Phase recovery based on signal modeling: application to audio source separation

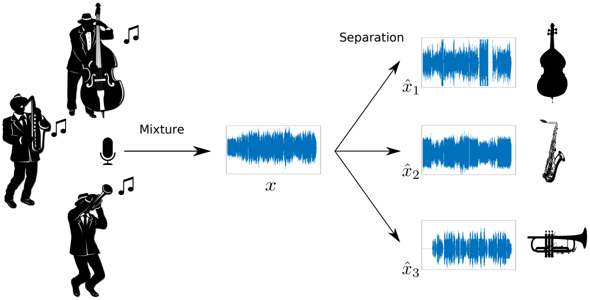
Paul Magron

Ph.D. defense

Mr. Yannis Stylianou	Examiner
Mr. Philippe Depalle	Reviewer
Mr. Laurent Girin	Reviewer
Mr. Jonathan Le Roux	Examiner
Mr. Roland Badeau	Ph.D. Supervisor
Mr. Bertrand David	Ph.D. Supervisor

02/12/2016

Source separation



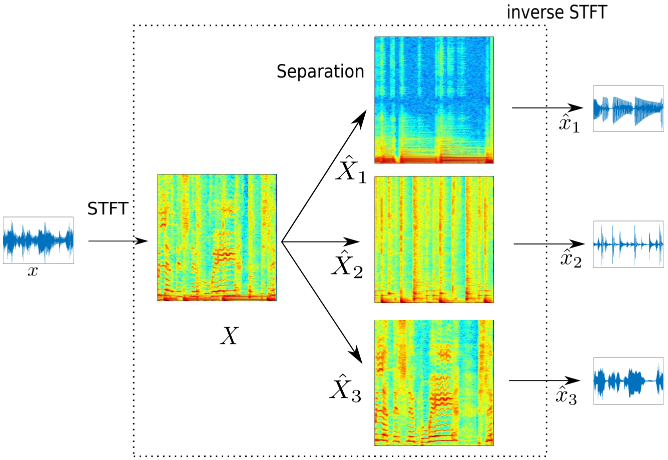
- Applications: karaoke, automatic transcription, denoising...

  $\rightarrow$

- Challenges: Reduction of **interference** and **artifacts**.

Short-Term Fourier Transform (STFT)

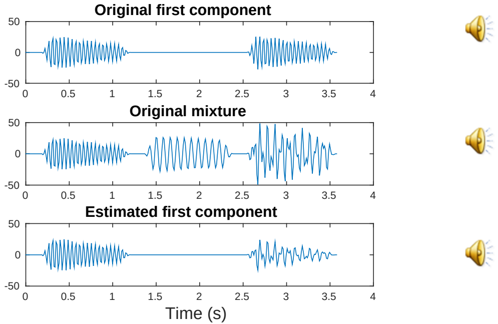
Exploit the particular structure of music signals.



Time-Frequency (TF) overlap

Source estimation:
Soft masking of the mixture's STFT: $\hat{X}_k = G_k \odot X$.

- ⊖ Issues when sources **overlap** in the TF domain:



- ⊖ $\hat{X}_k \neq$ STFT of a $\hat{x}_k$.

TELECOM ParisTech

Problem setting

Mixture model: $x(n) = \sum_k x_k(n)$.
STFT: $X(f,t) = \sum_{n=0}^{N_w-1} x(n+tS) w_a(n) e^{-2i\pi \frac{f}{F} n}$.

- Redundancy $\rightarrow$ an invertible transform;
- $X_k \in \mathbb{C}^{F \times T} \rightarrow X_k(f,t) = \underbrace{V_k(f,t)}_{\text{Magnitude}} e^{i \overbrace{\phi_k(f,t)}^{\text{Phase}}}$

Problem setting

Mixture model: $x(n) = \sum_k x_k(n)$.
STFT: $X(f,t) = \sum_{n=0}^{N_w-1} x(n+tS) w_a(n) e^{-2i\pi \frac{f}{F} n}$.

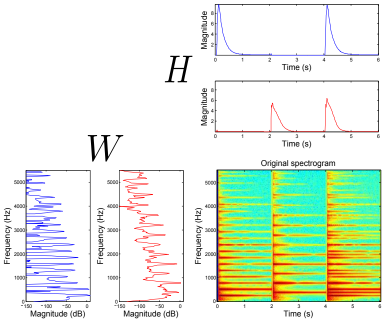
- Redundancy $\rightarrow$ an invertible transform;
- $X_k \in \mathbb{C}^{F \times T} \rightarrow X_k(f,t) = \underbrace{V_k(f,t)}_{\text{Magnitude}} e^{\overbrace{i\phi_k(f,t)}^{\text{Phase}}}$

Goal: compute an estimate $\hat{X}_k$ of X_k.

- Magnitude estimation;
- Phase reconstruction is necessary for time-domain synthesis;
- Joint estimation of amplitude and phase.

Nonnegative matrix factorization (NMF)

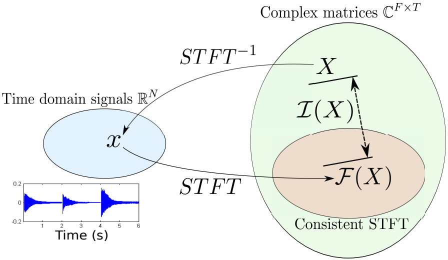
Model: $V \approx \hat{V} = WH$, where V, W and H are nonnegative.



- Estimation: minimization of $D(V, WH)$;
- Extensions: constraints (sparsity, harmonicity...), side-information (music score)...

Phase reconstruction

Wiener filtering: $\hat{X}_k = \frac{\hat{V}_k^{\odot 2}}{\sum_l \hat{V}_l^{\odot 2}} \odot X \rightarrow \phi$-source $= \phi$-mixture.



