ML-based modelling for LISA

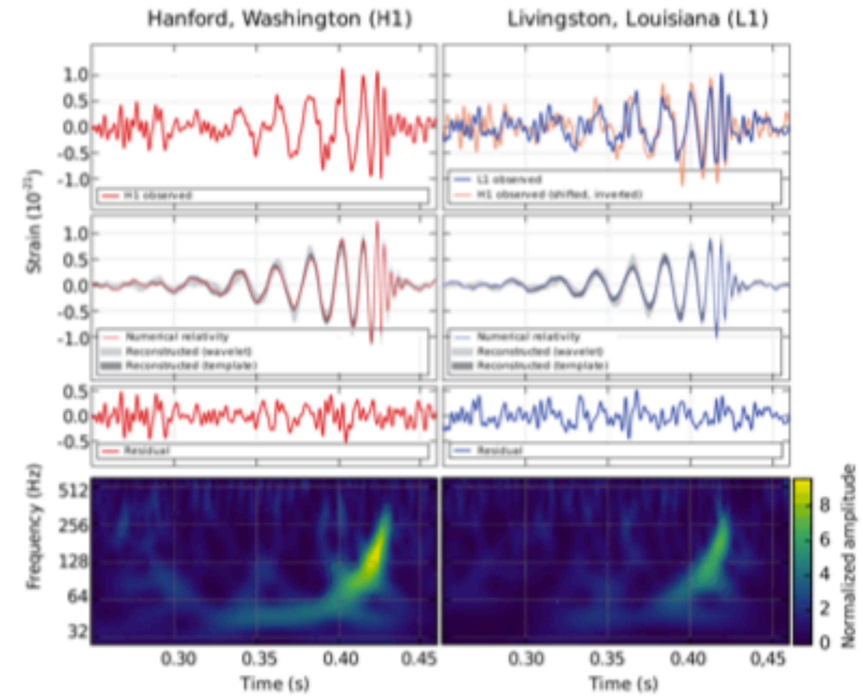
Towards a fast unmixing of the LISA data

Jérôme Bobin, H.Moutarde, A.Blelly, E.Leroy, S.Charpigny

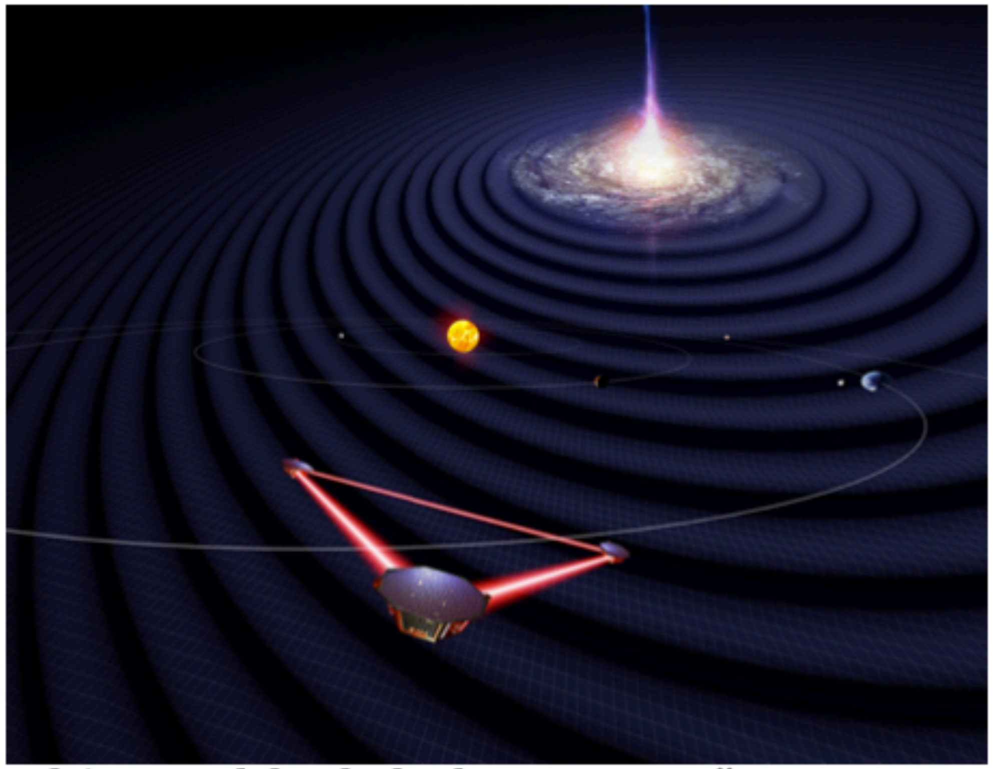
CEA/I, France

Quick context elements

- 2015 : First GW detection by LIGO-VIRGO ground detectors
- 2034 : LISA launch
- First **space interferometer** : 3 satelites in heliocentric orbit forming an interferometer with 2.5M km arm length
- **Data** : 4 years acquired through 3 channels containing millions of potential signals from multiple source types



1

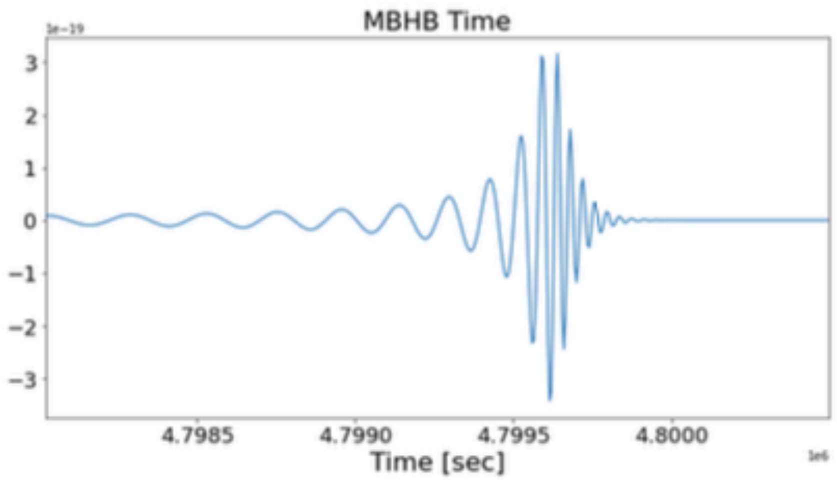


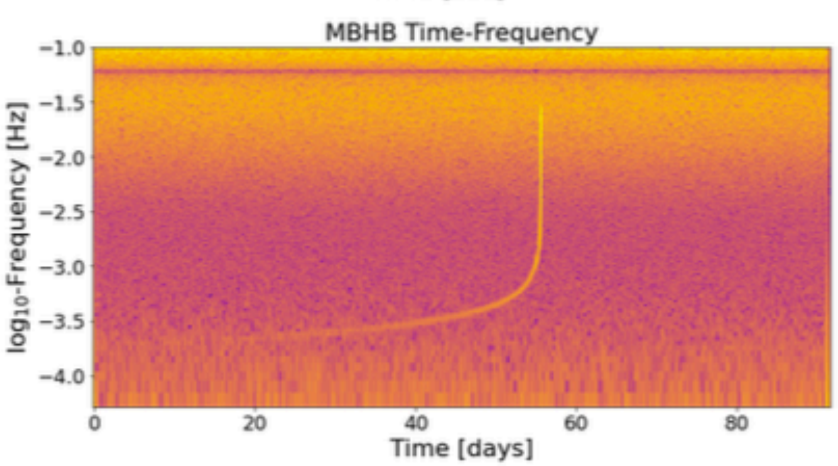
2

1. Abbott et al. "Observation of gravitational waves from a binary black hole merger." Physical review letters 116.6 (2016) : 061102.
2. University of Florida / Simon Barke (CC-BY 4.0)

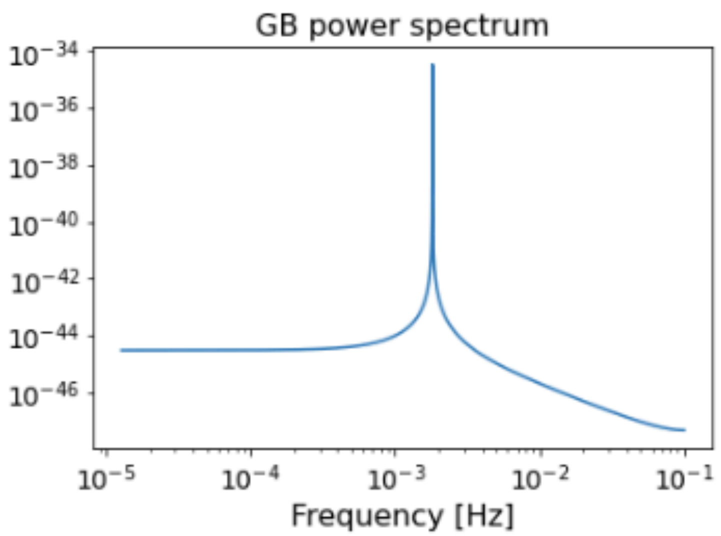
LISA: the complexity of GW event unmixing

Massive Black Hole Binaries (MBHB)
- Loudest source
- 1 every 3 days
- Transient "chirp"





