

Méthode de séparation de sources

Modèles et algorithmes

Applications en Astrophysique

Advanced sparse BSS

The modern way of tackling BSS

J.Bobin/C. Kervazo

jerome.bobin@cea.fr - christophe.kervazo@telecom-paris.fr

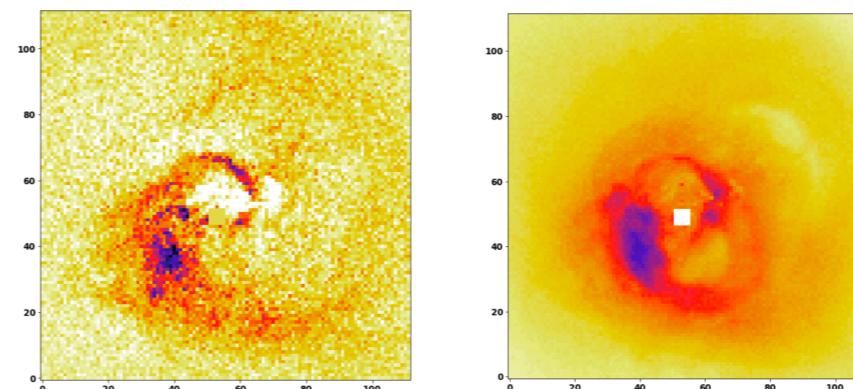
Challenges of BSS

Open problems in blind source separation include (but not restricted to):

- ▶ **The joint (and robust) estimation of the number of sources**
- ▶ **Dealing with more exotic/realistic noise contamination**
(e.g. Poisson, impulsive noise, etc.)

$$\min_{\mathbf{A}, \mathbf{S}} \mathcal{R}(\mathbf{A}) + \mathcal{J}(\mathbf{S}) + \boxed{\frac{1}{2} \|\mathbf{X} - \mathbf{AS}\|_F^2}$$

- ▶ *Adapt to data fidelity term*
- ▶ **Real data generally exhibit some partial correlations, which entails that morphological diversity does not hold**



Perseus galaxy cluster in X-ray

Open problems in blind source separation include (but not restricted to):

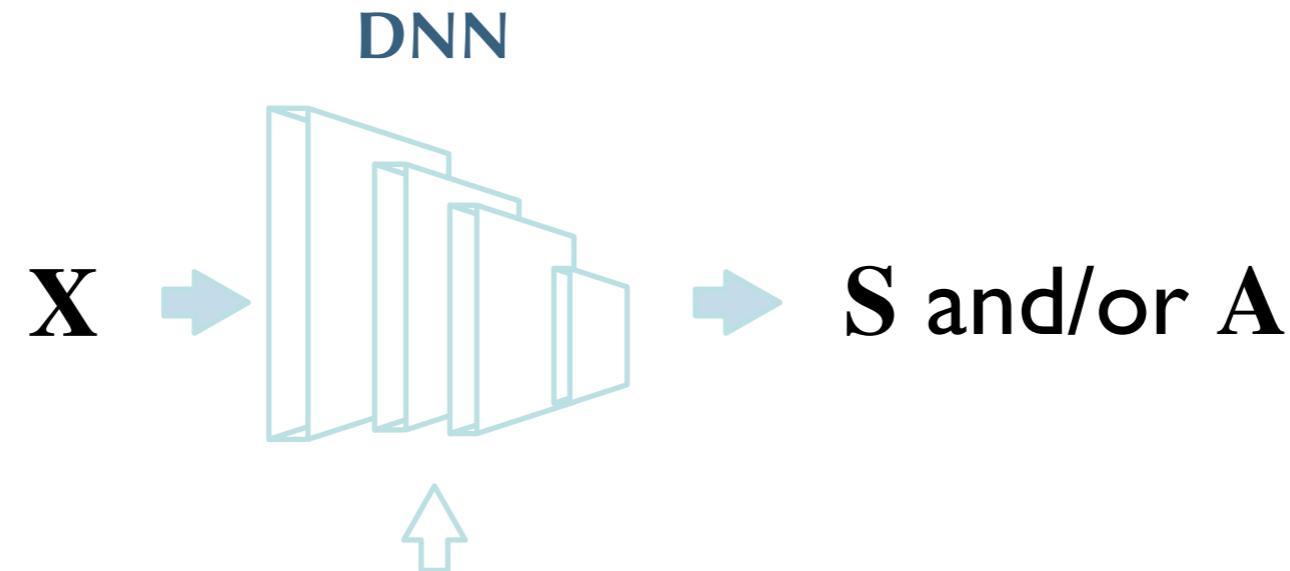
- ▶ **How to switch from linear mixture models to more realistic non-linear models ?**
 - ▶ *Course #2; some recent advances but far from real-world applications*
 - ▶ *Course #8 in the context of NMF*
- ▶ **Rank-1 components are generally a simplistic assumption**
 - ▶ *Dealing with non-stationary mixture models*
- ▶ **Interpretable results ??**
 - ▶ *Inject data-driven/physics knowledge in the unmixing process*
 - ▶ *Quantifying estimation uncertainties*

Hybrid sparse BSS

Data-driven regularizations

How to build a physics-informed unmixing process ?

- *Machine learning-based unmixing process*

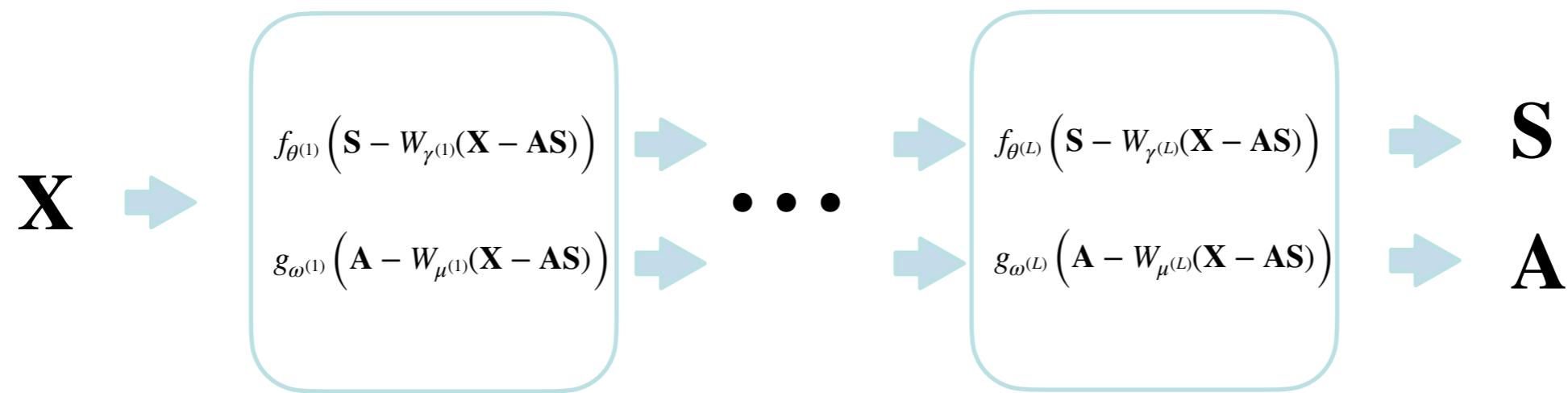


information is carried out by the training set

- **Does not allow to precisely account for the direct model**
Noise statistics, mixture model, etc
- **Training set building could be cumbersome** (*mixture dependent, etc*)
- **Not quite adaptive** (*highly problem/data dependent*)

Algorithm unrolling

Beyond building fast solvers, algorithm unrolling is also a way to inject information in the estimation process.



information is carried out by the training set

- Allows to precisely account for the direct model
Noise statistics, mixture model, etc
- Training set building could be cumbersome (*mixture dependent, etc*)
- Probably mildly adaptive (*highly problem/data dependent*)

Plug & Play methods

Let's go back to the basics,

$$\min_{\mathbf{X}} \lambda \|\mathbf{X}\|_1 + \frac{1}{2} \|\mathbf{Y} - \mathbf{X}\|_2^2$$

Classically gives

$$\mathbf{X} = \mathcal{S}_\lambda(\mathbf{Y})$$

Basically, the regularisation often boils down to a denoiser, or more generally to providing a low-dimensional approximation of the signal.

Plug & Play methods

Gist of plug&play methods (in this context) is to learn denoisers

Romano et al. 2016

$$\mathbf{X} = \mathcal{D}_\theta(\mathbf{Y})$$

and plug them in classical solvers:

$$\mathbf{S} \leftarrow \mathcal{D}_\theta \left(\mathbf{S} + \alpha \mathbf{A}^T (\mathbf{X} - \mathbf{A}\mathbf{S}) \right)$$

here in a forward-backward splitting algorithm

which could be interpreted as some regularization

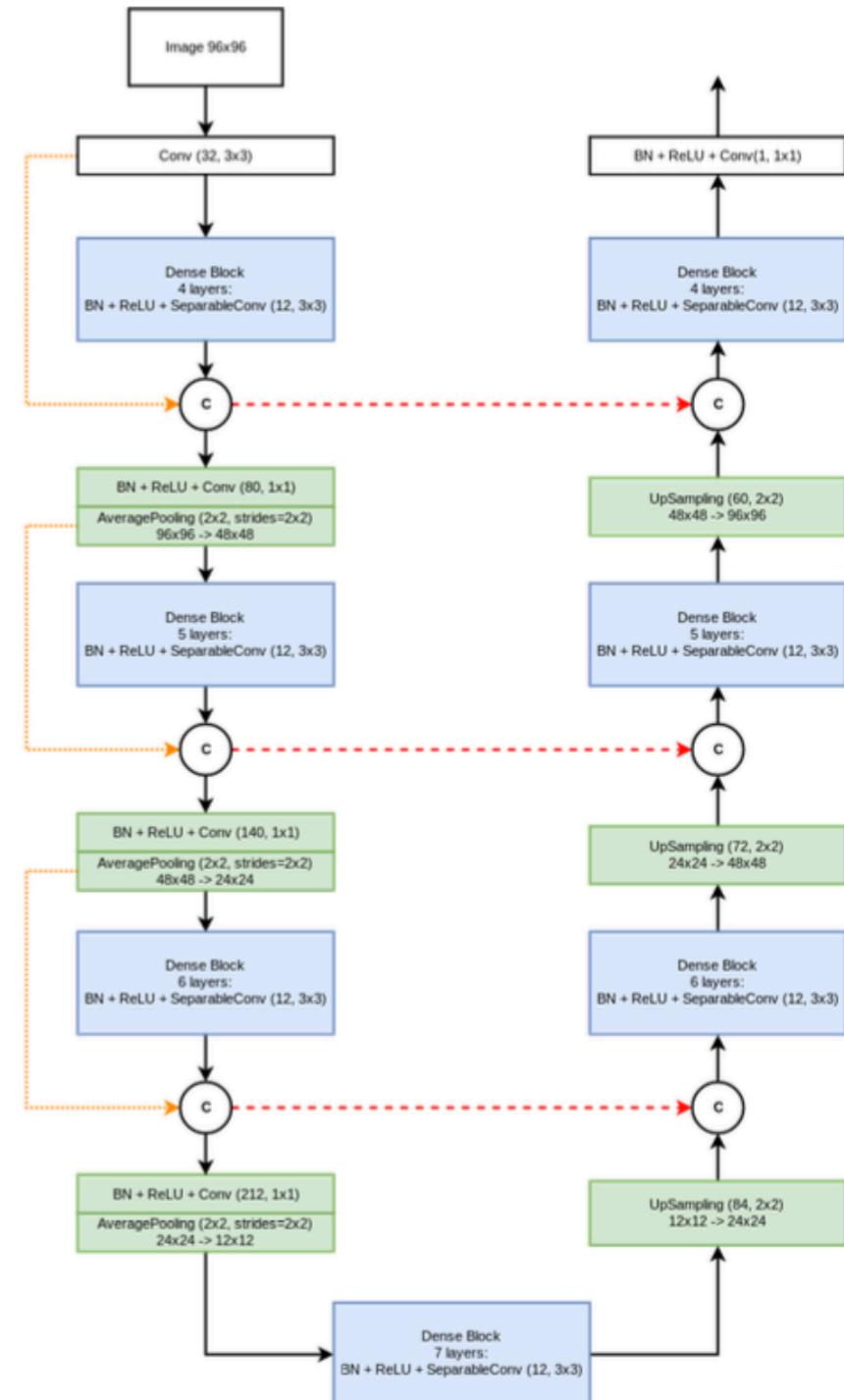
$$\min_{\mathbf{S}} \mathcal{J}_\theta(\mathbf{S}) + \frac{1}{2} \|\mathbf{X} - \mathbf{A}\mathbf{S}\|_2^2$$

Plug & Play methods

- Virtually any denoiser architecture can be used

- U-Net quite popular for image denoising
(Though being highly over-parameterized)

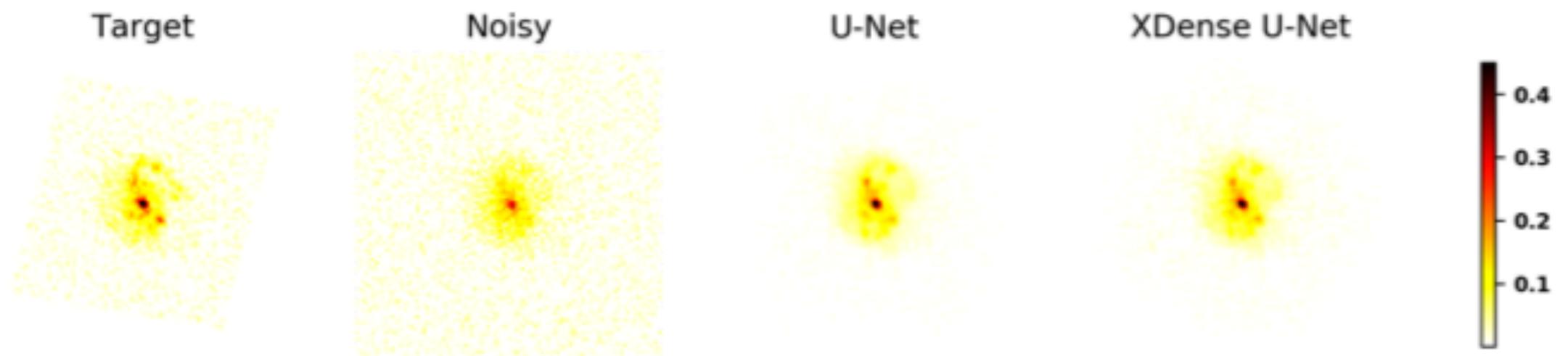
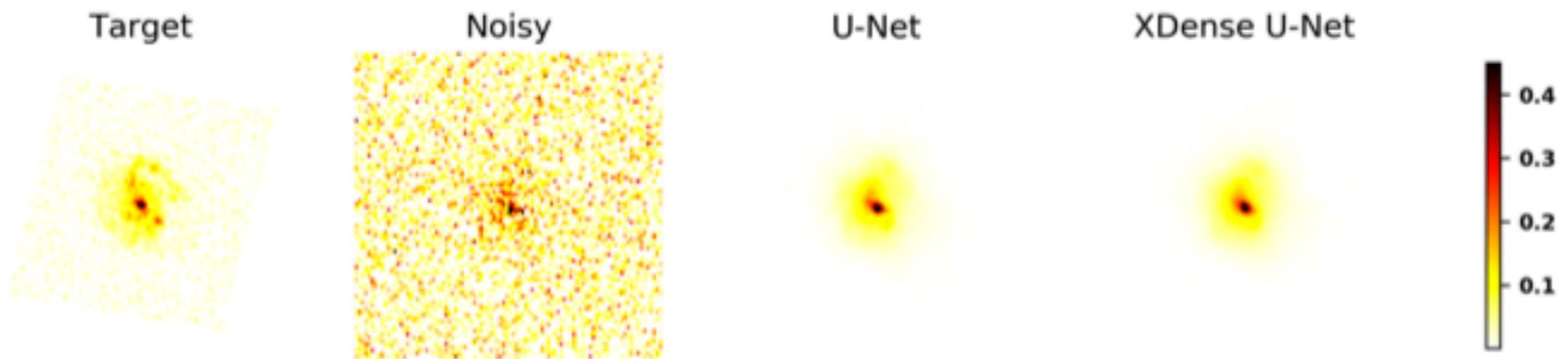
- Applied to denoising, deconvolution, tomographic reconstruction, etc. but not to BSS



