Chapter 9

TENSOR DISPLACEMENT STRUCTURES AND POLYSPECTRAL MATCHING

Victor S. Grigorascu

Phillip A. Regalia

9.1 INTRODUCTION

This chapter studies the extension of the notion of structured matrices to tensors. These are multi-indexed arrays, in contrast to matrices, which are two-indexed arrays. Such arrays arise while considering higher-order cumulants and the corresponding polyspectra in applications, particularly in blind model identification and approximation problems.

While matrices are adequate representations for second-order statistics, higher-order cumulants are more naturally (and more completely) studied in a tensor setting. In this chapter, we examine the displacement rank concept of Ch. 1 for tensors. After a semitutorial presentation of Tucker products and cumulant representations of linear systems, we show links between interpolation of polyspectral values by a linear model and the Tucker factorability of a certain Pick tensor. We also develop a particular higher-order extension of a Schur-type algorithm, based on a novel outer product of tensors. This leads to a pyramidal factorization approach for tensors, which specializes to triangular factorization in the matrix case.

9.2 MOTIVATION FOR HIGHER-ORDER CUMULANTS

Recent years have witnessed increasing interest in higher-order cumulants, which convey more information about an underlying stochastic process than second-order statistics, including non-Gaussianity, phase information, nonlinearities, and so forth.

An example arises in the blind identification problem, in which one considers a process $\{y(\cdot)\}$ generated by

$$y(n) = h_1\, u(n-1) + h_2\, u(n-2) + h_2\, u(n-3) + \cdots = \sum_{i=1}^{\infty} h_k\, u(n-i) \ ,$$

where $\{y(\cdot)\}$ is observable; $\{u(\cdot)\}$ is an unobserved but independent, identicially distributed (i.i.d.) stochastic process; and $\{h_i\}$ denote the impulse responses of an unknown channel. The transfer function associated with the channel is

$$H(z) = \sum_{i=1}^{\infty} h_i z^i, \qquad |z| < 1,$$

where in this chapter we are using z (instead of z^{-1}) to denote the delay operator, viz., $z[u(n)] = u(n-1)$. A basic problem in this setting is to estimate the impulse response $\{h_i\}$ or the transfer function $H(z)$, given the output sequence $\{y(\cdot)\}$.

Second-order statistics of the output process $\{y(\cdot)\}$ allow one to determine the magnitude $|H(e^{j\omega})|$ of the channel but not its phase, whereas higher-order statistics of the output process allow one to deduce both the magnitude and the phase of the channel frequency response. This added informational content served as the impetus for a revived interest in higher-order statistics in signal processing in the early 1990s [Men91].

Despite numerous intriguing developments in this field, including the ability to separate minimum-phase from nonminimum-phase components of a signal [GS90] and linear from nonlinear components [TE89], or the ability to locate more sources than sensors in an array processing context [Car90], practical interest in algorithm development for signal processing applications has dwindled rapidly. This phenomenon may be attributed to two basic obstacles underlying cumulant-based signal processing.

The first concerns the computational complexity of estimating cumulants from a given time series. Although empirical estimation formulas are available, they tend to be computationally expensive, and in some cases they converge more slowly than estimators for second-order statistics.

The second obstacle concerns the successful extraction of desired information from higher-order statistics. Although such statistics often are touted for carrying phase information, they also convey information on potential nonlinear mechanisms. This is problematic in applications where the underlying process $\{y(\cdot)\}$ *is* linear, since any estimation errors can result in higher-order (estimated) statistics that are suddenly incompatible with *any* linear process. (By a "linear process," we mean the output of a linear time-invariant system when driven by an i.i.d. process.)

To appreciate this problem further, let us turn momentarily to second-order statistics. Suppose $\{y(\cdot)\}$ is a real-valued wide-sense stationary stochastic process. We introduce a finite number of autocorrelation lags:

$$r_i = E[y(n)\, y(n-i)], \qquad i = 0, 1, \ldots, M.$$

Under the mild constraint that the Toeplitz matrix

$$R = \begin{bmatrix} r_0 & r_1 & \cdots & r_M \\ r_1 & r_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_1 \\ r_M & \cdots & r_1 & r_0 \end{bmatrix}$$

be positive definite, the familiar Yule–Walker equations, in which the unknowns $\{\sigma^2, a_1, \ldots, a_M\}$ are obtained according to

$$R \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_M \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

allow one to deduce a candidate linear model for the data $\{y(\cdot)\}$. In particular, choosing

$$\widehat{H}(z) = \frac{\sigma}{1 + a_1 z + \cdots + a_M z^M}$$

yields a stable autoregressive transfer function such that, when driven by unit-variance white noise, the resulting output sequence $\{\hat{y}(\cdot)\}$ is compatible with the given second-order statistics, i.e.,

$$E[\hat{y}(n)\, \hat{y}(n-i)] = r_i, \qquad i = 0, 1, \ldots, M.$$

Whether the initial process $\{y(\cdot)\}$ is autoregressive, or even linear, is irrelevant to the validity of this result.

Many attempts to generalize such relations to higher-order cumulants may be found in [GM89], [GS90], [JK92], [NM93], [SM90a], [SM90b], under the hypothesis that the underlying (non-Gaussian) process $\{y(\cdot)\}$ is linear and generated from a rational transfer function of known degree. In cases where the process *is* linear, but the degree of the underlying model is underestimated, the equations so solved do not in general lead to a model that replicates the cumulant values used for its determination. The incompatibility between the resulting model and the cumulant values used to determine it implies that such methods do not correctly capture the structure of the data.

One of the few results establishing compatibility of higher-order cumulants with a linear process is given by Tekalp and Erdem [TE89]. Introduce the kth-order cumulant lags of a process $\{y(\cdot)\}$ as

$$c_{i_1, i_2, \ldots, i_{k-1}} \triangleq \operatorname{cum}[y(n-i_1), y(n-i_2), \ldots, y(n-i_{k-1}), y(n)],$$

where $\operatorname{cum}[\cdots]$ denotes the cumulant value of the k random variables that form its argument. (A nice tutorial overview of cumulants in signal processing may be found in [Men91].) Since our process $\{y(\cdot)\}$ is assumed stationary, the cumulant value here depends only on the relative lags $i_1, \ldots, i_{k-1}$. The kth-order polyspectrum of the process $\{y(\cdot)\}$ is a $(k-1)$-dimensional z-transform of the sequence $\{c_{i_1, \ldots, i_{k-1}}\}$, defined as [Men91]

$$R(z_1, \ldots, z_{k-1}) \triangleq \sum_{i_1=-\infty}^{+\infty} \cdots \sum_{i_{k-1}=-\infty}^{+\infty} c_{i_1, \ldots, i_{k-1}}\, z_1^{i_1} \cdots z_{k-1}^{i_{k-1}},$$

whenever the infinite sum converges on the unit polycircle $|z_1| = \cdots = |z_{k-1}| = 1$. A well-known relation [Men91], [NP93] shows that whenever $\{y(\cdot)\}$ is the output process of a linear system with transfer function $H(z)$, which in turn is driven by an i.i.d. sequence, the polyspectrum assumes the form

$$R(z_1, \ldots, z_{k-1}) = \gamma_k \cdot H(z_1) \cdots H(z_{k-1}) H((z_1 \cdots z_{k-1})^{-1}),$$

where γ_k is the kth-order cumulant of the i.i.d. input sequence (assumed nonzero).

The polycepstrum is defined as the logarithm [NP93], [TE89] of the polyspectrum which, for the linear case, gives the separable structure

$$\log[R(z_1, \ldots, z_{k-1})] = \log \gamma_k + \log H(z_1) + \cdots + \log H(z_{k-1}) + \log H((z_1 \cdots z_{k-1})^{-1}).$$

Assuming $H(z)$ has no poles or zeros on the unit circle $|z| = 1$, we may develop a multidimensional z-transform expansion of the polycepstrum as

$$\log[R(z_1, \ldots, z_k)] = \sum_{i_1=-\infty}^{+\infty} \cdots \sum_{i_{k-1}=-\infty}^{+\infty} \hat{c}_{i_1, \ldots, i_{k-1}}\, z_1^{i_1} \cdots z_{k-1}^{i_{k-1}},$$

in which the terms $\{\hat{c}_{i_1,\ldots,i_{k-1}}\}$ are the polycepstral coefficients. The process $\{y(\cdot)\}$ is then linear if and only if the polycepstral coefficients are nonzero only on the principal axes (only one index nonzero) and the main diagonal ($i_1 = i_2 = \cdots = i_{k-1}$), with elementary symmetry relations connecting the nonzero coefficients [TE89]. Since this result involves infinitely many cumulant lags, its practical application is limited to cases where the cumulant lags are finite in duration or decay sufficiently rapidly in all indices as to render truncation effects negligible [NP93].

Similar in spirit to the cumulant matching approach of [Tug87], [Tug95], a generic problem statement that motivates the present work is the following: Given a finite number of cumulant lags, or possibly a finite number of evaluations of a polyspectrum, under what conditions can a linear process be fit to such values?

For second-order statistics, this problem is solved. Various formulations are possible, including the Yule–Walker equations, the closely connected Levinson recursion, the Kalman–Yakubovich–Popov lemma (e.g., [FCG79]), and deeper approaches connected with the Schur algorithm [DD84], Darlington synthesis [DVK78], and interpolation problems among the class of Schur functions [Dym89a]. Many of these approaches admit matrix analogues by way of matrix displacement structure theory [KS95a], and we examine candidate extensions of displacement structure relations to higher-order cumulants.

Many algorithmic contributions in recent years aim to manipulate cumulant information by way of basic matrix algebra. Since a matrix is a two-indexed structure, while higher-order cumulants involve more than two indices, a tensorial formulation for cumulants, where tensor here refers simply to a multi-indexed array, would seem a more natural setting for capturing cumulant-based structures [McC87].

Some recent works have reinforced the utility of tensorial representations of higher-order statistics. For example, Delathauwer, DeMoor, and Vandewalle [DMV99] have developed a multilinear singular value decomposition, in which a kth-order tensor is reindexed into k different "matrix unwindings," each of whose left singular vectors may be computed. The overall scheme is then equivalent to applying k unitary transformations (one for each index dimension) to the tensor to expose a core tensor—not, in general, diagonal—verifying certain norm and orthogonality properties.

Cardoso and Comon [CC96a] and Comon and Mourrain [CM96] have shown the role of independent component analysis in many signal processing problems, particularly source separation. This notion specializes to principal component analysis when applied to second-order statistics.

Cardoso [Car90], [Car95] has developed supersymmetric tensor diagonalization, motivated by earlier work involving quadricovariance structures defined from fourth-order cumulants. Just as a second-order tensor (or matrix) can be understood as an operator between first-order tensor (or vector) spaces, a fourth-order cumulant structure may be treated as an operator between matrix spaces, leading to many fruitful extensions of eigendecompositions familiar in matrix theory.

Our approach aims to exploit displacement structure in a multi-indexed setting, with the chapter organized as follows. Sec. 9.3 presents a brief overview of displacement structure in second-order statistical modeling, so that various higher-order extensions may appear more recognizable. Sec. 9.4 then presents a tutorial overview of a particular multilinear matrix product (sometimes called the Tucker product), followed by its relation to cumulant representations in system theory. Sec. 9.6 then introduces displacement structure for cumulant tensors, along with relations connecting displacement residues with polyspectra; the relations so studied are valid for all cumulant orders. From these relations we show in Sec. 9.7 that the existence of a linear model compatible with a given set of cumulant or polyspectral values implies the Tucker-factorability of a certain Pick tensor defined from the data. Sec. 9.8 then presents a candidate extension of a Schur algorithm to higher-order tensors, based on an apparently novel outer product involving tensors of successive degrees.

Concluding remarks are made in Sec. 9.9, including some open problems that arise throughout our presentation.

9.3 SECOND-ORDER DISPLACEMENT STRUCTURE

We present a brief review of displacement structure in second-order stochastic modeling to motivate subsequent extensions to higher-order arrays. Further details can be found in [KS95a] and the references therein as well as in Ch. 1 of this book.

Consider a wide-sense real stationary time series $\{y(\cdot)\}$ with autocorrelation coefficients

$$E[y(n)\,y(n-k)] = r_k$$

and the corresponding autocorrelation matrix

$$R = E\left\{\begin{bmatrix} y(n) \\ y(n-1) \\ y(n-2) \\ \vdots \end{bmatrix} \begin{bmatrix} y(n) & y(n-1) & y(n-2) & \cdots \end{bmatrix}\right\} = \begin{bmatrix} r_0 & r_1 & r_2 & \cdots \\ r_1 & r_0 & r_1 & \ddots \\ r_2 & r_1 & r_0 & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix}$$

of infinite dimensions for now, which assumes a celebrated Toeplitz structure.

Let Z denote the shift matrix with ones on the subdiagonal and zeros elsewhere. The matrix ZRZ^T relates to R by shifting all elements one position diagonally; the Toeplitz structure of R implies that the displacement residue

$$R - ZRZ^T = \begin{bmatrix} r_0 & r_1 & r_2 & \cdots \\ r_1 & 0 & 0 & \cdots \\ r_2 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{9.3.1}$$

vanishes except along the borders of the matrix.

