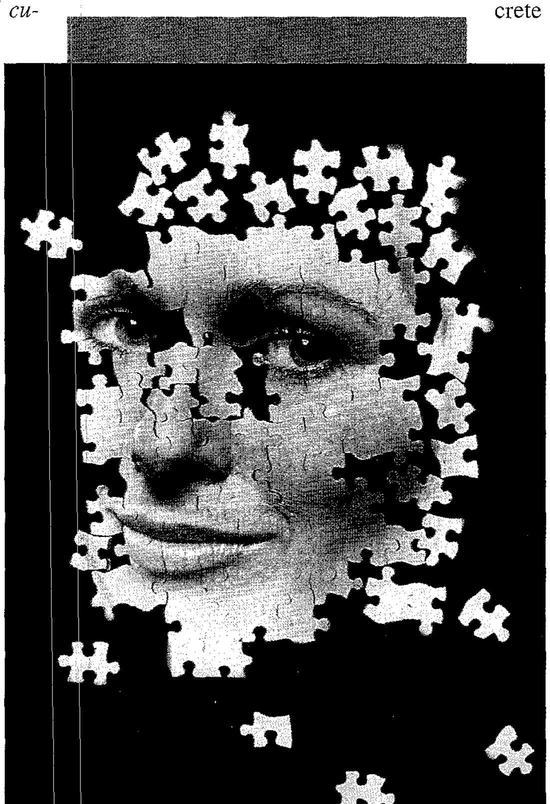
Blind Deconvolution via Cumulant Extrema

JAMES A. CADZOW

Classical *deconvolution* is concerned with the task of recovering an excitation signal, given the response of a known time-invariant linear operator to that excitation. In this article, deconvolution is discussed along with its more challenging counterpart, blind deconvolution, where no knowledge of the linear operator is assumed. This discussion focuses on a class of deconvolution algorithms based on higher-order statistics, and more particularly, cumulants. These algorithms offer the potential of superior performance in both the noise free and noisy data cases relative to that achieved by other deconvolution techniques. This article provides a tutorial description as well as presenting new results on many of the fundamental higher-order concepts used in deconvolution, with the emphasis on maximizing the deconvolved signal's *normalized cumulant*.



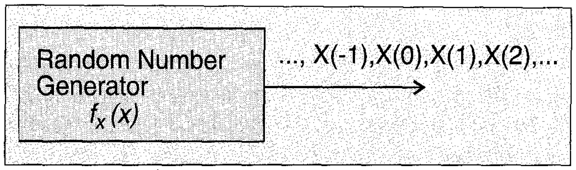
Alfred Gescheidt/The Image Bank

Overview

Most expositions of signal processing as applied to deconvolution begin with the assumption that the reader is conversant with the fundamental concepts of random time series and applicable higher order statistics (e.g., [4, 20, 21, 27, 28]). Unfortunately, many of these concepts are either unknown or only vaguely familiar to the potential user who wishes to apply them to real world problems. In order to make this important topic accessible to this audience, the first part of this article focuses on a review of appropriate concepts from higher-ordered statistics, which are then employed in characterizing stationary random time series. This review includes an overview of the density function associated with a random variable, and the corresponding moments and cumulants of that random variable as found in standard probability textbooks (e.g., [13, 16, 17, 26]). These concepts are then applied to the study of stationary random time series and the task of deconvolution.

In a typical signal processing application, one is concerned with the task of extracting information contained in a set of experimentally obtained data as designated by $\{x(n)\}$. It is useful to interpret the elements of this data set as being samples of a sequence of underlying random variables as denoted by $\{X(n)\}$. The indexing variable, n, exclusively takes on integer values and is frequently associated with *time*, although other descriptors may be more appropriate in a given application (e.g., distance, temperature). Thus, a discrete random variable sequence is often referred to as a *random time series*. In the language of probability theory, the nth element $x(n)$ of the data sequence is said to be a sample (or realization) of the associated random variable, $X(n)$, while the entire empirical time series $\{x(n)\}$ is a sample of the underlying random time series $\{X(n)\}$. We here adopt the convention of designating a random variable by an upper case Roman letter, and its realization by the corresponding lower case Roman letter. Further, in order to simplify the presentation, the random variables will be restricted to be real. Although an extension to the complex valued case is straightforward, it entails the introduction of additional notation, which only serves to cloud the fundamental issues being addressed.

Many of the more important results from signal proc-



1. Stationary random time series

essing are achieved by imposing a time invariance (i.e., stationary) behavior on the random time series being considered. The most basic stationary assumption is that each random element of the random time series is governed by the same probability density function as designated by $f_X(x)$. A random time series so characterized is said to be *stationary of order one* and may be interpreted as being a sequence of samples of an underlying random variable, X, as depicted in Fig. 1. Depending on the nature of the phenomenon represented, these samples may be stochastically dependent or independent. This indicates that the *stationary of order one* assumption implies no requirements concerning the nature of any statistical relationship that might exist between elements of the random time series. The underlying random variable X shall be referred to as the *generating random variable* for the random time series.

In order to impose a higher level of time invariance on a random time series, it is necessary to introduce higher orders of stationary. For example, the random time series $\{X(n)\}$ is said to be *stationary of order two* if the joint probability density function of the random variable pairs $(X(n_1), X(n_2))$ and $(X(n_1 + m), X(n_2 + m))$ is the same for all choices of the integers m, n_1 and n_2. Thus, the statistical nature of any two elements of a stationary of order two random time series is dependent only on their time difference, $n_2 - n_1$, and not on the specific values of n_1 and n_2. A stationary of order two random time series is also stationary of order one, but, the converse need not be true. We may extend this concept in a logical fashion to introduce higher orders of stationary. Thus, a random time series is stationary of order k if the joint density function of the sets $(X(n_1), X(n_2), ...X(n_k))$ and $(X(n_1 + m), X(n_2 + m),...,X(n_k + m))$ are the same for all choices of the integers $n_1, n_2, ...,n_k$ and m.

A class of time series that is of particular interest to theoretical studies has its elements composed of independent samples of an underlying generating random variable. A random time series so characterized is often referred to as *white noise*. The response of a time-invariant linear operator that is excited by a white noise input can be interpreted as being composed of dependent samples of another underlying generating random variable, Y. This response element dependency arises from the impact of the linear operation. It follows that the excitation and response time series so characterized are stationary of all orders. A random time series which is stationary of all orders shall hereafter be referred to simply as a *stationary* random time series. In the next few sections, we shall examine relevant characteristics of a random variable, with the ultimate objective of describing features of the generating random variable associated with a stationary random time series. These developments are then applied to the fundamental problem of blind deconvolution.

Deconvolution

In various applications, one seeks to identify the basic nature of a phenomenon by making empirical measurements of that phenomena. Due to factors such as instrumentation dynamics, environmental effects, and other considerations, these measurements actually take the form of a time-invariant linear convolution on a signal(s) that is of ultimate interest to the investigator. It is assumed that these measurements are a realization of a random time series governed by the time-invariant linear convolution operation

$$X(n) = \sum_{k \in K} f_k W(n-k) \tag{1}$$

where $\{X(n)\}$ is the excitation time series, $\{W(n)\}$, the stationary signal of interest, is the signal of interest, while the number sequence $\{f(n)\}$ designates the unit-impulse response of the distortion producing convolution operation. The index set, K, designates the set of integers over which the convolution operation is defined. This index set can be finite or infinite in size depending on the nature of the linear convolution operation. In the developments to follow, the index set associated with any time-invariant linear operation is implicitly implied but not explicitly written. Whatever the case, since this linear operation is time-invariant, it follows that the observed random time series $\{X(n)\}$ is also stationary.

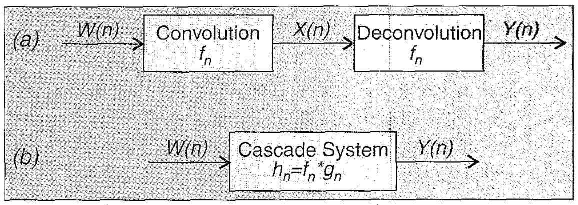
In the classical deconvolution problem, we desire to recover the excitation signal, $\{W(n)\}$, from a realization of the response time series $\{X(n)\}$. Since these two signals are related by a time-invariant linear operation, it follows that the required deconvolution must also take the form of a time-invariant linear operation, that is

$$Y(n) = \sum_{k} g_k X(n-k) \tag{2}$$

where $\{g_n\}$ is the unit-impulse response of the deconvolving operator. The summation index set of this deconvolution operation is generally different from that of convolution operation (Eq. 1). For example, if the index set of the convolution operator is finite and contains more than one integer then the index set of the deconvolution operator will contain an infinite number of integers. An *ideal deconvolution operation* is defined to be one for which the deconvolved random response time series, $\{Y(n)\}$ is a scaled and time-shifted image of $\{W(n)\}$, that is:

$$Y(n) = aW(n-m)$$

In this expression, a is an arbitrary nonzero scalar, and the time-shift parameter, m, is an arbitrary integer. If this identity holds for all input signals $\{W(n)\}$, the requirements for an *ideal deconvolution operation* are satisfied. A necessary and sufficient condition for satisfying this ideal deconvolution requirement is that the convolution of the unit-impulse responses of convolution operator (1) and the deconvolution operator (2) satisfy



2. Deconvolution: (a) convolution-deconvolution operation; (b) equivalent system.

$$\begin{aligned} h_n &= \sum_k f_k g_{n-k} \\ &= a\delta(n-m) \end{aligned} \tag{3}$$

where $\{\delta(n)\}$ designates the unit Kronecker delta sequence. as defined by:

$$\delta(n) = \begin{cases} 1 & n = 0 \\ 0 & \text{otherwise} \end{cases}$$

The validity of requirement (3) is readily confirmed by employing the z-transform and its associated properties [25]. In summary, an ideal deconvolution is achieved if and only if the cascading of the convolving and deconvolving operations results in a combined system whose unit-impulse response has precisely one nonzero element. When this requirement is met, the deconvolution operation effectively removes the distortion introduced by the original convolution operation. These concepts are illustrated in Fig. 2. If the unit-impulse response of the convolving operator (i.e., $\{f(n)\}$) is known, the data recovery process is commonly known as deconvolution. We shall be concerned with the more challenging task of *blind* deconvolution, in which knowledge of this unit-impulse information is not available.

Moments

As indicated in the previous section, a stationary random time series can be thought of as being generated by a sequence of statistically independent or dependent samples of an underlying *generating random variable*, here designated by X. The probability density function $f_X(x)$ governing this generating random variable X can be interpreted as a mathematical description of how a unit of mass is distributed along the x-axis. If this mass is distributed in a continuous fashion then X is said to be a continuous random variable and $f_X(x)$ is a continuous function of x. On the other hand, if the unit mass is located at only a finite or a countably infinite number of points on the x-axis, X is said to be a discrete random variable and $f_X(x)$ is composed of a sum of weighted displaced Dirac delta functions. When the unit mass is distributed in both a continuous and a discrete fashion, the random variable is mixed and the associated probability density function contains both continuous and Dirac delta components.

Whatever the nature of a random variable, it is possible to summarize the manner in which the unit mass is distributed in terms of a set of discrete parameters called *moments*. In particular, the nth order moment of random variable X is specified by

$$E\{X^n\} = \int_{-\infty}^{\infty} x^n f_X(x)dx \quad \text{for } n = 1,2,3,\cdots$$

where the symbol E denotes the expected value operator. If the nth order moment exists (i.e., is finite), it then follows that all moments of order smaller than n also exist. The first-order moment is commonly referred to as the *mean* value of random variable X and corresponds to the center of gravity of the unit mass distribution. To emphasize its importance, the special symbol $m_X = E\{X\}$ has been reserved for the mean value. It is one of the more important parameters describing a random variable.

Central Moments

The central moments provide a set of parameters that describe the manner in which a random variable's unit-mass is distributed about its mean value. The nth order central moment of random variable X is formally defined by

$$\begin{aligned} \mu_X(n) &= E\{[X - m_X]^n\} \\ &= \int_{-\infty}^{\infty} [x - m_X]^n f_X(x)dx \quad \text{for } n = 1,2,3,\cdots \end{aligned}$$

Clearly, a random variable's moments and central moments are identical when its mean value is zero. The second-order central moment is commonly referred to as the *variance* of random variable X and the special symbol $\sigma_X^2 = \mu_X(2)$ is reserved for its designation. Variance provides a measure of how dispersed the mass is about its center of gravity (mean).

