STATISTICAL SIGNAL PROCESSING, HIGHER ORDER TOOLS

Statistical (or stochastic) signal processing deals with random signals, their transformation by system operators, and their characterization using time- or frequency-domain statistical descriptors computed from measured data records. Signals are treated as random when no exact expression describes their evolution and the engineer has incomplete knowledge about their description or origin. Random signals are temporal or spatial and originate from sources that are man-made (e.g., binary communication signals) or natural (e.g., thermal or ambient noise). They may be continuous or discrete in their amplitude or index, but their processing is usually performed in the discrete-index (sampled if originally continuous) domain.

Statistical descriptors are deterministic quantities and reflect one's degree of knowledge or ignorance about randomness. Complete statistical description of random signals is provided by probability density and distribution functions. Gaussianity, for example, refers to a specific distribution that is characterized completely by its first- and second-order statistics (mean and variance). It is often encountered in practice because, thanks to the central limit theorem, averaging a sufficient number of random signal values (an operation often performed by, e.g., narrowband filtering) yields outputs that

are (at least approximately) distributed according to the Gaussian probability law. However, non-Gaussian signals are also encountered in sciences and engineering applications: exploration seismology, communications, sonar, radar, speech and image processing, fluid mechanics, atmospheric sciences, econometrics and medicine, to name only a few. The following example deals with real seismic reflectivity data and motivates characterization of non-Gaussian signals with higher-than second-order statistics.

Example 1. *Statistical analysis of non-Gaussian data (seismic reflectivity).* Figure 1(a) shows a sampled waveform, $\{x(n)\}_{n=0}^{N-1}$, received by a hydrophone in a marine oil exploration experiment. It is the output of an airgun's input signature fired in the ocean (the so called seismic wavelet) convolved with the earth's impulse response, which is known as the reflectivity sequence. The histogram in Fig. 1(b) reveals that $x(n)$ has mean $\hat{c}_{1x} = N^{-1} \sum_{n=0}^{N-1} x(n) = 0$ and variance $\hat{c}_{2x} = N^{-1} \sum_{n=0}^{N-1} x^2(n) = 1.0033 \times 10^3$. If we normalize $x(n)$ by its standard deviation (std) to obtain $\tilde{x}(n) = x(n)/\sqrt{\hat{c}_{2x}}$, which has zero mean and unit variance, then $\tilde{x}(n)$ is seen to have small third-order moment (skewness) $\hat{c}_{3\tilde{x}} = N^{-1} \sum_{n=0}^{N-1} \tilde{x}^3(n) = 0.1275$, but relatively large fourth-order cumulant (kurtosis) $\hat{c}_{4\tilde{x}} = N^{-1} \sum_{n=0}^{N-1} \tilde{x}^4(n) - 3\hat{c}_{2\tilde{x}}^2 = 2.0883$. Since the distribution is fairly symmetric, third-order statistics do not convey significant information about the system, and hence we favor the use of fourth-order statistics for seismic data processing.

MOTIVATION AND OVERVIEW

A histogram and its moments provide only marginal average information about a random process. Relationship among samples at different index points is conveyed by the notions of correlation and (in)dependence, which, together with the notion of stationarity, provide information about the signal's dynamical behavior and memory as it evolves in time or space. Stationary random signals have invariant probabilistic structure relative to the index (e.g., time) of reference. The autocorrelation sequence (ACS) of a stationary signal is time-invariant and—together with its Fourier transform (FT), known as the power spectral density (PSD)—offers sufficient time- and frequency-domain description for Gaussian processes. The PSD describes how average power (or signal variance) is distributed over frequencies, but for non-Gaussian processes, higher-order spectral densities (or polyspectra) are useful as well, because they capture distributions of higher-order signal moments across frequencies.

Statistical descriptors are also important when stationary input signals pass through linear time-invariant (LTI) systems to yield random outputs. If we denote the input by $w(n)$, system impulse response (IR) by $h(n)$, and output by $x(n)$, then we can express $x(n)$ as a linear convolution of $h(n)$ and $w(n)$:

$$x(n) = \sum_l h(l)w(n-l) = \sum_l w(l)h(n-l) \tag{1}$$

The FT of $h(n)$ is called the frequency response,

$$H(\omega) = \sum_n h(n)e^{-j\omega n} \tag{2}$$

and they both characterize the LTI system uniquely. The magnitude response $|H(\omega)|$ or phase response $\angle H(\omega)$ alone does not offer a complete description of the system in general.

The input–output auto- and cross-correlations and spectra characterize not only the random signals themselves but also the transformation induced by the underlying system. For example, when $w(n)$ is white, the output PSD can be shown to be $S_{2x}(\omega) = \sigma_w^2 |H(\omega)|^2$, where σ_w^2 is the variance of the input in (1). Therefore, the output PSD contains the magnitude but not the phase information of the system.

For parsimony, LTI systems are often modeled in the z-domain using finite-order parametric rational functions:

$$H(z) = \sum_{n=0}^{\infty} h(n)z^{-n} \tag{3}$$

$$= \frac{B(z)}{A(z)} = \frac{\sum_{n=0}^{q} b_n z^{-n}}{\sum_{n=0}^{p} a_n z^{-n}} \tag{4}$$

The roots of $A(z)$ and $B(z)$ are called the *poles* and *zeros,* respectively. If a system is causal and stable, then its poles must lie inside the unit circle. If in addition, the zeros are also positioned inside the unit circle, then the system is called *minimum-phase.* It is known that second-order statistics

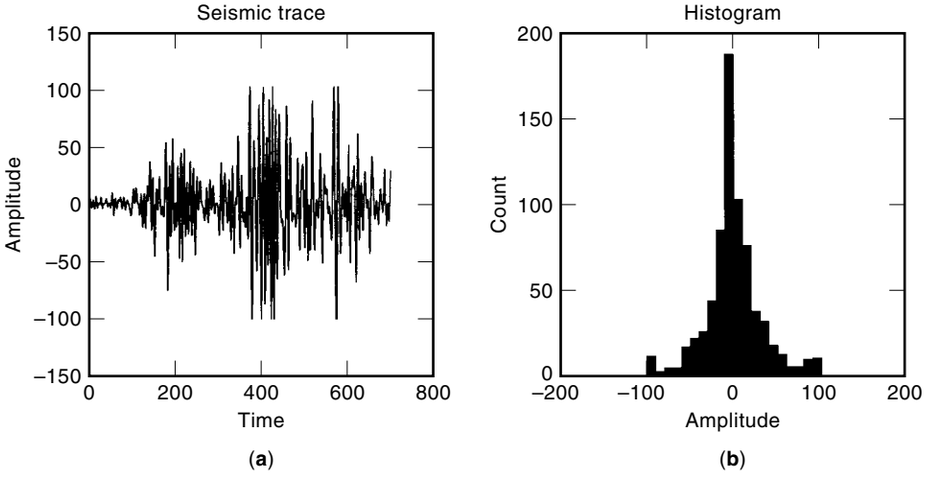

Figure 1. (a) Samples of a seismic waveform. (b) Histogram showing that the seismic signal has a symmetric but non-Gaussian distribution.

(SOS) of the output are sufficient to identify uniquely (modulo scale constants) minimum-phase, causal, and stable LTI systems (see Refs. 1 and 2). In order to assure system stability, we do not allow poles on the unit circle. Besides that, mixed-phase (or non-minimum-phase) and noncausal LTI systems can have poles and zeros that lie anywhere on the complex plane. Higher-order statistics (HOS) of the output convey complementary phase information and can be used to identify uniquely the underlying system (see Refs. 3–6). Nonminimum-phase and noncausal LTI system models appear with frequency-selective communication channels, optical point-spread and image-blurring functions, and seismic wavelets, and we shall elaborate on the latter in the ensuing example.

Example 2. *Nonminimum-phase and noncausality properties of linear time-invariant systems (seismic wavelets).* Figure 2(a) depicts the impulse response $\{h(n)\}$ of a minimum-phase wavelet with transfer function having five poles at 0.7466, $0.7996e^{\pm 0.9995j}$, $0.7553e^{\pm 0.7844j}$ and five zeros at 0.9973, $0.8762e^{\pm 1.2515j}$, $0.2600e^{\pm 1.9626j}$, where $j = \sqrt{-1}$. The correlation of the wavelet, $h_2(\tau) := \sum_n h(n)h(n+\tau)$, is shown in Fig. 2(b). It can be shown that the FT of $h_2(\tau)$ is $H_2(\omega) := |H(\omega)|^2$, which contains magnitude information of the system only. A nonminimum-phase wavelet obtained by replacing the poles at $0.7553e^{\pm 0.7844j}$ and the zeros at $0.2600e^{\pm 1.9626j}$ by their complex conjugate reciprocals has very different impulse response, as shown in Fig. 2(c); but its correlation, plotted in Fig. 2(d), (and hence its magnitude response) is identical with that of the minimum phase wavelet [cf. Fig. 2(b)]. Such wavelets are called *spectrally equivalent,* and their difference lies only in the transfer-function phase $\angle H(\omega)$ [compare Fig. 2(e,g)].

Second-order correlations are symmetric, which implies that the corresponding power spectra have zero phase; that is, they are *phase-blind* when it comes to discerning pole–zero transfer functions that differ in their phase characteristics. We are thus motivated to plot in Fig. 2(f,h) the third-order correlations,

$$h_3(\tau_1, \tau_2) := \sum_n h(n)h(n+\tau_1)h(n+\tau_2) \tag{5}$$

corresponding to the wavelets in Fig. 2(a,c), respectively.

The 2-D FT of $h_3(\tau_1, \tau_2)$,

$$H_3(\omega_1, \omega_2) := \sum_{\tau_1}\sum_{\tau_2} h_3(\tau_1, \tau_2)e^{-j\omega_1\tau_1}e^{-j\omega_2\tau_2} \tag{6}$$

can be shown to satisfy

$$H_3(\omega_1, \omega_2) = H(\omega_1)H(\omega_2)H^*(\omega_1 + \omega_2) \tag{7}$$

Comparing $H_3(\omega_1, \omega_2)$ with $H_2(\omega)$, we see that the former contains phase information whereas the latter does not. This is attributed to the fact that $h_3(\tau_1, \tau_2) \neq h_3(-\tau_1, -\tau_2)$ in general, although $h_2(\tau) = h_2(-\tau)\ \forall\tau$. It is this phase information that allows for the differentiation of the causal minimum-phase wavelet from its spectrally equivalent noncausal and nonminimum-phase counterpart [cf. Fig. 2(f) with Fig. 2(h)].

HOS play a significant role not only in characterizing LTI systems, but also in nonlinear transformation of random signals. Consider for example the second-order sample moment,

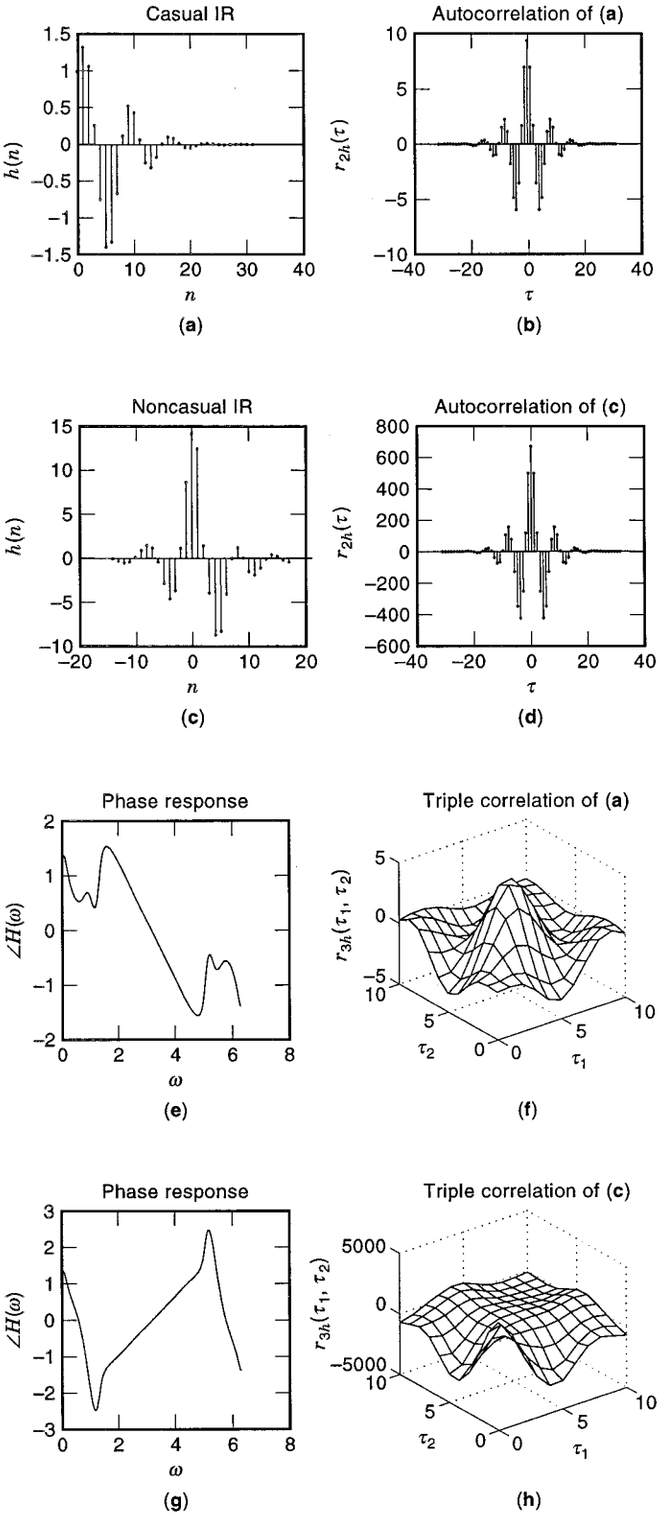

Figure 2. Two linear systems can have the same second-order statistics but different third-order statistics. (a) and (c) are two distinct impulse responses, and (b) and (d) are their autocorrelation functions. Except for a scale factor, the second-order statistics are identical, which implies that the two systems have identical magnitude responses. However, their phase responses are different as shown in (e) and (g). The triple correlation functions in (f) and (h) can characterize their difference in the phase.

$\hat{m}_{2x} := N^{-1} \sum_{n=0}^{N-1} x^2(n)$, of a square-law device $x(n) = w^2(n)$. Clearly, $\hat{m}_{2x}$ of the output corresponds to the fourth-order sample moment $\hat{m}_{4w}$ of the input $w(n)$. In addition, $x(n)$ is non-Gaussian even if $w(n)$ is Gaussian. Nonlinearities appear in diverse signal-processing problems ranging from high-power amplifiers operating in their saturation region, magnetic recording channels, ocean wave interactions, and ambient shipping noise, to physiological models of the auditory and nervous systems. HOS provide useful cues for determining the order and degree of nonlinear effects, for identifying nonlinear models of the Volterra type, and for detecting and characterizing harmonic coupling effects such as those described in the next example.

