Flow-Based Fast Multichannel Nonnegative Matrix Factorization for Blind Source Separation

Aditya Arie Nugraha[1,2] *Kouhei Sekiguchi*[1,2] *Mathieu Fontaine*[3,1] *Yoshiaki Bando*[4,1] *Kazuyoshi Yoshii*[2,1]

[1] Center for Advanced Intelligence Project, RIKEN, Japan
[2] Graduate School of Informatics, Kyoto University, Japan
[3] LTCI, Télécom Paris, Institut Polytechnique de Paris, France
[4] National Institute of Advanced Industrial Science and Technology, Japan

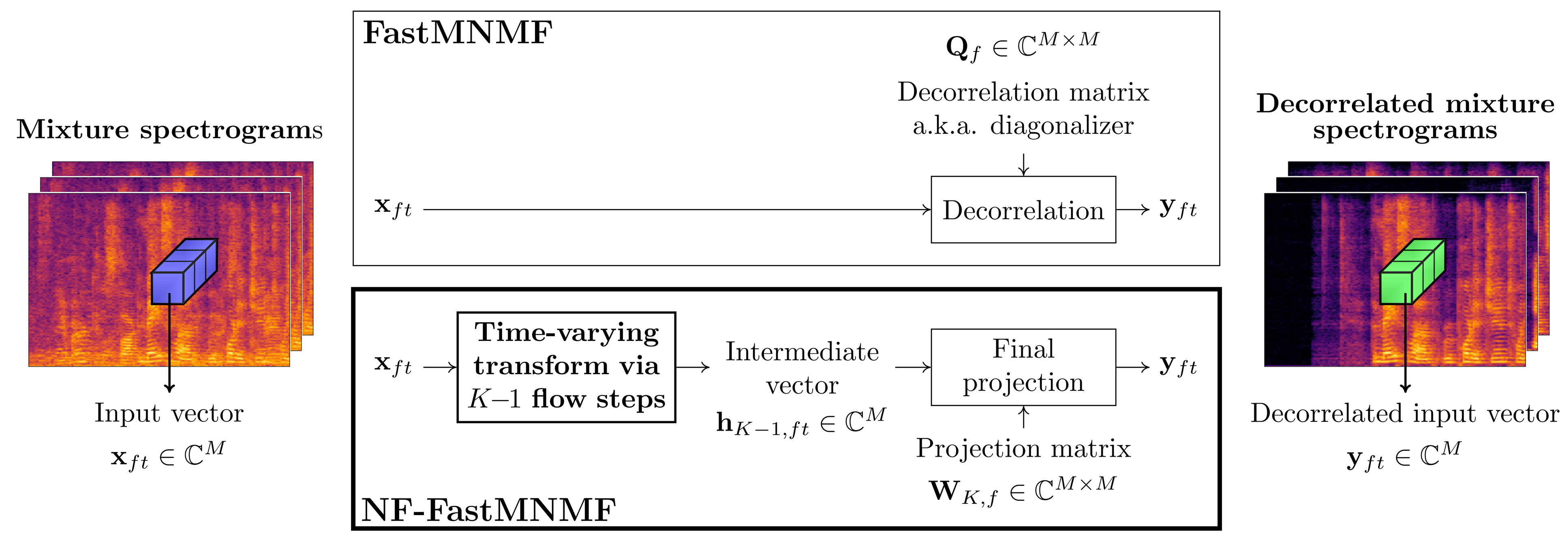
RIKEN AIP icassp 2022 Singapore

① MOTIVATION

- The time-invariant filtering in typical blind source separation (BSS) methods may have sub-optimal interference reduction due to the possible data variation in a mixture.
- NF-IVA [1] that performs time-varying demixing made possible by *normalizing flow (NF)* has been shown to outperform the standard independent vector analysis (IVA) with time-invariant demixing.
- We expect that time-varying transform by NF would also benefit other separation methods, but the NF has been limited to determined separation due to its bijective nature.

② KEY POINTS

- We show that the joint diagonalization technique in FastMNMF [2] *enables the NF to be applicable to non-determined separation.*
- The NF allows us to have *time-varying diagonalization transforms*, instead of time-invariant ones as in FastMNMF, that are expected to better cope with possible data variation.
- To increase the expressiveness, the NF includes neural networks (NNs) estimating *upper triangular transformation matrices*, rather than diagonal ones as in the NF-IVA.