The third-order central moment is typically used to measure the *skewness* of the density function about its mean value. For example, the skewness measure is zero if the density function is symmetric about it mean value (all odd central moments are zero as well in this case). The fourth-order central moment is often used to measure the *excess* or flatness (i.e., *kurtosis*) of the probability density function about its mean. Using the binomial theorem, it is found that a random variable's moments and central moments are interrelated by

$$\mu_X(n) = \sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} E\{X^{n-k}\} m_X^k$$

Moment Generating Functions

The Fourier transform serves as an important analysis and synthesis tool in mathematically based disciplines. It is therefore quite natural that its use in the study of random variables be considered. In particular, the Fourier transform of the probability density function of the random variable X is formally given by

$$\begin{aligned} \phi_X(\omega) &= \int_{-\infty}^{\infty} e^{j\omega x} f_X(x)dx \\ &= E\{e^{j\omega X}\} \end{aligned}$$

The function $\phi_X(\omega)$ actually corresponds to the complex conjugate of the probability density function's Fourier transform. In probability literature, this Fourier transform is commonly referred to as the *moment generating function* of random variable X. Thus, the moment generating function possesses all the properties associated with Fourier transforms. For example, at points of continuity the probability density function can be perfectly recovered from its associated moment generating function by means of the inverse Fourier transform relationship

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_X(\omega)e^{-j\omega x}d\omega$$

From this development, it is clear that the moment generating function conveys the same information concerning the underlying random variable as does the probability density function.

The moment generating function possesses a number of properties. From a signal processing perspective, three of its more important properties are given by:

$$\begin{aligned} Y = X + a &\Leftrightarrow \phi_Y(\omega) = e^{j\omega a}\,\phi_X(\omega) \\ Y = aX &\Leftrightarrow \phi_Y(\omega) = \phi_X(a\omega) \\ Y = X_1 + X_2 &\Leftrightarrow \phi_Y(\omega) = \phi_{X1}(\omega)\phi_{X2}(\omega) \end{aligned} \qquad (4)$$

where a is a scalar in properties one and two and X_1 and X_2 are independent random variables in property three.

The proofs of these properties follows directly from the moment generating function's definition. The first property indicates that adding a constant to a random variable only influences the phase behavior (in a linear fashion) but not the magnitude behavior of the associated moment generating function. The second property related to the scalar multiplication of a random variable shows that the corresponding moment generating function undergoes a scale change in the ω variable. Finally, the sum of two independent random variables has a moment generating function that is the product of the moment generating functions of the random variables being summed. This property is proven by noting that the moment generating function of Y can be decomposed as $E\{e^{j\omega Y}\} = E\{e^{j\omega X_1 + X_2}\} = E\{e^{j\omega X_1}\}E\{e^{j\omega X_2}\}$, where use of the statistical independence of the X_1 and X_2 random variables has been used in the last step. This property plays a prominent role in what is to follow.

In addition to possessing all the properties associated with the Fourier transform, the moment generating function possesses specialized properties due to the basic nature of a probability density function. Specifically, since a probability density function has unit area and is real it follows that

$$\begin{aligned} \phi_X(0) &= 1 \\ |\phi_X(\omega)| &\leq 1 \text{ for all } \omega \\ \overline{\phi}_X(\omega) &= \phi_X(-\omega) \end{aligned}$$

where the overbar symbol designates complex conjugation. Moreover, a sufficient condition for the Fourier transform of any function to exist is that the function be absolutely integrable. Since the density function is nonnegative real and has unit area, it follows that the moment generating function always exits.

Taylor Series Expansion

If the nth moment of random variable X exists, it is possible to make a truncated Taylor series expansion of the associated moment generating function about the origin ω=0. The kth coefficient of this expansion is formally obtained by evaluating the kth derivative of the moment generating function at ω= 0. From the definition of the moment generating function, the required coefficients are found to correspond to the moments of the random variable, that is

$$\left.\frac{d^k\phi_X(\omega)}{d\omega^k}\right|_{\omega=0} = (j)^k E\{X^k\} \quad \text{for } k = 1,2,3,\cdots,n$$

The desired series representation for the moment generating function is therefore

$$\phi_X(\omega) = \sum_{k=0}^{n}\frac{1}{k!}E\{X^k\}(j\omega)^k + o_n(\omega)$$

where the remainder function $o_n(\omega)$ is such that $o_n(\omega)/\omega^n$ goes to zero in the limit as ω approaches zero. If all the moments of X exist, then the moment generating function is exactly represented by the summation term in this expression, with the upper summation limit, n, replaced by plus infinity provided that this infinite summation is finite.

Cumulants

In probability applications directed toward the analysis of a linear combination of statistically independent random variables, it is shortly shown that there is much to be gained by considering the moment generating function's natural logarithm. This logarithm is commonly referred to as the *cumulant generating function* and is formally specified by

$$\psi_X(\omega) = \ln[\phi_X(\omega)] = \ln[\mathrm{E}\{e^{j\omega X}\}]$$

Properties of the moment generating function as specified in Eq. 4 carry directly over to the cumulant generating function, that is

$$\begin{aligned} Y = X + a &\Leftrightarrow \psi_Y(\omega) = j\omega a + \psi_X(\omega) \\ Y = aX &\Leftrightarrow \psi_Y(\omega) = \psi_X(a\omega) \\ Y = X_1 + X_2 &\Leftrightarrow \psi_Y(\omega) = \psi_{X1}(\omega) + \psi_{X2}(\omega) \end{aligned} \qquad (5)$$

in which a is a scalar in properties one and two and X_1 and X_2 are independent random variables in the third property.

Taylor Series Expansion

A revealing interpretation of the cumulant generating function is obtained by making a Taylor series expansion in the ω variable about the origin. The coefficient of the Taylor series term ω^k, multiplied by $(-j)^k$, is called the kth order cumulant and is formally given by

$$c_X(k) = (-j)^k \frac{d^k \psi_X(\omega)}{d\omega^k}\bigg|_{\omega=o} \quad \text{for } k = 1,2,3,\cdots \tag{6}$$

It is readily shown that these cumulants are real valued if the underlying random variable is real. Furthermore, if the nth moment of random variable X exists, the associated cumulants up to order n also exist. The cumulant generating function can be then represented by the truncated Taylor series expansion

$$\psi_X(\omega) = \sum_{k=1}^{n} \frac{1}{k!} c_X(k)(j\omega)^k + o_n(\omega) \tag{7}$$

where the remainder function, $o_n(\omega)$, is such that $o_n(\omega)/\omega^n$ goes to zero in the limit as ω approaches zero. If all the moments of X exist, then the cumulant generating function is exactly represented by the summation term in this expression, with the upper summation limit, n, replaced by plus infinity, provided that this infinite summation exists. It is subsequently shown that the concept of cumulants provides a particularly powerful means for characterizing the nature of stationary random time series.

It is possible to express the cumulants as functions of the mean and central moments of the random variable under analysis by using standard differentiation and identity (6). For example, the first eight cumulants as functions of the central moments are found to be [32]:

$$c(1) = m;\quad c(2) = \mu(2) = \sigma^2;\quad c(3) = \mu(3);\ c(4) = \mu(4) - 3\mu(2)^2;\ c(5) = \mu(5) - 10\,\mu(3)\mu(2);\quad c(6) = \mu(6) - 15\,\mu(4)\mu(2) - 10\text{-}\mu(3)^2 + 30\mu(2)^3\ \ c(7) = \mu(7) - 21\mu(5)\mu(2) - 35\,\mu(4)\mu(3) + 210\,\mu(3)\mu(2)^2\ \ c(8) = \mu(8) - 28\,\mu(6)\mu(2) - 56\mu(5)\mu(3) - 35\,\mu(4)^2 + 420\,\mu(4)\mu(2)^2 + 560\,\mu(3)^2\,\mu(2) - 630\,\mu(2)^4 \tag{8}$$

where, for purposes of clarity, the subscript X has been omitted in these expressions. In most practical applications, the probability density function of a random variable is unknown and the cumulants must be estimated from several realizations of the random variable. These estimates are typically obtained by first estimating the appropriate random variable's central moments and then inserting these estimates into relationship (8).

Linear Combinations of Statistically Independent Random Variables

In practical applications, it often happens that a random variable, Y, can be expressed as a linear combination of p statistically independent random variables, $\{W_1, W_2, \ldots, W_p\}$. This linear combination takes the form

$$Y = a_1 W_1 + a_2 W_2 + \ldots + a_p W_p$$

Due to the statistical independence of the W_k random variables, the moment generating function of Y is equal to the product of scaled versions of the moment generating functions associated with these constituent statistically independent random variables, that is:

$$\phi_Y(\omega) = \prod_{k=1}^{p} \phi_{W_k}(a_k \omega)$$

This identity directly follows from the second and third properties (4) associated with the scalar multiplication and sum of independent random variables. Upon taking the natural logarithm of this moment generating function, we obtain the associated cumulant generating function. Since the moment generating function is in the form of a product of terms, it follows that the cumulant generating function is equal to the sum of a_k scaled versions of the cumulant generating functions of the individual W_k random variable components, that is

$$\psi_Y(\omega) = \psi_{W_1}(a_1\omega) + \psi_{W_2}(a_2\omega) + \cdots + \psi_{W_p}(a_p\omega)$$

From the cumulant definition (6), the kth-order cumulant of random variable Y is obtained by evaluating the kth derivative of $\psi_Y = 0$ at $\omega = 0$. Carrying out this differentiation and evaluating the result at $\omega = 0$, we get

$$c_Y(k) = a_1^k c_{W_1}(k) + a_2^k c_{W_2}(k) + \cdots + a_p^k c_{W_p}(k)$$

in which $c_{W_m}(k)$ designates the kth order cumulant of random variable W_m. This result provides the basis for using cumulants in characterizing the effects of linear operations made on white noise time series.

Linear System Response, and Cumulants

A fundamental issue in signal processing is concerned with determining the relationships between the cumulants of a white noise excitation, and its associated time-invariant linear operator's response. In particular, it is assumed that the excitation random time series, $\{W(n)\}$, is composed of independent samples of an underlying generating random variable, W. Any time-invariant linear operation made on this random time series can always be represented as a convolution summation of form

$$Y(n) = \sum_{k \in K} h_k W(n-k) \tag{9}$$

where the sequence h_n corresponds to the linear operator's unit-impulse response defined on the integer set, K. Since this linear operator is time-invariant, it follows that the corresponding response time series, $\{Y(n)\}$, is also stationary. Thus, the response random time series also has a generating random variable, Y that characterizes its elements. The response random time series then corresponds to a sequence of generally *dependent* samples of the generating random variable, Y. The nature of this dependency is dictated by the behavior of the time-invariant linear operator's unit-impulse response, h_n. For the purposes of this article, it is useful to interpret the above linear operator's unit-impulse response h_n as being the unit-impulse response of the cascaded convolution and deconvolution operators as described earlier.

We shall now examine how the concept of cumulants may be employed to analyze the response of a time-invariant operation made on a white noise excitation. In order to make this analysis more accessible to a general audience, one of our primary objectives is that of presenting this development in a fashion that avoids the cumbersome notation often found in deconvolution articles. To begin this exposition, it is apparent from the results of the last section that when the excitation is composed of a sequence of samples of an independent random variable that the moment generating function of the response generating random variable is specified by

$$\phi_Y(\omega) = \prod_{k \in K} \phi_W(h_k \omega)$$

where $\phi_W(\omega)$ is the moment generating function of the excitation generating random variable, W. This response moment generating function is dependent on the linear operator's unit-impulse response elements, h_k. The associated cumulant generating function of the response generating random variable is obtained by taking the natural logarithm of this relationship for $\phi_Y(\omega)$, and results in:

$$\psi_Y(\omega) = \sum_{k \in K} \psi_W(h_k \omega) \tag{10}$$

where $\psi_W(\omega) = \ln \phi_W(\omega)$. Using this relationship, the following fundamental theorem is obtained.

