CASCADE: Channel-Aware Structured Cosparse Audio DEclipper

Clément Gaultier, Nancy Bertin, Rémi Gribonval

Univ Rennes, Inria, CNRS, IRISA, France

Problem definition

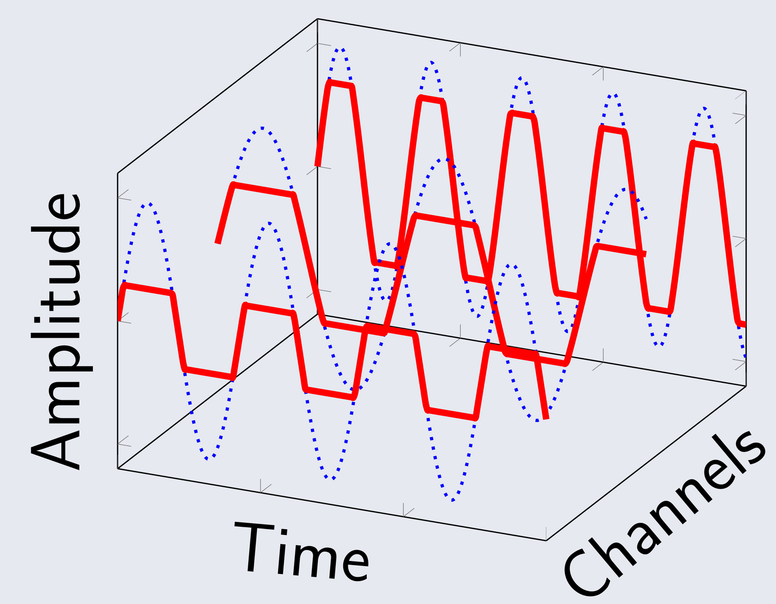
The problem:

Estimate a multichannel audio signal $\tilde{\mathbf{x}}$ from its saturated version $\tilde{\mathbf{y}}$.

- $\mathbf{Y} \in \mathbb{R}^{\mathsf{J} \times \mathsf{K}}$ an overlapping frame (J samples, K channels) of $\tilde{\mathbf{y}}$;
- $\mathbf{y}_{\mathsf{jk}}$ (resp. $\mathbf{x}_{\mathsf{jk}}$) the j^{th} sample recorded on the k^{th} channel from $\mathbf{Y}$ (resp. $\mathbf{X}$);
- τ_{k} the hard-clipping level in the k^{th} channel.

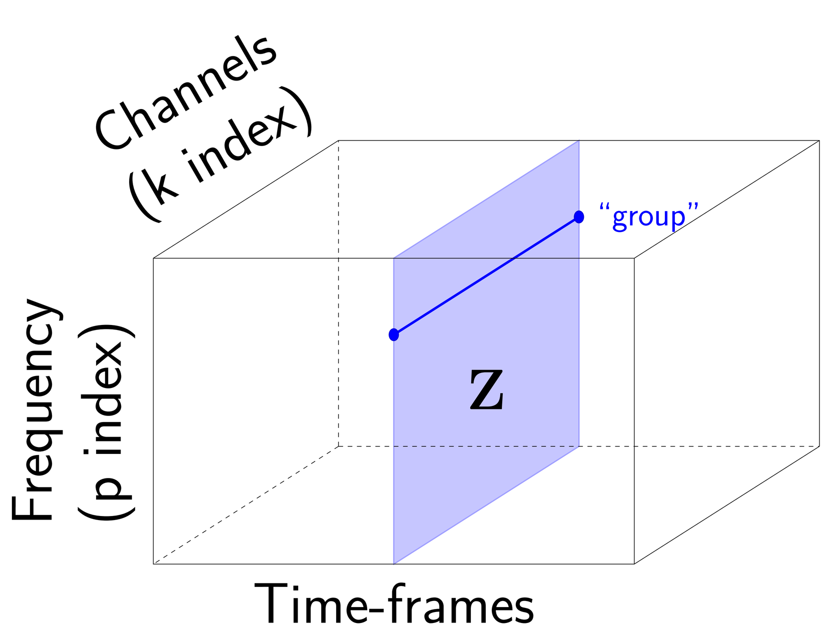
Clipping model:

$$\mathbf{y}_{\mathsf{jk}} = \begin{cases} \mathbf{x}_{\mathsf{jk}} & \text{for } |\mathbf{x}_{\mathsf{jk}}| \leq \tau_{\mathsf{k}}; \\ \operatorname{sgn}(\mathbf{x}_{\mathsf{jk}})\tau_{\mathsf{k}} & \text{otherwise.} \end{cases}$$



Channel-aware structured cosparse modeling

- $\mathbf{Z} \simeq \mathbf{A}\mathbf{X}$ a frequency representation of $\mathbf{X}$;
- $\mathbf{Z} \in \mathbb{C}^{\mathsf{P} \times \mathsf{K}}$ (P frequency bins);
- $\mathbf{A} \in \mathbb{C}^{\mathsf{P} \times \mathsf{J}}$;
- $\mathbf{Z}$ is sparse;
- $\mathbf{Z}$ is "structured across channels".