Sureau et al. 2019

Plug & Play methods

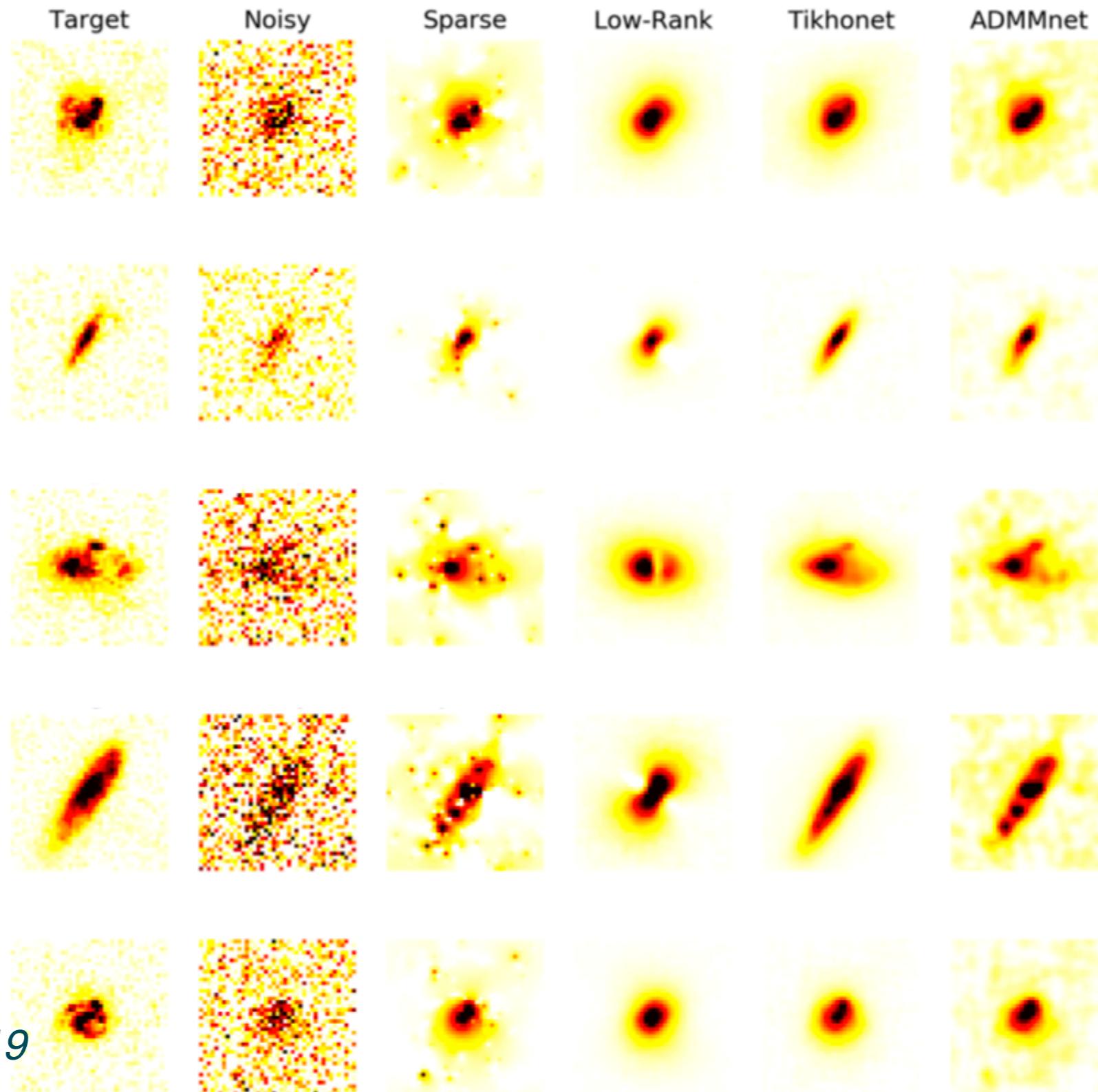
U-Net architectures for galaxy image denoising



Sureau et al. 2019

Plug & Play methods

U-Net architectures for galaxy image deconvolution



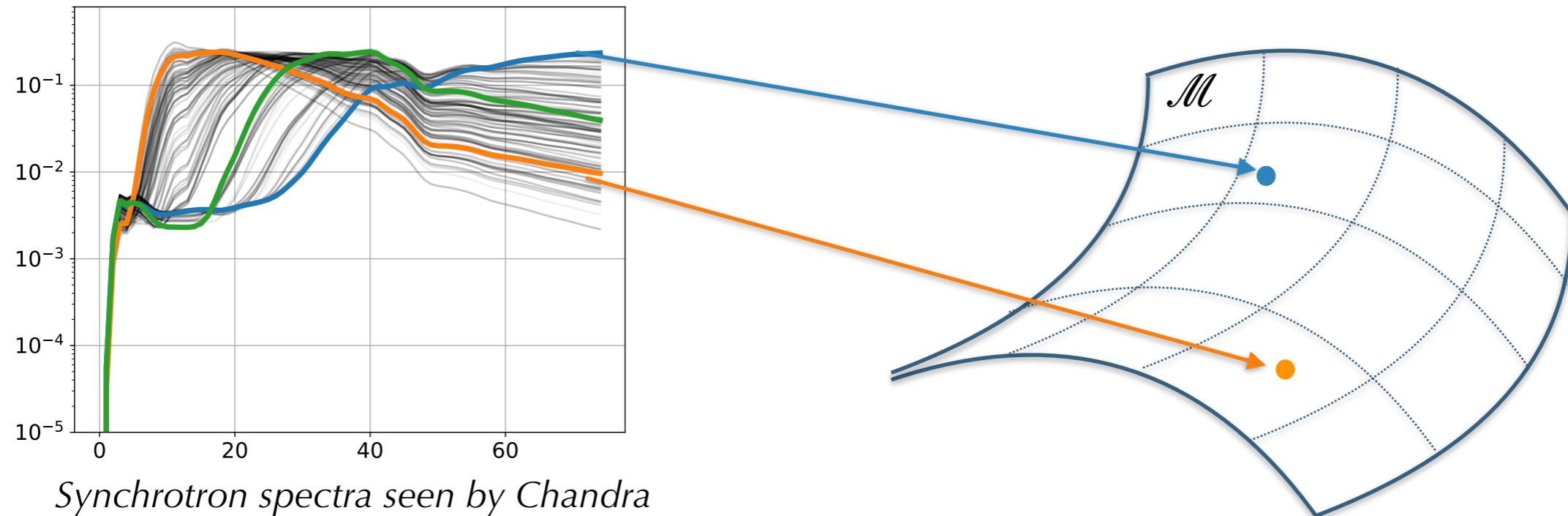
Sureau et al. 2019

Combining GMCA with manifold learning

(And what for ?)

The manifold hypothesis

**Physically relevant signals generally belong to
a smooth low-dimensional manifold**



State of the art (in the context of BSS):

- parametric models (*Irfan et al., 2019*): **not flexible and difficult to apply**
- learning-based methods: **not adapted and/or too complex**
 - ▶ end-to-end (*Krakel & Bengio, 2017; Kameoka et al., 2019*)
 - ▶ combined with variational approach (*Nugraha et al., 2016*)

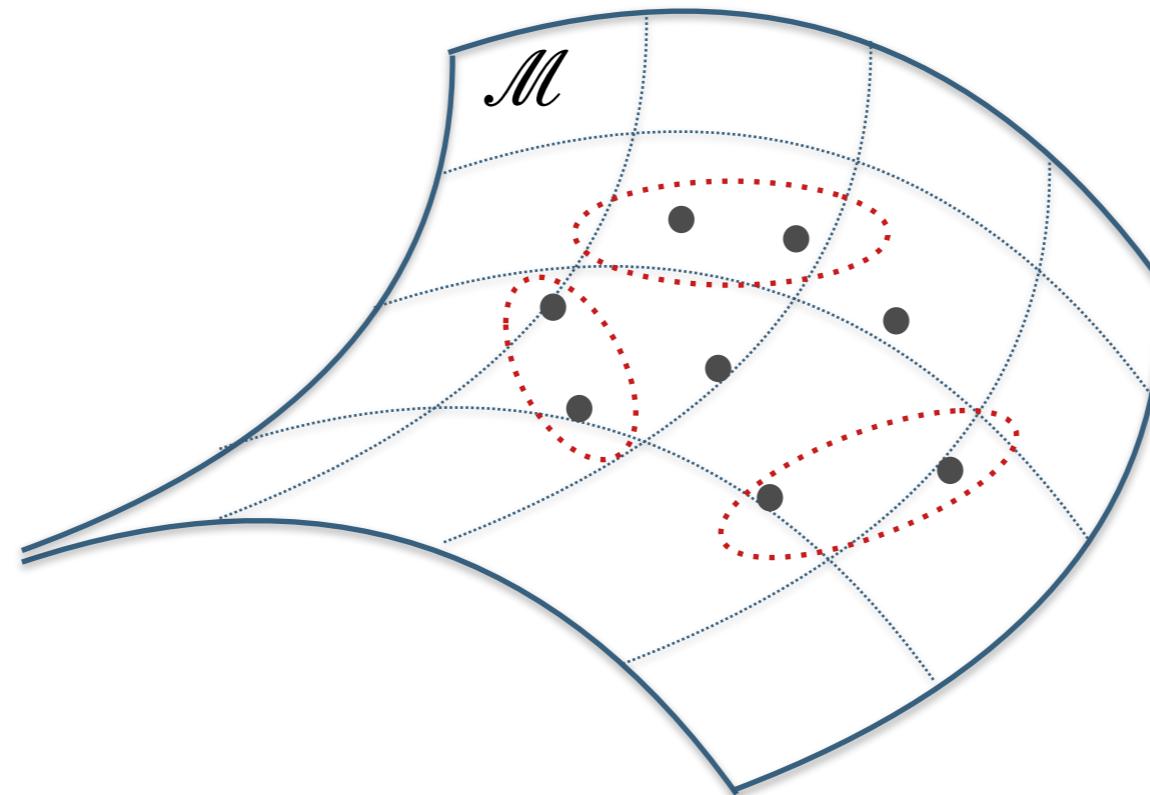
The manifold hypothesis

1st ingredient:

Learning how to transport points
on the manifold

- Culpepper & Olshausen, 2004
Transport learning on manifold

- Connor, Canal & Rozell, 2020
*Transport learning on manifold
+ generative model*

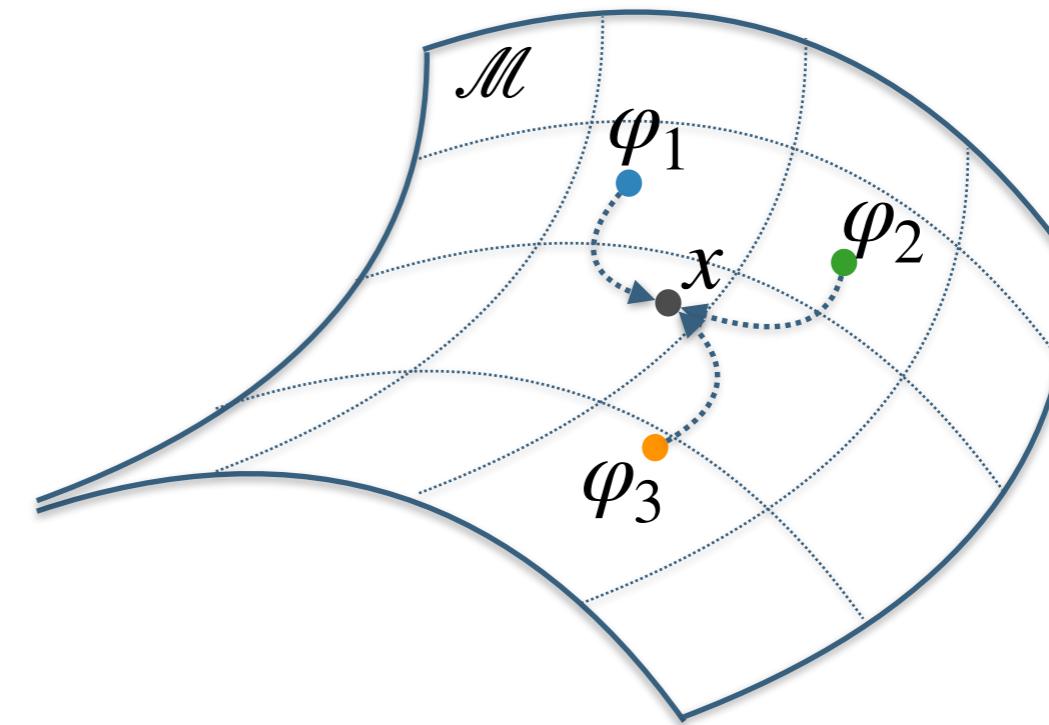
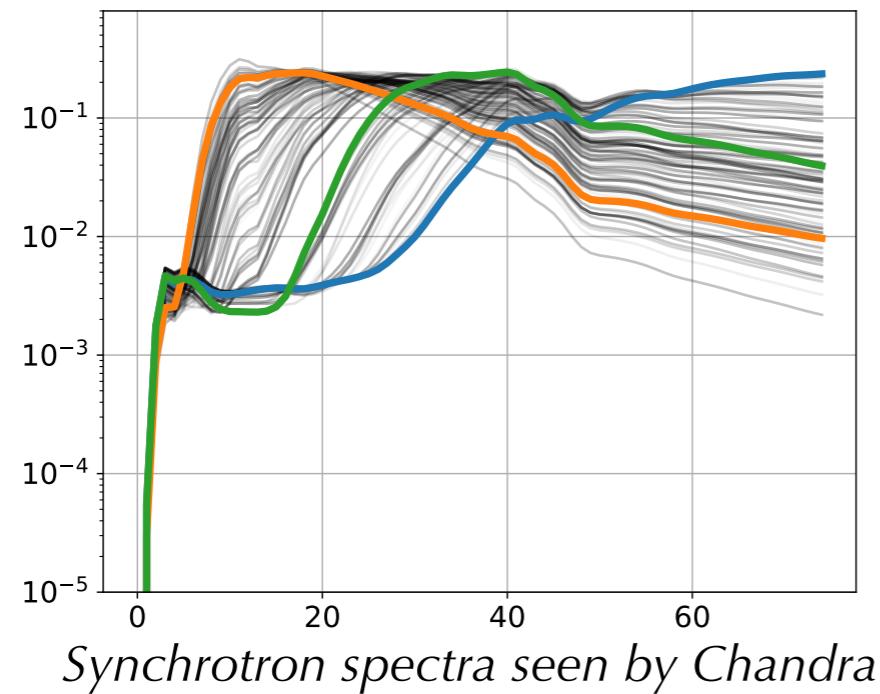


- Not adapted for scarce training samples
- Fixed metric formulation

A non-linear interpolation scheme

2nd ingredient:

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



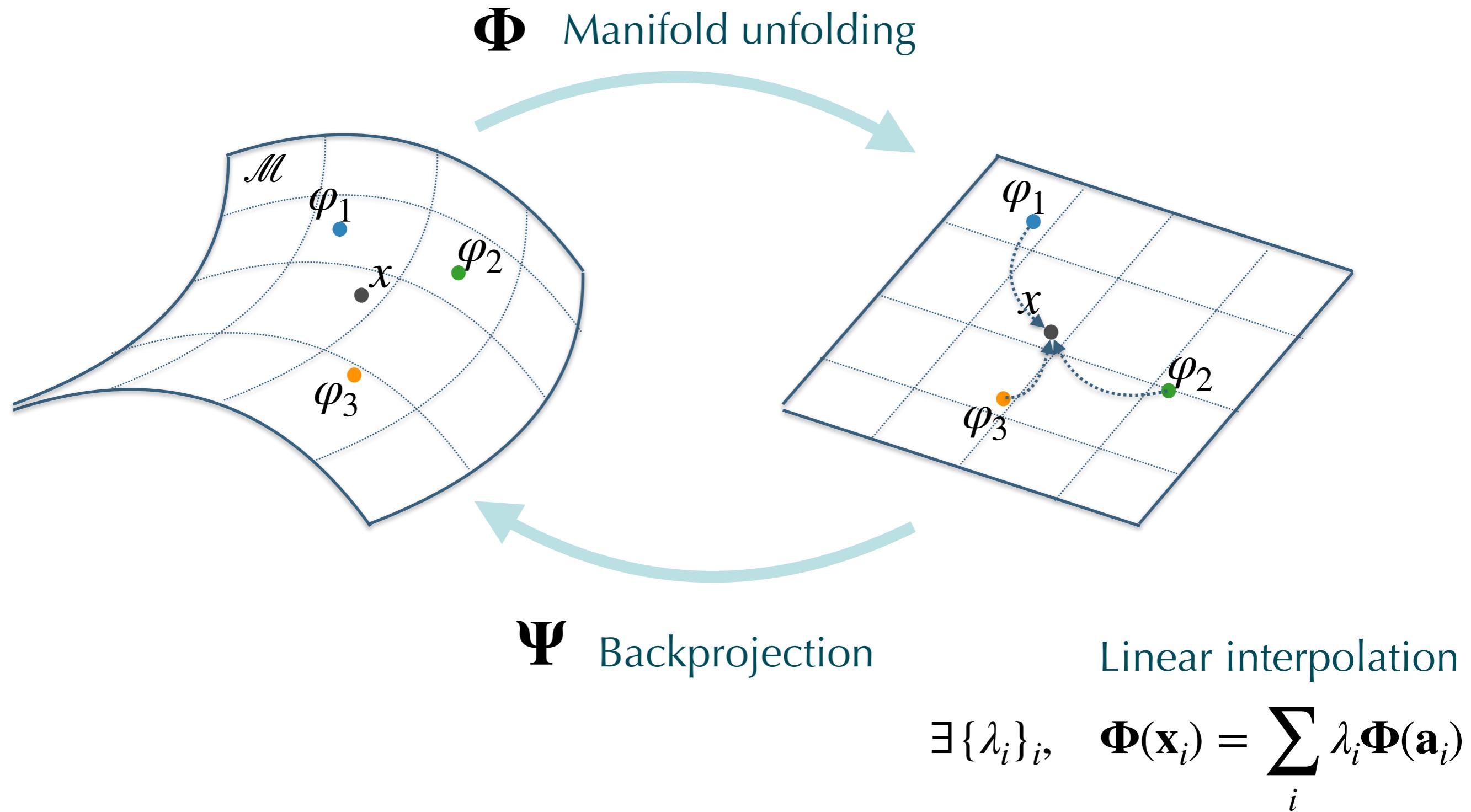
Define model-based signals as barycenters according to some metric ϕ

