Mathieu FONTAINE
mathieu.fontaine@riken.jp

Fast Multichannel Nonnegative Matrix Factorization with Gaussian Scale Mixture Distributions

KYOTO SAP Seminar

matfontaine.github.io

January 27th, 2021

Outline

I - FastMNMF for Multichannel Blind Speech Separation [Sek. 20]

II - Gaussian Scale Mixture

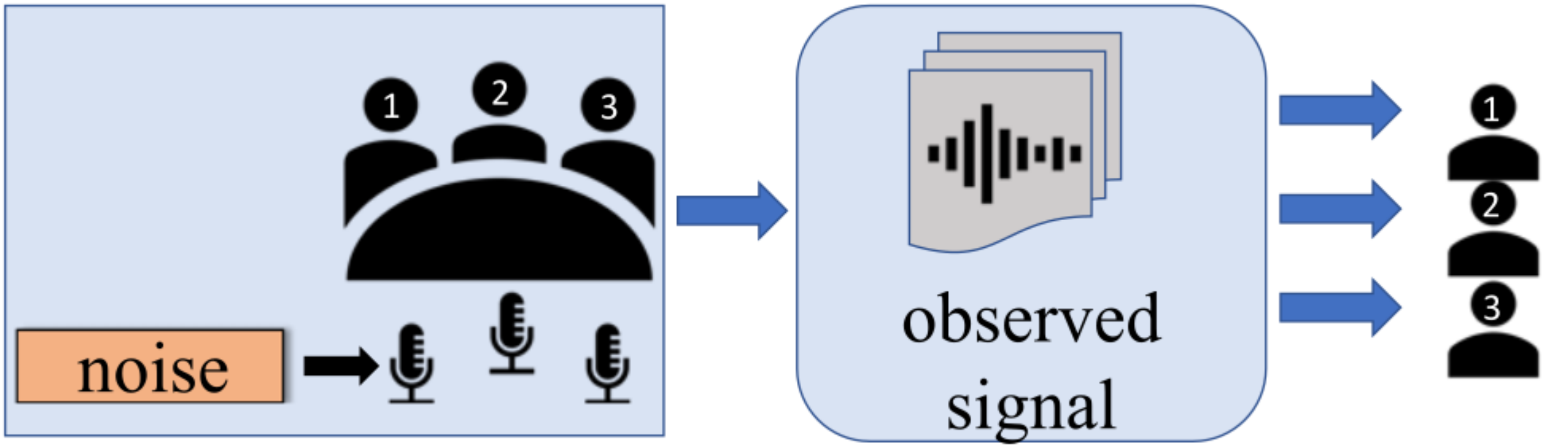
III - Variational Lower Bound Through Expectation-Maximization Algorithm : Application to GH-FastMNMF and β-FastMNMF

IV - Speech Enhancement and Speaker Separation Experiments

V - Conclusion and Future Works

▹ Sekiguchi, K. et al. (2020, TASLP). FastMNMF with Directivity-Aware Jointly-Diagonalizable Spatial Covariance Matrices for Blind Source Separation

Multichannel Blind Speech Separation?



$$\text{Goal: extract speakers 1,2,3}$$

In the Short-time Fourier transform (STFT) domain with $\mathbf{x}_{ft} \in \mathbb{C}^M$:

$$\underbrace{\mathbf{x}_{ft}}_{\text{observation}} = \underbrace{\sum_{n=1}^{N-1} \mathbf{x}_{nft}}_{\text{speakers}} + \underbrace{\mathbf{x}_{Nft}}_{\text{noise}}$$

M: number of channels

F : number of frequency bins

T : number of time frame

N: number of sources

Spatial Gaussian Model + MNMF

$$\forall n, \ \ \mathbf{x}_{nft} \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \lambda_{nft} \mathbf{G}_{nft}\right)$$

scale parameter
$\mathbb{R}_+$

spatial covariance matrix
$\mathbb{C}^{M \times M}$

$$\Downarrow$$

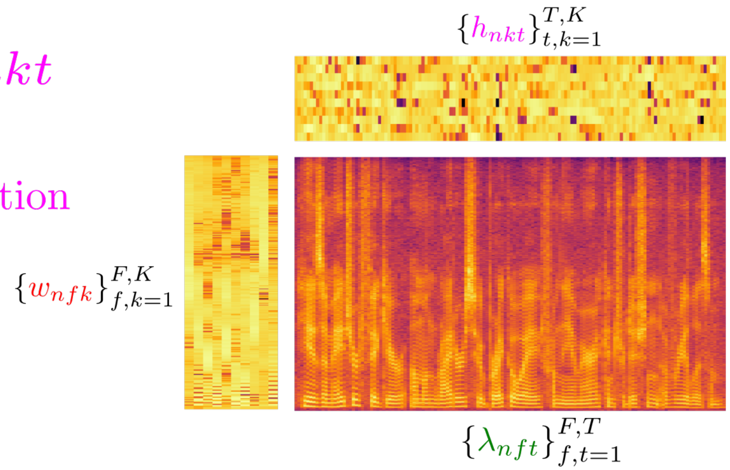
$$\mathbf{x}_{ft} = \sum_{n=1}^{N} \mathbf{x}_{nft} \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \sum_{n=1}^{N} \lambda_{nft} \mathbf{G}_{nft}\right)$$

$$\lambda_{nft} = \sum_{k=1}^{K} w_{nfk} h_{nkt}$$

frequency basis
$\mathbb{R}_+$

time activation
$\mathbb{R}_+$

MNMF Model



▹ Duong, N. et al. (2009, TASLP). Under-determined reverberant audio source separation using a full-rank spatial covariance model.

Fast Gaussian MNMF Models

Independent Low-Rank Matrix Analysis (ILRMA) [Kit. 18]

- $\mathbf{x}_{nft} = \mathbf{a}_{nf} s_{nft}$ (Direct sound propagation model)
- Then $\mathbf{x}_{nft} \sim \mathcal{N}_{\mathbb{C}}\left(\lambda_{nft}\mathbf{a}_{nf}\left(\mathbf{a}_{nf}\right)^{\mathrm{H}}\right)$ (Rank-1 SCM model)
- MNMF model for λ_{nft} parameters

Fast MNMF 2: a joint diagonalization (JD) technique [Sek. 19, 20]

- $\mathbf{x}_{nft} \sim \mathcal{N}_{\mathbb{C}}\left(\lambda_{nft}\underbrace{\mathbf{Q}_f^{-1}\mathrm{Diag}(\tilde{\mathbf{g}}_n)\mathbf{Q}_f^{-\mathrm{H}}}_{=\mathbf{G}_{nf}}\right)$ (JD - FastMNMF2)
- MNMF model for λ_{nft} parameters
- ILRMA $\subset$ FastMNMF2

