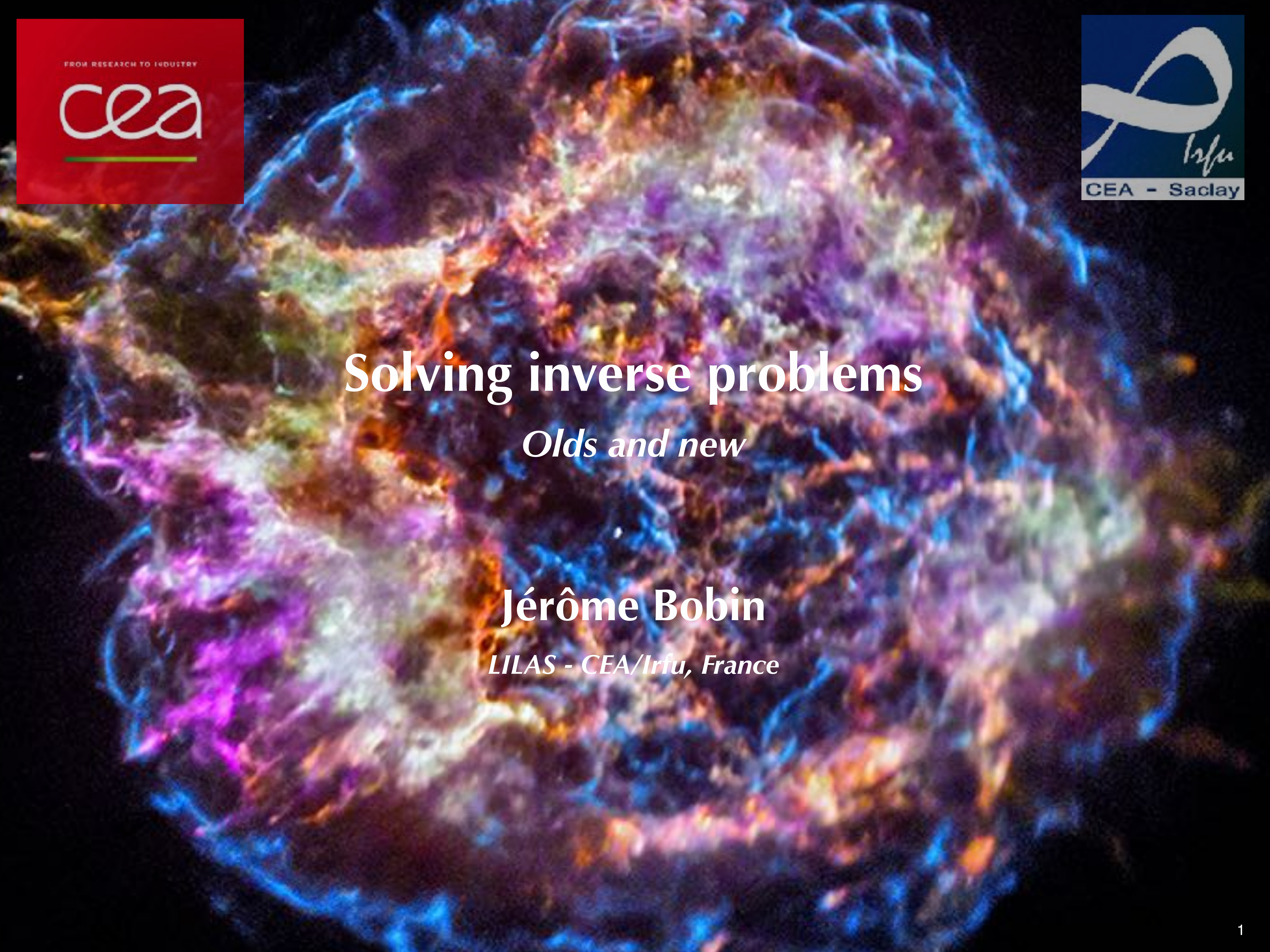


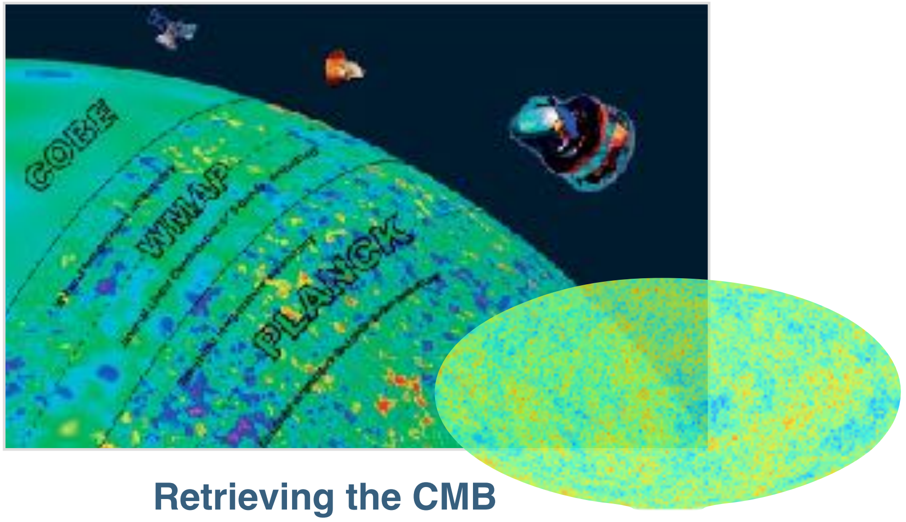
Solving inverse problems

Olds and new

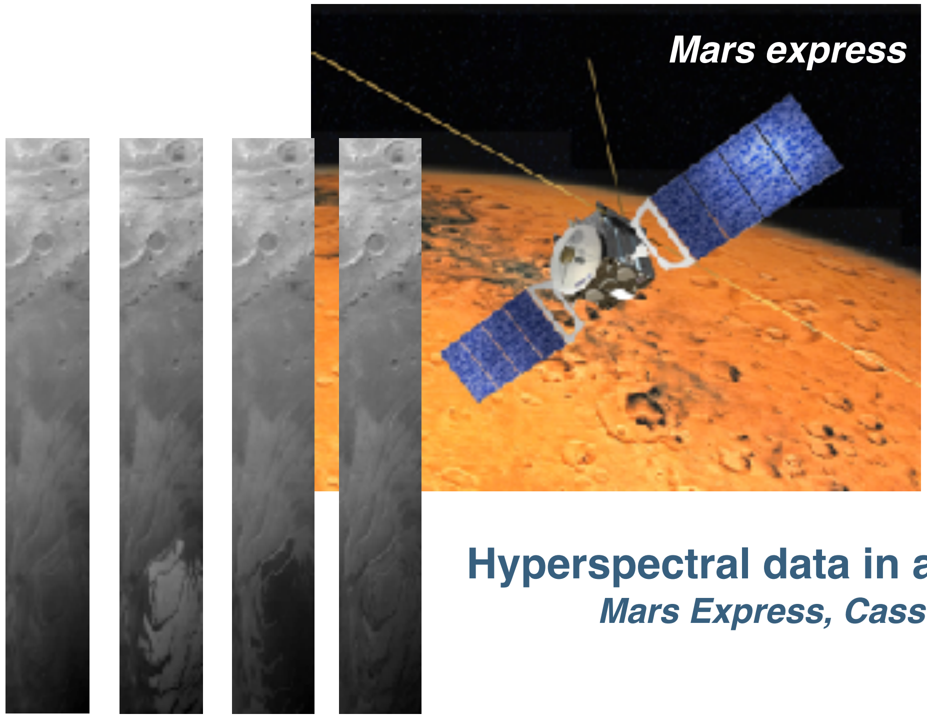
Jérôme Bobin

LILAS - CEA/Irfu, France

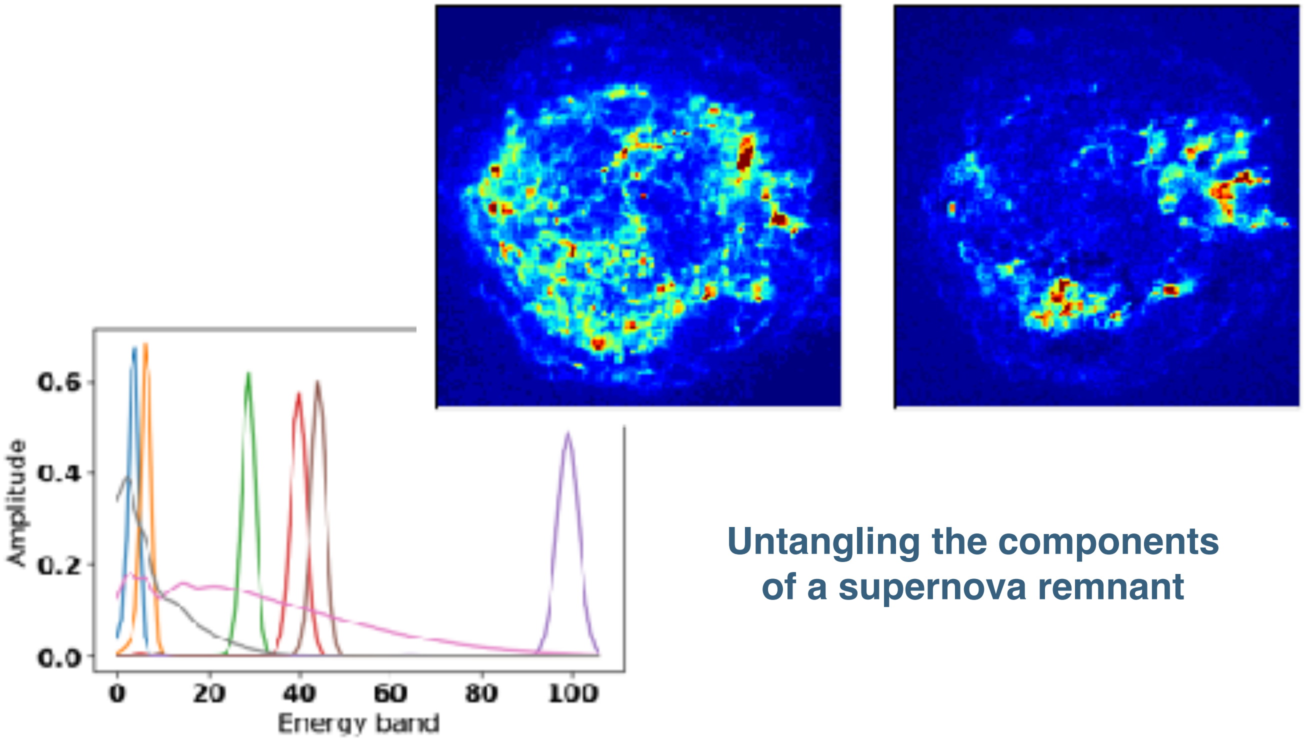
What is an inverse problem ?



Retrieving the CMB
From microwave observations

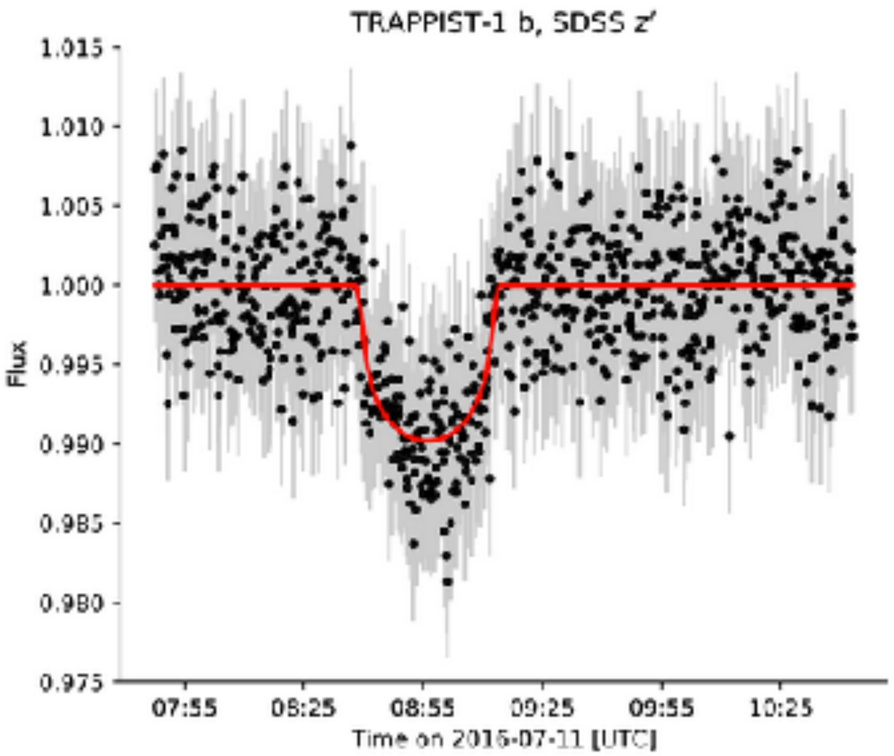


Hyperspectral data in astrophysics
Mars Express, Cassini, etc.

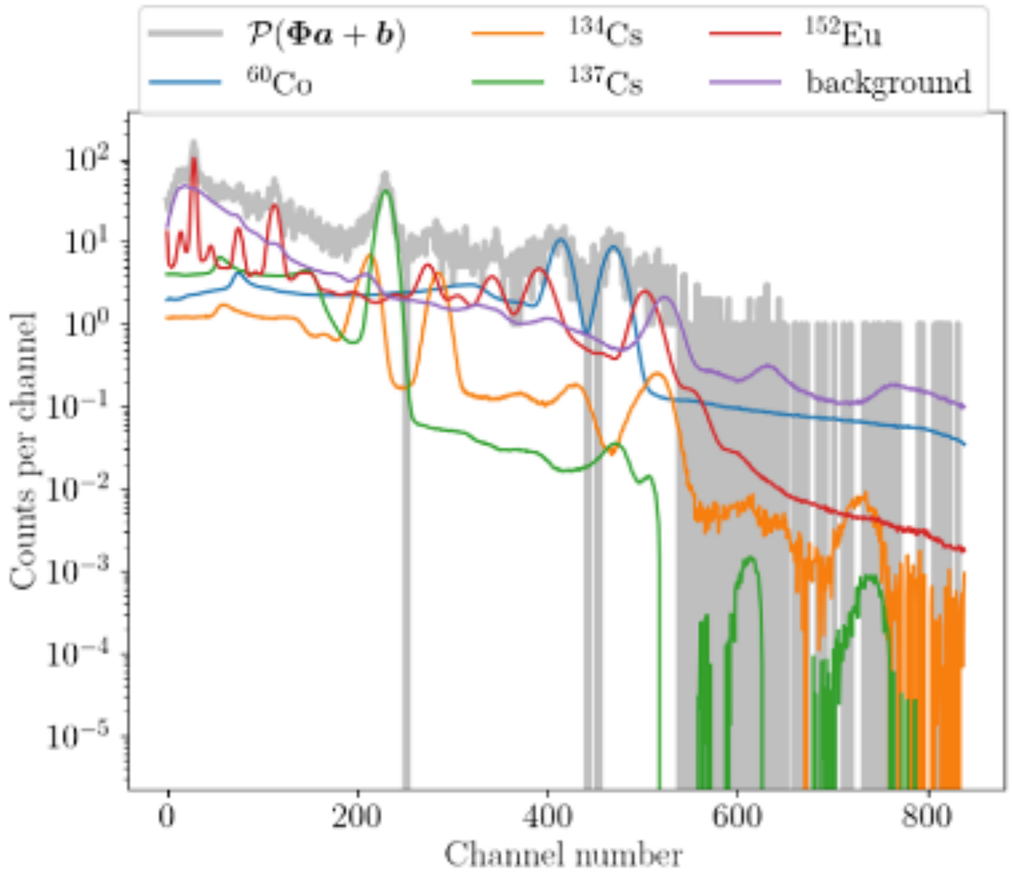


Untangling the components
of a supernova remnant

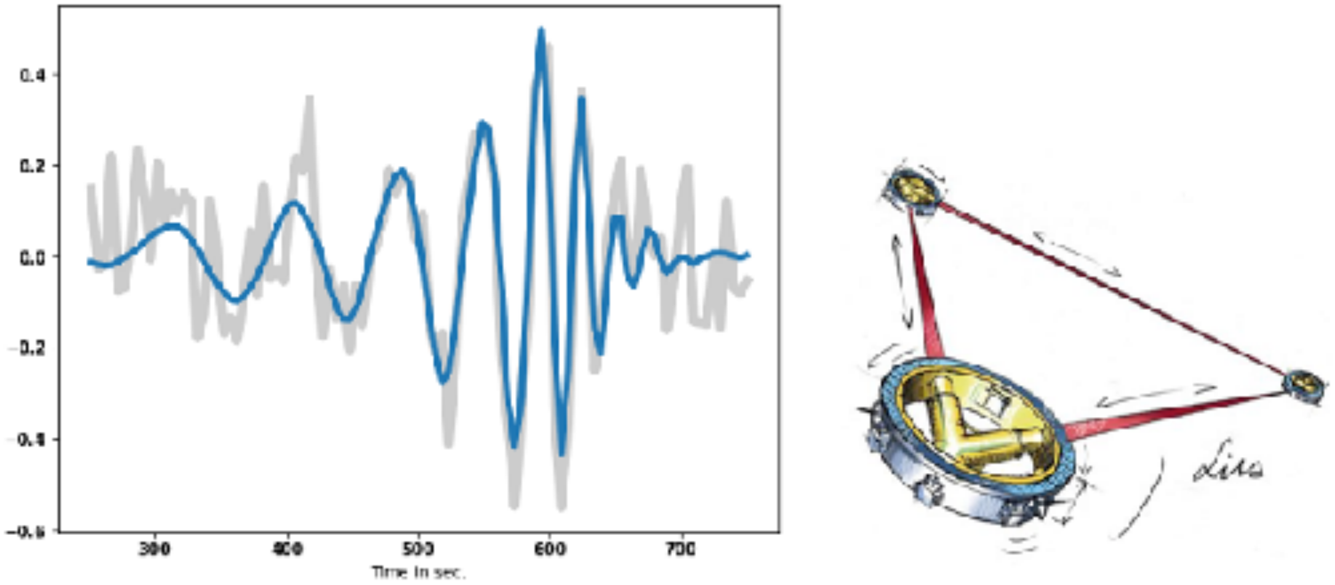
What is an inverse problem ?