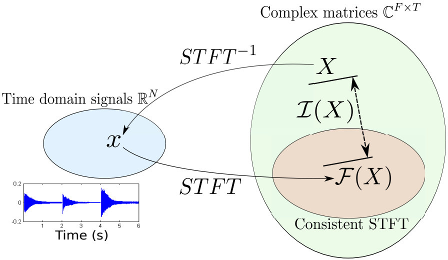
Inconsistency: $\mathcal{I}(X) = ||X - \mathcal{F}(X)||_F^2$, $\mathcal{F} = STFT \circ STFT^{-1}$.

[Griffin, 1984] Iteratively applying $\mathcal{F}$;

[Le Roux, 2008] Direct minimization of $\mathcal{I}$.

Phase reconstruction

Wiener filtering: $\hat{X}_k = \frac{\hat{V}_k^{\odot 2}}{\sum_l \hat{V}_l^{\odot 2}} \odot X \rightarrow \phi$-source $= \phi$-mixture.



Inconsistency: $\mathcal{I}(X) = ||X - \mathcal{F}(X)||_F^2$, $\mathcal{F} = STFT \circ STFT^{-1}$.

[Griffin, 1984] Iteratively applying $\mathcal{F}$;

[Le Roux, 2008] Direct minimization of $\mathcal{I}$.

Extensions

- Combine mixture phase/consistency constraint;
- Consistent Wiener filtering [Le Roux, 2013].

NMF with phase estimation

Complex NMF (CNMF) [Kameoka, 2009]

$$\hat{X}(f,t) = \sum_{k=1}^{K} \hat{X}_k = \sum_{k=1}^{K} \underbrace{W(f,k)H(k,t)}_{\text{NMF model}} e^{i\phi_k(f,t)}.$$

- Estimation by minimization of the Euclidean distance between X and $\hat{X}$ (+ sparsity).
- $\oplus$ Joint estimation of magnitude and phase.
- Needs to be constrained, e.g. consistency [Le Roux, 2009].

NMF with phase estimation

High Resolution NMF (HRNMF) [Badeau, 2014]
Modeling each frequency band by means of AR filtering:

$$\hat{X}_k(f,t) = b_k(f,t) + \sum_{p=1}^{P(k,f)} a_p(k,f)\hat{X}_k(f,t-p),$$

$b_k(f,t) \sim \mathcal{N}(0, \sigma_k(f,t)^2)$ where $\sigma_k(f,t)^2 = W(f,k)H(k,t)$

- The complex STFT components are directly estimated.
- $\oplus$ Naturally captures phase dependencies over time.

How "well" do those methods perform?

Performance measurement with BSS Eval [Vincent, 2006]:

- Signal to Distortion/Interference/Artifacts Ratios (SDR, SIR, SAR).

Comparison of NMF-based source separation techniques:

- It is mandatory to design novel phase recovery techniques;
- Consistency $\neq$ separation quality;
- HRNMF is promising $\rightarrow$ signal modeling.

How can we incorporate model-based phase information in a mixture model for audio source separation?

P. Magron, R. Badeau and B. David (2015).
Phase reconstruction in NMF for audio source separation: an insightful benchmark.
In *Proc. of IEEE ICASSP*.

Outline

1. Phase recovery by sinusoidal modeling
2. Onset phase reconstruction
3. Complex NMF under phase constraints
4. Probabilistic source models

Outline

1. Phase recovery by sinusoidal modeling

2. Onset phase reconstruction

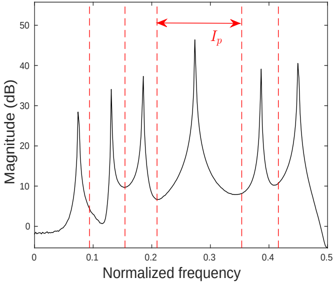
3. Complex NMF under phase constraints

4. Probabilistic source models

Sinusoidal model

A signal is modeled as a $\sum$ of sinusoids [McAuley, 1986]:

$$x(n) = \sum_p A_p e^{2i\pi\nu_p n + i\phi_{0,p}}.$$



- STFT's phase of the p-th partial:

$$\phi_p(f,t) = \phi_p(f,t-1) + 2\pi S\nu_p.$$

- In the p-th *region of influence*

$$\phi(f,t) = \angle X(f,t) = \phi_p(f,t).$$

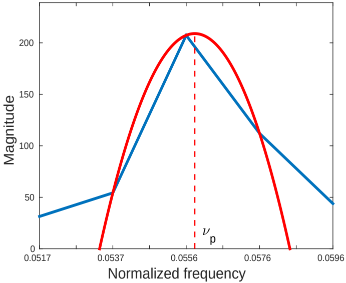
- **Phase unwrapping** (PU) relation:

$$\phi(f,t) = \phi(f,t-1) + 2\pi S\nu(f).$$

Frequency estimation

Most techniques use:

- the STFT's phase (e.g. phase vocoder [Laroche, 1999]);
- a harmonic model (e.g. Harmonic Spectral Product/Sum...).



$\rightarrow$ *Quadratic Interpolated FFT* (QIFFT).

- Each peak $\approx$ a parabola;
- Max. of the parabola $\rightarrow \nu_p$.

- Estimation within each time frame $\rightarrow$ slowly-varying sinusoids.
- A recursive relationship $\rightarrow$ initialization.

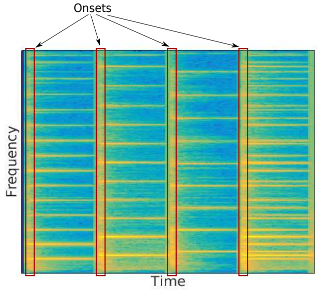
Phase recovery procedure

Tempogram Toolbox [Grosche, 2011]:

- Onset frames detection.

Initialize PU within onset frames:

- Assumed known (Oracle);
- Mixture phase (source separation).