▹ Kitamura, D. et al. (2018, Audio Source Separation, Springer). Determined blind source separation with independent low-rank matrix analysis.
▹ Sekiguchi, K. et al. (2020, TASLP) FastMNMF with Directivity-Aware Jointly-Diagonalizable Spatial Covariance Matrices for Blind Source Separation

Multiplicative update strategy

- Expectation-Maximization approach $\Longrightarrow$ minimization of the log-likelihood
- $w_{nfk} \leftarrow w_{nfk}\sqrt{\frac{\sum_{t,m=1}^{T,M} h_{nkt}\tilde{g}_{nm}\tilde{x}_{ftm}\tilde{y}_{ftm}^{-2}}{\sum_{t,m=1}^{T,M} h_{nkt}\tilde{g}_{nm}\tilde{y}_{ftm}^{-1}}}$; $h_{nkt} \leftarrow h_{nkt}\sqrt{\frac{\sum_{f,m=1}^{F,M} w_{nfk}\tilde{g}_{nm}\tilde{x}_{ftm}\tilde{y}_{ftm}^{-2}}{\sum_{f,m=1}^{F,M} w_{nfk}\tilde{g}_{nm}\tilde{y}_{ftm}^{-1}}}$;
- $\tilde{g}_{nm} \leftarrow \tilde{g}_{nm}\sqrt{\frac{\sum_{f,t,m=1}^{F,T,M} \lambda_{nft}\tilde{x}_{ftm}\tilde{y}_{ftm}^{-2}}{\sum_{f,t,m=1}^{F,T,M} \lambda_{nft}\tilde{y}_{ftm}^{-1}}}$ where $\tilde{\mathbf{g}}_n = [\tilde{g}_{n1}, \ldots, \tilde{g}_{nm}]^\top$
- $\tilde{x}_{ftm} = \left|\mathbf{q}_{fm}^{\mathrm{H}}\mathbf{x}_{ft}\right|$, $\tilde{y}_{ftm} = \sum_{n=1}^{N} \lambda_{nft}\tilde{g}_{nm}$

Iterative projection method [Ono 11]

- $\mathbf{q}_{fm} \leftarrow (\mathbf{Q}_f\mathbf{V}_{fm})^{-1}\mathbf{e}_m$; $\quad \mathbf{q}_{fm} \leftarrow (\mathbf{q}_{fm}^{\mathrm{H}}\mathbf{V}_{fm}\mathbf{q}_{fm})^{-\frac{1}{2}}\mathbf{q}_{fm}$

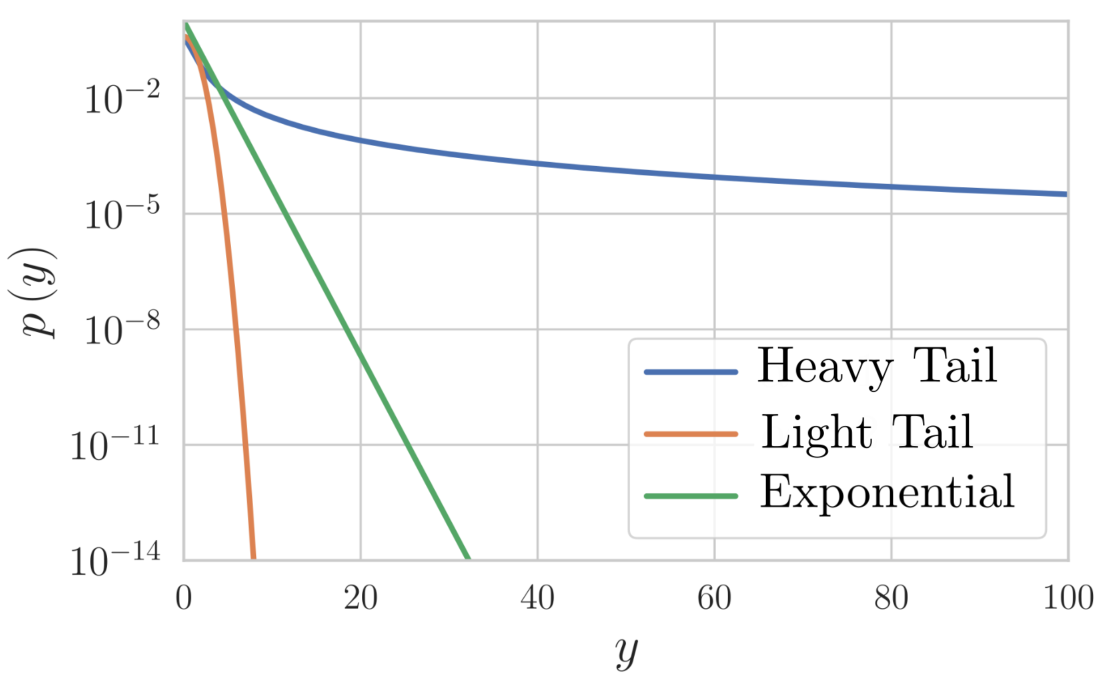
$$\mathbf{V}_{fm} = \tfrac{1}{T}\sum_{t=1}^{T}\mathbf{x}_{ft}\mathbf{x}_{ft}^{\mathrm{H}}y_{ftm}^{-1}$$

$$\mathbf{e}_m = [\delta_{1,m}\ldots,\delta_{M,m}]^\top \text{ with } \delta_{m,m'} = \begin{cases} 1 & \text{if } m = m' \\ 0 & \text{otherwise} \end{cases}$$

▹ Ono, N. (2011, WASPAA). Stable and fast update rules for independent vector analysis based on auxiliary function technique

Drawbacks of Gaussian MNMF Models

- The MNMF initialization is sometimes tricky [Bou. 08]
- Light tails $\implies$ less robust against impulsive noise or uncommon scenario



In [Sim. 19], suggest to use heavy-tailed models for gradient descent algorithm

▹ Boutsidis C. (2008, Pattern Recognition). SVD based initialization: A head start for nonnegative matrix factorization
▹ Simsekli U. (2019, Deep AI). A Tail-Index Analysis of Stochastic Gradient Noise in Deep Neural Networks

