be suitable for sub-Gaussian sources, is inconsistent for some of such sources (wherein $3.3567 < a < 3.7352$). With the only exception of $\varphi'(s) = |s|^3\text{sign}(s)$, whose consistency is always determined by the sign of kurtosis excess, other QML estimators require more delicate analysis in order to determine whether they are suitable or not for estimation of super- or sub-Gaussian sources. Generally, the answer is distribution-dependent. The main conclusion from this correspondence is that the nonlinear functions $\varphi_i(s)$ should be chosen with more caution.

ACKNOWLEDGMENT

The authors are grateful to the anonymous referees for their constructive criticism and valuable comments.

REFERENCES

[1] S.-I. Amari, A. Cichocki, and H. H. Yang, "Novel online adaptive learning algorithms for blind deconvolution using the natural gradient approach," in *Proc. SYSID*, Jul. 1997, pp. 1057–1062.

[2] S.-I. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," in *Proc. SPAWC*, Apr. 1997, pp. 101–104.

[3] E. Moulines, J.-F. Cardoso, and E. Gassiat, "Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5, Apr. 1997, pp. 3617–3620.

[4] D. Pham and P. Garrat, "Blind separation of a mixture of independent sources through a quasi-maximum likelihood approach," *IEEE Trans. Signal Process.*, vol. 45, no. 7, pp. 1712–1725, Jul. 1997.

[5] M. Zibulevsky, B. A. Pearlmutter, P. Bofill, and P. Kisilev, "Blind source separation by sparse decomposition," in *Independent Components Analysis: Principles and Practice*, S. J. Roberts and R. M. Everson, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2001.

[6] M. Zibulevsky, P. Kisilev, Y. Y. Zeevi, and B. A. Pearlmutter, "Blind source separation via multinode sparse representation," in *Proc. NIPS*, 2002.

[7] P. Kisilev, M. Zibulevsky, and Y. Zeevi, "Multiscale framework for blind source separation of linearly mixed signals," *J. Machine Learning Res.*, vol. 4, no. 7, 8, pp. 1339–1364, Oct.–Nov. 2004.

[8] A. M. Bronstein, M. M. Bronstein, and M. Zibulevsky. (2003, Oct.) Blind Deconvolution with Relative Newton Method. Technion, Haifa, Israel. [Online]. Available: http://visl.technion.ac.il/bron/alex

[9] ——, "Relative optimization for blind deconvolution," *IEEE Trans. Signal Process.*, vol. 53, no. 6, pp. 2018–2026, Jun. 2005.

[10] A. M. Bronstein, M. Bronstein, M. Zibulevsky, and Y. Y. Zeevi. (2003) Quasi-Maximum Likelihood Blind Deconvolution of Images Using Sparse Representations. Technion, Haifa, Israel. [Online]. Available: http://visl.technion.ac.il/bron/alex

[11] A. M. Bronstein, M. Bronstein, M. Zibulevsky, and Y. Y. Zeevi, "Blind deconvolution of images using optimal sparse representations," *IEEE Trans. Image Process.*, vol. 14, no. 6, pp. 726–736, Jun. 2005.

[12] H. Mathis, M. Joho, and S. Moschytz, "A simple threshold nonlinearity for blind source separation of sub-gaussian signals," in *Proc. ISCAS*, 2000, pp. 489–492.

[13] M. Joho and P. Schniter, "Frequency domain realization of a multichannel blind deconvolution algorithm based on the natural gradient," in *Proc. ICA*, Apr. 2003.

[14] A. M. Bronstein, M. M. Bronstein, M. Zibulevsky, and Y. Y. Zeevi. (2004, Jan.) Asymptotic Performance Analysis of MIMO Blind Deconvolution. Technion, Haifa, Israel. [Online]. Available: http://visl.technion.ac.il/bron/alex

[15] S.-I. Amari, T.-P. Chen, and A. Cichocki, "Stability analysis of learning algorithms for blind source separation," *Neural Networks*, vol. 10, no. 8, pp. 1345–1351, 1997.

[16] J.-F. Cardoso, "Blind signal separation statistical principles," in *Proc. IEEE; Special Issue on Blind Source Separation*, vol. 9, Oct. 1998, pp. 2009–2025.

[17] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 6, pp. 1081–1089, Sep. 1986.

[18] A. Gorokhov and J.-F. Cardoso, "Equivariant blind deconvolution of MIMO-FIR channels," in *Proc. SPAWC*, 1997, pp. 489–492.

[19] P. Stoica, J. Li, and T. Soederstroem, "On the inconsistency of IQML," *Signal Process.*, vol. 56, no. 2, pp. 185–190, 1997.

[20] J.-F. Cardoso, "On the stability of source separation algorithms," in *J. VLSI Signal Process. Syst.*, vol. 26, 2000, pp. 7–14.

On the Resolvability of Sinusoids With Nearby Frequencies in the Presence of Noise

Morteza Shahram and Peyman Milanfar

***Abstract*—This correspondence develops statistical algorithms and performance limits for resolving sinusoids with nearby frequencies in the presence of noise. We address the problem of distinguishing whether the received signal is a single-frequency sinusoid or a double-frequency sinusoid, with possibly unequal, and unknown, amplitudes and phases. We derive a locally optimal detection strategy that can be applied in a standalone fashion or as a refinement step for existing spectral estimation methods to yield improved performance. We further derive explicit relationships between the minimum detectable difference between the frequencies of two tones for any particular false alarm and detection rate and at a given SNR.**

***Index Terms*—Kullback–Leibler distance, MUSIC, Rayleigh limit resolution, resolution detection, spectral estimation, subspace.**

I. INTRODUCTION

Spectral estimation has a long history and significant applications in signal processing. In many areas of application, including the vast body of knowledge in array processing [1], resolving sinusoidal signals with nearby frequencies has been of special interest. In particular, the problem in array signal processing arises in several contexts, including direction-of-arrival estimation, when two incoherent plane waves are incident on a linear equispaced array of sensors [2]. In the past, the vast majority of the techniques in this area have been based on matrix decomposition methods. Notable instances of the relevant literature are found in [2]–[7].

These approaches are based principally on second-order statistical analysis, which relies on the covariance structure of the measured signal. Extensive work has been done to determine the performance of such methods [2], [8]–[15].

A common question in this area has been to investigate the relationship between resolution and SNR. Nearly all papers that have addressed this question, either directly or in a related framework, have been focused on the celebrated MUSIC algorithm [16] or its variants (e.g., root MUSIC [17]). The earliest related work was done by Kaveh and Barabell [2] to determine the (minimum) threshold SNR required to resolve two equipowered sinusoids in the asymptotic regime. In the context of array processing, recent work has employed Cramér–Rao bound analysis to investigate the relationship between resolvability and

Manuscript received August 13, 2003; revised November 10, 2004. This work was supported in part by the National Science Foundation under CAREER Award CCR-9984246 and the Air Force Office of Scientific Research under Grant F49620-03-1-0387. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jian Li.

The authors are with the Department of Electrical Engineering, University of California, Santa Cruz, CA 95064 USA (e-mail: shahram@ee.ucsc.edu; milanfar@ee.ucsc.edu).

Digital Object Identifier 10.1109/TSP.2005.845492

signal-to-noise ratio (SNR) [18]–[20]. While the methods employed are somewhat different, the results obtained are consistent both with our earlier work on establishing detection and estimation-theoretic bounds for resolution in imaging systems [21]–[23] and with the results reported in this paper.

Without being limited to subspace methods or to asymptotic regimes, a relatively similar question interests us in this paper. We employ a local model-based hypothesis-testing approach to determine the limits to the resolution of frequencies of nearby tones in signals measured in the presence of noise and over short observation intervals. Our approach is to precisely define a quantitative measure of resolution in statistical terms by addressing the following question: "What is the minimum separation between two frequencies of nearby tones (maximum attainable resolution) that is detectable at a given SNR and for pre-specified probabilities of detection and false alarm (P_d and P_f)?"

As we will demonstrate, in the process of addressing the above question, the machinery of the analysis will also suggest a corresponding detection strategy that can be applied in practice. In other words, the final computed performance limit is simply the result of employing these locally, uniformly, most powerful detectors. In order to illustrate the relevance of the results, we present comparisons against the general class of subspace methods (in particular, the MUSIC algorithm), which is perhaps the most commonly used subspace-based technique in practice. We demonstrate that the proposed detectors yield significantly improved performance in distinguishing frequencies of nearby tones.

We begin by defining the signal of interest as

$$s(x;\delta_1,\delta_2) = a_1 \sin\left(2\pi(f_c-\delta_1)x+\phi_1\right) + a_2 \sin\left(2\pi(f_c+\delta_2)x+\phi_2\right) \quad (1)$$

in the range $x \in [-B/2, B/2]$, where for convenience, we consider the two frequencies $f_c-\delta_1$ and $f_c+\delta_2$ to be around a "center" frequency[1] f_c. The measured signal is a sampled and noise-corrupted version of (1) as follows:

$$f(k;\delta_1,\delta_2) = s(k;\delta_1,\delta_2) + w(k) \quad (2)$$

$$= a_1 \sin\left(2\pi\left(f_c-\delta_1\right)\frac{k}{f_s}+\phi_1\right) + a_2 \sin\left(2\pi\left(f_c+\delta_2\right)\frac{k}{f_s}+\phi_2\right) + w(k) \quad (3)$$

where the sampling frequency is f_s (Hz), which is assumed to be sufficiently high to avoid aliasing, and the integer index k is in the range $k \in \{-(N-1)/2,\ldots,(N-1)/2\}$, where $N = Bf_s$. The term $w(k)$ is assumed to be a zero-mean Gaussian white noise process with variance σ^2.