Theorem 1 *Let $\{W_n\}$ be a white noise time series whose elements are independent identically distributed samples of a generating random variable, W. The response, $\{Y_n\}$, of a time-invariant linear operation with unit-impulse response $\{h_n\}$ made on this white noise time series is also stationary. This response time series is composed of generally dependent samples of an underlying response generating random variable, Y. Moreover, the cumulants of order p of the excitation and response generating random variables are related according to:*

$$c_Y(p) = c_W(p) \sum_k (h_k)^p \tag{11}$$

provided that $c_W(p)$ is finite.

This theorem is proven by employing relationship (10) and the cumulant definition (6). In particular, from the cumulant definition, it is seen that the response cumulant is formally given by $c_Y(p) = (-j)^p d^p \psi_Y(\omega)/d\omega^p |_{\omega=0}$. Applying the chain rule of differentiation to this response cumulant expression, we obtain relationship (11). The response cumulant of order p is seen to be the product of the excitation cumulant of order p with the sum of the linear operator's unit-impulse response elements raised to the pth power. It provides a particularly useful insight into the basic nature existent between the white noise excitation and response cumulants related by a time-invariant linear operation. This fundamental result was originally obtained by Bartlett [1], and later independently derived by Brillinger and Rosenblatt [4].

Upon examination of relationship (11), it is apparent that for any non-trivial time-invariant linear operation, the summation term appearing is always positive for even values of p. Thus, nonzero, even-order excitation cumulants may not be driven to zero by a time-invariant linear operation, and the sign of even-ordered response and excitation cumulants are always the same. On the other hand, the summation term for odd values of p need not be positive. It is conceivable that a linear operation can result in an odd order response cumulant(s) being zero or of the opposite sign of the associated excitation cumulants. For example, consider the first difference operator, with impulse response $h_n = \delta(n) - \delta(n-1)$, where $\delta(n)$ designates the Kronecker delta sequence. If a white noise time series is applied to this first difference operator, then according to Eq. (11), the corresponding response random time series has all its odd order cumulants equal to zero, while its even order cumulants are twice the value of those of the excitation.

The effect of a linear operation on a white noise time series in terms of the generating random variable is completely characterized by Eq. (11). In particular, in accordance with the Taylor series expansion (7), it follows that the response generating function has its associated cumulant generating function specified by

$$\psi_Y(\omega) = \sum_{k=1}^{n} \frac{1}{k!}\left[c_W(k) \sum_m (h_m)^k \right] (j\omega)^k + o_n(\omega) \tag{12}$$

where it has been assumed that the moments of the excitation's generating random variable, W, exist up to order n. If all the moments of W exist, then an exact representation of the response generating random variable is given by the summation term in this expression, with n replaced by infinity, provided that the infinite summation is finite.

It is found that in certain signal processing applications as exemplified by blind deconvolution, it is desired to select the unit-impulse response, h_n, so that the generating functions of the excitation and response time series are scalar multiples so that $Y = aW$, where a is an arbitrary nonzero number. In accordance with the second cumulant generating function property entry (5), this is equivalent to having the cumulant generating functions of random variables W and Y satisfy $\psi_Y(\omega) = \psi_W(a\omega)$. Using relationship (12), we conclude that the requirement $Y = aW$ is met if and only if

$$Y = aW \Leftrightarrow \sum_m (h_m)^k = (a)^k \quad \textit{for all k such that } c_W(k) \neq 0$$

A sufficient condition for satisfying this requirement is that all but one of the elements of the unit-impulse response $\{h_n\}$ be zero.

Normalized Cumulants

In the blind deconvolution approach under development, a desirable feature is that it be invariant to signal size. Unfortunately, the concept of cumulant as embodied in Eq. 6 is inherently dependent on the size of the generating random variable. In particular, it is readily shown that if a is an arbitrary scalar, then the cumulants of the random variables W and aW are related as

$$Y = aW \Leftrightarrow c_Y(k) = a^k c_W(k)$$

In order to make the concept of cumulant invariant to scalar multiplications, it is necessary to provide a normalization of the cumulant. With this in mind, let us consider the *normalized cumulant* of order (p, q) associated with random variable W as defined by

$$k_W(p,q) = \frac{c_W(p)}{|c_W(q)|^{p/q}} \tag{13}$$

in which it is tacitly assumed that cumulant $c_W(q)$ is nonzero. Typically, the order integer parameters are selected so that $p > q$, although this definition is applicable for any choice of these parameters. Using the definition of cumulant, it directly follows that this normalized cumulant is scalar invariant in the sense that $k_{aW}(p,q) = k_W(p,q)$ for any nonzero scalar, a. In many applications, the specific selection of $q = 2$ provides a logical choice, since $c_W(2) = \sigma_w^2$ is always nonzero for any non-trivial random variable. This particular selection yields the normalized cumulant relationship

$$k_W(p,2) = \frac{c_W(p)}{\sigma_W^p}$$

For development purposes, however, it is useful for us to employ the more general definition of normalized cumulant (13), while being aware that the specific choice $q = 2$ is often used. A number of authors have employed Eq. (13) for specific choices of the parameters p and q to solve the blind deconvolution problem. For instance, Tugnait [33] has employed it in the noise free cases $p = 3$, $q = 2$ and $p = 4$, $q = 2$, and in the additive Gaussian noise case $p = 6$, $q = 4$, while Shalvi and Weinstein [30] addressed the case $q = 2$. Other researchers have employed modified versions of this relationship for specific choices of p and q [8, 34].

Linear System: Response Normalized Cumulants

We now apply the above normalized cumulant definition to the results of Theorem 1 concerning the effects of a time-invariant operation made on a stationary white noise time series. In particular, the normalized cumulants of the excitation and response generating random variables are seen to be related by employing (13) and the response cumulant identity (11) to give:

$$k_Y(p,q) = \frac{\Sigma_k (h_k)^p}{|\Sigma_k (h_k)^q|^{p/q}} k_W(p,q) \tag{14}$$

As was the case for cumulants, this relationship indicates that the signs of the excitation and response normalized cumulants are the same for p an even integer. Furthermore, this normalized response cumulant measure also satisfies the desired scalar multiplication invariance property in the sense that the magnitudes of the normalized response cumulants for two systems with unit-impulse responses, $\{h_n\}$ and $\{ah_n\}$, are the same for any nonzero scalar, a. With this relationship serving as foundation, the following generalization of the Shalvi and Weinstein result, which treated the special case $q = 2$, is given.

Theorem 2 *Let a time-invariant linear system with unit-impulse response $\{h_n\}$ be excited by a white noise time series $\{W_n\}$ composed of independent identically distributed samples of an excitation generating random variable W. The corresponding random response time series, $\{Y_n\}$, is also stationary and is composed of dependent samples of a response generating random variable, Y. Furthermore, the magnitude of the normalized response cumulant $k_Y(p,q)$ for any positive even integer, q, is bounded above by*

$$|k_Y(p,q)| \leq |k_W(p,q)| \quad \textit{for all even and odd } p > q \tag{15}$$

and for p any even positive integer is bounded below by

$$|k_Y(p,q))| \geq |k_W(p,q)| \quad \textit{for all even and odd } q > p \tag{16}$$

Furthermore, for any normalized excitation cumulant $k_W(p,q)$ that is nonzero in either of these bound relationships, the inequality becomes an equality if and only all but one element of the unit-impulse response $\{h_n\}$ are zero.

The first bound of this theorem is proven by noting that for q, a positive even integer that the denominator term in (14) is equal to $[\|\underline{h}\|_q]^p$, where $\|\underline{h}\|_q = [\Sigma_k |h_k|^q]^{1/q}$ designates the standard ℓ_q norm. We therefore have $|k_Y(p,q)|=|k_W(p,q)|\,|\Sigma_k (h_k)^p|/(\|\underline{h}\|_q)^p \leq |k_W(p,q)|(\|\underline{h}\|_p / \|\underline{h}\|_q)^p$, and upon using one of several classical bounding identities, the required bound follows, with equality holding of and only if only one of the components of $\underline{h}$ is nonzero. A similar approach establishes the validity of the second bound for p, assuming it is a positive even integer.

Of great use in many signal processing applications, the above theorem indicates that any time-invariant linear operation made on a white noise time series results in a stationary random time series whose normalized cumulants all have magnitudes that are less than or equal to the magnitude of the excitation's normalized cumulant for all even $q < p$ (or all even $p < q$). This theorem, in conjunction with the necessary and sufficient condition established in (3) for perfect deconvolution, suggests a potentially important mechanism for solving the blind deconvolution problem. In particular, the required deconvolving operator must generate a response whose normalized cumulants have magnitudes that are the largest over the class of all linear operators for all even $q < p$ and smallest magnitudes for all even $p < q$. For purposes of brevity, we shall exclusively consider the case in which $p > q$ with q even, whereby it is desired to select h so as to maximize an estimate of $|k_Y(p,q)|$. A similar development for the case in which $p < q$, with p even, can readily be made in which maximums are replaced by minimums.

In certain applications, the signal being recovered is composed of a sequence of independent samples of an underlying generating random variable, which, in turn, is equal to a sum of independent random variables. With this possibility in mind, the following lemma provides the required characterization of the generating random variable associated with the signal being recovered.

Lemma 3 *Let the random time series $\{W_n\}$ be composed of a sequence of independent samples of a generating random variable, W, which is itself composed of the sum $W = V_1 + V_2 + \ldots + V_M$, in which the V_k random variables are independent. The normalized cumulant of order (p,q) of the generating random W is then specified by*

$$k_W(p,q) = \sum_{m=1}^{M} k_{V_m}(p,q)$$

where $k_{V_m}(p,q)$ designates the normalized cumulant of order (p,q) associated with random variable V_m.

This lemma is proven by using the previously established cumulant properties associated with the sum of independent random variables. It is therefore apparent that the principals embodied in Theorem 2 are applicable to the case in which the signal being deconvolved is equal to the sum of independent random signals. As an example, let the random variable $W = S + \mathcal{N}$, where S and $\mathcal{N}$ are independent generating random variables corresponding to the signal, and noise components, of a random time series, respectively.

Cumulant Based Deconvolution Procedure: CASE $p > q$, with q positive even

To begin the proposed procedure, we first select any integer value of p greater than q for which the signal being recovered has a nonzero cumulant $c_W(p)$. The required ideal deconvolution operation is then achieved in principal by maximizing the magnitude of the normalized response cumulant $k_Y(p,q;\underline{h})$, where q is any positive even integer less than p for which $c_W(q)$ is nonzero (e.g., $q = 2$). This maximization is to be made with respect to the unit-impulse response $\{h_n\}$ of the combined convolution-deconvolution operation. Since the unit-impulse response of the unknown linear convolution operator $\{f(n)\}$ is implicitly contained within the observed data $\{x(n)\}$, this maximization must be made with respect to the deconvolution operator's unit-impulse response $\{g_n\}$. The required maximization therefore takes the form

$$\max_{g}|k_Y(p,q)| = \max_{g}[k_Y(p,q)\operatorname{sgn}[k_Y(p,q)]] \tag{17}$$

where $\mathbf{g}$ is an appropriately dimensioned vector whose components are the elements of the unit-impulse response of the deconvolving operator. Using the fact that the cumulants of the deconvolved signal and the signal being recovered must be equal, it follows that the term $\operatorname{sgn}[k_Y(p,q)]$ may be replaced by $\operatorname{sgn}[k_W(p,q)]$ in this maximization problem. Clearly, the required deconvolution is achieved by either maximizing or minimizing the normalized response cumulant $k_Y(p,q)$, depending on whether the signal being recovered has a normalized cumulant $k_W(p,q)$ that is positive or negative, respectively. In either case, it is desired to find a global maximum of $|k_Y(p,q)|$.

Nonrecursive Deconvolution Operator

In any practical application, explicit knowledge of the normalized response cumulant $k_Y(p,q)$ is not available. This information must be estimated from the measured data being convolved and the deconvolution operator's unit-impulse response. In conformity with real world applications, this convolved data is of finite length hereby designated by the positive integer N, and without loss of generality the data's indexing is taken to begin at $n = 1$, that is

$$x(1), x(2), \ldots, x(N) \tag{18}$$

The first step in the deconvolution process is that of removing the sample mean and any obvious long term linear or higher order trends that may be present in the data so as to render a data set that has a stationary behavior. This process involves finding the best long term trend fit to the data and then subtracting that trend from the data. It is assumed that this preliminary detrending operation has been applied and results in the above data set.