Gaussian Scalar Mixture (GSM) Distribution

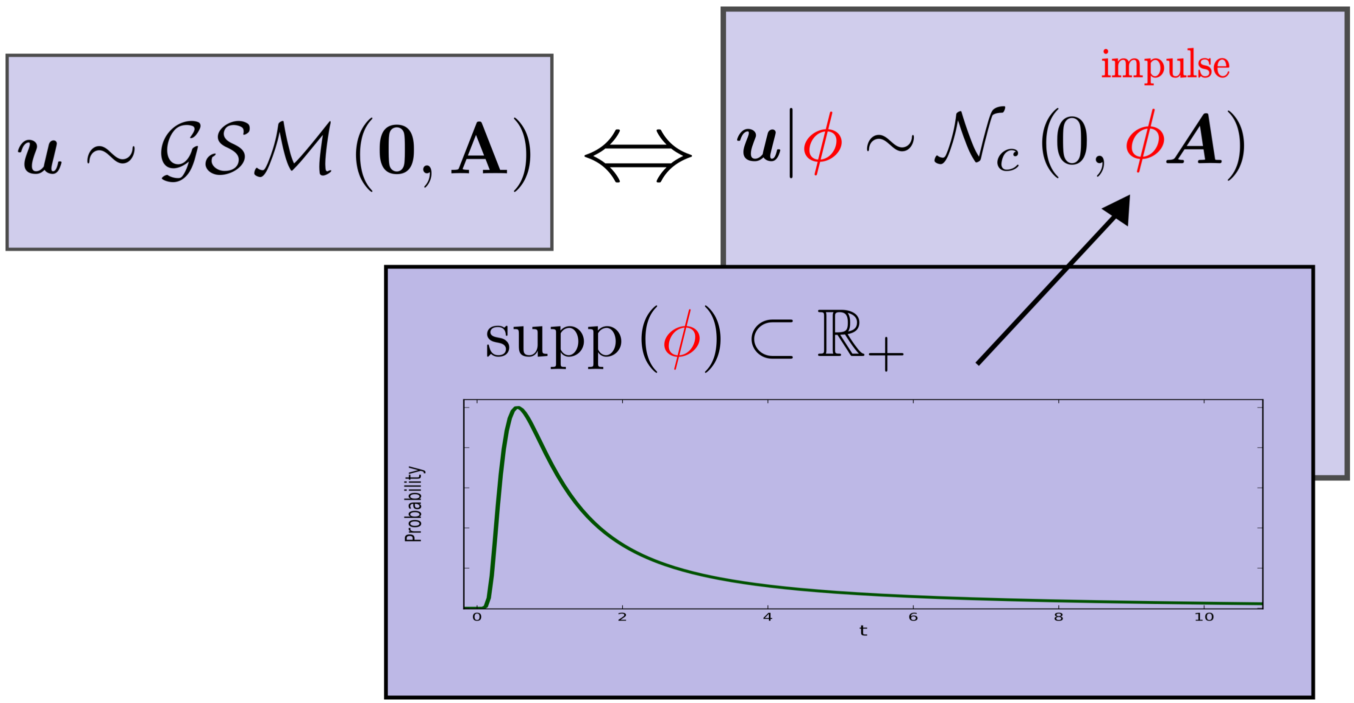
- Gaussian where the covariance is randomly perturbed

- If $\mathbf{u}$ is a GSM, then its PDF. is

$$p(\mathbf{u}) = \int_0^\infty p(\mathbf{u} \mid \phi)\, p(\phi)\, d\phi$$

with

$$\mathbf{u} \mid \phi \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \phi\mathbf{A})$$



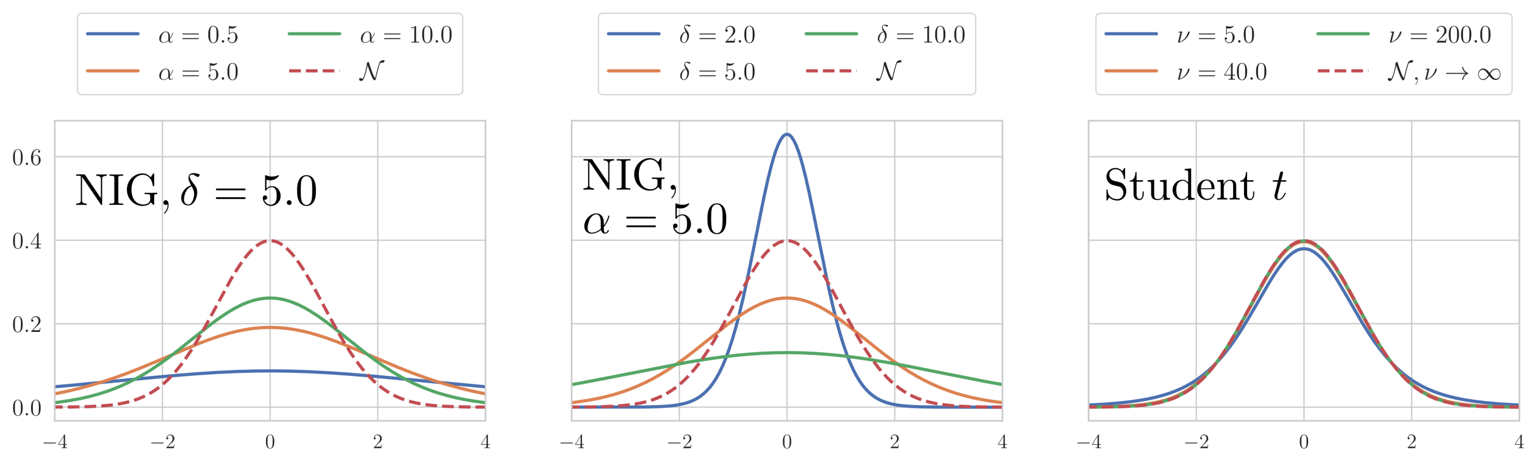
Examples of Gaussian Scale Mixture (1/2)

(Symmetric Isotropic) Generalized Hyperbolic (GH) distribution

- ϕ is known: inverse Gaussian distribution
- The PDF of $\mathbf{u}$ is given by:

$$p(\mathbf{u}) = C_{\eta,\alpha,\delta,\mathbf{A}} \left(\frac{2\mathbf{u}^{\mathrm{H}} \boldsymbol{A}^{-1} \mathbf{u} + 1}{(\delta\alpha)^2} \right)^{\frac{\eta - M}{2}} \mathcal{K}_{\eta - M} \left(\frac{\alpha}{\delta} \sqrt{2\mathbf{u}^{\mathrm{H}} \boldsymbol{A}^{-1} \mathbf{u} + 1} \right)$$

- η, α: controls the heaviness of the tails
- $\delta, \mathbf{A}$: shape ("scale") parameter and covariance matrix
- $e.g.$ Student t $(\eta = \frac{-\nu}{2}, \alpha = 0, \delta = \sqrt{\nu})$, Gaussian, Normal-Inverse Gaussian (NIG) $(\eta = -0.5, \alpha > 0, \delta > 0)$ are GH