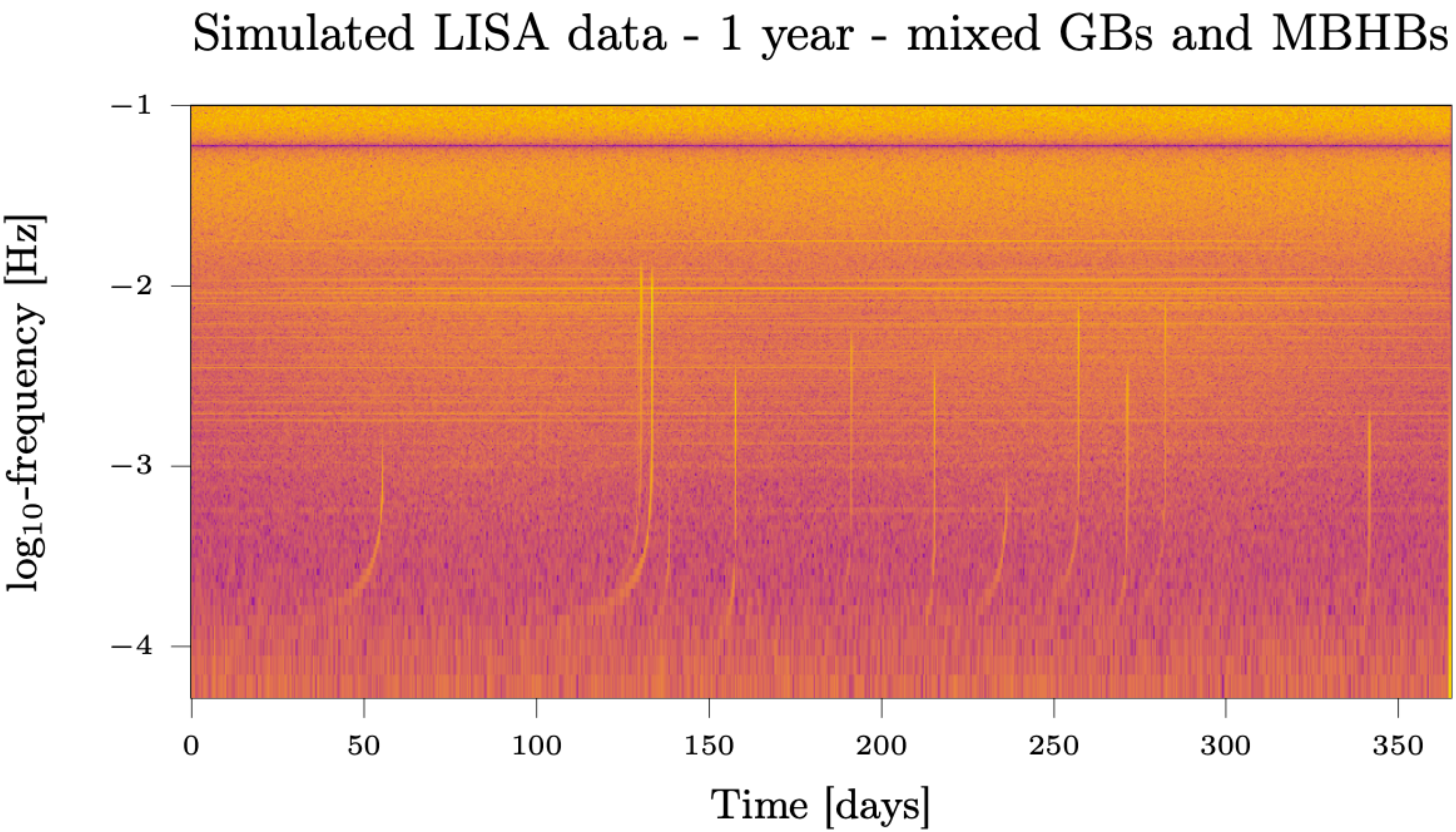
Galactic Binaries (GB)
- Millions of sources
- Thousands of detectable ones
- Stationary, monochromatic



Other sources
- Extreme Mass Ratio Inspirals
- Stochastic background
- Confusion background

Noise
- Instrumental
- Glitches
- Gaps

LISA: the complexity of GW event unmixing



Why non-parametric methods ?

The road towards "global analysis" might require a two-stage approach:

1- A <u>fast</u> and <u>robust</u> unmixing stage allowing to:

- Detect known and *<u>unknown</u>* GW events
- Provide a first *rough* estimate of individual source
- Provide a *rough* estimate of the parameter space region for each individual source
- Robustness w.t. gaps, glitches, source leakages, non-stationary noise, etc.

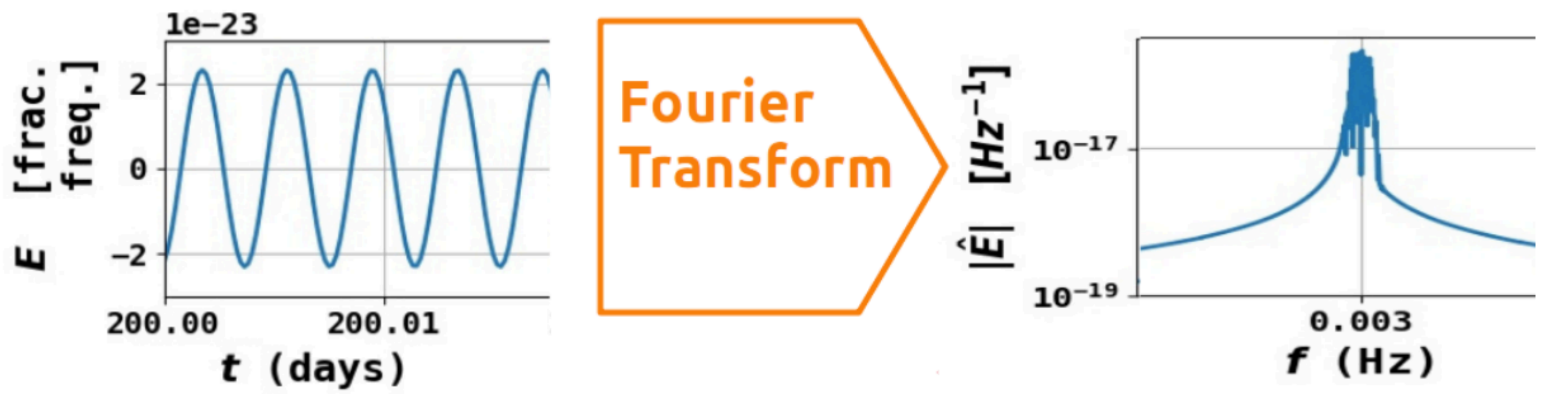
2- A precise refinement stage using more heavy-weight methods (e.g. MCMC)

Non-parametric methods *(that do not make direct use of analytic waveform sims codes)* may be a good candidate:

e.g. sparse modelling in wavelets (already popular in LIGO/VIRGO), Fourier, STFT, etc.

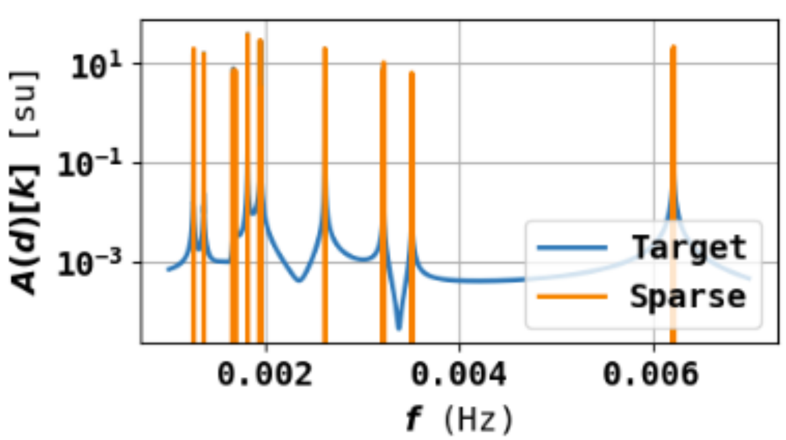
Non-parametric models in action

So far, during Aurore Blelly's PhD, we have investigated the use of sparsity-based methods to detect/estimate GB events:



GBs have quasi-periodic waveforms ⇒ **sparse in the Fourier domain**

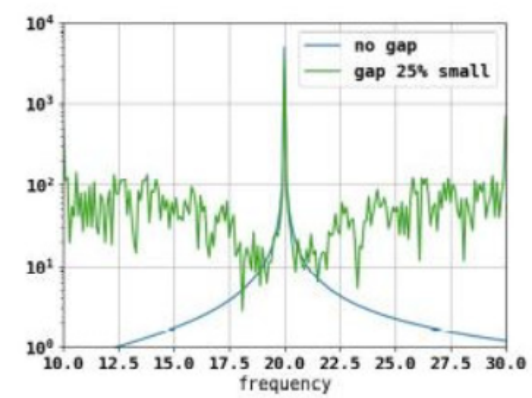
GB detection/estimation



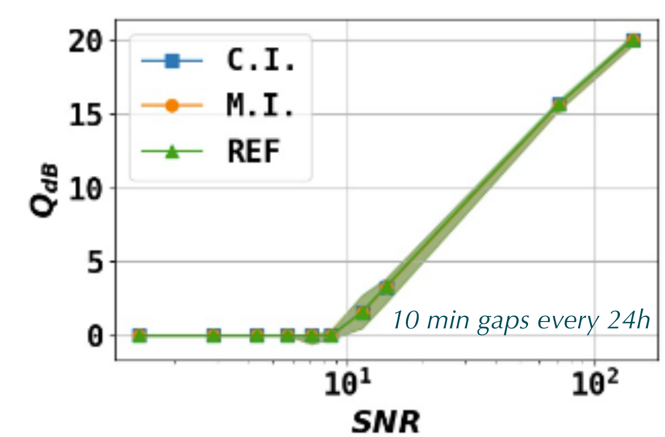
Detection/estimation results from LDC 1-3

