \dots & \lambda^{n+2} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda^{n+1} & \lambda^{n+2} & \dots & \lambda^{2n} \end{bmatrix}.$$
 (11)

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

$$\lambda_1(\Sigma) = \lim_{n \to \infty} \operatorname{tr}(\Sigma_n) = \lim_{n \to \infty} \sigma^2 \sum_{k=1}^n \lambda^{2k} = \frac{\sigma^2 \lambda^2}{1 - \lambda^2} \quad (12)$$

thus, in force of Theorem 3.1, y is an idiosyncratic sequence. Hence there are (non-stationary) q-PD sequences which are idiosyncratic.

This is a possibility which must clearly be excluded if the decomposition (6) has to be unique.<sup>2</sup> The question is which properties need to be satisfied by the functions  $f_1, f_2, \ldots f_q$  for y not to be an idiosyncratic sequence. One necessary condition is easily found: the  $f_i$  cannot be in  $\ell^2$  since otherwise any sequence of functionals  $\{a_n\}$  in  $\ell^2$  converging to zero would lead to

$$\lim_{n \to \infty} a_n^{\top} f_i = 0 \tag{13}$$

so that  $\lim_{n\to\infty} a_n^\top y = 0$  as well. This is clearly the problem with Example 4.1.

Proposition 4.1: If y is q-PD sequence with a uniformly bounded variance, then the  $f_i$ 's are uniformly bounded sequences; i.e., belong to the space  $\ell^{\infty}$ . If in addition y is not idiosyncratic the  $f_i$ 's belong to  $\ell^{\infty}$  but cannot belong to  $\ell^2$ .

*Proof:* The statement follows since  $\|\mathbf{y}(k)\|^2 \leq M^2$ , which is the same as  $\sum_{i=1}^q f_i(k)^2 \leq M^2$  implies that  $|f_i(k)| \leq M$  for all k's.

We now discuss conditions in terms of the covariance matrix. Definition 4.2: A q-PD sequence y is q-aggregate if  $\operatorname{rank}(\Sigma) = q$  and  $\lim_{n\to\infty} \lambda_k(\Sigma_n) = +\infty$  for  $k = 1, \ldots, q$ . In short, there are only q nonzero eigenvalues of  $\Sigma$  which are all infinite.

For q = 1 this condition just means that the (only) column of F has  $\ell^2$ -norm equal to infinity.

*Proposition 4.2:* A *q*-aggregate sequence y can be idiosyncratic only if it is the zero sequence.

*Proof:* This follows trivially from Theorem 3.1. If q > 0 the maximal eigenvalue of the covariance matrix of y is  $+\infty$  by definition.

 $^{2}$ Note for example that Definition 2 in [11, p. 1294] is not enough to guarantee uniqueness.

Hence, the condition guarantees some sort of uniqueness of the decomposition (6). Of course the question is under what conditions the q eigenvalues of  $\hat{\Sigma}$  may tend to infinity. The notion of strong linear independence introduced below provides an answer.

Definition 4.3: Let

$$\tilde{f}_i^n := f_i^n - \Pi\left[f_i^n | \mathcal{F}_i^n\right] \tag{14}$$

where  $\Pi$  is the orthogonal projection onto the Euclidean space  $\mathcal{F}_i^n = \text{span} \{f_j^n, j \neq i\}$  of dimension q - 1. The vectors  $f_i, i = 1, \dots, q$  in  $\mathbb{R}^\infty$  are **strongly linearly independent** if

$$\lim_{n \to \infty} \left\| \tilde{f}_i^n \right\|_2 = +\infty \qquad i = 1, \dots, q.$$
 (15)

In a sense, the tails of two strongly linearly independent vectors in  $\mathbb{R}^{\infty}$  cannot get "too close" asymptotically.

*Theorem 4.1:* Let y be a q-PD sequence, i.e., let

$$\mathbf{y}(k) = \sum_{i=1}^{q} f_i(k) \mathbf{x}_i, \qquad k \in \mathbb{Z}_+;$$
(16)

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

The proof is given in Appendix C.

*Example 4.2:* Consider the 2–PD sequence  $\mathbf{y}(k) := \sum_{i=1}^{2} f_i(k) \mathbf{x}_i$ , with

$$f_1(k) = 1$$
 for all  $k$ ,  $f_2(k) = 1 - \left(\frac{1}{2}\right)^k$ .

It is not difficult to check that this sequence does not satisfy condition (15). We shall show that this sequence is not 2-aggregate. The Gramian matrix of the functions  $f_1, f_2$  restricted to [1, n] is

$$F^{n^{\top}}F^{n} = \begin{bmatrix} \|f_{1}^{n}\|_{2}^{2} & \langle f_{1}^{n}, f_{2}^{n} \rangle_{2} \\ \langle f_{1}^{n}, f_{2}^{n} \rangle_{2} & \|f_{2}^{n}\|_{2}^{2} \end{bmatrix}$$

and it can be seen that as  $n \to \infty$ , the second eigenvalue converges to (5/3). Hence one eigenvalue of the covariance matrix of y is finite and the sequence is not 2-aggregate.  $\Box$ 

## V. GFA Representations: Existence and Uniqueness

Summing up Theorem 3.1, Theorem 4.1 and the uniqueness result in Proposition 4.2 we obtain conditions on the covariance  $\Sigma = \mathbb{E} \mathbf{y} \mathbf{y}^{\top}$  to describe processes admitting a GFA representation.

Definition 5.1: The covariance  $\Sigma$  has a *GFA decomposition* of rank q if it can be decomposed as the sum of a matrix  $\tilde{\Sigma}$ which is a bounded operator in  $\ell^2$ , and a rank q perturbation  $\hat{\Sigma} = FF^{\top}$ , namely

$$\Sigma = FF^{\top} + \tilde{\Sigma}, \quad \text{with } F = [f_1 \quad \dots \quad f_q], f_i \in \mathbb{R}^{\infty} \quad (17)$$

where  $F \in \mathbb{R}^{\infty \times q}$  has strongly linearly independent columns.

Theorem 5.1: The infinite covariance matrix  $\Sigma$  has a GFA decomposition of rank q if and only if for  $n \to \infty$ ,  $\Sigma_n$  has q unbounded eigenvalues and  $\lambda_{q+1}(\Sigma_n)$  stays bounded as  $n \to \infty$ . A GFA decomposition of  $\Sigma$  is unique, modulo right multiplication of F by a  $q \times q$  orthogonal matrix.

This result is close to Chamberlain and Rothschild [11, Theorem 4], where it is obtained via a quite different and rather lengthy series of arguments.

Note that there may well be sequences (of positive symmetric)  $\Sigma_n$  for which *all eigenvalues* tend to infinity. In this case there is no GFA decomposition. When it applies, the criterion can be seen as a limit of the well-known rule of separating "large" from "small" eigenvalues in Principal Components Analysis (PCA). Let  $f_i^n \in \mathbb{R}^n$ ;  $i = 1, \ldots, q$  be the eigenvectors corresponding to the q (ordered) eigenvalues of  $\Sigma_n$  which increase without bound when  $n \to \infty$ . We normalize these eigenvectors in such a way that  $F_n := [f_1^n \ldots f_q^n]$  yields  $\hat{\Sigma}_n = F_n F_n^\top$ . Then

$$\lim_{n \to \infty} F_n F_n^{\top} = F F^{\top}.$$
 (18)

See [11, Theorem 4 (ii), p. 1299] for a proof and a discussion. Then convergence of  $\{F_n\}$  can be interpreted as column space convergence in the gap metric, see [52, p. 260]. Although the usual orthogonality of the  $f_i^n$  in PCA does not make sense in infinite dimensions as the limit eigenvectors do not belong to  $\ell^2$ , one may however interpret the strong linear independence condition as a limit of the orthogonality holding for finite n. Hence, we can (asymptotically) get q and F by a limit PCA procedure on the sequence  $\Sigma_n$ .

Trivially, if a random sequence y admits a GFA representation then its covariance matrix has a GFA decomposition. On the other hand, assume we are given a GFA decomposition  $\hat{\Sigma} + \tilde{\Sigma}$  of an infinite covariance  $\Sigma$ . How do we find the hidden variables in the representation  $\mathbf{y} = F\mathbf{x} + \tilde{\mathbf{y}}$ ?

This question has also to do with uniqueness of the representation as there may be several non-equivalent choices of x and  $\tilde{y}$  compatible with a GFA decomposition of  $\Sigma$ . We shall show that there is an essentially unique choice, under the constraint that both x and  $\tilde{y}$  belong to H(y). Models of this kind are called *internal* in stochastic realization. The following definition from [23] is meant to generalize the phenomenon described in Example 3.1.

Definition 5.2: 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  $\lim_{n\to\infty} a_n^\top \mathbf{y} = \mathbf{z}$ . The set of all aggregate random variables in  $H(\mathbf{y})$  is a closed subspace denoted by  $\mathcal{F}(\mathbf{y})$  called the *aggregation subspace* of  $H(\mathbf{y})$ .

Clearly, if y is an idiosyncratic sequence then  $\mathcal{F}(\mathbf{y}) = \{0\}$ . One can then define an orthogonal decomposition of the type

$$\mathbf{y} = \mathbb{E}\left[\mathbf{y}|\mathcal{F}(\mathbf{y})\right] + \mathbf{u} \tag{19}$$

where  $\mathbb{E}[\cdot|\mathcal{F}(\mathbf{y})]$  is the orthogonal projection operator onto the subspace  $\mathcal{F}(\mathbf{y})$ , so that all components  $\mathbf{u}(k)$  are uncorrelated with  $\mathcal{F}(\mathbf{y})$ . The idea behind this decomposition is that, in case  $\mathcal{F}(\mathbf{y})$  is finite dimensional, say generated by a *q*-dimensional random vector  $\mathbf{x}$ , one may naturally identify  $\mathbf{u}$  as the idiosyncratic component and capture a unique decomposition of  $\mathbf{y}$  of

the type (6). This intention is probably behind the analogous decomposition in [23] but this idea cannot be pursued further unless some further technical requirements are imposed, which are so far unknown. There may be pathological situations in which  $\mathcal{F}(\mathbf{y})$  is finite dimensional, or in which  $\mathcal{F}(\mathbf{y}) = \{0\}$ , but the process  $\mathbf{u}$  is not idiosyncratic. Theorem 5.2 below asserts that in the special case of stationary sequences, the construction works if and only if the spectral density of  $\mathbf{u}$  is in  $L^{\infty}$ .

