A new data-dependent zero-memory preprocessing algorithm is proposed to mitigate additive heavy-tail noise or shot noise. Each sensor-array snapshot’s spatial data vector is adaptively normalized by its infinity-norm. Zero cross-correlation is preserved between the signal-subspace and the noise-space.

I. INTRODUCTION

This work proposes a new preprocessing technique for second-order-statistics-based algorithms to mitigate additive heavy-tail noise or shot noise.\(^1\)

This new scheme falls within the class of data-adaptive zero-memory (DA-ZM) algorithms, which 1) is “blind,” in not requiring any prior knowledge of the “heavy-tailed” noise’s statistics; 2) is versatile and robust, in being applicable to any impulsive-noise model (including the \(\alpha\)-stable model and the Gaussian-mixture model), in the sense that this algorithm is not tailored specifically to any specific impulsive-noise model; 3) is versatile and robust also in its applicability to noise processes of any temporal nonstationarity; 4) can adopt second-order-statistics-based algorithms to estimate parameters from impulse-noise-corrupted data; 5) is memoryless.

The proposed scheme adaptively normalizes each sensor-array snapshot’s spatial data vector by its infinity-norm, to construct a pseudocorrelation function (or pseudocovariance function) out of the impulse-noise impaired data. In contrast, earlier

\(^1\)Confusingly, many references label both “heavy-tailed noise” and/or “shot noise” as “impulse noise.” The present algorithm can handle either type of impulse noise. A shot-noise process 1) has a non-zero amplitude only intermittently and sporadically over all discrete-time samples, and 2) when it takes on non-zero amplitudes, such non-zero amplitudes may or may not have a finite mean, a finite variance, and finite higher order moments. In contrast, heavy-tailed noise process typically has a non-zero amplitude at all discrete-time samples. Incidentally, a noise-process may be both shot noise and heavy tailed.

Manuscript received February 14, 2007; revised September 11, 2007; released for publication March 4, 2008.

IEEE Log No. T-AES/44/3/929767.

Refereeing of this contribution was handled by E. Chornoboy.


0018-9251/08/$25.00 c 2008 IEEE
DA-ZM schemes preprocess each space-time scalar data sample separately and independently from other space-time scalar data samples. For example, [4], [7], [10], [12], and [15] process each scalar data sample nonlinearly using various forms of Gaussian-tailed (GT) self-normalizations. This scalar data-samples’ independent self-normalization in these references would generally corrupt the original data correlation-matrix’s signal subspace by mixing with the noise subspace. In contrast, the present algorithm adaptively normalizes each sensor-array snapshot’s spatial data vector by its infinity-norm. Zero cross-correlation is thereby preserved between the signal subspace and the noise space. In other words, the presently proposed \( \| \|_{\infty} \) -normalization preprocessing scheme aims to construct a pseudocorrelation function or a pseudocovariance function despite the impulse noises. Furthermore, in contrast to the above nonlinear schemes, the presently proposed preprocessing scheme is linear.

The sign covariance matrix MUSIC (SCM-MUSIC) [14], like the present scheme, normalizes linearly all space-time data samples in a snapshot together as a unit, by that snapshot vector’s second-order norm. This proposed scheme uses instead the infinity-norm, which is 1) more sensitive against impulse noises than any finite order, and 2) computationally simpler than any finite-order norm.\(^3\)

II. MATHEMATICAL DATA MODEL AND PROBLEM STATEMENT

The proposed preprocessing scheme is applicable to a wide variety of parameter-estimation problems in an impulse-noise environment. For illustrative purposes, consider the classic problem of direction finding of \( P \) spatio-temporally stationary narrowband far-field point-sources, emitting at the same center frequency. Collected by an \( L \)-sensor array at the time \( t \) is the \( L \times 1 \) data-vector:

\[
x(t) = \sum_{p=1}^{P} a(\theta_p)s_p(t) + n(t)
\]

where \( a(\cdot) \) represents the a priori known \( L \times 1 \) array manifold, and the entries in the vector \( \theta_p \) contain the \( p \)th source’s a priori unknown azimuth-elevation direction-of-arrival, polarization parameters, and other signal parameters. Moreover, \( s_p(t) \) symbolizes the \( p \)th source’s temporal signal of finite (but possibly unknown) variance,\(^2\) and \( n(t) \) refers to an \( L \times 1 \) vector of additive noise. The present problem is to estimate \( \{\theta_1, \theta_2, \ldots, \theta_P\} \) from a known \( N \) number of snapshots \( \{x(1), x(2), \ldots, x(N)\} \).