Blelly, Moutarde, Bobin, Phys. Rev. D, 2020

Dealing with gapped data



Gaps' effect on a GB waveform



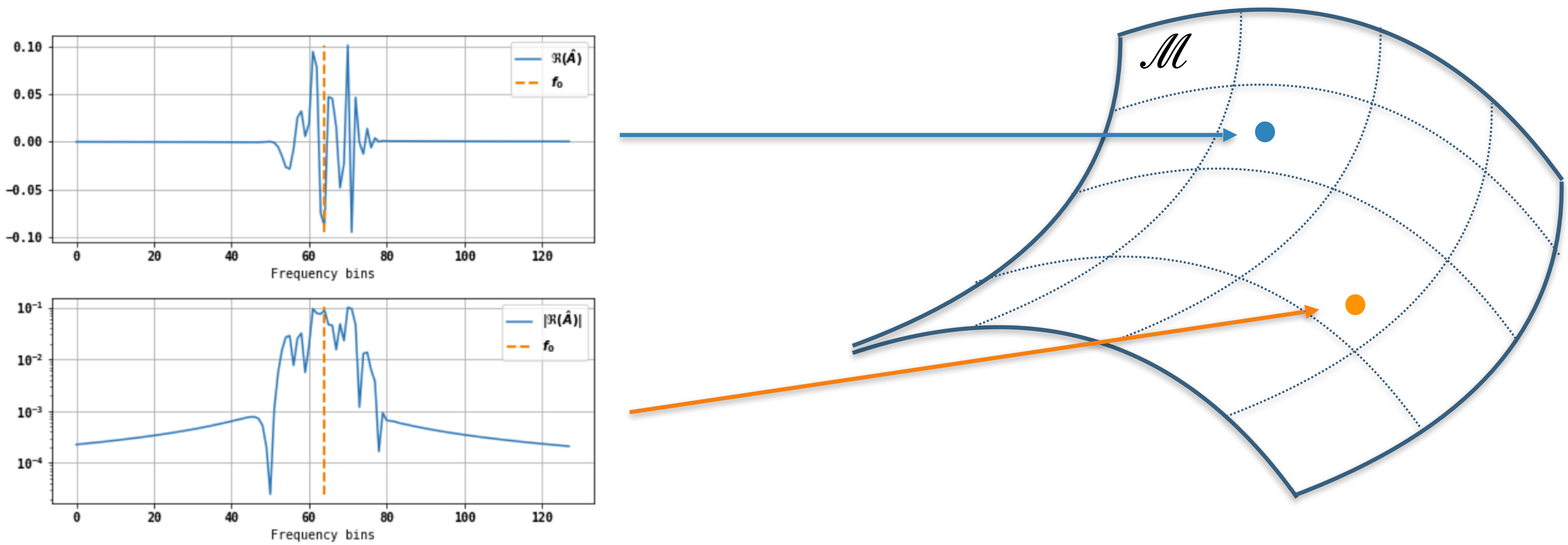
Detection: false positive rate/data SNR

Blelly, Bobin, Moutarde, MNRAS, 2021

LISA: the complexity of GW event unmixing

Standard sparse models rely on linear representations, which are not adapted to efficiently capture *morphologies* of most GW waveforms (e.g. GB, MBHB)

Such signals better live on a **low-dimensional manifold**.



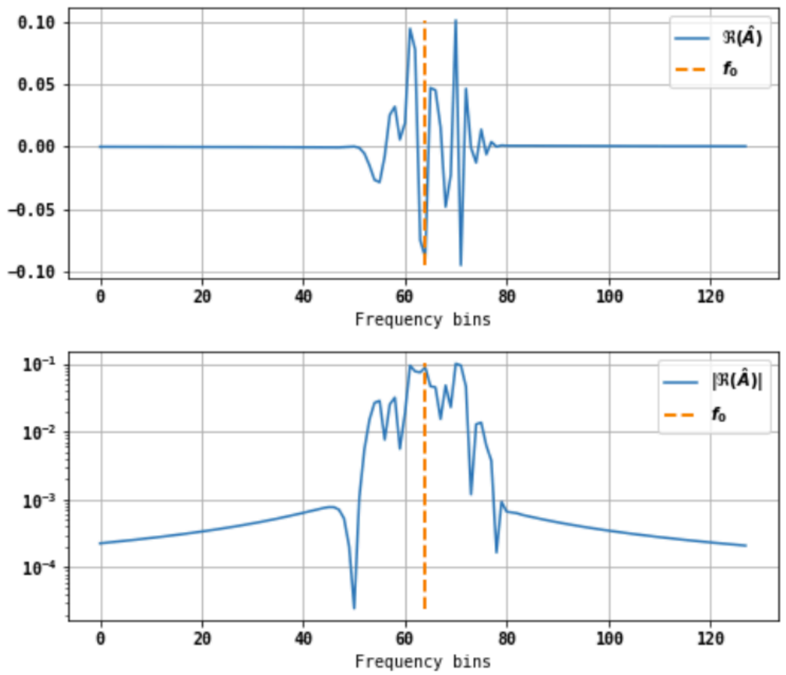
Two examples of GB waveforms in the Fourier domain

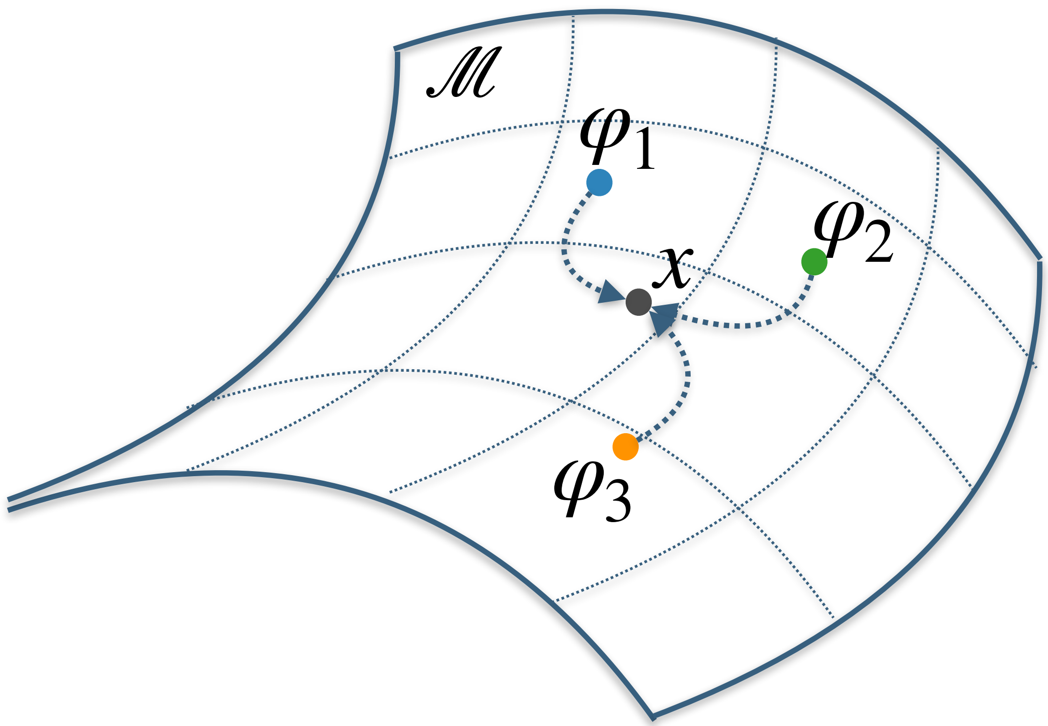
Learning a signal representation ≡ learning to "navigate" on the manifold

Transport learning on manifold - ***Culpepper & Olshausen, 2004*** - ***Connor, Canal & Rozell, 2020***

Learning manifold representations

Learning how to transport points on the manifold from **anchor points**



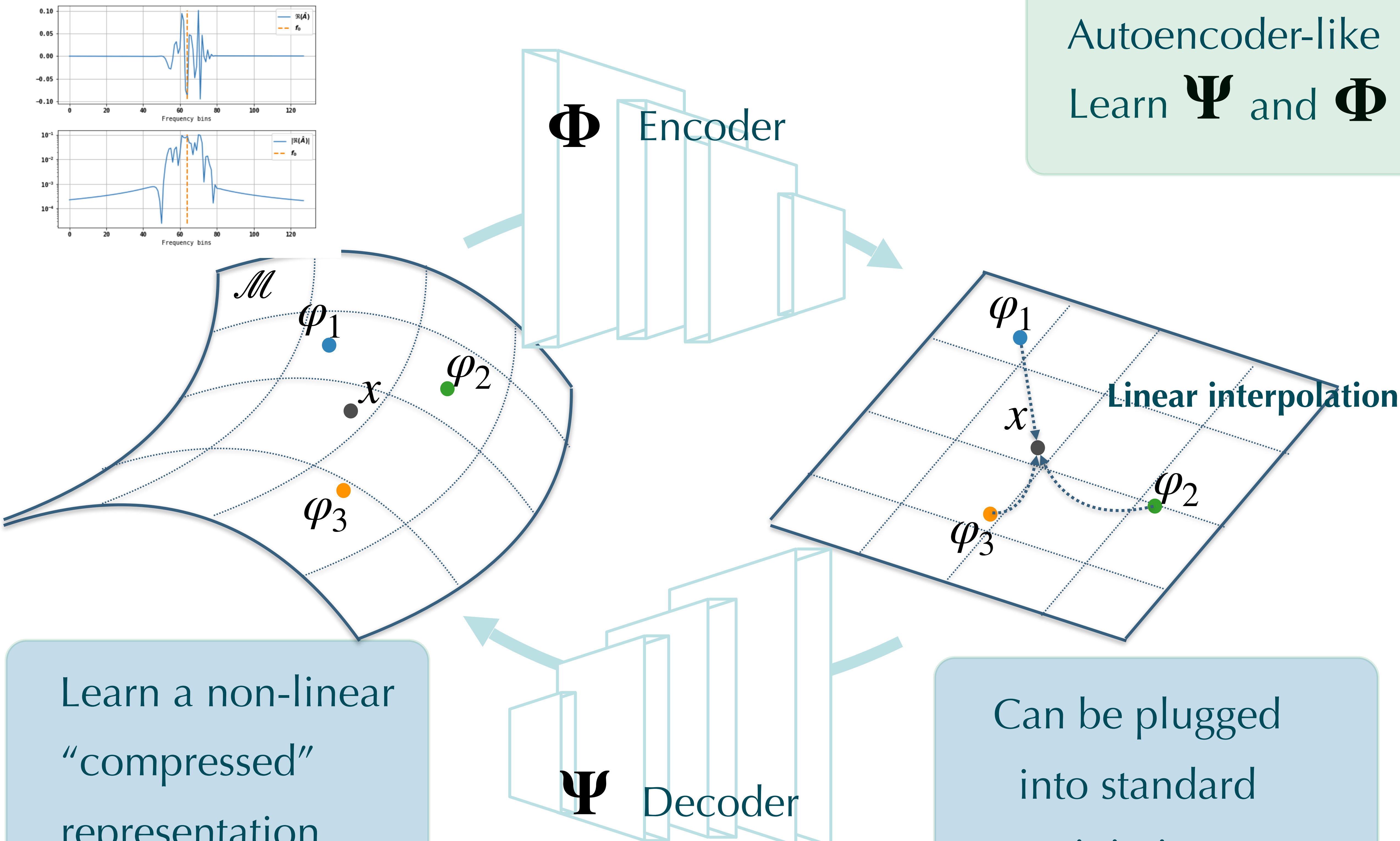


Define model-based signals as barycenters according to some metric ϕ

$$x = \operatorname{argmin}_{\mathbf{z}} \sum_{i=1}^{d} \lambda_i \phi(\mathbf{z}, \varphi_i)$$