Example 3. *Coupling effects in signals undergoing nonlinear transformations.* A signal consisting of two harmonics, $w(n) = \exp[j(\omega_1 n + \phi_1)] + 1.5 \exp[j(\omega_2 n + \phi_2)]$ with (normalized) frequencies $\omega_1 = 0.5$ and $\omega_2 = 1.5$, undergoes linear $x(n) = \sum_l h(l)w(n - l) + v(n)$ and nonlinear $x(n) = \sum_l h(l)w(n - l) + 0.8\ w^2(n) + v(n)$ transformations. Additive Gaussian noise $v(n)$ is present in $x(n)$, and the PSD estimates are obtained as $|X(\omega)|^2/N$, where $X(\omega)$ is the FT of $x(n)$ and N is the data length. The PSD estimate of the linear output is shown in Fig. 3(a) and is seen to peak only at the original input frequencies ω_1 and ω_2. The PSD estimate of the nonlinear output is plotted in Fig. 3(b). It exhibits spectral lines not only at ω_1, ω_2 due to the linear effect, but also at $2\omega_1$, $2\omega_2$ arising from the quadratic part, as well as at the coupled frequency $\omega_1 + \omega_2$. By inspecting the PSD only, it is not possible to determine quadratic coupling effects, simply because a frequency at (or close to) $\omega_1 + \omega_2$ may happen to be present in $w(n)$. However, as we shall see later, HOS are capable of discerning not only coupled frequencies but also coupled phases (presence of phase $\phi_1 + \phi_2$ for the $\omega_1 + \omega_2$ frequency component) in $x(n)$. Phase plus frequency couplings provide strong indication of nonlinear (in this example quadratic) interactions.

In a nutshell, HOS augment the traditional notion of correlations and spectra, and are well motivated for signals and systems topics dealing with non-Gaussianity, nonminimum-phase character, noncausality, and nonlinearity. In this article we focus on discrete-index stationary time series, and our goal is to offer basic background material, and then to delineate advantages and limitations of HOS in discerning non-Gaussian from Gaussian signals, recovering phase information, checking for and identifying nonlinearities, and separating sources in multichannel processing. Critical concluding remarks, topics not covered, and suggestions for future research directions are given in the final section. Recent tutorials in HOS-based statistical signal processing include Refs. 7–10, while nonrandom signals are treated in Ref. 11 (see also Ref. 12 for an updated and comprehensive bibliography).

BASIC BACKGROUND

In this section we generalize notions of ensemble correlations

$$m_{2x}(\tau) := E\{x(n)x(n+\tau)\} \tag{8}$$

(E denotes expectation) and the correlation spectrum [PSD if $x(n)$ has zero mean],

$$M_{2x}(\omega) := \sum_{\tau} m_{2x}(\tau) e^{-j\omega\tau} \tag{9}$$

to higher-order correlations, cumulants, and their Fourier transforms known as polyspectra, and we summarize certain properties of interest to statistical signal processing. Subsequently, we describe how to estimate such lag- and frequency-domain quantities from finite data samples, and highlight estimation accuracy issues. Finally, we focus on HOS of noisy linear non-Gaussian processes modeled as the output

$$x(n) = \sum_l h(l)w(n-1) + v(n) \tag{10}$$

of an LTI system. We assume that the impulse response $h(n)$ is absolutely summable, $\Sigma_n|h(n)| < \infty$; the input $w(n)$ is zero-mean independent identically distributed (i.i.d.) non-Gaussian and has finite moments $E\{w^k(n)\} < \infty$; and the additive noise $v(n)$ is zero-mean, stationary, and independent of $w(n)$.

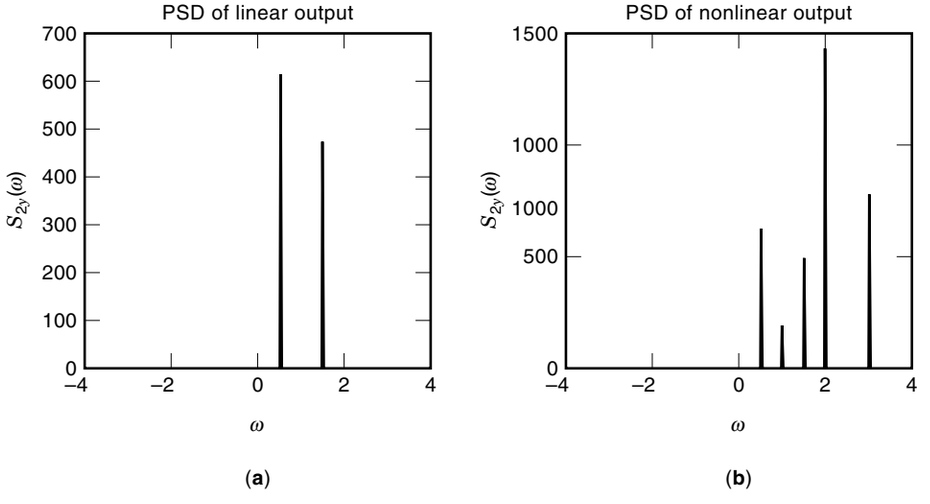

Figure 3. (a) When two harmonics pass through a linear system, the same frequencies emerge in the output. (b) When the system is nonlinear, additional frequency components are created.

Definitions and Properties

The kth-order moment of a real-valued discrete-time stationary random process $x(n)$ at the set of lags $\boldsymbol{\tau} := (\tau_1, \ldots, \tau_{k-1})$ is defined as

$$m_{kx}(\boldsymbol{\tau} := E\{x(n)x(n+\tau_1)\cdots x(n+\tau_{k-1})\} \tag{11}$$

The mean $m_{1x} := E\{x(n)\}$ and the ACS $m_{2x}(\tau)$ in Eq. (8) are obtained as special cases of Eq. (11) with $k = 1, 2$. The kth-order cumulant $c_{kx}(\boldsymbol{\tau})$ is expressed in terms of $m_{kx}(\boldsymbol{\tau})$ and combinations of lower-order moments m_{lx} with $l < k$ (see Ref. 13, p. 19). For $k = 1$, we have $c_{1x} = m_{1x}$, and the second-order cumulant is by definition the autocovariance,

$$c_{2x}(\tau) := m_{2x}(\tau) - m_{1x}^2$$

For zero-mean processes the third- and fourth-order cumulants are defined as

$$c_{3x}(\tau_1, \tau_2) = m_{3x}(\tau_1, \tau_2) \tag{12}$$

$$\begin{aligned} c_{4x}(\tau_1, \tau_2, \tau_3) = {} & m_{4x}(\tau_1, \tau_2, \tau_3) - m_{2x}(\tau_1)m_{2x}(\tau_2 - \tau_3) \\ & - m_{2x}(\tau_2)m_{2x}(\tau_3 - \tau_1) - m_{2x}(\tau_3)m_{2x}(\tau_1 - \tau_2) \end{aligned} \tag{13}$$

In general, the kth-order cumulant

$$c_{kx}(\boldsymbol{\tau}) := \mathrm{cum}_k\{x(n), x(n+\tau_1), \ldots, x(n+\tau_{k-1})\}$$

is defined as the coefficient of the $(-j)^k\, u_0 \ldots u_{k-1}$ term in the Taylor series expansion of the log characteristic function, $\ln [E\{\exp(j[u_0x(n) + u_1x(n+\tau_1) + \ldots + u_{k-1}x(n+\tau_{k-1})])\}]$. Because moments can also be defined via Taylor series expansion of the characteristic function, general relations expressing moments of any order k in terms of cumulants of orders $\leq k$ (and vice versa) are available and can be found in Refs. 13 and 14. The log characteristic function of Gaussian processes is a quadratic polynomial, which implies that a Gaussian distribution is completely characterized by its first- and second-order statistics. This leads to the first basic property of cumulants:

(P1). If $x_l(n)$, $l = 1, 2, \ldots$, are Gaussian, then

$$\mathrm{cum}_k[x_{a_0}(n), x_{a_1}(n+\tau_1), \ldots, x_{a_{k-1}}(n+\tau_{k-1})] \equiv 0, \qquad k \geq 3$$

for all $\boldsymbol{\tau}$ and $\{a_l\}_{l=0}^{k-1} \in \{1, 2, \ldots\}$.

This property is important even for a single random signal, because undesired noise or interference $v(n)$ is often modeled as Gaussian, and its cumulant $c_{kv} \equiv 0$ for $k \geq 3$. Hence, with algorithms based on (ensemble) high-order cumulants, one works in a high-SNR domain when additive Gaussian noise (AGN) $v(n)$ with unknown color is superimposed to the signal of interest (SOI) $s(n)$ to yield $x(n) = s(n) + v(n)$. Non-Gaussianity of $x(n)$ implies that $c_{kx}(\boldsymbol{\tau}) \neq 0$ for some $\boldsymbol{\tau}$ and $k \geq 3$.

Note that (P1) and the AGN suppression property hold for cumulants but not for moments. Separation of a non-Gaussian SOI from the AGN relies also on the second interesting property of cumulants:

(P2). If $x(n) = s(n) + v(n)$ with $s(n)$ independent of $v(n)$, then

$$c_{kx}(\boldsymbol{\tau}) = c_{ks}(\boldsymbol{\tau}) + c_{kv}(\boldsymbol{\tau})$$

Suppose that we wish to estimate the parameter vector $\boldsymbol{\theta}$ from data $x(n) = s(n; \boldsymbol{\theta}) + v(n)$, where the SOI $s(n; \boldsymbol{\theta})$ is non-Gaussian and the additive noise $v(n)$ is Gaussian. If $s(n)$ is independent of $v(n)$, then thanks to (P2), $c_{kx}(\boldsymbol{\tau}) = c_{ks}(\boldsymbol{\tau}) + c_{kv}(\boldsymbol{\tau})$, and thanks to (P1), $c_{kv}(\boldsymbol{\tau}) \equiv 0\ \forall k \geq 3$. Thus, $c_{kx}(\boldsymbol{\tau}) = c_{ks}(\boldsymbol{\tau})$, and recovery of $\boldsymbol{\theta}$ based on $c_{kx}(\boldsymbol{\tau})$ gains tolerance to AGN. Interestingly, $m_{kx}(\boldsymbol{\tau}) \neq m_{ks}(\boldsymbol{\tau}) + m_{kv}(\boldsymbol{\tau})$; that is, additivity does not hold for moments in general. Therefore, cumulants (but not moments) separate signals on the basis of their distributions and their symmetries. For example, if $v(n)$ is symmetrically distributed (hence nonskewed) while $s(n)$ is skewed, then $c_{kx}(\boldsymbol{\tau}) = c_{ks}(\boldsymbol{\tau})$ for $k = 3$.

What order of cumulant one should use depends on the application, but considerations of computational load and estimation accuracy, to be discussed later, suggest $k = 3$ or 4 for most practical scenarios. Clearly, one should also prefer the cumulant order for which the non-Gaussianity of the signal appears to be stronger. For example, for the symmetrically distributed seismic reflectivity histogram of Example 1, one should favor the use of fourth- over third-order cumulants.

Cumulants also measure statistical independence, which explains why absolute cumulant summability conditions appear in the laws of large numbers and the central limit theorem for correlated data (see Ref. 13). This feature stems from the third cumulant property:

(P3). If $\{x_1(n), \ldots, x_k(n)\}$ are independent of $\{y_1(n), \ldots, y_l(n)\}$, then

$$\mathrm{cum}_{k+l}\{x_1(n), \ldots, x_k(n); y_1(n), \ldots, y_l(n)\} \equiv 0$$

If $x(n)$ is i.i.d., then $m_{4x}(\tau_1, \tau_2, \tau_3) \neq \delta(\tau_1, \tau_2, \tau_3)$, but thanks to (P3) we have

$$c_{4x}(\tau_1, \tau_2, \tau_3) = \gamma_{4x}\delta(\tau_1, \tau_2, \tau_3), \qquad \gamma_{4x} := c_{4x}(0, 0, 0)$$

If $c_{kx}(\boldsymbol{\tau})$ is absolutely summable, that is,

$$\sum_{\boldsymbol{\tau}} |c_{kx}(\boldsymbol{\tau})| < \infty \qquad \text{for} \quad k = 1, 2, \ldots \tag{14}$$

then cumulants at large lags must decay to zero; that is, $\lim_{\tau \to \infty} c_{kx}(\boldsymbol{\tau}) = 0$, which amounts to approximate independence of the $x(n)$ samples that are well separated in time (see Ref. 13). Except for harmonic signals, which give rise to line spectra and infinite memory, most random signals in engineering applications have asymptotically vanishing memory and satisfy the so-called mixing condition expressed by the absolute cumulant summability stated in Eq. (14).

In addition to quantifying mixing conditions, Eq. (14) guarantees that kth-order moment and cumulant Fourier spectra (or polyspectra) exist. They are defined as the $(k-1)$-dimensional FT of $m_{kx}(\boldsymbol{\tau})$ and $c_{kx}(\boldsymbol{\tau})$ respectively:

$$M_{kx}(\boldsymbol{\omega}) := \sum_{\boldsymbol{\tau}=-\infty}^{\infty} m_{kx}(\boldsymbol{\tau})e^{-j\boldsymbol{\omega}\boldsymbol{\tau}'} \tag{15}$$

$$S_{kx}(\boldsymbol{\omega}) := \sum_{\boldsymbol{\tau}=-\infty}^{\infty} c_{kx}(\boldsymbol{\tau})e^{-j\boldsymbol{\omega}\boldsymbol{\tau}'} \tag{16}$$

In the above equations, $\boldsymbol{\omega} := (\omega_1 \ldots \omega_{k-1})$, prime stands for transpose, and $\sum_{\tau}$ is a shorthand notation for $\sum_{\tau_1} \cdots \sum_{\tau_k - 1}$.

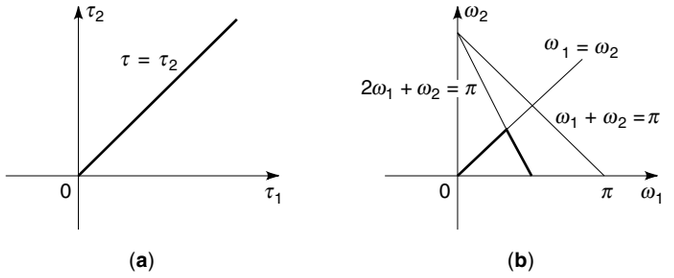

Figure 4. (a) The nonredundant region for the third-order cumulant is given by the wedge $\tau_1 \geq \tau_2 \geq 0$. (b) The nonredundant region for the bispectrum is given by the triangular region with ending points (0, 0), (π/2, 0), and (π/3, π/3). They are obtained by intersecting the lines defined by $\omega_2 = 0$, $\omega_1 = \omega_2$, and $2\omega_1 + \omega_2 = \pi$.