Exoplanet detection
From transit observations

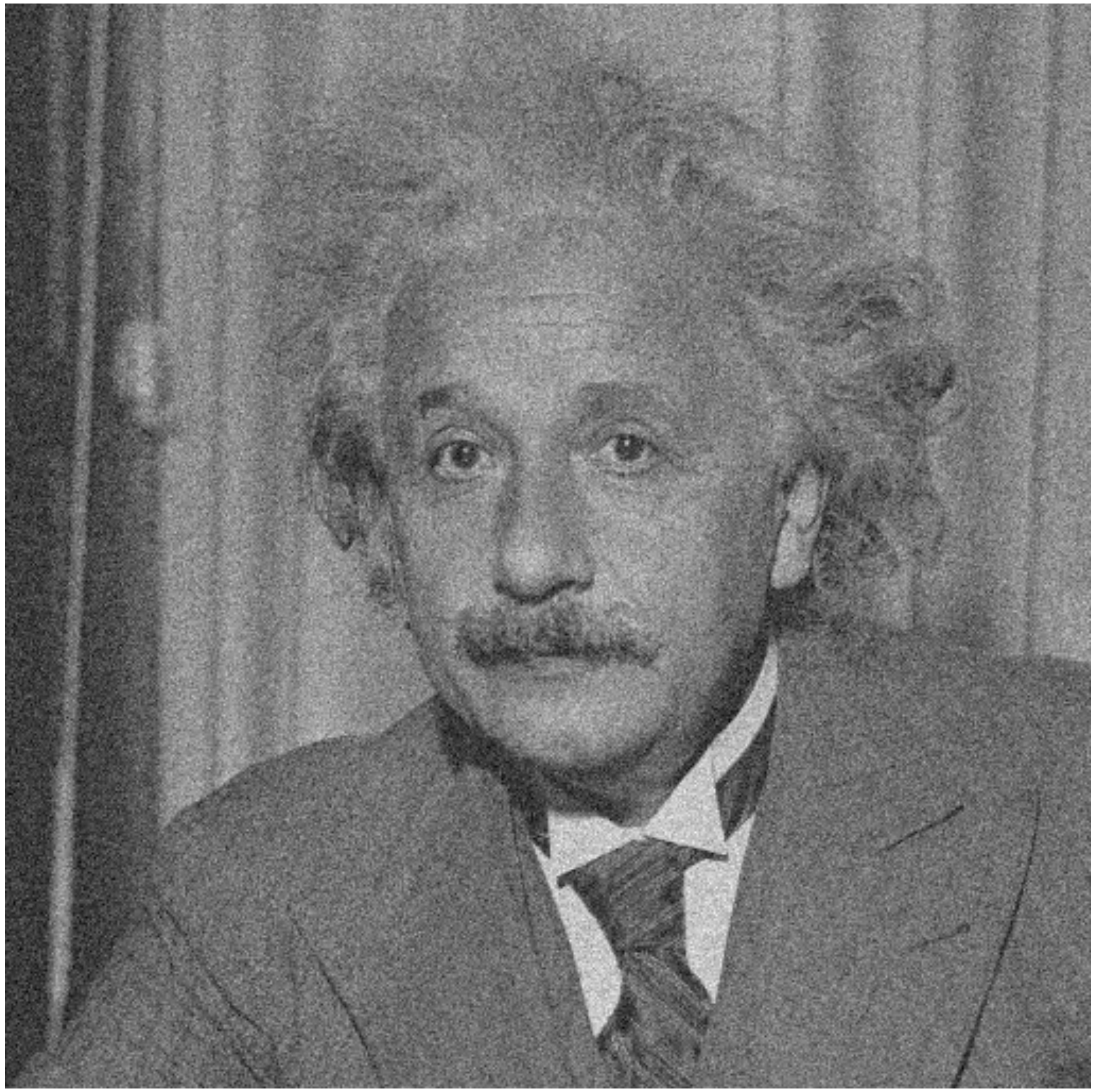


Spectral unmixing



Detection of a Massive Black Hole Binary signal from interferometric data

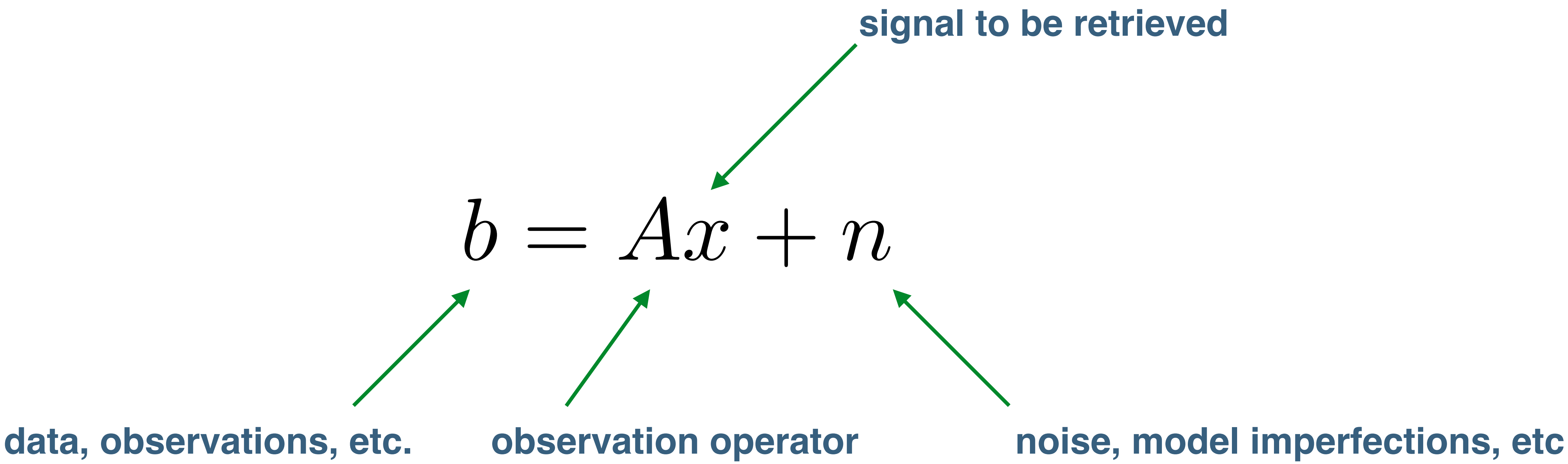
Let’s start from a simple example



$$b = x + n \qquad n \sim \mathcal{N}(0, \sigma^2)$$

Formalism

More generally, we will focus on linear inverse problems where :



signal to be retrieved

$$b = Ax + n$$

data, observations, etc.

observation operator

noise, model imperfections, etc

<u>This models many inverse problems arising in physics :</u>

- Denoising (A is the identity operator)
- Deconvolution (A is the convolution kernel)
- Inpainting/missing data interpolation (A is a binary mask)
- Tomographic reconstruction (A is the partial Radon transform)
- Radio-interferometric reconstruction (A is the partial Fourier transform)
- Compressed sensing
- Blind source separation

Where it started

There are many ways to tackle inverse problems (IP). So far, the vast majority of methods which have been proposed to solve (IP) boil down to finding some solution/estimator which minimizes some cost function:

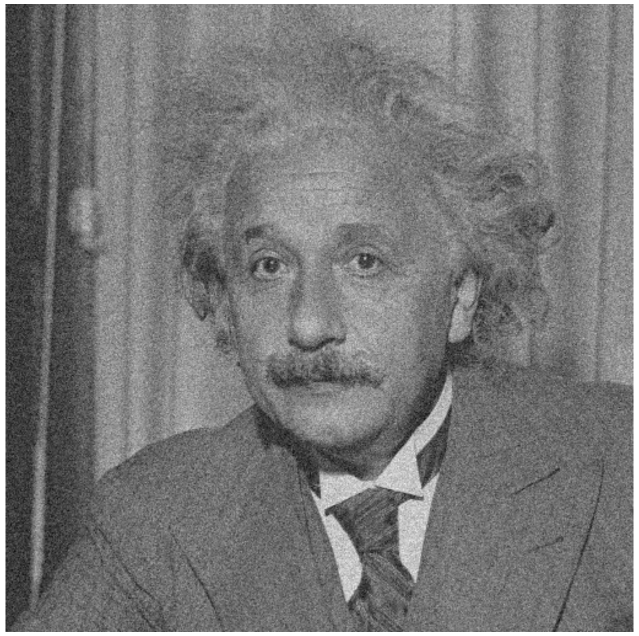
$$\hat{x} = \mathrm{Argmin}_x \quad \mathcal{J}(x)$$

Probably the most popular estimator in Physics is the least-square estimator :

$$\hat{x} = \mathrm{Argmin}_x \quad \|b - Ax\|_2^2$$

which minimizes the Euclidean norm between the observations and the model.

In the previous example: $\hat{x}_{LS} = b$



Remark: in case the noise is additive and Gaussian, the LS estimator is equivalent to the celebrated Maximum Likelihood estimator in statistics.

The need for a regularization

The least-square estimator does not assume any prior assumption about the signal to be retrieved. Including such kind of prior information can be done by penalizing/favoring certain desired signal properties in the estimate procedure. This is done by adding a prior/penalization term in the cost function:

$$\hat{x} = \mathrm{Argmin}_x \quad \mathcal{P}(x) + \mathcal{J}(x)$$

penalty term
(penalizes/favors certain signal properties)

data fidelity term
(measures how well the model fits the data)

Again, many penalization/penalty terms have been proposed in the literature ***(ex: energy, entropy, signal smoothness, positivity, etc)*****. Probably the simplest penalty is the one that penalizes high-energy solutions:**

$$\hat{x} = \mathrm{Argmin}_x \quad \lambda ||x||_2^2 + ||b - x||_2^2$$

Remark: in statistics, the use of prior knowledge arises naturally in the Bayesian inference framework. The aforementioned estimator is then better known as the maximum a posteriori estimator (MAP).

A simple application

The least-square estimator does not assume any prior assumption about the signal to be retrieved. Including such kind of prior information can be done by penalizing/favoring certain desired signal properties in the estimate procedure. This is done by adding a prior/penalization term in the cost function:

$$\hat{x} = \mathrm{Argmin}_x \quad \mathcal{P}(x) + \mathcal{J}(x)$$

penalty term
(penalizes/favors certain signal properties)

data fidelity term
(measures how well the model fits the data)

Again, many penalization/penalty terms have been proposed in the literature ***(ex: energy, entropy, signal smoothness, positivity, etc)*****. Probably the simplest penalty is the one that penalizes high-energy solutions:**

$$\hat{x} = \mathrm{Argmin}_x \quad \lambda \|x\|_2^2 + \|b - x\|_2^2$$

$$\hat{x} = \frac{1}{\lambda + 1} b$$

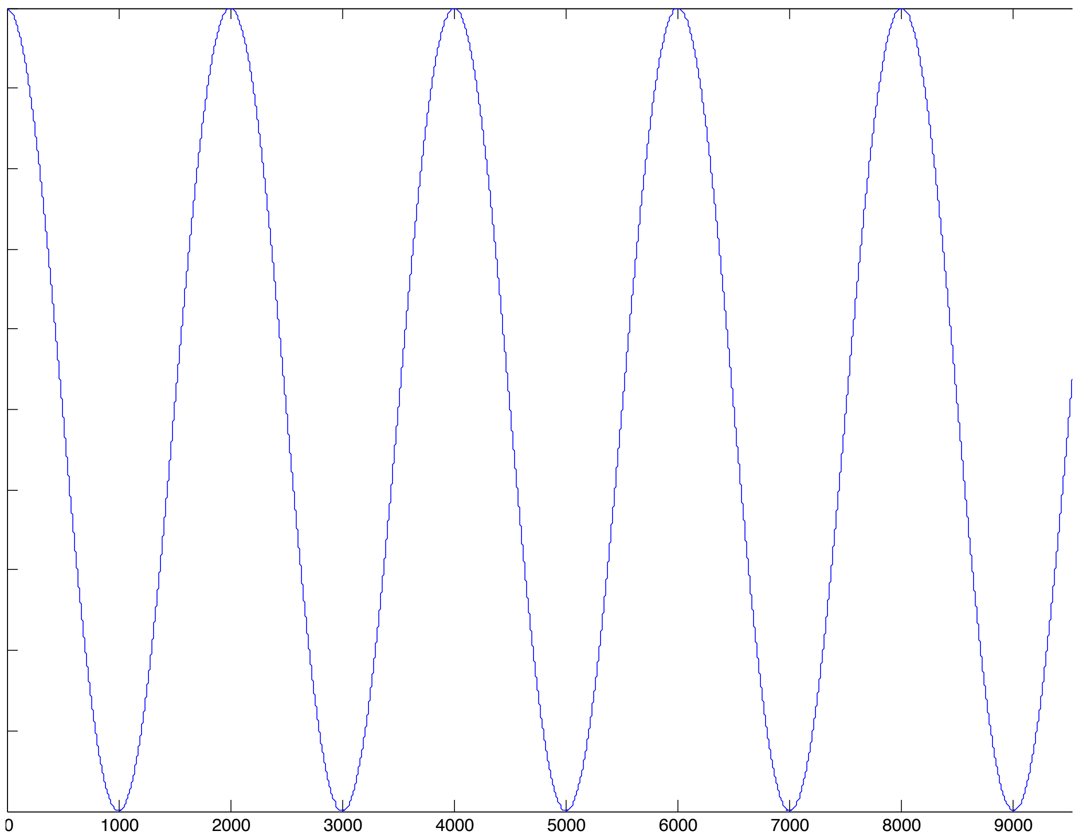
Remark: this is best known as the Wiener filter

Remark: in statistics, the use of prior knowledge arises naturally in the Bayesian inference framework. The aforementioned estimator is then better known as the maximum a posteriori estimator (MAP).

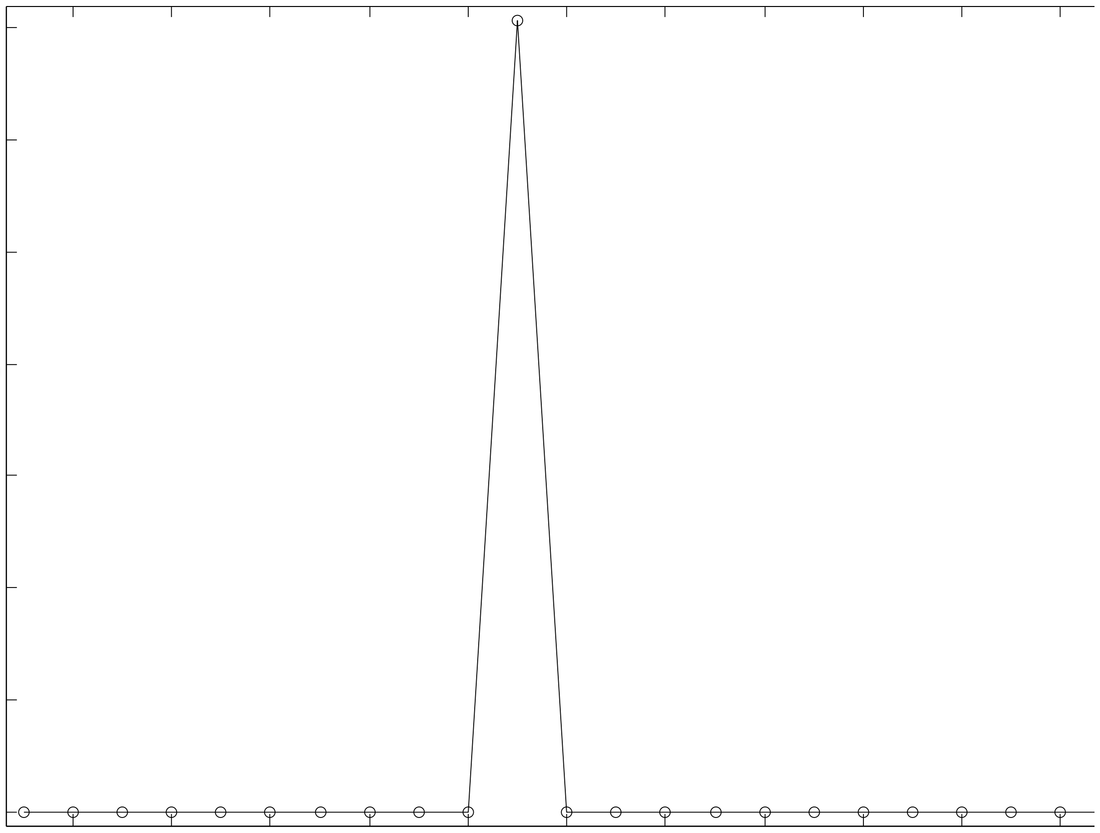
Sparsity and compressibility

In the last two decades, the most dramatic advances in signal estimation have focused on using prior information enforcing signal properties based on desired **geometrical/morphological properties**.

Gist of the sparsity : signals can be sparsely represented in representations (basis, etc.) that efficiently encode their geometrical/morphological properties.



$$x$$



$$\alpha = \mathbf{\Phi} x$$

Discrete cosine transform

Sparse modelling

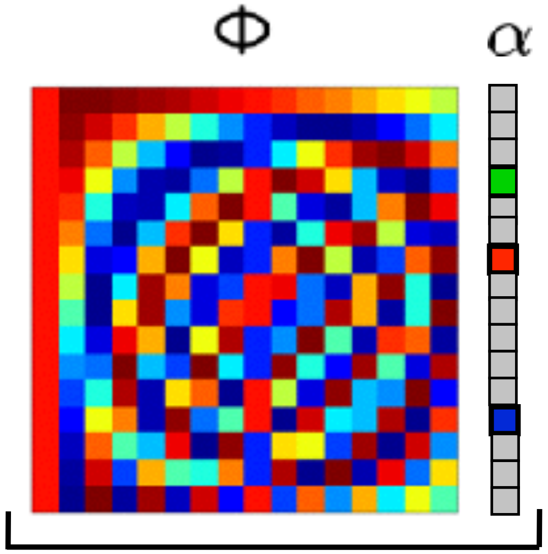
Basis, frame, dictionary
(Fourier, wavelets, curvelets ....)

$$\Phi = \{\phi_1, \cdots, \phi_t\}$$

coefficients

$$x = \sum_{j=1}^{t} \alpha_j \phi_j$$

$x =$



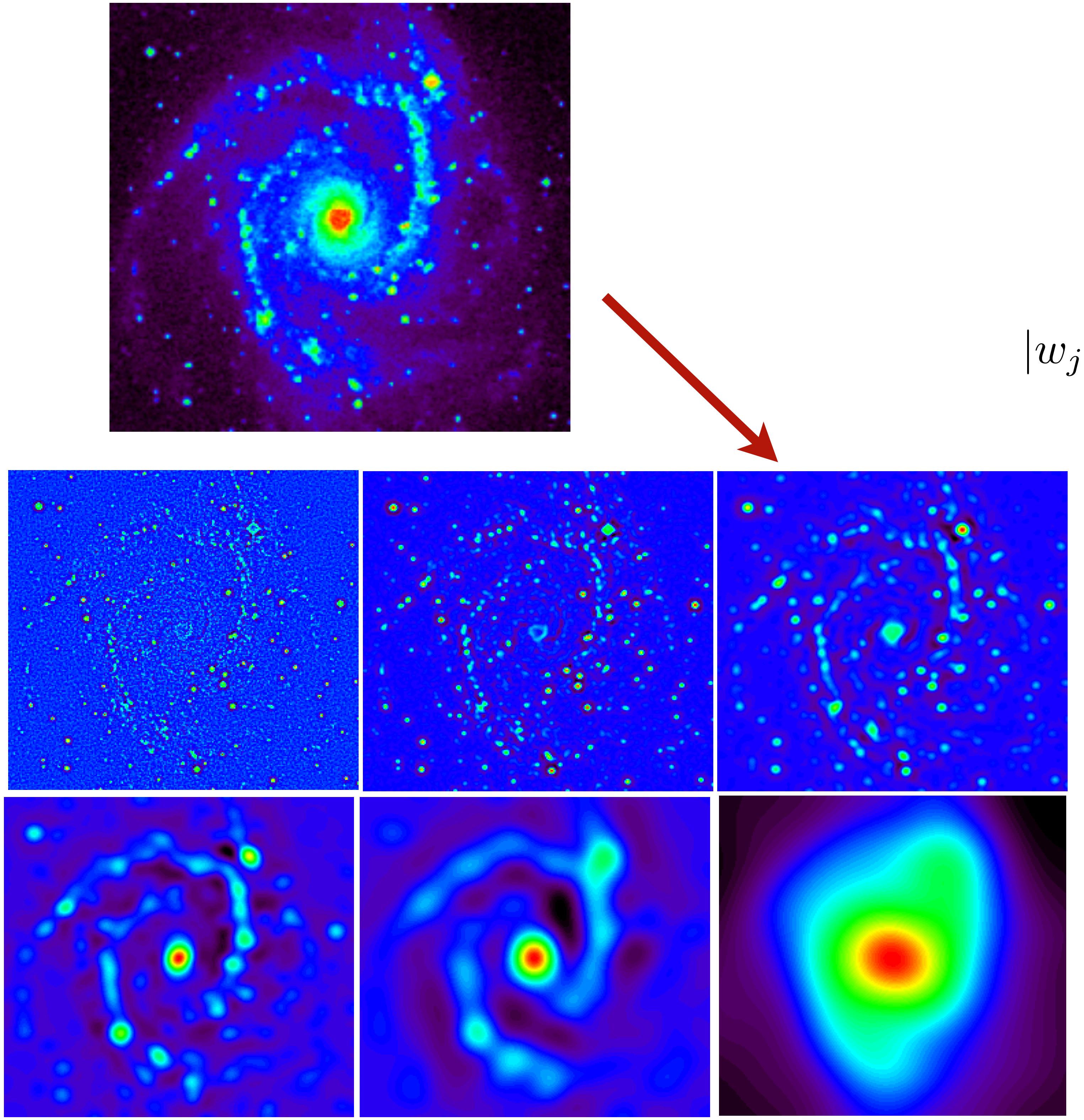
Prior: Data Representation

Sparse Model 1 :