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

To be learnt

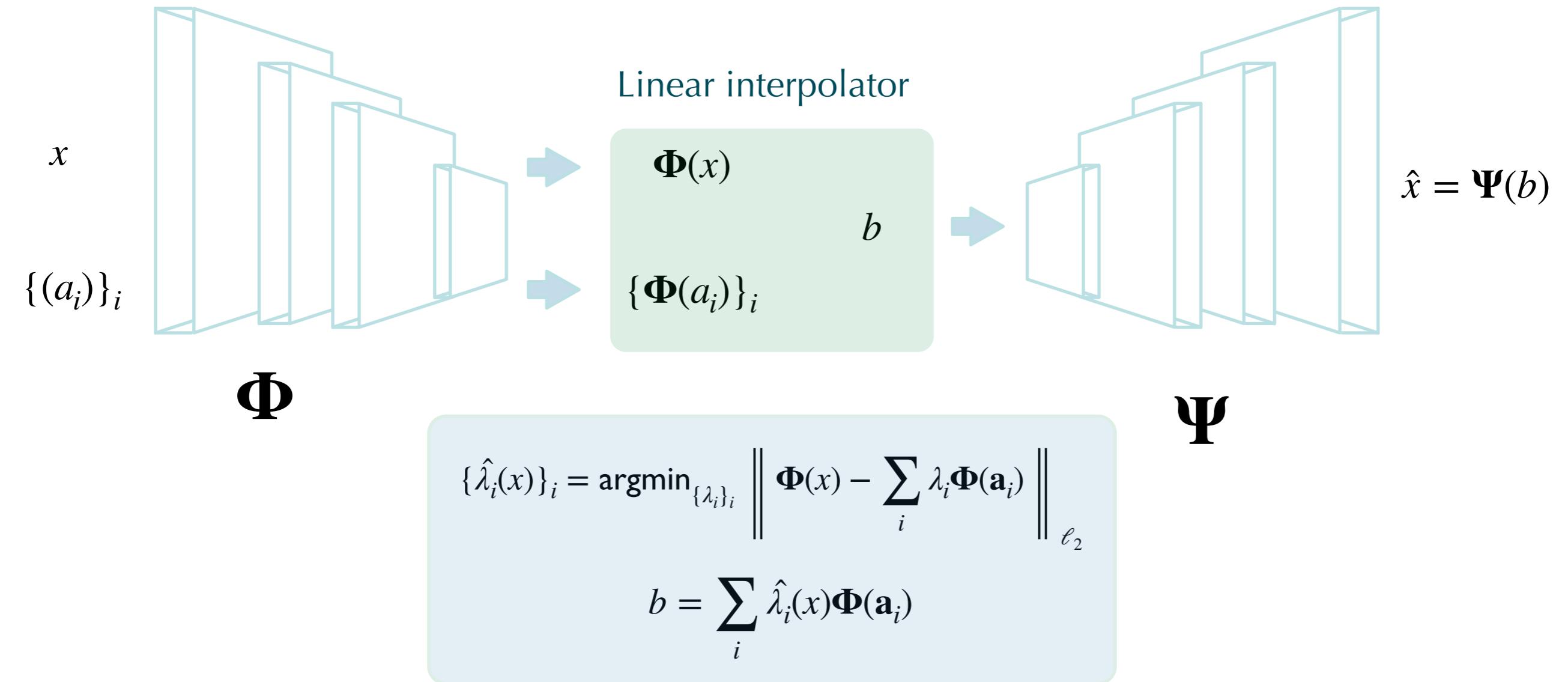
A non-linear interpolation scheme

A direct learning of the metric is a cumbersome task, we rather resort to a proxy:



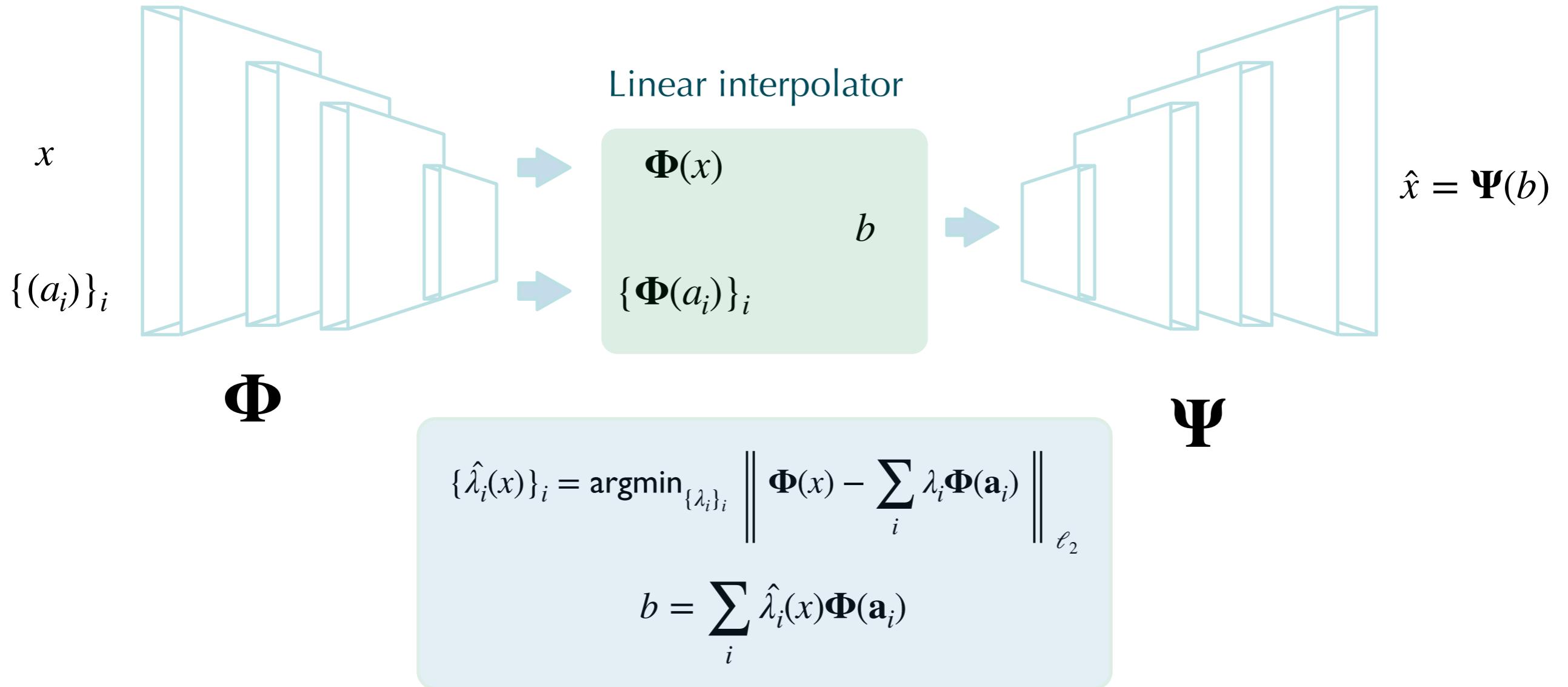
The interpolator AutoEncoder

Builds upon an auto encoder architecture:



The interpolator autoencoder

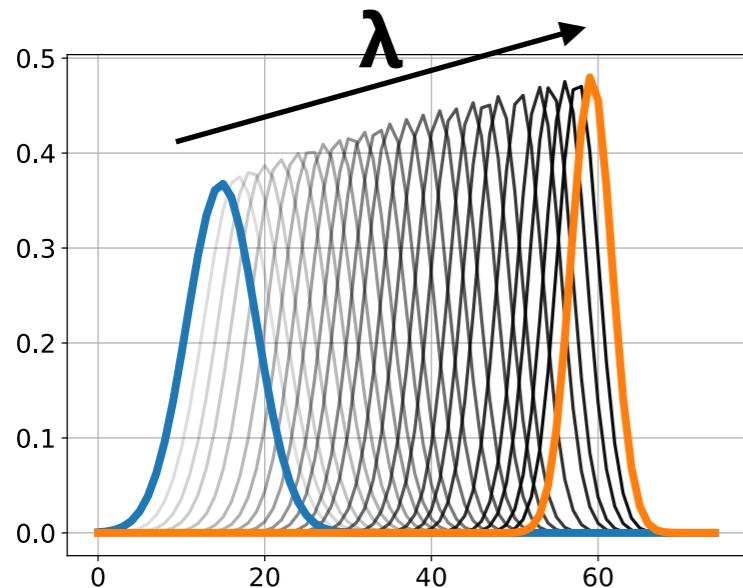
Builds upon an auto encoder architecture:



The parameters of the networks Φ and Ψ (e.g. *MLP*, etc.) are learnt by minimising the reconstruction error:

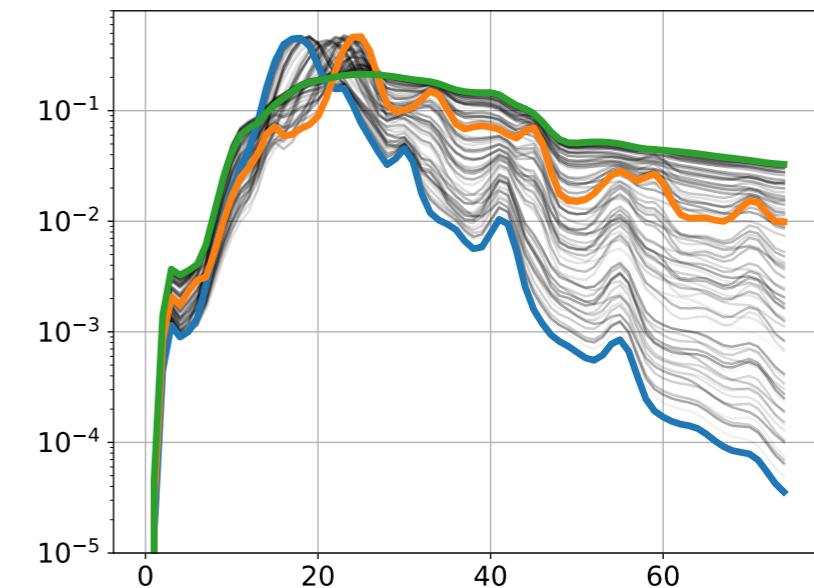
$$\min_{\Phi, \Psi} \sum_x \left\| x - \Psi \left(\sum_i \hat{\lambda}_i(x) \Phi(\varphi_i) \right) \right\|_{\ell_2}^2$$

The interpolator autoencoder

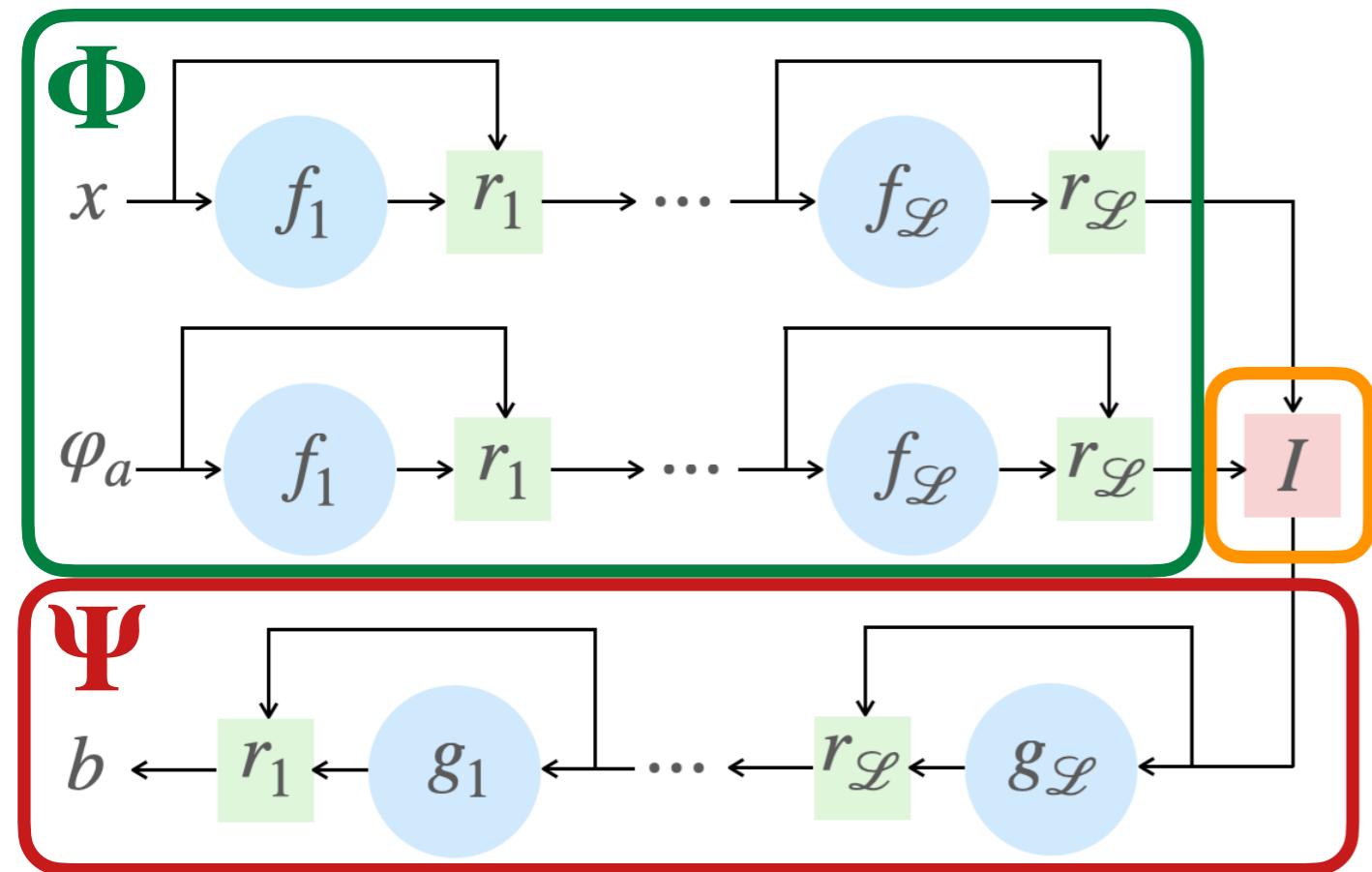


Simplistic emission lines as seen by Chandra

- × Element of the training set
- × Anchor points



Thermal emission lines as seen by Chandra



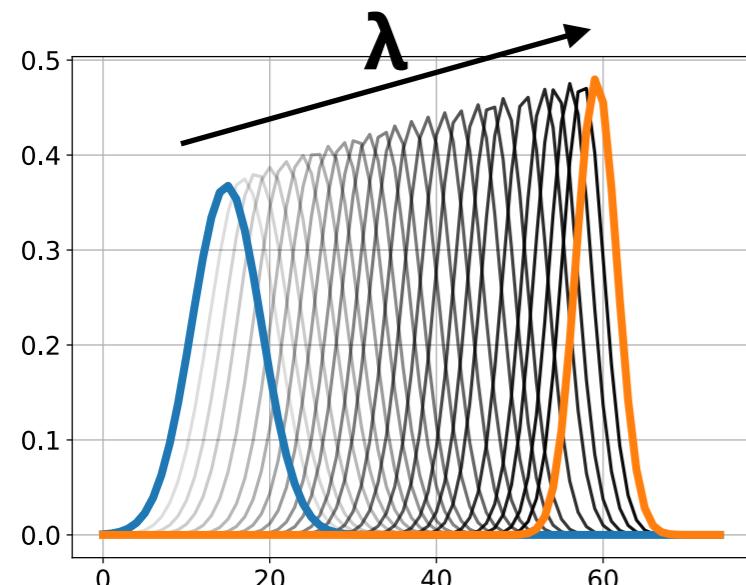
Φ and Ψ are modelled with feed-forward NN with residual short-cuts

$$h^{(l+1)} = f_l(\mathbf{W}^{(l)} h^{(l)} + z^{(l)}) + \rho_l h^{(l)}$$

$$b_{\Phi_\theta} = I(h^{(\mathcal{L}+1)}, h_{\varphi}^{(l+1)})$$

$$\tilde{h}^{(l+1)} = g_l(\tilde{\mathbf{W}}^{(l)} \tilde{h}^{(l)} + \tilde{z}^{(l)}) + \rho_l \tilde{h}^{(l)}$$

