



HAL
open science

On marginal markovianity in gaussian pairwise Markov models

Wojciech Pieczynski

► **To cite this version:**

Wojciech Pieczynski. On marginal markovianity in gaussian pairwise Markov models. 2023. hal-04296719

HAL Id: hal-04296719

<https://hal.science/hal-04296719v1>

Preprint submitted on 20 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On marginal markovianity in Gaussian pairwise Markov models

Wojciech Pieczynski

SAMOVAR Telecom Sudparis, Institut Polytechnique de Paris, Palaiseau, France
wojciech.pieczynski@telecom-sudparis.eu

Abstract— We consider two discrete-time processes X and Y , the pair (X, Y) being Gaussian, homogeneous and Markovian. Such models, called "Gaussian homogeneous pairwise Markov models" (GH-PMMs), extend the classical Gaussian homogeneous hidden Markov models (GH-HMMs), also called Gaussian homogeneous state space models. In GH-PMMs, neither X nor Y is necessarily Markovian, and the problem addressed is to find conditions on the GH-PMM parameters for X (or Y) to be Markovian. We give necessary and sufficient conditions for real-valued X and Y , and necessary conditions for the general multivariate case. The advantage of GH-PMMs over GH-HMMs is that they are more general and still allow various treatments such as smoothing, filtering or forecasting. A practical application of the proposed contributions is that when, for a given smoothing, filtering or forecasting problem, the specified conditions cannot be justified, GH-PMMs should be used rather than GH-HMMs.

Keywords— Kalman filter, Pairwise Markov models, hidden Markov models, forecasting, pairwise Kalman filter, marginal Markovianity

I. INTRODUCTION

Let us consider random processes $X_1^N = (X_1, \dots, X_N)$, $Y_1^N = (Y_1, \dots, Y_N)$, and $Z_1^N = (Z_1, \dots, Z_N)$, with $Z_n = (X_n, Y_n)$. We will use capital letters for designating random variables, and lowercase letters for designating their realisations. Different distributions will be written $p(x_1^N)$, $p(x_1^N, y_1^N)$, ... and $p(y_1^N | x_1^N)$, $p(x_1^N | y_1^N)$ for conditional distributions.

When considering the general problem of estimating X_1^N - or part of it - from Y_1^N - or part of it - Hidden Markov Models (HMMs) are very popular tools. Hundreds of papers are published every year. These include the pioneering papers [3], [45], and some recent books [5], [6], [11], [7], [8], [28], [47], [50]. In HMMs, the hidden sequence X_1^N is Markovian, and $p(y_1^N | x_1^N)$ are sufficiently simple to make $p(x_1^N | y_1^N)$ Markovian. The important points are that the Markovianity of $p(x_1^N | y_1^N)$ is sufficient to consider different Bayesian estimates of X_1^N from Y_1^N , and the markovianity of X_1^N is not necessary for that. This leads to pairwise Markov models (PMMs), in which only the Markovianity of the pair (X_1^N, Y_1^N) is assumed [42], [43]. In PMMs, $p(x_1^N | y_1^N)$ is Markovian and all Bayesian methods can be used in the same way as in HMMs.

PMMs are strictly more general than HMMs: an HMM is a PMM, while a PMM is not necessarily an HMM, because in a PMM X_1^N may or may not be Markov. This higher generality has proved more effective in many applications. In the case of discrete finite valued X_1^N , PMMs have mainly been applied to two problems: (i) image segmentation [1], [2], [9], [13], [14], [21], and (ii) multi-target tracking [30], [31], [32], [33], [34], [41], [48], [49]. Let us also mention promising introduction of variational inference in PMMs [35], [36]. Some theoretical results in the case of both X_1^N , Y_1^N finite valued can be seen in [24], [25], [26], [27]. In the case of X_1^N valued in \mathbb{R}^r and of Y_1^N valued in \mathbb{R}^s , it has been showed that Kalman filter remains possible and is as easy to construct in PMMs as in HMMs [16], [17], [18], [25], [26], [38], [39], [40], [43], [49], and to forecasting [12], [15]. It should be noted that if exact calculations are not feasible, particle filtering, similar to that used in HMMs, can be considered [29]. In all the applications studied, using PMMs instead of HMMs can be of significant interest. For example, in the case studied in [9], where the data clearly follows neither a PMM nor an HMM, using a PMM instead of an HMM halves the segmentation error ratio.

As the Markovianity of X_1^N is not necessary for its search from Y_1^N , the following question arises. How does the modelling power of PMMs decrease when X_1^N is assumed to be Markovian? How can the Markovianity of X_1^N be expressed in terms of constraints on the parameters of the PMM? Are these constraints justified in a given real application? These unnecessary implicit constraints have been little studied to date. To our knowledge, the only solutions to the problem in terms of necessary and sufficient conditions for the Markovianity of X_1^N exist in stationary invertible PMMs [42], [19], [20]. We study it here in the more general homogeneous Gaussian PMMs (HG-PMMs) defined below, which are not stationary in general. To our knowledge, no study relating to this case has been published.

More precisely, consider the following HG-PMM. The variables X_n, Y_n, Z_n take their values in \mathbb{R}^r , \mathbb{R}^s , and \mathbb{R}^{r+s} , respectively. Z_1^N is Markovian, Gaussian and homogeneous. For simplicity, we take all the variables $X_1, \dots, X_N, Y_1, \dots, Y_N$ to have zero mean. We call this model an homogeneous Gaussian pairwise Markov model (HG-PMM). It verifies

$$\begin{bmatrix} X_{n+1} \\ Y_{n+1} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} X_n \\ Y_n \end{bmatrix} + \begin{bmatrix} U_{n+1} \\ V_{n+1} \end{bmatrix}, \quad (1.1)$$

with A, B, C , and D , matrices of related sizes, and $\begin{bmatrix} U_2 \\ V_2 \end{bmatrix}, \dots, \begin{bmatrix} U_N \\ V_N \end{bmatrix}$ Gaussian, independent, zero mean, and equi-distributed.

Let us note

$$\begin{bmatrix} X_1 \\ Y_1 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_1 & \Lambda \\ \Lambda^T & \Sigma_2 \end{bmatrix}\right), \quad (1.2)$$

$$\begin{bmatrix} U_n \\ V_n \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_U & \Delta \\ \Delta^T & \Sigma_V \end{bmatrix}\right) \quad (1.3)$$

Our aim is to give conditions on $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ and $\begin{bmatrix} \Sigma_U & \Delta \\ \Delta^T & \Sigma_V \end{bmatrix}$ implied by the assumption that X_1^N is Markov for each $\begin{bmatrix} \Sigma_1 & \Lambda \\ \Lambda^T & \Sigma_2 \end{bmatrix}$. Due to the symmetry of the model, they will give conditions for Y_1^N to be Markov, and thus they will also give conditions for both X_1^N and Y_1^N to be Markov.

We present two original contributions:

- (i) in the case $s = r = 1$, where $A, B, C, D, \Sigma_U, \Sigma_V, \Sigma_1, \Sigma_2, \Lambda$, and Δ are real numbers, we show that X_1^N is Markov for each Σ_1, Σ_2 , and Λ , if and only if (iff) either $B = 0$, or $AD - BC = A\Delta - C\Sigma_U = B\Delta - D\Sigma_V = 0$;
- (ii) in the case of arbitrary s, r we show, under the additional hypothesis of existence of A^{-1} , that if X_1^N is Markov for each Σ_1, Σ_2 , and Λ , then

$$B(CA^{-1}B - D) = B(CA^{-1}\Sigma_U - \Delta^T) = 0. \quad (1.4)$$

The motivation in this particular situation is as follows. There are many problems, like filtering, smoothing, or forecasting, in which unobserved parts of X_1^N (or unobserved parts of Y_1^N) are estimated from observed parts of Y_1^N . The commonly used model is the Hidden State Space Markov Model, which we will refer to as GH-HMM in the following. It is defined by

$$X_{n+1} = FX_n + W_{n+1}; \quad (1.5)$$

$$Y_{n+1} = GX_{n+1} + R_{n+1}, \quad (1.6)$$

with $\begin{bmatrix} W_n \\ R_n \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_W & 0 \\ 0 & \Sigma_R \end{bmatrix}\right)$