- In frame t:
  1. Frequency estimation ν_p by QIFFT near each magnitude peak;
  2. Decomposition into regions of influence: $\forall f \in I_p$, $\nu(f,t) = \nu_p$;
  3. Phase unwrapping: $\phi(f,t) = \phi(f,t-1) + 2\pi S \nu(f,t)$.
- Proceed to next frame.

Comparison with Griffin Lim

- Onset phases are known;
- Magnitudes: known (Oracle) or NMF (semi-Oracle).

SDR results (in dB):

	Oracle		semi-Oracle	
	GL	PU	GL	PU
Piano	0.4	**5.8**	−0.2	**4.7**
Guitar	−0.5	**2.2**	−11.2	**−9.7**
Strings	−6.5	**0.4**	−8.9	**−4.7**
Speech	**1.1**	−1.8	−11.8	**−11.6**

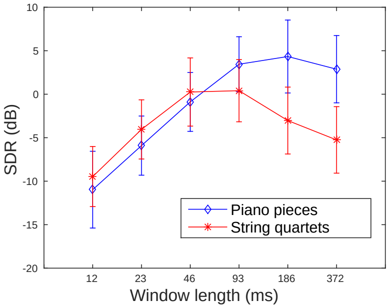
- Phase unwrapping (PU) > Griffin Lim (GL);
- Limits of the SDR.

Influence of the window length

Two main artifacts:

- Musical noise (short windows)
- Reverberation (long windows)

Need to find a trade-off!



Artifacts → cumulative error over time frames.

Applications:

- not many frames to recover (click removal);
- additional phase information: **source separation**.

Source separation - Problem setting

- Mixture model: $X = \sum_k X_k$ with known magnitudes.
- Goal: estimate $\hat{X}_k$.

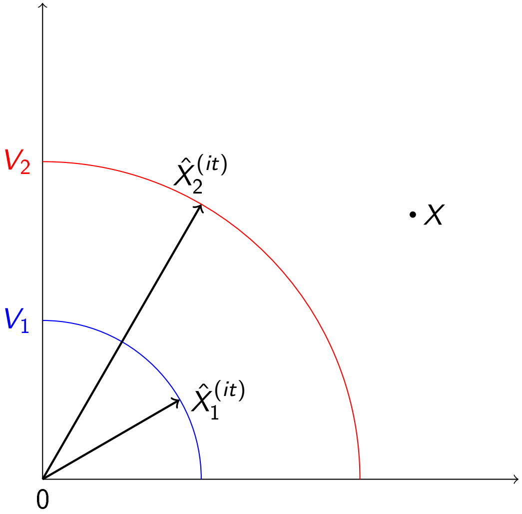
Problem:

$$\text{minimize } ||X - \sum_k \hat{X}_k||_F^2 \text{ s.t. } |\hat{X}_k| = V_k.$$

Proposed approach:

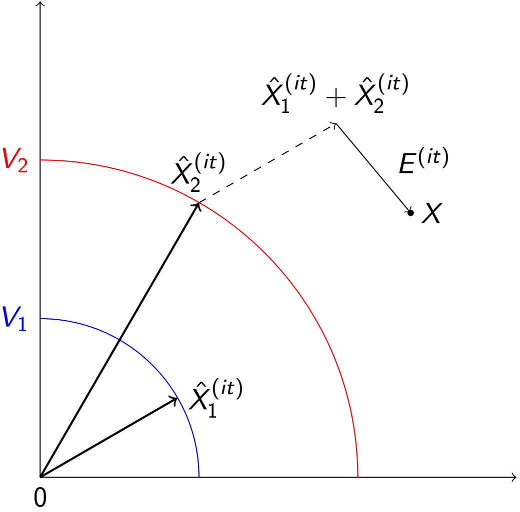
- Iterative procedure;
- Phase information **through the initialization**.

Source separation procedure



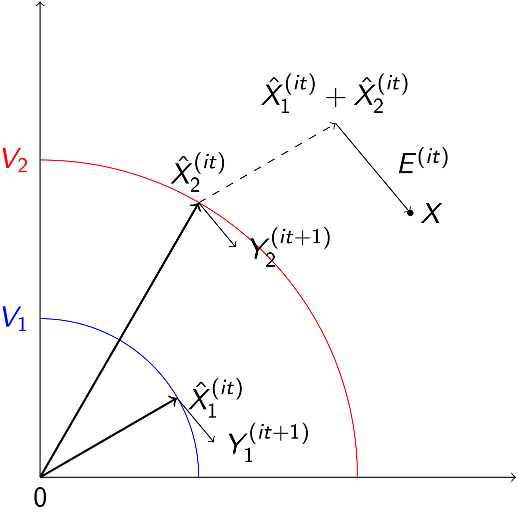
1. Initialize $\hat{X}_k$;

Source separation procedure



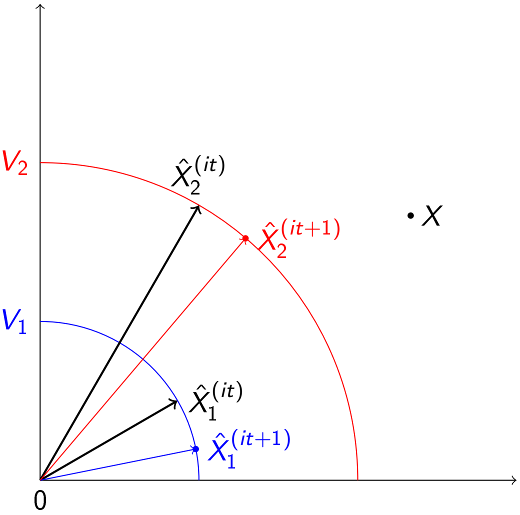
1. Initialize $\hat{X}_k$;
2. $E = X - \sum_k \hat{X}_k$;

