An Appraisal of Gardner Transform-Based Methods of Transient Multiexponential Signal Analysis

Abdussamad U. Jibia and Momoh-Jimoh E. Salami

Abstract—There are several problems in applied science in which experimental observations can be accurately represented by a sum of exponential decay functions in which the amplitudes, decay rates and number of components have different physical interpretations and need to be estimated. A parameter estimation technique of multicomponent exponential functions that has undergone many modifications is the Gardner transform in which a nonlinear transformation is used to convert the data signal into a convolution model containing the parameters of interest. Modifications of this early technique include modification of the original transform or deconvolution procedure and additional processing of the deconvolved data to obtain better estimates of the desired parameters. This paper presents an appraisal of Gardner transform and its variants. It discusses major modifications and their implications to the overall results of analysis.

Index Terms—Gardner transform, deconvolution, modeling, multiexponential.

I. INTRODUCTION

There are many problems in science and engineering in which the data approximates to a linear combination of exponentials of the form:

$$S(\tau) = \sum_{i=1}^{M} A_i \exp(-\lambda_i \tau) + n(\tau) \quad 0 < \tau < \infty \tag{1}$$

Such problems are common in semiconductor physics (deep level transient spectroscopy), biophysics (fluorescence decay analysis), nuclear physics and chemistry (radioactive decays, nuclear magnetic resonance), chemistry and electrochemistry (reaction kinetics) and medical imaging. In these and other problems, the parameters M, A_i and λ_i have different physical meanings. It is therefore not sufficient that the function approximates the data accurately; it is also important that these parameters are accurately estimated. This problem is a classical one and many approaches to its solution have been proposed. The problem is difficult and there has been no perfect magic way to its solution. All proposed approaches suffer from shortcomings. The difficulty in the analysis of this class of signals lies in the fact that we are dealing with a series of nonlinear equations, that the data are only approximating the function over a finite range in τ and the strong nonorthogonality of the exponential functions with real decay rates.

One of the classical approaches to the solution of the problem was proposed by Gardner et al [1] in which approach a nonlinear change of variables was used to convert the original signal into a convolution integral which could be deconvolved by Fourier transform technique. Gardner transform belongs to the more general class of spectroscopic methods [2] in which the decay is described by a continuous distribution of decay rates which may be considered as a spectral representation of the transient signal

$$S(\tau) = \int_{0}^{\infty} g(\lambda) \exp(-\lambda \tau) d\lambda \tag{2}$$

Analysis in this case involves the determination of the spectral function $g(\lambda)$ Eq. (2) reduces to (1) if the spectral function $g(\lambda)$ can be represented as a sum of M delta functions

$$g(\lambda) = \sum_{i=1}^{M} A_i \delta(\lambda - \lambda_i) \tag{3}$$

At the time it was introduced, Gardner transform did not attract the attention of many researchers basically because of the nonavailability of effective algorithms for the computation of Fourier integrals. This problem was later solved by Schlesinger [3] using the fast Fourier transform (FFT). Although successful in simplifying the numerical computation of the Fourier integrals, Schlesinger's method did not have a good filtering technique to eliminate the side ripples caused by FFT and noise. Several other research findings have since been published on different ways of improving the performance of Gardner transform. These modifications can generally be grouped into three:
1) Modifications in the nonlinear transformation.
2) Modifications in the deconvolution procedure.
3) Modifications involving additional processing of the deconvolved data.

In this paper, we critically examine these important modifications of the Gardner transform with special emphasis on our previous and current research in this area. Performance of the various approaches including their merits and shortcomings are examined.

In section II of this paper, an overview of Gardner transform is given along with its merits and the inherent problems in its implementation. Subsequent sections discuss the variants of the Gardner transform. Section III discusses the modification of the original nonlinear change of variables and its effect on sampling and noise reduction. Section IV presents methods that involve modification in the deconvolution procedure. In section V parametric techniques used in modeling the deconvolution results are reviewed.

II. GARDNER TRANSFORM

Starting with the multiexponential function

$$S(\tau) = \sum_{i=1}^{M} A_i \exp(-\lambda_i \tau) \tag{4}$$

We can rewrite it as

$$S(\tau) = \int_{0}^{\infty} g(\lambda) \exp(-\lambda \tau) d\lambda \tag{5}$$

where $g(\lambda) = \sum_{i=1}^{M} A_i \delta(\lambda - \lambda_i)$ and $\delta(\lambda)$ is the Diract delta function.

If $S(\tau)$ represents experimental data, $g(\lambda)$ will be a continuous spectrum whose maxima will give the values of λ_i and their heights will be proportional to A_i. Thus, determination of M, A_i and λ_i will be reduced to the determination of $g(\lambda)$.

Gardner transform is based on the inversion of the integral in (5) into a convolution integral. To achieve this, Gardner and co-researchers used the substitution $\tau = e^t$ and $\lambda = e^{-r}$. Thus equation (5) becomes

$$S(e^t) = \int_{-\infty}^{\infty} g(e^{-r}) \exp(-e^{(t-r)}) e^{-r} dr \tag{6}$$

Both sides of equation (6) are multiplied by e^t to give the convolution integral

$$e^t S(e^t) = \int_{-\infty}^{\infty} g(e^{-r}) e^{r(\alpha-1)} \exp(-e^{(t-r)}) e^{(t-r)} dr \tag{7}$$

This can be expressed as

$$y(t) = \int_{-\infty}^{\infty} x(\lambda) h(t-\lambda) d\lambda + v(t) \tag{8}$$

where

$y(t) = \exp(t) S(\exp(t))$, $x(t) = \exp((\alpha-1)t) g(e^{-t})$, $h(t) = \exp(t) \exp(-e^t)$ and $v(t) = \exp(\alpha t) n(e^t)$.