Reporting (1.5) into (1.6), we see that GH-HMM is a particular GH-PMM:

$$\begin{bmatrix} X_{n+1} \\ Y_{n+1} \end{bmatrix} = \begin{bmatrix} F & 0 \\ GF & 0 \end{bmatrix} \begin{bmatrix} X_n \\ Y_n \end{bmatrix} + \begin{bmatrix} W_{n+1} \\ GW_{n+1} + R_{n+1} \end{bmatrix}, \quad (1.7)$$

and thus

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} F & 0 \\ GF & 0 \end{bmatrix}, \quad (1.8)$$

$$\begin{bmatrix} \Sigma_U & \Delta \\ \Delta^T & \Sigma_V \end{bmatrix} = \begin{bmatrix} \Sigma_W & G\Sigma_W G^T \\ G\Sigma_W & G(\Sigma_W G^T + \Sigma_R) \end{bmatrix} \quad (1.9)$$

We see that PMM (1.1) is defined by eight matrices, while GH-HMM (1.5)-(1.6) is defined by four matrices, so that, *a priori*, the simplification is quite crude.

Note that we have four families of HG-PMMs of increasing generality:

- (F1) the classical HG-HMMs (1.4)-(1.5), included in the family of HG-PMMs obtained by setting $B = D = 0$;
- (F2) an extension of the classical HG-HMMs (1.4)-(1.5), obtained by setting $B = 0$. In these models X_1^N is Markov, but the components of Y_1^N can be correlated conditionally on X_1^N . In signal processing terminology, we say that “the measurement noise is correlated”;
- (F3) the extension of (F2) to all HG-PMMs with X_1^N Markov (with $B = 0$ or not);
- (F4) HG-PMMs.

Our goal is to characterize (F3) in terms of constraints on $A, B, C, D, \Sigma_U, \Sigma_V$, and Δ . That is, we answer the question “what conditions on $A, B, C, D, \Sigma_U, \Sigma_V$, and Δ are equivalent to (or, at least, implied by) the Markovianity of X_1^N for each Σ_1, Σ_2 , and Λ ?”

Note the wide variety of possible applications of PMMs (1.1) in general. We have the classic filtering, smoothing, and forecasting of unobserved components of X_1^N from observed components of Y_1^N . We can also mention the smoothing and forecasting of unobserved components of Y_1^N from its observed ones. In this case, X_1^N is an additional latent process and can be taken among any process correlated with the Y_1^N under consideration.

Remark that in classical GH-HMMs, there is an asymmetry between X_1^N and Y_1^N . In many potential applications, there is a symmetry and therefore no reason to consider X_1^N as Markovian and Y_1^N as non-Markovian, as is usually the case. There are countless examples in all fields: any pair of two time series correlated with each other is one of them. For example, in economics, suppose X_1^N is unemployment and Y_1^N is inflation. We may want to estimate X_1^N from Y_1^N , but we may also want to estimate Y_1^N from X_1^N . In signal processing terminology, Y_1^N can be thought of as a noisy version of X_1^N , but X_1^N can also be thought of as a noisy version of Y_1^N . PMMs are preferable to HMMs in such situations because they provide a single model, simultaneously extending the two different particular HMMs.

The paper is organised as follows. The next section is devoted to the real case: the variables X_n, Y_n take their values in \mathbb{R} . Section III deals with the general multivariate case, and the last section IV contains conclusions and prospects.

II. MARGINAL MARKOVIANITY IN REAL VALUED PROCESSES

Let $X_1^N = (X_1, \dots, X_N)$, $Y_1^N = (Y_1, \dots, Y_N)$, and $Z_1^N = (Z_1, \dots, Z_N)$, with $Z_n = (X_n, Y_n)$. Variables X_n, Y_n take their values in \mathbb{R} , Z_1^N is Markov, Gaussian, homogeneous, and zero mean: $E[X_n] = E[Y_n] = 0$ for $n = 1, \dots, N$. To distinguish this particular case from the multivariate case considered in the next section, let us consider the following notations:

$$\begin{bmatrix} X_{n+1} \\ Y_{n+1} \end{bmatrix} = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \begin{bmatrix} X_n \\ Y_n \end{bmatrix} + \begin{bmatrix} U_{n+1} \\ V_{n+1} \end{bmatrix} \text{ for } n = 1, \dots, N-1; \quad (2.1)$$

$$\begin{bmatrix} X_1 \\ Y_1 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \tau \\ \tau & \sigma_2^2 \end{bmatrix}\right); \quad (2.2)$$

$$\begin{bmatrix} U_{n+1} \\ V_{n+1} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_U^2 & \rho \\ \rho & \sigma_V^2 \end{bmatrix}\right) \text{ for } n = 1, \dots, N-1, \quad (2.3)$$

where $\alpha, \beta, \gamma, \delta, \sigma_U^2, \sigma_V^2, \rho, \sigma_1^2, \sigma_2^2, \tau$ are real numbers. The problem we consider is therefore to give necessary and sufficient conditions (NSCs) on $\alpha, \beta, \gamma, \delta, \sigma_U^2, \sigma_V^2, \rho$ so that X_1^N is Markov for each $\sigma_1^2, \sigma_2^2, \tau$.

We will use the following

Lemma 1

Let Z_1^3 verify (2.1)-(2.3) for $n = 1$, and

$$X_3 = \alpha_1 X_2 + \beta_1 Y_2 + U_3^*, \quad (2.4)$$

with U_3^* independent from (X_1, Y_1, X_2, Y_2) , and

$$E[U_3^*] = 0 \quad (2.5)$$

Then X_1^3 is Markov for each $\sigma_1^2, \sigma_2^2, \tau$, iff either

$$\beta_1 = 0, \quad (2.6)$$

or

$$\det \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} = \det \begin{bmatrix} \alpha & \sigma_U^2 \\ \gamma & \rho \end{bmatrix} = \det \begin{bmatrix} \beta & \sigma_V^2 \\ \delta & \rho \end{bmatrix} = 0. \quad (2.7)$$

Before developing the proof, let us note the following point. In the GH-PMMs considered in the paper, we have $\beta_1 = \beta$; however, in the proof of the theorem below we will use this Lemma twice: once for $\beta_1 = \beta$, and once for $\beta_1 \neq \beta$.

Proof

According to (2.1), (2.4), we have

$$X_3 = \alpha_1 X_2 + \beta_1 Y_2 + U_3^* = \alpha_1(\alpha X_1 + \beta Y_1 + U_2) + \beta_1(\gamma X_1 + \delta Y_1 + V_2) + U_3^*.$$

Thus

$$X_2 = \alpha X_1 + \beta Y_1 + U_2; \quad (2.8)$$

$$X_3 = (\alpha_1 \alpha + \beta_1 \gamma) X_1 + (\alpha_1 \beta + \beta_1 \delta) Y_1 + \alpha_1 U_2 + \beta_1 V_2 + U_3^* \quad (2.9)$$

which implies

$$E(X_1 X_2) = \alpha \sigma_1^2 + \beta \tau; \quad (2.10)$$

$$E(X_2 X_3) = \alpha(\alpha_1 \alpha + \beta_1 \gamma) \sigma_1^2 + \beta(\alpha_1 \beta + \beta_1 \delta) \sigma_2^2 + (\alpha \alpha_1 \beta + \alpha \beta_1 \delta + \beta \alpha_1 \alpha + \beta \beta_1 \gamma) \tau + \alpha_1 \sigma_U^2 + \beta_1 \rho \quad (2.11)$$

$$E(X_2 X_2) = \alpha^2 \sigma_1^2 + \beta^2 \sigma_2^2 + 2\alpha\beta\tau + \sigma_U^2; \quad (2.12)$$

$$E(X_1 X_3) = (\alpha_1 \alpha + \beta_1 \gamma) \sigma_1^2 + (\alpha_1 \beta + \beta_1 \delta) \tau. \quad (2.13)$$