According to the so-called Rayleigh criterion [10], the two peaks in the frequency domain corresponding to $f_c-\delta_1$ and $f_c+\delta_2$ are barely resolvable if

$$\delta_1+\delta_2 = \frac{1}{B}. \quad (4)$$

In this correspondence, we are interested in studying the scenario in which the two frequency components are, in this "classical" sense, *unresolvable*. In practice, this corresponds to the situation in which the main-lobe of the Fourier transform of the (sum of) two sinusoids is located in the same FFT bin. Therefore, in this context, what we mean by "signals with short observation interval" is simply those signals in which the values of B, δ_1, and δ_2 satisfy the inequality $\delta_1+\delta_2 < 1/B$.

[1]We note that this center frequency can be assumed to be known or estimated (see Appendix B), or the detection procedure can be repeated at various candidate center frequencies.

With the above framework in place, we treat the problem of resolution by formulating a hypothesis test. In particular, the corresponding hypotheses for this problem are

$$\begin{cases} \mathcal{H}_0: & \delta_1 = 0 \text{ and } \delta_2 = 0 \\ \mathcal{H}_1: & \delta_1 > 0 \text{ or } \delta_2 > 0 \end{cases} \quad (5)$$

where $\mathcal{H}_0$ embodies the case where only one spectral component is present, whereas $\mathcal{H}_1$ captures the case where two distinct frequencies are present.[2] We note here that in this framework, we consider δ_1 and δ_2 to be unknown to the detector; therefore, this is a composite hypothesis testing problem. Our approach in this work will be to take advantage of the small separation between the frequency components to effect an approximation that will yield a detector that is locally uniformly most powerful. Notably, this analysis will enable us to explicitly compute the relationship between minimum detectable frequency separation and SNR. We should note that the methodology we present here is quite similar to an approach we have recently advocated for determining resolution limits in optical imaging [21]–[23].

The organization of the paper is as follows. In Section II, we introduce our approach by first treating the special case of known and equal amplitudes and phases for the two sinusoids. Having conveyed the basic ideas and intuition, in Section III, we present the most general case, where the amplitudes and phases are unequal and unknown to the detector. In this section, we also develop the corresponding detection strategies and characterize their performance. Section IV presents some comparisons of the proposed method with existing subspace methods. Finally, in Section V, we summarize the results and present some concluding remarks.

II. Case of Equal and Known Amplitude and Phase

To gain maximum intuition and perspective from the foregoing analysis, we first consider a simple case with the following assumptions.

- $a_1 = a_2 = 1$.
- $\phi_1 = \phi_2 = 0$.
- $\delta_1 = \delta_2 = \delta$.

In this case, the measured signal model is given by

$$f(k;\delta) = s(k;\delta) + w(k) \quad (6)$$

$$= \sin\left(2\pi(f_c-\delta)\frac{k}{f_s}\right) + \sin\left(2\pi(f_c+\delta)\frac{k}{f_s}\right) + w(k). \quad (7)$$

Since the range of interest for the values of δ is small ($\delta < 1/2B$) (these representing one wide peak in the frequency domain), it is quite appropriate for the purposes of the our analysis (even in the more general case treated in the next section) to consider approximating the model of the signal around $\delta = 0$. The Taylor expansion of (6) about $\delta = 0$, with all other variables fixed, is

$$s(k;\delta) = h_0(k) + \delta^2 h_2(k) + O(\delta^4) \quad (8)$$

where

$$h_0(k) = 2\sin\left(\frac{2\pi f_c k}{f_s}\right) \quad (9)$$

$$h_2(k) = -\frac{4\pi^2 k^2}{f_s^2}\sin\left(\frac{2\pi f_c k}{f_s}\right). \quad (10)$$

By ignoring the $O(\delta^4)$ terms in (8), the approximate measured signal model can then be written as

$$\widetilde{f}(k;\delta) = h_0(k) + \delta^2 h_2(k) + w(k). \quad (11)$$

It is worth noting that in the above approximation, we elect to make explicit use of terms up to order 2 of the Taylor series since *no* linear

[2]Note that the hypothesis test in (5) is a one-sided test.

term in δ appears in the approximation. By neglecting higher order terms ($O(\delta^4)$), the hypotheses in vector form will be

$$\begin{cases} \mathcal{H}_0: & \widetilde{\mathbf{f}} = \mathbf{h}_0 + \mathbf{w} \\ \mathcal{H}_1: & \widetilde{\mathbf{f}} = \mathbf{h}_0 + \delta^2 \mathbf{h}_2 + \mathbf{w} \end{cases} \tag{12}$$

where

$$\widetilde{\mathbf{f}} = \left[\widetilde{f}\left(-\frac{N-1}{2}\right), \ldots, \widetilde{f}\left(\frac{N-1}{2}\right) \right]^T \tag{13}$$

and $\mathbf{h}_0, \mathbf{h}_2$, and $\mathbf{w}$ are similarly defined. Since $\mathbf{h}_0$ is a common (known) term in both hypotheses, we may simplify further:

$$\begin{cases} \mathcal{H}_0: & \mathbf{y} = \mathbf{w} \\ \mathcal{H}_1: & \mathbf{y} = \delta^2 \mathbf{h}_2 + \mathbf{w} \end{cases} \tag{14}$$

where we have defined $\mathbf{y} = \tilde{\mathbf{f}} - \mathbf{h}_0$, and the parameter δ is unknown. This is a problem of detecting a deterministic signal with an unknown parameter (δ^2). The general structure of composite hypothesis testing is involved when unknown parameters appear in the hypotheses [25, p. 186]. The Generalized Likelihood Ratio Test (GLRT) is a well-known approach to solving these types of problems. The GLRT uses the maximum likelihood (ML) estimates of the unknown parameters to form the standard Neyman–Pearson (NP) likelihood ratio detector. The (unconstrained) ML estimate for the parameter δ^2 is given by [24]

$$\widehat{\delta^2} = (\mathbf{h}_2^T \mathbf{h}_2)^{-1} \mathbf{h}_2^T \mathbf{y} \tag{15}$$

which leads to the following GLRT detector:

$$T(\mathbf{y}) = \frac{\widehat{\delta^2}^2}{\sigma^2} \mathbf{h}_2^T \mathbf{h}_2 = \frac{1}{\sigma^2} (\mathbf{h}_2^T \mathbf{h}_2)^{-1} \left(\mathbf{h}_2^T \mathbf{y} \right)^2 \tag{16}$$

where we decide $\mathcal{H}_1$ if the statistic exceeds a specified threshold $T(\mathbf{y}) > \gamma_1$. It is worth noting that $T(\mathbf{y})$ is in fact a quadratic form in a rank-one projection. While it may seem troublesome to use the unconstrained ML estimate to form the GLRT, in fact, due to the (known) positivity of δ^2, the detector structure is effectively producing a one-sided test and, hence, is in fact a Uniformly Most Powerful (UMP) detector [25, p. 194], [26, p. 124]. The detector can therefore be described simply as a normalized matched filter ($\mathbf{h}_2^T \mathbf{y}$), giving the best detection rate for a given false alarm rate and for all small values of δ^2. Hence, we can write

$$T'(\mathbf{y}) = \sqrt{T(\mathbf{y})} = \sqrt{\frac{1}{\sigma^2} (\mathbf{h}_2^T \mathbf{h}_2)^{-1}} \left(\mathbf{h}_2^T \mathbf{y} \right). \tag{17}$$

For any given data set $\mathbf{y}$, we decide $\mathcal{H}_1$ if the statistic exceeds a specified threshold[3]

$$T'(\mathbf{y}) > \gamma. \tag{18}$$

The choice of γ is motivated by the level of tolerable false alarm (or false-positive) in a given problem, but is typically kept very low. For this matched filter structure, the detection rate (P_d) and false-alarm rate (P_f) are related as

$$Q(P_d) = Q(\delta^2 \eta + \gamma) = Q\left(\delta^2 \eta + Q^{-1}(P_f) \right) \tag{19}$$

[3]Due to the known positivity of δ^2, the threshold (γ) should be positive. In fact, as one reviewer suggested, another way of writing the GLRT is $\max\{\widehat{\delta^2}, 0\}(\mathbf{h}_2^T \mathbf{h}_2/\sigma^2) > \gamma_2$. This will result in deciding $\mathcal{H}_0$ for any negative estimate of δ^2 (i.e., $\mathbf{h}_2^T \mathbf{y} < 0$).