The interpolator autoencoder



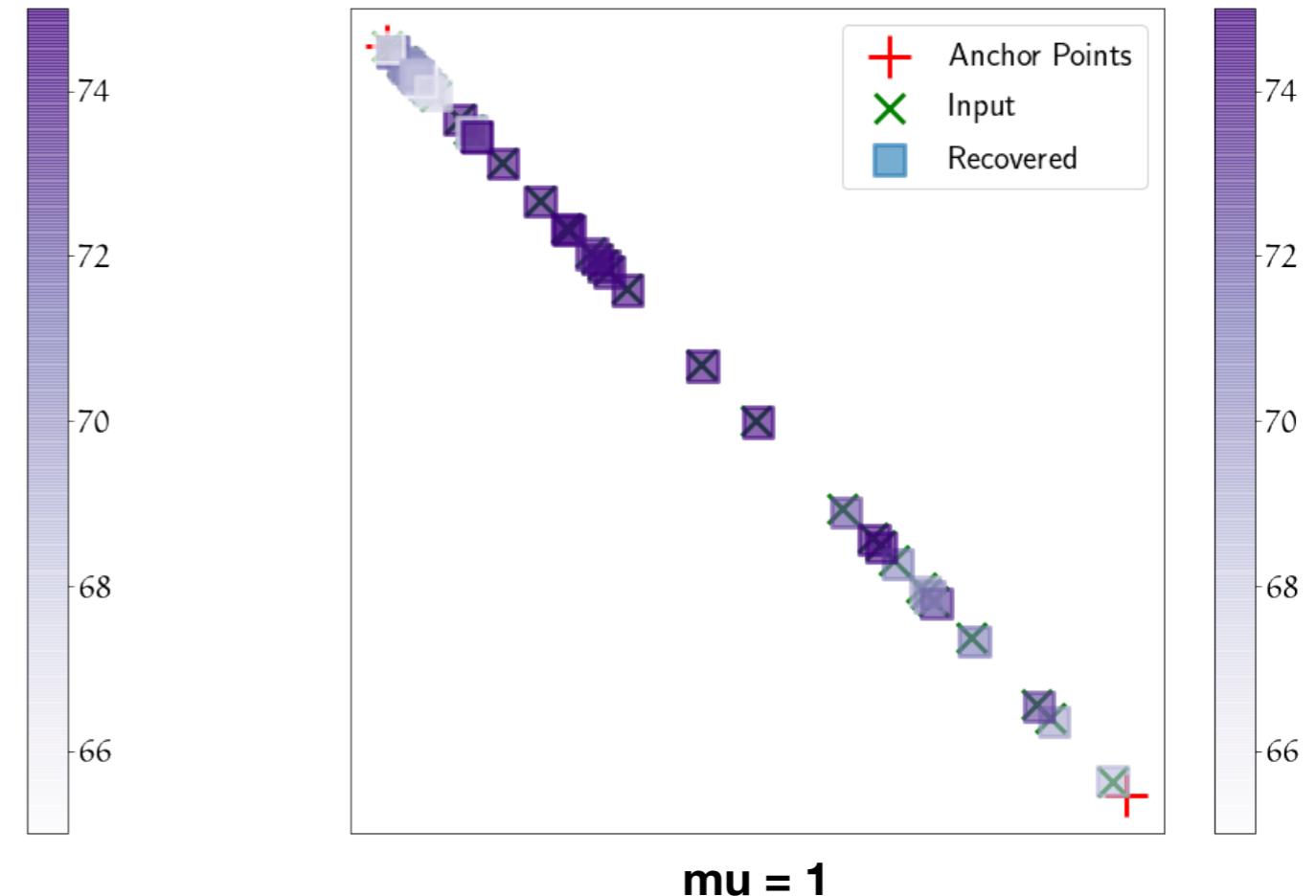
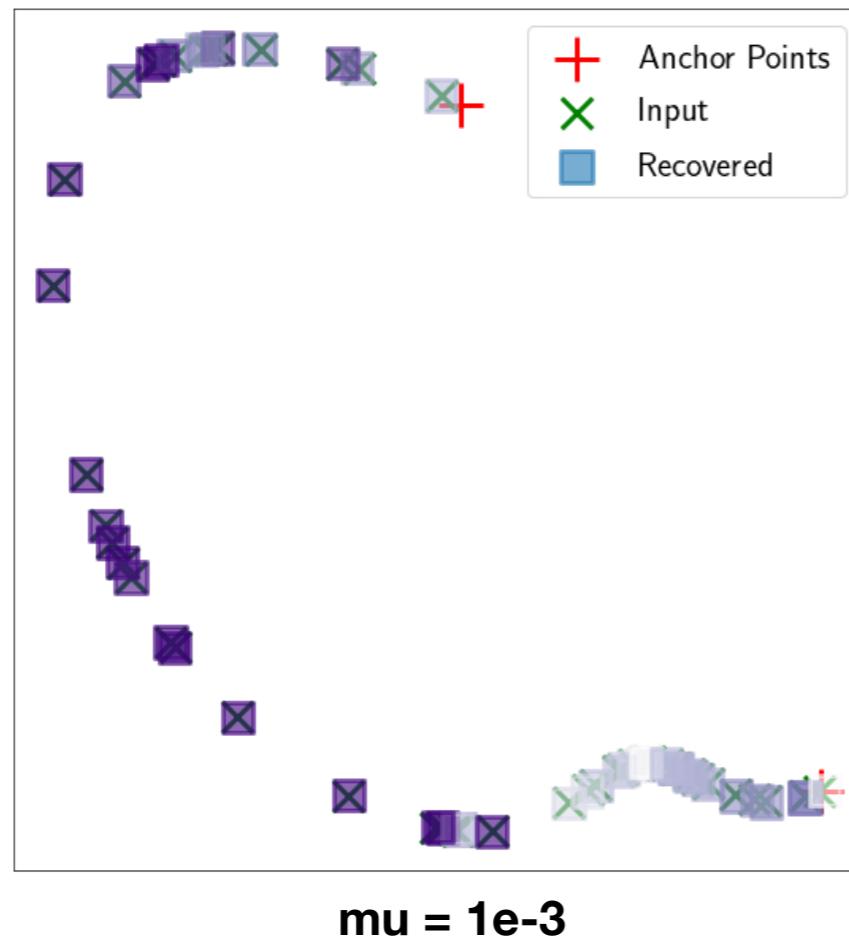
✗ Element of the training set

✗ Anchor points

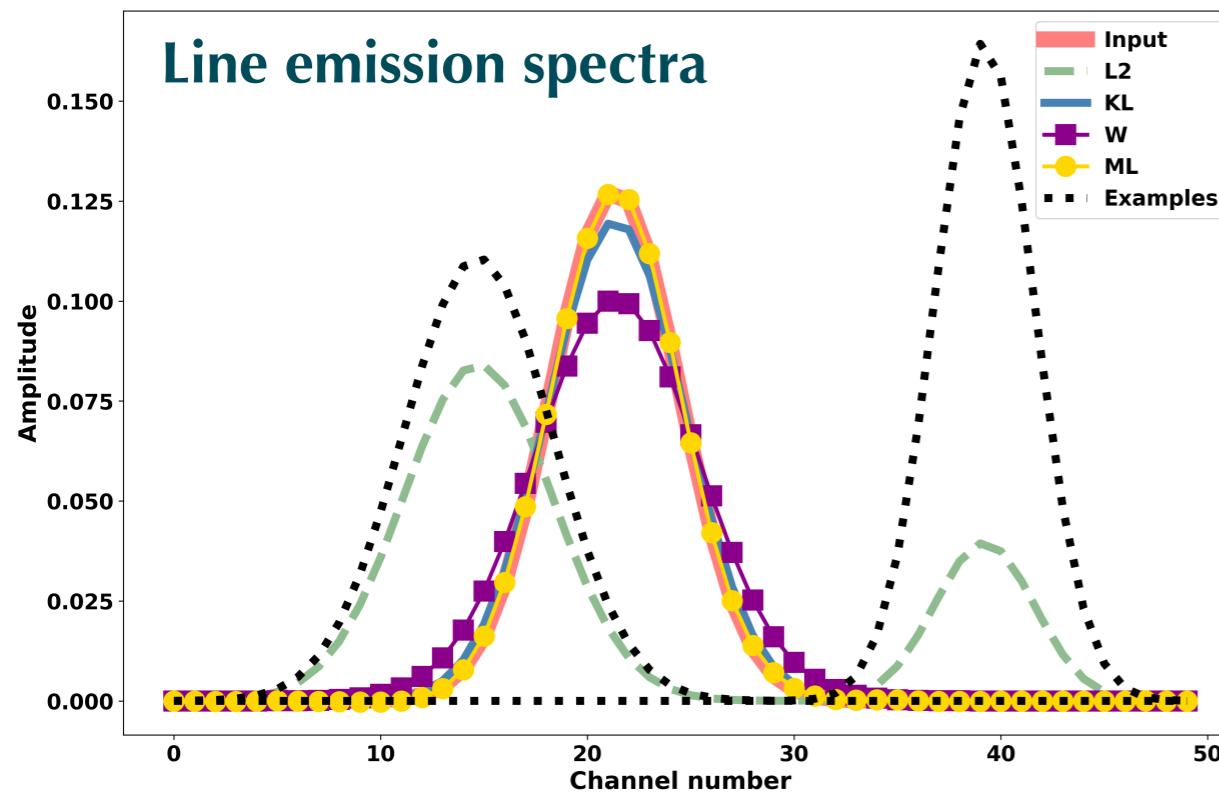
Simplistic emission lines as seen by Chandra

$$\min_{\Phi, \Psi} \sum_x \left\| x - \Psi \left(\sum_i \hat{\lambda}_i(x) \Phi(\varphi_i) \right) \right\|_{\ell_2}^2 + \mu \min_{\Phi, \Psi} \sum_x \left\| \Phi(x) - \sum_i \hat{\lambda}_i(x) \Phi(\varphi_i) \right\|_{\ell_2}^2$$

Add a regularisation term
In the latent domain



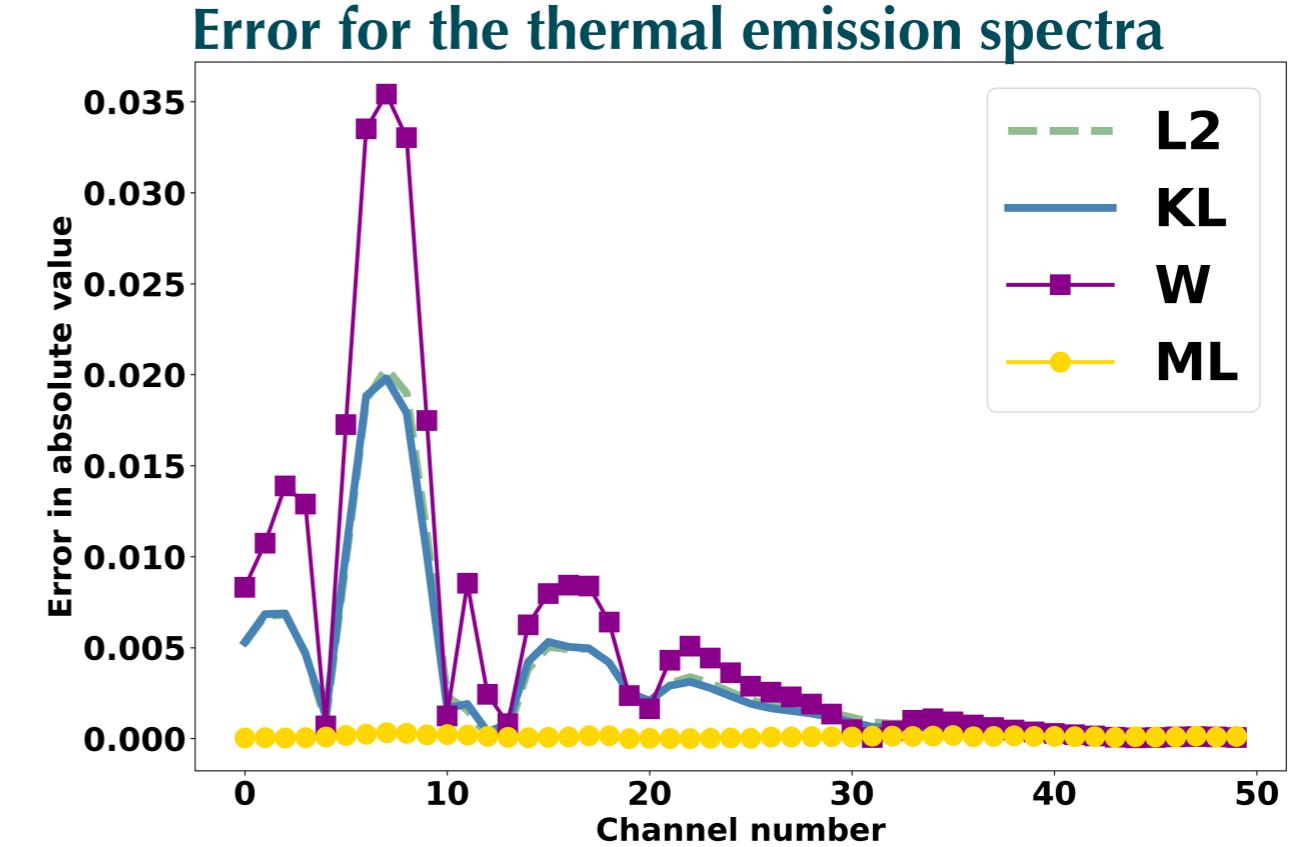
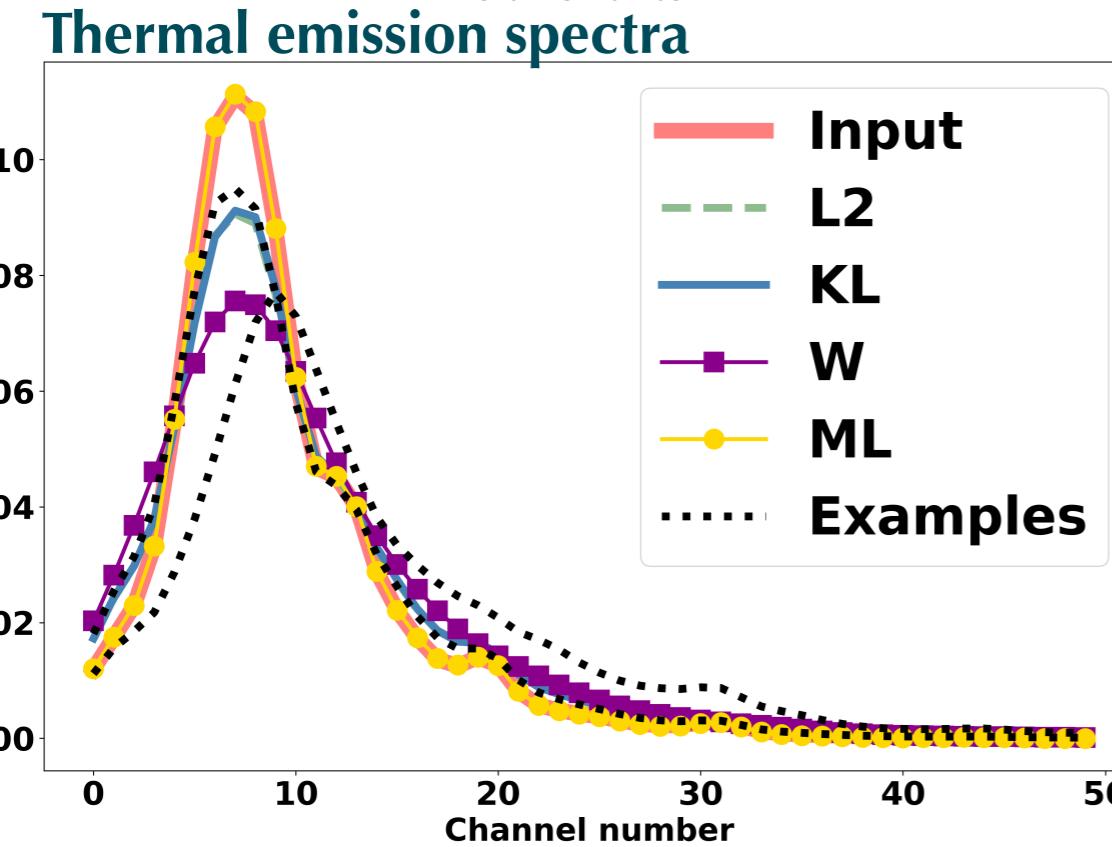
The interpolator autoencoder



Nb of layers

	$\mathcal{L} = 3$	4	5	Eucl.	Wass.	KL
G-shaped	6.7	39.3	4.6	-0.5	13.5	25.4
Thermal	40.5	47.2	21.8	13.5	10.6	13.7

MSE in dB of the reconstructed spectra



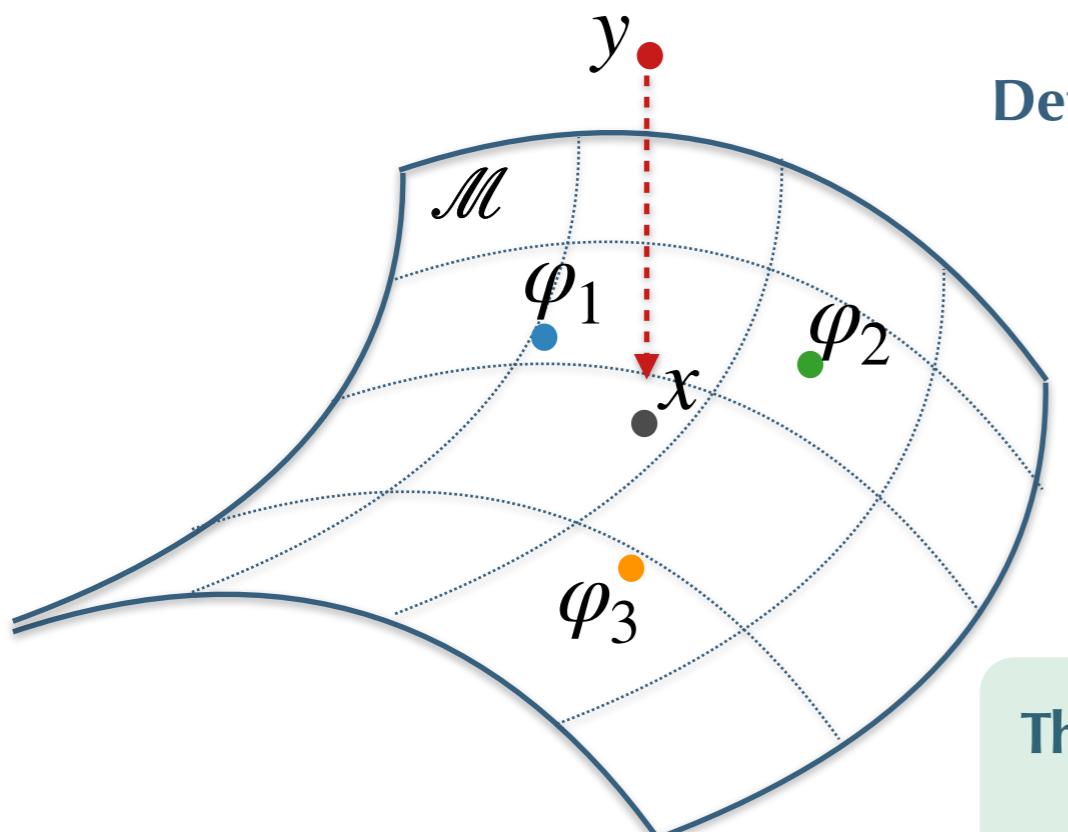
As a denoiser

Let's take a very simple denoising problem

$$\mathbf{y} = \mathbf{x} + \mathbf{n}$$

$$\min_{\mathbf{x}} \mathcal{R}(\mathbf{x}) + \frac{1}{2} \left\| \mathbf{y} - \mathbf{x} \right\|_2^2$$