X_1^3 is Markov for each σ_1^2 , σ_2^2 , and τ , if and only if $Cov(X_1, X_2)Cov(X_2, X_3) = Cov(X_1, X_3)Var(X_2)$, which is equivalent to $E(X_1 X_2)E(X_2 X_3) = E(X_1 X_3)E(X_2 X_2)$.

$$(2.14)$$

According to (2.10)-(2.13), (2.14) is written

$$[\alpha \sigma_1^2 + \beta \tau][\alpha(\alpha_1 \alpha + \beta_1 \gamma) \sigma_1^2 + \beta(\alpha_1 \beta + \beta_1 \delta) \sigma_2^2 + (\alpha \alpha_1 \beta + \alpha \beta_1 \delta + \beta \alpha_1 \alpha + \beta \beta_1 \gamma) \tau + \alpha_1 \sigma_U^2 + \beta_1 \rho] = [(\alpha_1 \alpha + \beta_1 \gamma) \sigma_1^2 + (\alpha_1 \beta + \beta_1 \delta) \tau][\alpha^2 \sigma_1^2 + \beta^2 \sigma_2^2 + 2\alpha \beta \tau + \sigma_U^2]. \quad (2.15)$$

Before developing (2.15), let us show that α_1 plays no role. Let us see what are terms products by α_1 . On the l.h.s. we have

$$\alpha_1 [\alpha \sigma_1^2 + \beta \tau][\alpha \sigma_1^2 + \beta \sigma_2^2 + \alpha \beta \tau + \beta \alpha \tau + \sigma_U^2];$$

and on the right-hand side we have

$$\alpha_1 [\alpha \sigma_1^2 + \beta \tau][\alpha^2 \sigma_1^2 + \beta^2 \sigma_2^2 + 2\alpha \beta \tau + \sigma_U^2].$$

As they are equal, we can remove them from (2.15), and (2.15) becomes

$$\beta_1 [\alpha \sigma_1^2 + \beta \tau][\gamma \alpha \sigma_1^2 + \delta \beta \sigma_2^2 + \alpha \delta \tau + \beta \gamma \tau + \rho] = \beta_1 [\gamma \sigma_1^2 + \delta \tau][\alpha^2 \sigma_1^2 + \beta^2 \sigma_2^2 + 2\alpha \beta \tau + \sigma_U^2]. \quad (2.16)$$

Thus it is equivalent to either $\beta_1 = 0$, or $\beta_1 \neq 0$ and

$$[\alpha \sigma_1^2 + \beta \tau][\gamma \alpha \sigma_1^2 + \delta \beta \sigma_2^2 + \alpha \delta \tau + \beta \gamma \tau + \rho] = [\gamma \sigma_1^2 + \delta \tau][\alpha^2 \sigma_1^2 + \beta^2 \sigma_2^2 + 2\alpha \beta \tau + \sigma_U^2]$$

Developing gives

$$\delta \beta \sigma_2^2 \alpha \sigma_1^2 + \rho \alpha \sigma_1^2 + \beta \gamma \tau \beta \tau + \rho \beta \tau = \beta^2 \sigma_2^2 \gamma \sigma_1^2 + \sigma_U^2 \gamma \sigma_1^2 + \alpha \beta \tau \delta \tau + \sigma_U^2 \delta \tau, \quad (2.17)$$

equivalent to

$$\beta(\delta \alpha - \beta \gamma) \sigma_1^2 \sigma_2^2 + (\rho \alpha - \sigma_U^2 \gamma) \sigma_1^2 + \beta(\beta \gamma - \alpha \delta) \tau \tau + (\rho \beta - \sigma_U^2 \delta) \tau = 0 \quad (2.18)$$

Since (2.18) is true for every σ_1^2 , σ_2^2 , τ , it is equivalent to

$$\beta(\delta \alpha - \beta \gamma) = \rho \alpha - \sigma_U^2 \gamma = \beta(\beta \gamma - \alpha \delta) = \rho \beta - \sigma_U^2 \delta = 0,$$

which is (2.7) since $\beta_1 \neq 0$, and ends the proof. ■

Theorem

Let (X_1^N, Y_1^N) be a HG-PMM (2.1)-(2.2). Then

(1) For $N \geq 3$, X_1^N is Markov for each σ_1^2 , σ_2^2 , τ , iff either

$$\beta = 0, \text{ or} \quad (2.19)$$

$$\det \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} = \det \begin{bmatrix} \alpha & \sigma_U^2 \\ \gamma & \rho \end{bmatrix} = \det \begin{bmatrix} \beta & \sigma_U^2 \\ \delta & \rho \end{bmatrix} = 0; \quad (2.20)$$

(2) Under (2.19), Markovian X_1^N is homogeneous and verifies:

$$X_1 \sim N(0, \sigma_1^2); \quad (2.21)$$

$$X_{n+1} = \alpha X_n + U_{n+1}, \quad (2.22)$$

and under (2.20) with $\beta \neq 0$, Markovian X_1^N is homogeneous and verifies:

$$X_{n+1} = (\alpha + \delta) X_n + W_{n+1}^X; \quad (2.23)$$

$$E[W_{n+1}^X W_{n+1}^X] = (1 - \delta^2) \sigma_U^2 + \beta^2 \sigma_V^2, \quad (2.24)$$

with W_2^X, \dots, W_N^X zero mean and independent.

Proof

(1) If X_1^N is Markov then X_1^3 is Markov, and we have (2.19) or (2.20) by virtue of Lemma 1 applied to $\beta_1 = \beta$.

Conversely, assume (2.19) or (2.20). For $\beta = 0$ we have $X_{n+1} = \alpha X_n + U_{n+1}$, and thus it appears directly that X_1^N is Markov and verifies (2.22). Assume (2.20) with $\beta \neq 0$. Note that X_1^N is Markovian iff for each $n = 2, \dots, N-1$, the triplet (X_{n-k}, X_n, X_{n+1}) is Markovian for $k = 1, \dots, n-1$. Indeed, $p(x_{n+1} | x_1^n) = p(x_{n+1} | x_n)$ iff X_{n+1} and X_1^{n-1} are independent conditionally on X_n and, in

Gaussian case we are, this is the case iff (X_{n-k}, X_n, X_{n+1}) is Markov for $1 \leq k \leq n-1$. Let $N-1 \geq n \geq 2$. We will show that Markovianity of (X_{n-k}, X_n, X_{n+1}) (for $2 \leq k \leq n-1$) implies Markovianity of $(X_{n-k-1}, X_n, X_{n+1})$. As this is verified for $k=1$, this will be verified for each $1 \leq k \leq n-1$, implying Markovianity of X_1^N .

Let us set

$$\begin{bmatrix} X_{n+1} \\ Y_{n+1} \end{bmatrix} = \begin{bmatrix} \alpha^* & \beta^* \\ \gamma^* & \delta^* \end{bmatrix} \begin{bmatrix} X_{n-k} \\ Y_{n-k} \end{bmatrix} + \begin{bmatrix} U_{n+1}^* \\ V_{n+1}^* \end{bmatrix}, \quad (2.25)$$

$$\begin{bmatrix} X_{n+1} \\ Y_{n+1} \end{bmatrix} = \begin{bmatrix} \alpha^{**} & \beta^{**} \\ \gamma^{**} & \delta^{**} \end{bmatrix} \begin{bmatrix} X_{n-k-1} \\ Y_{n-k-1} \end{bmatrix} + \begin{bmatrix} U_{n+1}^{**} \\ V_{n+1}^{**} \end{bmatrix} \quad (2.26)$$