where

$$\eta = \sqrt{\frac{\mathbf{h}_2^T \mathbf{h}_2}{\sigma^2}} \tag{20}$$

Q is the right-tail probability function for a standard Gaussian random variable (zero mean and unit variance), and Q^{-1} is the inverse of this function. From (19), we can write

$$\delta^2 \eta = Q^{-1}(P_f) - Q^{-1}(P_d). \tag{21}$$

The above expression is key in illuminating a very useful relationship between the SNR and the smallest δ, which can be detected with very high probability and very low false alarm rate. To see this, it is convenient to define the output (total) SNR as follows:

$$\text{SNR} = \frac{\|\mathbf{h}_0 + \delta^2 \mathbf{h}_2\|^2}{\sigma^2}. \tag{22}$$

Using (20) and (22), the relation between minimum resolvable δ (i.e., $\delta_{\min}$) and the required SNR can be made explicit. Namely, SNR can be computed as

$$\text{SNR} = \frac{\|\mathbf{h}_0 + \delta^2 \mathbf{h}_2\|^2}{\|\mathbf{h}_2\|^2} \times \eta^2. \tag{23}$$

By substituting the required value of η from (21), we have

$$\text{SNR} = (Q^{-1}(P_f) - Q^{-1}(P_d))^2 \left[\frac{\mathbf{h}_0^T \mathbf{h}_0}{\mathbf{h}_2^T \mathbf{h}_2} \frac{1}{\delta^4} + 2 \frac{\mathbf{h}_2^T \mathbf{h}_0}{\mathbf{h}_2^T \mathbf{h}_2} \frac{1}{\delta^2} + 1 \right]. \tag{24}$$

This is a fundamental relationship relating minimum resolvable δ to SNR. To make the expressions more explicit, the energy terms in (24) can be approximated by[4]

$$\begin{aligned} \mathbf{h}_0^T \mathbf{h}_0 &\approx 2N \\ \mathbf{h}_2^T \mathbf{h}_2 &\approx \frac{\pi^4 N^5}{10 f_s^4} = \frac{\pi^4 N B^4}{10} \\ \mathbf{h}_0^T \mathbf{h}_2 &\approx -\frac{\pi^2 N^3}{6 f_s^2} = -\frac{\pi^2 N B^2}{6}. \end{aligned} \tag{25}$$

With these approximations, it is readily seen that for the range of $2\delta B < 1$, the relation (24) can be properly summarized by

$$\text{SNR} \approx \frac{320}{\pi^4} \frac{(Q^{-1}(P_f) - Q^{-1}(P_d))^2}{(2\delta B)^4}. \tag{26}$$

A plot of (24) and its approximation (26) are shown in Fig. 1. The result clearly shows that the minimum resolvable separation is essentially proportional to the inverse of the SNR to the fractional power of 1/4 for the range of $2\delta B < 1$. Note that the frequencies here are separated by 2δ.

Looking at (24) or (26), one may study the effect of sampling rate on these relationships. It should be mentioned that the sampling rate is embedded inside the "SNR" on the left hand side of (24) and (26). For instance, for resolving a particular frequency separation (2δ), doubling the sampling rate does not change the required SNR but, rather, implies that the same detection performance can be achieved with twice the noise variance as compared with the original sampling rate.

III. General Case

With the results of the previous section in place, we now follow a similar analysis and extend the results in this section to the general signal model of (3) with unknown amplitudes, phases, and unknown

[4]See Appendix A for a justification of these approximations.

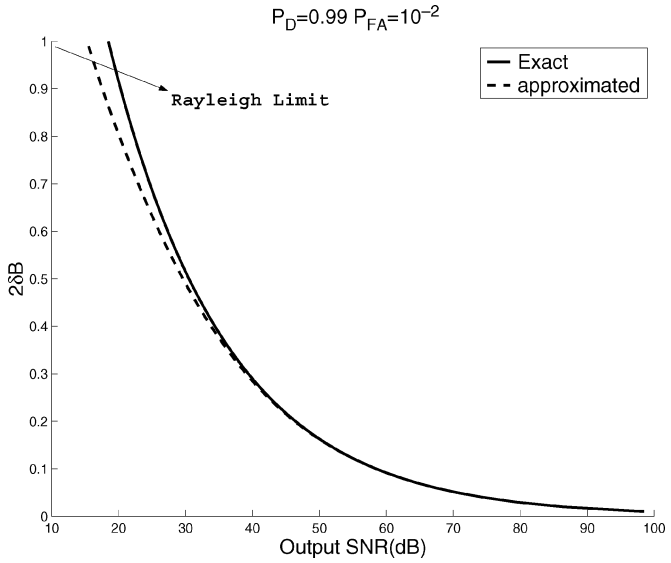

Fig. 1. $2\delta B$ versus required SNR.

frequency parameters δ_1 and δ_2.[5] The second-order Taylor expansion of the signal model around $(\delta_1, \delta_2) = (0, 0)$ is

$$s(k;\delta) \approx \alpha_0 p_0(k) + \beta_0 q_0(k) + \alpha_1 p_1(k) + \beta_1 q_1(k) + \alpha_2 p_2(k) + \beta_2 q_2(k) \quad (27)$$

where

$$p_i(k) = \left(\frac{k}{f_s}\right)^i \sin\left(2\pi f_c \frac{k}{f_s}\right) \quad (28)$$

$$q_i(k) = \left(\frac{k}{f_s}\right)^i \cos\left(2\pi f_c \frac{k}{f_s}\right) \quad (29)$$

and

$$\begin{aligned}
\alpha_0 &= a_1\cos(\phi_1) + a_2\cos(\phi_2)\\
\beta_0 &= a_1\sin(\phi_1) + a_2\sin(\phi_2)\\
\alpha_1 &= 2\pi(a_1\delta_1\sin(\phi_1) - a_2\delta_2\sin(\phi_2))\\
\beta_1 &= 2\pi(-a_1\delta_1\cos(\phi_1) + a_2\delta_2\cos(\phi_2))\\
\alpha_2 &= -2\pi^2(a_1\delta_1^2\cos(\phi_1) + a_2\delta_2^2\cos(\phi_2))\\
\beta_2 &= -2\pi^2(a_1\delta_1^2\sin(\phi_1) + a_2\delta_2^2\sin(\phi_2)).
\end{aligned} \quad (30)$$

Rewriting (27) in vector form will result in

$$\mathbf{s} \approx \alpha_0\mathbf{p}_0 + \beta_0\mathbf{q}_0 + \alpha_1\mathbf{p}_1 + \beta_1\mathbf{q}_1 + \alpha_2\mathbf{p}_2 + \beta_2\mathbf{q}_2. \quad (31)$$

Now, the hypotheses in (5) appear in the following form:

$$\begin{cases}\mathcal{H}_0: & \mathbf{z} = \alpha_0\mathbf{p}_0 + \beta_0\mathbf{q}_0 + \mathbf{w}\\ \mathcal{H}_1: & \mathbf{z} = \alpha_0\mathbf{p}_0 + \beta_0\mathbf{q}_0 + \alpha_1\mathbf{p}_1 + \beta_1\mathbf{q}_1 + \alpha_2\mathbf{p}_2 + \beta_2\mathbf{q}_2 + \mathbf{w}\end{cases} \quad (32)$$

where $\mathbf{z}$ denotes the approximate measured signal model. Equation (32) leads to a linear model for testing the parameter set θ defined as follows:

$$\mathbf{z} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w} \quad (33)$$

$$\mathbf{H} = [\mathbf{p}_0 \mid \mathbf{q}_0 \mid \mathbf{p}_1 \mid \mathbf{q}_1 \mid \mathbf{p}_2 \mid \mathbf{q}_2] \quad (34)$$

$$\boldsymbol{\theta} = [\alpha_0\ \beta_0\ \alpha_1\ \beta_1\ \alpha_2\ \beta_2]^T \quad (35)$$

[5]Another more general and perhaps more practical problem would be to consider the case where σ^2 is also unknown. This is clearly a more complicated scenario, and we have addressed it in [28].

where $\mathbf{H}$ and $\boldsymbol{\theta}$ are an $N \times 6$ matrix, and a 6×1 vector, respectively. The corresponding hypotheses are[6]

$$\begin{cases}\mathcal{H}_0: & \mathbf{A}\theta = \mathbf{0}\\ \mathcal{H}_1: & \mathbf{A}\theta \neq \mathbf{0}\end{cases} \quad (36)$$

where

$$\mathbf{A} = \begin{bmatrix}0&0&1&0&0&0\\0&0&0&1&0&0\\0&0&0&0&1&0\\0&0&0&0&0&1\end{bmatrix}. \quad (37)$$

The GLRT for (36) will be

$$T = \frac{1}{\sigma^2}\widehat{\theta}^T\mathbf{A}^T\left[\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T\right]^{-1}\mathbf{A}\widehat{\theta} \quad (38)$$

where

$$\widehat{\theta} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{z}. \quad (39)$$

From (38), the performance of this detector is characterized by

$$P_f = Q_{\chi_4^2}(\gamma) \quad (40)$$

$$P_d = Q_{\chi_4'^2(\lambda)}(\gamma) \quad (41)$$

$$\lambda = \frac{1}{\sigma^2}\theta^T\mathbf{A}^T\left[\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T\right]^{-1}\mathbf{A}\theta \quad (42)$$

where $Q_{\chi_4^2}$ is the right tail probability for a Central Chi-Squared PDF with 4 degrees of freedom, and $Q_{\chi_4'^2(\lambda)}$ is the right tail probability for a noncentral Chi-Squared PDF with 4 degrees of freedom and noncentrality parameter λ. For a specific desired P_d and P_f, we can compute the implied value for the noncentrality parameter from (40) and (41). We call this value of the noncentrality parameter $\lambda(P_f, P_d)$ and explicitly denote it as a function of desired probability of detection and false alarm rate. Meanwhile, similar to the simpler case in the previous section, the SNR in this case is given by

$$\text{SNR} = \frac{\theta^T\mathbf{H}^T\mathbf{H}\theta}{\sigma^2}. \quad (43)$$