An alternative definition to Eq. (15) is

$$M_{kx}(\boldsymbol{\omega}) := \lim_{N\to\infty} \frac{1}{N} E\left\{ X_N(\omega_1)\cdots X_N(\omega_{k-1}) X_N^*\left(\sum_{i=1}^{k-1}\omega_i\right)\right\} \quad (17)$$

where * denotes complex conjugation, and $X_N(\omega)$ is the FT of the length-N data,

$$X_N(\omega) := \sum_{n=0}^{N-1} x(n) e^{-j\omega n} \quad (18)$$

Cumulant spectra for $k = 3$ and 4 are called bispectra and trispectra, respectively, and generalize the ordinary PSD that is obtained when $k = 2$. It turns out that moment and cumulant spectra coincide except for the proper submanifold $\omega_1 + \cdots + \omega_{k-1} = 0 \pmod{2\pi}$ (Ref. 15)—a feature that will be useful when computing polyspectral estimates using the fast fourier transform (FFT).

Cumulants and polyspectra need not be known in their entire region of support, because symmetries (implied by stationarity) are present, and can be utilized if one knows their values over the so-called nonredundant support region. Symmetry in $c_{2x}(\tau) = c_{2x}(-\tau)$ implies that the nonredundant region of $c_{2x}(\tau)$ is $\tau \geq 0$. Similarly, for real processes it suffices to know $c_{kx}(\boldsymbol{\tau})$ over the nonredundant region $0 \leq \tau_{k-1} \leq \cdots \leq \tau_1$. In contrast to $c_{2x}(\tau)$, we generally have $c_{kx}(\tau) \neq c_{kx}(-\tau)$ for $k \geq 3$, and as a result, polyspectra are complex-valued. The shaded region in Fig. 4(a) depicts the $\tau_1 \geq \tau_2 \geq 0$ nonredundant region for the third-order cumulant $c_{3x}(\tau_1, \tau_2)$. As for the bispectrum $S_{3x}(\omega_1, \omega)$, we find the nonredundant region as the common intersection of the areas defined by $0 \leq \omega_2 \leq \omega_1 \leq \pi$, $\omega_1 + \omega_2 \leq \pi$, and $2\omega_1 + \omega_2 \leq \pi$. The result is shown as the shaded triangle in Fig. 4(b).

For complex processes, many definitions of cumulants and polyspectra arise depending on the number of (un)conjugated lagged signal copies, and additional symmetries may appear in the corresponding support regions. HOS of complex processes have been studied in the context of retrieving constant-amplitude complex harmonics in Ref. 16, and when dealing with (cyclo)stationary processes with mixed spectra in Refs. 17 and 18.

Sample Estimation

We now turn our attention to deterministic signals $d(n)$ and sample estimates of moments and cumulants when a realization $\{x(n)\}_{n=0}^{N-1}$ of a stationary random process is available. We have already seen in the preceeding section the third-order moment $h_3(\tau_1, \tau_2)$ of the deterministic impulse response $h(n)$. Generalizations to orders $k > 3$ are straightforward to construct. Moments and moment spectra are more appropriate than cumulants for deterministic signals in view of the following property:

(P4). If $d(n)$ is deterministic, its cumulant $c_{kd}(\boldsymbol{\tau}) \equiv 0$ for $k > 1$. Thus, if $d(n)$ is observed in the presence of stationary noise $v(n)$ [i.e., $x(n) = d(n) + v(n)$], then $c_{kx}(\boldsymbol{\tau}) = c_{kv}(\boldsymbol{\tau})$ for $k > 1$ and information about $d(n)$ is lost in $c_{kx}(\boldsymbol{\tau})$.

In view of (P4), cumulant-based deterministic signal reconstruction in stationary noise is impossible from a single data record. In the noise-free case, however, it is possible to reconstruct a finite-support $\{d(n)\}_{n=0}^{Q}$ uniquely (up to a scale) from its kth-order moment, for any $k \geq 3$. We illustrate this fact with the three signal copies $d(l)$, $d(l + Q)$, and $d(l + n)$ shown in Fig. 5. By having at least three signal copies (as in correlations of order greater than 2), one can utilize at least two copies in order to create a Kronecker delta as follows: fix the first copy, and shift the second copy to the left by Q positions so only $d(0)$ from the first sequence and $d(Q)$ from the second sequence overlap. The product is thus $d(l)d(l + Q) = d(0)d(Q)\delta(l)$, where $\delta(l)$ is the Kronecker delta function. Slide now the third copy over so the delta function selects, within a scale constant $d(0)d(Q)$, the signal values $d(n)$. Using Eq. (5) with $(\tau_1, \tau_2) = (Q, n)$ and $(\tau_1, \tau_2) = (Q, 0)$, we find

$$d_3(Q, n) = \sum_{l=0}^{Q} d(l)\, d(l+Q)\, d(l+n) = d(0)\, d(Q)\, d(n)$$
$$0 \leq n \leq Q$$

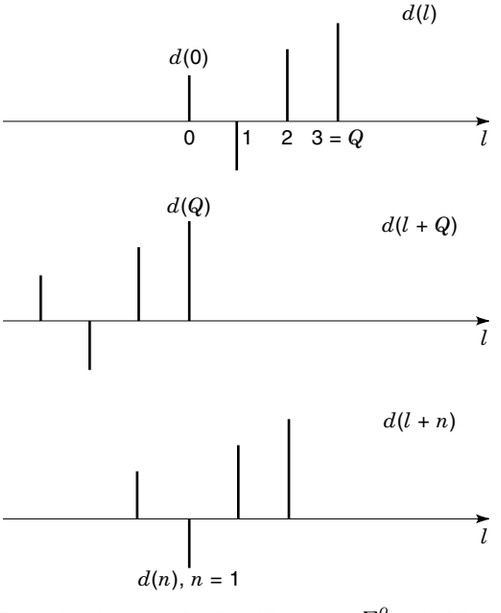

Figure 5. A simple example that illustrates $\sum_{l=0}^{Q} d(l)d(l + Q)d(l + n) = d(0)d(Q)d(n)$. Since $d(l)$ has length Q, it overlaps with $d(l + Q)$ by only one sample. The product of the two is a Kronecker delta function with amplitude $d(0)d(Q)$, and it has the ability to "single out" the $d(n)$ value as the third copy $d(l + n)$ slides through.

Hence, in closed form, we obtain (Ref. 19)

$$d(n) = d(0)\frac{d_3(Q, n)}{d_3(Q, 0)} \tag{19}$$

The one-to-one mapping here requires knowledge of Q and triple or higher-order correlations having at least two lags.

Given N samples of a stationary random signal $x(n)$, kth-order moments are estimated from time averages as

$$\hat{m}_{kx}(\boldsymbol{\tau}) = \frac{1}{N}\sum_{n=0}^{N-1} x(n)x(n+\tau_1)\cdots x(n+\tau_{k-1}) \tag{20}$$

Similarly to ensemble cumulants, kth-order sample cumulants $\hat{c}_{kx}(\boldsymbol{\tau})$ are computed in terms of $\hat{m}_{lx}$ with $l \le k$ (Ref. 13, p. 19). For example, using Eq. (20) with $k = 2$ and 4, we first find $\hat{m}_{2x}$ and $\hat{m}_{4x}$. Subsequent substitution into Eq. (13) yields $\hat{c}_{4x}$.

Under the condition (14), sample cumulants are mean-square-sense consistent with computable variance (see, e.g., Ref. 17); that is, the mean squared error (m.s.e.) between the sample and ensemble moments and cumulants can be made arbitrarily small by increasing the data length. However, the price paid for going higher is twofold: additional computations due to increased dimensionality, and larger variance for a given data length. Interestingly, the variance (and hence accuracy) of kth-order sample moments and cumulants depends upon ensemble moments and cumulants of up to order $2k$. For example, it turns out that the variance of $\hat{c}_{2x}$ is given by

$$\mathrm{var}\{\hat{c}_{2x}(\tau)\} = \frac{1}{N^2}\sum_{n_1=0}^{N-1}\sum_{n_2=0}^{N-1}\mathrm{cum}_2\{x(n_1)x(n_1+\tau), x(n_2)x(n_2+\tau)\}$$

which clearly entails c_{2x} and c_{4x}.

In the frequency domain, estimates of moment and cumulant spectra are formed by Fourier-transforming windowed (smoothed and truncated) versions of $\hat{m}_{kx}(\boldsymbol{\tau})$ and $\hat{c}_{kx}(\boldsymbol{\tau})$:

$$\hat{M}_{kx}(\boldsymbol{\omega}) = \sum_{\boldsymbol{\tau}=-M}^{M} w_k(\boldsymbol{\tau})\hat{m}_{kx}(\boldsymbol{\tau})e^{-j\boldsymbol{\omega}\boldsymbol{\tau}'} \tag{21}$$

$$\hat{S}_{kx}(\boldsymbol{\omega}) = \sum_{\boldsymbol{\tau}=-M}^{M} w_k(\boldsymbol{\tau})\hat{c}_{kx}(\boldsymbol{\tau})e^{-j\boldsymbol{\omega}\boldsymbol{\tau}'} \tag{22}$$

where $w_k(\boldsymbol{\tau})$ is a lag-domain window sequence with nonzero support over $[-M, M]$, and M is chosen to approximate the "effective memory" of the process beyond which $c_{kx}(\boldsymbol{\tau})$ is practically negligible. Alternatively, we can also smooth the sample version of Eq. (17) obtained after dropping the limit and the expectation operator:

$$I_{kx}(\boldsymbol{\omega}) = \frac{1}{N}X(\omega_1)\cdots X(\omega_{k-1})X^*(\omega_1+\cdots+\omega_{k-1}) \tag{23}$$

$$\hat{\tilde{M}}_{kx}\left(\frac{2\pi}{N}\boldsymbol{i}\right) = \frac{1}{N}\sum_{\boldsymbol{l}=\boldsymbol{0}}^{N-1} W_k\left(\frac{2\pi}{N}(\boldsymbol{i}-\boldsymbol{l})\right)I_{kx}\left(\frac{2\pi}{N}\boldsymbol{l}\right) \tag{24}$$

where $W_k((2\pi/N)\boldsymbol{i})$, $\boldsymbol{i} := (i_1, \ldots, i_{k-1})$, is a $(k-1)$-dimensional frequency-domain window with bandwidth $\approx 1/M$. Its role is to smooth the polyperiodogram $I_{kx}(\boldsymbol{\omega})$ over the discrete-frequency grid $2\pi\boldsymbol{i}/N$. The latter approach is adopted because $I_{kx}(\boldsymbol{\omega})$ is often obtained using computationally efficient FFT algorithms.

Both types of polyspectral estimates in Eqs. (21)–(24) parallel classical spectral estimation approaches and correspond to what one would call smoothed polycorrelogram and polyperiodogram estimates. The lag- and frequency-domain windows are designed to obey the symmetries present in the ensemble cumulants and polyspectra. Their choice is critical because windowing controls the bias–variance tradeoff encountered with sample polyspectral estimates. Without it, the variance of $\hat{M}_{kx}(\boldsymbol{\omega})$, $\hat{S}_{kx}(\boldsymbol{\omega})$, and $\hat{\tilde{M}}_{kx}(\boldsymbol{\omega})$ would not decrease to zero as $N \to \infty$ (see Refs. 13, 14, and 20) for rigorous statements and guidelines on the choice of windows). Consistent polyspectral estimates are obtained only with appropriate windowing, but similarly to cumulants' estimates, their variance increases with the order k. For example, it is well known that using a window with bandwidth M^{-1} and energy $\mathscr{E}_w$, PSD estimates [$k = 2$ in Eqs. (21)–(24)] have $\mathrm{var}\{\hat{S}_{2x}(\omega)\} \approx \mathscr{E}_w MN^{-1}S_{2x}^2(\omega)$ (see Ref. 13), whereas sample bispectra [$k = 3$ in Eqs. (21)–(24)] have for $\omega_1 + \omega_2 \neq 0 \pmod{2\pi}$ (see Ref. 15)

$$\mathrm{var}\{\hat{S}_{3x}(\omega_1, \omega_2)\} \approx \mathscr{E}_w\frac{M^2}{N}S_{2x}(\omega_1)S_{2x}(\omega_2)S_{2x}(\omega_1+\omega_2) \tag{25}$$

This implies that due to the factor M^2 (as opposed to M in the PSD estimate), an order of magnitude more data are required for bispectral estimates to achieve comparable accuracy.

No windowing is required if the process has line spectra. However, when the mixing condition [Eq. (14)] is satisfied, unsmoothed (or raw) polyperiodograms $I_{kx}(\boldsymbol{\omega})$ are inconsistent and exhibit erratic variations although they are asymptotically unbiased. In fact, $I_{kx}(\boldsymbol{\omega})$ values at distinct frequencies off the proper submanifold $\omega_1 + \cdots + \omega_{k-1} = 0 \pmod{2\pi}$ are asymptotically independent and Gaussian-distributed (Ref. 15)—a property assuring that polyspectra offer asymptotically efficient feature vectors for estimation and classification.

The following question arises with polyperiodogram extensions: Do we lose "lower-order information" by going higher? Consider Eq. (23) with $k = 3$ and $\omega_1 = \omega$, $\omega_2 = 0$ to verify that

$$I_{3x}(\omega, 0) \propto I_{2x}(\omega) = N^{-1}|X(\omega)|^2 \tag{26}$$

Therefore, higher- (than second-) order statistics generalize the notion of second-order statistics and their additional properties make them useful in signal analysis applications. Thanks to the next (so-called multilinearity) property, HOS play a major role in LTI system analysis as well.

(P5). If $\{x_1, \ldots, x_k\}$ and y are random variables and $\rho_1, \ldots, \rho_k$ are deterministic constants, then homogeneity and superposition hold with cumulants:

$$\begin{aligned}\mathrm{cum}_k\{\rho_1 x_1, \ldots, \rho_k x_k\} &= \rho_1\cdots\rho_k\,\mathrm{cum}_k\{x_1, \ldots, x_k\}\\ \mathrm{cum}_k\{x_1 + y, x_2, \ldots, x_k\} &= \mathrm{cum}_k\{x_1, x_2, \ldots, x_k\}\\ &\quad + \mathrm{cum}_k\{y, x_2, \ldots, x_k\}\end{aligned}$$

As we shall see next, (P5) will allow us to express the kth-order cumulant $c_{kx}(\boldsymbol{\tau})$ of the output $x(n)$ in Eq. (10) in terms of the kth-order moment $h_k(\boldsymbol{\tau})$ and the input kth-order cumulant $c_{ku}(\boldsymbol{\tau})$.