Let us assume that (X_{n-k}, X_n, X_{n+1}) is Markovian. Let's apply Lemma 1 with $\alpha, \beta, \gamma, \delta, \sigma_U^2$, and ρ replaced by $\alpha^*, \beta^*, \gamma^*, \delta^*, \sigma_U^{*2}$, and ρ^* , and β_1 replaced by β . As $\beta \neq 0$, Markovianity of (X_{n-k}, X_n, X_{n+1}) is equivalent to

$$\alpha^* \delta^* = \beta^* \gamma^*; \quad (2.27)$$

$$\alpha^* \rho^* = \gamma^* \sigma_U^{*2}. \quad (2.28)$$

The aim is to show that $(X_{n-k-1}, X_n, X_{n+1})$ is Markovian, which is equivalent to

$$\alpha^{**} \delta^{**} = \beta^{**} \gamma^{**}; \quad (2.29)$$

$$\alpha^{**} \rho^{**} = \gamma^{**} \sigma_U^{**2}. \quad (2.30)$$

Since

$$\begin{bmatrix} X_{n-k} \\ Y_{n-k} \end{bmatrix} = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \begin{bmatrix} X_{n-k-1} \\ Y_{n-k-1} \end{bmatrix} + \begin{bmatrix} U_{n-k} \\ V_{n-k} \end{bmatrix}, \quad (2.31)$$

(2.5) gives

$$\begin{aligned} \begin{bmatrix} X_{n+1} \\ Y_{n+1} \end{bmatrix} &= \begin{bmatrix} \alpha^* & \beta^* \\ \gamma^* & \delta^* \end{bmatrix} \left[\begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \begin{bmatrix} X_{n-k-1} \\ Y_{n-k-1} \end{bmatrix} + \begin{bmatrix} U_{n-k} \\ V_{n-k} \end{bmatrix} \right] + \begin{bmatrix} U_{n+1}^* \\ V_{n+1}^* \end{bmatrix} = \\ \begin{bmatrix} \alpha^* & \beta^* \\ \gamma^* & \delta^* \end{bmatrix} \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \begin{bmatrix} X_{n-k-1} \\ Y_{n-k-1} \end{bmatrix} + \begin{bmatrix} \alpha^* U_{n-k} + \beta^* V_{n-k} + U_{n+1}^* \\ \gamma^* U_{n-k} + \delta^* V_{n-k} + V_{n+1}^* \end{bmatrix} = \\ \begin{bmatrix} \alpha^{**} & \beta^{**} \\ \gamma^{**} & \delta^{**} \end{bmatrix} \begin{bmatrix} X_{n-k-1} \\ Y_{n-k-1} \end{bmatrix} + \begin{bmatrix} U_{n+1}^{**} \\ V_{n+1}^{**} \end{bmatrix} \end{aligned} \quad (2.32)$$

Then (2.29) is immediate since it is equivalent to $\det \begin{bmatrix} \alpha^{**} & \beta^{**} \\ \gamma^{**} & \delta^{**} \end{bmatrix} = 0$, which is true since

$$\det \begin{bmatrix} \alpha^{**} & \beta^{**} \\ \gamma^{**} & \delta^{**} \end{bmatrix} = \det \begin{bmatrix} \alpha^* & \beta^* \\ \gamma^* & \delta^* \end{bmatrix} \det \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} = 0.$$

It remains to show (2.30). According to (2.32):

$$\alpha^{**} = \alpha^* \alpha + \beta^* \gamma; \quad (2.33)$$

$$\gamma^{**} = \gamma^* \alpha + \delta^* \gamma; \quad (2.34)$$

$$\rho^{**} = E[(\alpha^* U_{n-k} + \beta^* V_{n-k} + U_{n+1}^*)(\gamma^* U_{n-k} + \delta^* V_{n-k} + V_{n+1}^*)] = \alpha^* \gamma^* \sigma_U^2 + \alpha^* \delta^* \rho + \beta^* \gamma^* \rho + \beta^* \delta^* \sigma_V^2 + \rho^*, \quad (2.35)$$

$$\sigma_U^{**2} = E[(\alpha^* U_{n-k} + \beta^* V_{n-k} + U_{n+1}^*)(\alpha^* U_{n-k} + \beta^* V_{n-k} + U_{n+1}^*)] = \alpha^{*2} \sigma_U^2 + 2\alpha^* \beta^* \rho + \beta^{*2} \sigma_V^2 + \sigma_U^{*2} \quad (2.36)$$

Then (2.30) is written

$$(\alpha^* \alpha + \alpha^* \gamma)(\alpha^* \gamma^* \sigma_U^2 + \alpha^* \delta^* \rho + \beta^* \gamma^* \rho + \beta^* \delta^* \sigma_V^2 + \rho^*) = (\gamma^* \alpha + \gamma^* \gamma)(\alpha^{*2} \sigma_U^2 + 2\alpha^* \beta^* \rho + \beta^{*2} \sigma_V^2 + \sigma_U^{*2}),$$

or equivalently

$$\begin{aligned} &\alpha^* \alpha \gamma^* \sigma_U^2 + \alpha^* \alpha \delta^* \rho + \alpha^* \alpha \beta^* \gamma^* \rho + \alpha^* \alpha \beta^* \delta^* \sigma_V^2 + \alpha^* \alpha \rho^* + \\ &\alpha^* \gamma^* \alpha \gamma^* \sigma_U^2 + \alpha^* \gamma^* \alpha \delta^* \rho + \alpha^* \gamma^* \beta^* \gamma^* \rho + \alpha^* \gamma^* \beta^* \delta^* \sigma_V^2 + \alpha^* \gamma^* \rho^* = \\ &\gamma^* \alpha \alpha^{*2} \sigma_U^2 + \gamma^* \alpha 2\alpha^* \beta^* \rho + \gamma^* \alpha \beta^{*2} \sigma_V^2 + \gamma^* \alpha \sigma_U^{*2} + \gamma^* \gamma \alpha^{*2} \sigma_U^2 + 2\gamma^* \gamma \alpha^* \beta^* \rho + \gamma^* \gamma \beta^{*2} \sigma_V^2 + \gamma^* \gamma \sigma_U^{*2} \end{aligned}$$

By rearranging products in alphabetical order and placing starred terms before non-starred terms, we gets:

$$\alpha^{*2} \gamma^* \alpha \sigma_U^2 + \alpha^{*2} \delta^* \alpha \rho + \alpha^* \beta^* \gamma^* \alpha \rho + \alpha^* \beta^* \delta^* \alpha \sigma_V^2 + \alpha^* \rho^* \alpha + \alpha^{*2} \gamma^* \gamma \sigma_U^2 + \alpha^{*2} \delta^* \gamma \rho + \alpha^* \beta^* \gamma^* \gamma \rho + \alpha^* \beta^* \delta^* \gamma \sigma_V^2 + \alpha^* \rho^* \gamma =$$

$$\alpha^* \gamma^* \alpha \sigma_U^2 + 2\alpha^* \beta^* \gamma^* \alpha \rho + \beta^* \gamma^* \alpha \sigma_V^2 + \sigma_U^2 \gamma^* \alpha + \alpha^* \gamma^* \gamma \sigma_U^2 + 2\alpha^* \beta^* \gamma^* \gamma \rho + \beta^* \gamma^* \gamma \sigma_V^2 + \sigma_U^2 \gamma^* \gamma \quad (3.37)$$

Replacing $\gamma^* \sigma_U^2$ with $\alpha^* \rho^*$ and $\alpha^* \delta^*$ with $\beta^* \gamma^*$ we find that (3.37) is verified, which completes the proof of (1). To show (2) let us temporarily assume that $N = 3$. According to (2.1), we have

$$X_2 = \alpha X_1 + \beta Y_1 + U_2 \quad (2.38)$$

$$X_3 = (\alpha^2 + \beta \gamma) X_1 + (\alpha \beta + \beta \delta) Y_1 + \alpha U_2 + \beta V_2 + U_3 \quad (2.39)$$