III. CUSTOMARY SECOND-ORDER-STATISTICS EIGENSTRUCTURE-BASED DIRECTION FINDING

Made in this work are the following “standard” data-model assumptions in direction finding:

\( \{s_p(t), \; p = 1, \ldots, P\} \) constitutes a set of zero-mean complex-valued stationary random processes\(^6\) not cross-correlated among themselves; \( P < L \); the elements in \( n(t) \) are spatio-temporally independent identically distributed complex-value \( L \times 1 \) vector random processes, whose entries all have a zero mean, a finite variance \( \sigma_n^2 \), and zero cross-correlation with any of the signals. Then, the data would be a zero-mean process with a finite-magnitude covariance-matrix \( R = E\{x(t)x^H(t)\} = ARAR^H + R_n \),

where \( E\{\cdot\} \) denotes the statistical expectation operator, the superscript \( H \) refers to the Hermitian operation, \( R_i = \text{diag}\{\sigma_{f_1}^2, \sigma_{f_2}^2, \ldots, \sigma_{f_P}^2\} \), \( R_n = \sigma_n^2 I \), \( \text{diag}\{\cdot\} \) symbolizes a diagonal matrix with the diagonal elements being the entries inside the brackets, and \( \sigma_p^2 = E\{|s_p(t)|^2\} \) denotes the \( p \)th signal’s power.

Exploiting the aforementioned second-order-statistics’ eigen-structure in \( R \), various subspace-based “high-resolution” parameter-estimation methods\(^7\) may be applied [9].

For illustration here and for reference later, the basic MUSIC algorithm is briefly stated here. Form the data-sample covariance-matrix \( \hat{R} = (1/N)\sum_{t=1}^{N} x(t)x^H(t) \). Further assume that the value of \( P \) is known or correctly estimated, and that the \( P \) steering vectors in \( \{a(\theta_p), \; p = 1, \ldots, P\} \) are linearly independent. Then, eigen-decompose \( \hat{R} \) to construct the \( L \times (L-P) \) “noise-subspace” matrix \( \hat{E}_n \), whose columns equal the \( L \times 1 \) eigenvectors associated with the \( L-P \) smallest eigenvalues of \( \hat{R} \). As the array manifold’s mathematical form \( a(\theta) \) is assumed as a priori known, form the MUSIC pseudospectrum scalar function: \( V(\theta) = 1/\|a^H(\theta)E_\theta\|_2^2 \). Those \( P \)

\( ^2\)For details how this time-space data sample independent normalization affects the signal subspace, please see the discussion at the end of Section IV.

\( ^3\)Reference [5] is another ZMNL-based algorithm, but it is for detection and it uses an inverse-Gaussian weighting.

\( ^4\)Also aiming to handle arbitrary impulsive noise of unknown statistics, but not within the aforementioned DA-ZM algorithmic class is the minimax estimation in [11], [16], [17], and [18]. These algorithms, unlike the present scheme, do not attempt to modify the impulse-noise-impaired data for use with second-order-statistics-based parameter-estimation methods like the aforementioned eigen-structure-based methods of MUSIC or ESPRIT.

\( ^5\)This finite-variance requirement is not very restrictive, as all communication signals, radar signals, and sonar signals have finite powers and finite autocorrelation functions.

\( ^6\)The exact mathematical form of \( \{s_p(t), \; p = 1, \ldots, P\} \) is immaterial.

\( ^7\)These include the well-known algorithms of “multiple signal classification” (MUSIC) [1] and “estimation of signal parameters via rotational invariance techniques” ESPRIT [3].
values of $\theta$ that maximize $V(\theta)$ would constitute MUSIC’s estimates of the $P$ sets of unknown signal parameters.

