IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 10, 2018, accepted March 25, 2018, date of publication April 23, 2018, date of current version May 24, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2823381

Blind Source Separation Using Temporal Correlation, Non-Gaussianity and Conditional Heteroscedasticity

SEYYED HAMED FOULADI[1], ILANGKO BALASINGHAM[1,2], (Senior Member, IEEE), KIMMO KANSANEN[1], AND TOR AUDUN RAMSTAD[1]

[1]Department of Electronics and Telecommunications, Norwegian University of Science and Technology, 7491 Trondheim, Norway
[2]Intervention Center, Oslo University Hospital, N-0027 Oslo, Norway

Corresponding author: Seyyed Hamed Fouladi (hamed.fouladi@ntnu.no)

This work was supported by the Research Council of Norway project the Medical Sensing, Localization, and Communication using Ultra Wideband Technology (Phase II) under Grant 225885.

ABSTRACT Independent component analysis separates latent sources from a linear mixture by assuming sources are statistically independent. In real world applications, hidden sources are usually non-Gaussian and have dependence among samples. In such case, both attributes should be considered jointly to obtain a successful separation. To capture sample dependence, a latent source is sometimes modeled by autoregressive or moving average models with an independent and identically distributed error or residual. However, these models are limited by assuming only linear dependence among a source's samples. This paper proposes a new blind source separation algorithm based on an autoregressive-autoregressive conditional heteroscedasticity (AR-ARCH) model, which captures linear correlations, non-Gaussianity, and squared residuals' dependence. The AR part of the AR-ARCH model captures the correlation among samples. The ARCH part of the model captures the non-Gaussianity and nonlinear dependence among samples. The ARCH model also assumes the time-varying conditional variances for sources. We derive the Cramér Rao lower bound (CRLB) for the mixing matrix based on the AR-ARCH model. We perform simulations on both synthetic and real data. The results show that the proposed method outperforms the baseline algorithms especially for a small number of samples and approaches the CRLB.

INDEX TERMS Blind source separation, independent component analysis, autoregressive conditional heteroscedasticity, maximum likelihood, Fisher's information matrix.

I. INTRODUCTION

Blind source separation (BSS) is a method applicable in various areas such as biomedical signal processing, image processing, speech, radar, and sonar signal processing [1]. In many situations of practical interest, observations are instantaneous linear mixtures of hidden source signals that the observations are written as

$$\boldsymbol{x}(n) = \boldsymbol{A}\boldsymbol{s}(n), \quad 1 \leq n \leq N \tag{1}$$

where $\boldsymbol{A}$ is a full rank unknown $L \times L$ mixing matrix, $\boldsymbol{x}(n) = [x_1(n)\; x_2(n)\; \dots\; x_L(n)]^T$ is the L-dimensional observation vector at time n, and $\boldsymbol{s}(n) = [s_1(n)\; s_2(n)\; \dots\; s_L(n)]^T$ is the source vector at time n. BSS models aim at estimating the mixing matrix $\boldsymbol{A}$ and the source vector $\boldsymbol{s}(n)$ using the observation vector $\boldsymbol{x}(n)$.

Independent component analysis (ICA) is a statistical approach to extract source vectors as independently as possible from an observed multidimensional random vector. Many ICA algorithms use sample dependency and the property that samples are not Gaussian distributed (non-Gaussianity) to separate the sources [1]. FastICA [2], JADE [3] and ICA based on entropy bound minimization (EBM) [4] exploit the non-Gaussian property without considering sample dependence within each source.

Some algorithms use sample dependence to separate sources. Examples are Autoregressive (AR) models [5]–[7] and moving-average (MA) models [8]. In [9]–[11], a Markov model takes into account linear sample dependence. However, sources with equal spectra shape cannot be separated based on algorithms exploiting linear source sample dependence only. In [12]–[14], mutual information rate is exploited for minimizing entropy rate (ERBM) in order to build a framework for derivation of algorithms that consider sample dependence and non-Gaussianity jointly.

Considering both temporal structure and non-Gaussianity improves source separation performance [12].

Another statistical property that can be used for source separation is nonstationarity of the variance [5], [7], [15], [16]. In these methods, the local variance of each independent source is assumed to change smoothly as a function of time. In [7] and [16], the local variance is obtained by the average of squared samples of the whitened signal over a time interval. This model in [7] is called the Unifying model. If we assume a specific filter for all signals as the Unifying model irrespective of different properties, the mismatch between the model and the signal gives poor estimation of local variance, which results in degradation in separation performance. Moreover, there are processes called autoregressive conditional heteroscedasticity (ARCH) whose conditional variances are time-varying while their unconditional variances are constant in time. In this situation, if the length of the filter estimating the local variance is large, the variation of conditional variance over time may not be captured. To overcome these deficiencies, we suggest a more general model, AR-ARCH, which captures time-varying conditional variances of sources and autocorrelations of squared residuals along with non-Gaussianity and linear correlation in sources' samples.

The conditional variance of a signal can be estimated using ARCH models, proposed by Engle in [17]. The ARCH model estimates conditional variance from past immediate information. This model was introduced for modeling heavy tailed financial time series. In addition to financial applications, considering conditional variance as a non-constant process has been found realistic in signal processing applications such as speech processing [18]–[21], radar [22]–[24], image processing [25]–[28], and denoising [29].

Parameter estimation is one of the steps in model based methods. The most popular approach for estimating the parameters of the ARCH model is quasi maximum likelihood estimation (QMLE) which needs numerical maximization because it does not admit a closed form expression [30]. Bose and Mukherjee proposed in [31] a two stage least squares (TSLS) method for estimating the parameters of the ARCH model by solving linear equations. This method has a better or at least the same performance as QMLE [31]. Furthermore, this method has very low computational cost compared to QMLE. Also, it has been demonstrated that the estimator performs better than QMLE for small sample sizes in [31]. We use the TSLS method to estimate the ARCH model parameters due to its benefit over QMLE.

In this paper, we propose a BSS method which exploits sample dependence and non-Gaussianity jointly. The proposed algorithm is based on an AR-ARCH model. This means the linear temporal correlations of sources are modeled by an AR model, and the conditional variance and autocorrelation of the squared residuals are estimated by ARCH process. The demixing matrix is estimated by maximizing the likelihood function corresponding to the AR-ARCH model. We analytically demonstrate why the conditional heteroscedasticity assumption for sources is realistic, and show that using the same predefined filter for estimating the conditional variance of all signals is not an accurate assumption. We show analytically that the estimated parameters of the ARCH model give us the capability of capturing the autocorrelation of squared residuals. In addition, we derive a CRLB for the AR-ARCH model.

The paper is organized as follows. In Section II, the AR-ARCH model is presented. Also, reasons for using this model, the conditional heteroscedasticity assumption, and considerations for applying different filters for estimating conditional variances are presented. The proposed separation algorithm based on the AR-ARCH model is explained in Section III. We discuss the separability condition in Section IV. The Cramér Rao lower bound (CRLB) for the demixing matrix is derived in Section V. Finally, simulations are performed on both synthetic and real data. The performance of the proposed algorithm is compared with existing algorithms along with CRLB in Section VI in order to evaluate the effectiveness of the proposed approach. Conclusions are given in Section VII.

II. THE AR-ARCH MODEL AND ITS BENEFITS OVER CONVENTIONAL MODELS

In this section, the benefits of the AR-ARCH model over conventional models are presented. The AR part of the AR-ARCH model captures the signal spectrum. The residual of AR-ARCH process, i.e. the ARCH part, is a white process but the squared ARCH process is a colored process. Since the squared residual of AR-ARCH is not white, we derive the autocorrelation of the squared residual of the ARCH process to show the capability of the model to capture the spectrum of the sources' squared residuals.

A. BACKGROUND ON AR-ARCH

The AR-ARCH process is introduced to capture linear dependence among the sources' samples and time varying conditional variance by AR and ARCH models respectively [17]. The AR model is defined as [32]

$$\begin{aligned} s_i(n) &= \sum_{k=1}^{P} a_{i,k} s_i(n-k) + \varepsilon_i(n), \\ \varepsilon_i(n) &= \sigma_i(n)\zeta_i(n), \quad \zeta_i(n) \sim f_i(\zeta_i). \end{aligned} \tag{2}$$

Here the signal component i at time n, $s_i(n)$, is recursively generated from its past P samples weighted by the coefficients $a_{i,k}$, $k = 1, 2, \ldots, P$ plus an additive and independent process, $\varepsilon_i(n)$. The AR coefficients $a_{i,k}$ are constant in time. $\varepsilon_i(n)$ is a product of $\sigma_i(n)$, the conditional standard deviation, and $\zeta_i(n)$, an independent and identically distributed (i.i.d) unit variance, zero-mean process with probability density function (PDF) $f_i(\zeta_i)$. As indicated, the conditional variance $\sigma_i^2(n)$ can also be time dependent and modeled by the ARCH model. A Q'th order ARCH process is written

$$\sigma_i^2(n) = \beta_{i,0} + \sum_{k=1}^{Q} \beta_{i,k}\varepsilon_i^2(n-k),$$
$$\boldsymbol{\beta}_i = \left[\beta_{i,0}, \beta_{i,1}, \ldots, \beta_{i,Q}\right]^T. \quad (3)$$

The coefficient vector $\boldsymbol{\beta}_i$ consists of positive components only. If $Q > 0$ and $\boldsymbol{\beta}_i \neq \mathbf{0}$, the process $s_i(n)$ and $\varepsilon_i(n)$ are called AR-ARCH process and ARCH process respectively. As it can be seen in (2), $s_i(n)$ is the filtered version of $\varepsilon_i(n)$.

If Q is equal to zero, then, $\varepsilon_i(n)$ is an i.i.d process, i.e.,

$$E\{\varepsilon_i(n)\varepsilon_i(m)\} = 0, \; n \neq m,$$

and $s_i(n)$ in (2) is an ordinary AR process.

Let us denote past immediate information as $\boldsymbol{\Psi}_i(n) = \{s_i(0), \ldots, s_i(n), \varepsilon_i(0), \ldots, \varepsilon_i(n)\}$. Using (2), the conditional mean of $s_i(n)$ is written as

$$\mu_i(n) = E\{s_i(n)|\boldsymbol{\Psi}_i(n-1)\} = \textstyle\sum_{k=1}^{P} a_{i,k}s_i(n-k).$$

Since $\sigma_i^2(n)$ in (3) and $\mu_i(n)$ at time n depend on the past information $\boldsymbol{\Psi}_i(n-1)$, they are deterministic if $\boldsymbol{\Psi}_i(n-1)$ is available. The process $\zeta_i(n)$ in (2) is equal to $(s_i(n) - \mu_i(n))/\sigma_i(n)$ when $\sigma_i(n)$ is nonzero. Therefore, the conditional PDF of $s_i(n)$ is

$$p(s_i(n)|\boldsymbol{\Psi_i(n-1)}) = \frac{1}{\sigma_i(n)} f_i\left(\frac{s_i(n) - \sum_{k=1}^{P} a_{i,k}s_i(n-k)}{\sigma_i(n)}\right). \quad (4)$$

The joint PDF of $s_i(n)$ for $1 \leq n \leq N$ is

$$p(s_i(N), s_i(N-1), \cdots, s_i(0)) = \prod_{n=1}^{N} p(s_i(n)|\boldsymbol{\Psi_i(n-1)}). \quad (5)$$

In Section III, we will use the conditional probability density function in (4) to form the likelihood function for estimating the demixing matrix. In the Section II-B, we show why the conditional variance in (3) is not well estimated by the Unifying model.

B. AUTOCORRELATION OF SQUARED ARCH PROCESS (ARCH AS A NONLINEAR MODEL)

In this section, we explain why $\varepsilon_i(n)$ in (2) should be modeled by the ARCH model. We demonstrate analytically that the autocorrelation of the squared residual can be captured by the ARCH model.

An ARCH process is uncorrelated. However, the ARCH process $\varepsilon_i(n)$ is not i.i.d and the essential characteristic of the ARCH model is that the covariance between $\varepsilon^2(n)$ and $\varepsilon^2(m)$ is not zero when $n \neq m$, i.e., $Cov(\varepsilon^2(n), \varepsilon^2(m)) \neq 0$.

In Appendix A, we derive the autocorrelation of the residual $\varepsilon_i^2(n)$ in (2). If we assume that $\varepsilon_i^2(n)$ is a wide sense stationary (WSS) process, the autocorrelation can be written as

$$r_{\varepsilon_i^2}(m) = \beta_{i,0}E\{\varepsilon_i^2(n)\} + \sum_{k=1}^{Q} \beta_{i,k} r_{\varepsilon_i^2}(m-k) \;\; \text{for } m \geq 0, \quad (6)$$

where

$$E\{\varepsilon_i^2(n)\} = \frac{\beta_{i,0}}{1 - \beta_{i,1} - \cdots - \beta_{i,Q}},$$

and $r_{\varepsilon_i^2}(m)$ is the autocorrelation function of $\varepsilon_i^2(m)$. Let a filter $h_i(k)$ be denoted by $h_i(k) = \delta(k) - \beta_{i,1}\delta(k-1) - \beta_{i,2}\delta(k-2) - \cdots - \beta_{i,Q}\delta(k-Q)$ and $H_i(z)$ is the $\mathcal{Z}$ transform of $h_i(n)$. Since $r_{\varepsilon_i^2}(m)$ is two-sided, its $\mathcal{Z}$-transform is

$$R_{\varepsilon_i^2}(z) = \beta_{i,0}E\{\varepsilon_i^2(n)\}\left(\frac{1}{H_i(z^{-1})} + \frac{1}{H_i(z)} - 1\right). \quad (7)$$

The filter $H_i(z)$ in (7) has Q zeros which result in $2Q$ poles of $R_{\varepsilon_i^2}(z)$. Therefore, it gives the capability to model the spectrum of the squared residual.