Using $\beta \gamma = \alpha \delta$, (2.38) and (2.39) imply

$$X_3 = \alpha(\alpha + \delta) X_1 + \beta(\alpha + \delta) Y_1 + \alpha U_2 + \beta V_2 + U_3,$$

which gives, using $\beta \rho = \delta \sigma_U^2$:

$$E(X_2 X_2) = \alpha^2 \sigma_1^2 + \beta^2 \sigma_2^2 + 2\alpha \beta \tau + \sigma_U^2;$$

$$E(X_2 X_3) = (\alpha + \delta)(\alpha^2 \sigma_1^2 + \beta^2 \sigma_2^2 + 2\alpha \beta \tau + \sigma_U^2),$$

and thus

$$X_3 = \frac{E(X_2 X_3)}{E(X_2 X_2)} X_2 + W_3 = \left(\frac{(\alpha + \delta)(\alpha^2 \sigma_1^2 + \beta^2 \sigma_2^2 + 2\alpha \beta \tau + \sigma_U^2)}{\alpha^2 \sigma_1^2 + \beta^2 \sigma_2^2 + 2\alpha \beta \tau + \sigma_U^2} \right) X_2 + W_3 = (\alpha + \delta) X_2 + W_3,$$

with

$$E[X_3 W_3] = 0,$$

$$E[W_3 W_3] = E[X_3 X_3] - (\alpha + \delta)^2 E[X_2 X_2] = (1 - \delta^2) \sigma_U^2 + \beta^2 \sigma_V^2.$$

We see that neither $\frac{E(X_2 X_3)}{E(X_2 X_2)}$ nor $E[W_3 W_3]$ depend on $\sigma_1^2, \sigma_2^2, \tau$, so we can replace $n = 2$ by any $n = 3, \dots, N - 1$, which gives (2.23)-(2.24) and completes the proof. ■

Remark 2.1

In certain real-life situations, the hidden X_1^N can be Markovian. According to the theorem there are, in the context of GH-PMs, four families with Markov X_1^N :

(F1) Classic GH-HMMs form the poorest family of models, since the distributions $p(x_{n+1}, y_{n+1} | x_n, y_n)$ depend on four parameters;

(F2) GH-PMs with $\beta = 0, \delta = 0$ contain (F1) and are strictly richer; indeed, $p(x_{n+1}, y_{n+1} | x_n, y_n)$ depend on five parameters;

(F3) GH-PMs with $\beta = 0$. This family obviously contain (F2) and is obviously strictly richer. In particular, $p(x_{n+1}, y_{n+1} | x_n, y_n)$ depend on six parameters;

(F4) GH-PMs with X_1^N Markov and $\beta \neq 0$. In (F4) $p(x_{n+1}, y_{n+1} | x_n, y_n)$ are defined by five free parameters (there So, if we know *a priori*, in a given case, that X_1^N is Markov, family (F3) seems the most interesting as it depends on the largest number of parameters. On the other hand, if β is not close to zero, (F4) might be better suited.

We illustrate the mutual positioning of (F1), ..., (F4) in Fig. 1.



Fig. 1 Four families of GH-PMs with X_1^N Markov. F1: classic GH-HMMs (four free parameters); F2: GH-PMs with $\beta = 0, \delta = 0$ (five free parameters); F3: GH-PMs with $\beta = 0$ (six free parameters); F4: GH-PMs with $\beta \neq 0$ (five free parameters).

Example 2.1

As mentioned in Remark 2.1, family (F3) seems to be the most interesting, since it is defined by the largest number of parameters. However, (F4) may be of interest for at least two reasons. Stationary GH-PMs were recently applied to forecasting in [12] and it turned out, in the real data cases considered, that β can be significantly different from zero. The second reason is the form of (2.23) and (2.24). They mean that if $\beta^2 \sigma_V^2 < \delta^2 \sigma_U^2$, we have $(1 - \delta^2) \sigma_U^2 + \beta^2 \sigma_V^2 < \sigma_U^2$, so that the Markov X_1^N is less noisy in (F4) than in

(F3). This can be important for forecasting, where the quality of forecasting x_{n+k} from y_1^n is measured by the variance of $p(x_{n+k}|y_1^n)$.

Since in a PMM X_1^N and Y_1^N play symmetrical roles, the theorem also gives NSCs for Markovianity of Y_1^N , and NSCs for Markovianity of both X_1^N and Y_1^N . Writing (2.2) in the form

$$\begin{bmatrix} Y_{n+1} \\ X_{n+1} \end{bmatrix} = \begin{bmatrix} \delta & \gamma \\ \beta & \alpha \end{bmatrix} \begin{bmatrix} Y_n \\ X_n \end{bmatrix} + \begin{bmatrix} V_{n+1} \\ U_{n+1} \end{bmatrix}, \quad (2.40)$$

We have

Corollary 1

Under the assumptions of theorem:

(1) Y_1^N is Markov for each $\sigma_1^2, \sigma_2^2, \tau$, iff one of two conditions :

(i) $\delta = 0$; (2.41)

(ii) $\det \begin{bmatrix} \delta & \gamma \\ \beta & \alpha \end{bmatrix} = \det \begin{bmatrix} \delta & \sigma_V^2 \\ \beta & \rho \end{bmatrix} = \det \begin{bmatrix} \gamma & \sigma_V^2 \\ \alpha & \rho \end{bmatrix} = 0$ (2.42)

is verified;

(2) Since (2.7), Markov Y_1^N is homogeneous and verify:

$$Y_1 \sim N(0, \sigma_2^2), \quad (2.43)$$

$$Y_{n+1} = (\alpha + \delta)Y_n + W_{n+1}^Y, \quad (2.44)$$

$$E[W_{n+1}^Y W_{n+1}^Y] = (1 - \alpha^2)\sigma_V^2 + \gamma^2\sigma_U^2 \quad (2.45)$$

Corollary 2

Under the assumptions of the theorem:

1. Both X_1^N and Y_1^N are Markovian iff one of the four conditions:

(C1) $\det \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} = \det \begin{bmatrix} \alpha & \sigma_U^2 \\ \gamma & \rho \end{bmatrix} = \det \begin{bmatrix} \delta & \sigma_V^2 \\ \beta & \rho \end{bmatrix} = 0,$

and $\beta \neq 0, \gamma \neq 0$; (2.46)

(C2) $\beta = 0$ and (2.42) ; (2.47)

(C3) $\gamma = 0$ and (2.20) ; (2.48)

(C4) $\beta = 0$ and $\delta = 0$. (2.49)

2. Under (2.46), we have

$$X_{n+1} = (\alpha + \delta)X_n + W_{n+1}^X; \quad (2.50)$$

$$E[W_{n+1}^X W_{n+1}^X] = \sigma_U^2; \quad (2.51)$$

$$Y_{n+1} = (\alpha + \delta)Y_n + W_{n+1}^Y; \quad (2.52)$$

$$E[W_{n+1}^Y W_{n+1}^Y] = \sigma_V^2; \quad (2.53)$$

Proof

As X_1^N is Markov, we have

$$X_{n+1} = (\alpha + \delta)X_n + W_{n+1}^X; \quad (2.54)$$

$$E[W_{n+1}^X W_{n+1}^X] = \sigma_U^2 - \delta^2\sigma_U^2 + \beta^2\sigma_V^2; \quad (2.55)$$

Furthermore, Markovianity of Y_1^N implies $\sigma_V^2 = \frac{\delta}{\beta}\rho$, and Markovianity of X_1^N implies $\sigma_U^2 = \frac{\beta}{\delta}\rho$. Setting them into (2.55) gives

$$E[W_{n+1}^X W_{n+1}^X] = \sigma_U^2 - \delta^2\sigma_U^2 + \beta^2\sigma_V^2 = \sigma_U^2 - \delta\beta\rho + \beta\delta\rho = \sigma_U^2.$$

Similarly, $E[W_{n+1}^Y W_{n+1}^Y] = \sigma_V^2$, which completes the proof. ■

Example 2.2

Consider the case of both X_1^N and Y_1^N Markovian with $\beta = 0$ and $\gamma = 0$. We have $X_{n+1} = \alpha X_n + U_{n+1}$, $Y_{n+1} = \delta Y_n + V_{n+1}$, and correlations $E[X_n Y_n]$ are given recursively by $E[X_1 Y_1] = \tau$, $E[X_{n+1} Y_{n+1}] = \alpha \delta E[X_n Y_n] + \rho$. Moreover, we have $(x_{n+1} | x_n, y_n) = p(x_{n+1} | x_n)$, $p(y_{n+1} | x_n, y_n) = p(y_{n+1} | y_n)$, but $p(x_{n+1} | x_n, y_n, y_{n+1}) \neq p(x_{n+1} | x_n)$ and $p(y_{n+1} | x_n, y_n, x_{n+1}) \neq p(y_{n+1} | y_n)$.