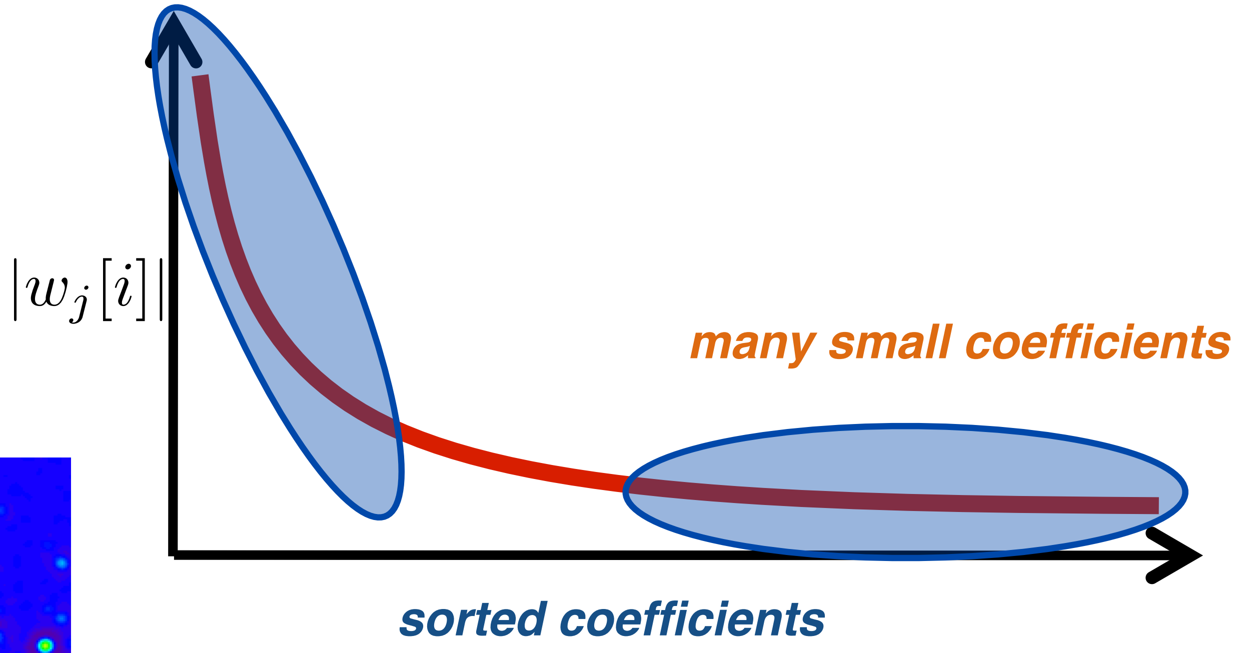
x is assumed to have only k non zero entries

x is said to be exactly k-sparse in Φ

Sparse modelling



Starlet transform
(isotropic undecimated wavelet transform)



for clarity $w_j[i] \to \alpha[k]$

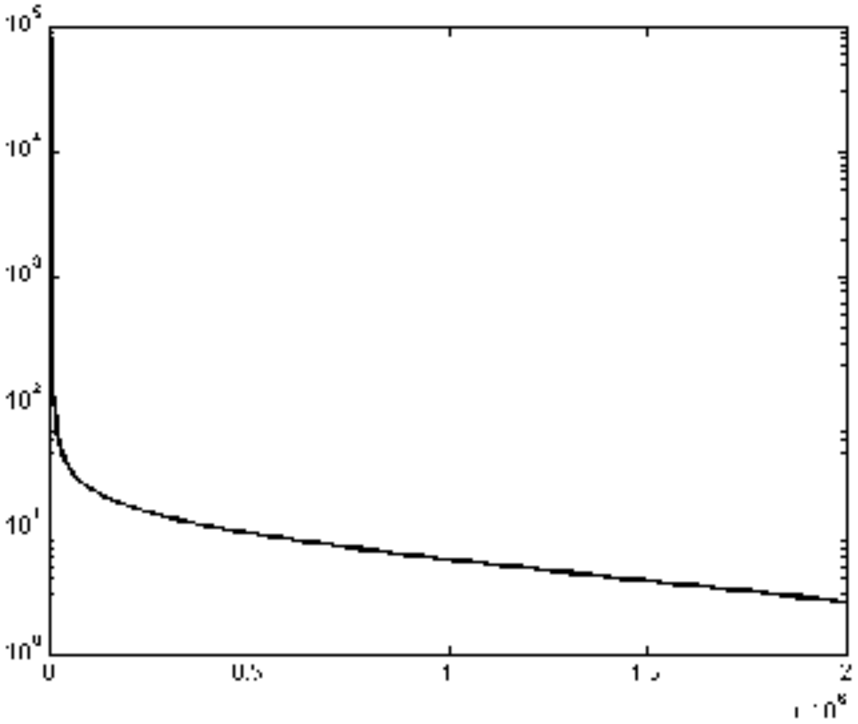
$$|\alpha[k]| \leq C k^{-1/q}$$

Sparse Model 2 :

x is approximately sparse in Φ

Sparse modelling

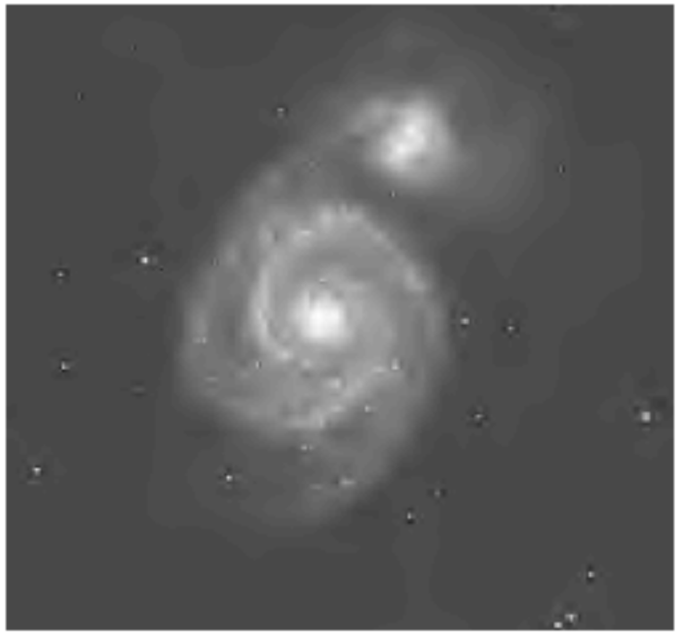
wavelet transform



sorted wavelet coefficients

1%

0.1%



0.01%

Sparse modelling

JPEG VS JPEG 2000

Original BMP

300x300x24

270056 bytes

JPEG 1:68

3983 bytes

JPEG2000 1:70

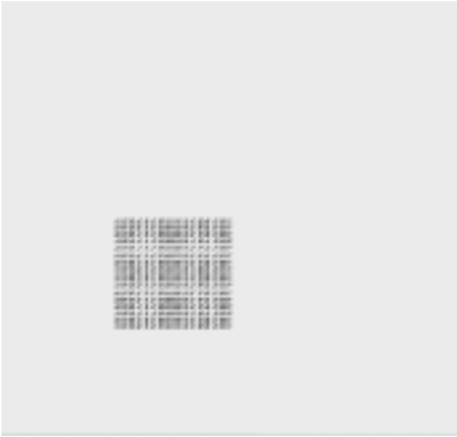
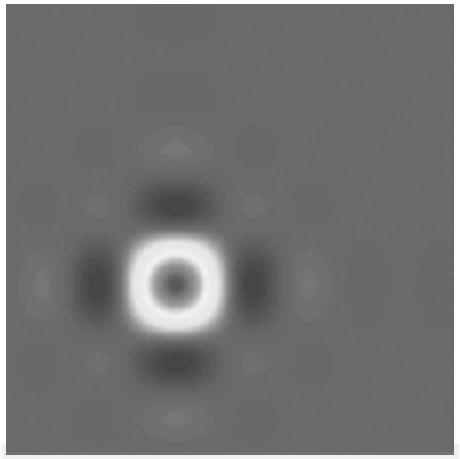
3876 bytes

Based on an harmonic basis (Discrete Cosine Transform)

Based on the wavelet transform

Sparse modelling

In general, sparse representations should be chosen based on the desired morphology

Cosine transform	Oscillating/periodic structures Textures	
Wavelets	Singularities, point-like structures	
Curvelets	Contours	
… and many more		

Sparse solutions to inverse problems

Let's assume x is sparse is some orthogonal basis: $\alpha = \mathbf{\Phi} x$

$$\hat{x} = \mathrm{Argmin}_{x = \mathbf{\Phi}\alpha} \quad \mathcal{P}(\alpha) + \|b - \mathbf{\Phi}\alpha\|_2^2$$

sparsity-enforcing penalty

data fidelity term

(measures how well the model fits the data)

Examples of penalty terms:

$$\mathcal{P}(\alpha) = \|\alpha\|_{\ell_0}$$

The 0-norm counts the number of nonzero elements

$$\mathcal{P}(\alpha) = \|\alpha\|_{\ell_1}$$

$$\|\alpha\|_{\ell_1} = \sum_i |\alpha[i]|$$

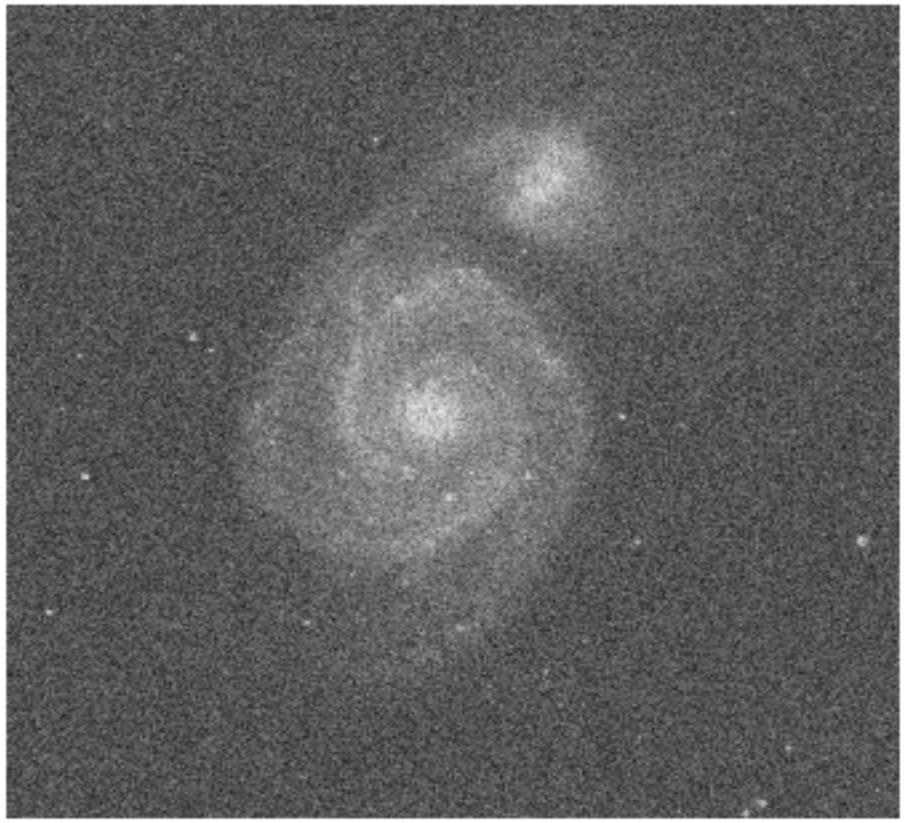
Sparse solutions to inverse problems

Denoising as a linear inverse problem:

$$b = x + n$$

The observation matrix A is the identity matrix.

The noise is assumed to be additive, white and Gaussian: $n \sim \mathcal{N}(0, \sigma^2)$



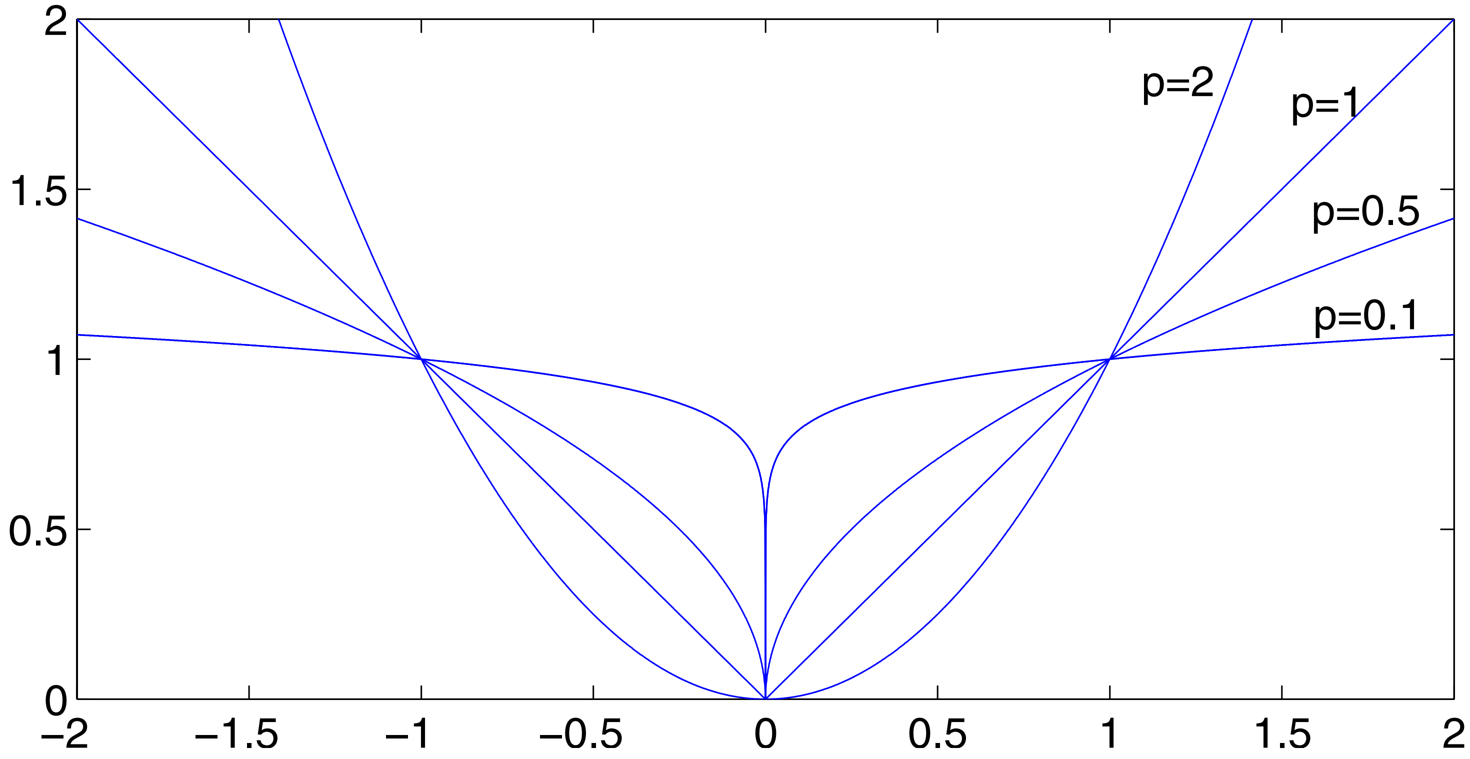
SNR = 1dB

Sparse solutions to inverse problems

It can be recovered from the noisy data by solving the following linear inverse problem:

$$\hat{x} = \mathrm{Argmin}_x \quad \lambda \|x\|_{\ell_p} + \frac{1}{2}\|b - x\|_2^2$$

with: $$\|x\|_{\ell_p} = \left(\sum_i |x[i]|^p\right)^{1/p}$$



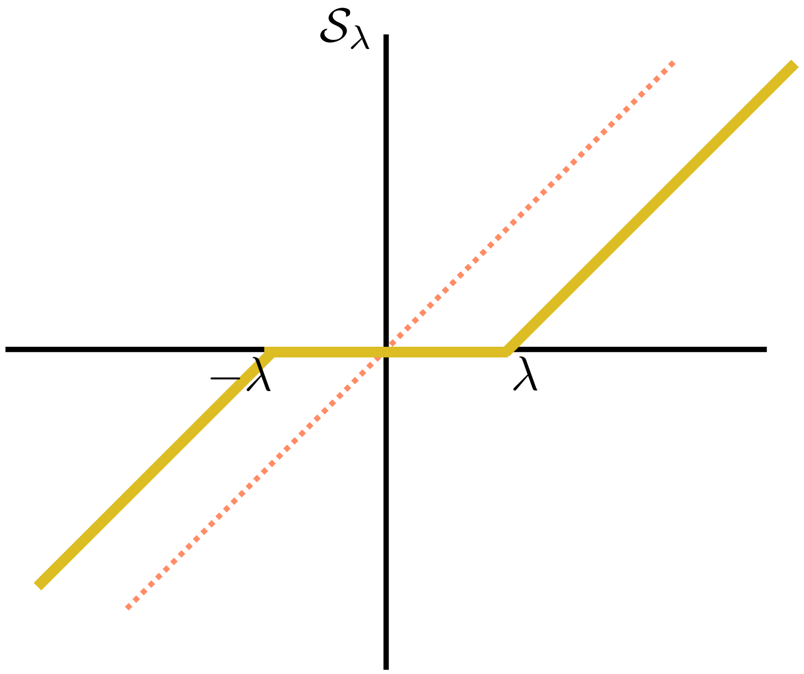
The solution amounts to a thresholding or shrinkage of b :

$$\hat{x} = \mathcal{T}_\lambda^{(p)}(b)$$

Sparse solutions to inverse problems

The most common sparse regularizers are the L1 norm and the L0-“pseudo” norm:

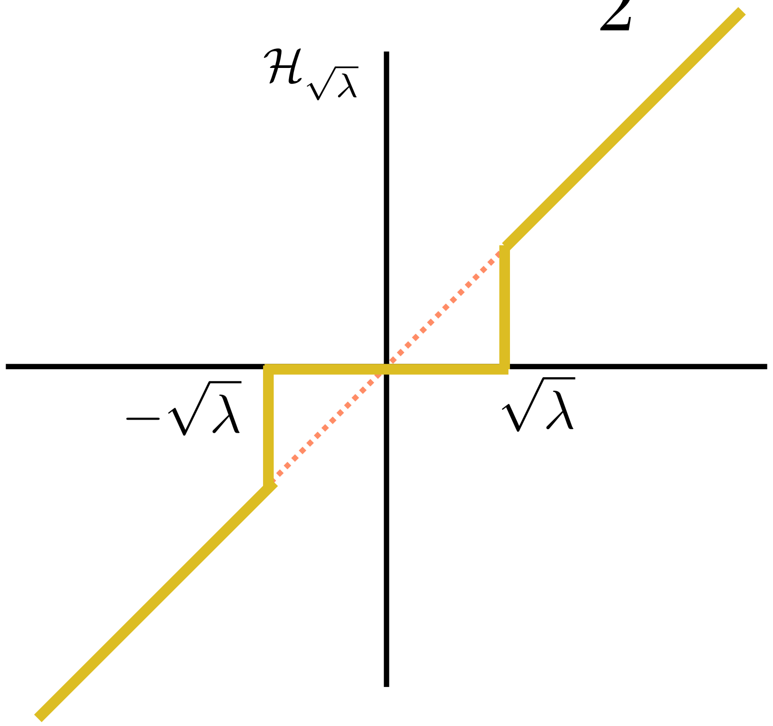
$$\hat{x} = \mathrm{Argmin}_x \quad \lambda \|x\|_{\ell_1} + \frac{1}{2}\|b - x\|_2^2$$



$$\hat{x} = \mathcal{S}_\lambda(b)$$

soft-thresholding operator

$$\hat{x} = \mathrm{Argmin}_x \quad \lambda \|x\|_{\ell_0} + \frac{1}{2}\|b - x\|_2^2$$



$$\hat{x} = \mathcal{H}_{\sqrt{\lambda}}(b)$$

hard-thresholding operator

Sparse solutions to inverse problems

x

$b = x + n$

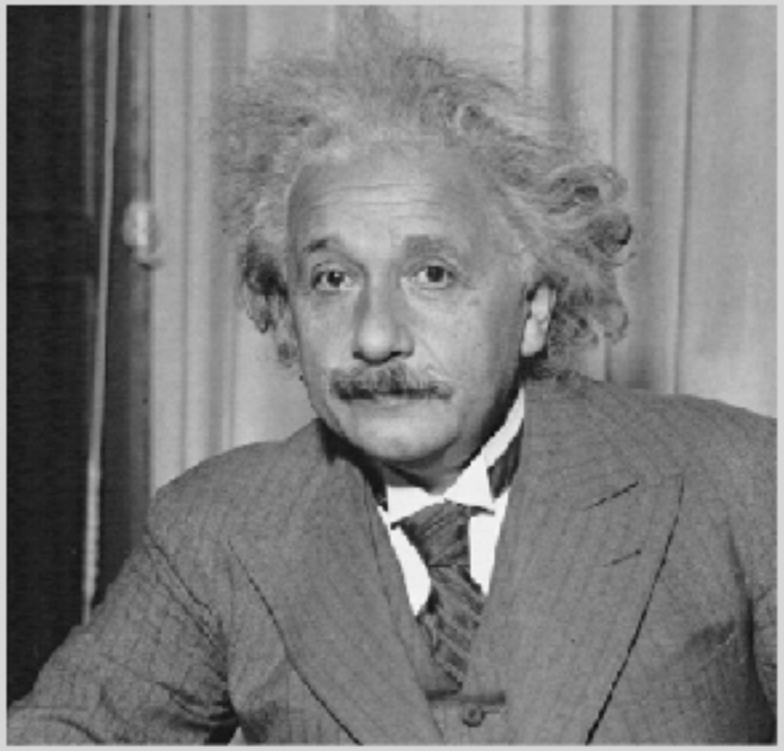
In case sparsity in enforced in some signal representation, the problem to be solved is the following:

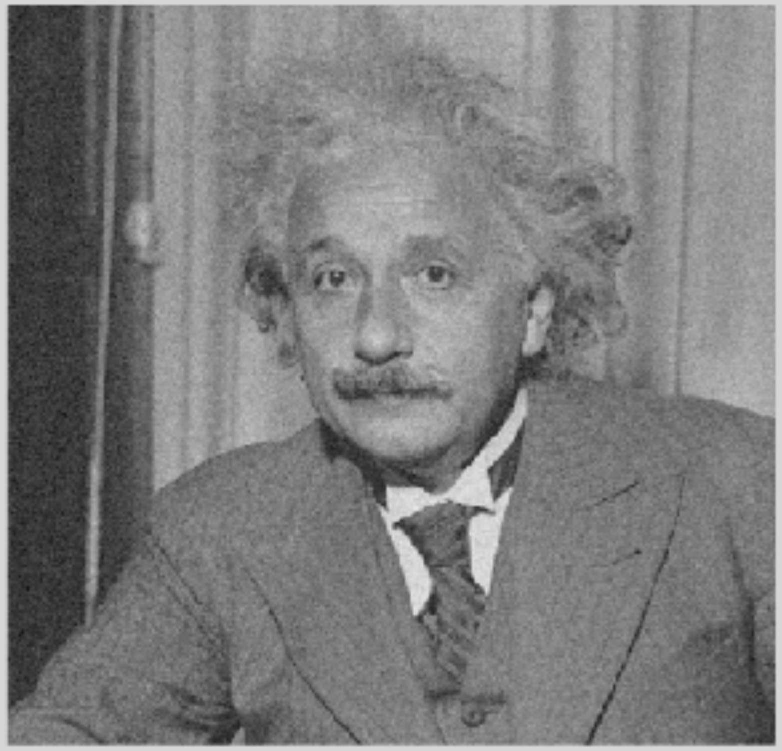
$$\hat{x} = \text{Argmin}_x \quad \lambda \|\mathbf{\Phi}^T x\|_{\ell_p} + \frac{1}{2}\|b - x\|_2^2 \xrightarrow[\text{orthogonal case}]{} \hat{x} = \text{Argmin}_{x = \mathbf{\Phi}\alpha} \quad \lambda \|\alpha\|_{\ell_p} + \frac{1}{2}\|\mathbf{\Phi}^T b - \alpha\|_2^2$$

$$\hat{x} = \mathbf{\Phi}\ \mathcal{T}_\lambda^{(p)}(\mathbf{\Phi}^T b)$$

Sparse solutions to inverse problems

x

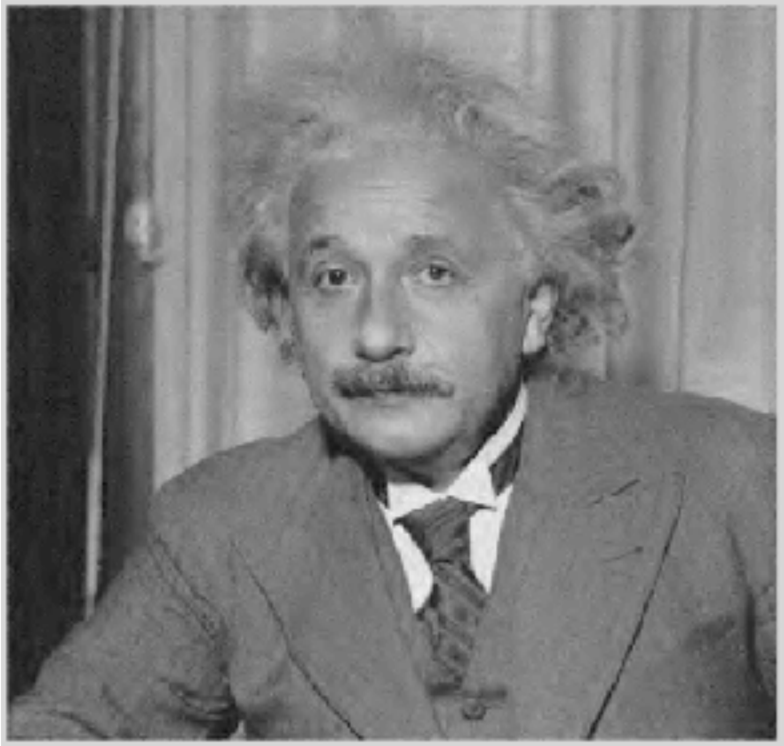


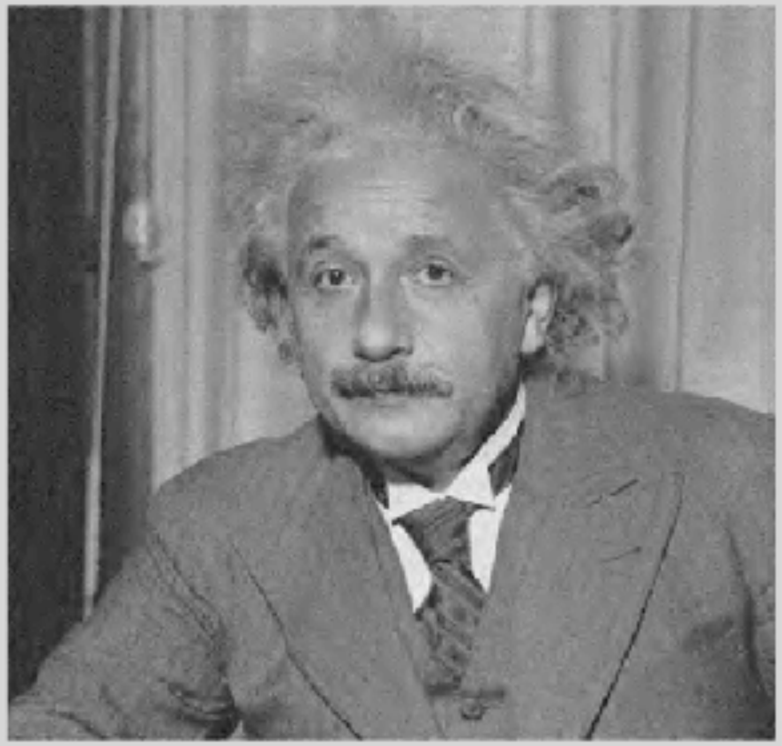


$b = x + n$

L1

$\hat{x} = \mathbf{\Phi}\mathcal{S}_\lambda(\mathbf{\Phi}^T b)$

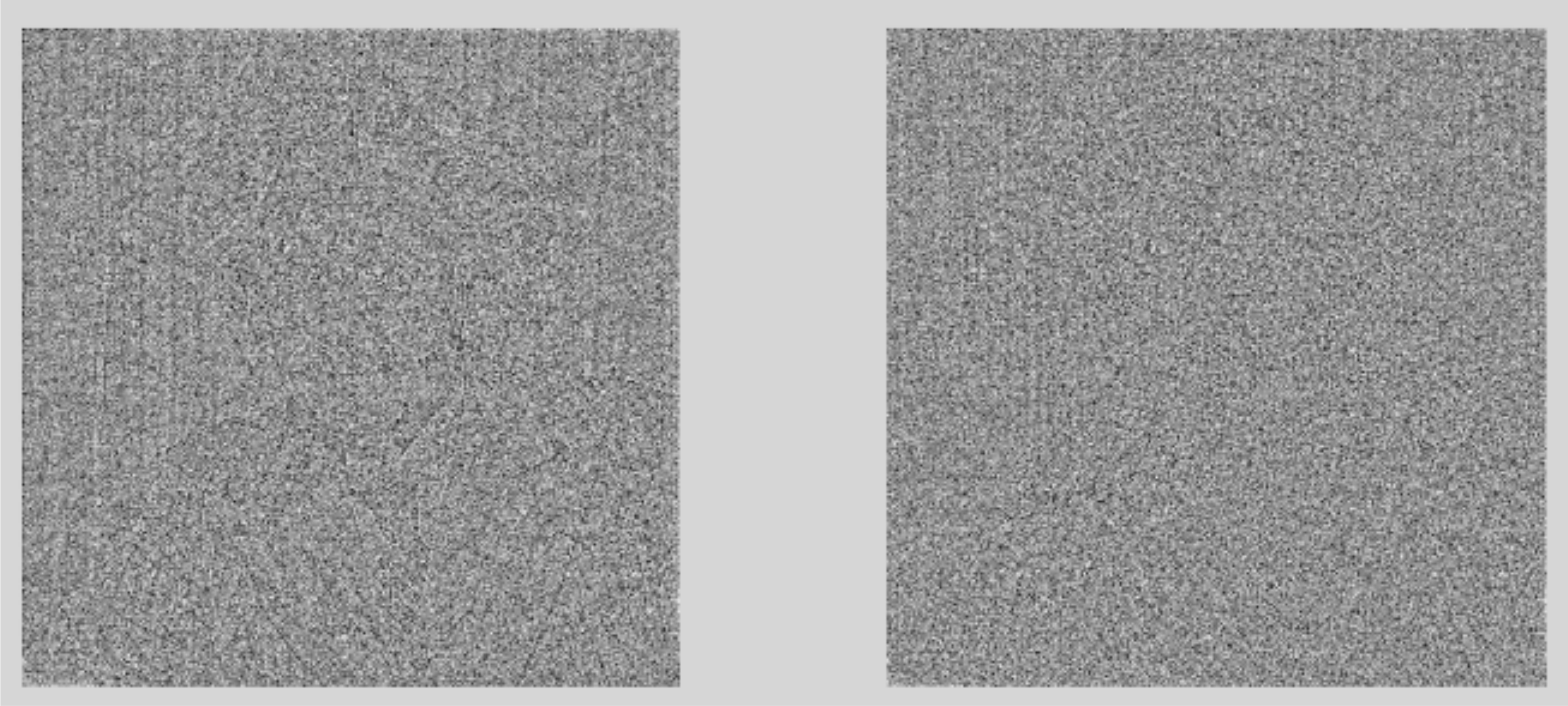




L0

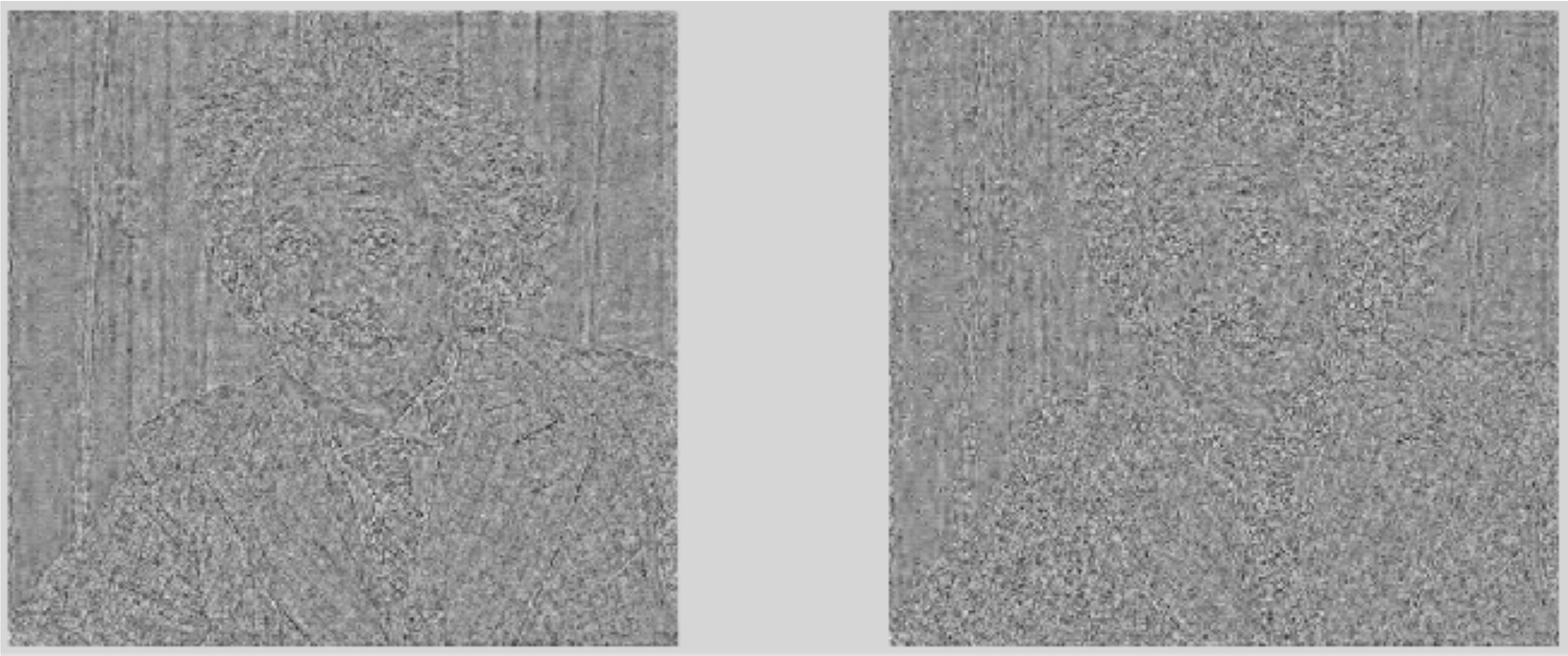
$\hat{x} = \mathbf{\Phi}\mathcal{H}_\lambda(\mathbf{\Phi}^T b)$