Gardner et al [1] deconvolved this integral using the Fourier transform

$$g(e^t) = \Im^{-1} \left[\frac{\Im[e^t f(e^t)]}{\Im[e^t \exp(e^{-t})]} \right] \tag{9}$$

where $\Im$ denotes the Fourier transform operation and $\Im^{-1}$ its inverse. A plot of $g(e^{-t})$ against λ shows maxima when $\lambda = \lambda_i$ with the amplitude being proportional to A_i / λ_i. Hence, A_i can be computed once λ_i is accurately located.

At the time of introduction of the Gardner transform, there was difficulty in performing fast numerical integration of the Fourier transform integral with enough precision. Furthermore, it is impossible to integrate from $-\infty$ to ∞. The integration interval was thus truncated on both sides by means of the cut-off points $\pm t_0$ for the integral in the direct Fourier transform and $\pm \omega_0$ in the inverse Fourier transform.

This truncation led to the emergence of spurious high-frequency components in the spectrum of the direct Fourier transform and in error ripples in the inverse Fourier transform. The effect of error ripples can be alleviated by an appropriate choice of ω_0. A very small value of ω_0 would however lead to loss of resolution in the resulting plot of $g(e^t)$ against t.

The accuracy of Gardner transform largely depends on the available data. Data with lower noise level usually give better accuracy than data with higher noise level. For example, a high level of resolution in the final results is possible when analyzing a mixture of radioactive isotopes. In contrast, certain types of chemical kinetic data or some measurements involving the rates at which injected materials disseminate in a living organism are usually obtained with poor accuracy. Accurate results are possible if the measurements are repeated as often as possible to get the average results.

Gardner transform has a number of advantages and disadvantages. Some of its advantages include:

- No a priori knowledge of the unknown parameters is required.
- All the parameters are identified simultaneously.
- Broadened peaks indicate the presence of exponential components with similar (unresolved) decay rates without interfering with the determination of other components.

Among its disadvantages are:

- The nonlinear transformation of the experimental data $S(\tau)$ into the signal $y(t)$ so as to derive the convolution integral equation.
- The necessity of approximating the infinite continuous Fourier transform by a finite DFT.
- Enhancement of high frequency noise by the deconvolution procedure.
- The nonlinear change of variables complicates the analysis especially for noisy signals. Even if the noise is derived from a stationary white noise process, there is no guarantee that the transformed noise would be stationary.

III. MODIFICATION IN THE NONLINEAR TRANSFORMATION

A. Swingler's Differentiation Technique

Swingler [4] proposed a different approach to the Gardner transform. He proposed the formation of a t-derivative $S'(e^t)$ as a starting point instead of $e^t S(e^t)$ of in the original Gardner transform and its earlier modifications. In this differential technique, the t-derivative is given by

$$\frac{dS(e^t)}{dt} = \sum_{i=1}^{M} A_i e^{(t-r)} \exp[-e^{(t-r)}]$$
$$= -e^t \exp(e^r) \sum_{i=1}^{M} A_i \delta(t-r) \tag{10}$$

Eq. (10) is approximated with a simple first-order difference equation. The remaining part of the procedure is similar to Schlesinger's FFT technique. Next $G(\omega)$ is

multiplied by a windowing function so as to reduce the sidelobes and error ripples in the final output.

The merits of Swingler's differential technique can be outlined as follows:

- It yields processed outputs whose peaks are proportional to A_i for better resolution display so that the problem of lost components in the spectrum is minimized.
- It has the ability to analyze function with direct current (DC) bias level or DC offset because derivative of a constant is zero.
- It requires only the first-order difference equation for approximating the t-derivative.

However, the performance of this technique is marred by sensitivity to noise due to the deconvolution process. An additional noise in the original signal even worsens the performance. A signal of up to four components ($M = 4$) has been analyzed using this method.

B. Modification by Weighting Factor Introduction

Nichols et al [5] modified the nonlinear transformation by introducing a weighting factor, α into it. To begin with, the original signal in Eq. (1) is assumed to be embedded in additive white Gaussian noise $n(\tau)$ leading to the following equations in place of equations (1) and (2) respectively.

$$S(\tau) = \sum_{i=1}^{M} A_i \exp(-\lambda_i \tau) + n(\tau) \qquad (11)$$

$$S(\tau) = \int_{0}^{\infty} g(\lambda) \exp(-\lambda \tau) d\lambda + n(\tau) \qquad (12)$$

Eq. (12) was now multiplied by τ^{α}, $\alpha > 0$ instead of τ and the transformation becomes $\tau^{\alpha} = e^{\alpha t}$ and $\lambda = e^{-r}$. This now leads to

$$e^{\alpha t} S(e^t) = \int_{-\infty}^{\infty} g(e^{-r}) e^{r(\alpha-1)} \exp(-e^{(t-r)}) e^{\alpha(t-r)} dr \qquad (13)$$
$$+ e^{\alpha t} n(e^t)$$

which simplifies to

$$y(t) = \int_{-\infty}^{\infty} x(\lambda) h(t-\lambda) d\lambda + v(t) \qquad (14)$$

where

$$y(t) = \exp(\alpha t) S(\exp(t)) \qquad (15a)$$
$$x(t) = \exp((\alpha-1)t) g(e^{-t}) \qquad (15b)$$
$$h(t) = \exp(\alpha t) \exp(-e^t) \qquad (15c)$$
$$v(t) = \exp(\alpha t) n(e^t) \qquad (15d)$$

and $-\infty < t < \infty$.

Salami [6] obtained an expression for the input distribution from the following analysis.