Source separation procedure



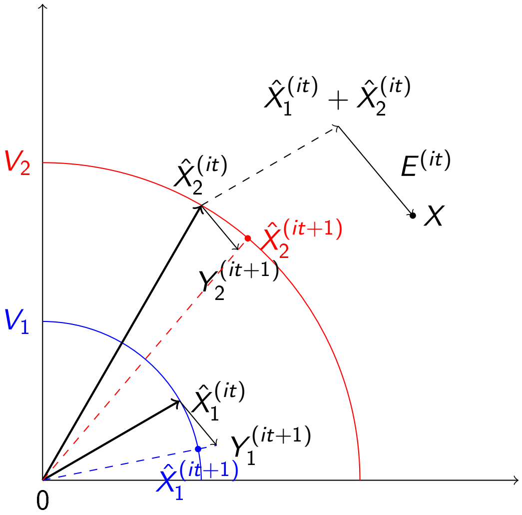
1. Initialize $\hat{X}_k$;
2. $E = X - \sum_k \hat{X}_k$;
3. $Y_k \leftarrow \hat{X}_k + \lambda_k E$;

Source separation procedure



1. Initialize $\hat{X}_k$;
2. $E = X - \sum_k \hat{X}_k$;
3. $Y_k \leftarrow \hat{X}_k + \lambda_k E$;
4. $\hat{X}_k \leftarrow \frac{Y_k}{|Y_k|} V_k$;

Source separation procedure

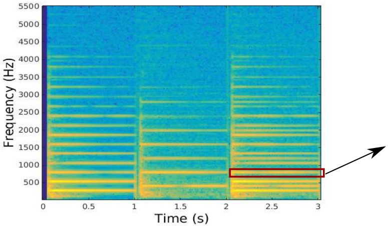


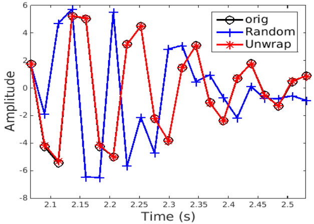
1. Initialize $\hat{X}_k$;
2. $E = X - \sum_k \hat{X}_k$;
3. $Y_k \leftarrow \hat{X}_k + \lambda_k E$;
4. $\hat{X}_k \leftarrow \frac{Y_k}{|Y_k|} V_k$;
5. Return to step 2.

Influence of the initialization

→ **Initialization with the PU technique.**

Mixtures of piano notes with TF overlap:





- +3.5 dB in SDR/SAR, +7.5 dB in SIR over a random initialization.

Source separation results

DSD100 database:

- 50 development songs + 50 test songs;
- 4 sources: `bass`, `drums`, `vocals` and `other`.

Magnitudes are known or estimated by NMF.

Method	SDR	SIR	SAR
Wiener	9.1	16.4	10.4
Consistent Wiener	**11.1**	19.7	**12.0**
Proposed	11.0	**22.3**	11.3

Example: mix bass

- Proposed procedure > Consistent Wiener filtering

Significant reduction of computational cost ($\approx \times 7$).

Phase recovery by sinusoidal modeling - Conclusion

Exploiting phase information based on sinusoidal modeling improves the source separation quality over a phase-unaware approach.

P. Magron, R. Badeau and B. David (2015).
Phase reconstruction of spectrograms with linear unwrapping: application to audio signal restoration.
In *Proc. of EUSIPCO*.

P. Magron, R. Badeau and B. David (2017).
STFT phase recovery by sinusoidal modeling for audio source separation.
submitted to the *IEEE Transactions on Audio, Speech and Language Processing*.

Outline

1. Phase recovery by sinusoidal modeling

2. Onset phase reconstruction

3. Complex NMF under phase constraints

4. Probabilistic source models

Onset phase

Why are onset phases important?

Onset phase

Why are onset phases important?

- Perceptive quality of the sound;
- Initialize the PU recursive relationship.

Onset phase

Why are onset phases important?

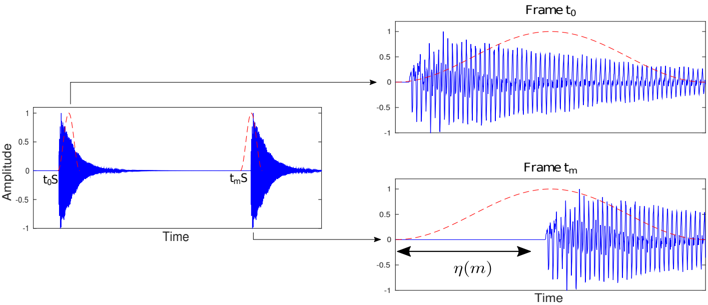
- Perceptive quality of the sound;
- Initialize the PU recursive relationship.

Approach:

- Model the signal within onset frames (e.g. impulse);
- Exploit the **repetition** of audio events $\rightarrow$ onset phase constraints.

Model of repeated audio events

Two onset signals are equal up to a gain factor and a delay:



$$X(f, t_m) \approx X(f, t_0)\rho e^{i\lambda(m)f}, \text{ with } \lambda(m) = \frac{2\pi\eta(m)}{F}.$$

$$\underbrace{\phi(f, t_m)}_{\text{phase within an onset frame}} \approx \underbrace{\psi(f)}_{\text{reference phase}} + \underbrace{\lambda(m)f}_{\text{offset}}.$$

TELECOM ParisTech

Onset mixture model

Onset matrix: $Y(f,m) = X(f,t_m)$.
Model within onset frames:

$$\tilde{Y}(f,m) = \sum_{k=1}^{K} V_k(f,t_m) e^{i\psi_k(f)} e^{i\lambda_k(m)f}.$$

Goal: estimate $\hat{Y}_k$. Minimization of:

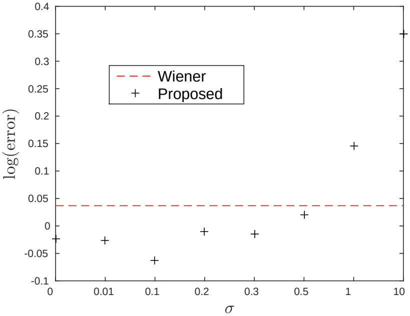
$$\mathcal{C}_r = ||Y - \sum_{k=1}^{K} \hat{Y}_k||_F^2 + \sigma \sum_{k=1}^{K} ||\tilde{Y}_k - \hat{Y}_k||_F^2.$$

- $\partial_\psi \mathcal{C}_r = 0$ (resp. $\partial_\phi \mathcal{C}_r = 0$) $\rightarrow$ update on ψ (resp. ϕ);
- Adaptation of the ESPRIT algorithm $\rightarrow$ update on λ.