Consider now the two-variable (generating function) form

$$\begin{bmatrix} 1 & z_1 & z_1^2 & \cdots \end{bmatrix} \begin{bmatrix} r_0 & r_1 & r_2 & \cdots \\ r_1 & r_0 & r_1 & \ddots \\ r_2 & r_1 & r_0 & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix} \begin{bmatrix} 1 \\ z_2 \\ z_2^2 \\ \vdots \end{bmatrix} = \sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} r_{|i_1-i_2|} z_1^{i_1} z_2^{i_2}$$

$$\triangleq R(z_1, z_2).$$

By way of the displacement residue equation (9.3.1), we see that the function $R(z_1, z_2)$ satisfies

$$(1 - z_1 z_2)\, R(z_1, z_2) = R_+(z_1) + r_0 + R_+(z_2), \tag{9.3.2}$$

where

$$R_+(z) \triangleq \sum_{i=1}^{\infty} r_i\, z^i.$$

Setting $z_2 = z_1^{-1}$ gives the power spectral density function

$$R_+(z_1) + r_0 + R_+(z_1^{-1}) = \sum_{i=-\infty}^{\infty} r_{|i|}\, z_1^i, \qquad |z_1| = 1.$$

Since this function along the unit circle $|z_1| = 1$ is simply the real part of $r_0 + 2R_+(z_1)$, the positivity of the power spectrum at (almost) all points on the unit circle reveals $r_0 + 2R_+(z_1)$ as a positive real function, i.e., one that may be continued analytically to all points in $|z_1| < 1$ with positive real part.

The spectral factorization problem may be advantageously treated by considering a dyadic decomposition of the displacement residue from (9.3.1), namely,

$$\begin{bmatrix} r_0 & r_1 & r_2 & \cdots \\ r_1 & 0 & 0 & \cdots \\ r_2 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} = \underbrace{\begin{bmatrix} \sqrt{r_0} \\ r_1/\sqrt{r_0} \\ r_2/\sqrt{r_0} \\ \vdots \end{bmatrix}}_{\triangleq a} [\cdot]^T - \underbrace{\begin{bmatrix} 0 \\ r_1/\sqrt{r_0} \\ r_2/\sqrt{r_0} \\ \vdots \end{bmatrix}}_{\triangleq b} [\cdot]^T, \tag{9.3.3}$$

in which $[\cdot]$ means "repeat the previous vector." The two-variable form (9.3.2) induced by the displacement residue (9.3.1) may then be rewritten as

$$(1 - z_1 z_2)\, R(z_1, z_2) = a(z_1)\, a(z_2) - b(z_1)\, b(z_2),$$

in which

$$b(z) = R_+(z)/\sqrt{r_0}, \qquad a(z) = \sqrt{r_0} + b(z). \tag{9.3.4}$$

According to a celebrated result, whose origins go back to the contributions of Toeplitz, Carathéodory, and Schur, the Toeplitz matrix R that induces $R(z_1, z_2)$ is positive (semi-) definite if and only if there exists a Schur function $S(z)$ (meaning that $S(z)$ is analytic in $|z| < 1$ and strictly bounded by unit magnitude there) that maps $a(z)$ into $b(z)$:

$$b(z) = S(z)\, a(z).$$

By way of (9.3.4), we see that $S(z)$ must relate to the positive real function $r_0 + 2R_+(z)$ according to

$$r_0 + 2R_+(z) = r_0\, \frac{1 + S(z)}{1 - S(z)},$$

which is simply the Cayley transform. By a well-known property of this transform, $S(z)$ will indeed be a Schur function if and only if $r_0 + 2R_+(z)$ is a positive real function.

The virtue of this approach is best appreciated if we consider the case in which only partial information on the power spectrum is available. To this end, suppose we know (or have estimated) r_0 as well as $R_+(z)$ at N distinct points $z = \lambda_1, \ldots, \lambda_N$ inside the unit disk $0 < |z| < 1$. With the convention $\lambda_0 = 0$, this then determines, again by way of (9.3.4), the value pairs

$$\begin{aligned} [a(\lambda_0), b(\lambda_0)] &= [\sqrt{r_0}, 0], \\ [a(\lambda_i), b(\lambda_i)] &= [\sqrt{r_0} + R_+(\lambda_i)/\sqrt{r_0}, R_+(\lambda_i)/\sqrt{r_0}], \quad i = 1, 2, \ldots, N. \end{aligned}$$

There then exists a Schur function $S(z)$ fulfilling the system of equations

$$b(\lambda_i) = S(\lambda_i)\, a(\lambda_i), \qquad i = 0, 1, \ldots, N,$$

if and only if a certain Pick matrix P, written elementwise as

$$P_{i_1,i_2} = \frac{a(\lambda_{i_1})\,a(\lambda^*_{i_2}) - b(\lambda_{i_1})\,b(\lambda^*_{i_2})}{1-\lambda_{i_1}\lambda^*_{i_2}} = R(\lambda_{i_1},\lambda^*_{i_2}), \tag{9.3.5}$$

is positive (semi-) definite. If positive semidefinite and singular, then $S(z)$ becomes a rational allpass function of degree equal to the rank of P. If positive definite, infinitely many solutions exist; they may be parametrized by various recursive constructive procedures (e.g., [DVK78], [Dym89a], [KS95a], [SKLC94]).

The recursive procedures for constructing $S(z)$ so cited also place in evidence a complementary Schur function, call it $Q(z)$, fulfilling

$$S(z)\,S(z^{-1}) + Q(z)\,Q(z^{-1}) = 1.$$

When P is positive definite, the solution set for $S(z)$ always includes choices fulfilling the constraint $1 - S(z) \neq 0$ for all $|z| = 1$. In this case, $S(z)$ and its complement $Q(z)$ yield

$$\widehat{H}(z) = \sqrt{r_0}\,\frac{Q(z)}{1-S(z)}$$

as a stable and causal function, providing a candidate model for the correlation data $\{r_0, R_+(\lambda_i)\}$. This means that if the system $\widehat{H}(z)$ is driven by unit-variance white noise, its output sequence $\{\hat{y}(\cdot)\}$ fulfills the correlation matching properties

$$\begin{aligned} E[\hat{y}^2(n)] &= r_0, \\ \sum_{k=1}^{\infty} E[\hat{y}(n)\,\hat{y}(n-k)]\,\lambda_i^k &= R_+(\lambda_i), \quad i = 1,2,\ldots,N. \end{aligned}$$

A higher-order extension of this problem will be addressed in Sec. 9.7.

9.4 TUCKER PRODUCT AND CUMULANT TENSORS

Second-order cumulants reduce to conventional second-order statistics [Men91], i.e., $\mathrm{cum}[y(n-i_1), y(n-i_2)] = E[y(n-i_1)\,y(n-i_2)]$ with $E[\cdot]$ the expectation operator. This quantity depends on only two indices i_1 and i_2 such that calculations involving second-order cumulants reduce to basic operations on two-indexed arrays (i.e., matrices). Because higher-order cumulants are multi-indexed quantities, they may be profitably treated using tools of multi-indexed arrays, or tensors. Cumulants are also multilinear functions of their arguments [Men91], so it is useful to introduce some basic concepts of multilinear algebra applied to multi-indexed arrays. A particularly useful tool in this regard, to be reviewed in this section, is a multilinear matrix product called the Tucker product, in view of its early application to three-mode factor analysis in [Tuc64], [Tuc66]. Illustrations of its utility in cumulant analysis of system theory are included as well. For notational convenience, all vectors, matrices, and tensors will be indexed starting from zero rather than one.

Consider k matrices $\{A_i\}_{i=1}^k$, each of dimensions $M_i \times L$. A kth-order tensor $\mathcal{D}$, of dimensions $M_1 \times M_2 \times \cdots \times M_k$, may be defined from a Tucker product as

$$\begin{aligned} \mathcal{D}_{i_1,i_2,\ldots,i_k} &= \sum_{l=0}^{L-1} (A_1)_{i_1,l}(A_2)_{i_2,l}\cdots(A_k)_{i_k,l} \\ &\triangleq A_1 \star A_2 \star \cdots \star A_k. \end{aligned}$$

Example 1. Suppose each matrix A_i reduces to a column vector. The summation in the above definition becomes superfluous, and $\mathcal{D}$ may be written elementwise as

$$\mathcal{D}_{i_1,i_2,\dots,i_k} = (A_1)_{i_1}(A_2)_{i_2}\cdots(A_k)_{i_k}.$$

The reader may think of this as a kth-order "outer product" since, if we consider the case $k=2$, the tensor $\mathcal{D}_{i_1,i_2}$ becomes a matrix and the relation $\mathcal{D}_{i_1,i_2} = (A_1)_{i_1}(A_2)_{i_2}$ implies that the matrix $\mathcal{D}$ is the outer product $A_1A_2^T$.

Example 2. Suppose instead that each matrix A_i is a row vector. The above definition yields the scalar

$$\mathcal{D}_{0,0,\dots,0} = \sum_{l=0}^{L-1}(A_1)_l(A_2)_l\cdots(A_k)_l.$$

The reader may think of this as a kth-order "inner product" since, for the case $k=2$, we recognize the sum as the the standard inner product of two real vectors A_1 and A_2.

◊

When all the factors A_i are matrices, the resulting tensor $\mathcal{D}$ may be considered as an array collecting all possible kth-order inner products of the rows of each factor or as the sum of L higher-order outer products.

Example 3. Consider $k=3$ matrices A, B, and C, each having $L=2$ columns. Partition these matrices columnwise as

$$A = [a_1 \ \ a_2], \quad B = [b_1 \ \ b_2], \quad C = [c_1 \ \ c_2].$$

Their third-order Tucker product may be written as

$$\mathcal{D} = A\star B\star C = (a_1\star b_1\star c_1) + (a_2\star b_2\star c_2)$$

involving $L=2$ vector outer products.

◊

One can also consider a weighted version, using a kth-order tensor $\mathcal{T}$ as a kernel. The $\mathcal{T}$-product of matrices A_i, $i=1,\dots,k$, is the kth-order tensor

$$\begin{aligned}\mathcal{D}_{i_1,i_2,\dots i_k} &= \sum_{l_1=0}^{L_1-1}\sum_{l_2=0}^{L_2-1}\cdots\sum_{l_k=0}^{L_k-1}(A_1)_{i_1,l_1}(A_2)_{i_2,l_2}\cdots(A_k)_{i_k,l_k}(\mathcal{T})_{l_1,l_2,\dots,l_k},\\ &\triangleq A_1 \overset{\mathcal{T}}{\star} A_2 \overset{\mathcal{T}}{\star}\cdots\overset{\mathcal{T}}{\star} A_k,\end{aligned}$$

where all dimensions are assumed compatible. One may check that if $\mathcal{T}$ is the identity tensor $[\mathcal{T}_{i_1,\dots,i_k} = \delta(i_1,\dots,i_k)]$, the weighted Tucker product reduces to the standard Tucker product.

Example 4. Suppose we are given three column vectors v, w, x containing random variables, whose third-order cross cumulants are collected into the tensor $\mathcal{T}$:

$$\mathcal{T}_{i_1,i_2,i_3} = \text{cum}[v_{i_1}, w_{i_2}, x_{i_3}].$$

Suppose each vector undergoes a linear transformation, using matrices A, B, and C:

$$v' = Av, \qquad w' = Bw, \qquad x' = Cx.$$

Let $\mathcal{D}$ be the new third-order cross-cumulant tensor, i.e.,

$$\mathcal{D}_{i_1,i_2,i_3} = \text{cum}[v'_{i_1}, w'_{i_2}, x'_{i_3}].$$

Cumulants are multilinear functions of their arguments, and the new tensor $\mathcal{D}$ relates to the old one $\mathcal{T}$ by the multilinear transformation

$$\mathcal{D} = A \overset{\mathcal{T}}{\star} B \overset{\mathcal{T}}{\star} C,$$

using the weighted Tucker product.

$\diamond$

Some further properties are summarized for the reader's convenience:

1. When specialized to a second-order tensor (or matrix) $\mathcal{T}$,

$$A_1 \overset{\mathcal{T}}{\star} A_2 = A_1 \mathcal{T} A_2^T,$$

in which T denotes matrix or vector transposition, with the usual matrix product interpretation on the right-hand side.

2. If $x = [x_1, x_2, \ldots]$ is a row vector and I is the identity matrix, then

$$\left(I \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} I \overset{\mathcal{T}}{\star} x\right)_{i_1,\ldots,i_{k-1}} = \sum_l x_l \, \mathcal{T}_{i_1,\ldots,i_{k-1},l}$$

yields a tensor of order $k-1$.

3. If $e_{i_1}, \ldots, e_{i_k}$ are k unit row vectors each having a 1 in the position indexed i_k, then

$$\mathcal{T}_{i_1,i_2,\ldots,i_k} = e_{i_1} \overset{\mathcal{T}}{\star} e_{i_2} \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} e_{i_k}.$$

4. If vec$(\cdot)$ is the operator which rearranges a tensor into a vector according to

$$\text{vec}(\mathcal{T}) = \begin{bmatrix} \mathcal{T}_{00\ldots0} \\ \mathcal{T}_{10\ldots0} \\ \vdots \\ \mathcal{T}_{01\ldots0} \\ \mathcal{T}_{11\ldots0} \\ \vdots \end{bmatrix},$$

then the equation

$$\mathcal{S} = A_1 \overset{\mathcal{T}}{\star} A_2 \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} A_k$$

is equivalent to

$$\text{vec}(\mathcal{S}) = \left(A_1 \otimes A_2 \otimes \cdots \otimes A_k\right) \text{vec}(\mathcal{T}),$$

where $\otimes$ denotes the conventional Kronecker product of matrices [Bre78], [RM89].