Multiplying each side of equation (11) by τ^{α}, $\alpha > 0$ gives

$$\tau^{\alpha} S(\tau) = \sum_{i=1}^{M} A_i e^{-\lambda_i \tau} \tau^{\alpha} + \tau^{\alpha} n(\tau) \qquad (16)$$

$$\tau^{\alpha} S(\tau) = \sum_{i=1}^{M} \frac{A_i}{\lambda_i^{\alpha}} (\lambda_i \tau)^{\alpha} e^{-\lambda_i \tau} + \tau^{\alpha} n(\tau) \qquad (17)$$

Letting $\tau = e^t$ and $\lambda_i = e^{\ln \lambda_i}$, then

$$\tau \lambda_i = e^{t + \ln \lambda_i} \qquad (18)$$

which when substituted into equation (17) gives

$$e^{\alpha t} S(e^t) = \sum_{i=1}^{M} \frac{A_i}{\lambda_i^{\alpha}} \varphi(t + \ln \lambda_i) + e^{\alpha t} n(e^t) \qquad (19)$$

where

$$\varphi(t + \ln \lambda_i) = \exp[\alpha(t + \ln \lambda_i) - \exp(t + \ln \lambda_i)].$$

Equations (15b) and (15c) can be used to simplify expression (19) to give

$$y(t) = \sum_{i=1}^{M} \frac{A_i}{\lambda_i^{\alpha}} \varphi(t + \ln \lambda_i) + v(t) \qquad (20)$$

Allowing $p'(\lambda_i t) = \varphi(t + \ln \lambda_i)$, then equation (20) can be written as

$$y(t) = \sum_{i=1}^{M} \frac{A_i}{\lambda_i^{\alpha}} p'(\lambda_i t) + v(t) \qquad (21)$$

where $p'(\lambda_i t)$ is the derivative of $p(\lambda_i t) = \exp(-\lambda_i t)$ with respect to t.

Comparing equations (14) and (21), it can be concluded that the unknown input distribution function is given by:

$$x(t) = \sum_{i=1}^{M} B_i \delta(t + \ln \lambda_i) \qquad (22)$$

where $B_i = A_i(\lambda_i)^{-\alpha}$.

The immediate effect of the weighting factor, α, was in noise reduction [6]. Recently, Jibia et al [7] related it to the bandwidth of the conceptual system response $h(t)$ by using the concept of noise equivalent bandwidth. If the NEB is given by

$$B = \frac{\int_{-\infty}^{\infty} |h(t)|^2}{2 \left[\int_{-\infty}^{\infty} h(t) dt \right]^2} \qquad (23)$$

By integrating and taking the limits to be 0 and some arbitrary t_{min} sufficiently large, they showed that the bandwidth is given by

$$B = \frac{\alpha}{2} \qquad (24)$$

For unaliased sampling, therefore, the sampling frequency, f_s, must satisfy

$$f_s \geq 2B = \alpha \qquad (25)$$

A very important result of introducing α is the spectral fading which takes place for components with small amplitudes (A_i). This means that certain components are consumed by the effect of noise and the division by λ^{α} in Eq. (22) when the value of α is raised high. Therefore, much as

α contributes to noise reduction, it is not advisable to raise its value very high because of the deleterious effect this may have on the detectability of certain components. A maximum value of $\alpha = 1$ would normally suffice.

IV. MODIFICATIONS IN THE DECONVOLUTION PROCEDURE

A. Schlesinger's FFT Technique

When the Gardner transform was proposed, it was not widely accepted due to the difficulty in evaluating Fourier integrals. Realizing this shortcoming, Schlesinger proposed a means of alleviating it. His proposal published in [3] consists of the replacement of the numerical integration of the Fourier transform and its inverse with the discrete Fourier transform and its fast Fourier transform algorithm of Cooley and Tukey [8].

Schlesinger used the FFT algorithm to separately evaluate the Fourier transform of y(t) and x(t) in equation (8) and the inverse discrete Fourier transform of $\dfrac{Y(\omega)}{X(\omega)}$ gives an expression similar to (9).

In most cases deconvolution takes the form of a lowpass filter, reducing the amplitudes of high frequency components in the signal. However, if the amplitudes of these components fall below the inherent noise of the system, the information in these frequencies is lost. Other sources of noise include sampling errors and noise in signal acquisition, leakage and aliasing errors in the Fourier transform and computational errors due to the computing machine.

Although successful in simplifying the numerical computation of the integrals, Schlesinger's method did not have a good filtering technique to eliminate the side ripples caused by FFT and noise. This limited its detection power to only three components ($M = 3$).

B. Digital Technique with Gaussian Filtering

In their paper, Cohn Sfetcu et al [9] observed that the inherent deconvolution accompanying Gardner transform enhances noise in the data by favouring high-frequency components, especially for low frequency signals. Further, they observed, even if the original data were noise-free, digital processing with the FFT was bound to produce some noise due to inherent computational inaccuracies. This noise is also enhanced by deconvolution. To alleviate the problem of noise enhancement, Cohn Sfetcu and co-authors proposed Gaussian filtering. In the Gaussian filtering, the impulse response function of the filter is expressed as follows:

$$H_g(\omega) = \exp\left(-\omega^2 / \omega_d^2\right) \qquad (26)$$

The dispersion parameter ω_d must be tuned to particular experimental conditions. The smaller the value of ω_d the better the SNR of the spectrum.

This filter was proposed as a compromise between noise reduction and loss of resolution when dealing with separate pulses and white noise. Unfortunately, for this approach to work properly a collection of interpolation of data points are required at increasingly large intervals for better results. However, it is a difficult problem to obtain a correct interpolation itself. Furthermore, a correct usage of the FFT techniques requires the evaluation of the required signals over a much larger range of data points. A larger number of points in the DFT would normally increase the resolution but at a cost of longer computational times and more computational noise from the FFT and interpolation.

C. Provencher's Modification of the FFT

Provencher [10] modified the Schlesinger's FFT technique by introducing two additional parameters β_c and ε into Eq. (5), making the solution more practicable despite a severely restricted range of data points. Parameter β_c is the convergence parameter and ε is the amplitude equalization parameter.