**Proposition 5.1:** Assume that its covariance matrix  $\Sigma$  has a GFA decomposition of rank q. Then  $\mathbf{y}$  has a GFA representation with q factors where both  $\mathbf{x}$  and  $\tilde{\mathbf{y}}$  have components in  $H(\mathbf{y})$ .

The Proof is in Appendix D.

# A. Interpretation: Short and Long Distance Interaction

Imagine a scenario of an ensemble of infinitely many agents distributed in space interacting randomly, producing as output measurements the random variables  $\mathbf{y}(k) = \hat{\mathbf{y}}(k) + \tilde{\mathbf{y}}(k); k = 1, 2, \dots$ 

The covariances  $\tilde{\sigma}(k, j) = \mathbb{E}\tilde{\mathbf{y}}(k)\tilde{\mathbf{y}}(j)$  measure the mutual correlation of the idiosyncratic fluctuations of neighboring agents  $\tilde{\mathbf{y}}(k), \tilde{\mathbf{y}}(j)$  located in positions k and j. Since  $\tilde{\Sigma}$  is a bounded operator in  $\ell^2$ , it is a known fact [1, Section 26] that  $\tilde{\sigma}(k, j) \to 0$  as  $|k - j| \to \infty$  so, in a sense the idiosyncratic component  $\tilde{\mathbf{y}}$  of a GFA representation models only *short range* interaction among the agents, as  $\tilde{\sigma}(k, j)$  is decaying to zero when the distance |k - j| tends to infinity.

Whenever an ensemble can be described by an idiosyncratic sequence, the agents which are far away from each other essentially do not resent of mutual influence. The statement holds in general, for every GFA model, although the decay of the elements  $\tilde{\sigma}(k, j)$  may be faster depending on the particular covariance structure. Just the opposite will be true for the sequence  $\hat{\mathbf{y}}$ .

On the other hand,  $\mathbb{E}\hat{\mathbf{y}}(k)\hat{\mathbf{y}}(j) = \sum_i f_i(k)f_i(j)$  and the elements of the column vectors  $f_i$  cannot be in  $\ell^2$ . In particular, as stated in Proposition 4.1,  $f_i \in \ell^\infty$  when the variances of the random variables  $\mathbf{y}(k)$  are uniformly bounded.

In any case, since the components  $f_i(k)$  do not decay with distance, the products  $f_i(k)f_i(j)$  generically cannot vanish when  $|k - j| \rightarrow \infty$ . Therefore the factor loadings describe "long range" correlation and the aggregate component  $\hat{\mathbf{y}}$  of  $\mathbf{y}$ can be interpreted as variables modeling *long range interaction* among the agents. In this sense  $\hat{\mathbf{y}}$  models an average *collective behavior* of the ensemble. This is in fact the core of the flocking structure that will emerge as soon as the  $\mathbf{x}_i$  are allowed to depend on time.

#### B. The Case of Stationary Sequences

The characterizations of GFA models discussed so far are for general second order sequences, that is for processes  $\mathbf{y}$  which may well be non-stationarity with respect to the cross sectional (space) index k. Much sharper results hold in the special case in which the sequence  $\mathbf{y}$  is (weakly) stationary; i.e.,  $\mathbb{E}\mathbf{y}(k)\mathbf{y}(j) = \sigma(k-j)$  for  $k, j \ge 0$ . A complete analysis of this case cannot be presented here and can be found in [7]. Here we shall just report the main result.

Let  $H_k(\mathbf{y})$  be the closed linear span of all random variables  $\{\mathbf{y}(s); s \ge k\}$ . Introducing the *remote future subspace* of  $\mathbf{y}$ :

$$H_{\infty}(\mathbf{y}) = \bigcap_{k \ge 0} H_k(\mathbf{y}) \tag{20}$$

the sequence of orthogonal wandering subspaces  $E_k := H_k(\mathbf{y}) \ominus H_{k+1}(\mathbf{y})$  and their orthogonal direct sum  $\check{H}(\mathbf{y}) = \bigoplus_{k \ge 0} E_k$ , it is well known, see, e.g., [19], [25], [49], that one has the orthogonal decomposition

$$\mathbf{y} = \hat{\mathbf{y}} + \check{\mathbf{y}}, \qquad \hat{\mathbf{y}}(k) \in H_{\infty}(\mathbf{y}), \quad \check{\mathbf{y}}(k) \in \check{H}(\mathbf{y})$$
(21)

for all  $k \in \mathbb{Z}_+$ , the component  $\hat{\mathbf{y}}$  being the purely deterministic (PD), while  $\check{\mathbf{y}}$  the purely non-deterministic (PND) components. The two sequences are orthogonal and uniquely determined. Furthermore, it is well known that  $\check{\mathbf{y}}$  has an absolutely continuous spectrum with a spectral density function, say  $S_y(\omega)$  satisfying the log-integrability condition  $\int \log S_y(\omega) d\omega > -\infty$ , while the spectral distribution of  $\hat{\mathbf{y}}$  is singular with respect to Lebesgue measure (for example consisting only of jumps) possibly together with a spectral density such that  $\int \log S_y(\omega) d\omega = -\infty$ , compare, e.g., [49].

Theorem 5.2: Assume that y is a stationary sequence with  $\dim H_{\infty}(\mathbf{y}) < \infty$  and an a.e. bounded spectral density. Then  $H_{\infty}(\mathbf{y}) \equiv \mathcal{F}(\mathbf{y})$ .

A stationary sequence admits a unique internal GFA representation (6) with q factors if and only if it has a bounded spectral density and the remote future space is of dimension q. The aggregate component  $\hat{\mathbf{y}}$  is the purely deterministic component of y while the idiosyncratic  $\tilde{\mathbf{y}}$  is the purely nondeterministic component.

Note that there are stationary processes with a finite dimensional remote future space, whose PND component has an unbounded spectral density. It follows from Szegö's theorem [49] that  $\tilde{\Sigma}$  is an unbounded operator and these processes are neither aggregate nor idiosyncratic.

In the papers [11], [23], stationarity with respect to the crosssectional index is not assumed. However, without stationarity, there may be random sequences which fail to satisfy the eigenvalue conditions of Theorem 5.1 and do not admit a generalized factor analysis representation. A precise characterization of which class of non-stationary sequences admits a GFA representation seems still to be an open problem.

# VI. DYNAMIC GFA MODELS

We come back to dynamic modeling and to the question raised in Section I-A namely when does a second order random field have a flocking component. We shall initially restrict to the case of processes which are stationary with respect to the time variable which is a natural assumption to make in view of statistical inference.

A time-dependent family  $\mathbf{y} := {\mathbf{y}(t); t \in \mathbb{Z}}$  of infinitedimensional zero-mean random vectors,  $\mathbf{y}(t)$ , whose covariance matrix,  $\Sigma(\tau) := \mathbb{E}\mathbf{y}(t+\tau)\mathbf{y}(t)^{\top}$  is (finite and) independent of t, will be called a *time-stationary (second order) random field*. The following definition extends and makes precise the finite-dimensional concepts introduced at the end of Section I. Definition 6.1: We shall say that a time-stationary random field has a *dynamic GFA representation* of rank q if it can be written as

$$\mathbf{y}(t) = F\mathbf{x}(t) + \tilde{\mathbf{y}}(t) \tag{22}$$

where  $F \in \mathbb{R}^{\infty \times q}$  has strongly linearly independent columns and  $\tilde{\mathbf{y}}(t)$  is an *idiosyncratic random field*; i.e the covariance matrix  $\tilde{\Sigma}(\tau) := \mathbb{E}\tilde{\mathbf{y}}(t+\tau)\tilde{\mathbf{y}}(t)^{\top}$  defines, for all  $\tau \in \mathbb{Z}$ , a bounded linear operator in  $\ell^2$ . The *q* dimensional *factor process*  $\mathbf{x}(t)$  and  $\tilde{\mathbf{y}}(t)$  are jointly stationary and uncorrelated, that is

$$\mathbb{E}\mathbf{x}_i(t)\tilde{\mathbf{y}}_j(t) = 0, \qquad i = 1, \dots, q; \ j = 1, 2, \dots, \quad t \in \mathbb{Z}.$$

Without loss of generality,  $\mathbf{x}(t)$  can be chosen with orthonormal components; i.e.,  $\mathbb{E}\mathbf{x}(t)\mathbf{x}(t)^{\top} = I_q$ .

Proposition 6.1: The stationary random field  $\mathbf{y} := {\mathbf{y}(t); t \in \mathbb{Z}}$  has a dynamic GFA representation (22) if and only if  $\mathbf{y}(0)$  has a static GFA representation with the same factor loading matrix  $F, \mathbf{x} \equiv \mathbf{x}(0)$  and  $\tilde{\mathbf{y}} \equiv \tilde{\mathbf{y}}(0)$ .

*Proof:* The proof of the direct implication is trivial. The converse is proven in Appendix E.  $\Box$ 

Incidentally, the proposition guarantees uniqueness of the dynamic representation (22). The following criterion for the existence of a flocking structure in a time-stationary random field follows directly from Theorem 5.1 and the proposition above.

Corollary 6.1: For a time-stationary random field, a flocking structure exists with q factors if and only if q eigenvalues of the steady state covariance matrix  $\Sigma_n$  of the *n*-dimensional random subvector  $\mathbf{y}^n(t)$  of  $\mathbf{y}(t)$ , tend to infinity with n while the others remain bounded.