5. *Composition property.* If $\mathcal{T} = B_1 \overset{\mathcal{D}}{\star} B_2 \overset{\mathcal{D}}{\star} \cdots \overset{\mathcal{D}}{\star} B_k$ with $\mathcal{D}$ some kth-order tensor, then $\mathcal{S} = A_1 \overset{\mathcal{T}}{\star} A_2 \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} A_k$ implies that

$$\mathcal{S} = (A_1 B_1) \overset{\mathcal{D}}{\star} (A_2 B_2) \overset{\mathcal{D}}{\star} \cdots \overset{\mathcal{D}}{\star} (A_k B_k). \tag{9.4.1}$$

9.5 EXAMPLES OF CUMULANTS AND TENSORS

We now show some simple examples relating the cumulants of the output process of a linear system to tensors constructed from the Tucker product. We set

$$y(n) = h_1\, u(n-1) + h_2\, u(n-2) + h_3\, u(n-3) + \cdots,$$

where $\{u(\cdot)\}$ is an i.i.d. sequence and the transfer function $H(z) = \sum_i h_i\, z^i$ is strictly causal. (If $H(z)$ were causal but not strictly causal, the system $zH(z)$ would be strictly causal and would generate the same output cumulants; the strictly causal constraint on $H(z)$ leads to simpler relations later on.)

Example 5. Consider a strictly causal system initially at rest. The output sequence becomes

$$\begin{bmatrix} y(0) \\ y(1) \\ y(2) \\ \vdots \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 0 & 0 & \cdots \\ h_1 & 0 & 0 & \ddots \\ h_2 & h_1 & 0 & \ddots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}}_{\triangleq H} \begin{bmatrix} u(0) \\ u(1) \\ u(2) \\ \vdots \end{bmatrix},$$

where H is the convolution matrix of the system. If $\{u(\cdot)\}$ is an i.i.d. sequence, its kth-order cumulants become

$$\operatorname{cum}[u(i_1), u(i_2), \ldots, u(i_k)] = \begin{cases} \gamma_k, & i_1 = i_2 = \cdots = i_k, \\ 0 & \text{otherwise.} \end{cases}$$

The cumulant tensor built from such an i.i.d. sequence is clearly $\gamma_k \mathcal{I}$ (where $\mathcal{I}$ is the identity tensor). The kth-order cumulant tensor $\mathcal{T}$ with elements indexed from zero,

$$(\mathcal{T}_1)_{i_1,i_2,\ldots,i_k} = \operatorname{cum}[y(i_1), y(i_2), \ldots, y(i_k)],$$

then becomes

$$\mathcal{T}_1 = \gamma_k \cdot H \star H \star \cdots \star H.$$

This tensor is, of course, symmetric (i.e., invariant to any permutation of the indices) since cumulants are symmetric functions of their arguments [Men91].

Example 6. Consider rewriting the input-output relation in the form

$$\begin{bmatrix} y(n) \\ y(n-1) \\ y(n-2) \\ \vdots \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & h_1 & h_2 & \cdots \\ 0 & 0 & h_1 & \cdots \\ 0 & 0 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix}}_{H^T} \begin{bmatrix} u(n) \\ u(n-1) \\ u(n-2) \\ \vdots \end{bmatrix}.$$

We suppose that n is sufficiently large for any initial conditions to have died out, thus yielding a stationary process for $\{y(\cdot)\}$. Taking now the kth-order output cumulant tensor as

$$(\mathcal{T}_2)_{i_1,i_2,\ldots,i_k} = \operatorname{cum}[y(n-i_1), y(n-i_2), \ldots, y(n-i_k)],$$

with elements again indexed from zero, we obtain

$$\mathcal{T}_2 = \gamma_k \cdot H^T \star H^T \star \cdots \star H^T.$$

Note that this tensor is Toeplitz (or invariant along any diagonal: $(\mathcal{T}_2)_{i_1,\ldots,i_k} = (\mathcal{T}_2)_{i_1+l,\ldots,i_k+l}$), due to the stationarity of the process $\{y(\cdot)\}$.

$\diamond$

We shall give special attention to the Toeplitz tensor of Ex. 6. Consideration of the structured tensor of Ex. 5, however, leads to the following interesting identity. For any cumulant of order k, we have [Gri96]

$$\begin{aligned}\left(\mathcal{T}_2 - \mathcal{T}_1\right)/\gamma_k &= H^T \star \cdots \star H^T - H \star \cdots \star H \\ &= \Gamma_H \star \cdots \star \Gamma_H, \end{aligned} \tag{9.5.1}$$

where Γ_H is the infinite Hankel matrix

$$\Gamma_H = \begin{bmatrix} h_1 & h_2 & h_3 & \cdots \\ h_2 & h_3 & h_4 & \cdots \\ h_3 & h_4 & h_5 & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}.$$

Hankel matrices take a special significance in system theory [AAK71], [Glo84]. For now we note that, since $H\star\cdots\star H$ vanishes along all faces (where any index equals zero), the faces of $H^T \star\cdots\star H^T$ coincide with those of $\Gamma_H \star\cdots\star \Gamma_H$, which will prove convenient in what follows.

When specialized to second-order arrays, the identity (9.5.1) reads as

$$H^T H - H H^T = \Gamma_H \Gamma_H, \quad \text{or} \quad H^T H = H H^T + \Gamma_H \Gamma_H.$$

This implies the existence of an orthogonal matrix Q (satisfying $QQ^T = Q^T Q = I$) fulfilling

$$\begin{bmatrix} H \\ \bigcirc \end{bmatrix} = Q \begin{bmatrix} H^T \\ \Gamma_H \end{bmatrix}.$$

One is naturally led to inquire whether there exist square matrices that appear "orthogonal" with respect to higher-order Tucker products, i.e., square matrices Q for which $Q\star Q\star\cdots\star Q = \mathcal{I}$. If Q is an infinite matrix, it is readily verified that the choice $Q = Z^T$ (the "up-shift" matrix) yields $Z^T \star\cdots\star Z^T = \mathcal{I}$, the kth-order identity tensor, for any order $k \geq 2$. Similarly, choosing Q as any permutation matrix likewise leads to $Q\star\cdots\star Q = \mathcal{I}$, for any order $k \geq 2$. And, for k even, choosing Q as any signed permutation matrix (i.e., having a sole entry of ± 1 in each row) still works.

The following result shows that, in finite dimensions, the list of "higher-order orthogonal" matrices is short.

Theorem 9.5.1 (Higher-Order Orthogonal Matrices). *A square matrix Q of finite dimensions fulfills the kth-order orthogonality*

$$\underbrace{Q\star Q\star\cdots\star Q}_{k\geq 3 \text{ terms}} = \mathcal{I} \tag{9.5.2}$$

if and only if Q is a permutation matrix (k odd) or a signed permutation matrix (k even).

Proof: Suppose Q has dimensions $L \times L$, with L arbitrary, and write out the matrix as

$$Q = \begin{bmatrix} q_1^T \\ q_2^T \\ \vdots \\ q_L^T \end{bmatrix} = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1L} \\ q_{21} & q_{22} & \cdots & q_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ q_{L1} & q_{L2} & \cdots & q_{LL} \end{bmatrix}.$$

If $v^T = [v_1, \ldots, v_L]$ and $w^T = [w_1, \ldots, w_L]$ are two row vectors, their Hadamard (or componentwise) product will be denoted as

$$v^T * w^T = [v_1 w_1, \ldots, v_L w_L].$$

Now, the constraint (9.5.2) can be written in vector inner product form as

$$\Big(\underbrace{q_{i_1} * \cdots * q_{i_m}}_{m \text{ terms}}\Big)^T \Big(\underbrace{q_{i_m+1} * \cdots * q_{i_k}}_{k-m \text{ terms}}\Big) = \delta(i_1, i_2, \ldots, i_k) \tag{9.5.3}$$

for any $1 \le m < k$. Upon choosing $i_1 = i_2 = \cdots = i_k$, we see immediately that none of the row vectors q_l^T can be the zero vector.

Let us show next that the row vectors must be linearly independent. First note from (9.5.3) that

$$\Big(\underbrace{q_1 * \cdots * q_1}_{k-1 \text{ terms}}\Big)^T q_1 = 1 \qquad \text{and} \qquad \Big(\underbrace{q_1 * \cdots * q_1}_{k-1 \text{ terms}}\Big)^T q_l = 0, \quad l = 2, 3, \ldots, L.$$

Suppose to the contrary that the vectors are linearly dependent. We can then find nonzero constants $\alpha_1, \ldots, \alpha_L$ such that

$$\alpha_1 q_1 + \cdots + \alpha_L q_L = 0.$$

If only one of the terms α_l were nonzero, the corresponding vector q_l would be zero, in contradiction with all vectors being distinct from the zero vector. Suppose then that two or more terms from $\alpha_1, \ldots, \alpha_L$ are nonzero. By permuting the indices if necessary, we may suppose that $\alpha_1 \neq 0$ and $\alpha_2 \neq 0$ (and possibly others as well). We may then write q_1 as

$$q_1 = -\frac{\alpha_2}{\alpha_1} q_2 - \frac{\alpha_3}{\alpha_1} q_3 - \cdots.$$

This yields a contradiction as

$$\begin{aligned} 1 &= \Big(\underbrace{q_1 * \cdots * q_1}_{k-1 \text{ terms}}\Big)^T q_1 \\ &= \Big(\underbrace{q_1 * \cdots * q_1}_{k-1 \text{ terms}}\Big)^T \Big(-\frac{\alpha_2}{\alpha_1} q_2 - \frac{\alpha_3}{\alpha_1} q_3 - \cdots\Big) = 0. \end{aligned}$$

Accordingly, the vectors $q_1, \ldots, q_L$ must be linearly independent.

Now, from (9.5.3) we can write

$$\text{for all } i \neq j, \qquad \Big(\underbrace{q_i * \cdots * q_i}_{k-n \text{ terms}} * \underbrace{q_j * \cdots * q_j}_{n-1 \text{ terms}}\Big)^T q_l = 0, \qquad l = 1, 2, \ldots, L,$$

for any choice of n between 2 and $k-1$. Since the vectors $q_1, \ldots, q_L$ are linearly independent, they span $\mathbb{R}^L$. The only vector in $\mathbb{R}^L$ orthogonal to $\mathbb{R}^L$ is, of course, the zero vector. The previous expression then implies that

$$\text{for all } i \neq j, \qquad \underbrace{q_i * \cdots * q_i}_{k-n \text{ terms}} * \underbrace{q_j * \cdots * q_j}_{n-1 \text{ terms}} = 0,$$

which reads componentwise as

$$(q_{im})^{k-n}(q_{jm})^{n-1} = 0 \qquad \text{for all } i \neq j.$$

This simplifies to

$$\text{if } q_{im} \neq 0 \text{ then } q_{jm} = 0 \qquad \text{for all } j \neq i.$$

As such, each column of Q can have only one nonzero entry. The same must now apply to each row of Q, for if a given row were to have two or more nonzero entries, then another row would be left with no nonzero entries, giving a zero vector. From the constraint

$$\sum_{l=1}^{L} q_{il}^k = 1,$$

it follows easily that the sole nonzero entry of each row q_i^T must be $+1$ if k is odd, or ± 1 if k is even, giving Q as a (signed) permutation matrix.

◇

9.6 DISPLACEMENT STRUCTURE FOR TENSORS

In this section we develop various relations that connect the displacement structure of a cumulant tensor to polyspectral functions. Relations to cumulant interpolation will follow in Sec. 9.7.

Let $\mathcal{T}$ be a given kth-order tensor and let Z still denote the shift matrix with ones on the subdiagonal and zeros elsewhere. The kth-order tensor

$$\mathcal{D} = \underbrace{Z \overset{\mathcal{T}}{\star} Z \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} Z}_{k \text{ terms}}$$

relates to $\mathcal{T}$ as

$$\mathcal{D}_{i_1,\ldots,i_k} = \begin{cases} 0 & \text{if } i_1 = 0 \text{ and/or } \ldots \text{and/or } i_k = 0, \\ \mathcal{T}_{i_1-1,\ldots,i_k-1} & \text{otherwise.} \end{cases}$$

Example 7. If we take for $\mathcal{T}$ the Toeplitz tensor of Ex. 6, then its displacement residue

$$\mathcal{T} - Z \overset{\mathcal{T}}{\star} Z \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} Z$$

will coincide with $\mathcal{T}$ along all faces (when at least one index equals zero) and will vanish at all interior points (where all indices are nonzero).

◇

If we instead consider a Hankel-based tensor, i.e.,

$$\mathcal{T} = \Gamma_H \star \Gamma_H \star \cdots \star \Gamma_H, \tag{9.6.1}$$

we can obtain an interesting relation for the up-shifted displacement residue, using the up-shift matrix Z^T.

Lemma 9.6.1 (Displacement Residue). *The tensor $\mathcal{T}$ from* (9.6.1) *fulfills*

$$\mathcal{T} - Z^T \overset{\mathcal{T}}{\star} Z^T \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} Z^T = h \star h \star \cdots \star h$$

with $h = [h_1, h_2, h_3, \ldots]^T$.