Together, the above yield the relation between the parameter set θ and the required SNR as follows:

$$\text{SNR} = \lambda(P_f, P_d)\left(\theta^T\mathbf{A}^T\left[\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T\right]^{-1}\mathbf{A}\theta\right)^{-1}\theta^T\mathbf{H}^T\mathbf{H}\theta. \quad (44)$$

It is instructive to simplify (44) by approximating the elements of the matrix $\mathbf{H}^T\mathbf{H}$. These approximations (again, justified in Appendix A) yield

$$\mathbf{H}^T\mathbf{H} \approx \begin{bmatrix}\frac{N}{2}&0&0&\frac{-N}{4}\mu&\frac{N^3}{24f_s^2}&0\\0&\frac{N}{2}&\frac{-N}{4}\mu&0&0&\frac{N^3}{24f_s^2}\\0&\frac{-N}{4}\mu&\frac{N^3}{24f_s^2}&0&0&\frac{-N^3}{16}\mu\\\frac{-N}{4}\mu&0&0&\frac{N^3}{24f_s^2}&\frac{-N^3}{16}\mu&0\\\frac{N^3}{24f_s^2}&0&0&\frac{-N^3}{16}\mu&\frac{N^5}{160f_s^4}&0\\0&\frac{N^3}{24f_s^2}&\frac{-N^3}{16}\mu&0&0&\frac{N^5}{160f_s^4}\end{bmatrix}$$

[6]Two inequalities constrain the values of the parameters in (35): $\alpha_0\alpha_2 \leq 0$ and $\beta_0\beta_2 \leq 0$. For the detector development in Section III, we have ignored these constraints. We note that ignoring these constraints will still yield a detector, whereas invoking the constraints will yield (slightly) better detection performance. At an operating point where very high P_d and low P_f are considered, the performance of the detector will not be affected much at all by applying these constraints. Indeed, in such cases, the implied high value of SNR will effectively enforce the constraints with very high likelihood. In other words, for high SNR cases, the probability of violating these inequality constraints is negligible.

where $\mu = \cos((2\pi f_c/f_s)N))/\sin(2\pi f_c/f_s)$. With this approximation, after some algebra and replacing N/f_s by B, (44) will result in

$$\text{SNR} \approx \lambda(P_f, P_d)\frac{E_1 + E_2\mu + E_3B^2 + E_4B^2\mu + E_5B^4}{F_1B^2 + F_2B^2\mu + F_3B^4} \quad (45)$$

where

$$\begin{aligned} E_1 &= 16(\alpha_0^2 + \beta_0^2) \\ E_2 &= -16(\alpha_0\beta_1 + \beta_0\alpha_1) \\ E_3 &= \frac{4}{3}\left[\alpha_1^2 + \beta_1^2 + 2\alpha_0\alpha_2 + 2\beta_0\beta_2\right] \\ E_4 &= -4(\alpha_1\beta_2 + \beta_1\alpha_2) \\ E_5 &= \frac{1}{5}\left(\alpha_2^2 + \beta_2^2\right) \\ F_1 &= \frac{4}{3}\left(\alpha_1^2 + \beta_1^2\right) \\ F_2 &= \frac{-8}{3}(\alpha_1\beta_2 + \beta_1\alpha_2) \\ F_3 &= \frac{4}{45}\left(\alpha_2^2 + \beta_2^2\right). \end{aligned} \quad (46)$$

It is useful to note that for the case where $\phi_1 \approx \phi_2$, the first two terms in the numerator of (45) dominate its size for small δ_1 and δ_2 (i.e., δ_1, $\delta_2 \ll 1/B$), as the other terms are $O(\delta_1^2)$ and $O(\delta_2^2)$. Hence, (45) can be further approximated to

$$\text{SNR} \approx \lambda(P_f, P_d)\frac{E_1 + E_2\mu}{F_1B^2 + F_2B^2\mu + F_3B^4}. \quad (47)$$

To gain further insight, we can consider yet another special case. By assuming $a_1\delta_1 \approx a_2\delta_2$, which results from a proper choice of the center frequency f_c (see Appendix B), the values E_2, F_1, F_2 are also negligibly small, and we have

$$\text{SNR} \approx \lambda(P_f, P_d)\frac{E_1}{F_3B^4} \quad (48)$$

$$= \lambda(P_f, P_d)\frac{16(\alpha_0^2 + \beta_0^2)}{\frac{4}{45}(\alpha_2^2 + \beta_2^2)B^4} \quad (49)$$

$$= \frac{45}{\pi^4}\frac{\lambda(P_f, P_d)}{B^4}\frac{a_1^2 + a_2^2 + 2a_1a_2\cos(\phi_1 - \phi_2)}{a_1^2\delta_1^4 + a_2^2\delta_2^4 + 2a_1a_2\delta_1^2\delta_2^2\cos(\phi_1 - \phi_2)}. \quad (50)$$

A plot of (50) is shown in Fig. 2 for the case of equal amplitude and for the case of $a_1 = 4a_2$. (In either case, the amplitudes and phases are not known to the detector.) As expected, the case of equal amplitudes produces better detection performance.

In order to compare (50) with (26), let us set $\delta_1 = \delta_2 = \delta$, $a_1 = a_2 = 1$, and $\phi_1 = \phi_2 = 0$ to get[7]

$$\text{SNR} \approx \frac{720}{\pi^4}\frac{\lambda(P_f, P_d)}{(2B\delta)^4}. \quad (51)$$

As an example, let $P_d = 0.99$ and $P_f = 10^{-2}$. Comparing (26) and (51) shows that the required SNR for the second case (general case) is increased by a multiplicative factor of 2.72.

IV. Comparison With Existing Subspace-Based Methods

A very significant question is how the above results compare with existing methods for spectral estimation. Since we claim that the proposed detector structures are optimal, we expect that, at least for the particular signal model studied here, we should observe improved performance over existing subspace-based methods. As we demonstrate below, this is indeed the case. The subspace methods (e.g., MUSIC) for spectral estimation are based on eigendecomposition of the autocorrelation matrix into orthogonal signal and noise subspaces [7]. In practice, however, since, typically, only the time series are available, one uses an estimate of the autocorrelation matrix derived from the signal samples.

[7]It should be noted that these parameter values are unknown to the detector in the general case, and therefore, we expect poorer performance, as observed.

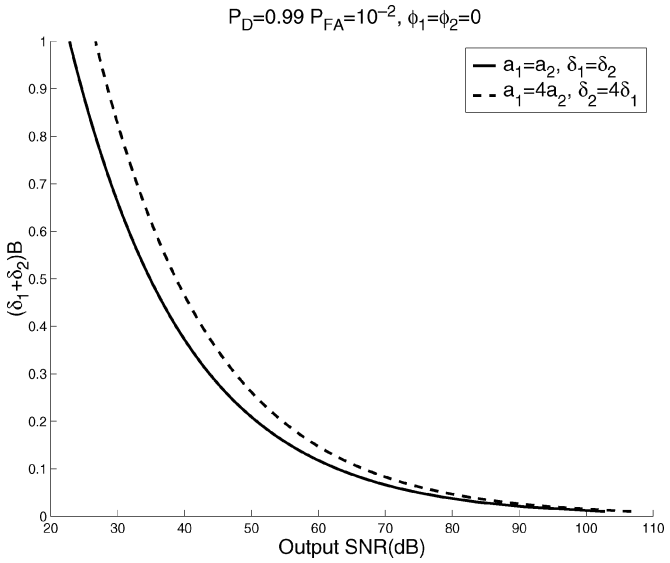

Fig. 2. $(\delta_1 + \delta_2)B$ versus required SNR for equal and unequal amplitudes.

In any event, much work has been done to study the performance and sensitivity of subspace methods (specifically MUSIC) [2], [8]–[10], [12]–[15]. Here, we make some comparisons to the existing methods. First, we consider the general class of subspace methods, in which we, very optimistically, assume that the exact autocorrelation matrix is known to the subspace detector under both hypotheses. As we will see, the proposed approach outperforms the subspace methods even in this (unrealistic) situation. Next, we present a comparison to the performance of the MUSIC algorithm in resolving sinusoids with nearby frequencies.

Throughout this section, we assume that $a_1 = a_2 = 1$ and that $\delta_1 = \delta_2 = \delta$. However, we will use our detector structure described in Section III, where we assume that amplitudes, frequencies, and phases in the signal model are unknown to the detector. Note that for subspace detectors, the phase is typically assumed to be a uniformly distributed random variable in $[0, 2\pi]$. Meanwhile, the "required SNR" computed in Section III is in general a function of the phases of the sinusoids. Thus, in order to set up a fair comparison to the subspace method, we perform the following averaging for the required SNR over the possible range of ϕ_1 and ϕ_2:

$$\text{SNR}_{\text{avg}} = \frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi}\text{SNR}\,\mathrm{d}\phi_1\mathrm{d}\phi_2 \quad (52)$$

where subscript "avg" denotes the averaged value, and the integrand (SNR) is the right-hand side of (44).

