MEASURES WITH ZEROS IN THE INVERSE OF THEIR MOMENT MATRIX

J. WILLIAM HELTON, JEAN B. LASSERRE, AND MIHAI PUTINAR

ABSTRACT. We investigate and discuss when the inverse of a multivariate truncated moment matrix of a measure μ has zeros in some prescribed entries. We describe precisely which pattern of these zeroes corresponds to independence, namely, the measure having a product structure. A more refined finding is that the key factor forcing a zero entry in this inverse matrix is a certain *conditional triangularity property* of the orthogonal polynomials associated with the measure μ.

1. Introduction

It is well known that zeros in off-diagonal entries of the inverse M^{-1} of a $n \times n$ covariance matrix M identify pairs of random variables that have no *partial correlation* (and so are conditionally independent in case of normally distributed vectors); see e.g. Wittaker[7, Cor. 6.3.4]. Allowing zeros in the off-diagonal entries of M^{-1} is particularly useful for Bayesian estimation of regression models in statistics, and is called Bayesian *covariance selection*. Indeed, estimating a covariance matrix is a difficult problem for large number of variables, and exploiting sparsity in M^{-1} may yield efficient methods for Graphical Gaussian Models (GGM). For more details, the interested reader is referred to Cripps et al. [3], and the many references therein.

The covariance matrix can be thought of as a matrix whose entries are second moments of a measure. This paper focuses on the truncated moment matrices, M_d, consisting of moments up to an order determined by d. First we describe precisely the pattern of zeroes of M_d^{-1} resulting from the measure having a product type structure. Next we turn to the study of a particular entry of M_d^{-1} being zero. We find that the key is what we call the *conditional triangularity* property of *orthonormal polynomials* (OP) up to degree $2d$, associated with the measure. To give the flavor of what

1991 *Mathematics Subject Classification.* 52A20 52A.

Key words and phrases. Moment matrix; orthogonal polynomials.

This work was initiated at Banff Research Station (Canada) during a workshop on Positive Polynomials, and completed at the Institute for Mathematics and its Applications (IMA, Minneapolis). All three authors wish to thank both institutes for their excellent working conditions and financial support.

J.W. Helton was partially supported by the NSF and Ford Motor Co., M. Putinar by the NSF and J.B. Lasserre by the (French) ANR.

this means, let for instance μ be the joint distribution μ of n random variables $X = (X_1, \ldots, X_n)$, and let $\{p_\sigma\} \subset \mathbb{R}[X]$ be its associated family of orthonormal polynomials. When $(X_k)_{k \neq i,j}$ is fixed, they can be viewed as polynomials in $\mathbb{R}[X_i, X_j]$. If in doing so they exhibit a triangular structure (whence the name conditional triangularity w.r.t. $(X_k)_{k\neq i,j}$), then entries of M_d^{-1} at precise locations vanish. Conversely, if these precise entries of M_d^{-1} vanish (robustly to perturbation), then the conditional triangularity w.r.t. $(X_k)_{k\neq i,j}$ holds. And so, for the covariance matrix case $d = 1$, this conditional triangularity property is equivalent to the zero partial correlation property well studied in statistics (whereas in general, conditional independence is *not* detected by zeros in the inverse of the covariance matrix).

Interestingly, in a different direction, one may relate this issue with a constrained *matrix completion* problem. Namely, given that the entries of M_d corresponding to marginals of the linear functional w.r.t. one variable at a time, are fixed, complete the missing entries with values that make M_d positive definite. This is a *constrained* matrix completion problem as one has to respect the moment matrix structure when filling up the missing entries. Usually, for the classical matrix completion problem with *no* constraint on M, the solution which maxmizes an appropriate entropy, gives *zeros* to entries of M^{-1} corresponding to missing entries of M. But under the additional constraint of respecting the moment matrix structure, the maximum entropy solution does not always fill in M_d^{-1} with zeros at the corresponding entries (as seen in examples by the authors). Therefore, any solution of this constrained matrix completion problem does not always maximize the entropy. Its "physical" or probabilistic interpretation is still to be understood.

We point out another accomplishment of this paper. More generally than working with a measure is working with a linear functional ℓ on the space of polynomials. One can consider moments with respect to ℓ and moment matrices. Our results hold at this level of generality.

2. Notation and definitions

For a real symmetric matrix $A \in \mathbb{R}^{n\times n}$, the notation $A \succ 0$ (resp. $A \succeq 0$) stands for A positive definite (resp. semidefinite), and for a matrix B, let B' or B^T denote its transpose.

2.1. **Monomials and polynomials.** We now discuss monomials at some length, since they are used in many ways, even to index the moment matrices which are the subject of this paper. Let $\mathbb{N}$ denote the nonnegative integers and $\mathbb{N}^n$ denote n tuples of them and for $\alpha = (\alpha_1, \alpha_2 \cdots \alpha_n) \in \mathbb{N}^n$, define $|\alpha| := \sum_i \alpha_i$. The set $\mathbb{N}^n$ sits in one to one correspondence with the monomials via

$$\alpha \in \mathbb{N}^n \ \sim \ X^\alpha := X_1^{\alpha_1} X_2^{\alpha_2} \cdots X_n^{\alpha_n}.$$

Recall also the standard notation

$$\deg X^\alpha = |\alpha| = \alpha_1 + ... + \alpha_n.$$

By abuse of notation we will freely interchange below X^α with α, for instance speaking about $\deg \alpha$ rather than $\deg X^\alpha$, and so on.

Define the the **fully graded partial order (FG order) " $\leq$ ",** on monomials, or equivalently for $\gamma, \alpha \in \mathbb{N}^n$ by $\gamma \leq \alpha$, iff $\gamma_j \leqslant \alpha_j$ for all $j = 1, \cdots, g$. Important to us is:

$$\alpha \leq \beta \qquad \text{iff} \qquad X^\alpha \text{ divides } X^\beta.$$

Define the the **graded lexicographic order (GLex) " $<_{gl}$ ",** on monomials, or equivalently for $\gamma, \alpha \in \mathbb{N}^n$ first by using $\deg \gamma \leq \deg \alpha$ to create a partial order. Next refine this to a total order by breaking ties in two monomials X^{m_1}, X^{m_2} of same degree $|m_1| = |m_2|$, as would a dictionary with $X_1 = a, X_2 = b, \cdots$. Beware $\gamma <_{gl} \alpha$ does not imply $\gamma \leqslant \alpha$; for example, $(1,1,3) <_{gl} (1,3,1)$ but $\leq$ fails. However, $\beta \leq \alpha$ and $\beta \neq \alpha$ implies $\beta <_{gl} \alpha$.