Proof: We recall that a Hankel matrix satisfies (by definition of Hankel) the shift equation

$$Z^T \Gamma_H = \Gamma_H Z.$$

Since $\mathcal{T} = \Gamma_H \star \cdots \star \Gamma_H$, we see by a direct calculation that

$$\begin{aligned} \mathcal{T} - Z^T \overset{\mathcal{T}}{\star} Z^T \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} Z^T &= \Gamma_H \star \cdots \star \Gamma_H - \Big((Z^T\Gamma_H) \star \cdots \star (Z^T \Gamma_H) \Big) \\ &= \Gamma_H \star \cdots \star \Gamma_H - \Big((\Gamma_H Z) \star \cdots \star (\Gamma_H Z) \Big) \\ &= \Gamma_H \star \cdots \star \Gamma_H - \Big(\Gamma_H \overset{\mathcal{J}}{\star} \cdots \overset{\mathcal{J}}{\star} \Gamma_H \Big), \end{aligned}$$

in which the final line comes from the composition property with $\mathcal{J} = Z \star \cdots \star Z$. This latter tensor reads elementwise as

$$(\mathcal{J})_{i_1,\ldots,i_k} = \begin{cases} 1, & i_1 = \cdots = i_k \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

This allows us to continue as

$$\Big(\Gamma_H \star \cdots \star \Gamma_H \Big) - \Big(\Gamma_H \overset{\mathcal{J}}{\star} \cdots \overset{\mathcal{J}}{\star} \Gamma_H \Big) = \Gamma_H \overset{\mathcal{I}-\mathcal{J}}{\star} \cdots \overset{\mathcal{I}-\mathcal{J}}{\star} \Gamma_H,$$

in which $\mathcal{I} - \mathcal{J}$ vanishes everywhere except in the leading entry, which equals one, thus giving $\mathcal{I} - \mathcal{J} = e_0^T \star \cdots \star e_0^T$, in which e_0^T is the unit column vector with a 1 in the leading entry. This then gives, again by the composition property,

$$\Gamma_H \overset{\mathcal{I}-\mathcal{J}}{\star} \cdots \overset{\mathcal{I}-\mathcal{J}}{\star} \Gamma_H = (\Gamma_H e_0^T) \star \cdots \star (\Gamma_H e_0^T) = h \star \cdots \star h$$

as claimed. ◇

9.6.1 Relation to the Polyspectrum

We return now to the Toeplitz tensor

$$\mathcal{T}_{i_1,i_2,\ldots,i_k} = \text{cum}[y(n-i_1), y(n-i_2), \ldots, y(n-i_k)]$$

with the assumption that $\{y(\cdot)\}$ is a stationary process, although not necessarily a linear process. Its cumulant lags of order k involve the relative lags of the first $k-1$ arguments with respect to the final argument, i.e.,

$$c_{i_1,i_2,\ldots,i_{k-1}} = \text{cum}[y(n-i_1), y(n-i_2), \ldots, y(n-i_{k-1}), y(n)].$$

These are simply the elements of the final face of $\mathcal{T}$, i.e.,

$$\Big(I \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} I \overset{\mathcal{T}}{\star} e_0 \Big)_{i_1,\ldots,i_{k-1}} = c_{i_1,\ldots,i_{k-1}}.$$

The kth-order polyspectrum is defined as the bilateral $(k-1)$-dimensional Fourier transform of the cumulants,

$$R(z_1, z_2, \ldots, z_{k-1}) \tag{9.6.2}$$

$$\triangleq \sum_{i_1=-\infty}^{+\infty} \sum_{i_2=-\infty}^{+\infty} \cdots \sum_{i_{k-1}=-\infty}^{+\infty} c_{i_1,i_2,\ldots,i_{k-1}}\, z_1^{i_1}, z_2^{i_2} \cdots z_{k-1}^{i_{k-1}},$$

whenever the sum converges along the unit polycircle $|z_1| = \cdots = |z_{k-1}| = 1$.

Introduce the infinite row vector

$$\mathbf{z}_i \triangleq [1 \;\; z_i \;\; z_i^2 \;\; z_i^3 \;\; \cdots], \qquad |z_i| < 1,$$

containing successive powers of the complex variable z_i. We may then introduce the k-variable scalar "generating function," whose coefficients are the elements of the tensor $\mathcal{T}$, as

$$T(z_1, z_2, \ldots, z_k) = \mathbf{z}_1 \overset{\mathcal{T}}{\star} \mathbf{z}_2 \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} \mathbf{z}_k = \sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} \cdots \sum_{i_k=0}^{\infty} \mathcal{T}_{i_1,i_2,\ldots,i_k} z_1^{i_1} z_2^{i_2} \cdots z_k^{i_k}.$$

Note that this function depends on k complex variables, whereas the polyspectrum from (9.6.2) involves only $k-1$ complex variables. We pursue now how to reconcile these two functions.

Introducing the displacement residue

$$\mathcal{S} = \mathcal{T} - Z \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} Z,$$

its multivariable function becomes

$$\begin{aligned} S(z_1, \ldots, z_k) &= \mathbf{z}_1 \overset{\mathcal{S}}{\star} \cdots \overset{\mathcal{S}}{\star} \mathbf{z}_k \\ &= \sum_{i_1=0}^{\infty} \cdots \sum_{i_k=0}^{\infty} \mathcal{S}_{i_1,\ldots,i_k} z_1^{i_1} \cdots z_k^{i_k} \\ &= (1 - z_1 z_2 \cdots z_k)\, T(z_1, z_2, \ldots, z_k). \end{aligned}$$

Because $\mathcal{T}$ is a Toeplitz tensor, its displacement residue $\mathcal{S}$ vanishes at all interior points. We shall call $S(z_1, \ldots, z_k)$ the polyspectral residue function, based on the following identity.

Lemma 9.6.2 (Polyspectral Residue Function). *With*

$$S(z_1, \ldots, z_k) = (1 - z_1 \cdots z_k)\, T(z_1, \ldots, z_k),$$

the polyspectrum is obtained by setting $z_k = (z_1 \cdots z_{k-1})^{-1}$*:*

$$R(z_1, \ldots, z_{k-1}) = S(z_1, \ldots, z_{k-1}, (z_1 \cdots z_{k-1})^{-1}).$$

Proof: This identity comes from exploiting various symmetry relations linking cumulants of stationary processes. We illustrate the proof for third-order cumulants, as the verification for higher-order cumulants is quite similar.

Introduce the constant, one-dimensional causal, and two-dimensional causal parts of $S(z_1, z_2, z_3)$ as

$$\begin{aligned} c_{00} &= S(0,0,0), \\ S_+(z_1) &= \sum_{i_1=1}^{\infty} c_{i_1,0}\, z_1^{i_1} = S(z_1, 0, 0) - c_{00}, \\ S_{2+}(z_1, z_2) &= \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{\infty} c_{i_1,i_2}\, z_1^{i_1} z_2^{i_2} = S(z_1, z_2, 0) - c_{00} - S_+(z_1) - S_+(z_2). \end{aligned}$$

Now, $S(z_1, z_2, z_3)$ is a symmetric function of the complex variables z_1, z_2, and z_3 and may be expressed as

$$S(z_1, z_2, z_3) = c_{00} + S_+(z_1) + S_+(z_2) + S_+(z_3) + S_{2+}(z_1, z_2) + S_{2+}(z_2, z_3) + S_{2+}(z_1, z_3).$$

Setting $z_3 = (z_1 z_2)^{-1}$ gives

$$\begin{aligned} S(z_1, z_2, (z_1 z_2)^{-1}) &= c_{00} + S_+(z_1) + S_+(z_2) + S_+((z_1 z_2)^{-1}) \\ &\quad + S_{2+}(z_1, z_2) + S_{2+}(z_2, (z_1 z_2)^{-1}) \\ &\quad + S_{2+}(z_1, (z_1 z_2)^{-1}). \end{aligned} \tag{9.6.3}$$

By exploiting stationarity, we see that

$$c_{i,0} = \mathrm{cum}[y(n-i), y(n), y(n)] = \mathrm{cum}[y(n), y(n+i), y(n+i)] = c_{-i,-i},$$

so that

$$S_+((z_1 z_2)^{-1}) = \sum_{i=1}^{\infty} c_{i,0} z_1^{-i} z_2^{-i} = \sum_{i=1}^{\infty} c_{-i,-i} z_1^{-i} z_2^{-i},$$

which coincides with the z-transform of the negative diagonal slice $c_{-i,-i}$.

In a similar way, stationarity again gives

$$\begin{aligned} c_{i_1,i_2} &= \mathrm{cum}[y(n-i_1), y(n-i_2), y(n)] \\ &= \mathrm{cum}[y(n-i_1+i_2), y(n), y(n+i_2)] = c_{i_2-i_1,-i_2}, \end{aligned}$$

so that

$$S_{2+}(z_1, (z_1 z_2)^{-1}) = \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{\infty} c_{i_1,i_2} z_1^{i_1} [(z_1 z_2)^{-1}]^{i_2} = \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{\infty} c_{i_2-i_1,-i_2} z_1^{i_1-i_2} z_2^{-i_2}$$

and, in the same way,

$$S_{2+}(z_2, (z_1 z_2)^{-1}) = \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{\infty} c_{-i_1,i_1-i_2} z_1^{-i_1} z_2^{i_2-i_1}.$$

Figure 9.1 illustrates the (i_1, i_2)-plane, in which each point represents a sample value c_{i_1,i_2} and the dashed lines indicate which samples enter into which sum from (9.6.3). The sum from (9.6.3) is seen to incorporate the doubly two-sided sequence c_{i_1,i_2}, i.e.,

$$S(z_1, z_2, (z_1 z_2)^{-1}) = \sum_{i_1=-\infty}^{+\infty} \sum_{i_2=-\infty}^{+\infty} c_{i_1,i_2} z_1^{i_1} z_2^{i_2},$$

yielding the bispectrum as claimed.

$\diamond$

Remark. The decomposition into $S_+(z_1)$ and $S_{2+}(z_1, z_2)$ can be considered, for this third-order case, a type of analytic continuation into the unit bidisk ($|z_1| < 1$, $|z_2| < 1$) of the bispectrum, with the bispectrum obtained along the boundary $|z_1| = |z_2| = 1$ by the symmetry relation (9.6.3). An open problem here is to determine the set of admissible functions for $S_+(z_1)$ and $S_{2+}(z_1, z_2)$ for which (9.6.3) yields a *valid* bispectrum, i.e., corresponding to *some* stationary process. An analogous open question applies to higher orders as well; for second-order statistics, admissibility reduces to a positive real constraint.

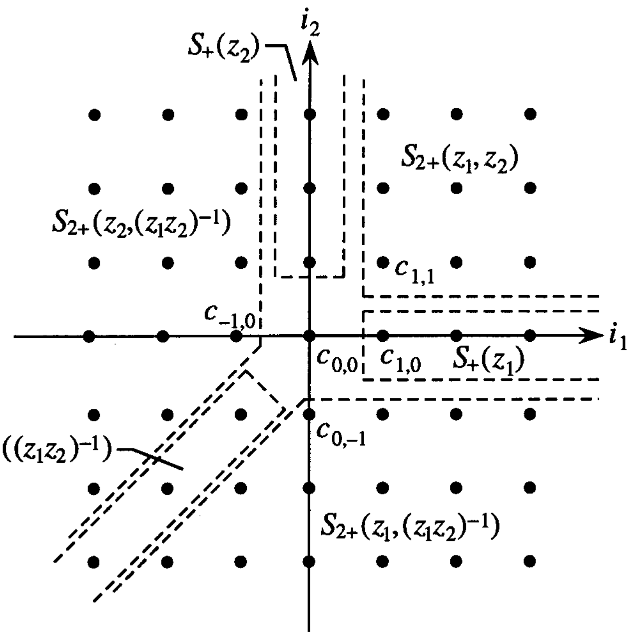

Figure 9.1. Illustrating cumulant sample values c_{i_1,i_2} in the (i_1, i_2)-plane and which terms enter into which sum from (9.6.3).

9.6.2 The Linear Case

In this section we further examine the structure of the polyspectral residue function for the special case in which the Toeplitz tensor $\mathcal{T}$ is obtained from the cumulants of a linear process. We recall from Ex. 6 that the Toeplitz tensor $\mathcal{T}$ is Tucker factorable as

$$\mathcal{T} = \gamma_k \cdot \underbrace{H^T \star \cdots \star H^T}_{k \text{ terms}},$$

in which H is the convolution matrix of the linear system and γ_k is the kth-order cumulant of the i.i.d. driving sequence to the system. We shall study the $(k-1)$-variable function obtained from $T(z_1, \ldots, z_k)$ by setting the final complex variable z_k to zero:

$$T(z_1, \ldots, z_{k-1}, 0) = \sum_{i_1=0}^{\infty} \cdots \sum_{i_{k-1}=0}^{\infty} c_{i_1,\ldots,i_{k-1}} z_1^{i_i} \cdots z_{k-1}^{i_{k-1}}.$$

Observe that this function involves unilateral z-transforms in each index and differs from the polyspectrum of (9.6.2), which involves bilateral z-transforms in each index. We seek a closed-form expression for $T(z_1, \ldots, z_{k-1}, 0)$ in terms of the system $H(z)$, when this transfer function is rational.

We exploit the fact that the faces of the Toeplitz tensor $\mathcal{T}$ coincide with those of the Hankel-based tensor $\gamma_k \cdot \Gamma_H \star \cdots \star \Gamma_H$ [cf. (9.5.1)], i.e.,

$$(\mathcal{T})_{i_1,\ldots,i_{k-1},0} = \left(\gamma_k \cdot H^T \star \cdots \star H^T\right)_{i_1,\ldots,i_{k-1},0} = \left(\gamma_k \cdot \Gamma_H \star \cdots \star \Gamma_H\right)_{i_1,\ldots,i_{k-1},0},$$

whose multidimensional z-transform is easier to treat.