It has been previously established that the deconvolution operator associated with a time-invariant linear convolution operation must itself by linear and time-invariant. To simplify the presentation of the principals of deconvolution, we first consider linear deconvolution operators that are nonrecursive (no feedback terms) in nature. In a later section, these concepts are extended to the case of linear recursive deconvolution operators. A linear nonrecursive operator of length $Q + 1$ is governed by an excitation-response relationship of form

$$y(n) = \sum_{k=0}^{Q} b_k x(n-k) \tag{19}$$

This response is seen to be dependent on the linear operator's $(Q \times 1)$ unit-impulse response vector as specified by

$$b = \begin{bmatrix} b_o \\ b_1 \\ \vdots \\ b_Q \end{bmatrix}$$

To explicitly recognize the dependency, of the response on the b_k coefficients, the response elements could have been denoted as $y(n;\mathbf{b})$. However, this unnecessarily complicates the resultant notation and is therefore not employed. Any time-invariant linear operator can be suitably approximated by a nonrecursive deconvolution operator of sufficient length.

The response of the nonrecursive deconvolution operator (19) to the given data (18) is next computed. Although the response elements are generally nonzero over the time interval $1 \le n \le N+Q$, only the response elements on the interval $Q+1 \le n \le N$ can be considered as steady state in behavior. The first and last Q response elements are transient in nature, since they entail excitation elements outside the given data range. With this in mind, the specific response terms $y(Q+1)$, $y(Q+2)$,... ,$y(N)$ are stationary in behavior and are used to form estimates of the deconvolution signal's central moments. These central moment estimates are in turn used to form estimates of the deconvolved signals normalized cumulants. The following theorem provides a characterization of these central moment estimates.

Lemma 4 *Let the response elements y(n) be generated according to the nonrecursive relationship (19) and let the standard central moment estimate of order k of this response sequence be obtained according to*

$$\hat{\mu}_Y(k) = \frac{1}{N-Q} \sum_{n=Q+1}^{N} [y(n) - \hat{m}_Y]^k \tag{20}$$

in which the index, k, is a positive integer and $\hat{m}_Y$ *designates the sampled mean of the deconvolution operator's response elements y(n) as given by*

$$\hat{m}_Y = \frac{1}{N-Q} \sum_{n=Q+1}^{N} y(n) \tag{21}$$

The components of the $(Q+1) \times 1$ *gradient vector* $\nabla_{\mathbf{b}}[\hat{\mu}_Y(k)]$ *of the central element* $\hat{\mu}_Y(k)$ *with respect to the deconvolving operator's* b_m *coefficients are given by*

$$\frac{\partial \hat{\mu}_Y(k)}{\partial b_m} = \frac{k}{N-Q} \sum_{n=Q+1}^{N} [y(n) - \hat{m}_Y]^{k-1} \times \left[x(n-m) - \frac{1}{N-Q} \sum_{k=Q+1}^{N} x(k-m) \right] \tag{22}$$

for m = 0, 1,...,Q. Furthermore, if the convolved data (18) has zero sampled mean and if N – Q is sufficiently large, then these gradient vector components may be accurately approximated by the simpler expression

$$\frac{\partial \hat{\mu}_Y(k)}{\partial b_m} \approx \frac{k}{N-Q} \sum_{n=Q+1}^{N} [y(n) - \hat{m}_Y]^{k-1} x(n-m) \tag{23}$$

for m = 0, 1 ,...,Q.

Formulation of Normalized Cumulant Estimates

Central moment estimates (20) may be used in conjunction with appropriate entrees from the list (8) to form estimates of the response cumulants $c_Y(p)$ and $c_Y(q)$, while equation (13) is employed to obtain an estimate of the response's normalized cumulant $k_Y(p,q)$. In the blind deconvolution algorithm to be developed, the following gradient of a central moment ratio is a direct consequence of the above lemma and plays a central role is subsequent developments:

$$\nabla_{\mathrm{b}} \left[\frac{\hat{\mu}_Y(k_1)^{m_1}}{\hat{\mu}_Y(k_2)^{m_2}} \right] = \frac{m_1 \hat{\mu}_Y(k_1)^{m_1-1}}{\hat{\mu}_Y(k_2)^{m_2}} \nabla_{\mathrm{b}}[\hat{\mu}_Y(k_1)] - \frac{m_2 \hat{\mu}_Y(k_1)^{m_1}}{\hat{\mu}_Y(k_2)^{m_2+1}} \nabla_g[\hat{\mu}_Y(k_2)] \tag{24}$$

In this expression, $\nabla_{\mathrm{b}}[\hat{\mu}_Y(k_Y)]$ designates the $(Q+1) \times 1$ gradient vector with components (see (22)), while k_1, k_2, m_1, and m_2 are arbitrary positive integers. Further, when implementing the deconvolution algorithms to be shortly described for large data lengths, we must compute these gradient element expressions in a computationally fast and memory efficient manner. Otherwise, the process of deconvolution can result in excessively long computer run times as well as unreasonable memory requirements. Depending on the software package being used (e.g., FORTRAN, MATLAB, MAPLE), these computations may be efficiently implemented using correlation or filtering operations.

Deconvolution via Kurtosis Extremization

To illustrate a general deconvolution procedure based on the concept of maximizing normalized cumulants, we first consider the important special case in which $p = 4$ and $q = 2$. The steps in this special case may then be appropriately modified for other choices for p and q, as shown in the next section. The normalized response cumulant of order (4,2) is called the *kurtosis* and is formally given by

$$k_Y(4,2) = \frac{c_Y(4)}{c_Y(2)^2} = \frac{\mu_Y(4)}{\mu_Y(2)^2} - 3 \tag{25}$$

where appropriate entries from the list (8) have been made. This kurtosis measure can be shown to exclusively take on values in the interval [-2, ∞], with the lower bound being meet by a random binary process whose elements exclusively take on the values $\pm M$, while the upper bound is approached by a impulse type random process. When invoking this kurtosis measure, we implicitly assume that the excitation kurtosis $k_W(4,2)$ is nonzero. For example, an excitation which is governed by a Gaussian generating random variable has a zero kurtosis and is thereby excluded in what is to follow.

A logical choice for an estimate of the kurtosis measure (25) is specified by:

$$\hat{k}_Y(4,2) = \frac{\frac{1}{N-Q}\sum_{n=Q+1}^{N}[y(n)-\hat{m}_Y]^4}{\left(\frac{1}{N-Q}\sum_{n=Q+1}^{N}[y(n)-\hat{m}_Y]^2\right)^2} - 3 \tag{26}$$

where expression (20) for $k = 2,4$ has been used to estimate the required central moments. This response kurtosis estimate is seen to be implicitly dependent on the parameter vector **b** through the deconvolution operator's response $\{y(n)\}$. It is found that this kurtosis estimate is bounded as $-2 \le \hat{k}_Y(4,2) \le N-Q-2$ with the upper bound being met if and only if $y(n)$ has one non-zero component, and the lower bound achieved if and only if all components of $y(n)$ are non-zero and have the same amplitude (i.e., are binary in nature). The fact that the lower bound is achieved only by binary type signals is to be noted in regards to the deconvolution of binary random sequences that have been distorted by a linear operation.

Since the ideal deconvolved signal is to be a shifted-scaled multiple of the signal $\{w_n\}$ being recovered, it follows that their respective kurtosis should be equal. With this in mind, and properly interpreting the above bounding interpretation of $\hat{k}_Y(4,2)$, we make the following comments on the fundamental time behavior of a signal and its associated kurtosis: (i) when the signal being recovered has a magnitude behavior $|w_n|$ vs n that is relatively constant, then its kurtosis tends to be negative and the kurtosis estimate (26) approaches the lower bound of minus two; (ii) when the signal to be recovered has a highly sporadic time behavior (i.e., widely spaced impulses) then its kurtosis tends to be positive and kurtosis estimate approaches the upper bound of N–Q–2.

As suggested in the previous section, a nonrecursive deconvolution operation is realized by selecting its nonrecursive unit-impulse response vector **b** so as to maximize the magnitude of kurtosis estimate as specified by

$$|\hat{k}_Y(4,2)| = \hat{k}_Y(4,2)\,\mathrm{sgn}[\hat{k}_Y(4,2)] \tag{27}$$

It is seen from (14) that since $p = 4$ and $q = 2$, the sign term as designated by $\mathrm{sgn}[k_Y(4,2)]$ may be replaced by its equivalent $\mathrm{sgn}[k_W(4,2)]$ in this magnitude expression. The decision as to whether $\mathrm{sgn}[k_W(4,2)]$ is set to plus or minus one can be based on the nature of the signal being recovered. In those cases where such information is not available, this sign function can be set equal to the sign of a kurtosis estimate of the observed data (18).

Nonlinear Programming Solution

It is now desired to select the unit-impulse response vector, **b**, of the deconvolving operator so as to maximize the kurtosis magnitude estimate (26). Due to the highly nonlinear manner in which this unit-impulse response vector enters this functional, however, it is necessary to employ nonlinear programming techniques to numerically compute the required maximum. The user is forewarned that a properly functioning nonlinear programming algorithm only ensures convergence to a relative maximum of the functional. Among the most popular of nonlinear programming algorithms for finding a functional's (relative) maximum are the class of *ascent* algorithms. In an ascent algorithm, the approximation of the desired solution at iteration m, as designated by $\mathbf{b}_m$ is perturbed in the additive fashion

$$\mathbf{b}_{m+1} = \mathbf{b}_m + \alpha\delta \tag{28}$$

where the δ corresponds to a *perturbation vector* and α is a positive *step-size* scalar. This perturbed vector is said to satisfy the *improvement condition* if the kurtosis magnitude estimate is increased, that is:

$$|\hat{k}_Y(4,2)^{(m+1)}| > |\hat{k}_Y(4,2)^{(m)}| \tag{29}$$

where $\hat{k}_Y(4,2)^{(m)}$ designates the value of the normalized cumulant estimate evaluated with nonrecursive parameter vector, $\mathbf{b}_m$.

The various ascent algorithms are distinguished by the manner in which the perturbation vector is chosen. Many of the more popular ascent algorithms explicitly employ the gradient vector of the functional being maximized. This gradient vector points in the direction in which the functional most rapidly increases. Thus, the gradient vector associated with the response kurtosis magnitude estimate is of interest and obtained by using standard differentiation. In particular, it follows from Eq. 24 with $k_1 = 4$, $m_1 = 1$, $k_2 = 2$, and $m_2 = 2$, that the components of the gradient of this kurtosis estimate relative to the b_m coefficients are specified by:

$$\begin{aligned}\frac{\partial\hat{k}_Y(4,2)}{\partial b_m} &= \frac{4}{(N-Q)\hat{\mu}_Y(2)^2}\sum_{n=Q+1}^{N}[y(n)-\hat{m}_Y]^3 \times \\ &\left[x(n-m) - \frac{1}{N-Q}\sum_{k=Q+1}^{N}x(k-m)\right] - \\ &\frac{4\hat{\mu}_Y(4)}{(N-Q)\hat{\mu}_Y(2)^3}\sum_{n=Q+1}^{N}[y(n)-\hat{m}_Y] \times \\ &\left[x(n-m) - \frac{1}{N-Q}\sum_{k=Q+1}^{N}x(k-m)\right]\end{aligned} \tag{30}$$

for $m = 0, 1, \ldots, Q$.