To begin with, the variables, λ and τ of Eq. (5) are replaced by $\lambda = e^{-z} - \beta_c$ and $t = e^t$. Then, both sides of equation (5) are multiplied by $\exp\left[(1+\varepsilon)t - \beta_c e^t\right]$ with $\varepsilon > -1$. Eq. (5) is rewritten as

$$g_y(t) = \int_{-\infty}^{\infty} g_s(z) g_m(t-z) dz \qquad (27)$$

where

$$g_y(t) = \exp\left[(1+\varepsilon)t - \beta_c e^t\right] S\left(e^t\right) \qquad (28)$$

$$g_s(z) = e^{\varepsilon z} g\left(e^{-z} - \beta_c\right) = e^{\varepsilon z} g(\lambda) \qquad (29)$$

and

$$g_m(t-z) = \exp\left[-\exp(t-z) + (1+\varepsilon)(t-z)\right] \qquad (30)$$

The Fourier transform of equation (27) is $G_y(\omega) = G_s(\omega) G_m(\omega)$. The formal solution is obtained by taking the inverse Fourier transform of $G_s(\omega)$.

The problem of error ripples in the Gardner transform is solved by adjusting the convergence parameter, β_c so that the last few data points to $g_y(0)$ is between 0.02 to 0.04 % of the total $g_y(0)$. The value of β_c should not be too large because large values of β_c will reduce the distance between the peaks in the frequency spectrum and hence limit the resolving power of this technique. Some of the problems of the Gardner transform technique are the loss of components in the spectrum with large λ_k and the divergence of solution for a baseline component with $\lambda_i = 0$. These problems are caused by the proportionality of the amplitude of the spectral peaks to the height of the peak, A_i / λ_i instead of A_i. The parameters $\beta_c > 0$ and $-1 < \varepsilon < 0$ are chosen to minimize these problems as much as possible.

In his paper, Provencher [10] concluded that his technique performs better than the previous modifications of the Gardner transform technique. However, the resolving power

of this technique decreases when the peaks are masked as the number of closely spaced components increases.

D. Optimal Compensation Deconvolution

The idea of optimal compensation deconvolution was originally proposed by Riad and Stafford [11] and introduced into multiexponential signal analysis by Salami and Sidek [12]. To start with, y(t) in equation (14) is sampled at the rate of $1/\Delta t$ Hz, resulting in the following discrete convolution model

$$y[n] = \sum_{m=-n_{min}}^{n_{max}} x[m]h[n-m] + v[n] \qquad (31)$$

where $N = n_{max} - n_{min} + 1$ and it represents the total number of samples for both $y[n]$ and $h[n]$. n_{max} and n_{min} represent, respectively, the upper and lower data cut-off points.

Taking the discrete Fourier transform (DFT) of (31) yields

$$Y(k) = H(k)X(k) + V(k) \qquad (32)$$

from which the deconvolved data can be generated according to

$$\hat{X}(k) = \frac{Y(k)}{H(k)} = X(k) + \frac{V(k)}{H(k)} \qquad (33)$$

for $0 \le k \le N-1$.

where Y(k), X(k), H(k) and V(k) represent respectively the DFT of y(n), x(n), h(n) and v(n).

Optimal compensation deconvolution essentially aims at modifying the system function H(k) for use in Eq. (33) so that $\hat{X}(k)$ is as close to $X(k)$ as possible for reasonably large values of k. If the compensated system model is denoted as $H_0(k)$, then $X(k) = Y(k)H_0(k)$ so that the error energy

$$E_x = \sum_{k=0}^{N_0} \left| X(k) - \hat{X}(k) \right|^2 \qquad (34)$$

is minimized for $N_0 \le N$; N_0 represents the maximum data point of interest. This minimization conditions yields $H_0 = \{h(k)\}^{-1}$ which results in an unbounded value for large values of k. To limit the high-frequency components, both $H_0(k)$ and $X(k)$ are constrained to be bounded. The second design requirement is thus E_b should be a finite energy, where

$$E_b = \sum_{k=0}^{N_0} \left| X(k)H(k) \right|^2 \qquad (35)$$

Thus, the problem of obtaining good estimate of $X(k)$ is equivalent to that of minimizing E_x while keeping E_b finite. Combining these two requirements lead to

$$E = E_x + \mu E_b, \qquad \mu > 0 \qquad (36)$$

where μ is the optimization parameter. If E is minimized then

$$H_0(k) = \frac{H^*(k)}{|H(k)|^2 + \mu} \qquad (37)$$

where the symbol $*$ denotes complex conjugate. This equation represents a compensated system function for performing inverse filtering. Combining it and Eq. (33) leads to

$$\hat{X}(k) = \frac{Y(k)H^*(k)}{|H(k)|^2 + \mu} = Y(k)D(k) \qquad (38)$$

For high SNR, μ should be small and of the same order of magnitude as the attenuation $H(k)$ at the cut-off frequency. However, as the SNR of the data decreases the choice of the optimum value of μ, for use in (38) is best determined by experimental testing.

The estimated input distribution in equation (22) is computed by Fourier transformation which gives

$$\hat{X}(k) = \sum_{i=1}^{M} B_i e^{j\frac{2\pi k}{N}\ln \lambda_i} + \varepsilon(k) \qquad (39)$$

for $k = 1,2,....,N$; $\varepsilon(k)$ is the new additive nonstationary noise.

As the spectral estimates resulting from the application of this deconvolution procedure is not satisfactory, Salami and Sidek used autoregressive moving average (ARMA) parametric model as described later.

E. Homomorphic Deconvolution

This modification recently proposed by Jibia et al [7] consists of a six-step procedure to extract the input signal from the discrete deconvolution model resulting from Gardner transform. The procedure as outlined in that paper is as follows:

- Convert the input signal y[n] to minimum phase by exponential weighting.
- Compute the cepstrum.
- Convert the real cepstrum to complex cepstrum.
- Apply the appropriate cepstral lifter.
- Compute the inverse complex cepstrum.
- Perform exponential unweighting.