We shall study a class of random fields described by linear evolution equations of the general form

$$\mathbf{y}(t+1) = A\mathbf{y}(t) + \mathbf{w}(t) \qquad t \in \mathbb{Z}$$
(23)

where  $\mathbf{w}$  is a string of uncorrelated stationary white noise processes and A is an infinite matrix (a linear operator) acting on infinite sequences. We assume that the evolution is stationary in time so that the steady state covariance matrix of  $\mathbf{y}(t)$  is a constant positive definite matrix  $\Sigma$ , which should satisfy an infinite dimensional Lyapunov equation

$$\Sigma = A\Sigma A^{\top} + Q \tag{24}$$

where Q is the variance matrix of the white noise which we assume an infinite diagonal matrix with uniformly bounded positive entries (it is actually no loss of generality assuming that Q is the identity matrix). In this case, a GFA model of y will also be stationary and the structure of the model can be inferred by analyzing the covariance matrix  $\Sigma$ . To this end consider the *n*-dimensional random sub-processes  $y^n(t)$  of y(t), obeying the equation

$$\mathbf{y}^{n}(t+1) = A_{n}\mathbf{y}^{n}(t) + \mathbf{w}^{n}(t), \quad n = 1, 2, \dots$$
 (25)

where  $A_n$  is the upper left  $n \times n$  submatrix of A and the input process  $\mathbf{w}^n(t)$  is the *n*-dimensional white noise with variance

 $\mathbb{E}\mathbf{w}^n(t)\mathbf{w}^n(s)^\top = Q_n \delta_{t,s}, Q_n$  being the upper left  $n \times n$  submatrix of Q, and study the behavior as  $n \to \infty$  of the covariance matrix of  $\mathbf{y}^n(t)$ , solution to the family of Lyapunov equations

$$\Sigma_n = A_n \Sigma_n A_n^\top + Q_n \qquad n = 1, 2, \dots$$
 (26)

The existence of a flocking component can be addressed by analyzing the asymptotic behavior of  $\Sigma_n$  when  $n \to \infty$ . Some families of matrices  $\{A_n\}_{n \in \mathbb{N}}$  are considered below.

## A. Autonomous Agents

In this scenario, the behavior of each agent is independent of the others, being just an autoregressive motion of the type

$$\mathbf{y}_k(t+1) = a_k \mathbf{y}_k(t) + \mathbf{w}_k(t), \qquad \sup_{k \in \mathbb{N}} |a_k| < 1.$$
(27)

In this case,  $A_n = \text{diag}\{a_1, \ldots, a_n\}$  and the family of Lyapunov (26) admits diagonal (nested) solutions with uniformly bounded elements. Hence, in this case the resulting sequence is idiosyncratic noise with uncorrelated components and there is no flocking structure.

#### B. Flocking by Following a Leader

As discussed in [51], flocking may be observed in hierarchical leadership models where the evolution of the first n agents influences that of the agents of index k > n but not conversely so the matrix of the operator A has a nested lower triangular structure of the type

$$A_{n+1} = \begin{bmatrix} A_n & 0\\ b_n^\top & a_{n+1} \end{bmatrix}$$
(28)

where  $|a_{n+1}| < 1$  to keep the asymptotic stability of  $A_n$  preserved.

A very simple instance is the following linear model where each agent, evolving with the same scalar random dynamics, wants to follow a leader by applying a proportional control law based on the measurement of its position with respect to the leader's  $y_1(t)$ 

$$\mathbf{y}_1(t+1) = a \mathbf{y}_1(t) + \mathbf{w}_1(t), \qquad |a| < 1 \mathbf{y}_k(t+1) = (1-a) \mathbf{y}_1(t) + a \mathbf{y}_k(t) + \mathbf{w}_k(t), \ k = 2, 3, \dots$$

The question is if following a leader should, under appropriate circumstances, produce a random flock. The steady-state co-variance matrices of  $y^n(t)$  solves the Lyapunov (26) for the model

$$\begin{bmatrix} \mathbf{y}_1(t+1) \\ \mathbf{y}_2(t+1) \\ \vdots \\ \mathbf{y}_n(t+1) \end{bmatrix} = \begin{bmatrix} a & 0 & \cdots & 0 \\ 1-a & a & \vdots \\ \vdots & 0 & \ddots & 0 \\ 1-a & \cdots & a \end{bmatrix} \begin{bmatrix} \mathbf{y}_1(t) \\ \mathbf{y}_2(t) \\ \vdots \\ \mathbf{y}_n(t) \end{bmatrix} + \begin{bmatrix} \mathbf{w}_1(t) \\ \mathbf{w}_2(t) \\ \vdots \\ \mathbf{w}_n(t) \end{bmatrix}$$

and it is possible to show that indeed a flocking structure is present.

Proposition 6.2: Assume for simplicity that  $Q_n = I_n$ . The solution of the Lyapunov (26) tends for  $n \to \infty$  to a covariance matrix of the form  $\Sigma = f f^\top + \tilde{\Sigma}$  where  $f \in \mathbb{R}^\infty$  has components

$$f_k = \begin{cases} \frac{a}{(1-a^4)^{\frac{1}{2}}}, & k = 1\\ \frac{(1+a^2)^{\frac{1}{2}}}{\left[(1+a)(1-a^2)^{\frac{1}{2}}\right]}, & k > 1 \end{cases}$$

and  $\tilde{\Sigma}$  is a bounded operator in  $\ell^2$ . Hence

$$\mathbf{y}(t) = f\mathbf{x}(t) + \tilde{\mathbf{y}}(t), \quad \mathbf{x}(t) = (1 - a^4)^{\frac{1}{2}} \mathbf{y}_1(t - 1),$$
$$\operatorname{Var} \tilde{\mathbf{y}}(t) = \tilde{\Sigma}.$$

The calculations and the structure of  $\tilde{\Sigma}$  are in Appendix F. Note that the infinite matrix A does not define a bounded operator on the whole space  $\ell^2$  since the first column is not square summable (it just belongs to  $\ell^{\infty}$ ) the domain being the linear subspace of all sequences in  $\ell^2$  having zero initial symbol.

## C. Infinite-Dimensional Distributed Average Consensus

Assume that the k-th agent adjusts its output in discrete time by a symmetric linear relation

$$\mathbf{y}_{k}(t+1) = a_{k}\mathbf{y}_{k}(t) + \sum_{j \in N_{k}} a_{k,j} \left(\mathbf{y}_{j}(t) - \mathbf{y}_{k}(t)\right) + \mathbf{w}_{k}(t) \quad (29)$$

where k = 1, 2, ... and the sum is over the set of neighbors  $N_k$  of each state k, which we assume to be a finite set. The overall motion can be described as

$$\mathbf{y}(t+1) = A\mathbf{y}(t) + \mathbf{w}(t) \tag{30}$$

starting at some initial state  $\mathbf{y}(0)$ . Here, A is a matrix with positive elements such that

$$A = A^\top \qquad A1 = 1$$

an infinite doubly stochastic matrix. The state of (30) is not stationary since it has a random walk component. We want to see if for some averaging sequence  $\{a_n\}$  the limit

$$\lim_{n \to \infty} a_n^\top \mathbf{x}(t)$$

is non-zero. This would imply the existence of a flocking component. Problems of this kind have been studied in the finite-dimensional setting in [58]. Here we study a slightly different model, obtained by modifying (29) so as to deal with an infinite number of agents. Let us assume that:

- for each n ≥ n<sub>0</sub>, where n<sub>0</sub> is a fixed initial integer, the symmetric doubly stochastic matrix A<sub>n</sub>, achieves consensus on the first n agents;
- 2) define a sequence of matrices  $\bar{A}_n := (1 (1/n))A_n$ , and assume that consensus is reached as  $n \to \infty$ .

Denoting by  $\overline{A}$  the limit of the sequence  $\{\overline{A}_n, n \in \mathbb{N}\}$ , the following result holds.

Proposition 6.3: The model

$$\mathbf{y}(t+1) = A\mathbf{y}(t) + \mathbf{w}(t), \quad Q = I \tag{31}$$

admits a flocking structure. The relative GFA decomposition has one (q = 1) latent factor.

The proof is in Appendix G.

## D. Generalizations

In both the above examples, the matrix A can be decomposed as the sum of a bounded operator in  $\ell^2$  plus an unbounded rank one perturbation in  $\ell^{\infty}$ . It then happens that the unbounded solution of the Lyapunov equation has exactly the same column space and the same rank as the unbounded perturbation of A. Although we do not have a rigorous proof, this seems likely to be a general fact. Let us assume for simplicity that A is symmetric and is a direct sum of a finite, rank q perturbation plus a bounded operator in  $\ell^2$ , defined on  $\mathcal{F} \oplus \ell^2$  where  $\mathcal{F}$ has dimension q, which are complementary invariant subspaces. For each finite n, we therefore have a block decomposition

$$A[F G] = [F G] \operatorname{diag}\{\hat{A}, \hat{A}\}$$

where for  $n \to \infty$  the q columns of F belong to  $\ell^{\infty}$  but not to  $\ell^2$  while G is a unitary operator in  $\ell^2$ . Writing formally T := [F G], we have  $A = T \operatorname{diag}\{\hat{A}, \tilde{A}\}T^{-1}$  and also

$$A^k = T \operatorname{diag}\{\hat{A}^k, \; \tilde{A}^k\}T^{-1}$$

Since Q is diagonal, letting  $T^{-1}QT^{-*} := \text{diag}\{\hat{Q}, \tilde{Q}\}$ , the solution of (24) can be written as

$$\Sigma = [F \ G] \operatorname{diag} \left\{ \sum_{k=0}^{+\infty} \hat{A}^k \hat{Q} \left[ \hat{A}^\top \right]^k, \sum_{k=0}^{+\infty} \tilde{A}^k \tilde{Q} \left[ \tilde{A}^\top \right]^k \right\} \begin{bmatrix} F^\top \\ G^* \end{bmatrix}$$
$$= F \hat{P} F^\top + G \tilde{P} G^*$$

where  $\hat{P} \in \mathbb{R}^{q \times q}$  and  $\tilde{P}$  are symmetric and positive with  $\tilde{P}$  bounded in  $\ell^2$ . Hence, when A has q eigenvectors in  $\ell^{\infty}$  (but not in  $\ell^2$ ), the steady-state covariance has a GFA decomposition.

