Time-frequency-masking-based determined BSS with application to Sparse IVA

Kohei Yatabe (Waseda University) and Daichi Kitamura (NIT, Kagawa College)

Introduction

- Many independence-based blind source separation (BSS) methods reduce to the following minimization problem:
  \[
  \min_{W[f]} \mathcal{P}(W[f]|x[t,f]) = \sum_{f=1}^{F} \log |\det(W[f])|
  \]
- Laplace-distribution-based independent component analysis (FDICA)
  \[
  \mathcal{P}(y[t,f]|x[t,f]) = C \sum_{m=1}^{M} T \sum_{t=1}^{T} |m[t,f]| \]
- Spherical-Laplace-distribution-based Independent vector analysis (IVA)
  \[
  \mathcal{P}(y[t,f]|x[t,f]) = C \sum_{m=1}^{M} T \sum_{t=1}^{T} |m[t,f]|^2 \]

- Proximal algorithm has been proposed for handling these models [1].
  - Proximity operator of the source model \( \text{prox}_g[\cdot] \) is required in the 6th line. This requirement is the major limitation of the algorithm as deriving a proximity operator for some source models may be complicated or impossible.
  - This paper heuristically extends this algorithm to deal with the limitation of the applicability.

Proposed Algorithm

- Proximity operator of the source model in the 6th line is replaced by time-frequency masking.
  - Any mask generator which may not be written as a mathematical formula such as rule-based one can be inserted to obtain a new BSS algorithm (convergence and stability of the algorithm should be checked experimentally).

- Proximity operator corresponds to the maximum a posteriori (MAP) estimator of the following form (Gaussian denoiser), where observed signals are assumed to be contaminated by additive Gaussian noise, and the source model is employed as the prior distribution:
  \[
  \text{prox}_{\mu,p}[z] = \arg \max_{\xi} \left( e^{-\frac{1}{2} \|z - \xi\|^2} e^{-\mathcal{P}(\xi)} \right)
  \]
- The proposed algorithm can be interpreted as replacement of the BSS problem by the denoising problem with the same prior distribution of source signals. This is important property because learning a Gaussian denoiser is much easier than learning a regressor of demixing matrices.

Experimental Evaluation of Sparse IVA

- Comparing to the ordinary IVA, SDR improved 3.3 dB in average by only requiring 1.2x computational efforts. As IVA can be recovered as a special case, Sparse IVA can be seen as an improved version of IVA.

Proximity Operators as T-F Masking

- Proximity operator is a map defined by the following optimization problem (whose solution is unique if the function is convex):
  \[
  \text{prox}_{\mu,p}[z] = \arg \min_{\xi} \left( \mathcal{P}(\xi) + \frac{1}{2\mu} \|z - \xi\|^2 \right)
  \]
- Some proximity operator related to sparsity has closed-form solution:
  - (bin-wise) soft-thresholding operator (corresponding to the 1-norm):
    \[
    \text{prox}_{\lambda}[z][m] = \left( 1 - \frac{\lambda}{z_m} \right)_+ z_m[1:t ф]
    \]
  - Group-thresholding operator (corresponding to the 2,1-mixed norm):
    \[
    \text{prox}_{\lambda}[\mathcal{M}_z][m] = \left( 1 - \frac{\lambda}{\sum_{t=1}^{T} |z_m[t,f]|^2} \right)_+ z_m[1:t ф]
    \]
  - These thresholding operators can be interpreted as time-frequency masking operators (bin-wise multiplication of scalars in \([0,1]):
    \[
    \mathcal{T}_z[m] = \lambda \mathcal{T}_z[m] \frac{\text{prox}_{\lambda}[\mathcal{M}_z][m]}{z_m[1:t ф]}
    \]
- Time-frequency masks are functions of inputted signals which may be obtained mathematically or as some procedures:
  \[
  \mathcal{M}_z[m] = \left( 1 - \frac{\lambda}{\sum_{t=1}^{T} |z_m[t,f]|^2} \right)_+
  \]

Application: Sparse IVA

- Whitening induces not only the positive effect but also the side effect illustrated on the right.
  - Frequency-wise treatment of the whitening distorted the group sparse structure of spectrogram of speech signals assumed in IVA.
  - We propose the following mask-generating function to improve IVA by recovering the group sparseness and enhancing the thresholder:
    \[
    \mathcal{M}[z][m] = \Xi_{\alpha} \left( \left( 1 - \frac{\lambda}{\sum_{t=1}^{T} |z_m[t,f]|^2} \right)_+ \right)
    \]
  - Frequency-wise weight for recovering the group sparse structure:
    \[
    \Xi_{\alpha}[x] = \Theta_{\Psi_{\alpha}[x]} \left( \sum_{t=1}^{T} |z_m[t,f]|^2 \right)_+ \left( \sum_{t=1}^{T} |z_m[t,f]|^2 \right)_+
    \]
- Firm-thresholder for enhancing the sparsity and reducing the bias:
  \[
  \Xi_{\alpha}[x][m] = \Xi_{\alpha} \left( 1 - \frac{\lambda}{\sum_{t=1}^{T} |z_m[t,f]|^2} \right)_+ \left( \sum_{t=1}^{T} |z_m[t,f]|^2 \right)_+
  \]