It is convenient to list all monomials as an infinite vector $v_\infty(X) := (X^\alpha)_{\alpha \in \mathbb{N}^n}$ where the entries are listed in GLex order, henceforth called the **tautological vector**; $v_d(X) = (X^\alpha)_{|\alpha| \leq d} \in \mathbb{R}^{s(d)}$ denotes the finite vector consisting of the part of $v_\infty(X)$ containing exactly the degree $\leq d$ monomials.

Let $\mathbb{R}[X]$ denote the ring of real polynomials in the variables $X_1, \ldots, X_n$ and let $\mathbb{R}_d[X] \subset \mathbb{R}[X]$ be the $\mathbb{R}$-vector space of polynomials of degree at most d. A polynomial is a linear combination of polynomials and $g \in \mathbb{R}[X]$ can be written

$$g(X) \;=\; \sum_{\alpha \in \mathbb{N}^n} g_\alpha X^\alpha = \langle \mathbf{g}, v_d(X) \rangle \tag{2.1}$$

for some real vector $\mathbf{g} = (g_\alpha)$, where the latter is the standard non-degenerate pairing between $\mathbb{R}^{s(d)}$ and $\mathbb{R}^{s(d)} \otimes_{\mathbb{R}} \mathbb{R}[X]$.

2.2. **Moment matrix.** Let $y = (y_\alpha)_{\alpha \in \mathbb{N}^n}$ (i.e. a sequence indexed with respect to the ordering GLex), and define the linear functional $L_y : \mathbb{R}[X] \to \mathbb{R}$ to be:

$$g \mapsto L_y(g) \;:=\; \sum_{\alpha \in \mathbb{N}^n} g_\alpha \, y_\alpha, \tag{2.2}$$

whenever g is as in (2.1).

Given a sequence $y = (y_\alpha)_{\alpha \in \mathbb{N}^n}$, the **moment matrix** $M_d(y)$ associated with y has its rows and columns indexed by $\alpha, |\alpha| \leq d$, and

$$M_d(y)(\alpha, \beta) \;:=\; L_y(X^\alpha X^\beta) \;=\; y_{\alpha+\beta}, \quad \forall\, \alpha, \beta \in \mathbb{N}^n \text{ with } |\alpha|, |\beta| \leq d.$$

For example, $M_2(y)$ is

$$M_2(y):\quad \begin{array}{rc|cccccc} & & 1 & X_1 & X_2 & X_1^2 & X_1X_2 & X_2^2 \\ 1 & \rightarrow & 1 & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\ X_1 & \rightarrow & y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ X_2 & \rightarrow & y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ X_1^2 & \rightarrow & y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\ X_1X_2 & \rightarrow & y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} \\ X_2^2 & \rightarrow & y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} \end{array}$$

A sequence y is said to have a representing measure μ on $\mathbb{R}^n$ if

$$y_\alpha = \int_{\mathbb{R}^n} x^\alpha \, \mu(dx), \qquad \forall\, \alpha \in \mathbb{N}^n,$$

assuming of course that the signed measure μ decays fast enough at infinity, so that all monomials are integrable with respect to its total variation. Note that the functional L_y produces for every $d \geq 0$ a positive semidefinite moment matrix $M_d(y)$ if and only if

$$L_y(P^2) \geq 0, \quad P \in \mathbb{R}[X].$$

The associated matrices $M_d(y)$ are positive definite, for all $d \geq 0$ if and only if $L_y(P^2) = 0$ implies $P = 0$. We will distinguish this latter case by calling L_y a **positive functional**.

3. Measures of product form

In this section, given a sequence $y = (y_\alpha)$ indexed by $\alpha, |\alpha| \leq 2d$, we investigate some properties of the the inverse $M_d(y)^{-1}$ of a positive definite moment matrix $M_d(y)$ when entries of the latter satisfy a product form property.

Definition 1. We say that the variables $X = (X_1, \ldots, X_n)$ are mutually independent, and the moment matrix $M_d(y) \succ 0$ has the **product form property**, if

$$L_y(X^\alpha) = \prod_{i=1}^{n} L_y(X_i^{\alpha_i}), \quad \forall \alpha \in \mathbb{N}^n, \, |\alpha| \leq 2d. \tag{3.1}$$

If $M_d(y)^{-1}(\alpha, \beta) = 0$ for every y such that $M_d(y) \succ 0$ has the product form property, then we say the pair (α, β) is a **congenital zero for d-moments**.

We can now state an intermediate result.

Theorem 1. *The pair α, β is a congenital zero for the d- moment problem if and only if the least common multiple X^η of X^α and X^β has degree bigger than d.*

The above result can be conveniently rephrased in terms the max operation defined for $\alpha, \beta \in \mathbb{N}^n$ by

$$\max(\alpha, \beta) := (\max(\alpha_j, \beta_j))_{j=1,\cdots,n}.$$

Set $\eta := \max(\alpha, \beta)$. Simple observations about this operation are:

(1) X^η is the least common multiple of X^α and X^β,
(2) X^α divides X^β iff $X^\beta = X^\eta$,
X^β divides X^α iff $X^\alpha = X^\eta$,
(3) $|\eta| = \sum_{j=1}^n \max(\alpha_j, \beta_j)$.

Thus, Theorem 1 asserts that the entry (α, β) does not correspond to a congenital zero in the matrix $M_d(y)^{-1}$ if and only if $\max(\alpha, \beta) \leq d$.

Later in Theorem 4 we show that this LCM (least common multiple) characterization of zeros in M_d^{-1} is equivalent to a highly triangular structure of orthonormal polynomials associated with product measures.

Example. In the case of $M_2^{-1}(y)$ in two variables X_1, X_2 we indicate below which entries $M_2(y)^{-1}(\alpha, \beta)$ with $|\beta| = 2$, are congenital zeroes. These (α, β) index the last three columns of $M_2(y)^{-1}$ and are

		X_1^2	X_1X_2	X_2^2
1	$\rightarrow$	$*$	$*$	$*$
X_1	$\rightarrow$	$*$	$*$	0
X_2	$\rightarrow$	0	$*$	$*$
X_1^2	$\rightarrow$	$*$	0	0
X_1X_2	$\rightarrow$	0	$*$	0
X_2^2	$\rightarrow$	0	0	$*$

Here $*$ means that the corresponding entry can be different from zero. Note each $*$ corresponds to X^α failing to divide X^β.

The proof relies on properties of orthogonal polynomials, so we begin by explaining in some detail the framework.