A. General Class of Subspace Methods; Completely Known Autocorrelation Matrix

We first consider the most idealistic subspace detector structure, to which the amplitudes $a_1 = a_2 = 1$ and frequency variables $\delta_1 = \delta_2 = \delta$ of the signal model $f(k, \delta_1, \delta_2)$ in (3) are known, and where ϕ_1 and ϕ_2 are assumed to be uniformly distributed random variables in the range of $[0, 2\pi]$. To decide whether the received signal contains a

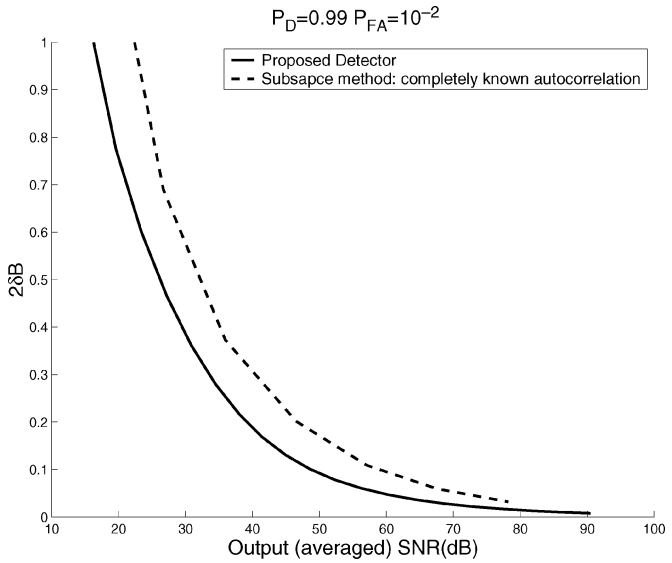

Fig. 3. $2\delta B$ versus required output SNR for the subspace detector with completely known autocorrelation.

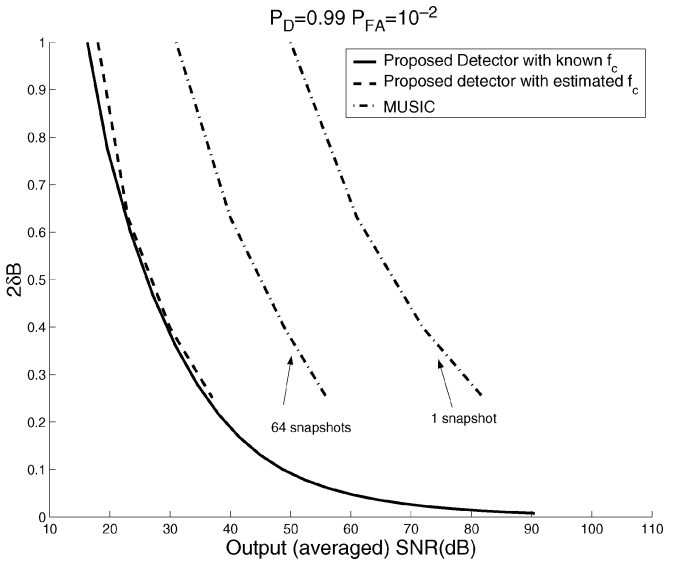

Fig. 4. $2\delta B$ versus required output SNR for the MUSIC algorithm.

single frequency component or two frequency components, we set up the following hypothesis test:

$$\begin{cases} \mathcal{H}_0: & \mathbf{f} \sim \mathcal{N}\left(0, \mathbf{R}_0 + \sigma^2\mathbf{I}\right) \\ \mathcal{H}_1: & \mathbf{f} \sim \mathcal{N}\left(0, \mathbf{R}_1 + \sigma^2\mathbf{I}\right) \end{cases} \tag{53}$$

where $\mathbf{R}_0$ and $\mathbf{R}_1$ are the autocorrelation matrices of the signal part in (3)[8]

$$\mathbf{R}_0 = \mathcal{RE}\left[\mathbf{r}(f_c)\mathbf{r}^H(f_c)\right] \tag{54}$$

$$\mathbf{R}_1 = \frac{1}{2}\mathcal{RE}\left[\mathbf{r}(f_c+\delta)\mathbf{r}^H(f_c+\delta)\right] + \frac{1}{2}\mathcal{RE}\left[\mathbf{r}(f_c-\delta)\mathbf{r}^H(f_c-\delta)\right] \tag{55}$$

where $\mathcal{RE}[\cdot]$ denotes the real part, and $\mathbf{r}(\cdot)$ is the vector form of

$$r(k; f_c) = \exp\left(j2\pi f_c \frac{k}{f_s}\right).$$

An NP detector for (53) decides $\mathcal{H}_1$ if

$$T_c(\mathbf{f}) = \mathbf{f}^T\left[\left(\mathbf{R}_1 + \sigma^2\mathbf{I}\right)^{-1} - \left(\mathbf{R}_0 + \sigma^2\mathbf{I}\right)^{-1}\right]\mathbf{f} > \gamma \tag{56}$$

where subscript "c" denotes the "completely known" case. The performance of this detector can be calculated analytically [25] or through Monte Carlo simulations, the result of which is shown in Fig. 3. For the purpose of simulation, the performance of (56) was computed by averaging over the possible range of ϕ_1 and ϕ_2 similar to (52). It is observed that the *required* SNR of this idealistic subspace method is generally between 5 and 10 dB higher than the required SNR for the proposed GLRT detector in (38). An interesting analysis related to the subspace framework is to compute the symmetric Kullback–Leibler Distance (KLD) or Divergence ($J(\cdot)$) [28, p. 26]. KLD is a measure of difficulty in discriminating between two hypotheses and is directly related to the performance figure of the subspace detector. More specifically, let $p(\mathbf{f}, 0)$ and $p(\mathbf{f}, \delta)$ be the PDFs of the measured signal under hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ in (53). Then, we will have (see Appendix C for proof)

$$J(\delta) = \int_{\mathcal{D}} [p(\mathbf{f}, \delta) - p(\mathbf{f}, 0)] \log\left(\frac{p(\mathbf{f}, \delta)}{p(\mathbf{f}, 0)}\right) \mathrm{d}\mathbf{f} \approx \frac{\delta^4}{8} tr\left[\left(\left[\mathbf{R}_1 + \sigma^2\mathbf{I}\right]^{-1} \left.\frac{\partial^2 \mathbf{R}_1}{\partial \delta^2}\right|_{\delta=0}\right)^2\right] \tag{57}$$

as $\delta \to 0$, where $tr[\cdot]$ is the trace operator, and $\mathcal{D}$ is the observation (signal) space. We note that the KLD measure behaves as the minimum detectable δ raised to the power of 4 (confirming the power law we have derived for the inverse of the required SNR in earlier sections). A comprehensive analysis of the relationship between divergence and resolution in a related framework can be found in [10].

B. Comparison With Music

For further comparison, we simulated the behavior of the MUSIC algorithm for resolving sinusoids with nearby frequencies. In simulation of MUSIC, the signal is declared to be resolvable if the output of MUSIC produces two distinct peaks within an interval around the true frequencies ($f_c \pm \delta$). The simulations for MUSIC are carried out for cases in which either a single snapshot, or multiple snapshots, are available. Naturally, we consider the output SNR in the latter case as the sum of SNRs of each snapshot.

Here, we develop two different comparison procedures. First, we compare the performance of MUSIC with the performance of the detector in (38), where we assume that the center frequency f_c, at which we perform the hypothesis test, is known *a priori*. Since this might be seen as an unfair comparison, we have put forward an alternative (perhaps more practical) scenario as well. In this scenario, we first seek assistance from MUSIC to estimate the center frequency and then apply the proposed detector in (38) centered at the peak estimated by MUSIC.

The results of these experiments are shown in Fig. 4. First, we observe that the proposed detector significantly outperforms MUSIC in both cases (using known or estimated center frequency). More interestingly, we see that the result of the proposed detector with estimated center frequency (provided by MUSIC) is very close to the performance of the same detector with known center frequency, the latter representing the ultimate performance bound. This implies that the

[8]Superscript "H" in (54) and (55) denotes conjugate transpose.

MUSIC algorithm does a very promising job in locating the center frequency (i.e., the candidate location where we can perform a refinement step using our proposed approach). Intuitively, the reason for this behavior is that for the case where a high probability of resolution (0.99) is considered, a fairly high value of SNR should be provided. This value of SNR will effectively guarantee a condition under which the MUSIC algorithm will produce the peak in its spectrum within the range of $[f_c-\delta, f_c+\delta]$. This observation is essentially in agreement with what has been observed in the past about the stability of MUSIC for single-sinusoid signals.[9] See, for example, [2] and [12].