Sparse solutions to inverse problems



$b - \hat{x}_{\ell_1}$

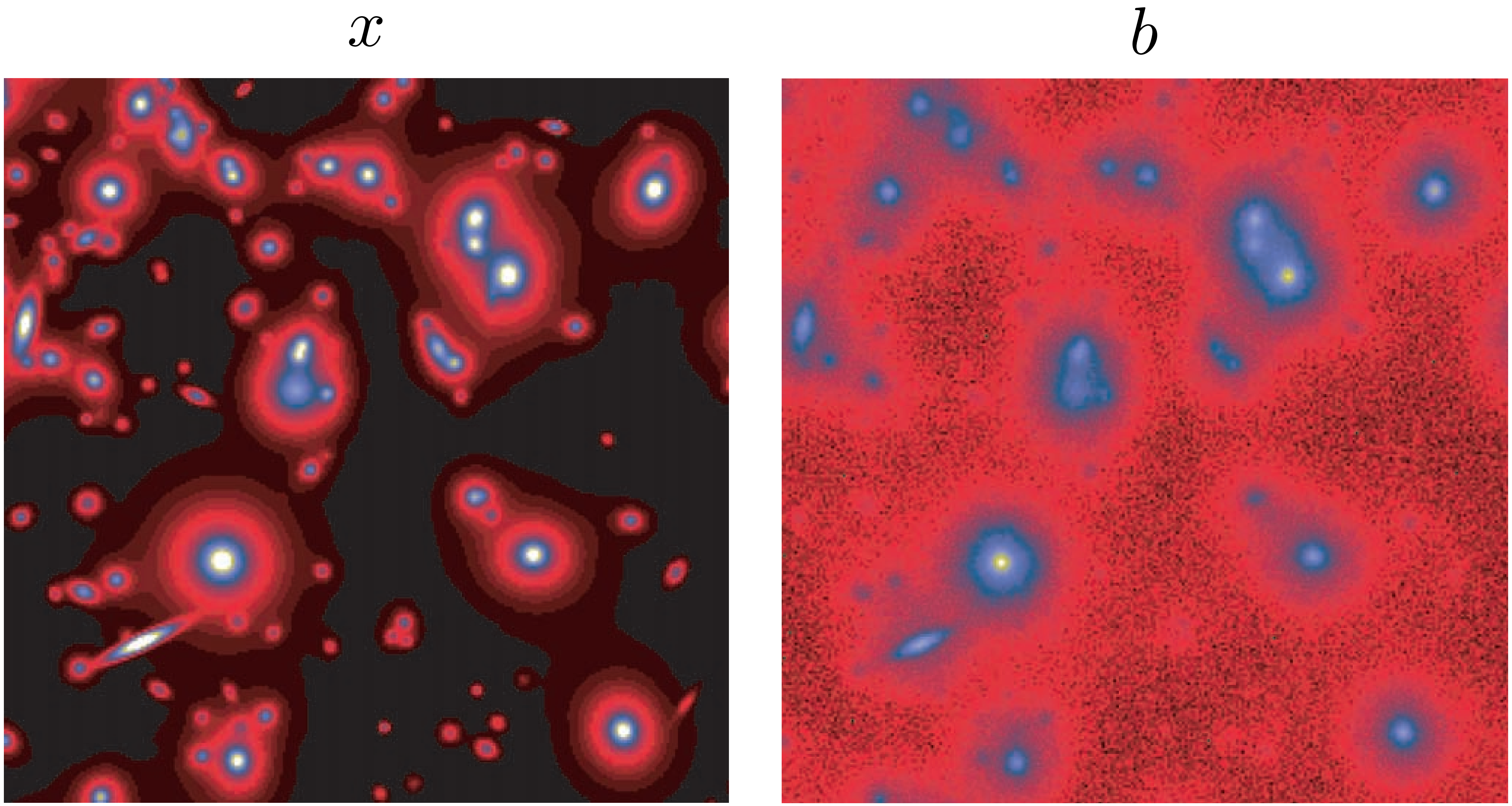
$b - \hat{x}_{\ell_0}$



$x - \hat{x}_{\ell_1}$

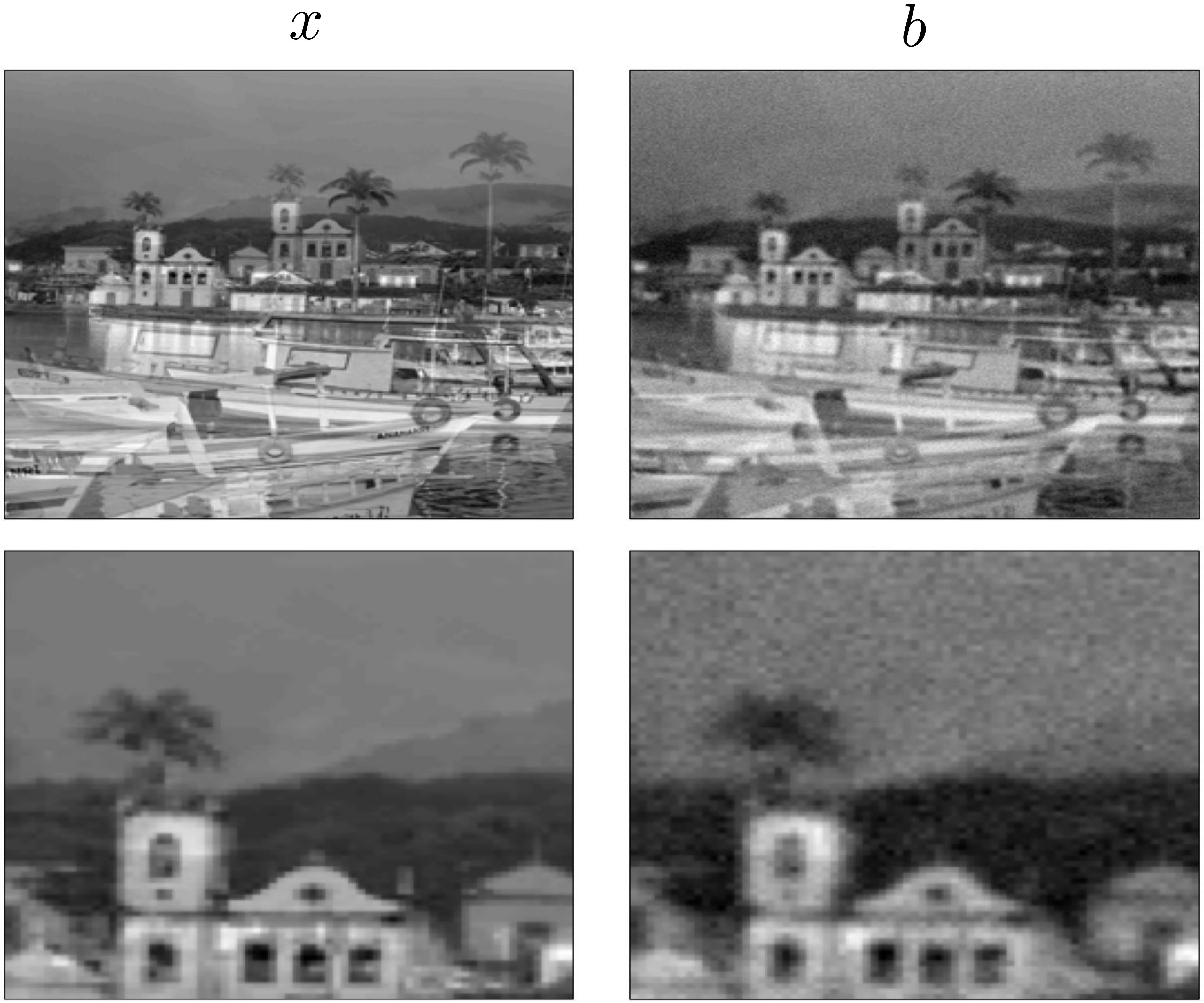
$x - \hat{x}_{\ell_0}$

Beyond denoising: deblurring/deconvolution



Simulations of image from the *Hubble Space Telescope*

Beyond denoising: deblurring/deconvolution



Beyond denoising: deblurring/deconvolution

In imaging science, the (spatial) resolution of the images/signals is limited by the instrument/sensor/… etc.

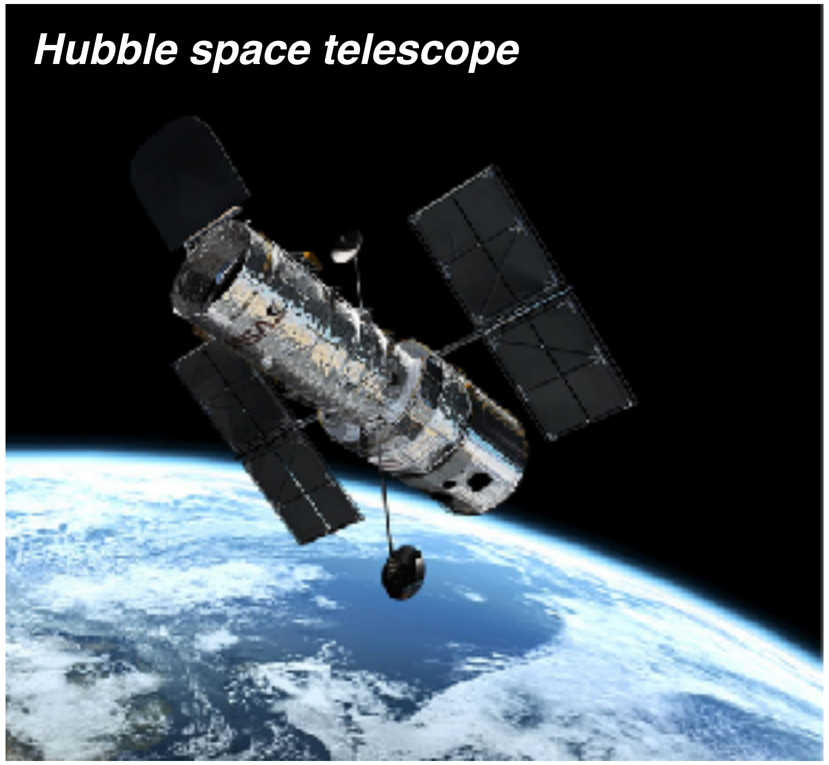
The loss of resolution is mathematically described by the convolution of the signal with the **point-spread-function (PSF)** of the optical device:

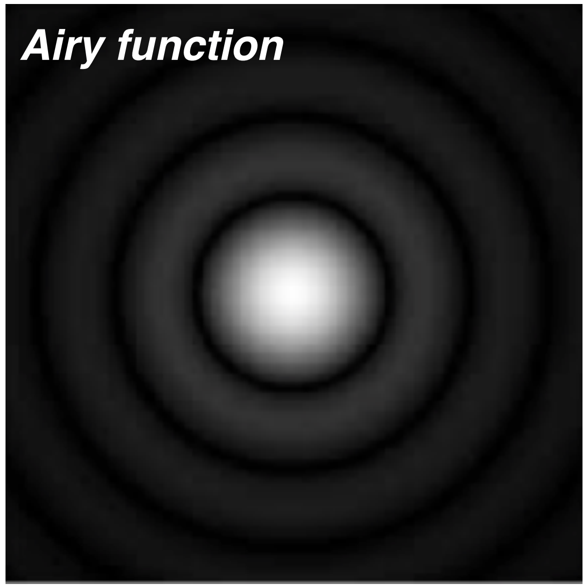
$$b = h \star x + n$$

Impulsive response/PSF

$$b = Hx + n$$

Toeplitz-circulant matrix


Hubble space telescope

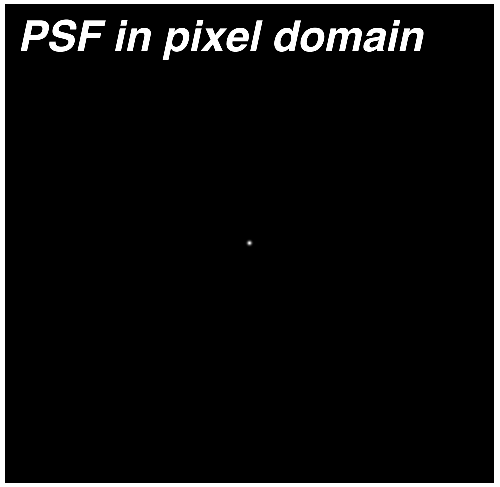

Airy function

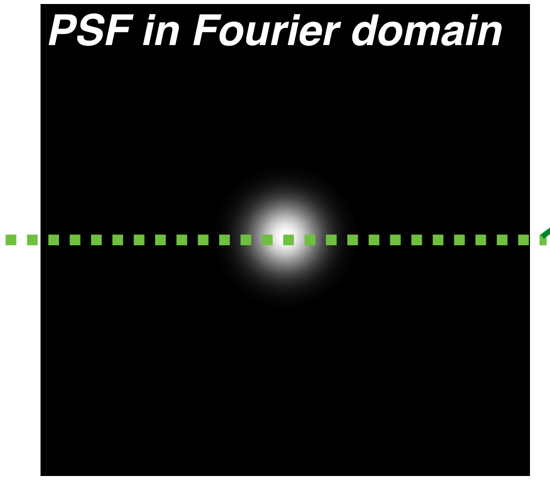
$$r \sim 1.22\,\frac{\lambda}{D}$$

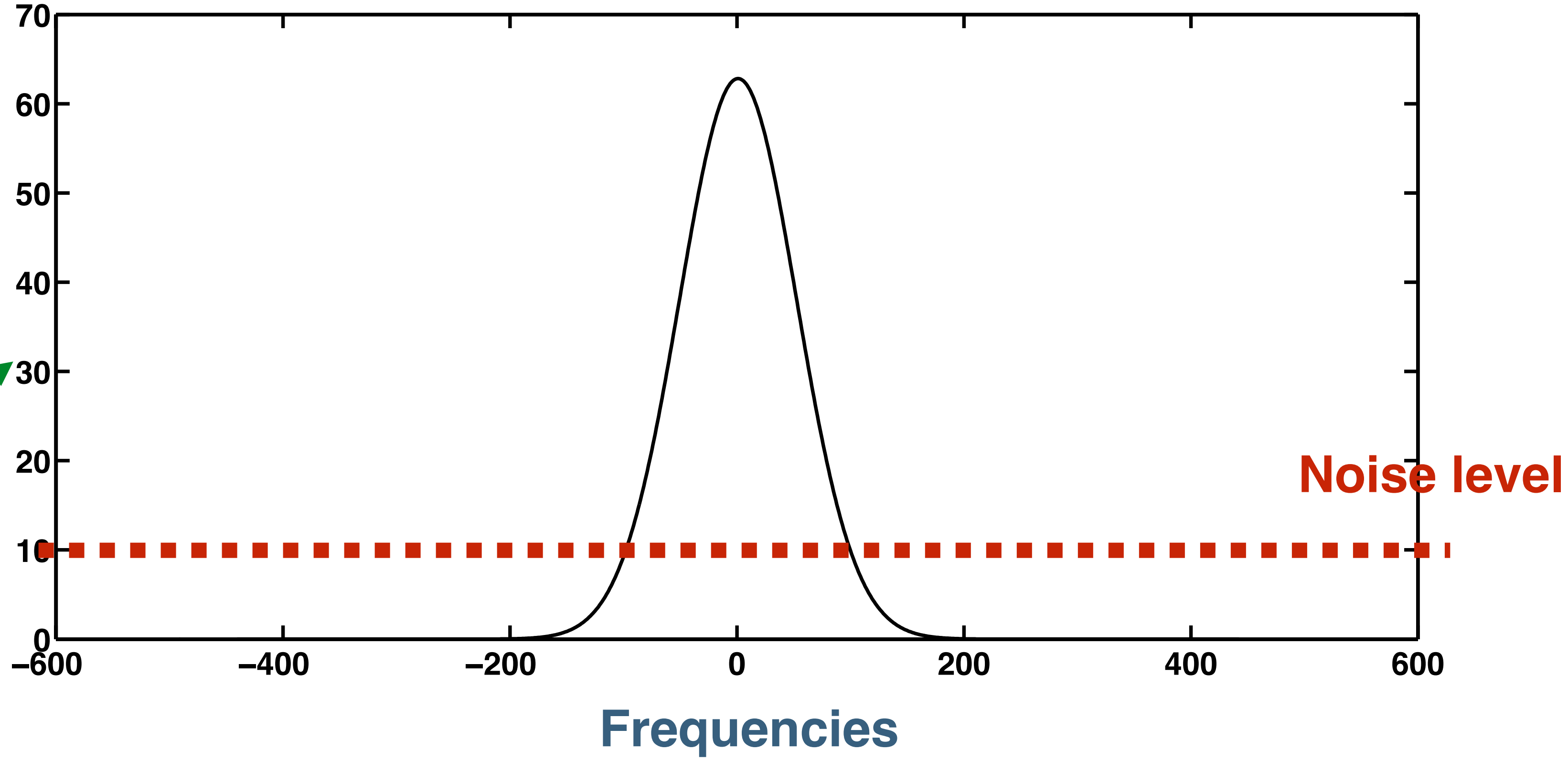
Beyond denoising: deblurring/deconvolution

GOOD NEWS: Toeplitz-circulant matrices are diagonalized by the Fourier transform:

$$b = Hx + n \xrightarrow[\text{Fourier transform}]{\mathcal{F}} \mathcal{F}b = \underset{\text{Diagonal matrix}}{D}\mathcal{F}x + \mathcal{F}n$$







Deblurring/deconvolution is an ill-posed inverse problem

Deblurring/deconvolution: ML estimator

Let’s turn to the classical maximum likelihood estimator. In case of additive Gaussian noise, it is fully equivalent to a least-square estimator:

$$\hat{x}_{\mathrm{ML}} = \mathrm{Argmin}_x \|b - Hx\|_{\ell_2}^2$$

which can be recast in the Fourier domain as follows:

$$\hat{x}_{\mathrm{ML}} = \mathcal{F}^{-1}\mathrm{Argmin}_u \ \|\mathcal{F}b - Du\|_{\ell_2}^2$$

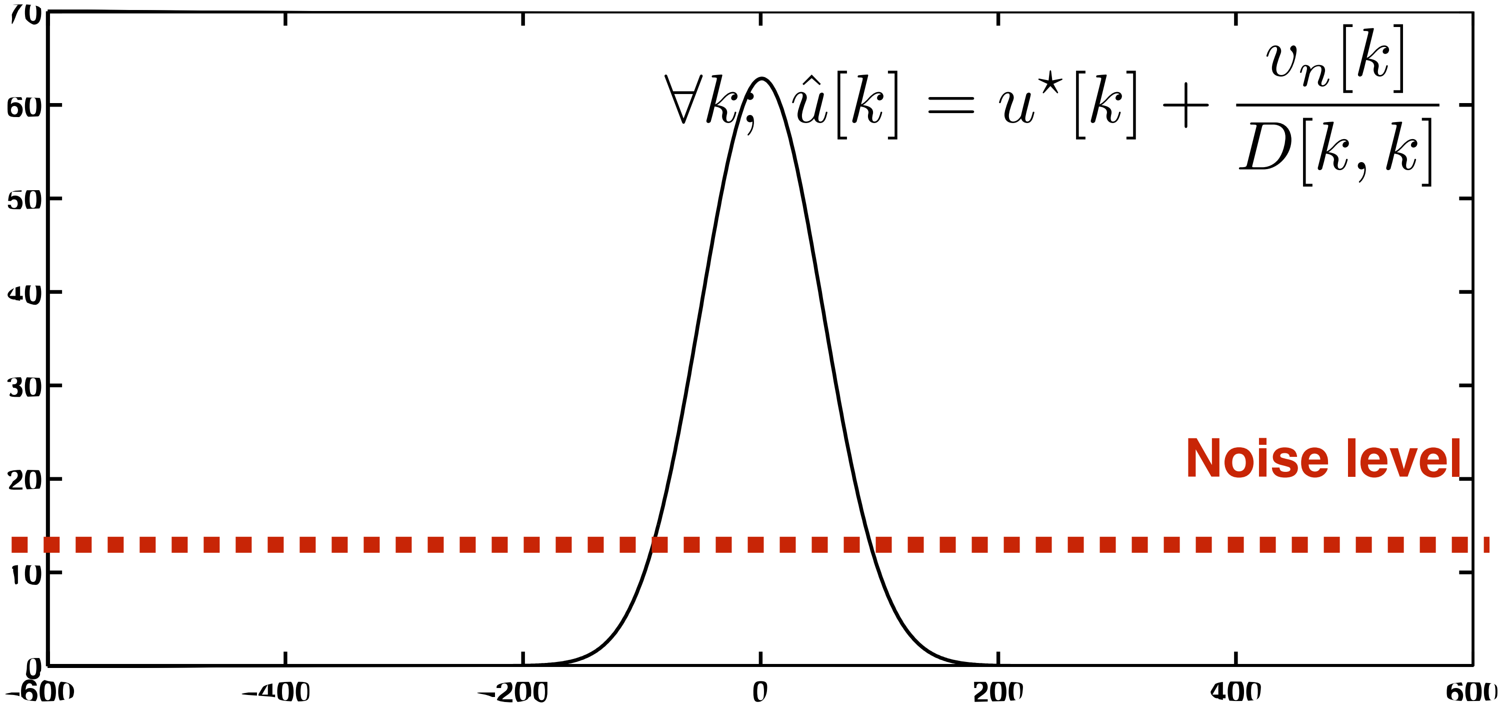
which now fully separable in u:

$$v = \mathcal{F}b$$

$$v_n = \mathcal{F}n$$

$$\forall k; \ \hat{u}[k] = \frac{v[k]}{D[k,k]}$$

$$\hat{x}_{\mathrm{ML}} = \mathcal{F}^{-1}\hat{u}$$



$$\forall k; \ \hat{u}[k] = u^\star[k] + \frac{v_n[k]}{D[k,k]}$$

Highly amplifies noise !

Beyond denoising: Wiener filter

One of the most common Bayesian estimator is the Wiener filter, which obtained when the signal x is assumed to be a Gaussian random field (GRF) that is described by its power spectrum P:

$$\hat{x}_{\text{Wiener}} = \text{Argmin}_x \ \frac{1}{2} x^T \mathcal{F}^T W^{-1} \mathcal{F} x + \frac{1}{2\sigma_n^2} \| b - Hx \|_{\ell_2}^2$$

Inverse covariance matrix
of x in the Fourier domain
GRF: diagonal matrix
W[k,k] = P[k]

Fourier transform

After some basic calculation, we can show that in the Fourier domain:

$$\forall k; \ \hat{u}_{\text{Wiener}}[k] = \frac{P[k]}{P[k] + \frac{\sigma_n^2}{D[k,k]^2}} \boxed{\frac{v[k]}{D[k,k]}}$$

ML estimator

Beyond denoising: deblurring/deconvolution

Similarly to the case of denoising, the state-of-art deconvolution methods are based on sparsity-constrained least-square solution.

It allows to better account for the sparsity of the signal x in some sparse representation (*e.g.* wavelet, … X-let) as well as noise through the data fidelity term and a choice of the regularization parameter.

$$\hat{x} = \mathrm{Argmin}_x \quad \lambda \|\mathbf{\Phi}^T x\|_{\ell_p} + \frac{1}{2}\|b - Hx\|_2^2$$

No explicit solution !