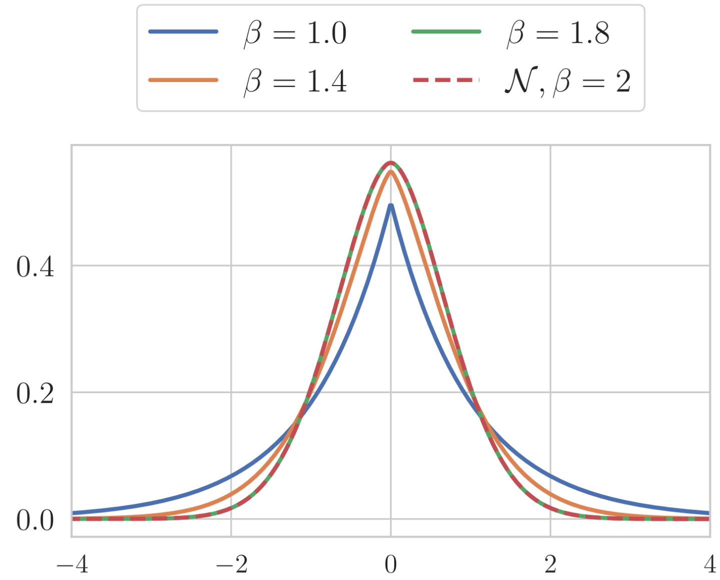
Examples of Gaussian Scale Mixture (2/2)

(Symmetric) Generalized Super-Gaussian Distribution

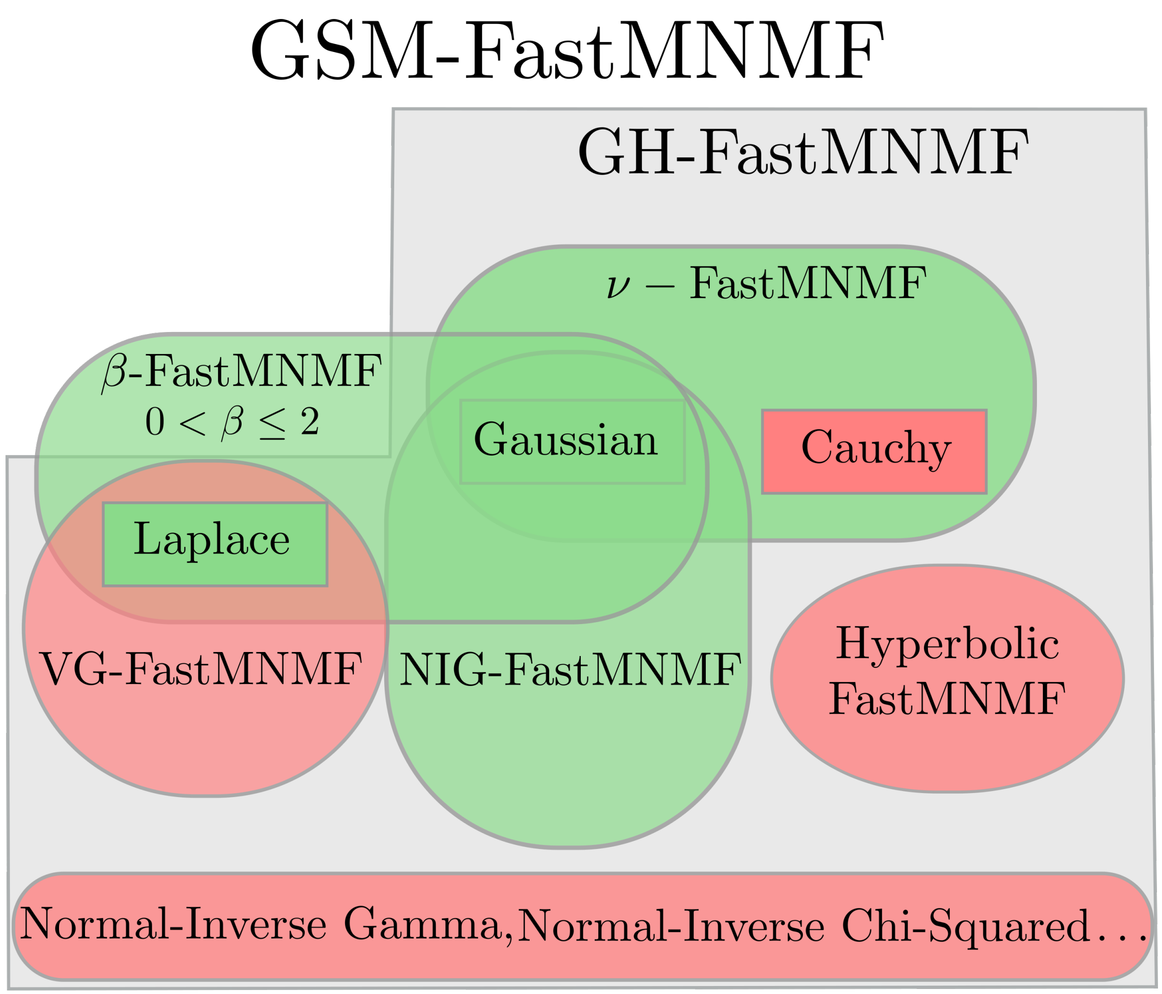
- for $0 < \beta \leq 2$, GSM. But ϕ unknown ! (except for $\beta = 1, \beta = 2$)
- The PDF of $\mathbf{u}$ is given by:

$$p(\mathbf{u}) = C_{\beta,\mathbf{A}} \exp\left(-\left[\mathbf{u}^{\mathrm{H}}\mathbf{A}^{-1}\mathbf{u}\right]^{\frac{\beta}{2}}\right)$$

- β: shape parameter controling the tail-index
- $\mathbf{A}$: shape matrix



A plethora of new FastMNMF models



How to derive a parameter technique that unify all those probabilistic models ?

Probabilistic Model

- We assume a GSM model on sources:

$$\forall n, \mathbf{x}_{nft} \mid \phi_{ft} \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \phi_{ft}\lambda_{nft}\mathbf{G}_{nft}\right)$$

- The covariance perturbation ϕ_{ft} is the same for all sources
- The mixing model becomes:

$$\mathbf{x}_{ft} \mid \phi_{ft} \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \phi_{ft}\sum_{n=1}^{N}\lambda_{nft}\mathbf{G}_{nft}\right)$$

GSM FastMNMF + Filtering Method

Weighted-shared JD model

- $$\mathbf{Y}_{ft} = \mathbf{Q}_f^{-1} \left(\underbrace{\phi_{ft} \sum_{n=1}^{N} \underbrace{\lambda_{nft} \mathrm{Diag}\left(\tilde{\mathbf{g}}_n\right)}_{=\mathrm{Diag}(\tilde{\mathbf{y}}_{nft})}}_{=\mathrm{Diag}(\tilde{\mathbf{y}}_{ft})} \right) \mathbf{Q}_f^{-\mathrm{H}}$$
- $\mathbf{x}_{ft} \mid \mathbf{\Phi} \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \mathbf{Y}_{ft}\right)$
- MNMF model for λ_{nft} parameters

