ISTITUTO DI ANALISI DEI SISTEMI ED INFORMATICA
"Antonio Ruberti"
CONSIGLIO NAZIONALE DELLE RICERCHE

F. Carravetta

NEAREST-NEIGHBOUR MODELLING OF RECIPROCAL CHAINS

R. 671 Novembre 2007

Francesco Carravetta – Istituto di Analisi dei Sistemi ed Informatica del CNR, Viale Manzoni 30, I-00185 Roma, Italy.
Email : `francesco.carravetta@iasi.cnr.it`. URL : `http://www.iasi.cnr.it/~adp`.

ISSN: 1128–3378

Collana dei Rapporti dell'Istituto di Analisi dei Sistemi ed Informatica "Antonio Ruberti", CNR

viale Manzoni 30, 00185 ROMA, Italy

tel. ++39-06-77161
fax ++39-06-7716461
email: iasi@iasi.cnr.it
URL: http://www.iasi.cnr.it

Abstract

This paper focuses on the class of finite-states, discrete-index, reciprocal processes (reciprocal chains). Such a class of processes seems to be a suitable setup in many applications, and in particular it appears well suited for image-processing. While addressing this issue, the aim is twofold: theoretic and practical. As to the theoretic purpose, some new results are provided: first, a general stochastic realization result is provided for reciprocal chains endowed with a known, but no matter how it is, distribution. Such model has the form of a fixed-degree *nearest-neighbour* polynomial model. Next, the polynomial model is shown to be *exactly linearizable*, which means it is equivalent to a nearest-neighbour *linear* model in a different set of variables. The latter model results to be *formally* identical to the Levi-Frezza-Krener linear model of a Gaussian reciprocal process, although actually nonlinear respect to the chain's values. As far as the practical purpose is concerned, in order to yield an example of application an estimation issue is addressed: a suboptimal (polynomial-optimal) solution is derived for the smoothing problem of a reciprocal chain partially observed under non-Gaussian noise. To this purpose, two kinds of boundary conditions (Dirichelet and Cyclic), specifying the reciprocal chain on a finite interval, are considered, and in both cases the model is shown to be *well-posed*, in a 'wide-sense'. Under this view, some well known representation results about Gaussian reciprocal processes, in a sense, extend to a 'non-Gaussian' case by reason of the present paper.

Key words: Reciprocal processes, Markov chains, Markov fields, Smoothing algorithms, Stochastic realization, Nearest-neighbour models.

1. Introduction

Reciprocal processes (1-D Markov random fields) are an appealing class of noncausal random processes occurring in many areas of science and engineering. These processes are usually indexed by space, instead of time, so that the usual Markovian/causality concepts no longer apply. In the last decades reciprocal processes have been studied in detail in different setups, such as discrete/continuous-index. We refer to [3, 29, 30, 16, 11, 12, 13, 21, 20, 32, 19, 1, 18, 28] and references therein as a meaningful selection of basic works about reciprocal processes. The reciprocity property is essentially one of 'time locality', and it was first formalized by Bernstein in [3] while following along with some ideas of Schrodinger, who in [29, 30] was attempting to develop a stochastic formulation of quantum mechanics in terms of Markov diffusions. Essentially, a stochastic process is said to be reciprocal if given a fixed interval, the information concerning the process in the interval, given by the process *outside* the interval, is actually redundant, and could be replaced just by the information given by the process values at both ends of the interval. Thus, while thinking to an interval where an end point goes to infinity, one could guess that reciprocal property is a generalization of Markov one. In fact, the class of reciprocal processes has been shown to include Markov processes, as well as other processes. Afterwards, in 1956, P. Levy defined the concept of Markov field [16] as a generalization of the above described reciprocity property to processes with a multi-dimensional indeces set. For this reason, nowadays 'Markov field' is the commonly used name in the literature to denote multi-dimensional reciprocal processes, even though a Markov field in one dimension is not indeed a Markov process in general.

A detailed study on reciprocal processes can be found in [11], where more insight has been given on the relashionship with Markov processes. In particular it has been shown that a reciprocal process is specified (in the Kolmogorov sense: i.e. all its finite-dimensional distributions are determined) by a family of *three-point reciprocal transitions* and by the joint probability of the two end-point values. This result allows us to view more sharply a reciprocal process as a generalization of Markov ones since, as well known, Markov processes are specified by the family of its *two-point* probability transitions and by the distribution of the process at the initial time. Moreover, in the same paper a necessary and sufficient condition is given for a reciprocal process to be Markovian. By the Jamison's work in [11] we can infer that any reciprocal process can be constructed by first *pinning* a Markov process at both ends of an interval, and then assigning an arbitrary probability distribution to the end points of the process.

In the above mentioned papers, reciprocal processes have been characterized by 'transitions' of probability, thus enhancing a likeness with Markov processes. In the papers [12, 13, 21, 20, 32] the attention was focused on the 'dynamic' properties of a reciprocal process. In particular, efforts have been devoted in characterizing it as a solution of some 'dynamic' equations. Such a topic can be seen as a *stochastic realization* problem for a reciprocal process, but the 'dynamic' structure cannot be chosen a-priori to be a general stochastic differential equation (SDE) with unknown coefficients, since the solution of a SDE is always a Markov process. For these reasons, Krener introduced in [12] the concept of *second-order* stochastic differential equation. As a matter of fact, he was able to prove that a Gaussian reciprocal process is a solution of a linear second-order stochastic differential equation driven by a locally correlated noise.

As to *discrete-index* reciprocal processes, among the most relevant contribution in modeling and estimation for Gaussian reciprocal processes we highlight the article of Levi, Frezza and Krener [19], where a number of topics concerning Gaussian reciprocal processes with discrete-index are addressed and solved. In their paper a complete characterization is given for the considered class of reciprocal processes in terms of a second-order *nearest-neighbour* model driven by a Gaussian locally correlated noise (namely $e(k)$):

$$M_0(k)x(k) - M_-(k)x(k-1) - M_+(k)x(k+1) = e(k), \tag{1}$$

where for any k, $M_0(k)$ is symmetric positive definite, and $M_+(k) = M_-^T(k+1)$, and model-coefficients as well as input-noise correlation function are shown to be computable by the process second-order statistics. Since this kind of model needs to be endowed with suitable boundary conditions, two important sets of boundary conditions, namely *Dirichlet* and *cyclic*, are considered. Then, by using a Green function, a kernel representation of the model is derived and next used to obtain a recursive solution to the smoothing problem for a Gaussian reciprocal process partially observed under additive Gaussian noise.

The Levi-Frezza-Krener model (1) represents indeed a solution of the stochastic realization problem for Gaussian reciprocal processes in the form of a *non-causal* linear model driven by a one-step correlated (Gaussian) noise. It has been derived, essentially, by using classical tools of stochastic processes and probability theory, in the setting of discrete-index processes, and starting from the basic ideas previously used by Krener in [12], in the more critical continuous-index setting. In general we can say that for a Gauss-Markov process the stochastic realization from the second order statistics (which are sufficient ones) is well known: the signal can be represented as the output of dynamical linear-system driven by a Gaussian white-noise. In this perspective the article [19], represents an extension of the above basic result to a wider class of processes than Gauss-Markov, i.e. the Gaussian-reciprocal ones.

The link with the general geometric stochastic realization theory was not investigated in [19], since it was beyond the scope of that paper. Nevertheless, this task has been carried out in the paper [28], where the relationship of the results in [19] with the stochastic realization theory, has been deeply investigated. The nearest-neighbour model (1), in the stationary case, and defined on the *circle* (in other words: with 'cyclic boundary conditions') has been derived newly, by following a different approach based upon geometric stochastic realization theory. To this purpose we point out the paper by Lindquist and Picci [22], which is a survey work collecting in one place a comprehensive theory of stochastic realization for continuous-time stationary Gaussian processes, and it includes the main definitions, issues and tools of a general geometric stochastic realization theory. In particular, the basic concept of 'splitting subspace' has been used in [28] in order to reformulate the results of [19] into the geometric language of [22]. Moreover, in the same paper [28] new results have been found concerning general structural properties of *reciprocal realizations* on the circle such as minimality as well as interior/exterior observability. The results of paper [28] are however derived under some restrictive hypotheses which seem quite difficult to relax. In fact, besides the stationarity hypothesis, and although some concepts introduced in the paper – such as the definition of 'family of reciprocal subspaces' – are independent by Gaussian hypothesis, the results of the paper are based upon the equivalence between the concepts of 'conditional orthogonality' and 'conditional independence', thus indeed valid only in the Gaussian case.

Even though non-causal, the Levi-Frezza-Krener model (1) has nice properties which allow to derive estimation algorithms. In particular, the driving-noise is uncorrelated with 'past' and 'future' states. This allowed, in [19], the computation of a smoothing algorithm running recursively twice over the smoothing-interval in the forward and backward directions.

As far as nearest-neighbour models for 2-D Markov fields are concerned (which are almost directly connected with image processing problems), extensions of the previous result can be found in [15, 33, 17]. We highlight image processing problems as a tipical application area where stochastic models with reciprocal processes reveal their usefulness. Indeed reciprocal processes seem to be particularly suited to describe the spatial correlation structure of textured images, and research efforts have been recently devoted in order to apply reciprocal-processes-modeling to texture analysis. In particular, in [7] a setup for texture analysis is adopted (a similar one was previously considered in [31]) where the signal (representing the image) is assumed to be a finite sequence of vectors where each vector collects all the intensity values in a row. Then all the problems concerning image processing are shown to be equivalent to realization and subspace identification problems for reciprocal processes defined on the discrete unit circle.

A purpose of the present paper is to offer a solution of the stochastic realization problem in the case of a *finite states* reciprocal process. The methods we will use are just the classical stochastic processes and probability theory for stochastic sequences (so we follows along with Levi, Frezza and Krener in [19]) as to the kind of problem-formulation adopted) as well as the *Kronecker algebra*, which will be a powerful tool in order to represent and work with polynomials of vectors.

It will be considered a particular class of non-Gaussian reciprocal processes, which seem to be interesting from an application point of view, i.e. the *finite-states* processes. More precisely we shall consider processes taking values in a finite set of elements, each one with a known probability. Following a well-known terminology widely used for the Markov case, we will call *reciprocal chains* this class of processes.

We are able to show that a nearest-neighbour model can be found even for non-Gaussian reciprocal chains. The model here presented is a nonlinear one, and in particular it is *polynomial.*

As an example of application of reciprocal chains we point out again image-processing and all related problems: modeling, estimation, recognition etc. As a matter of fact, the pixel intensity value of an image is often quantized in a finite set of gray-levels. Moreover, the probabilistic structure of quantized images can be often derived by off-line experiments. As an example, in the recognition problem of an object, one could use a set of different images of the same object, and for any pixel set the probability of each gray-level being equal to the measured occurrence of that gray-level in the set of images. On the other hand, by taking into account that such a class of processes are 1-D Markov-fields, the probabilistic structure of them can be inferred by the theory of *Gibbs fields* [10]. We will discuss in more detail this point later in §2.

As a first result, in the present paper it will be shown that a reciprocal chain x, taking values with probability one (w.p.1) in a finite subset of $\mathbb{R}^n$ is necessarily a solution of *any* of the following *polynomial stochastic equation* (for $h = 1, 2, ...$) :

$$x_{[h]}(k) = \sum_{\substack{i,j=0 \\ 0<i+j\leq\nu}}^{\nu} F_{i,j}^{(h)}(k)(x_{[i]}(k-1) \otimes x_{[j]}(k+1)) + F_{0,0}^{(h)}(k) + d_h(k) \tag{2}$$

for some suitable set of matrix coefficients $F_{\cdot,\cdot}^{(\cdot)}$, where d_h is a driving one-step correlated noise sequence, and $x_{[h]}$ denotes the *h-th Kronecker power* (reduced order) of the vector x. Moreover the matrix coefficients of representation (2), as well as the correlation function of the driving noise, are all computable by the process statistics. Equation 2 is a polynomial equation of degree ν, where ν is an integer depending on the *number of states* of the reciprocal chain x.

The second result concerns the existence of an *exact linearization* of eq. (2). The problem of exact linearization can be stated as follows. For $\mu \geq \nu$ let us define the vectors

$$\mathcal{X}(k) = \begin{bmatrix} x^{\{2\}}(k) \\ x^{\{3\}}(k) \\ \vdots \\ x^{\{\mu\}}(k) \end{bmatrix}; \qquad X(k) = \begin{bmatrix} x(k) \\ x_{[2]}(k) \\ \vdots \\ x_{[\mu]}(k) \end{bmatrix}; \tag{3}$$

where

$$x^{\{\gamma\}}(k) = \begin{bmatrix} x(k-1) \otimes x_{[\gamma-1]}(k+1) \\ x_{[2]}(k-1) \otimes x_{[\gamma-2]}(k+1) \\ \vdots \\ x_{[\gamma-1]}(k-1) \otimes x(k+1) \end{bmatrix}, \tag{4}$$

and $v_{[i]}$ denotes, for any nonnegative integer i, the ith *reduced order Kronecker power* of a vector v (see §2.1). It is quite intuitive that eq. (2) can be grouped together for $h = 1, ..., \mu$ as follows

$$\widetilde{X}(k) = d(k) + F(k)\widetilde{\mathcal{X}}(k) + \Phi(k)\widetilde{X}(k+1) + \Psi(k)\widetilde{X}(k-1) \tag{5}$$

where $\widetilde{X}, \widetilde{\mathcal{X}}$ denote $X - \mathbf{E}\{X\}$ and $\mathcal{X} - \mathbf{E}\{\mathcal{X}\}$ respectively. Equation (5) shows that the reciprocal chain x satisfies a polynomial equation which actually reduces to a linear one in the augmented variables $\widetilde{X}, \widetilde{\mathcal{X}}$, which is very close to the representation (1). However, if one considers eq. (5) being a linear equation in $\widetilde{X}, \widetilde{\mathcal{X}}$, this turns out to be under-determined. Thus one needs to find further equations (as much as the dimension of $\widetilde{\mathcal{X}}$ is) in order to reduce the polynomial representation (49) to a *linear* one (exact linearization).

As a matter of fact, in §4 of the present paper it will be shown that the vector $\widetilde{\mathcal{X}}$, collecting the *mixed powers* of the reciprocal chain x, can be expressed as a linear function of the process $\widetilde{X}$ and a *white noise* b:

$$\widetilde{\mathcal{X}}(k) = b(k) + H(k)\widetilde{X}(k+1) + L(k)\widetilde{X}(k-1). \tag{6}$$

Moreover, which is the main, and somewhat surprising result, one has $F(k) = 0$ in (5), which means that the process $\widetilde{X}$ (and hence x) is given by an equation independent of the mixed powers $\widetilde{\mathcal{X}}$, and indeed *formally identical* to (1). In this regard, what has to be stressed here is that knowing the matrix-coefficient $F(k)$ vanishes in eq. (5), besides the gain in theoretical insight, helps in reducing greatly the state-space dimension (we will turn on this issue, i.e. the increase of dimensionality, in §5).

Next, the *well-posedness* problem of the nearest-neighbour polynomial model is considered. We say that eq. (2) is well-posed if it has x as *unique* solution. Now, even though it can be reduced to the linearized model expressed by eqs. (5)–(6), we will see that well-posedness does not hold in general. Nevertheless we will show that our nonlinear model is *well-posed in a wide-sense*, meaning this that x is a solution of it, and any other solution has the same moments of x at least up to certain finite order. As we will see, the wide-sense well-posedness will be a sufficient condition in order to be the model useful for estimation purposes.

This will be the last issue of the present paper: we will present a suboptimal (in the sense below explained) smoothing algorithm for the estimation problem of a partially and noisy observed reciprocal chain x, and in this perspective the present paper aims to represent a contribution in the area of estimation problems for signals without a state-space description.

Let us comment on the latter topic which is a relevant one in many engineering areas. Usually the complete knowledge of the process-statistics is assumed, replacing the (more strong) requirement of a state-space model being available, and under various additional restrictive hypothesis solutions have been provided in the literature. Estimating signals from their statistics is such a problem which can be of course reduced to a state-estimation one in all cases when the stochastic realization problem can be solved. To this purpose, as already mentioned, in [19], a smoothing algorithm has been derived for a Gaussian reciprocal process, indeed a smoother which runs recursively twice over the smoothing-interval in the forward and backward directions. Following a different approach suboptimal smoothing algorithms have been directly derived without using stochastic realization at all, but using additional hypothesis on the statistics. To cite most recent ones, in [24] suboptimal estimation algorithms are derived for non-Gaussian discrete-time signals provided a separation-property is assumed for the covariance functions (semi-degenerated kernel form). Under the same hypothesis in [25] it is shown that a second-order polynomial recursive estimator can be defined provided a sufficient number of moments are given for the signal. Another approach in signal estimation is based on the assumption that, at a sufficiently wide neighbourhood of every time instant, the signal of interest is smooth enough to be approximated by a low-order polynomial. The optimal solution is based on the least-squares method. Signal reconstruction is implemented with a time-varying FIR filter whose coefficients are obtained in closed form via polynomial expansion. As an example of recent papers following the afore mentioned polynomial-filtering method we point out [14] and references therein. We stress that the word "polynomial-filter" is used in [14] with a different meaning than in [25]. In particular, in [25] polynomial-filtering is accomplished in the sense of Carravetta-Germani-Raimondi [5, 6], where for a signal described by a state-space model an algorithm was defined to recursively compute the mean-square optimal state-estimate within all estimates given by a fixed-degree polynomial function of the available observations. In the present paper we will define polynomial filters/smoothers in the sense of [5].

The paper is organized as follows. §2 is devoted to a brief recall of the background material used throughout the paper, the definition of the class of processes we will focus on is given, and some relevant probabilistic properties are described. In subsection 2.4 the *basic choice*, upon which the whole paper is next developped, is stated. Such a 'choice' will consists in the particular way we adopt to represent some – relevant for reciprocal processes – conditional expectations: that is a polynomial representation. The basic assumptions are all grouped togheter in subsection 2.5.

§3 and §4 include the main results of the paper, which we briefly introduced just before, while commenting on eqs. (2), (5), and (6). In particular eq. (2) is proven in §3, as well as algorithms to derive all the matrix-coefficients of (2) by the statistics, are given. even though much of the results in §4 include the ones presented in §3 as a particular case (thus, basically, one could replace §3 with §4) nevertheless §3 gives some basic results in a much more simple particular case than §4, thus we choose to present our method first in a more simple case in order to improve readability and the overall presentation. Moreover, we will see that the general method, given in §4, to calculate the system-coefficients actually blocks the view of some structural properties. We will return on this in the conclusive chapter.

§5 includes all topics concerning the well-posedness of our model, as well as the suboptimal smoothing algorithm. The definition of *polynomial-optimal estimate* (in the sense of [5]) is recalled, then the results of the previous sections are applied in order to build up an *any-order* polynomial-optimal smoother for a reciprocal chain partially observed under additive noise.

§6 includes some thoughts about the problem of model's growing in dimension. Some examples are given showing that the rice of dimensionality is only apparent in some meaningful cases. Also, the issue of mimimum-entropy-quantization is argued as the horizon towards which future

research should be addressed in order to overcome, in practical cases, dimensionality problems.
Finally, in §7 all main results are summed up and some concluding remarks are given.

2. Background and preliminaries

2.1. Block-matrix notation and Kronecker algebra

Let $A \in \mathbb{R}^{h \cdot l \times s \cdot r}$ be an $(hl \times sr)$-dimensioned matrix, where h, l, s, r are positive integers. Let us choose integers $i_1 < i_2 < .. < i_h$, $j_1 < j_2 < .. < j_h$, and denote by $A(i,j) \in \mathbb{R}^{l \times r}$ the (i,j)-th block, $i \in \{i_1, ..., i_h\}$, $j \in \{j_1, ..., j_s\}$, partitioning A as

$$A = \begin{bmatrix} A(i_1, j_1) & \dots & A(i_1, j_s) \\ \vdots & \ddots & \vdots \\ A(i_h, j_1) & \dots & A(i_h, j_s) \end{bmatrix}. \tag{7}$$

Throughout this paper we will often use block-matrices having some level of nesting, and we shall use the following – shorter and more easy to write – block-matrix notation in place of (7)

$$A = [A(i,j)]_{i=i_1,\dots,i_h}^{j=j_1,\dots,j_s} \quad . \tag{8}$$

In the case of a block-matrix composed by either a single block-row or a single block-column, the following slight modification of notation will be used (8)

$$A = \begin{cases} [A(i,j)]^{j=j_1,\dots,j_s}, & \text{for} \quad i = i_1 = i_2 ... = i_h; \\ \\ [A(i,j)]_{i=i_1,\dots,i_h}, & \text{for} \quad j = j_1 = j_2 ... = j_s. \end{cases} \tag{9}$$

In particular, for $a, b \in \mathbb{R}^{h \cdot l}$ being the following row- and column-vector respectively:

$$a = \begin{bmatrix} a(i_1) & \dots & a(i_h) \end{bmatrix}; \qquad b = \begin{bmatrix} b(i_1) \\ \vdots \\ b(i_h) \end{bmatrix}, \tag{10}$$

with $a(i), b(i) \in \mathbb{R}^l$ block-row- and block-column-entries respectively, we will rewrite the vectors in (10) as

$$a = [a(i)]^{i=i_1,\dots,i_h} \qquad b = [b(i)]_{i=i_1,\dots,i_h} \quad . \tag{11}$$

Nevertheless, we still use the traditional matrix and vector notation when the block-structure is not indexed. We will write $\dim(a)$ for the dimension of vector a, and A^T to denote the transpose of matrix A. For a square matrix A, the notation $A > 0$ will mean that A is positive definite. The symbol I_l will denote the identity matrix in $\mathbb{R}^l$, although it will be also used just the symbol I if dimension is clearly determined from context. The 'null-blocks' of a block-matrix will be possibly indicated by '0' without specifying dimensions. For a block-diagonal matrix, say (8) with $s = h$ and $A(i,j) = 0$ for $i \neq j$ we shall write

$$A = \operatorname{diag}\{A(i)\}_{i=i_1,\dots,i_h} \tag{12}$$

where $A(i) = A(i,i)$.

Let us consider two matrices $A \in \mathbb{R}^{l\times r}$, $B \in \mathbb{R}^{s\times h}$. We shall denote with $A \otimes B \in \mathbb{R}^{l\cdot s\times r\cdot h}$ the *Kronecker product*

$$A \otimes B = [A_{i,j} \cdot B]_{i=1,\dots,l}^{j=1,\dots,r} \quad , \tag{13}$$

where $A_{i,j}$ is the scalar (i,j)-entry of A and "$\cdot$" represents the multiplication between scalars and matrices. For any $h = 0, 1, \dots$ the *h-th Kronecker power* of A, namely $A^{[h]}$ is recursively defined as

$$A^{[h]} = A^{[h-1]} \otimes A, \qquad A^{[0]} = 1. \tag{14}$$

Let us consider $a = [a_i]_{i=1,\dots,l}$, and notice that, by definition, the vector $a^{[h]}$, for $h \geq 2$, has many repeated entries. We shall denote by $a_{[h]}$ the *reduced h-th Kronecker power* of a vector a, defined by

$$a_{[h]} = \widetilde{T}_h^{(l)} a^{[h]}, \tag{15}$$

where $\widetilde{T}_h^{(l)} \in \mathbb{R}^{\dim(a_{[h]})\times l^h}$ is a 0,1-matrix extracting the entries of $a^{[h]}$ so that the result is a vector without repeated components (see §4.1 of [5]). For $h = 0, 1$ the ordinary and reduced Kronecker powers of a are the same, and one has $a_{[1]} = a^{[1]} = a$ and $a_{[0]} = a^{[0]} = 1$ (the scalar unit). We define $T_h^{(l)}$ as the matrix performing the the inverse operation:

$$a^{[h]} = T_h^{(l)} a_{[h]}, \tag{16}$$

The reader is directed to [2], [27], [5] for more details about Kronecker algebra; here we just present a brief list of properties which are used in the present paper.

The Kronecker product is non commutative in general, nevertheless for any given pair of matrices $A \in \mathbb{R}^{r\times s}$, $B \in \mathbb{R}^{n\times m}$, it is

$$B \otimes A = \mathbf{K}_{r,n}^T (A \otimes B) \mathbf{K}_{s,m}. \tag{17}$$

where the commutation matrix $\mathbf{K}_{u,v}$ is the $(u \cdot v) \times (u \cdot v)$ *invertible* matrix such that its (h,l) entry is given by:

$$(\mathbf{K}_{u,v})_{h,l} = \begin{cases} 1, & \text{if} \quad l = (|h-1|_v)u + \left(\left[\frac{h-1}{v}\right] + 1\right); \\ 0, & \text{otherwise.} \end{cases} \tag{18}$$

Observe that $\mathbf{K}_{1,1} = 1$, hence in the vector case when $a \in \mathbb{R}^r$ and $b \in \mathbb{R}^n$, (18) becomes

$$b \otimes a = \mathbf{K}_{r,n}^T (a \otimes b). \tag{19}$$

Moreover, from (19), it readily follows that:

$$\mathbf{K}_{r,n}^{T\;-1} = \mathbf{K}_{n,r}^T. \tag{20}$$

If A, B, C, D are suitaly dimensioned matrices the following properties hold

$$\begin{aligned} (A+B) \otimes (C+D) &= A \otimes C + A \otimes D + B \otimes C + B \otimes D && (21) \\ A \otimes (B \otimes C) &= (A \otimes B) \otimes C && (22) \\ (A \cdot C) \otimes (B \cdot D) &= (A \otimes B) \cdot (C \otimes D) && (23) \\ (A \otimes B)^T &= A^T \otimes B^T. && (24) \end{aligned}$$

Finally, an extension of the classical Newton formula holds for Kronecker powers of vectors [5]: for any integer $h \geq 0$ and vectors $a, b \in \mathbb{R}^l$ there exists matrix coefficients $\{\mathbf{M}_0^h(l), ..., \mathbf{M}_h^h(l)\}$ (namely 'binomial matrices') such that

$$(a+b)^{[h]} = \sum_{k=0}^{h} \mathbf{M}_k^h(l)(a^{[k]} \otimes b^{[h-k]}) \tag{25}$$

and for $1 \leq j \leq h-1$ the binomial matrices can be calculated by the following recursive equations (extension to a vector case of the *Tartaglia's triangle*):

$$\mathbf{M}_h^h(l) \quad = \mathbf{M}_0^h(l) = I_{l^h} \ , \tag{26}$$

$$\mathbf{M}_j^h(l) \quad = (\mathbf{M}_j^{h-1}(l) \otimes I_l) + (\mathbf{M}_{j-1}^{h-1}(l) \otimes I_l) \cdot (I_{l^{j-1}} \otimes \mathbf{K}_{l,l^{h-j}}^T) \tag{27}$$

where $\mathbf{K}_{\cdot,\cdot}$ is the commutation matrix defined in (18) and I_s denotes the identity in $\mathbb{R}^s$.

