Online Implementation of Independent Vector Extraction based on Auxiliary function

Jakub Janský, Jaroslav Čmejla, Tomáš Kounovský

Separation of the sources from mixture is interesting field of audio processing. This paper is focusing only on extraction of the one source from mixture. We propose a novel approach how to achieve this task in online processing.

Key words: Online Blind source separation, Independent Vector Analysis, Auxiliary function

Mixing Models

The Short-time Fourier Transform (STFT) model of the observed signals can be described as a determined instantaneous mixture of M sources

$$\mathbf{X}_{k,\ell} = \mathbf{A}_k \mathbf{S}_{k,\ell}, \tag{1}$$

where $\mathbf{S}_{k,\ell}$ and $\mathbf{X}_{k,\ell}$ denote $(M \times 1)$ vectors of the original and mixed signals respectively. The $k = 1, \ldots, K$ denote frequency bin and $\ell = 1, \ldots, L$ denote frame index. $\mathbf{A}_k$ denotes $(M \times M)$ non-singular mixing matrix. To obtain separated sources from a mixture, we can employ Independent Vector Analysis (IVA) [1]. IVA assume mutual independence of original signals $\mathbf{S}_{k,\ell}$ on each frequency bin k, however assume mutual dependence between frequency within each signal.

To obtain separated sources from the mixture we are looking for demixing matrix $\mathbf{W}_k$, which fulfills

$$\mathbf{W}_k \mathbf{X}_{k,\ell} = \mathbf{W}_k \mathbf{A}_k \mathbf{S}_{k,\ell} \approx \mathbf{S}_k(\omega, \ell). \tag{2}$$

In determined BSS scenarios, the separation of all sources in mixture is provided. This can lead to obtaining unnecessary source signals. For example, in scenario with one speaker in noisy environment, we are interested only in speech signal and separation of noise signals is redundant. The proposition to separate one particular source, *Signal of Interest* (SOI), is referred as *Blind Source Extraction* (BSE). The model (1) can be rewritten in the terms of BSE as

$$\mathbf{X}_{k,\ell} = \mathbf{A}_k \mathbf{S}_{k,\ell} = \mathbf{a}_k \mathbf{s}_{k,\ell} + \mathbf{Y}_{k,\ell}, \tag{3}$$

where $\mathbf{a}_k$ is mixing vector for desirable source, $\mathbf{S}_{k,\ell}$ denotes SOI and $\mathbf{Y}_{k,\ell}$ is mixture of the rest of the sources. In the IVA we assume mutual independence between all sources in mixture. This assumption can be extended to the BSE, however only the independence between SOI and rest of the mixture $\mathbf{Y}_{k,\ell}$ is required. This approach is referred as Independent Component Extraction (IVE)[2].

To obtain SOI, the extracting vector $\mathbf{w}_k$ needs to be found. Let demixing matrix $\mathbf{W}_k$ be partitioned as

$$\mathbf{W}_k = \begin{pmatrix} \mathbf{w}_k^H \\ \mathbf{B}_k \end{pmatrix} = \begin{pmatrix} \overline{\beta_k} & \mathbf{h}_k^H \\ \mathbf{g}_k & -\gamma_k \mathbf{I}_{M-1} \end{pmatrix}, \tag{4}$$

where $\mathbf{w}_k$ is $(M \times 1)$ separating vector associated with SOI, $\cdot^H$ denotes hermitian transpose, $\mathbf{B}_k$ is $(M \times M-1)$ matrix and $\mathbf{I}_{M-1}$ $(M-1 \times M-1)$ identity matrix.

Because of relation $\mathbf{A}_k = \mathbf{W}_k^{-1}$, the mixing matrix $\mathbf{A}_k$ can be also partitioned as $\mathbf{A}_k = \begin{pmatrix} \mathbf{a}_k & \mathbf{D}_k \end{pmatrix}$, where $\mathbf{a}_k$ is mixing vector associated with and $\mathbf{D}_k$ is $(M \times M-1)$ mixing matrix associated with $\mathbf{Y}_{k,\ell}$. The mixing vector $\mathbf{a}_k$ and extraction vector $\mathbf{w}_k$ are linked through $\mathbf{w}_k^H \mathbf{a}_k = 1$. To guarantee $\mathbf{B}_k \mathbf{X}_{k,\ell}$ do not contain any contribution of , $\mathbf{B}_k$ needs to be orthogonal to $\mathbf{a}_k$, i.e., $\mathbf{B}_k \mathbf{a}_k = \mathbf{0}$.