V. Conclusion

In this correspondence, we have derived a performance bound for the minimum resolvable frequency separation between two tones in the presence of noise. We carried out the analysis in the context of locally optimal detectors and developed corresponding detection strategies that can in practice produce significant improvements over existing spectral estimation methods. For the task of bounding performance, we have answered a very practical question: "What is the minimum detectable frequency difference between two sinusoids at a given signal-to-noise ratio?" Equivalently: "What is the minimum SNR required to discriminate these two sinusoids?"

Compared to existing spectral estimation method, the proposed locally optimal detectors yield significantly improved detection of very nearby frequencies. It is worth noting that as a matter of implementation, one can always apply an existing method for spectral estimation (such as MUSIC etc.) to the given signal and then apply the proposed detector as a post-processing operation intended to further improve resolution. As discussed in Section IV-B, the application of such a detector, which uses (for example) MUSIC to estimate the center frequency as the test point, is nearly as effective as applying the proposed detector with a known center frequency.

In closing, we mention that the strategy for the analysis of resolution we have put forward here is very generally applicable to other types of signal models, such as damped sinusoidal signals. Once the signal model is decided upon, the same line of reasoning including approximations, and the development of locally optimal detectors can be carried out.

Appendix A
Computing the Energy Terms

In this Appendix, we explain the general process for the approximate computation of the energy terms. We will utilize the following identities for the calculation:

$$\sum_{k=0}^{L} x^k = \frac{1-x^{L+1}}{1-x} \tag{58}$$

$$\sum_{k=0}^{L} k^p x^k = \sum_{m=1}^{p} x^m \frac{\partial^m}{\partial x^m}\left(\frac{1-x^{L+1}}{1-x}\right) \tag{59}$$

$$\sum_k k^{p+1} \sin(xk)\cos(xk) = \frac{1}{2}\frac{\partial}{\partial x}\sum_k k^p \sin^2(xk). \tag{60}$$

[9]Although the signal in our case is a double sinusoid, since the frequencies are very close and we expect MUSIC to produce one peak, this is indeed a similar situation.

Instead of showing all the calculations, for the sake of brevity we discuss, as an example, the calculation of the term $\mathbf{h}_0^T\mathbf{h}_0$:

$$\begin{aligned}
\mathbf{h}_0^T\mathbf{h}_0 &= 4\sum_{k=-(N-1)/2}^{(N-1)/2} \sin^2\left(\frac{2\pi f_c}{f_s}k\right) \\
&= \sum_{k=-(N-1)/2}^{(N-1)/2} -\left[\exp\left(j\frac{2\pi f_c}{f_s}k\right)-\exp\left(-j\frac{2\pi f_c}{f_s}k\right)\right]^2 \\
&= \sum_{k=-(N-1)/2}^{(N-1)/2} 2-\exp\left(j\frac{4\pi f_c}{f_s}k\right)-\exp\left(-j\frac{4\pi f_c}{f_s}k\right) \\
&= 2N - 2\frac{1-\exp\left(j\frac{2\pi f_c}{f_s}(N+1)\right)}{1-\exp\left(j\frac{4\pi f_c}{f_s}\right)} \\
&\quad - 2\frac{1-\exp\left(-j\frac{2\pi f_c}{f_s}(N+1)\right)}{1-\exp\left(-j\frac{4\pi f_c}{f_s}\right)} + 2 \\
&= 2N + 2 \\
&\quad -2\frac{1-\cos\left(\frac{4\pi f_c}{f_s}\right)+\cos\left(\frac{2\pi f_c}{f_s}(N-1)\right)-\cos\left(\frac{2\pi f_c}{f_s}(N+1)\right)}{1-\cos\left(\frac{4\pi f_c}{f_s}\right)} \\
&= 2N - \underbrace{2\frac{\sin\left(\frac{2\pi f_c}{f_s}N\right)}{\sin\left(\frac{2\pi f_c}{f_s}\right)}}_{C}.
\end{aligned} \tag{61}$$

Since $\sin(x) \geq 1-|(2/\pi)x-1|$ for $0 \leq x \leq \pi$, and $2\pi f_c/f_s < \pi$, by upper and lower bounding the numerator and the denominator of $|C|$, respectively, we have

$$|C| = 2\left|\frac{\sin\left(\frac{2\pi f_c}{f_s}N\right)}{\sin\left(\frac{2\pi f_c}{f_s}\right)}\right| \leq \frac{2}{\sin\left(\frac{2\pi f_c}{f_s}\right)} \leq \frac{2}{1-\left|\frac{4f_c}{f_s}-1\right|}. \tag{62}$$

Thus, for the range of $\epsilon \leq 2f_c/f_s \leq 1-\epsilon$ (representing the range of f_s from just above the Nyquist rate ($2f_c$) to $1/\epsilon$ times the Nyquist rate), we will have $|C| < 1/\epsilon$, and therefore, for $\epsilon < 5/N$

$$\mathbf{h}_0^T\mathbf{h}_0 \approx 2N. \tag{63}$$

A similar approach can be followed to compute other energy terms.

Appendix B
Is $a_1\delta_1 \approx a_2\delta_2$ a Reasonable Assumption?

Suppose that $\phi_1 \approx \phi_2$. Then, the magnitude of the discrete-Time Fourier Transform of the signal is given by [29]

$$F(f,f_c,a_1,a_2,\delta_1,\delta_2) = S(f,f_c,a_1,a_2,\delta_1,\delta_2) + W(f) \tag{64}$$

$$\begin{aligned}
&= a_1\frac{\sin\left[\pi N\left(f-\frac{f_c-\delta_1}{f_s}\right)\right]}{\sin\left[\pi\left(f-\frac{f_c-\delta_1}{f_s}\right)\right]} \\
&\quad + a_2\frac{\sin\left[\pi N\left(f-\frac{f_c+\delta_2}{f_s}\right)\right]}{\sin\left[\pi\left(f-\frac{f_c+\delta_2}{f_s}\right)\right]} + W(f)
\end{aligned} \tag{65}$$

where $W(f)$ is the Fourier domain representation of $w(x_k)$. A reasonable way to find a good candidate center frequency (f_c), where we can perform our test, is to compute the correlation of the signal with the following window in the frequency domain:

$$G(f, f_x) = \frac{\sin\left[\pi N\left(f - \frac{f_x}{f_s}\right)\right]}{\sin\left[\pi\left(f - \frac{f_x}{f_s}\right)\right]} \tag{66}$$

and find the point where the correlation is maximum (this would yield a point near the stronger of the two peaks). Consider

$$\begin{aligned} &R_{SG}(|f_c - f_x|, a_1, a_2, \delta_1, \delta_2) \\ &\quad = \int_{-\infty}^{+\infty} [S(f, f_c, a_1, a_2, \delta_1, \delta_2) + W(f)]\, G(f, f_x)\mathrm{d}f \\ &\quad = a_1 R_{GG}(|f_c - \delta_1 - f_x|) + a_2 R_{GG}(|f_c + \delta_2 - f_x|) \\ &\qquad + R_{WG}(|f_x|) \end{aligned} \tag{67}$$

where R_{SG}, R_{WG}, and R_{GG} are the cross-correlation and autocorrelation functions defined as

$$R_{GG}(|f_c - \delta_2 - f_x|) = \int_{-\infty}^{+\infty} G(f, f_x) G(f, f_c - \delta_2)\mathrm{d}f. \tag{68}$$

$$R_{GG}(|f_c + \delta_2 - f_x|) = \int_{-\infty}^{+\infty} G(f, f_x) G(f, f_c + \delta_2)\mathrm{d}f. \tag{69}$$

$$R_{WG}(|f_c - \delta_1 - f_x|) = \int_{-\infty}^{+\infty} G(f, f_x) W(f)\mathrm{d}f. \tag{70}$$

Since $\delta_1, \delta_2 \ll 1/B$ and f_x is expected to be close to f_c, we can again use the Taylor expansion for (68) around $(f_c - \delta_1 + f_x, f_c + \delta_2 - f_x) = (0, 0)$

$$R_{GG}(|f_c - \delta_1 - f_x|) \approx \xi_0 + (|f_c - \delta_1 - f_x|)\xi_1 + (f_c - \delta_1 - f_x)^2 \xi_2 \tag{71}$$

$$R_{GG}(|f_c + \delta_2 - f_x|) \approx \xi_0 + (|f_c + \delta_2 - f_x|)\xi_1 + (f_c + \delta_2 - f_x)^2 \xi_2 \tag{72}$$

where ξ_0, ξ_1, and ξ_3 are some constant coefficients of the above Taylor expansion. In addition, it can be shown that $\xi_1 = 0$. Therefore, we can write (67) as follows:

$$\begin{aligned} R_{SG}(|f_c - f_x|, a_1, a_2, \delta_1, \delta_2) &\approx (a_1 + a_2)\xi_0 \\ &\quad + \left(a_1(f_c - f_x - \delta_1)^2 + a_2(f_c - f_x + \delta_2)^2\right)\xi_2. \end{aligned} \tag{73}$$

Taking the derivative of the right-hand side of (73) with respect to f_x and setting it to zero will result in

$$(a_1 + a_2)(f_c - f_x) \approx a_1\delta_1 - a_2\delta_2. \tag{74}$$

Hence, a proper selection of f_c (by using the above correlation-based approach) will lead to $a_1\delta_1 \approx a_2\delta_2$. It is worth noting that application of any subspace-based method to the data will also provide a reasonable candidate for the center frequency f_c, as discussed in Section IV-B.