In the special case of one-dimensional direction finding of sources sharing a known common carrier frequency using a linear array of isotropic sensors uniformly spaced at a known inter-sensor separation $D$, the vector $\theta_p$ degenerates to a scalar $\theta_p$. Moreover,

$$
a(\theta_p) = \left[1, \exp\left[-j2\pi\frac{D}{\lambda} \sin\theta_p\right], \ldots, \exp\left[-j2\pi(L-1)\frac{D}{\lambda} \sin\theta_p\right]\right]^T$$

(2)

where $\lambda$ refers to the $P$ sources’ common wavelength.

IV. PROPOSED INFINITY-NORM-NORMALIZATION PREPROCESSING METHOD

With impulse noise (e.g., non-Gaussian $\alpha$-stable noise), the second moment does not exist, thereby rendering the above covariance-based direction finding methods inapplicable in their original forms. This work proposes a normalization for the $L \times 1$ vector $n(t)$ by its $\infty$-norm at each $t$, in an admittedly ad hoc attempt to construct a pseudocovariance of sorts. That is, normalize the collected data $x(t)$ to give

$$
y(t) = w(t)x(t)$$

(3)

where the adaptive memoryless normalizing weight $w(t)$ is formed as the infinity-norm of $x(t)$ blindly with no prior information about the noise statistics:

$$
w(t) = \frac{1}{\|x(t)\|_\infty} = \frac{1}{\max\{|x_1(t)|, |x_2(t)|, \ldots, |x_L(t)|\}}$$

(4)

where $x_i(t)$ represents the $i$th entry in the vector $x(t)$.

The above $\|\cdot\|_\infty$ normalized $y(t)$ can substitute for $x(t)$ in customary second-order-statistics parameter-estimation methods, including the eigenstructure-based high-resolution algorithms such as MUSIC.

Relative to the nonlinear zero-memory techniques in [5], [14], [12], or [15], the above proposed $\|\cdot\|_\infty$ normalization linear preprocessing scheme’s better direction finding resolution and accuracy than the nonlinear zero-memory schemes. This may be intuitively explained by the infinity-norm’s higher sensitivity to outliers, over the other zero-memory schemes’ soft-clipping [5, 12, 15] or $\|\cdot\|_\infty$ normalization [14]). Moreover, the proposed method is more accurate than the “scalar-sample by scalar-sample” normalization schemes in [4, 7, 10, 12, 15], because the present one-whole-snapshot-at-a-time normalization preserves for the data spatial-correlation matrix its signal subspace’s and noise subspace’s rank, unlike the earlier DA-ZM “hole-punching” methods [4, 7, 10, 12, 15] that set to zero only the particular space-time data samples corrupted by the impulse noise. These latter methods cross-correlate the signal subspace and the noise subspace. To elaborate, consider the one-dimensional direction finding example given in [12] corresponding to (2). If the $\ell$ = 2th antenna’s datum is corrupted by impulse noise and zeros by GT self-normalization, the resulting steering vector

$$
a(\theta_p) = \left[1, 0, \exp\left[-j2\pi\frac{D}{\lambda} \sin\theta_p\right], \ldots, \exp\left[-j2\pi(L-1)\frac{D}{\lambda} \sin\theta_p\right]\right]^T$$

(5)

would be linearly independent from all incident sources’ steering vectors and would increase the signal subspace’s rank by one. In contrast, the proposed $\|\cdot\|_\infty$ normalization applies to the entire $L$-element spatial snapshot, thereby scaling down only that snapshot’s contribution to $R$ but not affecting the signal subspace. The impulse-noise’s influence stays in only the noise subspace; and no false cross-correlation is introduced between the signal subspace and the noise subspace.

---

8Reference [5] attempts to convert the heavy-tail noise to Gaussian density. This requires prior knowledge of the heavy-tail noise’s exact density; and this involves complex computations for each collected data. Hence, [5] would offer neither advantages (1) nor (4).

9Reference [14] normalizes each data-snapshot to unity Frobenius norm (i.e., $\|\cdot\|_2$), thereby keeping only that data-snapshot’s vectorial direction. This Frobenius-norm computation is more complex than the infinity-norm’s identification of the snapshot-vector’s largest magnitude element. Hence, [14] would be inferior to the proposed scheme with respect to advantage (4).

