Communications in Information and Systems
Volume 20, Number 2, 209–248, 2020

Stochastic models based on moment matching

Patrick Dewilde

In honor of Tom Kailath on his 85th birthday, with admiration and gratitude

The paper considers interpolating models for non-linear, non-Gauss stochastic variables and processes, given a well-ordered set of moments of increasing order. The proposed models use a characterization with *independent parameters*, much in the style of the Schur-Levinson parametrization for the linear, Gaussian case (a topic to which Tom Kailath made seminal contributions), but very different from it, given the different kind of structured matrices involved (Hankel-like instead of Toeplitz). The paper starts out with a review of the classical Hamburger-Akhiezer-Jacobi parametrization for one stochastic variable, using a (non-classical) dynamical system theory approach. Next, the paper generalizes these results to the multivariable case, and presents a detailed generalized Jacobi-like (independent) parametrization for two variables. Like in the Schur-Levinson case, such parametrizations succeed in characterizing models that interpolate the moment data (given the complexity of the issue, only the 2D case is treated in this paper, but using a method that generalizes to more variables).

Keywords and phrases: Nonlinear stochastic models, parametrization, Hankel and Jacobi matrices, dynamical system theory.

1. Introduction

Modeling stochastic processes has been a central piece of endeavor in signal processing, leading in particular to the famous Schur-Levinson model filters [13, 10, 6], in which a (zero means) stochastic process for which n specific covariance data items have been measured is modeled by an artificial linear filter of order n driven by (artificial) white noise. This order n stochastic modeling provides a 'best' possible model when the original process is known to be (zero mean) Gaussian, and where 'best' can be understood as reproducing (interpolating) the measured covariance data and being as unspecified as possible concerning not known covariances (i.e., being maximum

entropy for the unknown parameters, while satisfying the positivity of the covariance matrix). Connected to this type of model is a derived predictor or estimator, which, in the specific Gaussian case, is both the conditional expectation as well as the maximum likelihood estimate.

However, many processes are highly non-Gaussian so that higher degree moment information is needed if any modeling accuracy is to be obtained for them [18], and the modeling filter (also to be driven by an artificially constructed process) will hence have to be non-linear. Needless to say, the problem just stated has not been solved in all generality so far. In this paper, we look at the problem of matching a well-ordered set of higher order data (higher order moments or correlations), first of a single stochastic variable (the 1D problem) and then show how these results can be extended to multiple variables (as they occur in a stochastic process). The general strategy for this kind of modeling is to determine an adequate approximating joint pdf of the variables present, i.e., a pdf that is as simple as possible while matching the known data (in the Gaussian case, just data on the mean and the covariances determine the pdf and hence provide for 'sufficient statistics'). Once a joint pdf is settled on, then the best model filter can be taken as the conditional probability of the model variable with respect to the measured variables, and the estimation or prediction filter becomes a quantity derived from it (such as some average or maximum likelihood).

Motivation The author has heard claims (even in recent times) that the moment modeling problem can be solved using an extended version of the classical Schur interpolation problem. That is (perhaps unfortunately) but pertinently not true. The modeling problem based on covariances is essentially different from the one based on higher order moments, although both can be seen as solving an interpolation problem, each in its own ways. So some of the techniques used may be similar, but their algebraic content is very different (a positive real function on the complex plane is a very different object from a real positive function on the real line, like a pdf!). However, in both cases, the key to solving the modeling problem is parametrization. A *parametrization* of a problem is the determination of a set of independent parameters that characterize solutions uniquely—preferably all solutions, but 'good' characterizations are always useful. *Independent* parameters are parameters that can be chosen at will within a given collection of individually admissible values. In the case of the Schur-Levinson solution for the covariance problem, the parameters are *reflection coefficients* or *Schur parameters* and go back to Schur's seminal paper [15]. These reflection coefficients take

on values in the unit disc of the complex plane. Such a parametrization characterizes a positive definite Töplitz matrix. In the case of the (real) moment matching problem, the situation is very different. In the one variable case, a classical parametrization is available and known as *Jacobi parametrization* [1], which consists of two sets of values $\{a_i, b_i\}$ for integers $i \geq 0$ from $i = 0$ up to a maximum index (which may be infinite), whereby the a_i are arbitrary reals, and the b_i real > 0. The 1D Jacobi parametrization characterizes a positive definite Hankel matrix (in both cases a distinction has to be made between the non-singular case, for which there is an infinite number of parameters, and the (many possible) singular solutions, for which the number of parameters is restricted in a specific way).

The Schur parametrization can easily be extended from Toeplitz matrices to general positive definite matrices [5], and of course positive definite Toeplitz block Toeplitz (TBT) matrices. The case of the moment problem is more complex. We shall see that an nD moment problem characterizes a matrix that we shall call hierarchically Hankel, a notion that will be defined further on in the paper.

The classical scalar real moment problem of analytical function theory has been exhaustively treated by I.N. Akhiezer in [1], building on results from Hamburger [8]. There has been continuous interest in achieving extensions of the treatment in various directions, in particular to the complex case [4] and to the matrix case [7, 2, 14]. However, these extensions do not specifically address the stochastic moment problem (as will be made clear in this paper), but do provide for interesting methods and inspiration.

The paper starts out describing the general idea behind higher order stochastic modeling and a review of the classical single variable case, followed by an analysis of the multivariable stochastic modeling problem given a coherent set of correlation data. Our aim is system modeling, so we devote special attention to system theoretical properties. This is particularly relevant for the moment problem, because Hankel matrices play a central role in system theory. Classically, the parametrization problem is approached via the theory of orthogonal polynomials. The system theoretical, or matrix algebra approach presented in this paper is therefore substantially different from the classical Akhiezer approach based on orthogonal polynomials, but appears to serve our main aim, namely the extension of 1D properties to a multivariable, nD context. In order not to burden the exposition with too much notational details, the extension from one variable will be limited to two, using, however, a method for which a convincing case can be made that it extends to n variables, recursively on the number of variables (of course

with a considerable increase of complexity, but no more than what happens in numerical analysis, when one considers higher dimensional partial differential equations).

Notation

- $\delta(x-a)$ is the one-dimensional Dirac impulse of unit weight at the real point a; we shall also need two dimensional Dirac impulses, denoted as $\delta(x-a, y-b)$; shorthand notation: $\delta_k = \delta(k)$;
- MATLAB matrix notation is used informally, except when indicated otherwise; in particular, an accent is used to denote transposes (the paper uses real algebra throughout); when no confusion arrises with powers and to abbreviate the writing, the row index of a matrix is put as subscript and the column index as superscript: $H_{i,j} \leftarrow H_i^j$ (the notation A←B means 'replace A by B' or 'A becomes B');
- 'constructors' are used to build matrices: col[·] makes a column from the ordered list in its argument, Toe[·] and Han[·] construct Toeplitz and Hankel matrices from their respective ordered list arguments (we indicate an ordered list by putting it between square brackets [·]);
- We use some abbreviated notation for expectations: when X denotes a stochastic variable we write $\overline{X} = \mathbf{E}X$;
- We may simplify the MATLAB notation in some cases, or use the special structure of the matrix considered. E.g., the Hankel matrix $H_{0:k}^{0:k} \leftarrow H_{\mathbf{k}}$, considering that its first column is based on the moment data $\mu_{0:k}$, although the matrix depends on $\mu_{0:2k}$; a single ':' is also often suppressed: H_k indicates the k^{th} row of H and H^k the k^{th} column;
- The linear span of a collection of vectors x_k is indicated by $\bigvee[x_k]_{k=\dots}$. If X is a matrix, then $\bigvee X$ indicates the span of its columns.

2. The modeling principle

Given a multidimensional set of $n+1$ stochastic variables $Y_{0:n} = [Y_0 \cdots Y_n]$, then the best possible model in the pdf sense, is of course a set of (artificial) stochastic variables $X_{0:n}$ with the same overall distribution, and the best stochastic approximate (or model) to X_0, assuming the $[X_i]_{i=1:n}]$ known, is the *conditional estimate* $\widehat{X}_0 = X_0|_{[X_i]_{i=1:n}]}$. The conditional estimate is a stochastic variable in one dimension, with a specific pdf for each value of the n-tuple $X_{1:n}$.

The traditional modeling strategy, given a stationary stochastic process Y_i, is to assume that any Y_i is not directly dependent on the Y_ℓ for $\ell <$

$i-n$, i.e., $Y_i|_{Y_{i-n:i-1}} = Y_i|_{Y_{-\infty:i-1}}$, so that an n^{th} order model based on the n previous values $Y_{i-n:i-1}$ is adequate. In this case the model becomes a new stochastic process X_i which is such that the conditional estimate $\widehat{X_i}|_{X_{i-n:i-1}} = \widehat{X_i}|_{X_{-\infty:i-1}}$.

This means that the joint pdf of the relevant stochastic variables determines the model. Hence, the effort in the non-linear, non-Gaussian case will aim at determining joint pdf's that are consistent with the measured or known data. In the present paper we assume the known data to be fully ordered, both in sequence and in degree or power order (this means: we typically assume a few stochastic variables $\{X, Y, \cdots\}$ for which all the moment data $\mathbf{E}(X^a Y^b \cdots)$ is known up to a certain degree $n = a + b + \cdots$. Since the theory presented in this paper is a first approach to what turns out to be a very ambitious problem, we shall restrict ourselves to at most two variables (the 2D case), but use methods that can be extended further. Since only few data on the statistics are given or measurable in most cases, there will be a large collection of pdf's fitting the data, and one of the endeavors is to describe (i.e., parametrize if possible) all solutions, starting with the simplest(s) possible. It turns out that we can indeed find minimal complexity pdf's that fit the data, and that we can generate large collections of solutions, if not all, using interpolation of the given data as criterion (allowing other criteria on the parametrization to establish the choice of solution in the class of all possible solutions).

In the Gaussian stationary case, the conditional pdf is Gaussian as well, with mean $\mu_{\widehat{X}} = -\sum_{i=1:n} a_i X_i$, which is linearly dependent on the $X_{1:n}$ with Levinson coefficients $\{a_i\}$ [12], and variance found by solving the Wiener-Levinson equations. In the general (non-Gaussian) case the model will not depend linearly on known values, and the proposed method will have to produce a (hopefully as simple as possible) model which fits the desired or known moments and correlations. In case not a model is needed, but an estimation or prediction, some strategy is needed to select one representative value. This can be a mean, a median, maximum likelihood or any desirable average.

3. The 1-D case

In the one variable case, the Hamburger theorem [8] gives necessary and sufficient conditions for a moment series to belong to a pdf. Let $\{\mu_k\}_{k=0}^{\infty}$ ($\mu_0 = 1$) be the series of moments given, then there will exist a (generalized)

pdf $p(x)$ and a stochastic variable X that generates these moments as $\mu_k = \mathbf{E}X^k := \int_{-\infty}^{\infty} x^k p(x)dx$ iff, for all indices $k \geq 1$ the Hankel matrices

$$H_{\mathbf{k}} = \begin{bmatrix} 1 & \mu_1 & \mu_2 & \cdots & \mu_k \\ \mu_1 & \mu_2 & \iddots & \iddots & \mu_{k+1} \\ \mu_2 & \iddots & \iddots & \iddots & \mu_{k+2} \\ \vdots & \iddots & \iddots & \iddots & \vdots \\ \mu_k & \mu_{k+1} & \mu_{k+2} & \cdots & \mu_{2k} \end{bmatrix} \tag{1}$$
$$= \mathbf{E} \begin{bmatrix} 1 \\ X \\ X^2 \\ \vdots \\ X^k \end{bmatrix} \begin{bmatrix} 1 & X & X^2 & \cdots & X^k \end{bmatrix}$$

are positive (semi-)definite [1]. The pdf is the generalized derivative of a non-decreasing and left-continuous function $\sigma(x)$ such that $\mu_k = \int_{-\infty}^{\infty} x^k d\sigma(x)$ (Stieltjes integral). The singularities in $p(x) = d\sigma(x)/dx$ can only be positive Dirac impulses, derived from the jumps in $\sigma(x)$.[1] We shall call a distribution *atomic* of order n if $d\sigma(x)/dx$ consists of just a finite number n of Dirac impulses and is zero otherwise.

The modeling issue we consider here is in a sense simpler than the modeling of an infinite series (the problem most mathematics in this area considers). We are interested mostly in constructing finite models that fit the given data, hence in determining parameters that describe such fitting models adequately.

Proposition 1. *Suppose the series of (known) moments is given as* $\{\mu_k\}$ *for* $k = 0 : 2(n+1)$ *and some* n *with, as is necessary,* $H_{\mathbf{n+1}} \geq 0$ *and, in addition,* $H_{\mathbf{n+1}}$ *singular of rank* $\eta \leq n+1$. *Then the series has a unique extension to a moment series. This unique extension pertains to an atomic pdf of order* η. *No atomic pdf of lower order exists that generates the given partial moment series.*

Proof. Let us first assume $H_{\mathbf{n}}$ non-singular, and prove the theorem for that case. From classical dynamical system theory, and because the (singular) Hankel matrix $H_{\mathbf{n+1}}$ is symmetrical, we know that a transfer function $T(z) = \sum_{k=1}^{\infty} \mu_{k-1} z^k$ built on the given $\{\mu_i\}$ can be realized by a system of degree

[1] This is a somewhat imprecise statement, as limit points of singular atoms are also singular points, but we shall not encounter such singularities in the present theory. For more precise information on this, see a textbook on measure theory.