Linear Non-Gaussian Processes

If we view the impulse response coefficients in Eq. (10) as the constants in (P5), we arrive at

$$c_{kx}(\boldsymbol{\tau}) = \gamma_{kw} \sum_i h(l)h(l+\tau_1)\cdots h(l+\tau_{k-1}), \qquad \gamma_{kw} := c_{kw}(\mathbf{0}) \tag{27}$$

where we also assume that $v(n)$ has vanishing kth-order cumulant, which is true, for example, when $v(n)$ is AGN and $k \geq 3$ is used. Note that Eq. (27) generalizes the well-known output correlation expression $c_{2x}(\tau) = \sigma_w^2 \sum_l h(l)h(l+\tau)$, if we recall that $\sigma_w^2 := c_{2w}(0)$. The corresponding "amplitude-only" spectrum $S_{2x}(\boldsymbol{\omega}) = \sigma_w^2 |H(\boldsymbol{\omega})|^2$ becomes a special case of the kth-order output polyspectrum

$$S_{kx}(\boldsymbol{\omega}) = \gamma_{kw} H(\omega_1)\cdots H(\omega_{k-1})H^*(\omega_1+\cdots+\omega_{k-1}) \tag{28}$$

If $\phi_{kx}(\boldsymbol{\omega})$ denotes the phase of $S_{kx}(\boldsymbol{\omega})$ and $\phi_h(\omega)$ that of the transfer function, then generally,

$$\phi_{kx}(\boldsymbol{\omega}) = \phi_h(\omega_1)+\cdots+\phi_h(\omega_{k-1}) - \phi_h(\omega_1+\cdots+\omega_{k-1}) \tag{29}$$

This shows that for some $k \geq 3$, $\phi_{kx}(\boldsymbol{\omega}) \neq 0$, and thus—contrary to the phase-blind SOS, which has $\phi_{2x}(\omega) = \phi_h(\omega) - \phi_h(\omega) \equiv 0$—HOS convey phase information that is necessary for the recovery of $\angle H(\boldsymbol{\omega})$.

Since no lower-order information is lost by going higher, the combination of amplitude and phase information allows for one-to-one correspondence between $S_{kx}(\boldsymbol{\omega})$ [or $c_{kx}(\boldsymbol{\tau})$] and $H(\omega)$ [or $h(n)$], which we have established already for high-order correlations of deterministic signals in Eq. (19). We will elaborate further in the section "Recovering Phase Information," in the context of blind identification of pole–zero LTI systems, or equivalently, autoregressive moving-average (ARMA) modeling of non-Gaussian linear processes. We wrap up this section with an example illustrating the tradeoffs involved in SOS versus HOS estimation of a non-Gaussian linear moving-average (MA) process observed in AGN.

Example 4. *Estimation of cumulant statistics of a non-Gaussian information-bearing signal $s(n)$ in the presence of additive colored Gaussian noise $v(n)$.* We simulated i.i.d. exponentially distributed deviates $w(n)$ and passed them through a finite impulse response (FIR) system with impulse response $\boldsymbol{h} = [1, -1.3487, 1.0766, -0.6089, -0.0631, -0.0517]$. We superimposed on the resulting non-Gaussian MA(5) output process $s(n)$ the colored AGN $v(n)$, and obtained N noisy samples $x(n)$. Figure 6(a) depicts the mean (dashed line) ± standard deviation (std) bounds (dotted lines) of the sample correlation $\hat{c}_{2x}(\tau)$ estimated as in Eq. (20) with $k = 2$, $N = 256$, and by averaging 100 independent realizations. Compared to the ensemble SOI correlation $c_{2s}(\tau)$ (solid line), the sample estimate exhibits noticeable bias due to the presence of the noise correlation $c_{2v}(\tau)$. On the other hand, thanks to the ability of HOS to suppress AGN, the bias between the third-order sample cumulant $\hat{c}_{3x}(\tau_1, \tau_2)$ and the ensemble SOI cumulant $c_{3s}(\tau_1, \tau_2)$ is negligible, although the standard deviation bounds (dotted lines) are noticeably wider than those in Fig. 6(a). Hence, HOS improve AGN-induced bias, and from this viewpoint they provide a high-SNR domain in the presence of AGN, but they exhibit more variability, as we argued in the previous subsection. Nevertheless, they remain consistent, and as Fig. 6(c,d) show, the SOS bias remains but the HOS variance decreases as the data length is increased to $N = 4096$.

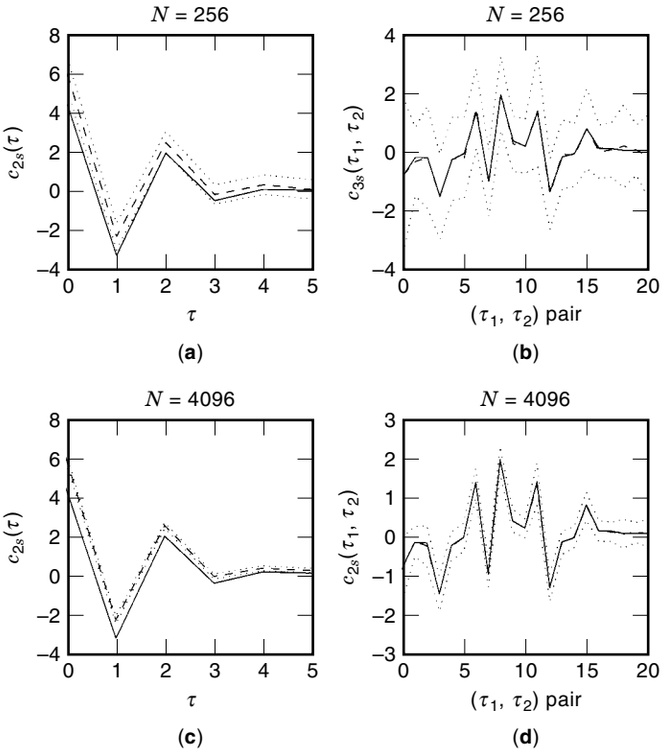

Figure 6. The received data model is $x(n) = s(n) + v(n)$, where $s(n)$ is the non-Gaussian signal with asymmetric distribution and $v(n)$ is the additive Gaussian noise. The true covariance function of $s(n)$ is plotted as solid lines in (a) and (c), and the true third-order cumulant function is plotted as solid lines in (b) and (d). The sample covariance [(a) and (c)] and third-order cumulant functions [(b) and (d)] of $x(n)$ are calculated and averaged over 100 independent realizations to generate the mean (dashed) and mean ± std (dotted) curves. The data length used was 256 for (a) and (b) and 4096 for (c) and (d). We observe that (i) the variance in the sample estimates decreases as the data length increases; (ii) the sample third-order cumulant $\hat{c}_{3x}(\tau_1, \tau_2)$ is unbiased for $c_{3s}(\tau_1, \tau_2)$, whereas the sample covariance $\hat{c}_{2x}(\tau)$ is biased for $c_{2s}(\tau)$. The tradeoff is that $\hat{c}_{3x}(\tau_1, \tau_2)$ has higher variance than $\hat{c}_{2x}(\tau)$.

Of course, caution must be exercised, because a significant increase in N may render the stationarity assumption invalid in real-life applications. Keeping the HOS order to no higher than four is a safe guideline. With sufficient data, the use of HOS in signal processing applications may be well justified, some of which will be elaborated further in the following sections.

DISCERNING NON-GAUSSIANITY

In view of properties (P1) and (P2), HOS quantify distance from Gaussianity and have thus been used for estimating, detecting, and classifying non-Gaussian signals in Gaussian noise, or vice versa (see Refs. 21 and 22).

Consider first a data model $x(n) = s(n) + v(n)$, where the SOI $s(n)$ is zero-mean Gaussian and the additive noise $v(n)$ is

zero-mean non-Gaussian with unknown color and is independent of $s(n)$. Such models appear when sonar and radar information-bearing signals are observed in impulsive noise or clutter. Our goal is to detect, model, or classify $s(n)$ according to its statistics. Since $s(n)$ is Gaussian, its only nonvanishing statistic is $c_{2s}(\tau) = c_{2x}(\tau) - c_{2v}(\tau)$ which is biased by the unknown noise color. Interestingly, we will show in the sequel that $c_{2v}(\tau)$ can be estimated free from the influence of $s(n)$ (at least theoretically) by going into the HOS domain. We shall model $v(n)$ as a linear process with input $w(n)$. According to Eq. (27), we find

$$c_{(k+1)v}(\boldsymbol{\tau}, \tau_k) = \gamma_{(k+1)w} \sum_n h(n)h(n+\tau_1)\ldots h(n+\tau_{k-1})h(n+\tau_k)$$

By summing $c_{(k+1)v}(\boldsymbol{\tau}, \tau_k)$ over τ_k and again recalling Eq. (27), we obtain

$$\sum_{\tau_k} c_{(k+1)v}(\boldsymbol{\tau}, \tau_k) = \frac{\gamma_{(k+1)w}}{\gamma_{kw}} H(0)\, c_{kv}(\boldsymbol{\tau})$$

Therefore, we arrive at the following order-recursive relation between $c_{(k+1)v}$ and c_{kv} by projecting $c_{(k+1)v}(\boldsymbol{\tau}, \tau_k)$ over the τ_k axis (Ref. 23):

$$c_{kv}(\boldsymbol{\tau}) = \frac{\gamma_{kw}}{H(0)\gamma_{(k+1)w}} \sum_{\tau_k=-Q}^{Q} c_{(k+1)v}(\boldsymbol{\tau}, \tau_k) \tag{30}$$

where Q denotes the effective memory of the process. In deriving Eq. (30), we have assumed that both $\gamma_{(k+1)w}$ and $H(0)$ are nonzero. We point out that modified versions of Eq. (30) are available by using $H(\omega) \neq 0$ for some ω (see Ref. 24).

Recall the frequency-domain relationship [Eq. (26)], which explains the notion that lower-order information is not lost by going higher. Interestingly, Eq. (30) makes the same statement in the time domain, and its practical use is exemplified next. Thanks to (P1) and (P2), we have that $c_{kx}(\boldsymbol{\tau}) = c_{kv}(\boldsymbol{\tau})$ for $k \geq 3$. Therefore, unbiased higher-order cumulant estimates of $v(n)$ can be obtained directly from the data, and their successive projections as in Eq. (30) give rise to an unbiased estimate for $c_{2v}(\tau)$. With $c_{2v}(\tau)$ available, the SOI correlation can then be computed as $c_{2s}(\tau) = c_{2x}(\tau) - c_{2v}(\tau)$. This is an interesting example that illustrates how one can benefit from a "detour" to the HOS domain in order to find lower-order information. Related ideas have been adopted for identification of LTI systems using input–output data, both of which are observed in colored AGN. Such errors-in-variables models find applications in time-delay estimation of sonar and radar signals (see Ref. 24).

Conversely, if $v(n)$ is the SOI, then the above unbiased estimate of $c_{2v}(\tau)$ constructed from $c_{3v}(\boldsymbol{\tau})$ or $c_{4v}(\boldsymbol{\tau})$ shows how one finds the SOS of interest without suffering from the AGN (Ref. 23). In practice, however, the presence of $s(n)$ is evidenced in the variance of $\hat{c}_{2v}(\tau)$, which nevertheless diminishes as the data length grows larger. HOS are also useful for detection and estimation problems entailing linear non-Gaussian (especially heavy-tailed distributed) noise. Results in this direction have been reported in Ref. 56.

When the probabilistic distribution of the data $x(n)$ is unknown, HOS are attractive features for pattern recognition and classification purposes, because they contain condensed but complete statistical information (Ref. 25). Especially when dealing with the classification of deterministic signals or non-Gaussian linear processes, the one-to-one correspondence between the IR $h(n)$ and its high-order correlations $h_k(\boldsymbol{\tau})$ mentioned earlier assures uniqueness even when spectrally equivalent signals are to be classified. Furthermore, the asymptotic (as $N \to \infty$) normality of the sample HOS estimates holds even for non-Gaussian data, and leads to asymptotic optimality of HOS-based classifiers. Specifically, suppose that we collect L lags of $\hat{c}_{kx}$, or $\hat{m}_{kx}$, or $\hat{S}_{kx}$, in a vector $\hat{\boldsymbol{g}}_N$. If $x(n)$ satisfies Eq. (14), then $\hat{\boldsymbol{g}}_N$ is approximately normal (for large N) with asymptotic covariance matrix $\boldsymbol{\Sigma}_g$. If class i is characterized by the vector $\boldsymbol{\theta}_i$, $i \in [1, I]$, with template HOS feature vector $\boldsymbol{g}(\boldsymbol{\theta}_i)$, then the following rule offers an asymptotically maximum-likelihood classifier in the HOS domain (22):

$$\hat{\boldsymbol{\theta}}_0 = \boldsymbol{\theta}_l \quad \text{iff} \quad J_N(\boldsymbol{\theta}_0, \boldsymbol{\theta}_l) \leq J_N(\boldsymbol{\theta}_0, \boldsymbol{\theta}_i) \tag{31}$$

$$J_N(\boldsymbol{\theta}_0, \boldsymbol{\theta}_i) := [\hat{\boldsymbol{g}}_N(\boldsymbol{\theta}_0) - \boldsymbol{g}_N(\boldsymbol{\theta}_i)]\boldsymbol{\Sigma}_g^{-1}(\boldsymbol{\theta}_i)[\hat{\boldsymbol{g}}_N(\boldsymbol{\theta}_0) - \boldsymbol{g}_N(\boldsymbol{\theta}_i)] \quad \forall i \neq l \tag{32}$$

If $\boldsymbol{\Sigma}_g = \mathbf{I}$, which is asymptotically the case when FFT-based frequency domain sample estimates are used, the classifier in Eq. (31) assigns labels on the basis of minimum HOS-error energies. In other words, one forms the HOS templates for each class during training and then uses the minimum-distance criterion to classify a candidate HOS vector as belonging to the "closest" HOS template.

With linear non-Gaussian ARMA processes, SOS features do not guarantee uniqueness in classification, since two distinct time series can have identical autocorrelations (see Example 2). As we shall see in Section 4, finite HOS lags guarantee identifiability of ARMA models and are thus capable of classifying even spectrally equivalent non-Gaussian signals observed in low SNR Gaussian noise with perhaps unknown color. In Ref. 25, HOS features have been employed for object recognition and classification of textured images (see also Ref. 26). Other possible applications include speech and radar signal detection and classification. HOS-based detectors of linear non-Gaussian processes $s(n)$ in AGN $v(n)$ convert the binary hypothesis-testing problem

$$\mathrm{H}_0 : x(n) = v(n) \quad \text{versus} \quad \mathrm{H}_1 : x(n) = s(n) + v(n)$$

into a HOS-based one that involves nonvanishing SOI cumulants or polyspectra. For example, with skewed SOI and $e_{3x}(\boldsymbol{\tau}) := \hat{c}_{3x}(\boldsymbol{\tau}) - c_{3x}(\boldsymbol{\tau})$ denoting the error between sample and ensemble third-order cumulants, the binary hypotheses become

$$\mathrm{H}_0 : \hat{c}_{3x}(\boldsymbol{\tau}) = e_{3x}(\boldsymbol{\tau}) \quad \text{versus} \quad \mathrm{H}_1 : \hat{c}_{3x}(\boldsymbol{\tau}) = c_{3s}(\boldsymbol{\tau}) + e_{3x}(\boldsymbol{\tau})$$

and the distribution-independent HOS-based detection statistic is $\sum_{\boldsymbol{\tau}=-Q}^{Q} |e_{3x}(\boldsymbol{\tau})|^2$, where $\boldsymbol{\tau} := (\tau_1, \tau_2)$, and Q denotes the effective memory of $s(n)$. Omitting irrelevant terms, the HOS-based detection test statistic becomes

$$D_k = \sum_{\boldsymbol{\tau}=-Q}^{Q} \hat{c}_{3x}(\boldsymbol{\tau}) c_{3s}(\boldsymbol{\tau}) - \frac{1}{2} \sum_{\boldsymbol{\tau}=-Q}^{Q} |c_{3s}(\boldsymbol{\tau})|^2 \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\lessgtr}} \mathscr{T} \tag{33}$$

The threshold $\mathscr{T}$ depends on the variance of the sample cumulant, and the false-alarm rate in a Neyman–Pearson test (see

Ref. 21 for details). Interestingly, the first sum in Eq. (33) can be implemented with a filter matched to $h(n)$, since the output $y(n)$ can be expressed as the convolution between $x(n)$ and $h(Q - n)$ and has third-order sample cumulant at lag $\boldsymbol{\tau} = \mathbf{0}$ given by (Ref. 22)

$$\hat{c}_{3y}(\mathbf{0}) = \gamma_{3w} \sum_{\boldsymbol{\tau}=-Q}^{Q} \hat{c}_{3x}(\boldsymbol{\tau}) c_{3s}(\boldsymbol{\tau}) = \frac{1}{N+Q} \sum_{n=0}^{N+Q-1} y^3(n)$$

Therefore, if $s(n)$ is modeled well as a linear process with known $h(n)$ but unknown non-Gaussian driving noise distribution, then HOS-based detection of $s(n)$ can be accomplished by matched filtering of the data, followed by a kth-order nonlinear transformation (that depends on the HOS-order used), and appropriate thresholding as indicated in Eq. (33).