Conceptually, the positive step-size scalar, α, in (28) is chosen so that the magnitude of the kurtosis takes on its maximum value along the perturbation vector direction δ. The *method of steepest ascent* corresponds to this optimum choice for the step size scalar when the perturbation vector is set equal to the gradient vector (30). Under conditions which are met in the maximization of the kurtosis magnitude estimate, this steepest ascent method produces a unit-impulse response vector sequence which converges to a relative maximum of the kurtosis estimate's magnitude. Unfortunately, the one-dimensional search required in determining the optimal step-size scalar is typically very demanding in a computational sense. This computational burden can be significantly eased by sequentially evaluating the kurtosis magnitude estimate at the decreasing step-size selections $\alpha = (\rho)^k$ for $k = 0,1,2,\ldots$ until improvement condition (29) is first met where ρ is a positive scalar selected in the interval $0 < \rho < 1$. An improvement is always ensured, provided that the point $\mathbf{b}_m$ is not already a relative maximum. The unconstrained gradient algorithm therefore takes the form:

$$\mathbf{b}_{m+1} = \mathbf{b}_m + (\rho)^k \nabla_{\mathbf{b}}[\hat{k}_Y(4,2)^{(m)}] \ \ \mathrm{sgn}[k_Y(4,2)^{(m)}] \tag{31}$$

where Eq. (30) is used to compute the required gradient vector, and k is sequenced through the values $0,1,2,\ldots$ until (29) is first satisfied. The unit-impulse response vector so obtained then serves as an improved guess of the desired optimum. This process is continued until either the gradient vector (30) is adequately close to zero (a necessary condition for a local maximum) or the kurtosis estimate magnitude does not change sufficiently from one iteration to the next.

This gradient based method leads to a linear convergent behavior that can be too slow for certain applications. Nonetheless, we shall employ this standard ascent algorithmic approach to illustrate the use of normalized cumulant maximization for solving the blind deconvolution problem. One can readily adapt these concepts to other nonlinear programming methods which offer the possibility of faster convergence generally at a higher computation cost per iteration. These alternate algorithms employ an other than gradient selection for the perturbation vector.

General Deconvolution Algorithm

A general method for achieving deconvolution which is based on the notion of maximizing a normalized response cumulant magnitude of order (p,q) is described in Theorem 2, in which it is tacitly assumed that the signal being recovered has nonzero cumulants of orders p and q. We applied this approach in the last section to the specific kurtosis case $p = 4$ and $q = 2$. In this section, these kurtosis based concepts are extended to more general selections of the normalized cumulant order (p,q). In the more general application of this theorem, the required ideal deconvolution operation is achieved by maximizing the normalized response cumulant magnitude estimate as symbolically specified by

$$\left|\hat{k}_Y(p,q)\right| = \hat{k}_Y(p,q)\,\mathrm{sgn}\left[\hat{k}_Y(p,q)\right] \tag{32}$$

where $\hat{k}_Y(p,q)$ designates an estimate of the normalized response cumulant of order (p,q). This estimate may be formed by using the appropriate entries from the list (8) and the central moment estimates (20).

For illustrative purposes, we shall employ the gradient based algorithm approach for numerically finding a (local) maximum of the normalized response cumulant magnitude estimate (32). The gradient based update algorithm takes the form

$$\mathbf{b}_{m=1} = \mathbf{b}_m + (\rho)^k \left(\nabla_{\mathbf{b}}\left[\hat{k}_Y(p,q)^{(m)}\right]\right) \mathrm{sgn}\left[\hat{k}_Y(p,q)^{(m)}\right] \tag{33}$$

where the parameter k is sequenced through the values $0,1,2,\ldots$ until an increase in the normalized response cumulant magnitude estimate over its previous value is first obtained. This algorithm is iterated until convergence is deemed to have occurred as suggested by: (i) the gradient vector being sufficiently close to zero, or, (ii) the relative change in normalized cumulant magnitude from one iteration to the next being sufficiently small.

Example 1 To illustrate the approach outlined here, let us consider the deconvolution operation based on maximizing the magnitude of an estimate of the normalized response cumulant of order (6,2). In accordance with the normalized cumulant definition (13) and the second and sixth entries of the list (8), it is seen that a logical estimate for kY(6,2) is given by

$$\hat{k}_Y(6,2) = \frac{\hat{\mu}_Y(6) - 15\hat{\mu}_Y(4)\hat{\mu}_Y(2) - 10\hat{\mu}_Y(3)^2 + 30\hat{\mu}_Y(2)^3}{\hat{\mu}_Y(2)^3} \tag{34}$$

where the central moment estimates $\hat{\mu}_Y(k)$ entries are obtained using (20). The gradient of this estimate is readily found by employing (24) to yield:

$$\nabla_{\mathbf{b}}[\hat{k}_Y(6,2)] = \frac{1}{\hat{\mu}_Y(2)^3}\nabla_{\mathbf{b}}[\hat{\mu}_Y(6)] - \frac{15}{\hat{\mu}_Y(2)^2}\nabla_{\mathbf{b}}[\hat{\mu}_Y(4)] - \frac{20\hat{\mu}_Y(3)}{\hat{\mu}_Y(2)^3}\nabla_{\mathbf{b}}[\hat{\mu}_Y(3)] + \left(\frac{30\hat{\mu}_Y(4)\hat{\mu}_Y(2) - 3\hat{\mu}_Y(6) + 30\hat{\mu}_Y(3)^2}{\hat{\mu}_Y(2)^4}\right)\nabla_b[\hat{\mu}_Y(2)] \tag{35}$$

This gradient expression is then substituted into (33) to form the basic step for the gradient based algorithm. The elements of this gradient vector are computed using identities in relationships 21-23 and gives rise to:

$$\frac{\partial \hat{k}_Y(6,2)}{\partial b_m} = \frac{6}{(N-Q)\hat{\mu}_Y(2)^3} \times \sum_{n=1}^{N-Q}\left\{\left[y(n) - \hat{m}_Y\right]^5 + \alpha_3\left[y(n) - \hat{m}_Y\right]^3 + \alpha_2\left[y(n) - \hat{m}_Y\right]^2 + \alpha_1\left[y(n) - \hat{m}_Y\right]\right\} x(n+Q-m) \tag{36}$$

for $m = 0,1,\ldots,Q$, where the α_k coefficients here appearing are given by

$$\alpha_1 = \frac{10\hat{\mu}_Y(4)\hat{\mu}_Y(2) - \hat{\mu}_Y(6) + 10\hat{\mu}_Y(3)^2}{\hat{\mu}_Y(2)},$$

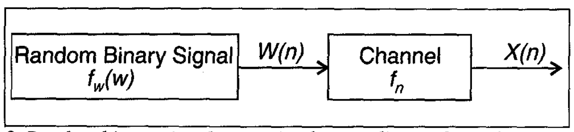
$$\alpha_2 = -10(3), \quad \alpha_3 = -10\hat{\mu}_Y(2)$$

Equalizing Convolved Random Binary Signals

Digital signals play an increasingly important role in contemporary communications. For the purposes of this section, the simplest digital signal is defined to be a number sequence whose elements exclusively take on the two values zero and M (e.g., $M = 1$). Such a signal is then a realization of a stationary random binary time series in which the underlying generating random variable has probability density function

$$f_W(w) = p\delta(w - M) + (1 - p)\delta(w) \qquad (37)$$

in which p denotes the probability that a given element of the time series is equal to M while $1 - p$ is the probability that the element is equal to zero. It is postulated that the elements of the random binary time series are independent. Upon a realization of the random binary time series, the resultant deterministic binary signal is transmitted over a communication channel which is assumed to be a governed by a time-invariant linear operator. The signal as received is then a convolved mapping of the transmitted binary signal and can be distinctly non-binary in appearance. In order to recover the original transmitted binary signal, it is necessary to perform a deconvolution of the received signal using techniques such as described in the previous sections. A depiction of this random binary communication scenario is given in Fig. 3.



3. Random binary signal transmitted over a linear channel.

To begin the analysis of the transmitted random binary signal, the moments of its associated generating random variable are first determined. It directly follows from the generating function's probability density function (37) that these moments are specified by

$$E\{W^n\} = pM^n \text{ for } n = 1,2,3,...$$

Moreover, using the binomial expansion relating moments and central moments given earlier, the associated central moments are found to be

$$\mu_W(n) = p(1 - p)[(1 - p)^{n-1} - (-p)^{n-1}]M^n \quad \text{for } n = 2,3,4,...$$

These central moment expressions are then inserted into (8) to obtain the corresponding random binary signals cumulants and normalized cumulants. A listing of these results are displayed in Table 1.

In applications requiring the deconvolution of random binary signals that have been distorted by a linear convolution operation, use can be made of these closed form expressions for the normalized cumulants. For example, if the kurtosis approach developed earlier is employed in the deconvolving operation, it is first noted that

$$k_W(4,2) = \begin{cases} < 0 & \textit{for } \frac{1}{2}(1 - 1/\sqrt{3}) < p < \frac{1}{2}(1 + 1/\sqrt{3}) \\ \geq 0 & \textit{otherwise} \end{cases}$$

Thus, the entity $\text{sgn}[k_Y(4,2)]$ appearing in blind deconvolution algorithm (31) is set to minus one if the probability parameter p is contained in the interval $(\frac{1}{2}(1 - 1/\sqrt{3}), \frac{1}{2}(1 + 1/\sqrt{3}))$, and is set to plus one otherwise.

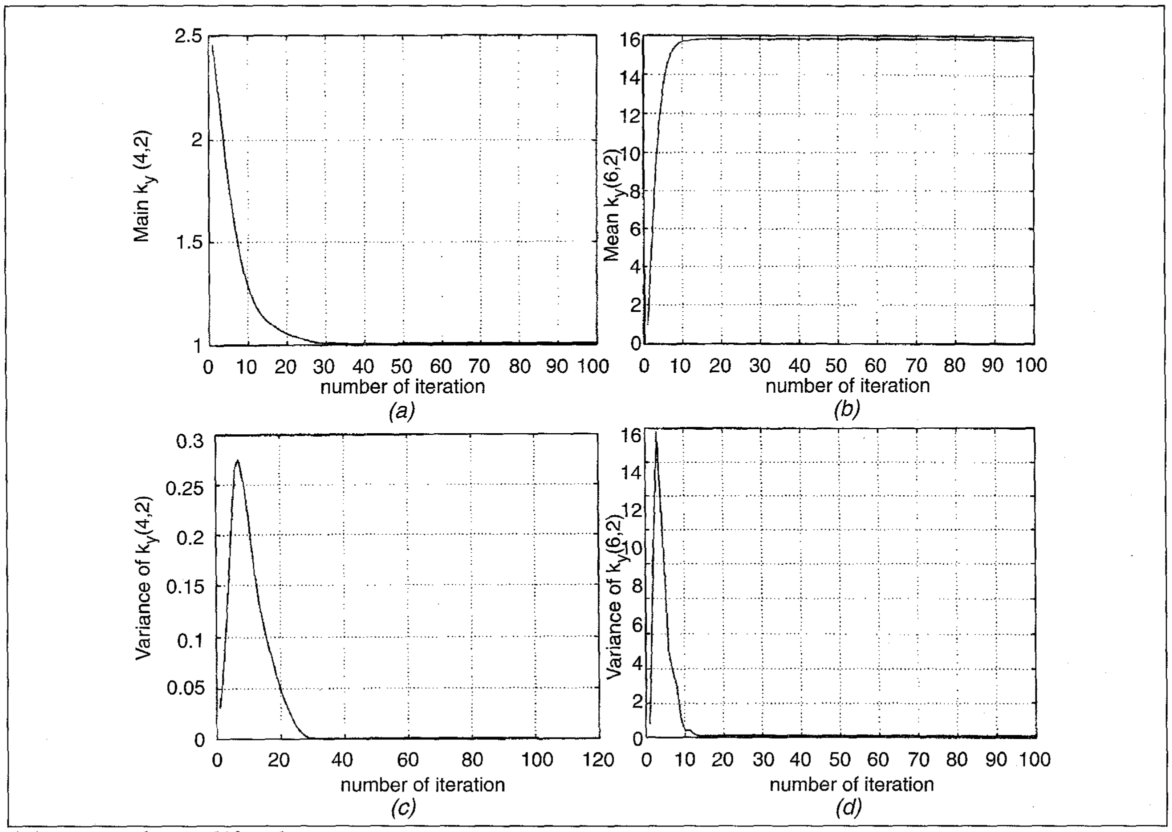
Example 2 In this example a demonstration of the effectiveness of the proposed deconvolution algorithms for random binary signals is given. A random binary sequence with $p = 0.5$ and $M = 1$ is first generated and then applied to an all-pass system with transfer function:

$$F(z) = \frac{-0.4 + z^{-1}}{1 - 0.4z^{-1}} \qquad (38)$$

which is seen to have an unstable causal inverse due to the zero at $z = 2.5$. After transient effects have decayed away, the resultant steady state response elements are recorded to yield an observed deconvolved data set of length $N = 500$. To begin the deconvolution operation, the sampled mean of this data is computed and then subtracted from each data element yielding a zero sample mean data set. In order to judge the