Experiments on mixtures of piano notes

Influence of σ:

- Estimation error $\frac{1}{K}\sum_k ||Y_k - \hat{Y}_k||_F$;



- Slight improvement over Wiener filtering.

Experiments on mixtures of piano notes

Source separation:

- Onset phase recovery + Iterative procedure with PU.

Onset phase	SDR	SIR	SAR
Mixture	20.4	27.1	21.5
Proposed model	21.0	27.9	22.1
Oracle	22.6	29.8	23.6

- Proposed onset phase model > mixture phase;
- Some room for further improvement.

Repetition model - Conclusion

A phase constraint based on a model of repeated audio events improves the separation over using the mixture phase.

P. Magron, R. Badeau and B. David (2015).
Phase reconstruction of spectrograms based on a model of repeated audio events.
In *Proc. of IEEE WASPAA*.

Outline

1. Phase recovery by sinusoidal modeling
2. Onset phase reconstruction
3. Complex NMF under phase constraints
4. Probabilistic source models

Complex NMF

Goal: Joint estimation of magnitude and phase.
Complex NMF model:

$$\hat{X}(f,t) = \sum_{k=1}^{K} \hat{X}_k = \sum_{k=1}^{K} \underbrace{W(f,k)H(k,t)}_{\text{NMF model}} e^{i\phi_k(f,t)}.$$

Needs to be constrained (*cf.* benchmark).

- Phase constraints based on **time signal properties**.

Complex NMF - Phase constraints

Phase unwrapping constraint (*cf.* [Bronson, 2014]) :

$$\mathcal{C}_u(\phi) = \sum_{f,k} \sum_{t \neq \text{onsets}} |X(f,t)|^2 |e^{i\phi_k(f,t)+} - e^{i\phi_k(f,t-1)+2i\pi S\nu_k(f)}|^2.$$

Phase repetition constraint within onset frames:

$$\mathcal{C}_r(\phi,\psi,\lambda) = \sum_{f,k} \sum_{t \in \text{onsets}} |X(f,t)|^2 |e^{i\phi_k(f,t)} - e^{i\psi_k(f)+i\lambda_k(t)f}|^2.$$

Complex NMF - Estimation

Complete cost function:

$$\mathcal{C}(\theta) = \underbrace{D(X, \hat{X})}_{\text{NMF}} + \sigma_u \underbrace{\mathcal{C}_u(\phi)}_{\text{Unwrapping}} + \sigma_r \underbrace{\mathcal{C}_r(\phi, \psi, \lambda)}_{\text{Repetition}} + \sigma_s \underbrace{\mathcal{C}_s(H)}_{\text{Sparsity}}$$

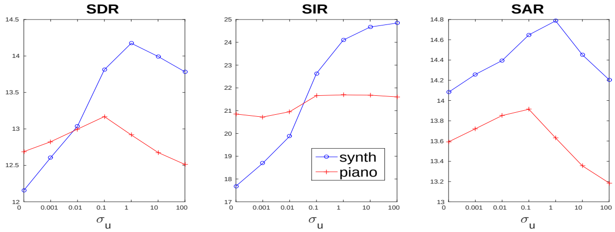
- $\theta = \{W, H, \phi, \psi, \lambda\}$;

Minimization of $\mathcal{C}$:

- Coordinate descent, auxiliary function method.

CNMF - Experiments on simple data

Influence of σ_u



Influence of σ_r

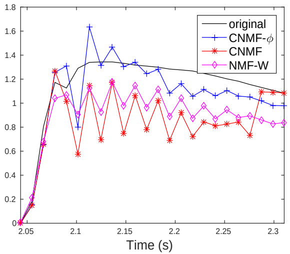
- The repetition constraint does not improve the results.

CNMF - Experiments on simple data

Source separation

Data	Method	SDR	SIR	SAR
Synthetic sinusoids	NMF-W	11.7	16.8	**13.5**
	CNMF	9.6	16.7	10.7
	CNMF-ϕ	**12.3**	**23.2**	12.8
Piano notes	NMF-W	**14.7**	18.5	**17.4**
	CNMF	13.3	19.6	14.9
	CNMF-ϕ	14.6	**22.1**	15.7

- CNMF-ϕ > CNMF.
- Reconstruction of a piano note partial:



CNMF - Experiments on DSD100

Optimal weights (learning database): $(\sigma_r, \sigma_u) \approx (0.1, 0.1)$.

Source separation (test database):

Method	SDR	SIR	SAR
NMF-W	**1.9**	10.2	**3.7**
CNMF	1.4	10.9	2.9
CNMF-ϕ	1.7	**12.2**	2.9

Mix

Voice

- $\sigma_u / \sigma_r \rightarrow$ trade-off between SDR, SIR and SAR.

Complex NMF - Conclusion

A promising approach for separating overlapping sources in the TF domain with improved interference rejection.

P. Magron, R. Badeau and B. David (2016).
Complex NMF under phase constraints based on signal modeling: application to audio source separation.
In *Proc. of IEEE ICASSP*.

Outline

Probabilistic source models

Why is the probabilistic framework useful?