2.2. Reciprocal processes, and the subclass here considered

Throughout this paper $\mathbf{E}$ will denote the expectation operator. If U is a random vector, the notation $\widetilde{U}$ will stand for $U - \mathbf{E}\{U\}$. The symbol $\mathrm{Cov}\{U, V\}$ will be often used to denote the mutual covariance of two random vectors U, V, i.e. $\mathbf{E}\{\widetilde{U}\widetilde{V}^T\}$. The notation $\xi = \{\xi(k), k \in \mathcal{I}\}$, will be used to denote a stochastic sequence defined on $\mathcal{I} \subset \{0, \pm 1, \pm 2, ...\}$. When $\mathcal{I} = [0, N]$, and $\xi(k) \in \mathbb{R}^n$, the sequence ξ will be identified with a random (column) vector, namely $\xi = \{\xi(k)]_{k=0,...,N}$, taking values in $\mathbb{R}^{(N+1)n}$. The underlying probability space will be always understood, and the sigma-algebra generated by a subset of random variables, namely $\mathbf{S} \subset \xi$ will be denoted by $\sigma(\mathbf{S})$, and often identified with the subset $\mathbf{S}$ itself. Thus, conditional expectation of U given $\sigma(\mathbf{S})$, will be denoted: $\mathbf{E}\{U/\mathbf{S}\}$.

As well known [11], a stochastic process x defined on $\mathbf{T} \subset \mathbb{R}$ is said to be a *reciprocal* process if, given any subinterval $[r, s] \subset \mathbf{T}$, $k \in (r, s)$, $l \in [r, s]^c$ ($[r, s]^c$ denoting the set-complement in $\mathbf{T}$), $x(k)$ and $x(l)$ are conditionally independent given $x(r), x(s)$. Following [11] we know that any Markov process is reciprocal (but the converse is not true in general). The family of three argument functions $P(t_1; t_2; t_3)$, $t_1 < t_2 < t_3$, expressing the conditional probability of $x(t_2)$ given $x(t_1)$ and $x(t_3)$ is said to be the family of *reciprocal probability transitions.* A reciprocal process is specified (i.e. all the joint finite-dimensional probabilities are determined) *i)* by a family of reciprocal transitions *and ii)* by a two-point joint probability (i.e. the joint distribution of $x(t_1), x(t_2)$ for a given pair $t_1, t_2 \in \mathbf{T}$). There are in general *many* reciprocal processes inducing the same family of reciprocal transitions – and in fact the process is univocally determined by the further condition *ii)* – so the collection of them is called a *reciprocal class.*

As to the definition above, of course it characterizes any reciprocal process, with either continuous or discrete index, and taking values in a general Hausdorff space (this is the setting for instance of [11]), however, in specific here we are only interested to discrete-index and finite-states reciprocal processes, that is *reciprocal chains* for short. In our setting $\mathbf{T}$ is a discrete interval named *indeces set.*

We say x is a "finite-states" process in the following sense. Let $\mathcal{S} \subset \mathbb{R}^n$ be a finite set of "states", namely $\mathcal{S} = \{\xi^{(1)}, ..., \xi^{(L)}\}$, and $\mathcal{S}^{\mathbf{T}}$ the set of all sequences: $\{s(k); k \in \mathbf{T}\}$ with $s(k) \in \mathcal{S}$. Denoting $\mathbb{R}^{n\mathbf{T}}$ the set of all maps $\mathbf{T} \to \mathbb{R}^n$, one has $\mathcal{S}^{\mathbf{T}} \subset \mathbb{R}^{n\mathbf{T}}$. Thus, by viewing x as an $\mathbb{R}^{n\mathbf{T}}$-valued random variable, we say that x is a finite-states process if x has a mass probability distribution concetrated on the elements of $\mathcal{S}^{\mathbf{T}}$.

The set of states may be even a general finite alphabet of symbols: $\mathcal{S}' = \{\alpha_1, ..., \alpha_L\}$, in which case we understood that it has been given a 'distance' $d'(\alpha_i, \alpha_j)$ between all symbols, that is any positive real function such that $d'(\alpha, \alpha) = 0$, and an invertible map $\phi : \mathcal{S}' \to \mathcal{S}$ such that the euclidean distance $d(\cdot, \cdot)$ in $\mathbb{R}^n$ is an extension of the function d', that is : $d(\phi(\alpha_i), \phi(\alpha_j)) = d'(\alpha_i, \alpha_j)$, $\forall i, j = 1, ..., L$.

Reciprocity is characterized in the discrete-index case at issue as follows. For any Borel function $f : \mathbb{R}^n \to \mathbb{R}^l$ (any l) one has

$$\mathbf{E}\{f(x(k))/x(i), i \neq k\} = \mathbf{E}\{f(x(k))/x(k-1), x(k+1)\}. \tag{28}$$

2.3. Reciprocal chains as Markov-fields in one dimension

It has been mentioned so far, in the introduction, that reciprocal processes are a particular case of *Markov fields* [10], i.e. they are Markov-fields in one dimension. Now, it is well known (Theorem of Hammersey-Clifford, see for instance [10]) that all Markov-fields are well defined subclasses of *Gibbs fields*, that is they can be characterized by a properly choosen *Gibbs potential*, namely a family of functions $\psi = \{\psi_A, A \in \mathcal{C}\}$, ψ_A being a scalar function of the field-configurations, where $\mathcal{C}$ is the family of cliques of the graph underlying the Markov-field. In other words, there is a Gibbs potential ψ which determines all the conditional-probabilities of the field (while using the same term as in [11]: ψ determines a reciprocal class). Cliques are sub-sets of *sites* collecting *nearest-neighbours*, so it is sufficient to define the potential ψ only on the cliques in order to specify a reciprocal class. In our case (one-dimensional) the set of sites is $\{0, ..., N\}$, whereas the nearest-neighbours are all the sets: $\{k, k+1\}$. Thus, given a Gibbs potential ψ, defined on the sets $\{k, k+1\}$, we can specify a set of conditional-probabilities (that is the reciprocal class of x) by using the general expression of a Gibbs distribution:

$$P\left\{x(k) = \xi^{(s)}/x(k-1), x(k+1)\right\}\Bigg|_{\substack{x(k-1)=\xi^{(i)} \\ x(k+1)=\xi^{(j)}}} = Z^{-1}\exp \sum_{l\in\{i,j\}} \psi(\xi^{(s)}, \xi^{(l)}) = G_{\xi^{(s)}}(k; \xi^{(i)}, \xi^{(j)}) \tag{29}$$

where

$$Z = \sum_{s=1}^{L} \exp \sum_{l\in\{i,j\}} \psi(\xi^{(s)}, \xi^{(l)}). \tag{30}$$

2.4. Finding an expression of the conditional expectation

We are dealt with the issue of finding 'dynamic' models for the reciprocal chain $x = \{x(k), k \in \mathbf{T}\}$. To this purpose we would know in advance the 'shape' of the function, say $f_{\bar{\phi}_k}(\cdot, \cdot)$, expressing the following conditional expectation (CE):

$$\mathbf{E}\{\bar{\phi}_k(x)/x(k-1), x(k+1)\} = f_{\bar{\phi}_k}(x(k-1), x(k+1)), \tag{31}$$

where $\bar{\phi}_k$ is some Borel function (choosen at k) $\bar{\phi}_k : \mathbb{R}^{n\mathbf{T}} \to \mathbb{R}^l$, with l a positive integer, and $\mathbb{R}^{n\mathbf{T}}$ is the standard set-theoretic notation denoting the set of all maps: $\mathbf{T} \to \mathbb{R}^n$.

To better explain this, consider first a less general expression than (31):

$$\mathbf{E}\{\phi(x(k))/x(k-1), x(k+1)\} = f_\phi(k; x(k-1), x(k+1)), \tag{32}$$

with ϕ being any Borel function $\phi : \mathbb{R}^n \to \mathbb{R}^l$. In the case of a *Gaussian* discrete-index reciprocal process x (i.e. not a finite-states process), if one takes the identity function ϕ: $\phi(x(k)) = x(k)$, and exploits the reciprocity property (28) (with f identity in $\mathbb{R}^n$), then eq. (32) rewrites

$$\mathbf{E}\left\{x(k)/x(i), i \neq k\right\} = \mathbf{l}(k; x(k-1), x(k+1)), \tag{33}$$

where $\mathbf{l}(k; \cdot, \cdot)$ is a *linear* function of the last two arguments. Equation (33), and the linearity of $f_\phi(k, \cdot, \cdot)$ are, essentially, the key properties which allows to derive, in [19], the linear model (1), for the Gaussian reciprocal process x. In the case we are here faced with, i.e. an $\mathcal{S}$-valued reciprocal chain x ($\mathcal{S}$ defined in subsection 2.1) the CE in (33) assumes, almost surely, the following L^2 values

$$\mathbf{E}\left\{\phi(x(k))/x(k-1), x(k+1)\right\}\Bigg|_{\substack{x(k-1)=\xi^{(i)} \\ x(k+1)=\xi^{(j)}}} . \qquad i, j = 1, ..., L. \tag{34}$$

Now, many equivalent ways – in the 'almost surely' sense – can be adopted to express the CE in (33). Indeed, to this purpose we could choose *any* function $f_\phi(k, \cdot, \cdot)$, *provided it is an exact fit through the L^2 points* given by (34). For instance, a ready choice could be to set up the following (almost sure) identity:

$$\mathbf{E}\{\phi(x(k))/x(k-1), x(k+1)\} = G^\phi(k; x(k-1), x(k+1)), \tag{35}$$

where

$$G^\phi(k; \cdot, \cdot) = \sum_{s=1}^{L} \phi(\xi^{(s)}) G_{\xi^{(s}}(k; \cdot, \cdot),$$

and $G_{\xi^{(s)}}$ is the Gibbs distribution given in (29). Indeed, once x belongs to the reciprocal class determined by the potential ψ, by (29), (30), the function $G^\phi(k; \cdot, \cdot)$ is obvioulsy an exact interpolator of the L^2 points (34). Then, reciprocal property implies that an equation such as (33) keeps on being valid, almost surely, while replacing the linear function $\mathbf{l}(k; \cdot, \cdot)$ with the nonlinear one $G^\phi(k; \cdot, \cdot)$. Also, besides identity (35), different choices are possible, which are even more convenient to our purposes: for instance one can choose *the finite degree polynomial which exactly interpolates the finite number of values* (34).

This is indeed the basic choice of the present paper. We will express (within stochastic equivalence) the CE (32) as a ν-degree polynomial:

$$\mathbf{E}\{\phi(x(k))/x(k-1), x(k+1)\} = \sum_{\substack{i,j=0 \\ i+j \leq \nu}}^{\nu} c_{i,j}^{(\phi)}(k) \left(x_{[i]}(k-1) \otimes x_{[j]}(k+1)\right). \tag{36}$$

In other words, since one has to interpolate only a finite number of CE values, there exist matrix-coefficients $c_{\cdot,\cdot}^{(\phi)}$ such that (36) holds with probability one. Note that ν in (36) is the *maximal degree* of the interpolating polynomial for an 'alphabet' $\mathcal{S}$ having L elements. Of course $\nu < +\infty$, and it does not depend of ϕ (indeed it only depends of L).

We stress that expressions (35) and (36) are *equivalent*, as both right hand sides in that expressions are *almost surely equal.* The ν-order polynomial in (36) and the function $G^\phi(k; \cdot, \cdot)$ are both exact interpolators of a finite set of points in the CE-graph where probability concentrates its mass. Therefore, the subset of the underlying probability space over which that expressions differ either from each other or from the CE in the left hand side of (35) and (36), is necessarily a zero-measure set. Although both expressions are equivalent, choosing the polynomial expression (36) as a version of the conditional expectation will allow us to set up a new approach, different to the classical Markov fields theory, and closer to the stochastic-system-theoretic approach just used in [19] for the Gaussian discrete-index case. As a matter of fact, if we set $\phi(\xi) = \xi_{[h]}$, with h a positive integer, and exploit the reciprocity property (28), by (36) we have

$$\mathbf{E}\{x_{[h]}(k)/x(i), i \neq k\} = \sum_{\substack{i,j=0 \\ i+j\leq\nu}}^{\nu} F_{i,j}^{(h)}(k) \left(x_{[i]}(k-1) \otimes x_{[j]}(k+1)\right), \tag{37}$$

where $F_{i,j}^{(h)}(k) = c_{i,j}^{(\phi)}(k)$, with ϕ choosen as above. From (37), we will be able to prove that x satisfies a polynomial 'dynamic' model as (49). Moreover we will show that, under usual assumptions, the coefficients $F_{\cdot,\cdot}^{(h)}$ – which will be called *projection matrices* – can be calculated, for any $h = 1, 2...$ by the following moments of x

$$\mathbf{E}\left\{x_{[i]}(k-1) \otimes x_{[j]}(k+1)\right\}, \qquad i,j = 0, ..., 2\nu, \quad i+j \leq 2\nu, \tag{38}$$

$$\mathbf{E}\left\{x_{[h]}\left(x_{[i]}(k-1) \otimes x_{[j]}(k+1)\right)^T\right\}, \quad i,j = 0, ..., \nu, \quad i+j \leq \nu. \tag{39}$$

A polynomial representation can be adopted for the general expression (31) as well

$$\mathbf{E}\{\bar{\phi}_k(x)/x(k-1), x(k+1)\} = \sum_{\substack{i,j=0 \\ i+j\leq\nu}}^{\nu} \mathbf{c}_{i,j}^{(\bar{\phi})}(k) \left(x_{[i]}(k-1) \otimes x_{[j]}(k+1)\right). \tag{40}$$

The polynomial-degree in (40) is the same – ν – as in eq. (36). However, since we are now concerned with $\bar{\phi}$, which is a Borel function of the *whole* process x, we cannot exploit the reciprocity property (28) to get an expression, similar to (37), for the whole conditional expectation $\mathbf{E}\{\bar{\phi}_k(x)/x(i), i \neq k\}$.

2.5. Basic assumptions

Throughout this paper, unless differently specified, we understand that the parameter set $\mathbf{T}$ is either a subinterval, finite of infinite, of the integers set $\{0, \pm1, \pm2, ...\}$ or it is the discrete *circle* $[0, N]$. The latter means that $[0, N]$ is a discrete interval, and for $k \in [0, N]$ the algebraic sum $k \pm i$ is interpreted modulo N for any integer i. We consider an $\mathcal{S}$-valued (the set $\mathcal{S}$ being defined in subsection 2.1) reciprocal chain x, defined on $\mathbf{T}$, such that all its moments are known.

Assuming the knowledge of *any-order* moments is not actually necessary to our purposes. In fact, as mentioned earlier, the projection matrices $c_{i,j}^{(\phi)}(k) = F_{i,j}^{(h)}(k)$ of the CE (40) are computable by a finite number of moments, i.e. (38), (39), therefore this assumption could be weakened. Anyway, this is not really a stronger hypothesis. Indeed, since x is a finite-states process, as defined in subsection 2.1, all the moments are finite, that is, for any j-tuple of non-negative integers $i_1, ..., i_j$

$$\mathbf{E}\left\{x_{[i_1]}(k_{i_1}) \otimes x_{[i_2]}(k_{i_2}) \otimes \cdots x_{[i_j]}(k_{i_j})\right\} < \infty \tag{41}$$

for any choice $k_{i_1}, ..., k_{i_j} \in \mathbf{T}$, where j is any integer. Therefore we only need to assume that moments (41) are known, and it is reasonable to think that, if moments are computable (or they can be estimated) up to a certain order, then they can be calculated for higher order as well. As an example of a situation where moments (41) are available, we point out the case of a Markov field associated to a given potential. From what has been argued in previous subsections it results that – as to a reciprocal class given by some Gibbs potential, say ψ, defined on the collection of nearest-neighbours in $\mathbf{T}$ – the knowledge of the joint probability distribution of $x(t_1)$ and $x(t_2)$, with any $t_1, t_2 \in \mathbf{T}$, determines a reciprocal chain x in that class. Then all the a-priori probabilities are computable by the transition probabilities given by the Gibbs distribution (29). Therefore, all the moments of x, that is (41), can be calculated as well.

The second our basic assumption is the *nonsingularity assumption*, which we state as follows. Let j a positive integer, and $k_1, ..., k_j \in \mathbf{T}$ such that $k_i+1, k_i-1 \in \mathbf{T}$, for any $i = 1, ..., j$. Notice that the above condition is always satisfied either $\mathbf{T}$ is the unit circle or it is the set of all integers, whereas when $\mathbf{T} = [0, N]$ a bounded discrete interval, the above condition is satisfied in the *interior* of $\mathbf{T}$ (in the 'discrete' sense, i.e.: $\{1, ..., N-1\}$). Now, define

$$\mathbf{X}^j(k_1, ..., k_j) = [x(k_i)]_{i=1,...,j}\,.$$

Then we state the following *nonsingularity assumption*

$$\mathrm{Cov}\left\{\left(\mathbf{X}^j(k_1, ..., k_j)\right)_{[h]}, \left(\mathbf{X}^j(k_1, ..., k_j)\right)_{[h]}\right\} > 0, \qquad \forall h = 0, 1, ... \tag{42}$$

for any choice of $j, k_1, ..., k_j$ as before defined.

We now introduce the processes X, and $\mathcal{X}$, defined on $\mathbf{T}$, which will play a basic role in the sequel. Let x be an $\mathcal{S}$-valued reciprocal chain satisfying the nonsingularity assumption (42), and let $\nu < +\infty$ be the *maximal* polynomial degree in the CE expression (40). Let p be any positive integer $p \geq \nu$. We define

$$X(p;k) = \left[x_{[i]}(k)\right]_{i=1,...,p}, \tag{43}$$

$$X'(p;k) = \left[x^{[i]}(k)\right]_{i=1,...,p}. \tag{44}$$

Moreover, for $k \in \mathbf{T}$ such that $k \pm 1 \in \mathbf{T}$

$$\mathcal{X}(p;k) = \left[x^{\{i\}}(k)\right]_{i=2,...,p}\quad, \tag{45}$$

where, for any $\gamma = 2, ..., p$, the vector $x^{\{i\}}(k)$ is

$$x^{\{\gamma\}}(k) = \left[x_{[i]}(k-1) \otimes x_{[\gamma-i]}(k+1)\right]_{i=1,...,\gamma-1}\quad. \tag{46}$$

As to the dimension of above vectors, we shall denote

$$\alpha = \dim\left(X(p;k)\right); \quad \alpha' = \dim\left(X'(p;k)\right); \quad \beta = \dim\left(\mathcal{X}(p;k)\right).$$

3. Nearest-neighbour polynomial representation

Let x be a nonsingular reciprocal chain satisfying all the assumptions given in §2.5. By the nonsingularity assumption, for any positive integer p, the matrix $P(p;k)$ defined as

$$P(p;k) = \mathrm{Cov}\{\mathcal{X}(p;k), \mathcal{X}(p;k)\}, \tag{47}$$

results in a definite positive matrix for any k. Then we can prove the following theorem.

Theorem 3.1. *For any $h = 1, 2...$ there exists an unique set of $(\nu^2 + 3\nu)/2$ matrices (projection matrices), namely $\mathcal{F}_h(k)$:*

$$\mathcal{F}_h(k) = \{F_{i,j}^{(h)}(k), i, j = 0, ..., \nu, 0 \leq i + j \leq \nu\}, \tag{48}$$

such that the reciprocal, finite-states, process x satisfies the following polynomial equation:

$$x_{[h]}(k) = \sum_{\substack{i,j=0 \\ 0<i+j\leq\nu}}^{\nu} F_{i,j}^{(h)}(k)(x_{[i]}(k-1) \otimes x_{[j]}(k+1)) + F_{0,0}^{(h)}(k) + d_h(k) \tag{49}$$

where $\{d_h, h = 1, 2...\}$ is a set of driving one-step-correlated noise-sequences, that is $\forall r, s = 1, 2...$

$$\mathbf{E}\{d_r(k)d_s^T(l)\} = 0, \qquad \text{for} \quad |k - l| > 1, \tag{50}$$

and satisfies the orthogonality property:

$$\mathbf{E}\{d_h(k)\phi^T(x(i))\} = 0, \quad i \neq k, \quad \forall h = 0, 1, ... \tag{51}$$

for any Borel map $\phi : \mathbb{R}^n \to \mathbb{R}^{\dim(d_h)}$.

Moreover, by arranging the projection matrices as follows:

$$F_\gamma^{(h)}(k) = \Big[F_{i,\gamma-i}^{(h)}(k)\Big]^{i=0,...,\gamma} \quad , \tag{52}$$

with $\gamma = 1, ..., \nu$, and

$$F^{(h)}(k) = \Big[F_i^{(h)}(k)\Big]^{i=1,...,\nu} \quad , \tag{53}$$

one has that the matrices in $\mathcal{F}_h(k) \setminus \{F_{0,0}^{(h)}(k)\}$ can be calculated for any $h = 1, ...$ by taking the (unique) solution of the linear equation

$$F^h(k)P(\nu; k) = Q_h(k), \tag{54}$$

where for any $h = 1, 2...$

$$Q_h(k) = \Big[\mathrm{Cov}\{x_{[h]}(k), x^{\{i\}}(k)\}\Big]^{i=1,...,\nu}$$

$x^{\{\gamma\}}, \gamma = 1, ..., \nu$ given in (46), and $P(\cdot, k)$ defined by (47). Next, $F_{0,0}^{(h)}(k)$ can be obtained by

$$F_{0,0}^{(h)}(k) = \mathbf{E}\left\{x_{[h]}(k)\right\} - \sum_{\substack{i,j=0 \\ 0<i+j\leq\nu}}^{\nu} F_{i,j}^{(h)}(k)\mathbf{E}\left\{x_{[i]}(k-1) \otimes x_{[j]}(k+1)\right\}. \tag{55}$$

Proof. For any $h = 1, 2...$ define

$$d_h(k) = x_{[h]}(k) - \mathbf{E}\{x_{[h]}(k)/x(i), i \neq k\}, \tag{56}$$

which results in a zero-mean process satisfying the orthogonality property (51). Moreover, from (51) it readily follows that, $\forall h, h' = 0, 1, ...$

$$\mathbf{E}\{d_h(k)d_{h'}^T(k)\} = \mathbf{E}\{d_h(k)x_{[h']}^T(k)\}, \quad i \neq k. \tag{57}$$

Now, by (37), the general expression of the process d_h, defined in (56), is

$$d_h(k)=x_{[h]}(k)-\sum_{\substack{i,j=0\\ i+j\le\nu}}^{\nu} F_{i,j}^{(h)}(k)(x_{[i]}(k-1)\otimes x_{[j]}(k+1)). \tag{58}$$

Thus, by rewriting (58), it has been shown that the process x has to satisfy the polynomial equation (49) for some set $\{F_{i,j}^{(h)}(k)\}$ of projection matrices. Also, it can be easily checked that the orthogonality property (51) is ensued by (50). Now, for any couple of integers $h'\ge 1$, $h"\ge 0$, by post-multiplying (49) by $(x_{[h']}(s)\otimes x_{[h"]}(l))^T$, $s,l=0,...,N$, with $l\ne k$:

$$\begin{aligned}&\mathbf{E}\left\{x_{[h]}(k)(x_{[h']}(s)\otimes x_{[h"]}(l))^T\right\}\\ &\quad=\sum_{\substack{i,j=0\\ 0<i+j\le\nu}}^{\nu} F_{i,j}^{(h)}(k)\mathbf{E}\left\{(x_{[i]}(k-1)\otimes x_{[j]}(k+1))(x_{[h']}(s)\otimes x_{[h"]}(l))^T\right\}\\ &\qquad+F_{0,0}^{(h)}(k)\mathbf{E}\left\{(x_{[h']}(s)\otimes x_{[h"]}(l))^T\right\}+\mathbf{E}\left\{d_h(k)d_{h'}^T(k)\otimes x_{[h"]}^T(l)\right\}\delta(k-s)\end{aligned} \tag{59}$$

While deriving eq. (59) the following equality has been used

$$\mathbf{E}\left\{d_h(k)(x_{[h']}(s)\otimes x_{[h"]}(l))^T\right\}=\mathbf{E}\left\{d_h(k)d_{h'}^T(k)\otimes x_{[h"]}^T(l)\right\}\delta(k-s),\quad \text{for}\quad l\ne k. \tag{60}$$

In order to prove (60), first note that, by (51),

$$\mathbf{E}\left\{d_h(k)\mathbf{E}\left\{x_{[h']}^T(k)\otimes x_{[h"]}^T(l)/x(i),i\ne k\right\}\right\}=0,$$

thus

$$\begin{aligned}&\mathbf{E}\left\{d_h(k)(x_{[h']}(k)\otimes x_{[h"]}(l))^T\right\}\\ &\quad=\mathbf{E}\left\{d_h(k)\Big(x_{[h']}^T(k)\otimes x_{[h"]}^T(l)-\mathbf{E}\left\{x_{[h']}^T(k)\otimes x_{[h"]}^T(l)/x(i),i\ne k\right\}\Big)\right\}\\ &\quad=\mathbf{E}\left\{d_h(k)\Big(x_{[h']}^T(k)-\mathbf{E}\left\{x_{[h']}^T(k)/x(i),i\ne k\right\}\Big)\otimes x_{[h"]}^T(l)\right\}\\ &\quad=\mathbf{E}\left\{(d_h(k)d_{h'}^T(k))\otimes x_{[h"]}^T(l)\right\}\end{aligned}$$

where, by recalling that $l\ne k$, the measurability of $x_{[h"]}^T(l)$ w.r.t. $\{x(i),i\ne k\}$ has been exploited. Now, if both $s,l\ne k$ one has $\mathbf{E}\left\{d_h(k)(x_{[h']}(s)\otimes x_{[h"]}(l))^T\right\}=0$, by orthogonality. Thus, for $s,l=0,...,N$ and $l\ne k$, equality (60) ensues.