To be learnt

Learning manifold representations



Autoencoder-like

Learn Ψ and Φ

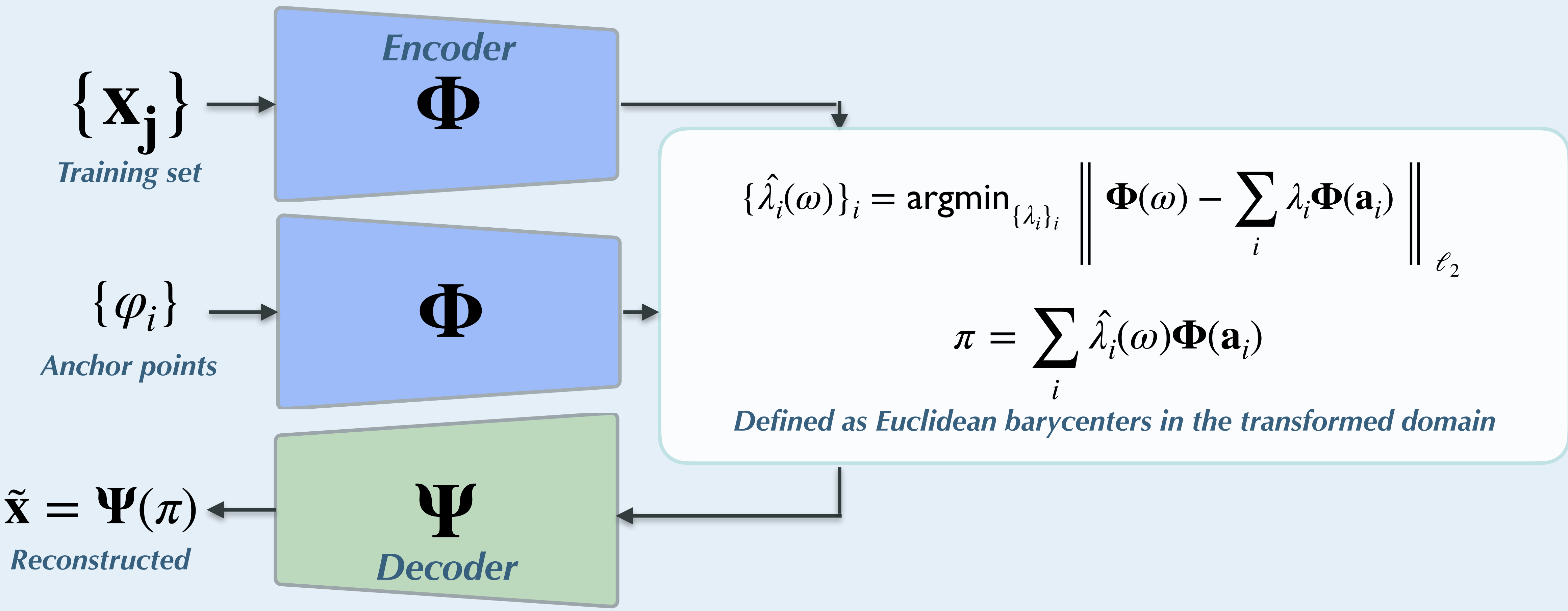
Learn a non-linear "compressed" representation

Can be plugged into standard minimisers

Bobin, Carloni-Gertosio, Bobin, Thiam, 2021

Learning manifold representations

Interpolatory AutoEncoder (IAE)



The parameters of the networks $\mathbf{\Phi}$ and $\mathbf{\Psi}$ *(e.g. MLP, CNN etc.)* are learnt by minimising the reconstruction error:

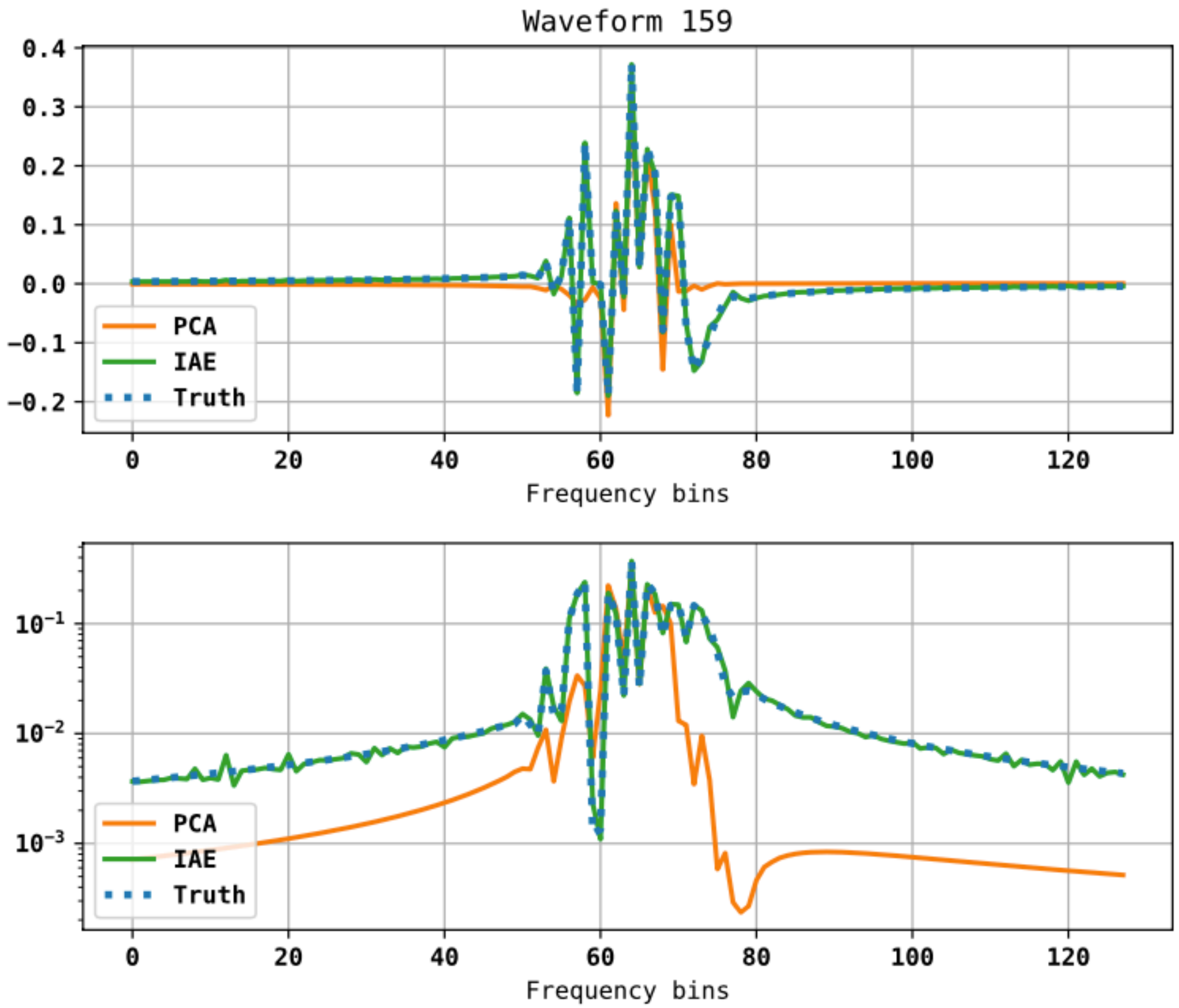
$$\min_{\mathbf{\Phi},\mathbf{\Psi}} \sum_{x_j \in \mathcal{T}} \left\| x_j - \mathbf{\Psi}\left(\sum_i \hat{\lambda}_i(x_j)\mathbf{\Phi}(\varphi_i) \right) \right\|_{\ell_2}^2$$

Choosing the can be made automatically by imposing a sparsity constraint in the latent domain

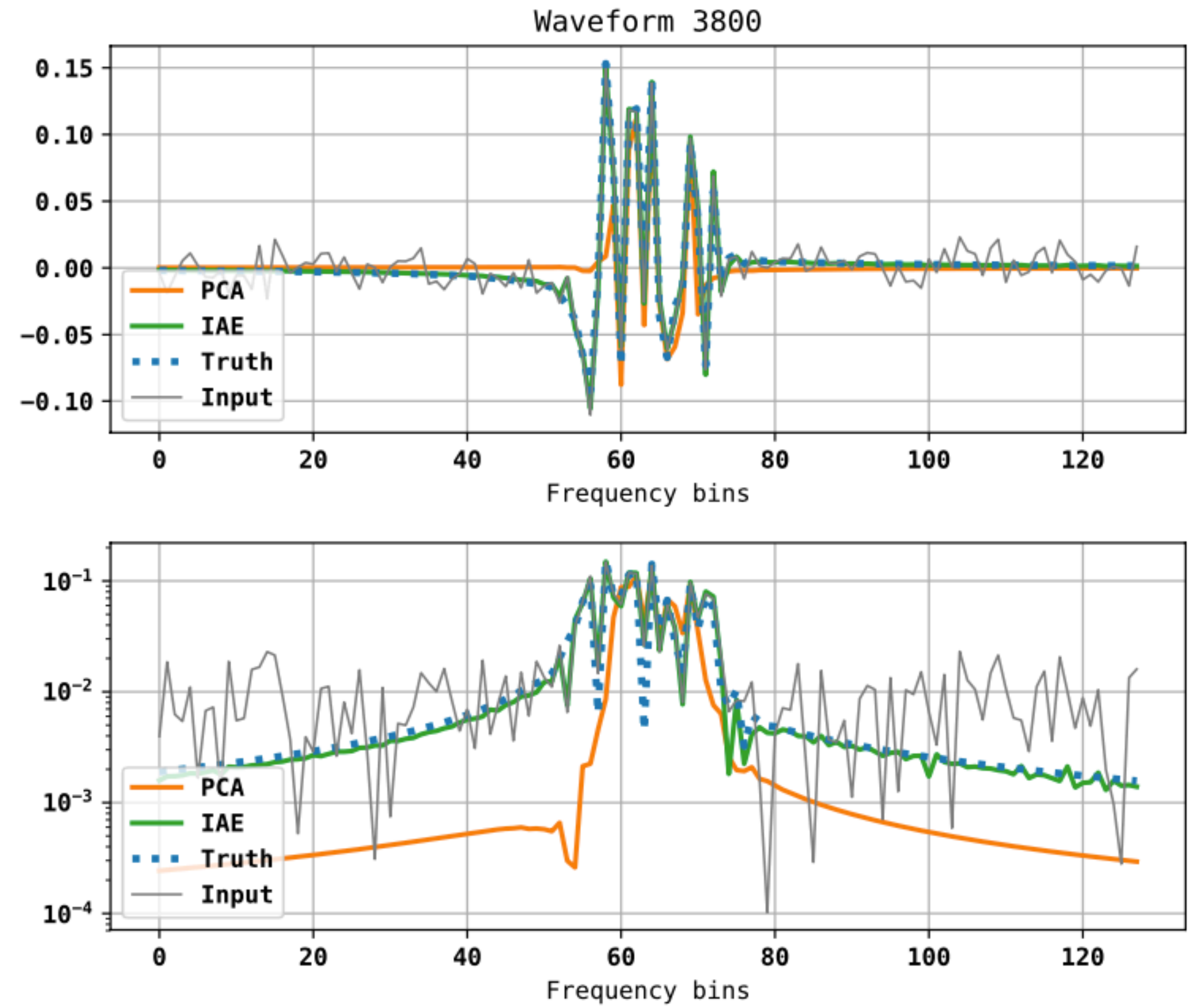
Training set: 5000 FastGB sims, uniform parameters' distribution