Changing basis in (23) by letting  $\mathbf{y}(t) = [F \ G] \begin{bmatrix} \hat{\mathbf{x}}(t) \\ \tilde{\mathbf{x}}(t) \end{bmatrix}$  so that

$$\begin{bmatrix} \hat{\mathbf{x}}(t+1) \\ \tilde{\mathbf{x}}(t+1) \end{bmatrix} = \operatorname{diag}\{\hat{A}, \ \tilde{A}\} \begin{bmatrix} \hat{\mathbf{x}}(t) \\ \tilde{\mathbf{x}}(t) \end{bmatrix} + \begin{bmatrix} \hat{\mathbf{w}}(t) \\ \tilde{\mathbf{w}}(t) \end{bmatrix}$$

we end up with a GFA decomposition  $\mathbf{y}(t) = \hat{\mathbf{y}}(t) + \tilde{\mathbf{y}}(t)$ where the two components

$$\hat{\mathbf{y}}(t) = F\hat{\mathbf{x}}(t), \qquad \tilde{\mathbf{y}}(t) = G\tilde{\mathbf{x}}(t)$$

are the flocking and the idiosyncratic parts of  $\mathbf{y}(t)$ . Note that the noise components  $\hat{\mathbf{w}}(t)$  and  $\tilde{\mathbf{w}}(t)$  are mutually uncorrelated and hence so are  $\hat{\mathbf{y}}(t)$  and  $\tilde{\mathbf{y}}(t)$ .

#### E. Separable Space-Time Processes

Random fields which are often encountered in geostatistics, hydrology, marine wave models, meteorology, and environmen-

tal applications, see e.g., [38] and the references therein, belong to the class of so-called *separable space-time processes* 

$$\mathbf{y}(k,t) = \sum_{i=1}^{m} \mathbf{v}_i(k) \mathbf{u}_i(t)$$
(32)

represented as the product of a space,  $\mathbf{v}(k) := [\mathbf{v}_1(k)\mathbf{v}_2(k)\ldots\mathbf{v}_m(k)]$ , and time component,  $\mathbf{u}(t) := [\mathbf{u}_1(t)\mathbf{u}_2(t)\ldots\mathbf{u}_m(t)]^\top$ , both zero mean and with finite variance. In general, one should take  $m = \infty$  [56] but finite dimensional approximations are often enough. To discuss these models, one should generalize the static theory in the preceding sections to *m*-vector-valued processes. Although this is quite straightforward, involving no new concepts but just more notations, for the sake of clarity we shall restrain to the scalar case m = 1.

The model (32) needs to be specified probabilistically, as the dynamics of the "time" process  $\{\mathbf{u}(t)\}\$  may well be space dependent and dually, the distribution of  $\mathbf{v}(k)$  may be *a priori* time-dependent. The following assumption specifies in probabilistic terms the multiplicative structure (32) 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

$$\mathbb{E}\left\{\mathbf{v}(k_1)\mathbf{v}(k_2)|\mathbf{u}(t_1)\mathbf{u}(t_2)\right\} = \mathbb{E}_{\mathbf{v}}\left\{\mathbf{v}(k_1)\mathbf{v}(k_2)\right\}$$
(33)

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 the multiplicative uncorrelation (33) one gets

$$\mathbb{E} \left\{ \mathbf{v}(k_1)\mathbf{v}(k_2)\mathbf{u}(t_1)\mathbf{u}(t_2) \right\} = \mathbb{E} \left\{ \mathbf{v}(k_1)\mathbf{v}(k_2) \right\} \mathbb{E} \left\{ \mathbf{u}(t_1)\mathbf{u}(t_2) \right\}$$
$$= \sigma_{\mathbf{v}}(k_1, k_2)\sigma_{\mathbf{u}}(t_1, t_2)$$
(34)

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 separable structure of the process. If  $\mathbf{v}$  and  $\mathbf{u}$ are jointly Gaussian, the multiplicative uncorrelation property follows if the two components are uncorrelated; namely their joint covariance is separable. This is a structure which is often assumed in the literature, see [35] and references therein. Assume now that the space process has a nontrivial GFA representation with q factors

$$\mathbf{v}(k) = \sum_{i=1}^{q} f_i(k) \mathbf{z}_i + \tilde{\mathbf{v}}(k)$$
(35)

where  $\hat{\mathbf{v}}(k) := \sum_i f_i(k)\mathbf{z}_i$  is the aggregate and  $\tilde{\mathbf{v}}(k)$  the idiosyncratic component 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 represent the random field (32) by a dynamic GFA model

$$\mathbf{y}(k,t) = \sum_{i=1}^{q} f_i(k) \mathbf{x}_i(t) + \tilde{\mathbf{y}}(k,t) := \hat{\mathbf{y}}(k,t) + \tilde{\mathbf{y}}(k,t).$$
(36)

**Proposition 6.4:** If the processes  $\mathbf{v}$  and  $\mathbf{u}$  are multiplicatively uncorrelated then the two terms  $\hat{\mathbf{y}}(k,t)$  and  $\tilde{\mathbf{y}}(h,s)$  in the GFA model (36) are uncorrelated for all k, h and t, s. Hence a separable random field satisfying the multiplicative

uncorrelation property has a flocking component if and only if its space process  $\mathbf{v}$  has a nontrivial aggregate component.

*Proof:* We have

$$\mathbb{E}\left\{\hat{\mathbf{y}}(k,t)\tilde{\mathbf{y}}(h,s)\right\} = \sum_{i=1}^{q} f_i(k)\mathbb{E}\left\{\mathbf{z}_i\mathbf{u}(t)\tilde{\mathbf{v}}(h)\mathbf{u}(s)\right\} \quad (37)$$

where the expectation in the last term can be written as

$$\mathbb{E}\left\{\mathbf{z}_{i}\tilde{\mathbf{v}}(h)\mathbf{u}(t)\mathbf{u}(s)\right\} = \mathbb{E}\left\{\mathbb{E}_{\mathbf{v}}\left[\mathbf{z}_{i}\tilde{\mathbf{v}}(h)|\mathbf{u}(t)\mathbf{u}(s)\right]\mathbf{u}(t)\mathbf{u}(s)\right\}$$
$$= \mathbb{E}\left\{\mathbb{E}_{\mathbf{v}}\left[\mathbf{z}_{i}\tilde{\mathbf{v}}(h)\right]\mathbf{u}(t)\mathbf{u}(s)\right\} = 0 \quad (38)$$

since the  $\mathbf{z}_i$ 's are random variables in  $H(\hat{\mathbf{v}})$  and  $\tilde{\mathbf{v}}(h)$  is orthogonal to this space. The last statement then follows directly.  $\Box$ 

Here is probably the simplest nontrivial example of decomposition (36).

*Example 6.1 (Exchangeable Space Processes):* Consider a *(weakly) exchangeable* space process **v**; i.e., a process whose second order statistics are invariant with respect to all index permutations of locations (k, j). Clearly the covariances  $\sigma_{\mathbf{v}}(k, j) = \mathbb{E}\mathbf{v}(k)\mathbf{v}(j)$  must be independent of k, j for  $k \neq j$  and  $\sigma_{\mathbf{v}}(k, k) = \sigma^2 > 0$  must be independent of k, see [2]. Letting  $\rho := \sigma_{\mathbf{v}}(k, j), k \neq j$ , one has

$$\Sigma_{\mathbf{v}} = \begin{bmatrix} \sigma^2 & \rho & \rho & \rho & \dots \\ \rho & \sigma^2 & \rho & \rho & \dots \\ \dots & \ddots & \dots \end{bmatrix}$$
(39)

where  $\sigma^2 > |\rho|$  for positive definiteness. Letting f denote an infinite column vector with components all equal to  $\rho$ , one can decompose  $\Sigma_{\mathbf{v}}$  as

$$\Sigma_{\mathbf{v}} = f f^{\top} + (\sigma^2 - \rho) I \tag{40}$$

where I denotes an infinite identity matrix. This is a Factor Analysis decomposition of rank q = 1 of  $\Sigma_{\mathbf{v}}$  with  $\tilde{\Sigma}_{\mathbf{v}}$  a diagonal matrix. Hence a weakly exchangeable space process is a 1-factor process with an idiosyncratic component which is actually white. In the GFA representation (35), there is just one factor  $\mathbf{z}$  and the factor loading vector f does not depend on the space coordinate.

Consider a random field with the multiplicative structure (32), then the flocking component

$$\hat{\mathbf{y}}(k,t) = f\mathbf{x}(t), \qquad \mathbf{x}(t) = \mathbf{z}\mathbf{u}(t)$$

describes a constant, space independent, configuration moving randomly in time.

## VII. CONCLUSION

We have proposed a new modeling paradigm for large dimensional aggregates of random systems based on the theory of Generalized Factor Analysis. We have discussed in some depth static GFA representations and characterized in a rigorous way their properties, especially the nature of the *idiosyncratic* and *aggregate* components and provided new conditions guaranteeing uniqueness of the representation. We have shown that the model splits the output y of the system into two components describing the short- and long- range interaction among the agents of the ensemble. For wide-sense stationary ensembles, the nature and existence of these components can be clarified in the light of the Wold decomposition. For time-dependent evolutions, the aggregate component provides the core structure of the (random) flocking component. A detailed analysis of interesting classes of random fields, such as the linear evolution equation in (23), by using the decomposition of the steady state covariance has just been touched upon. Visibly, there is here ample room for further research on specific structures. Also the statistical identification had regrettably to be left out and will be considered in forthcoming publications.

## APPENDIX

#### A. Proof of Lemma 3.1

Let  $\lambda_1(\Sigma_n)$  be the maximal eigenvalue of  $\Sigma_n$ . Since

$$\Sigma_n \le \lambda_1(\Sigma_n) I_n \le \lambda_1(\Sigma) I_n \tag{41}$$

where  $I_n$  is the  $n \times n$  identity matrix and  $\lambda_1(\Sigma) < \infty$  by assumption, it follows that for all sequences  $x, y \in \ell^2$ 

$$x^{n}\Sigma_{n}y^{n} \leq \lambda_{1}(\Sigma) \|x^{n}\|_{2} \|y^{n}\|_{2}, \qquad n = 1, 2, \dots$$
 (42)

Then the result follows from the theorem in [1, p. 53].