Taking expectations in (58) and keeping away the term $F_{0,0}^{(h)}(k)$ from the summation is ensued by eq. (55). By substituting (55) in (59) the following covariance expression is readily derived:

$$\begin{aligned}&\mathrm{Cov}\left\{x_{[h]}(k),x_{[h']}(s)\otimes x_{[h"]}(l)\right\}\\ &\quad=\sum_{\substack{i,j=0\\ 0<i+j\le\nu}}^{\nu} F_{i,j}^{(h)}(k)\mathrm{Cov}\left\{x_{[i]}(k-1)\otimes x_{[j]}(k+1),x_{[h']}(s)\otimes x_{[h"]}(l)\right\}\\ &\qquad+\mathbf{E}\left\{\big(d_h(k)d_{h'}^T(k)\big)\otimes x_{[h"]}^T(l)\right\}\delta(k-s)\end{aligned} \tag{61}$$

which holds $\forall k \neq l$. In particular, for $s = k-1$, $l = k+1$, $h' + h" = p$ with $p = 1, 2...$, and by rewriting the summation with respect to the new index $\gamma = 1, ..., \nu$, eq. (61) can be rewritten as

$$\mathrm{Cov}\left\{x_{[h]}(k), x_{[h']}(k-1) \otimes x_{[p-h']}(k+1)\right\}$$
$$= \sum_{\gamma=1}^{\nu}\sum_{i=0}^{\gamma} F^{(h)}_{i,\gamma-i}(k)\mathrm{Cov}\left\{x_{[i]}(k-1) \otimes x_{[\gamma-i]}(k+1), x_{[h']}(k-1) \otimes x_{[p-h']}(k+1)\right\} \quad (62)$$

with $h' = 0, ..., p$. Equations (62) can be aggregated for $p = 1, ..., \nu$ so that one writes down eq. (54) . Thus, the theorem is proven as soon as it is noticed that, on account of nonsingularity assumption, $P(k)$ is symmetric positive-definite for any k. ∎

Now, let be given, for any fixed integer $\mu \geq \nu$, for $h = 1, ..., \mu$, the sets of projection matrices $\mathcal{F}_h(k)$, (48). Then the noise-variance, and the one-step correlation, are computable as shown in the following theorem.

Theorem 3.2. *Let us define*

$$d(k) = [d_i(k)]_{i=1,...,\mu}\,, \quad (63)$$

and denote

$$D_{r,s}(k,l) = \mathbf{E}\{d_r(k)d_s^T(l)\}, \quad (64)$$
$$D(k,l) = \mathbf{E}\{d(k)d^T(l)\} = [D_{r,s}(k,l)]^{s=1,...,\mu}_{r=1,...,\mu}\,, \quad (65)$$

the auto-correlation function of the "input noise" d. Then

$$D(k,l) = 0, \qquad \text{for} \quad |k-l| > 1, \quad (66)$$

and for $h, h' = 1, ..., \mu$:

$$D_{h,h'}(k,k) = \mathrm{Cov}\left\{x_{[h]}(k), x_{[h']}(k)\right\} - \sum_{\substack{i,j=0 \\ 0<i+j\leq\nu}}^{\nu} F^{(h)}_{i,j}(k)\mathrm{Cov}\left\{x_{[i]}(k-1)\otimes x_{[j]}(k+1), x_{[h']}(k)\right\}. \quad (67)$$

Moreover, the one-step correlation $D(k, k+1)$ is given by ($r, s = 1, ..., \mu$):

$$\begin{aligned} D_{r,s}(k,k+1) &= -\sum_{\substack{i,j=0 \\ i+j\leq\nu}}^{\nu}\left(D_{r,i}(k,k)\otimes\mathbf{E}\left\{x^T_{[j]}(k+2)\right\}\right)F^{(s)^T}_{i,j}(k+1) \\ &= -\sum_{\substack{i,j=0 \\ i+j\leq\nu}}^{\nu} F^{(r)^T}_{i,j}(k)\left(\mathbf{E}\left\{x^T_{[i]}(k-1)\right\}\otimes D_{j,s}(k+1,k+1)\right) \end{aligned} \quad (68)$$

Proof. Theorem 3.1 immediately lead to eq. (66). Given the $F^{(h)}_{i,j}(k)$'s, by setting $h" = 0$ and $k = s$ in (61) we do find eq. (67). Substituting $d_s(k+1)$ – while using (58) to compute it – and by property (51), yields

$$D_{r,s}(k,k+1) = -\sum_{\substack{i,j=0 \\ i+j\leq\nu}}^{\nu}\mathbf{E}\left\{d_r(k)(x_{[i]}(k)\otimes x_{[j]}(k+2))^T\right\}F^{(s)^T}_{i,j}(k+1). \quad (69)$$

Since $x_{[j]}(k+2)$ is measurable w.r.t. (the σ-algebra generated by) $\{x(i), i \neq k\}$, one has

$$\mathbf{E}\left\{x_{[i]}(k) \otimes x_{[j]}(k+2) \Big/ x(i), i \neq k\right\} = (x_{[i]}(k) - d_i(k)) \otimes x_{[j]}(k+2) \tag{70}$$

and by taking into account that the conditional expectation in (70) is orthogonal – by (51) – to $d_r(k)$, using (70) in (69) results in

$$D_{r,s}(k, k+1) = -\sum_{\substack{i,j=0 \\ i+j\leq\nu}}^{\nu} \mathbf{E}\left\{d_r(k)(d_i(k) \otimes x_{[j]}(k+2))^T\right\} {F_{i,j}^{(s)}}^T(k+1). \tag{71}$$

By using property (23) it is

$$(d_r(k) \otimes 1) \cdot (d_i^T(k) \otimes x_{[j]}^T(k+2)) = d_r(k)d_i^T(k) \otimes 1 \cdot x_{[j]}^T(k+2),$$

where "1" has to be understood being the 1×1-matrix which has the only element 1. Using the latter formula in (71), while accounting of the mutual independence of $d_.(k)$ and $x(k+2)$, gives

$$D_{r,s}(k, k+1) = -\sum_{\substack{i,j=0 \\ i+j\leq\nu}}^{\nu} \mathbf{E}\left\{d_r(k)d_i^T(k)\right\} \otimes \mathbf{E}\left\{x_{[j]}^T(k+2)\right\} {F_{i,j}^{(s)}}^T(k+1). \tag{72}$$

Thus, expression (68) has been proven, where in particular – by substituting the expression of $d_r(k)$, given by (58), in $D_{r,s}(k, k+1) = \mathbf{E}\{d_r(k)d_s^T(k+1)\}$ – the latter expression is similarly derived. ∎

By eq. (68) the computation of any $D_{r,s}(k, k+1)$ requires the klowledge of the two sets of projection matrices $\mathcal{F}_s(k+1)$ and $\mathcal{F}_r(k)$, thus, the couple of equations (68), (67) allows the recursive computation of the whole correlation function of the input noise $d_h(k)$, for any $h = 1, ..., \mu$, for any fixed integer $\mu \geq \nu$, provided the joint-moments of process x are available up to a degree 2μ in any subinterval $[k-1, k+1] \subset \mathbf{T}$.

4. The exactly-linearized nearest-neighbour model

Throughout this section $x = \{x(k), k \in \mathbf{T}\} \subset \mathbb{R}^n$ denotes a reciprocal chain satisfying all the assumptions given in §2. Let us fix a positive integer $\mu \geq \nu$ and define, for any $k \in \mathbf{T}$

$$X(k) = X(\mu, k), \tag{73}$$

and for any $k \in \mathbf{T}$, such that $k \pm 1 \in \mathbf{T}$:

$$\mathcal{X}(k) = \mathcal{X}(\mu, k), \tag{74}$$

where $X(\cdot, k), \mathcal{X}(\cdot, k)$ has been defined in (43), (45) respectively. From now onwards, within the present section, as we state that some property holds for any $k \in \mathbf{T}$ we will mean that it actually holds where it is well defined; for example, we say that vector (74) is defined for any $k \in \mathbf{T}$ even though actually it is only defined for $k \in \mathbf{T}$ such that $k \pm 1 \in \mathbf{T}$. It makes no difference being $\mathbf{T}$ either the unit circle or the set of all integers, but in the case of a finite discrete interval $[0, N]$, when it denotes only the interior of the interval.

The present section includes the main result (Theorem 4.2), where it is shown that x owns a polynomial representation which is *exactly linearizable*. We have so far discussed such issue briefly in the introduction, where it has been pointed out that, by using the vectors (73), (74), the polynomial representation (49) can be put in the form (5), which is a linear representation in the augmented variables. However, we cannot say that (5) is a *linearized* model of x because, by theorem 3.1, we've got an equation for each power $x_{[h]}(k)$, whereas we should need to write down additional equations even for the *mixed* powers, i.e. : $x_{[i]}(k-1) \otimes x_{[j]}(k+1)$.

Now, the main result of this section is that any reciprocal chain x, provided it satisfies the mild hypotheses given in §2.5, own the following property: in the polynomial representation of x, given by eq. (49), one has, for any h, $F_{i,j}^{(h)}(k) = 0$ whatever $(i,j) \neq (0,0)$. In other words, the projection matrices *associated to mixed-powers* are zero. Therefore, the matrix $F(k)$ in (5) vanishes, and eq. (5) is actually an *exactly linearized* model, as though being a nonlinear (polynomial) model, anyhow it can be rewritten as a nearest-neighbour *linear* model with respect to the new variable $X(k)$. This model will be next (corollary 4.4) put in a 'normalized' version, formally similar to the Levi-Frezza-Krener linear model (1) describing the Gaussian reciprocal processes.

Before proving this *representation theorem*, we need to prove the following result (Theorem 4.1) which generalize Theorem 3.1. Representation (49) was derived, essentially, by first exploiting the reciprocity property of x – applying (28) by choosing $f(\xi) = \xi_{[h]}$ – and next by using the polynomial representation of the conditional expectation given in (36). On the other hand, the mixed-power process: $x_{[i]}(k-1) \otimes x_{[j]}(k+1)$ *is not* reciprocal, thus the results of Theorem 3.1 cannot be straightforwardly extended to mixed-powers. Nevertheless, we can show that the processes ξ^+ and ξ^-, defined as

$$\xi_{h,r}^{\pm}(k) = x_{[h]}(k\pm 2) \otimes x_{[r]}(k)$$

are (in the subset of $\mathbf{T}$ where they are defined) 'reciprocal' in a wide sense. Indeed, they are *reciprocal respect to* x

$$\mathbf{E}\left\{\xi_{h,r}^{\pm}(k)/x(i), i \neq k\right\} = \mathbf{E}\left\{\xi_{h,r}^{\pm}(k)/x(k+1), x(k-1)\right\}, \tag{75}$$

for any couple of positive integers h, r. This will be the key which allows us, in the forthcoming theorem, to get the desired polynomial equations even for mixed-powers.

Theorem 4.1. *For any couple of integers h, r, with $h \geq 0$, and $r > 0$, there exist sets of projection matrices, namely $\mathcal{G}_{h,r}^{-}(k)$, $\mathcal{G}_{h,r}^{+}(k)$, having each a number $(\nu^2 + 3\nu)/2$ of matrices*

$$\mathcal{G}_{h,r}^{\pm}(k) = \{G_{i,j}^{(h,r)^{\pm}}(k), i, j = 0, ..., \nu, 0 \leq i + j \leq \nu\}, \tag{76}$$

such that x satisfies the following polynomial stochastic equation (almost surely)

$$x_{[h]}(k\pm 2) \otimes x_{[r]}(k) = \sum_{\substack{i,j=0 \\ 0<i+j\leq\nu}}^{\nu} G_{i,j}^{(h,r)^{\pm}}(k)(x_{[i]}(k-1) \otimes x_{[j]}(k+1)) + G_{0,0}^{(h,r)^{\pm}}(k) + d_{h,r}^{\pm}(k) \tag{77}$$

where $d_{h,r}^{+}$ and $d_{h,r}^{-}$, are, for any $h, r = 1, 2...$ driving one-step-correlated, zero mean, noise-sequences. In specific, denoting for any $h, h', r, r' = 1, 2...$

$$D_{h,h'}^{r,r'^{\pm}}(k,l) = \mathbf{E}\left\{d_{h,r}^{\pm}(k) {d_{h',r'}^{\pm}}^{T}(l)\right\} \tag{78}$$

it is

$$\mathbf{E}\left\{d^{\pm}_{h,r}(k)\right\} = 0, \qquad \forall k, \tag{79}$$

$$D^{r,r'\,\pm}_{h,h'}(k,l) = 0 \qquad \text{for} \quad |l-k| > 1. \tag{80}$$

$$D^{r,r'\,\pm}_{h,h'}(k,k) = \mathbf{E}\left\{x_{[h]}(k\pm2)x^T_{[h']}(k\pm2)\right\}\otimes D_{r,r'}(k,k), \tag{81}$$

$$D^{r,r'\,\pm}_{h,h'}(k,k-1) = \mathbf{E}\left\{x_{[h]}(k\pm2)x^T_{[h']}(k-1\pm2)\right\}\otimes D_{r,r'}(k,k-1), \tag{82}$$

$$D^{r,r'\,\pm}_{h,h'}(k,k+1) = \mathbf{E}\left\{x_{[h]}(k\pm2)x^T_{[h']}(k+1\pm2)\right\}\otimes D_{r,r'}(k,k+1), \tag{83}$$

where $D_{\cdot,\cdot}(\cdot,\cdot)$ *is the correlation function of the process* (63)*, given by the recursive equations* (69), (70)*. All the processes* $d_{h,r}$ *satisfy the orthogonality property, i.e. for any Borel map* $\phi : \mathbb{R}^n \to \mathbb{R}^{1\times n^{h+r}}$

$$\mathbf{E}\{d^{\pm}_{h,r}(k)\phi(x(i))\} = 0, \quad i \neq k, \quad \forall h, r = 1,2,... \tag{84}$$

Finally, one has

$$d^{+}_{0,r}(k) = d^{-}_{0,r}(k) = d_r(k), \quad \forall r = 1,2,... \tag{85}$$

Proof. Let us define, for any couple of integers $h \geq 0$, $r > 0$

$$d^{\pm}_{h,r}(k) = x_{[h]}(k\pm2)\otimes x_{[r]}(k) - \mathbf{E}\left\{x_{[h]}(k\pm2)\otimes x_{[r]}(k)\big/x(i), i\neq k\right\}. \tag{86}$$

Equation (86) reduces to (56) for $h = 0$, thus eq. (85) immediately ensues. From measurability of the random variables $x_{[h]}(k\pm2)$ respect to $\{x(i), i \neq k\}$, and on account of the reciprocity of $x_{[r]}$ it follows that

$$\begin{aligned}\mathbf{E}\left\{x_{[h]}(k\pm2)\otimes x_{[r]}(k)\big/x(i), i\neq k\right\} &= x_{[h]}(k\pm2)\otimes\mathbf{E}\left\{x_{[r]}(k)\big/x(i), i\neq k\right\}\\ &= x_{[h]}(k\pm2)\otimes\mathbf{E}\left\{x_{[r]}(k)\big/x(k-1), x(k+1)\right\}.\end{aligned} \tag{87}$$

On the other hand, for any random vector ϕ_k, taking values in $\mathbb{R}^{n^{h+r}}$, and measurable respect to
$\{x(k-1), x(k+1)\}$, it is

$$\begin{aligned}&\mathbf{E}\left\{\left(x_{[h]}(k\pm2)\otimes x_{[r]}(k) - x_{[h]}(k\pm2)\otimes\mathbf{E}\left\{x_{[r]}(k)\big/x(k-1),x(k+1)\right\}\right)\phi^T_k\right\}\\ &\qquad = \mathbf{E}\left\{\left(x_{[h]}(k\pm2)\otimes\left(x_{[r]}(k) - \mathbf{E}\left\{x_{[r]}(k)\big/x(i), i\neq k\right\}\right)\right)\phi^T_k\right\}.\end{aligned} \tag{88}$$

Denoting $x^l_{[h]}(k\pm2)$ the l-th (scalar) component of vector $x_{[h]}(k\pm2)$, by definition of Kronecker product, the second line of (88) is the aggregate, for $l =, 1..., n^h$, of the following blocks:

$$\mathbf{E}\left\{\left(x_{[r]}(k) - \mathbf{E}\left\{x_{[r]}(k)\big/x(i), i\neq k\right\}\right)x^l_{[h]}(k\pm2)\phi^T_k\right\} = 0, \tag{89}$$

which vanish, as $x^l_{[h]}(k\pm2)\phi_k^T$ is $\sigma\{x(i), i\neq k\}$-measurable. Thus the first term in (88) vanishes for any ϕ_k $\sigma\{x(i), i\neq k\}$-measurable, therefore

$$x_{[h]}(k\pm2)\otimes \mathbf{E}\left\{x_{[r]}(k)\Big/x(k-1), x(k+1)\right\} = \mathbf{E}\left\{x_{[h]}(k\pm2)\otimes x_{[r]}(k)\Big/x(k-1), x(k+1)\right\}, \quad (90)$$

so the comparison with (87) is ensued by

$$\mathbf{E}\left\{x_{[h]}(k\pm2)\otimes x_{[r]}(k)\Big/x(i), i\neq k\right\} = \mathbf{E}\left\{x_{[h]}(k\pm2)\otimes x_{[r]}(k)\Big/x(k-1), x(k+1)\right\}. \quad (91)$$

Now, by reason of (40) with

$$\bar{\phi}_k(x) = x_{[h]}(k\pm 2)\otimes x_{[r]}(k), \quad (92)$$

there exists matrices $G^{(\cdot,\cdot)^{\pm}}_{\cdot,\cdot}(k)$ – indeed $G^{(h,r)}_{i,j}(k) = \mathbf{c}^{(\bar{\phi})}_{i,j}(k)$ with $\bar{\phi}$ as in (92) – such that

$$\mathbf{E}\left\{x_{[h]}(k\pm2)\otimes x_{[r]}(k)\Big/x(k-1), x(k+1)\right\} = \sum_{\substack{i,j=0\\ 0\leq i+j\leq \nu}}^{\nu} G^{(h,r)^{\pm}}_{i,j}(k)(x_{[i]}(k-1)\otimes x_{[j]}(k+1)), \quad (93)$$

Thus, by using the above expression and (91) in (86), eq. (77) is derived.

Let us prove now that the processes $d^{\pm}_{\cdot,\cdot}$ satisfy (79)–(83) and the orthogonality relation (84). Properties (79) and (84) immediately follow from definition (86). Furthermore, by (86), (87):

$$\begin{aligned} d^{\pm}_{h,r}(k) &= x_{[h]}(k\pm2)\otimes\left(x_{[r]}(k) - \mathbf{E}\left\{x_{[r]}(k)\Big/x(k-1), x(k+1)\right\}\right)\\ &= x_{[h]}(k\pm2)\otimes d_r(k) \end{aligned} \quad (94)$$

where, by Theorem 2.1, $d_h, h = 1, 2...$ is the set of driving one-step-correlated noise sequences in eq. (49), defined in eq. (56). By using the expression of $d_{h,r}$ given by (94) one has

$$\begin{aligned} \mathbf{E}\left\{d^{\pm}_{h,r}(k)d^{\pm}_{h',r'}(l)\right\} &= \mathbf{E}\left\{(x_{[h]}(k\pm2)\otimes d_r(k))(x^T_{[h']}(l\pm2)\otimes d^T_{r'}(l))\right\}\\ &= \mathbf{E}\left\{\left(x_{[h]}(k\pm2)x^T_{[h']}(l\pm2)\right)\otimes\left(d_r(k)d^T_{r'}(l)\right)\right\}. \end{aligned} \quad (95)$$

Now, recall that $d_h(k)$, by definition given in (56), for any h and k is equal to some Borel function of $x(k-1), x(k), x(k+1)$, and it is independent of $\{x(i), i\neq k\}$. Thus, for $l = k$ since $d_{\cdot}(k)$ is independent of $x(k\pm2)$, eq. (81) ensues. Also, since both $d_r(k)$ and $d_{r'}(k-1)$ are independent of $\{x(k\pm2), x(k-1\pm2)\}$, the product $d_r(k)d^T_{r'}(k-1)$ is independent as well, therefore eq. (95) for $l = k-1$ implies eq. (82). Similarly, $d_r(k)d^T_{r'}(k+1)$ is independent of $\{x(k\pm2), x(k+1\pm2)\}$ hence eq. (95) for $l = k+1$ is ensued by eq. (83). In order to prove (80) consider first $l < k-1$, in which case $d_r(k)$ is independent of $\{x(k-2), x(l-2), d_{r'}(l)\}$, and $d_{r'}(l)$ is independent of $\{x(k+2), x(l+2), d_r(k)\}$, therefore eq. (95) rewrites

$$\begin{aligned} \mathbf{E}\left\{d^{-}_{h,r}(k)d^{-}_{h',r'}(l)\right\} &= \mathbf{E}\left\{\left(x_{[h]}(k-2)x^T_{[h']}(l-2)\otimes\mathbf{E}\left\{d_r(k)\right\}d^T_{r'}(l)\right)\right\} = 0,\\ \mathbf{E}\left\{d^{+}_{h,r}(k)d^{+}_{h',r'}(l)\right\} &= \mathbf{E}\left\{\left(x_{[h]}(k+2)x^T_{[h']}(l+2)\otimes d_r(k)\mathbf{E}\left\{d^T_{r'}(l)\right\}\right)\right\} = 0. \end{aligned}$$

For $l > k+1$, since $d_{r'}(l)$ is independent of $\{x(k-2), x(l-2), d_r(k)\}$, and $d_r(k)$ is independent of $\{x(k+2), x(l+2), d_{r'}(l)\}$, eq. (95) becomes

$$\begin{aligned} \mathbf{E}\left\{d^{-}_{h,r}(k)d^{-}_{h',r'}(l)\right\} &= \mathbf{E}\left\{\left(x_{[h]}(k-2)x^T_{[h']}(l-2)\otimes d_r(k)\mathbf{E}\left\{d^T_{r'}(l)\right\}\right)\right\} = 0,\\ \mathbf{E}\left\{d^{+}_{h,r}(k)d^{+}_{h',r'}(l)\right\} &= \mathbf{E}\left\{\left(x_{[h]}(k+2)x^T_{[h']}(l+2)\otimes\mathbf{E}\left\{d_r(k)\right\}d^T_{r'}(l)\right)\right\} = 0, \end{aligned}$$

thus, the proof is endowed. ∎

Equation (77) includes eq. (49) as a particular case (i.e. by setting $h = 0$ in (77)). Hence, a subset of the projection matrices (76) has been just calculated by eq. (54) of theorem 3.1, i.e. for any $h = 1, ...,$

$$G_{i,j}^{0,h^-}(k) = G_{i,j}^{0,h^+}(k) = F_{i,j}^{h}(k), \tag{96}$$

$$G_{0,0}^{0,h^-}(k) = G_{0,0}^{0,h^+}(k) = F_{0,0}^{h}(k), \tag{97}$$

We can now state and prove the *representation theorem.*

Theorem 4.2. *(Representation theorem). For any positive integer $\mu \geq \nu$, and for any $k \in \mathbf{T}$, there exist matrices $\Phi(k), \Psi(k), H(k), L(k)$ such that the processes X, and $\mathcal{X}$ defined in (73) and (74) satisfy the following equations:*

$$\widetilde{X}(k) = d(k) + \Phi(k)\widetilde{X}(k+1) + \Psi(k)\widetilde{X}(k-1), \tag{98}$$

$$\widetilde{\mathcal{X}}(k) = b(k) + H(k)\widetilde{X}(k+1) + L(k)\widetilde{X}(k-1), \tag{99}$$

where d is the one-step correlated noise defined in (56), and b is a zero-mean, white sequence, uncorrelated with d, that is for any k:

$$\mathbf{E}\{b(k)\} = 0, \tag{100}$$

$$\mathbf{E}\left\{b(k)b^T(l)\right\} = 0, \qquad l \neq k, \tag{101}$$

$$\mathbf{E}\left\{d(k)b^T(l)\right\} = 0, \qquad \forall l. \tag{102}$$

The covariance of $b(k)$, for any k, is given by:

$$B(k) = \mathbf{E}\left\{b(k)b^T(k)\right\} = \left[B_{p,p'}(k,k)\right]_{p=1,\dots,\nu}^{p'=1,\dots,\nu} \tag{103}$$

$$B_{p,p'}(k,k) = \left[D_{h,h'}^{p-h,p'-h'^-}(k+1,k+1)\right]_{h=0,\dots,p}^{h'=0,\dots,p} \tag{104}$$

where $D_{\cdot,\cdot}^{\cdot,\cdot^-}(\cdot,\cdot)$ is defined in (81). Moreover, denoting $D(k) = D(k,k)$ the covariance of d ($D(\cdot,\cdot)$ being defined in (65)), we have for any k

$$D(k) > 0, \tag{105}$$

$$B(k) > 0. \tag{106}$$

Finally, d and b satisfy the following orthogonality properties:

$$\mathbf{E}\left\{d(k)X^T(l)\right\} = 0, \qquad l \neq k, \tag{107}$$

$$\mathbf{E}\left\{d(k)\mathcal{X}^T(l)\right\} = 0, \qquad l \neq k-1, k+1, \tag{108}$$

$$\mathbf{E}\left\{b(k)X^T(l)\right\} = 0, \qquad \forall l. \tag{109}$$

$$\mathbf{E}\left\{b(k)\mathcal{X}^T(l)\right\} = 0, \qquad l \neq k, \tag{110}$$

and the non-zero terms of the mutual covariance between $\{d, b\}$ and $\{X, \mathcal{X}\}$ are given by

$$\mathbf{E}\left\{d(k)\widetilde{X}^T(k)\right\} = \mathbf{E}\{d(k)d^T(k)\} = D(k), \tag{111}$$

$$\mathbf{E}\left\{d(k)\widetilde{\mathcal{X}}^T(k-1)\right\} = D(k)H^T(k-1), \tag{112}$$

$$\mathbf{E}\left\{d(k)\widetilde{\mathcal{X}}^T(k+1)\right\} = D(k)L^T(k+1), \tag{113}$$

$$\mathbf{E}\left\{b(k)\widetilde{\mathcal{X}}^T(k)\right\} = \mathbf{E}\left\{b(k)b^T(k)\right\} = B(k). \tag{114}$$

