Low-rankness of complex-valued spectrogram and its application to phase-aware audio processing

WASEDA University 早稲田大学

Yoshiki Masuyama*, Kohei Yatabe, Yasuhiro Oikawa (Waseda University, Japan) *Email: mas-03151102@akane.waseda.jp

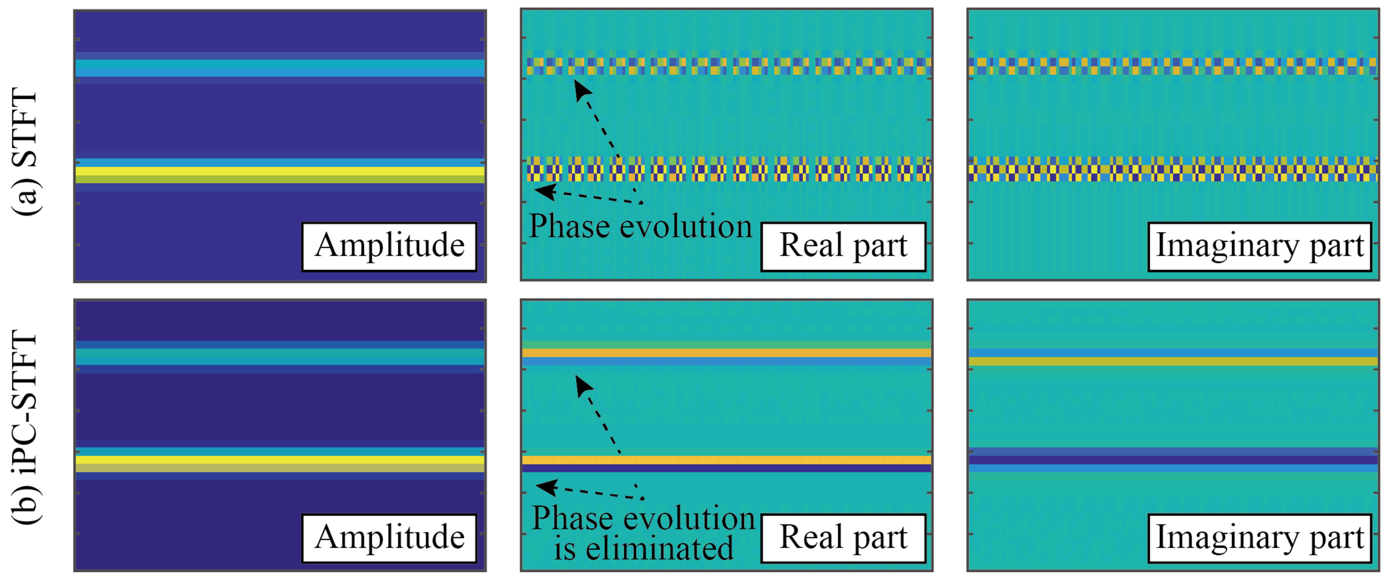
Introduction

- Low-rankness of amplitude spectrograms has been effectively used in audio signal processing methods including NMF.
- **Problem**: Such methods suffer from the problem of phase reconstruction because they focused on the amplitude treatment.
- **Proposal 1**: **A complex-valued spectrogram** is modeled by **a low-rank complex-valued matrix** by modifying its phase.
- **Proposal 2**: **A prior emphasizing harmonic signals** is presented based on the proposed low-rank modeling.
- **Results**: In audio denoising, an optimization-based method using the proposed prior outperformed NMF and its variants.