#### B. Proof of Theorem 3.1

Assume first that  $\lim_{n\to\infty} \lambda_1(\Sigma_n) = +\infty$ . Since  $\Sigma_n > 0$  is symmetric it has a spectral representation

$$U_n^{\top} \Sigma_n U_n = D_n \tag{43}$$

where  $U_n$  is orthonormal and  $D_n = \text{diag}\{\lambda_1(\Sigma_n), \ldots, \lambda_n(\Sigma_n)\}$ . Consider the first column of  $U_n$ , say  $u_1^n$ , which is the eigenvector of  $\lambda_1(\Sigma_n)$  and define the sequence of elements in  $\ell^2 \cap \ell^2(\Sigma)$  constructed as

$$a_n := \frac{1}{\sqrt{\lambda_1(\Sigma_n)}} \begin{bmatrix} (u_1^n)^\top & 0 & \dots \end{bmatrix}^\top, \qquad n = 1, 2, \dots$$
(44)

Since  $\lim_{n\to\infty} \lambda_1(\Sigma_n) = +\infty$ , this is an AS, for which

$$\left\|\boldsymbol{a}_{n}^{\mathsf{T}}\boldsymbol{\mathbf{y}}\right\|^{2} = \frac{1}{\lambda_{1}(\Sigma_{n})} \left(\boldsymbol{u}_{1}^{n}\right)^{\mathsf{T}} \Sigma_{n} \boldsymbol{u}_{1}^{n} = 1$$
(45)

for every n and hence the sequence y cannot be idiosyncratic.

Conversely, suppose that  $\lambda_1(\Sigma) < +\infty$  and again use the diagonalization (43). Let  $a_n$  be an arbitrary AS and consider the random variable  $\mathbf{z} = \lim_{n \to \infty} a_n^\top \mathbf{y} = \lim_{n \to \infty} a_n^{n^\top} \mathbf{y}^n$ , which has variance

$$\operatorname{var}[\mathbf{z}] = \lim_{n \to \infty} \left( a_n^n \right)^\top U_n D_n U_n^\top a_n^n := \left( d_n^n \right)^\top D_n d_n^n \qquad (46)$$

where the vector  $d_n^n := U_n^\top a_n^n$  is used to form the first n elements of an infinite string, say  $d_n$ , whose remaining entries are taken equal to those of  $a_n$ ; i.e.,  $d_n(k) = a_n(k)$  for k > n. Clearly  $d_n$  is an AS.

Since 
$$(d_n^n)^{\top} D_n d_n^n = \sum_{k=1}^n \lambda_k(\Sigma_n) d_n(k)^2$$
, one can write

$$\operatorname{var}[\mathbf{z}] = \lim_{n \to \infty} \sum_{i=1}^{n} \lambda_k(\Sigma_n) d_n(k)^2 \le \lim_{n \to \infty} \lambda_1(\Sigma) \sum_{k=1}^{n} d_n(k)^2$$
$$= \lim_{n \to \infty} \lambda_1(\Sigma) \|d_n\|_2^2 = 0$$

which shows that y is idiosyncratic.

#### C. Proof of Theorem 4.1

First we prove the sufficiency of condition (15). Let k be a fixed positive constant and let  $f_1$  be such that

$$\lim_{n \to \infty} \|f_1^n - \Pi [f_1^n | \mathcal{F}_1^n]\|_2 = k^{\frac{1}{2}} < +\infty.$$
(47)

Let

$$\tilde{f}_1^n = f_1^n - \Pi \left[ f_1^n | \mathcal{F}_1^n \right] = f_1^n - \alpha_2^n f_2^n - \dots - \alpha_q^n f_q^n; \quad (48)$$

whence, defining  $\tilde{F}^n := [\tilde{f}_1^n f_2^n \dots f_q^n]$ , one can write  $\tilde{F}^n = F^n T^n$ , with  $T^n$  is a full rank matrix of the form

$$T^{n} = \begin{bmatrix} 1 & 0\\ -\alpha_{n} & I_{q-1} \end{bmatrix}$$
(49)

where  $\alpha_n := [\alpha_2^n \dots \alpha_q^n]^\top$ . Since  $\tilde{f}_1^n \perp f_i^n$ ,  $i \neq 1$ , the Gramian matrix of  $\tilde{F}^n$  is block diagonal,

$$\tilde{F}^{n\top}\tilde{F}^n = \begin{bmatrix} \left\| \tilde{f}_1^n \right\|^2 & 0\\ 0 & A_n \end{bmatrix}$$
(50)

where  $A_n$  is a positive definite matrix whose eigenvalues tend to infinity as n increases. Note that the spectrum of  $\tilde{F}^{n\top}\tilde{F}^n$ contains the eigenvalue  $\|\tilde{f}_1^n\|^2$ , which, for  $n \to \infty$ , converges to  $k < +\infty$ . Now, let us compute the trace of both sides of the identity  $T^n(\tilde{F}^n^\top \tilde{F}^n)^{-1}T^{n\top} = (F^n^\top F^n)^{-1}$  obtaining

$$\operatorname{tr}\left[(F^{n^{\top}}F^{n})^{-1}\right] = \operatorname{tr}\left[T^{n}(\tilde{F}^{n^{\top}}\tilde{F}^{n})^{-1}T^{n^{\top}}\right] \\ = \operatorname{tr}\left[T^{n^{\top}}T^{n}(\tilde{F}^{n^{\top}}\tilde{F}^{n})^{-1}\right] \\ = \operatorname{tr}\left[\begin{array}{cc}1+\|\alpha_{n}\|^{2} & -\alpha_{n}^{\top}\\ -\alpha_{n} & I_{q-1}\end{array}\right]\left[\begin{array}{cc}k^{-1} & 0\\ 0 & A_{n}^{-1}\end{array}\right] \\ = \operatorname{tr}\left[\begin{array}{cc}k^{-1}(1+\|\alpha_{n}\|^{2}) & -\alpha_{n}^{\top}A_{n}^{-1}\\ -\alpha_{n}k^{-1} & A_{n}^{-1}\end{array}\right] \\ = k^{-1}\left(1+\|\alpha_{n}\|^{2}\right) + \operatorname{tr}\left[A_{n}^{-1}\right].$$
(51)

Since the eigenvalues of  $A_n$  tend to infinity, those of  $A_n^{-1}$  tend to zero, while, for every n we have  $k^{-1}(1 + ||\alpha_n||^2) > 0$ . Thus, one eigenvalue of  $(F^{n\top}F^n)^{-1}$  is bounded below by a fixed constant as n tends to infinity. Hence, we conclude that one eigenvalue of  $F^{n\top}F^n$  remains bounded as n tends to infinity, which is a contradiction.

For the necessity, we define  $f_i^{n_1,n_2} := [f_i(n_1) \dots f_i(n_2)]^\top$ and observe that condition (15) implies that

$$\lim_{n \to \infty} \|f_i^{n_1, n} - \Pi [f_i^{n_1, n} | \mathcal{F}_i^{n_1, n}]\|_2 = +\infty$$
 (52)

for every index i = 1, ..., q and natural number  $n_1$ . Moreover, by definition of limit, we have that for every  $n_1 \in \mathbb{N}$  and  $K \in \mathbb{R}_+$  there exists an integer  $n_2$  such that the inequality (with an obvious meaning of the symbols)

$$\|f_i^{n_1,n_2} - \Pi \left[f_i^{n_1,n_2} | \mathcal{F}_i^{n_1,n_2}\right]\|_2^2 \ge K$$
(53)

holds for every  $i = 1, \ldots, q$ .

Now, consider the sequence generated by the q-th eigenvalue of the matrix  $F^{n\top}F^n$ , say  $\{\lambda_q^n; n \in \mathbb{N}\}$ . Our goal is to show that for every natural  $n_1$  and arbitrary constant c > 0 there exists a natural number  $n_2$  such that  $\lambda_q^{n_2} \ge \lambda_q^{n_1} + c$ , so that  $\lim_{n\to\infty} \lambda_q^n = +\infty$ . To this end, fix c and, for a generic  $n_1$ , consider the normalized eigenvector of the q-th eigenvalue of the matrix  $F^{n_2 \top}F^{n_2}$ , say  $v_q^{n_2}$ . Since for every  $n_2 > n_1$  it holds that

$$F^{n_2 \top} F^{n_2} = F^{n_1 \top} F^{n_1} + F^{n_1, n_2 \top} F^{n_1, n_2}$$
(54)

we can write

$$\lambda_q^{n_2} = v_q^{n_2 \top} F^{n_1 \top} F^{n_1} v_q^{n_2} + v_q^{n_2 \top} F^{n_1, n_2 \top} F^{n_1, n_2} v_q^{n_2}.$$
 (55)

Consider the first term on the right side of this identity; expressing  $v_q^{n_2}$  as a linear combination of the eigenvectors of  $F^{n_1 \top} F^{n_1}$ , i.e.,  $v_q^{n_2} = \alpha_1 v_1^{n_1} + \cdots + \alpha_q v_q^{n_1}$ , the orthogonality of these eigenvectors implies that

$$v_{q}^{n_{2}\top}F^{n_{1}\top}F^{n_{1}}v_{q}^{n_{2}} = \lambda_{1}^{n_{1}}\alpha_{1}^{2} + \dots + \lambda_{q}^{n_{1}}\alpha_{q}^{2} \ge \lambda_{q}^{n_{1}}\sum_{i=1}^{q}\alpha_{i}^{2} = \lambda_{q}^{n_{1}}$$
(56)

so that

$$\lambda_q^{n_2} \ge \lambda_q^{n_1} + v_q^{n_2 \top} F^{n_1, n_2 \top} F^{n_1, n_2} v_q^{n_2}.$$
 (57)