- Model **uncertainty**;
- Incorporate **prior** information;
- **Conservative** estimators (e.g. posterior expectation);
- Novel estimation **techniques**.

Probabilistic source models

Why is the probabilistic framework useful?

- Model **uncertainty**;
- Incorporate **prior** information;
- **Conservative** estimators (e.g. posterior expectation);
- Novel estimation **techniques**.

Gaussian model [Févotte, 2005]: $X = \sum_k X_k$ with $X_k \sim \mathcal{N}(0, \sigma_k^2)$

$$\Leftrightarrow X_k = V_k e^{i\phi_k} \text{ with } V_k \sim \underbrace{\mathcal{R}(\sigma_k)}_{\text{Rayleigh}} \text{ and } \phi_k \sim \underbrace{\mathcal{U}_{[0,2\pi[}}_{\text{Uniform}}.$$

Probabilistic source models

Why is the probabilistic framework useful?

- Model **uncertainty**;
- Incorporate **prior** information;
- **Conservative** estimators (e.g. posterior expectation);
- Novel estimation **techniques**.

Gaussian model [Févotte, 2005]: $X = \sum_k X_k$ with $X_k \sim \mathcal{N}(0, \sigma_k^2)$

$$\Leftrightarrow X_k = V_k e^{i\phi_k} \text{ with } V_k \sim \underbrace{\mathcal{R}(\sigma_k)}_{\text{Rayleigh}} \text{ and } \phi_k \sim \underbrace{\mathcal{U}_{[0,2\pi[}}_{\text{Uniform}}.$$

Proposed approach:

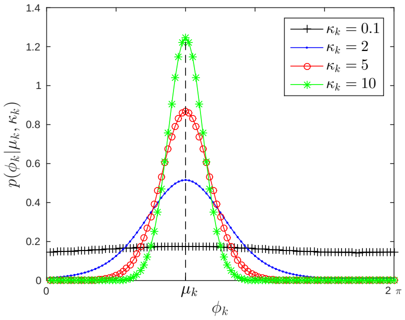
- A **non-uniform** phase model.
- A **robust** magnitude model.

Von Mises phase

Mixture model: $X = \sum_k V_k e^{i\phi_k}$ with constant magnitudes.

- A prior phase μ_k can be obtained.

$\phi_k \sim$ **Von Mises** (VM) with location μ_k.



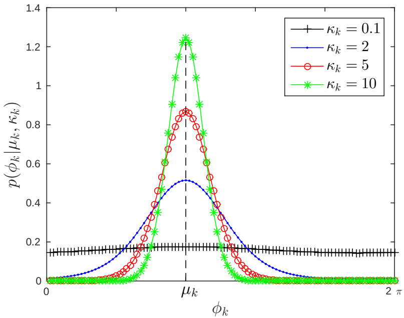
$$p(\phi_k|\mu_k,\kappa_k) = \frac{e^{\kappa_k \cos(\phi_k - \mu_k)}}{2\pi I_0(\kappa_k)}.$$

Von Mises phase

Mixture model: $X = \sum_k V_k e^{i\phi_k}$ with constant magnitudes.

- A prior phase μ_k can be obtained.

$\phi_k \sim$ **Von Mises** (VM) with location μ_k.



$$p(\phi_k|\mu_k,\kappa_k) = \frac{e^{\kappa_k \cos(\phi_k - \mu_k)}}{2\pi I_0(\kappa_k)}.$$

Drawback: a non-tractable model.

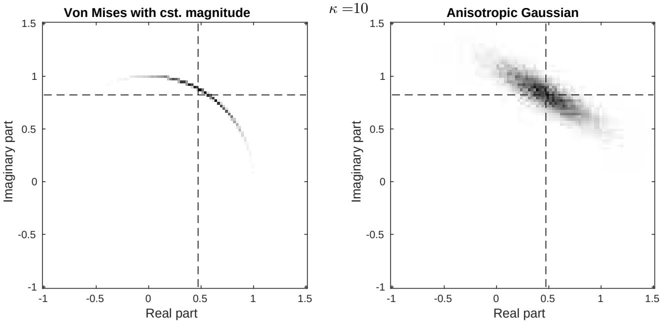
$\rightarrow$ **Approximate the VM model by a Gaussian model which keeps the phase dependencies.**

Anisotropic Gaussian (AG) model

Mixture model: $X = \sum_k X_k$ with complex Gaussian variables:

$$X_k \sim \mathcal{N}(\underbrace{m_k}_{\text{Mean}}, \underbrace{\gamma_k}_{\text{Variance}}, \underbrace{c_k}_{\text{Relation}}), \Gamma_k = \begin{pmatrix} \gamma_k & c_k \\ \bar{c}_k & \gamma_k \end{pmatrix}.$$

Key idea: the moments are the same ones in VM and AG models.



MMSE estimator of the sources

$$\hat{X}_k = \mathbb{E}(X_k|X).$$

For Gaussian mixtures:

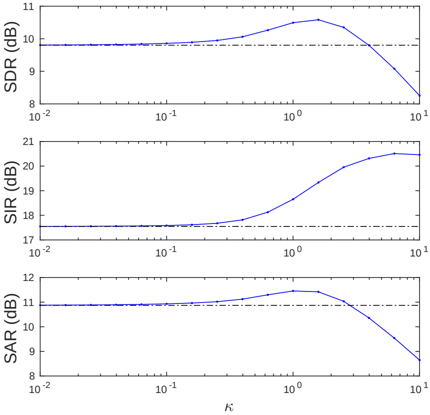
$$\underline{\hat{X}}_k = \underline{m}_k + \Gamma_k \Gamma_X^{-1} (\underline{X} - \underline{m}_X) \text{ where } \underline{u} = \begin{pmatrix} u \\ \bar{u} \end{pmatrix}.$$

- Conservative: $\sum_k \hat{X}_k = X$;
- When $\kappa \to 0$: Wiener filtering $\frac{V_k^2}{\sum_l V_l^2} X$!