To start with, the discrete convolution model of Eq. (31) is converted to a minimum phase expression. This is achieved by application of the concept of exponential weighting which involves the multiplication of the discrete convolution model by a real number ξ^n:

$$y'[n] = \xi^n y[n] \qquad 0 < \xi < 1 \qquad (40)$$

Relationship between the cepstrum and complex cepstrum is then obtained by noting that, if y[n] is real, then $\log|Y(e^{j\omega})|$ is real and even function of ω. This implies that

$$c_y'[n] = \frac{\hat{y}'[n] + \hat{y}'[-n]}{2} \qquad (41)$$

In other words, the cepstrum is the even part of the complex cepstrum.

The complex cepstrum can thus be expressed as:

$$\hat{y}[n] = c_y[n]u_+[n] \tag{42}$$

where

$$u_+(n) = \begin{cases} 1, & n=0 \\ 2, & n>0 \\ 0, & n<0 \end{cases} \tag{43}$$

When DFT is used instead of DTFT, the cepstrum is time-aliased and given by

$$c_p(n) = \sum_{k=-\infty}^{\infty} c_y(n+kN) \tag{44}$$

To compute the complex cepstrum, we can then write

$$\hat{y}'(n) = \begin{cases} c_p(n) \ for \ n=0 \\ 2c_p(n) \ for 1 \le n \le N_0 \\ 0 \ for \ N_0 +1 \le n \le N-1 \end{cases} \tag{45}$$

N_0 is the cepstrum cut-off point which must be carefully selected to get good results. The choice of N_0 is critical to the performance of this method.

Next, homomorphic filtering is achieved by forming a modified complex cepstrum

$$\hat{y}_l = l[n]\hat{y}'[n] \tag{46}$$

where $l[n]$ is a window (or lifter) which selects a portion of the complex cepstrum for inverse processing. This is achieved as follows:

For $n \ge 1$

$$q(1) = \hat{y}'(1)$$
$$q(n) = \gamma\hat{y}'(n) + (1-\gamma)q(n-1) \tag{47}$$

and for $n \le N_0$

$$w(N_0) = \hat{y}'(N_0)$$
$$w(n) = \gamma\hat{y}'(n) + (1-\gamma)w(n+1) \tag{48}$$

so that the output variable would be

$$y_l[n] = 0.5(q(n)+w(n)) \tag{49}$$

γ is the variable filter coefficient such that $0 \le \gamma \le 1$.

Taking the forward DFT of $\hat{y}_l[n]$ followed by exponentiation, inverse DFT and exponential unweighting yields $x[n]$ with dominant peaks at $\ln \lambda_i$, $i = 1,2,.........,M$.

The main advantage of this technique is computational simplicity and elimination of data length constraint encountered in [12]. This method also successfully detects the number of components, M. However, its accuracy in determining the decay rates is very poor.

V. MODIFICATION BY ADDITIONAL PROCESSING OF THE DECONVOLVED DATA

A. Stationarity of the Deconvolved Data

Application of Gardner transform complicates the analysis and interpretation of $\varepsilon(k)$. Even if $n(\tau)$ is derived from a stationary white noise process, there is no guarantee that the noise $\varepsilon(k)$ will be equally stationary. The division of $V(k)$ by $H(k)$ make $\hat{X}(k)$ contain a nonstationary noise component that may complicate analysis. To make the result of any mathematical operation on $\hat{X}(k)$ meaningful, $\hat{X}(k)$ must be stationary. It is noteworthy that while $\hat{X}(k)$ may be nonstationary, the desired information data $X(k)$ is stationary and deterministic. Consequently, simple procedures for preprocessing nonstationary sequences can be used to convert $\hat{X}(k)$ into a locally stationary sequence. Some of the techniques are windowing, prefiltering and differencing [13].

For the stationarization of $\hat{X}(k)$ a rectangular window is used here as it weighs the data uniformly and has a very good spectral resolution at the expense of high variance. Consider a symmetric window, W_k, of length $N_d = 2N_0 +1$, so that

$$W_k = \begin{cases} 1, & |k| \le N_0 \\ 0, & otherwise \end{cases} \tag{50}$$

then $\hat{x}(k)$ is computed from $\hat{X}(k)$ according to the relation

$$\hat{x}(k) = W_k\hat{X}(k) \tag{51}$$

It is well-known from the properties of the DFT that a shift of $\hat{X}(k)$ does not destroy its spectral characteristics. Therefore $\hat{X}(k)$ is shifted through the length N_0 so as to derive the complex sequence $\hat{x}(k)$, where

$$\hat{x}(k) = \hat{X}(k-N_0), \qquad 1 \le k \le 2N_0 +1 \tag{52}$$

Eq. (39) thus becomes

$$\hat{X}(k) = \sum_{i=1}^{M} B_i e^{j\frac{2\pi k}{N_d}\ln \lambda_i} + e(k) \tag{53}$$

with $k = 1,2,....,N$. $e(k)$ is the new stationary white noise.

The spectral estimates of $\hat{x}(k)$ would therefore provide the values M, A_i and λ_i that characterize the transient multiexponential signals.

B. ARMA Modeling Technique

This method was used by Salami and Sidek [12] to model the output of the optimal compensation deconvolution filter. The deconvolved data is first windowed using an appropriate length N_0 essentially determined by trial and adjustment depending on the signal resolution, the number of components and the SNR. The truncated data is then modelled as the output of an ARMA model whose input is a

complex white noise $e(k)$ so that

(54)

$$\sum_{n=0}^{p} a_n x(k-n) = \sum_{n=0}^{q} b_n e(k-n), \qquad a_0 = 1 \qquad (54)$$

where a_n and b_n represent respectively the AR and moving average (MA) model coefficients and p and q are AR and MA model orders respectively. The above equation is then multiplied by $x^*(k-m)$, the expectation of the resulting expression is

$$R_{dd}(k) = -\sum_{n=1}^{p} a(n) R_{dd}(k-n)$$

$$+ \sum_{n=0}^{q} b(n) h(k-m) \qquad (55)$$

where $R_{dd}(k)$ is the autocorrelation function of $x(k)$ and $h(k)$ is the impulse response function of the ARMA model. Next, the AR portion of equation (55) leads to the modified Yule-Walker equation