Should enforce x to belong to \mathcal{M}



Defines the “orthogonal” projection onto the barycentric span:

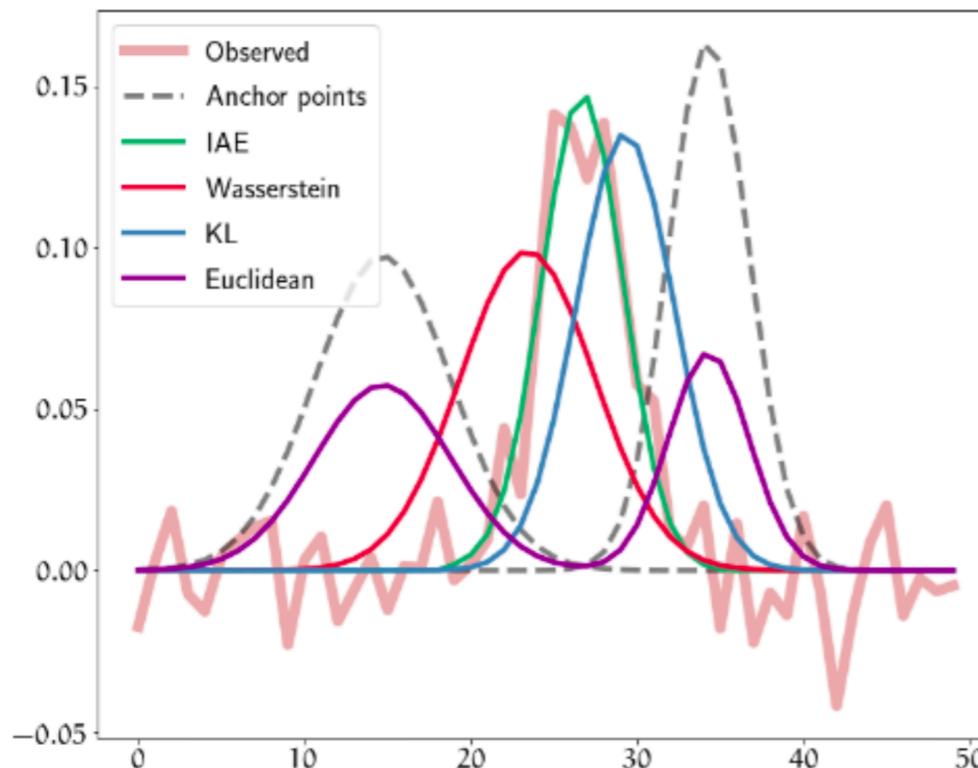
$$\mathbb{B}_\theta(\{\varphi_a\}) = \left\{ \mathbf{x}; \exists \{\lambda_a\} \in \mathcal{S}_d, \mathbf{x} = \Psi_\theta \left(\sum_a \lambda_a \Phi_\theta(\varphi_a) \right) \right\}$$

$$\{\hat{\lambda}_a(\mathbf{u})\}_a = \operatorname{Argmin}_{\lambda_a \in \mathcal{S}_d} \left\| \mathbf{u} - \Psi_\theta \left(\sum_a \lambda_a \Phi_\theta(\varphi_a) \right) \right\|_2^2$$

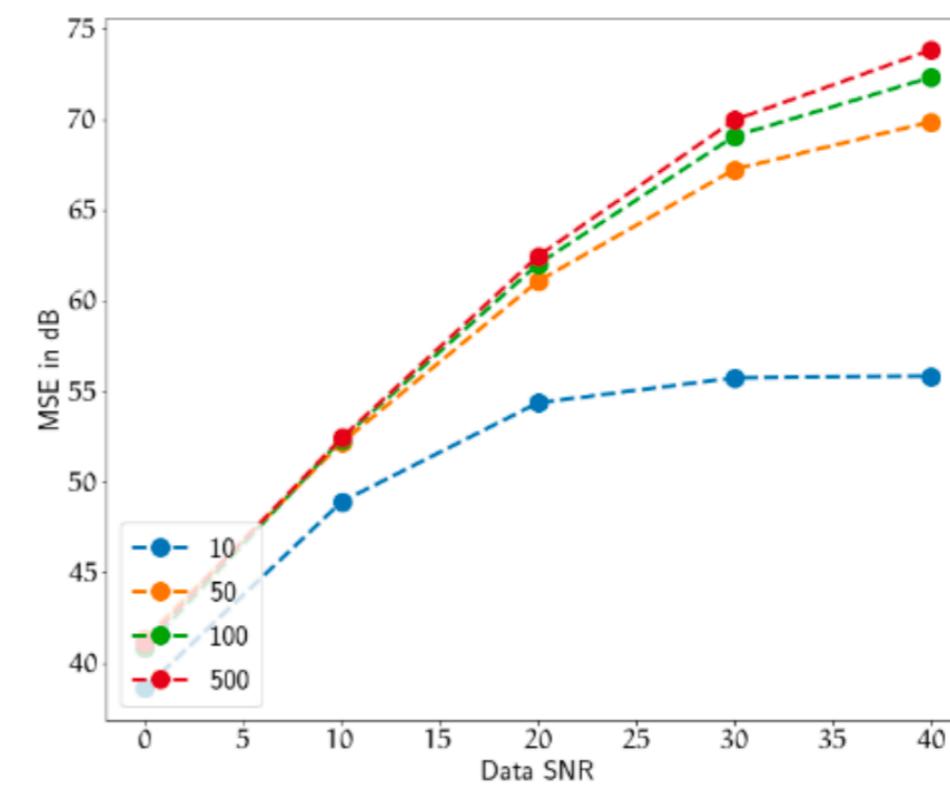
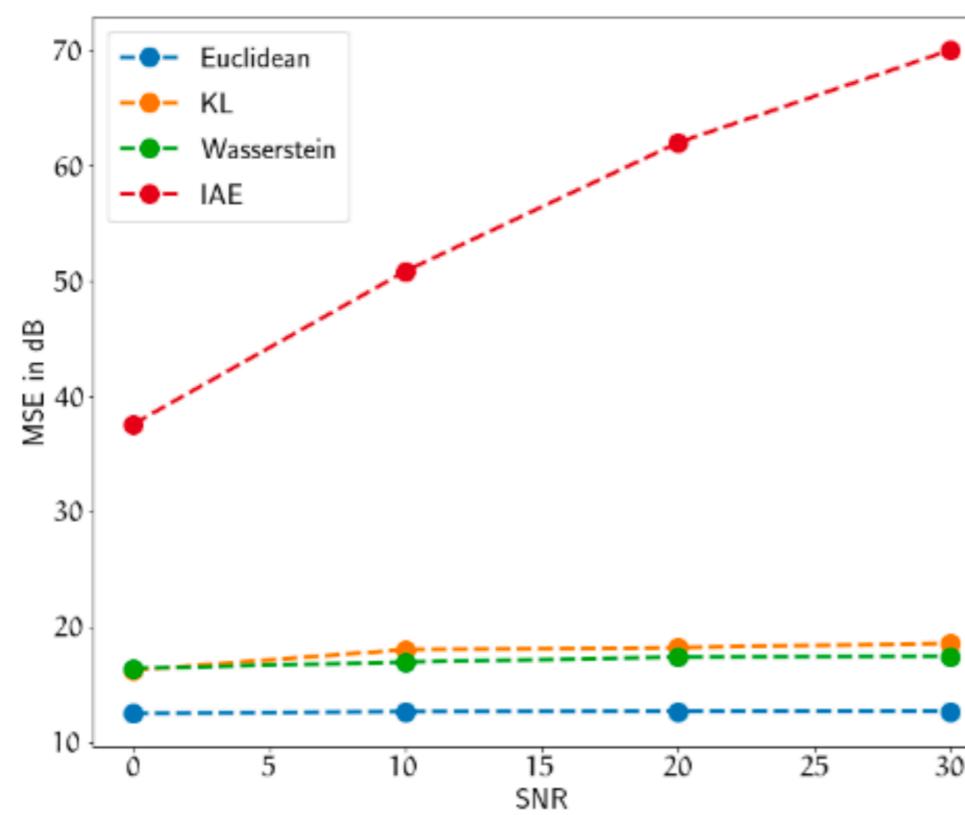
The projection/proximal operator write:

$$\operatorname{prox}_{\mathcal{R}}(\mathbf{u}) = \Psi_\theta \left(\sum_a \hat{\lambda}_a(\mathbf{u}) \Phi_\theta(\varphi_a) \right)$$

As a denoiser



Experiment with 5 layers
Gaussian-shaped signals
Varying noise level



Application to single-observation unmixing

- Industries



Radiation Portal Monitor
Screening for detection of illicit sources

- Health

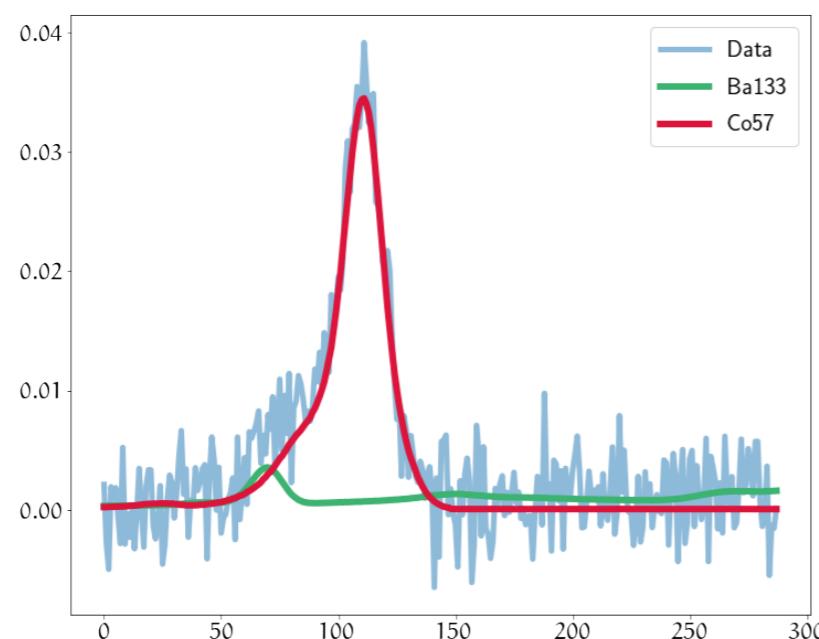


Dose calibrator with ionization chamber used in hospitals

- Environment



Spectrometric probe gamma contamination monitoring in air/soil/water



Example of gamma-ray spectrum

Application to single-observation unmixing

Single observation of a gamma-ray spectrum (*3" x 3" NaI scintillation detector*)

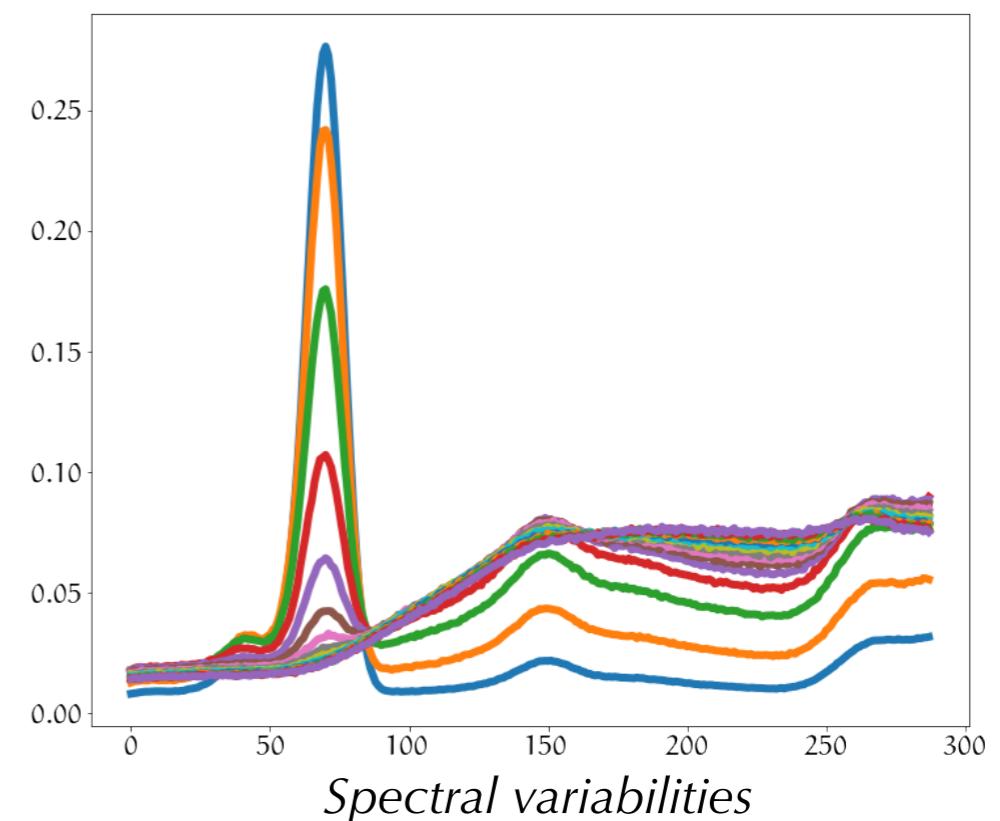
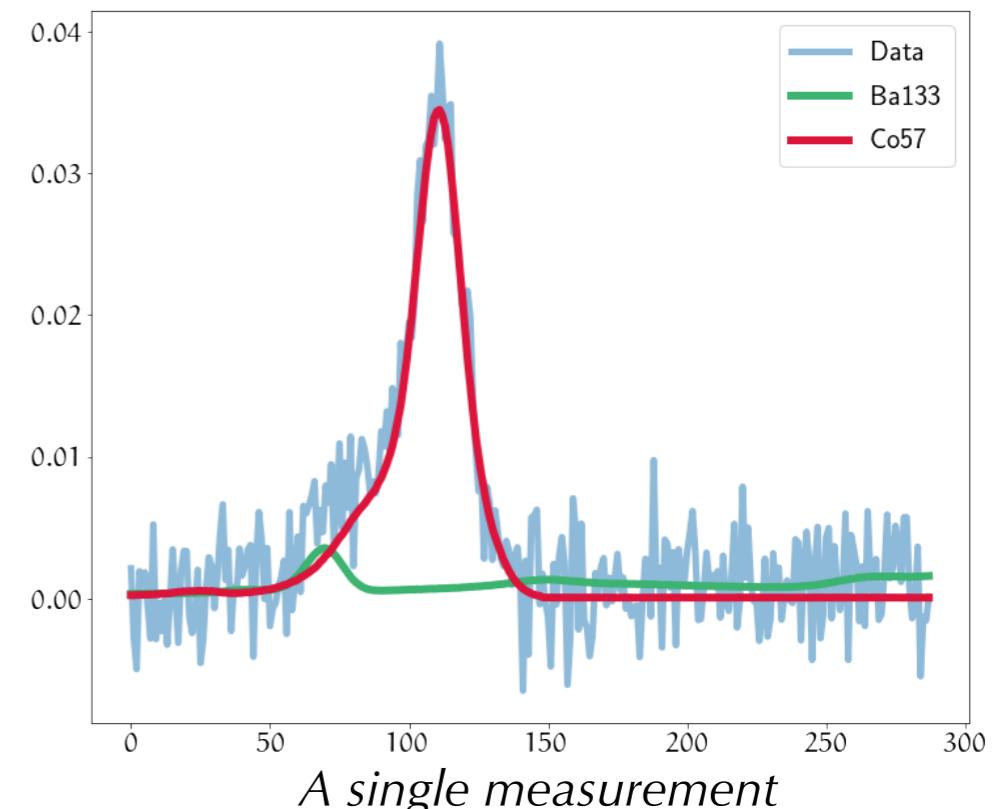
Linear mixture of radionuclides (*here ^{57}Co and baryum ^{133}Ba*)

Challenges:

- High noise level
- Variability of the spectral signatures

$$\min_{\mathbf{x}, \mathbf{A}} \mathcal{R}(\mathbf{x}) + \frac{1}{2} \left\| \mathbf{y} - \mathbf{Ax} \right\|_2^2$$

**Ill-posed/non-convex
inverse problem**



Application to single-observation unmixing

Build an IAE model for each radionuclide from only 15 samples

$$\min_{\mathbf{x}, \mathbf{A}} \chi_{\{\cdot \geq 0\}}(\mathbf{x}) + \sum_{i=1}^p \chi_{\mathbb{B}_{\phi_i}}(\mathbf{A}^i) + \frac{1}{2} \left\| \mathbf{y} - \mathbf{Ax} \right\|_2^2$$