Fully connected residual network, with 3 layers, #hidden units/layer equal for all layers to the input dimension

PCA vs IAE - dimension of 9 (#of APs or PCs)



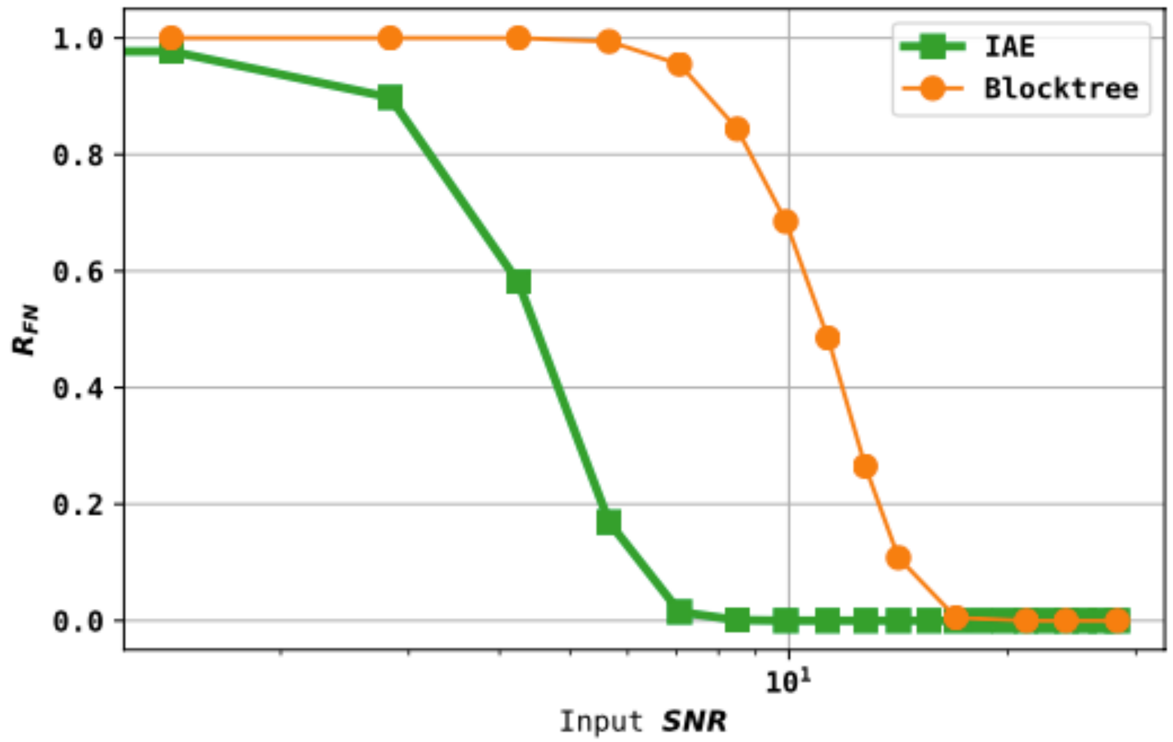
Noiseless examples



Noisy examples (LISA PSD as of LDC1-3)

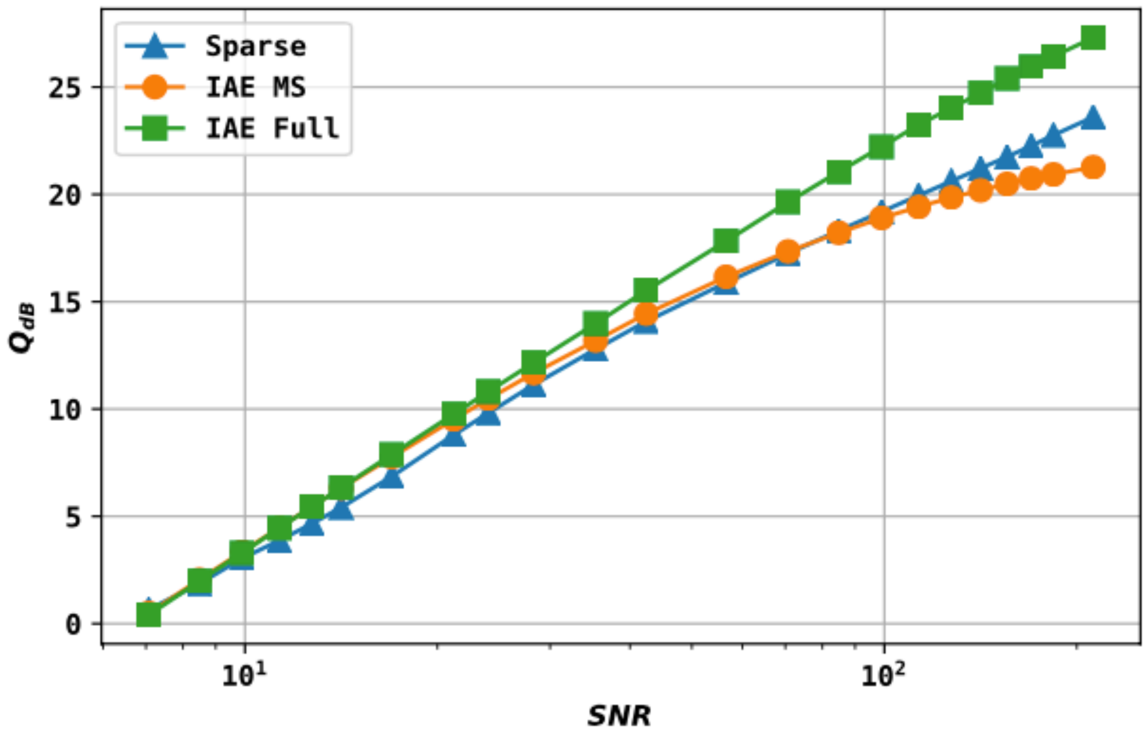
Results

Detection



False negative rate/input SNR

Reconstruction

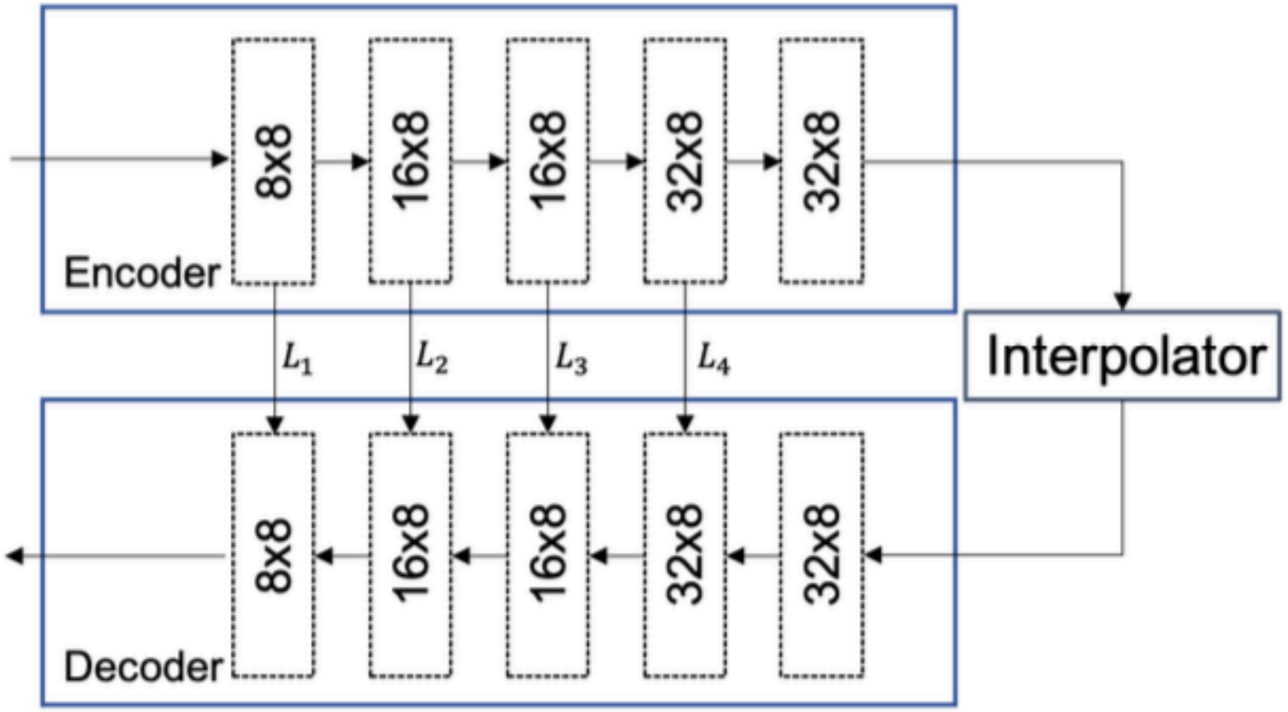


Recovery MSE in dB/input SNR

Slight improvement in recovery MSE

Significant improvement in recovery FNR

Preliminary application to MBHBs



- Convolutional IAE
- Lateral connections [5] : convolution & interpolation

Training

- 1000 centered MBHB signals, 500 points
- 30 randomly selected anchor points

Testing

- 250 MBHB signals with multiple noise levels each
- Extraction and detection performance

5. Rasmus, A., Raiko, T., & Valpola, H. (2014). Denoising autoencoder with modulated lateral connections learns invariant representations of natural images.