Structured sparse prior on $\mathbf{Z}$ and cosparse prior on $\mathbf{X}$ [1].

Method

How to get a declipped frame estimate $\hat{\mathbf{X}}$?
Design an iterative algorithm based on ADMM [2] that alternatively projects $\hat{\mathbf{X}}$ on:

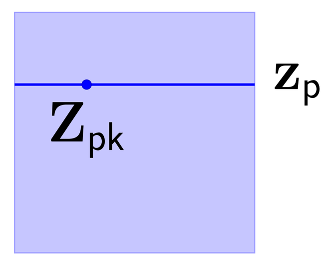
1. The modeling constraint thanks to a specific sparsifying operator;
2. The declipping constraint thanks to a data-fidelity projection.

Tools

Sparsifying operator

Group-Empirical Wiener [3]:

$$\mathcal{S}_{\mu}(\mathbf{Z})_{\mathsf{pk}} = \mathbf{Z}_{\mathsf{pk}} \cdot \left(1 - \frac{\mu^2}{\|\mathbf{z}_{\mathsf{p}}\|_2^2}\right)_+$$



Data-fidelity projection

Optimization problem:

$$\underset{\mathbf{X} \in \Theta}{\text{minimize}} \ \|\mathbf{A}\mathbf{X} - \mathbf{Z}\|_{\mathsf{F}}^2 \qquad (1)$$

$$\Theta = \left\{ \mathbf{X} \mid \begin{array}{l} \mathbf{X}_{\Omega_r} = \mathbf{Y}_{\Omega_r}; \\ \mathbf{X}_{\Omega_+} \succcurlyeq \mathbf{Y}_{\Omega_+}; \\ \mathbf{X}_{\Omega_-} \preccurlyeq \mathbf{Y}_{\Omega_-}. \end{array} \right\}$$

- Ω_r set of reliable indices;
- Ω_+, (resp. Ω_-) set of positively (negatively) clipped indices;
- $\succcurlyeq, \preccurlyeq$ component-wise comparisons.

Closed form solution for (1):

$$\hat{\mathbf{X}}_{\mathsf{jk}} = \begin{cases} \mathbf{Y}_{\mathsf{jk}} & \text{if } \mathsf{jk} \in \Omega_r; \\ (\mathbf{A}^{\mathsf{H}}\mathbf{Z})_{\mathsf{jk}} & \text{if } \begin{cases} \mathsf{jk} \in \Omega_+, (\mathbf{A}^{\mathsf{H}}\mathbf{Z})_{\mathsf{jk}} \geq \tau_{\mathsf{k}}; \\ \text{or} \\ \mathsf{jk} \in \Omega_-, (\mathbf{A}^{\mathsf{H}}\mathbf{Z})_{\mathsf{jk}} \leq -\tau_{\mathsf{k}}; \end{cases} \\ \operatorname{sgn}(\mathbf{Y}_{\mathsf{jk}})\tau_{\mathsf{k}} & \text{otherwise.} \end{cases}$$

CASCADE Algorithm

Sparsification step:

$$\mathbf{Z}^{(i)} = \mathcal{S}_{\mu^{(i-1)}}(\mathbf{A}\hat{\mathbf{X}}^{(i-1)} + \mathbf{U}^{(i-1)})$$

Projection step on the declipping constraint:

$$\hat{\mathbf{X}}^{(i)} = \underset{\mathbf{X}}{\operatorname{argmin}} \|\mathbf{A}\mathbf{X} - \mathbf{Z}^{(i)} + \mathbf{U}^{(i-1)}\|_{\mathsf{F}}^2 \text{ subject to } \mathbf{X} \in \Theta$$

Update step:

$$\mu^{(i)} = \alpha\mu^{(i-1)}, \ \mathbf{U}^{(i)} = \mathbf{U}^{(i-1)} + \mathbf{A}\hat{\mathbf{X}}^{(i)} - \mathbf{Z}^{(i)}$$