Independent Vector Extraction

The main assumption in IVE is that $\mathbf{s}_k$ and $\mathbf{Z}_{k,\ell}$ are mutually independent. The goal is to find $\mathbf{w}_k$ and $\mathbf{a}_k$ such that $\mathbf{w}_k^H \mathbf{X}_{k,\ell}$ and $\mathbf{B}_k \mathbf{X}_{k,\ell}$ will be independent as possible. The probability density function (pdf) of the mixed signals for one block is

$$p_{x,b}(\mathbf{X}_{k,\ell}) = p_{s,b}(\mathbf{w}_k^H \mathbf{X}_{k,\ell}) \cdot p_{z,b}(\mathbf{B}_k \mathbf{X}_{k,\ell}) \cdot |\det \mathbf{W}_k|^2. \tag{5}$$

To obtain $\mathbf{w}_k$ and $\mathbf{a}_k$, we can apply log-likelihood principle on (5). Log-likelihood function for our problem reads

$$\mathcal{L}(\mathbf{w}_k, \mathbf{a}_k) = \log p_{s,b}(\mathbf{w}_k^H \mathbf{X}_{k,\ell}) \\ \log p_{z,b}(\mathbf{B}_k \mathbf{X}_{k,\ell}) + \log |\det \mathbf{W}_k|^2. \tag{6}$$

Because the separation of the $\mathbf{Z}_{k,\ell}$ is not expected, we can assume $p_{z,b}(\cdot)$ to be circular Gaussian with covariance matrix $\hat{\mathbf{C}}_\mathbf{z} = E[\mathbf{Z}_{k,\ell}^H \mathbf{Z}_{k,\ell}]$. The pdf $p_{s,b}(\cdot)$ is usually unknown, therefore is replaced by model pdf $f(\cdot)$. The expression $|\det \mathbf{W}_k|^2$ can be replaced by $|\gamma|^{2(d-2)}$. The log-likelihood function (6) is optimized in variables $\mathbf{a}_k$ and $\mathbf{w}_k$. However, $\mathbf{a}_k$ and $\mathbf{w}_k$ are linked through distortion-less response $\mathbf{w}_k^h \mathbf{a}_k = 1$. Unfortunately, this connection is too weak and estimated $\mathbf{a}_k$ and $\mathbf{w}_k$ can be associated with different sources. To overcome this problem, orthogonal constraint (OC) is introduced. The relation of $\mathbf{a}_k$ and $\mathbf{w}_k$ through OC reads

$$\mathbf{a}_k = \frac{\mathbf{C}_\mathbf{x} \mathbf{w}_k}{\mathbf{w}_k^H \mathbf{C}_\mathbf{x} \mathbf{w}_k} \tag{7}$$

where $\mathbf{C}_\mathbf{x} = E_\ell \left[\mathbf{X}_{k,\ell}^H \mathbf{X}_{k,\ell} \right]$. Utilizing OC, the contrast function can be derived form the log-likelihood function (6)

$$\mathbf{J}(\mathbf{w}_k) = \mathsf{E}\left[f(\mathbf{w}_k^H \mathbf{X}_{k,\ell}) \right] \\ - E\left[\mathbf{X}_{k,\ell}^H \mathbf{B}_k^H \hat{\mathbf{C}}_\mathbf{z}^{-1} \mathbf{B}_k \mathbf{X}_{k,\ell} \right] + (d-2) \log |\gamma_k|^2, \tag{8}$$

In [2] is proposed to minimize the function (8) by steepest descend method.

Online AuxIVE

The selection of proper step size in steepest descent algorithm can be a difficult task. In this paper we propose to avoid step size completely by employing auxiliary function technique. The object function is not minimized directly, but alternatively in original object function and in auxiliary function. The proper selection of the auxiliary function can provide fast and stable update rule. The general idea is to find function which is easy to minimize, preferably to have first derivative closed solution. In [3] is proposed to use following theorem to find suitable auxiliary function

Theorem 1 *For $f(z) = f_R(\|z\|_2) \in S_f$ holds*

$$f(z) \leq \frac{f'_R(r_0)}{2r_0}\|z\|_2^2 + \left(f_R(r_0) - \frac{r_0 f'_R(r_0)}{2}\right) \tag{9}$$

where S_f is defined as

$$S_f = \{f(z)|f(z) = f_R(\|z\|_2)\} \tag{10}$$

and f_R is continuous and differentiable function of a real variable r satisfying that $f'_R(r)/r$ is continuous everywhere and it is monotonically decreasing in $r \geq 0$. The equality sign in (9) is satisfied if and only if $r_0 = \|z\|_2$.