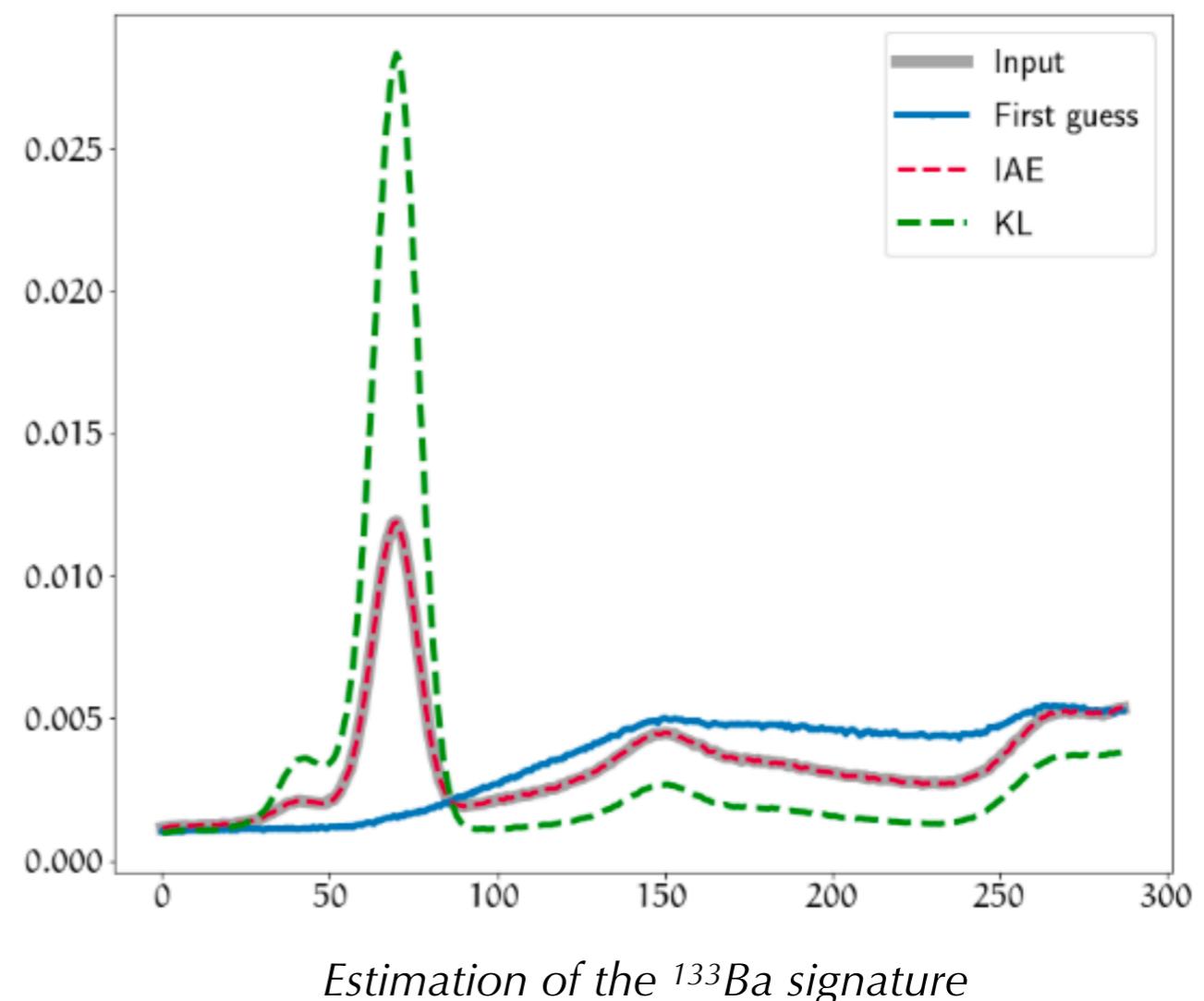
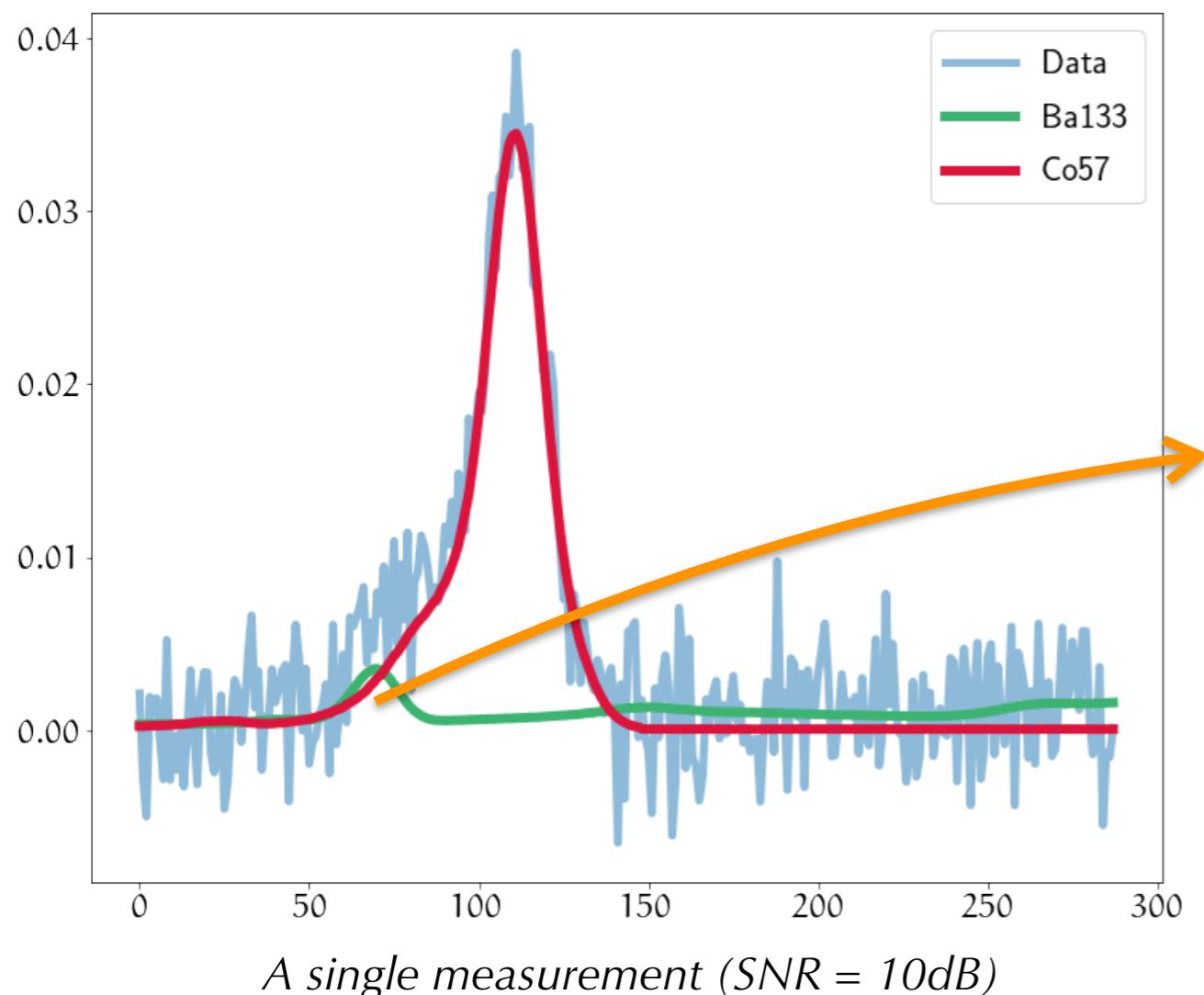
*Imposes the
non-negativity
of x*



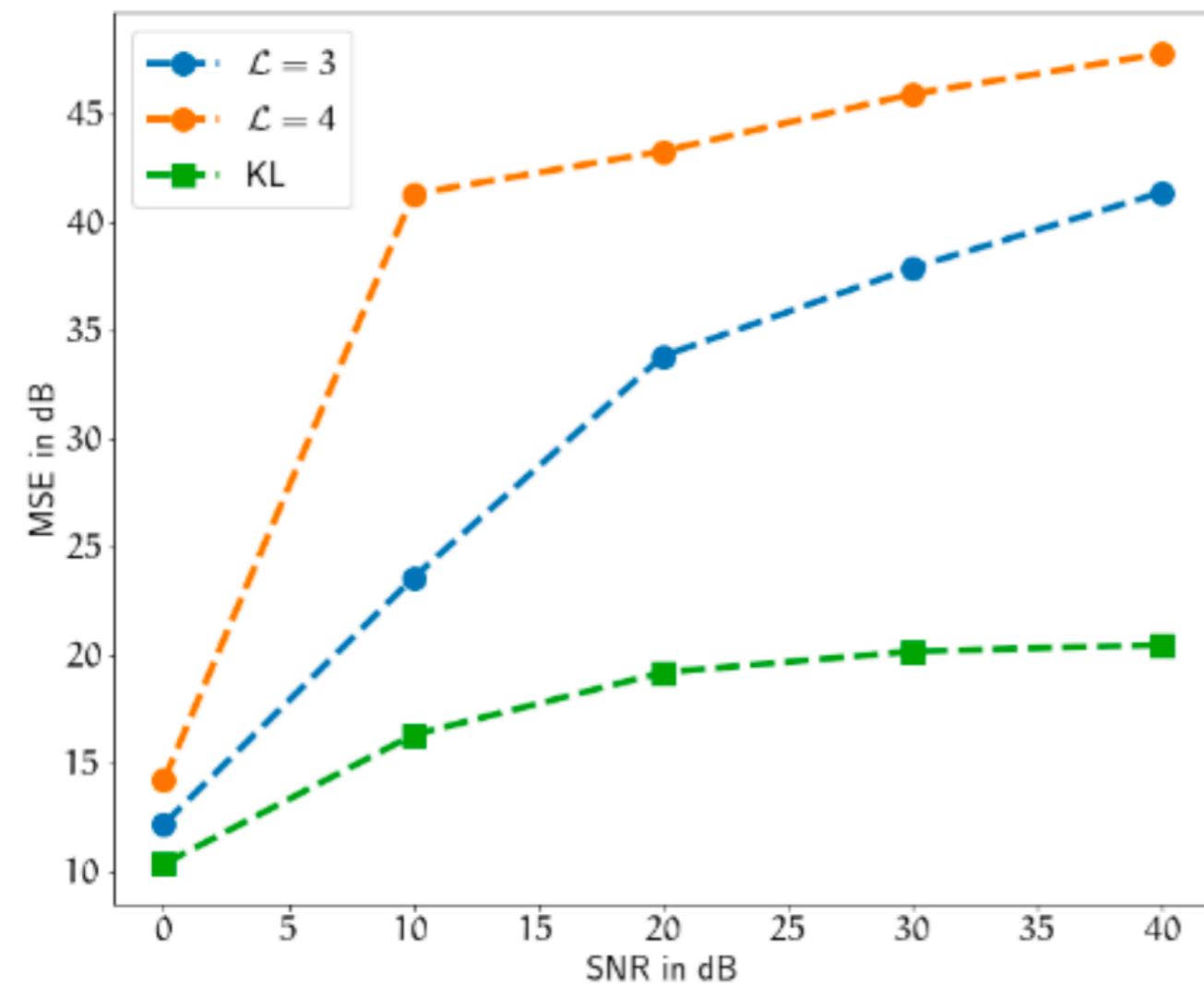
*Characteristic
function related
to each model*

**Non-convex/multi-convex
optimisation problem**

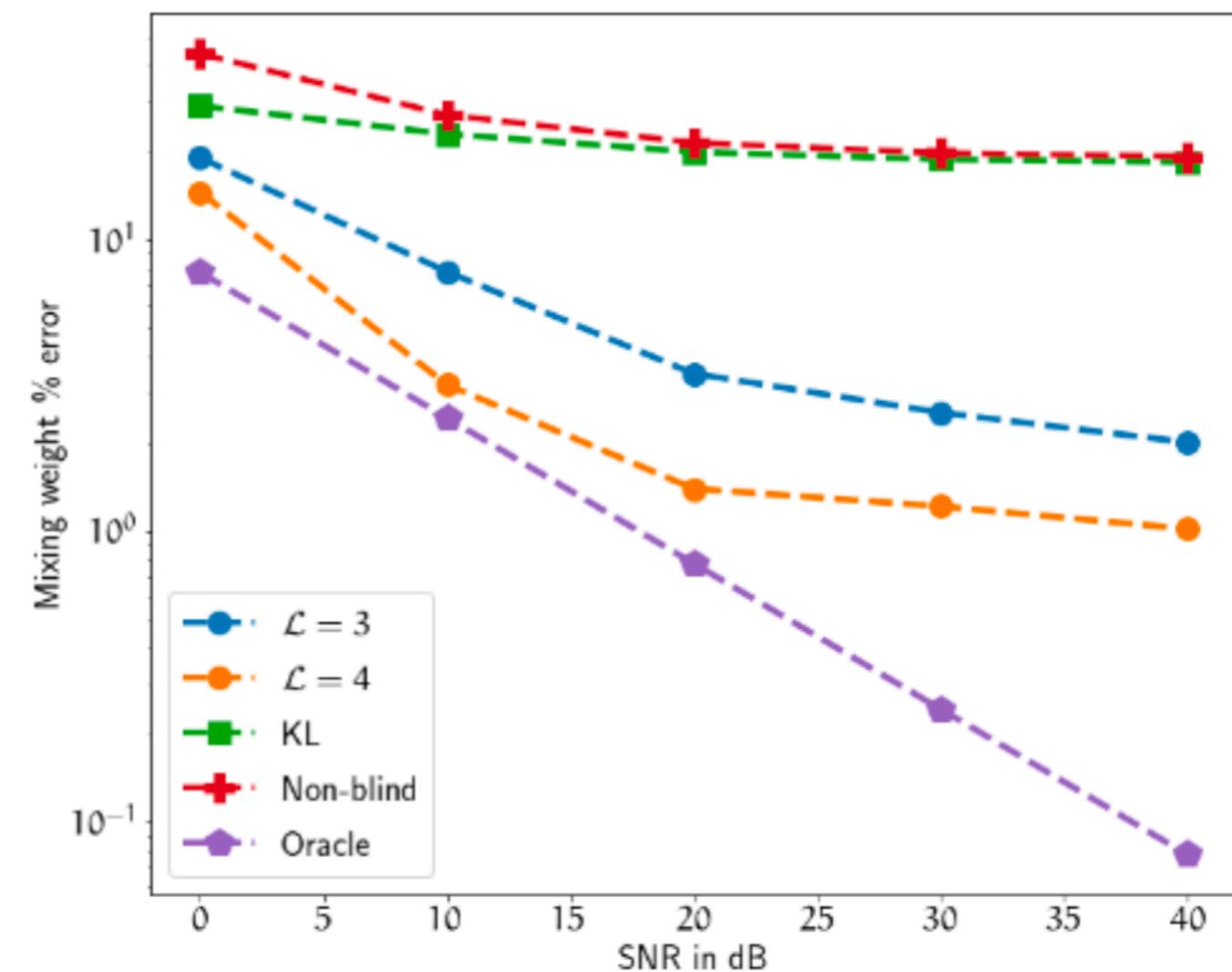
Application to single-observation unmixing



Application to single-observation unmixing



Spectral signature estimation

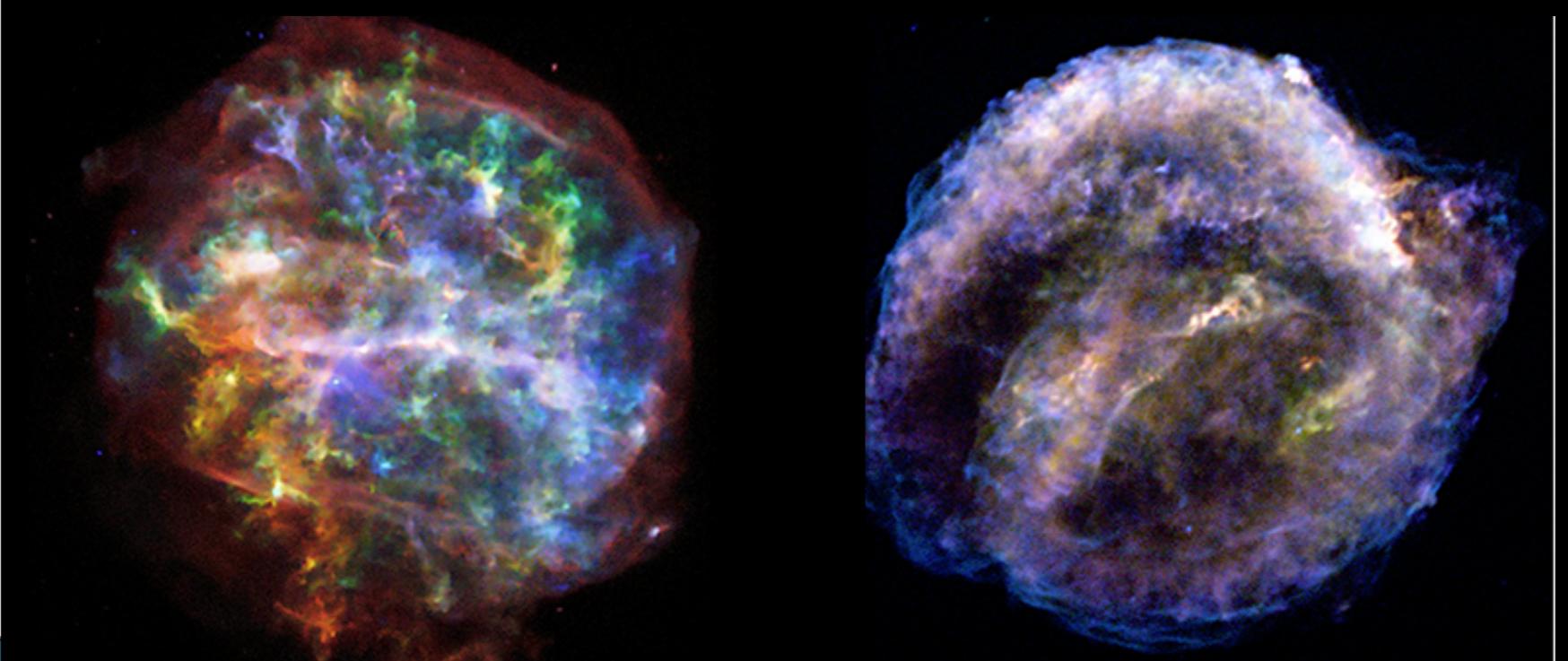


Mixing weight estimation

Application to BSS in X-ray imaging



Supernova remnants as seen in X-rays



Semi-blind sparse BSS

sGMCA : a semi-blind sparse unsupervised matrix factorization method

$$\operatorname{argmin}_{\mathbf{A}, \mathbf{S}} \|\Lambda \odot \mathbf{SW}\|_1 + \sum_{j \neq \mathcal{J}} \iota_{\mathcal{S}_m}(\mathbf{A}^j) + \sum_{j \in \mathcal{J}} \iota_{\mathcal{B}_{\phi_j}(\{\varphi_{\mathcal{M}_j}^e\})}(\mathbf{A}^j) + \frac{1}{2} \|\mathbf{X} - \mathbf{AS}\|_F^2$$