Proof. For any h, r, integers, and $k \in \mathbf{T}$, let us allow the indeces i, j of the projection matrices $G_{i,j}^{(h,r)^{\pm}}(k)$ to run up to $\mu \geq \nu$ – indeed the projection matrices in (76) was defined only for $i, j = 0, \ldots\nu$ – by setting $G_{i,j}^{(h,r)^{\pm}}(k) = 0$, for $\nu < i, j \leq \mu$. Thus, by (96), (97), the definition of the projection matrices of (48) results to be extended as well. Then, we can define a set of nine matrices, namely $F(k), \Phi(k), \Psi(k), G_{\pm}(k), H_{\pm}(k), L_{\pm}(k)$, as the following block-aggregation of the projection matrices (76):

$$F(k) = \left[F^{(h)}(k)\right]_{h=1,\ldots,\mu} \qquad \Phi(k) = \left[\Phi^{(h)}(k)\right]_{h=1,\ldots,\mu} \qquad \Psi(k) = \left[\Psi^{(h)}(k)\right]_{h=1,\ldots,\mu} \tag{115}$$

$$\begin{aligned} F_{\gamma}^{(h)}(k) &= \left[F_{i,\gamma-i}^{(h)}(k)\right]^{i=1,\ldots,\gamma-1} & F^{(h)}(k) &= \left[F_{\gamma}^{(h)}(k)\right]^{\gamma=2,\ldots,\mu} \\ \Phi^{(h)}(k) &= \left[F_{0,\gamma}^{(h)}(k)\right]^{\gamma=1,\ldots,\mu} & \Psi^{(h)}(k) &= \left[F_{\gamma,0}^{(h)}(k)\right]^{\gamma=1,\ldots,\mu} \end{aligned} \tag{116}$$

$$G_{\pm}(k) = \left[G_{\pm}^{(p)}(k)\right]_{p=1,\ldots,\mu} \qquad H_{\pm}(k) = \left[H_{\pm}^{(p)}(k)\right]_{p=1,\ldots,\mu} \qquad L_{\pm}(k) = \left[L_{\pm}^{(p)}(k)\right]_{p=1,\ldots,\mu} \tag{117}$$

$$\begin{aligned} G_{+}^{(p)}(k) &= \left[K_{n^{p-h},n^h}^{T} G_{+}^{(p-h,h)}(k)\right]_{h=1,\ldots,p-1} & H_{+}^{(p)}(k) &= \left[K_{n^{p-h},n^h}^{T} H_{+}^{(p-h,h)}(k)\right]_{h=1,\ldots,p-1} \\ L_{+}^{(p)}(k) &= \left[K_{n^{p-h},n^h}^{T} L_{+}^{(p-h,h)}(k)\right]_{h=1,\ldots,p-1} \end{aligned} \tag{118}$$

$$\begin{aligned} G_{-}^{(p)}(k) &= \left[G_{-}^{(h,p-h)}(k)\right]_{h=1,\ldots,p-1} & H_{-}^{(p)}(k) &= \left[H_{-}^{(h,p-h)}(k)\right]_{h=1,\ldots,p-1} \\ L_{-}^{(p)}(k) &= \left[L_{-}^{(h,p-h)}(k)\right]_{h=1,\ldots,p-1} \end{aligned} \tag{119}$$

$$\begin{aligned} G_{\gamma,\pm}^{(h,r)}(k) &= \left[G_{i,\gamma-i}^{(h,r)^{\pm}}(k)\right]^{i=1,\ldots,\gamma-1} & G_{\pm}^{(h,r)}(k) &= \left[G_{\gamma,\pm}^{(h,r)}(k)\right]^{\gamma=2,\ldots,\mu} \\ H_{\pm}^{(h,r)}(k) &= \left[G_{0,\gamma}^{(h,r)^{\pm}}(k)\right]^{\gamma=1,\ldots,\mu} & L_{\pm}^{(h,r)}(k) &= \left[G_{\gamma,0}^{(h,r)^{\pm}}(k)\right]^{\gamma=1,\ldots,\mu} \end{aligned} \tag{120}$$

Equations (49) and (77) can be assembled in such a way to obtain the following expressions

$$\widetilde{X}(k) = d(k) + F(k)\widetilde{\mathcal{X}}(k) + \Phi(k)\widetilde{X}(k+1) + \Psi(k)\widetilde{X}(k-1), \tag{121}$$

$$\widetilde{\mathcal{X}}(k) = b^{-}(k) + G_{-}(k+1)\widetilde{\mathcal{X}}(k+1) + H_{-}(k)\widetilde{X}(k+1) + L_{-}(k)\widetilde{X}(k-1), \tag{122}$$

$$\widetilde{\mathcal{X}}(k) = b^{+}(k) + G_{+}(k-1)\widetilde{\mathcal{X}}(k-1) + H_{+}(k)\widetilde{X}(k+1) + L_{+}(k)\widetilde{X}(k-1), \tag{123}$$

where (122), and (123) will be named the 'backward' and 'forward' equations respectively, of $\mathcal{X}$. In (122), (123), d is the process defined in (63), and $b^{\pm}$ are the stochastic sequences

$$b^{\pm}(k) = \left[b^{(p)^{\pm}}(k)\right]_{p=1,\ldots,\mu} \tag{124}$$

$$b^{(p)^{\pm}}(k) = \left[b_{h,p-h}^{\pm}(k)\right]_{h=1,\ldots,p-1} \tag{125}$$

$$b_{h,r}^{+}(k) = \mathbf{K}_{n^r,n^h}^{T} d_{r,h}^{+}(k-1), \tag{126}$$

$$b_{h,r}^{-}(k) = d_{h,r}^{-}(k+1), \tag{127}$$

where $\mathbf{K}_{\cdot,\cdot}$ is the commutation matrix defined in (19), $d^{\pm}$ are the forcing noises in the polynomial equation (77).

Let us denote, for any h, k, r, s,

$$\overline{x_{[h]}(k)\otimes x_{[r]}(s)} = x_{[h]}(k)\otimes x_{[r]}(s) - \mathbf{E}\{x_{[h]}(k)\otimes x_{[r]}(s)\}.$$

In order to derive eq. (121), first of all perform the change of index: $\gamma = i+j$ and rewrite eq. (55) as

$$\mathbf{E}\left\{x_{[h]}(k)\right\} = F_{0,0}^{(h)}(k) + \sum_{\gamma=1}^{\nu}\sum_{i=0}^{\gamma} F_{i,\gamma-i}^{(h)}(k)\mathbf{E}\left\{x_{[i]}(k-1)\otimes x_{[\gamma-i]}(k+1)\right\} \tag{128}$$

Subtracting (128) to eq. (49) results in

$$\begin{aligned}
x_{[h]}(k) - \mathbf{E}\left\{x_{[h]}(k)\right\} \quad &= d_h(k) + \sum_{\gamma=1}^{\nu}\sum_{i=0}^{\gamma} F_{i,\gamma-i}^{(h)}(k)\overline{x_{[i]}(k-1)\otimes x_{[\gamma-i]}(k+1)} \\
&= d_h(k) + \sum_{\gamma=2}^{\nu}\sum_{i=1}^{\gamma-1} F_{i,\gamma-i}^{(h)}(k)\overline{x_{[i]}(k-1)\otimes x_{[\gamma-i]}(k+1)} \\
&\quad + \sum_{\gamma=1}^{\nu} F_{0,\gamma}^{(h)}(k)\left(x_{[\gamma]}(k+1) - \mathbf{E}\left\{x_{[\gamma]}(k+1)\right\}\right) \\
&\quad + \sum_{\gamma=1}^{\nu} F_{\gamma,0}^{(h)}(k)\left(x_{[\gamma]}(k-1) - \mathbf{E}\left\{x_{[\gamma]}(k-1)\right\}\right) \\
&= d_h(k) + \sum_{\gamma=2}^{\nu} F_{\gamma}^{(h)}(k)\left(x^{\{\gamma\}}(k) - \mathbf{E}\left\{x^{\{\gamma\}}(k)\right\}\right) \\
&\quad + \sum_{\gamma=1}^{\nu} F_{0,\gamma}^{(h)}(k)\left(x_{[\gamma]}(k+1) - \mathbf{E}\left\{x_{[\gamma]}(k+1)\right\}\right) \\
&\quad + \sum_{\gamma=1}^{\nu} F_{\gamma,0}^{(h)}(k)\left(x_{[\gamma]}(k-1) - \mathbf{E}\left\{x_{[\gamma]}(k-1)\right\}\right),
\end{aligned} \tag{129}$$

where in last equality the block-matrix $F_{\gamma}^{(h)}$ defined in (116) has been substituted, as well as the block-vector $x^{\{\gamma\}}$ defined in (46). Now, by using the block-matrices (116) equation (129) can be rewritten:

$$\begin{aligned}
x_{[h]}(k) - \mathbf{E}\left\{x_{[h]}(k)\right\} \quad &= d_h(k) + F^{(h)}(k)\left(\mathcal{X}(k) - \mathbf{E}\left\{\mathcal{X}(k)\right\}\right) \\
&\quad + \Phi^{(h)}(k)\left(X(k+1) - \mathbf{E}\left\{X(k+1)\right\}\right) \\
&\quad + \Psi^{(h)}(k)\left(X(k-1) - \mathbf{E}\left\{X(k-1)\right\}\right).
\end{aligned} \tag{130}$$

Finally, using (115) equation (130) rewrites as representation (121).

Let us derive now the backward and forward equations (122), (123). To this purpose rewrite eq. (77) by eliminating $G_{0,0}^{(h,r)\pm}$ (take expectations in (77) then subtract the result to (77)), and perform the change of index $\gamma = i{+}j$, the result is

$$\begin{aligned}\overline{x_{[h]}(k\pm2)\otimes x_{[r]}(k)} &= d^{\pm}_{h,r}(k) + \sum_{\gamma=1}^{\nu}\sum_{i=0}^{\gamma} G^{(h,r)\,\pm}_{i,\gamma-i}(k)\overline{x_{[i]}(k-1)\otimes x_{[\gamma-i]}(k+1)} \\ &= d^{\pm}_{h,r}(k) + \sum_{\gamma=2}^{\nu}\sum_{i=1}^{\gamma-1} G^{(h,r)\,\pm}_{i,\gamma-i}(k)\overline{x_{[i]}(k-1)\otimes x_{[\gamma-i]}(k+1)} \\ &\quad + \sum_{\gamma=1}^{\nu} G^{(h,r)\,\pm}_{0,\gamma}(k)\left(x_{[\gamma]}(k+1) - \mathbf{E}\left\{x_{[\gamma]}(k+1)\right\}\right) \\ &\quad + \sum_{\gamma=1}^{\nu} G^{(h,r)\,\pm}_{\gamma,0}(k)\left(x_{[\gamma]}(k-1) - \mathbf{E}\left\{x_{[\gamma]}(k-1)\right\}\right).\end{aligned} \tag{131}$$

Setting $r = p{-}h$ in equation (131), and using the block-matrices (120), results in

$$\begin{aligned}\overline{x_{[h]}(k\pm2)\otimes x_{[p-h]}(k)} &= d^{\pm}_{h,p-h}(k) + \sum_{\gamma=2}^{\nu} G^{(h,p-h)}_{\gamma,\pm}(k)\left(x^{\{\gamma\}}(k) - \mathbf{E}\left\{x^{\{\gamma\}}(k)\right\}\right) \\ &\quad + \sum_{\gamma=1}^{\nu} G^{(h,p-h)\,\pm}_{0,\gamma}(k)\left(x_{[\gamma]}(k+1) - \mathbf{E}\left\{x_{[\gamma]}(k+1)\right\}\right) \\ &\quad + \sum_{\gamma=1}^{\nu} G^{(h,p-h)\,\pm}_{\gamma,0}(k)\left(x_{[\gamma]}(k-1) - \mathbf{E}\left\{x_{[\gamma]}(k-1)\right\}\right) \\ &= d^{\pm}_{h,p-h}(k) + G^{(h,p-h)}_{\pm}(k)\left(\mathcal{X}(k) - \mathbf{E}\left\{\mathcal{X}(k)\right\}\right) \\ &\quad + H^{(h,p-h)}_{\pm}(k)\left(X(k+1) - \mathbf{E}\left\{X(k+1)\right\}\right) \\ &\quad + L^{(h,p-h)}_{\pm}(k)\left(X(k-1) - \mathbf{E}\left\{X(k-1)\right\}\right).\end{aligned} \tag{132}$$

Now, by grouping together (132), with the minus sign, for $h = 1, ..., p{-}1$ we get

$$\begin{aligned}x^{\{p\}}(k-1) - \mathbf{E}\left\{x^{\{p\}}(k-1)\right\} &= b^{(p)^-}(k-1) + G^{(p)}_{-}(k)(\mathcal{X}(k) - \mathbf{E}\left\{\mathcal{X}(k)\right\}) \\ &\quad + H^{(p)}_{-}(k)(X(k+1) - \mathbf{E}\left\{X(k+1)\right\}) \\ &\quad + L^{(p)}_{-}(k)(X(k-1) - \mathbf{E}\left\{X(k-1)\right\}),\end{aligned} \tag{133}$$

with $b^{(p)^-}$ defined in (125) and the matrices (119) having been used. Also, let us rewrite eq. (132), with the plus sign, by commutating the left hand side using (19), and by performing the change of index $h' = p - h$

$$\begin{aligned}\mathbf{K}^T_{n^{h'},n^{p-h'}}\overline{x_{[h']}(k)\otimes x_{[p-h']}(k+2)} &= G^{(p-h',h')}_{+}(k)\left(\mathcal{X}(k) - \mathbf{E}\left\{\mathcal{X}(k)\right\}\right) + d^{+}_{p-h',h'}(k). \\ &\quad + H^{(p-h',h')}_{+}(k)\left(X(k+1) - \mathbf{E}\left\{X(k+1)\right\}\right) \\ &\quad + L^{(p-h',h')}_{+}(k)\left(X(k-1) - \mathbf{E}\left\{X(k-1)\right\}\right).\end{aligned} \tag{134}$$

Multiplying eq. (134) by ${\mathbf{K}^T_{n^{h'},n^{p-h'}}}^{-1}$ (which, by (20) is equal to $\mathbf{K}^T_{n^{p-h'},n^{h'}}$), and assembling for $h' = 1, ..., p-1$ results in

$$\begin{aligned} x^{\{p\}}(k+1) - \mathbf{E}\left\{x^{\{p\}}(k+1)\right\} &= G_+^{(p)}(k)(\mathcal{X}(k) - \mathbf{E}\left\{\mathcal{X}(k)\right\}) + {b^{(p)}}^+(k+1), \\ &+ H_+^{(p)}(k)(X(k+1) - \mathbf{E}\left\{X(k+1)\right\}), \\ &+ L_+^{(p)}(k)(X(k-1) - \mathbf{E}\left\{X(k-1)\right\}), \end{aligned} \tag{135}$$

${b^{(p)}}^+$ being defined in (125) and the matrices (118) having been substituted. Finally, representations (122), (123) readily follows from assembling (133), (135) for $p = 1, ..., \mu$, and by using matrices (117).

Now, define

$$B_-(k) = \mathbf{E}\{b^-(k), {b^-}^T(k)\}, \tag{136}$$

that is the covariance of the process b^- defined by eqs. (124), (125), and (127). By the backward equation (122), $B_-(k)$ is nonsingular for any k. Otherwise we could find a vector, say $\alpha \neq 0$, such that $\alpha^T b^-(k) = 0$, and this would imply, by (122) that a linear combination (with at least one non-zero coefficient) of the vectors $\widetilde{\mathcal{X}}(k)$, $\widetilde{\mathcal{X}}(k+1)$, $\widetilde{X}(k+1)$, and $\widetilde{X}(k-1)$ is zero. But this contradicts the nonsingularity assumption. Thus it is

$$B_-(k) > 0. \qquad \forall k. \tag{137}$$

Similarly, the nonsingularity assumption implies

$$\mathbf{E}\{d(k)d^T(k)\} = D(k) > 0, \qquad \forall k, \tag{138}$$

otherwise by eq. (121), we would find a vanishing linear combination of $\widetilde{\mathcal{X}}(k)$, $\widetilde{X}(k+1)$, and $\widetilde{X}(k-1)$ with at least one non-zero coefficient.

By definition of d (given in Theorem 3.2), by the equations (127)–(124) defining $b^\pm$, and by (79), (80), we readily get, for any couple of integers k, l, the following relations

$$\mathbf{E}\{d(k)\} = \mathbf{E}\{b^+(k)\} = \mathbf{E}\{b^-(k)\} = 0, \tag{139}$$

$$\mathbf{E}\left\{b^\pm(k){b^\pm}^T(l)\right\} = 0, \quad |k-l| > 1, \tag{140}$$

Furthermore, an application of (80) for $h = 0$ (recall that $d_h = d^\pm_{0,h}$) yields

$$\mathbf{E}\left\{d(k){b^\pm}^T(l)\right\} = 0, \quad |k \pm 1 - l| > 1. \tag{141}$$

Let us consider the orthogonality property (84). On account of (127), (126), it can be written in terms of the sequences $b^\pm$ as follows

$$\mathbf{E}\left\{b^-(k)\phi(x(i))\right\} = 0, \quad \forall i \neq k+1, \tag{142}$$

$$\mathbf{E}\left\{b^+(k)\phi(x(i))\right\} = 0, \quad \forall i \neq k-1, \tag{143}$$

where ϕ is any suitably dimensioned vector Borel function on $\mathbb{R}^n$. By the definition of X and $\mathcal{X}$, given in (45), it is

$$\mathbf{E}\left\{d(k)X^T(i)\right\} = 0, \quad \forall i \neq k, \tag{144}$$

$$\mathbf{E}\left\{d(k)\mathcal{X}^T(i)\right\} = 0, \quad \forall i \neq k-1, k+1, \tag{145}$$

which is straightforward by (84), whereas equations (142), (143) are ensued by

$$\mathbf{E}\left\{b^-(k)X^T(i)\right\} = 0, \quad \forall i \neq k+1, \tag{146}$$
$$\mathbf{E}\left\{b^+(k)X^T(i)\right\} = 0, \quad \forall i \neq k-1, \tag{147}$$
$$\mathbf{E}\left\{b^-(k)\mathcal{X}^T(i)\right\} = 0, \quad \forall i \neq k, k+2, \tag{148}$$
$$\mathbf{E}\left\{b^+(k)\mathcal{X}^T(i)\right\} = 0, \quad \forall i \neq k, k-2. \tag{149}$$

By using eq. (121) and (144), (145) one has

$$\begin{aligned}\mathbf{E}\{d(k)\widetilde{X}^T(k)\} \quad &= \mathbf{E}\left\{b^-(k)\Big(d(k) + F(k)\widetilde{\mathcal{X}}(k) + \Phi(k)\widetilde{X}(k+1) + \Psi(k)\widetilde{X}(k-1),\Big)^T\right\}\\ &= \mathbf{E}\{d(k)d^T(k)\} = D(k).\end{aligned} \tag{150}$$

By using the backward equation (122)

$$\begin{aligned}\mathbf{E}\{b^-(k)\widetilde{\mathcal{X}}^T(k+2)\} \quad &= \mathbf{E}\Big\{b^-(k)\Big(b^-(k+2) + G_-(k+3)\widetilde{\mathcal{X}}(k+3)\\ &\qquad + H_-(k+2)\widetilde{X}(k+3) + L_-(k+2)\widetilde{X}(k+1)\Big)^T\Big\}\\ &= \mathbf{E}\left\{b^-(k)\widetilde{X}^T(k+1)\right\}L_-^T(k+2),\end{aligned} \tag{151}$$

$$\begin{aligned}\mathbf{E}\{d(k)\widetilde{\mathcal{X}}^T(k+1)\} \quad &= \mathbf{E}\Big\{d(k)\Big(b^-(k+1) + G_-(k+2)\widetilde{\mathcal{X}}(k+2)\\ &\qquad + H_-(k+1)\widetilde{X}(k+2) + L_-(k+1)\widetilde{X}(k)\Big)^T\Big\}\\ &= \mathbf{E}\{d(k)b^{-T}(k+1)\} + \mathbf{E}\{d(k)\widetilde{X}^T(k)\}L_-^T(k+1),\end{aligned} \tag{152}$$

where orthogonality relations has been used, namely (148) and (146) in (151), and (144), (145) in (152). Using representation (121), and properties (148) and (146) again, gives

$$\begin{aligned}\mathbf{E}\{b^-(k)\widetilde{X}^T(k+1)\} \quad &= \mathbf{E}\Big\{b^-(k)\Big(d(k+1) + F(k+1)\widetilde{\mathcal{X}}(k+1)\\ &\qquad + \Phi(k+1)\widetilde{X}(k+2) + \Psi(k+1)\widetilde{X}(k),\Big)^T\Big\}\\ &= \mathbf{E}\{b^-(k)d^T(k+1)\}.\end{aligned} \tag{153}$$

Similarly, by the forward equation (123) we readily prove

$$\mathbf{E}\{b^+(k)\widetilde{\mathcal{X}}^T(k-2)\} \quad = \mathbf{E}\{b^+(k)\widetilde{X}^T(k-1)\}H_+^T(k-2), \tag{154}$$
$$\mathbf{E}\{d(k)\widetilde{\mathcal{X}}^T(k-1)\} \quad = \mathbf{E}\{d(k)b^{+T}(k-1)\} + \mathbf{E}\{d(k)\widetilde{X}^T(k)\}H_+^T(k-1), \tag{155}$$

where orthogonality relations have been used, in specific: (149) and (147), in (154), and (144) and (145) in (155). On the other hand it can be used, instead of the backward one, the forward equation to eliminate $\widetilde{\mathcal{X}}(k{+}1)$ in (152), and the backward equation (instead of the forward one) to eliminate $\widetilde{\mathcal{X}}(k{-}1)$ in (155); with the help of the usual orthogonality relations the result is

$$\mathbf{E}\{d(k)\widetilde{\mathcal{X}}^T(k{+}1)\} \quad = \mathbf{E}\{d(k)b^{+^T}(k{+}1)\} + \mathbf{E}\{d(k)\widetilde{X}^T(k)\}L_+^T(k{+}1), \tag{156}$$

$$\mathbf{E}\{d(k)\widetilde{\mathcal{X}}^T(k{-}1)\} \quad = \mathbf{E}\{d(k)b^{-^T}(k{-}1)\} + \mathbf{E}\{d(k)\widetilde{X}^T(k)\}H_-^T(k{-}1). \tag{157}$$

Moreover, using representation (121) yields

$$\mathbf{E}\{b^+(k)\widetilde{X}^T(k{-}1)\} = \mathbf{E}\{b^+(k)d^T(k{-}1)\}. \tag{158}$$

Now, $\mathbf{E}\{d(k)b^{-^T}(k{+}1)\} = \mathbf{E}\{d(k)b^{+^T}(k{-}1)\} = 0$, by (141), so we can rewrite eqs. (152), (155) as

$$\mathbf{E}\{d(k)\widetilde{\mathcal{X}}^T(k{+}1)\} = D(k)L_-^T(k{+}1) \tag{159}$$

$$\mathbf{E}\{d(k)\widetilde{\mathcal{X}}^T(k{-}1)\} = D(k)H_+^T(k{-}1), \tag{160}$$

where (150) has been used. On the other hand eqs. (156), (157), by (150), can be rewritten

$$\mathbf{E}\{d(k)\widetilde{\mathcal{X}}^T(k{+}1)\} = \mathbf{E}\left\{d(k)b^{+^T}(k{+}1)\right\} + D(k)L_+^T(k{+}1), \tag{161}$$

$$\mathbf{E}\{d(k)\widetilde{\mathcal{X}}^T(k{-}1)\} = \mathbf{E}\left\{d(k)b^{-^T}(k{-}1)\right\} + D(k)H_-^T(k{-}1). \tag{162}$$

Consider eq. (159) and use the forward equation to eliminate $\widetilde{\mathcal{X}}(k{+}1)$. Then some terms vanish by (144), (145), so the result is

$$D(k)L_-^T(k{+}1) = \mathbf{E}\{d(k)\widetilde{\mathcal{X}}^T(k{+}1)\} = \mathbf{E}\{d(k)b^{+^T}(k{+}1)\} + \mathbf{E}\{d(k)\widetilde{\mathcal{X}}^T(k)\}L_-^T(k{+}1), \tag{163}$$

from which, $D(k)$ having been replaced by r.h.s. of (150), it results that

$$\mathbf{E}\{d(k)b^{+^T}(k{+}1)\} = 0. \tag{164}$$

Similarly as in (163), but now using the backward equation in (160) to eliminate $\widetilde{\mathcal{X}}(k{-}1)$ we readily prove

$$\mathbf{E}\{d(k)b^{-^T}(k{-}1)\} = 0. \tag{165}$$

Thus, by (164), eqs. (161) and (159) imply

$$D(k)\left(L_-^T(k{+}1) - L_+^T(k{+}1)\right) = 0, \tag{166}$$

whereas by using (165) in (162) a resembling with (160) gives

$$D(k)\left(H_-^T(k{-}1) - H_+^T(k{-}1)\right) = 0, \tag{167}$$

and since $D(k) > 0$ for any k, it follows:

$$L_-^T(k) = L_+^T(k), \qquad \forall k, \tag{168}$$

$$H_-^T(k) = H_+^T(k), \qquad \forall k. \tag{169}$$

Moreover eqs. (164), (165) merged into eq. (141) result in

$$\mathbf{E}\left\{d(k)b^{\pm^T}(l)\right\} = 0, \quad l \neq k, k \pm 2. \tag{170}$$