If linear non-Gaussian SOI modeling with known templates $h(n)$ is impossible, binary HOS-based detectors reduce to (non-)Gaussianity tests (see Refs. 27–29). For third-order cumulants the latter address the binary hypotheses testing problem

$$\mathrm{H}_0 : \hat{c}_{3x}(\boldsymbol{\tau}) = 0 \quad \text{versus} \quad \mathrm{H}_1 : \hat{c}_{3x}(\boldsymbol{\tau}) \neq 0$$

and declare non-Gaussianity when the null is rejected.

For HOS with multiple lags, we have under H_0 that $D_g := \hat{\boldsymbol{g}}_N' \, \Sigma_g^{-1} \hat{\boldsymbol{g}}_N \approx 0$, while under H_1, $D_g \neq 0$ (29). In other words, nonzeroness of the detection statistic implies non-Gaussianity [cf. (P1)]. If the entries of $\hat{\boldsymbol{g}}_N$ contain smoothed polyperiodograms, then we test $D_G = \Sigma_{\boldsymbol{\omega}} |\hat{\beta}_{kx}(\boldsymbol{\omega})|^2$, where $\hat{\beta}_{kx}(\boldsymbol{\omega})$ denotes the sample kth-order coherence (Ref. 27):

$$\hat{\beta}_{kx}(\omega_1, \ldots, \omega_{k-1}) = \frac{\hat{S}_{kx}(\omega_1, \ldots, \omega_{k-1})}{\sqrt{\hat{S}_{2x}(\omega_1) \ldots \hat{S}_{2x}(\omega_{k-1}) \hat{S}_{2x}(\omega_1 + \cdots + \omega_{k-1})}} \tag{34}$$

With $k = 3$, for example, the denominator in Eq. (34) is the standard deviation of the sample bispectrum of the numerator for frequencies off the proper submanifold [see also Eq. (25)]. Such a standardization procedure facilitates statistical testing in the sample bispectral (as opposed to the sample cumulant) domain, because the matrix $\boldsymbol{\Sigma}_g$ becomes diagonal when frequency-domain sample statistics are used.

With $k = 3$, the imaginary part of the bicoherence can also be used to test for symmetry in the distribution of a stationary stochastic process. Indeed, asymmetric cumulants in the lag domain give rise to polyspectra with nonzero imaginary parts. This property of HOS suggests a cumulant-based definition of time-reversibility (Ref. 15). The use of HOS for testing symmetry in the probabilistic structure of time series is well motivated, since temporal symmetry cannot be judged from the ACS, which is a symmetric quantity itself. Note that time-irreversible processes are non-Gaussian, but the converse is not necessarily true. Outputs of symmetric noncausal pole–zero LTI systems are time-reversible, and SOS are sufficient for their identification even if the driving input is non-Gaussian. On the other hand, evidence of time-irreversibility in the data favors HOS-based analysis. A specific chi-square time-reversibility test can be developed by testing for the statistical nonzeroness of $D_{\mathrm{TR}} := \hat{c}_{kx}(\boldsymbol{\tau}) - \hat{c}_{kx}(-\boldsymbol{\tau})$ (see Ref. 29).

Note that if the test statistics D_G and D_{TR} are declared to be zero with $k = 3$, then one needs to test them also for $k = 4, 5, \ldots$, in order to infer Gaussianity and time-reversibility. From this viewpoint, such statistical tests should be called non-Gaussianity and time-irreversibility tests instead of Gaussianity and time-reversibility tests, because they are conclusive only when nonzero. However, for most processes, tests based on third- and fourth-order HOS are sufficiently descriptive of the distribution shape.

In general, Gaussianity and time-reversibility are two fundamental properties of random signals that must be validated prior to applying statistical signal-processing algorithms that make such assumptions. Based on these tests, textured image models have been validated as non-Gaussian (Ref. 26), and real seismic data have been classified as non-Gaussian and nonskewed (Ref. 29).

RECOVERING PHASE INFORMATION

In this section, we focus on two related problems, namely (i) reconstructing deterministic signals $h(n)$ from their higher-order correlations or spectra, defined as in Eqs. (5) and (7), and (ii) recovering LTI models $h(n)$ from stationary and non-Gaussian time series $x(n)$ in Eq. (10) using lag- or frequency-domain HOS quantities computed via the sample estimates in Eq. (20) or in Eqs. (21)–(24). Because the magnitude response can be retrieved (at least in the noise-free case) from PSD estimates, our emphasis will be on recovering the phase $\angle H(\boldsymbol{\omega})$. Depending on whether reconstruction targets $H(\boldsymbol{\omega})$ at finite frequencies (e.g., the FFT grid $\{2\pi i/N\}_{i=0}^{N-1}$) or the parameters that describe the pole–zero model of $H(\boldsymbol{\omega})$, the resulting algorithms are classified as nonparametric (see the following subsection) or parametric (see the subsection after). Both approaches have found diverse applications, which include reconstruction of astronomical objects from photon-limited frames and imaging through turbulence, as well as solving inverse problems such as those encountered in blind deconvolution of seismic traces and self-recovering equalization of frequency-selective communication channels (see the last subsection of this section).

Nonparametric Approaches

Nonparametric algorithms rely on frequency-domain HOS, and their starting point is Eq. (29) for phase recovery and Eq. (28) for AGN-resistant magnitude reconstruction of signals and LTI systems. The last two differ only in the constant γ_{kw}, so we will concentrate on identifying LTI models $H(\boldsymbol{\omega})$ from the non-Gaussian $x(n)$. For notational simplicity, we will assume that $\gamma_{3w} \neq 0$ and use the log bispectrum, $\ln S_{3x}(\omega_1, \omega_2)$, whose amplitude and phase at frequencies $\{\omega_i = 2\pi_i/N\}_{i=0}^{N-1}$ are given by

$$\mu_{3x}(\omega_1, \omega_2) = \ln |\gamma_{3w}| + \mu_h(\omega_1) + \mu_h(\omega_2) + \mu_h(\omega_1 + \omega_2) \tag{35}$$

$$\phi_{3x}(\omega_1, \omega_2) = \phi_h(\omega_1) + \phi_h(\omega_2) - \phi_h(\omega_1 + \omega_2) \tag{36}$$

with

$$\mu_{3x}(\omega_1, \omega_2) := \ln |S_{3x}(\omega_1, \omega_2)|, \qquad \phi_{3x}(\omega_1, \omega_2) := \angle S_{3x}(\omega_1, \omega_2)$$
$$\mu_h(\omega) := \ln |H(\omega)|, \qquad \phi_h(\omega) := \angle H(\omega)$$

The nonredundant equations at bifrequencies $(\omega_1, \omega_2) \in \{2\pi i/N\}_{i=0}^{N-1}$ can be inferred from the nonredundant support re-

gions discussed in the "Basic Background" section. Because for a nonzero constant λ and a shift τ, pairs $\{h(n), w(n)\}$ and $\{\lambda h(n+\tau), w(n+\tau)/\lambda\}$ give rise to the same $x(n)$ in Eq. (10), it is well known that output-only (i.e., blind) reconstruction algorithms can recover the system up to a scale and/or shift ambiguity. To fix the scale, one can assume without loss of generality that $\gamma_{3w} = 1$ and omit the corresponding term in Eq. (35). Concatenating equations that result from Eqs. (35) and (36) at distinct bifrequencies, we arrive at two systems of linear equations with unknown vectors $\boldsymbol{\mu}_h := [\mu_h(\omega_0) \cdots \mu_h(\omega_{N-1})]'$, and $\boldsymbol{\phi}_h := [\phi_h(\omega_0) \cdots \phi_h(\omega_{N-1})]'$, respectively (Ref. 38):

$$\mathbf{A}_\mu \boldsymbol{\mu}_h = \boldsymbol{\mu}_{3x}, \qquad \mathbf{A}_\phi \boldsymbol{\phi}_h = \boldsymbol{\phi}_{3x} \tag{37}$$

where the matrices $\mathbf{A}_\mu$ and $\mathbf{A}_\phi$ can be made full-rank (with appropriate selection of equations) and are sparse with entries -1, 0, 1, 2, and 3.

Given the $\{x(n)\}_{n=0}^{N-1}$ samples, we first estimate the bispectrum as in Eqs. (21)–(24), extract its log-magnitude and phase, use them in the right-hand side (r.h.s.) of the matrix equations in Eq. (37), and then solve for the log-magnitude and phase responses of the system. Afterward, the complete frequency response $H(\omega)$ and the corresponding IR $h(n)$ can be obtained.

Note that we do not impose minimum- or maximum-phase assumptions, and hence the zeros of $H(\omega)$ are allowed to lie inside or outside the unit circle. If $H(\omega)$ has zeros on the unit circle, then $\ln |H(\omega)|$ is not well defined at certain frequencies. However, this can be circumvented, because the zeros are isolated and zero padding [prior to FFT in Eq. (24)] creates enough additional bifrequencies (and thus enough equations) to avoid the system nulls while assuring identifiability of $\ln |H(\omega)|$. One more issue complicating the phase recovery is the fact that Eq. (36) holds for the unwrapped phases while HOS phase estimates are wrapped and thus an additive (but unknown) factor $2\pi k$ is needed in the sample version of Eq. (36). Phase-unwrapping issues, performance analysis, and weighted least-squares solutions of the nonparametric identification approach can be found in Ref. 30. Additional nonparametric schemes include those in Ref. 3 and the polycepstral approaches of Nikias et al. (see Ref. 9). The latter are of interest because sample (poly)cepstra are known to be variance-stabilized versions of polyspectra whose estimation variance is (poly)frequency-independent (see also Ref. 13). But especially in the presence of AGN, they have problems when the underlying system has zeros on (or close to) the unit circle.

Bispectrum-based reconstruction of deterministic signals in AGN requires averaging over multiple records in order to estimate the bispectrum. Recall that due to (P5), the signal vanishes with single-record sample averages. Even with approximate knowledge of the signal support, say $[0, Q]$, one can apply a 2-D window to the noisy sample cumulant over the support $0 \le \tau_1 \le \tau_2 \le Q$ prior to estimating the sample bicorrelogram as in Eq. (22), and then apply the reconstruction algorithm of Eq. (37).

Parametric Approaches

Nonparametric HOS-based algorithms are advantageous when no parametric models are clearly evident. But computing the required frequency-domain HOS entails delicate selection of the smoothing windows in Eqs. (22)–(24) to control the bias–variance tradeoffs mentioned in the "Basic Background" section. On the other hand, parametric approaches impose finite parametrization on the LTI model. But if their degrees of freedom (numbers of parameters or poles and zeros in the rational transfer function) and parameters are estimated judiciously, parsimonious description of the data leads to more accurate model estimates when the required number of parameters is considerably smaller than the data length. HOS of $x(n)$ are capable of identifying uniquely (up to scale and shift ambiguities inherent in blind identification algorithms) all-pole AR(p) (autoregressive of order p) models, all-zero MA(q) (moving-average of order q) models, and general pole–zero ARMA(p, q) models without restricting the system to be minimum- or maximum-phase. Thus, contrary to SOS, HOS guarantee identifiability of (non-)minimum-phase and noncausal models (see Refs. 3–6 and 22).

If $x(n)$ is described by the ARMA(p, q) model

$$\sum_{i=0}^{p} a(i)x(n-i) = \sum_{i=0}^{q} b(i)w(n-i) \tag{38}$$

then the AR parameters of a non-Gaussian causal ARMA(p, q) process are obtained by solving the following equations:

$$\sum_{i=0}^{p} a(i)c_{kx}(\tau_1 - i, \tau_2, 0, \ldots, 0) = 0, \qquad a(0) = 1 \tag{39}$$

for $\tau_1 = q+1, \ldots, q+p$, $\tau_2 = q-p, \ldots, q$. Normalization with $a(0) = 1$ is used, without loss of generality, to fix the scale ambiguity. Equation (39) can be used with $k = 2$ as well, but $k \ge 3$ in Eq. (39) offers suppression of AGN and can also be modified for noncausal ARMA models as reported in Ref. 6. Note also that pure (non)causal AR(p) models can be handled as a special case of ARMA(p, q), with $q = 0$.

Having estimated the AR parameters, we can inverse filter $x(n)$ to obtain

$$\tilde{x}(n) := \sum_{i=0}^{p} a(i)x(n-i) \tag{40}$$

which is an AR-compensated MA(q) process. The MA parameters $\{b(i)\}_{i=0}^{q}$ can be found from the following linear equations (see Ref. 5):

$$\frac{\sigma_w^2}{\gamma_{kw}} \sum_{i=0}^{q} b(i)c_{k\tilde{x}}(\tau_1, \ldots, \tau_{k-2}; l+i) = \sum_{i=0}^{q} b(i)\ldots b(i+\tau_{k-2})c_{2\tilde{x}}(i+l) \tag{41}$$

which relate second- with higher-order cumulants of linear processes for $k \ge 3$. Note that with $x(n)$ replacing $\tilde{x}(n)$ Eq. (41) applies also to pure MA(q) models.

Determining the ARMA orders p and q is crucial, and HOS-based model order selection methods are available to complement the computationally simpler linear parameter estimation methods of Eqs. (39) and (41) (see Refs. 6 and 57). Especially for MA parameter estimation, accurate order determination is instrumental in implementing the closed-form solution [Eq. (19)]—an even simpler alternative to Eq. (41)

from a computational point of view. If, on the other hand, statistical estimation accuracy and efficiency are one's concern, computationally more intensive nonlinear methods can improve upon the linear equation ARMA parameter estimates $\boldsymbol{\theta} := [a(1), \ldots, a(p); b(0), \ldots, b(q)]$.