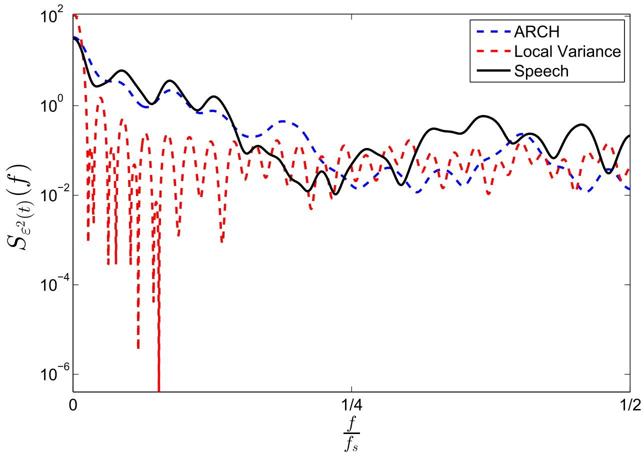

FIGURE 1. Spectrum of the squared residual of a speech signal with the duration 0.3 second and the sampling rate $f_s = 8$ kHz.

An example is provided to clarify why conventional AR model and local variance estimation may not be able to model natural signals. In order to demonstrate that the residual is not independent in time, a speech signal with the duration 0.3 second and sampling rate $f_s = 8$ kHz is modeled by an AR(10) process. Fig. 1 shows the spectrum of the squared residual of the speech signal, the fitted ARCH process, and the Unifying model in [7] which estimates local variance. As it can be seen from Fig. 1, the spectrum of the squared residual is not constant over frequencies, which illustrates that conventional AR model is not suitable for capturing the spectral density of the speech signal, and also, the spectral densities related to local variance estimation is not able to model the autocorrelation of the squared residual. The ARCH process with parameters estimated from the data is a closer fit to the data than the Unifying model.

III. BLIND SOURCE SEPARATION METHOD BASED ON THE AR-ARCH MODEL

A BSS approach based on the AR-ARCH model is presented in this section. The proposed iterative algorithm estimates sources' parameters and unmixing matrix. First, source parameter estimation is presented. Second, the maximum likelihood (ML) estimate of the unmixing matrix based on the AR-ARCH model is obtained. The parameters and unmixing matrix are estimated iteratively.

A. SOURCE PARAMETERS

1) AR COEFFICIENTS

In each iteration, the AR coefficients of sources are estimated using the estimated $\hat{s}_i(n)$

$$\hat{s}_i(n) = \hat{\boldsymbol{w}}_i^T \boldsymbol{x}(n), \tag{8}$$

where $\hat{\boldsymbol{w}}_i$ is the ith column of the estimated unmixing matrix $\hat{\boldsymbol{W}}$ and $\hat{s}_i(n)$ is the ith estimated source. The estimated sources are used to find the AR coefficients $\hat{a}_i(k)$ by solving Yule-Walker equations [33].

2) ARCH PARAMETERS

The ARCH parameters can be obtained by ML estimator. However, it was shown that TSLS algorithm outperforms ML estimator [31]. We thus exploit the TSLS ARCH parameter estimation in [31] for the proposed algorithm.

Each source residual is estimated in all iterations. The estimated residual, $\hat{\varepsilon}_i(n)$, is obtained using $\hat{s}_i(n)$, $\hat{a}_i(k)$ and (2),

$$\begin{aligned}\varepsilon_i(n) &= s_i(n) - \sum_{k=1}^{P} a_{i,k} s_i(n-k) \\ &= \boldsymbol{w}_i^T \boldsymbol{x}(n) - \sum_{k=1}^{P} a_{i,k} \boldsymbol{w}_i^T \boldsymbol{x}(n-k).\end{aligned} \tag{9}$$

Now let random processes z_i, $\boldsymbol{Z}_i$, and $\nu_i(n)$ be defined as $z_i(n) = \varepsilon_i^2(n)$ and

$$\begin{aligned}\boldsymbol{Z}_i(n) &= [1, \varepsilon_i^2(n-1), \ldots, \varepsilon_i^2(n-q)]^T \\ &= [1, z_i(n-1), \ldots, z_i(n-q)]^T, \\ \nu_i(n) &= \zeta_i^2(n) - 1.\end{aligned} \tag{10}$$

Since $\zeta_i(n)$ is a zero-mean and i.i.d process with unit variance, the auxiliary process $\nu_i(n)$ is zero-mean and uncorrelated. Then, one can write using (2), (3) and (10),

$$\begin{aligned}\sigma_i^2(n) &= \boldsymbol{Z}_i^T(n)\boldsymbol{\beta}_i \\ z_i(n) &= \boldsymbol{Z}_i^T(n)\boldsymbol{\beta}_i + \sigma_i^2(n)\nu_i(n).\end{aligned} \tag{11}$$

If the dependence between $\sigma_i^2(n)$ and $\boldsymbol{Z}_i^T(n)$ is ignored, (11) will be a linear regression model with zero-mean error. Therefore, one can solve the equation by ignoring the randomness of $\sigma_i^2(n)$. A preliminary ordinary least-square estimator of regression parameters in (11) can be written

$$\hat{\boldsymbol{\beta}}_i^{pr} = \left(\boldsymbol{Z}_i^T \boldsymbol{Z}_i\right)^{-1} \boldsymbol{Z}_i^T \boldsymbol{Y}_i, \tag{12}$$

where $\boldsymbol{Z}_i$ is $N \times (1+q)$ matrix with the nth row equal to $\boldsymbol{Z}_i^T(n)$ and N is the number of samples $(1 < n < N)$. $\boldsymbol{Y}_i$ is the vector with nth entry $z_i(n)$. The conditional variance obtained by $\hat{\boldsymbol{\beta}}_i^{pr}$ is

$$\tilde{\sigma}_i^2(n) = \boldsymbol{Z}_i^T(n)\boldsymbol{\beta}_i^{pr}. \tag{13}$$

A linear equation is approximated by dividing (11) with $\tilde{\sigma}_i^2(n)$,

$$\frac{z_i(n)}{\tilde{\sigma}_i^2(n)} \approx \frac{\boldsymbol{Z}_i^T(n)}{\tilde{\sigma}_i^2(n)}\boldsymbol{\beta}_i + \nu_i(n). \tag{14}$$

The final estimate by the least squares method is

$$\hat{\boldsymbol{\beta}}_i = \left[\sum_{n=1}^{N}\left(\frac{\boldsymbol{Z}_i(n)\boldsymbol{Z}_i^T(n)}{\tilde{\sigma}_i^4(n)\hat{\boldsymbol{\beta}}_i^{pr}}\right)\right]^{-1}\left[\sum_{n=1}^{N}\left(\frac{\boldsymbol{Z}_i(n)\varepsilon_i^2(n)}{\tilde{\sigma}_i^4(n)\hat{\boldsymbol{\beta}}_i^{pr}}\right)\right]. \tag{15}$$

Here, $\hat{\boldsymbol{\beta}}_i$ is the ARCH parameter vector for the i'th source.

B. UNMIXING MATRIX

In this subsection, we develop the algorithm to estimate the unmixing matrix $\boldsymbol{W}$ based on AR-ARCH model to consider the time-varying conditional variance in (3). The source vector is given by

$$\boldsymbol{s}(n) = \boldsymbol{W}\boldsymbol{x}(n). \tag{16}$$

Each source $s_i(n)$ is modeled by an AR-ARCH model. As a result, the vector $\boldsymbol{x}(n)$ is described as a multivariate AR-ARCH model. We assume that $\boldsymbol{W}$ is orthogonal in order to avoid that $\boldsymbol{W}$ becomes zero or infinity, and the observed data is whitened. It should be noted that the prewhitening step degrades the performance of the BSS, if the linear mixtures of the sources are contaminated with noise [34]. From (4), the log likelihood of $\boldsymbol{W}$ is

$$\begin{aligned}\boldsymbol{L}(\boldsymbol{W}) &= \sum_{n=1}^{N}\sum_{i=1}^{L}\log\left(p\left(s_i(n)|\Psi_i(n-1)\right)\right) \\ &= \sum_{n=1}^{N}\sum_{i=1}^{L} F_i\left(\frac{\boldsymbol{w}_i^T\boldsymbol{x}(n) - \sum_{k=1}^{P} a_{i,k}\boldsymbol{w}_i^T\boldsymbol{x}(n-k)}{\sigma_i(n)}\right) \\ &\quad - \log(\sigma_i(n)),\end{aligned} \tag{17}$$

where the function F_i is the logarithm of the probability density function of ζ_i. The gradient of (17) is given by

$$\begin{aligned}&\nabla_{\boldsymbol{w}_i}\boldsymbol{L}(\boldsymbol{W}) \\ &= \sum_{n=1}^{N}\left(\frac{\boldsymbol{x}(n) - \sum_{k=1}^{P} a_{i,k}\boldsymbol{x}(n-k)}{\sigma_i(n)}\right) \\ &\quad \times F_i'\left(\frac{\boldsymbol{w}_i^T\boldsymbol{x}(n) - \sum_{k=1}^{P} a_{i,k}\boldsymbol{w}_i^T\boldsymbol{x}(n-k)}{\sigma_i(n)}\right) \\ &\quad - \frac{\nabla_{\boldsymbol{w}_i}\sigma_i(n)}{\sigma_i(n)}\left(1 + \left(\frac{\boldsymbol{w}_i^T\boldsymbol{x}(n) - \sum_{k=1}^{P} a_{i,k}\boldsymbol{w}_i^T\boldsymbol{x}(n-k)}{\sigma_i(n)}\right)\right. \\ &\quad \left.\times F_i'\left(\frac{\boldsymbol{w}_i^T\boldsymbol{x}(n) - \sum_{k=1}^{P} a_{i,k}\boldsymbol{w}_i^T\boldsymbol{x}(n-k)}{\sigma_i(n)}\right)\right),\end{aligned} \tag{18}$$

where F_i' is the derivative of F_i. When the unmixing matrix $\boldsymbol{W}$ is updated, the estimated source parameters $a_{i,k}$ and $\boldsymbol{\beta_i}$ are fixed in each step of the iteration. Using (11), the conditional standard deviation $\sigma_i(n)$ is given by

$$\sigma_i(n) = \sqrt{\boldsymbol{Z}_i^T(n)\boldsymbol{\beta}_i}. \tag{19}$$

In [7], it was shown that the second term in (18) is negligible in terms of the first term. Therefore, (18) can be approximated by

$$\nabla_{w_i} \boldsymbol{L}(\boldsymbol{W}) \approx \sum_{n=1}^{N} \left(\frac{\boldsymbol{x}(n) - \sum_{k=1}^{P} a_{i,k}\boldsymbol{x}(n-k)}{\sigma_i(n)} \right) \times F_i' \left(\frac{\boldsymbol{w}_i^T \boldsymbol{x}(n) - \sum_{k=1}^{P} a_{i,k}\boldsymbol{w}_i^T \boldsymbol{x}(n-k)}{\sigma_i(n)} \right). \quad (20)$$

The function F_i' can be derived from the distribution of $\zeta_i(n)$. However, when $\zeta_i(n)$ is super-Gaussian, implying a sharper peak and heavier tail than the corresponding Gaussian probability density function, it is suitable to choose the Generalized Gaussian distribution (GGD)

$$f_i(\zeta_i(n); \gamma_i, \lambda_i) = \frac{\gamma_i}{2\lambda_i \Gamma(\frac{1}{\gamma_i})} \exp\left(-|\zeta_i(n)|^{\gamma_i}\right), \quad (21)$$

where $\gamma_i > 0$ and λ_i are shape and scale parameters, and $\Gamma(.)$ defines the gamma function. The function, $F'(\zeta_i)$, would be

$$F_i'(\zeta_i(n); \gamma_i) = -(\gamma_i)\text{sgn}(\zeta_i(n))|\zeta_i(n)|^{(\gamma_i - 1)}. \quad (22)$$

We estimate the shape parameter γ_i using the method in [35]. The logarithm likelihood function can be solving by updating rule

$$\gamma_i \leftarrow \gamma_i - \frac{\mathcal{G}(\gamma_i)}{\mathcal{G}'(\gamma_i)}, \quad (23)$$

where,

$$\mathcal{G}(\gamma_i) = 1 + \frac{\psi(\frac{1}{\gamma_i})}{\gamma_i} - \frac{\sum_{n=1}^{N} |\zeta_i(n)|^{\gamma_i} \log|\zeta_i(n)|}{\sum_{n=1}^{N} |\zeta_i(n)|^{\gamma_i}} + \frac{\log(\frac{\gamma_i}{N}\sum_{n=1}^{N} |\zeta_i(n)|^{\gamma_i})}{\gamma_i}, \quad (24)$$

and,

$$\mathcal{G}'(\gamma_i) = -\frac{\psi(\frac{1}{\gamma_i})}{\gamma_i^2} - \frac{\psi'(\frac{1}{\gamma_i})}{\gamma_i^3} + \frac{1}{\gamma_i^2} - \frac{\sum_{n=1}^{N} |\zeta_i(n)|^{\gamma_i} (\log|\zeta_i(n)|)^2}{\sum_{n=1}^{N} |\zeta_i(n)|^{\gamma_i}} + \frac{\left(\sum_{n=1}^{N} |\zeta_i(n)|^{\gamma_i} \log|\zeta_i(n)|\right)^2}{\left(\sum_{n=1}^{N} |\zeta_i(n)|^{\gamma_i}\right)^2} + \frac{\sum_{n=1}^{N} |\zeta_i(n)|^{\gamma_i} \log|\zeta_i(n)|}{\gamma_i \sum_{n=1}^{N} |\zeta_i(n)|^{\gamma_i}} - \frac{\log\left(\frac{\gamma_i}{N}\sum_{i=1}^{N} |\zeta_i(n)|^{\gamma_i}\right)}{\gamma_i^2}, \quad (25)$$

where $\psi(.)$ and $\psi'(.)$ denote the digamma and the trigamma functions respectively.