3.1. **Orthonormal polynomials.** A functional analytic viewpoint to polynomials is expeditious, so we begin with that. Let $s(d) := \binom{n+d}{d}$ be the dimension of vector space $\mathbb{R}_d[X]$. Let $\langle \ , \ \rangle_{\mathbb{R}^{s(d)}}$ denote the standard inner product on $\mathbb{R}^{s(d)}$. Let $f, h \in \mathbb{R}[X]$ be the polynomials $f(X) = \sum_{|\alpha|=0}^{s(d)} f_\alpha X^\alpha$ and $h(X) = \sum_{|\alpha|=0}^{s(d)} h_\alpha X^\alpha$. Then,

$$\langle f, h \rangle_y := \langle \mathbf{f}, M_d(y)\mathbf{h} \rangle_{\mathbb{R}^{s(d)}} \ = \ L_y(f(X)h(X)),$$

defines a *scalar product* in $\mathbb{R}_d[X]$, provided $M_d(y)$ is positive definite.

With a given $y = (y_\alpha)$ such that $M_d(y) \succ 0$, one may associate a unique family $(p_\alpha)_{|\alpha|=0}^{s(d)}$ of **orthonormal polynomials**. That is, the p_α's satisfy:

$$
\left\{\begin{array}{l}
p_\alpha \in \text{lin.span}\{X^\beta;\ \beta \leq_{gl} \alpha\}, \\
\langle p_\alpha, p_\beta\rangle_y \;=\; \delta_{\alpha\beta}, \qquad\quad |\alpha|, |\beta| \leq d, \\
\langle p_\alpha, X^\beta\rangle_y = 0,\ \text{if}\ \beta <_{gl} \alpha, \\
\langle p_\alpha, X^\alpha\rangle_y > 0,
\end{array}\right. \tag{3.2}
$$

Note $\langle p_\alpha, X^\beta\rangle_y = 0$, if $\beta \leq \alpha$ *and* $\alpha \neq \beta$, since the latter implies $\beta <_{gl} \alpha$.

Existence and uniqueness of such a family is guaranteed by the Gram-Schmidt orthonormalization process following the GLex order on the monomials, and by the positivity of the moment (covariance) matrix, see for instance [1, Theorem 3.1.11, p. 68].

Computation. Suppose that we want to compute the orthonormal polynomials p_σ for some index σ. Then proceed as follows: build up the sub-moment matrix $M^{(\sigma)}(y)$ with columns indexed by all monomials $\beta \leq_{gl} \sigma$, and rows indexed by all monomials $\alpha <_{gl} \sigma$. Hence, $M^{(\sigma)}(y)$ has one row less than columns. Next, complete $M^{(\sigma)}(y)$ with an additional last row described by $[M^{(\sigma)}(y)]_{\sigma,\beta} = X^\beta$, for all $\beta \leq_{gl} \sigma$. Then up to a normalizing constant, p_σ is nothing less than $\det(M^\sigma(y))$.

To see this, let $\gamma <_{gl} \sigma$. Then

$$\int X^\gamma p_\sigma(X)\, d\mu(X) \;=\; \det(B^\sigma)(y),$$

where the matrix $B^\sigma(y)$ is the same as $M^\sigma(y)$ except for the last row which is now the vector $(L_y(X^{\gamma+\alpha}))_{\alpha\leq_{gl}\sigma}$, already present in one of the rows above. Therefore, $\det(B^\sigma)(y) = 0$. Next, writing

$$p_\sigma(X) \;=\; \sum_{\beta\leq_{gl}\sigma} \rho_{\sigma\beta} X^\beta,$$

its coefficient $\rho_{\sigma\beta}$ is just (again up to a normalizing constant) the cofactor of the element $[M^\sigma(y)]_{1,\beta}$ in the square matrix $M^\sigma(y)$ with rows and columns both indexed with $\alpha \leq_{gl} \sigma$.

Further properties. Now we give further properties of the orthonormal polynomials. Consider first one variable polynomials. The orthogonal polynomials p_k have, by their very definition a "triangular" form, namely

$$p_k(X_1) := \sum_{\ell\leqslant k} \rho_{k\ell}\; X_1^\ell.$$

The orthonormal polynomials inherit the product form property of $M_r(y)$, assuming that the latter holds. Namely, each orthonormal polynomial p_α is a product

$$p_\alpha(X) \;=\; p_{\alpha_1}(X_1)p_{\alpha_2}(X_2)\cdots p_{\alpha_n}(X_n) \tag{3.3}$$

of orthogonal polynomials $p_{\alpha_j}(X_j)$ in one dimension. Indeed, by the product property

$$\langle p_{\alpha_1}(X_1)p_{\alpha_2}(X_2)\cdots p_{\alpha_n}(X_n), X^\beta\rangle_y = \prod_{j=1}^n \langle p_{\alpha_j}(X_j), X_j^{\beta_j}\rangle_y,$$

whence the product of single variable orthogonal polynomials satisfies all requirements listed in 3.2.

"Triangularity" in one variable and the product form property (3.3) forces p_α to have what we call a **fully triangular** form:

$$p_\alpha(X) := \sum_{\gamma \leqslant \alpha} \rho_{\alpha\gamma}\ X^\gamma, \quad |\alpha| \leq d. \tag{3.4}$$

Also note that for any $\gamma \leqslant \alpha$ there is a linear functional L_y of product type making $\rho_{\alpha\gamma}$ not zero. In order to construct such a functional we consider first the Laguerre polynomials in one variable $L_k^{(\sigma)}$, that is

$$L_k^{(\sigma)}(x) = \frac{e^x x^{-\sigma}}{k!}\frac{d^k}{dx^k}(e^{-x}x^{n+\sigma}) = \sum_{j=0}^k \begin{pmatrix} k+\sigma \\ k-j \end{pmatrix} \frac{(-x)^j}{j!}.$$

These polynomials satisfy an orthogonality condition on the semi-axis $\mathbb{R}_+$, with respect to the weight $e^{-x}x^\sigma$, see for instance [4]. Remark that the degree in σ of the coefficient of x^j is precisely $k - j$.

Thus the functional

$$L_y(p) = \int_{\mathbb{R}_+{}^n} p(x_1, ..., x_n)e^{-x_1-...-x_n}x_1^{\sigma_1}...x_n^{\sigma_n}dx_1...dx_n, \quad p \in \mathbb{R}[X],$$

will have the property, that is the associated (Laguerre$_\sigma$) orthogonal polynomials are $\Lambda_\alpha(X) = \prod_{j=1}^n L_{\alpha_j}^{(\sigma_j)}(X_j)$.

We formalize a simple observation as a lemma because we use it later.

Lemma 2. *The coefficients $\rho_{\alpha,\beta}$ in the decomposition*

$$\Lambda_\alpha(X) = \sum_{\beta \leq \alpha} \rho_{\alpha,\beta}X^\beta$$

are polynomials in $\sigma = (\sigma_1, ..., \sigma_n)$, viewed as independent variables, and the multi-degree of $\rho_{\alpha,\beta}(\sigma)$ is $\alpha - \beta$.