$n+1$ with state space representation $\{A, b, b'\}$ as $T(z) = zb'(1-zA)^{-1}b$, in which $A = A'$ is of dimension $n+1$ and symmetrical. For the sake of later reference, here are the arguments:
1. The Hankel property assures that for all n, $[H_{\mathbf{n+1}}]_{1:(n+1)}^{0:n} = [H_{\mathbf{n+1}}]_{0:n}^{1:(n+1)}$.
2. With $H_{\mathbf{n}}$ non-singular and symmetrical, we dispose of two factorizations: $H_{\mathbf{n}} = L_n L_n'$ and $H_{\mathbf{n+1}} = L_{\mathbf{n+1}} L_{\mathbf{n+1}}'$, with $L_{\mathbf{n}}$ square non-singular and $L_{\mathbf{n+1}} = \begin{bmatrix} L_{\mathbf{n}} \\ r_{n+1} \end{bmatrix}$ for some vector r_{n+1} (because $H_{\mathbf{n+1}}$ is assumed singular).
(Note: we do not have to assume here that $L_{\mathbf{n}}$ is lower triangular (i.e., Cholesky), any non-singular factor will do.)
3. Let $L_{\mathbf{n+1}}^{\uparrow} = [L_{\mathbf{n+1}}]_{1:(n+1)}^{0:n}$ be the 'restricted upward shift' of $L_{\mathbf{n+1}}$ and similarly for the restricted left shift, then we have (by cancellation of an appropriate row and column)

$$H_{\mathbf{n+1}}^{\uparrow} = [H_{\mathbf{n+1}}]_{1:(n+1)}^{0:n} = L_{\mathbf{n+1}}^{\uparrow} L_{\mathbf{n}}' = [H_{\mathbf{n+1}}]_{0:n}^{1:(n+1)} = L_{\mathbf{n}} (L_{\mathbf{n+1}}^{\uparrow})' = H_{\mathbf{n+1}}^{\leftarrow}. \tag{2}$$

Now define $b := [L_{\mathbf{n}}']_{0:n}^{0}$ (first column of $L_{\mathbf{n}}'$) and $A = L_{\mathbf{n}}^{-1} L_{\mathbf{n+1}}^{\uparrow} = (L_{\mathbf{n+1}}^{\uparrow})' L_{\mathbf{n}}^{-\prime}$, then it follows, by $AL_{\mathbf{n}}' = (L_{\mathbf{n+1}}^{\leftarrow})'$, that $L_{\mathbf{n+1}}' = \begin{bmatrix} b & Ab & \cdots & A^n b \end{bmatrix}$. A is symmetrical by definition (eq. (2)), and for all $0 \le k \le 2n+2$ we find $\mu_k = b' A^k b$.
4. The construction given is unique up to a unitary transformation, given the moments $\mu_k : 0 \le k \le 2n+1$, that is, up to an $n \times n$ unitary matrix U such that an alternative 'realization' with the same interpolation properties is given by $\widehat{b} = Ub$, $\widehat{A} = UAU^{-1}$ (notice: $\mu_{2(n+1)}$ is fixed by the singularity condition on $H_{\mathbf{n+1}}$). It also follows that all subsequent moments μ_k for $k \ge 2(n+1)$ are determined by $\mu_k = b' A^k b$ (this requires a bit of proof).

Since A is symmetrical, there exists a unitary transformation making A diagonal; in fact, we can just assume that A is diagonal to begin with: $A = \mathrm{diag}(\alpha_i)$, $i = 0 \cdots n$. Let the corresponding $b' = \begin{bmatrix} \beta_0 & \cdots & \beta_n \end{bmatrix}$, then it results that all $\mu_k = b' A^k b = \sum_{i=0}^{n} \beta_i^2 \alpha_i^k$, which is the moment series of the atomic pdf $p(x) = \sum_{i=0}^{n} \beta_i^2 \delta(x - \alpha_i)$. The $\{\alpha_i\}$, the corresponding weights $\{\beta_i^2\}$ and the resulting atomic pdf is hence fully determined by the moments up to and including μ_{2n+1}.

In the case also $H_{\mathbf{n}}$ is singular, the construction can be done with a lower order model (say of degree $\eta < n$), and will also be essentially unique, given the moments up to and including $2\eta + 1$, for the same reasons. □

Remark One may wonder whether the rank of the Hankel matrix can go up again, after having been stationary for some $k \ge \eta$. That is indeed possible

in general realization theory, but not in the present case where the Hankel matrix is required to be positive definite. The next proposition shows this.

Proposition 2. *Suppose a power series* $\{\mu_k\}$, $k = 1, \cdots, 2n$ *is given with* $H_{\mathbf{n}}$ *non-singular. Then there is a one-parameter collection of corresponding atomic pdf's of degree* $n+1$ *parametrized by the single moment* μ_{2n+1}*, which may be chosen arbitrarily.*

Proof. From the construction in the previous proposition, we see that μ_{2n+1} can be chosen arbitrarily, but to make $H_{\mathbf{n+1}}$ singular, $\mu_{2(n+1)}$ is then fixed. More precisely,

$$H_{\mathbf{n+1}} = \begin{bmatrix} H_{\mathbf{n}} & \mu_{n+1:2n+1} \\ (\mu_{n+1:2n+1})' & \mu_{2(n+1)} \end{bmatrix} \tag{3}$$

has to be singular, hence $\mu_{2(n+1)} = [\mu_{n+1:2n+1}]' H_n^{-1} \mu_{n+1:2n+1}$. By the construction of the previous theorem, there are degree n realizations $\{A, b, b'\}$ generating the moments up to $\mu_{2(n+1)}$ by the rule $\mu_k = b' A^k b$, $0 \leq k \leq 2n$. The series can be augmented further step by step, keeping the degree equal to n: first the already known vector $\mu_{n+2:2n+2}$ is linearly dependent on the columns of $H_{\mathbf{n}}$, determining μ_{2n+3}, then the linear dependence of $\mu_{n+3:2n+3}$ determines μ_{2n+4} etc. recursively. In fact: all these quantities equal $b' A^k b$ for values of $k > 2n$. □

There is an important stochastic interpretation to the properties presented so far, which follows from the following general principle. Let a covariance matrix $C := \overline{UU'}$ with $U := U_{1:n}$ an n-dimensional stochastic vector, and suppose that C is singular, with $a'C = 0$ for some (non-stochastic, i.e., regular) vector a. Then the stochastic variable $a'U \equiv 0$, because $\overline{(a'U)^2} = a'Ca = 0$ (strictly speaking, $a'U$ is almost everywhere zero on its stochastic σ-algebra, which we do not need to specify further in this context). Applied to the singular $H_{\mathbf{n+1}}$ of the previous theorem, there exists a single constant vector a (of dimension $n+2$) such that $a' H_{\mathbf{n+1}} = 0$, and hence $a'[X^k]_{k=0:n+1} = 0$ with $a_{n+1} = 1$ — i.e., X^{n+1} depends linearly on $(X^k)_{k=0:n}$. This has profound consequences for all sub-Hankels based on the X^k. For example, for any positive integers i and k, $H_{i:i+n+1}^{k:k+j} = \overline{[X_k]_{k=0:n+1} X^{i+k} [X_k]'_{k=0:j}}$ and hence singular as well with $a' H_{i:i+n+1}^{k:k+j} = 0$.

Generating function Assuming the existence of moments up to the order needed (i.e., integrability with the pdf), let us define[2] the *generating function*

[2]This definition may differ from definitions in the literature, relating to a different context!

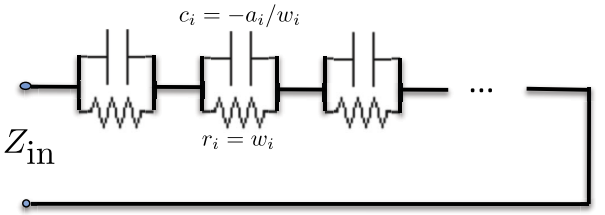

Figure 1: Realization of the generating function as the input impedance of a Foster circuit.

$G_\sigma(z) = \int_{-\infty}^{\infty} \frac{1}{1-zx} d\sigma(x) = \sum_{i=0}^{\infty} \mu_i z^i$ (this being merely a unilateral series in z, which may not converge anywhere in the complex plane[3] or where terms may become infinite from a certain degree on). The definition is motivated by the fact that atomic pdf's with a finite number of atoms possess moments of all degree—hence there will always be infinite continuations of a given series. The generating function of an order n atomic realization can be written (using the definition of α_i and β_i given earlier)

$$zG(z) = \sum_{i=0}^{n} \beta_i^2 z(1 - z\alpha_i)^{-1}, \tag{4}$$

with $\sum_{i=0}^{n} \beta_i^2 = 1$ and all α_i real. Hence $G(z) = \sum_{i=0}^{n} \beta_i^2 (1 + z(-\alpha_i))^{-1}$. In electrical circuit theory, one may interpret this expression as an impedance, and realize it as a (Foster) series of shunt RC-circuits with $R = \beta_i^2$ and $C = -\alpha_i/\beta_i^2$ shown in Fig. 1. Notice that the circuit contains both positive and negative capacitances, testimony to the double-sidedness of the corresponding atomic pdf $\sum_{i=0}^{n} \beta_i^2 \delta(x - \alpha_i)$. Also, the residues corresponding to negative real frequencies are positive, while those corresponding to positive ('unstable') frequencies are negative.

Following the terminology in Willems [17] we may call a system with transfer function of the type given by $G(z)$, namely a rational transfer function with corresponding positive definite Hankel matrix, of *generalized relaxation type* or a *GR system* (Willems only considers systems with poles in the left half z-plane and calls them *relaxation systems*). In the present case there is no pole at infinity (which is possible in the classical relaxation

[3]Which is the case for Gaussian or normal distributions. E.g., a zero-mean Gaussian distribution of variance σ^2 has moments $\mathbf{E}(X^n) = (n-1)!!\sigma^n$ with $(n-1)!! = 1.3.\cdots.(n-1)$ for n even and otherwise zero, so that the radius of convergence of its generating function is zero.

case), and, moreover, the zero component $g_0 = \sum \beta_i^2 = 1$. We call this case accordingly of *normalized general relaxation type* or NGR.

Jacobi realization

Proposition 3. *A factorization of the Hamburger matrix $H_{\mathbf{n}} = L_{\mathbf{n}} L'_{\mathbf{n}}$ with $L_{\mathbf{n}}$ a non-singular lower staircase of rank $m+1 \leq n+1$ produces two important objects:*

1. *A basis $\mathbf{P}_m := L_{\mathbf{m}}^{-1} \mathbf{X}_m$, orthonormal with respect to the Gram matrix $H_{\mathbf{m}}$ for $m \leq n$, and*
2. *A special 'realization' $\{J_{\mathbf{m}}, e_0, e_0'\}$ with $J_{\mathbf{m}}$ a characteristic Jacobi matrix*

$$J_{\mathbf{m}} := \begin{bmatrix} a_0 & b_0 & & & \\ b_0 & a_1 & b_1 & & \\ & \ddots & \ddots & \ddots & \\ & & b_{m-2} & a_{m-1} & b_{m-1} \\ & & & b_{m-1} & a_m \end{bmatrix} \tag{5}$$

and e_0 the first natural unit vector ($e_0' = \begin{bmatrix} 1 & 0 & \cdots \end{bmatrix}$), by which is meant that $\mu_k = e_0' J^k e_0$ for all $k = 0{:}\,\infty$.

Proof. First, if $H_{\mathbf{n}}$ is non-singular, then $L_{\mathbf{n}}$ will be square non-singular and lower triangular. Otherwise there will be an index $m < n$ such that all $H_{\mathbf{k}}$ for $m < k \leq n$ are singular, and $L_{\mathbf{n}}$ necessarily has a staircase form with a non-singular upper triangular part. In any case there will be $L_{\mathbf{m}}$ with $m \leq n$ such that

$$H_{\mathbf{m}} = L_{\mathbf{m}} L'_{\mathbf{m}} \tag{6}$$

and $L_{\mathbf{n}} = \begin{bmatrix} L_{\mathbf{m}} \\ M_{m+1:n} \end{bmatrix}$ where M completes the staircase (and disappears when $m = n$). Let us first observe that the elements of $L_{\mathbf{m}}^{-1} \mathbf{X}_m$ form an orthonormal basis for $\mathbf{X}_m = [X^k]_{k=0:m}$, which we denote with $\mathbf{P}_m := L_{\mathbf{m}}^{-1} \mathbf{X}_m$. Indeed, $\overline{\mathbf{P}_m \mathbf{P}'_m} = L_{\mathbf{m}}^{-1} \overline{\mathbf{X}_m \mathbf{X}'_m} L_{\mathbf{m}}^{-\prime} = I$. Secondly, if $m < n$, then by our previous theorems, there exist degree $m+1$ symmetric realizations $\{A, b, b'\}$ such that $[H_{\mathbf{n}}]_i^j = b_i A^{i+j} b_j$. Any such choice leads to a ('square-root') factorization of $H_{\mathbf{n}} = \mathbf{O}_n \mathbf{O}_n'$ with the 'observability' matrix $\mathbf{O}_n = [b' A^k]_{k=0:n}$ in which A has dimension $(m+1) \times (m+1)$. Conversely, every such realization can be obtained from a (minimal) square root factorization (see classical realization theory, e.g., [11]). In the case of a staircase factorization, this leads to $b = e_0$ (since anyway $[L'_{\mathbf{m}}]_{0:m}^0 = e_0$) and for $'A'$ the usual

and earlier derived formula $A = L_{\mathbf{m}}^{-1}[L_{\mathbf{m+1}}]_{1:m+1}^{0:m} = L_{\mathbf{m}}L_{\mathbf{m+1}}^{\uparrow}$, resulting in a singular extension of $L_{\mathbf{m}}$ beyond m (i.e., a set of solutions parametrized by the choice of μ_{m+1} — see Proposition 2 — let's call it $J_{\mathbf{m}}$ at this point of the proof and show that it is a Jacobi matrix indeed. Notice that $J_{\mathbf{m}}$ will automatically be symmetric, but it is also tri-diagonal, which can be seen as follows. Let

$$\mathbf{P}_m(X) = \begin{bmatrix} P_0(X) \\ P_1(X) \\ \vdots \\ P_m(X) \end{bmatrix} := L_{\mathbf{m}}^{-1}\mathbf{X}_m \tag{7}$$

then $\overline{\mathbf{P}_m X \mathbf{P}_m'} = L_{\mathbf{m}}^{-1}\overline{\mathbf{X}_m X \mathbf{X}_m'}L_{\mathbf{m}}^{-\prime} = J_{\mathbf{m}}$, because $\overline{\mathbf{X}X\mathbf{X}'} = H^{\leftarrow}$, hence $\overline{\mathbf{X}_m X \mathbf{X}_m'} = [e_0' J_{\mathbf{m}}^{i+j} e_0]_{0:m}^{1:(m+1)} = L_{\mathbf{m}}J_{\mathbf{m}}L_{\mathbf{m}}'$ (consequence of the realization theory). But we also have $\overline{P_k(X)X_j} = 0$ as soon as $j \leq k-1$ because $\bigvee \mathbf{P}_k = \bigvee \mathbf{X}_k$ for $k \leq m$, and hence also $\overline{P_k X P_j} = [J_{\mathbf{m}}]_k^j = 0$ for $j < k-2$. From this it follows that $J_{\mathbf{m}}$ is symmetric tri-diagonal, hence a Jacobi matrix. Reconstituting $L_{\mathbf{m}} = [e_0'[J_{\mathbf{m}}]^k]_{k=0:m}$ one finds that the diagonal elements are $[L_{\mathbf{m}}]_{k,k} = 1 \cdot b_0 \cdots b_{k-1}$, which may be chosen positive together (there is an ambiguity in the choice of the sign of diagonal elements in a Cholesky factorization, which is removed by choosing them all positive). Conversely, given any $(m+1)$-dimensional Jacobi matrix $J_{\mathbf{m}}$ with non-zero b_k, $L_{\mathbf{m}} := [e_0'(J_{\mathbf{m}})^k]_{k=0:m}$ will be a lower, non-singular triangular matrix such that $L_{\mathbf{m}}L_{\mathbf{m}}'$ is Hankel — an obvious property from the form of $L_{\mathbf{m}}$ in terms of e_0' and $J_{\mathbf{m}}$. □

Remarks

- the non-zero off-diagonal elements $[b]_{0:m-1}$ may all be chosen strictly positive without loss of generality (this choice makes the construction unique or 'canonical');
- once $J_{\mathbf{m}}$ is known, then the complete $\mathbf{X}$ up to infinity is known as well (this is because $J_{\mathbf{m}}$ fixes $L_{m+1}^{0:m}$ and singularity then forces $H_{m+1}^{m+1} = L_{m+1}^{0:m}[L_{m+1}^{0:m}]'$ — a quantity that we call the *Schur residue* S_{m+1}—in this singular case the Schur complement is zero and hence $S_{m+1} = e_0'[J_{\mathbf{m}}]^{2m+2}e_0$).