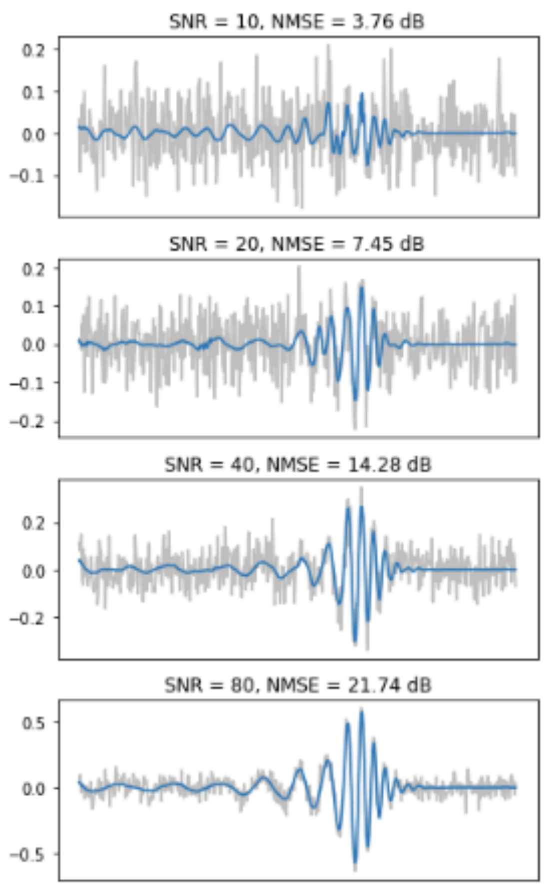
Preliminary application to MBHBs

Fast Interpolation

Orthogonal projection

minimize$_{(\lambda_i)}$

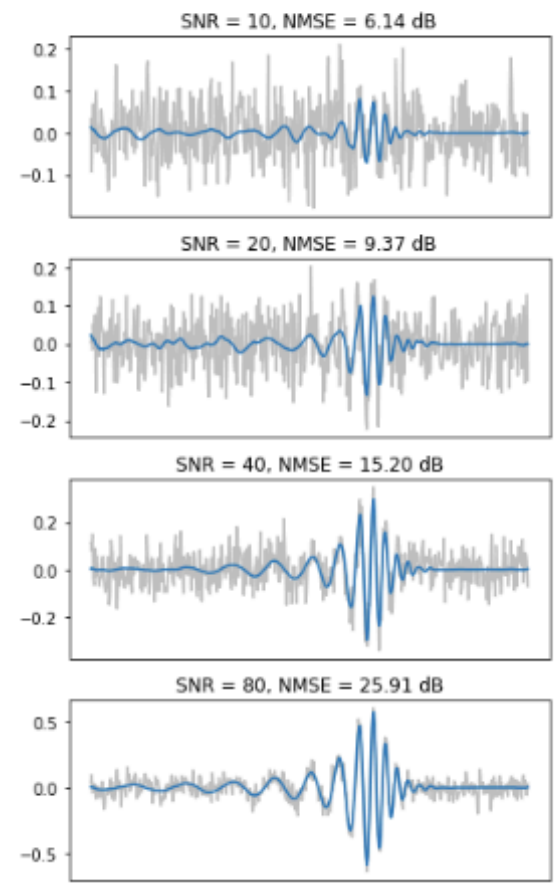
$$||\Phi(X_{in}) - \sum_i \lambda_i \Phi(a_i)||_2^2$$



Barycentric Span Projection

minimize$_{(\lambda_i)}$

$$||X_{in} - \Psi(\sum_i \lambda_i \Phi(a_i))||_2^2$$

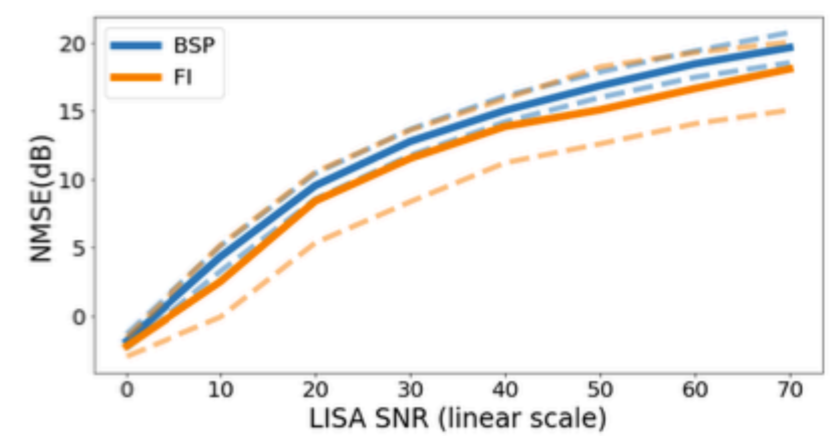


Simplex constraint

$$\sum_i \lambda_i = 1$$

Performance

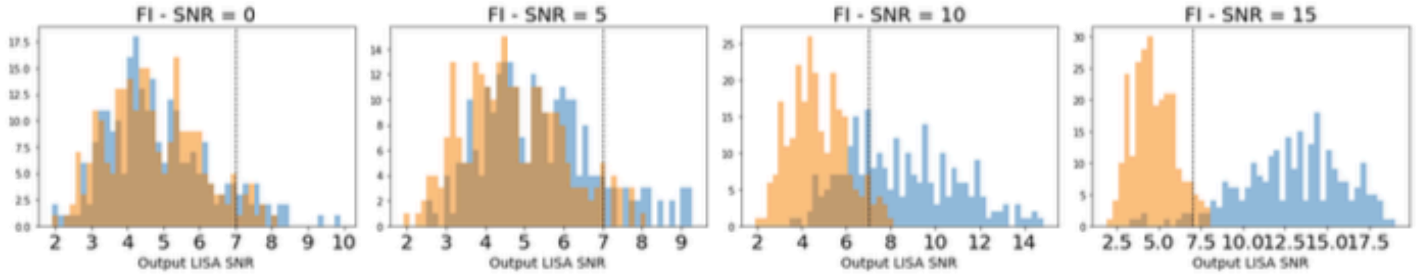
$$-20 \log_{10} \left(\frac{||X_{out} - X^*||_2}{||X^*||_2} \right)$$



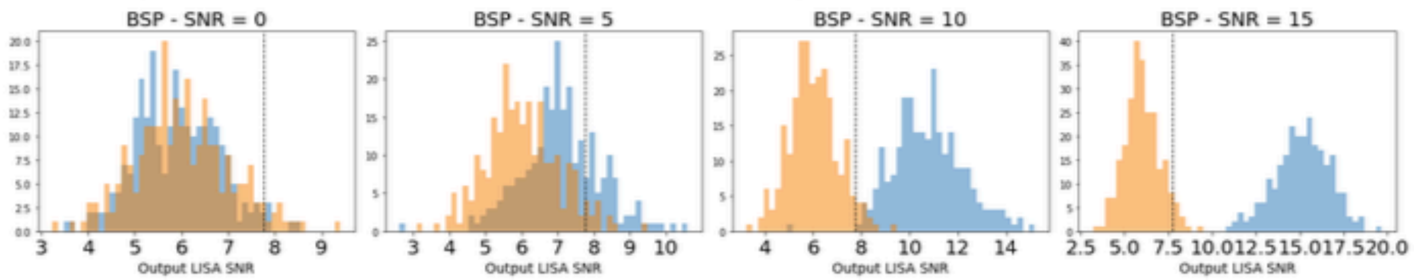
Methodology : Hypothesis testing

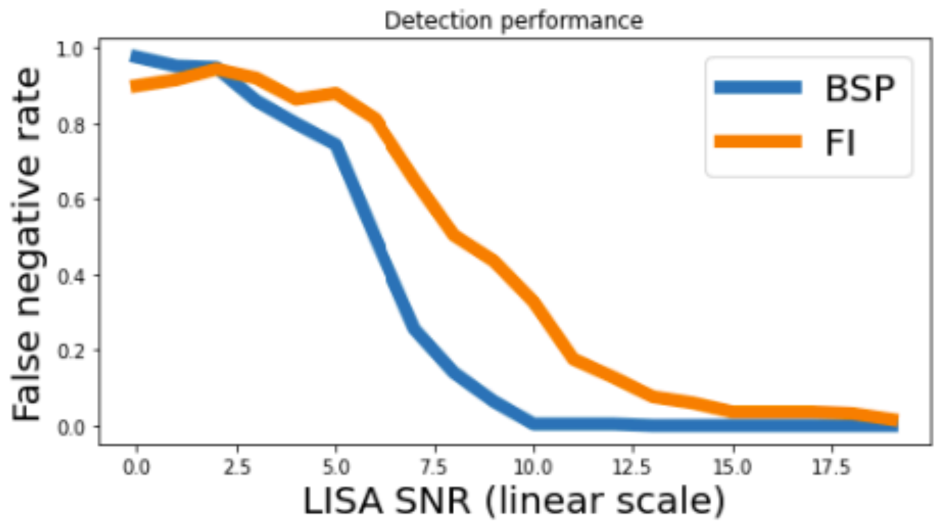
- Generate MBHB+noise and noise-only signals
- H_0 : "X_{in} contains only noise"
- Attempt to extract MBHB and compute a metric on X_{out}
- Thresholding based on fixed acceptable false positive rate