Note that for an appropriate choice of the parameters $\sigma_j,\ 1 \leq j \leq n$, the coefficients $\rho_{\alpha,\beta}$ in the decomposition

$$\Lambda_\alpha(X) = \sum_{\beta \leq \alpha} \rho_{\alpha,\beta}X^\beta$$

are linearly independent over the rational field, and hence non-null. To prove this evaluate σ on an n-tuple of algebraically independent transcendental real numbers over the rational field.

3.2. **Proof of Theorem 1.**

Proof. Let L_y be a linear functional for which $M_d(y) \succ 0$ and let (p_α) denote the family of orthogonal polynomials with respect to L_y. Orthogonality in (3.2) for expansions (3.4) reads

$$\delta_{\alpha\beta} = \langle p_\alpha, p_\beta \rangle_y = \sum_{\gamma \leqslant \alpha,\ \sigma \leqslant \beta} \rho_{\alpha\gamma}\rho_{\beta\sigma}\langle X^\gamma, X^\sigma\rangle_y.$$

In matrix notation this is just

$$I = D\, M_d(y)\, D^T$$

where D is the matrix $D = (\rho_{\alpha\gamma})_{|\alpha|,|\gamma| \leqslant d}$. Its columns are indexed (as before) by monomials arranged in GLex order, and likewise for its rows. That $\rho_{\alpha\gamma} = 0$ if $\gamma \not\leqslant \alpha$, implies that $\rho_{\alpha\gamma} = 0$ if $\alpha <_{gl} \gamma$ which says precisely that D is lower triangular. Moreover, its diagonal entries $\rho_{\beta\beta}$ are not 0, since p_β must have X^β as its highest order term. Because of this and triangularity, D is invertible. Write

$$M_d(y) = D^{-1}(D^T)^{-1} \quad \text{and} \quad M_d(y)^{-1} = D^T D.$$

Our goal is to determine which entries of $M_d(y)^{-1}$ are forced to be zero and we proceed by writing the formula $Z := M_d(y)^{-1} = D^T D$ as

$$z_{\alpha\beta} = \sum_{|\gamma| \leq d} \rho_{\gamma\alpha}\rho_{\gamma\beta} = \sum_{\beta \leqslant \gamma,\ \alpha \leqslant \gamma,\ |\gamma| \leq d} \rho_{\gamma\alpha}\rho_{\gamma\beta} = \sum_{\max(\alpha,\beta) \leqslant \gamma,\ |\gamma| \leq d} \rho_{\gamma\alpha}\rho_{\gamma\beta} \tag{3.5}$$

We emphasize (since it arises later) that this uses only the full triangularity of the orthogonal polynomials rather than that they are products of one variable polynomials. If full triangularity were replaced by triangularity w. r. to $<_{gl}$, then the first two equalities in (3.5) would be the same except that $\beta \leqslant \gamma,\ \alpha \leqslant \gamma,\ |\gamma| \leq d$ would be replaced by $\beta \leq_{gl} \gamma,\ \alpha \leq_{gl} \gamma,\ |\gamma| \leq d$.

To continue with our proof, consider (α, β) and set $\eta := \max(\alpha, \beta)$. If $|\max(\alpha, \beta)| > d$, then $z_{\alpha,\beta} = 0$, since the sum in equation (3.5) is empty. This is the forward side of Theorem 1.

Conversely, when $|\max(\alpha, \beta)| \leq d$ the entry $z_{\alpha\beta}$ is a sum of one or more products $\rho_{\gamma\alpha}\rho_{\gamma\beta}$ and so is a polynomial in σ. If this polynomial is not identically zero, then some value of σ makes $z_{\alpha\beta} \neq 0$, so (α, β) is not a congenital zero. Now we set out to show that $z_{\alpha\beta}$ as a polynomial in σ is not identically 0.

Lemma 2 tells us each product $\rho_{\gamma,\alpha}\rho_{\gamma,\beta}$ is a polynomial whose multi-degree in σ is exactly $2\gamma - \alpha - \beta$. The multi-index γ is subject to the constraints $\max(\alpha, \beta) \leq \gamma$ and $|\gamma| \leq d$. We fix an index, say $j = 1$ and choose

$$\hat{\gamma} = \max(\alpha, \beta) + (d - |\max(\alpha, \beta)|, 0, ..., 0).$$

Note the product $\rho_{\hat{\gamma}\alpha}\rho_{\hat{\gamma}\beta}$ is included in the sum (3.5) for $z_{\alpha,\beta}$ and it is a polynomial of degree $2d - 2|\max(\alpha,\beta)| + 2\max(\alpha_1,\beta_1) - \alpha_1 - \beta_1$ with respect to σ_1. By the extremality of our construction of $\hat{\gamma}$, every other term $\rho_{\gamma\alpha}\rho_{\gamma\beta}$ in $z_{\alpha\beta}$ will have smaller σ_1 degree. Hence $\rho_{\hat{\gamma}\alpha}\rho_{\hat{\gamma}\beta}$ can not be cancelled, proving that $z_{\alpha\beta}$ is not the zero polynomial. □

4. Partial independence

In this section we consider the case where only a **partial independence** property holds. We decompose the variables into disjoint sets $X = (X(1),...,X(k))$ where $X(1) = (X(1)_1,...,X(1)_{d_1})$, and so on. The linear functional L_y is said to satisfy a partial independence property (w.r. to the fixed grouping of variables), if

$$L_y(p_1(X(1))...p_k(X(k))) = \prod_{j=1}^{k} L_y(p_j(X(j)),$$

where p_j is a polynomial in the variables from the set $X(j)$, respectively.

Denote by $\deg_{X(j)} Q(X)$ the degree of a polynomial Q in the variables $X(j)$. Assuming that L_y is a positive functional, one can associate in a unique way the orthogonal polynomials $p_\alpha,\quad \alpha \in \mathbb{N}^n$. Note that the lexicographic order on $\mathbb{N}^n$ respects the grouping of variables, in the sense

$$(\alpha_1,...,\alpha_k) <_{gl} (\beta_1,...,\beta_k)$$

if and only if, either $\alpha_1 <_{gl} \beta_1$, or, if $\alpha_1 = \beta_1$, then, either $\alpha_2 <_{gl} \beta_2$, and so on.

Let $\alpha = (\alpha_1,...,\alpha_k)$ be a multi-index decomposed with respect to the groupings $X(1),...,X(k)$. Then, the uniqueness property of the orthonormal polynomials implies

$$p_\alpha(X) = p_{\alpha_1}(X(1))...p_{\alpha_k}(X(k)),$$

where $p_{\alpha_j}(X(j))$ are orthonormal polynomials depending solely on $X(j)$, and arranged in lexicographic order within this group of variables.