Such methods require nonlinear optimization programs that are initialized by the linear estimates and optimize lag- or frequency-domain quadratic HOS-matching criteria such as (Refs. 3, 4, 22)

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{\boldsymbol{\tau}} [\hat{c}_{kx}(\boldsymbol{\tau}) - c_{kx}(\boldsymbol{\tau}; \boldsymbol{\theta})]^2 \tag{42}$$

Asymptotically optimal (minimum variance) modifications of Eq. (42) are possible using weighted inverse covariance of cumulant estimators (see Refs. 1 and 22 for a unifying treatment). Asymptotic efficiency in Eq. (42) is traded for the computational complexity relative to that involved in Eqs. (39), (41), and in the presence of local minima in the nonconvex objective function in Eq. (42). Sufficient lags to guarantee identifiability and thus consistency are discussed in Refs. 6 and 22, and references therein.

Inverse Modeling

In several applications such as seismic deconvolution and channel equalization, the inverse model is sought in order to undo the effects of $h(n)$, which is often assumed to obey an all-zero (i.e., FIR) model. The inverse model $g(n)$ allows deconvolution of $x(n)$ in Eq. (10) and thus recovery of the unobserved information-bearing input $w(n)$—the ultimate goal in the aforementioned applications. For computational efficiency, $g(n)$ is designed to be FIR in practice. One deconvolution approach is to simply invert the "direct model" $H(\omega)$ found using the (non)parametric methods of the previous subsections [i.e., $G(\omega) = 1/H(\omega)$] and then truncate the resulting IIR $g(n)$ to obtain the deconvolver. When the signal-to-noise ratio (SNR) is known and is sufficiently low, one can also regularize the inverse by incorporating the SNR in the design of the so-called minimum-mean-squared-error (m.m.s.e., or Wiener) inverse model (see Refs. 1, 2, 31).

An alternative to the linear two-step (direct first, inverse next) approaches is to seek the inverse system directly using, for example, the polycepstral methods of Ref. 9 (see also Ref. 32) or adopt the parametric methods of the first subsection of this section applied to the *inverse cumulants* $c^I_{kx}(\boldsymbol{\tau})$. The latter are estimated by inverse Fourier transformation of $1/\hat{S}_{kx}(\boldsymbol{\omega})$ as

$$\hat{c}^I_{kx}(\boldsymbol{\tau}) = \sum_{\boldsymbol{i}=0}^{N-1} \frac{1}{\hat{S}_{kx}(2\pi \boldsymbol{i}/N)} e^{j(2\pi/N)\boldsymbol{i}\boldsymbol{\tau}'}$$

Note that "inverse cumulants" of the direct MA model are "direct cumulants" of the inverse (AR) model, and thus the linear equations [Eq. (39)] can be applied directly to $\hat{c}^I_{kx}(\boldsymbol{\tau})$ in order to estimate the deconvolver coefficients.

Nonlinear HOS-based criteria have also been developed for estimating directly the inverse model $\{g(i)\}_{i=d_0}^{K+d_0}$ (see, e.g., Ref. 33). They rely upon the idea that the input $w(n)$, and thus the deconvolved sequence $\hat{w}(n) = \sum_{i=d_0}^{K+d_0} g(i)x(n-i)$, must be i.i.d. as well, and according to property (P3), its cumulants should be delta sequences. With $\boldsymbol{g}_0$ denoting the deconvolver coefficient vector [normalized to assure that $g_0(0) = 1$], one method of enforcing the delta-like structure is to suppress the cumulant energy away from the zero lags using the criterion

$$\hat{\boldsymbol{g}}_0 = \arg\min_{\boldsymbol{g}_0} \sum_{\boldsymbol{\tau} \neq \boldsymbol{0}} |c_{k\hat{w}}(\boldsymbol{\tau})|^2$$

The constraint $g_0(0) = 1$ prevents the convergence to the trivial solution $\hat{\boldsymbol{g}}_0 = \boldsymbol{0}$. Potential convergence to local minima becomes less likely when the nonlinear search (steepest-descent or Gauss–Newton) is initialized with estimates obtained from the linear inverse modeling methods.

Instead of fixing $g_0(0) = 1$, an alternative constraint is to fix the norm of $\boldsymbol{g}$, or optimize cumulant ratios such as the one proposed for blind equalization of communication channels in Ref. 33:

$$\hat{\boldsymbol{g}} = \arg\min_{\boldsymbol{g}} \frac{|c_{4\hat{w}}(\boldsymbol{0})|}{c^2_{2\hat{w}}(0)} \tag{43}$$

With $f(n) = \sum_l h(l)g(n-l)$ denoting the convolver–deconvolver cascade, the ratio in Eq. (43) is proportional to $\sum_n |f(n)|^4/[\sum_n |f(n)|^2]^2$, which is maximized when $f(n) = \alpha\, \delta(n)$. Thus, the coefficient vector $\hat{\boldsymbol{g}}$ that solves Eq. (43) must also equalize the channel $h(n)$.

In all inverse methods described so far, it is assumed that $H(\omega)$ has no zeros on the unit circle. To alleviate this assumption, one usually starts with the linear m.m.s.e. inverse mentioned earlier (see Refs. 1 and 31), before switching to a nonlinear approach. In both cases, where one centers the deconvolver (i.e., the delay d_0) affects the deconvolution performance, and common practice suggests the choice $d_0 = -[K/2]$.

If $w(n)$ satisfies additional properties such as finite alphabet or constant modulus, then approaches that capitalize on such *a priori* knowledge often lead to more reliable HOS-based algorithms, because the exploitation of extra information can help reduce the high variance associated with the HOS-based methods. It turns out that the criterion [Eq. (43)] leads to the well-known constant-modulus algorithm (CMA) for QPSK sources $w(n)$; see references in Ref. 33. CMA forces the CM constraint on the equalized $\hat{w}(n)$ by selecting the equalizer that minimizes $E\{|\hat{w}(n)|^2 - \rho\}^2$, where $\rho := |m_{4w}(\boldsymbol{0})|/|m_{2w}(0)|$.

Both direct and inverse deconvolution methods can be applied in a block-by-block fashion. If real-time processing is desired, they can be recast in an adaptive format as well. Adaptive deconvolution updates the inverse model coefficients as each new datum becomes available, and is well motivated not only from a computational perspective [matrix inversions involved in Eqs. (39) and (41) are replaced by scalar divisions], but also when the underlying convolver is slowly time-varying and should be tracked. A major tradeoff is between parameter estimation accuracy, which favors longer data records, and tracking capability, which calls for smaller data windows (see Refs. 1, 9, and 34 for further reading on HOS-based adaptive algorithms).

If $w(n)$ comes from a finite-alphabet source, a nonlinear inverse method that bypasses the no-unit-circle-zeros constraint and exhibits excellent performance when initialized properly is the decision-feedback equalization (DFE) approach. HOS-based DFEs quantize (or project) the linear

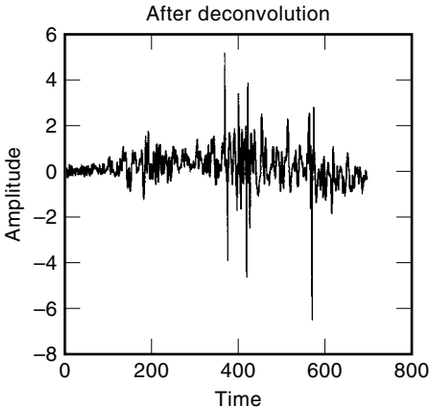

Figure 7. A non-causal and non-minimum-phase model was found for the seismic wavelet and was used to deconvolve the received data to yield the reflectivity sequence. Success of deconvolution is revealed by the more spiky appearance of the estimated reflectivity sequence.

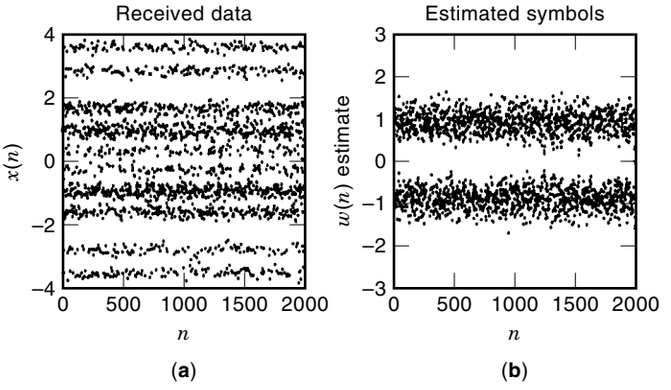

Figure 8. I.i.d. information symbols $w(n) \in \{1, -1\}$ pass through a non-minimum-phase FIR channel, and additive Gaussian noise is also present in the received data as shown in (a). The FIR channel is estimated using fourth-order statistics of the data, and deconvolution is performed to generate the symbol estimates as shown in (b).

equalizer output onto the finite-alphabet set and use it to generate an updated channel/equalizer estimate for the next iteration using input–output correlations. Convergence of DFEs is a current research topic and goes beyond the scope of this section. The following two examples illustrate the main points in the context of HOS-based seismic and communications data processing.

Example 5. *Deconvolution of seismic data.* Using fourth-order statistics calculated from multiple seismic traces [see Fig. 1(a) for one of the traces], we obtained ARMA(5,5) parameter estimates $\boldsymbol{a} = [1, -4.0784, 7.6817, -8.0893, 4.7562, -1.5012]$ and $\boldsymbol{b} = [1, 1.2239, 11.7788, -20.7298, 26.0287, -19.2216]$ for the seismic wavelet, which renders the model both non-causal and non-minimum-phase. An FIR Wiener inverse was calculated and used to deconvolve the received data to produce an estimate of the reflectivity sequence shown in Fig. 7. Success of the algorithm is revealed by the spikier appearance of the after-deconvolution result [compare with Fig. 1(a)].

Example 6. *Equalization of a non-minimum-phase communication channel.* Consider a channel $\boldsymbol{h} = [1, 0.3768, -1.2499, -0.9379]$, which is non-minimum-phase. The input symbols $w(n)$ are i.i.d., and take on ± 1 values with equal probability. AGN $v(n)$ with SNR = 20 dB is also present. Figure 8(a) shows $N = 2000$ received data samples, from which the distribution of $w(n)$ cannot be readily recognized. We first used fourth-order statistics of $x(n)$ to find the channel estimate $\hat{\boldsymbol{h}} = [1, 0.4487, -1.2378, -0.8664]$. Afterwards, we obtained symbol estimates $\hat{w}(n)$ as shown in Fig. 8(b). The correct symbol pattern was revealed.

CHECKING AND IDENTIFYING NONLINEARITIES

If $x(n)$ is a linear process obeying the model in Eq. (10), then it follows that the theoretical kth-order coherence $\beta_{kx}(\boldsymbol{\omega})$ [ensemble counterpart of Eq. (34)] satisfies

$$|\beta_{kx}(\boldsymbol{\omega})|^2 = \frac{\gamma_{kw}^2}{\sigma_w^{2k}} \qquad \forall \boldsymbol{\omega} \tag{44}$$

Therefore, checking for constancy in the kth-order coherence with respect to $\boldsymbol{\omega}$ amounts to testing for linearity (see Ref. 27). Hence, if a stationary non-Gaussian process $x(n)$ has

$$D_L := \sum_{\alpha,\beta} [\hat{\beta}_{kx}(\boldsymbol{\omega}_\alpha) - \hat{\beta}_{kx}(\boldsymbol{\omega}_\beta)]^2 \neq 0$$

then $x(n)$ is nonlinear. Statistical nonlinearity tests are reported in Ref. 28 and have been applied for characterization of local shipping (ambient) noise using the bicoherence ($k = 3$) in Ref. 35. The following example shows how the bicoherence can be used as an exploratory tool to discern nonlinearity in a random process.

Example 8. *Bicoherence of linear and nonlinear processes.* Consider an input process $w(n)$ that is zero-mean, i.i.d., and exponentially distributed. It passes through an LTI system to yield output $x_1(n)$. A second process is formed as $x^2(n) = x_1(n) + w_2(n) + 0.5\, w(n)w(n-2)$ and is thus nonlinear. Available are 16,384 samples for both processes, and bicoherence estimates are obtained according to Eq. (34) with $k = 3$. The resulting $|\hat{\beta}_{3x}(\omega_1, \omega_2)|^2$ plots are shown as Fig. 9(a,b). Although

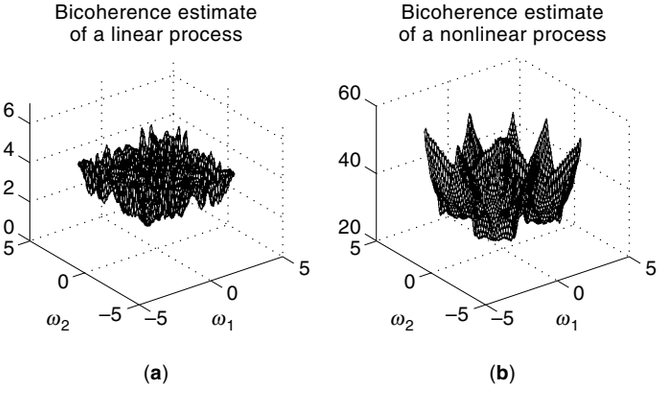

Figure 9. The bicoherence estimate is seen to be fairly constant in (a) for a linear process, but not so constant in (b) for a nonlinear process.

the theory predicts that the bicoherence of the linear process $x_1(n)$ should be flat, its sample estimate shown in Fig. 9(a) is only approximately constant. The bicoherence for $x_2(n)$, however, is by no means constant. Statistical tests can be employed to check on the constancy of $\hat{\beta}_{kx}(\boldsymbol{\omega})$ (see Ref. 28).

As we saw in the introductory section, if the harmonic process $w(n) = \exp[j(\omega_1 n + \phi_1)] + \exp[j(\omega_2 n + \phi_2)]$ goes through a memoryless linear–quadratic transformation $x(n) = w(n) + \alpha w^2(n)$, then harmonics with frequencies $2\omega_1$, $2\omega_2$, and $\omega_1 + \omega_2$ with corresponding phases $2\phi_1$, $2\phi_2$, and $\phi_1 + \phi_2$ appear in $x(n)$ besides those already present in $w(n)$. In general, presence of the sum frequency $\sum_{i=1}^{k}\omega_i$ and the sum phase $\sum_{i=1}^{k}\phi_i$ is referred to as kth-order coupling and is a strong indication that a nonlinear transformation may have taken place. Detection and estimation of nonlinearly interacting harmonic components using polyspectra have found diverse applications in areas such as EEG analysis (Ref. 26) and fluid dynamics (Ref. 27).