Table 1. Cumulants and Normalized Cumulants of a Random Binary Variable

n	$c_W(n)$	$k_W(n;2)$
2	$p(1\text{-}p)M^2$	1
3	$p(1\text{-}p)(1\text{-}2p)M^3$	$\frac{1-2p}{\sqrt{p(1-p)}}$
4	$p(1\text{-}p)(1\text{-}3p+3p^2)M^4$	$\frac{1-6p+6p^2}{p(1-p)}$
5	$p(1\text{-}p)(1\text{-}4p+6p^2\text{-}4p^3)M^5$	$\frac{1-14p+36p^2-24p^3}{[p(1-p)]^{3/2}}$
6	$p(1\text{-}p)(1\text{-}5p+10p^2\text{-}10p^3+5p^4)M^6$	$\frac{1-30p+150p^2-240p^3+120p^4}{p^2(1-p)^2}$

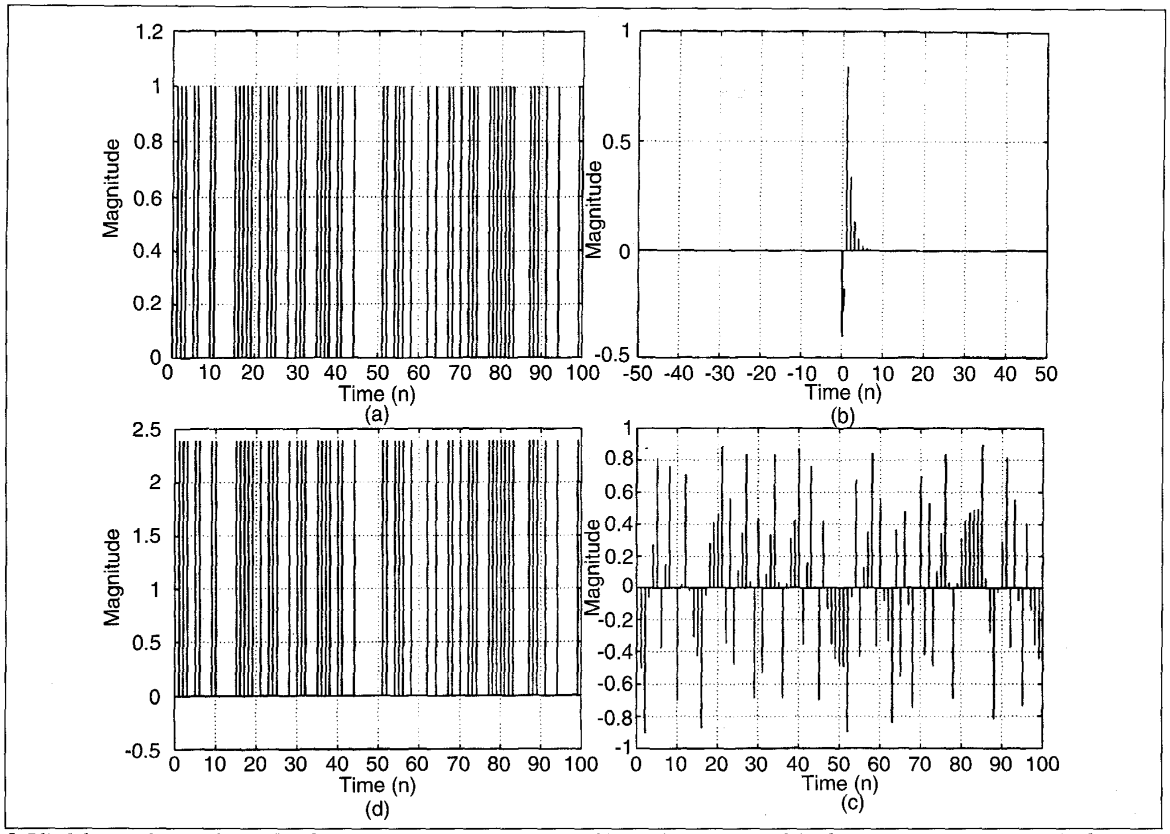


4. Average results over 100 trials.

relative effectiveness of extremizing different order normalized cumulant estimates for achieving blind deconvolution, the unconstrained gradient algorithm is then used to maximize both the magnitudes of $\hat{k}_W$ (4,2;**b**) and $\hat{k}_W$ (6,2;**b**), respectively. It is first noted that for $p = 0.5$, Table 1 indicates that k_W (4,2;**b**) = –2, and, k_W(6,2;**b**) = 16. If it is assumed that the data processor has *a priori* knowledge that a random binary sequence is to be recovered, the term $\mathrm{sgn}[\hat{k}_Y(p,2;\mathbf{b})]$ appearing in unconstrained gradient algorithm (33) may be replaced by minus one for cumulant order $p = 2$ and plus one for $p = 6$. This a priori knowledge is assumed in this example although the deconvolution algorithm works in a comparable fashion without this substitution.

To begin this numerical demonstration, let us consider the problem of maximizing $[\hat{k}_Y(4,2)]$ in which unconstrained gradient algorithm (31) is used with ρ = 0.5. The deconvolution operator is taken to have length parameter $Q = 9$. To initiate the algorithm for maximizing $|\hat{k}_Y(4,2)|$, the components of the initial unit-impulse response vector $\mathbf{b}_0$ are obtained by making Q+1 independent samples of a zero mean-unit variance Gaussian random variable. Starting at this random initial guess, fifty iterations of the unconstrained gradient algorithm are made. The algorithm was declared to have converged to the correct solution if $\hat{k}_Y(4,2)$ lies in the interval [–1.9,–2.1] (its theoretical value is –2). In order to gain a statistical insight into the behavior of the proposed deconvolution algorithm, one hundred independent trials of this experiment were taken in which both the convolved binary signal and initial random $\mathbf{b}_0$ are generated in a random manner for each trial run. It was found that 84 out of the 100 trials resulted in the convergence criterion being satisfied. In the remaining 16 trial runs, a new random $\mathbf{b}_0$ was used a produced a convergent result 13 out of the 16 trial runs. This process was continued until convergence was satisfied in all one hundred trial runs.

A plot of the sampled mean and sampled variance of the convergent values of the one hundred estimates $\hat{k}_Y(4,2)$ is shown in Fig. 4a and 4c. The sample mean of the one hundred convergent values of $\hat{k}_Y(4,2)$ was found to be –1.9921, with near zero sample variance. This mean value closely corresponds to the theoretical value –2 with convergence occurring in approximately thirty iterations while the variance goes to zero with increasing iterations. A segment of a typical random binary sequence, the unit-impulse response of the corrupting convolving system (38), the convolved data, and, the resultant deconvolved data are displayed in Fig. 5a-d. The data shift in the deconvolved signal has been removed by correlation techniques. It is apparent upon comparing Figs. 5a and 5d that a near perfect deconvolution has been attained.



5. Blind deconvolution of a random binary sequence: (a) excitation; (b) impulse response of the distortion system; (c) convolved response; (d) recovered sequence.

Unconstrained gradient based deconvolution algorithm (33) with $p = 6$, $q = 2$ was next used in maximizing $|\hat{k}_Y(6,2)|$ along the same lines described in the two preceding paragraphs for the $\hat{k}_Y(4,2)$ kurtosis approach. The same data was used in each of the one hundred trial runs and the results of these trial runs are displayed in Figs. 4b and 4d. From these plots it is seen that convergence typically occurred in about twelve iterations which was significantly faster than that obtained in maximizing $|\hat{k}_Y(4,2)|$. A similar conclusion has been observed when testing other simulated and real world data, thereby suggesting that the maximizing of $|\hat{k}_Y(6,2)|$ should be considered in any deconvolution task.

Linear Recursive Deconvolution Operator

Although the linear nonrecursive deconvolution operator examined in the previous sections performs in a satisfactory fashion in many applications, it is an inappropriate model when the convolving operator's transfer function as denoted by $F(z)$ is rational and has zeros that lie close to the unit-circle. In this case, the desired deconvolving filter has a rational transfer function $G(z)= F(z)^{-1}$, which has poles located close to the unit-circle, thereby giving rise to a unit-impulse response that is excessively long. Although a nonrecursive filter could be used to approximate this required deconvolution operation, its order q would necessarily have to be made large for an adequately accurate approximation. This, in turn, leads to an excessively large heavy computational load. Further, in the likely case where $F(z)$ is a nonminimum phase, the inverse transfer function, $G(z)$, will have poles located both inside and outside the unit-circle. To obtain a stable deconvolution operation, it is then necessary to implement the inverse filter in a mixed causal-anticausal fashion. In this section, we address the ramifications of employing a recursive deconvolving operator.

As in the nonrecursive case, the data being deconvolved is represented by the finite length set

$$x(1), x(2), \ldots, x(N) \tag{39}$$

while the deconvolution operation is implemented by the stable linear autoregressive-moving average (ARMA) deconvolution operator of order (P,Q), as governed by

$$\sum_{k=0}^{P} a_k y(n-k) = \sum_{k=0}^{Q} b_k x(n-k) \tag{40}$$

As indicated above, the transfer function of this ARMA operator as designated by $G(z) = B(z)/A(z)$ will have poles located both inside and outside the unit-circle. A stable deconvolution operation is conceptually obtained by decomposing this transfer function by a partial fraction expansion as $G(z) = G_c(z) + G_a(z)$, where $G_c(z)$ and $G_a(z)$ have their poles located exclusively inside and outside the unit-circle, respectively.

The component transfer functions $G_c(z)$ and $G_a(z)$ can then be implemented by stable causal and anticausal recursive operations, respectively. Instead of explicitly decomposing the deconvolution operator in this manner, however, one can employ a discrete Fourier transform method for effecting this decomposition in an automatic fashion [7]. Although most available deconvolution algorithms are based on using a moving average (MA) operator (i.e., $P = 0$), some authors have successfully appealed to a recursive deconvolution solution (e.g., [33]).

In using the ARMA expression (40) to compute the response elements, the left most response elements (due to causal effects) and the right most response elements (due to anticausal effects) are found to be distinctly nonstationary in behavior due to the dynamical effects of the deconvolution operation. Once this transient effect has been dissipated, however, the middle most response elements will behave in a stationary manner. With this in mind, it is assumed that the set of response elements are stationary over the index integers $L_1+1 \le n \le L_2$, where the user selects the transient response length parameters L_1 and L_2 by inspection of the response elements. The entities L_1 and $N - L_2$ correspond to the time it takes the causal and anticausal components of the deconvolving filter's unit-impulse response to be approximately zero. For the above recursive deconvolution operator, the response elements over the stationary interval $L_1+1 \le n \le L_2$ are then used to form estimates of the deconvolution signal's central moments, and these estimates are in turn used to form estimates of the deconvolved signals normalized cumulants. The following theorem provides a characterization of these central moment estimates.