Property of complex-valued spectrogram

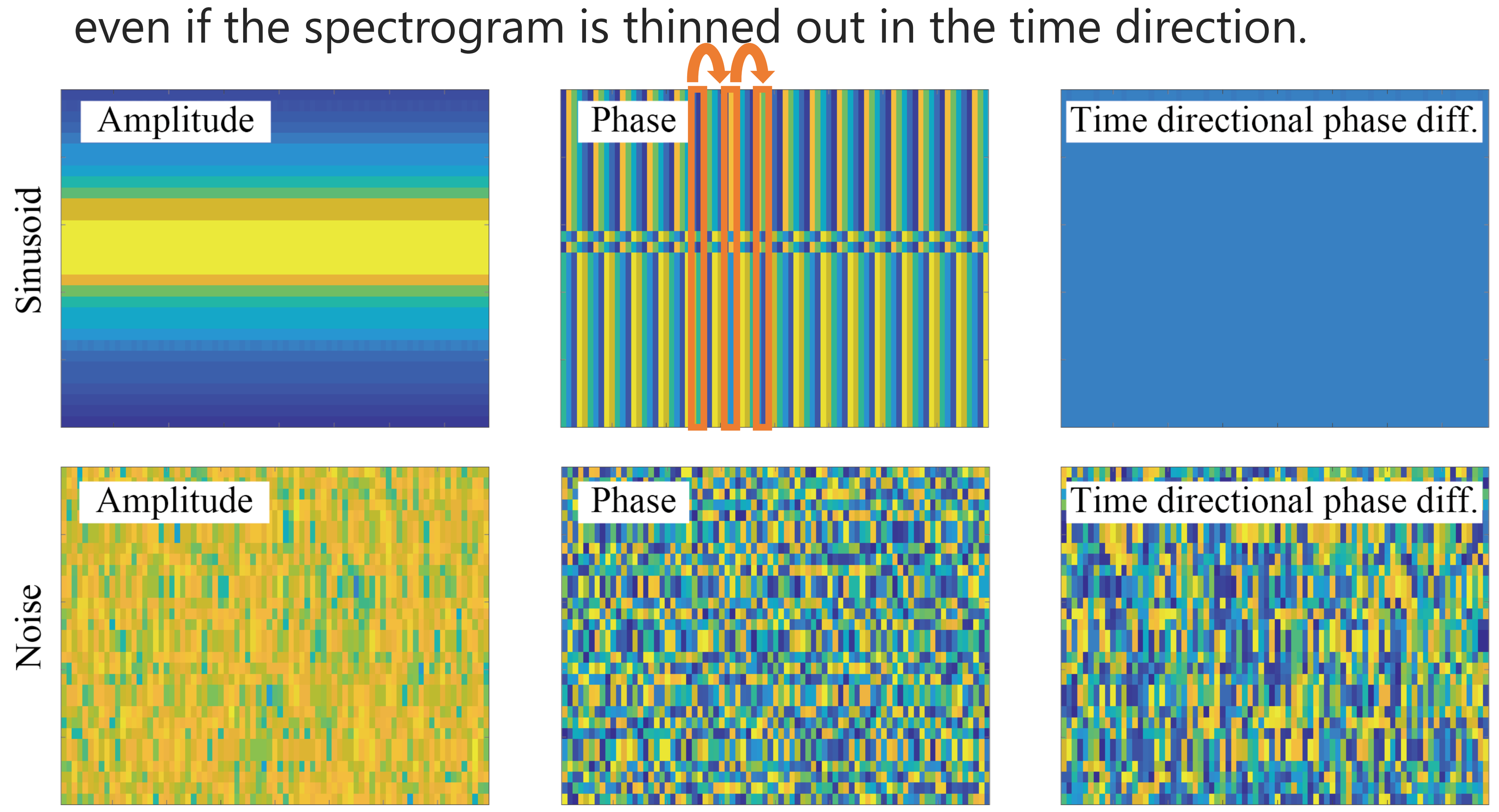
Rank of complex-valued spectrograms

- Previous work proved that the complex-valued spectrogram of a sum of r complex exponentials becomes rank-r in some conditions.
- A complex-valued spectrogram of a sum of sinusoids is not low-rank when the number of sinusoids increases.
- **Direct low-rank modeling of a complex-valued spectrogram can not be considered for audio signals**.



Property of time-difference of phase

- The time directional phase difference of a complex exponential is constant, and thus **the phase in a future frame is predictable** even if the spectrogram is thinned out in the time direction.



Proposed low-rank modeling method of complex-valued spectrograms

Instantaneous phase corrected STFT (iPC-STFT)

- If the phase evolution of each sinusoid is eliminated, the complex-valued spectrograms of a sum of sinusoids can be low-rank.
- Assuming each sinusoid is sufficiently separated, its phase spectrogram has the following relation.

$$\phi[\xi,\tau+1] = \phi[\xi,\tau] + 2\pi a v[\xi,\tau]/L$$

- To cancel the phase evolution, we use **the instantaneous phase corrected STFT (iPC-STFT)**.

$$\mathcal{G}^{\mathbf{w}}_{\mathrm{iPC}}(\mathbf{x})[\xi,\tau] = \textstyle\prod_{\eta=0}^{\tau-1} e^{-2\pi j a v[\xi,\eta]/L}\mathcal{G}^{\mathbf{w}}(\mathbf{x})[\xi,\tau]$$

- The instantaneous frequency can be estimated from the observed signal.

$$v[\xi,\tau] = b\xi - \mathrm{Im}[\mathcal{G}^{\mathbf{w}'}(\mathbf{x})[\xi,\tau]/\mathcal{G}^{\mathbf{w}}(\mathbf{x})[\xi,\tau]]$$

- Once the instantaneous frequency is estimated, iPC-STFT is defined as an invertible linear transform
- iPC-STFT also eliminated the phase evolution of a locally stational signal whose phase spectrogram satisfies the above relation.