Here, we simply define a stopping criterion

$$\|\nabla_{\boldsymbol{W}}\|_2 < \delta, \quad (26)$$

where δ is a positive threshold.

Algorithm 1 lists the steps of the proposed algorithm.

The computation of steps 3b and 3c dominates the complexity of the algorithm. $\mathcal{O}(L^2N)$ and $\mathcal{O}\left(((Q+1)^2 + L^2)N\right)$ multiplications are required for 3b and 3c, respectively. Therefore, the complexity of the algorithm for one iteration is $\mathcal{O}\left(((Q+1)^2 + L^2)N\right)$ if $L, Q \ll N$.

Algorithm 1 The Steps of the Proposed Algorithm.

Require: $\boldsymbol{x}(n)$, δ, it_{max}, μ
Initialization: $\hat{\boldsymbol{W}} \leftarrow$ *normal distribution*, $\gamma_i = 1$, $it = 0$
Whitening data: $\boldsymbol{x}(n) \leftarrow$ whitened $\boldsymbol{x}(n)$
  while $\|\nabla_{\boldsymbol{W}}\|_2 < \delta$ or $it < it_{max}$ **do**
    $it \leftarrow it + 1$
    $\hat{s}_i(n) \leftarrow \hat{\boldsymbol{w}}_i^T \boldsymbol{x}(n)$
  AR coefficient estimation:
    $\boldsymbol{r}_{i_{(j)}} \leftarrow \frac{1}{N}\sum_{n=1}^{N} \hat{s}_i(n)\hat{s}_i(n-j), \ 1 \le j \le P$
    $\boldsymbol{R}_{i_{(j,k)}} \leftarrow \frac{1}{N}\sum_{n=1}^{N} \hat{s}_i(n)\hat{s}_i(n-|j-k|), \ 1 \le j,k \le P$
    $\hat{\boldsymbol{a}}_i \leftarrow \boldsymbol{R}_i^{-1}\boldsymbol{r}_i$
  Residual estimation:
    $\hat{\varepsilon}_i(n) \leftarrow \hat{s}_i(n) - \sum_{k=1}^{P} \hat{\boldsymbol{a}}_{i_{(k)}}\hat{s}_i(n-k)$
  ARCH coefficient estimation:
    $\boldsymbol{Y}_i \leftarrow [\hat{\varepsilon}_i^2(1), \hat{\varepsilon}_i^2(2), \cdots, \hat{\varepsilon}_i^2(N)]^T$
    $\boldsymbol{Z}_i(n) \leftarrow [1, \hat{\varepsilon}_i^2(n-1), \ldots, \hat{\varepsilon}_i^2(n-q)]^T$
    The nth row of $\boldsymbol{Z}_i \leftarrow \boldsymbol{Z}_i^T(n)$
    $\hat{\boldsymbol{\beta}}_i^{pr} \leftarrow \left(\boldsymbol{Z}_i^T\boldsymbol{Z}_i\right)^{-1}\boldsymbol{Z}_i^T\boldsymbol{Y}_i$
    $\tilde{\sigma}_i^2(n) \leftarrow \boldsymbol{Z}_i^T(n)\boldsymbol{\beta}_i^{pr}$
    $\hat{\boldsymbol{\beta}}_i \leftarrow \left[\sum_{n=1}^{N}\left(\frac{\boldsymbol{Z}_i(n)\boldsymbol{Z}_i^T(n)}{\tilde{\sigma}_i^4(n)\hat{\boldsymbol{\beta}}_i^{pr}}\right)\right]^{-1}\left[\sum_{n=1}^{N}\left(\frac{\boldsymbol{Z}_i(n)\varepsilon_i^2(n)}{\tilde{\sigma}_i^4(n)\hat{\boldsymbol{\beta}}_i^{pr}}\right)\right]$
    $\hat{\sigma}_i^2(n) \leftarrow \boldsymbol{Z}_i^T(n)\hat{\boldsymbol{\beta}}_i$
    $\hat{\zeta}_i(n) \leftarrow \frac{\hat{\varepsilon}_i(n)}{\hat{\sigma}_i(n)}$
    $\gamma_i \leftarrow \gamma_i - \frac{g(\gamma_i)}{g'(\gamma_i)}$
  Demixing matrix estimation:
    $\nabla_{\hat{\boldsymbol{W}}} \leftarrow \sum_{n=1}^{N}\left(\frac{\boldsymbol{x}(n)-\sum_{k=1}^{P}\hat{\boldsymbol{a}}_{i_{(k)}}\boldsymbol{x}(n-k)}{\sigma_i(n)}\right) \times F_i'\left(\hat{\zeta}_i(n)\right)$
    $\hat{\boldsymbol{W}} \leftarrow \hat{\boldsymbol{W}} + \mu\nabla_{\hat{\boldsymbol{W}}}$
  Orthogonalization:
    $\hat{\boldsymbol{W}} \leftarrow \left(\hat{\boldsymbol{W}}\hat{\boldsymbol{W}}^T\right)^{-\frac{1}{2}}\hat{\boldsymbol{W}}$
  end while

IV. SEPARABILITY OF ARCH PROCESSES

In this section, we discuss the condition under which the model of (1) is separable when sources are ARCH processes. Each of the properties, sample dependence, non-Gaussianity, and time-varying conditional variance, can be used to separate sources if the respective assumption is met. If sources i and j have similar spectral shapes, then this implies non-identifiability of the sources [1, Ch. 7.3]. Therefore, if two processes have the same $a_{i,k}$, $i = 1, \cdots, P$, then the linear combination will have the same regression coefficients. Therefore, the sources will not be separable using the regression coefficients as a result.

If there is at most one $\varepsilon_i(n)$, Gaussian for $i \in \{1, \cdots, L\}$, sources will be separable [36]. The following theorem describes the condition which should be satisfied for ARCH processes to be separable where $\zeta_i(n)$s in (3) are Gaussian processes.

Theorem 1: If sources are ARCH processes with Gaussian distributed innovation processes $\zeta_i(n)$ in (2) and the fourth order moment of $\varepsilon_i(n)$ exists, the model of (1) is separable if the mixing matrix $\boldsymbol{A}$ is full rank and at most one source has parameters vector $\boldsymbol{\beta}_i = [\beta_i^0 \; \mathbf{0}_{1\times Q}]^T$.

Proof: See Appecdix VII. □

Theorem 1 shows that sources are separable if the parameters $\beta_{i,k}$ for $1 \leq k \leq Q$ are nonzero even if the innovation process $\zeta_i(n)$ in (2) is a Gaussian process. We assumed that the fourth order moment exists for the ARCH process $\varepsilon_i(n)$, i.e., the condition $3 \times \sum_{k=1}^{Q} (\beta_{i,k})^2 < 1$ needs to be satisfied (See Appendix B).

The condition on the fourth order moment is restrictive. The following theorem is a result on the summation of the two ARCH processes with the same parameter vectors with no assumption on existence of the fourth moment for $\varepsilon_i(n)$.

Theorem 2: A linear mixture of two ARCH processes $\varepsilon_i(n)$, $i = 1, 2$, with the same parameter vectors $\boldsymbol{\beta}_i = \left[\beta_{i,0}, \beta_{i,1}, \ldots, \beta_{i,Q}\right]^T$, will not result in the same ARCH parameters, $\beta_{i,1:Q}$, except for the case that $\varepsilon_i(n)$, $i = 1, 2$, are not conditionally heteroscedastic, where $\zeta_i(n)$ is a Gaussian process and $\boldsymbol{\beta}_i = [\beta_{i,0} \; \mathbf{0}_{1\times Q}]^T$ is the vector with the constant parameter $\beta_{i,0}$.

Proof: See Appecdix VII. □

In order to find $\boldsymbol{W}$, one may maximize the likelihood function $\boldsymbol{L}$,

$$\boldsymbol{L}(\boldsymbol{W}) = -\sum_{i=1}^{L} \log\left(p\left(\boldsymbol{w}_i^T \boldsymbol{x}(N), \cdots, \boldsymbol{w}_i^T \boldsymbol{x}(1)\right)\right) \quad \text{s.t. } \boldsymbol{x}(n) = \boldsymbol{A}\boldsymbol{s}(n) \tag{27}$$

The Log-likelihood function in (27) is minimized if the estimated sources $\hat{s}_i = \hat{\boldsymbol{w}}_i \boldsymbol{x}$ are samples of the joint PDFs of the primary sources s_i for all $i \in \{1, \cdots, L\}$. Theorem 2 shows that the mixture of two ARCH processes with the same parameter vector is not a process with the same parameter vectors or the same joint PDF as the primary sources. Therefore, $\hat{s}_i$ is equal to s_i if the condition in Theorem 2 is satisfied and the minimum of the Log-likelihood function is achieved.

V. PERFORMANCE ANALYSIS

The CRLB for $\boldsymbol{W}$ can be obtained by the Fisher information matrix (FIM) and can be derived as in [1, Ch. 7.3]. The covariance error of the parameter estimator $\hat{\boldsymbol{\theta}}$ is bounded by

$$E\left\{\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)^T\right\} \geq \boldsymbol{J}^{-1}. \tag{28}$$

where the parameter $\boldsymbol{\theta}$ corresponds to

$$\begin{aligned}\boldsymbol{\theta} = [&\boldsymbol{H}_{1,2}, \boldsymbol{H}_{2,1}, \cdots, \boldsymbol{H}_{L-1,L}, \boldsymbol{H}_{L,L-1},\\ &a_{1,1}, \cdots, a_{i,P}, \cdots, a_{L,1}, \cdots, a_{L,P},\\ &\beta_{1,0}, \cdots, \beta_{1,Q}, \beta_{L,0}, \cdots, \beta_{L,Q}]^T.\end{aligned} \tag{29}$$

where $\boldsymbol{H} = \boldsymbol{W}\boldsymbol{A}$ and $\boldsymbol{H}_{i,j}$ is the element in the i'th row and the j'th column of the matrix $\boldsymbol{H}$. We obtain the lower band when $\boldsymbol{A} = \boldsymbol{I}$. This lower bound can be used other invertible matrices due to equivariance property [37]. Then, the lower bound obtained for $\boldsymbol{H}$ can be exploited for $\boldsymbol{W}$. The diagonal elements of matrix $\boldsymbol{H}$ is assumed to be 1. $N(N-1)$ elements of $\boldsymbol{H}$ would be considered in the vector parameters $\boldsymbol{\theta}$. The FIM, $\boldsymbol{J}$, is given by

$$\boldsymbol{J} = E\left\{\left(\frac{\partial(\log(f(\boldsymbol{x}|\boldsymbol{\theta})))}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial(\log(f(\boldsymbol{x}|\boldsymbol{\theta})))}{\partial\boldsymbol{\theta}}\right)^T\right\}. \tag{30}$$

The FIM $\boldsymbol{J}$ can be formed

$$\boldsymbol{J} = \begin{bmatrix} \boldsymbol{J}_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \boldsymbol{J}_{\frac{L(L-1)}{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \boldsymbol{J}_{a1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \boldsymbol{J}_{aN} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \boldsymbol{J}_{\beta 1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boldsymbol{J}_{\beta N} \end{bmatrix}, \tag{31}$$

where $\boldsymbol{J}_i$ for $i \in \{1, 2, \cdots, L(L-1)/2\}$ is a 2 by 2 matrix. The matrices $\boldsymbol{J}_{ai}$ and $\boldsymbol{J}_{\beta i}$, are P by P and $(Q+1)$ by $(Q+1)$ matrices respectively. In Appendix D, the elements of FIM are obtained.

The lower bound for the mixing matrix elements is given by (See Appendix D).

$$E\{\boldsymbol{H}_{i,j}^2\} \geq \kappa_{i,j}(\boldsymbol{\beta}_i, \boldsymbol{\beta}_j)$$

$$\kappa_{i,j}(\boldsymbol{\beta}_i, \boldsymbol{\beta}_j) = \frac{1}{N}\frac{\lambda_j \sigma_{\varepsilon_i}^2 E\left\{\frac{1}{\sigma_j^2(n)}\right\}}{\lambda_j \sigma_{\varepsilon_i}^2 E\left\{\frac{1}{\sigma_j^2(n)}\right\} \lambda_i \sigma_{\varepsilon_j}^2 E\left\{\frac{1}{\sigma_i^2(n)}\right\} - 1}. \tag{32}$$

where $\sigma_{\varepsilon_i}^2 = \frac{\beta_{i,0}}{1-\beta_{i,1}-\cdots-\beta_{i,Q}}$ and λ_i is $E\left\{\left(F_i'(\zeta_i(n))\right)^2\right\}$.

Proposition 1: If no conditional heteroscedasticity exists, i.e., $\boldsymbol{\beta}_i = \boldsymbol{\beta}_j = [1 \; \mathbf{0}_{1\times q}]^T$, the term on right hand side in (32) equals to $\frac{1}{N}\frac{\lambda_j \frac{\sigma_{\varepsilon_i}^2}{\sigma_{\varepsilon_j}^2}}{\lambda_j\lambda_i - 1}$ which leads to the same result in [38] when $\varepsilon_i(n)$ is assumed an i.i.d process.