Marginalized Wiener filter

- Using the conditional Gaussian model we have:

$$\mathbb{E}_{\phi}\left[\mathbb{E}\left[\mathbf{x}_{nft} \mid \mathbf{\Theta}, \phi, \mathbf{x}_{ft}\right]\right] = \mathbf{Q}_f^{-1} \mathrm{Diag}\left(\tilde{\mathbf{y}}_{nft}\right) \mathrm{Diag}\left(\tilde{\mathbf{y}}_{ft}\right)^{-1} \mathbf{Q}_f^{-\mathrm{H}} \mathbf{x}_{ft}.$$

- Where $\mathbf{\Theta} = \left\{\mathbf{W}, \mathbf{H}, \tilde{\mathbf{G}}, \mathbf{Q}\right\}$

The filtering technique is equivalent to the classical Multichannel Wiener filter

Variational lower-bound

We develop a Majorization-Minimization algorithm for the parameter estimation

- Consider the LL: $\log p(\mathbf{X}|\boldsymbol{\Theta}) = \log \int p(\mathbf{X}|\boldsymbol{\Theta}, \boldsymbol{\phi})p(\boldsymbol{\phi})d\boldsymbol{\phi}$
- Using the fact that $\mathbf{X}$ is a GSM, we get:

$$\log p(\mathbf{X}|\boldsymbol{\Theta}) \overset{c}{\geq}$$
$$\sum_{n,f,t,k,m=1}^{N,F,T,K,M} \left(-\omega_{ftm}^{-1}\tilde{g}_{nm}w_{nkf}h_{nkt} + \tilde{x}_{ftm}\pi_{nkftm}^{2}\tilde{g}_{nm}^{-1}w_{nkf}^{-1}h_{nkt}^{-1}\mathbb{E}_{q(\theta)}\left[\phi_{ft}^{-1}\right]\right)$$
$$-T\sum_{f=1}^{F}\log\left|\mathbf{Q}_f\mathbf{Q}_f^{\mathrm{H}}\right| - \mathrm{KL}\left[q\left(\phi_{ft}\right) \mid\mid p\left(\phi_{ft}\right)\right]$$

- KL denotes the Kullback-Leibler divergence
- $\tilde{x}_{ftm} = \left|\mathbf{q}_{fm}^{\mathrm{H}}\mathbf{x}_{ft}\right|$
- $\omega_{ftm}, \pi_{nkftm}$ are auxiliary variables that depends on $\boldsymbol{\Theta}$ to satisify the equality
- $q\left(\theta\right)$ satisfy the equality with the LL when $q(\theta) = p\left(\phi \mid \mathbf{X}, \boldsymbol{\Theta}\right)$

How to compute $\mathbb{E}_{q(\theta)}\left[\phi_{ft}^{-1}\right]$?

E-Step: computation of $\mathbb{E}_{p(\phi|\mathbf{X},\mathbf{\Theta})}\left[\phi_{ft}^{-1}\right]$

Thanks to the GSM assumption, it can be shown that:

$$\frac{d \log p(\mathbf{x}_{ft})}{d\mathbf{x}_{ft}^{\mathrm{H}}} = -\sum_{m=1}^{M} \mathbf{2q}_{fm}^{\mathrm{H}}\mathbf{x}_{ft}\mathbf{q}_{fm}\tilde{y}_{ftm}^{-1}\mathbb{E}_{p(\phi|\mathbf{X},\mathbf{\Theta})}\left[\phi_{ft}^{-1}\right]$$

- $\tilde{y}_{ftm} = \sum_{n,k=1}^{N,K} \tilde{g}_{nm} w_{nfk} h_{nkt}$
- Only the knowledge of the log PDF is required
- The knowledge of the law of ϕ_{ft} is not necessary !

M-Step

Multiplicative update rules (MUR)

- Let assume that $\tilde{\phi}_{ft}^{-1} \triangleq \mathbb{E}_{p(\phi|\mathbf{X},\boldsymbol{\Theta})}\left[\phi_{ft}^{-1}\right]$ are known
- As in FastMNMF, the MURs are given as the Itakura-Saito (IS) minimization:
- Minimization-Maximization approach $\Longrightarrow$ minimization of the log-likelihood
- $w_{nfk} \leftarrow w_{nfk}\sqrt{\frac{\sum_{t,m=1}^{T,M} \tilde{\phi}_{ft}^{-1} h_{nkt}\tilde{g}_{nm}\tilde{x}_{ftm}\tilde{y}_{ftm}^{-2}}{\sum_{t,m=1}^{T,M} \tilde{\phi}_{ft}^{-1} h_{nkt}\tilde{g}_{nm}\tilde{y}_{ftm}^{-1}}}; h_{nkt} \leftarrow h_{nkt}\sqrt{\frac{\sum_{f,m=1}^{F,M} \tilde{\phi}_{ft}^{-1} w_{nfk}\tilde{g}_{nm}\tilde{x}_{ftm}\tilde{y}_{ftm}^{-2}}{\sum_{f,m=1}^{F,M} \tilde{\phi}_{ft}^{-1} w_{nfk}\tilde{g}_{nm}\tilde{y}_{ftm}^{-1}}};$
- $\tilde{g}_{nm} \leftarrow \tilde{g}_{nm}\sqrt{\frac{\sum_{f,t,m=1}^{F,T,M} \tilde{\phi}_{ft}^{-1} \lambda_{nft}\tilde{x}_{ftm}\tilde{y}_{ftm}^{-2}}{\sum_{f,t,m=1}^{F,T,M} \tilde{\phi}_{ft}^{-1} \lambda_{nft}\tilde{y}_{ftm}^{-1}}}$ where $\tilde{\mathbf{g}}_n = \left[\tilde{g}_{n1}, \ldots, \tilde{g}_{nm}\right]^\top$

Iterative projection method

- $\mathbf{q}_{fm} \leftarrow (\mathbf{Q}_f\mathbf{V}_{fm})^{-1}\mathbf{e}_m; \quad \mathbf{q}_{fm} \leftarrow (\mathbf{q}_{fm}^{\mathrm{H}}\mathbf{V}_{fm}\mathbf{q}_{fm})^{-\frac{1}{2}}\mathbf{q}_{fm}$

$$\mathbf{V}_{fm} = \frac{1}{T}\sum_t^T \tilde{\phi}_{ft}^{-1}\mathbf{x}_{ft}\mathbf{x}_{ft}^{\mathrm{H}}y_{ftml}^{-1}$$

$$\mathbf{e}_m = \left[\delta_{1,m} \ldots, \delta_{M,m}\right]^\top \text{ with } \delta_{m,m'} = \begin{cases} 1 & \text{if } m = m' \\ 0 & \text{otherwise} \end{cases}$$