Next I claim: the *Jacobi realization* $G(z) = e_0'(I - zJ_{\mathbf{m}})^{-1}e_0 (= 1 + e_0'(I - zJ_{\mathbf{m}})^{-1}zJe_0)$, for which $\mu_k = e_0'J_{\mathbf{m}}^k e_0$ becomes a cascade of sections, with the k^{th} section shown in Fig. 2. The inputs and the state of section k are vectors $U_k := u_k^i$, $Y_{k+1} := y_{k+1}^i$, $X_k := x_k^i$ where $i = 0, 1, \cdots$ is the time

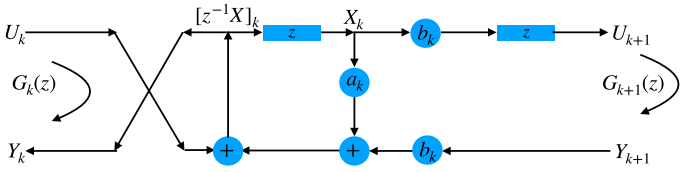

Figure 2: A Jacobi section.

index and $x_k^0 = 0$ for all sections; similarly for the outputs. This section stands for the detailed equations

$$\left\{ \begin{array}{rcl} x_k^{i+1} & = & a_k(x_k^i + u_k^i) + b_k y_{k+1}^i \\ u_{k+1}^i & = & b_k(x_k^{i-1} + u_k^{i-1}) \\ y_k^i & = & x_k^i + u_k^i \end{array} \right. \tag{8}$$

Derivation of the Jacobi section

Proposition 4. *Let $\{a, b\}$ be the first Jacobi coefficients of the normalized generating function $G(z)$ of relaxation type with $b \neq 0$, then*

$$G(z) = \left[1 - z(a + b^2 z G_L(z))\right]^{-1} \tag{9}$$

for an arbitrary normalized generating function $G_L(z)$ of relaxation type.

Proof. Let $G_L(z)$ be a normalized generating function of relaxation type, with Jacobi matrix J_L. We have $G_L(z) = e_0'(I - zJ_L)^{-1}e_0$. Let J be the Jacobi matrix for $G(z) = e_0'(I - zJ)^{-1}e_0'$, with $J = \begin{bmatrix} a & be_0' \\ e_0 b & J_L \end{bmatrix}$, then

$$e_0'(I - zJ)^{-1}e_0 = \left(\begin{bmatrix} 1 - za & -zbe_0' \\ -e_0 zb & I - zJ_L \end{bmatrix}^{-1} \right)_{0,0} \tag{10}$$

Using formal LU-decomposition on the right hand side, and abbreviating $\beta := e_0 z b$ and $\Gamma := I - Jz - \beta(1 - za)^{-1}\beta'$, we find

$$e_0'(I - zJ)^{-1}e_0 = (1 - za)^{-1} + (1 - za)^{-1}\beta'\Gamma^{-1}\beta(1 - za)^{-1} \tag{11}$$

Now using a standard argument (also applicable to our formal calculus of one-sided series),

$$\begin{array}{rl} \beta'\Gamma^{-1}\beta & = \beta' \left[I - (I - J_L)^{-1}\beta(1 - za)^{-1}\beta'\right]^{-1} (I - zJ_L)^{-1}\beta \\ & = \beta'(I - zJ_L)^{-1}\beta \left[1 - (1 - za)^{-1}\beta'(I - zJ_L)^{-1}\beta\right]^{-1} \end{array} \tag{12}$$

Now, $\beta'(I - zJ_L)^{-1}\beta = b^2z^2G_L(z)$. Working out further we find

$$G(z) = \frac{1}{1-za}\left(1 + \frac{b^2z^2G_L(z)}{1 - za - z^2b^2G_L(z)}\right) \tag{13}$$

and the claimed result. □

From this a realization for $G(z)$ in terms of $G_L(z)$ follows directly, using auto-regression:

$$Y = G(z)U \to Y = U + (a + zb^2G_L(z))zY \tag{14}$$

This means: after having produced some Y_k, one puts it in memory and uses it as input for the next stage. A diagram representing this is what Fig. 2 shows.

A finite realization of order n is obtained by putting $b_n = 0$ (and no further coefficients.) However, notice that in that case the impulse response remains infinite, as the system keeps on 'ringing' (for example, if $b_0 = 0$ and $u^0 = \mathrm{col}[1, 0, \cdots]$ then $y^0 = \mathrm{col}[1, a_0, a_0^2, a_0^3, \cdots]$, corresponding to $G(z) = 1/(1 - za_0)$.)

From Cholesky to Jacobi and vice-versa

The Jacobi matrix can easily be computed recursively from a Cholesky factorization of the respective augmented Hankel. Assuming temporarily that $H_{\mathbf{m+1}}$ is non-singular, we have, from $L_{\mathbf{m}}\mathbf{P}_m = \mathbf{X}_m$ that $[H_{\mathbf{m+1}}]^{1:m+1} = \mathbf{X}_m X \mathbf{X}_m' = L_{\mathbf{m}}J_{\mathbf{m}}L_{\mathbf{m}}' = L_{\mathbf{m}}[L'^{\leftarrow}]_{\mathbf{m}}$, and hence

$$J_{\mathbf{m}}L_{\mathbf{m}}' = \begin{bmatrix} L_{1,0} & L_{2,0} & \cdots & L_{m,0} & L_{m+1,0} \\ L_{1,1} & L_{2,1} & \cdots & L_{m,1} & L_{m+1,1} \\ 0 & L_{2,2} & \cdots & & \\ \vdots & \ddots & \ddots & & \\ 0 & \cdots & \cdots & L_{m,m} & L_{m+1,m} \end{bmatrix} \tag{15}$$

The left-hand side induces a recursion on the sub-diagonal and diagonal entries:

$$\begin{cases} [L_{\mathbf{m}}]_{i,i} &= b_0 \cdot b_1 \cdots b_{i-1} \\ [L_{\mathbf{m}}]_{i,i-1} &= (a_0 + a_1 + \cdots + a_{i-1})[L_{\mathbf{m}}]_{i-1,i-1} \end{cases} \tag{16}$$

and

$$\begin{cases} b_{i-1} &= [L_{\mathbf{m}}]_{i,i}/[L_{\mathbf{m}}]_{i-1,i-1} \\ \sum_{k=0:i-1} a_k &= [L_{\mathbf{m}}]_{i,i-1}/[L_{\mathbf{m}}]_{i-1,i-1} \end{cases} \tag{17}$$

from which also a_{i-1} can be deduced recursively. Notice that both $L_{\mathbf{m}}$ and $J_{\mathbf{m}}$ are determined by the $[L_{\mathbf{m}}]_i^i$ and $[L_{\mathbf{m}}]_i^{i-1}$, for $i = 1{:}(m+1)$, but there is some indeterminacy at the end: supposing $L_{\mathbf{m}}$ non-singular and $L_{\mathbf{m+1}}$ singular, we have a_m free (arbitrary) and $b_m = 0$, because g_{2m+1} can be chosen arbitrarily, actually

$$a_m = \frac{1}{[L_{\mathbf{m}}]_{m,m}}([M_{m+1}]_m - [L_{\mathbf{m}}]_{m,m-1}b_{m-1}) \tag{18}$$

since in this case $M_{m+1} = e_0'[J_{\mathbf{m}}]^{m+1} = [L_{\mathbf{m}}]_{m,0:m}J_{\mathbf{m}}$ and also $[M_{m+1}]_m$ is free together with g_{2m+1} in the original Hankel.

An important property concerning the relation between the orthonormal polynomials and the Jacobi matrices is the following.[4]

Proposition 5. *Let $P_{n+1}(z)$ be the orthonormal polynomial of degree $n+1$, assuming H_{n+1} non-singular and $J_{\mathbf{n}}$ the Jacobi matrix of dimension $(n+1)^{\times 2}$, then the eigenvalues of $J_{\mathbf{n}}$ are the zeros of the polynomial $P_{n+1}(z)$ with eigenvector $\mathbf{P}_n(\alpha)$.*

Proof. The Jacobi recursion up to order n with non-zero b_n (necessary to have $H_{\mathbf{n+1}}$ non-singular) gives

$$z\mathbf{P}_n(z) = J_{\mathbf{n}}\mathbf{P}_n(z) + e_n P_{n+1}(z)b_n \tag{19}$$

Putting $z = \alpha$ gives the result. □

Notice that $P_{n+1}(z) = b_n^{-1}(zP_n(z) - b_{n-1}P_{(n-1)}(z) - a_nP_n(z))$, so that b_n appears to be a normalizing factor on $P_{n+1}(z)$. In the finite index case of order n, we have that $\mathbf{P}_{\mathbf{n}}(z)$ depends only on the Jacobi parameters $\{a_{0:(n-1)}, b_{0:(n-1)}\}$, and P_{n+1} can be chosen monic and is then the characteristic polynomial of $\lambda I - J_{\mathbf{n}}$.

Extension theory

The Jacobi parametrization allows for a universal extension strategy on a given set of moments to be interpolated. Let us consider two 'Jacobi systems': $\{J_{\mathbf{m}}^{(1)}, \delta_{0:m}^0\}$ and $\{J_{\mathbf{n}}^{(2)}, \delta_{0:n}^0\}$ (with $\delta_k^j := \delta(k-j)$ the discrete Dirac function), then we can extend the (1) system with the (2) system just by

[4] I am grateful to dr. Sankar Basu for pointing out this fact and proof!

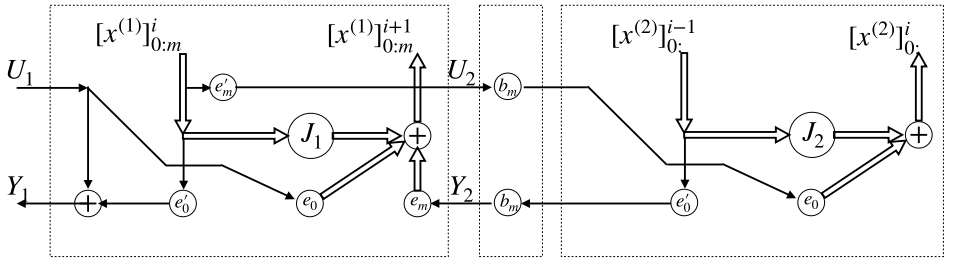

Figure 3: Extension of a Jacobi realization with another one (using the notation $e_0 = \delta_{0:} = \mathrm{col}[1, 0, \cdots]$, $e_m = \delta_{-m:0} = \mathrm{col}[0, \cdots, 0, 1]$).

introducing a free 'connecting' b_m and defining the total as the Jacobi system

$$\left\{ J^{(t)}_{m+n+2}, \delta^0_{0:(m+n+1)} \right\} := \left\{ \left[\begin{array}{c|c} J^{(1)}_{\mathbf{m}} & \delta^0_{-m:0} b_m \delta^{0:n}_0 \\ \hline \delta^0_{0:n} b_m \delta^{-m:0}_0 & J^{(2)}_{\mathbf{n}} \end{array} \right], \delta^0_{0:(m+n+1)} \right\}. \tag{20}$$

Splitting the state vector $x = \mathrm{col}[x_{0:m}, x_{(m+1):}]$ produces the general cascade shown in Fig. 3.