To clarify the role of polyspectra in checking for nonlinearities, consider $x(n) = \sum_{l=1}^{L} A_l \exp[j(\omega_l n + \phi_l)] + v(n)$, and define the scaled moment polyspectrum as

$$\overline{M}_{k+1}(\boldsymbol{\lambda}) := \lim_{N\to\infty} \frac{1}{N^{k+1}} E\{X(\lambda_1)\ldots X(\lambda_k)X^*(\lambda_1+\cdots+\lambda_k)\} \tag{45}$$

where $\boldsymbol{\lambda} := (\lambda_1, \ldots, \lambda_k)$, and $X(\lambda)$ is the FT of $x(n)$. If the ϕ_l's obey the so-called phase randomization assumption [i.e., they are i.i.d., uniformly distributed, and independent of $v(n)$], then it turns out that (Ref. 18)

$$\overline{M}_{(k+1)x}(\boldsymbol{\lambda}) = \sum_{l_1\cdots l_{k+1}=1}^{L} \left(\prod_{i=1}^{k+1} A_{l_i}\right) E\{e^{j(-\phi_{k+1}+\sum_{i=1}^{k}\phi_{l_i})}\} \times \delta(\lambda_1-\omega_{l_1})\ldots\delta(\lambda_k-\omega_{l_k})\delta(\omega_{l_1}+\cdots+\omega_{l_k}-\omega_{l_{k+1}}) \tag{46}$$

where $\delta(\lambda) = 1$ for $\lambda = 0 \bmod 2\pi$ and zero elsewhere. Both frequency coupling, $\omega_{l_k+1} = \sum_{i=1}^{k}\omega_{l_i}$, and phase coupling, $\phi_{l_k+1} = \sum_{i=1}^{k}\phi_{l_i}$, are needed for $\overline{M}_{(k+1)x}$ in Eq. (46) to be nonzero and peak at $(\omega_{l_1}, \ldots, \omega_{l_k})$. However, if the ϕ_l's are assumed deterministic (which is the case when only a single record is available), frequency coupling alone suffices for $\overline{M}_{(k+1)x}(\boldsymbol{\lambda})$ to peak (Ref. 18). Phase randomization causes lack of ergodicity, but it is possible to detect frequency and phase coupling from a single data record under the deterministic assumption (Ref. 18). Estimation of quadratically coupled frequencies is reported in Ref. 38, and asymptotics of $\hat{\overline{M}}_{(k+1)x}$ for uncoupled harmonics with constant or random amplitudes can be found in Ref. 18.

Harmonic processes do not satisfy the mixing condition Eq. (14), and their (poly)spectral estimation should also be conceptually different. It is explained in Ref. 18 that when it comes to (poly)spectral estimation of harmonic processes, no windowing is necessary and the raw (poly)periodograms are themselves consistent. However, proper scaling by the data length, similar to the one used in Eq. (45), is often necessary to ensure that the quantities have finite values. Sample estimates of $\overline{M}_{(k+1)x}(\boldsymbol{\lambda})$ are obtained by simply dropping the limit and the expected-value operators in Eq. (45).

To avoid the increase in dimensionality and thus computations when dealing with nonlinearities of increasing orders, one may test for the presence of self-coupling, which amounts to frequency and phase pairs of the form $(\omega_0, k\omega_0)$ and $(\phi_0, k\phi_0)$, respectively. Such pairs give rise to peaks at $\lambda = \omega_0$ in the 1-D slice (Ref. 39)

$$\overline{M}_{(k+1)x}(\lambda, \ldots, \lambda) := \lim_{N\to\infty} \frac{1}{N^{k+1}} X^k(\lambda)X^*(k\lambda) \tag{47}$$

The following example explains the points we have made so far in this section.

Example 7. *Polyspectra and coupled harmonics.* Suppose that we have available a noisy harmonic (cosinusoidal) signal

$$x(n) = \sum_{i=1}^{4} \cos(\omega_i n + \phi_i) + v(n)$$

where $v(n)$, is zero-mean white Gaussian with variance $\sigma_v^2 = 0.1$. Figure 10(a) shows the bispectral estimate of $x(n)$ obtained from the (scaled) biperiodogram [sample version of Eq. (45) with $k = 2$], and Fig. 10(b) is its contour plot. From the peak positions (0.5, 0.5) and (1.2, 1.2) that occur along the diagonal, we infer that frequencies 0.5 and 1.2 and their respective doubles 1 and 2.4 must be present in $x(n)$. The diagonal slice statistic $\overline{M}_{3x}(\lambda)$ plotted in Fig. 10(c) reveals the same information. The above assessment agrees with the true

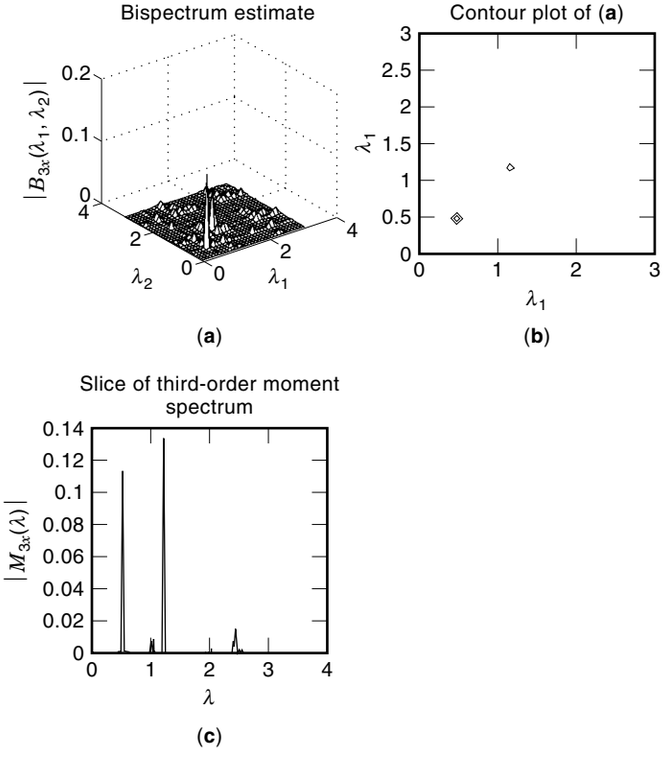

Figure 10. The signal consists of four cosinusoids with frequencies 0.5, 1, 1.2, and 2.4 and additive Gaussian noise is also present. The bispectrum estimate (a) and its contour plot (b) show peaks along the diagonal axis at $(\lambda_1, \lambda_2) = (0.5, 0.5)$ and (1.2, 1.2). The diagonal slice statistic (c) reveals the same information but is easier to compute.

model parameters that were used in generating $x(n)$: $\omega_1 = 0.5$, $\phi_1 = 0.4$, $\omega_2 = 1$, $\phi_2 = -0.5$, $\omega_3 = 1.2$, $\phi_3 = 0.5$, $\omega_4 = 2.4$, $\phi_3 = 1$.

In addition to checking for nonlinearities, (cross)polyspectra play an important role in identifying nonlinear (e.g., Volterra) models (see Refs. 28, 40 and 41). As an example, consider the linear–quadratic system

$$x(n) = \sum_{\tau} h_1(\tau) w(n-\tau) + \sum_{\tau_1,\tau_2} h_2(\tau_1,\tau_2) w(n-\tau_1) w(n-\tau_2) \tag{48}$$

Define 1-D and 2-D transfer functions as

$$H_1(\omega) = \sum_{\tau} h_1(\tau) e^{-j\omega\tau}$$
$$H_2(\omega_1,\omega_2) = \sum_{\tau_1}\sum_{\tau_2} h_2(\tau_1,\tau_2) e^{-j\omega_1\tau_1} e^{-j\omega_2\tau_2}$$

Next consider the second- and third-order cross cumulants

$$c_{xw}(\tau) = \text{cum}\{x(n), w(n-\tau)\}$$
$$c_{xww}(\tau_1,\tau_2) = \text{cum}\{x(n), w(n-\tau_1), w(n-\tau_2)\}$$

and their FTs,

$$S_{xw}(\omega) = \sum_{\tau} c_{xw}(\tau) e^{-j\omega\tau}$$
$$S_{xww}(\omega_1,\omega_2) = \sum_{\tau_1}\sum_{\tau_2} c_{xww}(\tau_1,\tau_2) e^{-j\omega_1\tau_1} e^{-j\omega_2\tau_2}$$

When $w(n)$ is Gaussian, then using (P1), we can prove that the kernels can be decoupled and identified via (see Ref. 40)

$$H_1(\omega) = \frac{S_{xw}(\omega)}{S_{2w}(\omega)} \quad \text{and} \quad H_2(\omega_1,\omega_2) = \frac{S_{xww}(-\omega_1,-\omega_2)}{S_{2w}(\omega_1) S_{2w}(\omega_2)} \tag{49}$$

If in addition, $w(n)$ is white, then we obtain directly in the time domain

$$h_1(\tau) = \frac{c_{xw}(\tau)}{\sigma_w^2}, \qquad h_2(\tau_1,\tau_2) = \frac{c_{xww}(\tau_1,\tau_2)}{\sigma_w^4} \tag{50}$$

In general, $(k+1)$st-order cross-spectra are required to identify kth-order kernels. Nonlinear distortions appear when high-power amplifiers operate in their saturation region, and Volterra kernel estimation is an important step in constructing pre- or postcompensators (approximate inverses) of nonlinear channels. Nonlinear compensators linearize amplifier characteristics and are very useful, especially for high-efficiency low-power satellite communication systems.

SEPARATING MULTIPLE SOURCES

Following the HOS-based harmonic retrieval ideas of Ref. 16, a number of frequency estimation and narrowband array-processing algorithms have been developed with the sole motivation of suppressing additive Gaussian noise of unknown color and correlation from sensor to sensor. The same motivation in the wideband case prompted HOS-based time-delay and differential delay–Doppler estimation algorithms (see Refs. 42 and 43). Despite the high variance associated with the HOS estimators, it is possible to achieve processing gains at low SNRs for strongly non-Gaussian wideband signals observed in correlated Gaussian noise. For harmonic signals, however, the mixing condition [Eq. (14)] is not met. It is pointed out in Refs. 18 and 44 that HOS may not offer advantage over SOS when uncoupled harmonics are retrieved in the presence of stationary and mixing noise, regardless of its color and distribution.

In sensor array processing, HOS have played a major role in identifying the mixing matrix $\mathbf{A}$ and separating the sources $s(n)$ in the vector model

$$\boldsymbol{x}(n) = \mathbf{A}(\boldsymbol{\theta})\boldsymbol{s}(n) + \boldsymbol{v}(n)$$

where the $N_x \times 1$ vectors $\{\boldsymbol{x}(n)\}_{n=0}^{N-1}$ are the N snapshots collected at N_x sensors, and the ith entry of the $N_s \times 1$ vector $\boldsymbol{s}(n)$ is the ith source signal $s_i(n)$. When any of the plane wave, narrowband, far-field, and whiteness assumptions about the sources does not hold true, the matrix $\mathbf{A}$ is not of the Vandermonde type and SOS-based direction of arrival (DOA) algorithms are not applicable. Relying upon the independence among non-Gaussian sources but allowing for temporal source correlation, HOS-based eigenapproaches have been developed for blind identification of $\mathbf{A}$, source separation, and deconvolution (see Refs. 45–47, and references therein).

HOS-based DOA algorithms have been motivated by the need to suppress additive (perhaps correlated and colored) Gaussian noise, but, as we explained earlier, such a gain may not always be feasible. A strong motivation, however, is the overdeterminacy offered by HOS, which allows for estimation of more sources with less sensors (Ref. 48), or equivalently creates "virtual" sensors for a given number of sources (Ref. 49). Taking advantage of this overdeterminacy, HOS-based nonlinear (Ref. 48) and ESPRIT-type methods have been developed [the latter obviate the need for a full copy of the array (Ref. 49)]. More recent advances include applications to speaker separation and multiple wideband signal separation (Ref. 50). A simple yet motivating example on HOS-based source separation is given next.

Example 9. *Blind source separation.* Consider a two-input, two-output system

$$x_1(n) = w_1(n) + \alpha w_2(n) + v_1(n) \tag{51}$$
$$x_2(n) = \beta w_1(n) + w_2(n) + v_2(n) \tag{52}$$

where $w_1(n) \in \{1, -1\}$ and $w_2(n) \in \{1, -1, 3, -3\}$ are the i.i.d. and mutually independent inputs. The crosstalk coefficients are $\alpha = -0.2$ and $\beta = 0.4$, and $v_1(n)$ and $v_2(n)$ are additive Gaussian noise processes that are independent of each other and independent of $w_1(n)$ and $w_2(n)$. The signal-to-interference-and-noise ratio (SINR) is 5 dB for $x_1(n)$ and 14 dB for $x_2(n)$. Given $N = 1000$ samples of $x_1(n)$ and $x_2(n)$ as shown in Fig. 11(a,b), we would like to estimate the inputs $w_1(n)$ and $w_2(n)$.

Capitalizing on the independence between $w_1(n)$ and $w_2(n)$ and using (P3), one can verify that the output processes sat-

isfy the following cumulant-based equations:

$$\begin{aligned}
\text{cum}\{x_1(n), x_1(n), x_1(n), x_1(n)\} &= \gamma_{4w_1} + \alpha^4 + \gamma_{4w_2} \\
\text{cum}\{x_2(n), x_2(n), x_2(n), x_2(n)\} &= \beta^4 + \gamma_{4w_1} + \gamma_{4w_2} \\
\text{cum}\{x_1(n), x_1(n), x_2(n), x_2(n)\} &= \beta^2\gamma_{4w_1} + \alpha^2 + \gamma_{4w_2} \\
\text{cum}\{x_1(n), x_2(n), x_2(n), x_2(n)\} &= \beta^3\gamma_{4w_1} + \alpha\gamma_{4w_2}
\end{aligned}$$

Therefore, one can solve the four equations for the four unknowns α, β, γ_{4w_1}, and γ_{4w_2}. If prior knowledge about the inputs is available, then γ_{4w_1} and γ_{4w_2} are known and α and β can be solved for in a straightforward manner using only the first two equations. Once α and β are found, the $\hat{w}_1(n)$, $\hat{w}_2(n)$ estimates are obtained as shown in Fig. 11(c,d), and the original symbol patterns are revealed. A final decoding stage will follow to quantize these symbol estimates. To examine the performance of the above fourth-order cumulant-based algorithm, we varied the data length from 100 to 4000 in increments of 200 and calculated the m.s.e., $N^{-1}\sum_{n=0}^{N-1}[\hat{w}_i(n) - w_i(n)]^2$. The resulting m.s.e. based on 500 independent realizations is shown in Fig. 12. We observe that for both channels, the performance of the algorithm improves as the data length N increases.