This can be solved using an iterative thresholded Landweber scheme (Bertero 98):

Adjoint of H, here transpose-conjuguate

$$x^+ = \underbrace{\mathbf{\Phi}\mathcal{S}_\lambda\left[\mathbf{\Phi}^T\right.}\left.(x^- + H^\star(b - Hx^-))\right]$$

thresholding in the sparse domain **thresholding in the sparse domain**

Beyond denoising: deblurring/deconvolution

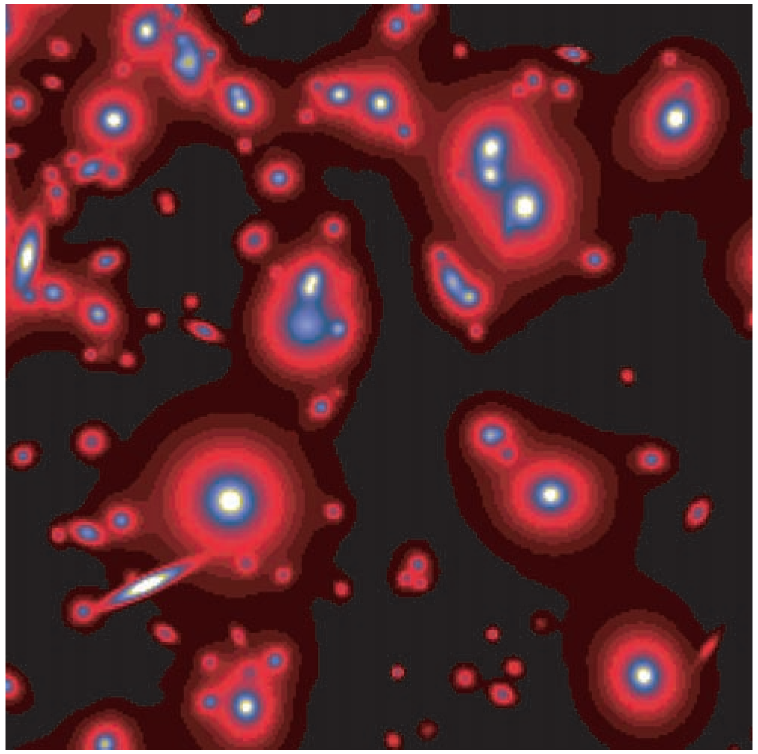
x

b

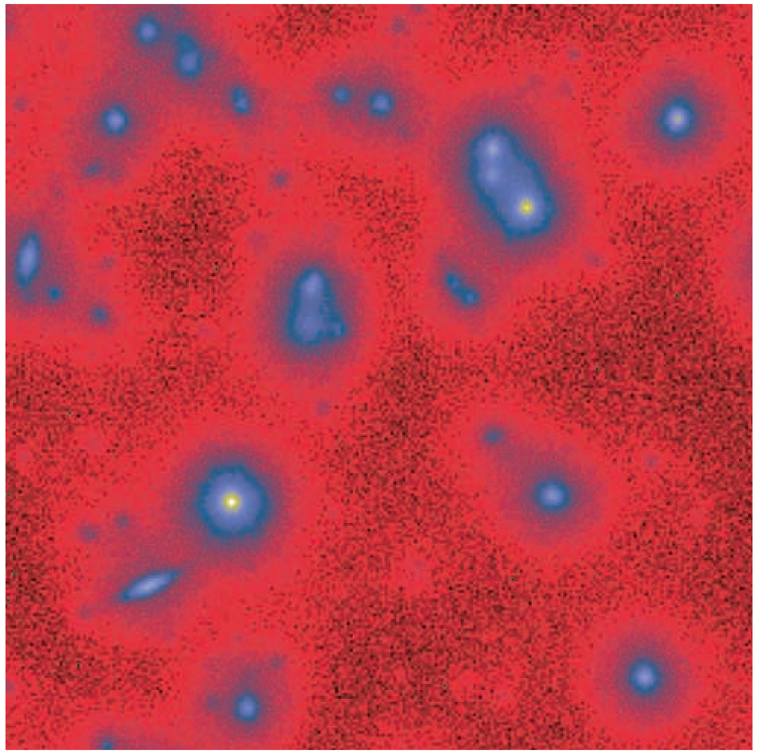
$\hat{x}$

Sparsity-based deconvolution

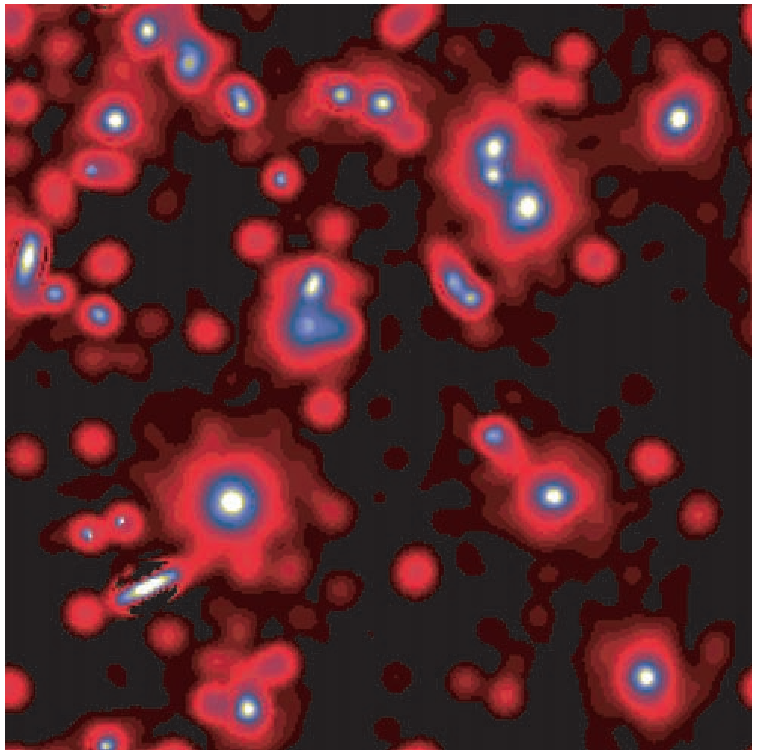
Beyond denoising: deblurring/deconvolution



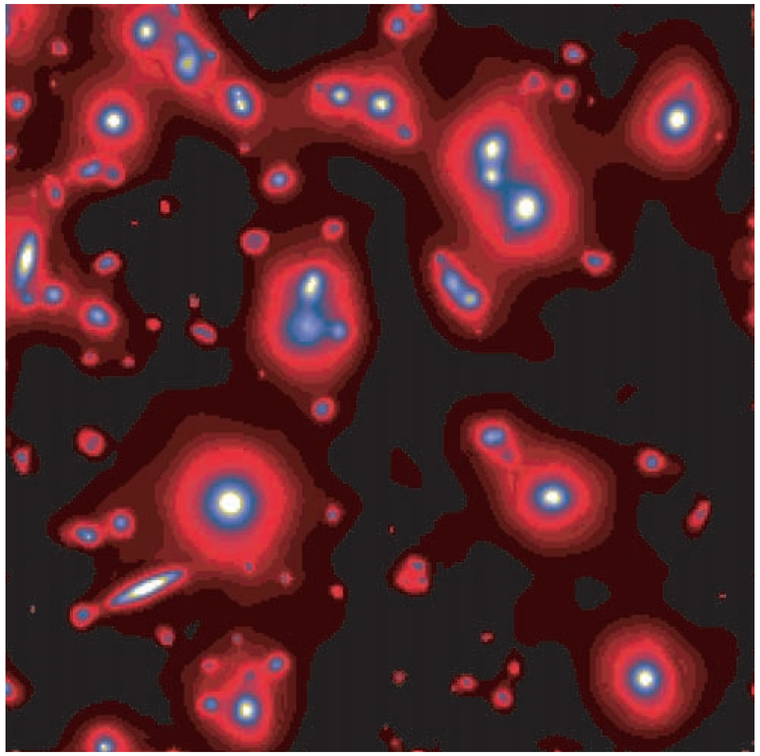
Input



Observation



Richardson-Lucy



Wavelets

Beyond denoising: inpainting

Beyond denoising: inpainting

Inpainting has been tackled by solving a L1-penalized least-square problem of the form:

mask recast as a diagonal matrix

$$\hat{x} = \mathrm{Argmin}_{x=\mathbf{\Phi}\alpha}\ \lambda\|\alpha\|_{\ell_1} + \frac{1}{2}\|b - \mathbf{M}\mathbf{\Phi}\alpha\|_{\ell_2}^2$$

convex but not differentiable

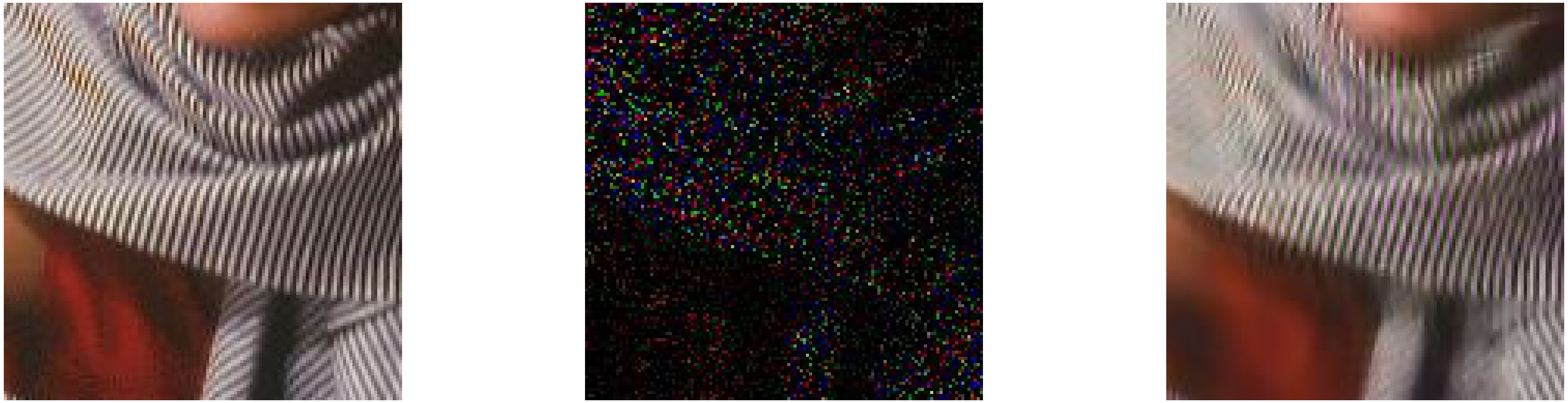
convex and differentiable with 1-Lipschitz gradient

The forward-backward algorithm then reads:

$$\alpha^{(t+1)} = \mathrm{prox}_{\gamma f}\left(\alpha^{(t)} + \gamma\mathbf{\Phi}^T(b - \mathbf{M}\mathbf{\Phi}\alpha)\right)$$

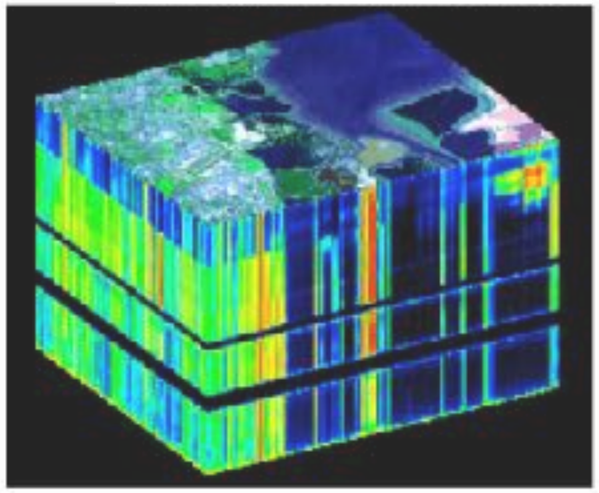
Beyond denoising: deblurring/deconvolution

Φ = [Curvelets, Local DCT]



Blind source separation

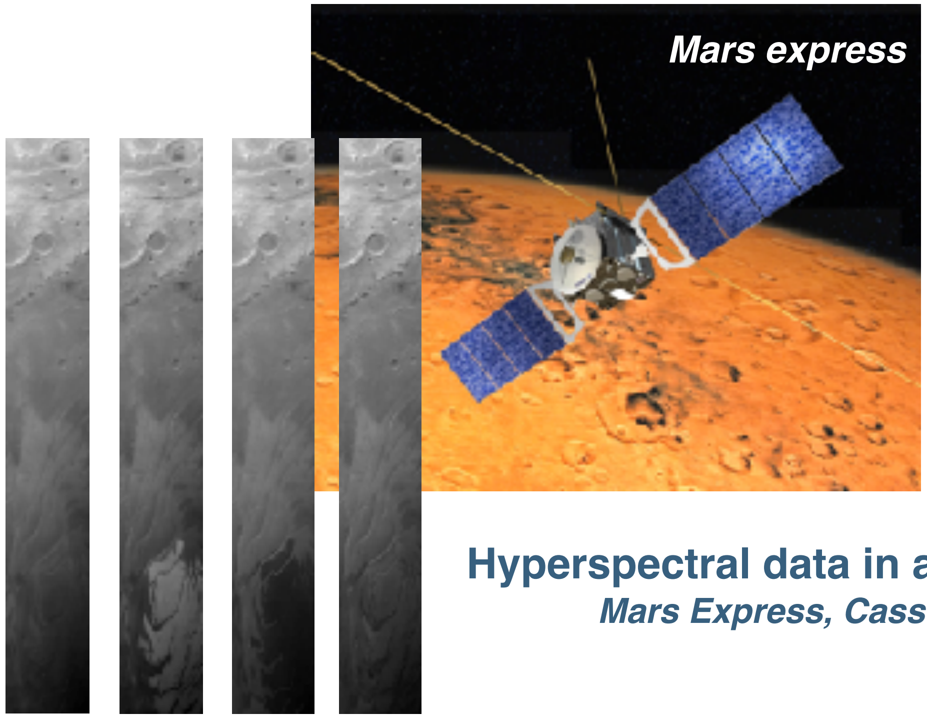
Analysing multispectral data



Courtesy of M. Lennon

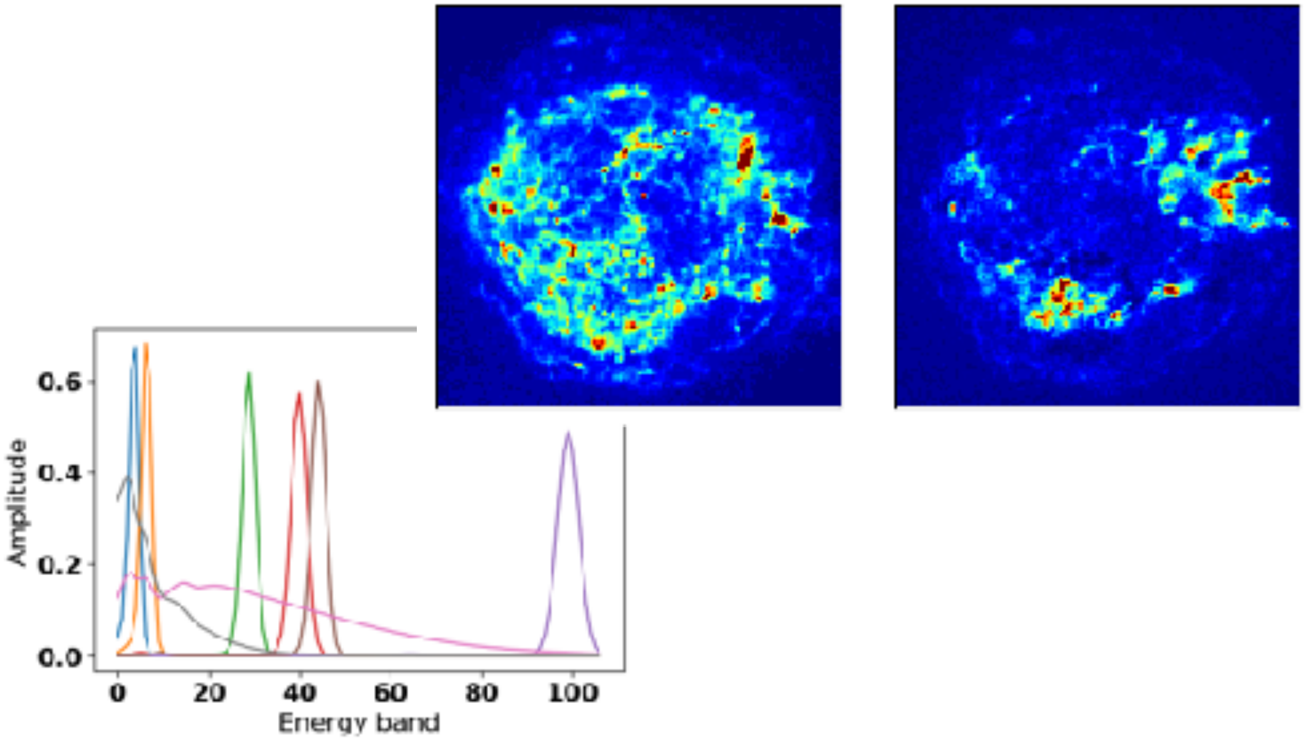
Hyperspectral data

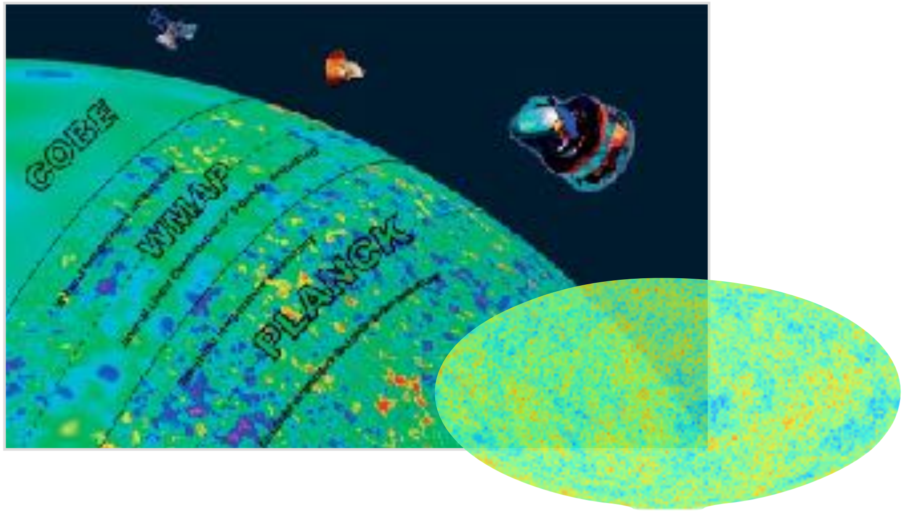
remote sensing, aerial data, etc.



Hyperspectral data in astrophysics

Mars Express, Cassini, etc.

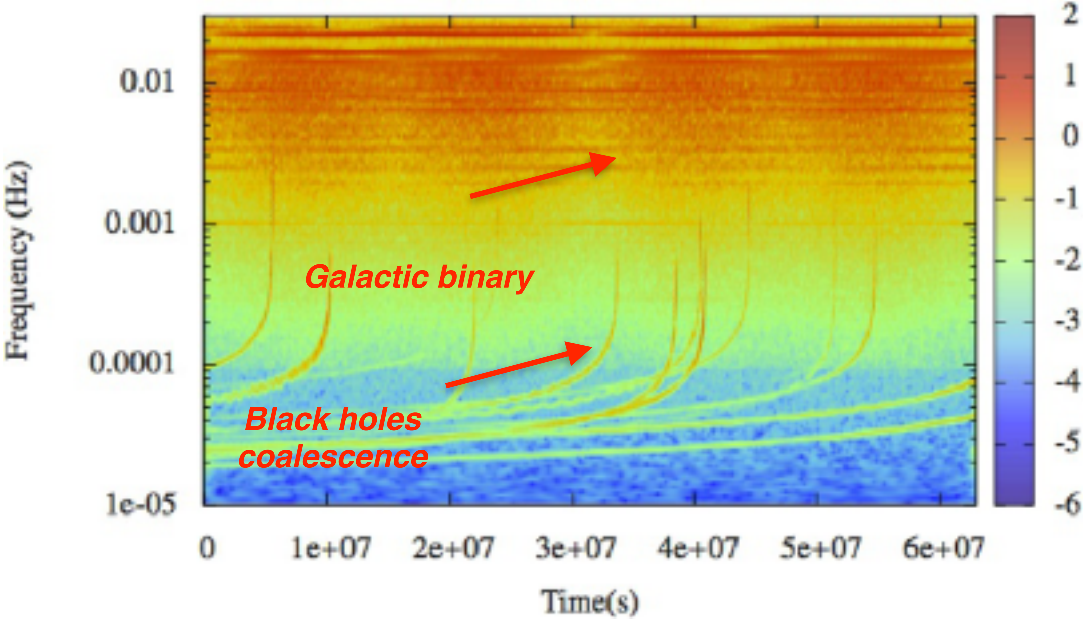




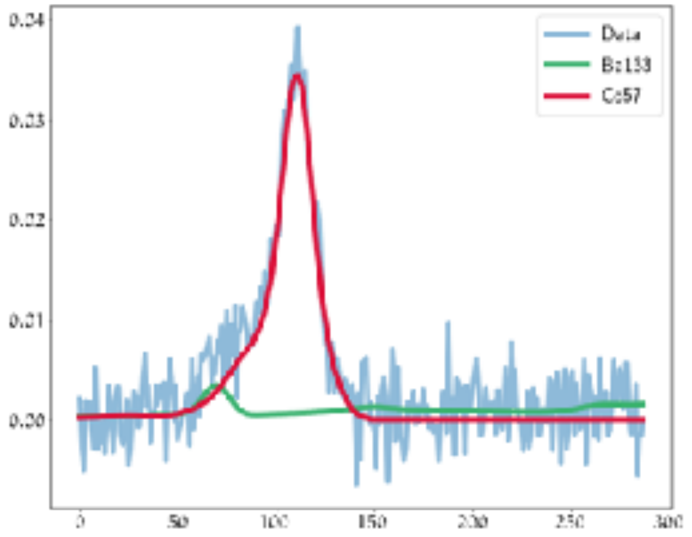
Multispectral data in astrophysics

Planck, Fermi, radio-interferometry (Lofar/SKA/…), etc.