Theorem 3. *Let L_y be a positive linear functional satisfying a partial independence property with respect to the groups of variables $X = (X(1),...,X(k))$. Let $\alpha = (\alpha_1,...,\alpha_k), \beta = (\beta_1,...,\beta_k)$ be two multi-indices decomposed according to the fixed groups of variables, and satisfying $|\alpha|, |\beta| \le d$.*

Then the (α,β)-entry in the matrix $M_d(y)^{-1}$ is congenitally zero if and only if for every $\gamma = (\gamma_1,...,\gamma_k)$ satisfying $\gamma_j \ge_{gl} \alpha_j, \beta_j,\quad 1 \le j \le k$, we have $|\gamma| > d$.

The proof will repeat that of Theorem 1, with the only observation that

$$p_\alpha(X) = \sum_{\substack{1\le j\le k \\ \gamma_j \le_{gl} \alpha_j}} c_{\alpha,\gamma} X^\gamma.$$

5. Full triangularity

A second look at the proof of Theorem 1 reveals that the only property of the multivariate orthogonal polynomial we have used was the full triangularity form (3.4). In this section we provide an example of a non-product measure which has orthogonal polynomials in full triangular form, and, on the other hand, we prove that the zero pattern appearing in our main result, in the inverse moment matrices $M_r(y)^{-1}$, $r \leq d$, implies the full triangular form of the associated orthogonal polynomials. Therefore, *zeros in the inverse M_d^{-1} are coming from some triangularity property of orthogonal polynomials rather than from a product form of M_d.*

Example 1. We work in two real variables (x, y), with the measure $d\mu = (1 - x^2 - y^2)^t dxdy$, restricted to the unit disk $x^2 + y^2 < 1$, where $t > -1$ is a parameter.

Let $P_k(u; s)$ denote the orthonormalized Jacobi polynomials, that is the univariate orthogonal polynomials on the segment $[-1, 1]$, with respect to the measure $(1 - u^2)^s du$, with $s > -1$.

According to [6, Example 1, Chapter X], the orthonormal polynomials associated to the measure $d\mu$ on the unit disk are:

$$Q_{m+n,n}(x, y) = P_m(x, t + n + 1/2)(1 - x^2)^{n/2} P_n(\frac{y}{\sqrt{1 - x^2}}; t)$$

and

$$Q_{n,m+n}(x, y) = P_m(y, t + n + 1/2)(1 - y^2)^{n/2} P_n(\frac{x}{\sqrt{1 - y^2}}; t).$$

Remark that these polynomials have full triangular form, yet the generating measure is not a product of measures.

Theorem 4 (Full triangularity theorem)**.** *Let $y = (y_\alpha)_{\alpha \in \mathbb{N}^n}$ be a multi-sequence, such that the associated Hankel forms $M_d(y)$ are positive definite, where d is a fixed positive integer. Then the following holds.*

For every $r \leq d$, the (α, β)-entry in $M_r(y)^{-1}$ is 0 whenever $|\max(\alpha, \beta)| > r$ if and only if the associated orthogonal polynomials P_α, $|\alpha| \leq d$, have full triangular form.

Proof. The proof of the forward side is exactly the same as in the proof of Theorem 1. The converse proof begins by expanding each orthogonal polynomial p_α as

$$p_\alpha(X) := \sum_{\gamma \leq_{gl} \alpha} \rho_{\alpha\gamma}\, X^\gamma, \quad |\alpha| \leq d \tag{5.1}$$

where we emphasize that $\gamma \leq_{gl} \alpha$ is used as opposed to (3.4) and we want to prove that $\beta \leq_{gl} \alpha$ and $\beta \not\leq \alpha$ implies $\rho_{\alpha\beta} = 0$. Note that when $|\alpha| = d$ the inequality

$$\beta \not\leq \alpha \text{ is equivalent to } |\max(\alpha, \beta)| > d. \tag{5.2}$$

The first step is to use the zero locations of $M_d(y)^{-1}$ to prove that

$$\rho_{\alpha\beta} = 0 \ \text{ if } \ |\max(\alpha, \beta)| > d, \tag{5.3}$$

only for $|\alpha| = d$. Once this is established, then we apply the same argument to prove (5.2) for $|\alpha| = d'$ where d' is smaller.

If $n = 1$ there is nothing to prove. Assume $n > 1$ and decompose $\mathbb{N}^n$ as $\mathbb{N} \times \mathbb{N}^{n-1}$. The corresponding indices will be denoted by $(i, \alpha), i \in \mathbb{N}, \alpha \in \mathbb{N}^{n-1}$. We shall prove (5.3) by descending induction on the graded lexicographic order applied to the index (i, α) the following statement:

Let $i + |\alpha| = d$ and assume that $(j, \beta) <_{gl} (i, \alpha)$ and $(j, \beta) \not\leq (i, \alpha)$. Then $\rho_{(i,\alpha),(j,\beta)} = 0$.

The precise statement we shall use is equivalent (because of (5.2)) to:

The induction hypothesis: Suppose that

$$\rho_{(i',\alpha'),(j,\beta)} = 0 \ \text{ if } \ \max(i', j) + |\max(\alpha', \beta)| > d, \tag{5.4}$$

holds for all indices $(i', \alpha') >_{gl} (i, \alpha)$, with $i' + |\alpha'| = i + |\alpha| = d$.

We want to prove $\rho_{(i,\alpha),(j,\beta)} = 0$. Since we shall proceed by induction from the top, we assume that $i + |\alpha| = d$ that $(j, \beta) \not\leq (i, \alpha)$ that $(j, \beta) <_{gl} (i, \alpha)$ and let (i, α) denote the largest such w. r. to $<_{gl}$ order. Clearly, $i = d$. There is only one corresponding term in the graded lexicographic sequence of indices of length less than or equal to d, namely $(d, 0)$. We shall prove that for every $\beta \in \mathbb{N}^{n-1}$, $(j, \beta) \leq_{gl} (d, 0)$, $|\beta| > 0$, we have

$$\rho_{(d,0),(j,\beta)} = 0.$$

Since the corresponding entry in $M_d(y)^{-1}$, denoted henceforth as before by z_{**}:

$$z_{(d,0),(j,\beta)} = 0,$$

is zero by assumption, and because

$$z_{(d,0),(j,\beta)} = \rho_{(d,0),(d,0)}\rho_{(d,0),(j,\beta)}$$

we obtain $\rho_{(d,0),(j,\beta)} = 0$.