To derive suitable auxiliary function from (8) utilizing theorem (9) only the first part of the function **J** is needed to be taken to the consideration. This leads to

$$\begin{aligned} E\left[f(\mathbf{w}_k^H \mathbf{X}_{k,\ell})\right] &\leq E\left[\frac{f'(r)}{2r}\mathbf{w}_k^H \mathbf{X}_{k,\ell}\mathbf{X}_{k,\ell}^H \mathbf{w}_k\right] + G = \\ &= \frac{1}{2}\mathbf{w}_k^H E\left[\frac{f'(r)}{r}\mathbf{X}_{k,\ell}\mathbf{X}_{k,\ell}^H\right]\mathbf{w}_k + G = \frac{1}{2}\mathbf{w}_k^H \mathbf{V}_k \mathbf{w}_k + G, \end{aligned} \tag{11}$$

where $\mathbf{V}_k = E\left[\frac{f'(r)}{r}\mathbf{X}_{k,\ell}\mathbf{X}_{k,\ell}^H\right]$ is auxiliary variable and G is constant term independent of $\mathbf{w}_k$. The equation (8) is then rewritten to the form

$$\begin{aligned} \mathbf{Q}(\mathbf{w}_k, \mathbf{V}_k) &= \frac{1}{2}\mathbf{w}_k^H \mathbf{V}_k \mathbf{w}_k \\ &- E\left[\mathbf{X}_{k,\ell}^H \mathbf{B}_k^H \mathbf{C}_{z,b}^{-1} \mathbf{B}_k \mathbf{X}_{k,\ell}\right] + (d-2)\log|\gamma_b|^2. \end{aligned} \tag{12}$$

The function $\mathbf{Q}(\mathbf{w}_k, \mathbf{V}_k)$ have to be minimized in $\mathbf{V}_k$ and $\mathbf{w}_k$ alternatively. The minimization of (12) in $\mathbf{V}_k$ is easily obtained. In the online scenario, the separation is done only over one or a few batch of frames. Thus function (12) have to be computed only over b number of the frames. To avoid extreme change in computed demixing vector $\mathbf{w}_k$, the past computation is included. The update rule of $\mathbf{Q}(\mathbf{w}_k, \mathbf{V}_k)$ for $\mathbf{w}_k$ is obtained by solving derivation of (12) respect to $\mathbf{w}_k^H$ equals to 0. This leads to the following update rule

$$\hat{\mathbf{C}}_{\mathbf{x},b} = \alpha \hat{\mathbf{C}}_{\mathbf{x},b-1} + (1-\alpha)E_b\left[\mathbf{X}_{k,b}^H \mathbf{X}_{k,b}\right], \tag{13}$$

$$r_b = \|\mathbf{w}_k^H \mathbf{X}_{k,b}\|_2, \tag{14}$$

$$\mathbf{V}_{k,b} = \alpha \mathbf{V}_{k,b-1} + (1-\alpha)E_b\left[\frac{f(r_b)'}{r_b}\mathbf{X}_{k,b}^H \mathbf{X}_{k,b}\right], \tag{15}$$

$$\mathbf{w}_k^H = \mathbf{a}_k^H \mathbf{V}_{k,b}^{-1}, \tag{16}$$

where α is forgetting factor controlling change of the computed demixing vector $\mathbf{w}_k$.

Experiments

The experiment was provided over simulated mixture recorded on 5 microphones. One signal was 10s recording of female speaker situated 1 meter in front of microphone field. The second signal was white noise signal on the 70 degree position from microphone field in 2 meter distance. The setting for the method were $\alpha = 0.96$ and number of frames 1,20,50 and 80 in block.

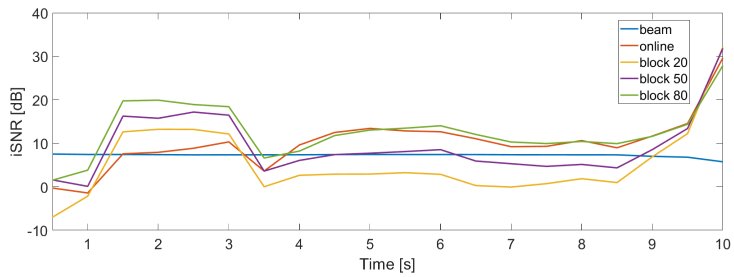

Figure 1: my caption of the figure

In the figure 1 the improvement of Signal-to-Noise ration are shown for each setting. The best method is with 80 block size, however the computation time is 10.7 seconds. In other hand online method need only 1.7 seconds.

Conclusion

The online version of the Independent Vector Analysis based on Auxiliary function is presented in this paper. In the experimental part we compare online proposed method with batch online method with variable frame block size. The online method shown promising result. However more research is needed in this area.

Acknowledgment

This work is supported by SGS project.

References

[1] A. Hiroe, "Solution of permutation problem in frequency domain ICA, using multivariate probability density functions," in *Independent Component Analysis and Blind Signal Separation*, Berlin, Heidelberg, 2006, pp. 601–608, Springer Berlin Heidelberg.

[2] P. Tichavský Z. Koldovský, "Gradient algorithms for complex non-gaussian independent component/vector extraction, question of convergence," in *IEEE Transactions on Signal Processing*, 2019, vol. 67, pp. 1050–1064.

[3] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2011, pp. 189–192.