Fast Interpolation



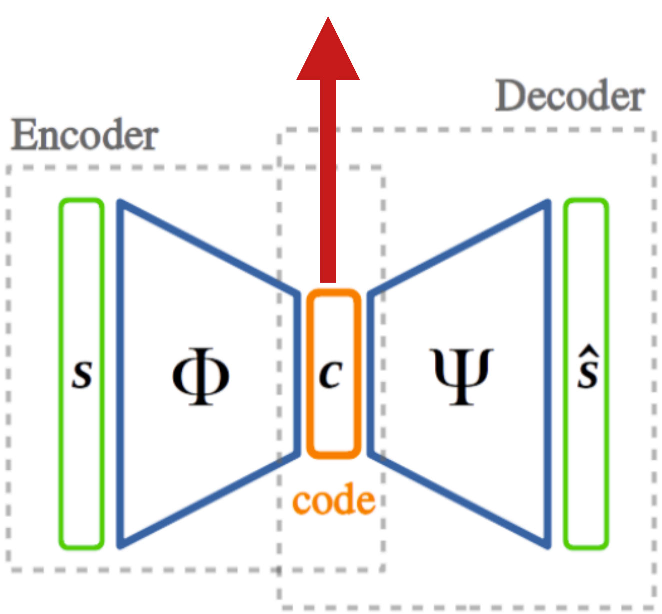
Barycentric Span Projection



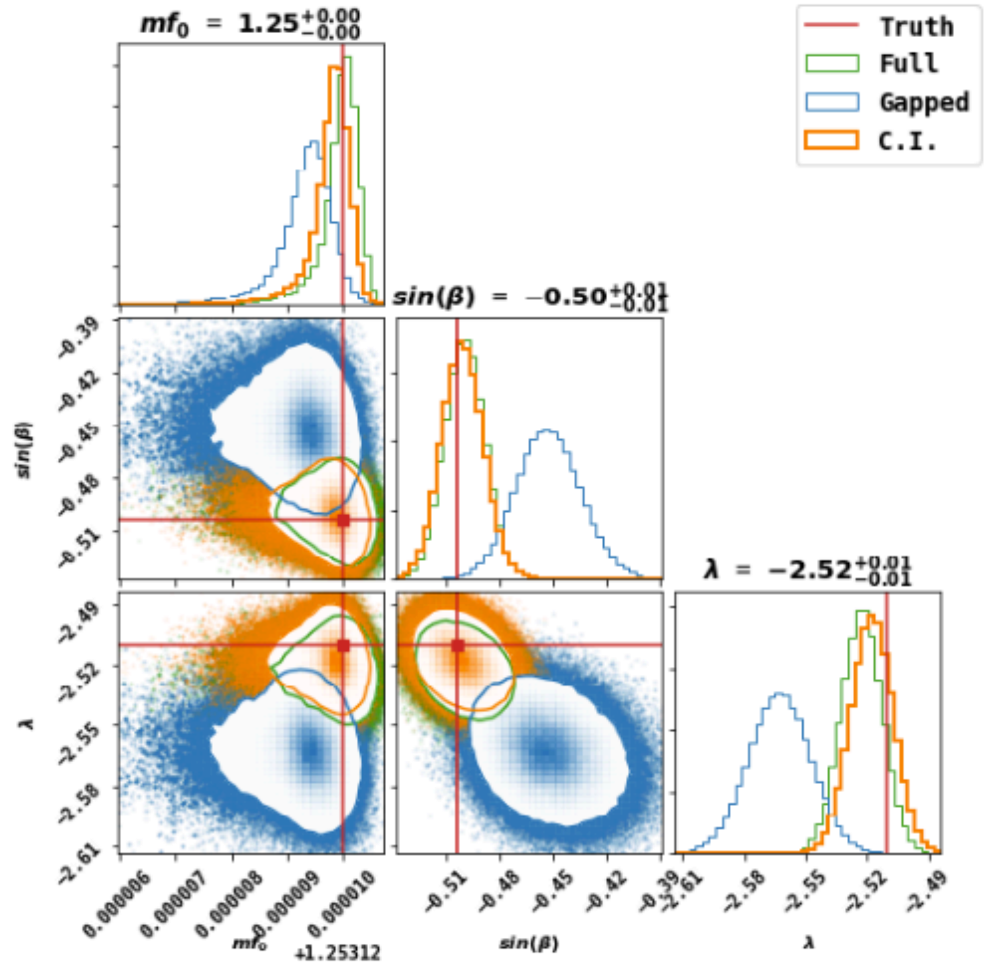


What's next for GBs ?

- Going towards a full fast GB pipeline
    - Estimation of individual events from the galaxy, discrimination resolved/unresolved sources
    - Improving the architecture for GB models
    - Plugging in the learned models in inpainting algorithm to deal with gaps
    - One of the building blocks for unmixing
    - Can we derive the parameters ?



Internship of S.Charpigny



What's next for MBHBs and beyond ?

- Going towards a full fast MBHB pipeline
    - Building an hybrid inspiral/coalescence model for MBHBs *(PhD work of E.Leroy)*
    - Plugging in the learned models in inpainting algorithm to deal with gaps, application to Spritz
    - Can we derive the parameters ?

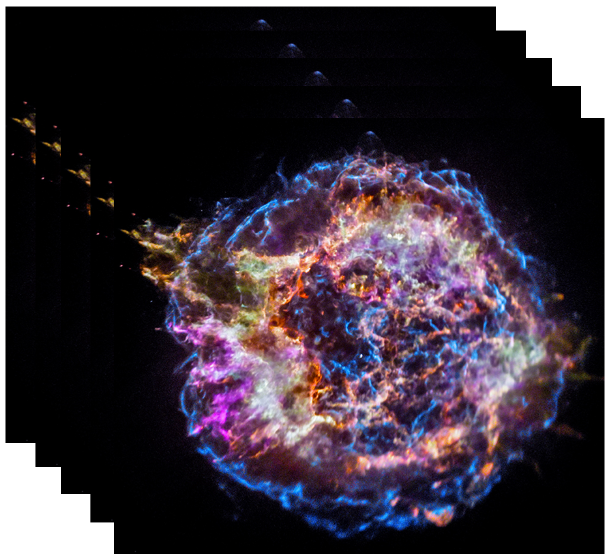
- Towards a first unmixing algorithm
    - Together with Aurore's work, second basic ingredient for unmixing, application to Sangria

- Beyond
    - Applications to other "sources", glitches ?
    - Combining with "purely" non-parametric methods for unknown sources

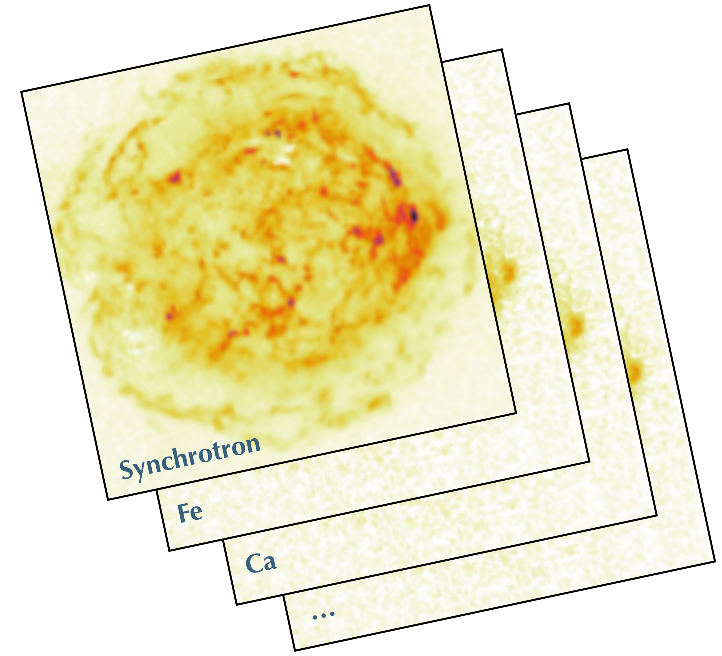
Not so different topic …
Learning to optimise

General unmixing problems

A linear mixture model



$= \sum$



Multispectral observations

Sources or components

Spatial distribution

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{N}$$

Emission spectra

Noise

Ill-posed problem:
Infinite number of solutions

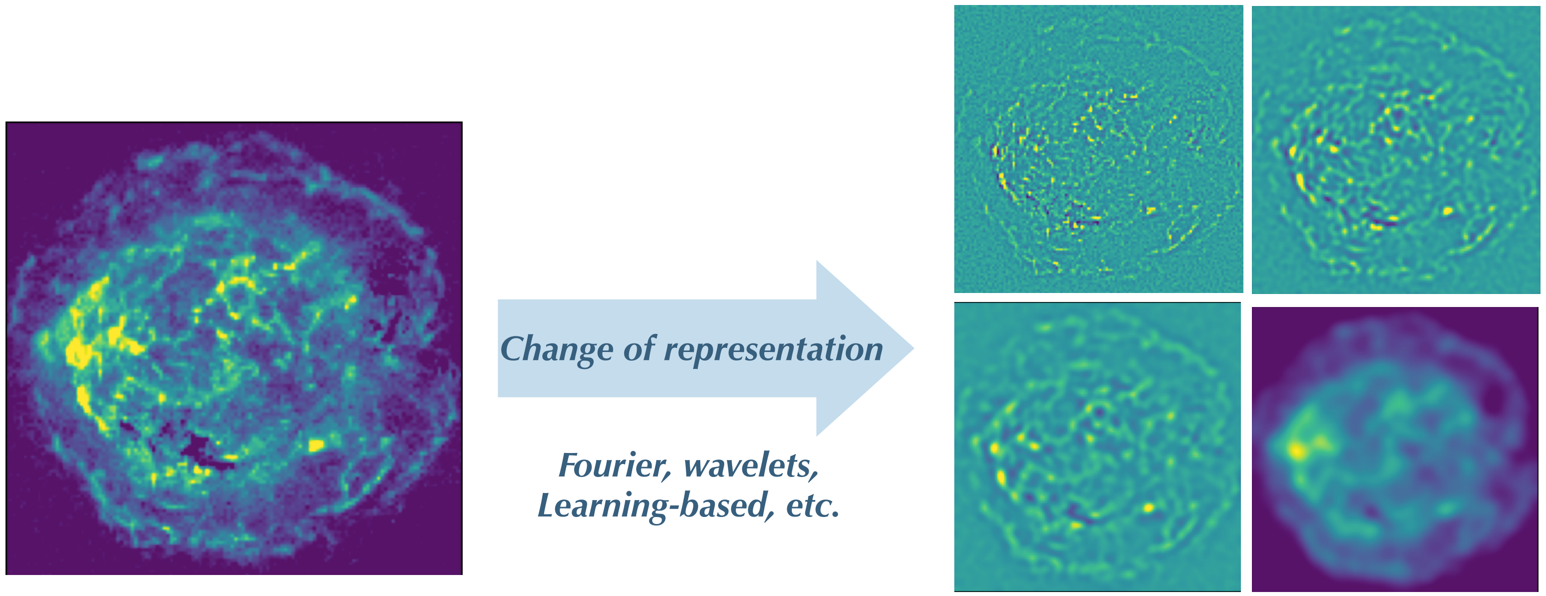
Unsupervised Source Separation:
Estimation both A and S from X only