Proof: Since $\boldsymbol{\beta}_i = [1 \; \mathbf{0}_{1\times q}]^T$, then, $\sigma_i^2(n) = 1$. Therefore,

$$\sigma_{\varepsilon_i}^2 E\left\{\frac{1}{\sigma_i^2(n)}\right\} = \sigma_{\zeta_i}^2 = 1 \tag{33}$$

where $\sigma_{\zeta_i}^2$ is the variance of $\zeta_i(n)$. By substituting (33) in (32), the term $\kappa_{i,j}(\boldsymbol{\beta}_i, \boldsymbol{\beta}_j)$ would be equal to $\frac{1}{N}\frac{\lambda_j \frac{\sigma_{\varepsilon_i}^2}{\sigma_{\varepsilon_j}^2}}{\lambda_j\lambda_i - 1}$. □

Proposition 2: The following inequality holds for $\kappa_{i,j}(\boldsymbol{\beta}_i, \boldsymbol{\beta}_j)$ in (32)

$$\kappa_{i,j}(\boldsymbol{\beta}_i, \boldsymbol{\beta}_j) \leq \frac{1}{N}\frac{\lambda_j \frac{\sigma_{\varepsilon_i}^2}{\sigma_{\varepsilon_j}^2}}{\lambda_j\lambda_i - 1}. \tag{34}$$

Proof: See Appendix E. □

The maximum of $\kappa_{i,j}(\boldsymbol{\beta}_i, \boldsymbol{\beta}_j)$ is $\frac{1}{N}\frac{\lambda_j \frac{\sigma^2_{\varepsilon_i}}{\sigma^2_{\varepsilon_j}}}{\lambda_j\lambda_i - 1}$ that is equal to the case with conditionally homoscedastic sources as in Proposition 1. The result of Proposition 2 in (34) shows that if the conditional heteroscedasticity exists in sources, a better performance can be achieved compared to the case that $\varepsilon_i(n)$ is an i.i.d process.

Proposition 3: If the distribution of $\zeta_i(n)$ is Gaussian and $\boldsymbol{\beta}_i \rightarrow [\beta_{i,0}\ \mathbf{0}_{1\times q}]^T$, $\boldsymbol{\beta}_j \rightarrow [\beta_{j,0}\ \mathbf{0}_{1\times q}]^T$, then, $\kappa_{i,j} \rightarrow \infty$.

Proof: When the distribution of $\zeta_i(n)$ is Gaussian for all i, $\lambda_i = 1$ and $\lambda_j = 1$. In this case, the numerator in (32) is nonzero. Therefore, it is sufficient to show that the denominator goes to zero. We can write

$$\lim_{\boldsymbol{\beta}_{i,j} \rightarrow \begin{bmatrix} \beta^0_{i,j} \\ i \neq j \end{bmatrix}} \sigma^2_{\varepsilon_i} E\left\{\frac{1}{\sigma^2_i(n)}\right\} = \sigma^2_{\zeta_i}. \tag{35}$$

If we use the limit obtained in (35) for the denominator in (32), it goes to zero and then $\kappa_{i,j} \rightarrow \infty$. □

In order to show the behavior of $\kappa_{i,j}$, we generate 10^7 samples of two ARCH processes in (2) where $\zeta_i(n)$ and $\zeta_j(n)$ are Gaussian processes with zero mean and unit variance. The sources' parameter vectors are $\boldsymbol{\beta}_i = [1\ \beta_{i,1}]^T$ and $\boldsymbol{\beta}_j = [1\ \beta_{j,1}]^T$ where $\beta_{i,1}$ and $\beta_{j,1}$ are equal to β and increase from 0 to 0.9. In order to obtain $\kappa_{i,j}$ in (32), the values for $E\left\{\frac{1}{\sigma^2_i(n)}\right\}$ and $E\left\{\frac{1}{\sigma^2_j(n)}\right\}$ are calculated numerically. In Fig. 2, the value of $\kappa_{i,j}$ is plotted. As it can be seen in Fig. 2 and Proposition 3, as $\boldsymbol{\beta}_i$ and $\boldsymbol{\beta}_j$ go to $\boldsymbol{\beta}_i = [1\ 0]^T$ and $\boldsymbol{\beta}_i = [1\ 0]^T$, the value of $\kappa_{i,j}$ goes to infinity, which means that the sources are not separable.

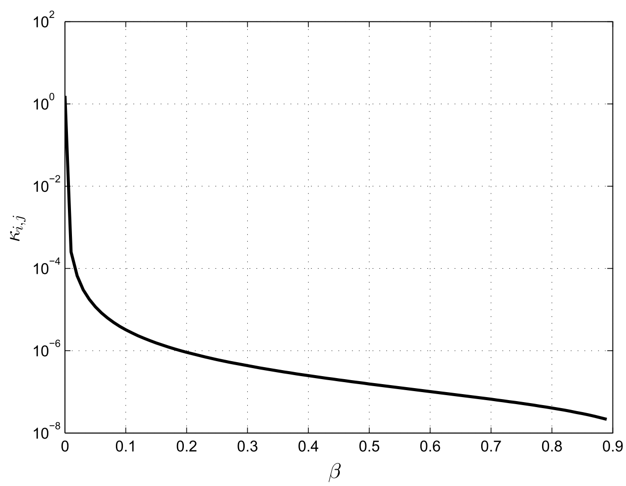

FIGURE 2. $\kappa_{i,j}$ behavior in terms of β.

VI. EXPERIMENTAL RESULTS

We study the performance of different algorithms including AR-ARCH, the Unifying model [7], ERBM, SOBI [6], FASTICA and JADE. FASTICA and JADE separate the sources to be as non-Gaussian as possible while SOBI considers sample dependence. The ERBM approach takes into account both non-Gaussianity and sample dependence, and the Unifying model also assumes local variances of sources smoothly change over time. For the Unifying model it is assumed that the distribution of $\zeta_i(n)$ is Laplacian as in [7]. We consider AR(3)-ARCH(8) model for our approach in experimental results. The normalized interference to source ratio (ISR) is used to compare the approaches. Simulated and natural data are used to study the performance. For simulated data, the performance is compared with the CRLB to demonstrate that the assumed model is efficient when the data follows the model which captures linear and non-linear dependence of samples and non-Gaussianity. We note that the CRLB values plotted are actually the induced Cramér-Rao lower bound (iCRLB) [39] because it is a bound on $\boldsymbol{H}$ rather than $\boldsymbol{W}$. In the experiments,

$$\text{ISR} = \frac{1}{L(L-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^{L} E\{\boldsymbol{H}^2_{i,j}\}. \tag{36}$$

The term $E\{\boldsymbol{H}^2_{i,j}\}$ is calculated using averaging over all trials for each algorithms.

The experiments have been carried out using the MATLAB software on an Intel Core i7 CPU 2.7 GHz processor and 16 GB RAM.

A. EXPERIMENT 1 (SOURCES GENERATED BY THE ASSUMED MODEL)

We study the performance of different algorithms as a function of of number of samples compared with the corresponding CRLB. In order to show that the proposed algorithm is efficient when the data follows the assumed model, three sources are generated based on AR(1)-ARCH(5) with Laplacian distributed $\zeta_i(n)$ by using (2) and (3). The AR coefficients of the sources are selected as $a_{1,1} = 0.9$, $a_{2,1} = 0.5$ and $a_{3,1} = 0$. By the chosen AR coefficients, different linear correlations in time are obtained for sources. The ARCH parameters are chosen as $\boldsymbol{\beta_1} = [1\ 0\ 0\ 0\ 0.4\ 0.5]$, $\boldsymbol{\beta_2} = [1\ 0.5\ 0.4\ 0\ 0\ 0\ 0]$ and $\boldsymbol{\beta_3} = [1\ 0\ 0\ 0.5\ 0.4\ 0\ 0]$ to describe the conditional variances of the sources with different lags of squared residual. For each source, the summation of the ARCH parameters is 0.9 which makes the sources heteroscedastic and the summation is less than 1 which generates wide sense stationary sources. By the chosen ARCH parameters, the conditional variances of sources depend on different time lags of the squared residual. The number of samples is in the range from 100 to 10000. The unconditional variances of the sources are set to be 1. The sources are mixed by a randomly generated mixing matrix with entries drawn i.i.d from the normal distribution. ISR is calculated for each source and is averaged over 100 runs.

Fig. 3 shows the performance of the algorithms along with iCRLB. We observe that AR-ARCH converges close to the iCRLB as the number of samples increases. ERBM is flexible on modeling non-Gaussianity in sources, therefore, a better

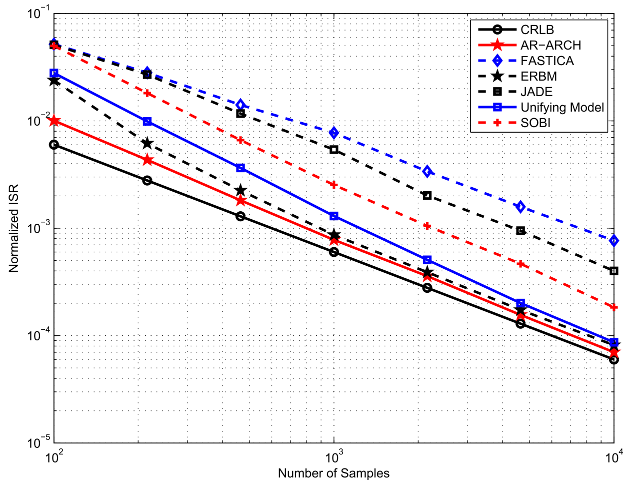

FIGURE 3. ISR as function of samples size in the separation of three heteroscedastic sources with Laplacian distributed $\zeta_i(n)$.

modeling is obtained by increasing the number of samples. The Unifying model also provides a desirable result when the number of samples becomes larger. However, since this model cannot capture the autocorrelation of squared residual of the source well, its performance is 19% worse than AR-ARCH model at 10^4 samples. The performance of FASTICA and JADE is poor due the fact that temporal correlation deteriorates the performance of ICA algorithms. The main superiority of AR-ARCH method over other approaches is its performance when a low number of samples is available. This advantage is due to the accuracy of TSLS approach in estimation of ARCH parameters for each source using a small number of samples.

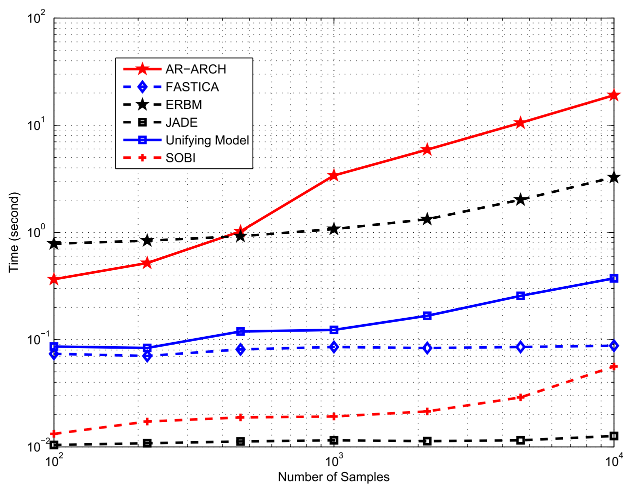

FIGURE 4. CPU time (in seconds) in the separation of three heteroscedastic sources with Laplacian distributed $\zeta_i(n)$ in terms of number of samples.

Fig. 4 shows the processing time of all of the algorithms. As it can be seen from the figure, the AR-ARCH approach has the second highest time consumption after ERBM for low number of samples. By increasing the number of samples, the time consumption for the AR-ARCH algorithm increases and have the highest time consumption for the sample sizes larger than 500. At sample size 1000 or higher, the AR-ARCH approach has the highest processing time. Thus, the drawback of the proposed approach is that it is not efficient for large sample size.

B. EXPERIMENT 2 (MODEL MISMATCH)

We evaluate the algorithm under the condition when there is a mismatch between the generated sources and the model. We generate the sources using AR(1)-ARCH(6) where the AR and ARCH parameters are the same as Experiment 1 except $\zeta_i(n)$ which is Gaussian distributed in this experiment. For the proposed algorithm and the Unifying model, we assume a Laplacian distribution for $\zeta_i(n)$ for all i which provides the model mismatch. The number of samples is in the range from 100 to 10000.

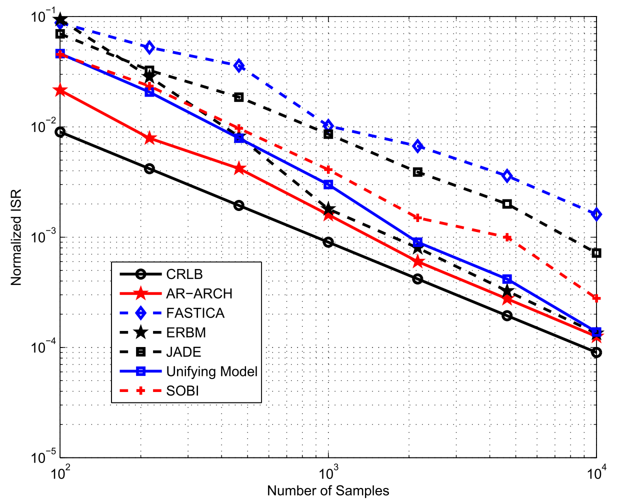

FIGURE 5. ISR as function of samples size in the separation of three heteroscedastic sources with Gaussian distributed $\zeta_i(n)$.

Fig. 5 shows the ISR values as function of number of samples. As observed in Fig. 5, in general, AR-ARCH approach has the best overall performance among all these algorithms especially for the small sample size case. ERBM does not perform well for a small number of samples because it cannot capture sample dependence and non-Gaussianity with low sample size.

C. EXPERIMENT 3 (NATURAL DATA)

We perform BSS experiments with natural data containing six speech sources.[1] The mixing matrix is generated randomly by normal distribution in each trial. The ISR is calculated for each algorithm and is averaged over 100 runs. The sample size is between 100 to 10000.