Appendix C
Computing the Kullback–Leibler Distance in (57)

Directly using the results in [28, p. 26], we can obtain the following expression for KLD:

$$J(\delta) \approx \delta^2 I(0) \tag{75}$$

where $I(\delta)$ is the Fisher Information measure [24, p 40],

$$I(\delta) = -E\left[\frac{\partial^2 \ln p(\mathbf{f}, \delta)}{\partial \delta^2}\right] = \frac{1}{2} tr\left[\left(\left[\mathbf{R}_1 + \sigma^2\mathbf{I}\right]^{-1}\frac{\partial \mathbf{R}_1}{\partial \delta}\right)^2\right] \tag{76}$$

where $E[\cdot]$ is the expectation operator. However, for the hypothesis test of interest in (53), $I(0)$ is zero, and (75) is not directly applicable. Here, we extend the approach in [28, p. 26] by considering higher order terms. Consider the following Taylor expansion:

$$J(\delta) = \int_{\mathcal{D}} [p(\mathbf{f}, \delta) - p(\mathbf{f}, 0)] \log\left(\frac{p(\mathbf{f}, \delta)}{p(\mathbf{f}, 0)}\right)\mathrm{d}\mathbf{f} \tag{77}$$

$$\begin{aligned} &= J(0) + \delta\left.\frac{\partial J}{\partial \delta}\right|_{\delta=0} + \frac{\delta^2}{2}\left.\frac{\partial^2 J}{\partial \delta^2}\right|_{\delta=0} + \frac{\delta^3}{6}\left.\frac{\partial^3 J}{\partial \delta^3}\right|_{\delta=0} \\ &\quad + \frac{\delta^4}{24}\left.\frac{\partial^4 J}{\partial \delta^4}\right|_{\delta=0} + O(\delta^6). \end{aligned} \tag{78}$$

Noting that[10]

$$\left.\frac{\partial^i p(\mathbf{f}, \delta)}{\partial \delta^i}\right|_{\delta=0} = 0 \quad i = 1, 3 \tag{79}$$

we will have (80)–(84), shown at the top of the next page. As a result, we can write (77) as

$$J(\delta) \approx \frac{\delta^4}{4}\int_{\mathcal{D}} \left.\frac{\left[\frac{\partial^2 p(\mathbf{f}, \delta)}{\partial \delta^2}\right]^2}{p(\mathbf{f}, \delta)}\right|_{\delta=0} \mathrm{d}\mathbf{f}. \tag{85}$$

On the other hand, we see that we have (86)–(88), shown at the top of the next page. Therefore

$$J(\delta) \approx \frac{\delta^4}{8}\left.\frac{\partial^2 I(\delta)}{\partial \delta^2}\right|_{\delta=0} \tag{89}$$

$$= \frac{\delta^4}{16} tr\left[\left.\frac{\partial^2}{\partial \delta^2}\left(\left[\mathbf{R}_1 + \sigma^2\mathbf{I}\right]^{-1}\frac{\partial \mathbf{R}_1}{\partial \delta}\right)^2\right|_{\delta=0}\right] \tag{90}$$

$$= \frac{\delta^4}{16} tr\left[\left.2\left(\frac{\partial \mathbf{R}}{\partial \delta}\right)^2 + \frac{\partial^2 \mathbf{R}}{\partial \delta^2}\mathbf{R} + \frac{\partial^2 \mathbf{R}}{\partial \delta^2}\mathbf{R}\right|_{\delta=0}\right] \tag{91}$$

where

$$\mathbf{R} = \left[\mathbf{R}_1 + \sigma^2\mathbf{I}\right]^{-1}\frac{\partial \mathbf{R}_1}{\partial \delta} \tag{92}$$

$$\frac{\partial \mathbf{R}}{\partial \delta} = -\left(\left[\mathbf{R}_1 + \sigma^2\mathbf{I}\right]^{-1}\frac{\partial \mathbf{R}_1}{\partial \delta}\right)^2 + \left[\mathbf{R}_1 + \sigma^2\mathbf{I}\right]^{-1}\frac{\partial^2 \mathbf{R}_1}{\partial \delta^2} \tag{93}$$

$$\begin{aligned} \frac{\partial^2 \mathbf{R}}{\partial \delta^2} &= 2\left(\left[\mathbf{R}_1 + \sigma^2\mathbf{I}\right]^{-1}\frac{\partial \mathbf{R}_1}{\partial \delta}\right)^3 \\ &\quad - 2\left[\mathbf{R}_1 + \sigma^2\mathbf{I}\right]^{-1}\frac{\partial \mathbf{R}_1}{\partial \delta}\left[\mathbf{R}_1 + \sigma^2\mathbf{I}\right]^{-1}\frac{\partial^2 \mathbf{R}_1}{\partial \delta^2} \\ &\quad - \left[\mathbf{R}_1 + \sigma^2\mathbf{I}\right]^{-1}\frac{\partial^2 \mathbf{R}_1}{\partial \delta^2}\left[\mathbf{R}_1 + \sigma^2\mathbf{I}\right]^{-1}\frac{\partial \mathbf{R}_1}{\partial \delta} \\ &\quad + \left[\mathbf{R}_1 + \sigma^2\mathbf{I}\right]^{-1}\frac{\partial^3 \mathbf{R}_1}{\partial \delta^3}. \end{aligned} \tag{94, 95}$$

[10] $p(\mathbf{f}, \delta)$ is an even (and differentiable) function around $\delta = 0$.

$$J(0) = 0, \tag{80}$$

$$\left.\frac{\partial J}{\partial \delta}\right|_{\delta=0} = \int_{\mathcal{D}} \frac{\partial p(\mathbf{f},\delta)}{\partial \delta} \log\left(\frac{p(\mathbf{f},\delta)}{p(\mathbf{f},0)}\right) + [p(\mathbf{f},\delta) - p(\mathbf{f},0)] \left.\frac{\frac{\partial p(\mathbf{f},\delta)}{\partial \delta}}{p(\mathbf{f},\delta)}\right|_{\delta=0} \mathrm{d}\mathbf{f} = 0 \tag{81}$$

$$\left.\frac{\partial^2 J}{\partial \delta^2}\right|_{\delta=0} = \int_{\mathcal{D}} \left.\frac{2\left[\frac{\partial p(\mathbf{f},\delta)}{\partial \delta}\right]^2}{p(\mathbf{f},\delta)}\right|_{\delta=0} \mathrm{d}\mathbf{f} = 0 \tag{82}$$

$$\left.\frac{\partial^3 J}{\partial \delta^3}\right|_{\delta=0} = \int_{\mathcal{D}} \left.\frac{6\frac{\partial p(\mathbf{f},\delta)}{\partial \delta}\frac{\partial^2 p(\mathbf{f},\delta)}{\partial \delta^2}}{p(\mathbf{f},\delta)} - \frac{3\left[\frac{\partial p(\mathbf{f},\delta)}{\partial \delta}\right]^3}{[p(\mathbf{f},\delta)]^2}\right|_{\delta=0} \mathrm{d}\mathbf{f} = 0 \tag{83}$$

$$\begin{aligned}\left.\frac{\partial^4 J}{\partial \delta^4}\right|_{\delta=0} &= \int_{\mathcal{D}} \left.\frac{8\frac{\partial p(\mathbf{f},\delta)}{\partial \delta}\frac{\partial^3 p(\mathbf{f},\delta)}{\partial \delta^3} + 6\left[\frac{\partial^2 p(\mathbf{f},\delta)}{\partial \delta^2}\right]^3}{p(\mathbf{f},\delta)} - \frac{18\left[\frac{\partial p(\mathbf{f},\delta)}{\partial \delta}\right]^2\frac{\partial^2 p(\mathbf{f},\delta)}{\partial \delta^2}}{[p(\mathbf{f},\delta)]^2} + \frac{8\left[\frac{\partial p(\mathbf{f},\delta)}{\partial \delta}\right]^4}{[p(\mathbf{f},\delta)]^3}\right|_{\delta=0} \mathrm{d}\mathbf{f} \\ &= \int_{\mathcal{D}} \left.\frac{6\left[\frac{\partial^2 p(\mathbf{f}!,\delta)}{\partial \delta^2}\right]^2}{p(\mathbf{f},\delta)}\right|_{\delta=0} \mathrm{d}\mathbf{f}.\end{aligned} \tag{84}$$

$$\left.\frac{\partial^2 I(\delta)}{\partial \delta^2}\right|_{\delta=0} = -\frac{\partial^2}{\partial \delta^2}\int_{\mathcal{D}} \left.\frac{\partial^2 \ln p(\mathbf{f},\delta)}{\partial \delta^2} p(\mathbf{f},\delta)\right|_{\delta=0} \mathrm{d}\mathbf{f} \tag{86}$$

$$= \frac{\partial^2}{\partial \delta^2}\int_{\mathcal{D}} \left.\frac{\left[\frac{\partial \ln p(\mathbf{f},\delta)}{\partial \delta}\right]^2}{p(\mathbf{f},\delta)}\right|_{\delta=0} \mathrm{d}\mathbf{f} \tag{87}$$

$$\begin{aligned}&= \int_{\mathcal{D}} \left.\frac{2\frac{\partial p(\mathbf{f},\delta)}{\partial \delta}\frac{\partial^3 p(\mathbf{f},\delta)}{\partial \delta^3} + 2\left[\frac{\partial^2 p(\mathbf{f},\delta)}{\partial \delta^2}\right]^3}{p(\mathbf{f},\delta)} - \frac{5\left[\frac{\partial p(\mathbf{f},\delta)}{\partial \delta}\right]^2\frac{\partial^2 p(\mathbf{f},\delta)}{\partial \delta^2}}{[p(\mathbf{f},\delta)]^2} + \frac{2\left[\frac{\partial p(\mathbf{f},\delta)}{\partial \delta}\right]^4}{[p(\mathbf{f},\delta)]^3}\right|_{\delta=0} \mathrm{d}\mathbf{f} \\ &= \int_{\mathcal{D}} \left.\frac{2\left[\frac{\partial^2 p(\mathbf{f},\delta)}{\partial \delta^2}\right]^2}{p(\mathbf{f},\delta)}\right|_{\delta=0} \mathrm{d}\mathbf{f}.\end{aligned} \tag{88}$$