On the other hand, since k is arbitrary, eqs. (164), (165) imply that right hand sides of (153), and (158) vanish:

$$\mathbf{E}\{b^+(k)\widetilde{X}^T(k-1)\} = 0, \tag{171}$$
$$\mathbf{E}\{b^-(k)\widetilde{X}^T(k+1)\} = 0. \tag{172}$$

Now, by eliminating $\widetilde{\mathcal{X}}(k)$ via forward equation, and exploiting properties (147)–(149) the following expressions are derived:

$$\begin{aligned}\mathbf{E}\{b^+(k)\widetilde{\mathcal{X}}^T(k)\} &= \mathbf{E}\{b^+(k){b^+}^T(k)\} + \mathbf{E}\{b^+(k)\widetilde{X}^T(k-1)\}L_+^T(k)\\ &= \mathbf{E}\{b^+(k){b^+}^T(k)\} + \mathbf{E}\{b^+(k)d^T(k-1)\}L_+^T(k)\end{aligned} \tag{173}$$

$$\begin{aligned}\mathbf{E}\{b^-(k)\widetilde{\mathcal{X}}^T(k)\} &= \mathbf{E}\{b^-(k){b^+}^T(k)\} + \mathbf{E}\{b^-(k)\widetilde{X}^T(k+1)\}H_+^T(k)\\ &= \mathbf{E}\{b^-(k){b^+}^T(k)\} + \mathbf{E}\{b^-(k)d^T(k+1)\}H_+^T(k)\end{aligned} \tag{174}$$

where in the second line of (173) eq. (158) has been used, as well as eq. (153) in (174). Similarly, by exploiting this once the backward equation to eliminate $\widetilde{\mathcal{X}}(k)$

$$\begin{aligned}\mathbf{E}\{b^+(k)\widetilde{\mathcal{X}}^T(k)\} &= \mathbf{E}\{b^+(k){b^-}^T(k)\} + \mathbf{E}\{b^+(k)\widetilde{X}^T(k-1)\}L_-^T(k)\\ &= \mathbf{E}\{b^+(k){b^-}^T(k)\} + \mathbf{E}\{b^+(k)d^T(k-1)\}L_-^T(k)\end{aligned} \tag{175}$$

$$\begin{aligned}\mathbf{E}\{b^-(k)\widetilde{\mathcal{X}}^T(k)\} &= \mathbf{E}\{b^-(k){b^-}^T(k)\} + \mathbf{E}\{b^-(k)\widetilde{X}^T(k+1)\}H_-^T(k)\\ &= \mathbf{E}\{b^-(k){b^-}^T(k)\} + \mathbf{E}\{b^-(k)d^T(k+1)\}H_-^T(k)\end{aligned} \tag{176}$$

By (168), (169) (or simply by (164), (165), on account of the arbitrarily of k) eqs. (173), (175) imply:

$$\mathbf{E}\{b^+(k){b^+}^T(k)\} = \mathbf{E}\{b^+(k){b^-}^T(k)\}, \tag{177}$$

and eqs. (174), (176) imply:

$$\mathbf{E}\{b^-(k){b^+}^T(k)\} = \mathbf{E}\{b^-(k){b^-}^T(k)\}, \tag{178}$$

thus $\mathbf{E}\{b^-(k){b^+}^T(k)\} = \mathbf{E}\{b^+(k){b^-}^T(k)\}$, and by recalling (136) we get

$$B_-(k) = \mathbf{E}\{b^-(k)b^-(k)^T\} = \mathbf{E}\{b^+(k)b^-(k)^T\} = \mathbf{E}\{b^-(k)b^+(k)^T\} = \mathbf{E}\{b^+(k)b^+(k)^T\}. \tag{179}$$

By (164), (165) and (168), (169), by defining

$$L(k) = L_\pm(k), \qquad H(k) = H_\pm(k), \tag{180}$$

eqs. (159)–(160) can be rewritten

$$\mathbf{E}\{d(k)\widetilde{\mathcal{X}}^T(k+1)\} = D(k)L^T(k+1), \tag{181}$$
$$\mathbf{E}\{d(k)\widetilde{\mathcal{X}}^T(k-1)\} = D(k)H^T(k-1), \tag{182}$$

and (151), (154), on account of (171), (172) can be rewritten

$$\mathbf{E}\{b^-(k)\widetilde{\mathcal{X}}^T(k+2)\} = 0, \tag{183}$$
$$\mathbf{E}\{b^+(k)\widetilde{\mathcal{X}}^T(k-2)\} = 0. \tag{184}$$

By (179), by substituting $B_-(k)$ and while taking into account of (164), (165), eqs. (173) and (174) reduce to

$$\mathbf{E}\{b^+(k)\widetilde{\mathcal{X}}^T(k)\} = B_-(k) \quad (185)$$
$$\mathbf{E}\{b^-(k)\widetilde{\mathcal{X}}^T(k)\} = B_-(k). \quad (186)$$

Now, by the forward equation (123), and property (147)

$$\begin{aligned} 0 = \mathbf{E}\{b^+(k)\widetilde{\mathcal{X}}^T(k{+}1)\} \quad &= \mathbf{E}\Big\{b^+(k)\Big(b^+(k{+}1) + G_+(k)\widetilde{\mathcal{X}}(k) \\ &\qquad +H_+(k{+}1)\widetilde{X}(k{+}2) + L_+(k{+}1)\widetilde{X}(k)\Big)^T\Big\} \\ &= \mathbf{E}\{b^+(k){b^+}^T(k{+}1)\} + \mathbf{E}\{b^+(k)\widetilde{\mathcal{X}}^T(k)\}G_+^T(k), \end{aligned} \quad (187)$$

and similarly

$$0 = \mathbf{E}\{b^+(k)\widetilde{\mathcal{X}}^T(k{-}1)\} = \mathbf{E}\{b^+(k){b^+}^T(k{-}1)\} + \mathbf{E}\{b^+(k)\widetilde{\mathcal{X}}^T(k{-}2)\}G_+^T(k{-}2). \quad (188)$$

Similarly, by the backward equation (122), and on account of (146)

$$0 = \mathbf{E}\{b^-(k)\widetilde{\mathcal{X}}^T(k{+}1)\} = \mathbf{E}\{b^-(k){b^-}^T(k{+}1)\} + \mathbf{E}\{b^-(k)\widetilde{\mathcal{X}}^T(k{+}2)\}G_-^T(k{+}2), \quad (189)$$

$$0 = \mathbf{E}\{b^-(k)\widetilde{\mathcal{X}}^T(k{-}1)\} = \mathbf{E}\{b^-(k){b^-}^T(k{-}1)\} + \mathbf{E}\{b^-(k)\widetilde{\mathcal{X}}^T(k)\}G_-^T(k) \quad (190)$$

By using relations (183)–(186) in eqs. (187)–(190), the following identities are derived:

$$\mathbf{E}\{b^+(k){b^+}^T(k{-}1)\} = 0, \quad (191)$$
$$\mathbf{E}\{b^+(k){b^+}^T(k{+}1)\} = -B_-(k)G_+^T(k), \quad (192)$$
$$\mathbf{E}\{b^-(k){b^-}^T(k{+}1)\} = 0, \quad (193)$$
$$\mathbf{E}\{b^-(k){b^-}^T(k{-}1)\} = -B_-(k)G_-^T(k). \quad (194)$$

Now, eqs. (194) and (192) vanish, i.e.: $B_-(k)G_+^T(k) = B_-(k)G_-^T(k) = 0$, as eqs. (191) and (193) holds for any k, and by (137) it is

$$G_+(k) = G_-(k) = 0, \qquad \forall k. \quad (195)$$

By (195) and (168), (169), the backward and forward equations imply: $b^+(k) = b^-(k)$, $\forall k$, thus, defining

$$b(k) = b^+(k) = b^-(k), \quad (196)$$

and recalling (180) imply that the backward and forward equations result actually in a single equation:

$$\widetilde{\mathcal{X}}(k) = b(k) + H(k)\widetilde{X}(k{+}1) + L(k)\widetilde{X}(k{-}1). \quad (197)$$

By (195), (196) eqs. (191)–(194) reduce to $E\{b(k)b(k{\pm}1)\} = 0$, thus by recalling (140) one has

$$\mathbf{E}\left\{b(k)b^T(l)\right\} = 0. \qquad l \neq k, \quad (198)$$

Now, by using equation (197) and (144), (145)

$$
\begin{aligned}
0 \quad &= \mathbf{E}\{d(k)\widetilde{\mathcal{X}}^T(k+2)\} \\
&= \mathbf{E}\left\{d(k)\Big(b(k+2) + H(k+2)\widetilde{X}(k+3) + L(k+2)\widetilde{X}(k+1)\Big)^T\right\} \\
&= \mathbf{E}\{d(k)b^T(k+2)\},
\end{aligned}
\tag{199}
$$

and similarly, using equation (99) to eliminate $\widetilde{\mathcal{X}}(k)$ and $\widetilde{\mathcal{X}}(k-2)$ gives

$$0 = \mathbf{E}\{d(k)\widetilde{\mathcal{X}}^T(k)\} = \mathbf{E}\{d(k)b^T(k)\}, \tag{200}$$

$$0 = \mathbf{E}\{d(k)\widetilde{\mathcal{X}}^T(k-2)\} = \mathbf{E}\{d(k)b^T(k-2)\}. \tag{201}$$

Thus, (170), (196), and (199)–(201) results in

$$\mathbf{E}\left\{d(k)b^T(l)\right\} = 0. \quad \forall k, l. \tag{202}$$

In conclusion, equation (99) is the afore proven eq. (197). Equations (101), and (102) are just eqs. (198), and (202) respectively. Property (105) has been proven (eq. (138)), and (106) is proven as well, as $B(k) = B_-(k)$, and property (137) has been shown. Using the definition of b^- (given by eqs. (124), (125) and (127)) and eq. (78) lead to the expression of $B(k)$ given in (103), provided of course taking into account of identity (196).

As to the orthogonality properties, eqs. (107), (108) are just eqs. (144), (145). Since $b = b^+ = b-$, eqs. (146) and (147) merge into eq. (109), as well as eqs. (148), (149) imply (110). Moreover, eq. (111) is just eq. (150), whereas eqs. (112), (113), follow from eqs. (159), (160) resembling eq. (180). By using (196) in (185) (or in (186)) we get eq. (114), as $B(k) = B_-(k)$.

Finally, it remains only to prove eq. (98). For, let us post-multiply eq. (121) by $b^T(k)$ and take expectations. Then, using (202), (109), and (114), gives $F(k)B(k) = 0$, $\forall k$ and by the nonsingularity of $B(k)$ it follows that

$$F(k) = 0, \qquad \forall k. \tag{203}$$

Thus, after substituting (203) in (121), eq. (98) immediately ensues. ∎

The next result is a corollary of Theorem 4.2 where it is shown how all the matrix coefficients in eqs. (98), (99) can be calculated by a subset of known statistics of the process x. Notice that formulas have just been given as to the computation of some coefficients. In fact, the matrix coefficients of eq. (98) are an aggregation of a subset of $\cup_{h=1}^{\mu}\mathcal{F}_h(k)$, where $\mathcal{F}_h(k)$ is the set of projection matrices defined in (48), therefore such coefficients result by eq. (54). Also, the noises covariances $B(k)$ and $D(k)$ can be calculated by the formulas given in Theorems 4.1 and 3.2. On the other hand, by the representation theorem just proven, we know in advance of a number of matrix-coefficients in the sets $\mathcal{F}_h(k)$ being zero, i.e. $F_{i,j}^{(h)}(k) = 0$ for any $(i, j) \neq (0, 0)$. So, we now turn to consider again the problem of coefficients computation, as Theorem 4.2 allows us to derive in a very simple way a formula giving all the matrix coefficients of both eqs. (98), (99), as well as the covariances $B(k)$ and $D(k)$. We show this in the following corollary.

Corollary 4.3. *Let us denote*

$$R_{1,1}(r, s) = \mathbf{E}\left\{\widetilde{X}(r)\widetilde{X}^T(s)\right\}, \tag{204}$$

$$R_{1,2}(r,s) = \mathbf{E}\left\{\widetilde{X}(r)\widetilde{\mathcal{X}}^T(s)\right\}, \tag{205}$$

$$R_{2,1}(r,s) = \mathbf{E}\left\{\widetilde{\mathcal{X}}(r)\widetilde{X}^T(s)\right\}, \tag{206}$$

$$R_{2,2}(r,s) = \mathbf{E}\left\{\widetilde{\mathcal{X}}(r)\widetilde{\mathcal{X}}^T(s)\right\}. \tag{207}$$

Then the matrix coefficients of representation (98), (99) can be calculated as the unique solution of the following system of matrix equations

$$\begin{bmatrix} \Phi(k) & \Psi(k) \end{bmatrix} P(k) = \begin{bmatrix} R_{1,1}(k,k-1) & R_{1,1}(k,k+1) \end{bmatrix}; \tag{208}$$

$$\begin{bmatrix} H(k) & L(k) \end{bmatrix} P(k) = \begin{bmatrix} R_{2,1}(k,k-1) & R_{2,1}(k,k+1) \end{bmatrix}; \tag{209}$$

where $P(k)$ is the (invertible, by nonsingularity assumption) matrix:

$$P(k) = \begin{bmatrix} R_{1,1}(k+1,k-1) & R_{1,1}(k-1,k-1) \\ R_{1,1}(k+1,k+1) & R_{1,1}(k-1,k+1) \end{bmatrix}. \tag{210}$$

The noise covariances $D(k)$ and $B(k)$ are given by

$$D(k) = R_{1,1}(k,k) - \Phi(k)R_{1,1}(k+1,k) - \Psi(k)R_{1,1}(k-1,k), \tag{211}$$

$$B(k) = R_{2,2}(k,k) - H(k)R_{1,2}(k+1,k) - L(k)R_{1,2}(k-1,k). \tag{212}$$

Moreover, the non-zero terms of the cross correlation $D(\cdot,\cdot)$, defined in (65) satisfy

$$\begin{aligned} D(k,k+1) &= -\Phi(k)D(k+1) \\ &= -D(k)\Psi^T(k+1). \end{aligned} \tag{213}$$

Proof. Post-multiply eqs. (98), (99) first by $\widetilde{X}^T(k+1)$, take expectations, and exploit (107), (108). Then, while multiplying now by $\widetilde{X}^T(k-1)$, repeat such manipulation, and after grouping together the resulting expressions, eqs. (208), (209) ensue. Equations (211) and (212) follow from post-multiplying eqs. (98) and (99) respectively, by $\widetilde{X}^T(k)$ and taking expectations. Finally, multiplying (98) by $d^T(k+1)$ and exploiting (107), (111) is ensued by the first one of (213), whereas the second one is derived in a similar way, by multiplying by $d^T(k)$ while considering (98) at $k+1$. ∎

We can rewrite the representation (98), (99) in a normalized version, where the forcing noises have an unitary covariance. We recall the notion of *conjugate process* [23], [26], which will be used in the statement of the following corollary. The conjugate process of a discrete-index nonsingular process $z = \{z(k), k \in \mathbf{T}\} \subset \mathbb{R}^p$ is the *unique* process e such that:

$$\mathbf{E}\{z(k)e^T(l)\} = I_p\delta(k-l), \qquad \forall k,l \in \mathbf{T}. \tag{214}$$

In our case, by the nonsingularity assumption the processes X and $\mathcal{X}$ involved in Theorem 4.2 are nonsingular, so one may guess that the conjugate processes of X and $\mathcal{X}$ can be obtained by normalizing the processes d and b of eqs. (98), and (98) respectively.

Corollary 4.4. *For any $\mu \geq \nu$, and for any $k \in \mathbf{T}$, there exist matrices $M_0(k)$, $M_+(k)$, $M_-(k)$, where*

$$M_0(k) = M_0^T(k) > 0, \tag{215}$$

$$M_+(k) = M_-^T(k+1), \tag{216}$$

and $S_0(k), S_+(k), S_-(k)$ *with*

$$S_0(k) = S_0^T(k) > 0, \tag{217}$$

such that the processes X *and* $\mathcal{X}$ *satisfy the following equations*

$$M_0(k)\widetilde{X}(k) - M_+(k)\widetilde{X}(k+1) - M_-(k)\widetilde{X}(k-1) = \xi(k), \tag{218}$$

$$S_0(k)\widetilde{\mathcal{X}}(k) - S_+(k)\widetilde{X}(k+1) - S_-(k)\widetilde{X}(k-1) = \eta(k), \tag{219}$$

where $\xi \in \mathbb{R}^\alpha$ *is a zero-mean, one-step correlated noise*

$$\mathbf{E}\{\xi(k)\} = 0, \qquad \forall k, \tag{220}$$

$$\mathbf{E}\{\xi(k)\xi^T(l)\} = 0, \qquad \text{for} \quad |k-l| > 1, \tag{221}$$

$$\mathbf{E}\{\xi(k)\xi^T(k)\} = M_0(k), \tag{222}$$

$$\mathbf{E}\{\xi(k)\xi^T(k+1)\} = -M_+(k) = -M_-^T(k+1). \tag{223}$$

and $\eta \in \mathbb{R}^\beta$ *is a zero-mean white noise*

$$\mathbf{E}\{\eta(k)\} = 0, \qquad \forall k, \tag{224}$$

$$\mathbf{E}\{\eta(k)\eta^T(l)\} = S_0(k)\delta(k-l). \tag{225}$$

ξ *and* η *are the conjugate processes of* X, *and* $\mathcal{X}$ *respectively i.e.*

$$\mathbf{E}\{\xi(k)\widetilde{X}^T(l)\} = I_\alpha\delta(k-l), \tag{226}$$

$$\mathbf{E}\{\eta(k)\widetilde{\mathcal{X}}^T(l)\} = I_\beta\delta(k-l). \tag{227}$$

Moreover η *is uncorrelated to* X

$$\mathbf{E}\{\eta(k)\widetilde{X}^T(l)\} = 0, \quad \forall k, l \tag{228}$$

and ξ *is correlated to* $\mathcal{X}$ *by the following formula*

$$\mathbf{E}\{\xi(k)\widetilde{\mathcal{X}}^T(l)\} = H^T(l)\delta(k-1-l) + L^T(l)\delta(k+1-l). \tag{229}$$

Proof. Let us define

$$\xi(k) = D^{-1}(k)d(k); \qquad \eta(k) = B^{-1}(k)b(k). \tag{230}$$

Multiplying eqs. (98), (99) by $D^{-1}(k)$ gives eqs. (218), (219) with

$$M_0(k) = D^{-1}(k); \qquad S_0(k) = B^{-1}(k); \tag{231}$$

$$M_+(k) = D^{-1}(k)\Phi(k); \qquad M_-(k) = D^{-1}(k)\Psi(k); \tag{232}$$

$$S_+(k) = B^{-1}(k)H(k); \qquad S_-(k) = B^{-1}(k)L(k). \tag{233}$$

Therefore, properties (220)–(223) readily follow from the definition of ξ and d, whereas the definitions of η and b lead to (224)–(225). The last equality in (223) follows, by direct calculation, on using (232) and (213)

$$\begin{aligned}\mathbf{E}\{\xi(k)\xi^T(k+1)\} &= D^{-1}(k)D(k,k+1)D^{-1}(k+1)\\ &= -D^{-1}(k)\Phi(k) = -M_+(k)\\ &= -\Psi(k+1)D^{-1}(k+1) = -M_-^T(k+1).\end{aligned}$$

Since (107), (111) and the first one of (230) hold, identity (226) readily ensues, whereas (227) is straightforward by (109). Finally, eq. (227) readily follows from (110) and (114), and eq. (229) from (112), (113). ∎

To conclude the present section we briefly sum up the results obtained up to now and comment on them. Theorem 4.2 states: any reciprocal chain $x = \{x(k), k \in \mathbf{T}\} \subset \mathbb{R}^n$, satisfying all assumptions given in §2.5, is *a solution* of the nearest-neighbour, non-causal, stochastic model (98), (99). While saying 'it is a solution' we mean that x satisfies both eqs. (98), (99) almost surely. Equations (98), (99), or their normalized version (218), (219), are *polynomial* equations in the variable x, though linear in the augmented variables X and $\mathcal{X}$ (i.e. they are 'exactly linearized' equations). The polynomial degree of eqs. (98), (99), is the integer μ we have fixed so far, at the beginning of §4, to some value greater or equal to ν, the maximal degree, defined in §2.4, of the polynomial espressing the conditional expectation (40). Therefore, eqs. (98), (99), actually represent a family of models, one for each μ, all satisfied by the same reciprocal chain x.

On the other hand, eqs. (98), (99) being *given* – where b and d are any pair of white and one-step correlated processes respectively, satisfying (100)–(106) and (64), (65) – is not ensued in general by x being the *unique* reciprocal chain satisfying (98), (99) and the orthogonality properties (107)–(114).

As a matter of fact, Corollary 4.3 tells us that the matrix-coefficients of the model (98), (99), are determined by the covariances (204)–(207). Now, by the definition of vectors $X(k), \mathcal{X}(k)$, given in (73), (74), it is easily recognized that these covariances are the collection of the following moments,

$$\mathbf{E}\left\{x_{[i]}(k-1) \otimes x_{[j]}(k) \otimes x_{[l]}(k+1)\right\}, \quad i, j, l = 0, ..., 2\mu, \quad i+j+l \leq 2\mu. \tag{234}$$

all joined for $k \in \mathbf{T}$ (such that $k \pm 1 \in \mathbf{T}$). Therefore, all the nonsingular reciprocal chains x which have the same moments up to the order 2μ are solutions of eqs. (98), (99) as well. Let us denote by $\mathcal{R}(\mathbf{T}; \mu)$ the set of all nonsingular reciprocal chains x having the same moments (234) for any $k \in \mathbf{T}$ such that $k \pm 1 \in \mathbf{T}$.

Now, denote $\mathcal{R}(\mathbf{T})$ the reciprocal class of x. By the definition given so far, in §2.2, $\mathcal{R}(\mathbf{T})$ is the set of all reciprocal chains on $\mathbf{T}$ owning the same family of reciprocal probability transitions. One has $\mathcal{R}(\mathbf{T}) \neq \mathcal{R}(\mathbf{T}; \mu)$, as the moments (234) are not sufficient in general to determine all the reciprocal conditional probabilities of process x. Thus, eqs. (98), (99) do not determine in general a reciprocal class. Only in the *particular* case of a probability distribution uniquely determined by the first 2μ moments (i.e. a *Gaussian-like* distribution), we can say $\mathcal{R}(\mathbf{T}) = \mathcal{R}(\mathbf{T}; \mu)$, i.e. the set of solutions of eqs. (98), (99) constitutes a reciprocal class.

We point out the close similarity between eq. (218) and eq. (1), which was found first in [19] as the equation completely describing any Gaussian nonsingular reciprocal process. Besides the different nature of the involved processes – a reciprocal chain for (218), (219), whereas eq. (1) concerns a reciprocal process with a probability density in $\mathbb{R}^n$ – in a sense we can say that eqs. (218), (219) reduce to (1) when the process is 'Gaussian'. Let x be a reciprocal chain such that, similarly to Gaussian case, all conditional probability distributions are determined by the first 2 moments of it (i.e. suppose the moments in (234), with $\mu = 1$, determine the family of conditional distribution of x). Indeed, for such a finite-states process, since $\mu = 1$, we see that eq. (219) is void, and eq. (218) is satisfied by $X(k) = x(k)$. Moreover, by the discussion above we have $\mathcal{R}(\mathbf{T}) = \mathcal{R}(\mathbf{T}; 2)$. Therefore, such a 'Gaussian' reciprocal chain satisfies *the same* model (1) of a Gaussian reciprocal process owning a density in $\mathbb{R}^n$, and such a model represents the reciprocal class of x as well.

In conclusion: given any nonsingular reciprocal chain x, we have it satisfies model (218), (219). However, unlike either the Gaussian case, or the 'Gaussian-like' case above described, the set of all solutions of (218), (219) is in general different by the reciprocal class of x.

5. well-posedness (in a wide sense) of our representation, and an application to smoothing problems

In the present section the following results are given: first it will be shown that the model (98), (99) is *well posed in a wide sense* (the meaning of this will be made clear in a moment) under suitable *boundary conditions*. Then, an example of application of our model to smoothing problems is presented. From now onwards we shall consider the indeces set $\mathbf{T}$ to be either the *finite* discrete interval $[0, N]$, or the *unit circle*. Also, we focus on the (equivalent) normalized version of model (98), (99), that is eqs. (218), (219).

We have seen, in the comment at the end of the previous section, that the nonsingular reciprocal chain x satisfies model (218), (219), however it is not the unique solution of such set of equations. Moreover, the set of solutions of eqs. (218), (219), is not in general the reciprocal class of x. Now, we know by [11] (see also §2.2) that a reciprocal process on $\mathbf{T} = [0, N]$, is determined by: 1) the family of all reciprocal probability transitions 2) the joint probability distribution of $x(0)$ and $x(N)$. In other words, it is specified by its reciprocal class *and* by the knowledge of the boundary random variables $x(0)$ and $x(N)$. Thus, we could guess that, if x is 'Gaussian-like' – in the sense depicted earlier at the end of §4 – then a 'Dirichelet' boundary condition, specifying the distribution of the process at the end points of $\mathbf{T}$, is sufficient to specify the whole process on $\mathbf{T}$, in other words to determine x as the *unique* solution of eqs. (218), (219). This is indeed what happens for the Gaussian case described in [19], where it has been shown that eq. (1) has a unique solution, provided a Dirichelet boundary condition is assigned at both ends of the definition interval. However, if x is any nonsingular reciprocal chain (not a 'Gaussian-like' one as above) since eqs. (218), (219) do not specify its reciprocal class, we cannot expect these equations to have a unique solution, even under Dirichelet boundary conditions.