③ NF-FASTMNMF: NORMALIZING FLOW × FAST MULTICHANNEL NONNEGATIVE MATRIX FACTORIZATION

FastMNMF:

$$\mathbf{x}_{nft} \sim \mathcal{N}_{\mathbb{C}}^{M}\Bigg(\mathbf{0},\ \overbrace{\left(\textstyle\sum_{k=1}^{K} u_{ncf} v_{nct}\right)}^{\text{PSD (via NMF) } \lambda_{nft}}\ \overbrace{\mathbf{Q}_f^{-1}\mathrm{Diag}(\tilde{\mathbf{g}}_n)\mathbf{Q}_f^{-\mathsf{H}}}^{\text{SCM } \mathbf{G}_{nf}}\Bigg)$$

$$\mathbf{y}_{ft} \triangleq \mathbf{Q}_f\mathbf{x}_{ft} \sim \mathcal{N}_{\mathbb{C}}^{M}\left(\mathbf{0}, \textstyle\sum_{n=1}^{N}\lambda_{nft}\mathrm{Diag}(\tilde{\mathbf{g}}_n)\right)$$

$$\iff \{\mathbf{y}_{ft}\}_m \sim \mathcal{N}_{\mathbb{C}}\left(0, \sigma_{mft}^2 \triangleq \textstyle\sum_{n=1}^{N}\lambda_{nft}\{\tilde{\mathbf{g}}_n\}_m\right)$$

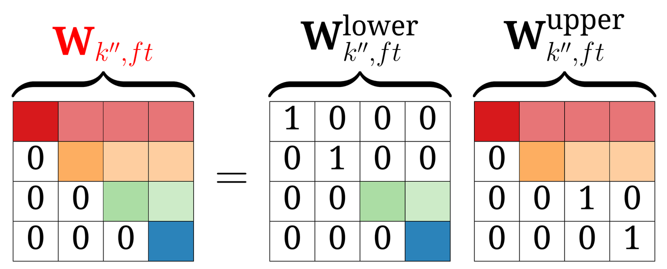
NF-FastMNMF:

$$\mathbf{y}_{ft} = \overbrace{\mathbf{W}_{K,f}\ \underbrace{\mathbf{W}_{K-1,ft}\mathbf{W}_{K-2,f}}_{\text{the } L\text{-th flow block}}\ \cdots\ \underbrace{\mathbf{W}_{2,ft}\mathbf{W}_{1,f}}_{\text{the first flow block}}}^{\text{the mixture decorrelation by an NF}}\ \mathbf{x}_{ft}$$

Separation by Wiener filtering for (NF-)FastMNMF:

$$\widehat{\mathbf{x}}_{nft} \triangleq \mathbb{E}[\mathbf{x}_{nft}|\mathbf{x}_{ft}] = \mathbf{Q}_f^{-1}\mathrm{Diag}\left(\frac{\lambda_{nft}\tilde{\mathbf{g}}_n}{\sum_{n'=1}^{N}\lambda_{n'ft}\tilde{\mathbf{g}}_{n'}}\right)\mathbf{Q}_f\mathbf{x}_{ft}$$

One flow block includes 2 steps: *affine coupling* and *projection*.



$$\mathbf{W}_{k'',ft}^{\text{lower}} \leftarrow \text{NN}\ \boldsymbol{\Omega}_{k'',f}^{\text{lower}},\ \mathbf{W}_{k'',ft}^{\text{upper}} \leftarrow \text{NN}\ \boldsymbol{\Omega}_{k'',f}^{\text{upper}},$$

$$\mathbf{W}_{k',f} \in \mathbb{C}^{M\times M},\ k'' \in \mathbb{K}^{\text{even}},\ k' \in \mathbb{K}^{\text{odd}}$$

In addition to orthogonalizing $\mathbf{W}_{k',f}$ during the forward pass, the log-likelihood function is computed with a volume-preserving (VP) constraint [1]:

$$\ln p(\mathbf{X}) = -\sum_{m,f,t=1}^{M,F,T}\left(\frac{|\{\mathbf{y}_{ft}\}_m|^2}{\sigma_{mft}^2} + \ln\sigma_{mft}^2\right) + T\sum_{k'\in\mathbb{K}^{\text{odd}}}\sum_{f=1}^{F}\ln|\mathbf{W}_{k',f}|^2$$

$$+\sum_{k''\in\mathbb{K}^{\text{even}}}\sum_{m,f,t=1}^{M,F,T}\ln\left|\{\mathbf{W}_{k'',ft}\}_{mm}\right|^2 + \sum_{k''\in\mathbb{K}^{\text{even}}}\sum_{f,t=1}^{F,T}\underbrace{\left(\sum_{m=1}^{M}\ln|\{\mathbf{W}_{k'',ft}\}_{mm}|\right)^2}_{\text{VP-related regularization term}}$$