E-Step for GH-FastMNMF and β-FastMNMF

We apply the following formula in the case of a GH model and a β-FastMNMF:

$$\frac{d\log p(\mathbf{x}_{ft})}{d\mathbf{x}_{ft}^{\mathrm{H}}} = -\sum_{m=1}^{M} \mathbf{2q}_{fm}^{\mathrm{H}} \mathbf{x}_{ft} \mathbf{q}_{fm} \tilde{y}_{ftm}^{-1} \mathbb{E}_{p(\phi|\mathbf{X},\mathbf{\Theta})} \left[\phi_{ft}^{-1}\right]$$

GH-FastMNMF

We get the following result for $\tilde{\phi}_{ft}^{-1} \triangleq \mathbb{E}_{p(\phi|\mathbf{X},\mathbf{\Theta})} \left[\phi_{ft}^{-1}\right]$:

$$\tilde{\phi}_{ft}^{-1} = \left(\frac{2(M-\eta)}{\gamma_{ft}} + \alpha \frac{\mathcal{K}_{\eta-M+1}\left(\alpha^2\gamma_{ft}\right)}{\gamma_{ft}\mathcal{K}_{\eta-M}\left(\alpha^2\gamma_{ft}\right)}\right)$$

where $\gamma_{ft} \triangleq 1 + \frac{2}{\delta^2}\sum_m \tilde{x}_{ftm}\tilde{y}_{ftm}^{-1}$

- The results coincide with the already proposed ν-FastMNMF and $\mathcal{N}$-FastMNMF

β-FastMNMF

We get the following result for $\tilde{\phi}_{ft}^{-1} \triangleq \mathbb{E}_{p(\phi|\mathbf{X},\mathbf{\Theta})} \left[\phi_{ft}^{-1}\right]$:

$$\tilde{\phi}_{ft}^{-1} = \frac{\beta}{2}\left(\sum_{m=1}^{M} \tilde{x}_{ftm} y_{ftm}^{-1}\right)^{\frac{\beta-2}{2}}$$

Setting for Speech Enhancement

Dataset description

- REVERB CHALLENGE dataset sampled at 16 kHz recorded with 8 microphones
- RT_{60} are either 0.25, 0.5 or 0.7s
- 3 Signal to noise ratio level: $0, 5, 10$ dB
- 2 distances: "near" (50cm between mic and speaker) & "far" ($\simeq$ 2m)
- $M \in \{2, 5, 8\}$ are considered with $N = M$ (to include ILRMA)
- 100 utterances for the first experiment (dev set)
- 200 utterances for the second experiment (test set)

Scores

- Signal to Distorsion Ratio (SDR), Perceptual Evaluation of Speech Quality (PESQ) (higher is better)

Methods

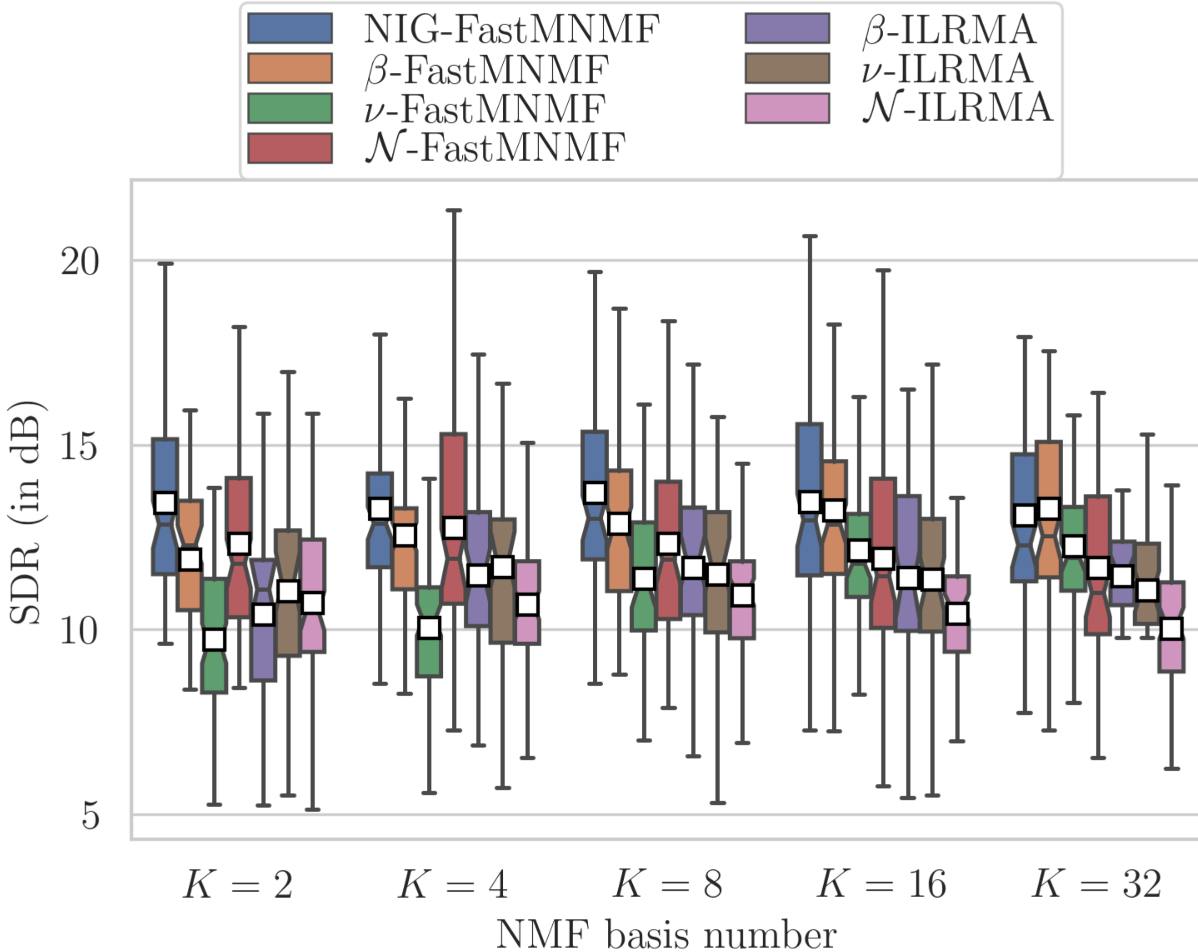
- $\mathcal{N}, \nu$-FastMNMF: Gaussian and Student t weighted-shared JD SCM decomposition + MNMF
- $\mathcal{N}, \beta, \nu$-ILRMA: Gaussian, Super-Gaussian and Student t rank 1 constrained SCM + direct sound propagation model + MNMF
- NIG-FastMNMF: proposed method with hyperparameters (α, δ)
- β-FastMNMF: proposed Super Gaussian FastMNMF with $0 < \beta \leq 2$