Theorem 5 *Let the response elements y(n) be generated according to the ARMA(P, Q) relationship (40), in which the transient parameters are specified by L_1 and L_2. The standard response kth central moment estimate is given by:*

$$\hat{\mu}_Y(k) = \frac{1}{L_2 - L_1} \sum_{n=L_1+1}^{L_2} \left[y(n) - \hat{m}_Y\right]^k \tag{41}$$

in which the index, k, is a positive integer value and $\hat{m}_Y$ designates the sampled mean of the deconvolution operator's response elements, as computed by:

$$\hat{m}_Y = \frac{1}{L_2 - L_1} \sum_{n=L_1+1}^{L_2} y(n) \tag{42}$$

The components of the (Q+1) x 1 gradient vector of $\hat{\mu}_Y(k)$ with respect to the deconvolving operator's b_m coefficients are given by:

$$\frac{\partial \hat{\mu}_Y(k)}{\partial b_m} = \frac{k}{L_2 - L_1} \sum_{n=L_1+1}^{L_2} \left[y(n) - \hat{m}_Y\right]^{k-1} \times \left[s_b(n-m) - \frac{1}{L_2 - L_1} \sum_{k=L_1+1}^{L_2} s_b(k-m) \right] \tag{43}$$

for m = 0, 1 ,...,Q. In this expression, the elements of the response sensitivity sequence $s_b(n) = \partial y(n) / \partial b_0$ are computed according to the mixed causal-anticausal related AR(p) relationship

$$\sum_{k=0}^{P} a_k s_b(n-k) = x(n) \tag{44}$$

where the input, {x(n)}, corresponds to the given data (39) and zero initial conditions are imposed in computing the response $\{s_b(n)\}$. Similarly, the components of the P x 1 gradient vector of $\hat{\mu}_Y(k)$ with respect to the deconvolving operator's a_m coefficients are given by:

$$\frac{\partial \hat{\mu}_Y(k)}{\partial a_m} = \frac{-k}{L_2 - L_1} \sum_{n=L_1+1}^{L_2} \left[y(n) - \hat{m}_Y\right]^{k-1} \times \left[s_a(n-m) - \frac{1}{L_2 - L_1} \sum_{k=L_1+1}^{L_2} s_a(k-m) \right] \tag{45}$$

for m = 1,2,...,P. In this expression, the elements of the response sensitivity sequence $s_a(n) = \partial y(n) / \partial a_1$ are computed according to the related mixed causal-anticausal AR(p) relationship:

$$\sum_{k=0}^{P} a_k s_a(n-k) = y(n-1) \tag{46}$$

where {y(n)} are the response elements computed according to Eq. 40, and zero initial conditions are imposed in computing the response $\{s_a(n)\}$.

Formulation of Normalized Cumulant Estimates

The central moment estimates (41) may be used in conjunction with appropriate entrees from list (8) to form estimates of the response cumulants $c_Y(p)$ and $c_Y(q)$, while (13) is employed to obtain an estimate of the normalized response cumulant $k_Y(p,q)$. When we implement a gradient based recursive blind deconvolution algorithm along the same lines of the previously described nonrecursive deconvolution operator, the following related gradient of central moment ratio expression plays a central role:

$$\nabla_c \left[\frac{\hat{\mu}_Y(k_1)^{m_1}}{\hat{\mu}_Y(k_2)^{m_2}} \right] = \frac{m_1 \hat{\mu}_Y(k_1)^{m_1-1}}{\hat{\mu}_Y(k_2)^{m_2}} \nabla_c\left[\hat{\mu}_Y(k_1)\right] - \frac{m_2 \hat{\mu}_Y(k_1)^{m_1}}{\hat{\mu}_Y(k_2)^{m_2+1}} \nabla_c\left[\hat{\mu}_Y(k_2)\right] \tag{47}$$

where $\nabla_{\mathbf{c}}[\hat{\mu}_Y(k_Y)]$ designates the gradient vector with components in (45), or (43), depending on whether $\mathbf{c} = \mathbf{a}$ or $\mathbf{b}$, respectively; while k_1, k_2, m_1 and m_2 are arbitrary positive integers. Depending on the software package being used (e.g., FORTRAN, MATLAB, MAPLE), these computations may be implemented using either correlation or filtering operations.

The example on the equalization of random binary signals as developed in the previous section was approximated by a nonrecursive deconvolution operator of order $Q = 9$. The true deconvolution operator, however, has the linear recursive transfer function $F(z)^{-1} = (1-0.4z^{-1})/(-0.4+z^{-1})$ of order (1,1). In view of this observation, the recursive deconvolution algorithm as described in this section was used to deconvolve the same data set in which an ARMA(1,1) deconvolution model was employed. This recursive deconvolution performed in a superior fashion to that achieved by its nonrecursive counterpart. In particular, it took on the order of twelve iterations for the recursive model to converge, in contrast to thirty iterations for the nonrecursive model.

Constrained Kurtosis Extremizing

The approach of maximizing normalized cumulants in solving the blind deconvolution problem is appealing in that it provides a convenient platform for introducing various solution procedures. An alternate method for the direct maximization procedure is now described that is dependent on the previously mentioned scalar invariant property possessed by the normalized cumulant $k_Y(p,q) = c_Y(p)/|c_Y(2)|^{p/q}$. In particular, from the cumulant definition, it was established that for any nonzero scalar, α, the identity $|k_Y(p,q)| = |k_{\alpha Y}(p,q)|$ holds. This scalar invariant property, in turn, implies that the response's normalized cumulant magnitude is invariant for all linear systems that have unit-impulses that are scalar multiples of one another. This suggests that the blind deconvolution problem of the unconstrained maximization of $|k_Y(p,q)|$ over all unit-impulse selections is equivalent to maximizing the kurtosis numerator term $|c_Y(p)|$, subject to the constraint that the kurtosis denominator term satisfy $|c_Y(q)| = 1$. Although the approach to be taken is applicable to a general selection of the p and q parameters, for the purpose of this presentation, only the kurtosis case $p = 4$ and $q = 2$ is treated. The constrained optimization problem for the kurtosis case therefore takes the form

$$\max_{c_Y(2)=1} |c_Y(4)| \tag{48}$$

The cumulants involved in this constrained optimization are exclusively dependent on the two central moments, $\mu_Y(2)$ and $\mu_Y(4)$. For the purposes of this development, the deconvolution operation employed is taken to be governed by the nonrecursive operator

$$y(n;\mathbf{b}) = \sum_{k=0}^{Q} b_k x(n-k) \tag{49}$$

This response's dependency on the b_k coefficients is here explicitly recognized by the appearance of the $(Q+1) \times 1$ coefficient parameter vector, $\mathbf{b}$, as an argument. This operator's response to the given data set $x(1), x(2), \ldots, x(N)$, is first computed and the steady state response elements $y(Q+1)$, $y(Q+2), \ldots, y(N)$ are used to form estimates of the response's mth central moment as specified by:

$$\hat{\mu}_Y(m;\mathbf{b}) = \frac{1}{N-Q} \sum_{n=Q+1}^{N} [y(n;\mathbf{b}) - \hat{m}_y(\mathbf{b})]^m$$

where $\hat{m}_Y(\mathbf{b})$ designates the standard *sampled mean* of the response sequence computed for $Q+1 \le n \le N$. These estimates for $m = 2,4$, are, in turn, inserted into the formulas for the second and fourth order cumulants found in the list (8) to form the estimates $\hat{c}_Y(2;\mathbf{b})$ and $\hat{c}_Y(4;\mathbf{b})$. In accordance with the constrained optimization problem (48), it is assumed that the current deconvolution operators $\{b_k\}$ coefficients have been chosen to satisfy the imposed constraint $\hat{c}_Y(2;\mathbf{b}) = 1$, that is,

$$\hat{c}_Y(2;\mathbf{b}) = \frac{1}{N-Q} \sum_{n=Q+1}^{N} [y(n;\mathbf{b}) - \hat{m}_y(\mathbf{b})]^2 = 1$$

while the term whose magnitude is being maximized is specified by

$$\hat{c}_Y(4;\mathbf{b}) = \frac{1}{N-Q} \sum_{n=Q+1}^{N} [y(n;\mathbf{b}) - \hat{m}_y(\mathbf{b})]^4 - 3$$

The prevailing deconvolving operator's $\{b_k\}$ coefficients are now perturbed in the additive fashion to $b_k + \delta_k$, where the perturbation vector, δ_k, is typically small in size. In accordance with the comments made in the opening paragraph, this perturbation vector is selected so as to satisfy the improving condition $|\hat{c}_Y(4;\mathbf{b}+\delta)| > |\hat{c}_Y(4;\mathbf{b})|$, while maintaining the imposed constraint, $\hat{c}_Y(2;\mathbf{b}+\delta) = 1$. Upon substituting the perturbed coefficient vector with elements $b_k + \delta_k$ into (49), we get the steady state perturbed response:

$$y(n;\mathbf{b}+\delta) = y(n;\mathbf{b}) + \sum_{k=0}^{Q} \delta_k x(n-k)$$
$$n = Q+1, Q+2, \cdots, N \tag{50}$$

which is seen to be linear in the δ_k elements. The sampled mean of this perturbed response is then given by:

$$\hat{m}_Y(\mathbf{b}+\delta) = \hat{m}_Y(\mathbf{b}) + \frac{1}{N-Q} \sum_{n=Q+1}^{N} \left[\sum_{k=0}^{Q} \delta_k x(n-k) \right]$$

An mth order estimate of the perturbed central moment, $\mu_Y(m;\mathbf{b}+\delta)$, is formally obtained by subtracting this sample mean from each side of (50), raising that difference to the mth power, summing this result over the integer range $Q+1 \le n \le N$, and, then dividing that sum by $N - Q$. A linear approximation of this estimate is then obtained by making a Taylor series expansion in the δ_k variables, in which only the constant and linear terms are retained. This linear approximation of the mth central moment is found to be:

$$\hat{\mu}_Y(m;\mathrm{b}+\delta)=\hat{\mu}_Y(m;\mathrm{b})+\rho_m(\delta)$$

where the linear term, $\rho_m(\delta)$, is specified by

$$\rho_m(\delta)=\frac{m}{N-Q}\sum_{k=0}^{Q}\delta_k \times$$
$$\left\{\sum_{n=Q+1}^{Q}\left[y(n;\mathrm{b})-\hat{m}_Y(\mathrm{b})\right]^{m-1}\left[x(n-k)-\frac{1}{N-Q}\sum_{i=Q+1}^{N}x(i-k)\right]\right\} \tag{51}$$

This approximation can be made as accurate as desired by restricting the δ_k perturbations to be sufficiently small in magnitude.

We now desire to select the unit-impulse response perturbation vector δ components so as satisfy the improvement condition $|\hat{c}_Y(4,\mathbf{b}+\delta)|>|\hat{c}_Y(4,\mathbf{b})|$, subject to the constraint $\hat{c}_Y(2,\mathbf{b}+\delta)=1$. From linear approximation considerations (51) with $m = 4$, we see that

$$\hat{c}_Y(4,\mathrm{b}+\delta)=\hat{\mu}_/(4;\mathrm{b}+\delta)-3$$
$$=\hat{c}_Y(4,\mathrm{b})+\rho_4(\delta)$$

A logical choice for the perturbation δ is to set it equal to the gradient of $\rho_4(\delta)$ with respect to δ, multiplied by sgn$[c_Y(4;\mathbf{b})]$. The elements of this gradient vector are therefore given by

$$\frac{\partial\rho_4(\delta)}{\partial\delta_k}=\frac{4}{N-Q}\sum_{n=Q+1}^{N}\left[y(n;\mathrm{b})-\hat{m}_Y(\mathrm{b})\right]^3\times$$
$$\left[x(n-k)-\frac{1}{N-Q}\sum_{i=Q+1}^{N}x(i-k)\right]$$

for $k=0,1,\ldots,Q$. Unfortunately, although this vector points in the direction of maximum increase of $\hat{c}_Y(4,\mathbf{b}+\delta)$, it generally causes the imposed constraint $\hat{c}_Y(2,\mathbf{b}+\delta)=1$ to be invalid. This is a direct consequence of the fact that the functional $\rho_2(\delta)$ is typically nonzero. To maintain the desired general direction of this gradient direction while satisfying the imposed constraint, we simply set the δ perturbation vector equal to that component of $\nabla_\delta[\rho_4(\delta)]\mathrm{sgn}[k_W(4,2)]$, which is orthogonal to the gradient vector $\nabla_\delta[\rho_2(\delta)]$, whose components are given by

$$\frac{\partial\rho_2(\delta)}{\partial\delta_k}=\frac{2}{N-Q}\sum_{n=Q+1}^{N}\left[y(n;\mathrm{b})-\hat{m}_Y(\mathrm{b})\right]\times$$
$$\left[x(n-k)-\frac{1}{N-Q}\sum_{i=Q+1}^{N}x(i-k)\right]$$

for $k = 0, 1,\ldots,Q$. It is readily shown that the perturbation vector that satisfies this requirement is given by

$$\delta^0=\left\{\nabla_\delta[\rho_4(\delta)]-\frac{\nabla_\delta[\rho_4(\delta)]^T\nabla_\delta[\rho_2(\delta)]}{\nabla_\delta[\rho_2(\delta)]^T\nabla_\delta[\rho_2(\delta)]}\nabla_\delta[\rho_2(\delta)]\right\}\times$$
$$\mathrm{sgn}[k_Y(4,2;\mathrm{b})] \tag{52}$$

As in the gradient approach to blind deconvolution, the components of the gradient vectors $\nabla_\delta[\rho_4(\delta)]$ and $\nabla_\delta[\rho_4(\delta)]$ are most efficiently computed using correlation techniques previously described. Moreover, it will be necessary to appropriately scale the perturbed unit-impulse response vector $\mathbf{b}+\delta^o$, since although $\rho_2(\delta^o) = 0$, this satisfies only the linear approximation of the imposed constraint. This scaling is easily achieved by first computing the perturbed response $y(n;\mathbf{b}+\delta^o)$ elements and then evaluating $\hat{c}_Y(2,\mathbf{b}+\delta)$, which will generally not be equal to one. It is a simple matter to show that the scaled perturbed unit-impulse response vector given by $\tilde{\mathbf{b}}=[\mathbf{b}+\delta^o]/\sqrt{\hat{c}_Y(2,\mathbf{b}+\delta}$ satisfies the imposed constraint.