Suppose now that $H(z)$ is rational; this means that we have a realization of the form

$$\begin{aligned} x(n+1) &= Ax(n) + bu(n), \\ y(n) &= cx(n), \end{aligned}$$

in which $H(z) = zc(I - zA)^{-1}b$. We let M denote the state vector dimension. The Hankel matrix Γ_H can then be decomposed as

$$\Gamma_H = \underbrace{\begin{bmatrix} c \\ cA \\ cA^2 \\ \vdots \end{bmatrix}}_{\mathcal{O}} \underbrace{\begin{bmatrix} b & Ab & A^2 & \cdots \end{bmatrix}}_{\mathcal{C}}$$

in terms of the infinite horizon observability and controllability matrices, $\mathcal{O}$ and $\mathcal{C}$, respectively. The full kth-order Hankel-based tensor becomes

$$\begin{aligned} \gamma_k \cdot \Gamma_H \star \cdots \star \Gamma_H &= \gamma_k \cdot (\mathcal{O}\mathcal{C}) \star \cdots \star (\mathcal{O}\mathcal{C}) \\ &= \mathcal{O} \overset{\mathcal{P}}{\star} \cdots \overset{\mathcal{P}}{\star} \mathcal{O} \end{aligned}$$

using the composition property in the second line, in which $\mathcal{P}$ is the kth-order tensor

$$\mathcal{P} = \gamma_k \cdot \mathcal{C} \star \cdots \star \mathcal{C}. \tag{9.6.4}$$

The following lemma is in direct analogy with the second-order case.

Lemma 9.6.3 (Higher-Order Lyapunov Equation). *The kth-order tensor $\mathcal{P}$ from* (9.6.4) *fulfills the (higher-order) Lyapunov equation*

$$\mathcal{P} - \underbrace{A \overset{\mathcal{P}}{\star} A \overset{\mathcal{P}}{\star} \cdots \overset{\mathcal{P}}{\star} A}_{k \text{ terms}} = \gamma_k \cdot \underbrace{b \star b \star \cdots \star b}_{k \text{ terms}}. \tag{9.6.5}$$

If $\lambda_1, \ldots, \lambda_M$ are the eigenvalues of A, then this equation admits a unique solution $\mathcal{P}$ provided

$$\lambda_{i_1} \lambda_{i_2} \cdots \lambda_{i_k} \neq 1 \qquad \text{for all } i_1, \ldots, i_k. \tag{9.6.6}$$

With respect to the state equation $x(n+1) = Ax(n) + bu(n)$, with A stable ($|\lambda_i| < 1$), $\mathcal{P}$ is the state cumulant tensor

$$\mathcal{P}_{i_1,\ldots,i_k} = \text{cum}[x_{i_1}(n), \ldots, x_{i_k}(n)].$$

Proof: The verification follows closely that familiar from second-order statistics [AM79] and is included for completeness. To begin, with $\mathcal{P} = \gamma_k \cdot \mathcal{C} \star \cdots \star \mathcal{C}$, the composition property gives

$$A \overset{\mathcal{P}}{\star} \cdots \overset{\mathcal{P}}{\star} A = \gamma_k \cdot (A\mathcal{C}) \star \cdots \star (A\mathcal{C}).$$

Here we observe that

$$\begin{aligned} A\mathcal{C} &= A\,[\,b \quad Ab \quad A^2b \quad \cdots\,] \\ &= [\,b \quad Ab \quad A^2b \quad \cdots\,]\, Z = \mathcal{C}\, Z, \end{aligned}$$

so that

$$\begin{aligned}\mathcal{P} - A \overset{\mathcal{P}}{\star} \cdots \overset{\mathcal{P}}{\star} A &= \gamma_k \cdot \Big(\mathcal{C} \star \cdots \star \mathcal{C}\Big) - \gamma_k \Big((\mathcal{C}Z) \star \cdots \star (\mathcal{C}Z)\Big) \\ &= \gamma_k \cdot \mathcal{C} \overset{\mathcal{J}}{\star} \cdots \overset{\mathcal{J}}{\star} \mathcal{C}, \end{aligned} \qquad (9.6.7)$$

in which $\mathcal{J} = \mathcal{I} - Z \star \cdots \star Z$. Since $\mathcal{J}$ has a 1 in the leading entry and vanishes elsewhere, we can write

$$\mathcal{J} = e_0^T \star \cdots \star e_0^T,$$

where e_0^T is the unit column vector with a 1 in its leading entry. This combines with (9.6.7) to give

$$\mathcal{P} - A \overset{\mathcal{P}}{\star} \cdots \overset{\mathcal{P}}{\star} A = \gamma_k \cdot (\mathcal{C}e_0^T) \star \cdots \star (\mathcal{C}e_0^T) = \gamma_k \cdot b \star \cdots \star b,$$

yielding the Lyapunov equation (9.6.5).

For existence and uniqueness, the vectorized tensor $\mathrm{vec}(\mathcal{P})$ fulfills

$$(I - A \otimes \cdots \otimes A)\,\mathrm{vec}(\mathcal{P}) = \gamma_k \cdot b \otimes \cdots \otimes b.$$

With $\{\lambda_i\}$ denoting the eigenvalues of A, those of its k-term Kronecker product $A \otimes \cdots \otimes A$ become $\lambda_{i_1} \cdots \lambda_{i_k}$ as the indices $i_1, \ldots, i_k$ range over all M^k possibilities. The relation (9.6.6) is then equivalent to invertibility of $I - A \otimes \cdots \otimes A$, which in turn is equivalent to existence and uniqueness of a solution $\mathrm{vec}(\mathcal{P})$ and hence of $\mathcal{P}$ itself.

For the final part, (asymptotic) stationarity implies that

$$\mathrm{cum}[x_{i_1}(n+1), x_{i_2}(n+1), \ldots, x_{i_k}(n+1)] = \mathrm{cum}[x_{i_1}(n), x_{i_2}(n), \ldots, x_{i_k}(n)].$$

It suffices to show that these values build a state cumulant tensor which indeed satisfies the given Lyapunov equation. Now, $x(n)$ depends only on past values $u(n-1)$, $u(n-2), \ldots$ of the input. By the i.i.d. assumption on $\{u(\cdot)\}$, cross cumulants involving $x(n)$ and $u(n)$ vanish, and a simple calculation shows that the state cumulant tensor satisfies the given Lyapunov equation.

$\diamond$

Let $\mathcal{T}^{(0)}$ denote the face of the Toeplitz tensor $\mathcal{T}$, i.e.,

$$\mathcal{T}^{(0)}_{i_1,\ldots,i_{k-1}} = \mathcal{T}_{i_1,\ldots,i_{k-1},0} = c_{i_1,\ldots,i_{k-1}}.$$

This face coincides with that of its Hankel-based counterpart, giving

$$\mathcal{T}^{(0)} = \mathcal{O} \overset{\mathcal{P}}{\star} \cdots \overset{\mathcal{P}}{\star} \mathcal{O} \overset{\mathcal{P}}{\star} c.$$

In particular, an expression for each output cumulant contained on the face $\mathcal{T}^{(0)}$ becomes

$$c_{i_1,i_2,\ldots,i_{k-1}} = \mathcal{T}^{(0)}_{i_1,i_2,\ldots,i_{k-1}} = (cA^{i_1}) \overset{\mathcal{P}}{\star} (cA^{i_2}) \overset{\mathcal{P}}{\star} \cdots \overset{\mathcal{P}}{\star} (cA^{i_{k-1}}) \overset{\mathcal{P}}{\star} c.$$

Analogous formulas are found in [SM90b], using conventional Kronecker product formalisms.

The multidimensional z-transform of the face of $\mathcal{T}$ is then readily computed as

$$\begin{aligned} T(z_1, \ldots, z_{k-1}, 0) &= \sum_{i_1=0}^{\infty} \cdots \sum_{i_{k-1}=0}^{\infty} c_{i_1,\ldots,i_{k-1}} z_1^{i_1} \cdots z_{k-1}^{i_{k-1}} \\ &= [c(I - z_1 A)^{-1}] \overset{\mathcal{P}}{\star} \cdots \overset{\mathcal{P}}{\star} [c(I - z_{k-1}A)^{-1}] \overset{\mathcal{P}}{\star} c. \end{aligned}$$

This shows, in particular, that whenever $\{y(\cdot)\}$ is a rational process, the polyspectral function $T(z_1, \ldots, z_{k-1}, 0)$ will likewise be rational with a separable denominator of the form $A(z_1) \cdots A(z_{k-1})$, where $A(z) = \det(I - zA)$ is the denominator of $H(z)$. This fact has been used in numerous works [GM89], [JK92], [SM90a], [SM92] to extract the poles of the model from a one-dimensional cumulant "slice," obtained by varying only one index. The z-transform of such a slice appears as

$$\sum_{i_1=0}^{\infty} c_{i_1,i_2,\ldots,i_k} z_1^{i_1} = \frac{\partial^{i_2}}{i_2!(\partial z_2)^{i_2}} \cdots \frac{\partial^{i_{k-1}}}{i_{k-1}!(\partial z_{k-1})^{i_{k-1}}} S(z_1, \ldots, z_{k-1}, 0)\Big|_{z_2=0;\ldots;z_{k-1}=0}.$$

This function may suffer pole-zero cancellations in certain cases [Men91], such that certain system poles are hidden from the cumulant slice in question and in some cases no matter which cumulant slice is taken [Men91]. As shown in [RG94], however, this phenomenon is limited to very special classes of systems $H(z)$.

9.7 POLYSPECTRAL INTERPOLATION

We consider some explicit solutions to higher-order Lyapunov equations and how such equations relate to specific evaluations of the polyspectral residue function $S(z_1, \ldots, z_k)$. We then establish a necessary condition for the existence of a linear process compatible with these polyspectral evaluations.

Let F be a square $(M \times M)$ matrix with all eigenvalues in the open unit disk and g an $M \times 1$ column vector, and suppose (F, g) is a controllable pair, i.e., the M rows of the infinite horizon controllability matrix

$$[g \quad Fg \quad F^2g \quad \cdots]$$

are linearly independent. We suppose in what follows that the eigenvalues of F are distinct for ease of presentation, although the various relations to follow extend readily to the case of repeated eigenvalues. By controllability of the pair (F, g), there exists an invertible matrix W, which renders the transformed pair $(W^{-1}FW, W^{-1}g)$ in canonic parallel form [Kai80], i.e.,

$$W^{-1}FW = \Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_M \end{bmatrix}, \qquad W^{-1}g = \mathbf{1} \triangleq \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

The controllability matrix in this parallel coordinate system is simply

$$[\mathbf{1} \quad \Lambda\mathbf{1} \quad \Lambda^2\mathbf{1} \quad \cdots] = \begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots \\ 1 & \lambda_2 & \lambda_2^2 & \cdots \\ \vdots & \vdots & \vdots & \cdots \\ 1 & \lambda_M & \lambda_M^2 & \cdots \end{bmatrix}. \tag{9.7.1}$$

A simple calculation then shows that the solution to the Lyapunov equation

$$\mathcal{D} - \underbrace{\Lambda \overset{\mathcal{D}}{\star} \cdots \overset{\mathcal{D}}{\star} \Lambda}_{k \text{ terms}} = \underbrace{\mathbf{1} \star \cdots \star \mathbf{1}}_{k \text{ terms}}$$

is given by (indexed from one)

$$\mathcal{D}_{i_1,i_2,\ldots,i_k} = \sum_{l=0}^{\infty}(\lambda_{i_1}\lambda_{i_2}\cdots\lambda_{i_k})^l = \frac{1}{1-\lambda_{i_1}\lambda_{i_2}\cdots\lambda_{i_k}}.$$

Since

$$[g \quad Fg \quad F^2g \quad \cdots] = W[\mathbf{1} \quad \Lambda\mathbf{1} \quad \Lambda^2\mathbf{1} \quad \cdots],$$

the solution to the Lyapunov equation $\mathcal{P} - F \overset{\mathcal{P}}{\star} \cdots \overset{\mathcal{P}}{\star} F = g \star \cdots \star g$ relates to $\mathcal{D}$ by the congruence transformation

$$\mathcal{P} = W \overset{\mathcal{D}}{\star} W \overset{\mathcal{D}}{\star} \cdots \overset{\mathcal{D}}{\star} W.$$

Note that although the elements of $\mathcal{P}$ are real whenever F and g are real, the elements of $\mathcal{D}$ will in general be complex.

Consider now the state recursion

$$\xi(n+1) = F\xi(n) + gy(n),$$

where $\{y(\cdot)\}$ is a stationary (possibly nonlinear) process, F is a stable matrix, and the time index n is sufficiently large for $\xi(\cdot)$ to be a stationary vector process. With $\bar{\xi}(n) = W^{-1}\xi(n)$, we have an equivalent parallel realization of the form

$$\bar{\xi}(n+1) = \Lambda\bar{\xi}(n) + \mathbf{1}\,y(n).$$

We examine in the remainder of this section state cumulant tensors from such recursions and how they relate to evaluations of the polyspectral residue function $S(z_1, \ldots, z_k)$ obtained from $\{y(\cdot)\}$ at specific points in the open unit polydisk $|z_1| < 1$, $\ldots$, $|z_k| < 1$. This will lead to a necessary condition in Thm. 9.7.1 for the existence of a linear process which is compatible with (or replicates) a set of polyspectral evaluations. We begin with the following identity.