We will show in the present section that, under suitable boundary conditions, an *explicit representation* of the reciprocal chain x there exists. For the case of Dirichelet boundary conditions, this means that we can find linear maps $\Psi_d(k; \cdot, \cdot, \cdot, \cdot) : \mathbb{R}^\alpha \times \mathbb{R}^\alpha \times \mathbb{R}^{\alpha\mathbf{T}} \times \mathbb{R}^{\beta\mathbf{T}} \to \mathbb{R}^{\alpha+\beta}$ such that, with probability one:

$$\mathcal{Z}(k) = \Psi_d(k; X(0), X(N), \xi, \eta), \qquad \forall k \in \mathbf{T}, \tag{235}$$

where, $\mathcal{Z}^T(k) = [X^T(k), \mathcal{X}^T(k)]^T$, and ξ, η are the 'forcing' processes of eqs. (218), (219). Notice that the existence of map (235) is not in general equivalent to well-posedness, because the processes ξ, η depend on x. Indeed the processes ξ, η are normalized versions of the processes d, b in eqs. (98), (99), and looking at the proofs of Theorems 4.1, and 4.2, one can check that b, by identities (124)–(127), depends on processes $d^\pm$ which, by (85) include process d, and are defined by (86). Thus, we cannot say *in general* that eq. (235) is an explicit expression of the stochastic process x. Only in the case of a 'Gaussian-like' reciprocal chain x, i.e. when all the reciprocal probability transitions of x are completely determined by a finite number of moments (see discussion at the end of previous section) one has that processes ξ, η *are completely determined by their* 2*nd-order statistics* given by (220)–(225). Therefore the distribution of x is univocally determined, in this case, by eq. (235), and we can say that eqs. (218), (219) are well-posed. We point out that the latter situation indeed occurs for a reciprocal process owning a Gaussian density, i.e. the case considered in [19] where, in order to prove well-posedness, it was sufficient to prove the existence of an explicit formula, actually a particular case of (235).

Even though the existence of map (235) does not imply well-posedness, (i.e. does not imply the unicity of the solution of eqs. (218), (219)), nevertheless being it linear implies that *any* pair of processes ξ, η having the statistics (220)–(225), by reason of (235) yield a process x belonging to $R(\mathbf{T}, \mu)$. In other words, if we use any process ξ' and η', provided they have the same moments of ξ and η up to the 2th-order, then the process x' given by (235) is a solution of (218), (219), and x' has the same moments as the 'true' reciprocal chain x up to the 2μ-order. We refer to this property by saying that eqs. (218), (219) are *well-posed in a wide sense.*

As we will see later, the wide sense well-posedness just defined, allows us to make representation (218), (219) useful from an application point of view, for instance in the smoothing problem we will present at the end of this section.

In the above discussion we considered 'Dirichelet boundary conditions', however the same thoughts are true even for cyclic boundary conditions. We state now formally these two kinds of boundary condition as follows.

Dirichelet boundary conditions (DBC): there exist two random variables $x_i, x_f \in \mathbb{R}^n$, having a joint probability distribution concentrated on $\mathcal{S} \times \mathcal{S}$ (recall $\mathcal{S}$ is the finite set of states defined in §2.2) such that:

$$x(0) = x_i; \qquad x(N) = x_f. \tag{236}$$

Cyclic boundary conditions (CBC): x is defined on the discrete unit circle $\mathbf{T} = [0, N]$, characterized by $k \pm i = |k \pm i|_{N+1}$, $\forall k \in [0, N]$, $\forall i$ integer, where $|n_1|_{n_2}$ denotes, for any n_1, n_2 integers, the remainder of n_1/n_2 . Therefore, x satisfies

$$x(N+1) = x(0); \qquad x(-1) = x(N). \tag{237}$$

The wide-sense well-posedness we above defined in the case of DBC, is expressed for CBC as follows: there exists a linear map $\Psi_c(k; \cdot, \cdot) : \mathbb{R}^{\alpha\mathbf{T}} \times \mathbb{R}^{\beta^{\mathbf{T}}} \to \mathbb{R}^{\alpha+\beta}$ such that, with probability one:

$$Z(k) = \Psi_c(k; \xi, \eta), \qquad \forall k \in \mathbf{T}, \tag{238}$$

where ξ, η are the 'forcing' processes of eqs. (218), (219), and $\mathbf{T}$ is understood to be the discrete unit circle.

5.1. Characterization of the solution's space for a general linear non-causal model.

Let us consider over the indeces set $\mathbf{T}$, two sequences of $\mathbb{R}^\alpha$-valued vectors, say u, v, and let $\mathcal{M}_0, \mathcal{M}_\pm$ be sequences of $\alpha \times \alpha$ matrices. Let us consider the following equation

$$\mathcal{M}_0(k)u(k) - \mathcal{M}_+(k)u(k+1) - \mathcal{M}_-(k)u(k+1) = v(k). \tag{239}$$

Let be given on $\mathbf{T}$, an $\alpha \times \alpha$-dimensioned sequences of matrices, $\mathcal{M}_0, \mathcal{M}_\pm$, such that

$$\mathcal{M}_0(k) > 0, \tag{240}$$
$$\mathcal{M}_+(k) = \mathcal{M}_-^T(k+1), \tag{241}$$

Let us denote by $\mathcal{N}(\mathcal{M}_0, \mathcal{M}_-, \mathcal{M}_+)$ the set of stochastic sequences ϕ, such that:

$$\begin{aligned}
\mathbf{E}\{\phi(k)\phi^T(l)\} &= 0, \qquad \text{for} \quad |k-l|>1, && (242)\\
\mathbf{E}\{\phi(k)\phi^T(k)\} &= \mathcal{M}_0(k), && (243)\\
\mathbf{E}\{\phi(k)\phi^T(k{+}1)\} &= -\mathcal{M}_+(k) \\
&= -\mathcal{M}_-^T(k{+}1). && (244)
\end{aligned}$$

Let us introduce the following matrix $\mathcal{F}_d$:

$$\mathcal{F}_d(\mathcal{M}_0, \mathcal{M}_-, \mathcal{M}_+) = \begin{bmatrix} I & 0 & & \cdots & & 0 \\ -\mathcal{M}_-(1) & \mathcal{M}_0(1) & -\mathcal{M}_+(1) & 0 & \cdots & 0 \\ 0 & -\mathcal{M}_-(2) & \mathcal{M}_0(2) & -\mathcal{M}_+(2) & & 0 \\ \vdots & & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -\mathcal{M}_-(\bar{N}) & \mathcal{M}_0(\bar{N}) & -\mathcal{M}_+(\bar{N}) \\ 0 & & \cdots & & 0 & I \end{bmatrix} \tag{245}$$

where $\bar{N} = N-1$, and the *cyrculant* matrix $\mathcal{F}_c$:

$$\mathcal{F}_c(\mathcal{M}_0, \mathcal{M}_-, \mathcal{M}_+) = \begin{bmatrix} \mathcal{M}_0(0) & -\mathcal{M}_+(0) & 0 & \cdots & -\mathcal{M}_-(0) \\ -\mathcal{M}_-(1) & \mathcal{M}_0(1) & -\mathcal{M}_+(1) & 0 & \cdots \\ 0 & -\mathcal{M}_-(2) & \mathcal{M}_0(2) & -\mathcal{M}_+(2) & \\ \vdots & & & \ddots & 0 \\ -\mathcal{M}_+(N) & \cdots & 0 & -\mathcal{M}_-(N) & \mathcal{M}_0(N) \end{bmatrix} \tag{246}$$

The matrices $\mathcal{F}_d$, and $\mathcal{F}_c$, are square block-matrices constituted by $(N{+}1)\times(N{+}1)$ blocks in $\mathbb{R}^{\alpha\times\alpha}$. Moreover, define $\mathcal{F}$ as the matrix obtained by deleting the first and last block-rows in $\mathcal{F}_d$ or in $\mathcal{F}_c$:

$$\mathcal{F}(\mathcal{M}_0, \mathcal{M}_+, \mathcal{M}_-) = \begin{bmatrix} -\mathcal{M}_-(1) & \mathcal{M}_0(1) & -\mathcal{M}_+(1) & 0 & \cdots & 0 \\ 0 & -\mathcal{M}_-(2) & \mathcal{M}_0(2) & -\mathcal{M}_+(2) & & 0 \\ \vdots & & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -\mathcal{M}_-(\bar{N}) & \mathcal{M}_0(\bar{N}) & -\mathcal{M}_+(\bar{N}) \end{bmatrix} \tag{247}$$

and finally, let $\mathcal{F}'$ be the matrix which results from erasing the first and last block columns in $\mathcal{F}$

$$\mathcal{F}'(\mathcal{M}_0, \mathcal{M}_+, \mathcal{M}_-) = \begin{bmatrix} -\mathcal{M}_0(1) & \mathcal{M}_+(1) & 0 & \cdots & & 0 \\ \mathcal{M}_-(2) & -\mathcal{M}_0(2) & \mathcal{M}_+(2) & & & 0 \\ \vdots & & & \ddots & \ddots & \vdots \\ 0 & \cdots & & 0 & \mathcal{M}_-(\bar{N}) & -\mathcal{M}_0(\bar{N}) \end{bmatrix} \tag{248}$$

Finally, let us consider the two cases of *deterministic* Dirichelet and cyclic boundary conditions for the difference equation (239):

Deterministic Dirichelet boundary condition, (DDBC) :

$$u(0) = u_i, \qquad u(N) = u_f, \tag{249}$$

Deterministic cyclic boundary condition, (DCBC) : the indeces set $\mathbf{T} = [0, N]$ is the unit circle, therefore the sequence u satisfies

$$u(N+1) = u(0), \qquad u(-1) = u(N). \tag{250}$$

Now, let be given sequences of $(\alpha \times \alpha)$-dimensioned matrices, $\mathcal{M}_0, \mathcal{M}_\pm$, satisfying (240), (241). We can state the following lemmas.

Lemma 5.1. *Equation (239) is equivalent to the following matrix equations*
Case DDBC :

$$\mathcal{F}_d u = \theta, \tag{251}$$

where

$$v' = [v(k)]_{k=1,...,N-1} \tag{252}$$

$$\theta^T = \left[\begin{array}{ccc} u_i^T & v'^T & u_f^T \end{array} \right]. \tag{253}$$

Case DCBC :

$$\mathcal{F}_c u = v, \tag{254}$$

where $v = [v(k)]_{k=0,...,N}$.

Proof. In both cases the thesis readily follows from assembling eq. (239) for $k = 0, ..., N$, while accounting of the definitions of the involved vectors and matrices. ∎

Lemma 5.2. *Suppose that DDBC are imposed and the matrix $\mathcal{F}_d(\mathcal{M}_0, \mathcal{M}_+, \mathcal{M}_-)$ is nonsingular, then there exists an unique solution $u = [u(k)]_{k=0,...,N}$ for eq. (239) given by*

$$u(k) = \sum_{l=1}^{N-1} \mathcal{G}_d(k,l) v(l) + \mathcal{G}_d(k,0) u_i + \mathcal{G}_d(k,N) u_f, \tag{255}$$

where the matrices $\mathcal{G}_d(k,l) \in \mathrm{IR}^{\alpha \times \alpha}$, are defined by

$$\mathcal{F}_d^{-1}(\mathcal{M}_0, \mathcal{M}_+, \mathcal{M}_-,) = [\mathcal{G}_d(k,l)]_{k=0,...,N}^{l=0,...,N}, \tag{256}$$

Also, suppose that DCBC are imposed and the matrix $\mathcal{F}_c(\mathcal{M}_0, \mathcal{M}_+, \mathcal{M}_-)$ is nonsingular, then there exists an unique solution $u = [u(k)]_{k=0,...,N}$ for eq. (239) given by

$$u(k) = \sum_{l=0}^{N} \mathcal{G}_c(k,l) \phi(l), \tag{257}$$

where the matrices $\mathcal{G}_c(k,l) \in \mathrm{IR}^{\alpha \times \alpha}$, are defined by

$$\mathcal{F}_c^{-1}(\mathcal{M}_0, \mathcal{M}_+, \mathcal{M}_-,) = [\mathcal{G}_c(k,l)]_{k=0,...,N}^{l=0,...,N}. \tag{258}$$

Proof. The equivalent representation (251) being invertible by hypotheses is ensued by

$$u = \mathcal{F}_d^{-1}\theta, \tag{259}$$

and similarly the equivalent representation (254) gives

$$u = \mathcal{F}_c^{-1}v. \tag{260}$$

Thus, by (259), (260), while accounting of the definition of θ, and v, and by decomposing $\mathcal{F}_d^{-1}$, $\mathcal{F}_c^{-1}$ in $(\alpha \times \alpha)$-blocks, as in (256) and (258) respectively, expressions (255), (257) are readily derived. ∎

Lemma 5.3. *Suppose that, for a given choice of matrices $\mathcal{M}_0(k), \mathcal{M}_-(k), \mathcal{M}_+(k)$ the matrices $\mathcal{F}_d, \mathcal{F}_c$ are nonsingular. Then the matrix $\mathcal{F}'$ admits the following decomposition*

$$\mathcal{F} = \mathcal{L}(\mathcal{A})\mathcal{Q}^{-1}\mathcal{H}(\mathcal{A}), \tag{261}$$

where $\mathcal{L}(\mathcal{A}), \mathcal{Q}, \mathcal{H}(\mathcal{A})$ are the following block matrices:

$$\mathcal{L}(\mathcal{A}) = \begin{bmatrix} I & -\mathcal{A}^T(1) & 0 & \dots & 0 \\ 0 & I & -\mathcal{A}^T(2) & & \\ \vdots & & & \ddots & \\ 0 & & & I & -\mathcal{A}^T(N-1) \end{bmatrix}, \tag{262}$$

$$\mathcal{H}(\mathcal{A}) = \begin{bmatrix} -\mathcal{A}(0) & I & 0 & \dots & & 0 \\ 0 & -\mathcal{A}(1) & I & & & \\ & & -\mathcal{A}(2) & & & \vdots \\ \vdots & & & & I & 0 \\ 0 & & & & -\mathcal{A}(N-1) & I \end{bmatrix}, \tag{263}$$

$$\mathcal{Q} = \mathrm{diag}\,\{\mathcal{Q}(k)\}_{k=0,\dots,N-1} \tag{264}$$

where $\mathcal{A}$ is a sequence of matrices defined by

$$\mathcal{A}(k) = \mathcal{Q}(k)\mathcal{M}_+^T(k), \qquad k = 0, ..., N-1 \tag{265}$$

and $\mathcal{Q}(k) > 0$, for $k = 0, ..., N-1$ satisfies the following backward recursive matrix equation

$$\mathcal{Q}^{-1}(k-1) = \mathcal{M}_0(k) - \mathcal{M}_+(k)\mathcal{Q}(k)\mathcal{M}_+^T(k), \qquad \mathcal{Q}(N) = 0. \tag{266}$$

Proof. The proof can be carried out with a standard argument, just used for instance in [19], thus we only sketch it. By hypothesis one has $\mathcal{F}'(\mathcal{M}_0, \mathcal{M}_+, \mathcal{M}_-) > 0$, because it is a symmetric matrix, and it is a principal minor of both the nonsingular matrices $\mathcal{F}_d$, $\mathcal{F}_c$. Therefore, it can be decomposed by Cholesky method as

$$\mathcal{F}' = \mathcal{L}(\mathcal{A})\mathcal{Q}^{-1}\mathcal{L}^T(\mathcal{A}), \tag{267}$$

and the existence of solutions for eq. (266) follows by noticing that it can be viewed as the standard backward Cholesky recursion for the above factorization. By using eqs. (265), (266), decomposition (261) can be readily verified by direct calculation. ∎

Theorem 5.4. *Let us consider the following stochastic equation on* $\mathbf{T}$*:*

$$\mathcal{M}_0(k)z(k) - \mathcal{M}_+(k)z(k+1) - \mathcal{M}_-(k)z(k-1) = \phi(k), \tag{268}$$

under the (stochastic) CBC (237), (with x being replaced by z), or the DBC (236), with $x_i = z_i$, $x_f = z_f$ where z_i, z_f are random vectors such that

$$\mathbf{E}\{z_i\} = 0, \quad \mathbf{E}\{z_i z_i^T\} > 0, \tag{269}$$
$$\mathbf{E}\{z_f\} = 0, \quad \mathbf{E}\{z_f z_f^T\} > 0. \tag{270}$$

Suppose that $\mathcal{F}_d(\mathcal{M}_0, \mathcal{M}_+, \mathcal{M}_-)$ or $\mathcal{F}_c(\mathcal{M}_0, \mathcal{M}_+, \mathcal{M}_-)$ are nonsingular, depending on whether DBC or CBC respectively are imposed. Then for any process ϕ in the set $\mathcal{N}(\mathcal{M}_0, \mathcal{M}_+, \mathcal{M}_-)$, there exists an unique, nonsingular, reciprocal, and samplewise solution – namely z_d^ϕ, z_c^ϕ for DBC and CBC respectively – of the stochastic equation (268), and ϕ is the conjugate process of the corresponding solution i.e.:

$$\mathbf{E}\{\phi(k) z_d^{\phi^T}(l)\} = \mathbf{E}\{\phi(k) z_c^{\phi^T}(l)\} = I\delta(k-l), \qquad \forall k, l \in \mathbf{T}. \tag{271}$$

Moreover, denoting $\mathcal{Z}_d, \mathcal{Z}_c$ the sets of the solutions of (268):

$$\mathcal{Z}_d = \left\{ z_d^\phi : \quad \phi \in \mathcal{N}(\mathcal{M}_0, \mathcal{M}_+, \mathcal{M}_-) \right\}$$
$$\mathcal{Z}_c = \left\{ z_c^\phi : \quad \phi \in \mathcal{N}(\mathcal{M}_0, \mathcal{M}_+, \mathcal{M}_-) \right\}$$

one has, given any $\zeta_1, \zeta_1' \in \mathcal{Z}_d$ and $\zeta_2, \zeta_2' \in \mathcal{Z}_c$:

$$\mathbf{E}\{\zeta_i\} = \mathbf{E}\{\zeta_i'\}; \quad \mathrm{Cov}\{\zeta_i, \zeta_i\} = \mathrm{Cov}\{\zeta_i', \zeta_i'\}, \qquad i = 1, 2. \tag{272}$$

Proof. Consider first the case of Dirichelet boundary conditions, by applying Lemma 5.1 samplewise to eq. (268) we have

$$\mathcal{F}_d z = \theta, \tag{273}$$

where

$$\phi' = [\phi(k)]_{k=1,\dots,N-1} \tag{274}$$
$$\theta^T = \left[\begin{array}{ccc} z_i^T & \phi'^T & z_f^T \end{array} \right]. \tag{275}$$

By using the block-matrix (247) we can extract from (273) the following equation

$$\mathcal{F} z = \phi'. \tag{276}$$

Since, by hypothesis, $\mathcal{F}_d(\mathcal{M}_0, \mathcal{M}_+, \mathcal{M}_-)$ is nonsingular eq. (273) gives the (unique) solution

$$z = \mathcal{F}_d^{-1}\theta. \tag{277}$$

By the definition of ϕ we easily recognize that

$$\mathbf{E}\{\phi\phi^T\} = \mathcal{F}', \tag{278}$$

where $\mathcal{F}'$ is the matrix defined in (248), which is nonsingular because it is a principal minor of the nonsingular matrix $\mathcal{F}_d$. From (277) we have:

$$\mathbf{E}\{zz^T\} = \mathcal{F}_d^{-1}\mathbf{E}\{\theta\theta^T\}\mathcal{F}_d^{-1T}, \tag{279}$$

where $\mathbf{E}\{\theta\theta^T\} > 0$ as it has the blocks $\mathbf{E}\{z_i z_i^T\} > 0$, $\mathbf{E}\{z_f z_f^T\} > 0$, and $\mathcal{F}' > 0$ on the block diagonal, thus we follow through on proving the nonsingularity of the solution. Also, defining $z' = [z(k)]_{k=1,\ldots,N-1}$ eq. (276) rewrites

$$\mathcal{F}'z' = \phi, \tag{280}$$

from which, on account of (278) it follows that

$$\mathbf{E}\{z\phi^T\} = I,$$

that is, ϕ is the conjugate process of z.

As to the case of cyclic boundary conditions, define the vector

$$\phi = [\phi(k)]_{k=0,\ldots,N}\,. \tag{281}$$

By an application of Lemma 5.1 to each sample of eq. (268) we obtain

$$\mathcal{F}_c z = \phi. \tag{282}$$

where $\mathcal{F}_c$ is the block-matrix defined in (246). By (282), and since $\mathcal{F}_c$ is invertible by hypothesis, we have that the process z

$$z = \mathcal{F}_c^{-1}\phi, \tag{283}$$

is the unique solution of (268) with cyclyc boundary conditions. By properties (242)–(244), once k is interpreted modulo N one has

$$\mathbf{E}\{\phi\phi^T\} = \mathcal{F}_c,$$

thus, eq. (283) directly gives

$$\mathbf{E}\{zz^T\} = \mathcal{F}_c^{-1},$$

which proves the nonsingularity of z, and

$$\mathbf{E}\{z\phi^T\} = I,$$

which proves that ϕ is the conjugate process of z in the CBC case.

As far as eq. (272) is concerned, since the processes in the set $\mathcal{N}(\mathcal{M}_0, \mathcal{M}_+, \mathcal{M}_-)$ have all the same correlation function, described by eqs. (242)–(244), eq. (272) readily follows from the linearity of eqs. (277) and (283).

It remains to prove that the above found solutions are reciprocal. As a matter of fact, denote z any of the two solutions for the cases DBC and CBC. We have just shown before that any $\phi \in \mathcal{N}$ gives place to a solution z of (268) such that ϕ is the conjugate process of z. Then, one has

$$\mathbf{E}\left\{\phi(k)|z(i), i \neq k\right\} = 0 = \mathbf{E}\left\{\phi(k)|z(k-1), z(k+1)\right\}, \tag{284}$$

therefore

$$\begin{aligned}\mathbf{E}\left\{z(k)|z(i), i \neq k\right\} \quad &= \mathcal{M}_0^{-1}(k)\mathcal{M}_+(k)z(k+1) + \mathcal{M}_0^{-1}(k)\mathcal{M}_-(k)z(k-1) \\ &= \mathbf{E}\left\{z(k)|z(k-1), z(k+1)\right\}.\end{aligned}$$

∎

5.2. Explicit representations for the polynomial non-causal model

We will show now that the polynomial representation (218), (219), is indeed wide-sense well-posed in both the DBC and CBC cases (236), (237). We remind that 'well-posed in a wide sense' here means that representation (218), (219) admits an explicit representation in the form of (235), or (238).

Before showing this, a further comment should be done. The well-posedness problem, even in a wide-sense, for the representation (218), (219), actually arise only for the first equation, which rules the evolution of X. As a matter of fact, (218) is a *non-causal* equation, whereas by (219) we see that $\mathcal{X}$, the collection of the *mixed powers* (at different k's) of the process x, is determined (in the stochastic sense) by a white noise η and by the process X playing the role of an external stochastic forcing term. Thus, once the (wide-sense) well-posedness has been proven as to eq. (218), it results proven as to eq. (219) as well.

Let us briefly remind our setting: we are supposing that a nonsingular reciprocal chain x on $\mathbf{T} = [0, N]$ there exists, such that it satisfies the DBC (236), or the CBC (237). Then we can build up, using moments of process x up to order μ, the covariances (204)–(207) from which we compute the matrices $D(k), B(k)$ (by (211), (212)). Next, we can compute the matrix-coefficients $M_0, M_\pm, S_\pm$ (by (231)–(233), and (208)–(210)), and the function $D(\cdot,\cdot)$ by eq. (213). Last, we can consider eqs. (218), (219) with $\xi(k) = D^{-1}(k)d(k), \eta(k) = B^{-1}(k)b(k)$ where d, b are *any* couple of zero-mean, mutually uncorrelated processes, with covariances $D(k), B(k)$ respectively, b white, d one-step correlated with $\mathbf{E}\{d(r)d^T(s)\} = D(r, s)$ and satisfying the orthogonality relations (107)–(114). As we have before remarked, in the present non-Gaussian case, there are in general *many* processes b, d (or, equivalently, ξ, η) satisfying the above construction, and the set of them is indeed $\mathcal{N}(M_0, M_-, M_+)$.

Now we are in a position to show the following theorems, whose proofs follow along with [19], even though in the present case we are working with a *substantially* different representation. As a matter of fact, in the Gaussian case considered in [19], an equation formally identical to (218), satisfied by x directly, was proven with a *Gaussian*, one-step correlated forcing noise ξ. In that situation there is only one choice for a Gaussian one-step correlated noise with a fixed correlation. Counterwise, in our non-Gaussian situation there are in general many noises ξ (and η) belonging to the class $\mathcal{N}$. Nevertheless, we prove now that explicit representations formally similar to the Gaussian case can be obtained, but these should be thought in a 'wide sense' manner.

Recall that $\widetilde{U} = U - \mathbf{E}\{U\}$ for any random vector U. Moreover, let us introduce the following notation: for a block-matrix $V = [V_{i,j}]_{i=0,\dots,N}^{j=0,\dots,N}$, $V_{i,j} \in \mathbb{R}^{\alpha\times\alpha}$ we denote by

$$(V)^* = [V_{i,j}]_{i=1,\dots,N-1}^{j=0,\dots,N} \quad , \tag{285}$$

the matrix which results from erasing the first and last block-rows in V.