Low-rankness of iPC-STFT spectrogram

- Based on the property of iPC-STFT, the following relation is obtained for **the iPC-STFT spectrogram of a sum of sinusoids**, and thus it **becomes rank-1**.

$$\begin{aligned}\mathcal{G}^{\mathbf{w}}_{\mathrm{iPC}}(\mathbf{s})[\xi,\tau+1] &= \prod_{\eta=0}^{\tau} e^{-2\pi j v[\xi,\eta]a/L}\mathcal{G}^{\mathbf{w}}(\mathbf{s})[\xi,\tau+1]\\ &= \prod_{\eta=0}^{\tau-1} e^{-2\pi j v[\xi,\eta]a/L}\mathcal{G}^{\mathbf{w}}(\mathbf{s})[\xi,\tau]\\ &= \mathcal{G}^{\mathbf{w}}(\mathbf{s})[\xi,0] \quad (= \mathcal{G}^{\mathbf{w}}_{\mathrm{iPC}}(\mathbf{s})[\xi,0])\end{aligned}$$

Harmonic signal prior based on proposed low-rank modeling

- For applying the proposed low-rank modeling to audio signal processing, **we propose a prior of harmonic signals, named iPC low-rankness (iPCLR)** with the instantaneous frequency estimated in advance.

$$\mathcal{P}_{\mathrm{iPCLR}}(\mathbf{x}) = \|\mathcal{G}^{\mathbf{w}}_{\mathrm{iPC}}(\mathbf{x})\|_*$$

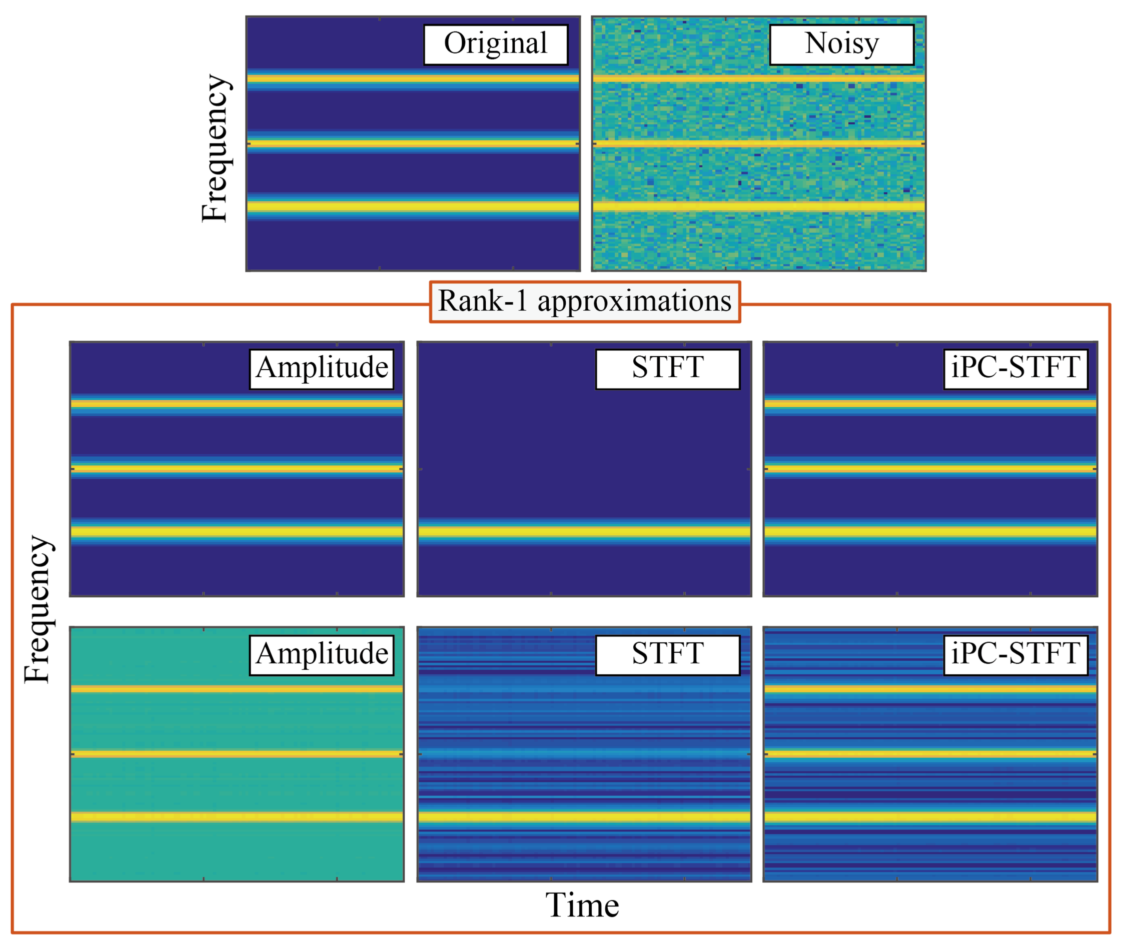
- Using the proposed iPCLR, the following convex optimization problem is considered for audio denoising.