Fig. 3 exhibits a general cascade connection, with a connecting system having the character of a generalized kind of 'immitance'. In order not to confuse notation, and stay in line with other types of cascade structures, let us denote the transfers in the first section as mapping (see the figure)

$$\left[\begin{array}{c} U_2 \\ Y_1 \end{array} \right] = S \left[\begin{array}{c} U_1 \\ Y_2 \end{array} \right] \tag{21}$$

then we find (chasing the diagram, and replacing J_1 by J)

$$S = \left[\begin{array}{c} e'_m \\ e'_0 \end{array} \right] (I - zJ)^{-1} \left[\begin{array}{cc} e_0 & e_m \end{array} \right] \tag{22}$$

loaded into $b_m^2 z G_L(z)$, where G_L represents a normalized general relaxation function. We find

Proposition 6. *Given a sequence of* $2m$ *moments* $\{\mu_{0:2m}\}$*, then any interpolating moment series is given by the generating function*

$$G(z) = e'_0 \left[I - z(J + e_m z b_m^2 G_L(z) e'_m) \right]^{-1} e_0, \tag{23}$$

in which $\{e_0', J, e_0\}$ *is the normalized order* m *interpolant with Jacobi matrix* J*, and* $G_L(z)$ *is an arbitrary normalized general relaxation function.*

Proof. (Abbreviating somewhat) we have, as usual

$$G = S_{2,1} + S_{2,2}(I - zb_m^2 G_L S_{1,2})^{-1} z b_m^2 G_L S_{1,1}, \tag{24}$$

and, filling in the values for this case,

$$\begin{aligned} G = e_0'[&(I - zJ)^{-1} + (I - zJ)^{-1} e_m \cdot \\ &\cdot [I - zb_m^2 G_L e_m'(I - zJ)^{-1} e_m]^{-1} z b_m^2 G_L e_m'(I - zJ)^{-1}] e_0, \end{aligned} \tag{25}$$

which produces the result after 'shift throughs' of the type $(I - ab')^{-1}a = a(I - b'a)^{-1}$, with $zb_m^2 G_L e_m'(I - zJ)^{-1} \leftarrow a$ and $e_m \leftarrow b$, assuming existence of the inverses (true in the one-sided algebra). □

The formula obviously generalizes the previous one step recursion, in which we had $m = 0$, $J = a_0$, $e_0 = e_m = 1$.

Referring to Fig. 3, one may use individual eigenstates for the subsystems, and derive an updated secular equation for the total. Indeed, assuming diagonalizations $J^{(i)} = U_i' A_i U_i$ for the original systems, and putting $\beta^{(i)} = [U_i']e_0$, $\gamma^{(1)} := [U_1']e_m$, we find as partially diagonalized global transition matrix

$$\{A^{(t)}, b^{(t)}\} := \left\{ \left[\begin{array}{c|c} A_1 & \gamma^{(1)} b_m [\beta^{(2)}]' \\ \hline \beta^{(2)} b_m [\gamma^{(1)}]' & A_2 \end{array}\right], \left[\begin{array}{c} \beta^{(1)} \\ \hline 0 \end{array}\right] \right\}. \tag{26}$$

The previous extension formula remains valid in terms of the eigenvalue-based state space variables:

$$G(z) = [\beta^{(1)}]' \left[I - zA - \gamma^{(1)} z^2 b_m^2 G_L(z) [\gamma^{(1)}]' \right]^{-1} \beta^{(1)} \tag{27}$$

One may choose for $\{A_2, \beta^{(2)}\}$ any convenient 'sampling' system, possibly a discretization around zero reflecting the desired fineness of the mesh. The combined system will produce a discretized pdf that interpolates the chosen set of moments (as produced by the system $\{A_1, w_1\}$). An interpolation of this new system will yield a continuous pdf that may approximate the desired result reasonably well (although the exact interpolation is based on the discretized version). This can be seen as follows.

Given a moment series $\mu_{0:}$ there is the corresponding spectrum (now with $j := \sqrt{-1}$)

$$M(j\omega) = \overline{e^{j\omega X}} = \sum_{k=0:} \frac{\mu_k}{k!}(j\omega)^k, \tag{28}$$

and the pdf is of course the reverse Fourier transform of $M(j\omega)$. While the generating function typically does not converge, the spectrum often does (except, among others, for processes for which moments do not exist), and this can be pretty fast as well thanks to the weights $k!$. (For example, for a Gaussian, zero mean pdf we have $\frac{\mu_k}{(2k)!} = \frac{\sigma^{2k}}{k!2^k}$.) This means that a relative small number of μ_k's may pretty well determine the pdf up to a reasonable accuracy, so that one might suffice with an expansion of – say – a couple of tens of terms. This approximation issue has not been studied extensively to my knowledge.

Here are few more observations or conjectures:

- when one peals Jacobi sections off a zero-mean Gaussian pdf, one destroys the Gaussian property. The first Jacobi coefficient a_0 determines the means and b_0 determines the variance, since $\mu_2 = a_0^2 + b_0^2$. After this information has been taken out, what remains is information on higher order moments. This effect demonstrates why one can expect loss of the most obvious information when one proceeds with 'pealing off' Jacobi coefficients;
- it seems the further one goes, the more the information gets 'pointed'. For example, when starting out with an originally order n atomic system, at the last step in the recursion, just a single impulse remains;
- one can use Cuppen's method [3] to do the join (perhaps it is possible to enhance the method nicely with an order 2 update, instead of the order one update of Cuppen's method which modifies the oddity of the constituents). The order one update of the secular equation interleaves the join of the eigenvalues. Assuming the interpolation forced by system (1) to be of relatively low order m and spread over an interval $[p_1, p_2]$, then system (2) could be a uniform discretization over the same region (say N atoms with weights $1/N$). The joint system would then have $N + m + 1$ atoms somewhat extending (with two atoms) over the borders set by system (1), but distributed in between the uniform distribution forced by system (2). This would assure a non-uniform, but still fairly evenly distributed final result. Assuming the total order sufficiently large to make the spectral series converge, this method finally yields a continuous distribution that interpolates the given data to a good accuracy (to be investigated further!).

Algebraic models

One may also wonder whether there is a natural stochastic variable connected to system (2) of the previous section, given the total system. For the construction given, this is not necessary. Connected to Hankel matrices, one may construct a purely deterministic polynomial algebra with an inner product defined by the Hankel operator [1, 16]. Using the classical strategy to deal with stochastic variables, related polynomial representations and their geometry, one can identify three, isometric, approaches. Let H be the global moment matrix ($H = \overline{\mathbf{X}\mathbf{X}'}$), which may or may not be singular (often finite singularity is ruled out as a spurious case, but singular cases are highly important in modeling theory, where one works up to a maximal index for which H is not singular, and then derives a computable dicretization).

Approach 1: vectors. Let a and b be vectors padded with zeros to make their indices run to infinity, one defines the inner product $< a, b >_H := a'Hb = (a'L)(b'L)'$ where $H = LL'$ is taken as a minimal factorization of H.

Approach 2: polynomials. Connected to a vector a one may define a polynomial $A(z) = \sum_{k=0:} a_k z^k$ and the inner product $< A(z), B(z) >_H = a'Hb$.

Approach 3: stochastics. This is the approach we have followed so far, where the inner product is defined by expectations: $< a'\mathbf{X}, b'\mathbf{X} > = \overline{a'\mathbf{X}\mathbf{X}'b}$.

Needless to say, the three approaches are equivalent and define the same inner product.

Hankel symmetry

An important characterization of the Hankel symmetry works on one step extensions of the Hankel matrix. An infinite Hankel matrix H with positive definite sub-Hankels $H_{\mathbf{n}}$ can be Cholesky factorized for as much as one wants to go, just by extending the Cholesky factorization of finite restrictions $H_{\mathbf{n}}$ of H. So, when $H_{\mathbf{m}} = L_{\mathbf{m}} L'_{\mathbf{m}}$ and similarly $H_{\mathbf{n}} = L_{\mathbf{n}} L'_{\mathbf{n}}$ are Cholesky factorizations with $m < n$, then $L_{\mathbf{m}}$ is a submatrix of $L_{\mathbf{n}}$, and $L_{\mathbf{n}}$ can be obtained recursively by just extending $H_{\mathbf{m}}$ and $L_{\mathbf{m}}$ recursively (as is usually done in algorithms for Cholesky factorization). In this way we may consider the infinite Hankel H and its (formal) Cholesky factorization $H = LL'$, although these infinite matrices are only defined numerically and will typically be

very unbounded (as is shown by the Gaussian example). Calculus on such matrices is allowed so long as it remains unilateral, meaning: recursively increasing. This works for the determination of corresponding orthonormal polynomials as well, and the inner products on which they depend, for as long as one deals with finitely supported vectors and matrices. Let then $\sigma_{k=0:,\ell=0:}$ with $[\sigma]_{k,\ell} = \delta(k+1-\ell)$ be the infinitely supported upper shift matrix then the Hankel symmetry on the global Hankel series can simply be written as $\sigma H = H\sigma'$. Suppose now that $H = LL'$ and L, being lower triangular, is recursively invertible, then the Hankel symmetry can be characterized in terms of L as well:

Proposition 7. *A recursively invertible (infinite dimensional) L is the lower Cholesky factor of a strictly positive definite Hankel matrix iff $J = L^{-1}\sigma L$ is a Jacobi matrix, and, if so, it is the Jacobi matrix that parametrizes H.*

Proof. One checks that equating the rows in $JL^{-1} = L^{-1}\sigma$ produces the three terms recursion which defines the Jacobi parameters. Uniqueness of these parameters (assuming the b_k strictly positive) does the rest. □

The property can of course be refined further for finite positive definite Hankel matrices, but one has to take account of the fact that J and σ are not lower triangular, leading to some border effects that have to be accounted for. Exploiting the tri-diagonal nature of J one only has the partial Jacobi recursion (for compactness of formulas we now adopt the notations[5] $A_{i,j} \leftarrow A_i^j$ and $[]_{0:\,n}^{0:\,n} \leftarrow \mathbf{n}$)

$$J_{0:n}^{0:n+1} p_{0:n+1}^{0:n} = p_{1:n+1}^{0:n}, \tag{29}$$

in which $J_{0:n}^{0:n+1}$ and $p_{0:n+1}^{0:n}$ are not square. However, a finite, non singular and positive definite Hankel matrix $H_{\mathbf{n}}$ of order n is still characterized by the symmetry of $J_{\mathbf{n}} = L_{\mathbf{n}}^{-1}(\sigma L)_{\mathbf{n}} = L_{\mathbf{n}}^{-1}[L^{\uparrow}]_{\mathbf{n}}$. This equation shows that $L_{n+1}^{0:n}$ enters in the definition of $J_{\mathbf{n}}$. Given only a finite (normalized) $H_{\mathbf{n}} = L_{\mathbf{n}}[L_{\mathbf{n}}]'$, the quantities in $L_{n+1}^{0:n}$ are still to be determined: they depend (as we shall see) partially on $L_{\mathbf{n}}$, but are also partially undefined (and can be characterized further).

As we shall have to characterize finite Hankel matrices in the 2D treatment, let us specialize to characterizations using just an order $n+1$ Jacobi matrix $J = J_{\mathbf{n}}$ (assuming given $H = H_{\mathbf{n}}$ non-singular). Such a Jacobi matrix generates an (infinite) moment series $g_k := e_0' J^k e_0$, whereby $[L_{\mathbf{n}}]_{0:n} =$

[5] We also suppress the single full range indicator ':': $\lambda_n \equiv \lambda_n^:$ specifies the n^{th} row of λ.

$[e_0' J^k]_{0:n}$ *and* $L_{n+1}^{0:n} = L_n J_{\mathbf{n}}$ (in general, for $k > n$: $H_{\mathbf{k}} = L_{0:k}^{0:n}[L_{0:k}^{0:n}]'$, singular of rank $n+1$). This establishes the connection Jacobi matrix $\Rightarrow$ Cholesky factor. We see that $[J_{\mathbf{n}}]_k$ is always a step ahead of $[L_{\mathbf{n}}]_k$: $L_0 = e_0'$, $L_1 = L_0 J_{\mathbf{n}} = \begin{bmatrix} a_0 & b_0 & 0 & \cdots \end{bmatrix}$, $L_{n+1} = L_n J_{\mathbf{n}}$ etc... In the other direction, the Hankel symmetry regulates the extension of $L_{\mathbf{n}}$ up to and beyond n (actually: the Jacobi recursion specifies the relation). More precisely:

(30)
$$\begin{aligned} L_{n+1} &= L_n J_{\mathbf{n}} \\ &= \begin{bmatrix} L_n^0 a_0 + L_n^1 b_0 & L_n^0 b_0 + L_n^1 a_1 + L_n^2 b_2 & \cdots & L_n^{n-1} b_{n-1} + L_n^n a_n \end{bmatrix} \end{aligned}$$

and we see that only a_n enters in the definition L_{n+1}^n and vice versa, or, to put it differently: one may choose either one as parameter. On the other hand, a_n does not enter in the definition of $L_{\mathbf{n}}$ and has no influence on $L_{\mathbf{n}}$ nor on the series $\mu_{0:2n}$. A specific choice will determine a specific $J_{\mathbf{n}}$, which can be used to characterize the series up to $2n$ in a univocal way. We might choose $a_n = 0$, but a more self-contained choice is to make $J_{\mathbf{n}}$ singular, i.e., forcing an atom at zero, which also makes the degree of the generating function minimal (i.e., n). Alternative choices (e.g., making the generating function 'balanced') are of course possible. Hence the

Definition 1. *We call the singular Jacobi matrix* $J_{\mathbf{n}}$ *that characterizes* $H_{\mathbf{n}}$ **canonical**. *This Jacobi matrix has the characteristic property that the corresponding generating function is of degree* n *and is the unique one with that property. We call the corresponding singular extension* $H_{\mathbf{n+1}}$ *and the resulting Cholesky factor* $L_{0:n+1}^{0:n}$ **canonical** *as well.*

Moreover (and extending the results to the non-normalized situation), let an eigen decomposition of a canonical $J_{\mathbf{n}} = U_{\mathbf{n}} \alpha U_{\mathbf{n}}'$ with eigenvalues $\alpha = \mathrm{diag}[\{0, \alpha_{1:n}]$, and let $\beta = U_{\mathbf{n}}' e_0$, then $\{\alpha, \beta\sqrt{g_0}, \sqrt{g_0}\beta'\}$ produces a diagonalized realization for $\widehat{G}(z)$ with $n+1$ distinct eigenvalues and weights $[\beta_k]^2 g_0$; the full state transformation matrix $U_{\mathbf{n}}$ can be generated from J_n, hence from β and α as well, actually

$$\begin{bmatrix} \beta \\ \beta\alpha \\ \vdots \\ \beta\alpha^n \end{bmatrix} = L_{\mathbf{n}} U_{\mathbf{n}} \tag{31}$$

is an LQ-factorization, which can be computed recursively, if so desired.