Now we have to show that we can always find an integer  $n_2$  such that the quantity

$$v_q^{n_2 \top} F^{n_1, n_2 \top} F^{n_1, n_2} v_q^{n_2}$$

can be chosen arbitrarily large, i.e., greater or equal to the previously fixed constant c. To this end, take  $n_2$  such that for every  $i = 1, \ldots, q$  the inequality (53) holds, with  $K = c\sqrt{q}$ . Then, there is an index i such that the i-th component of the norm one vector  $v_q^{n_2} = [w_1 \ldots w_q]^{\top}$ , satisfies the inequality  $w_i \ge (1/\sqrt{q})$ . Without loss of generality we may and shall assume that i = 1. Let  $\alpha_2 \ldots \alpha_q$  be defined as in (48) and set

$$\tilde{f}_1^{n_1,n_2} := f_1^{n_1,n_2} - \alpha_2 f_2^{n_1,n_2} - \dots - \alpha_q f_q^{n_1,n_2}$$
(58)

so that we have

$$v_{q}^{n_{2}\top}\!F^{n_{1},n_{2}\top}\!F^{n_{1},n_{2}}v_{q}^{n_{2}} = v_{q}^{n_{2}\top}T^{n\top} \begin{bmatrix} \|\hat{f}_{1}^{n_{1},n_{2}}\|^{2} & 0\\ 0 & A_{n} \end{bmatrix} T^{n}v_{q}^{n_{2}}$$
(59)

where  $T^n$  has the same structure as in (49). Now, observe that

$$T^n v_q^{n_2} = \begin{bmatrix} w_1 & -\alpha_2 w_1 + w_2 & \dots & -\alpha_q w_1 + w_q \end{bmatrix}^\top$$
 (60)

which implies that (59) is equal to  $w_1^2 \|\tilde{f}_1^{n_1,n_2}\|^2 + Q$ , where Q is a positive constant. Hence, from (58) we have  $v_q^{n_2 \top} F^{n_1,n_2 \top} F^{n_1,n_2} v_q^{n_2} > c$  and hence, recalling (57)

$$\lambda_q^{n_2} \ge \lambda_q^{n_1} + c \tag{61}$$

which proves the theorem.

#### D. Proof of Proposition 5.1

By a standard Q-R factorization we can orthogonalize the columns of  ${\cal F}_n$ 

$$\begin{bmatrix} f_1^n & f_2^n & \dots & f_q^n \end{bmatrix}$$

$$= \begin{bmatrix} g_1^n & g_2^n & \dots & g_q^n \end{bmatrix} \begin{bmatrix} 1 & r_{1,2} & r_{1,3} & \dots & r_{1,q} \\ 0 & 1 & r_{2,3} & \dots & r_{2,q} \\ 0 & 0 & 1 & \ddots & r_{3,q} \\ \dots & \dots & \dots & \ddots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$
(62)

which we shall write compactly as

$$F_n = Q_n R_n \tag{63}$$

where  $Q_n := [g_1^n g_2^n \dots g_q^n]$  has orthogonal columns. It is wellknown that each  $g_i^n$  can be obtained by a sequential Gram-Schmidt orthogonalization procedure as the difference of  $f_i^n$ with its projection onto the subspace  $\operatorname{span}\{f_j^n, j < i\} \subset \mathcal{F}_i^n$ . Hence  $\|g_i^n\| \ge \|\tilde{f}_i^n\|$  and hence, by assumption, tends to  $\infty$ when  $n \to \infty$ .

Next, define

$$a_{i,n}^{\top} := \frac{1}{\|g_i^n\|_2^2} \begin{bmatrix} g_i^n(1) & g_i^n(2) & \dots & g_i^n(n) & 0 & \dots \end{bmatrix}$$
(64)

where the  $g_i^n$ 's are as defined above. Since  $||g_i^n||_2 \to \infty$  with n, we have  $||a_{i,n}||_2 = 1/||g_i^n||_2 \to 0$  as  $n \to \infty$ . Hence  $a_{i,n}$  is an AS.

Note that we can express each  $f_i^n$  as  $f_i^n = g_i^n + \sum_{j=1}^{i-1} r_{j,i} g_j^n$  so that

$$a_{i,n}^{\top} f_i = \frac{1}{\|g_i^n\|_2^2} \|g_i^n\|_2^2 = 1$$
(65)

for all *n* large enough and by a similar calculation one can easily check that  $a_{i,n}^{\top} f_j = 0$ , for all j < i. With these  $a_{i,n}$  construct a sequence of  $q \times \infty$  matrices

$$A_n := \begin{bmatrix} a_{1,n}^\top \\ \vdots \\ a_{q,n}^\top \end{bmatrix}$$
(66)

which provides an asymptotic left-inverse of F, in the sense that  $\lim_{n\to\infty} A_n F = R$ , where R is the limit of a sequence of  $q \times q$ matrices all of which are upper triangular with ones on the main diagonal. Next, define the random vector  $\mathbf{z}_n := A_n \mathbf{y}$  which converges as  $n \to \infty$  to a q-dimensional  $\mathbf{z}$  whose components must belong to  $\mathcal{F}(\mathbf{y})$ . These q components form in fact a basis for  $\mathcal{F}(\mathbf{y})$  as the covariance  $\mathbb{E}\mathbf{z}_n\mathbf{z}_n^{\top}$  converges to  $RR^{\top}$  which is non singular. From this, one can easily get an orthonormal basis  $\mathbf{x}$ , in  $H(\hat{\mathbf{y}})$ . Hence, since F is known, we can form  $\hat{\mathbf{y}} = F\mathbf{x}$ and letting  $\tilde{\mathbf{y}} := \mathbf{y} - \hat{\mathbf{y}}$  does yield a GFA representation of  $\mathbf{y}$ inducing the given GFA decomposition of  $\Sigma$ . Uniqueness is then guaranteed in force of Proposition 4.2.

## E. Proof of Proposition 6.1

For infinite covariance matrices we have the positive semidefinite ordering  $\Sigma_1 \leq \Sigma_2$  if and only if  $a^{\top}(\Sigma_1 - \Sigma_2)a \leq 0$  for all finite support sequences  $a \in \mathbb{R}^{\infty}$ . Let  $\mathbf{y}(t)$  be a time-stationary random field with matrix covariance function  $\Sigma(\tau) := \mathbb{E}\mathbf{y}(t + \tau)\mathbf{y}(t)^{\top}$ . For any finite support sequence  $a \in \mathbb{R}^{\infty}$  the scalar covariance function  $\sigma_{\mathbf{z}}(\tau)$  of the process  $\mathbf{z}(t) := a^{\top}\mathbf{y}(t)$  satisfies the well-known (Schwartz) inequality  $\sigma_{\mathbf{z}}(\tau) \leq \sigma_{\mathbf{z}}(0)$ ; hence the matrix covariance function of a stationary process satisfies  $\Sigma(\tau) \leq \Sigma(0)$ . It follows that if  $\Sigma(0)$  is a bounded operator in  $\ell^2$ then all covariances  $\Sigma(\tau)$  must also be bounded. The following lemma is a straightforward consequence of this fact.

*Lemma A.1:* A time-stationary random field  $\mathbf{y}(t)$  is idiosyncratic; that is

$$\lim_{n \to \infty} a_n^{\top} \mathbf{y}(t) = 0, \quad \text{for any} \quad t \in \mathbb{Z}$$

for all AS's  $a_n$ , if and only if  $\mathbf{y}(0)$  is an idiosyncratic sequence.

The lemma above implies in particular that a covariance function  $\Sigma(\tau)$  is the covariance of an idiosyncratic stationary random field if and only if  $\Sigma(0)$  is a bounded operator on  $\ell^2$ .

Now assume that  $\Sigma(0)$  has a static GFA decomposition of rank q and let x and  $\tilde{y}$  be constructed as in the proof of Proposition 5.1 so that the vector y(0) has a GFA representation

$$\mathbf{y}(0) = F\mathbf{x} + \tilde{\mathbf{y}}$$

where  $\mathbf{x}$  and  $\tilde{\mathbf{y}}$  have uncorrelated components belonging to  $H(\mathbf{y}(0))$ . Let  $\mathbf{H}(\mathbf{y})$  denote the closed linear span of the scalar components of the random field  $\mathbf{y}$ ; i.e.,

$$\mathbf{H}(\mathbf{y}) := \text{closure of } \left\{ \sum_{k,t} a_{k,t} \mathbf{y}(k,t); k = 0, 1, 2, \dots; t \in \mathbb{Z} \right\}$$

where the real numbers  $a_{k,t}$  are arbitrary but non zero only for a finite set of values of the indices. Let  $U : \mathbf{H}(\mathbf{y}) \to \mathbf{H}(\mathbf{y})$  be the forward shift operator of the process defined, for all finite support vectors a, by

$$Ua^{\top}\mathbf{y}(t) = a^{\top}\mathbf{y}(t+1), \qquad t \in \mathbb{Z}.$$

It is well known that U can be extended to the whole of  $\mathbf{H}(\mathbf{y})$  as a unitary operator [49] and that every scalar random variable  $\mathbf{z} \in \mathbf{H}(\mathbf{y})$  can be propagated in time by the action of the shift as  $\mathbf{z}(t) := U^t \mathbf{z}$  to form a stationary scalar process. This unitary propagation can in fact be applied to vector random variables of arbitrary dimension.

It follows hat  $\mathbf{x}(t) := U^t \mathbf{x}$  and  $\tilde{\mathbf{y}}(t) := U^t \tilde{\mathbf{y}}$  have uncorrelated components for all t. Moreover, by Lemma A.1 the stationary process  $\tilde{\mathbf{y}}(t)$  is idiosyncratic and  $\hat{\mathbf{y}}(t) := F\mathbf{x}(t)$  is a flocking process since the columns of F are strongly linearly independent.