Parameters to be optimized:

$$\Psi \triangleq \{\mathbf{W}_{k',f}, \boldsymbol{\Omega}_{k'',f}^{\text{upper}}, \boldsymbol{\Omega}_{k'',f}^{\text{lower}}, u_{ncf}, v_{nct}, \tilde{\mathbf{g}}_n | \forall k', \forall k'', \forall n, \forall f, \forall t, \forall c\}$$

Parameter initialization:

- Init. $\mathbf{W}_{k',f}, \boldsymbol{\Omega}_{k'',f}^{\text{upper}}, \boldsymbol{\Omega}_{k'',f}^{\text{lower}}$ such that the NF has the identity transform
- Init. u_{ncf}, v_{nct} randomly, and $\tilde{\mathbf{g}}_n$ with the circular initialization [2]

Parameter updates:

- $\mathbf{W}_{k',f}, \boldsymbol{\Omega}_{k'',f}^{\text{upper}}, \boldsymbol{\Omega}_{k'',f}^{\text{lower}}$: gradient descent with backprop. by Adam
  - for the first 512 epochs, these parameters are optimized as those in NF-IVA [1] for warming-up purpose
- $u_{ncf}, v_{nct}, \tilde{\mathbf{g}}_n$: multiplicative update rules in [2]

④ EVALUATION

Test cases: underdetermined, determined, overdetermined

- Sources: 3 speech signals + 1 environmental noise signal
  - Speech from WSJ0 dataset, while noise from DEMAND dataset (living room, office room, cafetaria)
  - Ratio of speech mixture and noise ∈ {6, 12, 18} dB
  - Room size: 6 x 6 x 3 m – Reverberation: RT60 $\sim \mathcal{U}[0.2s, 0.6s]$
- Number of mics: 3, 4, 7 microphones (of 7-microphone array)
- Total number of mixtures: 270
- Sampling rate: 16 kHz – STFT: 1024-point w/ 75% overlap
- Source conditions: *stationary* and *non-stationary*
  - *Non-stationary* condition: the speakers move at 2 random time instances to simulate the movement when someone shifts the body weight sideways.

Table 1. The median performance scores of the different separation methods on the *stationary* and *non-stationary* datasets. $\mathbf{W}_{k'',ft}$ is either a diagonal (`diag`) or an upper triangular (`triu`) matrix. A higher value is better for all performance metrics. Boldface numbers show the top performances taking into account the 95% confidence interval over the best performances that are indicated by the star symbol ⋆.