Summarizing the properties derived so far for the finitely indexed case, we may state for further reference:

Proposition 8. *Related to a strictly positive definite Hankel matrix* $H_{\mathbf{n}} = \text{Han}[g_{0:2n}]$ *of dimensions* $(n+1)^{\times 2}$*, there is a canonical singular Jacobi matrix* $J_{\mathbf{n}}$*, with canonical Jacobi parameters* $\{a_{0:n}, b_{0:n-1}\}$*, real distinct eigenvalues* $\{\alpha_0 = 0, \alpha_{1:n}\}$*, canonical, strictly positive weights* $\{w_{0:n} = \beta_{0:n}^2 g_0\}$ *with* $[\beta_k]_{k=0:n} > 0$ *and* $\sum_{k=0:n} \beta_k^2 = 1$*, and a canonical generating function* $\widehat{G}_{\mathbf{n}}(z)$ *such that*

1. $\widehat{G}_{\mathbf{n}}(z)$ *interpolates the coefficients* $g_{0:2n}$*,*
2. $\widehat{G}_{\mathbf{n}}(z) = \sum_{k=0:n} \frac{w_k}{1-\alpha_k z}$*,*
3. $\widehat{G}_{\mathbf{n}}(z)$ *is rational of degree* n *(corresponding to an atomic density function with* $n+1$ *atoms* $\delta(x-\alpha_k)$ *of weights* w_k*),*
4. $g_i = \sum_{k=1:n} w_k \alpha_k^i$ *for* $i \geq 1$ *and* $g_0 = \sum_{k=0:n} w_k$*.*

Hankel symmetry in the finite case

A compact way of expressing the Hankel symmetry in the finite case is by utilizing the unilateral infinite shift matrix σ and a specific pseudo-inverse for the Cholesky factor $L := L_{0:}^{0:n} = [e_0' J_{\mathbf{n}}^k]_{k=0:\infty}]$ of dimensions $(0\colon \infty) \times (0\colon n)$, in which $J_{\mathbf{n}}$ may be any order n Jacobi matrix (but which we typically will take canonical), namely

$$L^+ := \left[\begin{array}{c|c} L_{\mathbf{n}}^{-1} & 0 \end{array}\right] \tag{32}$$

The Hankel symmetry in terms of L is then simply

Proposition 9. $H = LL'$ *will be Hankel iff* $L^+\sigma L$ *is symmetrical, and* $J = L^+\sigma L$*.*

Proof. only if: the Hankel property is easily seen to be equivalent to $\sigma H = H\sigma'$, which, after applying the pseudo-inverse shown gives $L^+\sigma L = L'\sigma'(L^+)'$ (notice that all quantities are well-defined and products are finite);
if: because $L^+\sigma L = L_{\mathbf{n}}^{-1} L_{1:n+1}^{0:n} = J$, resulting in $L := L_{0:}^{0:n} = [e_0' J_{\mathbf{n}}^k]_{k=0:}]$, so that LL' is necessarily Hankel. □

4. Generalized moment matrices in two variables

A full moment matrix in two dimensions is $H = \overline{\mathbf{X}\mathbf{X}'}$, in which

$$\mathbf{X} := \text{col}\left[\begin{array}{c|cc|ccc|c} 1 & X & Y & X^2 & XY & Y^2 & \cdots \end{array}\right],$$

$$H_3 = \left[\begin{array}{c|cc|ccc|ccc|c} 1 & \mu_{10} & \mu_{01} & \mu_{20} & \mu_{11} & \mu_{02} & \mu_{30} & \mu_{21} & \mu_{12} & \mu_{03} \\ \hline \mu_{10} & \mu_{20} & \mu_{11} & \mu_{30} & \mu_{21} & \mu_{12} & \mu_{40} & \mu_{31} & \mu_{22} & \mu_{13} \\ \mu_{01} & \mu_{11} & \mu_{02} & \mu_{21} & \mu_{12} & \mu_{03} & \mu_{31} & \mu_{22} & \mu_{12} & \mu_{04} \\ \mu_{20} & \mu_{30} & \mu_{21} & \mu_{40} & \mu_{31} & \mu_{22} & \mu_{50} & \mu_{41} & \mu_{32} & \mu_{23} \\ \mu_{11} & \mu_{21} & \mu_{12} & \mu_{31} & \mu_{22} & \mu_{13} & \mu_{41} & \mu_{32} & \mu_{23} & \mu_{14} \\ \mu_{02} & \mu_{12} & \mu_{03} & \mu_{22} & \mu_{13} & \mu_{04} & \mu_{32} & \mu_{23} & \mu_{14} & \mu_{05} \\ \hline \mu_{30} & \mu_{40} & \mu_{31} & \mu_{50} & \mu_{41} & \mu_{32} & \mu_{60} & \mu_{51} & \mu_{42} & \mu_{33} \\ \mu_{21} & \mu_{31} & \mu_{22} & \mu_{41} & \mu_{32} & \mu_{23} & \mu_{51} & \mu_{42} & \mu_{33} & \mu_{24} \\ \mu_{12} & \mu_{22} & \mu_{13} & \mu_{32} & \mu_{23} & \mu_{14} & \mu_{42} & \mu_{33} & \mu_{24} & \mu_{15} \\ \mu_{03} & \mu_{13} & \mu_{04} & \mu_{23} & \mu_{14} & \mu_{05} & \mu_{33} & \mu_{24} & \mu_{15} & \mu_{06} \end{array}\right]$$

Figure 4: The 2D Hankel structure in H_3. Some identical blocks and elements are highlighted.

and X and Y are stochastic variables. Such matrices (and also in higher dimensions) have a generalized Hankel structure, in the sense that blocks on anti-diagonals contain the same elements and are internally Hankel as well, with compatible ordering—see H_3 represented in Fig. 4. Connected to the 2D case we can define a 2D generating function, very much in the same way as in the 1D case, but now introducing two variables z_1 and z_2, summed by order and then as $\mu(I-i,i)_{i=0:I} = (\overline{X^{I-i}Y^i})_{i=0:I}$ within order I:

$$G(z_1,z_2) = \sum_{I=0:} \left(\sum_{i=I:0} \mu_{i,I-i} z_1^i z_2^{I-i} \right) \tag{33}$$

which can also be seen as the unilateral expansion in 2D of $\mathbf{E}\frac{1}{(1-z_1X)(1-z_2Y)}$. We now have almost immediately:

Theorem 1. *A rational generating function $G(z_1,z_2)$ corresponds to a 2D atomic distribution.*

Proof. Consider marginals: one obtains the X marginal by putting $z_2 = 0$ in the generator, and similarly $z_1 = 0$ for the Y marginal. If $G(z_1,z_2)$ is rational, then so are these marginals, and they correspond to an atomic distribution each. The resulting 2D distribution can only have contributions for x values and y values corresponding to the location of the atoms in the marginal distributions, because the latter add the contributions (which are necessarily of the Dirac type due to the monotone character of the distribution function) on their respective abscis or ordinate, and hence have to be atomic as well with atoms located on the intersection of the coordinates of the atoms in the marginal distribution. □

Notation To simplify the writing, we use three related systems of notation. First, the (totally ordered) two digit notation

$$(ij) \in \left[\ 00 \ \middle|\ 10 \quad 01 \ \middle|\ 20 \quad 11 \quad 02 \ \middle|\ \cdots\ \right];$$

next, just positive integers to indicate the global order $i+j$ of an index (ij) or just a normal numerical index; and finally a fat bold face $\mathbf{k} = 0{:}k$ to indicate a global vector or matrix of a given order. So we write $\mathbf{X_k}$ for $\mathbf{X}_{0:k}$, not to be confused with $\mathbf{X}_k = \mathrm{col}[X^k \ X^{k-1}Y \ \cdots \ Y^k]$, which is the vector of index and degree k in $\mathbf{X}$. In the case of matrices, we often use the upper index to indicate a column and the lower to indicate a row, as in $A^{kl}_{ij} = A_{ij,kl}$.

A first example of extension

To introduce the topic, let us look at a simple example first. Suppose we want to find a minimal 2D pdf that matches $\overline{X} = \overline{Y} = 0, \overline{X^2} = \overline{Y^2} = 1, \overline{XY} = 0, \overline{X^3} = \overline{XY^2} = \overline{Y^3} = 0$ *and* $\overline{X^2Y} = d$ for some large d ($|d| > 1$, otherwise the problem is simpler). In other words: extend

$$H_{\mathbf{2}} = \left[\begin{array}{c|cc|cc|c} 1 & 0 & 0 & 1 & 0 & 1 \\ \hline 0 & 1 & 0 & 0 & d & 0 \\ 0 & 0 & 1 & d & 0 & 0 \\ \hline 1 & 0 & d & ? & ? & ? \\ 0 & d & 0 & ? & ? & ? \\ 1 & 0 & 0 & ? & ? & ? \end{array}\right], \tag{34}$$

to a full moment series, or equivalently, the corresponding infinite H. First, we try to do that as singularly as possible, respecting the extended Hankel structure and the positivity. Assuming $H_{\mathbf{1}}$ non-singular, the extension to $H_{\mathbf{2}}$ will be positive definite iff the Schur complement of H_2^2, namely $H_2^2 - H_2^{0:1}(H_{\mathbf{1}})^{-1}H_{0:1}^2$ is positive definite. We call $S_2 = H_2^{0:1}(H_{\mathbf{1}})^{-1}H_{0:1}^2$ the *Schur residue* at index 2, which, in this case, is $S_2 = \begin{bmatrix} 1+d^2 & 0 & 1 \\ 0 & d^2 & 0 \\ 1 & 0 & 1 \end{bmatrix}$ and is *not* Hankel. A minimal solution will be obtained when $H_2^2 - S_2$ is rank deficient of order 2, making $H_{\mathbf{2}}$ rank deficient of order 2 as well. We shall see that this leads to a minimal, atomic pdf, and hence will define the whole moment series. E.g., we can choose $H_2^2 = \begin{bmatrix} 2d^2 & 0 & d^2 \\ 0 & d^2 & 0 \\ d^2 & 0 & d^2 \end{bmatrix}$ in which case $H_2^2 - S_2 = \begin{bmatrix} d^2-1 & 0 & d^2-1 \\ 0 & 0 & 0 \\ d^2-1 & 0 & d^2-1 \end{bmatrix} \geq 0$. Expressing the singu-

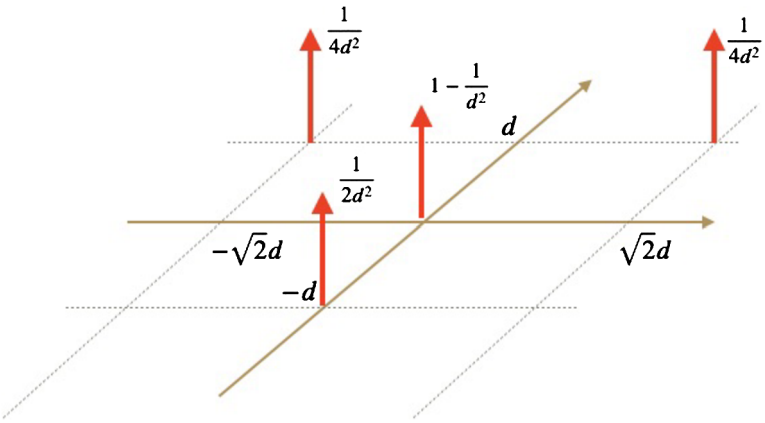

Figure 5: The atomic distribution for the 2D example.

larity, we find $\left\{ \begin{array}{rcl} -dX + XY & = & 0 \\ -dY + X^2 - Y^2 & = & 0 \end{array} \right.$. The first equation says that either $X = 0$ or $Y = d$. When $X = 0$, the second equation says that $Y(Y + d) = 0$ and hence either $Y = 0$ or $Y = -d$. In the second case of $Y = d$, $X^2 = 2d^2$ and hence $X = \pm\sqrt{2}d$. So this means that the corresponding distribution has to be atomic with Dirac impulses at the positions $(0, 0), (0, -d), (-\sqrt{2}d, d), (\sqrt{2}d, d)$ with appropriate positive weights. These are found by solving for the moments (a linear set of equations that can be solved with a simple strategy), and we find as final result the 2D atomic pdf $p(x, y) = (1 - \frac{1}{d^2})\delta(x, y) + \frac{1}{2d^2}\delta(x, y + d) + \frac{1}{4d^2}\delta(x + \sqrt{2}d, y - d) + \frac{1}{4d^2}\delta(x - \sqrt{2}d, y - d)$. Remark that all subsequent $H_{\mathbf{k}}$ for $\mathbf{k} \geq 3$ are now defined and will have the same rank as $H_{\mathbf{2}}$ (4 in this case). However, and as already observed in the 1D case, we probably wish to compute extensions of much larger order, keeping the subsequent generalized Hankels $H_{\mathbf{k}}$ non-singular, until a sufficient high degree is obtained corresponding to a fine grid of atoms and allowing a good approximation of an eventual continuous pdf (this is an issue that we shall not consider in the present paper).

General 2D extension strategy

The considerations given so far lead to a general extension strategy: assuming non-singularity up to some level n, obtained by determining new fitting data H^k_{k+1} and H^{k+1}_{k+1} for $k = 2 \cdots$ of Hankel type and such that H^{k+1}_{k+1} is strictly larger that the $(k+1)^{\text{st}}$ order Schur residue S_{k+1}, until a sufficient level of sampling accuracy is obtained, at which time a new H^n_n has to be

added making $H_{\mathbf{n}}$ rank deficient (it will turn out that deficiency 2 will be possible and sufficient at that point, see the discussion of the singular case further on). When that is the case, then all subsequent Hankels will turn out to be determined, based on an atomic distribution with a fine enough maze (hopefully). In the present paper, we concentrate on generating infinite non-singular extensions through parametrization, leaving the resulting approximation theory for future papers.

Jacobi representations in the 2D case

Jacobi representations may be derived also in 2D and nD cases, but the full parametrization property of the 1D case is lost and has to be restored by removing redundancies (as we shall see). The Jacobi 'coefficients' become matrices in a block-tridiagonal hierarchical matrix, which still have some internal structure that has to be characterized further parametrically. The main result of this paper is that a full parametrization is indeed possible (remember: parameters have to be independent), and an algorithm will be given to determine a full parametrization.