The memoryless mixture $\mathbf{A}(\boldsymbol{\theta})\boldsymbol{s}(n)$ often encountered in array processing is a special case of a multichannel ARMA

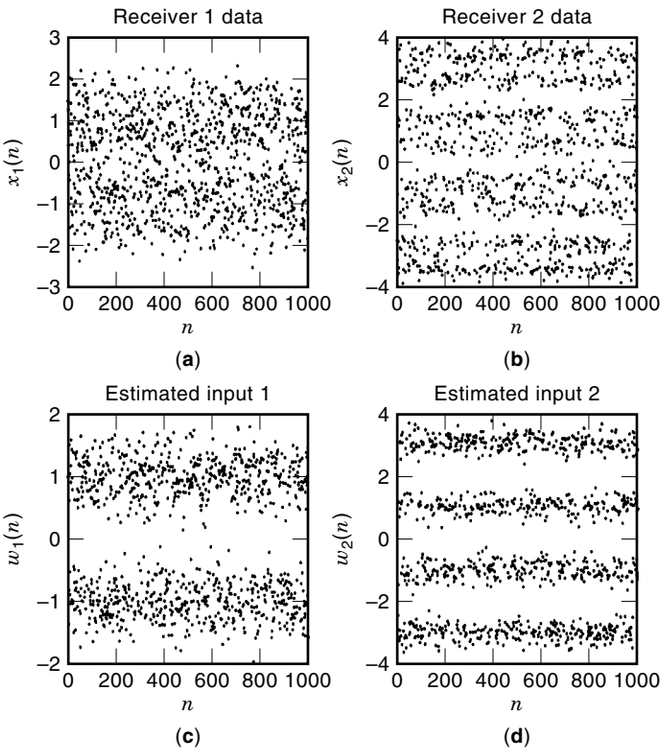

Figure 11. Two sensors receive symbols from two independent sources in the presence of additive Gaussian noise. Source 1 transmits i.i.d. symbols $w_1(n) \in \{1, -1\}$, whereas source 2 sends $w_2(n) \in \{1, -1, 3, -3\}$. (a) and (b) show that data received by sensors 1 and 2, respectively. Fourth-order statistics are utilized to yield source symbol estimates as shown in (c) and (d) and correct symbol patterns are revealed.

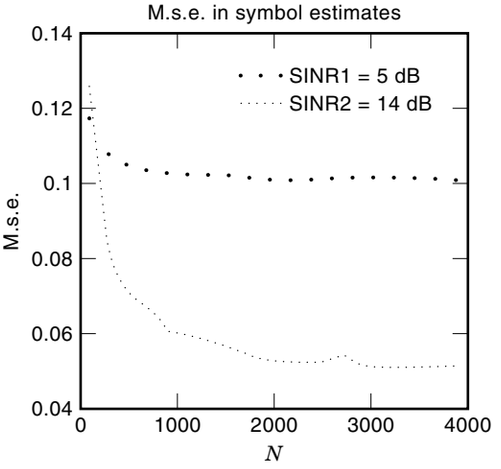

Figure 12. At a given signal-to-interference-and-noise ratio (SINR), the mean squared error in the symbol estimates, prior to decoding, is seen to improve as the data length is increased.

model

$$\sum_{i=0}^{p}\mathbf{A}(i)\boldsymbol{x}(n-i) = \sum_{i=0}^{q}\mathbf{B}(i)\boldsymbol{s}(n-i)$$

provided that the components of $\boldsymbol{s}(n)$ are assumed spatially (but not necessarily temporally) independent. If the input is i.i.d. in both time and space, it is shown in Refs. 51 and 52 that up to multiplication by a diagonal and a permutation matrix (which amounts to shuffling the inputs), it is possible to identify the ARMA matrix coefficients using output HOS only—a significant improvement over SOS-based methods that assume minimum-phase matrix transfer functions and entail ambiguity corresponding to multiplication by a unitary matrix (see Ref. 46).

CONCLUSION AND FUTURE DIRECTIONS

Drawbacks of the HOS methods include increased dimensionality, heavier computational load, and increase in the variance of their sample estimates. The higher variance associated with higher orders can be understood from an information-theoretic viewpoint because quantities that *vary* have the potential of offering a *variety* of information. Also, from a numerical viewpoint, errors associated with triple or quadruple products of quantized numbers are often larger than those present in double products.

As a rule of thumb, $(k + 1)$st-order sample estimates require approximately an order of magnitude more data, in order to achieve comparable accuracy, than kth-order sample estimates. Such comparison clearly depends on the application as well. For example, a HOS-based detector requires less data than a HOS-based parameter estimation algorithm dealing with a number of parameters. Especially with noise suppression claims, one has to be particularly careful because the gain in the mean offered by HOS may not be enough to offset the deterioration in the variance. The success of HOS-based methods also depends on the degree of non-Gaussianity and nonlinearity involved, and in this context, Gaussianity and

linearity tests are a useful first step. Long data records help to reduce the variability of HOS, but the user should also be conscious about the nonstationarity that may arise as additional data become available—a typical case with images. In multidimensional signal and image processing, HOS approaches have focused on reconstruction from noisy frame(s), linear nonparametric and ARMA modeling, blur identification, object recognition, texture classification, texture model validation, and image motion estimation (see Refs. 11, 25, 26, 53, and 54, and references therein).

Relatively unexplored areas of HOS research include efficient algorithms for nonlinear signal and system analysis with emphasis on robust estimation methods and potential applications to chaotic time series. Truly adaptive algorithms that exploit the advantages of HOS and their performance analyses are yet to be addressed even for linear processes. Theory and HOS-based applications of point processes have been investigated in Ref. 55, but many open problems remain. Man-made communication signals are clearly non-Gaussian. Besides the success of HOS in tackling blind equalization problems, there is ample room for research, especially in topics that are related to synchronization, multiuser systems, and sensor arrays.

Finally, HOS extensions to nonstationary and especially cyclostationary processes are worth further study in view of the fact that real-life signals and systems are often only locally stationary and time-invariant.

ACKNOWLEDGMENT

This work was supported in part by ONR Grant N0014-93-1-0485 and NSF Grant MIP-9703312.

BIBLIOGRAPHY

1. B. Porat, *Digital Processing of Random Signals: Theory and Methods,* Englewood Cliffs, NJ: Prentice-Hall, 1994.
2. B. Porat, *A Course in Digital Processing,* New York: Wiley, 1997.
3. K. S. Lii and M. Rosenblatt, Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes, *Ann. Statist.,* **10**: 1195–1208, 1982.
4. J. K. Tugnait, Identification of linear stochastic systems via second- and fourth-order cumulant matching, *IEEE Trans. Inf. Theory,* **33**: 393–407, 1987.
5. G. B. Giannakis and J. M. Mendel, Identification of non-minimum phase systems using higher-order statistics, *IEEE Trans. Acoust. Speech Signal Process.,* **37**: 360–377, 1989.
6. G. B. Giannakis and A. Swami, Identifiability of general ARMA processes using linear cumulant-based estimators, *Automatica,* **38**: 771–779, 1992.
7. C. L. Nikias and M. R. Raghuveer, Bispectrum estimation: A digital signal processing framework, *Proc. IEEE,* **75**: 869–891, 1987.
8. J. M. Mendel, Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications, *Proc. IEEE,* **79**: 278–305, 1991.
9. C. L. Nikias and A. P. Petropulu, *Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework,* Englewood Cliffs, NJ: Prentice-Hall, 1993.
10. C. L. Nikias and J. M. Mendel, Signal processing with higher-order spectra, *IEEE Signal Process. Mag.,* **10** (3): 10–37, 1993.
11. A. W. Lohmann and B. Wirnitzer, Triple correlations, *Proc. IEEE,* **72**: 889–901, 1984.
12. A. Swami, G. B. Giannakis, and G. Zhou (eds.), *Bibliography on Higher Order Statistics,* Signal Process., 60, Amsterdam: Elsevier, 1997, pp. 65–126.
13. D. R. Brillinger, *Time Series, Data Analysis and Theory.* New York: McGraw-Hill, 1981.
14. M. Rosenblatt, *Stationary Sequences and Random Fields,* Boston: Birkhäuser, 1985.
15. D. R. Brillinger and M. Rosenblatt, Asymptotic theory of estimates of *k*th-order spectra, in B. Harris (ed.), *Spectral Analysis of Time Series,* New York: Wiley, 1967, pp. 153–188.
16. A. Swami and J. M. Mendel, Cumulant-based approach to the harmonic retrieval and related problems, *IEEE Trans. Signal Process.,* **39**: 1099–1109, 1991.
17. A. V. Dandawate and G. B. Giannakis, Asymptotic theory of mixed time averages and *k*th-order cyclic-moment and cumulant statistics, *IEEE Trans. Inf. Theory,* **41**: 216–232, 1995.
18. G. Zhou and G. B. Giannakis, Polyspectral analysis of mixed processes and coupled harmonics, *IEEE Trans. Inf. Theory,* **42**: 943–958, 1996.
19. G. B. Giannakis, Cumulants: A powerful tool in signal processing, *Proc. IEEE,* **75**: 1333–1334, 1987.
20. A. V. Dandawate and G. B. Giannakis, Nonparametric polyspectral estimators for *k*th-order (almost) cyclostationary processes, *IEEE Trans. Inf. Theory,* **40**: 67–84, 1994.
21. G. B. Giannakis and M. K. Tsatsanis, Signal detection and classification using matched filtering and higher-order statistics, *IEEE Trans. Acoust. Speech Signal Process.,* **38**: 1284–1296, 1990.
22. G. B. Giannakis and M. K. Tsatsanis, A unifying maximum-likelihood view of cumulant and polyspectral measures for non-Gaussian signal classification and estimation, *IEEE Trans. Inf. Theory,* **38**: 386–406, 1992.
23. G. B. Giannakis and A. Delopoulos, Cumulant based autocorrelation estimates of non-Gaussian linear processes, *Signal Process.,* **47** (1): 1–17, 1995.
24. A. Delopoulos and G. B. Giannakis, Consistent identification of stochastic linear systems with noisy input–output data, *Automatica,* **30**: 1271–1294, 1994.
25. M. K. Tsatsanis and G. B. Giannakis, Object and texture classification using higher-order statistics, *IEEE Trans. Pattern Anal. Mach. Intell.,* **14**: 733–750, 1992.
26. T. E. Hall and G. B. Giannakis, Bispectral analysis and model validation of texture images, *IEEE Trans. Image Process.,* **4**: 996–1009, 1995.
27. D. R. Brillinger, An introduction to polyspectra, *Ann. Math. Stat.,* **36**: 1351–1374, 1965.
28. T. S. Rao and M. M. Gabr, An introduction to bispectral analysis and bilinear time series models, in *Lecture Notes in Statist.,* New York: Springer-Verlag, 1984, p. 24.
29. G. B. Giannakis and M. K. Tsatsanis, Time-domain tests for Gaussianity and time-reversibility, *IEEE Trans. Signal Process.,* **42**: 3460–3472, 1994.
30. M. Rangoussi and G. B. Giannakis, FIR modeling using log-bispectra: Weighted least-squares algorithms and performance analysis, *IEEE Trans. Circuits Syst.,* **38**: 281–296, 1991.
31. B. Porat and B. Friedlander, Blind equalization of digital communication channels using high-order moments, *IEEE Trans. Signal Process.,* **39**: 522–526, 1991.
32. D. Hatzinakos and C. L. Nikias, Blind equalization using a tricepstrum based algorithm, *IEEE Trans. Commun.,* **39**: 669–682, 1991.
33. O. Shalvi and E. Weinstein, New criteria for blind deconvolution of nonminimum phase systems (channels), *IEEE Trans. Inf. Theory,* **36**: 312–321, 1990.

34. G. B. Giannakis (guest ed.), Adaptive signal processing using higher-order statistics, Special issue, *Int. J. Adaptive Control and Signal Process.,* March, 1996.

35. M. J. Hinich, D. Marandino, and E. J. Sullivan, Bispectrum of ship radiated noise, *J. Acoust. Soc. Am.,* **85**: 1512–1517, 1989.

36. P. J. Huber et al., Statistical methods for investigating phase relations in stationary stochastic processes, *IEEE Trans. Audio Electroacoust.,* **19**: 78–86, 1971.

37. Y. C. Kim and E. J. Powers, Digital bispectral analysis of self-excited fluctuation spectra, *Phys. Fluids,* **21**: 1452–1453, 1978.

38. D. R. Brillinger, The comparison of least squares and third order periodogram procedures in the estimation of bifrequency, *J. Time Ser. Anal.,* **1**: 95–102, 1980.

39. G. Zhou and G. B. Giannakis, Retrieval of self coupled harmonics, *IEEE Trans. Signal Process.,* **43**: 1173–1186, 1995.

40. L. J. Tick, The estimation of transfer functions of quadratic systems, *Technometrics,* **3**: 563–567, 1961.

41. D. R. Brillinger, The identification of a particular nonlinear time series systems, *Biometrika,* **65**: 509–515, 1977.

42. C. L. Nikias and R. Pan, Time delay estimation in unknown Gaussian spatially correlated noise, *IEEE Trans. Acoust. Speech Signal Process.,* **36**: 1706–1714, 1988.

43. A. V. Dandawate and G. B. Giannakis, Differential delay–Doppler estimation using second and higher order ambiguity functions, *IEE Proc. Part F,* **140**: 410–418, 1993.

44. G. Zhou and G. B. Giannakis, On estimating random amplitude modulated harmonics using higher-order spectra, *IEEE J. Oceanic Eng.,* **19**: 529–539, 1994.

45. J. F. Cardoso and A. Souloumiac, Blind beamforming for non-Gaussian signals, *IEE Proc. Part F,* **140**: 362–370, 1993.

46. P. Comon, Independent component analysis: a new concept?, *Signal Process.,* **36**: 287–314, 1994.

47. L. Tong, Y. Inouye, and R. Liu, Waveform-preserving blind estimation of multiple independent sources, *IEEE Trans. Signal Process.,* **41**: 2461–2470, 1993.

48. S. Shamsunder and G. B. Giannakis, Modeling of non-Gaussian array data using cumulants: DOA estimation of more sources with less sensors, *Signal Process.,* **30**: 279–297, 1993.

49. M. C. Dogan and J. M. Mendel, Cumulant-based blind optimum beamforming, *IEEE Trans. Aerosp. Electron. Syst.,* **30**: 722–741, 1994.

50. S. Shamsunder and G. B. Giannakis, Multichannel blind signal separation and reconstruction, *IEEE Trans. Speech Audio Process.,* **5**: 515–528, 1997.

51. G. B. Giannakis, Y. Inouye, and J. M. Mendel, Cumulant based identification of multichannel moving-average models, *IEEE Trans. Autom. Control,* **34**: 783–787, 1989.

52. A. Swami, G. B. Giannakis, and S. Shamsunder, Multichannel ARMA processes, *IEEE Trans. Signal Process.,* **42**: 898–913, 1994.

53. A. Swami, G. B. Giannakis, and J. M. Mendel, Linear modeling of multidimensional non-Gaussian processes using cumulants, *J. Multidimens. Signals Syst.,* **1**: 11–37, 1990.

54. B. M. Sadler and G. B. Giannakis, Shift and rotation invariant object reconstruction using the bispectrum, *J. Opt. Soc. Amer. A,* **9**: 57–69, 1992.

55. D. R. Brillinger, Distributions of particle displacements via higher order moment functions, *IEE Proc., Part F,* **140**: 390–394, 1993.

56. B. Sadler, G. B. Giannakis, and K.-S. Lii, Estimation and detection in the presence of non-Gaussian noise, *IEEE Trans. Signal Process.,* **42**: 2729–2741, 1994.

57. S. Shamsunder and G. B. Giannakis, Detection and estimation of non-Gaussian sources using higher-order statistics, *IEEE Trans. Signal Process.,* **42**: 1145–1155, 1994.

GEORGIOS B. GIANNAKIS
University of Virginia

G. TONG ZHOU
Georgia Institute of Technology

STATISTICS, FUZZY. See FUZZY STATISTICS.

STATISTICS WITH FUZZY DATA. See FUZZY STATISTICS.

STEALTH. See RADAR CROSS-SECTION.