Theorem 5.5. *Let x a reciprocal chain on $\mathbf{T}$, specified by DBC or CBC. Let us suppose that x is nonsingular and therefore it satisfies the polynomial representation (218). Then, the following matrces have full-rank:*

$$\mathbf{F} = \mathcal{F}(M_0, M_+, M_-), \tag{286}$$

$$\mathbf{F}_d = \mathcal{F}_d(M_0, M_+, M_-), \tag{287}$$

$$\mathbf{F}_c = \mathcal{F}_c(M_0, M_+, M_-), \tag{288}$$

where $\mathcal{F}, \mathcal{F}_d, \mathcal{F}_c$, are the matrix functions defined in (247), (245),(246). Moreover, x can be expressed by the following explicit representations, for any $k \in \mathbf{T}$:

$$x(k) = \mathbf{E}\{x(k)\} + \mathcal{E}^n \widetilde{X}(k), \tag{289}$$

$$\widetilde{X}(k) = \sum_{l=1}^{N-1} \Gamma_d(k,l)\xi(l) + \Gamma_d(k,0)\widetilde{X}_i + \Gamma_d(k,N)\widetilde{X}_f, \tag{290}$$

$$X_i = \left[x_{i[h]}\right]_{h=1,\dots,\nu}, \qquad X_f = \left[x_{f\,[h]}\right]_{h=1,\dots,\nu}, \tag{291}$$

for DBC, and

$$x(k) = \mathbf{E}\{x(k)\} + \mathcal{E}^n \widetilde{X}(k), \tag{292}$$

$$\widetilde{X}(k) = \sum_{l=0}^{N} \Gamma_c(k,l)\xi(l), \tag{293}$$

for CBC. In (289), (292), $\mathcal{E}^n \in \mathbb{R}^{n\times\alpha}$ is the matrix extracting the first n entries of its argument, whereas in (290), (293), the matrices $\Gamma_d(k,l) \in \mathbb{R}^{\alpha\times\alpha}$, and $\Gamma_c(k,l) \in \mathbb{R}^{\alpha\times\alpha}$ are defined by

$$\mathbf{F}_d^{-1} = [\Gamma_d(k,l)]_{k=0,\dots,N}^{l=0,\dots,N}, \tag{294}$$

$$\mathbf{F}_c^{-1} = [\Gamma_c(k,l)]_{k=0,\dots,N}^{l=0,\dots,N}. \tag{295}$$

Finally, defining

$$X = [X(k)]_{k=0,\dots,N}, \tag{296}$$

$$\mathbf{R}_X = \mathbf{E}\left\{\widetilde{X}\widetilde{X}^T\right\}, \tag{297}$$

one has for DBC:

$$(\mathbf{R}_X^{-1})^* = \mathbf{F}, \tag{298}$$

and for CBC:

$$\mathbf{R}_X^{-1} = \mathbf{F}_c, \tag{299}$$

where in (298) notation (285) has been used.

Proof. First of all, notice that the Dirichelet conditions (236) imply

$$\widetilde{X}(0) = \widetilde{X}_i, \qquad \widetilde{X}(N) = \widetilde{X}_f, \tag{300}$$

where X_i, X_f are the vectors defined in (291). Now define the vectors

$$\xi = [\xi(k)]_{k=1,\dots,N-1}; \tag{301}$$

$$\theta^T = \left[\begin{array}{ccc} \widetilde{X}_i^T & \xi^T & \widetilde{X}_f^T \end{array}\right]. \tag{302}$$

By using the block-matrices (286) and (287), and the vector (296), eq. (218) can be assembled for $k = 0, \dots, N$ as follows

$$\mathbf{F}_d\widetilde{X} = \theta, \tag{303}$$

and for $k = 1, \dots, N-1$

$$\mathbf{F}\widetilde{X} = \xi. \tag{304}$$

From the latter, by the nonsingularity assumption, and by the orthogonality property (226) eq. (298) follows.

For cyclic boundary conditions, define the vector

$$\xi = [\xi(k)]_{k=0,\dots,N}\,. \tag{305}$$

Since the sum operation in $\mathbf{T}$ is interpreted modulo N, eq. (218) can be aggregated, by using (305), for $k = 0, ..., N$ as follows

$$\mathbf{F}_c \widetilde{X} = \xi. \tag{306}$$

where $\mathbf{F}_c$ is the block-matrix defined in (288). Equation (306), the nonsingularity assumption, and the orthogonality property: $\mathbf{E}\{\xi(k)\widetilde{X}(l)\} = I\delta(k-l)$ are ensued by eq. (299). Thus, we followed through on showing that $\mathbf{F}_c$ is invertible, which in turn implies that $\mathbf{F}$ (which results from $\mathbf{F}_c$ while deleting the first and last block-rows) is full-rank under cyclic boundary condition as well.

Equations (298), (299) imply that the hypotheses of Lemma 5.2 are fulfilled as to the sequence of matrices $\mathcal{M}_0 = M_0$, $\mathcal{M}_\pm = M_\pm$. Therefore the theorem's proof is achieved by using eqs. (255), (257) for the couple of conjugate processes $(\widetilde{X}, \xi)$. ∎

5.3. Polynomial-optimal smoothers.

In the present subsection we show how the explicit representations given in Theorem 5.5 can be used in order to solve smoothing problems for a nonsingular reciprocal chain x.

Let us consider the reciprocal chain x partially and noisy observed by

$$y(k) = C(k)x(k) + w(k), \tag{307}$$

where w is a non-Gaussian, white noise sequence in $\mathbb{R}^m$ with $E\{w(k)w^T(k)\} = R(k) > 0$. We wish to find filtering and smoothing estimates, over the interval $[0, N]$, of the process x in both cases of DBC and CBC. We will show that a kind of *suboptimal* recursive-smoother is achievable, namely a polynomial-optimal one, which is a recursive algorithm giving the μ-polynomial-optimal estimate of $x(k)$, for *any* positive integer μ, for any $k \in [0, N]$.

To this purpose let us recall some basic definitions. Let us consider two random vectors, namely $U \in \mathbb{R}^n$, $V \in \mathbb{R}^m$, with finite second order moments, i.e.

$$\mathbf{E}\{U_{[2]}\}, \mathbf{E}\{V_{[2]}\} < +\infty. \tag{308}$$

As is well known, for any pair of random vectors U, V which verify condition (308), it is well defined the *optimal-linear estimate of U given V*, namely $\lambda_U(V)$:

$$\lambda_U(V) = \arg\min_{u \in \mathcal{L}(V)} \mathbf{E}\{\|U - u\|^2\}, \tag{309}$$

where

$$\mathcal{L}(V) = \{u = \lambda(V) \ / \quad \lambda : \mathbb{R}^m \to \mathbb{R}^n; \qquad \lambda \text{ is linear}\}\,. \tag{310}$$

Suppose there exists a positive integer μ such that $\mathbf{E}\{V_{[2\mu]}\} < +\infty$. For the random vector $\mathcal{V} = \{V_{[i]}\}_{i=1,\dots\mu}$ one has $\mathbf{E}\{\mathcal{V}_{[2]}\} < +\infty$. We define the *$\mu$-polynomial-optimal estimate of U given V*, as the optimal-linear estimate of U given $\mathcal{V}$, that is $\lambda_U(\mathcal{V})$.

Let μ be any positive integer. Up to now we have considered $X(k) = X(\mu;k)$, with $X(\cdot;k)$ defined in (43) for $\mu > \nu$ only. Now let us slightly modify the definition of the vector $X(k)$ as follows:

$$X(k) = \begin{cases} \left[x_{[h]}\right]_{h=1,\dots,\mu}, & \text{if} \quad \mu > \nu; \\ \left[x_{[h]}\right]_{h=1,\dots,\nu}, & \text{if} \quad \nu \geq \mu, \end{cases} \tag{311}$$

and define $\mathcal{E}^{\alpha}_{\mu,\nu}$ as the matrix performing the following *extraction*:

$$X(\mu;k) = \mathcal{E}^{\alpha}_{\mu,\nu} X(k). \tag{312}$$

Notice that, from (311), for $\mu \geq \nu$ one has: $\mathcal{E}^{\alpha}_{\mu,\nu} = I_\alpha$. Moreover, let us define the *augmented observation*

$$Y(k) = \left[y_{[i]}(k)\right]_{i=1,\dots,\mu}. \tag{313}$$

Theorem 5.6. *The process Y satisfies the following recursive equation*

$$\widetilde{Y}(k) = \mathcal{C}(k)\widetilde{X}(k) + W(k), \tag{314}$$

where W is a zero mean orthogonal sequence, uncorrelated to X, defined as

$$W(k) = \widetilde{T}^{(m)} W'(k), \tag{315}$$

$$\widetilde{T}^{(m)} = \text{diag}\{\widetilde{T}^{(m)}_i\}_{1,\dots,\mu} \quad , \tag{316}$$

$\widetilde{T}^{(m)}_i$ being reduction matrices, as defined in (15), and

$$W'^{T}(k) = [h_1^T(k) \cdots h_\mu^T(k)], \tag{317}$$

$$h_i(k) = \sum_{l=0}^{i-1} \mathbf{M}^i_{i-l}(I_{m^{i-l}} \otimes C^{[l]}(k))(w^{[i-l]}(k) - \mathbf{E}\{w^{[i-l]}(k)\}) \otimes I_{n^l})x^{[l]}(k). \tag{318}$$

$$\mathcal{C}(k) = \widetilde{T}^{(m)}\mathcal{C}'(k)T^{(n)}\mathcal{E}^{\alpha}_{\mu,\nu} \qquad T^{(n)} = \text{diag}\{T^{(n)}_i\}_{1,\dots,\mu} \quad , \tag{319}$$

$$\mathcal{C}'(k) = \begin{bmatrix} C(k) & 0 & \dots & 0 \\ L_{2,1}(k) & C^{[2]}(k) & \dots & \\ \vdots & & \ddots & \vdots \\ L_{\mu,1}(k) & L_{\mu,2}(k) & \dots & C^{[\nu]}(k) \end{bmatrix}, \tag{320}$$

$$L_{i,l}(k) = \mathbf{M}^i_{i-l}(I_{m^{i-l}} \otimes C^{[l]}(k))(\mathbf{E}\{w^{[i-l]}(k)\} \otimes I_{n^l}) \tag{321}$$

The matrices $T^{(n)}_i$ in (319) has been defined in (16), whereas $\mathcal{E}^{\alpha}_{\mu,\nu}$ is the extraction matrix above defined by eq. (312). The matrices $\mathbf{M}^i_j$ appearing in (318) and (321) *are the "binomial matrices" defined in* [5], *given by the recursive equations (26), (27).*

Proof. Denote

$$Y'(k) = \left[y^{[i]}(k)\right]_{i=1,\dots,\mu} \qquad U(k) = \left[\mathbf{E}\{w^{[i]}(k)\}\right]_{i=1,\dots,\mu}$$

$$X'(k) = \begin{cases} \left[x^{[h]}\right]_{h=1,\dots,\mu}, & \text{if} \quad \mu > \nu; \\ \left[x^{[h]}\right]_{h=1,\dots,\nu}, & \text{if} \quad \nu \geq \mu, \end{cases}$$

Then, an application of Lemma 3.3.1 in [5] results in the following equation:

$$Y'(k) = \mathcal{C}'(k)X'(\mu;k) + U(k) + W'(k), \tag{322}$$

with $\mathcal{C}', W'$ given by (320) and (317) respectively, and $X'(\cdot;k)$ defined by (44). Lemma 3.3.1 of [5] assures that W' is a zero-mean white sequence ortoghonal to x and all of its Kronecker powers as well. On account of $\mathbf{E}\{W'(k)\} = 0$ (hence, by (322), $U(k) = \mathbf{E}\{Y'(k)\} - \mathcal{C}'(k)\mathbf{E}\{X'(\mu;k)\}$), eq. (322) results in

$$Y'(k) - \mathbf{E}\{Y'(k)\} = \mathcal{C}'(k)(X'(\mu;k) - \mathbf{E}\{X'(\mu;k)\}) + W'(k). \tag{323}$$

Finally, multiplying (323) by $\widetilde{T}^{(m)}$, using $X'(\mu;k) = T^{(n)}X(\mu;k)$, where $X(\cdot,k)$ is defined in (43), and taking into account of (312) result in eq. (314). ∎

Let us denote by $\mathbf{R}_W(k), \mathbf{R}_{W'}(k)$ the covariances of the zero-mean processes W, W', mutually related by eq. (315). Since

$$W'(k) = T^{(m)}W(k),$$

with $T^{(m)} = \text{diag}\{T_i^{(m)}\}_{i=1,\dots,\mu}$, one has

$$\mathbf{R}_W(k) = \widetilde{T}^{(m)}\mathbf{R}_{W'}(k)T^{(m)}. \tag{324}$$

The covariance $\mathbf{R}_{W'}(k)$ can be calculated from moments of $w(k)$ and $x(k)$ (here supposed known) by using a formula which can be found in [5] (in specific: formula (3.3.12)). We omit here that expression as unessential to our present purposes. Let us define the processes X, W, Y, by the usual notation:

$$X = [X(k)]_{k=0,\dots,N} \qquad W = [W(k)]_{k=0,\dots,N} \qquad Y = [Y(k)]_{k=0,\dots,N}, \tag{325}$$

where $W(k), Y(k)$ are the noise and output samples in the output equation (314), and $X(k)$ has been defined in (311).

We are now in a position to state the following theorem, which provides, for any positive integer μ, the algorithm of the μ-polynomial-optimal *smoothing estimate* of $x(k)$ given the observations (307), namely

$$\hat{x}(k) = \lambda_{x(k)}(Y), \tag{326}$$

where the reciprocal chain x is specified on the finite discrete interval $\mathbf{T}$ by some Dirichelet or cyclic boundary condition, and in the DBC case (236) we suppose that the random variables x_i, x_f, assigned to the end points of $\mathbf{T}$, can be measured without noise.

Theorem 5.7. *Let x be a nonsingular reciprocal chain on $\mathbf{T}$. Assume the output-noise-covariance $\mathbf{R}_W = \mathbf{E}\left\{WW^T\right\}$ is nonsingular. Then, the matrices $\mathcal{F}_d(\overline{M}_0, M_+, M_-)$ and $\mathcal{F}_c(\overline{M}_0, M_+, M_-)$ are nonsingular, where for $k = 0, ..., N$:*

$$\overline{M}_0(k) = M_0(k) + K(k)C(k), \tag{327}$$

$$K(k) = \mathcal{C}^T(k)\mathbf{R}_W^{-1}(k), \tag{328}$$

and $M_0, M_\pm$ are the matrix coefficients of the normalized nearest-neighbour model (218). The μ-degree polynomial-optimal smoothing estimate of $x(k)$, namely $\hat{x}(k)$ defined in (326), for any $k \in [0, N]$, is given by the following algorithm

$$\hat{x}(k) = \mathbf{E}\{x(k)\} + \mathcal{E}^n \hat{\widetilde{X}}(k), \tag{329}$$

where

$$\hat{\widetilde{X}} = \lambda_{\widetilde{X}}(Y), \tag{330}$$

with X, Y *defined in (325),* $\lambda.(\cdot)$ *defined in (309). The process* $\hat{\widetilde{X}}$ *satisfies, for the DBC case:*

$$\hat{\widetilde{X}}(k) = Z(k) + \bar{\Gamma}_d(k,N)(\widetilde{X}_f - Z(N)), \tag{331}$$

with

$$X_i = \left[x_{i[h]}\right]_{h=1,\dots,\max\{\nu,\mu\}}, \quad \widetilde{X}_i = X_i - \mathbf{E}\{X_i\}, \tag{332}$$

$$X_f = \left[x_{f[h]}\right]_{h=1,\dots,\max\{\nu,\mu\}}, \quad \widetilde{X}_f = X_f - \mathbf{E}\{X_f\} \tag{333}$$

and for the CBC case:

$$\begin{aligned}\hat{\widetilde{X}}(k) = Z(k) &+ \bar{\Gamma}_c(k,0)(K(0)\widetilde{Y}(0) + M_+(0)Z(1) - M_0(N)Z(N)) \\ &+ \bar{\Gamma}_c(k,N)(K(N)\widetilde{Y}(N) + M_-(N)Z(N-1) - M_-(0)Z(N)),\end{aligned} \tag{334}$$

with the set of matrices $\bar{\Gamma}_d(k,j)$*, and* $\bar{\Gamma}_c(k,j)$*,* $k = 0,...,N$*,* $j = 0, N$ *defined by the relations*

$$\mathcal{F}_d^{-1}(\overline{M}_0, M_+, M_-) = \left[\bar{\Gamma}_d(k,l)\right]_{k=0,\dots,N}^{l=0,\dots,N}, \tag{335}$$

$$\mathcal{F}_c^{-1}(\overline{M}_0, M_+, M_-) = \left[\bar{\Gamma}_c(k,l)\right]_{k=0,\dots,N}^{l=0,\dots,N}. \tag{336}$$

The process Z *is given by the forward recursive equation*

$$Z(k) = A(k-1)Z(k-1) + Q(k-1)\gamma(k), \qquad Z(0) = Z_i, \tag{337}$$

where $Z_i = \widetilde{X}_i$ *or* $Z_i = 0$ *depending on whether it is the case of Dirichelet or cyclic boundary conditions respectively. The process* γ *is given by the backward observations-driven equation*

$$\gamma(k) = A^T(k)\gamma(k+1) + K(k)\widetilde{Y}(k), \qquad \gamma(N-1) = 0, \tag{338}$$

with A *sequence of matrices defined by*

$$A(k) = Q(k)M_+^T(k), \tag{339}$$

and Q *solution of the backward matrix equation*

$$Q^{-1}(k-1) = \overline{M}_0(k) - M_+(k)Q(k)M_+^T(k). \qquad Q(N) = 0. \tag{340}$$

Proof. Let us group together eqs. (314) for $k = 0,...,N$, the result is

$$\widetilde{Y} = \mathcal{C}\widetilde{X} + W, \tag{341}$$

where X, W, Y are defined in (325), and

$$\mathcal{C} = \mathrm{diag}\{\mathcal{C}(k)\}_{k=0,\ldots,N}.$$

Now, by its definition, given in (330), $\widehat{\widehat{X}}$ is the L^2-orthogonal projection onto the space linearly spanned by Y, namely $\mathcal{L}(Y)$ defined in (310). Since $\mathcal{L}(Y) = \mathcal{L}(\widetilde{Y})$ we can compute $\widehat{\widehat{X}}$ by the optimal linear estimation formula applied to $(\widetilde{X}, \widetilde{Y})$, and the result is

$$\mathbf{R}_X^{-1}\widehat{\widehat{X}} = \mathcal{C}^T\mathbf{R}_W^{-1}(\widetilde{Y} - \mathcal{C}\widehat{\widehat{X}}). \tag{342}$$

Thus, using notation (285), eq. (342) implies

$$\left((\mathbf{R}_X^{-1})^* + (K\mathcal{C})^*\right)\widehat{\widehat{X}} = K^*\widetilde{Y}^*, \tag{343}$$

where

$$\begin{aligned} K(k) &= \mathcal{C}^T(k)\mathbf{R}_W^{-1}(k), \qquad & \widetilde{Y}^* &= \{\widetilde{Y}(k)]_{k=1,\ldots,N-1} \\ K &= \mathrm{diag}\{K(k)\}_{k=0,\ldots,N} \qquad & K^* &= \mathrm{diag}\{K(k)\}_{k=1,\ldots,N-1} \end{aligned}$$

It results $(\mathbf{R}_X^{-1})^* = \mathbf{F}$, by (298), therefore by (343), and looking at the structure of of $\mathbf{F}$ given by (286), for $k = 1, \ldots, N-1$, we realize that

$$\overline{M}_0(k)\widehat{\widehat{X}}(k) - M_-(k)\widehat{\widehat{X}}(k-1) - M_+(k)\widehat{\widehat{X}}(k+1) = K(k)\widetilde{Y}(k), \tag{344}$$

with

$$\overline{M}_0(k) = M_0(k) + K(k)C(k). \tag{345}$$

Therefore eq. (343) can be rewitten

$$\mathbf{\Lambda}\widehat{\widehat{X}} = K^*\widetilde{Y}^*, \tag{346}$$

with

$$\mathbf{\Lambda} = \mathcal{F}(\overline{M}_0, M_+, M_-). \tag{347}$$

and $\mathcal{F}$ being the matrix function defined in (247). Now, by Theorem 5.5, $\mathbf{F}_d$ is invertible, so its principal minor, given by $\mathcal{F}^*(M_0, M_+, M_-)$ ($\mathcal{F}^*$ defined by (248)) is invertible as well. Also, it is positive definite as symmetric with block diagonal elements $M_0(k) > 0$. Now, one has

$$\mathcal{F}^*(\overline{M}_0, M_+, M_-) = \mathcal{F}^*(M_0, M_+, M_-) + (K\mathcal{C})^{**},$$

Where

$$(K\mathcal{C})^{**} = \mathrm{diag}\{K(k)\mathcal{C}(k)\}_{k=1,\ldots,N-1} \quad .$$

Thus, $\mathcal{F}^*(\overline{M}_0, M_+, M_-) > 0$ because $(K\mathcal{C})^{**} \geq 0$. Therefore, by its structure, $\mathcal{F}_d(\overline{M}_0, M_+, M_-)$ is nonsingular as well. Then, an application of Lemma 5.2 to *each sample* of the difference equation (344) yields

$$\widehat{\widehat{X}}(k) = \sum_{l=1}^{N-1}\bar{\Gamma}_d(k,l)K(l)\widetilde{Y}(l) + \bar{\Gamma}_d(k,0)\widehat{\widehat{X}}(0) + \bar{\Gamma}_d(k,N)\widehat{\widehat{X}}(N), \tag{348}$$

where the matrices $\bar{\Gamma}(k,l)$ are given by

$$\bar{\mathbf{F}}_d^{-1} = \left[\bar{\Gamma}_d(k,l)\right]_{k=0,\dots,N}^{l=0,\dots,N}, \tag{349}$$

with

$$\bar{\mathbf{F}}_d = \mathcal{F}_d(\overline{M}_0, M_+, M_-). \tag{350}$$

For cyclic boundary conditions, on account of (299), eq. (342) rewrites

$$(\mathbf{F}_c + K\mathcal{C})\,\widehat{\widehat{X}} = K\widetilde{Y}, \tag{351}$$

and one has

$$\mathbf{F}_c + K\mathcal{C} = \mathcal{F}_c(\overline{M}_0, M_+, M_-). \tag{352}$$

Since $\mathbf{F}_c$ is symmetric and positive definite, and $K\mathcal{C}$ symmetric and non-negative, eq. (352) implies $\mathcal{F}_c(\overline{M}_0, M_+, M-) > 0$. Therefore, by applying again lemma 5.2 (as to the case of cyclic boundary conditions) to eq. (344), it is

$$\widehat{\widehat{X}}(k) = \sum_{l=0}^{N} \bar{\Gamma}_c(k,l)K(l)\widetilde{Y}(l) \tag{353}$$

$$\tag{354}$$

where the $\bar{\Gamma}(k,l)$'s are given by

$$\bar{\mathbf{F}}_c^{-1} = \left[\bar{\Gamma}_c(k,l)\right]_{k=0,\dots,N}^{l=0,\dots,N}, \tag{355}$$

with

$$\bar{\mathbf{F}}_c = \mathcal{F}_c(\overline{M}_0, M_+, M_-). \tag{356}$$

Now, on account of the hypoteses of lemma 5.3, which are indeed satisfied with $\mathcal{M}_0 = \overline{M}_0$ and $\mathcal{M}_\pm = M_\pm$, we can apply it in order to decompose the following matrix $\mathbf{\Lambda}$

$$\mathbf{\Lambda} = \mathcal{F}(\overline{M}_0, M_+, M_-). \tag{357}$$

The result is

$$\mathbf{\Lambda} = \mathbf{L}\mathbf{Q}^{-1}\mathbf{\Omega}, \tag{358}$$

where

$$\mathbf{L} = \mathcal{L}(A), \qquad \mathbf{\Omega} = \mathcal{H}(A), \tag{359}$$

$$\mathbf{Q} = \mathrm{diag}\{Q(k)\}_{k=0,\dots,N-1}, \tag{360}$$

where A is the sequence of matrices defined by

$$A(k) = Q(k)M_+^T(k), \qquad k = 0,\dots,N-1, \tag{361}$$

and Q is the solution of the backward matrix equation

$$Q^{-1}(k-1) = \overline{M}_0(k) - M_+(k)Q(k)M_+^T(k). \qquad Q(N) = 0. \tag{362}$$

Define the sequence $\gamma = \{\gamma(k)]_{k=0,\dots,N-1}$ as

$$\gamma = \mathbf{Q}^{-1}\mathbf{\Omega}\widehat{\widehat{X}}. \tag{363}$$

Then, by using the decomposition (358) in (346), we have that γ satisfies

$$\mathbf{L}\gamma = K^*\widetilde{Y}^*. \tag{364}$$

It is easily recognized, by the structure of involved matrices, that eq. (363) is equivalent to the following forward recursive equation

$$\widehat{\widehat{X}}(k+1) = A(k)\widehat{\widehat{X}}(k) + Q(k)\gamma(k), \qquad k = 0, ..., N-1 \tag{365}$$

whereas eq. (364) is equivalent to the following backward recursive equation

$$\gamma(k-1) = A^T(k)\gamma(k) + K(k)\widetilde{Y}(k), \qquad k = 1, ..., N-1. \tag{366}$$

Let γ_0 be the solution of (366), for $\gamma(N-1) = 0$.