As observed in Fig. 6, ERBM, Unifying model, AR-ARCH model successfully separate sources. Since FASTICA did not converge in many of the trials, its result has been excluded from the figure. In general, AR-ARCH and ERBM have the

[1] [Online]. Available: http://ecs.utdallas.edu/loizou/speech/noizeus/

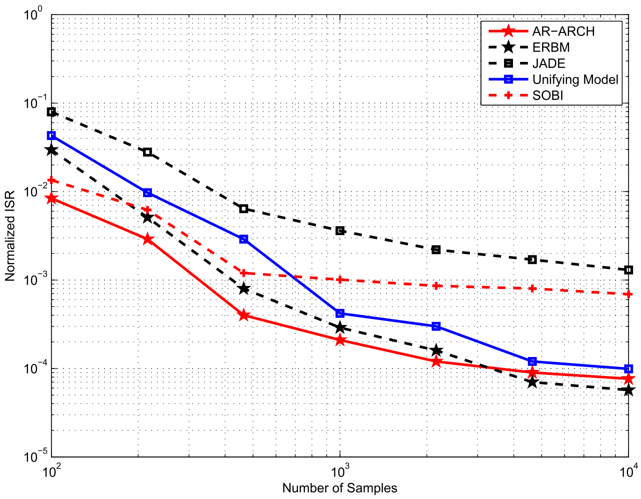

FIGURE 6. ISR as function of samples size in the separation of artificial mixtures of six speech sources.

best overall performance among all these algorithms. For small sample sizes, AR-ARCH outperforms ERBM. For large sample size, ERBM performs better than AR-ARCH model. Since, a long interval of speech is non-stationary in nature, the ARCH parameters are not well estimated. Therefore, the AR-ARCH performance is worse than ERBM for large sample size.

VII. CONCLUSION

A blind source separation method was proposed based on AR-ARCH modeling. The algorithm used time-varying conditional variance, autocorrelation, and non-Gaussianity of residuals for extracting sources. The conditional variance of each source was considered as an AR process, and the ARCH parameters were estimated to find the conditional variance. The key part was modeling the autocorrelation of the squared residual of the sources using ARCH, which showed how to model the joint PDF of the sources. We used the TSLS approach to estimate the ARCH parameters which performed well where few samples were available. It was shown that if the conditional heteroscedasticity existed in the sources, they would be separable even if $\zeta_i(n)$s were Gaussian. Finally, experimental results were presented and showed that the proposed method outperformed the conventional algorithms especially for a low number of samples and converged to the iCRLB as the number of samples increases. However, the proposed algorithm was slower than other approaches. In each iteration, the estimation of ARCH parameters was time consuming and grew by increasing the number of samples. The proposed algorithm improved the source separation performance at the cost of processing time.

APPENDIX A
AUTOCORRELATION OF SQUARED RESIDUAL OF AN ARCH PROCESS

We derive the autocorrelation of residual $\varepsilon_i^2(n)$. If we assume that $\varepsilon_i^2(n)$ is WSS, then, the autocorrelation of $\varepsilon_i^2(n)$ can be written using (2) and (3) as

$$r_{\varepsilon_i^2}(k) = E\{\varepsilon_i^2(n)\varepsilon_i^2(n-k)\} = E\{\sigma_i^2(n)\zeta_i^2(n)\varepsilon_i^2(n-k)\}. \tag{37}$$

Due to independence of $\zeta_i(n)$ from the other terms in (37),

$$\begin{aligned} r_{\varepsilon_i^2}(m) &= E\{\zeta_i^2(n)\}E\{\sigma_i^2(n)\varepsilon_i^2(n-m)\} \\ &= E\{\sigma_i^2(n)\varepsilon_i^2(n-m)\} \\ &= \beta_{i,0}E\{\varepsilon_i^2(n-k)\} \\ &\quad + \sum_{k=1}^{Q}\beta_{i,k}E\{\varepsilon_i^2(n-k)\varepsilon_i^2(n-m)\} \\ &= \beta_{i,0}E\{\varepsilon_i^2(n)\} + \sum_{k=1}^{Q}\beta_{i,k}r_{\varepsilon_i^2}(m-k) \\ &\quad \text{for } m \geq 0. \end{aligned} \tag{38}$$

If we define $h_i(k) = \delta(k) - \beta_{i,1}\delta(k-1) - \beta_{i,2}\delta(k-2) - \cdots - \beta_{i,Q}\delta(k-Q)$, then, the $\mathcal{Z}$-transform of (38) is obtained by (7).

The terms $E\{\varepsilon_i^2(n)\}$ and $H_i(z)$ are the unconditional variance of $\varepsilon_i(n)$ and the $\mathcal{Z}$-transform of $h_i(k)$ respectively. The unconditional variance of $\varepsilon_i(n)$ can be derived based on (2) and (3)

$$\begin{aligned} E\{\varepsilon_i^2(n)\} = &\quad E\{\sigma_i^2(n)\zeta_i^2(n)\} = E\{\sigma_i^2(n)\} \\ = &\quad \beta_{i,0} + \sum_{k=1}^{Q}\beta_{i,k}E\{\varepsilon_i^2(n-k)\}. \end{aligned} \tag{39}$$

Therefore, the unconditional variance of $\varepsilon_i(n)$ equals to

$$\begin{aligned} \sigma_{\varepsilon_i}^2 &= E\{\varepsilon_i^2(n)\} \\ &= \frac{\beta_{i,0}}{1-\beta_{i,1}-\cdots-\beta_{i,Q}}. \end{aligned} \tag{40}$$

Also, if $\varepsilon_i(n)$ is the error of the AR process $s_i(n)$ in (2), the unconditional variance of $s_i(n)$ is given by

$$\begin{aligned} \sigma_{s_i}^2 &= \sum_{k=1}^{P}a_{i,k}r_{s_i}(k) + \sigma_{\varepsilon_i}^2 \\ &= \sum_{k=1}^{P}a_{i,k}r_{s_i}(k) + \frac{\beta_{i,0}}{1-\beta_{i,1}-\cdots-\beta_{i,Q}}. \end{aligned} \tag{41}$$

According to (7), the coefficients of $H_i(z)$ give us freedom to model the autocorrelation of $\varepsilon_i^2(n)$.

APPENDIX B
PROOF OF THEOREM 1

Proof: In this appendix, we find the condition for separability if $\zeta_i(n)$s are Gaussian. Based on [36, Th. 4], if at most one source is Gaussian, then, the model of (1) is separable. The kurtosis of the ARCH process $\varepsilon_i(n)$, is given by

$$\text{Kurt}(\varepsilon_i(n)) = \frac{E\{\varepsilon_i^4(n)\}}{\sigma_{\varepsilon_i}^4}, \tag{42}$$

where the $E\{\varepsilon_i^4(n)\}$ is given by

$$E\{\varepsilon_i^4(n)\} = E\{\sigma_i^4(n)\}E\{\zeta_i^4(n)\} = 3E\{\sigma_i^4(n)\}, \quad (43)$$

and $E\{\sigma_i^4(n)\}$ is given by

$$\begin{aligned} E\left\{\sigma_i^4(n)\right\} &= E\left\{\left(\beta_{i,0}+\sum_{k=1}^{Q}\beta_{i,k}\varepsilon_i^2(n-k)\right)^2\right\} \\ &= E\Big\{(\beta_{i,0})^2+\sum_{k=1}^{Q}(\beta_{i,k})^2\varepsilon_i^4(n-k) \\ &\quad +\sum_{\substack{k_1=1\\k_1\neq k_2}}^{Q}\sum_{k_2=1}^{Q}\beta_{i,k_1}\beta_{i,k_2}\varepsilon_i^2(n-k_1)\varepsilon_i^2(n-k_2) \\ &\quad +2\beta_{i,0}\sum_{k=1}^{Q}\beta_{i,k}\varepsilon_i^2(n-k)\Big\} \\ &= (\beta_{i,0})^2+\sum_{k=1}^{Q}(\beta_{i,k})^2E\{\varepsilon_i^4(n-k)\} \\ &\quad +\sum_{\substack{k_1=1\\k_1\neq k_2}}^{Q}\sum_{k_2=1}^{Q}\beta_{i,k_1}\beta_{i,k_2}r_{\varepsilon_i^2}(k_1-k_2) \\ &\quad +2\beta_{i,0}\sum_{k=1}^{Q}\beta_{i,k}\sigma_{\varepsilon_i}^2. \end{aligned} \quad (44)$$

Using (43) and (44), $E\left\{\sigma_i^4(n)\right\}$ is given by

$$E\left\{\sigma_i^4(n)\right\} = \frac{(\beta_{i,0})^2+2\beta_{i,0}\sum_{k=1}^{Q}\beta_{i,k}\sigma_{\varepsilon_i}^2}{1-3\times\sum_{k=1}^{Q}(\beta_{i,k})^2} + \frac{\sum_{\substack{k_1,k_2=1\\k_1\neq k_2}}^{Q}\beta_{i,k_1}\beta_{i,k_2}r_{\varepsilon_i^2}(k_1-k_2)}{1-3\times\sum_{k=1}^{Q}(\beta_{i,k})^2}. \quad (45)$$

The kurtosis of $\varepsilon_i(n)$ using (45) and (42) is given by

$$\mathrm{Kurt}(\varepsilon_i(n)) = 3\left(\frac{\beta_{i,0}^2+2\beta_{i,0}\sum_{k=1}^{Q}\beta_{i,k}\sigma_{\varepsilon_i}^2}{\sigma_{\varepsilon_i}^4\left(1-3\times\sum_{k=1}^{Q}\beta_{i,k}^2\right)} + \frac{\sum_{\substack{k_1,k_2=1\\k_1\neq k_2}}^{Q}\beta_{i,k_1}\beta_{i,k_2}r_{\varepsilon_i^2}(k_1-k_2)}{\sigma_{\varepsilon_i}^4\left(1-3\times\sum_{k=1}^{Q}(\beta_{i,k})^2\right)}\right). \quad (46)$$

In order to obtain a positive value for the right hand side of (46), the term, $1-3\times\sum_{k=1}^{Q}(\beta_{i,k})^2$, needs to be positive. If $\mathrm{Kurt}(\varepsilon_i(n))$ is not equal to 3, the process $\varepsilon_i(n)$ is non-Gaussian. We prove the term in the parenthesis in the numerator of (46) is larger than denominator of (46). We use the fact that the minimum of $r_{\varepsilon_i^2}(m)$ happens when $\varepsilon_i^2(n)$ and $\varepsilon_i^2(n-m)$ are independent or $m\to+\infty$.

$$r_{\varepsilon_i^2}(m) \geq \lim_{m\to+\infty} r_{\varepsilon_i^2}(m) = E\{\varepsilon_i^2(n)\}E\{\varepsilon_i^2(n-m)\} = \sigma_{\varepsilon_i}^4. \quad (47)$$

If the minimum of $r_{\varepsilon_i^2}(m)$ obtained in (47) is substituted for $r_{\varepsilon_i^2}(m)$ in (46) and some mathematical calculations are done, we obtain that $\mathrm{Kurt}(\varepsilon_i(n)) > 3$ if the following condition is satisfied

$$\sum_{k=1}^{Q}\beta_{i,k}+\sum_{k=1}^{Q}\beta_{i,k}^2 > 0. \quad (48)$$

The condition in (48) is always satisfied unless $\beta_{i,k}=0$ for all k which makes $\varepsilon_i(n)$ a Gaussian process. □

APPENDIX C
PROOF OF THEOREM 2

Proof: We want to show the process resulted by summation of two independent ARCH processes with same parameters will not have the same parameters with the two mixed processes. Therefore, if the summations of the ARCH processes with same parameters are observed, the sources are separable. Assume that $\varepsilon_1(n)$ and $\varepsilon_2(n)$ are ARCH processes as given in (2) and (3) with $\boldsymbol{\beta} = [\beta_0\cdots\beta_Q]^T$. Let $\varepsilon_3(n) = \alpha_1\varepsilon_1(n)+\alpha_2\varepsilon_2(n)$. The conditional variance of $\varepsilon_3(n)$ conditioned on past immediate information $\Psi(n-1) = \{\varepsilon_3(n-1), \varepsilon_3(n-2), \cdots, \varepsilon_3(0)\}$ is

$$\begin{aligned} &E\{\varepsilon_3^2(n)|\Psi(n-1)\} \\ &\quad = E\{\alpha_1^2\varepsilon_1^2(n)+\alpha_2^2\varepsilon_2^2(n)+2\alpha_1\alpha_2\varepsilon_1(n)\varepsilon_2(n)|\Psi(n-1)\} \\ &\quad = E\{\alpha_1^2\sigma_1^2(n)|\Psi(n-1)\}E\{\zeta_1^2(n)|\Psi(n-1)\} \\ &\qquad +E\{\alpha_2^2\sigma_2^2(n)|\Psi(n-1)\}E\{\zeta_2^2(n)|\Psi(n-1)\} \\ &\qquad +2\alpha_1\alpha_2E\{\sigma_1(n)\sigma_2(n)|\Psi(n-1)\} \\ &\qquad \times E\{\zeta_1(n)\zeta_2(n)|\Psi(n-1)\} \\ &\quad = E\{\alpha_1^2\sigma_1^2(n)+\alpha_2^2\sigma_2^2(n)|\Psi(n-1)\}, \end{aligned} \quad (49)$$

where $\sigma_i^2(n)$ can be written as

$$\sigma_i^2(n) = \beta_0+g(n+1)*\varepsilon_i^2(n-1), \quad i=1,2, \quad (50)$$

where filter $g(n)$ is defined as $g(n)=\sum_{k=1}^{Q}\beta_k\delta(n-k)$ and the notation $*$ is the convolution operator. Therefore, (49) can be rewritten as

$$\begin{aligned} &E\{\varepsilon_3^2(n)|\Psi(n-1)\} \\ &= (\alpha_1^2+\alpha_2^2)\beta_0 \\ &\quad +E\{g(n+1)*\left(\alpha_1^2\varepsilon_1^2(n-1)+\alpha_2^2\varepsilon_2^2(n-1)\right)|\Psi(n-1)\} \\ &= (\alpha_1^2+\alpha_2^2)\beta_0 \\ &\quad +E\{g(n+1)*\big((\alpha_1\varepsilon_1(n-1)+\alpha_2\varepsilon_2(n-1))^2 \\ &\quad -2\alpha_1\alpha_2\varepsilon_1(n-1)\varepsilon_2(n-1)\big)|\Psi(n-1)\} \\ &= (\alpha_1^2+\alpha_2^2)\beta_0+g(n+1)*E\{\varepsilon_3^2(n-1)|\Psi(n-1)\} \\ &\quad -2\alpha_1\alpha_2g(n+1)*E\{\big(\varepsilon_1(n-1)\varepsilon_2(n-1)\big)|\Psi(n-1)\} \\ &= (\alpha_1^2+\alpha_2^2)\beta_0+g(n+1)*\varepsilon_3^2(n-1) \\ &\quad -2\alpha_1\alpha_2g(n+1)*E\{\big(\varepsilon_1(n-1)\varepsilon_2(n-1)\big)|\Psi(n-1)\}. \end{aligned} \quad (51)$$

The second term in (51), $g(n+1)*\varepsilon_3^2(n-1)$, is the same as the term in the conditional variances, $\sigma_i^2(n), i = 1,2$. The term $-2\alpha_1\alpha_2\, g(n+1)*E\{\big(\varepsilon_1(n-1)\varepsilon_2(n-1)\big)|\Psi(n-1)\}$

makes the conditional variance parameters of $\varepsilon_3(n)$ different from the conditional variance parameters of $\varepsilon_1(n)$ and $\varepsilon_2(n)$.