Method	$\mathbf{W}_{k'',ft}$	Blocks (L)	3 mics *(underdetermined case)* SDR	SIR	SAR	PESQ	STOI	4 mics *(determined case)* SDR	SIR	SAR	PESQ	STOI	7 mics *(overdetermined case)* SDR	SIR	SAR	PESQ	STOI
Stationary dataset																	
IVA-BP	n/a	0	n/a	n/a	n/a	n/a	n/a	5.7	7.8	**15.2**	1.50	0.81	7.0	10.7	⋆**17.5**	1.80	0.87
NF-IVA	`diag`	1	n/a	n/a	n/a	n/a	n/a	5.8	7.6	⋆**15.6**	1.52	**0.83**	6.9	10.5	16.7	1.73	0.88
NF-IVA	`diag`	2	n/a	n/a	n/a	n/a	n/a	5.9	7.7	**15.4**	1.58	**0.84**	6.9	10.6	16.3	1.71	0.88
NF-IVA	`triu`	1	n/a	n/a	n/a	n/a	n/a	5.9	7.8	**15.4**	1.57	**0.83**	7.1	10.8	16.9	1.74	0.88
NF-IVA	`triu`	2	n/a	n/a	n/a	n/a	n/a	5.8	7.7	**15.3**	1.56	**0.83**	7.2	11.2	**17.1**	1.82	**0.89**
FastMNMF-BP	n/a	0	4.7	9.0	**9.2**	1.34	**0.75**	6.6	**9.8**	13.2	1.57	0.80	7.0	11.2	15.7	1.86	0.82
NF-FastMNMF	`diag`	1	4.2	7.5	8.6	1.36	0.70	6.8	**10.0**	13.2	1.57	⋆**0.85**	**8.5**	11.8	16.1	1.79	**0.90**
NF-FastMNMF	`diag`	2	4.6	9.2	8.6	1.38	0.71	**7.3**	**10.3**	13.5	**1.68**	**0.84**	8.3	12.0	16.2	1.85	**0.90**
NF-FastMNMF	`triu`	1	⋆**5.6**	⋆**10.3**	⋆**9.3**	**1.44**	**0.76**	⋆**7.5**	**10.3**	13.6	⋆**1.70**	**0.84**	**8.7**	**12.6**	16.2	1.81	**0.90**
NF-FastMNMF	`triu`	2	**5.3**	**10.1**	**9.1**	⋆**1.46**	⋆**0.76**	**6.9**	⋆**10.5**	13.2	**1.65**	**0.84**	⋆**9.2**	⋆**13.2**	16.3	⋆**2.07**	⋆**0.91**
Non-stationary dataset																	
IVA-BP	n/a	0	n/a	n/a	n/a	n/a	n/a	5.5	7.2	⋆**14.2**	1.46	0.79	6.1	9.7	⋆**15.7**	1.69	0.84
NF-IVA	`diag`	1	n/a	n/a	n/a	n/a	n/a	4.9	6.7	13.6	1.46	0.80	5.8	9.6	14.8	1.59	0.84
NF-IVA	`diag`	2	n/a	n/a	n/a	n/a	n/a	5.4	7.1	**13.8**	1.49	0.80	6.0	9.7	14.7	1.62	0.85
NF-IVA	`triu`	1	n/a	n/a	n/a	n/a	n/a	5.3	7.2	**13.8**	1.49	0.79	5.7	9.4	14.8	1.64	0.84
NF-IVA	`triu`	2	n/a	n/a	n/a	n/a	n/a	5.3	7.1	**14.0**	**1.50**	0.80	6.2	10.0	15.0	1.69	0.84
FastMNMF-BP	n/a	0	**4.6**	8.7	**8.4**	**1.33**	0.72	6.0	9.6	11.2	1.45	0.78	6.3	10.2	13.2	1.67	0.82
NF-FastMNMF	`diag`	1	4.0	7.9	7.7	1.31	0.71	**6.2**	9.3	11.3	**1.50**	**0.83**	**7.3**	**10.9**	14.6	1.71	⋆**0.88**
NF-FastMNMF	`diag`	2	4.3	8.8	7.7	**1.32**	0.71	⋆**6.7**	⋆**10.4**	11.8	⋆**1.55**	⋆**0.83**	**7.6**	⋆**11.7**	14.9	**1.77**	**0.86**
NF-FastMNMF	`triu`	1	**4.6**	8.7	⋆**8.5**	**1.34**	**0.72**	**6.5**	**10.1**	11.7	**1.55**	**0.82**	**7.1**	**11.1**	14.3	1.68	0.85
NF-FastMNMF	`triu`	2	⋆**5.0**	⋆**9.9**	**8.3**	⋆**1.35**	⋆**0.75**	5.7	8.9	10.9	**1.54**	0.81	⋆**7.8**	**11.4**	14.1	⋆**1.84**	0.86

⑤ CONCLUSION

- Mixture decorrelation in FastMNMF as an NF optimization.
- NF can now be applied to non-determined separation thanks to the joint diagonalization technique from FastMNMF.
- Performance: NF-FastMNMF > FastMNMF-BP > NF-IVA ≈ IVA-BP.
- The upper triangular affine coupling construction improves the separation performance.
- Audio samples @ https://aanugraha.github.io/demo/nffastmnmf/.

⑥ REFERENCES

[1] A. A. Nugraha, K. Sekiguchi, M. Fontaine, Y. Bando, and K. Yoshii, "Flow-based independent vector analysis for blind source separation," IEEE SPL, vol. 27, pp. 2173–2177, 2020.

[2] K. Sekiguchi, Y. Bando, A. A. Nugraha, K. Yoshii, and T. Kawahara, "Fast multichannel nonnegative matrix factorization with directivity-aware jointly-diagonalizable spatial covariance matrices for blind source separation," IEEE/ACM TASLP, vol. 28, pp. 2610–2625, 2020.