Now, denote Z_d the solution of (365) for $\gamma = \gamma_0$ and $Z_d(0) = \widetilde{X}_i$. With $\mathbf{\Phi}_A(k, j)$ being the transition function associated to the forward recursive equation (365), we have $Z_d(N) = Z_N$

$$Z_N = \mathbf{\Phi}_A(N, 0)\widetilde{X}_i + \sum_{s=0}^{N-1} \mathbf{\Phi}_A(N, s+1)Q(s)\gamma_0(s). \tag{367}$$

Therefore the process Z_d is trivially a solution of eq. (344), with Dirichelet boundary conditions: $Z_d(0) = \widetilde{X}_i$, $Z_d(N) = Z_N$, then it satisfies expression (348) as well, with $\widehat{\widehat{X}}(0) = \widetilde{X}_i$ and $\widehat{\widehat{X}}(N) = Z_d(N)$

$$Z_d(k) = \sum_{l=1}^{N-1} \bar{\Gamma}_d(k, l)K(l)\widetilde{Y}(l) + \bar{\Gamma}_d(k, N)\widetilde{X}_i + \bar{\Gamma}_d(k, N)Z_d(N), \tag{368}$$

whereas the solution $\widehat{\widehat{X}}$ we are searching for, is given by (348) with the *true* boundary conditions: $\widehat{\widehat{X}}(0) = \widetilde{X}_i$, $\widehat{\widehat{X}}(N) = \widetilde{X}_f$, thus

$$\widehat{\widehat{X}}(k) = Z_d(k) + \bar{\Gamma}_d(k, N)(\widetilde{X}_f - Z_d(N)). \tag{369}$$

Now, let us denote Z_c the solution of (365) for $\gamma = \gamma_0$ and $Z_c(0) = 0$. Z_c satisfies eq. (344) as well (or the matrix version of it, i.e. eq. (346)). Looking to the structure of matrices $\bar{\mathbf{F}}_c$, defined in (356), and $\mathbf{\Lambda}$, whose definition is given by (347), we realize that the following equation holds

$$\bar{\mathbf{F}}_c Z_c = \begin{bmatrix} -M_+(0)Z_c(1) + M_0(N)Z_c(N) \\ (K\widetilde{Y})^* \\ -M_-(N)Z_c(N-1) + M_-(0)Z_c(N) \end{bmatrix}, \tag{370}$$

where $Z_c(0) = 0$ has been used, the first and last block-equations are identities, and the block-equations between them simply form eq. (346). Therefore, multiplying both sides of (370) by $\bar{\mathbf{F}}_c^{-1}$, and using the matrix $\bar{\Gamma}_c$ defined in (294), one has

$$\begin{aligned} Z_c(k) \quad &= \sum_{l=1}^{N-1} \bar{\Gamma}_c(k, l)K(l)\widetilde{Y}(l) + \bar{\Gamma}_c(k, 0)(- M_+(0)Z_c(1) + M_0(N)Z_c(N)) \\ &\qquad + \bar{\Gamma}_c(k, N)(- M_-(N)Z_c(N-1) + M_-(0)Z_c(N)). \end{aligned} \tag{371}$$

Now, subtracting (371) by (353) results in

$$\widehat{\widetilde{X}}(k) = Z_c(k) + \bar{\Gamma}_c(k,0)(K(0)\widetilde{Y}(0) + M_+(0)Z_c(1) - M_0(N)Z_c(N)) \\ +\bar{\Gamma}_c(k,N)(K(N)\widetilde{Y}(N) + M_-(N)Z_c(N-1) - M_-(0)Z_c(N)) \quad (372)$$

which allows the computation of $\widehat{\widetilde{X}}$ under cyclic boundary conditions. ∎

We point out that in (337), in the DBC case it is not really necessary to set the 'initial' condition to $\widetilde{X}_i$. One could set $Z(0) = 0$ even for DBC, provided eq. (331) is modified by adding the term $\bar{\Gamma}_d(k,0)\widetilde{X}_i$ in the left hand side.

By summing up, the μ-optimal polynomial estimate – for any μ positive integer – of a partially observed reciprocal chain x can be obtained by the following so called 'double sweep' algorithm: first, one builds up the vectors $Y(k)$ $k = 0, ..., N$, by calculating the Kronecker powers $y_{[i]}(k)$, for $i = 1, ..., \mu$, and uses them in the *backward* recursive equation (338) thus she obtains the $\gamma(k)$'s for $k = 0, ..., N$. Then, the *forward* recursive equation (337) can be used to compute $Z(k)$, and jointly eq. (331) (for DBC) or eq. (334) (for CBC), to compute the optimal estimate of $X(k)$, for $k = 0, ..., N$, from which, finally one extracts the desired estimate $\hat{x}(k)$.

It should be noticed that, even thought x takes values with probability one in a finite subset $\mathcal{S}$ of $\mathbb{R}^n$, the estimate $\hat{x}(k)$ is not a finite-state process, and may well take every value in $\mathbb{R}^n$ with a non-zero probability. Nevertheless, one can reasonably assume the actual estimate of $x(k)$ to be, for instance, the element of $\mathcal{S}$ closest to $\hat{x}(k)$. In the case of a reciprocal chain x taking values, with probability one, in a general finite alphabet of symbols, the algorithm still works and yields the μ-order polynomial-optimal estimate as before, provided the alphabet $\mathcal{S}$ satisfies the assumptions we described in §2.2.

6. Some remarks about the problem of dimensionality

It would be of practical interest to have some idea about the influence of growing model's dimension on the usability of possible algorithms one may build up while basing on that model. Of course, the model of reciprocal chain described by the representation theorem (Theorem 4.2), evolves in an augmented Euclidean space whose dimension is growing as either the dimension of the *original space* of values or the *number* of such values increase (i.e. n and L respectively in the symbology introduced in §2.2).

6.1. General thoughts

In this regard, what had to be first highlighted is that, generally speaking, the wider the a-priori krowledge is about the process the more one could expect a 'small' model's dimension. In our case, all we know about the process are: (i) it is a finite state process (ii) it is a reciprocal process. This means that we have only a 'weak' information about the probability distribution's process to exploit: as far as the *values* of the process, at a given k, are concerned we have no information at all (we know, of course, the value's probablities, but *any shape* could occur for their distribution); counterwise knowing that the process is reciprocal gives us a stronger restraint on the *whole* process distribution (i.e. the distribution over the whole interval T the index k belongs to). This means that, we can reasonably expect only to reduce dimensionality along the 'time' index k, whereas the lack of (probabilistic) information about the process values does not allows us to expect similar reductions in the dimension of state space at a fixed k. This is exactly what the nearest neighbour model of Theorem 4.2 does: such model indeed *strongly reduces* the dimensionality of the problem along the k-axis, whereas *the general setting* assumed for the distribution of model's values is necessarily ensued by a ricing in dimension of the augmented state $\mathcal{X}(k)$.

This point deserves to be better clarified, and in order to do so let us consider a case where it is apparent that a further information about the process values distribution does cause the model stopping in the rice of dimensionality. Suppose that the probability distribution of the chain x at k has some kind of regularity, for instance let it be such that the conditional expectation: $\mathbf{E}\{x(k)/x(k-1), x(k+1)\}$ is a *linear* function of its argument. This would mean that all the L^2 values of the couple $(x(k-1), x(k+1))$ are exactly interpolated by an hyperplane, that is a 1-degree polynomial. As a result, eq. (98) would have a degree equal to 1, so *there is no rice which occurs in the state-space model* in this example whatever the number L is. The reason is that we have a kind of *regularity* occurring to the distribution. This behavior is similar to what happening for continuous-valued Gaussian processes. In such cases, in spite of the *infinite* 'number of elements' of the process, one has no increase in the dimensionality of the model, so, in this case, one has the Levi-Frezza-Krener model of Gaussian reciprocal processes, which maintain the original state-space dimension. The reason of this comes from the assumption of a Gaussian distribution, which 'links'(in a probabilistic sense) the process values together, and bears a probabilistic 'information' which turns out in the *same* linear behavior of the conditional expectation we described above for a reciprocal *chain*. Also, different kind of distribution's regularity might occur than the linear one, it might be the case, for instance, of a 2-th degree interpolating polynomial and so forth, regardless of the number L of states; in all these cases the model defined by Theorem 4.2 still has not a rice in dimension, as the dimension of the augmented state X will be bounded from the above by the interpolating polynomial's degree, while the number L of chain's states might be considerably higher.

With that being said, one main concept can be summed up: while facing with a *fixed* reciprocal process distribution, one can try to exploit regularities within, if any; but while considering a reciprocal process *in general*, as it is the present case, i.e. provided *any* distribution for its values at any fixed k, and only accounting of the reciprocity property, then a rising in dimensionality is, in a sense, to be expected in the state-space. This is indeed the behavior of the nearest-neighbour model for reciprocal chains of Theorem 4.2, which reduces dimensionality along the k-axis while exploiting reciprocity, and increases the state-space dimension while accounting of generality in distribution. Nevertheless, it should be noted that, in any case, the rice in state-space dimension is finitely bounded by the number L of chain's states, and furthermore, while occuring some regularity-information about the distribution, the model is able to account of that and stops the ricing of dimensionality regardless to the number of states, as explained so far.

6.2. Sketch of a quantitative example: image processing

In order to add further insight, and to show how the nearest-neighbour model at issue can be useful from an application point of view, let us consider an *image-processing* setting. Image-processing has been the main motivation of the recent interest in reciprocal processes (some references about reciprocal processes in connection with image-processing has been given so far in the introduction of the present paper). As a matter of fact, while modelling a monochromatic *image* – i.e. a two dimensional distribution of gray-levels – Markov property fails, as mutual predictability between grey-levels might be improved by other pixels in the same direction. Now, suppose we have modelled an image as a finite set of *rows*, each row being identified with a reciprocal *scalar* process $x(k)$, $0 \le k \le N$. Moreover, suppose for the sake of simplicity to neglect correlation between rows and only to account of the statistical properties along the k-axis, which is done by reciprocity assumption. Furthermore, suppose the gray-level is recorded into a *byte*, which means we have 256 gray-levels, and x is a reciprocal *chain* with $L = 256$. In order to fix ideas suppose we want to solve a noise-reduction task over an image with a number of rows equal to N, and $N = 1000$ (an usual computer-resolution). In principle such issue might be solved by the linear estimation formula i.e.:

$$\widehat{X} = \mathbf{E}\{XY^T\}\mathbf{E}\{YY^T\}^{-1}Y,$$

where X, Y denotes vector-aggregates of original and noisy signal respectively, which have 10^6 elements each, but we realize that a prohibitive computational burden would occur with such setting, as it involves matrices with 10^{12} elements which should be multiplied and inverted as well.

Counterwise, suppose to adopt the model at issue. Then, looking at the structure of the state vectors (see for instance eqs. (3), (4)) one has for $\widetilde{X}(k)$ a dimension of 256, as x is scalar, thus $x_{[i]} = x^i \in \mathrm{I\!R}$, and the *maximal* degree μ is 255, as in this case $\mu = L - 1$ (an at most $(L-1)$-degree polynomial exacly interpolates L points). Similar calculations readily yields, for $\widetilde{\mathcal{X}}(k)$, a dimension of 32640, thus for the *whole* state vector, $(\widetilde{X}(k), \widetilde{\mathcal{X}}(k))$ we have a dimension of 32896. What has to be stressed now is that the two equations (98), (99) are *decoupled*, and only the self-governing eq. (98), describing the evolution of $\widetilde{X}(k)$, is actually involved in the estimation algorithms which have been developed in §5. Thus, the estimation task can be accomplished via a 'dynamical' model with a state space of dimension equal to 256 only, which is surely within modern computation device's reach. Also, note that, by reason of the 'decoupling' result Theorem 4.2 yields, one is no more concerned with the effect of mutual correlation between the nearest neighbours of $x(k)$ (i.e. $x(k-1)$ and $x(k+1)$) and the state space dimension is actually reduced from 32896 to 256, a 99% reduction.

Of course, in the image example above, model dimension gets growing fast as soon as each data set associated to a 'pixel' becomes more than one dimensional. However, it should be stressed that the hypothesis of a scalar 'field' is not unrealistic for an image-processing issue, as gray-levels are indeed scalar. Also, even for chromatic images, which are representable by *three* sets of gray-levels, one could, at least in a first approximation setting, suppose that the three gray-level configurations are mutually uncorrelated as the three basic colours occur independently of each other in a given image. Thus one is able to consider them separately, while reducing this way the issue to three scalar cases. In cases such that the basic process is not a scalar field, one has to resort to some other kind of simplifying hypothesis, or to some way of more efficient *coding* of the signal, as below argued.

6.3. Data quantization

Another important topic has to be stressed here, which comes from the fact that a reciprocal chain often occurs in practice as a model of a *quantized signal.* As we have seen, an image – as well as any kind of signal – is actually represented with a finite set of *bits* in modern devices. In other word, it is a *coded* signal, and in particular it ever undergoes to such coding usually known as *quantization.* Now, the so called *density* of the adopted quantization, is clearly relevant to the nearest-neighbour model at issue, as it determines the cardinality L of the alphabet the chain is taking its values into. To this purpose, we point out here the papers [9], [4], which, among those addressing the quantization issue, are the more pertaining to our present arguing.

Recall that, in the model at issue in the present paper, model dimension is upper bounded by a polynomial of L (for instance, in the image's example above given, the augmented state vector was the aggregation of the first $L-1$ powers of the original state), thus the overall dimensionality decreases as L decreases. In the image's example above, the cardinality of the alphabet was $L = 256$, which corresponds to a uniform gray-level quantization, but indeed other choices could be adopted, which, depending on the *entropy* of the signal, possibly are even more efficient.

In order to clarify this point, we mention that a lot of research work has been performed in past decades while addressing the issue of *minumum entropy quantization* (for fixed bound on distorsion). The reader is referred to [8] and to the wide list of references within the articles [9], [4] as more insight is needed about quantization and the definition of entropy in Information and Stochastic Processes Theory. Roughly speaking, the entropy-rate of a signal is a measure for its capability to be transmitted throughout a noiseless channel (as well as recorded in a memory device) *without distorsion.* In other words the signal's entropy measures the minimum amount of information which needs, besides the noise effects, to exactly reproduce it. It turns out that while accepting some *distortion* in the reproduction, a lower entropy could be attained.

With that being said, and while turning to the nearest neighbour model for reciprocal chains at issue in this paper, one might discover, as it often occurs, that the process can be partitioned in 'segments', where each segment needs a considerably smaller number of symbols than L, the cardinality of the alphabet, in order to reproduce it within a fixed distorsion. For instance, in the scalar image's example, one might need of a number of bits lower than the 256-bits per pixel therein assumed. Such method is known as *dynamic quantization*, and in this regard we point out again the article [4], where such issue has been addressed and some general results has been proven for the Markov case. Also, in the same paper a general setting is provided for dynamic quantization. To the author knowlegde, a generalization of the results in [4] to the case of reciprocal processes is up to now unavailable, and under this view the present paper aims to offer further motivation in addressing the dynamic quantization issue for reciprocal processes.

It turns out that, in the situation above outlined, the nearest neighbour model here at issue can be of course strongly reduced as to its state-space dimension. Also, no matter L is choosen large, when a low signal entropy occurs in some k-interval, the matrix parameters $\Phi(k), \Psi(k)$ in the model equation (98) *have necessarily zero-entries* in the columns corresponding to *powers in excess* of X (remind X has the structure defined in (3)). In fact, occurring that signal entropy is lower than $\log_2(L)$ (throughout some k-interval) is just the case that the conditional expectation $\mathbf{E}\{x(k)/x(k-1), x(k+1)\}$ can be exactly interpolated by a polynomial having a degree *lower* than L (as so far described in this section). Thus, assuming L being the maximum power degree in the state vector $X(k)$ yields a computational burden which is *only apparent* in the k-interval at issue, as the involved matrices will have actually non-zero values only in a lower dimensional top left block. Indeed the computational burden will be concetrated in such that portions of the k-axis where the signal entropy is higher (think, for instance, to a portion of an image having a small level of 'smoothness').

In conclusion, as to the dimension's growt of the reciprocal-chain's nearest-neighbour model at issue in this paper, it has been argued that (i) such large model's dimension essentially follows on being the model *a very general one*, i.e. being it a kind of *universalistic* model, which holds for any kind of chain's probability distribution; (ii) the problem of reducing dimension and/or computational burden have to be posponed until some new information becomes available about the chain's distribution. Just in cases such that a certain particular distribution has been given for the chain is indeed possible to start trying to exploit the possible 'nice' properties of that distribution (of course, this depending also on the particular task one has in mind while using a reciprocal-chain based model). (iii) Meaningful examples occur in practice, where the model has a tractable dimension (image-processing, where the 'chain' is embedded in a one-dimensional space of 'gray-levels'). (iv) (we highlight this point in particular) the problem of dimensionality and the *quantization* problem are closely connected. In fact, the 'alphabet' a chain takes values into is often determined, in practice, by a foregoing quantization process. It has been argued before that such issue is in turn connected with the *entropy* of the reciprocal chain. Such issue deserves to be better clarified from a quantitative point of view, but this goes beyond the purposes of the present paper, and we look forward to future papers in order to perform such research work.

7. Conclusions

We stress that the main result of the present paper, Theorem 4.2, allows us to conclude that for the interpolating polynomial of the C.E., given in eq. (40), all matrix-coefficients of the *mixed powers* are necessarily zero. In other words, for any reciprocal chain x, the polynomial function which exactly fits the C.E. values is actually the sum of two polynomials: i.e. a ν-degree polynomial for each one of the nearest neighbours $x(k+1)$, and $x(k-1)$). Theorem 4.2 gives the required representation result for reciprocal chains: a 'dynamic' model has been found, which consists, for any k in a polynomial equation involving just nearest neighbours values of x. This polynomial equation can be *exactly linearized*, that is it can be put in the form of a linear equation in the vectors $X(k)$ and $\mathcal{X}(k)$ defined in (73), (74). Moreover, these equations can be decoupled, which results in eqs. (98), (99), where the latter equation directly gives the vector $\mathcal{X}(k)$ by process X (X thus playing the role of an 'external' input).

Corollary 4.3 allows us to compute the matrix-coefficients of the model (98), (99), by a subset of 2μ order statistics, i.e. the covariances (204)–(207). Therefore, in this perspective, we can say that eqs. (98), (99) represent a solution of the stochastic realization problem for nonsingular reciprocal chains.

The results of §4 generalize Theorems 3.1 and 3.2 of §3. In particular, Theorem 4.1 is a more general result than 3.1, as well as Corollary 4.3 allows us to compute all system-coefficients including the projection matrices $F_{i,j}^{(h)}(k)$ given by Theorem 3.1, and the noise covariance $D(k,l)$ given by Theorem 3.2. Nevertheless, the results of §3 show us some additional relationships between matrix coefficients. Indeed we realize that the computation of the projection matrices associated to X and the correlation of process d can be carried out independently of the remaining matrix coefficients associated to vector $\mathcal{X}$ (collecting the mixed powers). In particular, the covariance of process b in eq. (99) can be performed of course by using eq. (212), which involves the matrix parameters $H(k), L(k)$ of eq. (99), nevertheless we see by (104) and (81)–(83) that $B(k)$ can be computed by using just the results of Theorem 3.2.

By Corollary 4.4 we get the 'normalized' model (218), (219), where in particular eq. (218) is formally similar (and owns similar properties) to the eq. (1) just proven (in [19]) for the case of a discrete-index Gaussian reciprocal process, but the powers aggregation X replacing the reciprocal process x. In this perspective, our model (218), (219) – besides the different nature of the involved processes, the former being a finite-states process the latter being a process which induces a probability density – represents an extension to a non-Gaussian case of the result in [19]. The insight difference between eq. (218) and eq. (1) consists in: provided suitable boundary conditions, a reciprocal Gaussian process x is the unique (within stochastic equivalence) solution of (1) ('well-posedness' of the equation), whereas a reciprocal chain x is not the unique solution (even within stochastic equivalence, and provided the same boundary conditions) of (218), but the particular case where x induces a family of conditional probability transitions that can be uniquely determined by a finite number of moments.

Even though we cannot say, in general, that eq. (218) is 'well-posed', nevertheless in Theorem (5.5) it is proven that under Dirichelet boundary conditions eq. (218) can be put in the *explicit* form (289)–(290), and with cyclic boundary conditions in the explicit form (292)–(293). The existence of such (linear) explicit representations, allows us to say that, besides x, any other solution of eq. (218), endowed with boundary conditions, has the same moments (234), up to 2μ order, of x. For this reason we have called the existence of such explicit representations the 'well-posedness in a wide-sense' property of the polynomial model (218).

Indeed, the wide-sense well-posedness, reveals itself to be a sufficient property to derive *suboptimal* smoothing algorithms. As a matter of fact, in Theorem (5.7) the μ-th degree polynomial-optimal smoother has been derived in the form of the backward/forward recursive equations (338)/(338), endowed with eqs. (331), (334) which adjust the estimate depending on either DBC or CBC are imposed. The algorithm allows the computation of the estimates in a 'double sweep', in a very similar way as in the (optimal) smoother proposed in [19] for the Gaussian case. It should be stressed that, even though suboptimal, the estimate given by the smoothing algorithm here proposed can be in principle calculated for any integer μ, thus reasonably 'approaching' the optimum by the well known approximation properties of polynomials [5]. However, the latter topic goes beyond our present purposes, since the smoothing algorithm here presented is just an example of application of the nonlinear nearest-neighbour model (98), (99).

As a concluding remark, we stress that all the results of the present paper can be applied to the *particular* case (nevertheless of great interest by itself) of a *Markov chain.* Moreover, as to possible further developments of the present research, we point out the connection of the dimensionality issue, as it involves the model presented in this paper, with the minimum-entropy-quantization problem for reciprocal chains. Such relationship has been here argued in §6, but such issue deserves a deepest investigation, which might be the subject of future research work.

References

[1] J. Abrahams and J. B. Thomas, "Some comments on conditionally Markov and reciprocal Gaussian processes," *IEEE Trans. Inform. Theory*, vol. 27, pp. 523–525, 1981.

[2] R. Bellman, *Introduction to matrix analysis.* New York: Mc-Graw Hill, 1970.

[3] S. Bernstein, "Sur le laisons entre les grandeurs aleatoires", in *Proceedings of Int. Cong. of Math., Zurich*, pp. 288–309, 1932.

[4] V. S. Borkar, S. K. Mitter, and S. Tatikonda, "Optimal sequential vector quantization of Markov sources". *SIAM J. Control Optim.*, vol. 40, pp. 135–148, 2001.

[5] F. Carravetta, A. Germani, and M. Raimondi, "Polynomial filtering for linear discrete-time non-Gaussian systems", *SIAM J. Control Optimization*, vol 34, pp. 1666–1690, 1996.

[6] F. Carravetta, A. Germani, and M. Raimondi, "Polynomial filtering of discrete time stochastic linear systems with multiplicative state noise", *IEEE Transactions on Automatic Control*, vol. 42, pp. 1106-1126, 1997.

[7] A. Chiuso, A. Ferrante, A., and G. Picci, “Reciprocal realization and modeling of textured images”, in *Proc. of the 44-th Conf. on Decision and Control, and European Contr. Conf., Seville, Spain*, December 12-15, 2005.

[8] T. Cover, and J. Thomas, *Elments of Information Theory.* New York: Wiley, 1991.

[9] N. Elia, and S. Mitter, “Stabilization of linear systems with limited information”, *IEEE Trans. Autom. Control*, vol. 46, pp. 384–400, 2001.

[10] X. Guyon, *Random Fields on a Network: Modeling, Statistics, and Applications.* New York: Springer Verlag, 1991.

[11] B. Jamison, B., “Reciprocal processes”, *Wahrscheinlichkeitstheorie verw. Gebiete*, vol. 30, pp. 65–86, 1974.

[12] A. J. Krener, “Reciprocal diffusions and stochastic differential equation of second order”, *Stochastics,* vol. 24, pp. 393–422, 1988.

[13] A. J. Krener, “Reciprocal diffusions in flat space”, *Probab. Theory and Related Fields,* vol. 107, pp. 243–281, 1997.

[14] T. I. Laakso, A. Tarczynski, N. P. Murphy, and V. Vlimki, “Polynomial filtering approach to reconstruction and noise reduction of nonuniformly sampled signals”, *Signal Processing,* vol. 80, pp. 567–575, 2000.

[15] B. C. Levy, “Noncausal estimation for discrete Gauss-Markov random fields” in *Proc. Int. Symp. Math. Theory Networks Syst. (MTNS-89), Birkhauser, Boston*, Dec. 1990.

[16] P. Levy, “A special problem of Brownian motion, and a general theory of Gaussian random functions”. in *Proc. 3rd Berkeley Symp. on Mathematical Statistics and Probability*, vol. 2, pp. 133–175, 1956.

[17] B. C. Levy, M. B. Adams, and A. S. Willsky, “Solution and linear estimation of 2-D nearest-neighbor models”, *Proc. IEEE,* vol. 78, pp. 627–641, 1990.

[18] B. C. Levy, and A. Ferrante, “Characterization of stationary discrete-time Gaussian reciprocal processes over a finite interval”, *SIAM J. Matrix Anal. Appl.,* vol. 24, pp. 334–355, 2002

[19] B. Levy, R. Frezza, and A. J. Krener, “Modelling and estimation of discrete-time Gaussian reciprocal processes”, *IEEE Trans. on Aut. Cont.*, vol. 35, pp. 1013–1022, 1990.

[20] B. C. Levy, R. Frezza, R., and A. J. Krener, “Gaussian reciprocal processes and self-adjoint stochastic differential equations of second order”, *Stochastics and stochastic reports,* vol. 34, pp. 29–56, 1991.

[21] B. Levy, and A. J. Krener, “Dynamics and kinematics of reciprocal diffusions”, *J. Math. Phys,* vol. 34, pp. 1846–1875, 1993.

[22] A. Lindquist, and G. Picci, “Realization theory for multivariate Stationary Gaussian processes”, *SIAM J. Control Optimization,* vol. 23, pp. 809–857, 1985.

[23] P. Masani, "The prediction theory of multivariate stochastic processes, III: unbounded spectral densities", *Acta Mathematica,* vol. 104, pp. 141–162, 1960.

[24] S. Nakamori, "Estimation technique using covariance information with uncertain observations in linear discrete-time systems", *Signal Processing,* vol. 58, pp. 309-317, 1997.

[25] S. Nakamori, R. Caballero-guila, A. Hermoso-Carazo, and J. Linares-Prez, "Second-order polynomial estimators from uncertain observations using covariance information", *Applied Mathematics and Computation,* vol. 143, pp. 319–338, 2003.

[26] M. Pavon, "The conjugate process in stochastic realization theory", *Math. Programming Study,* vol. 18, pp. 12–26, 1982.

[27] G. S. Rodgers, *Matrix derivatives*, Marcel Dekker (New York), 1980.

[28] J. A. Sand, "Reciprocal realizations on the circle", *SIAM J. Control Optimization*, vol. 32, pp. 507–520, 1996.

[29] E. Schrodinger, "Uber die Umkehrung der Naturgesteze", *Phys. Math.,* vol. 144, Sitz. Ber. der Preuss. Akad. Wissen., Berlin, 1931.

[30] E. Schrodinger, "Theorie relativiste de l'electron et l'interpretation de la mechanique quantique", *Ann. Inst. H. Poincare,* vol. 2, pp. 269–310, 1932.

[31] M. Sznaier, O. Camps O., and C. Mazzaro, "Finite horizon model reduction of a class of neutrally stable systems with application to texture synthesis and recognition". In: *Proc. of the 43-th Conf. on Decision and Control*, Paradise Island, Bahamas, December 14-17, 2004.

[32] M. Thieullen, "Reciprocal diffusions and symmetries of parabolic PDE: The nonflat case", *Potential Analysis,* vol. 16, pp. 1-28, 2002.

[33] J. M. Woods, "Two-dimensional discrete-Markovian fields", *IEEE Trans. Inform. Theory,* vol. 18, pp. 232–240, 1972.