When $\varepsilon_i(n),\ i=1,2$, are not conditionally heteroscedastic, where $\zeta_i(n)$ is a Gaussian process and $g(n)=0$, then, (51) can be written as

$$E\{\varepsilon_3^2(n)|\Psi(n-1)\} = (\alpha_1^2+\alpha_2^2)\beta_0. \tag{52}$$

Therefore $\varepsilon_i(n),\ i=1,2,3$, are Gaussian processes. □

APPENDIX D
Cramér Rao Lower Bound For AR-ARCH model
A. KNOWN PARAMETERS

In this Appendix, we obtain a lower bound based on CRLB for the parameter vector $\boldsymbol{\theta}$ including mean square errors of the off-diagonal elements of the contamination matrix $\boldsymbol{H}$, where $\boldsymbol{H} = \boldsymbol{WA}$. We assume that the parameters of the sources are known in deriving the lower bound. The covariance error of the parameter estimator $\hat{\boldsymbol{\theta}}$ is bounded by the inverse of FIM in (30).

The equivariance of the lower bound is used in this proof where the mixing matrix $\boldsymbol{A}$ equals to unit matrix $\boldsymbol{I}$. The lower bound is found at the point $\boldsymbol{A} = \boldsymbol{I}$ can be used to derive a lower bound for other invertible matrices due to equivariance property [37]. The diagonal elements of matrix $\boldsymbol{H}$ is assumed to be 1. $N(N-1)$ elements would be considered as parameters. Therefore, the likelihood $\boldsymbol{L}$ can be written as

$$\boldsymbol{L} = \sum_{n=1}^{N}\sum_{i=1}^{L} F_i\left(\frac{\boldsymbol{e}_n^T\boldsymbol{H}(\boldsymbol{x}(n)-\sum_{k=1}^{P}a_i(k)\boldsymbol{x}(n-k))}{\sigma_i(n)}\right) + N\log(|\boldsymbol{H}|), \tag{53}$$

where $\boldsymbol{e}_i$ is ith column of a $L\times L$ unit matrix. The derivative of the likelihood function obtained from (18) can be written as

$$\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{H}_{i,j}} \approx \sum_{n=1}^{N}\sum_{i=1}^{L}\frac{\partial}{\partial \boldsymbol{H}_{i,j}}F_i\left(\frac{\boldsymbol{e}_n^T\boldsymbol{H}(\boldsymbol{x}(n)-\sum_{k=1}^{P}a_{i,k}\boldsymbol{x}(n-k))}{\sigma_i(n)}\right) + N\frac{\partial}{\partial \boldsymbol{H}_{i,j}}\log(|\boldsymbol{H}|). \tag{54}$$

Equation (54) is simplified by substituting the matrix $\boldsymbol{H}=\boldsymbol{I}$ and considering $\frac{\partial \boldsymbol{H}}{\partial \boldsymbol{H}_{i,j}} = \boldsymbol{e}_n\boldsymbol{e}_m^T$ and $\frac{\partial|\boldsymbol{H}|}{\partial \boldsymbol{H}_{i,j}} = \boldsymbol{M}_{i,j}^{\boldsymbol{H}}$, where $\boldsymbol{M}_{i,j}^{\boldsymbol{H}}$ is minor determinant matrix of $\boldsymbol{H}$.

$$\begin{aligned}\left.\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{H}_{i,j}}\right|_{\boldsymbol{H}=\boldsymbol{I}} &= \sum_{n=1}^{N} F_i'\left(\frac{\boldsymbol{e}_i^T(\boldsymbol{s}(n)-\sum_{k=1}^{P}a_{i,k}\boldsymbol{s}(n-k))}{\sigma_i(n)}\right) \times \boldsymbol{e}_j^T\left(\frac{\boldsymbol{s}(n)-\sum_k a_{j,k}\boldsymbol{s}(n-k)}{\sigma_i(n)}\right) + N\delta_{i,j}\\ &= \sum_{n=1}^{N} F_i'\left(\boldsymbol{e}_i^T\boldsymbol{\zeta}(n)\right)\boldsymbol{e}_j^T\frac{\boldsymbol{\varepsilon}(n)}{\sigma_i(n)}\\ &= \sum_{n=1}^{N} F_i'(\zeta_i(n))\frac{\varepsilon_j(n)}{\sigma_i(n)},\end{aligned} \tag{55}$$

where $\delta_{i,j}$ is delta function which is 0 when $i\neq j$. The expectation of the second-order mixed derivatives of $\boldsymbol{L}$ with respect to off-diagonal elements of the matrix $\boldsymbol{H}$, $E\left\{\frac{\partial^2 \boldsymbol{L}}{\partial \boldsymbol{H}_{i,j}\partial \boldsymbol{H}_{u,v}}\right\}$, at the point $\boldsymbol{\theta}_0$ is non-zero when $i=v,\ j=u$ and $i=u,\ j=v$.

$$E\left\{\frac{\partial^2 \boldsymbol{L}}{\partial \boldsymbol{H}_{i,j}\partial \boldsymbol{H}_{u,v}}\right\} = \sum_{n_1=1}^{N}\sum_{n_2=1}^{N}E\left\{F_i'(\zeta_i(n_1))\frac{\varepsilon_j(n_1)}{\sigma_i(n_1)}F_u'(\zeta_u(n_2))\frac{\varepsilon_v(n_2)}{\sigma_u(n_2)}\right\}. \tag{56}$$

If $i=u$ and $j=v$, then

$$\begin{aligned}E\left\{\frac{\partial^2 \boldsymbol{L}}{\partial \boldsymbol{H}_{i,j}\partial \boldsymbol{H}_{i,j}}\right\} &= \sum_{n_1=1}^{N}\sum_{n_2=1}^{N}E\left\{F_i'(\zeta_i(n_1))\frac{\varepsilon_j(n_1)}{\sigma_i(n_1)}F_i'(\zeta_i(n_2))\frac{\varepsilon_j(n_2)}{\sigma_i(n_2)}\right\}\\ &= \sum_{n=1}^{N}E\left\{\left(F_i'(\zeta_i(n))\right)^2\frac{\varepsilon_j^2(n)}{\sigma_i^2(n)}\right\}\\ &= \sum_{n=1}^{N}E\left\{\left(F_i'(\zeta_i(n))\right)^2\right\}E\left\{\varepsilon_j^2(n)\right\}E\left\{\frac{1}{\sigma_i^2(n)}\right\},\end{aligned} \tag{57}$$

if $\lambda_i = E\left\{\left(F_i'(\zeta_i(n))\right)^2\right\}$, then,

$$E\left\{\frac{\partial^2 \boldsymbol{L}}{\partial \boldsymbol{H}_{i,j}\partial \boldsymbol{H}_{i,j}}\right\} = N\lambda_i\left\{\frac{1}{\sigma_i^2(n)}\right\}. \tag{58}$$

If $i=v$ and $j=u$, then

$$\begin{aligned}E\left\{\frac{\partial^2 \boldsymbol{L}}{\partial \boldsymbol{H}_{i,j}\partial \boldsymbol{H}_{j,i}}\right\} &= \sum_{n_1=1}^{N}\sum_{n_2=1}^{N}E\left\{F_i'(\zeta_i(n_1))\frac{\varepsilon_j(n_1)}{\sigma_i(n_1)}F_j'(\zeta_j(n_2))\frac{\varepsilon_i(n_2)}{\sigma_j(n_2)}\right\}\\ &= \sum_{n=1}^{N}E\left\{F_i'(\zeta_i(n))\zeta_i(n)\right\}E\left\{F_j'(\zeta_j(n))\zeta_j(n)\right\} = N.\end{aligned} \tag{59}$$

Using (58) and (59), $E\{\boldsymbol{H}_{i,j}^2\}$ is larger than

$$E\{\boldsymbol{H}_{i,j}^2\} \geq \frac{1}{N}\frac{\lambda_j\sigma_{\varepsilon_i}^2 E\left\{\frac{1}{\sigma_j^2(n)}\right\}}{\lambda_j\sigma_{\varepsilon_i}^2 E\left\{\frac{1}{\sigma_j^2(n)}\right\}\lambda_i\sigma_{\varepsilon_j}^2 E\left\{\frac{1}{\sigma_i^2(n)}\right\} - 1},$$

where $\sigma_{\varepsilon_i}^2$ can be substituted by (40).

Using the lower bound for $\boldsymbol{H}_{i,j}$, a lower bound for MSE can be derived. The relation between MSE and $E\left\{\boldsymbol{H}_{i,j}^2\right\}$ is given by

$$\begin{aligned}\text{MSE} &= E\{\|\hat{\boldsymbol{s}}(n)-\boldsymbol{s}(n)\|^2\}\\ &= E\left\{\text{Tr}\left((\hat{\boldsymbol{s}}(n)-\boldsymbol{s}(n))(\hat{\boldsymbol{s}}(n)-\boldsymbol{s}(n))^T\right)\right\}\\ &= E\{\text{Tr}((\boldsymbol{WA}-\boldsymbol{I})\boldsymbol{s}(n)\boldsymbol{s}^T(n)(\boldsymbol{WA}-\boldsymbol{I})^T)\}\end{aligned}$$

$$= E\{\mathrm{Tr}((\boldsymbol{H}-\boldsymbol{I})\boldsymbol{s}(n)\boldsymbol{s}^T(n)(\boldsymbol{H}-\boldsymbol{I})^T)\}$$
$$= E\{\mathrm{Tr}((\boldsymbol{H}-\boldsymbol{I})\boldsymbol{\Sigma}^2(\boldsymbol{H}-\boldsymbol{I})^T)\}$$
$$= \sum_{\substack{i,j=1\\ i\neq j}}^{L} \sigma_i^2 E\left\{\boldsymbol{H}_{i,j}^2\right\}. \tag{60}$$

B. FISHER INFORMATION MATRIX FOR MIXING MATRIX AND AR-ARCH MODEL PARAMETERS

We obtain Fisher's Information matrix for the case which the vector $\boldsymbol{\theta}$ includes the off diagonal elements of $\boldsymbol{H}$, the AR parameters $\boldsymbol{a}_i = [a_{i,1}, \cdots, a_{i,P}]$ and ARCH parameters $\boldsymbol{\beta}_i = [\beta_{i,0}, \cdots \beta_{i,Q}]^T$. The aim is to find the Fisher information matrix $\boldsymbol{J}$ at the point $\boldsymbol{\theta}_0$

$$\begin{aligned}\boldsymbol{H} &= \boldsymbol{I},\\ \boldsymbol{a}_i &= \boldsymbol{a}_i',\\ \boldsymbol{\beta}_i &= \boldsymbol{\beta}_i'.\end{aligned} \tag{61}$$

The derivatives of the likelihood function in (53) with respect to $a_{i,k}$ is given by

$$\frac{\partial \boldsymbol{L}}{\partial a_{i,k}} = \sum_{n=1}^{N} F_i'\left(\frac{\boldsymbol{w}_i^T\boldsymbol{x}(n) - \sum_{k=1}^{P} a_{i,k}\boldsymbol{w}_i^T\boldsymbol{x}(n-k)}{\sigma_i(n)}\right) \times -\frac{\boldsymbol{w}_i^T\boldsymbol{x}(n-k)}{\sigma_i(n)}. \tag{62}$$

The value of (62) at point $\boldsymbol{\theta}_0$ is

$$\left.\frac{\partial \boldsymbol{L}}{\partial a_{i,k}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \sum_{n=1}^{N} F_i'(\zeta_i(n))\frac{-s_i(n-k)}{\sigma_i(n)}. \tag{63}$$

The derivative of likelihood respect to $\beta_{i,0}$ is

$$\frac{\partial \boldsymbol{L}}{\partial \beta_{i,0}} = \sum_{n=1}^{N} F_i'\left(\frac{\boldsymbol{w}_i^T\boldsymbol{x}(n) - \sum_{k=1}^{P} a_{i,k}\boldsymbol{w}_i^T\boldsymbol{x}(n-k)}{\sigma_i(n)}\right) \times \frac{-\boldsymbol{w}_i^T\boldsymbol{x}(n-k)}{2\sigma_i^3(n)} - \frac{1}{2\sigma_i^2(n)}. \tag{64}$$

The value of (64) at point $\boldsymbol{\theta}_0$ is

$$\left.\frac{\partial \boldsymbol{L}}{\partial \beta_{i,0}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \sum_{n=1}^{N} F_i'(\zeta_i(n))\frac{-\varepsilon_i(n)}{2\sigma_i^3(n)} - \frac{1}{2\sigma_i^2(n)}. \tag{65}$$