10References [12] and [15] each employs a different Gaussian-tailed normalization of the heavy-tail noise. This involves soft amplitude-clipping, using an exponential function dependent on a real-time estimate of the heavy-tail noise’s rescaled median absolute deviation (RMAD). This RMAD estimate and the exponential soft-clipping function require significantly more computation than the proposed $\|\cdot\|_\infty$ scheme. Hence, [12] and [15] would not have advantages 3 and 4.
V. STATISTICAL ANALYSIS OF THE $\| \cdot \|_\infty$ NORMALIZED IMPULSE-NOISE

This section shows that the proposed $\| \cdot \|_\infty$ normalization will preserve the input-noise’s zero spatial cross-correlation, when the prenormalized noise has a spherically symmetrical distribution on the complex plane, a mean, and a variance.\(^{11}\) Consider the statistically independent, identically and symmetrically distributed components in the $L \times 1$ complex-valued noise vector $\mathbf{n}(t)$. With no signal (i.e., $\mathbf{x}(t) = \mathbf{n}(t)$), define $\mathbf{v}(t) = \mathbf{y}(t)$. Then, the proposed $\| \cdot \|_\infty$ normalization would give

\[
E\{\mathbf{v}(t)\} = \mathbf{0} \quad (6)
\]

\[
E\{\mathbf{v}(t)\mathbf{v}^H(t)\} = \mathbf{D} \quad (7)
\]

where $\mathbf{0}$ symbolizes a vector of all zeroes, and $\mathbf{D}$ denotes an $L \times L$ diagonal matrix with finite entries.

To prove (6): Because $w(t)$ is real-valued, $w(t)\mathbf{n}(t)$ is spherically symmetric. Moreover, $w(t) = \max\{n_1(t), n_2(t), \ldots, n_L(t)\}^{-1} \leq |n_i(t)|^{-1}$ implies that $0 < w(t)|n_i(t)| \leq 1$, $\forall t$. Because the real-valued number $E\{w(t)n_i(t)\} \leq E\{|w(t)|\} \leq \infty$ and because $w(t)n_i(t)$ is spherically symmetric, $E\{w(t)n_i(t)\} = 0$. This proves (6).

To prove (7): First consider each component’s autocorrelation. $E\{v_i(t)v_i^*(t)\} = E\{w(t)n_i(t)n_j^*(t)w^*(t)\} = E\{w^2(t)n_i(t)n_j^*(t)\} \leq E\{|n_i(t)|^{-2}n_i(t)n_j^*(t)\} = 1, \forall t = 1, 2, \ldots, L$. Each entry in $\{v_i(t), \forall t = 1, 2, \ldots, L\}$ thus has a finite variance not exceeding unity. Incidentally $\forall k, \ell \in \{1, 2, \ldots, L\}$ such that $n_k(t) \neq 1/w(t) \neq n_\ell(t)$, it is true that $E\{w^2(t)n_k(t)n_\ell^*(t)\} = E\{|w^2(t)n_k(t)n_\ell^*(t)\}$. Consider now the spatial cross-correlation. $\forall k, \ell \in \{1, 2, \ldots, L\}$ where $\ell \neq k, E\{v_i(t)v_j^*(t)\} = E\{w(t)n_i(t)n_j^*(t)w^*(t)\}$, where $w(t)n_i(t)n_j^*(t)w^*(t)$ is spherically symmetric. Hence, the above expectation must equal zero. The above conclusions prove (7).

VI. SIMULATION ASSESSMENT OF THE PROPOSED SCHEME’S EFFICACY

The proposed $\| \cdot \|_\infty$ normalization preprocessing scheme is compared against the GT normalization preprocessing by Kozick & Sadler [12], the GT normalization preprocessing by Swami & Sadler [15], the SCM-based preprocessing [14], and the FLOM-based approaches [13]. These various preprocessing schemes are applied to a one-dimensional direction finding problem using the eigen-based parameter-estimation method of MUSIC. The impulse noise will be spatio-temporally uncorrelated and modeled as either SoS impulsive noise or contaminated-Gaussian impulsive noise. Subsequent discussion uses the label “IN-MUSIC” to refer to this proposed infinity-norm normalization preprocessing scheme when coupled with MUSIC. FLOM-MUSIC will have its FLOM parameter set at $p = 1$, corresponding to the nearly optimum setting found in [13].