$$R_{dd}(k) = \sum_{n=1}^{p} a(n) R_{dd}(k-n) = 0; \quad k \geq q+1 \quad (56)$$

Note that (56) may not hold exactly in practice because both p and q are unknown prior to analysis and $R_{dd}(k)$ has to be estimated from noisy data. This problem is solved using an SVD algorithm. This algorithm provides consistent and accurate estimates of the AR parameters with minimal numerical problem. Furthermore, the SVD algorithm is a powerful computational procedure for matrix analysis especially for solving overdetermined system of equations. Equation (56) can be expressed in matrix form as

$$Ra = e \qquad (57)$$

with R having elements $r(i,l) = R_{dd}(q_e + 1 + i - l)$, where $1 \leq i \leq t; \quad 1 \leq l \leq p_e + 1$. Note that both p_e and q_e are the guess values of the AR and MA model order respectively, and e is an $t \times 1$ error vector with $t > p_e$.

The singular value decomposition of R yields

$$R = U\Sigma V^H \qquad (58)$$

where U is an $r \times (p_e + 1)$ unitary matrix, V is a $(p_e + 1) \times (p_e + 1)$ unitary matrix and $\Sigma = diag(\sigma_1, \sigma_2, \ldots, \sigma_{p_e+1})$ is a diagonal matrix whose elements are ordered as $\sigma_1 > \sigma_2 > \ldots > \sigma_{p_e+1} > 0$. H is a complex conjugate and transpose operator. For high SNR signals only the largest M singular values will be nonzero so that $\sigma_{M+1} = \sigma_{M+2} = \ldots = \sigma_{p_e+1} = 0$. However, for noisy signals, $\sigma_{M+1} \neq \sigma_{M+2} \neq \ldots \neq \sigma_{p_e+1} \neq 0$. The problem is solved by constructing a lower rank matrix R_L from R using the first singular values. That is,

$$R_L = U_M \Sigma_M V_M^H = \sum_{n=1}^{M} \sigma_n u_n v_n^H \qquad (59)$$

where U_M, Σ_M and V_M are the truncated versions of U, Σ and V respectively. The AR parameters are then estimated by the relation $a = -R_L^+ r$, where **r** corresponds to the first column of R_L and R_L^+ is the Moore-Penrose generalized inverse given as

$$R^+ = V_M \begin{pmatrix} \Sigma_M & 0 \\ 0 & 0 \end{pmatrix} U_M^H \qquad \begin{matrix}(60)\\(60)\end{matrix}$$

Next, the estimated AR parameters are used to generate the residual error sequences

$$\beta(k) = \sum_{l=0}^{p_e} \sum_{m=0}^{p_e} a[l] a^*[m] R_{dd}(k+m-l) \qquad (61)$$

from which the actual MA parameters are obtained. An exponential window is applied to $\beta(k)$ to ensure that the MA spectra derived from the error sequences are positive definite. Next, the ARMA spectrum is computed from

$$S_f(z) = \frac{\sum_{k=-p_e}^{p_e} \beta(k) z^{-k}}{|A(z)|^2} \qquad (62)$$

and the desired power distribution of $x(t)$, denoted as $P_x(t)$ is obtained by evaluating $S_f(z)$ on the unit circle $z = \exp\left(\dfrac{j2\pi t}{N\Delta t}\right)$; that is

$$P_x(t) = S_f(z)\Big|_{z=\exp\left(\frac{j2\pi t}{N\Delta t}\right)} = \sum_{k=1}^{M} B_k^2 \delta(t - \ln \lambda_k) \quad (63)$$

Eventually, M and $\ln \lambda_i$ are obtained from $P_x(t)$.

The problem with this method is that the procedure for computing the MA coefficient is very complicated and it requires a too high model order before desirable results are obtained.

C. Transient Error Method

The transient error method [15] is based on the premise that each data point can be expressed as a linear combination of M previous ones according to:

$$\hat{x}_k = \sum_{i=1}^{M} a_i \hat{x}_{k-i}, \quad M \leq k \leq 2N_0 \qquad (64)$$

where a_i are the AR coefficients.

Denoting the output of the AR filter as e_k and knowing that $\hat{x}_k$ is equal to zero for $k < 0$, then the output of transient values $e_0, e_1, \ldots \ldots e_{M-1}$ will be nonzero. These error coefficients $\{e_k\}$ correspond to the MA coefficients and are used to enhance resolvability of the exponents. From the preceding statement and equation (64), it follows that

$$e_k = \sum_{i=0}^{M} a_i \hat{x}_{k-i} \qquad (65)$$

Taking the z-transform of (65) and solving for $\hat{X}(z)$ gives the ARMA model

$$\hat{X}(z) = \frac{E(z)}{A(z)}$$

The distribution, $\hat{x}(t)$ is then computed from

$$\hat{x}(t) = Z^{-1}\left[\frac{E(z)}{A(z)}\right] \qquad (66)$$

The remaining part of the procedure is similar to the ARMA modeling.

The major problem with the transient error method is that it requires M to be specified and does not give good estimates of λ_i for relatively large values of the variance of the error sequence. It also has a very high detection SNR threshold.

D. Data Extension Method

The M complex sinusoids resulting from the application of Gardner transform and the accompanying deconvolution (see Chapter three) can be represented by the Mth order difference equation

$$x_k = \sum_{i=0}^{M} a_i x_{k-i} = 0 \; a_0 = 1 \qquad (67)$$

where the parameters $\{a_i, i = 1,2,...., M\}$ are complex coefficients.