A complex problem to be tackled

$$\min_{\mathbf{A},\mathbf{S}} \underbrace{\mathscr{R}(\mathbf{A}) + \mathscr{J}(\mathbf{S})}_{\text{Regularization Terms}} + \underbrace{\mathscr{D}\left(\mathbf{X}, \mathbf{AS}\right)}_{\text{Data fidelity term}}$$

Data fidelity term:
- measures a discrepancy between the data and the model
- allows to account for the noise statistics
- general formulation for various mixture models

Regularization terms:
- make “better”-posed an ill-posed problem
- favour solution properties for increased interpretability



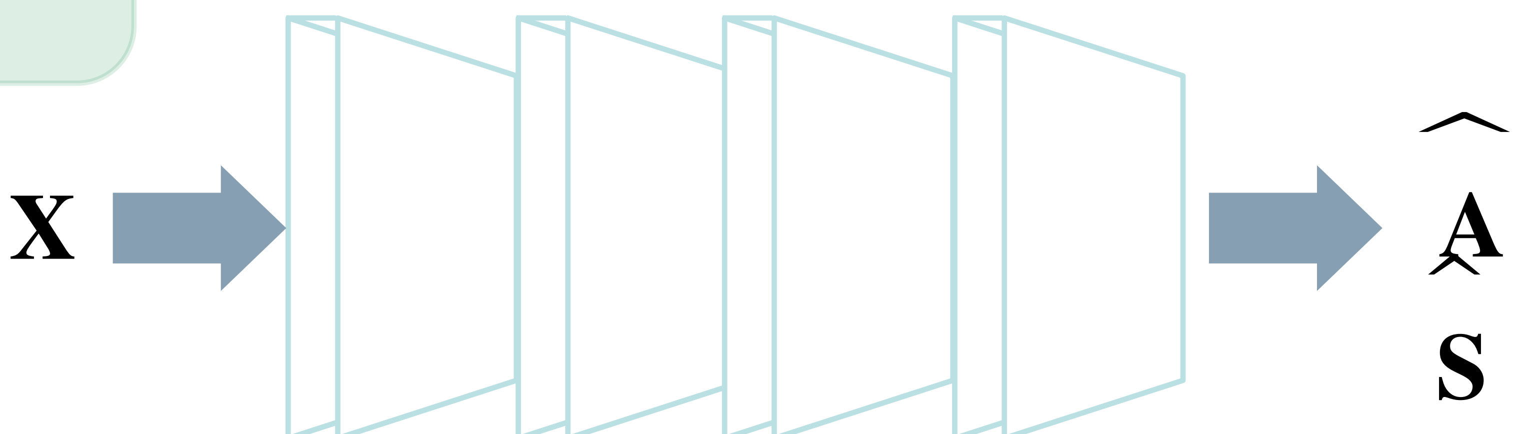
Sparse/compressible Representation

$$\min_{\mathbf{A},\mathbf{S}} \mathcal{R}(\mathbf{A}) + \mathcal{J}(\mathbf{S}) + \mathcal{D}\big(\mathbf{X}, \mathbf{AS}\big)$$

Costly
Iterative algorithms

$$\mathbf{S}^{(+)} \leftarrow \mathsf{P}_{\alpha\mathcal{J}}\Big[\mathbf{S}^{(-)} - \gamma \nabla_{\mathbf{S}} \mathcal{D}\big(\mathbf{X}, \mathbf{AS}^{(-)}\big)\Big]$$
$$\mathbf{A}^{(+)} \leftarrow \cdots$$

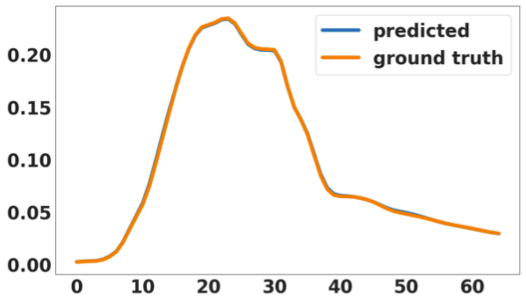
Algorithm unrolling
Each layer "mimics"
a single iteration



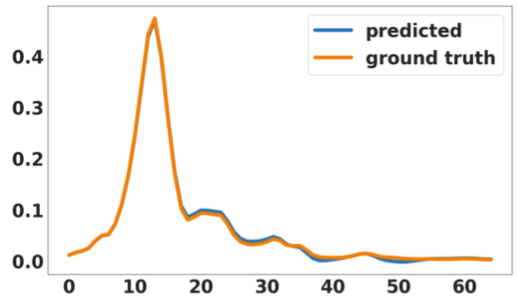
Recurrent neural network

Learn "algorithmic" hyperparameters
e.g. gradient stepsize, Hessian …

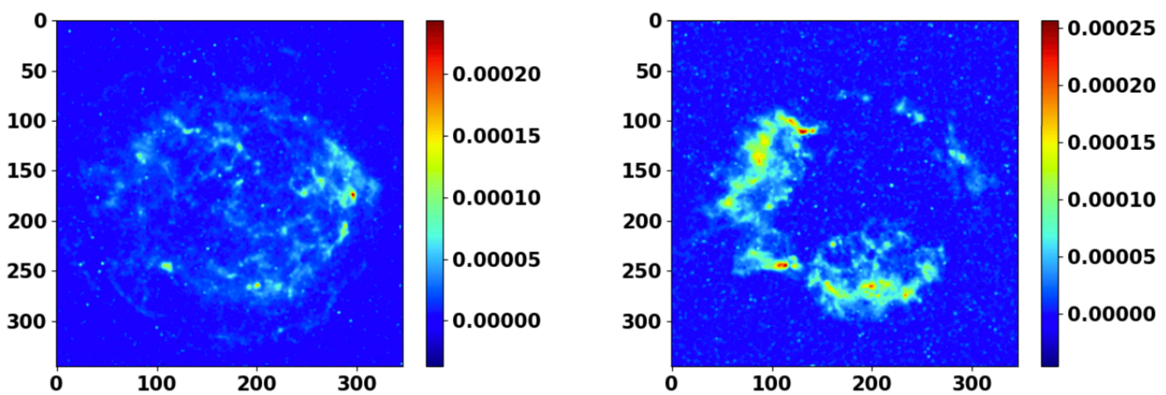
Speeding-up unmixing algorithms



(a)



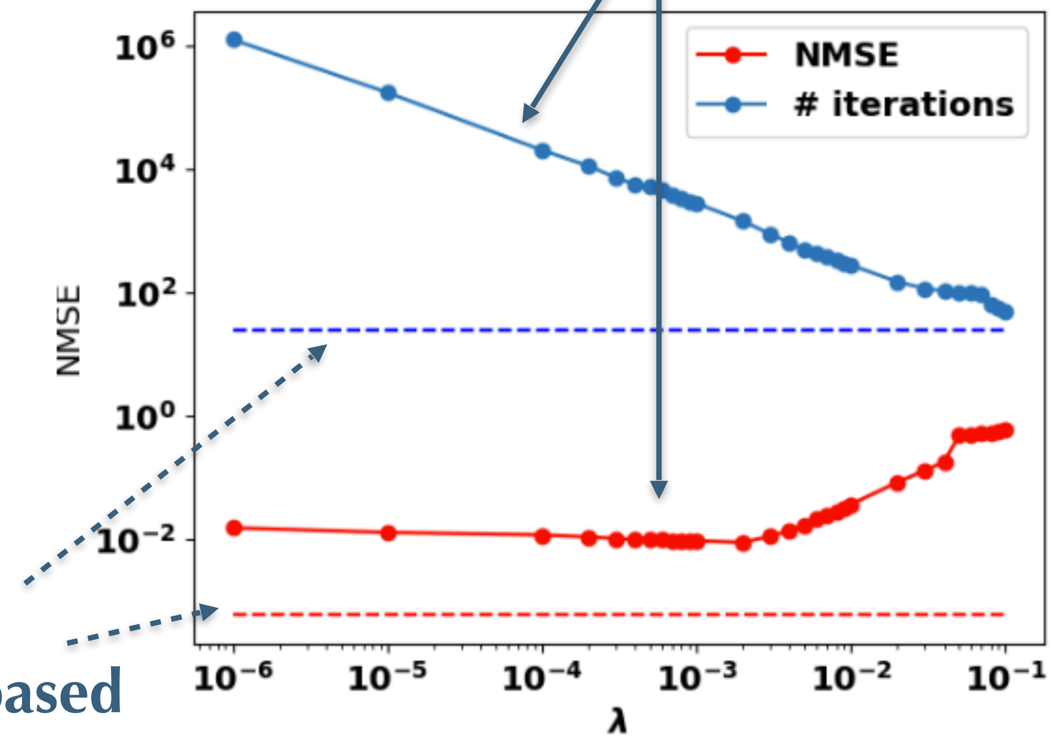
(b)



Results on simulated Chandra

RNN with 15 layers

Standard algorithm (PALM)



Learning-based (L-PALM)

Regularisation parameter

Faster *(up to 4 orders/magnitudes)*

Better *(implicit regularisation)*

than standard algorithms

Fahes, Kervazo, Bobin, Tupin, ICLR 2022

What to take away

Towards hybrid non-parametric/ML-based models/methods

- for fast analysis

- flexible framework to deal with/account for systematics/artefacts

- discovery/recovery of badly/unknown events

***IAE:** https://github.com/jbobin/IAE*

***GB codes :** https://github.com/GW-IRFU/gw-irfu*

To be explored:

- Speed up/improve the recovery based on algorithm unrolling

- Speed up param. est. (e.g simulation-based inference ?)

- Uncertainty quantification (e.g Bayesian deep learning, etc)