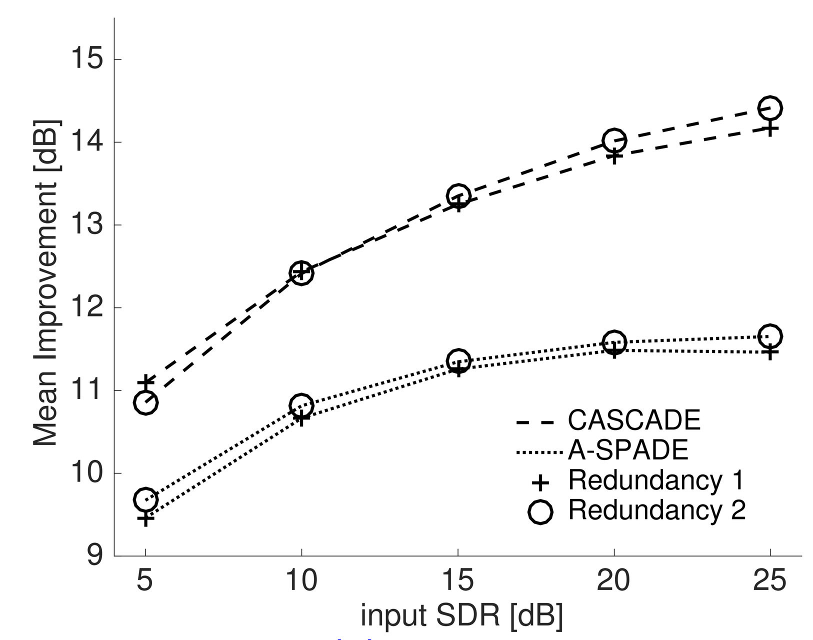
Conclusion

The joint use of cosparse and structured sparsity models is particularly efficient on music and speech multichannel data. CASCADE numerically outperforms state-of-the-art simple cosparse A-SPADE algorithm [1] by 1 dB to more than 3 dB while retaining very limited runtime overcost.

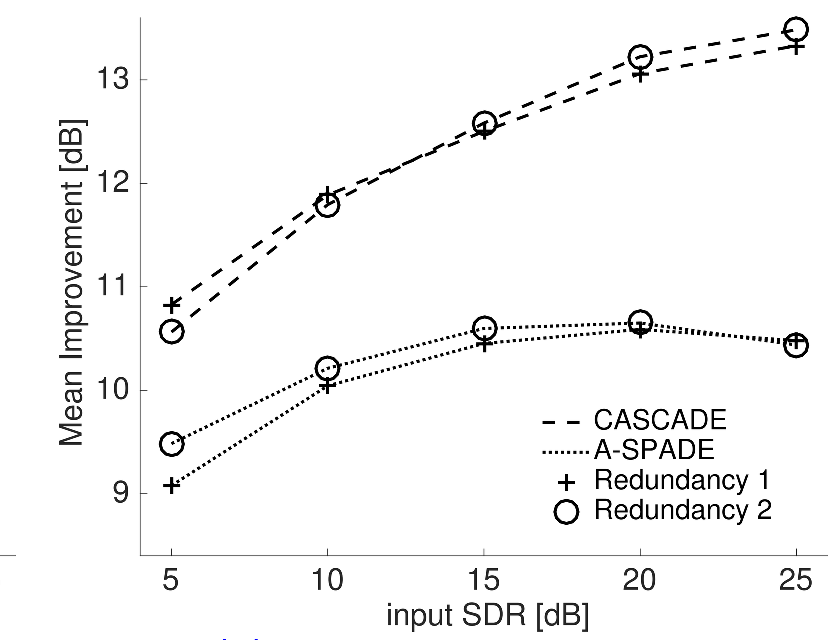
Experiments

8-channels recordings excerpts from the VoiceHome2 Corpus [4].
477 different examples artificially saturated at 5 SDR.

SDR improvements



(a) Speech



(b) Mixed Speech and Music

Runtime

	Algorithm	CASCADE		A-SPADE	
	Redundancy	R=1	R=2	R=1	R=2
Input SDR	5	167	398	73	190
	10	120	265	59	148
	15	80	177	42	103
	20	54	119	29	72
	25	37	78	20	50

(a) Runtime (ratio to realtime processing)

	Algorithm	CASCADE		A-SPADE	
	Redundancy	R=1	R=2	R=1	R=2
Input SDR	5	11.11	10.76	9.31	9.63
	10	12.39	12.45	10.57	10.79
	15	13.31	13.39	11.20	11.37
	20	14.01	14.32	11.67	11.79
	25	14.40	14.44	11.73	11.68

(b) Corresponding improvements (ΔSDR)

References:

[1] S. Kitić, N. Bertin, and R. Gribonval, "Sparsity and cosparsity for audio declipping: a flexible non-convex approach," in *Latent Variable Analysis and Signal Separation (LVA/ICA)*, pp. 243–250, Liberec, Czech Republic: Springer, 2015.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[3] C. Févotte and M. Kowalski, "Hybrid sparse and low-rank time-frequency signal decomposition," in *23rd European Signal Processing Conference (EUSIPCO)*, pp. 464–468, IEEE, 2015.

[4] N. Bertin, E. Camberlein, R. Lebarbenchon, E. Vincent, S. Sivasankaran, I. Illina, and F. Bimbot, "VoiceHome-2, an extended corpus for multichannel speech processing in real homes," *Speech Communication*, 2018.

clement.gaultier@inria.fr, nancy.bertin@irisa.fr, remi.gribonval@inria.fr