Equation (67) forms the basis of the data extension method that was developed by Arunachalam [14] and further developed by Smith and Nichols [16]. In this method, the value of x_k is extended beyond the $2N_0$ point by a recursive application of the unit step prediction operator determined by the Burg algorithm. This technique is justified because $x(t)$ is made up of a train of impulses of different amplitudes. The Fourier transform of an impulse has an infinite bandwidth so that if a good value of x_k can be established, then extrapolation can be applied to extend its range from $2N_0 + 1$ to N_1, where N_1 is the length of the extrapolated deconvolved data. Smith and Nichols [16] extended this procedure by applying appropriate weighting function to the extrapolated data. Both methods work well only if the data is noiseless and their algorithms do not give a criterion for selecting the good portion which is to be extrapolated. Also their algorithms may give an incorrect value of M since its determination is based on the minimum residual error energy.

E. Eigenvector Methods

MUSIC (multiple signal classification) and minimum norm algorithms were separately proposed by Schmidt [17] and Kumerasan and Tufts [18] respectively for the analysis of complex exponentials. These methods were used by Jibia and Salami [19] to model the output of optimal compensation deconvolution filter, essentially a sum of complex exponentials as in equation (39). These methods belong to and form the core of what is known as eigenvector algorithms.

The signal in Eq. (53) can be expressed in compact form as follows

$$x = Sb + e \qquad (68)$$

where

$$b = \begin{bmatrix} B_1 & B_2 ... B_M \end{bmatrix}^T \qquad (69)$$

$$S = \begin{bmatrix} s_1 & s_2 ... s_M \end{bmatrix} \qquad (70)$$

$$s_i = \begin{bmatrix} e^{\frac{j2\pi \ln \lambda_i}{N_d}} & e^{\frac{j4\pi \ln \lambda_i}{N_d}}, ..., e^{j2\pi \ln \lambda_i} \end{bmatrix} \qquad (71)$$

$N_d = 2N_0 + 1$ is the number of samples after truncation of the deconvolved data, λ_i is the decay rate of the ith component, B_i is the amplitude of the ith component and e is the noise after stationarization.

The autocorrelation matrix of the noisy signal can be written as

$$R_{xx} = E\left[xx^H\right] = R_{ss} + R_{nn} = SBS^H + \sigma^2 I \qquad (72)$$

where E denotes the expectation, H denotes the Hemittian transpose and $B = \begin{bmatrix} bb^H \end{bmatrix}$ is the diagonal matrix. Further, $R_{ss} = SBS^H$ and $R_{nn} = \sigma^2 I$ are correlation matrices of the signal and noise processes respectively. This can be expressed as

$$R_{ss} = \sum_{k=1}^{N_d} \lambda_k v_k v_k^H \qquad (73)$$

$$R_{nn} = \sigma^2 \sum_{k=1}^{N_d} v_k v_k^H \qquad (74)$$

where λ_k and v_k are the eigenvalues and eigenvectors of the matrix R_{ss}.

Thus, the autocorrelation of the noisy signal is

$$R_{xx} = \sum_{k=1}^{N_d} \lambda_k v_k v_k^H + \sigma^2 \sum_{k=1}^{N_d} v_k v_k^H = \sum_{k=1}^{N_d} \mu_k v_k v_k^H \qquad (75)$$

where $\mu_k = \lambda_k + \sigma^2$ are the eigenvalues of the matrix R_{xx} and are real.

All eigenvectors satisfy

$$\mu_1 \geq \mu_2 \geq ... \geq \mu_M \geq \mu_{M+1}$$
$$\geq \mu_{M+2} = ... = \mu_{N_d} = \sigma^2 \qquad (76)$$

The MUSIC pseudospectrum is expressed as

$$P_{MUSIC} = \frac{1}{\sum_{k=M+1}^{N_d} \left| s^H (\ln \lambda) v_k \right|^2} \qquad (77)$$

$$= \frac{1}{s^H (\ln \lambda) VV^H s(\ln \lambda)}$$

where $s(\ln \lambda)$ is the complex sinusoidal vector and $V = \begin{bmatrix} v_{M+1} ... v_{N_d} \end{bmatrix}$ is the matrix of eigenvalues of the noise space.

This is the MUSIC pseudospectrum. The term pseudospectrum is actually a misnomer since the domain is not frequency but $\ln \lambda$.

Strictly, Eq. (77) is valid if and only if the noise is white or assumed to be so. But due to Gardner transform, the signal has gone through many stages to arrive at $\varepsilon(k)$ for the noise. Although truncation is supposed to stationarize the deconvolved data and noise, the eigenvalues of the noise in Eq. (53) may still not be equal. Thus, a modification is necessary to account for the variation in eigenvalues. This is achieved by normalizing the pseudospectrum of each eigenvector by its corresponding eigenvalue.

$$P_{MUSIC} = \frac{1}{\sum_{k=M+1}^{N_d} \frac{1}{\mu_k} \left| s^H (\ln \lambda) v_k \right|^2} \qquad (78)$$

where μ_k is the eigenvalue associated with the eigenvector v_k.

The minimum norm is essentially a variant of MUSIC. The pseudospectrum (Eq.(77)) tests the orthogonality of the signal vector with respect to all the eigenvectors in the noise subspace. The minimum norm method uses only one arbitrary vector

$$d = \begin{bmatrix} d(1) & d(2) \ldots \ldots d(N_d) \end{bmatrix} \qquad (79)$$

constrained to lie in the noise subspace.

The minimum norm seeks to minimize the norm of d in order to avoid spurious peaks in the pseudospectrum. The norm of the vector contained in the noise subspace is

$$\|d\|^2 = d^H d = d^H P_n d \qquad (80)$$

where P_n is the matrix that projects an arbitrary vector on the noise subspace [22].

Since an unconstrained minimization of this norm will produce the zero vector, the first element of d is constrained to be unity, i.e.

$$\delta_1^H d = 1 \qquad (81)$$

The solution to this can be found by using Lagrange multipliers [23] as

$$d_{mn} = \frac{P_n \delta_1}{\delta_1^H P_n \delta_1} \qquad (82)$$

The $\ln \lambda$ are then obtained from the peaks in the pseudospectrum of the minimum norm vector

$$P_{mn} = \frac{1}{\left| s^H (\ln \lambda) d_{mn} \right|^2} \qquad (83)$$

Research on improving the results of MUSIC and minimum norm is still underway and it is hoped that they will provide an excellent solution to the problem of multiexponential analysis. Some of the options being considered include alternative deconvolution procedures, integrating them with the homomorphic deconvolution method and the use of ESPRIT algorithm [24].