The derivative of likelihood respect to β_i^q is

$$\frac{\partial \boldsymbol{L}}{\partial \beta_i^q} = \sum_{n=1}^{N} F_i'\left(\frac{\boldsymbol{w}_i^T\boldsymbol{x}(n) - \sum_{k=1}^{P} a_{i,k}\boldsymbol{w}_i^T\boldsymbol{x}(n-k)}{\sigma_i(n)}\right) \times -\frac{\boldsymbol{w}_i^T\boldsymbol{x}(n-k)}{\sigma_i(n)} \times \frac{(\boldsymbol{w}_i^T\boldsymbol{x}(n-q))^2}{2\sigma_i^2(n)} - \frac{(\boldsymbol{w}_i^T\boldsymbol{x}(n-q))^2}{2\sigma_i^2(n)}. \tag{66}$$

The value of (66) at point $\boldsymbol{\theta}_0$ is

$$\left.\frac{\partial \boldsymbol{L}}{\partial \beta_{i,q}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \sum_{n=1}^{N} F_i'(\zeta_i(n))\frac{-\varepsilon_i(n)}{\sigma_i^2(n)} \times \frac{\varepsilon_i^2(n-q)}{2\sigma_i(n)} - \frac{\varepsilon_i^2(n-q)}{2\sigma_i^2(n)}. \tag{67}$$

In order to find the matrix $\boldsymbol{J}$, the expectation of the terms $\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{H}_{i,j}\partial a_{u,k}}$, $\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{H}_{i,j}\partial \beta_{u,0}}$, $\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{H}_{i,j}\partial \beta_{u,q}}$, $\frac{\partial \boldsymbol{L}}{\partial a_{i,k}\partial \beta_{j,0}}$, $\frac{\partial \boldsymbol{L}}{\partial a_{i,k}\partial \beta_{j,q}}$, $\frac{\partial \boldsymbol{L}}{\partial \beta_{i,0}\partial \beta_{j,q}}$, $\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{H}_{i,j}\partial \boldsymbol{H}_{u,v}}$, $\frac{\partial \boldsymbol{L}}{\partial a_{i,k_1}\partial a_{j,k_2}}$, $\frac{\partial \boldsymbol{L}}{\partial \beta_{i,0}\partial \beta_{j,0}}$, $\frac{\partial \boldsymbol{L}}{\partial \beta_{i,q_1}\partial \beta_{j,q_2}}$ should be calculated.

The expectation of (55), (63), (64) and (67) at the point $\boldsymbol{\theta}_0$ is calculated

$$E\{\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{H}_{i,j}}\} = \sum_{n=1}^{N} E\{F_i'(\zeta_i(n))\}E\{\zeta_j(n)\} = 0, \tag{68}$$

$$E\{\frac{\partial \boldsymbol{L}}{\partial a_{i,k}}\} = \sum_{n=1}^{N} E\{F_i'(\zeta_i(n))\}E\{\frac{-\varepsilon_i(n-k)}{\sigma_i(n)}\} = 0, \tag{69}$$

$$E\{\frac{\partial \boldsymbol{L}}{\partial \beta_{i,0}}\} = \sum_{n=1}^{N} E\{F_i'(\zeta_i(n))\zeta_i(n)\}E\{\frac{-1}{2\sigma_i^2(n)}\} - E\{\frac{1}{2\sigma_i^2(n)}\} = 0, \tag{70}$$

$$E\{\frac{\partial \boldsymbol{L}}{\partial \beta_{i,q}}\} = \sum_{n=1}^{N} E\{F_i'(\zeta_i(n))\zeta_i(n)\}E\left\{-\frac{\varepsilon_i^2(n-q)}{2\sigma_i^2(n)}\right\} - E\left\{\frac{\varepsilon_i^2(n-q)}{2\sigma_i^2(n)}\right\} = 0. \tag{71}$$

Due to (68), (69), (70) and (71), the second-order mixed derivatives of $\boldsymbol{L}$ with respect to the parameters of different sources are zero. The second-order mixed derivatives of $\boldsymbol{L}$ with respect to the parameters of the sources and off-diagonal elements of the matrix $\boldsymbol{H}$ are derived at the point $\boldsymbol{\theta}_0$ as follow

$$\begin{aligned}E&\left\{\frac{\partial^2 \boldsymbol{L}}{\partial \boldsymbol{H}_{i,j}\partial \boldsymbol{H}_{u,v}}\right\}\\ &= E\left\{\sum_{n_1=1}^{N}\sum_{n_2=1}^{N} F_i'(\zeta_i(n_1))\zeta_j(n_1)F_u'(\zeta_u(n_2))\zeta_v(n_2)\right\}\\ &= \sum_{n=1}^{N} E\{F_i'(\zeta_i(n))\zeta_j(n)F_u'(\zeta_u(n))\zeta_v(n)\}\end{aligned} \tag{72}$$

The term in (72) is non-zero when $i = v,\ j = u$,

$$E\left\{\frac{\partial^2 \boldsymbol{L}}{\partial \boldsymbol{H}_{i,j}\partial \boldsymbol{H}_{u,v}}\right\} = \sum_{n=1}^{N} E\{(F_i'(\zeta_i(n)))^2\} = N\lambda_i E\left\{\frac{1}{\sigma_i^2(n)}\right\} \tag{73}$$

and when and $i = u,\ j = v$,

$$E\left\{\frac{\partial^2 \boldsymbol{L}}{\partial \boldsymbol{H}_{i,j}\partial \boldsymbol{H}_{u,v}}\right\} = \sum_{n=1}^{N} E\{F_i'(\zeta_i(n))\zeta_i(n)\} \times E\{F_j'(\zeta_j(n))\zeta_j(n)\} = N. \tag{74}$$

The term $E\left\{\frac{\partial^2 \boldsymbol{L}}{\partial \boldsymbol{H}_{i,j}\partial a_{u,k}}\right\}$ is given by

$$E\left\{\frac{\partial^2 \boldsymbol{L}}{\partial \boldsymbol{H}_{i,j}\partial a_{u,k}}\right\} = \sum_{n_1=1}^{N}\sum_{n_2=1}^{N} E\{F_i'(\zeta_i(n_1))\zeta_j(n_1) \times F_u'(\zeta_u(n_2))\frac{\varepsilon_u(n_2-k)}{\sigma_u(n_2)}\} = 0. \quad (75)$$

The term $E\left\{\frac{\partial^2 \boldsymbol{L}}{\partial \boldsymbol{H}_{i,j}\partial \beta_{u,0}}\right\}$ is given by

$$E\left\{\frac{\partial^2 \boldsymbol{L}}{\partial \boldsymbol{H}_{i,j}\partial \beta_{u,0}}\right\} = \sum_{n_1=1}^{N}\sum_{n_2=1}^{N} E\{F_i'(\zeta_i(n_1))\zeta_j(n_1) \times F_u'(\zeta_u(n))\frac{-\varepsilon_u(n)}{2\sigma_u^3(n)} - \frac{1}{2\sigma_u^2(n)}\} = 0. \quad (76)$$

The term $E\left\{\frac{\partial^2 \boldsymbol{L}}{\partial \boldsymbol{H}_{i,j}\partial \beta_{u,q}}\right\}$ is given by

$$E\left\{\frac{\partial^2 \boldsymbol{L}}{\partial \boldsymbol{H}_{i,j}\partial \beta_{u,q}}\right\} = \sum_{n_1=1}^{N}\sum_{n_2=1}^{N} E\{F_i'(\zeta_i(n_1))\zeta_j(n_1) \times (F_u'(\zeta_u(n_2))\zeta_u(n_2)\frac{\varepsilon_u^2(n_2-q)}{\sigma_u^2(n_2)} - \frac{\varepsilon_u^2(n_2-q)}{2\sigma_u^2(n_2)})\} = 0. \quad (77)$$

The term $E\left\{\frac{\partial^2 \boldsymbol{L}}{\partial a_{i,k_1}\partial a_{i,k_2}}\right\}$ is given by

$$E\left\{\frac{\partial^2 \boldsymbol{L}}{\partial a_{i,k_1}\partial a_{i,k_2}}\right\} = \sum_{n_1=1}^{L}\sum_{n_2=1}^{L} E\left\{F_i'(\zeta_i(n_1))\frac{\varepsilon_i(n_1-k_1)}{\sigma_i(n_1)} F_i'(\zeta_i(n_2))\frac{\varepsilon_i(n_2-k_2)}{\sigma_i(n_2)}\right\} = \sum_{n=1}^{L}\lambda_i E\left\{\frac{\varepsilon_i(n-k_1)\varepsilon_i(n-k_2)}{\sigma_i^2(n)}\right\}. \quad (78)$$

The term $E\{\frac{\varepsilon_i(n-k_1)\varepsilon_i(n-k_2)}{\sigma_i^2(n)}\}$ can be estimated by $\frac{1}{N}\sum_{n=1}^{N}\frac{\varepsilon_i(n-k_1)\varepsilon_i(n-k_2)}{\sigma_i^2(n)}$.

$$E\left\{\frac{\partial^2 \boldsymbol{L}}{\partial a_{i,k}\partial \beta_i^0}\right\} = \sum_{n_1=1}^{N}\sum_{n_2=1}^{N} E\left\{F_i'(\zeta_i(n_1))\frac{\varepsilon_i(n_1-k)}{\sigma_i(n_1)} \left(F_i'(\zeta_i(n_2))\zeta_i(n_2) \times \frac{1}{\sigma_i^2(n_2)} - \frac{1}{2\sigma_i^2(n_2)}\right)\right\} = \sum_{n=1}^{N} E\left\{F_i'(\zeta_i(n))\frac{\varepsilon_i(n-k)}{\sigma_i(n)} \left(F_i'(\zeta_i(n))\zeta_i(n) \times \frac{1}{\sigma_i^2(n)} - \frac{1}{2\sigma_i^2(n)}\right)\right\} = 0. \quad (79)$$

$$E\left\{\frac{\partial^2 \boldsymbol{L}}{\partial a_{i,k}\partial \beta_{i,q}}\right\} = \sum_{n_1=1}^{N}\sum_{n_2=1}^{N} E\left\{F_i'(\zeta_i(n_1))\frac{\varepsilon_i(n_1-k)}{\sigma_i(n_1)} \left(F_i'(\zeta_i(n_2))\zeta_i(n_2) \times \frac{\varepsilon_i^2(n_2-q)}{\sigma_i^2(n_2)} - \frac{\varepsilon_i^2(n_2-q)}{2\sigma_i^2(n_2)}\right)\right\} = \sum_{n=1}^{N} E\left\{F_i'(\zeta_i(n))\frac{\varepsilon_i(n-k)}{\sigma_i(n)} \left(F_i'(\zeta_i(n))\zeta_i(n) \times \frac{\varepsilon_i^2(n-q)}{\sigma_i^2(n)} - \frac{\varepsilon_i^2(n-q)}{2\sigma_i^2(n)}\right)\right\} = 0. \quad (80)$$

The term $E\left\{\frac{\partial^2 \boldsymbol{L}}{\partial \beta_{i,0}\partial \beta_{i,0}}\right\}$ is given by

$$E\left\{\frac{\partial^2 \boldsymbol{L}}{\partial \beta_{i,0}\partial \beta_{i,0}}\right\} = \sum_{n_1=1}^{N}\sum_{n_2=1}^{N} E\left\{\left(F_i'(\zeta_i(n_1))\zeta_i(n_1) \times \frac{1}{2\sigma_i^2(n_1)} - \frac{1}{2\sigma_i^2(n_1)}\right)\left(F_i'(\zeta_i(n_2))\zeta_i(n_2) \times \frac{1}{2\sigma_i^2(n_2)} - \frac{1}{2\sigma_i^2(n_2)}\right)\right\} \approx \sum_{n=1}^{N} E\left\{\left(F_i'(\zeta_i(n))\zeta_i(n) \times \frac{1}{2\sigma_i^2(n)} - \frac{1}{2\sigma_i^2(n)}\right)^2\right\} = \sum_{n=1}^{N} E\left\{\frac{1}{4\sigma_i^4(n)}\right\} E\left\{\left(F_i'(\zeta_i(n))\zeta_i(n)\right)^2\right\}. \quad (81)$$

The term $E\{\frac{1}{4\sigma_i^4(n)}\}$ can be obtained by $\frac{1}{N}\sum_{n=1}^{N}\frac{1}{4\sigma_i^4(n)}$.

$$\frac{\partial^2 \boldsymbol{L}}{\partial \beta_{i,0}\partial \beta_{i,q}} = \sum_{n_1=1}^{N}\sum_{n_2=1}^{N} E\left\{\left(F_i'(\zeta_i(n_1))\zeta_i(n_1) \times \frac{1}{2\sigma_i^2(n_1)} - \frac{1}{2\sigma_i^2(n_1)}\right)\left(F_i'(\zeta_i(n_2))\zeta_i(n_2) \times \frac{\varepsilon_i^2(n_2-q)}{2\sigma_i^2(n_2)} - \frac{\varepsilon_i^2(n_2-q)}{2\sigma_i^2(n_2)}\right)\right\} \approx \sum_{n=1}^{N} E\left\{\frac{\varepsilon_i^2(n-q)}{4\sigma_i^4(n)}\left(F_i'(\zeta_i(n))\zeta_i(n) - 1\right)^2\right\}. \quad (82)$$

The term $E\{\frac{\varepsilon_i^2(n-q)}{4\sigma_i^4(n)}\}$ can be estimated consistently by $\frac{1}{N}\sum_{n=1}^{N}\frac{\varepsilon_i^2(n-q)}{4\sigma_i^4(n)}$.

$$\frac{\partial^2 \boldsymbol{L}}{\partial \beta_{i,q_1}\partial \beta_{i,q_2}} = \sum_{n_1=1}^{N}\sum_{n_2=1}^{N} E\left\{\left(F_i'(\zeta_i(n_1))\zeta_i(n_1)\frac{\varepsilon_i^2(n-q_1)}{2\sigma_i^2(n_1)} - \frac{\varepsilon_i^2(n-q_1)}{2\sigma_i^2(n_1)}\right)\left(F_i'(\zeta_i(n_2))\zeta_i(n_2)\frac{\varepsilon_i^2(n_2-q_2)}{2\sigma_i^2(n_2)} - \frac{\varepsilon_i^2(n_2-q_2)}{2\sigma_i^2(n_2)}\right)\right\} \approx \sum_{n=1}^{N} E\left\{\frac{\varepsilon_i^2(n-q_2)\varepsilon_i^2(n-q_1)}{4\sigma_i^4(n)}\left(F_i'(\zeta_i(n))\zeta_i(n) - 1\right)^2\right\}. \quad (83)$$

The term $E\left\{\frac{\varepsilon_i^2(n-q_2)\varepsilon_i^2(n-q_1)}{4\sigma_i^4(n)}\right\}$ can be estimated by $\frac{1}{N}\sum_{n=1}^{N} \frac{\varepsilon_i^2(n-q_2)\varepsilon_i^2(n-q_1)}{4\sigma_i^4(n)}$.