↑ ↑ ↑ ↑
Sparsity regularization *Oblique constraint* *Barycentric span constraint* *Data fidelity*

- With the exception of the barycentric constraint, the problem is multi-convex
- In practice, the barycentric constraint seems to behave like a convex constraint
- For its robustness, the minoration scheme is based on an alternate least-squares

Semi-blind sparse BSS

Pseudo-code for a single iteration

Update of S

(1 – a) Least-square estimation: $\mathbf{S}' = \mathbf{A}^{(k)^+} \mathbf{X}$

(1 – b) Thresholding : $\mathbf{S}^{(k+1)} = \mathcal{T}_\Lambda(\mathbf{S}' \mathbf{W}) \mathbf{W}^T$

(2 – a) Least-square estimation : $\mathbf{A}' = \mathbf{X} \mathbf{S}^{(k+1)^+}$

(2 – b) $\forall j \in \mathcal{J}, \mathbf{A}^{j(k+1)} = P_{\mathcal{B}_{\phi_j}(\{\varphi_{\mathcal{M}_j}^e\})}(\mathbf{A}'^j)$ *Projection onto the barycentric span*

$$\{\hat{\lambda}_e\}_e = \operatorname{argmin}_{\{\lambda_e\}_e} \left\| \Phi(\mathbf{A}'^j) - \sum_e \lambda_e \Phi(\varphi_e) \right\|_{\ell_2}$$

Find the barycentric weights

Update of A

$$\mathbf{A}^j = \Psi \left(\Phi \left(\sum_e \hat{\lambda}_e \Phi(\varphi_e) \right) \right)$$

Reconstruct

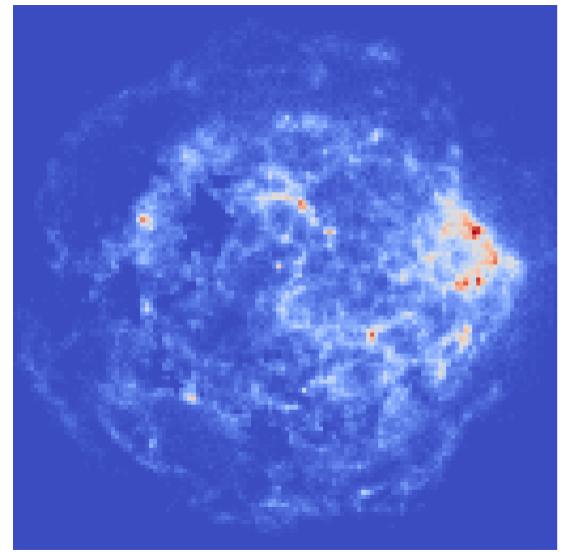
(2 – c) Normalize all columns of \mathbf{A}

Stopping criterion : relative variation on \mathbf{A} < $1e^{-6}$

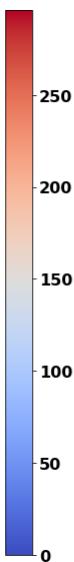
The choice of the thresholds is based on the noise level (Bobin et al. 2015)

For the next experiments, stabilisation is reached in less than 1000 iterations

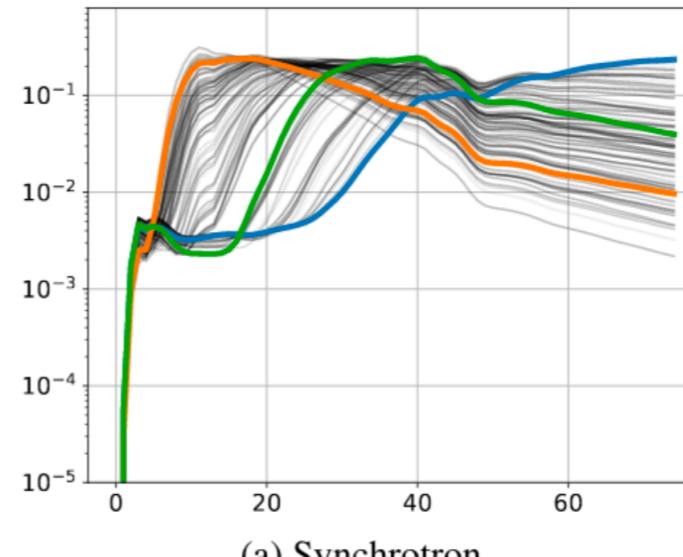
Semi-blind sparse BSS



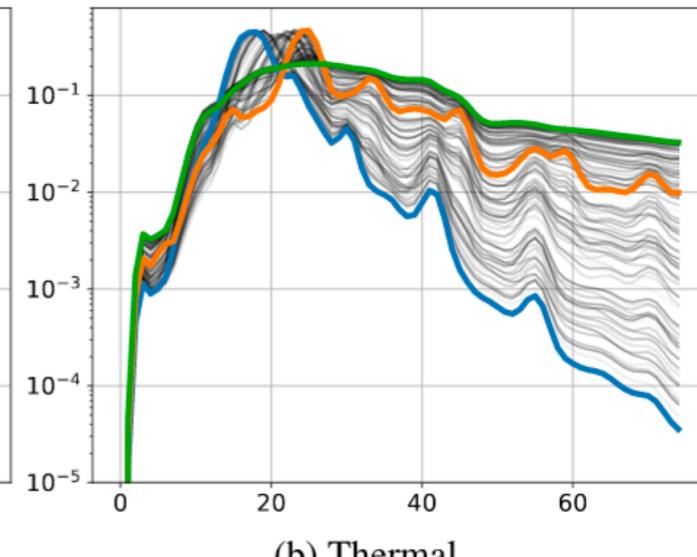
Observation at 5548eV



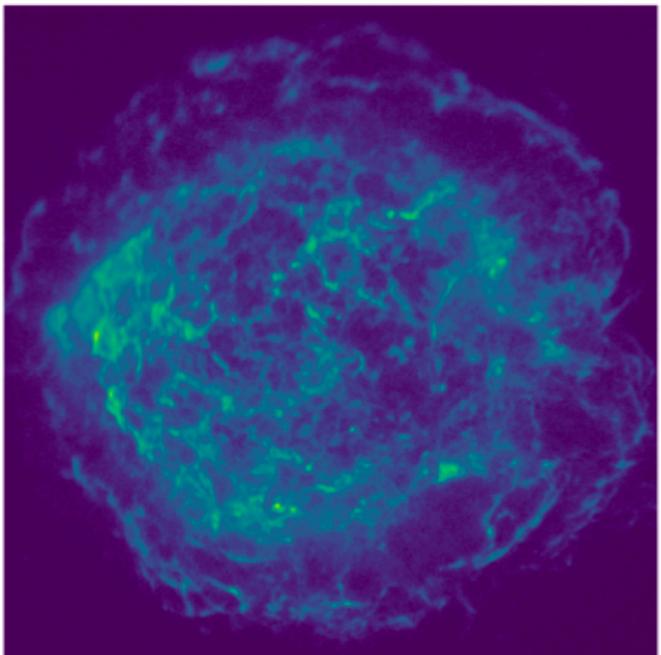
50 observations
128x128 pixels
128x128 pixels
4 sources



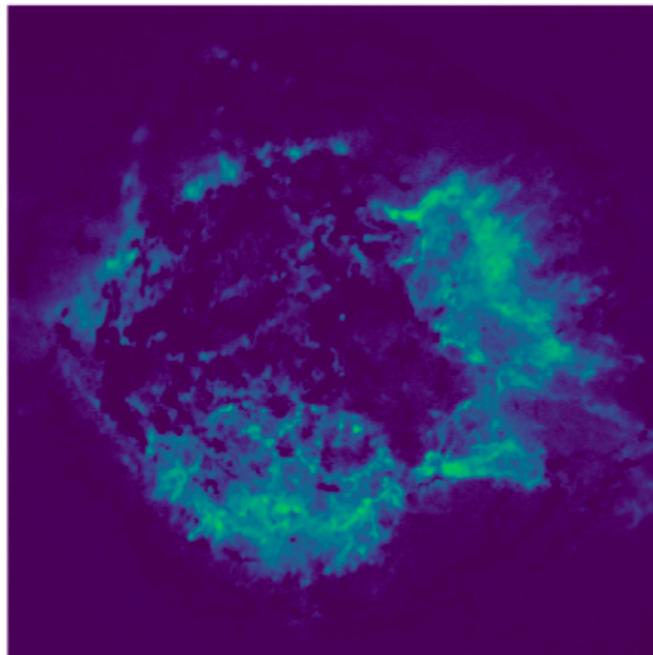
(a) Synchrotron



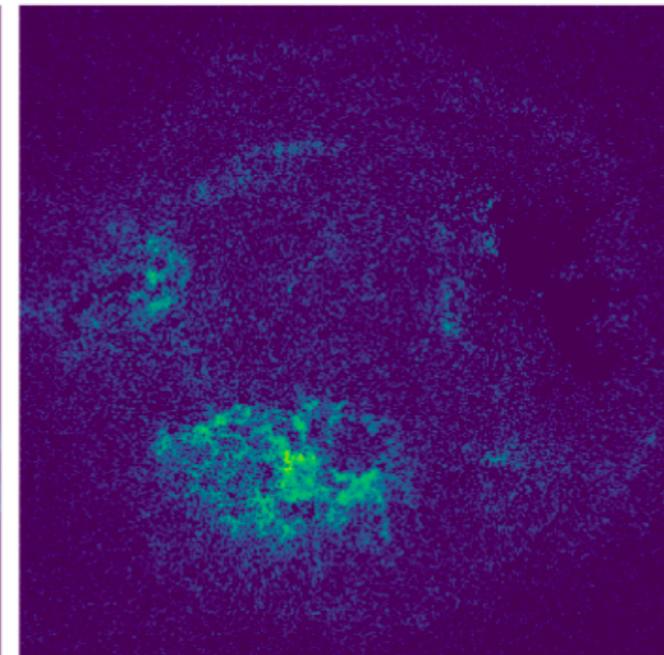
(b) Thermal



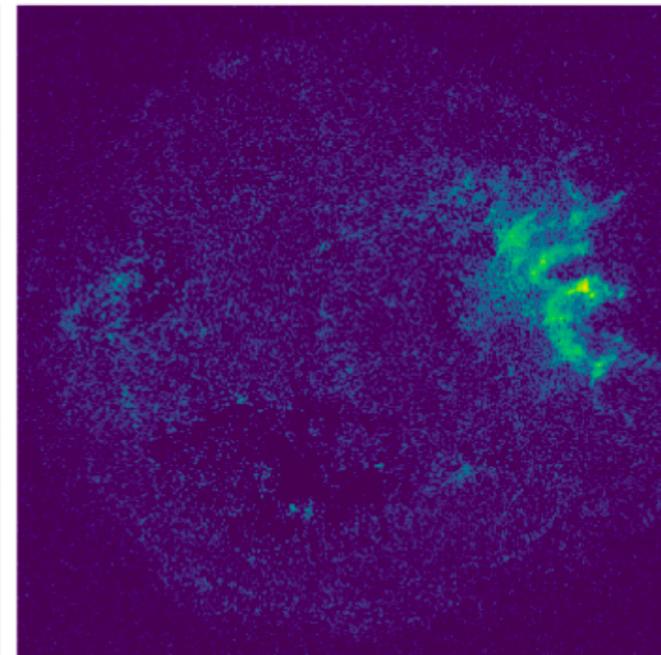
(a) Synchrotron



(b) Thermal



(c) Gaussian I

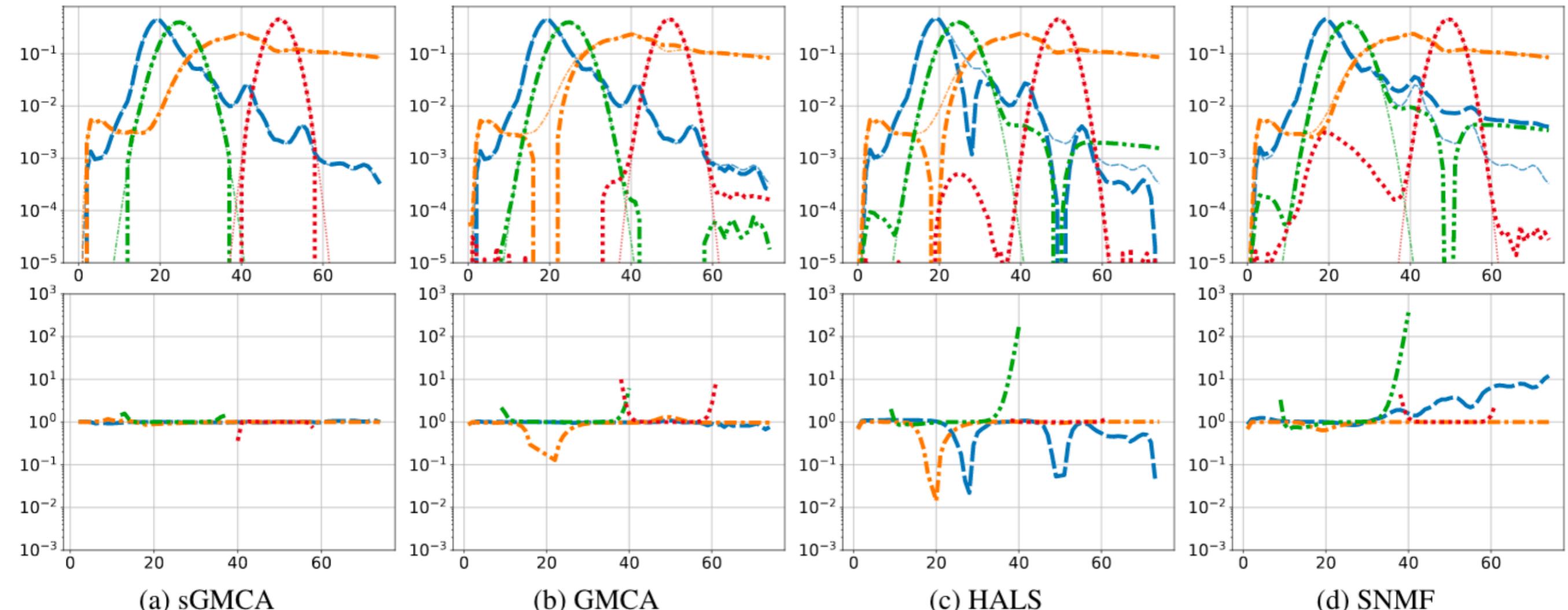


(d) Gaussian II

These simulations are representative of supernovae remnants in the range 5keV - 6keV