Finally, since

$$\left.\frac{\partial^i \mathbf{R}_1}{\partial \delta^i}\right|_{\delta=0} = 0 \quad i = 1, 3 \tag{96}$$

we will have

$$J(\delta) = \frac{\delta^4}{8} tr\left[\left(\left[\mathbf{R}_1 + \sigma^2\mathbf{I}\right]^{-1} \left.\frac{\partial^2 \mathbf{R}_1}{\partial \delta^2}\right|_{\delta=0}\right)^2\right]. \tag{97}$$

As we see from (89) and (97), the divergence for the underlying hypothesis testing problem is directly related to the second derivative of the Fisher information matrix evaluated at $\delta = 0$.

ACKNOWLEDGMENT

The authors wish to thank Prof. L. Scharf for his valuable comments on the first draft of this work. They also thank the anonymous reviewers for their very fruitful comments and suggestions.

REFERENCES

[1] D. H. Johnson and D. E. Dugeon, *Array Signal Processing: Concepts and Techniques*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[2] M. Kaveh and A. J. Barabell, "The statistical performance of the MUSIC and the minimum-norm algorithms in resolving plane waves in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 2, pp. 331–341, Apr. 1986.

[3] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2146–2157, Aug. 1994.

[4] J. A. Cadzow, Y. S. Kim, and D. C. Shiue, "General direction-of-arrival estimation: A signal subspace approach," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 25, no. 1, pp. 31–47, Jan. 1989.

[5] J. A. Cadzow, "Direction finding: A signal subspace approach," *IEEE Trans Syst., Man, Cybern.*, vol. 21, no. 5, pp. 1115–1124, Sep.-Oct. 1991.

[6] S. M. Kay, *Modern Spectral Estimation: Theory and Application*. Englewood Cliffs, NJ: Prentice-Hall, 1988.

[7] ——, "Spectrum analysis: A modern perspective," *Proc. IEEE*, vol. 69, pp. 1380–1418, 1981.

[8] V. U. Reddy and L. S. Biradar, "SVD-based information theoretic criteria for detection of the number of damped/undamped sinusoids and their performance analysis," *IEEE Trans. Signal Process.*, vol. 41, no. 9, pp. 2872–2881, Sep. 1993.

[9] L. T. McWhorter and L. L. Scharf, "Cramer–Rao bounds for deterministic modal analysis," *IEEE Trans. Signal Process.*, vol. 41, no. 5, pp. 1847–1862, May 1993.
[10] L. L. Scharf and P. H. Moose, "Information measures and performance bounds for array processors," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 11–21, Jan. 1976.
[11] H. V. Hamme, "A stochastical limit to the resolution of least squares estimation of the frequencies of a double complex sinusoid," *IEEE Trans. Signal Process.*, vol. 39, no. 12, pp. 2652–2658, Dec. 1991.
[12] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramér-Rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, pp. 720–741, May 1989.
[13] ——, "MUSIC, maximum likelihood, and Cramér–Rao bound: Further results and comparisons," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 12, pp. 2140–2150, Dec. 1990.
[14] A. Weiss and B. Friedlander, "Effects of modeling error on the resolution threshold of the MUSIC algorithm," *IEEE Trans. Signal Process.*, vol. 42, no. 1, pp. 147–155, Jan. 1994.
[15] H. B. Lee and M. S. Wegrovitz, "Resolution threshold of beamspace MUSIC for two closely spaced emitters," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 9, pp. 1545–1559, Sep. 1990.
[16] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. AP-34, pp. 276–280, 1986.
[17] A. J. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *Proc. ICASSP*, Boston, MA, Apr. 1983, pp. 336–339.
[18] S. T. Smith, "Statistical resolution limits and the complexified Cramér-Rao bound," *IEEE Trans. Signal Processing*, vol. 53, no. 5, pp. 1597–1609, May 2005.
[19] M. Zatman and S. T. Smith, "Resolution and ambiguity bounds for interferometric-like systems," in *Proc. 32nd Asilomar Conf. Signals, Syst., Comput.*, 1998.
[20] S. T. Smith, "Cramér–Rao and resolution bounds for adaptive sensor array processing," in *Proc. IEEE Statist. Signal Array Process. Workshop*, Portland, OR, 1998, p. 3740.
[21] P. Milanfar and A. Shakouri, "A statistical analysis of diffraction-limited imaging," in *Proc. Int. Conf Image Process.*, Rochester, NY, Sep. 2002, pp. 864–867.
[22] M. Shahram and P. Milanfar, "Imaging below the diffraction limit. A statistical analysis," *IEEE Trans. Image Process.*, vol. 13, no. 5, pp. 677–689, May 2004.
[23] ——, "A statistical analysis of achievable resolution in incoherent imaging," in *Proc. SPIE Annu. Meet., Signal Data Process. Small Targets*, San Diego, CA, Aug. 2003.
[24] S. M. Kay, *Fundamentals of Statistical Signal Processing, Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1998.
[25] ——, *Fundamentals of Statistical Signal Processing, Detection Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1998.
[26] L. L. Scharf, *Statistical Signal Processing, Detection, Estimation, and Time Series Analysis*. Reading, MA: Addison-Wesley, 1991.
[27] M. Shahram and P. Milanfar, "Local detectors for high-resolution spectral analysis: Algorithms and performance," *Digit. Signal Process.*, vol. 15, no. , pp. 305–316, 2005.
[28] S. Kullback, *Information Theory and Statistics*. New York: Dover, 1968.
[29] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

Nonlinear System Identification in Impulsive Environments

Binwei Weng, *Student Member, IEEE,* and
Kenneth E. Barner, *Senior Member, IEEE*

***Abstract*—Nonlinear system identification has been studied under the assumption that the noise has finite second and higher order statistics. In many practical applications, impulsive measurement noise severely weakens the effectiveness of conventional methods. In this paper, α-stable noise is used as a noise model. In such case, the minimum mean square error (MMSE) criterion is no longer an appropriate metric for estimation error due to the lack of finite second-order statistics of the noise. Therefore, we adopt minimum dispersion criterion, which in turn leads to the adaptive least mean pth power (LMP) algorithm. It is shown that the LMP algorithm under the α-stable noise model converges as long as the step size satisfies certain conditions. The effect of p on the performance is also investigated. Compared with conventional methods, the proposed method is more robust to impulsive noise and has better performance.**

***Index Terms*—Adaptive Volterra filter, α-stable process, impulsive noise, LMP algorithm, nonlinear system identification.**

I. Introduction

System identification is an important subject that plays a central role in many practical problems. In essence, one wants to establish a mathematical model for an unknown system through the input–output relationship. Linear system identification has been extensively studied for several decades [1]. In the physical world, however, many systems exhibit certain degrees of nonlinearity and are not well modeled as a linear system. In such situations, nonlinear system identification is necessary for further analysis.

Nonlinear systems are those that do not possess the superposition property. Several types of nonlinear models exist in system identification arena [1], [2]. Among them, the Volterra system [3] is one of the most commonly used models due to its roots in Taylor's series expansion of nonlinear functions with memory. The Volterra system is also known as the polynomial system and has applications in fields such as communications [4], [5], speech processing [6], image processing [7], and biomedical engineering [8].

An important issue in system identification is the effect of measurement noise on the results. The measurement noise is often assumed to be a random process with finite second-order statistics (SOS), making the MSE an appropriate metric for estimation error. In some circumstances, however, the noise may have heavier tails and even may not possess finite SOS. This is particularly true for impulsive noise, which occurs frequently in applications such as seismology, geophysics and astrophysics, biomedicine, communications, and underwater acoustics [9], [10]. In such cases, identification methods based on the conventional noise assumption have poor performance. Therefore, alternative methods must be sought to identify nonlinear systems in impulsive noise environments.

Many noise processes that are impulsive in nature can be modeled as α-stable process [11], [12]. This is based on the *Generalized Central*

Manuscript received March 19, 2004; revised August 17, 2004. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Chong-Yung Chi.

The authors are with the Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716 USA (e-mail: weng@ece.udel.edu; barner@ece.udel.edu).

Digital Object Identifier 10.1109/TSP.2005.849213