Gauss-Newton Based Kurtosis Deconvolution

In any direct algorithmic approach for maximizing a functional, the parameter vector to be optimized is perturbed in a systematic fashion so as to increase the functional at each iteration. It is possible, however, to use an indirect approach for achieving the same objective. A well known, necessary condition for a point to be a maximum, minimum, or saddle point of a functional is that the functional's gradient vector, evaluated at that point, must equal the zero vector. With this in mind, let us consider the case in which the deconvolution operator is governed by the nonrecursive relationship

$$y(n,\mathbf{b})=\sum_{k=0}^{Q}b_k x(n-k)$$

We shall now examine the case in which it is desired to extremize the kurtosis $k_Y(4,2;\mathbf{b})$. As suggested above, a necessary condition for this extremization is that the gradient of $k_Y(4,2;\mathbf{b})$with respect to $\mathbf{b}$ be equal to the zero vector. Using straightforward differentiation of $k_Y(4,2;\mathbf{b})$ with respect to the b_m coefficients, this necessary condition takes the form:

$$\mu_Y(2;\mathbf{b})\nabla_\mathbf{b}[\mu_Y(4;\mathbf{b})]=2\mu_Y(4;\mathbf{b})\nabla_\mathbf{b}[\mu_Y(2;\mathbf{b})] \tag{53}$$

Upon substituting estimates of the central moments $\mu_Y(2;\mathbf{b})$ and $\mu_Y(4;\mathbf{b})$ into this expression for a given $\mathbf{b}$, we generally find that these equations are not satisfied and give rise to the nonzero error vector

$$\mathrm{e}(\mathrm{b})=\hat{\mu}_Y(2;\mathrm{b})\nabla_\mathrm{b}[\hat{\mu}_Y(4;\mathrm{b})]-2\hat{\mu}_Y(4;\mathrm{b})\nabla_\mathrm{b}[\hat{\mu}_Y(2;\mathrm{b})]$$

It is then desirable to perturb the unit impulse response vector from $\mathbf{b}$ to $\mathbf{b}+\delta$ in a manner to drive the perturbed error vector, $\mathbf{e}(\mathbf{b}+\delta)$, closer to the zero vector than $\mathbf{e}(\mathbf{b})$. Using a Taylor series expansion, this perturbed error vector is approximated by the linear model

$$\mathbf{e}(\mathbf{b} + \delta) \cong \mathbf{e}(\mathbf{b}) + L_{\mathbf{b}}\delta \tag{54}$$

In this expression, the $(Q+1) \times (Q+1)$ matrix, $L_{\mathbf{b}}$, is given by

$$L_g = \hat{\mu}_Y(2;\mathbf{b})H(\mathbf{b}) - 2\hat{\mu}_Y(4;\mathbf{b})H(\mathbf{b}) +$$

$$\nabla_{\mathbf{b}}[\hat{\mu}_Y(4;\mathbf{b})]\nabla_{\mathbf{b}}[\hat{\mu}_Y(2;\mathbf{b})]^T - 2\nabla_{\mathbf{b}}[\hat{\mu}_Y(2;\mathbf{b})]\nabla_{\mathbf{b}}[\hat{\mu}_Y(4;\mathbf{b})]^T$$

where the (i,j)th element of the $(Q+1) \times (Q+1)$ matrix $H(\mathbf{b})$ is given by:

$$H(\mathbf{b})_{ij} = \frac{k(k-1)}{N-Q}\sum_{n=Q+1}^{N}[y(n;\mathbf{b}) - m_Y(\mathbf{b})]^{k-2}x(n-i)x(n-j)$$

$$0 \le i,j \le Q$$

The perturbed vector is then chosen so as to make the perturbed error approximation (54) equal to the zero vector. The required perturbation is therefore obtained by solving the linear system of equations

$$\delta^o = -\mathbf{L_b}^{-1}\mathbf{e}(\mathbf{b})$$

The vector $\mathbf{b}+\delta$ then serves as the next approximating solution to the necessary condition expression (53), where the step size parameter, α, may be selected in the sequential decreasing fashion previously described. If the initial choice for $\mathbf{b}$ is sufficiently close to the desired solution , then it is known that this algorithmic method converges quadratically to the solution. Unfortunately, it is necessary to solve a system of linear equations at each iteration for the perturbation vector. This task can be very computationally demanding for moderate to large values of the deconvolution operator's order Q.

Conclusion

Several algorithmic approaches for solving the blind deconvolution problem have been presented that are based on the concept of maximizing the magnitudes of the response's normalized cumulant estimates. This approach is predicated on the theoretical fact that the required deconvolution operator is one that maximizes the magnitudes of all nonzero response normalized cumulants. In this approach, one can either maximize a specific response normalized cumulant magnitude, or, a combination of response normalized cumulants. If a specific response normalized cumulant magnitude estimate is maximized, the confidence (i.e., validation) in the resultant deconvolution can be gauged by observing the behavior of other response normalized cumulant magnitude estimates as the deconvolution unit-impulse response evolves. In particular, one would have a great deal of confidence in the resultant deconvolution operator for those situations where the estimated magnitudes of the other response-normalized cumulants appear to approach a maximum as the specific response normalized cumulant being extremized approaches its magnitude maximum. On the other hand, a lack of confidence would follow if this behavior is not observed, since the optimum deconvolution operator should maximize the magnitudes of all nonzero response normalized cumulant estimates. As a final note, the development in this article was concerned with the noise free data case, in order to emphasis the principals of cumulant extrema. These algorithms, however, are effective in the case of moderate additive noise.

Acknowledgment

This work was sponsored in part by the Air Force Office of Scientific Research under Grant #F49620-93-1-0268SDIO/IST and the SDIO/IST, and managed by the Office of Naval Research under Grant #N00014-92-J-1995.

James A. Cadzow is Centennial Professor of Electrical Engineering in the Department of Electrical Engineering, Vanderbilt University.

References

1. M.S. Bartlett, *An Introduction to Stochastic Processes*, London, Cambridge University Press, 1955.

2. A. Benveniste, M. Goursat, and G. Rouget, "Robust identification of a non-minimum phase system: blind adjustment of a linear equalizer in data communications," *IEEE Transactions on Automatic Control*, vol. AC-25, no. 3, pp. 385-399, June 1980.

3. A. Benveniste and M. Goursat, "Blind Equalizers," *IEEE Transactions on Communications*, vol. COM-32, no. 8, pp. 871-883, August 1984.

4. D. R. Brillinger, *Time Series, Data Analysis, and Theory*, New York, Holt, Rinehart, and Winston, 1975.

5. D. R. Brillinger and M. Rosenblatt, "Computation and interpretation of kth-order spectra," in *Spectral Analysis of Time Series*, B. Harris, Editor, NYC, Wiley, pp. 189-232, 1967.

6. J.A. Cadzow and Xingkang Li, "Blind deconvolution," *Digital Signal Processing Journal*, John Wiley, vol. 5, no. 1, January, 1995, pp.3-20.

7. J.A. Cadzow, "Stable realization of mixed causal-anticausal linear recursive operators via the DFT," submitted for publication.

8. D. L. Donoho, "On minimum entropy deconvolution," *Applied Time Series Analysis, II*," D. F. Findley, editor, New York, Academic Press, 1981.

9. G. B. Giannakis and J.M. Mendel, "Identification of nonminimum phase system using higher order statistics," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, pp. 360-377, March 1989.

10. G. B. Giannakis and J. M. Mendel, "Cumulant-based order determination of non-gaussian ARMA models," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-38, no. 8, pp. 1411-1423, August 1990.

11. G. B. Giannakis and A. Swami, "On estimating noncausal nonminimum phase ARMA models of non-gaussian processes," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-38, no. 3, pp. 478-495, March 1990.

12. D. Godard, "Self recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Transactions on Communications*, vol. COM-28, no. 11, pp. 1867-1875, November 1980.

13. R. M. Gray and L. D. Davisson, *Random Processes: A Mathematical Approach for Engineers*, Englewood, NJ, Prentice-Hall, 1986.

14. D. Hatzinakos and C.L. Nikias, "Blind equalization using a tricepstrum-based algorithm," *IEEE Transactions on Communications,* vol. COM-39, no. 5, pp. 669-883, May 1991.

15. *Blind Deconvolution*, S. Haykin, editor, Englewood Cliffs, NJ, Prentice Hall, 1994.

16. H. L. Larson and B. O. Shubert, *Probabilistic Models in Engineering Sciences: Random Variable and Stochastic Processes*, Vol. I, New York, John Wiley and Sons, 1979.

17. H.L. Larson and B.O. Shubert, *Probabilistic Models in Engineering Sciences*, Vol. II, New York, John Wiley and Sons, 1979.

18. J. Longbottom, A.T. Walden and R.E. White, "Principles and application of maximum kurtosis phase estimation," *Geophysical Prospecting*, vol. 36, pp. 115-138, February 1988.

19. L. Ljung, "Consistency of least squares identification method," *IEEE Transactions on Automatic Control*, vol. AC-21, pp. 791-781, October, 1976.

20. J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications," *Proceedings of the IEEE*, vol. 79, pp. 278-305, March 1991.

21. C. L. Nikias, "ARMA bispectrum approach to nonminimum phase system identification," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-36, no. 4, pp. 513-524, April 1988.

22. C. L. Nikias and J. M. Mendel, "Signal processing with higher order spectra," *Signal Processing Magazine*, pp. 10-37, July, 1993.

23. C. L. Nikias and A.P. Petropulu, *Higher-Order Spectral Analysis*, Englewood Cliffs, NJ, Prentice Hall, 1993.

24. M. Ooe and T.J. Ulrych, "Minimum entropy deconvolution with an exponential transformation," *Geophys. Prosp.*, vol. 27, pp. 458-473, 1979.

25. A. V. Oppenheim and R.W. Schafer, *Digital Signal Processing*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1975.

26. A. Papoulis, *Probability, Random Varialbe and Stochastic Processes*, New York, McGraw-Hill Book Co., 1965.

27. M. Raghuveer and C.L. Nikias, "Bispectrum estimation: a parametric approach," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, pp. 1213-1230, 1985.

28. M. Rosenblatt, *Stationary Sequences and Random Fields*, Boston, MA, Birkhauser, 1985.

29. O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems," *IEEE Transactions on Information Theory*," vol. 36, no. 2, pp. 312-321, March 1990.

30. O. Shalvi and E. Weinstein, Universal Methods for Blind Deconvolution. In: *Blind Deconvolution*, S. Haykin, ed., Prentice Hall, Englewood Cliffs, NJ, 1994.

31. Y. Sato, "A method of self-recovering equalization for multilevel amplitude modulation," *IEEE Transactions on Communications*, pp. 679-682, June 1975.

32. A. Stuart and J.K. Ord, *Kendall's Advanced Theory of Statistics, Vol I, Distribution Theory*, John Wiley, New York.

33. J. K. Tugnait, "Estimation of linear parametric models using inverse filter criteria and higher order statistics," *IEEE Transactions on Signal Processing*, vol. 41, no. 11, pp. 3196-3199, November 1993.

34. R. A. Wiggins, "On minimum entropy deconvolution," *Geoexploration*, vol. 16, pp. 21-35, 1978.

35. A. Ziolkowski, *Deconvolution*, Boston, International Human Resources Development Corporation, 1984.