Settings

- 300 iterations for the EM algorithm is considered
- NMF coefficients are randomly initialized
- Demixing matrix in ILRMA and $\mathbf{Q}_f$ are initialized as identity matrix $\forall f$
- The Matrix $[\tilde{\mathbf{g}}_1, \ldots, \tilde{\mathbf{g}}_N]^\top \in \mathbb{R}^{N \times M}$ is initialized as the circulant matrix

$$\begin{pmatrix} 1 & \epsilon & \ldots & \epsilon & 1 & \epsilon & \ldots \\ \epsilon & 1 & \ldots & \epsilon & \epsilon & 1 & \ldots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots \\ \epsilon & \epsilon & \ldots & 1 & \epsilon & \epsilon & \ldots \end{pmatrix}$$

SDR performance along the number of NMF basis K



- $M = 8$ and best hyperparameters for all methods
- NIG-FastMNMF seems better for a small K compare to ν-FastMNMF

PESQ results for different settings

Distance	SNR	M	FastMNMF variants				ILRMA variants		
			NIG	β	ν	$\mathcal{N}$	β	ν	$\mathcal{N}$
Near	0 dB	2	1.9 ($\pm$0.6)	**2.0 ($\pm$0.7)**	1.9 ($\pm$0.6)	1.8 ($\pm$0.6)	1.8 ($\pm$0.6)	1.7 ($\pm$0.6)	1.7 ($\pm$0.6)
		5	**2.4 ($\pm$0.7)**	2.4 ($\pm$0.7)	2.4 ($\pm$0.7)	2.3 ($\pm$0.7)	2.0 ($\pm$0.7)	2.1 ($\pm$0.7)	2.0 ($\pm$0.7)
		8	**2.6 ($\pm$0.8)**	2.6 ($\pm$0.8)	2.5 ($\pm$0.7)	2.4 ($\pm$0.7)	2.3 ($\pm$0.7)	2.3 ($\pm$0.7)	2.2 ($\pm$0.8)
	5 dB	2	2.2 ($\pm$0.7)	**2.2 ($\pm$0.7)**	2.2 ($\pm$0.6)	2.1 ($\pm$0.7)	2.0 ($\pm$0.7)	2.1 ($\pm$0.7)	2.0 ($\pm$0.6)
		5	2.7 ($\pm$0.7)	2.7 ($\pm$0.6)	**2.7 ($\pm$0.6)**	2.6 ($\pm$0.6)	2.5 ($\pm$0.8)	2.4 ($\pm$0.8)	2.4 ($\pm$0.7)
		8	**2.9 ($\pm$0.7)**	2.9 ($\pm$0.7)	2.8 ($\pm$0.7)	2.8 ($\pm$0.6)	2.6 ($\pm$0.8)	2.6 ($\pm$0.8)	2.5 ($\pm$0.7)
	10 dB	2	2.5 ($\pm$0.6)	**2.5 ($\pm$0.6)**	2.4 ($\pm$0.6)	2.3 ($\pm$0.6)	2.2 ($\pm$0.8)	2.3 ($\pm$0.8)	2.2 ($\pm$0.7)
		5	**3.0 ($\pm$0.5)**	3.0 ($\pm$0.5)	3.0 ($\pm$0.5)	2.8 ($\pm$0.5)	2.7 ($\pm$0.9)	2.7 ($\pm$0.9)	2.7 ($\pm$0.6)
		8	**3.2 ($\pm$0.5)**	3.2 ($\pm$0.5)	3.1 ($\pm$0.5)	3.0 ($\pm$0.5)	2.9 ($\pm$0.9)	2.9 ($\pm$0.9)	2.8 ($\pm$0.6)
Far	0 dB	2	1.7 ($\pm$0.4)	**1.7 ($\pm$0.4)**	1.7 ($\pm$0.4)	1.6 ($\pm$0.4)	1.5 ($\pm$0.4)	1.5 ($\pm$0.4)	1.5 ($\pm$0.4)
		5	2.1 ($\pm$0.6)	**2.1 ($\pm$0.5)**	2.0 ($\pm$0.5)	1.9 ($\pm$0.5)	1.9 ($\pm$0.5)	1.8 ($\pm$0.5)	1.8 ($\pm$0.5)
		8	**2.3 ($\pm$0.7)**	2.2 ($\pm$0.6)	2.2 ($\pm$0.6)	2.1 ($\pm$0.6)	1.9 ($\pm$0.5)	1.9 ($\pm$0.5)	1.9 ($\pm$0.6)
	5 dB	2	**1.9 ($\pm$0.4)**	1.9 ($\pm$0.4)	1.9 ($\pm$0.4)	1.7 ($\pm$0.4)	1.7 ($\pm$0.5)	1.7 ($\pm$0.5)	1.7 ($\pm$0.4)
		5	**2.2 ($\pm$0.4)**	2.2 ($\pm$0.5)	2.2 ($\pm$0.4)	2.1 ($\pm$0.5)	2.1 ($\pm$0.5)	2.0 ($\pm$0.6)	2.0 ($\pm$0.5)
		8	2.3 ($\pm$0.5)	**2.4 ($\pm$0.6)**	2.3 ($\pm$0.5)	2.1 ($\pm$0.5)	2.2 ($\pm$0.6)	2.2 ($\pm$0.6)	2.1 ($\pm$0.6)
	10 dB	2	**2.0 ($\pm$0.4)**	2.0 ($\pm$0.4)	2.0 ($\pm$0.4)	1.8 ($\pm$0.4)	1.8 ($\pm$0.5)	1.8 ($\pm$0.5)	1.8 ($\pm$0.4)
		5	**2.3 ($\pm$0.4)**	2.3 ($\pm$0.4)	2.3 ($\pm$0.4)	2.1 ($\pm$0.4)	2.1 ($\pm$0.6)	2.2 ($\pm$0.6)	2.1 ($\pm$0.5)
		8	**2.5 ($\pm$0.5)**	2.5 ($\pm$0.5)	2.5 ($\pm$0.5)	2.3 ($\pm$0.5)	2.3 ($\pm$0.6)	2.4 ($\pm$0.6)	2.3 ($\pm$0.5)

- NMF basis number selected from previous SDR results
- NIG is not always the best PESQ
- The Modifed Bessel function computation for $\tilde{\phi}_{ft}$ could induces outliers