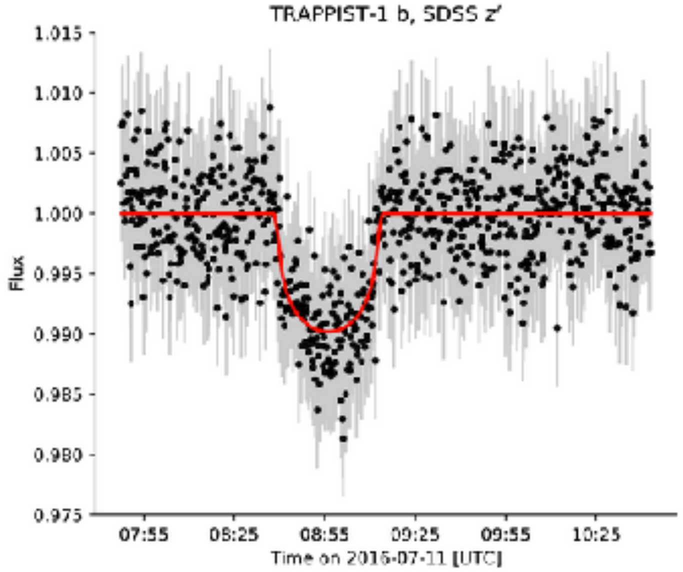
Analysing multispectral data



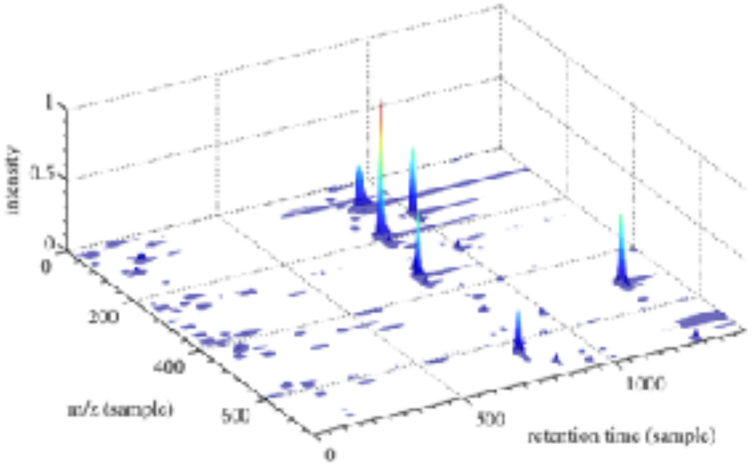
Unmixing gravitational wave signals
From the LISA data



Unmixing γ-ray spectra
to recover radionuclides'activities



Exoplanet detection
From transit observations



Biology
Retrieving metabolite MS spectra from LC/MS data

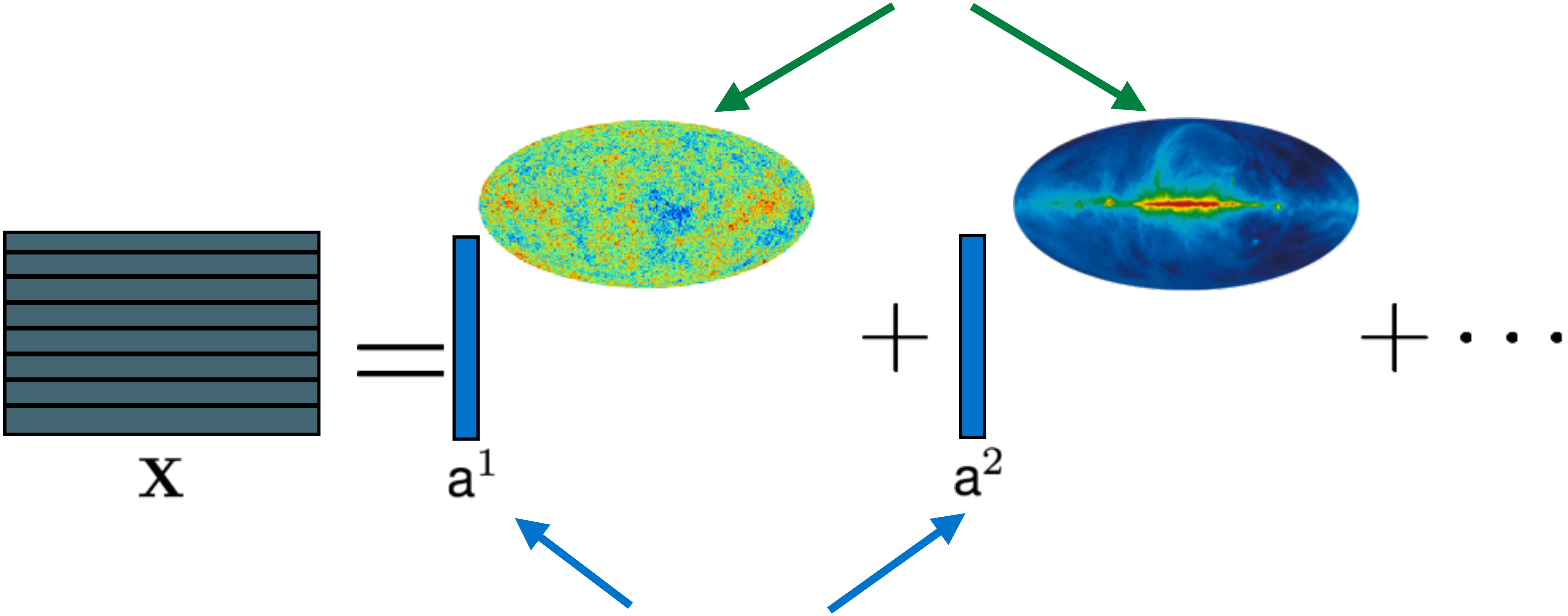
Different scientific fields but …

common problems: mixtures of elementary signals or sources

The underlying mixture model

The linear mixture model

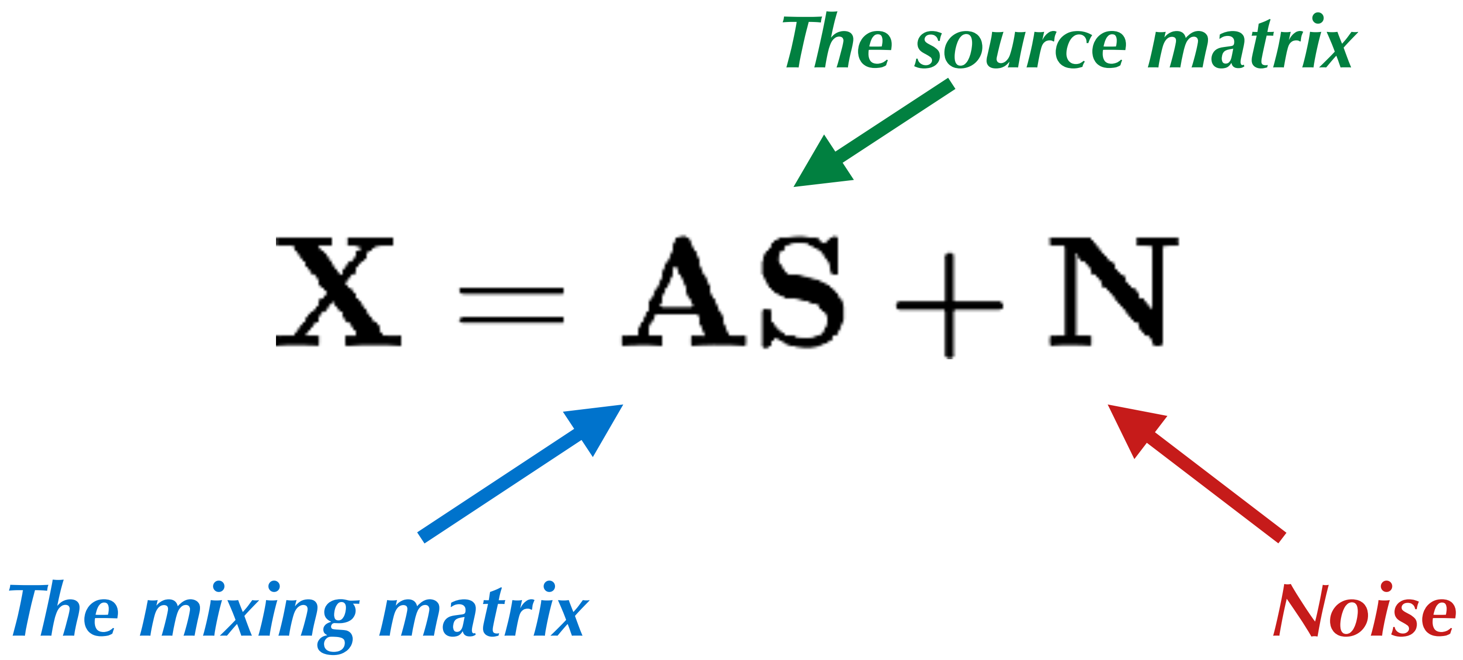
The sources



The mixture weights

e.g. spectral signatures, electromagnetic spectra, etc.

Unsupervised matrix factorisation



The source matrix

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

The mixing matrix

Noise

Blind Source Separation:

Estimation both A and S from X only

This is an ill-posed matrix factorization problem

Non-negative Matrix Factorization, Dictionary Learning, …

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

Instantaneous mixture, non-stationary mixture (e.g. Planck),
Joint convolution/mixing (radio), non-linear mixtures, …

Regularization terms:

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

Sparse BSS - a building block

$$\min_{\mathbf{A},\mathbf{S}} \underbrace{\mathcal{R}(\mathbf{A}) + \mathcal{J}(\mathbf{S})}_{\text{Regularization Terms}} + \frac{1}{2}\underbrace{\|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2}_{\text{Data fidelity term}}$$

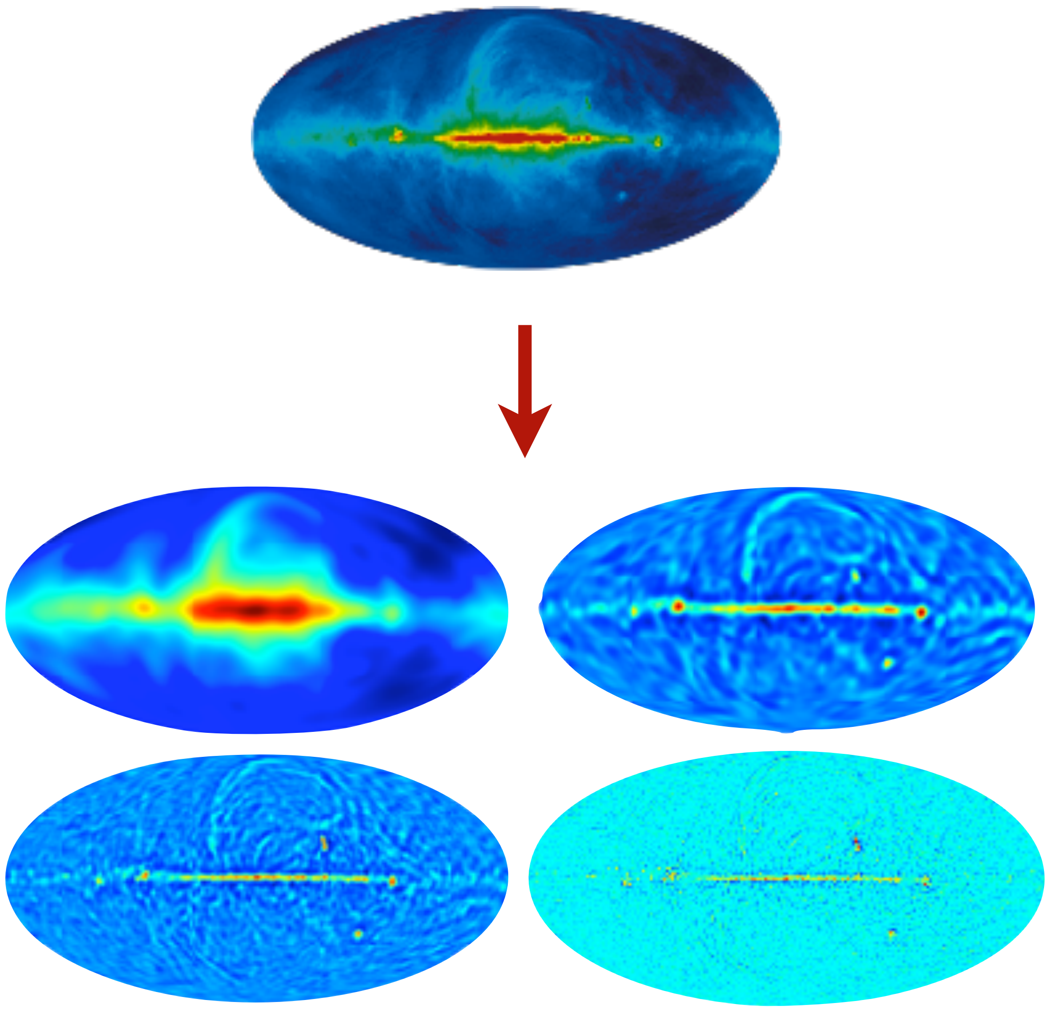
This is an ill-posed matrix factorization problem

Regularization terms: - sparsity of the sources in some signal representation

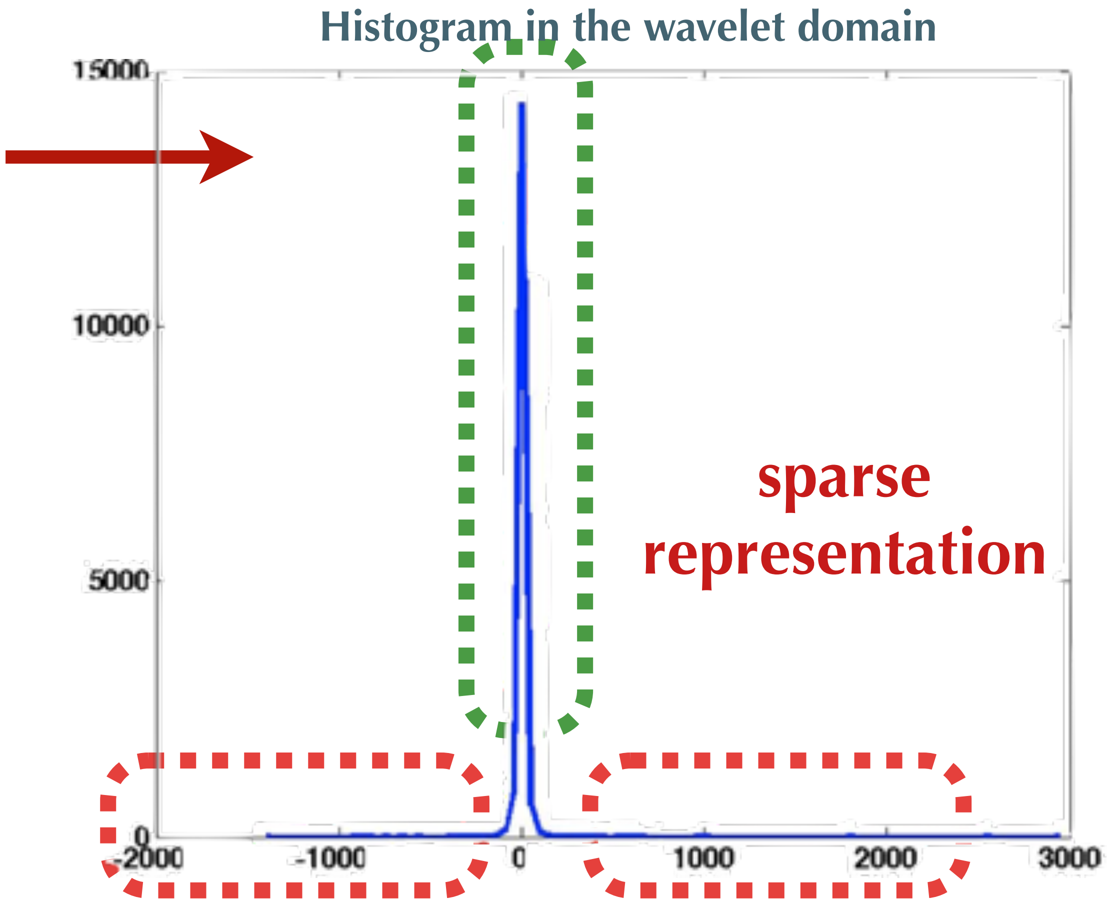
- scaling of the mixing matrix is constrained

Sparse source separation

Changing the way the sources are represented to get a sparse/compressed representation



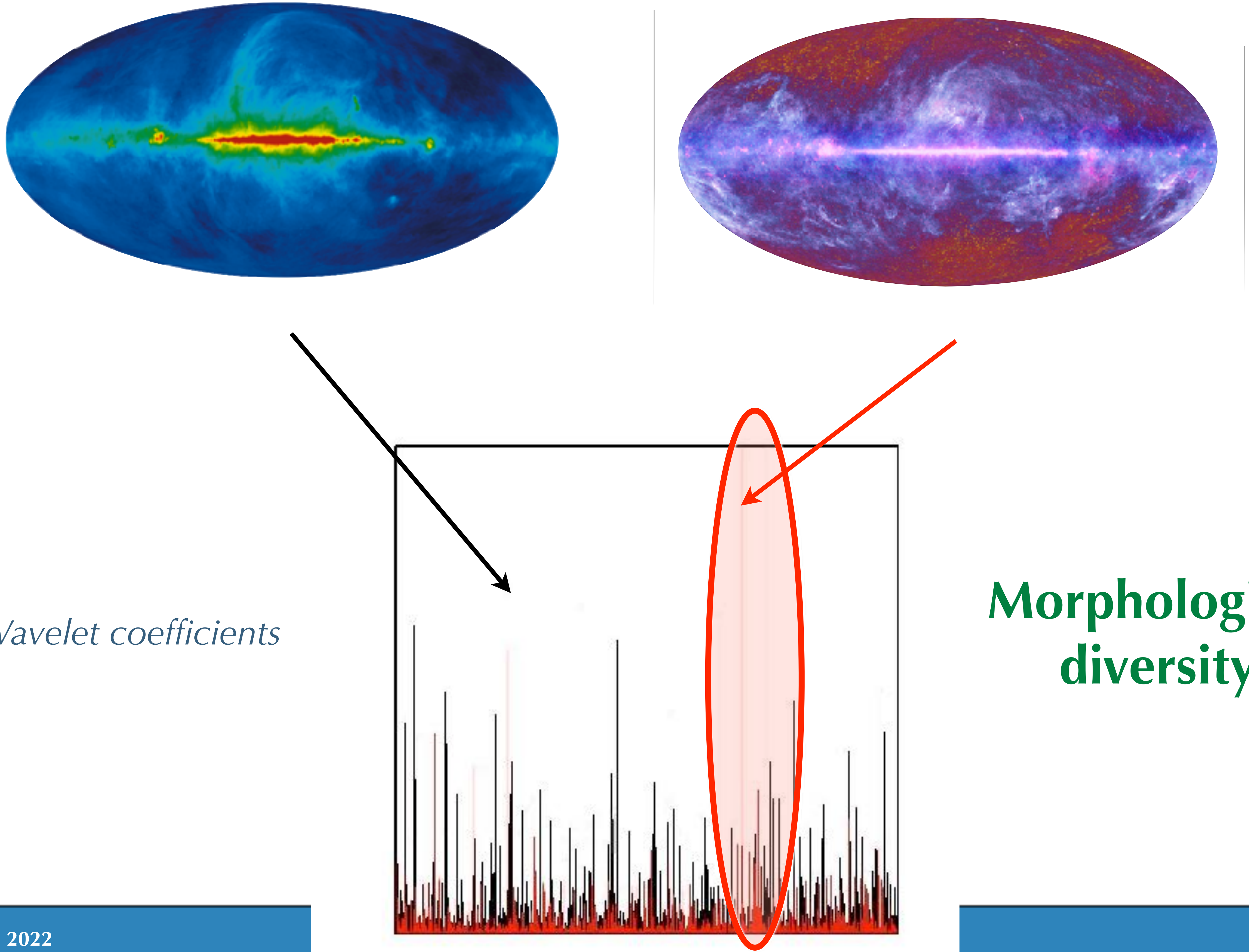
Wavelet transform for spherical data



Histogram in the wavelet domain

sparse representation

Sparse source separation



Wavelet coefficients

Morphological diversity

Sparse source separation

Gist: looking for the sparsest sources

Regularization params., weight matrix, etc.