Returning to the ordered stochastic vector

$$[\mathbf{X}_{\mathbf{n}}]_{ij} \text{ for } ij \in [00 \mid 10\ 01 \mid 20\ 11\ 02 \mid \cdots]$$

up to whatever order n, a recursive orthonormalization yields a modified, orthonormal basis $[\mathbf{P}_n]$ in the same order, with $P_{00} = e_0'$ (of whatever dimension needed), and orthonormal with respect to the Gramian $H_{\mathbf{n}}$ of sufficiently high order n. Let, for as far as one wants to go, $H_{\mathbf{n}} = L_{\mathbf{n}} L_{\mathbf{n}}'$. Connected to the orthonormal basis, we shall have 2D polynomials

$$P_{ij}(z_1, z_2) = L_{ij}^{00} + L_{ij}^{10} z_1 + L_{ij}^{01} z_2 + L_{ij}^{20} z_1^2 + L_{ij}^{11} z_1 z_2 + \cdots \tag{35}$$

in which the general term is $L_{ij}^{kl} z_1^k z_2^l$ with (kl) following the same ordered series as (ij). Just as in the 1D case, the orthogonality structure gets propagated recursively via shifts. In the 2D case (figure 6) there are two shifts: multiplication with X and with Y, resulting in the following propagation schema for the shifted polynomials (e.g.,

$$P_{21} \perp 1, (X\cdot)1, Y, (X\cdot)X,\ (X\cdot)Y, Y^2, (X\cdot)X^2,$$

hence $P_{21}(X) \perp \{1, X, Y, X^2\}$ where I have highlighted the X that is taken out to the other side of the inner product by bracketing it).

$\mathbf{X}$	$\mathbf{P}$	$P_{ij}X \perp$ on	$P_{ij}Y \perp$ on
1	1		
X	P_{10}	$\{\}$	$\{\}$
Y	P_{01}	$\{1\}$	$\{\}$
X^2	P_{20}	$\{1\}$	$\{1\}$
XY	P_{11}	$\{1, X\}$	$\{1\}$
Y^2	P_{02}	$\{1, X, Y\}$	$\{1, X\}$
X^3	P_{30}	$\{1, X, Y\}$	$\{1, X, Y\}$
X^2Y	P_{21}	$\{1, X, Y, X^2\}$	$\{1, X, Y\}$
XY^2	P_{12}	$\{1, X, Y, X^2, XY\}$	$\{1, X, Y, X^2\}$
Y^3	P_{03}	$\{1, X, Y, X^2, XY, Y^2\}$	$\{1, X, Y, X^2, XY\}$
		etc.	

Figure 6: Shift orthogonalities in the 2D case.

Noticing $\mathbf{X}_{n+1} = \begin{bmatrix} \mathbf{X}_n X \\ [\mathbf{X}_n]_n Y \end{bmatrix}$, this then results in the following hierarchical Jacobi recursion, exemplified up to order 3 (other choices are possible, see the discussion on symmetries further):

$$
\begin{bmatrix} P_{00}X \\ P_{00}Y \\ \hline P_{10}X \\ P_{01}X \\ P_{01}Y \\ \hline P_{20}X \\ P_{11}X \\ P_{02}X \\ P_{02}Y \end{bmatrix}
=
\left[\begin{array}{c|cc|ccc|cccc}
* & a & & & & & & & & \\
* & * & c & & & & & & & \\ \hline
a & * & h & d & & & & & & \\
0 & h & * & e & f & & & & & \\
c & * & * & * & * & g & & & & \\ \hline
0 & d & e & * & i & j & * & & & \\
0 & 0 & f & i & * & k & * & * & & \\
0 & 0 & 0 & j & k & * & * & * & * & \\
0 & 0 & g & * & * & * & * & * & * & *
\end{array}\right]
\begin{bmatrix} P_{00} \\ \hline P_{10} \\ P_{01} \\ \hline P_{20} \\ P_{11} \\ P_{02} \\ \hline P_{30} \\ P_{21} \\ P_{12} \\ P_{03} \end{bmatrix}
\tag{36}
$$

in which '$*$' indicates some entry, and repeated entries are annotated as 'a, b, c etc.'. The way the schema builds up is as follows:

1. There is a total order in the indexing series: $(kl) \prec (ij)$ iff $k+l < i+j$, or, if $k+l = i+j$ and $k > i$. The linear span $[\mathbf{X}]_{(kl)\preceq(ij)} = [\mathbf{P}]_{(kl)\preceq(ij)}$ always contains any span with a smaller (ij)—this being the central purpose of the definition of the subsequent entries (more solutions are possible!).
2. Any $P_{ij}X$ or $P_{ij}Y$ is orthogonal on any P_{kl} with $k+l \leq i+j-2$. This is obvious from the fact that $\overline{(P_{ij}X)P_{kl}} = \overline{P_{ij}(XP_{kl})}$ and XP_{kl} sits in the span of the P_{mn} with $m+n \leq i+j-1$. This makes the Jacobian matrix

block tridiagonal (*notice:* it is not square anymore: only the blocks in the first upper diagonal are square).

3. There is an evident progression of orthogonalities. P_{01} is orthogonal on $P_{00}X$, since the latter is in the span of $\{P_{00}, P_{10}\} = \bigvee\{1, X\}$. This makes all the blocks in the upper diagonal lower triangular. Moreover, these blocks may all be endowed with positive 'leading coefficients' $\{a, c, d, \cdots\}$.
4. The symmetries in the main diagonal blocks are of course obvious.

To analyze the structure further, let us introduce some notation for the 2D generalized Jacobi matrix, using a simplifying global indexing notation

$$J_{\mathbf{n}} = \begin{bmatrix} A_0 & B_0^u & & & & \\ B_0^\ell & A_1 & B_1^u & & & \\ & B_1^\ell & A_2 & B_2^u & & \\ & & \ddots & \ddots & \ddots & \\ & & & B_{n-2}^\ell & A_{n-1} & B_{n-1}^u \\ & & & & B_{n-1}^\ell & A_n \end{bmatrix} \tag{37}$$

until order n is reached and included. B_k^ℓ is determined by B_k^u and symmetries mentioned above have to be satisfied. The determination of the basis from the Jacobi matrices then proceeds like in the 1D case, but is a bit more complex. Eq. (36) defines $\sigma\mathbf{P}$, where the shift matrix now has the form (in the present case, easily generalizable, see further)

$$\sigma = \left[\begin{array}{c|cc|c|c} X & & & & \\ Y & & & & \\ \hline & X & & & \\ & & X & & \\ & & Y & & \\ \hline & & & X & \\ & & & & X \\ & & & & Y \end{array}\right] \tag{38}$$

and we have $J_{\mathbf{2}} = \overline{\sigma\mathbf{P_2}\mathbf{P_2'}}$ by definition of orthogonality. Now $\mathbf{P_2'}L_{\mathbf{2}}' = \mathbf{X_2'}$, hence

$$J_{\mathbf{2}}L_{\mathbf{2}}' = \overline{\sigma\mathbf{P_2}\mathbf{X_2'}} = \overline{(\sigma L_{\mathbf{2}}^{-1}\mathbf{X_2})\mathbf{X_2'}} \tag{39}$$

But $\sigma L_{\mathbf{2}}^{-1}$ is nothing but a stack of rows from $L_{\mathbf{2}}^{-1}$ (some repeated), which each get multiplied wholly with either X or Y. This multiplication, when transferred to $\mathbf{X_2}$, produces column shifts. More precisely, denoting the rows

of $L_{\mathbf{2}}^{-1}$ by ℓ_{ij} conformal to the indexing of $\mathbf{X_2}$ and, similarly the columns by ℓ^{ij} (i.e., we drop the comma in (i,j) for convenience), we find (specializing to maximum order 2 without impairing generality)

(40)

$$(\sigma L_{\mathbf{2}}^{-1}\mathbf{X_2}) = \left[\begin{array}{c} \ell_{00}\mathbf{X}X \\ \ell_{00}\mathbf{X}Y \\ \hline \ell_{10}\mathbf{X}X \\ \ell_{01}\mathbf{X}X \\ \ell_{01}\mathbf{X}Y \\ \hline \ell_{20}\mathbf{X}X \\ \ell_{11}\mathbf{X}X \\ \ell_{02}\mathbf{X}X \\ \ell_{02}\mathbf{X}Y \end{array}\right]$$

$$= \left[\begin{array}{cc|ccc|cccc} 1 & & & & & & & & \\ 0 & 1 & & & & & & & \\ \hline \ell_{10}^{00} & 0 & \ell_{10}^{10} & & & & & & \\ \ell_{01}^{00} & 0 & \ell_{01}^{10} & \ell_{01}^{01} & & & & & \\ 0 & \ell_{01}^{00} & 0 & \ell_{01}^{10} & \ell_{01}^{01} & & & & \\ \hline \ell_{20}^{00} & 0 & \ell_{20}^{10} & \ell_{20}^{01} & 0 & \ell_{20}^{20} & & & \\ \ell_{11}^{00} & 0 & \ell_{11}^{10} & \ell_{11}^{01} & 0 & \ell_{11}^{20} & \ell_{11}^{11} & & \\ \ell_{02}^{00} & 0 & \ell_{02}^{10} & \ell_{02}^{01} & 0 & \ell_{02}^{20} & \ell_{02}^{11} & \ell_{02}^{02} & \\ 0 & \ell_{02}^{00} & 0 & \ell_{02}^{10} & \ell_{02}^{01} & 0 & \ell_{02}^{20} & \ell_{02}^{11} & \ell_{02}^{02} \end{array}\right] \left[\begin{array}{c} X \\ Y \\ \hline X^2 \\ XY \\ Y^2 \\ \hline X^3 \\ X^2Y \\ XY^2 \\ Y^3 \end{array}\right] := \ell_{\mathbf{2}}^{\sigma}\mathbf{X}_{\mathbf{3}}^{\uparrow}$$

in which $\ell_{\mathbf{2}}^{\sigma}$ is the 'augmented' inverse of $L_{\mathbf{2}}$ shown, $\mathbf{X}_{\mathbf{3}}^{\uparrow} = [\mathbf{X}]_{10:03}$, and $L_{\mathbf{3}}^{\uparrow} := [L_{\mathbf{3}}]_{10:03}$. It follows that

$$J_{\mathbf{2}} = \ell_{\mathbf{2}}^{\sigma} L_{\mathbf{3}}^{\uparrow}. \tag{41}$$

More generally and using a generalization of the notation, we have, with an appropriately column-shifted ℓ^{σ}

$$J_{\mathbf{m}} = \ell_{\mathbf{m}}^{\sigma} L_{\mathbf{m+1}}^{\uparrow} \tag{42}$$

provided $L_{\mathbf{m}}$ is non-singular, and since $L_{\mathbf{m}}$ is lower triangular, its inverse contains the inverse of $L_{\mathbf{m-1}}$ as a diagonal top block. This expression resembles the expression for the 1D-case, except for the second block row, which represents an extra 'mixing' term involving the relation between X and Y (notice that in the 1D case we simply have $J_{\mathbf{m}} = L_{\mathbf{m}}^{-1} L_{\mathbf{m+1}}^{\uparrow}$.)

The recursion for the blocks in $L_{\mathbf{m}}$ follows from the Jacobi matrices, and conversely:

$$(\ell_{\mathbf{m}}^{\sigma})^{-1} J_{\mathbf{m}} = L_{\mathbf{m+1}}^{\uparrow} \tag{43}$$

Let the relevant permutation operator be

$$P^\sigma := \left[\begin{array}{c|c|c|c|c|c|c} 1 & & & & & & \\ \hline & & 1 & & & & \\ & & & 1 & & & \\ \hline & & & & 1 & & \\ & & & & & 1 & \\ & & & & & & 1 \\ \hline & 1 & & & & & \\ & & & 1 & & & \\ & & & & & & 1 \end{array}\right] \tag{44}$$

then

$$P^\sigma \ell^\sigma_{\mathbf{m}} P^\sigma = \left[\begin{array}{c|c|c|c|c|c||c|c|c} 1 & & & & & & & & \\ \hline \ell^{00}_{10} & \ell^{10}_{10} & & & & & & & \\ \ell^{00}_{01} & \ell^{10}_{01} & \ell^{01}_{01} & & & & & & \\ \hline \ell^{00}_{20} & \ell^{10}_{20} & \ell^{01}_{20} & \ell^{20}_{20} & & & & & \\ \ell^{00}_{11} & \ell^{10}_{11} & \ell^{01}_{11} & \ell^{20}_{11} & \ell^{11}_{11} & & & & \\ \ell^{00}_{02} & \ell^{10}_{02} & \ell^{01}_{02} & \ell^{20}_{02} & \ell^{11}_{02} & \ell^{02}_{02} & & & \\ \hline\hline 0 & 0 & 0 & 0 & 0 & 0 & 1 & & \\ 0 & 0 & \ell^{10}_{01} & 0 & 0 & 0 & \ell^{00}_{01} & \ell^{01}_{01} & \\ 0 & 0 & \ell^{10}_{02} & 0 & \ell^{20}_{02} & \ell^{11}_{02} & \ell^{00}_{02} & \ell^{01}_{02} & \ell^{02}_{02} \end{array}\right], \tag{45}$$

which may be written as

$$P^\sigma \ell^\sigma_{\mathbf{m}} P^\sigma = \left[\begin{array}{cc} [L_{\mathbf{2}}]^{-1} & \\ m_{\mathbf{2}} & \ell^{0,0:2}_{0,0:2} \end{array}\right] \tag{46}$$

in which

$$m_{\mathbf{2}} := \left[\begin{array}{c|c|c|c|c|c} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \ell^{10}_{01} & 0 & 0 & 0 \\ 0 & 0 & \ell^{10}_{02} & 0 & \ell^{20}_{02} & \ell^{11}_{02} \end{array}\right] \tag{47}$$

Hence

$$[\ell^\sigma_{\mathbf{2}}]^{-1} = P^\sigma \left[\begin{array}{cc} L_{\mathbf{2}} & \\ -[\ell^{0,0:2}_{0,0:2}]^{-1} m_{\mathbf{2}} L_{\mathbf{2}} & [\ell^{0,0:2}_{0,0:2}]^{-1} \end{array}\right] P^\sigma \tag{48}$$

and

$$P^\sigma \left[\begin{array}{cc} L_{\mathbf{2}} & \\ -[\ell^{0,0:2}_{0,0:2}]^{-1} m_{\mathbf{2}} L_{\mathbf{2}} & [\ell^{0,0:2}_{0,0:2}]^{-1} \end{array}\right] P^\sigma J_{\mathbf{2}} = L^{\uparrow}_{\mathbf{3}}. \tag{49}$$