Lemma 9.7.1 (An Identity). *Let $\mathcal{D}$ be the $M \times M \times \cdots \times M$ tensor (indexed from one)*

$$\mathcal{D}_{i_1,i_2,\ldots,i_k} = \operatorname{cum}[\bar{\xi}_{i_1}(n+1), \bar{\xi}_{i_2}(n+1), \ldots, \bar{\xi}_{i_k}(n+1)]. \tag{9.7.2}$$

Then $\mathcal{D}$ can be written as

$$\mathcal{D}_{i_1,i_2,\ldots,i_k} = \frac{S(\lambda_{i_1}, \lambda_{i_2}, \ldots, \lambda_{i_k})}{1-\lambda_{i_1}\lambda_{i_2}\cdots\lambda_{i_k}}. \tag{9.7.3}$$

Proof: To verify the identity (9.7.3), we have that

$$\bar{\xi}(n+1) = \underbrace{[\mathbf{1} \quad \Lambda\mathbf{1} \quad \Lambda^2\mathbf{1} \quad \cdots]}_{\mathcal{C}_p} \begin{bmatrix} y(n) \\ y(n-1) \\ y(n-2) \\ \vdots \end{bmatrix},$$

in which $\mathcal{C}_p$ denotes the controllability matrix in the parallel coordinate system. The cumulant tensor $\mathcal{D}$ from (9.7.2) then becomes

$$\mathcal{D} = \mathcal{C}_p \overset{\mathcal{I}}{\star} \cdots \overset{\mathcal{I}}{\star} \mathcal{C}_p,$$

in which $\mathcal{T}$ is the Toeplitz cumulant tensor constructed from the stationary process $\{y(\cdot)\}$. Consider now the displaced tensor

$$\mathcal{E} = \Lambda \overset{\mathcal{D}}{\star} \cdots \overset{\mathcal{D}}{\star} \Lambda,$$

which reads elementwise as

$$\mathcal{E}_{i_1,i_2,\ldots,i_k} = \lambda_{i_1}\lambda_{i_2}\cdots\lambda_{i_k}\cdot\mathcal{D}_{i_1,i_2,\ldots,i_k}. \tag{9.7.4}$$

Using the relation $\Lambda\mathcal{C}_p = \mathcal{C}_p Z$, we see that

$$\begin{aligned}\mathcal{D}-\mathcal{E} = \mathcal{D} - \Lambda \overset{\mathcal{D}}{\star} \cdots \overset{\mathcal{D}}{\star} \Lambda &= \left(\mathcal{C}_p \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} \mathcal{C}_p\right) - \left((\mathcal{C}_p Z) \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} (\mathcal{C}_p Z)\right)\\ &= \mathcal{C}_p \overset{\mathcal{S}}{\star} \cdots \overset{\mathcal{S}}{\star} \mathcal{C}_p,\end{aligned}$$

in which

$$\mathcal{S} = \mathcal{T} - Z \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} Z$$

is the displacement residue of the Toeplitz tensor $\mathcal{T}$. Now, the multidimensional z-transform of the elements of $\mathcal{S}$ gives $S(z_1,\ldots,z_k)$. Considering the special structure of $\mathcal{C}_p$ exposed in (9.7.1), we see that $\mathcal{D}-\mathcal{E}$ may be written elementwise as

$$(\mathcal{D}-\mathcal{E})_{i_1,i_2,\ldots,i_k} = \sum_{n_1=0}^{\infty}\cdots\sum_{n_k=0}^{\infty} \mathcal{S}_{n_1,\ldots,n_k}\lambda_{i_1}^{n_1}\cdots\lambda_{i_k}^{n_k} = S(\lambda_{i_1},\lambda_{i_2},\ldots,\lambda_{i_k}).$$

The relation (9.7.4) connecting $\mathcal{E}$ to $\mathcal{D}$ then allows us to solve for $\mathcal{D}$ as in (9.7.3).

◇

Remark. If we let

$$\mathcal{P}_{i_1,\ldots,i_k} = \text{cum}[\xi_{i_1}(n+1),\ldots,\xi_{i_k}(n+1)]$$

be the state cumulant tensor in the original (real) coordinate system, then again it relates to $\mathcal{D}$ by the congruence transformation

$$\mathcal{P} = W \overset{\mathcal{D}}{\star} \cdots \overset{\mathcal{D}}{\star} W$$

with $\mathcal{P}$ now containing all real entries. The reader may wish to check that, for the second-order case ($k=2$), the expression (9.7.3) reduces to the Pick matrix from (9.3.5).

◇

Note that the M eigenvalues $\lambda_1, \ldots, \lambda_M$ contained in Λ lead in fact to M^k evaluations of the polyspectral residue function $S(z_1,\ldots,z_k)$. A natural question is whether these evaluations might display some form of redundancy, allowing their determination from a reduced set. The following lemma shows that setting successive complex variables to zero (which simplifies the corresponding z-transform evaluations) suffices for retrieving the M^k evaluations of the previous lemma.

Lemma 9.7.2 (Determining Polyspectral Values). *The polyspectral values*

$$S(\lambda_{i_1},\ldots,\lambda_{i_k})$$

are uniquely determined from the set

$$S(\lambda_{i_1},0,\ldots,0),\ S(\lambda_{i_1},\lambda_{i_2},0,\ldots,0),\ \cdots,\ S(\lambda_{i_1},\ldots,\lambda_{i_{k-1}},0), \tag{9.7.5}$$

as each index i_l ranges over its M possibilities.

Remark. This reduces the number of evaluations of $S(z_1, \ldots, z_k)$ from M^k to $M + M^2 + \cdots + M^{k-1}$.

Proof: Set $\bar{\xi}_0(n) = y(n)$ and introduce the augmented vector $[\begin{smallmatrix} y(n) \\ \bar{\xi}(n) \end{smallmatrix}]$; its cumulant tensor (now indexed from zero) is denoted $\overline{\mathcal{D}}$. Its interior elements (all indices greater than zero) yield the elements of $\mathcal{D}$ from (9.7.2), since by stationarity

$$\begin{aligned} \overline{\mathcal{D}}_{i_1,\ldots,i_k} &= \operatorname{cum}[\bar{\xi}_{i_1}(n), \ldots, \bar{\xi}_{i_k}(n)] \\ &= \operatorname{cum}[\bar{\xi}_{i_1}(n+1), \ldots, \bar{\xi}_{i_k}(n+1)] = \mathcal{D}_{i_1,\ldots,i_k} \end{aligned} \tag{9.7.6}$$

whenever all indices are greater than zero. The elements on the faces of $\overline{\mathcal{D}}$ become

$$\mathcal{D}_{i_1,\ldots,i_l,0,\ldots,0} = \operatorname{cum}[\bar{\xi}_{i_1}(n), \ldots, \bar{\xi}_{i_l}(n), \underbrace{y(n), \ldots, y(n)}_{k-l \text{ times}}]$$

$$= \sum_{n_1=1}^{\infty} \cdots \sum_{n_l=1}^{\infty} \operatorname{cum}[y(n-n_1), \ldots, y(n-n_l), y(n), \ldots, y(n)] \lambda_{i_1}^{n_1} \cdots \lambda_{i_l}^{n_l}$$

$$= \lambda_{i_1} \cdots \lambda_{i_l} S(\lambda_{i_1}, \ldots, \lambda_{i_l}, 0, \ldots, 0), \tag{9.7.7}$$

which is one of the values from the set (9.7.5) multiplied by the scale factor $\lambda_{i_1} \cdots \lambda_{i_l}$.

Observe now that

$$\begin{bmatrix} 0 \\ \bar{\xi}(n+1) \end{bmatrix} = \begin{bmatrix} 0 & 0^T \\ 1 & \Lambda \end{bmatrix} \begin{bmatrix} y(n) \\ \bar{\xi}(n) \end{bmatrix} \tag{9.7.8}$$

so that the cumulant tensor formed from $\begin{bmatrix} 0 \\ \bar{\xi}(n+1) \end{bmatrix}$ relates to $\overline{\mathcal{D}}$ by the congruence transformation

$$\begin{bmatrix} 0 & 0^T \\ 1 & \Lambda \end{bmatrix} \overset{\overline{\mathcal{D}}}{\star} \cdots \overset{\overline{\mathcal{D}}}{\star} \begin{bmatrix} 0 & 0^T \\ 1 & \Lambda \end{bmatrix}.$$

This tensor, in turn, has interior elements coinciding with those of $\overline{\mathcal{D}}$ in view of (9.7.6) and vanishes on all faces because the leading entry of the vector (9.7.8) is zero. We deduce that $\overline{\mathcal{D}}$ satisfies the Lyapunov equation

$$\overline{\mathcal{D}} - \begin{bmatrix} 0 & 0^T \\ 1 & \Lambda \end{bmatrix} \overset{\overline{\mathcal{D}}}{\star} \cdots \overset{\overline{\mathcal{D}}}{\star} \begin{bmatrix} 0 & 0^T \\ 1 & \Lambda \end{bmatrix} = \partial \mathcal{D},$$

in which $\partial\mathcal{D}$ vanishes at all interior points and has faces coinciding with those of $\overline{\mathcal{D}}$. Since $[\begin{smallmatrix} 0 & 0 \\ 1 & \Lambda \end{smallmatrix}]$ is a stable matrix, the existence claim of Lemma 9.6.3 implies that $\overline{\mathcal{D}}$ is uniquely determined from the evaluation points $\{\lambda_i\}$, which build up Λ, and the face values gathered in (9.7.5), which generate $\partial\mathcal{D}$.

◇

Theorem 9.7.1 (A Pick Condition). *Given the polyspectral values $S(\lambda_{i_1}, \ldots, \lambda_{i_k})$, a linear process may be fit to these values only if the Pick tensor from* (9.7.3) *is Tucker factorable.*

Proof: To see this, suppose that

$$\hat{y}(n) = \hat{h}_1 u(n-1) + \hat{h}_2 u(n-2) + \cdots$$

is a linear process whose polyspectral function $\widehat{S}(z_1,\ldots,z_k)$ fulfills

$$\widehat{S}(\lambda_{i_1},\ldots,\lambda_{i_k}) = S(\lambda_{i_1},\ldots,\lambda_{i_k}).$$

We know that the cumulant tensor $\widehat{\mathcal{D}}$ generated from $\widehat{S}(\lambda_{i_1},\ldots,\lambda_{i_k})$ can be written as

$$\widehat{\mathcal{D}} = \mathcal{C}_p \overset{\widehat{\mathcal{T}}}{\star} \cdots \overset{\widehat{\mathcal{T}}}{\star} \mathcal{C}_p,$$

where $\widehat{\mathcal{T}}$ is the cumulant tensor built from the sequence $\{\hat{y}(\cdot)\}$. By design, $\widehat{\mathcal{T}}$ is Tucker factorable as

$$\widehat{\mathcal{T}} = \hat{\gamma}_k \cdot \widehat{H}^T \star \cdots \star \widehat{H}^T$$

because the process $\{\hat{y}(\cdot)\}$ is linear, and this gives $\mathcal{D} = \widehat{\mathcal{D}}$ as

$$\mathcal{D} = (\mathcal{C}_p\widehat{H}^T) \star \cdots \star (\mathcal{C}_p\widehat{H}^T),$$

which is Tucker factorable.

◇

One may be tempted to conjecture a result in the converse direction; let us pinpoint a difficulty here. As remarked at the end of Sec. 9.6.1, admissibility conditions for $S(z_1,\ldots,z_k)$ to be a valid polyspectral residue function are not known. Without this, one is confronted with the difficulty of how to verify whether a candidate solution is valid.

9.8 A SCHUR-TYPE ALGORITHM FOR TENSORS

We develop now a candidate procedure for a Schur-type algorithm adapted to higher-order arrays. For the benefit of the nonexpert, we review first the algorithm for second-order arrays, which consists of subtracting off vector outer products to expose the Cholesky factor of a symmetric factorable (or positive-definite) matrix.

We then introduce a tensorial outer product which replicates the faces of a symmetric tensor in terms of a tensor of one lower dimension. Successive subtraction operations annihilate successive faces of a tensor, leading to a type of pyramidal decomposition that reduces to a Cholesky decomposition in the second-order case. The relations with displacement residues conclude this section.

9.8.1 Review of the Second-Order Case

Consider a symmetric matrix R, assumed positive definite and hence factorable in the form $R = L \star L = LL^T$. Let us partition R in the form

$$R = \begin{bmatrix} r_0 & r^T \\ r & R_1 \end{bmatrix},$$

where r_0 is a scalar, r is a column vector, and R_1 is a submatrix. Upon setting

$$a = \begin{bmatrix} \sqrt{r_0} \\ r/\sqrt{r_0} \end{bmatrix}, \tag{9.8.1}$$

one verifies that the outer product aa^T coincides with R on the borders of the matrix and when subtracted from R reveals

$$R - aa^T = \begin{bmatrix} 0 & 0^T \\ 0 & R_1 - rr^T/r_0 \end{bmatrix},$$

in which the lower right block contains the Schur complement $R_1 - rr^T/r_0$ with respect to the leading entry (see also Sec. 1.6.1).

Now, if R is positive definite, we may write $R = LL^T$, in which L is a lower triangular Cholesky factor of R. The vector a from (9.8.1) is simply the first column of L.

The determination of successive columns of the Cholesky factor L has an intimate connection with the displacement structure of R and the Schur algorithm (see, e.g., Ch. 1 and [KS95a]). Consider the displacement residue of a matrix R, written as the sum and difference of vector dyads:

$$\begin{aligned} R - Z \overset{R}{\star} Z = R - ZRZ^T &= \sum_{i=1}^{p} a_i a_i^T - \sum_{i=1}^{q} b_i b_i^T \\ &= \underbrace{\begin{bmatrix} a_1 & \cdots & a_p & b_1 & \cdots & b_q \end{bmatrix}}_{G} \underbrace{\begin{bmatrix} I_p & \\ & -I_q \end{bmatrix}}_{J} G^T. \end{aligned}$$

The generator vectors $\{a_i\}$ and $\{b_i\}$ are said to be proper (see, e.g., [LK84] and Ch. 1) provided that a_1 is the sole vector that is nonzero in its leading entry. One checks readily that the border of R must then be $a_1 a_1^T$, such that a_1 yields the first column of the Cholesky factor of R.