#### F. Proof of Proposition 6.2

Consider first the case n = 3 and write the solution to the related Lyapunov equation as

$$\Sigma_3 = \begin{bmatrix} p_1 & p_2 & p_3 \\ p_2 & p_4 & p_5 \\ p_3 & p_5 & p_6 \end{bmatrix}.$$
 (67)

Then, simple calculations show that

$$p_{1} = \frac{1}{1-a^{2}}, \qquad p_{2} = p_{3} = \frac{a}{(1+a)(1-a^{2})}$$

$$p_{4} = p_{6} = \frac{1}{1-a^{2}} + \frac{1}{(1+a)^{2}} + 2\frac{a^{2}}{(1+a)^{2}(1-a^{2})}$$

$$p_{5} = \frac{1}{(1+a)^{2}} + 2\frac{a^{2}}{(1+a)^{2}(1-a^{2})}.$$
(68)

Now assume that, for a given  $n \ge 3$ , the solution to the equation  $X_n - A_n X_n A_n^{\top} = I_n$  has the form

$$\Sigma_{n} = \begin{bmatrix} p_{1} & p_{3} & p_{3} & p_{3} & \dots & p_{3} \\ p_{3} & p_{4} & p_{5} & p_{5} & \dots & p_{5} \\ p_{3} & p_{5} & p_{4} & p_{5} & \dots & p_{5} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ p_{3} & p_{5} & \dots & p_{5} & p_{4} & p_{5} \\ p_{3} & p_{5} & \dots & p_{5} & p_{5} & p_{4} \end{bmatrix};$$
(69)

our goal is to show that  $\Sigma_{n+1}$  has an analogous structure, that is

$$\Sigma_{n+1} = \begin{bmatrix} \Sigma_n & p \\ p^\top & p_4 \end{bmatrix}$$
(70)

where  $p = [p_3 \ p_5 \ \dots \ p_5]^{\top}$ . To this end, express the variable  $X_{n+1}$  as

$$X_{n+1} = \begin{bmatrix} X_n & z \\ z^\top & u \end{bmatrix}$$

and the matrix  $A_{n+1}$  as

$$A_{n+1} = \begin{bmatrix} A_n & 0\\ b^\top & a \end{bmatrix}$$

where  $b = [1 - a \ 0 \ \dots \ 0]^{\top}$ . Then the related Lyapunov equation has the form

$$\begin{bmatrix} X_n & z \\ z^{\top} & u \end{bmatrix} - \begin{bmatrix} A_n & 0 \\ b^{\top} & a \end{bmatrix} \begin{bmatrix} X_n & z \\ z^{\top} & u \end{bmatrix} \begin{bmatrix} A_n^{\top} & b \\ 0 & a \end{bmatrix} = I_{n+1} \quad (71)$$

which can be rewritten as

$$\begin{bmatrix} X_n - A_n X_n A_n^{\top} & (I_n - aA_n)z - A_n X_n b\\ z^{\top} (I_n - aA_n^{\top}) - b^{\top} X_n A_n^{\top} & (1 - a^2) u - b^{\top} X_n b - 2ab^{\top} z \end{bmatrix} = \begin{bmatrix} I_n & 0\\ 0 & 1 \end{bmatrix}.$$
(72)

The top-left block of (72) admits the solution given by (69). Then, by inserting this into the top-right block, one then gets z = p. Finally, by exploiting the former findings, from the bottom-right block one has  $u = p_4$ , and hence the solution is exactly (70). Hence, one can easily observe that the matrix  $\overline{\Sigma}_n$ , obtained by discarding the first row and column from  $\Sigma_n$ , has the structure

$$\begin{bmatrix} p_5 & p_5 & \dots \\ p_5 & p_5 & \\ \vdots & \ddots \end{bmatrix} + \operatorname{diag}\{p_4 - p_5, \dots, p_4 - p_5\}$$
(73)

that is, it admits a rank-one plus diagonal decomposition, where the vector generating the rank-one matrix is  $\bar{f} = [\sqrt{p_5} \sqrt{p_5} \dots]$ , with  $\sqrt{p_5} = (1 + a^2)^{(1/2)}/((1 + a)(1 - a^2)^{(1/2)})$ , while the elements of the diagonal matrix are  $p_4$  –

 $p_5 = 1/(1 - a^2)$ . Now, to complete the proof we need to show that also the matrix  $\Sigma_n$  admits a similar decomposition, i.e.,

$$\Sigma_n = \begin{bmatrix} f_0\\ \bar{f} \end{bmatrix} + \operatorname{diag}\left\{\sigma_0^2, \ \frac{1}{(1-a^2)}, \dots, \frac{1}{(1-a^2)}\right\}$$

This can be done be observing that, for any integer k > 0, it has to be  $p_3 = f_0 \bar{f}(k)$ , and so  $f_0 = a/(1 - a^4)^{(1/2)}$ . Moreover,  $\sigma_0^2$ is easily found by computing  $\sigma_0^2 = p_1 - f_0^2 = 1$ . Finally, since by comparing the leader dynamics

$$\mathbf{y}_0(t) = a\mathbf{y}_0(t-1) + \mathbf{w}_0(t-1)$$

with its GFA decomposition

$$\mathbf{y}_0(t) = f_0 \mathbf{x}(t) + \tilde{\mathbf{y}}_0(t)$$

where both  $\tilde{\mathbf{y}}_0(t)$  and  $\mathbf{w}_0(t-1)$  are white noise with the same variance, it has to be that  $\mathbf{x}(t) = (1-a^4)^{(1/2)}\mathbf{y}_0(t-1)$ .

# G. Proof of Proposition 6.3

For  $n \ge n_0$ , consider the Lyapunov equation  $\Sigma_n = \bar{A}_n \Sigma_n \bar{A}_n^\top + I_n$ , whose solution can be written

$$\Sigma_n = \sum_{j=0}^{\infty} \bar{A}_n^j \left( \bar{A}_n^j \right)^\top.$$
(74)

Since  $\bar{A}_n$  is symmetric, for every *j* the decomposition

$$\bar{A}_n^j \left( \bar{A}_n^j \right)^T = U_n S_n^{2j} U_n^\top$$

holds, with  $S_n$  being the matrix of the singular values of A and  $U_n$  a unitary matrix whose columns are the (normalized) eigenvectors of  $\overline{A}_n$ . Note that one of such singular values is  $(1 - (1/n))^2$  and the relative eigenvector is  $(1/\sqrt{n})\mathbb{1}_n$ , i.e., the normalized vector of all 1's in  $\in \mathbb{R}^n$ . The other eigenvalues are strictly stable. Then we can express  $\Sigma_n$  as

$$\Sigma_{n} = U_{n} \left( \sum_{j=0}^{\infty} S_{n}^{2j} \right) U_{n}^{\top}$$

$$= \frac{1}{\sqrt{n}} \left( \sum_{j=0}^{\infty} \left( 1 - \frac{1}{n} \right)^{2j} \right) \frac{1}{\sqrt{n}}^{\top} + \tilde{U}_{n} \left( \sum_{j=0}^{\infty} \tilde{S}_{n}^{2j} \right) \tilde{U}_{n}^{\top}$$

$$= 1 \frac{n}{2n+1} 1^{\top} + \tilde{U}_{n} \left( \sum_{j=0}^{\infty} \tilde{S}_{n}^{2j} \right) \tilde{U}_{n}^{\top}$$
(75)

where  $\tilde{U}_n$  and  $\tilde{S}_n$  are obtained from  $U_n$  and  $S_n$  by removing the parts related to the eigenvalue  $(1 - (1/n))^2$ . Now, take the averaging sequence (7)

$$a_n = \frac{1}{n} \left[ \mathbb{1}_n^{\mathsf{T}} 0 \dots \right], \quad \mathbb{1}_n \in \mathbb{R}^n$$
(76)

and apply it to  $\Sigma_n$ , that is, compute  $(1/n)\mathbb{1}_n^{\top}\Sigma_n\mathbb{1}_n(1/n)$ . Then, letting  $n \to \infty$ , the second term on the right hand side of (75) vanishes, while the first term gives

$$\frac{\mathbb{1}_{n}^{+}\mathbb{1}_{n}\mathbb{1}_{n}^{+}\mathbb{1}_{n}}{n(2n+1)} = \frac{n}{2n+1}$$
(77)

which converges asymptotically to a finite value. One can easily verify that the averaging sequence (76) is the only sequence converging to nonzero values.