Semi-blind sparse BSS

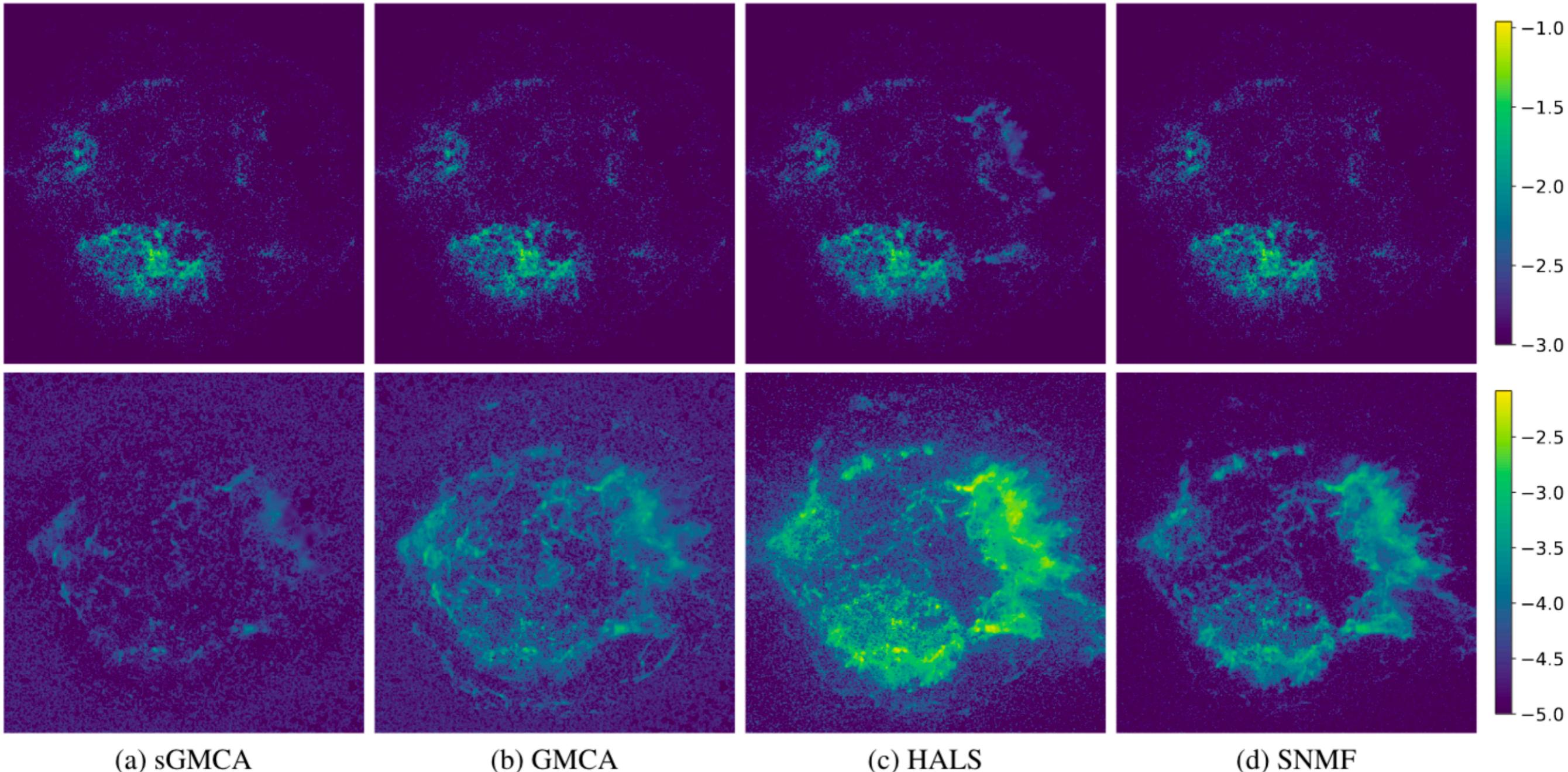
Examples of reconstructed spectra



Residual errors

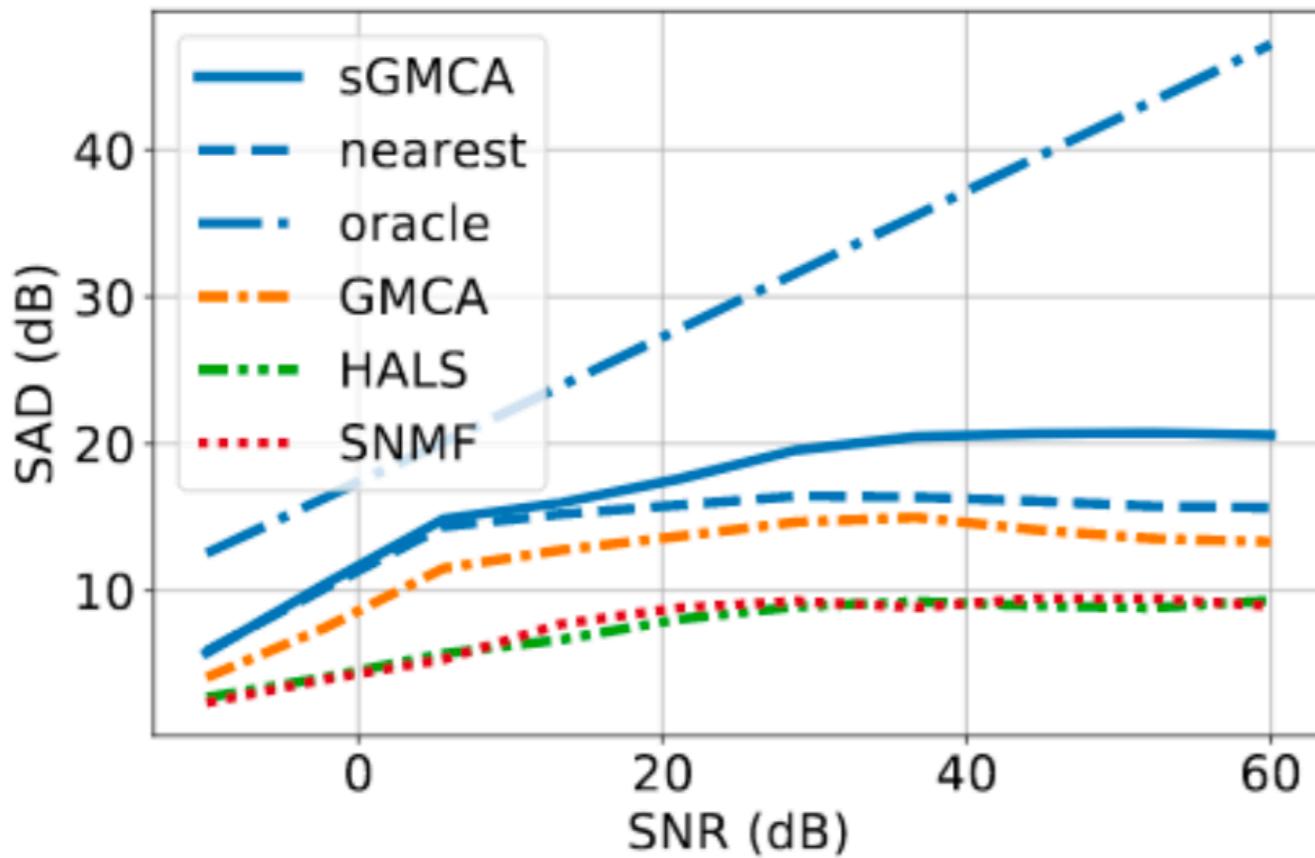
Semi-blind sparse BSS

Reconstructed thermal emission



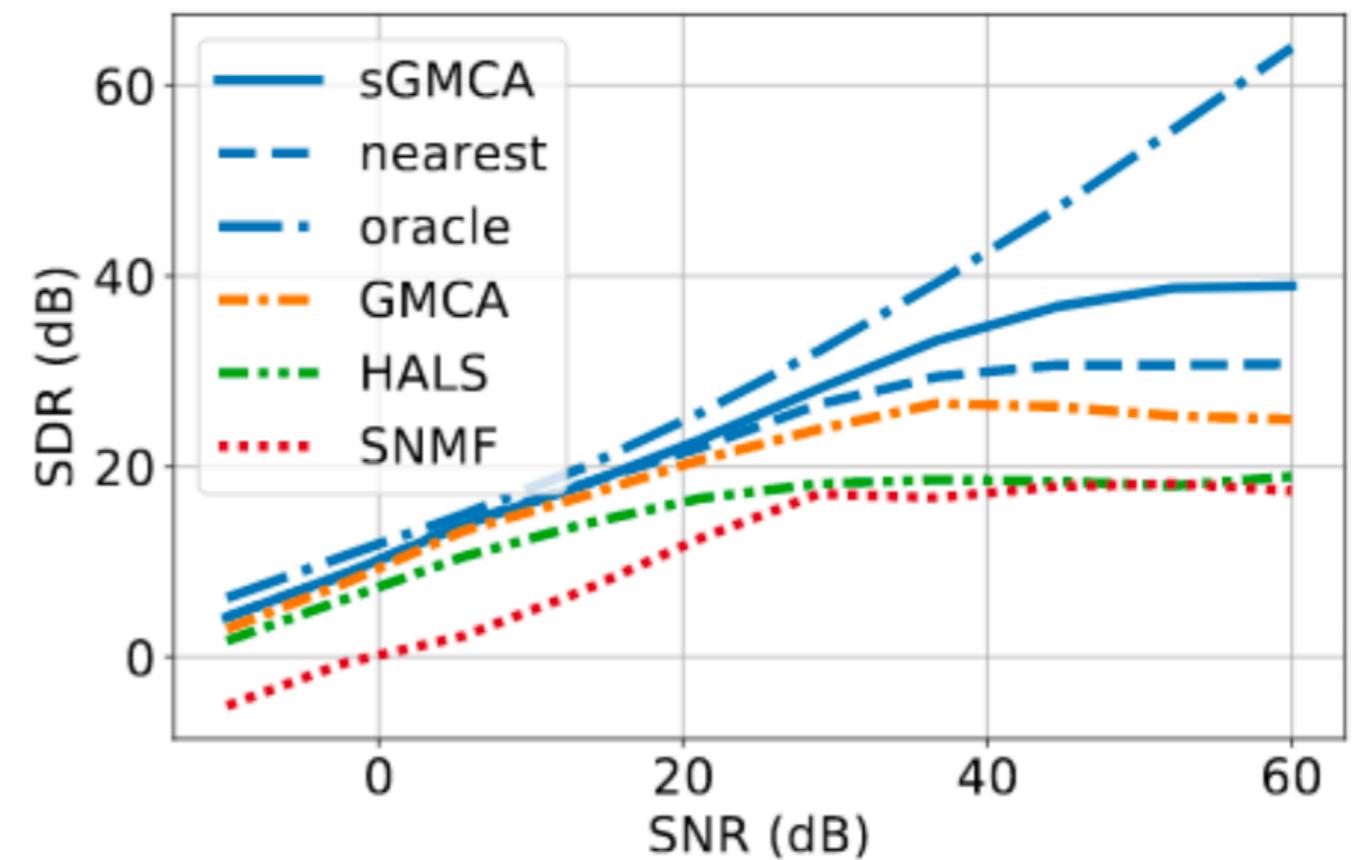
Residual error

Semi-blind sparse BSS



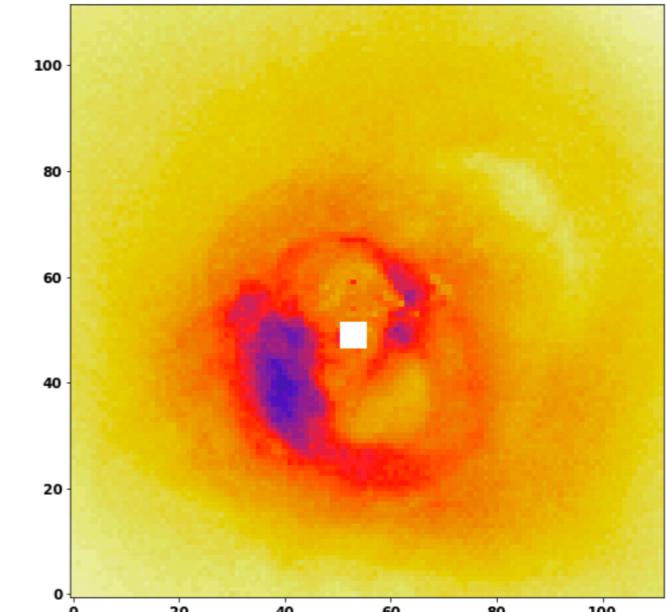
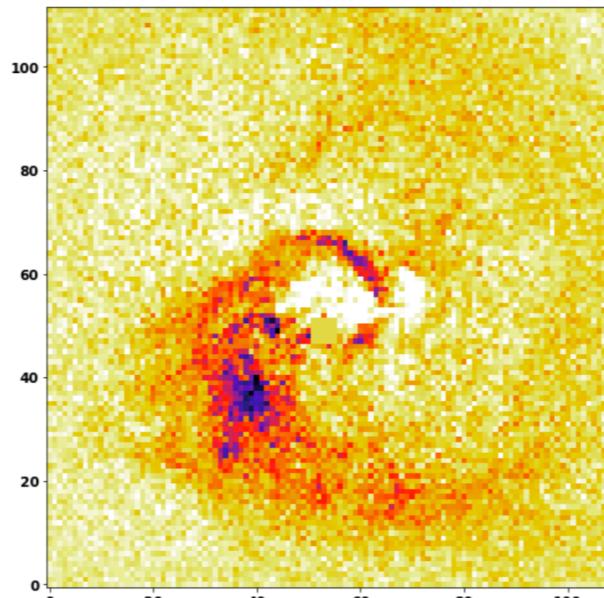
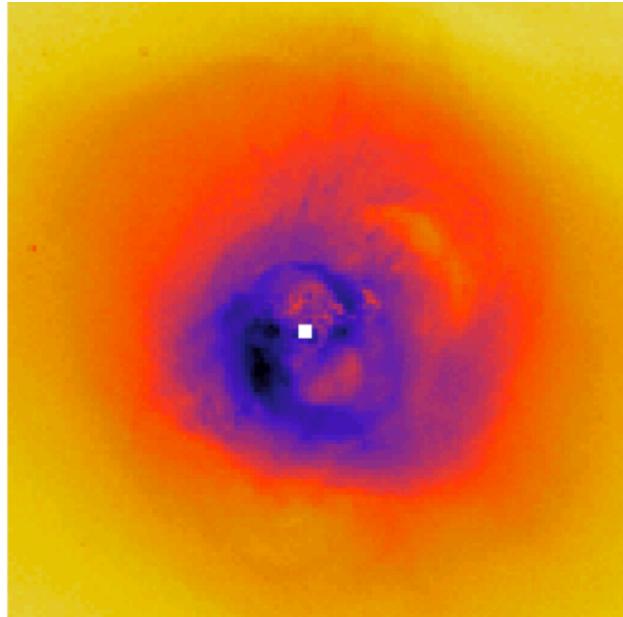
SAD: Spectral Angle Distance
Measures the mixing matrix estimation quality

SDR: Signal-to-Distortion Ratio
Measures the source's estimation quality

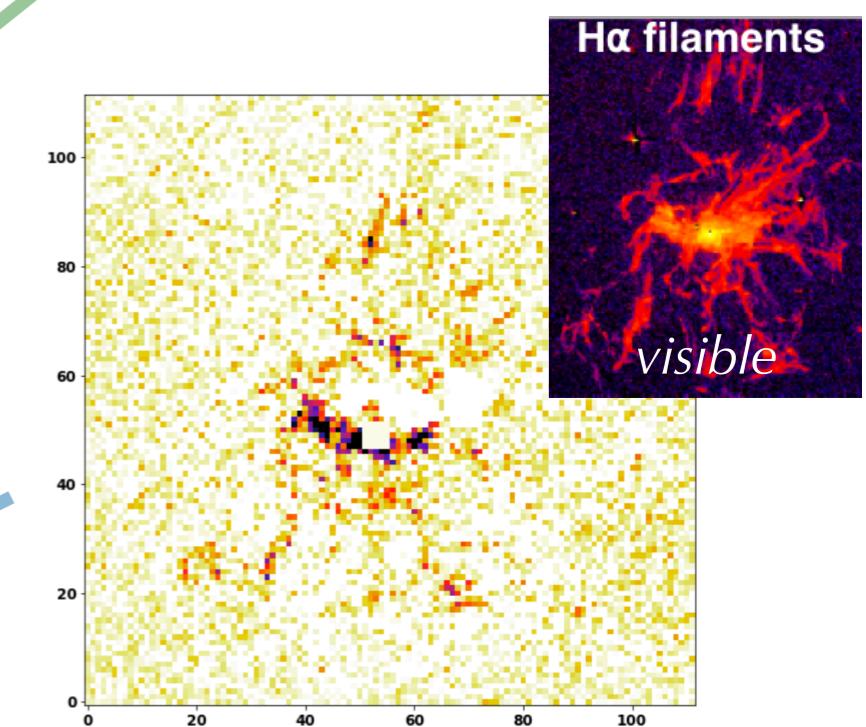
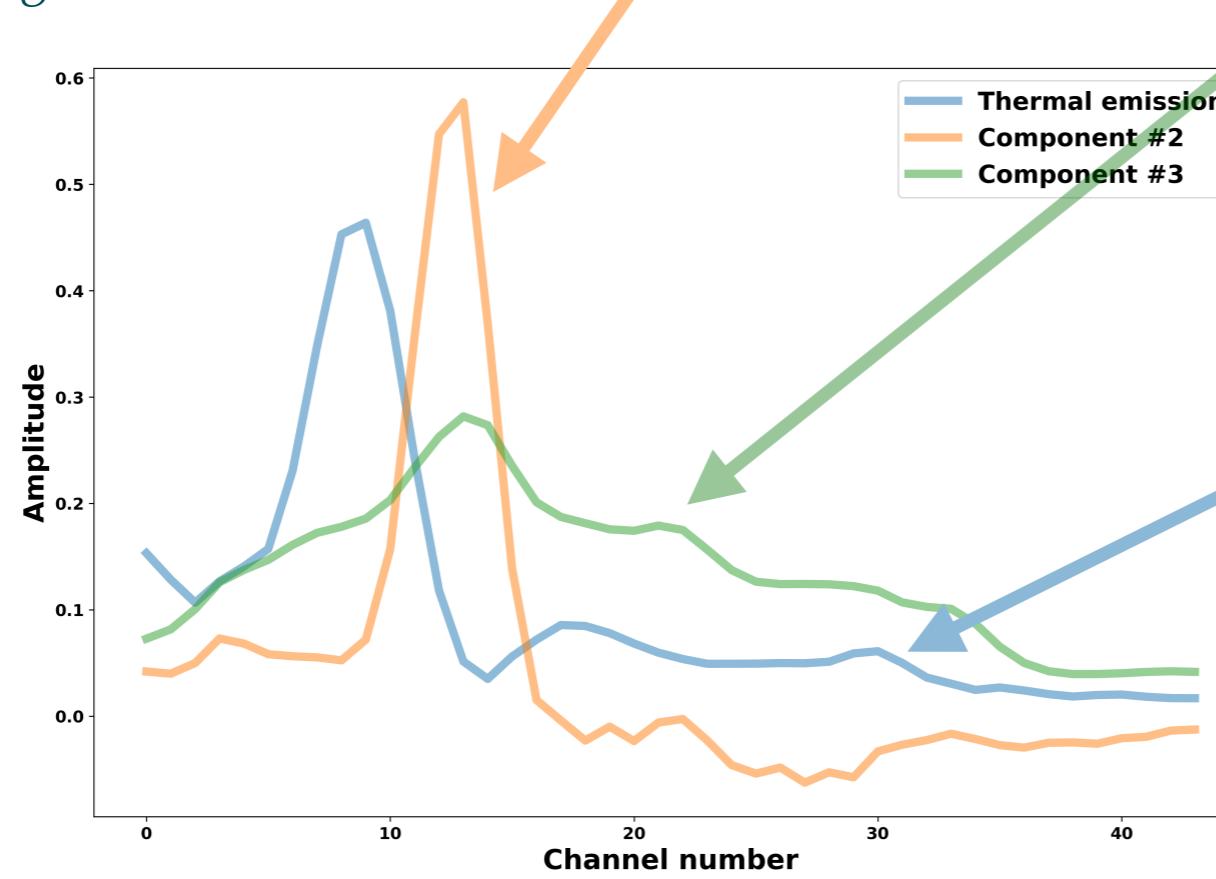


Semi-blind sparse BSS

Perseus galaxy cluster



Integrated image 0.5 - 8keV



X-ray filaments have $\sim 50\text{-}100$ counts
buried under 10^4 counts
Finding features with contrast $< 1\%$