Reversing the permutations and writing

$$-[\ell^{0,0:2}_{0,0:2}]^{-1} m_{\mathbf{2}} L_{\mathbf{2}} := \left[\begin{array}{c|cc|ccc} 0 & 0 & 0 & 0 & 0 & 0 \\ n^{00}_{01} & n^{10}_{01} & n^{01}_{01} & 0 & 0 & 0 \\ n^{00}_{02} & n^{10}_{02} & n^{01}_{02} & n^{20}_{02} & n^{11}_{02} & n^{02}_{02} \end{array}\right] \tag{50}$$

we obtain

$$(\ell^{\sigma}_{\mathbf{2}})^{-1} = \left[\begin{array}{cc|ccc|cccc} 1 & & & & & & & & \\ 0 & 1 & & & & & & & \\ \hline L^{00}_{10} & 0 & L^{10}_{10} & & & & & & \\ L^{00}_{01} & 0 & L^{10}_{01} & L^{01}_{01} & & & & & \\ n^{00}_{01} & L^{10}_{01} & n^{10}_{01} & n^{01}_{01} & L^{01}_{01} & & & & \\ \hline L^{00}_{20} & 0 & L^{10}_{20} & L^{01}_{20} & 0 & L^{20}_{20} & & & \\ L^{00}_{11} & 0 & L^{10}_{11} & L^{01}_{11} & 0 & L^{20}_{11} & L^{11}_{11} & & \\ L^{00}_{02} & 0 & L^{10}_{02} & L^{01}_{02} & 0 & L^{20}_{02} & L^{11}_{02} & L^{02}_{02} & \\ n^{00}_{02} & L^{00}_{02} & n^{10}_{02} & n^{01}_{02} & L^{01}_{02} & n^{20}_{02} & n^{11}_{02} & n^{02}_{02} & L^{02}_{02} \end{array}\right] := L^{\sigma}_{\mathbf{2}} \tag{51}$$

The n^{lm}_{ik} are 'new' correlation values connecting the Y variable with X. They involve both L and its inverse ℓ but can be computed recursively at the end of each stage, since ℓ depends recursively on L. The recursion works both ways, as we shall see now.

Jacobi matrices from factorization and vice versa

For the first right block diagonal we have immediately from (43)

$$[L^{\sigma}]^i_i B^u_i = L^{i+1}_{i+1}, \text{ for } i = 0: \tag{52}$$

This fully defines the $B^u_i = ([L^{\sigma}]^i_i)^{-1} L^{i+1}_{i+1}$ in terms of the L^i_i as long as these are invertible (which we simply assume at this point; when they become singular, a different strategy as explained before has to be followed by construction of the corresponding kernel). The converse works as well, but is slightly more complex. Assuming recursively $[L^{\sigma}]^i_i$, we may construct L^{i+1}_{i+1} by computing the corresponding n^{kl}_{0i} for $\{kl\} = i0 : 0i$:

$$n^{i0:0i}_{0i} = \left([L^{\sigma}]^i_i\right)^{i0:0i}_{0i} \tag{53}$$

in which $[\ell^{\sigma}]^i_i$ is constructed from $\ell^i_i = [L^i_i]^{-1}$ by shifting the last row as shown in eq. (40). Remark that in contrast of the 1D case, the diagonal blocks of both the factorization and the Jacobi representation increase in dimension. Next,

$$A_i = [\ell^{\sigma}]^i_i \left(L^i_{i+1} - [L^{\sigma}]^{i-1}_i B^u_{i-1}\right), \text{ for } i = 1: \tag{54}$$

with $A_0 = L_1^0$ as starting value. This determines A_i in terms of L_i^i (and its diagonal predecessors), and L_{i+1}^i — notice that this necessitates the computation of $[L^\sigma]_i^{i-1}$, which uses both L_i^i and L_i^{i-1}. Conversely, one reconstructs both L_m, L_m^σ and ℓ_m^σ from the recursion that produces the orthonormal polynomials starting from the given Jacobi matrices (equation (36)). In detail, the recursion proceeds as follows:

$$P_1 = \begin{bmatrix} P_{10} \\ P_{01} \end{bmatrix} = (B_0^u)^{-1} \left(\begin{bmatrix} X \\ Y \end{bmatrix} - A_0 \right) \text{ (since } P_0 = 1!)$$

$$P_2 = \left[\begin{array}{c} P_{20} \\ P_{11} \\ \hline P_{02} \end{array}\right] = (B_1^u)^{-1} \left(\left[\begin{array}{c} P_1 X \\ \hline P_{01} Y \end{array}\right] - B_0^\ell(P_0) - A_1 P_1 \right)$$

etc.

$$P_m = \left[\begin{array}{c} P_{m0} \\ P_{(m-1)1} \\ \vdots \\ \hline P_{0m} \end{array}\right] = (B_{m-1}^u)^{-1} \left(\left[\begin{array}{c} P_{m-1} X \\ \hline P_{0(m-1)} Y \end{array}\right] - B_{m-2}^\ell P_{m-2} - A_{m-1} P_{m-1} \right)$$

(this is a generalized 'three terms' recursion! The 'n-coefficients' are generated at the singularized bottom row). Next, the various powers $X^i Y^j$ can be recovered from the so-defined polynomials, actually each of them has the next $X^i Y^j$ in the total $\{ij\}$ order as its next highest order term.

Detailed parametrization

As remarked earlier, the 'global' parametrization A_i and B_i^u (B_i^ℓ is directly derived from B_i^u) is (largely) incomplete: it does not account for the 'local' Hankel-like structure of the constitutive blocks of the increasing series of $H_{\mathbf{i}}$'s. Assume stage i (i.e., order i) has been reached. At stage $i+1$ the new elements added to the Hankel matrix are H_{i+1}^i of dimensions $(i+2) \times (i+1)$ and H_{i+1}^{i+1} of dimensions $(i+2)^{\times 2}$. Both matrices are Hankel. This means that the algebraic freedom is, in the first case $2i+2$ and $2i+3$ respect., but the total of $4i+5$ real quantities are of course not independent of each other, because they have to make the total new Hankel $H_{\mathbf{i}+1}$ positive definite. Nonetheless, the odd moments in H_{i+1}^i can be chosen independently, constraining the choice[6] of H_{i+1}^{i+1}.

The first aim is now to translate these considerations to the Cholesky factor $L_{\mathbf{i}+1}$, assuming $L_{\mathbf{i}}$ (and of course also $\ell_{\mathbf{i}}$, which produces the orthogonal polynomials $\ell_{\mathbf{i}} \mathbf{X}_{\mathbf{i}}$) known. Assuming the free choice of moments in H_{i+1}^i

[6] This is like in the scalar case. It is important to notice that this property is dependent on the existence of infinite continuations, a fact that we just assume at this point.

made, also $L^{0:i}_{i+1}$ are known (we have of course $L^k_{i+1} = L^{0:k-1}_{i+1}(L_{0:k-1})'\ell^k_k$ for $k = 0 : i-1$). New is only L^i_{i+1}, which can be parametrized by H^i_{i+1}, but can also be parametrized in its own right (we shall discuss this point later and relate it to a parametrization of Jacobi matrices).

Let now the $(i+1)^{\text{st}}$ Schur residue be

$$S_{i+1} := L^{0:i}_{i+1}(L^{0:i}_{i+1})', \tag{55}$$

and consider the Schur complement at index $i+1$: $H^{i+1}_{i+1} - S_{i+1}$. The total $H_{\mathbf{i+1}}$ will be strictly positive definite, iff $H^{i+1}_{i+1} \gg S_{i+1}$, and $L^{i+1}_{i+1}{L^{i+1}_{i+1}}' = H^{i+1}_{i+1} - S_{i+1}$. I claim

Proposition 10. *A non-singular L^{i+1}_{i+1} may be parametrized by strictly positive numbers on its first diagonal, and arbitrary numbers on its first subdiagonal, i.e., $2i+3$ elements all together.*

Proof. Suppose L^{i+1}_{i+1} parametrized as proposed, one has to show that both the (strictly) positive definite Hankel matrix H^{i+1}_{i+1}, and the continuation of L^{i+1}_{i+1} to further lower diagonals, are well defined. H^{i+1}_{i+1} will be Hankel iff (1) it is symmetric, and (2) the submatrix $[H^{i+1}_{i+1}]^{0:i}_{1:i+1}$ is symmetric as well (proof of this fact is by tracing equalities: $(2,0) \to (0,2) \to (1,1)$ etc.). The first condition is automatically satisfied. As to the second, let for any square matrix X, $\Delta(X) := X - X'$ (a skew matrix), then H will be Hankel iff $\Delta([L_{i+1}L'_{i+1}]^{0:i}_{1:i+1}) = -\Delta([S_{i+1}]^{0:i}_{1:i+1})$, which is assumed known at this point of the recursion. This equation can be fulfilled recursively by solving $\Delta([L_{i+1}L'_{i+1}]^{0:k}_{1:k+1}) = -\Delta([S_{i+1}]^{0:k}_{1:k+1})$ for $k = 2 : i$. Putting for brevity, $L_{i+1} \leftarrow L$ and $[S_{i+1}] \leftarrow S$, we have $[LL']^{0:k}_{1:k+1} = L^{0:k+1}_{1:k+1}[L']^{0:k}_{0:k+1} = L^{0:k}_{1:k+1}[L']^{0:k}_{0:k}$ and we have to require $\Delta(L^{0:k}_{1:k+1}[L']^{0:k}_{0:k}) = -\Delta(S^{0:k}_{1:k+1})$. Writing out:

$$\left[\begin{array}{c|c} L^{0:k-1}_{1:k} & 0 \\ \hline \boxed{L^{0:k-1}_{k+1}} & L^k_{k+1} \end{array}\right] \left[\begin{array}{c|c} [L']^{0:k-1}_{0:k-1} & [L']^k_{0:k-1} \\ \hline 0 & L^k_k \end{array}\right] - \tag{56}$$
$$-\left[\begin{array}{c|c} [L]^{0:k-1}_{0:k-1} & 0 \\ \hline [L]^{0:k-1}_k & L^k_k \end{array}\right] \left[\begin{array}{c|c} [L']^{1:k}_{0:k-1} & [L']^{k+1}_{0:k-1} \\ \hline 0\cdots 0\ L^k_k & L^k_{k+1} \end{array}\right] \stackrel{?}{=} -\Delta(S^{0:k}_{1:k+1})$$

and we have to satisfy, at stage k,

$$\boxed{L^{0:k-1}_{k+1}}[L']^{0:k-1}_{0:k-1} = L^{0:k-1}_k[L']^{0:k-1}_{1:k} + [0\cdots 0\ (L^k_k)^2] - \Delta(S^{0:k}_{1:k+1}), \tag{57}$$

which define the remaining entries in L_{k+1}, since $[L']_{0:k-1}^{0:k-1}$ is known and invertible, and both L_{k+1}^k and L_{k+1}^{k+1} are given. □

Example. In the example of eq. (34), let's choose the diagonal and subdiagonal for $(L_2^2 \leftarrow L) := \begin{bmatrix} \epsilon & & \\ 0 & 1 & \\ ? & 0 & 1 \end{bmatrix}$, in which ϵ is a small positive number. We find $\Delta(S) = \begin{bmatrix} 0 & d^2-1 \\ 1-d^2 & 0 \end{bmatrix}$, $\Delta(LL') = \begin{bmatrix} 0 & 1-\epsilon L_2^0 \\ \epsilon L_2^0 - 1 & 0 \end{bmatrix}$, hence $L = \begin{bmatrix} \epsilon & & \\ 0 & 1 & \\ \frac{d^2}{\epsilon} & 0 & 1 \end{bmatrix}$, and finally

$$S + LL' = \begin{bmatrix} 1+d^2+\epsilon & 0 & 1+d^2 \\ 0 & 1+d^2 & 0 \\ 1+d^2 & 0 & 2+\frac{d^4}{\epsilon^2} \end{bmatrix}. \tag{58}$$

This example shows that a small choice for $(L_2^2)_0^0$ may (often) result in a very large H_2^2—compare this to the non-existence of a solution when $(L_2^2)_0^0$ is chosen zero. The singular case will be discussed later.

The Schur residue S_{i+1} appears to be recursively computable from the Jacobi recursion, and must be included in the determination of the next L_{i+1}^{i+1}, the point being that this will not influence the parametrization. We have

Proposition 11. *The Schur residue S_{i+1} is obtained by putting $B_i^u = 0$ in the generalized Jacobi recursion, i.e.,*

$$S_{i+1} = L_{i+1}^S (L_{i+1}^S)' \tag{59}$$

where

$$L_{i+1}^S = L_{\mathbf{i}}^\sigma J_{\mathbf{i}} \tag{60}$$

Proof. We have $J_{\mathbf{i}}^{\mathbf{i}+1} = \ell_{\mathbf{i}}^\sigma [L_{\mathbf{i}}^\uparrow]^{\mathbf{i}+1}$ — see formula (42). In the case of the Schur residue, the last column of $[L_{\mathbf{i}}^\uparrow]^{\mathbf{i}+1}$ is put equal to zero, making L_{i+1} solely dependent on $L_{\mathbf{i}}^\sigma = (\ell_{\mathbf{i}}^\sigma)^{-1}$. □

Remark Putting the last block column in $[L_{\mathbf{i}}^\uparrow]^{\mathbf{i}+1}$ to zero will not result in S_{i+1} being Hankel, as observed before.

The parametrization via L^i_{i+1} and L^{i+1}_{i+1} can fairly easily be transformed into a parametrization of the Jacobi matrices B^u_i and A_i (equations (52) and (54)), taking the structure of these matrices into account.