Unless otherwise stated subsequently, $P = 2$ is the number of far-field point-sources emitted from the directions-of-arrival of $10^\circ$ and $16^\circ$ upon a linear array of $L = 8$ isotropic sensors spaced uniformly at half a wavelength. The emitted signals are narrowband, temporally white, not cross-correlated, zero-mean, complex-value, Gaussian-distributed stochastic processes. Each subsequent graph’s data-point consists of $K = 500$ Monte Carlo independent trials, each of which involves $N = 200$ evenly time-sampled snapshots.

Two performance metrics are used: the root-mean-square error (RMSE), of the $P$ direction-of-arrival estimates, is defined as

\[
\text{RMSE} = \frac{1}{P} \sum_{p=1}^{P} \sqrt{\frac{1}{K} \sum_{k=1}^{K} (\hat{\theta}_p(k) - \theta_p)^2} \quad (8)
\]

where $\hat{\theta}_p(k)$ symbolizes the $k$th Monte Carlo trial’s estimate for the $p$th source’s direction-of-arrival. The second metric is the “resolution probability” [6], with “resolution” herein defined to have occurred between two incident sources if:

\[
2V\left(\frac{\theta_1 + \theta_2}{2}\right) < V(\theta_1) + V(\theta_2) \quad (9)
\]

where $V(\theta)$ represents each concerned algorithm’s own uniquely defined pseudospectrum. For MUSIC, $V(\theta) = 1/\mathbf{a}^H(\theta)\mathbf{E}_n\mathbf{E}_n^H\mathbf{a}(\theta)$.

A. IN-MUSIC Versus FLOM-MUSIC, GT-MUSIC and SCM-MUSIC in SoS Heavy-Tail Noise

Figs. 1 and 2, with SoS heavy-tail noise, verify that the proposed $\| \cdot \|_\infty$ normalization preprocessing scheme (i.e., IN-MUSIC) has a higher probability of resolution and a lower RMSE over FLOM-MUSIC [13], GT-MUSIC (Kozick-Sadler) [12], GT-MUSIC (Swami-Sadler) [15], and SCM-MUSIC [14].

Fig. 1 is plotted against $\alpha \in [0.5, 1.0]$ at generalized SNR (GSNR) = 30 dB. Fig. 1 omits FLOM-MUSIC, because FLOM-MUSIC is inapplicable to the $\alpha < 1$ there. Fig. 2 is plotted against various GSNRs at $\alpha = 1.2$.

The aforementioned GSNR [8] for the $p$th incident source is defined as

\[
\text{GSNR}_p = 10\log \frac{\sum_{n=1}^{N} |x_p(n)|^2}{\gamma_N} \quad (10)
\]
Simulations here all have incident sources at 10 dB power and the SoS impulse noise’s dispersion at \( \gamma = 1 \); hence, GS NR \( \rho = \text{GSNR} \), \( \forall \rho \).

**B. IN-MUSIC Versus GT-MUSIC & SCM-MUSIC in Gaussian-Mixture Heavy-Tail Noise**

Figs. 3 and 4, under Gaussian-mixture noise, demonstrate the proposed \( \| \| \infty \) normalization preprocessing scheme’s (i.e., IN-MUSIC’s) superior performance over GT-MUSIC (Kozick-Sadler) [12], GT-MUSIC (Swami-Sadler) [15], and SCM-MUSIC [14]. The Gaussian-mixture noise here has \( \epsilon_2 \in [0,0.5] \), \( \sigma_1^2 = 1 \), and \( \sigma_2^2 = 100 \). Figs. 3 and 4 each plot against \( \epsilon_2 \), with each incident source’s individual power at 10 dB; and Fig. 4 plots the corresponding entities versus the incident signal power, at \( \epsilon_1 = 0.1 \). Figs. 3 and 4 both demonstrate the proposed \( \| \| \infty \) data-dependent zero-memory preprocessing scheme’s (i.e., IN-MUSIC’s) higher probability of resolution and lower RMSE over its three competitors.