III. MARGINAL MARKOVIANITY IN MULTIVARIATE CASE

In this section we consider the general multi-variate case, valid for any dimensions r, s of state space and observation space, respectively. Unlike the previous case $r = s = 1$, conditions for the Markovianity of X_1^N we present, valid in the case of invertible A , are only necessary. The interest of the results presented in filtering, smoothing, or forecasting, is similar to its interest in the one-dimensional case of the previous section: it shows what model restrictions are imposed by unnecessarily assuming the Markovianity of the hidden chain. Note that insofar as the proposed conditions are only necessary, there may be additional parameter restrictions not specified by the proposal below, which reinforce its interest.

Proposition 1

Let us consider random processes X_1^N, Y_1^N , and Z_1^N verifying (1.1)-(1.2), with A invertible. Markovianity of X_1^N implies

$$B(CA^{-1}B - D) = B(CA^{-1}\Sigma_U - \Delta^T) = 0; \quad (3.1)$$

Proof

Let us first show that (3.1) is equivalent to the Markovianity of X_{n-1}^{n+1} for every $n = 2, \dots, N-1$ and every $\Sigma_1, \Sigma_2, \Lambda$. Due to homogeneity, it suffices to show this for $n = 2$. The Markovianity of X_1^3 is equivalent to

$$E[X_3 X_1^T] = E[X_3 X_2^T] [E[X_2 X_2^T]]^{-1} E[X_2 X_1^T] \quad (3.2)$$

According to (1.1) we have

$$\begin{aligned} X_2 &= AX_1 + BY_1 + U_2, \\ X_3 &= (A^2 + BC)X_1 + (AB + BD)Y_1 + AU_2 + BV_2 + U_3, \end{aligned}$$

and thus

$$E[X_2 X_1^T] = A\Sigma_1 + B\Lambda^T; \quad (3.3)$$

$$E[X_2 X_2^T] = A\Sigma_1 A^T + B\Sigma_2 B^T + B\Lambda^T A^T + A\Lambda B^T + \Sigma_U; \quad (3.4)$$

$$E[X_3 X_2^T] = ((AA + BC)\Sigma_1 + (AB + BD)\Lambda^T)A^T + ((AB + BD)\Sigma_2 + (AA + BC)\Lambda)B^T + A\Sigma_U + B\Delta^T;$$

$$E[X_3 X_1^T] = (A^2 + BC)\Sigma_1 + (AB + BD)\Lambda^T \quad (3.6)$$

Reporting (3.3)-(3.6) into (3.2) we get

$$(A^2 + BC)\Sigma_1 + (AB + BD)\Lambda^T = [((AA + BC)\Sigma_1 + (AB + BD)\Lambda^T)A^T + \quad (3.7)$$

$$((AB + BD)\Sigma_2 + (AA + BC)\Lambda)B^T + A\Sigma_U + B\Delta^T] \times [A\Sigma_1 A^T + B\Sigma_2 B^T + B\Lambda^T A^T + A\Lambda B^T + \Sigma_U]^{-1} [A\Sigma_1 + B\Lambda^T]$$

(3.7) holds for every Λ ; by setting $\Lambda = 0$, this implies:

$$(A^2 + BC)\Sigma_1 = [((AA + BC)\Sigma_1)A^T + ((AB + BD)\Sigma_2)B^T + A\Sigma_U + B\Delta^T] \times [A\Sigma_1 A^T + B\Sigma_2 B^T + \Sigma_U]^{-1} [A\Sigma_1] \quad (3.8)$$

Multiplying both sides of (3.8) by $[A\Sigma_1]^{-1} A\Sigma_1 A^T + B\Sigma_2 B^T + \Sigma_U$, we obtain

$$(A + BCA^{-1})(A\Sigma_1 A^T + B\Sigma_2 B^T + \Sigma_U) = [(AA + BC)\Sigma_1 A^T + (AB + BD)\Sigma_2 B^T + A\Sigma_U + B\Delta^T] \quad (3.9)$$

Expanding (3.9) gives successively

$$AA\Sigma_1 A^T + AB\Sigma_2 B^T + A\Sigma_U + BCA^{-1}A\Sigma_1 A^T + BCA^{-1}B\Sigma_2 B^T + BCA^{-1}A\Sigma_U =$$

$$AA\Sigma_1 A^T + BC\Sigma_1 A^T + AB\Sigma_2 B^T + BD\Sigma_2 B^T + A\Sigma_U + B\Delta^T,$$

$$BCA^{-1}B\Sigma_2 B^T + BCA^{-1}A\Sigma_U = BD\Sigma_2 B^T + B\Delta^T,$$

and finally

$$B(CA^{-1}B - D)\Sigma_2 B^T + B(CA^{-1}\Sigma_U - \Delta^T) = 0 \quad (3.10)$$

Since (3.10) holds for all Σ_2 , it implies $B(\Delta^T - CA^{-1}\Sigma_U) = 0$.

Indeed, the result is obtained by considering any parametrized family $(\Sigma_2^\varepsilon)_{\varepsilon > 0}$ verifying $\Sigma_2^\varepsilon \xrightarrow{\varepsilon \rightarrow 0} 0$.

Finally, $B(D - CA^{-1}B) = 0$ is obtained applying Lemma 2 below to $\Sigma_2 = \Sigma$, F any line of $B(D - CA^{-1}B)$, and G any column of B^T , which completes the proof. ■

Lemma 2

Let $F, G \neq 0$ be elements of \mathbb{R}^s . If $F^T \Sigma G = 0$ for any positive definite matrix Σ of size $s \times s$, then $F = 0$.

Proof

Let $F^T = (f_1, \dots, f_s)$, $G = (g_1, \dots, g_s)$. Let $\Sigma = (\lambda_{ij})_{1 \leq i, j \leq s}$ be positive definite verifying $\lambda_{ij} \neq 0$ for all $1 \leq i, j \leq s$. Let $1 \leq j \leq s$ be such that $g_j \neq 0$. To show that all f_1, \dots, f_s are zero, assume that there exists $f_i \neq 0$. Let $\varepsilon > 0$, and $\lambda_{ij}^\varepsilon = \lambda_{ij} + \varepsilon$. Let Σ_ε have all terms equal to those in Σ , except λ_{ij} which is replaced by λ_{ij}^ε . We then have $F^T \Sigma_\varepsilon G = F^T \Sigma G + 2\varepsilon f_i g_j = 2\varepsilon f_i g_j \neq 0$. Since for $\varepsilon > 0$ sufficiently small Σ_ε is positive definite, this shows the impossibility of $F \neq 0$, which completes the proof. ■

We remark that for $r = s = 1$, $\alpha \neq 0$, and $\Sigma_U \neq 0$, condition (3.1) is condition (2.7) of the theorem.

Example 3.1

Let's recall, as an application example, the Kalman filter. It was proposed in [43], and taken up in [16], [17], [18], [32], [33], [34], [38], [39], [40] in PMMs defined as follows

$$\begin{bmatrix} X_{n+1} \\ Y_n \end{bmatrix} = \begin{bmatrix} A' & B' \\ C' & D' \end{bmatrix} \begin{bmatrix} X_n \\ Y_{n-1} \end{bmatrix} + \begin{bmatrix} U_n \\ V_n \end{bmatrix}, \quad (3.10)$$

Let us recall how ‘‘pairwise’’ Kalman filter (PKF) runs in PMM considered in the form (1.1). Both forms are equivalents; however, the form (1.1) is more concise and better suited for our purpose, since we wish to have a model symmetrical with respect to X_1^N and Y_1^N . The very first PKF formulated in (1.1) form was written in stationary PMM in [4], and taken up in [22], [23], [12]. PKF runs as follows:

Proposition 2

Let X_1^N, Y_1^N satisfy (1.1)-(1.3), and let $p(x_n | y_1^n) \sim N(M_n, \Gamma_n)$. Then

$$p(x_{n+1} | y_1^{n+1}) \sim N(M_{n+1}, \Gamma_{n+1}), \quad (3.11)$$

with

$$M_{n+1} = M_{n+1}^1 + \Sigma_{n+1}^2 [\Sigma_{n+1}^4]^{-1} (y_{n+1} - M_{n+1}^2); \quad (3.12)$$

$$\Gamma_{n+1} = \Sigma_{n+1}^1 - \Sigma_{n+1}^2 [\Sigma_{n+1}^4]^{-1} \Sigma_{n+1}^3, \quad (3.13)$$

where

$$\begin{bmatrix} M_{n+1}^1 \\ M_{n+1}^2 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} M_n \\ y_n \end{bmatrix}; \quad (3.14)$$

$$\begin{bmatrix} \Sigma_{n+1}^1 & \Sigma_{n+1}^2 \\ \Sigma_{n+1}^3 & \Sigma_{n+1}^4 \end{bmatrix} = \begin{bmatrix} A \Gamma_n A^T + \Sigma_U & A \Gamma_n C^T + \Delta \\ C \Gamma_n A^T + \Delta^T & C \Gamma_n C^T + \Sigma_V \end{bmatrix}. \quad (3.15)$$

Proof

It is sufficient to show that (3.14) and (3.15) are the mean and variance of the Gaussian $p(x_{n+1}, y_{n+1} | y_1^n)$. Indeed, since $p(x_{n+1} | y_1^{n+1}) = p(x_{n+1} | y_{n+1}, y_1^n)$, (3.12) and (3.13) are obtained from (3.14) and (3.15) by classical Gaussian conditioning. We use the total expectation formula: for every random vectors W, Q we have $E[W] = E[E[W|Q]]$. Applying it to the conditional expectation $E[\cdot | y_1^n]$, and vectors $W = Z_{n+1} = \begin{bmatrix} X_{n+1} \\ Y_{n+1} \end{bmatrix}$, $Q = X_n$, and setting $T = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$, we can write

$$E[Z_{n+1} | y_1^n] = E[E[Z_{n+1} | X_n, y_n] | y_1^n] = E[T \begin{bmatrix} X_n \\ y_n \end{bmatrix} | y_1^n] = T \begin{bmatrix} E[X_n | y_1^n] \\ y_n \end{bmatrix} = T \begin{bmatrix} M_n \\ y_n \end{bmatrix},$$

which is (3.14). Similarly

$$\begin{aligned} E[Z_{n+1} Z_{n+1}^T | y_1^n] &= E[E[Z_{n+1} Z_{n+1}^T | X_n, y_n] | y_1^n] = \\ &= E \left[T \begin{bmatrix} X_n X_n^T & X_n y_n^T \\ y_n X_n^T & y_n y_n^T \end{bmatrix} T^T + \begin{bmatrix} \Sigma_U & \Delta \\ \Delta^T & \Sigma_V \end{bmatrix} | y_1^n \right] = T \begin{bmatrix} \Gamma_n + M_n M_n^T & M_n y_n^T \\ y_n M_n^T & y_n y_n^T \end{bmatrix} T^T + \begin{bmatrix} \Sigma_U & \Delta \\ \Delta^T & \Sigma_V \end{bmatrix} \end{aligned}$$

Then

$$E[Z_{n+1} Z_{n+1}^T | y_1^n] - E[Z_{n+1} | y_1^n] [E[Z_{n+1} | y_1^n]]^T = T \begin{bmatrix} \Gamma_n + M_n M_n^T & M_n y_n^T \\ y_n M_n^T & y_n y_n^T \end{bmatrix} T^T + \begin{bmatrix} \Sigma_U & \Delta \\ \Delta^T & \Sigma_V \end{bmatrix} - T \begin{bmatrix} M_n \\ y_n \end{bmatrix} \begin{bmatrix} M_n \\ y_n \end{bmatrix}^T T^T =$$

$$\mathbf{T} \begin{bmatrix} \Gamma_n & 0 \\ 0 & 0 \end{bmatrix} \mathbf{T}^T + \begin{bmatrix} \Sigma_U & \Delta \\ \Delta^T & \Sigma_V \end{bmatrix} = \begin{bmatrix} A\Gamma_n A^T + \Sigma_U & A\Gamma_n C^T + \Delta \\ C\Gamma_n A^T + \Delta^T & C\Gamma_n C^T + \Sigma_V \end{bmatrix}$$

which completes the proof. ■

IV. CONCLUSIONS AND PERSPECTIVES

We considered Gaussian Homogeneous Pairwise Markov Models (GH-PMMs) $Z_1^N = (Z_1, \dots, Z_N)$, where for $n = 1, \dots, N$, $Z_n = (X_n, Y_n)$ and $X_1^N = (X_1, \dots, X_N)$, $Y_1^N = (Y_1, \dots, Y_N)$ are valued in \mathbb{R}^r , \mathbb{R}^s , respectively. In classic Gaussian Homogeneous Hidden Markov Models (GH-HMMs) both X_1^N and Z_1^N are Markovian, whereas in GH-PMMs only Z_1^N is assumed Markovian. In the case $r = s = 1$, we have provided necessary and sufficient conditions on the parameters defining $p(z_1^N)$, for X_1^N (or Y_1^N) to be Markovian. In the general case, we provided necessary conditions for X_1^N (or Y_1^N) to be Markovian. One possible application of the proposed results relates to the problem of estimating unobserved X_1^N from observed Y_1^N . Since GH-PMMs enable Bayesian estimation methods (filtering, smoothing, forecasting, ...) similar to those associated with GH-HMMs, a practical interest of the proposed results for real applications is that when, for the given problem, the specified conditions on the parameters cannot be justified, it is appropriate to use GH-PMMs rather than GH-HMMs.

Extending the results presented to the homogeneous non-Gaussian case or searching for sufficient – as weak as possible – conditions of marginal Markovianity in the general case are possible perspectives for further work. In particular, switching systems are of interest in numerous applications, among which tracking. In classical models, the computational complexity increases exponentially with the number of switches [37], [10], [46], among others. The problem can be solved by considering the Markovianity of the pair (Switches, Observations), resulting in the "Markov Conditional Switching Hidden Linear Models" (CMSHLMs) proposed in [44]. It would therefore be interesting to understand how the modelling power of general models - which also encompass the classical switching GH-HMMs - decreases when one assumes the latter marginal Markovianity