VI. Conclusion

In this paper, an appraisal of the Garner transform method along with its modifications have been given. The modifications have been grouped into three. Some modifications are in the nonlinear transformation, others are in the accompanying deconvolution procedure and the rest involve parametric modeling of the deconvolved data. The use of parametric models improves the result of Gardner transform but introduces data length truncation which makes them difficult to use for real time analysis.

Research is quite active in improving the performance of Gardner transform-based methods. Current research is aimed at providing improved estimation accuracy over very low SNRs for signals with large number of components and high resolution. Specifically, these researchers are currently looking at the possibility of an automated procedure for the selection of data truncation point.

Another effort is on the way to design a better cepstral lifter for the homomorphic deconvolution method. It is hoped that a better windowing lifter will give the desired solution for the real-time analysis of multiexponential signals using Gardner transform-based techniques. Finally, as it has been pointed out, research is also active on improving the performance of eigenvector algorithms in the analysis of transient multiexponential data.

References

[1] D. G. Gardner, J. C. Gardner, G. Lush, and W.R. Ware, "Method for the analysis of multicomponent exponential decay curves," *Journ. Chem. Phys.*, vol. 31, 1959, pp. 978-986.

[2] A. A. Istratov and O. F. Vyvenko, "Exponential analysis in physical phenomena," *Rev. Sci. Instruments.*, vol. 70 no. 2, pp. 1233-1257, 1999.

[3] J. Schlesinger, "Fit to experimental data with exponential functions using the fast Fourier transform," *Nuclear instruments and methods*, vol. 106, no. 3 pp. 503-508, 1973.

[4] D. N. Swingler, "A differential technique for the Fourier transform processing of multicomponent exponential functions," *IEEE transactions on biomedical engineering*, vol. 24 no. 4, pp. 408-410, 1977.

[5] S. T. Nichols, M. R. Smith, and M. J. E. Salami, "High resolution estimates in multicomponent signal analysis," Tech. Report, Dept. Elec. Engg., University of Calgary, Alta, Sept., 1983.

[6] M. J. E. Salami, "ARMA models in multicomponent signal analysis," Ph.D. Dissertation, Dept. of Electr. Engng., University of Calgary, Calgary, Canada, 1985.

[7] A. U. Jibia, M. J. E. Salami, O. O. Khalifa, and A. M. Aibinu, "Analysis of transient multiexponential signals using cepstral deconvolution," *Applied and Computational Harmonic Analysis*, vol. 29, no. 1, pp. 88-96, 2010.

[8] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex fourier series," *Math. Computat.*, vol. 19, pp. 297-301, 1965.

[9] S. Cohn-Sfetcu, M. R. Smith, and S. T. Nichols, "A digital technique for analyzing a class of multicomponent signals," *Proceedings of IEEE* vol. 63, pp. 1460-1467, 1975.

[10] S. W. Provencher, "A Fourier method for the analysis of exponential decay curves," *Biophysical journal*, vol. 16 no. 1 pp. 27-39, 1976.

[11] S. M. Riad, R.B. Stafford, "Impulse response evaluation using frequency domain optimal compensation deconvolution," P*roc. 23rd Midwest symposium on circuits and systems*, pp. 521-525, 1980.

[12] M. J. E. Salami and S. N. Sidek, "Parameter estimation of multicomponent transient signals using deconvolution and ARMA modeling techniques," *Journal of Mechanical systems and signal processing*, vol. 17, no. 6, pp. 1201 - 1218, 2003.

[13] J. D. Hamilton, Time series analysis, Printiceton University Press, New Jersey, 1994.

[14] V. Arunachalam, "Multicomponent signal analysis," PhD dissertation, Dept. of Electrical Engineering, The University of Calgary, Alberta, Canada, 1980.

[15] M. J. E. Salami, S. T. Nichols, and M. R. Smith, "SVD-based transient error method for analyzing noisy multicomponent exponential signals," *Proc. International Conference of Acoustics, Speech and Signal Processing* (ICASSP 1987), pp. 677 680, 1987.

[16] M. R. Smith and S.T. Nichols, "Improved resolution in the analysis of multicomponent exponential signals," *Nuclear Instr. Methods*, vol. 205 no 3 pp. 479-483, 1983.

[17] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas and Propagation*, vol. 34 276-290, 1986.

[18] R. Kumerasan and D. W. Tufts, "Estimating the angles of arrival of multiple plane waves," *IEEE Transactions on Aerospace and electronic Systems*, vol. 19, pp. 134-139, 1983.

[19] A. U. Jibia and M. J. E., Salami, "Performance evaluation of MUSIC and minimum norm eigenvector algorithms in resolving noisy multiexponential signals," *International Journal of Computer Science*, Vol. 2 no 4, pp. 235- 239, 2007.

[20] D. H. Johnson and S. R. DeGraaf, "Improving the resolution of bearing in passive sonar arrays," *IEEE Trans. Acoustics, Speech and Signal Processing,* vol. 30 no. 4, pp. 638-647, 1982.

[21] A. U. Jibia, M. J. E. Salami, and O. O. Khalifa, "Analysis of multicomponent transient signals using music superresolution technique," *Proc.2nd International Conference on Computer and Communication Engineering* (ICCCE08), Kuala Lumpur, pp. 104-108, 2008.

[22] M. H. Hayes, "Statistical digital signal processing and modeling," John Wiley and Sons, Inc., New York, 1996.

[23] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, "Statistical and adaptive signal processing," Artech House, Inc., Norwood, 2005.

[24] R. Roy and T. Kailath, "ESPRIT - estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37 no. 7, pp. 984 - 995, 1989.