#### REFERENCES

- [1] N. I. 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, vol. 1117, 1985, ser. Springer Lecture Notes in Mathematics, pp. 1-198.
- [3] B. D. O. Anderson and M. Deistler, "Generalized linear dynamic factor models-A structure theory," in Proc. of the IEEE Decision and Control Conference (CDC), 2008.
- [4] D. J. Bartholomew, "The foundations of factor analysis," Biometrika, vol. 71, no. 2, pp. 221-232, 1984.
- [5] P. A. Bekker and J. de Leeuw, "The rank of reduced dispersion matrices," Psychometrika, vol. 52, no. 1, pp. 125-135, 1987.
- [6] 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. Pattern Analysis and Machine Intelligence, vol. 29, no. 11, pp. 1958-1972, 2007.
- [7] G. Bottegal and G. Picci, "A note on generalized factor analysis models," in Proc. 50th Decision and Control Conf. (CDC), Orlando, FL, 2011, pp. 1485-1490.
- [8] G. Bottegal and G. Picci, "Modeling random flocks through generalized factor analysis," in Proc. European Control Conf. (ECC13), Zürich, Switzerland, 2013, pp. 2421–2426.
- [9] R. W. Brockett, "On the control of a flock by a leader," in Differential Equations and Topology. I, Volume 268 of Tr. Mat. Inst. Steklova, MAIK CINauka/Interperiodica, Moscow, Russia, 2010, pp. 56-63.
- [10] G. Chamberlain, "Funds, factors and diversification in arbitrage pricing models," Econometrica, vol. 51, no. 5, pp. 1305-1324, 1983.
- [11] G. Chamberlain and M. Rothschild, "Arbitrage, factor structure and meanvariance analysis on large asset markets," Econometrica, vol. 51, no. 5, pp. 1281-1304, 1983.
- [12] V. Chandrasekaran, S. Sanghavi, P. Parrilo, and A. Willsky, "Ranksparsity incoherence for matrix decomposition," SIAM J. Optimiz., vol. 21, pp. 572-596, 2011.
- [13] A. Chiuso and G. Picci, "Some identification techniques in computer vision," in Proc. IEEE Decision and Control Conf. (CDC), 2008.
- [14] H. Cramèr, "On some classes of non-stationary stochastic processes," in Proc. IV Berkeley Symp. Math. Statistics and Probability, 1961, vol. II, pp. 57–77, Univ. California Press.
- [15] F. Cucker and S. Smale, "Emergent behavior in flocks," IEEE Trans. Autom. Control, pp. 852-862, 2007.
- [16] M. Deistler, B. D. O. Anderson, W. Chen, and A. Filler, "Modelling high dimensional time series by generalized factor models, (semi-plenary lecture delivered at the MTNS 2010 Symposium," in Proc. Mathematical Theory of Networks and Systems Symp. (MTNS 2010), Budapest, Hungary, 2010, pp. 323-329.
- [17] M. Deistler, B. D. O. Anderson, A. Filler, C. Zinner, and W. Chen, "Generalized linear dynamic factor models: An approach via singular autoregressions," Eur. J. Control, vol. 3, pp. 211-224, 2010.
- [18] M. Deistler and C. Zinner, "Modelling high-dimensional time series by generalized dynamic factor models: An introductory survey," Commun. on Inform. and Syst., vol. 7, no. 2, pp. 153-166, 2007.
- [19] J. L. Doob, Stochastic Processes. New York: Wiley Classics Library, 1990, Reprint of the 1953 original.
- [20] G. Doretto, A. Chiuso, S. Soatto, and Y. N. Wu, "Dynamic textures," Int. J. Comput. Vis., vol. 51, no. 2, pp. 91–109, Feb. 2003. [21] F. Fagnani and S. Zampieri, "Randomized consensus algorithms over
- large scale networks," IEEE J. Select. Areas Commun., vol. 26, pp. 634-649,2008
- [22] M. Forni, M. Hallin, M. Lippi, and L. Reichlin, "The generalized dynamic factor model: Identification and estimation," The Rev. of Econom. and Statist., vol. 65, pp. 453-473, 2000.
- [23] M. Forni and M. Lippi, "The generalized dynamic factor model: Representation theory," Econom. Theory, vol. 17, pp. 1113-1141, 2001.
- [24] J. Geweke, "The dynamic factor analysis of economic time series," in Latent Variables in Socio-Economic Models, D. J. Aigner and A. S. Goldberger, Eds. North-Holland, 1977, pp. 365-383.
- [25] P. R. Halmos, "Shifts on Hilbert spaces," J. für Reine und Angewandte Mathematik, vol. 208, pp. 102-112, 1961.
- [26] P. Hartman and A. Wintner, "The spectra of Toeplitz matrices," Amer. J. Mathemat., vol. 76, pp. 867-882, 1954.

- [27] C. K. Hemelrijk and H. Hildenbrandt, "Some causes of the variable shape of flocks of birds," PLoS One, vol. 6, no. 8, p. e22479, Aug. 2011.
- Y. Hu and R. Chou, "On the Peña-Box model," J. Time Series Analy., vol. 25, pp. 811-830, 2004.
- [29] A. Jadbabaie, J. Lin, and S. Morse, "Coordination of groups of mobile agents using nearest neighbor rule," IEEE Trans. Autom. Control, vol. AC-48, pp. 988-1001, 2003.
- [30] R. E. Kalman, "Identifiability and modeling in econometrics," in Developments in Statistics, Vol. 4, vol. 4. New York: Academic Press, 1983, pp. 97-136.
- [31] R. E. Kalman, "Identifiability and problems of model selection in econometrics," in Advances in Econometrics, W. Hildebrandt, Ed. Cambridge, U.K.: Cambridge University Press, 1983.
- [32] D. N. Lawley and A. E. Maxwell, Factor Analysis as a Statistical Method, 2nd ed. London: Butterworths, 1971.
- W. Ledermann, "On the rank of the reduced correlation matrix in multiple factor analysis," *Psychometrika*, vol. 2, pp. 85–93, 1937.
- [34] W. Ledermann, "On a problem concerning matrices with variable diagonal elements," in Proc. Royal Soc. Edinburgh, 1939, vol. XL, pp. 1-17.
- [35] B. Li, M. G. Genton, and M. Sherman, "Testing the covariance structure of multivariate random fields," Biometrika, vol. 95, pp. 813-829, 2008.
- [36] K. Lin and D. Husmeier, "Mixtures of factor analyzers for modeling transcriptional regulation," in *Learning and Inference in Computational* Systems Biology, Lawrence et al., Ed. Cambridge, MA: MIT Press, 2010, ch. 7.
- [37] A. Lindquist and G. Picci, Linear Stochastic Systems: A Geometric Approach, 2011, In preparation, to be published.
- [38] C. Ma, "Stationary random fields in space and time with rational spectral densities," IEEE Trans. Inform. Theory, vol. 53, pp. 1019-1029, 2007.
- [39] L. Ning and T. T. Georgiou, "Sparse factor analysis via likelihood and l<sup>1</sup>-regularization," in Proc. 50th IEEE Conf. Decision and Control, Orlando, FL, 2011, pp. 5188-5192.
- [40] L. Ning, T. T. Georgiou, A. Tannenbaum, and S. P. Boyd, "Linear models based on noisy data and the Frisch scheme," CoRR, 2013, abs/1304.3877.
- [41] R. Olfati-Saber, "Flocking for multi-agent dynamic systems: Algorithms and theory," IEEE Trans. Autom. Control, vol. 51, p. 401420, 2006.
- [42] R. Olfati-Saber, J. A. Fax, and R. Murray, "Consensus and cooperation in networked multi-agent systems," in Proc. IEEE, 2007, vol. 95, pp. 215-233.
- [43] D. Peña and G. Box, "Identifying a simplifying structure in time series," J. Amer. Stat. Ass., vol. 82, pp. 836–843, 1987. [44] D. Peña and P. Poncela, "Nonstationary dynamic factor analysis,"
- J. Statist. Plann. and Infer., vol. 136, pp. 1237-1257, 2006.
- [45] G. Picci, "Parametrization of factor analysis models," J. Economet., vol. 41, pp. 17-38, 1987.
- [46] G. Picci and S. Pinzoni, "Dynamic factor-analysis models for stationary processes," IMA J. Math. Control and Inform., vol. 3, pp. 185-210, 1986.
- [47] I. Pournara and L. Wersnich, "Factor analysis for gene regulatory networks and transcription factor activity profiles," BMC Bioinform., vol. 8, no. 61, 2007.
- [48] C. W. Reynolds, "Flocks, herds and schools: A distributed behavioral model," in Proc. 14th Annu. Conf. Computer Graphics and Interactive Techniques SIGGRAPH 87, New York, 1987, p. 2534, ACM Press.
- [49] Y. A. Rozanov, Stationary Random Processes: Holden Days, 1967.
- [50] J. Saunderson, V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky, "Diagonal and low-rank matrix decompositions, correlation matrices, ellipsoid fitting," SIAM J. Matrix Anal. and Applic., vol. 33, no. 4, pp. 1395-1416, 2012.
- [51] J. Shen, "Cucker-smale flocking under hierarchical leadership," SIAM J. Appl. Math, vol. 68, pp. 694-719, 2007.
- [52] G. W. Stewart and J. G. Sun, Matrix perturbation Theory. New York: Academic Press, 1990.
- [53] A. Tahbaz-Salehi and A. Jadbabaie, "Consensus over ergodic stationary graph processes," IEEE Trans. Autom. Control, vol. 55, pp. 225-230, 2010.
- [54] H. G. Tanner, A. Jadbabaie, and G. J. Pappas, "Flocking in fixed and switching networks," IEEE Trans. Autom. Control, vol. 52, pp. 863-868, 2007.
- [55] J. J. P. Veerman, J. S. Caughman, G. Lafferriere, and A. Williams, "Flocks and formations," J. Stat. Phys., vol. 5-6, no. 121, pp. 901-936, 2005.
- [56] D. Venturi, "A fully symmetric nonlinear biorthogonal decomposition theory for random fields," Physica D, vol. 240, pp. 415-425, 2011.
- [57] T. Vicsek, A. Czirók, E. Ben-Jacoba, I. Cohen, and O. Schochet, "Novel type of phase transition in a system of self-driven particles," Phys. Rev. Lett., vol. 75, p. 12261229, 1995.
- [58] L. Xiao, S. Boyd, and S. J. Kim, "Distributed average consensus with least-mean-square deviation," J. Parallel and Distrib. Comput., vol. 67, no. 1, pp. 33-46, 2007.



**Giulio Bottegal** received the Master's degree in automation engineering (*cum laude*) and the Ph.D. degree in information engineering, both from the University of Padova, Padova, Italy, in 2009 and 2013, respectively.

Currently, he holds is a Research Fellow (postdoc) at the School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden. His research interests include stochastic systems, latent variables models, and regularization techniques for system identification.



**Giorgio Picci** (S'67–M'70–SM'91–F'94–LF'08) received the Dr.Eng. degree from the University of Padua, Padua, Italy, in 1967.

Currently, he is Professor Emeritus with the Department of Information Engineering, University of Padua, Padua, Italy. He has held several long-term visiting appointments with various American, Japanese, and European universities including Brown University, MIT, the University of Kentucky, Arizona State University, the Center for Mathematics and Computer Sciences (CWI) in

Amsterdam, the Royal Institute of Technology, Stockholm, Sweden, Kyoto University, and Washington University, St. Louis, MO. He has been contributing to systems and control mostly in the area of modeling, estimation, and identification of stochastic systems and published over 100 papers and edited three books in this area. He has been involved in various joint research projects with industry and state agencies.

Dr. Picci is a Fellow of IFAC and a foreign member of the Swedish Royal Academy of Engineering Sciences. He has been Chairman of the IFAC Technical Committee on Stochastic Systems, past member of the EUCA council, Project Manager of the Italian team for the Commission of the European Communities Network of Excellence System Identification (ERNSI), and General Coordinator of the Commission of European Communities IST project RECSYS, in the fifth Framework Program.