$$\mathbf{x}^\star = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{x}-\mathbf{d}\|_2^2 + \lambda\,\mathcal{P}_{\mathrm{iPCLR}}(\mathbf{x})$$

Toy example

Rank-1 approximations of clean/noisy sinusoids

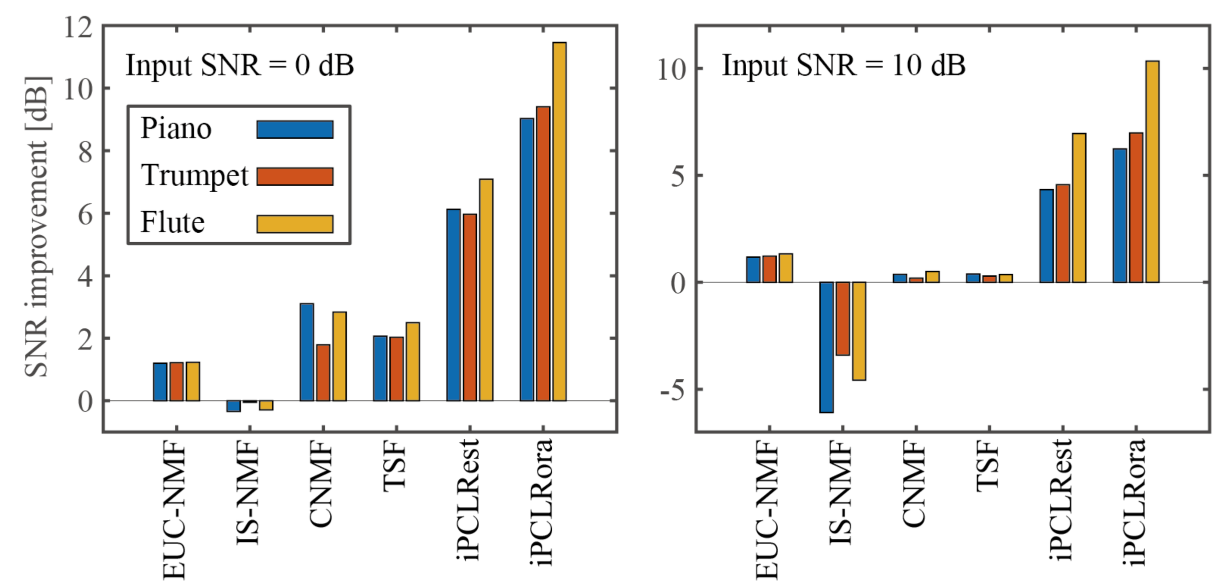
- In both cases, **the rank-1 approximation of iPC-STFT spectrogram represented all sinusoids**, while that of STFT represented only one sinusoid.
- In the noisy case, **the rank-1 approximation of iPC-STFT spectrogram effectively reduced the noise**, while that of amplitude did not.



Application to audio denoising

- The proposed method was compared with NMF and its phase-aware extensions using three melodies played by different musical instruments.
- **iPCLRest** (using instantaneous frequency estimated from the noisy signal) **outperformed the conventional methods**, and iPCLRora (using instantaneous frequency estimated from the clean signal) achieved the highest SNR.

Experimental condition	
Window	Hann
Window length	93 ms
Shift length	23 ms
# of bases	30



Conclusion

- We showed that the rank of a complex-valued spectrogram can be as low as its amplitude by applying iPC under mild assumptions.
- Based on the low-rankness, we further proposed a prior of harmonic signals.
- Seeking further applications of iPCLR remains as future works.