$\rightarrow$ Optimal combination of prior and mixture phases.

Experiments on DSD100

Influence of the concentration parameter:



Source separation:

- Wiener < Proposed MMSE < Consistent Wiener;
- Mix 🔈 Bass 🔈 Consistent Wiener 🔈 Proposed 🔈.

TELECOM ParisTech

Iterative procedure vs. AG model

So... which technique should I use?

P. Magron, R. Badeau and B. David (2017).
Phase-dependent anisotropic Gaussian model for audio source separation.
submitted to the *Proc. of IEEE ICASSP*.

Iterative procedure vs. AG model

So... which technique should I use?

Tests on the DSD100 database:

	Oracle			semi-Oracle		
	SDR	SIR	SAR	SDR	SIR	SAR
Iter	**10.0**	**20.5**	**10.4**	5.3	**15.0**	5.9
MMSE	9.0	16.7	9.9	**7.4**	13.9	**8.6**

- Magnitude estimate $\approx$ Oracle $\rightarrow$ iterative procedure;
- Alternatively $\rightarrow$ MMSE estimator from AG model.

P. Magron, R. Badeau and B. David (2017).
Phase-dependent anisotropic Gaussian model for audio source separation.
submitted to the *Proc. of IEEE ICASSP*.

TELECOM ParisTech

Towards a robust magnitude model

Common distributions (Poisson, Rayleigh) are not heavy-tailed.

- *Stable* distributions: additivity and robustness to outliers...
- ... not nonnegative in general.

$\rightarrow$ **A robust nonnegative data model based on the stable distribution family for source separation.**

Positive stable (PαS) distributions

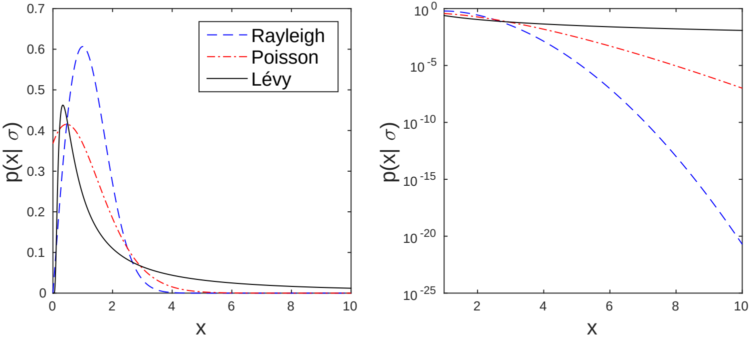
- PαS: subclass of the stable family with a nonnegative support.

Positive stable (PαS) distributions

- PαS: subclass of the stable family with a nonnegative support.
- Probability density function (PDF) cannot be expressed in closed-form...

Positive stable (PαS) distributions

- PαS: subclass of the stable family with a nonnegative support.
- Probability density function (PDF) cannot be expressed in closed-form...
- ... except for the **Lévy** distribution.



TELECOM ParisTech

Positive stable (PαS) distributions

- PαS: subclass of the stable family with a nonnegative support.
- Probability density function (PDF) cannot be expressed in closed-form...
- ... except for the **Lévy** distribution.

Lévy NMF model:

- $X = \sum_k X_k$ where X_k is Lévy-distributed: $X_k \sim \mathcal{L}(\sigma_k)$.

$$\rightarrow X \sim \mathcal{L}(\sigma) \text{ with } \sigma^{\odot 1/2} = \sum_k \sigma_k^{\odot 1/2}.$$

- NMF model: $\sigma^{\odot 1/2} = WH$.

Model estimation

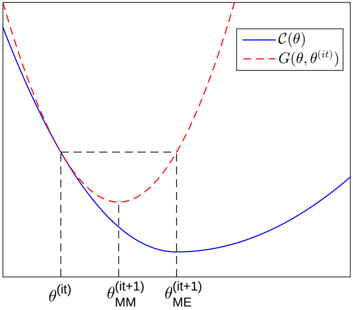
Maximum Likelihood (ML) estimation:

$$L(W, H) \propto -d_{IS}([WH]^{\odot 2}, X),$$

where $d_{IS}(a, b) = \frac{a}{b} - \log \frac{a}{b} - 1$.

Minimization of d_{IS}:

- Naïve approach;
- Majorize-Minimization (MM);
- Majorize-Equalization (ME).



Source separation

- For any PαS distribution:

$$\hat{X}_k = \mathbb{E}(X_k|X) = \frac{{\sigma_k}^{\odot\alpha}}{\sum_l {\sigma_l}^{\odot\alpha}} \odot X.$$

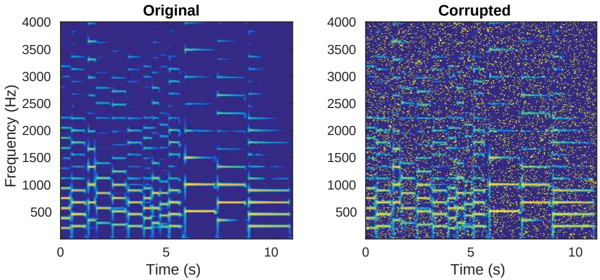
$\rightarrow$ **Generalized Wiener filtering**

- Lévy NMF model:

$$\hat{X}_k = \frac{W_k H_k}{\sum_l W_l H_l} \odot X.$$

Application: music spectrogram inpainting

Guitar spectrograms are corrupted by impulsive TF noise.