**C. IN-MUSIC Versus MUSIC & SCM-MUSIC in Gaussian Noise**

Fig. 5, under zero-mean Gaussian noise, shows that the proposed \( \| \| \infty \) normalization preprocessing scheme (i.e., IN-MUSIC) out-performs SCM-MUSIC [14] but underperforms MUSIC [1] especially at low SNRs. (FLOM-MUSIC degenerates into MUSIC for \( p = \alpha = 2 \) here.) Fig. 5’s Cramer-Rao lower bound is computed using the formula in [2, Sect. IV], which is applicable for joint estimation of the sources’ one-dimensional DOA and noise power, under zero-mean white Gaussian noise.

**VII. CONCLUSIONS**

This above \( \| \| \infty \) normalization DA-ZM linear preprocessing scheme is found to offer better direction
Fig. 3. Proposed $\| \cdot \|_\infty$ normalization zero-memory data-adaptive preprocessed direction-finding algorithm’s estimation performance, compared against three other competing algorithms under Gaussian-mixture additive noise.

Fig. 4. Proposed $\| \cdot \|_\infty$ normalization zero-memory data-adaptive preprocessed direction-finding algorithm’s estimation performance, compared against three other competing algorithms under Gaussian-mixture additive noise. $\epsilon = 0.1$.

Fig. 5. Proposed $\| \cdot \|_\infty$ normalization zero-memory data-adaptive preprocessed direction-finding algorithm’s estimation performance, compared against two other competing algorithms and Cramer-Rao bound under zero-mean white Gaussian additive noise.
finding resolution and accuracy than the nonlinear DA-ZM techniques in [15] (which use various forms of GT normalization), or SCM-MUSIC [14] (which discards the data’s magnitude but keeps only the sign). The present \( || \cdot ||_{\infty} \) normalization’s superior estimation accuracy over GT normalization is due to the preservation of the signal subspace’s noncorrelation from the noise space. The present \( || \cdot ||_{\infty} \) normalization’s better estimation accuracy over SCM-MUSIC is because the present scheme’s infinity-norm is more sensitive against impulse noise than SCM-MUSIC’s Frobenius-norm. The proposed scheme’s above-mentioned advantages exist despite the scheme’s algorithmic and computational simplicity. This \( || \cdot ||_{\infty} \) normalization preprocessing technique could be used for beamforming, detection, classification, and other estimation problems in impulse environments encountered in radar, sonar, wireless, or wired communications. Nonetheless, when the heavy-tail noise’s statistics are at least partially known, the alternate model-specific algorithms in Section I may offer better accuracy and resolution. Moreover, the proposed preprocessing scheme could degrade MUSIC’s performance when the additive noise is actually not heavy-tail but Gaussian.

JIN HE

ZHONG LIU

Dept. of Electronic Engineering
Nanjing University of Science & Technology
Nanjing, Jiangsu 210094
China

KAINAM THOMAS WONG

Dept. of Electronic & Information Engineering
Hong Kong Polytechnic University
Hung Hom, Kowloon
Hong Kong

E-mail: (ktwong@ieee.org)

REFERENCES

[1] Schmidt, R. O.
Multiple emitter location and signal parameter estimation.
IEEE Transactions on Antennas and Propagation, 34, 3


ESPRIT—Estimation of signal parameters via rotational invariance techniques.

Detection in impulsive noise based on robust whitening.

Detection in alpha-stable noise environments based on nonlinear prediction.

[6] Zhang, Q. T.
Probability of resolution of the MUSIC algorithm.

Adaptive robust impulse noise filtering.

[8] Tsakalides, P., and Nikias, C. L.
The robust covariation-based MUSIC (ROC-MUSIC) algorithm for bearing estimation in impulsive noise environments.

Two decades of array signal processing research: The parametric approach.

[10] Sadler, B. M.
Detection in correlated impulsive noise using fourth-order cumulants.

A new concept of adaptive beamforming for moving sources and impulse noise environment.

[12] Kozick, R. J., and Sadler, B. M.
Maximum-likelihood array processing in non-Gaussian noise with Gaussian mixtures.

A subspace-based direction finding algorithm using fractional lower order statistics.

Subspace-based direction-of-arrival estimation using nonparametric statistics.

On some detection and estimation problems in heavy-tailed noise.

[16] Lee, M. S., Katkovnik, V., and Kim, Y. H.
Robust approximate median beamforming for phased array radar with antenna switching.

Performance study of the minimax robust phased array for wireless communications.

CORRESPONDENCE