To obtain the remaining columns of the Cholesky factor, let $P = R_1 - rr^T/r_0$ denote the Schur complement. It turns out, when the generator G is proper, that P satisfies an analogous displacement equation [KS95a], viz.,

$$\begin{aligned} &\begin{bmatrix} 0 & 0^T \\ 0 & P \end{bmatrix} - Z \begin{bmatrix} 0 & 0^T \\ 0 & P \end{bmatrix} Z^T \\ = &\underbrace{\begin{bmatrix} Za_1 & a_2 & \cdots & a_p & b_1 & \cdots & b_q \end{bmatrix}}_{G'} J(G')^T, \end{aligned} \tag{9.8.2}$$

whose generator matrix G' relates to G by a down-shift operation on the leading column. (The matrix G' now has zeros on the top row.)

Now, if Θ is any J-unitary matrix, i.e., fulfilling $\Theta J \Theta^T = J$, then $G'\Theta$ remains a generator matrix for the displacement structure of (9.8.2). Whenever the leading entry of the Schur complement P is positive, then one can always determine a J-unitary matrix Θ which renders $G'\Theta$ in proper form, i.e., for which only the first column vector of $G'\Theta$ has a nonzero leading entry (occuring now in the second position); see Ch. 1. The leading column of the resulting $G'\Theta$ then generates the next column of the Cholesky factor L. Successive shift and rotate operations are then applied to yield the successive columns of L. We refer the reader to Ch. 1 and [KS95a], and the references therein, for more detail and applications of these recursions.

We turn now to a candidate extension of this procedure for higher-order tensors.

9.8.2 A Tensor Outer Product

A vector outer product, in the form $a_1 a_2^T = D$, generates a matrix D (i.e., a second-order tensor) obtained from two vectors a_1 and a_2 (i.e., first-order tensors), by projecting the indices onto the respective entries: $R_{i_1,i_2} = (a_1)_{i_1}(a_2)_{i_2}$. An analogous operation is to consider k tensors of order $k-1$, to generate a kth-order tensor by the formulation in the following definition.

Definition 9.8.1 (Tensor Outer Product). *Given k tensors $\mathcal{E}_1, \ldots, \mathcal{E}_k$, each of order $k-1$, their outer product, denoted*

$$\mathcal{D} = \mathcal{E}_1 \odot \mathcal{E}_2 \odot \cdots \odot \mathcal{E}_k,$$

is the kth-order tensor with elements

$$\mathcal{D}_{i_1,i_2,\ldots,i_k} = (\mathcal{E}_1)_{i_2,\ldots,i_k}(\mathcal{E}_2)_{i_1,i_3,\ldots,i_k} \cdots (\mathcal{E}_k)_{i_1,\ldots,i_{k-1}}.$$

$\diamond$

If $\mathcal{E}_1$ and $\mathcal{E}_2$ are column vectors, then $(\mathcal{E}_1 \odot \mathcal{E}_2)_{i_1,i_2} = (\mathcal{E}_1)_{i_2}(\mathcal{E}_2)_{i_1} = (\mathcal{E}_2\mathcal{E}_1^T)_{i_1,i_2}$, which reduces to the conventional outer product of vectors. If $\mathcal{E}_1$, $\mathcal{E}_2$, and $\mathcal{E}_3$ are three matices, the generation of the third-order tensor $\mathcal{E}_1 \odot \mathcal{E}_2 \odot \mathcal{E}_3$ is as illustrated in Fig. 9.2. Each matrix is set adjacent to a face of the cube, and each element of the cube is obtained by projecting the coordinates onto the three faces and multiplying the matrix elements occuring on the three faces.

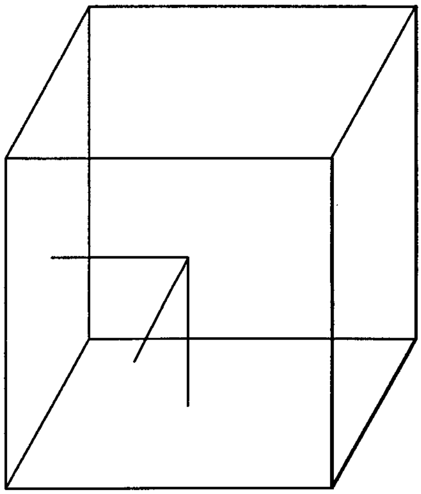

Figure 9.2. Geometric interpretation of the outer product of three matrices: Each matrix occupies one face of the cube (front, left, and bottom), and each interior element of the tensor is obtained by projecting its coordinates onto the three faces and multiplying the resulting elements.

The utility of this product can be appreciated by returning to the Toeplitz tensor $\mathcal{T}$ containing output cumulants. Consider first the third-order case, i.e.,

$$\mathcal{T}_{i_1,i_2,i_3} = \text{cum}[y(n-i_1), y(n-i_2), y(n-i_3)],$$

with $\mathcal{T}_{i_1,i_2,0} = c_{i_1,i_2}$ (the third-order cumulant with lags i_1 and i_2). The displacement residue $\mathcal{T} - Z \overset{\mathcal{T}}{\star} Z \overset{\mathcal{T}}{\star} Z$ coincides with $\mathcal{T}$ along the faces and vanishes at all interior points. The displacement residue can be expressed as

$$\mathcal{T} - Z \overset{\mathcal{T}}{\star} Z \overset{\mathcal{T}}{\star} Z = \mathcal{E}_1 \odot \mathcal{E}_1 \odot \mathcal{E}_1 - \mathcal{E}_2 \odot \mathcal{E}_2 \odot \mathcal{E}_2,$$

in which

$$(\mathcal{E}_1)_{i_1,i_2} = \begin{cases} \sqrt[3]{c_{0,0}}, & i_1 = i_2 = 0, \\ \sqrt{\frac{c_{i_1,0}}{(\mathcal{E}_1)_{00}}}, & i_1 > 0 \text{ and } i_2 = 0, \\ \sqrt{\frac{c_{0,i_2}}{(\mathcal{E}_1)_{00}}}, & i_1 = 0 \text{ and } i_2 > 0, \\ \frac{c_{i_1,i_2}}{(\mathcal{E}_1)_{i_1,0}(\mathcal{E}_1)_{0,i_2}} & \text{elsewhere,} \end{cases}$$

and

$$(\mathcal{E}_2)_{i_1,i_2} = \begin{cases} 0, & i_1 = 0 \text{ and/or } i_2 = 0, \\ (\mathcal{E}_1)_{i_1,i_2} & \text{elsewhere.} \end{cases}$$

The first term $\mathcal{E}_1 \odot \mathcal{E}_1 \odot \mathcal{E}_1$ generates the faces of $\mathcal{T}$, whereas the second term $\mathcal{E}_2 \odot \mathcal{E}_2 \odot \mathcal{E}_2$ vanishes along each face while replicating the interior terms of $\mathcal{E}_1 \odot \mathcal{E}_1 \odot \mathcal{E}_1$. Observe that $\mathcal{E}_1$ also can be written directly in terms of cumulant values c_{i_1,i_2} as

$$(\mathcal{E}_1)_{i_1,i_2} = c_{i_1,i_2} \frac{\sqrt[3]{c_{00}}}{\sqrt{c_{i_1,0}\, c_{0,i_2}}}.$$

This decomposition extends readily to high-order tensors as well. If $\mathcal{T}$ is a kth-order Toeplitz cumulant tensor, its displacement residue can be decomposed as

$$\mathcal{T} - Z \overset{\mathcal{T}}{\star} Z \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} Z = \underbrace{\mathcal{E}_1 \odot \cdots \odot \mathcal{E}_1}_{k \text{ terms}} - \underbrace{\mathcal{E}_2 \odot \cdots \odot \mathcal{E}_2}_{k \text{ terms}}, \tag{9.8.3}$$

where now

$$(\mathcal{E}_2)_{i_1,\dots,i_{k-1}} = \begin{cases} 0, & i_1 = 0 \text{ and/or } \dots i_{k-1} = 0, \\ (\mathcal{E}_1)_{i_1,\dots,i_{k-1}} & \text{elsewhere,} \end{cases} \tag{9.8.4}$$

and [Gri96]

$$\begin{aligned} (\mathcal{E}_1)_{i_1,\dots,i_{k-1}} &= c_{i_1,\dots,i_{k-1}} \\ &\times \left[c_{i_1,0,\dots,0}\, c_{i_2,0,\dots,0} \cdots c_{i_{k-1},0,\dots,0}\right]^{-1/2} \\ &\times \left[c_{i_1,i_2,0,\dots,0} c_{i_2,i_3,0,\dots,0} \cdots c_{i_{k-1},i_1,0,\dots,0}\right]^{1/3} \\ &\times \cdots \times \left[c_{0,\dots,0}\right]^{(-1)^{k+1}/k}. \end{aligned} \tag{9.8.5}$$

When specialized to the matrix case, the decomposition reduces to the outer product decomposition of the displacement structure of a Toeplitz matrix, as illustrated in (9.3.3). Some potential weaknesses of this higher-order extension, however, are worth noting. First, the generation of $\mathcal{E}_1$ involves various division and rooting operations, and a potential division by zero is not to be discarded immediately, indicating that this decomposition need not always exist. For the second-order case by contrast [cf. (9.3.3)], all divisions are by $\sqrt{r_0}$, whose positivity is assured whenever the stochastic process $\{y(\cdot)\}$ is nontrivial. Second, even if $\mathcal{T}$ has all real entries, those of $\mathcal{E}_1$ and $\mathcal{E}_2$ may be complex, due to the various radicals involved in their generation. Finally, the transform domain relation is more complicated with this product. To illustrate, consider the third-order case, for which $\mathcal{E}_1$ becomes a matrix. Let us set

$$\mathcal{E}(z_1, z_2) = \sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} \mathcal{E}_{i_1,i_2}\, z_1^{i_1} z_2^{i_2}$$

and $\mathcal{D} = \mathcal{E} \odot \mathcal{E} \odot \mathcal{E}$. The function $\mathcal{D}(z_1, z_2, z_3) = \sum \mathcal{D}_{i_1,i_2,i_3} z_1^{i_1} z_2^{i_2} z_3^{i_3}$ may then be expressed as the convolutional formula [Gri96]

$$\mathcal{D}(z_1, z_2, z_3)$$

$$= \frac{1}{(2\pi j)^3} \oint_{|w_1|=1} \oint_{|w_2|=1} \oint_{|w_3|=1} \mathcal{E}(w_1, w_2)\, \mathcal{E}\left(\frac{z_3}{w_3}, w_1\right) \mathcal{E}\left(\frac{z_1}{w_1}, \frac{z_2}{w_2}\right) \frac{dw_1}{w_1} \frac{dw_2}{w_2} \frac{dw_3}{w_3}.$$

For the second-order case, by contrast, one finds that $\mathcal{D} = \mathcal{E} \odot \mathcal{E}$ induces the much simpler multiplicative structure $\mathcal{D}(z_1, z_2) = \mathcal{E}(z_1)\, \mathcal{E}(z_2)$.

Despite these complications, some interesting relations with displacement generators do fall out, which we now address.

9.8.3 Displacement Generators

It is natural to consider a "Schur complement" of a tensor as that obtained by subtracting from a given symmetric tensor an outer product of lower-order tensors that eliminates the face elements. Although the expressions to be developed likewise apply to nonsymmetric tensors [Gri96], the symmetric case pursued here affords simpler notations.

For third-order symmetric tensors, this Schur complement operation appears as

$$(\mathcal{T}^{(1)})_{i_1,i_2,i_3} = (\mathcal{T})_{i_1,i_2,i_3} - (\mathcal{T}_{000} \mathcal{T}_{i_100}^{-1} \mathcal{T}_{0i_20}^{-1} \mathcal{T}_{00i_3}^{-1}) \mathcal{T}_{0i_2i_3} \mathcal{T}_{i_10i_3} \mathcal{T}_{i_1i_20}$$

or

$$\mathcal{T}^{(1)} = \mathcal{T} - \mathcal{E} \odot \mathcal{E} \odot \mathcal{E},$$

in which

$$\mathcal{E}_{i_1,i_2} = \mathcal{T}_{i_1,i_2,0} \frac{\sqrt[3]{\mathcal{T}_{000}}}{\sqrt{\mathcal{T}_{i_1,0,0} \mathcal{T}_{i_2,0,0}}},$$

thereby eliminating the faces of $\mathcal{T}$. For higher-order tensors, the operation appears as

$$(\mathcal{T}^{(1)})_{i_1,i_2,\ldots,i_k} = (\mathcal{T})_{i_1,i_2,\ldots,i_k} - \prod_{l=0}^{k} \left(\prod_{\sigma} \mathcal{T}_{j_1,j_2,\ldots,j_l,0,\ldots,0} \right)^{(-1)^{l+k+1}},$$

where $\sigma = \binom{i_1\ i_2\ \cdots\ i_l\ i_{l+1}\ \cdots\ i_k}{j_1\ j_2\ \cdots\ j_l\ j_{l+1}\ \cdots\ j_k}$ is a circular permutation of indices.

Suppose now that the symmetric tensor $\mathcal{T}$ has as displacement structure the outer product decomposition

$$\mathcal{T} - Z \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} Z = \sum_{l=1}^{L} \pm \mathcal{E}_l^{\odot k}, \tag{9.8.6}$$

in which $\mathcal{E}_l^{\odot k}$ denotes the k-term product $\mathcal{E}_l \odot \cdots \odot \mathcal{E}_l$. In analogy with the second-order case, the generator tensors $\{\mathcal{E}_l\}$ will be termed proper provided that $\mathcal{E}_1$ is the only generator that is nonzero on its faces. The displacement decomposition (9.8.3)–(9.8.5) for a Toeplitz tensor is proper, for example. The face elements of $\mathcal{T}$ then come from the contribution of $\mathcal{E}_1$ alone, such that the Schur complement of $\mathcal{T}$, by the above definition, becomes $\mathcal{T}^{(1)} = \mathcal{T} - \mathcal{E}_1^{\odot k}$.