Hankel symmetries in the 2D case

Consider two (restricted) shift matrices:

$$\sigma_1 = \left[\begin{array}{c|cc|ccc|c} 0 & 1 & 0 & & & & \\ \hline & 0 & & 1 & 0 & 0 & \\ & & 0 & 0 & 1 & 0 & \\ \hline & & & \ddots & & & \ddots \end{array}\right], \quad \sigma_2 = \left[\begin{array}{c|cc|ccc|c} 0 & 0 & 1 & & & & \\ \hline & 0 & & 0 & 1 & 0 & \\ & & 0 & 0 & 0 & 1 & \\ \hline & & & \ddots & & & \ddots \end{array}\right]. \tag{61}$$

Proposition 12. *Let $H_\infty = LL'$ be the Cholesky factorization of the 2D positive definite H_∞ matrix and let L be (formally) invertible, then H_∞ will be generalized Hankel iff both $L^{-1}\sigma_1 L$ and $L^{-1}\sigma_2 L$ are symmetric matrices.*

Proof. H_∞ is generalized Hankel iff both equations $\sigma_i H_\infty = H_\infty \sigma_i'$ $(i = 1, 2)$ are satisfied, much as in the scalar case. From this follows the proposition, provided L is formally invertible. □

Notice that the two equations overlap. It would suffice that one of the two is satisfied (say the first one), and that, in addition, the marginal for Y is scalar Hankel—this is exactly what we did before. Notice also that all the matrices in the products $L^{-1}\sigma_i L$ have dimensions $[1 \oplus 2 \oplus 3 \oplus \cdots]^{\times 2}$.

The symmetries expressed in terms of the Cholesky factor lead, in turn, to the symmetries the Jacobi matrices have to satisfy (this point can be worked out further, but is beyond the scope of the present paper).

The singular case

The full singular 2D (or nD) case has not been explored so far to the best of my knowledge. It is important because 1., it serves to terminate extensions once a sufficient number of additional moments has been reached and 2., it allows for variations in the relations between the stochastic variables. Nonetheless, the method used in the non-singular parametrization allows us to generate singular solutions of higher order in a systematic way (sufficient to terminate extensions) and, in addition, shows the way towards further generalization, thereby necessitating an extension of the theory beyond the hierarchical Hankel matrices considered so far. I briefly indicate how to generate singular solutions based on the parametrization.

With reference to the key parametrization step given in Proposition 10 and the simplified notation used there, one remarks that the next global

moment matrix $H_{\mathbf{i+1}}$ becomes singular when $L^{0(i+1)}_{0(i+1)} = 0$, and singular of rank 2 when both $L^{1i}_{0(i+1)} = 0$ and $L^{1i}_{1i} = 0$ in addition. Such choices are actually generic, but there are special cases where other choices are possible: as stated in the proposition, the diagonal and sub-diagonal elements of L^{i+1}_{i+1} can be chosen freely in the non-singular case, but the numbers on the diagonal have to be strictly non-zero. However, only the last two columns of L^{i+1}_{i+1} can be made zero generically without affecting the columns $0 \cdots i-1$, for the determination of which a division with the diagonal element (the pivot!) is called for (see also the example of (34)). The same holds for the other diagonal elements up to $i-1$. Hence, only the elements i and $i+1$ on the main diagonal can be made zero generically, while the resulting L^{i+1}_{i+1} will have a kernel of rank 2 only when the element $L^{1i}_{0(i+1)} = 0$ as well—but there is a small (and discrete) collection of cases where the two-dimensional co-kernel of L^{i+1}_{i+1} cannot be spanned by vectors whose projection on the last two components is non-singular. The reasoning that follows will nonetheless apply to such cases as well—actually one could make due with a factorization of H^{i+1}_{i+1} that is not lower-upper.

The introduction of a double singularity in the determination of $H_{\mathbf{i+1}}$ then effectively makes any *whole hierarchical Hankel H* extending $H_{\mathbf{i+1}}$ atomic. This is like in the 1D case, where extensions of a singular $H_{\mathbf{i}}$ are unique (Proposition 1); in 2D there are more possibilities, but generically there is only a single case, when, as assumed, $H_{\mathbf{i}}$ is non-singular. This can be seen as follows.

Proposition 13. *Suppose, at step* $i+1$, $\begin{bmatrix} a \\ b \end{bmatrix} H_{\mathbf{i}+1} = 0$ *while* $H_{\mathbf{i}}$ *is non singular, and* a, b *independent (row) vectors. Then all subsequent* H^k_{i+1} *are uniquely determined for all orders* $k > i+1$.

Proof. The equation $\begin{bmatrix} a \\ b \end{bmatrix} H_{\mathbf{i}+1} = 0$, when written out in terms of the moments in $H_{\mathbf{i+1}}$ results in what is called a Macaulay system of equations (see e.g., [9] and the references therein on this type of equation). With respect to H^k_{i+1} for $k \geq i+1$, given $H_{\mathbf{i}}$ and splitting $\left[\begin{array}{c} a \\ b \end{array}\right] = \left[\begin{array}{c|c} a^{0:i} & a^{i+1} \\ b^{0:i} & b^{i+1} \end{array}\right]$ we have $\begin{bmatrix} a^{i+1} \\ b^{i+1} \end{bmatrix} H^k_{i+1} = -\begin{bmatrix} a^{0:i} \\ b^{0:i} \end{bmatrix} H^k_{0:i}$ for all $k \geq i+1$ recursively defining each successive H^k_{i+1} starting with the known H^{i+1}_{i+1}. It turns out that the system $\begin{bmatrix} a^{i+1} \\ b^{i+1} \end{bmatrix} H^{i+1}_{i+1} = -\begin{bmatrix} a^{0:i} \\ b^{0:i} \end{bmatrix} H^{i+1}_{0:i}$ is redundant of order 2 in this case

(by counting equations and variables), and we show now that it is also non-singular (notice that this also results in $\begin{bmatrix} a \\ b \end{bmatrix} \mathbf{X_{i+1}} = 0$). This is because any $x \neq 0$ that would make $x \begin{bmatrix} a^{i+1} \\ b^{i+1} \end{bmatrix} = 0$ also results in $x \begin{bmatrix} a^{0:i} \\ b^{0:i} \end{bmatrix} \mathbf{X_i} = 0$ with $x \begin{bmatrix} a^{0:i} \\ b^{0:i} \end{bmatrix} \neq 0$ and hence $H_\mathbf{i}$ singular (for both $x \begin{bmatrix} a^{0:i} \\ b^{0:i} \end{bmatrix}$ and $x \begin{bmatrix} a^{i+1} \\ b^{i+1} \end{bmatrix}$ cannot be zero at the same time, since a and b have been chosen independent of each other). Subsequently, all the following $H^k_{\mathbf{i+1}}$ are determined for all $k \geq i+1$ and all $H_\mathbf{k}$ are of the same rank as $H_{\mathbf{i+1}}$ — a characteristic property of an atomic system. □

There are of course many other possibilities to extend $H_\mathbf{i}$ to higher orders: one can first introduce a single singularity at some $k_1 > i$, and then another one at a larger order $k_2 > k_1$, thereby introducing a differentiation in the resulting degrees of X and Y. All these possibilities still have to be explored, as well as their effects on approximations of pdf's that interpolate the original data.

Once a singular extension of order two has been obtained—say at order n—the further moment matrices $\bigvee H_\mathbf{k}$ $(k \geq n)$ will all be of the same rank, and the resulting atomic distribution can be obtained fairly easily by first evaluating the marginals, and next filling in the detailed distributions per abscis or ordinate, as was shown by the examples.

Toward the nD case

The theory extends to the nD case just by appending more variables in the list:

$$\mathbf{X} := \mathrm{col}\left[\ 1 \ \middle|\ X \quad Y \quad Z \quad \cdots \ \middle|\ X^2 \quad XY \quad XZ \quad Y^2 \quad \cdots \ \middle|\ \text{etc...} \ \right]. \tag{62}$$

This leads again to a global generalized Hankel $H = \overline{\mathbf{X}\mathbf{X}'}$, with individual blocks that exhibit 'local' generalized Hankel structures in turn. Also in this case one may define a generating matrix $G(z_1, z_2, \cdots)$ as before, and if this generating matrix is rational, it will lead to atomic distributions in the same way as in the 2D case. To extend the 2D parametrization to the nD case will require more work than done for the 2D case: the local matrices are not simply Hankel any more, but a block Hankel mixture of Hankel blocks. It seems possible to develop a hierarchical theory that generalizes the 2D theory presented, because this theory already shows how to handle a two tier Hankel hierarchy, and the ideas generalize.

5. Conclusions

1. The main result of the paper is a method to establish a Jacobi-like parametrization for all possible extensions of a given, fully ordered set of increasing moments of two coupled stochastic variables, and this for the non-singular case. At stage i of the parametrization, corresponding to the inclusion of degree $2(i+1)$, $4i+5$ independent parameters are added to the parametrization, which also corresponds to the total number of moments of degree $2i+1$ and $2i+2$ added (these moments are of course not independent, and hence do not amount to a parametrization). Although the algebra has not been fully worked out yet, the method seems to extend to n variables, modulo a stark increase in complexity.
2. Although the singular case has not been fully worked out in this paper, the paper shows how it can be tackled for more than one stochastic variable. Singular extensions lead to atomic or partially atomic realizations of pdf's that match the given number of (fully ordered) moments. In particular, it appears that at any stage of the parametrization, the parametrization can be stopped, resulting in an overall atomic solution. In the case of two variables, It suffices to create an extension that results in a total singularity of rank two for the full continuation of the Hankel matrix to be determined. This is because the number of conditions introduced by the singular vectors equals the number of free parameters at each stage of an extension.
3. The theory presented rests on (1) a generalization of the Akhiezer orthonormalization method for polynomials that are orthogonal with respect to a positive definite Hankel Gramian and (2) specific structural properties of generalized Hankel matrices of the type related to stochastic moment series (these are not block Hankel, but have strong Hankel-like block symmetries with blocks that vary in dimensions).
4. Many issues concerning the method presented are open. Here is a short list (to the best of my knowledge):
   - parametrizations for the general nD case (hopefully with a more streamlined way of handling symmetries than presented in this paper);
   - the development of a functional approximation theory for the pdf's resulting from the parametrizations and the derivation of simplified, approximate models from them;
   - (perhaps also) the parametrization for other cases of sufficient statistics other than Gaussian (e.g., Levi-Pareto derived statistics);

– all the necessary numerics has to be developed as well, not to mention
– various applications.

Acknowledgments

The author wishes to acknowledge many very stimulating and helpful discussions on the non-linear moment problem with Sankar Basu, Jan Zarzycki, Robert Hossa and Agnieszka Wielgus, and to Tom Kailath on the topics of System Theory and Stochastic Signal Processing. He is also very grateful to the Signal Processing group at Wroclaw University of Technology, the Institute of Advanced Study of the Technische Universität München, and especially to Tom Kailath privately for multiple and extensive periods of hospitality.

References

[1] N. I. Akhiezer. *The Classical Moment Problem*. Oliver and Boyd, Edinburgh, 1965.

[2] N. K. Bose and S. Basu. Two-dimensional matrix pade approximants: Existence, nonuniqueness, and recursive computation. *IEEE Trans. on Automatic Control*, AC-25:509–514, June 1980. MR0571758

[3] J. J. M. Cuppen. A divide and conquer method for the symmetric tridiagonal eigenvalue problem. *Numerische Mathematik*, 36:177–195, 1981. MR0611491

[4] Raúl E. Curto and Lawrence A. Fialkow. Flat extensions of positive moment matrices: Relations in analytic or conjugate terms. *Operator Theory: Advances and Applications*, OT 104:59–82, 1998. MR1639649

[5] P. Dewilde and E. F. Deprettere. Approximate inversion of positive matrices with application to modeling. In *Modeling, Robustness and Sensitivity Reduction in Control Systems*, pages 212–238. Springer, NATO ASI Series, 1987. MR0912138

[6] P. Dewilde, A. C. Vieira, and T. Kailath. On a generalized Szegö-Levinson realization algorithm for optimal linear predictors based on a network synthesis approach. *IEEE Trans. Circuits Syst.*, 25(9):663–675, September 1978. MR0510691

[7] Fred L. Fontaine and Sankar Basu. The partial realization problem for complex hamburger series and complex lossless multiport networks. *IEEE Trans. on CAS-I*, 45:161–177, January 1999. MR1683246

[8] H. Hamburger. Ueber eine erweiterung des stieltjesschen momentproblems. *Math. Ann.*, 81,82:235–319, 120–164, 1920–21. MR1511978

[9] Ph. Dreesen K. Batselier and B. De Moor. On the null spaces of the macaulay matrix. *LAA*, 460:259–289, 2014. MR3250542

[10] T. Kailath. *Lectures on Linear Least-Squares Estimation*, CISM Courses and Lectures, No. 140. Springer Verlag, Wien, New York, 1976.

[11] T. Kailath. *Linear Systems*. Prentice Hall, New York, 1980. MR0569473

[12] T. Kailath. *Lectures on Wiener and Filtering*, CISM Courses and Lectures, No. 140. Springer Verlag, Wien, New York, 1981.

[13] N. Levinson. The Wiener RMS error criterion in filter design and prediction. *J. Math. Phys.*, 25:261–278, 1947. MR0019257

[14] L. Miranian. Matrix-valued orthogonal polynomials on the real line: some extensions of the classical theory. *Journal of Physics A*, 38:5731–5749, 2005. MR2168386

[15] I. Schur. Uber Potenzreihen, die im Innern des Einheitskreises beschränkt sind, I. *J. Reine Angew. Math.*, 147:205–232, 1917. Eng. Transl. *Operator Theory: Adv. Appl.*, vol. 18, pp. 31–59, Birkhäuser Verlag, 1986. MR1580948

[16] Barry Simon. The classical moment problem as a self-adjoint finite difference operator. *Advances in Mathematics*, 137:82–203, 1998. MR1627806

[17] J. C. Willems. Dissipative dynamical systems – Part II: Linear systems with quadratic supply rates. *Archive for Rational Mechanics and Analysis*, 45:352–393, 1972. MR0527463

[18] Jan Zarzycki. Multidimensional non-linear schur parametrization of nongaussian stochastic signals. Parts I–III. *Multidimensional Systems and Signal Processing*, 15:217–241, 243–275, 313–340, 2004. MR2075151

Patrick Dewilde
Systems of Signal Processing Department, Faculty of Electronics
Wroclaw University of Technology
Wroclaw
Poland
Institute of Advanced Study
Technische Universität München
Garching
Germany
E-mail address: p.dewilde@me.com

Received January 29, 2020