Now we turn to proving the main induction step. Assuming (5.4), we want to prove $\rho_{(i,\alpha),(j,\beta)} = 0$. Let $(j, \beta) \leq_{gl} (i, \alpha)$ subject to the condition $\max(i, j) + |\max(\alpha, \beta)| > d$. Then by hypothesis $0 = z_{(i,\alpha),(j,\beta)}$, so the GLex version of expansion (3.5) gives

$$0 = \rho_{(i,\alpha),(i,\alpha)}\rho_{(i,\alpha),(j,\beta)} + \sum_{i'=i,\ \alpha' >_{gl} \alpha,\ i'+|\alpha'|=d} \rho_{(i',\alpha'),(i,\alpha)}\rho_{(i',\alpha'),(j,\beta)}$$

$$+ \sum_{i'>i,\ i'+|\alpha'|=d} \rho_{(i',\alpha'),(i,\alpha)}\rho_{(i',\alpha'),(j,\beta)}.$$

We will prove that the two summations above are zero.

Indeed, if $i' > i$, then $\max(i,i') + |\max(\alpha',\alpha)| > i + |\alpha| = d$, and the induction hypothesis implies $\rho_{(i',\alpha'),(i,\alpha)} = 0$. Which eliminates the second sum.

Assume $i' = i$, so that $|\alpha| = |\alpha'|$. Then if $\max(\alpha,\alpha')$ equals to either α or α', we get $\alpha = \alpha'$, but this can not be, since $\alpha' >_{gl} \alpha$. Thus $i+|\max(\alpha,\alpha')| > d$ and the induction hypothesis yields in this case $\rho_{(i',\alpha'),(i,\alpha)} = 0$. Which eliminates the first sum.

In conclusion

$$\rho_{(i,\alpha),(i,\alpha)}\rho_{(i,\alpha),(j,\beta)} = 0,$$

which implies

$$\rho_{(i,\alpha),(j,\beta)} = 0,$$

as desired.

Our induction hypothesis is valid and we initialized it successfully, so we obtain the working hypothesis:

$$\rho_{(i,\alpha),(j,\beta)} = 0,$$

whenever $(j,\beta) <_{gl} (i,\alpha)$, $|(i,\alpha)| = d$, and $\max(\ (j,\beta),(i,\alpha)\ ) > d$. By (5.2) this is full triangularity under the assumption $i + |\alpha| = d$.

□

In contrast to the above full triangularity criteria, the product decomposition of a potential truncated moment sequence $(y_\alpha)_{\alpha_i \le d}$ can be decided by elementary linear algebra. Assume for simplicity that $n = 2$, and write the corresponding indices as $\alpha = (i,j) \in \mathbb{N}^2$. Then there are numerical sequences $(u_i)_{i\le d}$ and $(v_j)_{j\le d}$ with the property

$$y_{(i,j)} = u_i v_j, \quad 0 \le i,j \le d,$$

if and only if

$$\mathrm{rank}(y_{(i,j)})_{i,j=0}^d \le 1.$$

A similar rank condition, for a corresponding multilinear map, can be deduced for arbitrary n.

6. Conditional Triangularity

The aim of this section is to extend the full triangularity theorem of §5, in a more general context.

6.1. **Conditional triangularity.** In this new setting we consider two tuples of variables

$$X = (x_1, ..., x_n), \quad Y = (y_1, ..., y_m),$$

and we will impose on the concatenated tuple (X,Y) the full triangularity only with respect to the set of variables Y. The conclusion is as expected: this assumption will reflect the appearance of some zeros in the inverse M_d^{-1} of the associated truncated moment matrix. The proof below is quite similar to that of Theorem 4 and we indicate only sufficiently many details to make clear the differences.

We denote the set of indices by (α, β), with $\alpha \in \mathbb{N}^n$, $\beta \in \mathbb{N}^m$, equipped with the graded lexicographic order " $<_{gl}$ ". In addition, the set of indices $\beta \in \mathbb{N}^m$ which refers to the set of variables Y, is also equipped with the full graded order " $\leq$ ".

Let $y = (y_\alpha)_{\alpha \in \mathbb{N}^n}$ be a multi-sequence, such that the associated Hankel forms $M_d(y)$ are positive definite, where d is a positive integer. As before we denote:

$$p_{(\alpha,\beta)}(X, Y) = \sum_{(\alpha,\beta) \geq_{gl} (\alpha',\beta')} \rho_{(\alpha,\beta),(\alpha',\beta')} X^{\alpha'} Y^{\beta'}$$

the associated orthogonal polynomials, and by

$$z_{(\alpha,\beta),(\alpha',\beta')} = \sum_{(\gamma,\sigma) \geq_{gl} (\alpha,\beta),(\alpha',\beta')} \rho_{(\gamma,\sigma),(\alpha,\beta)} \rho_{(\gamma,\sigma),(\alpha',\beta')}$$

the entries in $M_d(y)^{-1}$.

Definition 2. *(Conditional triangularity)* The orthonormal polynomials $p_{\alpha,\beta} \in \mathbb{R}[X, Y]$, with $|\alpha + \beta| \leq 2d$, satisfy the *conditional triangularity with respect to* X, if when X is fixed and considered as a parameter, the resulting family denoted $\{p_{\alpha,\beta}|X\} \subset \mathbb{R}[Y]$ is in full triangular form with respect to the Y variables. More precisely, the following $(O)_d$ condition below holds.

$$(O)_d : \left[(\alpha', \beta') \leq_{gl} (\alpha, \beta),\ |(\alpha, \beta)| \leq d, \ \text{ and } \ \beta' \not\leq \beta \right] \Rightarrow \rho_{(\alpha,\beta),(\alpha',\beta')} = 0$$

Next, for a fixed degree $d \geq 1$ we will have to consider the following *zero in the inverse condition*

Definition 3. *(zero in the inverse condition $(V)_d$)*
Let $(\alpha', \beta') \leq_{gl} (\alpha, \beta)$ with $|(\alpha, \beta)| \leq d$, be fixed, arbitrary.
If $|(\gamma, \sigma)| > d$ whenever $(\gamma, \sigma) \geq_{gl} (\alpha, \beta), (\alpha', \beta')$ and $\sigma \geq \max(\beta, \beta')$), then $z_{(\alpha,\beta),(\alpha',\beta')} = 0$.

The main result of this paper asserts that both conditions $(V)_d$ and $(O)_d$ are in fact equivalent.

Theorem 5 (Conditional triangularity)**.** *Let $y = (y_\alpha)_{\alpha \in \mathbb{N}^n}$ be a multi-sequence and let d be an integer such that the associated Hankel form $M_d(y)$ is positive definite.*

Then the zero in the inverse condition $(V)_r$, $r \leq d$, holds if and only if $(O)_d$ holds, i.e., if and only if the orthonormal polynomials satisfy the conditional triangularity with respect to X.