APPENDIX E
PROOF OF PROPOSITION 2

Proof: The bound $\kappa_{i,j}$ in (32) is a monotonically decreasing rational function in $E\left\{\frac{1}{\sigma_j^2(n)}\right\}$ if

$$\lambda_j\sigma_{\varepsilon_j}^2 E\left\{\frac{1}{\sigma_j^2(n)}\right\} \geq \frac{1}{\lambda_i\sigma_{\varepsilon_i}^2 E\left\{\frac{1}{\sigma_i^2(n)}\right\}}. \tag{84}$$

Also, if $E\left\{\frac{1}{\sigma_i^2(n)}\right\}$ increases, $\kappa_{i,j}$ will decrease. Therefore, lower values for $E\left\{\frac{1}{\sigma_j^2(n)}\right\}$ and $E\left\{\frac{1}{\sigma_i^2(n)}\right\}$ result a higher value for $\kappa_{i,j}$. We can use Jensen's inequality to a find lower bound for $E\left\{\frac{1}{\sigma_j^2(n)}\right\}$ and $E\left\{\frac{1}{\sigma_i^2(n)}\right\}$. Since the function $T(x) = 1/x$ for $x > 0$ is a convex function, one can simplify $E\left\{\frac{1}{\sigma_i^2(n)}\right\}$ based on Jensen's inequality

$$E\left\{\frac{1}{\sigma_i^2(n)}\right\} \geq \frac{1}{E\left\{\sigma_i^2(n)\right\}}. \tag{85}$$

Since $\zeta_i(n)$ and $\sigma_i^2(n)$ are independent, one can write

$$\begin{aligned}\sigma_{\varepsilon_i}^2 = E\left\{\varepsilon_i^2(n)\right\} &= E\left\{\sigma_i^2(n)\right\} \\ &= \sigma_{\varepsilon_i}^2. \end{aligned} \tag{86}$$

By substituting $E\frac{1}{\{\sigma_i^2(n)\}} = \frac{1}{\sigma_{\varepsilon_i}^2}$ and $E\frac{1}{\{\sigma_j^2(n)\}} = \frac{1}{\sigma_{\varepsilon_j}^2}$ in (32), the right hand side of the inequality (34) is obtained. □

REFERENCES

[1] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. San Francisco, CA, USA: Academic, 2010.

[2] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, May 1999.

[3] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," in *IEE Proc. F-Radar and Signal Process.*, vol. 140, no. 6, 1993, pp. 362–370, Dec. 1993.

[4] X.-L. Li and T. Adalı, "Independent component analysis by entropy bound minimization," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5151–5164, Oct. 2010.

[5] K. E. Hild, II, H. T. Attias, and S. S. Nagarajan, "An expectation-maximization method for spatio-temporal blind source separation using an AR-MOG source model," *IEEE Trans. Neural Netw.*, vol. 19, no. 3, pp. 508–519, Mar. 2008.

[6] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[7] A. Hyvarinen, "A unifying model for blind separation of independent sources," *Signal Process.*, vol. 85, no. 7, pp. 1419–1427, 2005.

[8] H. Attias, "Independent factor analysis with temporally structured sources," in *Proc. Adv. Neural Inf. Process. Syst.*, 1999, pp. 386–392.

[9] S. Hosseini, C. Jutten, and D. T. Pham, "Markovian source separation," *IEEE Trans. Signal Process.*, vol. 51, no. 12, pp. 3009–3019, Dec. 2003.

[10] M. M. Ichir and A. Mohammad-Djafari, "Hidden Markov models for wavelet-based blind source separation," *IEEE Trans. Image Process.*, vol. 15, no. 7, pp. 1887–1899, Jul. 2006.

[11] R. Guidara, S. Hosseini, and Y. Deville, "Blind separation of nonstationary Markovian sources using an equivariant Newton–Raphson algorithm," *IEEE Signal Process. Lett.*, vol. 16, no. 5, pp. 426–429, May 2009.

[12] G.-S. Fu, R. Phlypo, M. Anderson, X.-L. Li, and T. Adalı, "Blind source separation by entropy rate minimization," *IEEE Trans. Signal Process.*, vol. 62, no. 16, pp. 4245–4255, Aug. 2014.

[13] X.-L. Li and T. Adalı, "Blind spatiotemporal separation of second and/or higher-order correlated sources by entropy rate minimization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2010, pp. 1934–1937.

[14] G. S. Fu, R. Phlypo, M. Anderson, and T. Adalı, "Complex independent component analysis using three types of diversity: Non-Gaussianity, non-whiteness, and noncircularity," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 794–805, Feb. 2015.

[15] D.-T. Pham and J. F. Cardoso, "Blind separation of instantaneous mixtures of nonstationary sources," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 1837–1848, Sep. 2001.

[16] K. Zhang and A. Hyvarinen. (2012). "Source separation and higher-order causal analysis of MEG and EEG." [Online]. Available: https://arxiv.org/abs/1203.3533

[17] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation," *Econometrica*, vol. 50, no. 4, pp. 987–1007, 1982.

[18] S. Mousazadeh and I. Cohen, "AR-GARCH in presence of noise: Parameter estimation and its application to voice activity detection," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 4, pp. 916–926, May 2011.

[19] I. Cohen, "Speech spectral modeling and enhancement based on autoregressive conditional heteroscedasticity models," *Signal Process.*, vol. 86, no. 4, pp. 698–709, 2006.

[20] I. Cohen, "Modeling speech signals in the timeâĂŞfrequency domain using GARCH," *Signal Process.*, vol. 84, no. 12, pp. 2453–2459, 2004.

[21] H. K. Solvang, K. Ishizuka, and M. Fujimoto, "Voice activity detection based on adjustable linear prediction and GARCH models," *Speech Commun.*, vol. 50, no. 6, pp. 476–486, Jun. 2008.

[22] J. P. Pascual, N. von Ellenrieder, M. Hurtado, and C. H. Muravchik, "Adaptive radar detection algorithm based on an autoregressive GARCH-2D clutter model," *IEEE Trans. Signal Process.*, vol. 62, no. 15, pp. 3822–3832, Aug. 2014.

[23] J. P. Pascual, N. von Ellenrieder, M. Hurtado, and C. H. Muravchik, "Radar detection algorithm for GARCH clutter model," *Digit. Signal Process.*, vol. 23, no. 4, pp. 1255–1264, 2013.

[24] Y. Zhang, J. Shi, Z. Deng, and P. Pan, "Detection in sea clutter based on nonlinear ARCH model," *Digit. Signal Process.*, vol. 50, pp. 162–170, 2016.

[25] M. Amirmazlaghani and H. Amindavar, "A novel sparse method for despeckling SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 12, pp. 5024–5032, Dec. 2012.

[26] A. Noiboar and I. Cohen, "Anomaly detection based on wavelet domain GARCH random field modeling," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1361–1373, May 2007.

[27] S. Mousazadeh and I. Cohen, "Anomaly detection in sonar images based on wavelet domain noncausal AR-ARCH random field modeling," in *Proc. IEEE 26th Conv. Elect. Electron. Eng. Israel (IEEEI)*, Nov. 2010, pp. 306–309.

[28] S. Mousazadeh and I. Cohen, "Two dimensional noncausal AR-ARCH model: Stationary conditions, parameter estimation and its application to anomaly detection," *Signal Process.*, vol. 98, pp. 322–336, May 2014.

[29] S. H. Fouladi, M. Hajiramezanali, H. Amindavar, J. A. Ritcey, and P. Arabshahi, "Denoising based on multivariate stochastic volatility modeling of multiwavelet coefficients," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5578–5589, Nov. 2013.

[30] T. G. Andersen, R. A. Davis, J.-P. Kreiss, and T. V. Mikosch, *Handbook of Financial Time Series*. Springer, 2009.

[31] A. Bose and K. Mukherjee, "Estimating the ARCH parameters by solving linear equations," *J. Time Ser. Anal.*, vol. 24, no. 2, pp. 127–136, 2003.

[32] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. Hoboken, NJ, USA: Wiley, 2009.

[33] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering, and Array Processing*, vol. 46. Norwood, MA, USA: Artech House, 2005.

[34] J.-F. Cardoso *et al.*, "On the performance of orthogonal source separation algorithms," in *Proc. EUSIPCO*, vol. 94. Edinburgh, U.K., 1994, pp. 776–779.
[35] M. K. Varanasi and B. Aazhang, "Parametric generalized Gaussian density estimation," *J. Acoust. Soc. Amer.*, vol. 86, no. 4, pp. 1404–1415, Oct. 1989.
[36] J. Eriksson and V. Koivunen, "Identifiability, separability, and uniqueness of linear ICA models," *IEEE Signal Process. Lett.*, vol. 11, no. 7, pp. 601–604, Jul. 2004.
[37] Y. Lomnitz and A. Yeredor, "A blind-ML scheme for blind source separation," in *Proc. IEEE Workshop Statist. Signal Process.*, Oct. 2003, pp. 581–584.
[38] Z. Koldovský, P. Tichavsky, and E. Oja, "Cramer–Rao lower bound for linear independent component analysis," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 3. Mar. 2005, pp. 581–584.
[39] A. Yeredor, "Blind separation of Gaussian sources with general covariance structures: Bounds and optimal estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5057–5068, Oct. 2010.

SEYYED HAMED FOULADI received the B.S. degree in electrical engineering from Shahed University, Tehran, Iran, in 2009, and the M.S. degree in electrical engineering from the Amirkabir University of Technology, Tehran, Iran, in 2012. He is currently pursuing the Ph.D. degree in signal processing with the Norwegian University of Science and Technology. His research interests are sparse recovery, source separation, machine learning, and signal and image processing.

ILANGKO BALASINGHAM (SM'12) received the M.Sc. and Ph.D. degrees in signal processing from the Department of Electronics and Telecommunications, Norwegian University of Science and Technology (NTNU), Trondheim, Norway, in 1993 and 1998, respectively. He performed the master's degree thesis with the Department of Electrical and Computer Engineering, University of California at Santa Barbara, Santa Barbara, USA. From 1998 to 2002, he was a Research Engineer developing image and video streaming solutions for mobile handheld devices with the Fast Search & Transfer ASA, Oslo, Norway, which is now a part of Microsoft Inc. Since 2002 he has been with the Intervention Center, Oslo University Hospital, Oslo, Norway, as a Senior Research Scientist, where he heads the Wireless Sensor Network Research Group. He was appointed as a Professor in signal processing in medical applications with NTNU in 2006. From 2016 to 2017 he was a Professor by courtesy with the Frontier Institute, Nagoya Institute of Technology, Japan. He has authored or co-authored over 200 journal and conference papers, seven book chapters, 42 abstracts, five patents, and 16 articles in popular press. His research interests include super robust short range communications for both in-body and on-body sensors, body area sensor network, microwave short range sensing of vital signs, short range localization and tracking mobile sensors, and nanoscale communication networks. He has given 16 invited/keynotes at the international conferences. In addition he is active in organizing conferences (Steering Committee Member of ACM NANOCOM from 2018 to 2021; a General Chair: the 2019 IEEE International Symposium of Medical ICT and the 2012 Body Area Networks Conference; and a TPC Chair of the 2015 ACM NANOCOM), and Editorial Board (has been an Area Editor of Elsevier *Nano Communication Networks* since 2013).

KIMMO KANSANEN received the M.Sc. (EE) and Dr.Tech. degrees from the University of Oulu, Finland, in 1998 and 2005, respectively. He was a Research Scientist and a Project Manager with the Centre for Wireless Communications, University of Oulu. Since 2006 he has been with the Norwegian University of Science and Technology, Trondheim, Norway, a Full Professor since 2016. His research interests are within wireless communications and signal processing. He is an Associate Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.

TOR AUDUN RAMSTAD was born in Norway. He received the Siv.Ing. and Dr.Ing. degrees from the Norwegian Institute of Technology, in 1968 and 1971, respectively. From 1982 to 1983 he was a Visiting Associate Professor with the University of California at Santa Barbara, Santa Barbara (UCSB). He has held various positions with the Department of Electronic Systems, Norwegian University of Science and Technology, where he has been a Full Professor of telecommunications since 1983 and has been a Professor Emeritus since 2014. He was with the Georzia Institute of Technology from 1989 to 1990 as a Visiting Adjunct Professor, and with UCSB as a Visiting Professor from 1997 to 1998 and from 2008 to 2009. From 2003 to 2004, he was a Visiting Professor with EURECOM, France. He has published over 270 journals and full conference papers and has supervised over 35 PhDs and 320 Master students. His research interests include multi-rate signal processing, speech, image, and video communications, with emphasis on joint source-channel coding. His current research focus is on sensor network applications.