Extraction of the next Schur complement appears as

$$\mathcal{T}^{(2)} = \mathcal{T}^{(1)} - (\mathcal{E}_1^{(1)})^{\odot k},$$

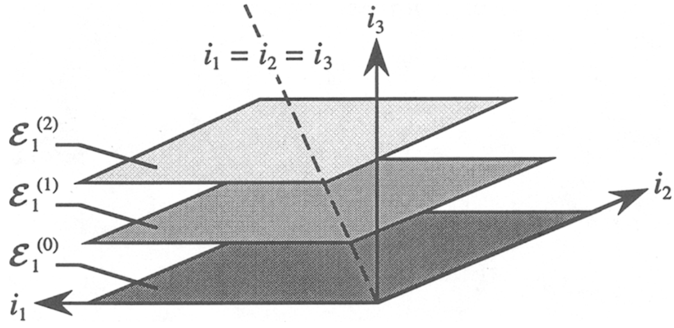

Figure 9.3. Illustrating a pyramidal structure for a third-order tensor.

in which $(\mathcal{E}_1^{(1)})^{\odot k}$ generates each first nonzero face of $\mathcal{T}^{(1)}$, such that $\mathcal{T}^{(2)}$ now vanishes in its first subfaces as well. Upon iterating this process, we generate successive $(k-1)$-dimensional tensors $\mathcal{E}_1^{(l)}$, vanishing in the first l subfaces and such that the tensor difference

$$\mathcal{T} - \sum_{l=1}^{n} (\mathcal{E}_1^{(l)})^{\odot k}$$

vanishes whenever all indices are less than n. A new kth-order tensor, defined as

$$\mathcal{D}_{i_1,\ldots,i_{k-1},l} = (\mathcal{E}_1^{(l)})_{i_1,\ldots,i_{k-1}} \tag{9.8.7}$$

for successive values of l, then assumes a "pyramidal" structure, which reduces to a triangular (or Cholesky) structure when specialized to second-order tensors (or matrices). Figure 9.3 illustrates the pyramidal structure for third-order tensors.

We now relate successive subtensors of $\mathcal{D}$ from (9.8.7) to the displacement structure of (9.8.6). Thus, suppose that the generator tensors in (9.8.6) are proper, i.e., only $\mathcal{E}_1$ is nonzero along any face. Define a shifted first generator $\mathcal{E}_1'$ by shifting all elements of $\mathcal{E}_1$ one position along its main diagonal, i.e.,

$$\mathcal{E}_1' = \underbrace{Z \overset{\mathcal{E}_1}{\star} \cdots \overset{\mathcal{E}_1}{\star} Z}_{k-1 \text{ terms}}.$$

We then have the following result.

Lemma 9.8.1 (Structure of Schur Complements). *The Schur complement* $\mathcal{T}^{(1)} = \mathcal{T} - \mathcal{E}_1 \odot \cdots \odot \mathcal{E}_1$ *satisfies the displacement equation*

$$\mathcal{T}^{(1)} - Z \overset{\mathcal{T}^{(1)}}{\star} \cdots \overset{\mathcal{T}^{(1)}}{\star} Z = (\mathcal{E}_1')^{\odot k} + \sum_{l=2}^{L} \pm \mathcal{E}_l^{\odot k}$$

obtained by shifting the first generator of a proper set.

Proof: Note that shifting each entry of $\mathcal{E}_1$ by one index simply shifts each entry of the outer product $\mathcal{E}_1^{\odot k}$ by one index, so that

$$(\mathcal{E}_1')^{\odot k} = \underbrace{Z \overset{\mathcal{E}_1^{\odot k}}{\star} \cdots \overset{\mathcal{E}_1^{\odot k}}{\star} Z}_{k \text{ terms}}.$$

We then find by direct calculation that

$$\begin{aligned} \mathcal{T}^{(1)} - Z \overset{\mathcal{T}^{(1)}}{\star} \cdots \overset{\mathcal{T}^{(1)}}{\star} Z &= \left(\mathcal{T} - Z \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} Z\right) - \left(\mathcal{E}_1^{\odot k} - Z \overset{\mathcal{E}_1^{\odot k}}{\star} \cdots \overset{\mathcal{E}_1^{\odot k}}{\star} Z\right) \\ &= \left(\mathcal{E}_1^{\odot k} + \sum_{l=2}^{L} \pm \mathcal{E}_l^{\odot k}\right) - \left(\mathcal{E}_1^{\odot k} - (\mathcal{E}_1')^{\odot k}\right) \\ &= (\mathcal{E}_1')^{\odot k} + \sum_{l=2}^{L} \pm \mathcal{E}_l^{\odot k} \end{aligned}$$

as claimed.

$\diamond$

The next step in the algorithm is to transform the resulting generators $\mathcal{E}_1'$ and $\{\mathcal{E}_l\}_{l \geq 2}$ into a proper set, since the leading generator tensor would then be identified as the next subtensor $\mathcal{E}_1^{(1)}$ of the pyramidal factor $\mathcal{D}$ from (9.8.7).

At this point a major difference arises compared to the second-order (or matrix) case. In the second-order case, the transformation amounts to arranging the generator vectors into a matrix, which is then multiplied by a J-unitary transformation to zero out preassigned terms. The higher-order analogy of this operation instead involves nonlinear operations, which we shall illustrate with a simplified example adapted from [Gri96].

Suppose we begin with a third-order Toeplitz tensor for $\mathcal{T}$, and let $\nabla\mathcal{T}^{(1)} = \mathcal{T}^{(1)} - Z \overset{\mathcal{T}^{(1)}}{\star} Z \overset{\mathcal{T}^{(1)}}{\star} Z$ be the displacement residue from the first Schur complement. By construction, this tensor vanishes along all faces, so we consider only nonzero indices. For the Toeplitz case considered, the displacement equation involves only two generators, i.e.,

$$\nabla\mathcal{T}^{(1)} = (\mathcal{E}_1')^{\odot 3} - \mathcal{E}_2^{\odot 3}.$$

We now seek a pair of *proper* generators $\mathcal{F}_1$ and $\mathcal{F}_2$, compatible with $\nabla\mathcal{T}^{(1)}$, i.e.,

$$\nabla\mathcal{T}^{(1)} = \mathcal{F}_1^{\odot 3} - \mathcal{F}_2^{\odot 3}.$$

We require that $\mathcal{F}_2$ vanish along its first two faces and hence that only $\mathcal{F}_1$ be nonzero along its first face. From the equation

$$(\nabla\mathcal{T}^{(1)})_{111} = (\mathcal{E}_1')_{11}^3 - (\mathcal{E}_2)_{11}^3 = (\mathcal{F}_1)_{11}^3,$$

we deduce the first nonzero element of $\mathcal{F}_1$ as

$$(\mathcal{F}_1)_{11} = \sqrt[3]{(\mathcal{E}_1')_{11}^3 - (\mathcal{E}_2)_{11}^3}\,.$$

The conditions

$$(\nabla\mathcal{T}^{(1)})_{1,1,i} = (\mathcal{E}_1')_{1,1}(\mathcal{E}_1')_{1,i}^2 - (\mathcal{E}_2)_{1,1}(\mathcal{E}_2)_{1,i}^2 = (\mathcal{F}_1)_{1,1}(\mathcal{F}_1)_{1,i}^2$$

give the elements of the first nonzero row and column of $\mathcal{F}_1$ as

$$(\mathcal{F}_1)_{1,i} = (\mathcal{F}_1)_{i,1} = \sqrt{\frac{(\mathcal{E}_1')_{1,1}(\mathcal{E}_1')_{1,i}^2 - (\mathcal{E}_2)_{1,1}(\mathcal{E}_2)_{1,i}^2}{(\mathcal{F}_1)_{1,1}}}.$$

To obtain the remaining elements of $\mathcal{F}_1$, let us set

$$\alpha_i = \frac{(\mathcal{E}_1')_{1,i}}{(\mathcal{F}_1)_{1,i}} \quad \text{and} \quad \beta_i = \frac{(\mathcal{E}_2)_{1,i}}{(\mathcal{F}_1)_{1,i}}.$$

Using the general relation

$$
\begin{aligned}
(\nabla \mathcal{T}^{(1)})_{i_1,i_2,1} &= (\mathcal{E}_1')_{i_1,1}(\mathcal{E}_1')_{1,i_2}(\mathcal{E}_1')_{i_1,i_2} - (\mathcal{E}_2)_{i_1,1}(\mathcal{E}_2)_{1,i_2}(\mathcal{E}_2)_{i_1,i_2} \\
&= (\mathcal{F}_1)_{i_1,1}(\mathcal{F}_1)_{1,i_2}(\mathcal{F}_1)_{i_1,i_2},
\end{aligned}
$$

we obtain

$$
\begin{aligned}
\mathcal{F}_{i_1,i_2} &= \frac{(\mathcal{E}_1')_{i_1,1}(\mathcal{E}_1')_{1,i_2}(\mathcal{E}_1')_{i_1,i_2} - (\mathcal{E}_2)_{i_1,1}(\mathcal{E}_2)_{1,i_2}(\mathcal{E}_2)_{i_1,i_2}}{(\mathcal{F}_1)_{i_1,1}(\mathcal{F}_1)_{1,i_2}} \\
&= \alpha_{i_1}\,\alpha_{i_2}(\mathcal{E}_1')_{i_1,i_2} - \beta_{i_1}\,\beta_{i_2}(\mathcal{E}_2)_{i_1,i_2}.
\end{aligned}
$$

A similar procedure may be carried out for $\mathcal{F}_2$, whose first nonzero elements begin with indices 2 or greater, since it must vanish along its first two faces. Starting now with the pivot $(\nabla \mathcal{T}^{(1)})_{2,2,2}$, we have that

$$
(\nabla \mathcal{T}^{(1)})_{2,2,2} = (\mathcal{F}_1)_{2,2}^3 - (\mathcal{F}_2)_{2,2}^3
$$

or

$$
(\mathcal{F}_2)_{2,2} = \sqrt[3]{(\mathcal{F}_1)_{2,2}^3 - (\nabla \mathcal{T}^{(1)})_{2,2,2}}\,,
$$

which gives the leading entry. From the relation

$$
(\nabla \mathcal{T}^{(1)})_{2,2,i} = (\mathcal{F}_1)_{2,2}(\mathcal{F}_1)_{2,i}^2 - (\mathcal{F}_2)_{2,2}(\mathcal{F}_2)_{2,i}^2
$$

we deduce the elements of the first nonzero row and column of $\mathcal{F}_2$ as

$$
(\mathcal{F}_2)_{2,i} = (\mathcal{F}_2)_{i,2} = \sqrt{\frac{(\mathcal{F}_1)_{2,2}(\mathcal{F}_1)_{2,i}^2 - (\nabla \mathcal{T}^{(1)})_{2,2,i}}{(\mathcal{F}_2)_{2,2}}}.
$$

Finally, from the general relation

$$
(\nabla \mathcal{T}^{(1)})_{i_1,i_2,2} = (\mathcal{F}_1)_{i_1,2}(\mathcal{F}_1)_{2,i_2}(\mathcal{F}_1)_{i_1,i_2} - (\mathcal{F}_2)_{i_1,2}(\mathcal{F}_2)_{2,i_2}(\mathcal{F}_2)_{i_1,i_2}
$$

we obtain the formula for the remaining elements of $\mathcal{F}_2$ as

$$
(\mathcal{F}_2)_{i_1,i_2} = \frac{(\mathcal{F}_1)_{i_1,2}(\mathcal{F}_1)_{2,i_2}(\mathcal{F}_1)_{i_1,i_2} - (\nabla \mathcal{T}^{(1)})_{i_1,i_2,2}}{(\mathcal{F}_2)_{i_1,2}(\mathcal{F}_2)_{2,i_2}}.
$$

We observe that obtaining a proper generator set from a nonproper set does not reduce to linear transformations. Whether a more expedient procedure to that outlined above may be obtained, and whether some physically relevant interpretation may be attached to the intermediate terms of this procedure, are topics which require further study.

9.9 CONCLUDING REMARKS

We have presented a semitutorial account of tensorial representations for cumulants in system theory, with special emphasis on Tucker factorability for linear process. The key result concerning the Tucker factorability of a Pick tensor for the linear case is apparently one of the few results establishing compatibility of cumulants with a linear process given only partial information. Usable conditions guaranteeing Tucker factorability of a given tensor require further study, as do methods for reliably deducing such a factor when one exists. Further study is likewise required to establish "admissibility" conditions for higher-order spectra.

Another avenue worthy of further study relates to the faces of a Toeplitz cumulant tensor coinciding with its Hankel-based counterpart in the linear case. Given thus the faces of a Toeplitz tensor, a useful query asks how to reconstruct candidate interior points compatible with a Hankel-based tensor. If such a procedure could be rendered successful, then Lemma 9.6.1 shows that a simple displacement operation reveals a rank 1 tensor built from the system impulse response.

A candidate Schur algorithm for pyramidal factorization of higher-order tensors has also been proposed which, when specialized to second-order arrays, reduces to a recursive algorithm for triangular factorization. Whether the given procedure may have some utility in checking for Tucker factorability is not immediately clear. Similarly, whether deeper connections with modeling filter synthesis in the linear case [LK84] are inherited by this procedure remains to be investigated.

Acknowledgments

The authors would like to thank P. Comon and L. Delathauwer for their critical comments on some earlier versions of this work. They are also grateful to Professors A. H. Sayed and T. Kailath for their assistance in adapting this work to the present volume.