Proof. Clearly, from its definition, $(O)_d$ implies $(O)_r$ for all $r \leq d$.
One direction is obvious: Let $r \leq d$ be fixed, arbitrary. If condition $(O)_r$ holds, then a pair $((\alpha', \beta') \leq_{gl} (\alpha, \beta))$ subject to the assumptions in $(V)_r$ will leave not a single term in the sum giving $z_{(\alpha,\beta),(\alpha',\beta')}$.

Conversely, assume that $(V)_d$ holds. We will prove the vanishing statement $(O)_d$ by descending induction with respect to the graded lexicographical order. To this aim, we label all indices (α,β), $|(\alpha,\beta)| = d$ in decreasing graded lexicographic order:

$$(\alpha_0,\beta_0) >_{gl} (\alpha_1,\beta_1) >_{gl} (\alpha_2,\beta_2) \ldots .$$

In particular $d = |\alpha_0| \geq |\alpha_1| \geq \ldots$ and $0 = |\beta_0| \leq |\beta_1| \leq \ldots$.

To initialize the induction, consider $(\alpha',\beta') \leq_{gl} (\alpha_0,\beta_0) = (\alpha_0,0)$, with $\beta' \nleq 0$, that is $\beta' \neq 0$. Then

$$0 = z_{(\alpha_0,\beta_0),(\alpha',\beta')} = \rho_{(\alpha_0,\beta_0),(\alpha_0,\beta_0)}\rho_{(\alpha_0,\beta_0),(\alpha',\beta')}.$$

Since the leading coefficient $\rho_{(\alpha_0,\beta_0),(\alpha_0,\beta_0)}$ in the orthogonal polynomial is non-zero, we infer

$$[(\alpha',\beta') <_{gl} (\alpha_0,\beta_0), \quad \beta' \nleq \beta_0] \;\Rightarrow \rho_{(\alpha_0,\beta_0),(\alpha',\beta')} = 0,$$

which is exactly condition $(O)_d$ applied to this particular choice of indices.

Assume that $(O)_d$ holds for all (α_j,β_j), $j < k$. Let $(\alpha',\beta') <_{gl} (\alpha_k,\beta_k)$ with $\beta' \nleq \beta_k$, that is, $|\max(\beta',\beta_k)| > |\beta_k|$. In view of (V_d),

$$z_{(\alpha_k,\beta_k),(\alpha',\beta')} = \rho_{(\alpha_k,\beta_k),(\alpha_k,\beta_k)}\rho_{(\alpha_k,\beta_k),(\alpha',\beta')} + \sum_{j=0}^{k-1} \rho_{(\alpha_j,\beta_j),(\alpha_k,\beta_k)}\rho_{(\alpha_j,\beta_j),(\alpha',\beta')}.$$

Note that the induction hypothesis implies that for every $0 \leq j \leq k-1$, at least one factor $\rho_{(\alpha_j,\beta_j),(\alpha_k,\beta_k)}$ or $\rho_{(\alpha_j,\beta_j),(\alpha',\beta')}$ vanishes. Thus,

$$0 = z_{(\alpha_k,\beta_k),(\alpha',\beta')} = \rho_{(\alpha_k,\beta_k),(\alpha_k,\beta_k)}\rho_{(\alpha_k,\beta_k),(\alpha',\beta')},$$

whence $\rho_{(\alpha_k,\beta_k),(\alpha',\beta')} = 0$.

Once we have exhausted by the above induction all indices of length d, we proceed similarly to those of length $d-1$, using now the zero in the inverse property $(V)_{d-1}$, and so on. □

Remark that the ordering (X,Y) with the tuple of full triangular variables Y on the second entry is important. A low degree example will be considered in the last section.

We call Theorem 5 the conditional triangularity theorem because when the variables X are *fixed* (and so can be considered as parameters), then the orthogonal polynomials now considered as elements of $\mathbb{R}[Y]$, are in full triangular form, rephrased as *in triangular form conditional to X is fixed.*

6.2. **The link with partial correlation.** Let us specialize to the case $d = 1$. Assume that the underlying joint distribution on the random vector $X = (X_1,\ldots,X_n)$ is *centered*, that is, $\int X_i \, d\mu = 0$ for all $i = 1,\ldots,n$; then M_d reads

$$M_d = \begin{bmatrix} 1 & | & 0 \\ - & & - \\ 0 & | & R \end{bmatrix} \quad \text{and} \quad M_d^{-1} = \begin{bmatrix} 1 & | & 0 \\ - & & - \\ 0 & | & R^{-1} \end{bmatrix}$$

where R is just the so-called *covariance* matrix. Partitionning the random vector in (Y, X_i, X_j) with $Y = (X_k)_{k \neq i,j}$, yields

$$R = \begin{bmatrix} \mathrm{var}\,(Y) & \mathrm{cov}\,(Y, X_i) & \mathrm{cov}\,(Y, X_j) \\ \mathrm{cov}\,(Y, X_i) & \mathrm{var}\,(X_i) & \mathrm{cov}\,(X_i, X_j) \\ \mathrm{cov}\,(Y, X_j) & \mathrm{cov}\,(X_i, X_j) & \mathrm{var}\,(X_j) \end{bmatrix},$$

where *var* and *cov* have obvious meanings. The *partial covariance* of X_i and X_j given Y, denoted $\mathrm{cov}\,(X_i, X_j \,|\, Y)$ in Wittaker [7, p. 135], satisfies:

$$\text{(6.1)} \quad \mathrm{cov}\,(X_i, X_j \,|\, Y) \;:=\; \mathrm{cov}\,(X_i, X_j) - \mathrm{cov}\,(Y, X_i)\mathrm{var}\,(Y)^{-1}\mathrm{cov}\,(Y, X_j).$$

After scaling R^{-1} to have a unit diagonal, the *partial correlation* between X_i and X_j (partialled on Y) is the negative of $\mathrm{cov}\,(X_i, X_j \,|\, Y)$, and as already mentioned, $R^{-1}(i, j) = 0$ if and only if X_i and X_j have *zero partial correlation*, i.e., $\mathrm{cov}\,(X_i, X_j \,|\, Y) = 0$. See e.g. Wittaker [7, Cor. 5.8.2 and 5.8.4].

Corollary 6. *Let $d = 1$. Then $R^{-1}(i, j) = 0$ if and only if the orthonormal polynomials of degree up to 2, associated with M_1, satisfy the conditional triangularity with respect to $X = (X_k)_{k \neq i,j}$.*

Proof. To recast the problem in the framework of §6.1, let $Y = (X_i, X_j)$ and rename $X := (X_k)_{k \neq i,j}$. In view of Definition 3 with $d = 1$, we only need consider pairs $(\alpha', \beta') \leq_{gl} (\alpha, \beta)$ with $\alpha = \alpha' = 0$ and $\beta' = (0, 1)$, $\beta = (1, 0)$. But then $\sigma \geq \max[\beta, \beta'] = (1, 1)$ implies $|(\gamma, \sigma)| \geq 2 > d$, and so as $R^{-1}(i, j) = 0$, the zero in the inverse condition $(V)_d$ holds. Equivalently, by Theorem 5, $(O)_d$ holds. □

Corollary 6 states that the pair (X_i, X_j) has zero partial correlation if and only if the orthonormal polynomials up to degree 2, satisfy the conditional triangularity with respect to $X = (X_k)_{k \neq i,j}$. That is, partial correlation and conditional triangularity are *equivalent*.