References

- [1] M. Ameur, N. Idrissi, and C. Daoui, Color images segmentation using Pairwise Markov Chain, International Conference on Multimedia Computing and Systems (ICMMCS2018), 2018.
- [2] A. K., Atiampo and G. L. Loum, Unsupervised image segmentation with Pairwise markov chains based on nonparametric estimation of copula using orthogonal polynomials, *International Journal of Image and Graphics*, vol. 16, no. 4, 2016.
- [3] L. E. Baum and T. Petrie, Statistical Inference for Probabilistic Functions of Finite state Markov Chains, *Annals of Mathematical Statistics*, 37, pp. 1554–1563, 1966.
- [4] D. Benboudjema, M. El Bouchraya Malainin, and W. Pieczynski, Exact Kalman filtering in pairwise Gaussian switching systems, Applied Stochastic Models and Data Analysis, ASMDA 2011, June 7-10, Rome, Italy, 2011.
- [5] R. Bhar and S. Hamori, *Hidden Markov models. Applications to Financial Economic*, Kluwer Academic Publishers, 2004.
- [6] N. Bouguila and M. Amayri, *Hidden Markov models and applications*, Springer, 2022.
- [7] O. Cappé, E. Moulines, and T. Ryden, *Inference in Hidden Markov Models; Springer Series in Statistics*, Springer, 2005.
- [8] J. Dannemann, *Inference for Hidden Markov Models and related Models*, Cuvillier Verlag, 2010.
- [9] S. Derrode, and W. Pieczynski, Signal and image segmentation using pairwise Markov chains, *IEEE Trans. on Signal Processing*, vol. 52, no. 9, 2004.
- [10] A. Doucet, N.J. Gordon, and V. Krishnamurthy, Particle filters for state estimation of jump Markov linear systems, *IEEE Trans. on Signal Processing*, vol. 49, no. 3, pp. 613-624, 2001.
- [11] P. Dymarski, *Hidden Markov models*, IntechOpen, 2011.
- [12] M. Escudier, C. Fernandes, I. Abdelkefi, and W. Pieczynski, Forecasting with Pairwise Gaussian Markov Models, 8th IEEE International Conference on Mathematics and Computers in Sciences and Industry (CMCSI 23), Athens, Greece, October 14-16, 2023.
- [13] H. Gangloff, K. Morales, and Y. Petetin, A general parametrization framework for Pairwise Markov models : an application to unsupervised image segmentation, IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2021), 2021.
- [14] H. Gangloff, K. Morales, and Y. Petetin, Deep parameterizations of pairwise and triplet Markov models for unsupervised classification of sequential data, *Computational Statistics and Data Analysis*, 180, 2023.
- [15] I. Gorynin, E. Monfrini, and W. Pieczynski, Pairwise Markov models for stock index forecasting. 25th European Signal Processing Conference, EUSIPCO 2017, pp. 2041-2045, 2017.
- [16] M. V. Kulikova, Gradient-based parameter estimation in pairwise linear Gaussian system, *IEEE Trans. on Automatic Control*, vol. 62, no. 3, pp. 1511–1517, 2017.
- [17] M. V. Kulikova, J. V. Tsyganova, and G. Yu. Kulikov, UD-based Pairwise and MIMO Kalman-like filtering for estimation of econometric model structures, *IEEE Trans. on Automatic Control*, vol. 65, no. 10, pp. 4472 - 4479, 2020.
- [18] M. V. Kulikova, J. V. Tsyganova, and G. Yu. Kulikov, SVD-based state and parameter estimation approach for generalized Kalman filtering with application to GARCH-in-Mean estimation, *Journal of Computational and Applied Mathematics*, vol. 387, 112487, 2021.
- [19] K. Kuljus and J. Lember, Pairwise Markov models and hybrid segmentation approach, *Methodology and Computing in Applied Probability*, May 2023.
- [20] P. Lanchantin, J. Lapuyade-Lahorgue and W. Pieczynski, Unsupervised segmentation of randomly switching data hidden with non-Gaussian correlated noise, *Signal Processing*, vol. 91, no. 2, pp. 163-175, 2011.
- [21] S. Le Cam, F. Salzenstein, and C. Collet, Fuzzy pairwise Markov chain to segment correlated noisy data, *Signal Processing*, vol. 88, no. 10, pp. 2526-2541, 2008.
- [22] F. Lehmann and W. Pieczynski, Suboptimal Kalman filtering in triplet Markov models using model order reduction, *IEEE Signal Processing Letters*, vol. 27, pp. 1100-1104, 2020.

- [23] F. Lehmann and W. Pieczynski, Reduced-dimension filtering in triplet Markov models, *IEEE Trans. on Automatic Control*, Vol. 67, No. 2, pp. 605-618, 2022.
- [24] J. Lember, H. Matzinger, J. Sova, and F. Zucca, Lower bounds for moments of global scores of pairwise Markov chains, *Stochastic Processes and their Applications*, vol. 128, no. 5, pp. 1678-1710, 2017.
- [25] J. Lember and J. Sova, Existence of infinite Viterbi paths for pairwise Markov models, *Stochastic Processes and their Applications*, vol. 130, issue 3, pp. 1388-1425, 2020.
- [26] J. Lember and J. Sova, Exponential forgetting of smoothing distributions for pairwise Markov models, *Electronic Journal of Probability*, 26, 2021.
- [27] Lember, Local Viterbi property in decoding, *Information and Inference: A Journal of the IMA*, vol. 12, no. 3, 2023.
- [28] J. Li and R. M. Gray, *Image segmentation and compression using Hidden Markov Models*, Springer Science, 2012.
- [29] J. Liu, C. Wang, W. Wang, and Z. Li, Particle probability hypothesis density filter based on pairwise Markov chains, *Algorithms*, vol. 12, no. 2, 2019.
- [30] J. Liu, C. Wang, W. Wang, SMC-CBMeMber filter based on pairwise Markov chains, *Xi Tong Gong Cheng Yu Dian Zi Ji Shu/Systems Engineering and Electronics*, vol. 41, no. 8, pp. 1686-1691, 2019.
- [31] J. Liu and C. Wang, Cardinalized probability hypothesis density filter based on pairwise Markov chains, *Xi Tong Gong Cheng Yu Dian Zi Ji Shu/Systems Engineering and Electronics*, vol. 41, no. 2, pp. 492-497, 2019.
- [32] R. Mahler, Tracking targets with Pairwise-Markov dynamics. *18th International Conference on Information Fusion*, 280-286, 2015.
- [33] R. Mahler, The pairwise-Markov Bernoulli filter, *IEEE Access*, 8, pp. 168229-168245, 2020.
- [34] R. Mahler, Bernoulli filters for multiple correlated sensors, *IEEE Access*, 9, pp. 2310 - 2316, 2021.
- [35] K. Morales and Y. Petetin, Variational Bayesian inference for pairwise Markov models. *IEEE Workshop on Statistical Signal Processing Proceedings*, pp. 251-255, 2023.
- [36] K. Morales, and Y. Petetin, "A probabilistic semi-supervised approach with triplet Markov chains", IEEE International Workshop on Machine Learning for Signal Processing (MLSP), 2023.
- [37] E. Masor, A. Averbuch, Y. Bar-Shalom, and J. Dayan, Interacting multiple model methods in target tracking: A survey, *IEEE Trans on Aerospace and Electronic Systems*, vol. 34, no. 1, pp. 103-123, 1998.
- [38] V. Nemesin and S. Derrode, Robust blind pairwise Kalman algorithms using QR decompositions, *IEEE Trans. Signal Processing.*, vol. 61, no. 1, pp. 5-9, 2013.
- [39] V. Nemesin and S. Derrode, Robust partial-learning in linear Gaussian systems, *IEEE Trans. on Automatic Control*, vol. 60, no. 9, pp. 2518-2523, 2015.
- [40] V. Nemesin, *Apprentissage non-supervisé dans les modèles linéaires gaussiens : Application à la biométrie dynamique de l'iris*, Presses Académiques Francophones, 2015.
- [41] Y. Petetin and F. Desbouvries, Bayesian multi-object filtering for pairwise Markov chains, *IEEE Trans. on Signal Processing*, vol. 61, no. 18, pp. 4481-4490, 2013.
- [42] W. Pieczynski, Pairwise Markov chains. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 634-639, 2003.
- [43] W. Pieczynski and F. Desbouvries, Kalman Filtering using Pairwise Gaussian Models, International Conference on Acoustics, Speech and Signal Processing (ICASSP), Hong-Kong, April 2003.
- [44] W. Pieczynski, Exact filtering in conditionally Markov switching hidden linear models, *Comptes Rendus Mathématique*, vol. 349, no. 9-10, pp. 587-590, May 2011.
- [45] L. Rabiner and B. Juang, An Introduction to Hidden Markov Models, *IEEE ASSP Magazine*, 3, pp. 4-16, 1986.
- [46] C. Shan, W. Zhou, Z. Jiang, and H. Shan, A new Gaussian approximate filter with colored non-stationary heavy-tailed measurement noise, *Digital Signal Processing*, p. 103358, 2021
- [47] I. Visser and M. Speekenbrink, *Mixture and Hidden Markov models with R*, Springer, 2022.
- [48] G.-H. Zhang, C.-Z. Han, F. Lian, and L.-H. Zeng, Cardinality balanced multi-target multi-Bernoulli filter for pairwise Markov model, *Zidonghua Xuebao/Acta Automatica Sinica*, vol. 43, no. 12, pp. 2100-2108, 2017.
- [49] G.-H. Zhang, J. Lan, L. Zhang, F. He, and S. Li, Filtering in Pairwise Markov Model with Student's t non-stationary noise with application to target tracking *IEEE Trans. on Signal Processing* 69, pp. 1627-1641, 2021.
- [50] W. Zucchini and I. L. MacDonald, *Hidden Markov models for time series*, CRC Press, Chapman & Hall, 2008