Settings for Speaker Separation

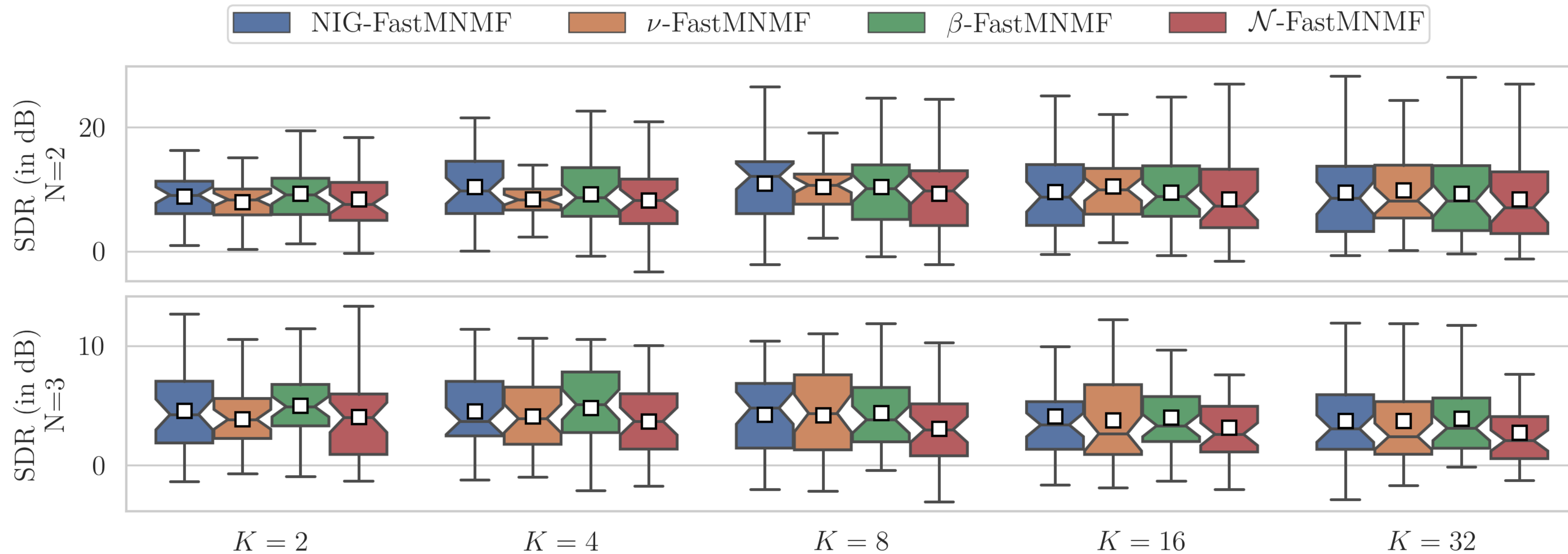
Dataset description

- spatialized WSJ0-2,3mix dataset sampled at 16 kHz recorded with 8 microphones
- RT$_{60}$ ranging from $0.2s$ to $0.6s$
- $N = 2$ or $N = 3$ speakers and $M = N, 5, 8$ (determined/overdetermined case)
- 100 utterances for the first experiment (dev set)
- 200 utterances for the second experiment (test set)

Scores

- Signal to Distorsion Ratio (SDR), Signal to Artifiact Ratio (PESQ) and Signal to Inference Ratio (SIR)(higher is better)

SDR performance along the number of NMF basis K



- NIG outperforms other methods for $K = 8$
- β-FastMNMF is also performant for a small number K
- β-FastMNMF seems to be more performant for 3 speakers

SDR,SAR,SIR performances

N	M	score	FastMNMF variants			
			NIG	β	ν	$\mathcal{N}$
2	2	SDR	**3.9** (**±3.4**)	3.6 (±3.3)	2.8 (±2.9)	2.8 (±3.4)
		SAR	11.4 (±2.6)	11.6 (±2.6)	**12.9** (**±2.7**)	10.0 (±2.6)
		SIR	**7.3** (**±4.3**)	7.0 (±4.1)	6.0 (±3.6)	6.5 (±4.1)
	5	SDR	8.6 (±5.8)	**8.7** (**±5.6**)	8.0 (±5.2)	7.3 (±5.1)
		SAR	17.2 (±5.0)	17.0 (±4.8)	**18.0** (**±4.4**)	14.9 (±4.1)
		SIR	**13.4** (**±7.1**)	13.4 (±6.8)	12.3 (±6.3)	12.0 (±6.0)
	8	SDR	**9.4** (**±5.6**)	8.9 (±5.8)	8.3 (±4.9)	7.7 (±5.1)
		SAR	19.0 (±4.8)	18.7 (±5.1)	**19.2** (**±4.2**)	16.7 (±4.3)
		SIR	**14.3** (**±7.3**)	14.0 (±7.6)	12.8 (±6.6)	12.7 (±6.7)
3	3	SDR	**1.5** (**±2.3**)	1.3 (±2.1)	1.0 (±2.1)	1.2 (±2.0)
		SAR	9.9 (±1.6)	10.0 (±1.5)	**11.3** (**±1.7**)	8.6 (±2.0)
		SIR	**3.7** (**±2.4**)	3.4 (±2.9)	3.0 (±2.8)	3.7 (±3.0)
	5	SDR	**3.5** (**±3.2**)	3.3 (±3.1)	3.1 (±3.4)	2.8 (±3.2)
		SAR	12.9 (±2.7)	12.8 (±2.4)	**14.1** (**±2.8**)	11.1 (±2.9)
		SIR	**6.5** (**±4.4**)	6.2 (±4.3)	5.9 (±4.2)	6.1 (±4.3)
	8	SDR	**5.1** (**±3.7**)	5.0 (±3.8)	4.5 (±3.6)	4.5 (±3.8)
		SAR	15.7 (±3.4)	15.6 (±3.4)	**16.0** (**±3.2**)	13.8 (±3.6)
		SIR	8.5 (±5.1)	**8.6** (**±5.2**)	7.6 (±4.9)	8.3 (±5.0)

- NIG-FastMNMF SAR performance is maybe due to Modified Bessel Function
- The best SDR is globally got by NIG-FastMNMF
- Surprisely the best SAR is in general for ν-FastMNMF

2 speakers Mix

0:00 / 0:03

Clean Speech 1

0:00 / 0:03

Clean Speech 2

0:00 / 0:03

$\mathcal{N}$-FastMNMF

Speaker 1

0:00 / 0:03

Speaker 2

0:00 / 0:03

ν-FastMNMF

Speaker 1

0:00 / 0:03

Speaker 2

0:00 / 0:03

β-FastMNMF

Speaker 1

0:00 / 0:03

Speaker 2

0:00 / 0:03

NIG-FastMNMF

Speaker 1

0:00 / 0:03

Speaker 2

0:00 / 0:03

3 speakers Mix | Clean Speech 1 | Clean Speech 2 | Clean Speech 3

$\mathcal{N}$-FastMNMF

Speaker 1 | Speaker 2 | Speaker 3

ν-FastMNMF

Speaker 1 | Speaker 2 | Speaker 3

β-FastMNMF

Speaker 1 | Speaker 2 | Speaker 3

NIG-FastMNMF

Speaker 1 | Speaker 2 | Speaker 3

Conclusion & Future Works

Conclusion

- Extension of Gaussian FastMNMF to GSM-FastMNMF
- Outperforms the state-of-the-art given the good set of parameters
- Easy to implement

Future works

- Improve NIG by smoothing the parameters
- Why NIG theoretically seems to work better ?

Thank you for your attention ! Questions ?