Example 2. Let $d = 1$, and consider the case of three random variables (X, Y, Z) with (centered) joint distribution μ. Then suppose that the orthonormal polynomials up to degree $d = 1$ satisfy the conditional triangularity property $(V)_1$ w.r.t. X. That is, $p_{000} = 1$ and

$$p_{100} = \alpha_1 + \beta_1 X;\; p_{010} = \alpha_2 + \beta_2 X + \gamma_2 Y;\; p_{001} = \alpha_3 + \beta_3 X + \gamma_3 Z,$$

for some coefficients $(\alpha_i \beta_i \gamma_i)$. Notice that because of $(O)_1$, we cannot have a linear term in Y in p_{001}. Orthogonality yields that $\langle X^\gamma, p_\alpha \rangle = 0$ for all $\gamma <_{gl} \alpha$, i.e., with $\mathbf{E}$ being the expectation w.r.t. μ,

$$\begin{aligned} \beta_2 \mathbf{E}(X^2) + \gamma_2 \mathbf{E}(XY) &= 0 \\ \beta_3 \mathbf{E}(X^2) + \gamma_3 \mathbf{E}(XZ) &= 0 \\ \beta_3 \mathbf{E}(XY) + \gamma_3 \mathbf{E}(YZ) &= 0. \end{aligned}$$

Stating that the determinant of the last two linear equations in (β_3, γ_3) is zero yields

$$\mathbf{E}(YZ) - \mathbf{E}(X, Y)\mathbf{E}(X^2)^{-1}\mathbf{E}(X, Z) \;=\; 0,$$

which is just (6.1), i.e., the zero partial correlation condition, up to a multiplicative constant.

This immediately raises two questions:

(i) What are the distributions for which the orthonormal polynomials up to degree 2 satisfy the conditional triangularity with respect to a given pair (X_i, X_j)?

(ii) Among such distributions, what are those for which conditional independence with respect to $X = (X_{k \neq i,j})$ also holds?

An answer to the latter would characterize distributions for which zero partial correlation imply conditional independence (like for the normal distribution).

6.3. **Conditional independence and zeros in the inverse.** We have already mentioned that in general, conditional independence is *not* detected from zero entries in the inverse of R^{-1} (equivalently, M_1^{-1}), except for the normal joint distribution, a common assumption in Graphical Gaussian Models. Therefore, a natural question of potential interest is to search for conditions on when conditional independence in the non gaussian case is related to the *zero in the inverse* property $(V)_d$, or equivalently, the conditional triangularity $(O)_d$, not only for $d = 1$ but also for $d > 1$.

A rather negative result in this direction is as follows. Let d be fixed, arbitrary, and let $M_d = (y_{ijk}) \succ 0$ be the moment matrix of an arbitrary joint distribution μ of three random variables (X, Y_1, Y_2) on $\mathbb{R}$. As we are considering only finitely many moments (up to order $2d$), by Tchakaloff's theorem there exists a measure φ finitely supported on, say s, points $(x^{(l)}, y_1^{(l)}, y_2^{(l)}) \subset \mathbb{R}^3$ (with associated probabilities $\{p_l\}$), $l = 1, \ldots, s$, and whose all moments up to order $2d$ match those of μ; see e.g. Reznick [5, Theor. 7.18].

Let us define a sequence $\{\varphi_t\}$ of probability measures as follows. Perturbate each point $(x^{(l)}, y_1^{(l)}, y_2^{(l)})$ to $(x^{(l)} + \epsilon(t,l), y_1^{(l)}, y_2^{(l)})$, $l = 1, \ldots, s$, in such a way that no two points $x^{(l)} + \epsilon(t,l)$ are the same, and keep the same weights $\{p_l\}$. It is clear that φ_t satisfies:

$$\begin{aligned} 1 &= \text{Prob}[Y = (y_1^{(l)}, y_2^{(l)}) \,|\, X = x^{(l)} + \epsilon(t,l)] \\ &= \text{Prob}[Y_1 = y_1^{(l)} \,|\, X = x^{(l)} + \epsilon(t,l)]\, \text{Prob}[Y_2 = y_2^{(l)} \,|\, X = x^{(l)} + \epsilon(t,l)] \end{aligned}$$

for all $l = 1, \ldots, s$. That is, conditional to X, the variables Y_1 and Y_2 are independent. Take a sequence with $\epsilon(t,l) \to 0$ for all l, as $t \to \infty$, and consider the moment matrix $M_d^{(t)}$ associated with φ_t. Clearly, as $t \to \infty$,

$$\int X^i Y_1^j Y_2^k \, d\varphi_t \;\to\; \int X^i Y_1^j Y_2^k \, d\mu \;=\; y_{ijk}, \qquad \forall\, i,j,k : \; i+j+k \leq 2d,$$

i.e., $M_d^{(t)} \to M_d$.

Therefore, if the zero in the inverse property $(V)_d$ does not hold for M_d, then by a simple continuity argument, it cannot hold for any $M_d^{(t)}$ with sufficiently large t, and still the conditional independence property holds for each φ_t. One has just shown that for every fixed d, one may easily construct examples of measures with the conditional independence property, which violate the zero in the inverse property $(V)_d$.

References

[1] C.F. Dunkl, Yuan Xu. *Orthogonal Polynomials of Several Variables*, Cambridge University Press, Cambridge, 2001.

[2] C. Berg. Fibonacci numbers and orthogonal polynomials. Preprint.

[3] E. Cripps, C. Carter, R. Kohn. Variable selection and covariance selection in multiregression models. Technical report #2003-14, Statistical and Applied Mathematical Sciences Institute, Research Triangle Park, NC, 2003.

[4] W. Magnus, F. Oberhettinger, *Formulas and Theorems for the Functions of Mathematical Physics*, Chelsea reprint, New York, 1949.

[5] B. Reznick. *Sums od Even Powers of real Linear Forms*, Amer. Math. Soc. Memoirs vol. 96, Amer. Math. Soc. Providence, R.I., 1992.

[6] P. K. Suetin, *Orthogonal polynomials in two variables* (in Russian), Nauka, Moscow, 1988.

[7] J. Wittaker. *Graphical Models in Applied Mathematical Analysis*, Wiley, New York, 1990.