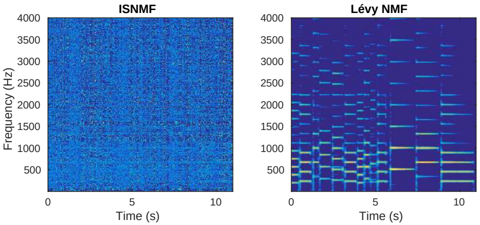
The models are learned directly on the corrupted data:

- ISNMF [Févotte,2009], KLNMF [Lee, 1999], Cauchy NMF [Liutkus, 2015];
- Lévy NMF;
- Robust PCA (RPCA) [Candès, 2011];
- Weighted ISNMF [Limem, 2013];

The noise location is unknown (except for Weighted ISNMF).

Application: music spectrogram inpainting

Method	log(KL)
ISNMF	9.0
KLNMF	6.2
Cauchy NMF	3.4
RPCA	3.6
Lévy NMF	**3.2**
Weighted ISNMF	3.8



- IS/KL NMFs: very poor results;
- Lévy NMF compares favorably with other robust methods.

P. Magron, R. Badeau and A. Liutkus (2017).
Lévy NMF for robust nonnegative source separation.
submitted to the *IEEE Signal Processing Letters*.

Conclusion

Contributions

- Novel phase recovery $\rightarrow$ improved source separation;
- Phase unwrapping: interference rejection (deterministic iterative procedure / probabilistic framework);
- Onsets: accounting for a repetition property for modeling the phase;
- A phase-constrained CNMF framework;
- A robust nonnegative source separation model.

Conclusion

Perspectives

- Sinusoidal model-based constraints in multi-resolution transforms;
- Refined modeling of onsets;
- Multichannel modeling;
- Nonnegative models (inverse-gamma, PαS...);
- A complete probabilistic model: Rayleigh or PαS magnitude + Von Mises phase.

Publications

Journal articles

- P. Magron, R. Badeau and A. Liutkus (2017).
  Lévy NMF for robust nonnegative source separation.
  submitted to the *IEEE Signal Processing Letters*.
- P. Magron, R. Badeau and B. David (2017).
  STFT phase recovery by sinusoidal modeling for audio source separation.
  submitted to the *IEEE Transactions on Audio, Speech and Language Processing*.

Conference proceedings

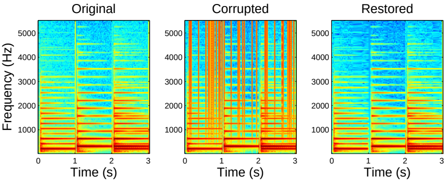
- P. Magron, R. Badeau and B. David (2017).
  Phase-dependent anisotropic Gaussian model for audio source separation.
  submitted to the *Proc. of IEEE ICASSP*.
- P. Magron, R. Badeau and B. David (2016).
  Complex NMF under phase constraints based on signal modeling: application to audio source separation.
  In *Proc. of IEEE ICASSP*.
- P. Magron, R. Badeau and B. David (2015).
  Phase reconstruction of spectrograms based on a model of repeated audio events.
  In *Proc. of IEEE WASPAA*.
- P. Magron, R. Badeau and B. David (2015).
  Phase reconstruction of spectrograms with linear unwrapping: application to audio signal restoration.
  In *Proc. of EUSIPCO*.
- P. Magron, R. Badeau and B. David (2015).
  Phase reconstruction in NMF for audio source separation: an insightful benchmark.
  In *Proc. of IEEE ICASSP*.

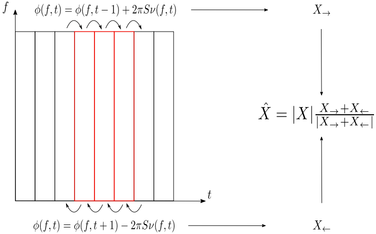
Click removal

Signals are corrupted with clicks in the time-domain.
Restoration in the TF domain:

Phase: forward + backward unwrapping.

Magnitude: linear interpolation.





Orig Corrupted AR HRNMF Proposed

Lévy NMF vs. ISNMF

Updates rules if $K = 1$ and $W(f) = 1 \ \forall f$:

$$H_{\mathrm{IS}}(t) \leftarrow \frac{1}{F}\sum_f X(f,t), \ H_{\mathrm{Lévy}}(t) \leftarrow \sqrt{\frac{F}{\sum_f \frac{1}{X(f,t)}}}.$$

- ISNMF $\rightarrow$ **arithmetic** mean;
- Lévy NMF $\rightarrow$ **harmonic** mean (and $\surd$).

If $X(f,t) = 1$ except for one entry: $X(f_0, t_0) = 10^8$, then:

$$H_{\mathrm{IS}}(t_0) \leftarrow 10^7, \ H_{\mathrm{Lévy}}(t_0) \leftarrow 1.05.$$

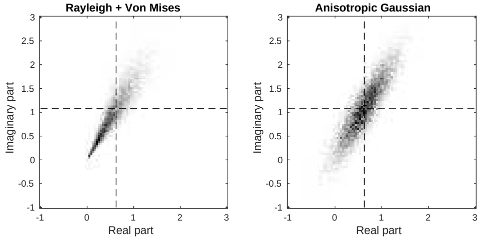
$\rightarrow$ **Lévy NMF is robust to outliers.**

Rayleigh magnitude + Von Mises phase

Proposed model: $X = \sum_k X_k = \sum_k V_k e^{i\phi_k}$ with:

$$V_k(f,t) \sim \underbrace{\mathcal{R}(\sigma_k(f,t))}_{\text{Rayleigh}} \text{ and } \phi_k(f,t) \sim \underbrace{\mathcal{VM}(\mu_k(f,t), \kappa_k(f,t))}_{\text{Von Mises}}.$$

- NMF model: $\sigma_k(f,t)^2 = W(f,k)H(k,t)$;
- Markov chain prior on $\mu_k \rightarrow$ Phase Unwrapping;
- Non-tractability $\rightarrow$ Anisotropic Gaussian approximation;



- SAGE algorithm $\rightarrow$ "Complex ISNMF".