$$\min_{\mathbf{A},\mathbf{S}} \underbrace{\|\Lambda \odot \mathbf{SW}\|_p}_{\text{Sparse regularization}} + \frac{1}{2} \underbrace{\|\mathbf{X} - \mathbf{AS}\|_F^2}_{\text{Data fidelity term}}$$

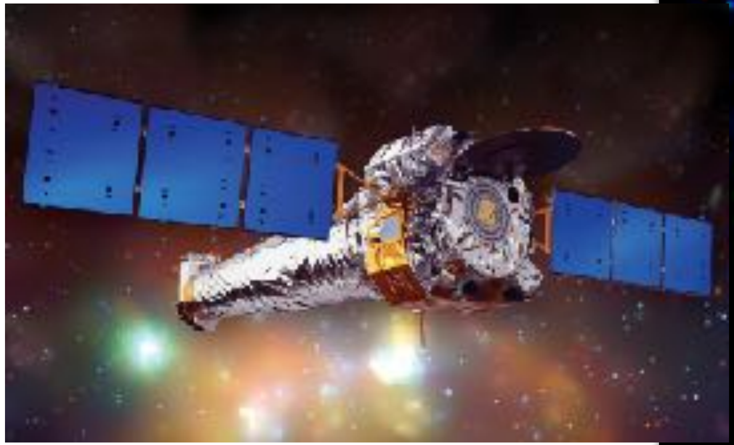
Generalized Morphological Component Analysis (GMCA):

- *S-BSS with redundant sparse representations*
- *Iterative soft/hard thresholding algorithm*
- *Thresholding strategy, robustness to Gaussian noise/local stationary points*
- *No parameters to tune*

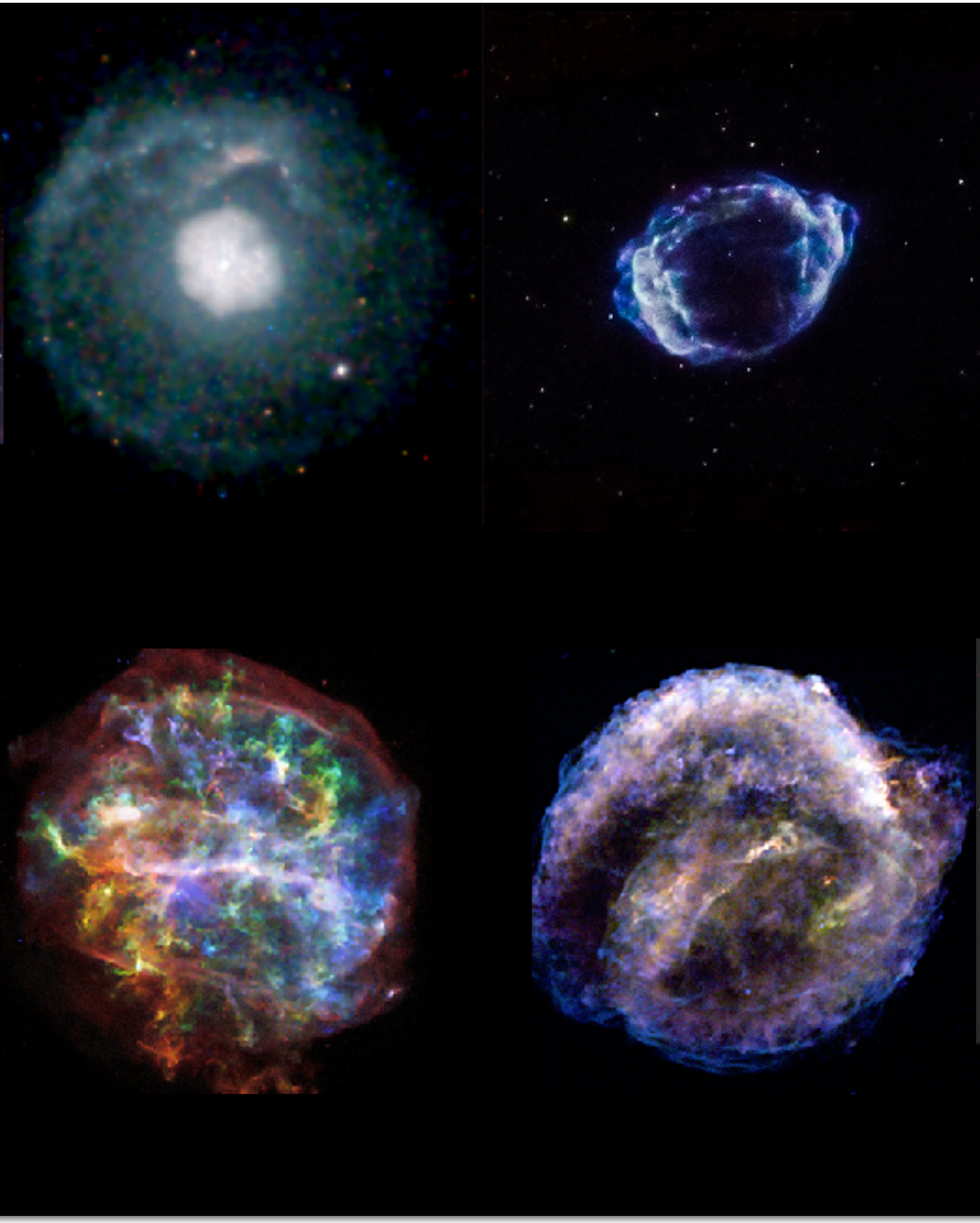
Bobin, Starck, Fadili, and Moudden, Sparsity, Morphological Diversity and Blind Source Separation, IEEE Trans. on Image Processing, Vol 16, No 11, pp 2662 - 2674, 2007.

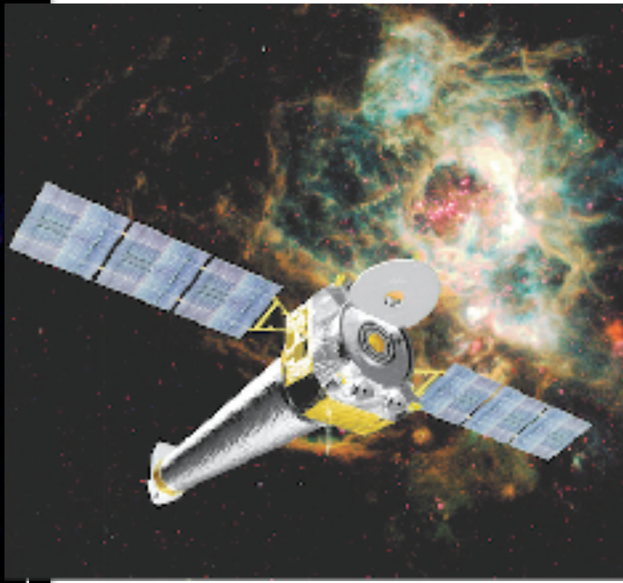
Bobin, Starck, Fadili, and Moudden, Blind Source Separation: The Sparsity Revolution, Advances in Imaging and Electron Physics , Vol 152, pp 221 -- 306, 2008.

Unmixing X-ray images



NASA - Chandra

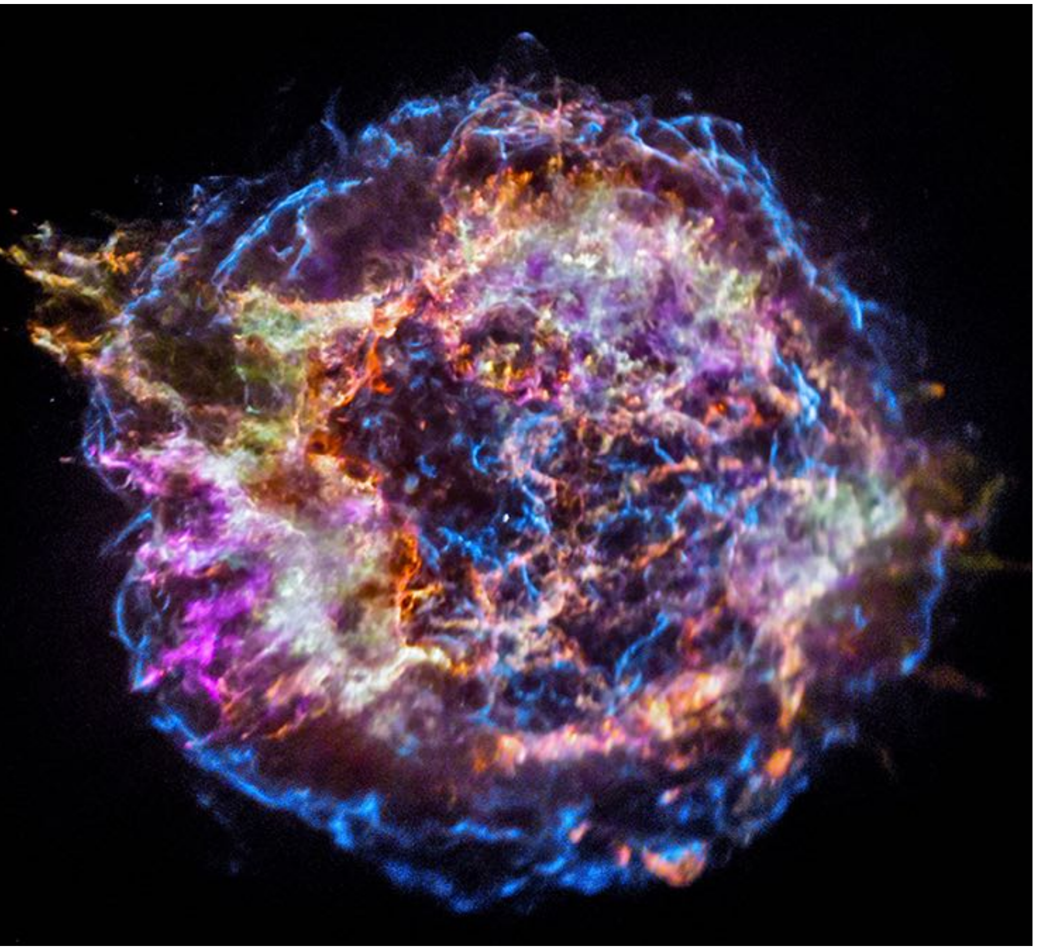


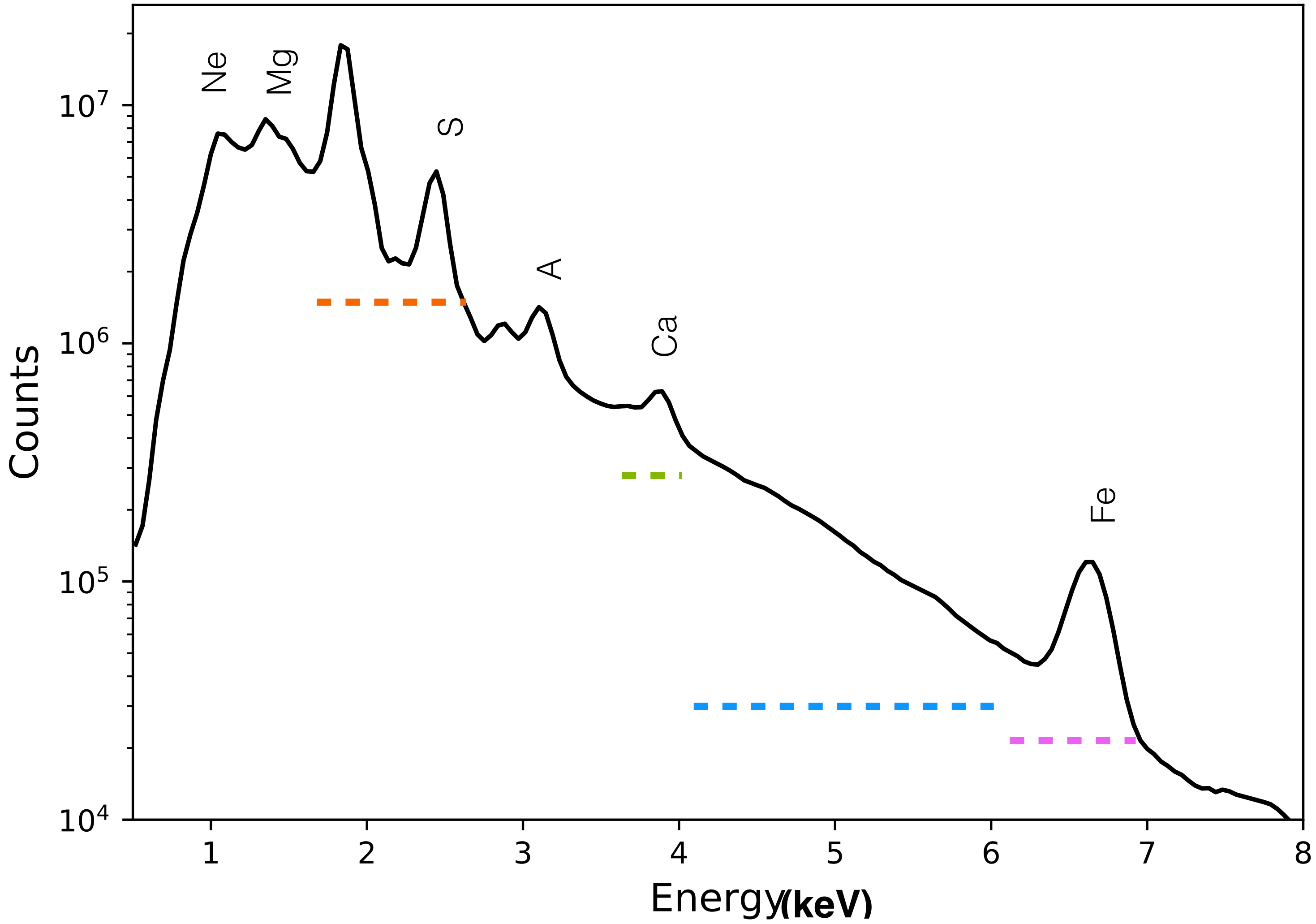


ESA - Athena
launch in 2034

Unmixing X-ray images

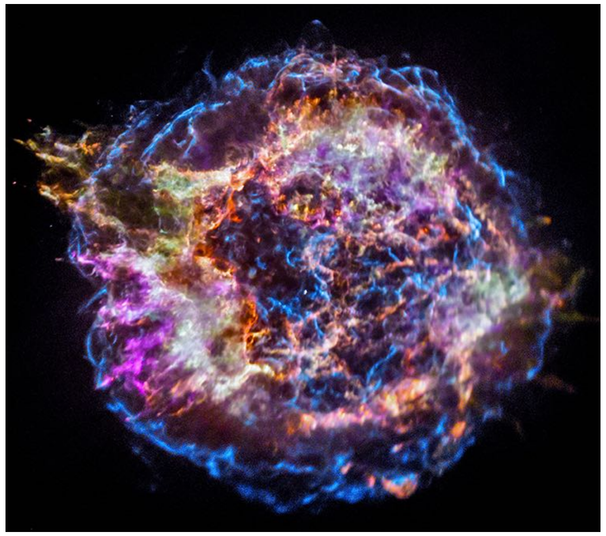
CasA with Chandra
1 Ms observation
~1 billions counts !!

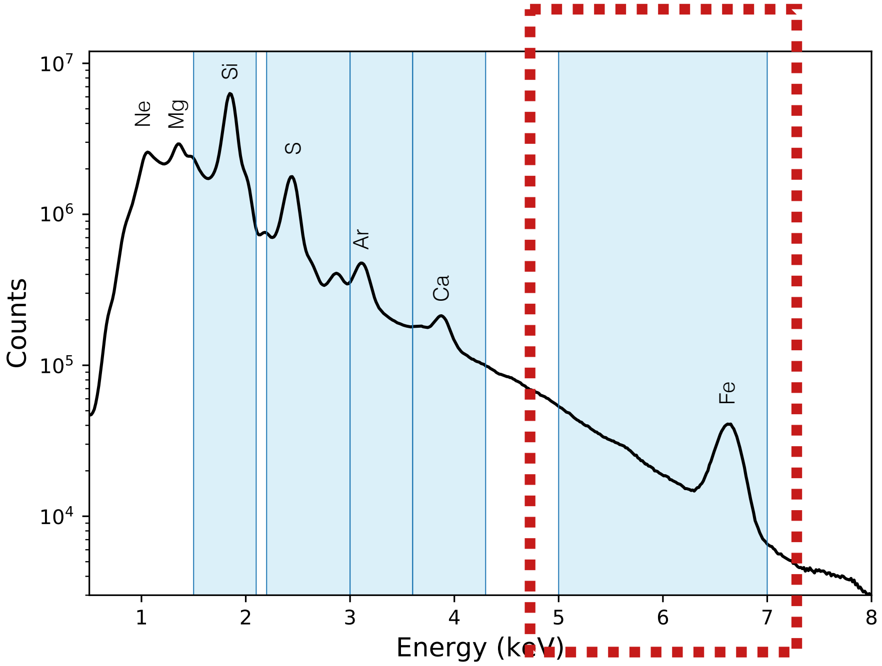


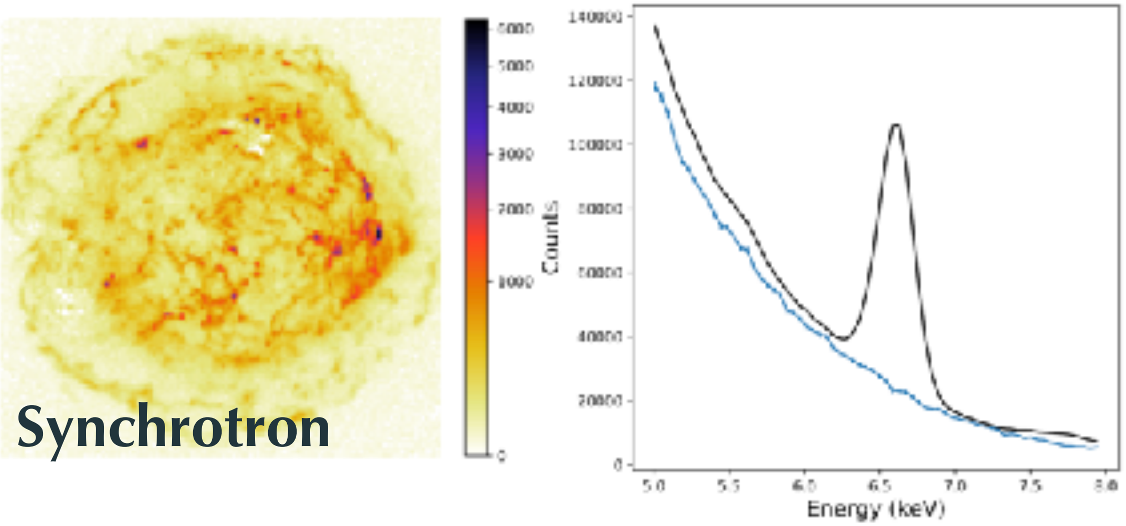


- **Ejecta thermal emission gives insight on :**
  - **Individual elements distribution**
  - **Morphology, asymmetries**
  - **Velocities**

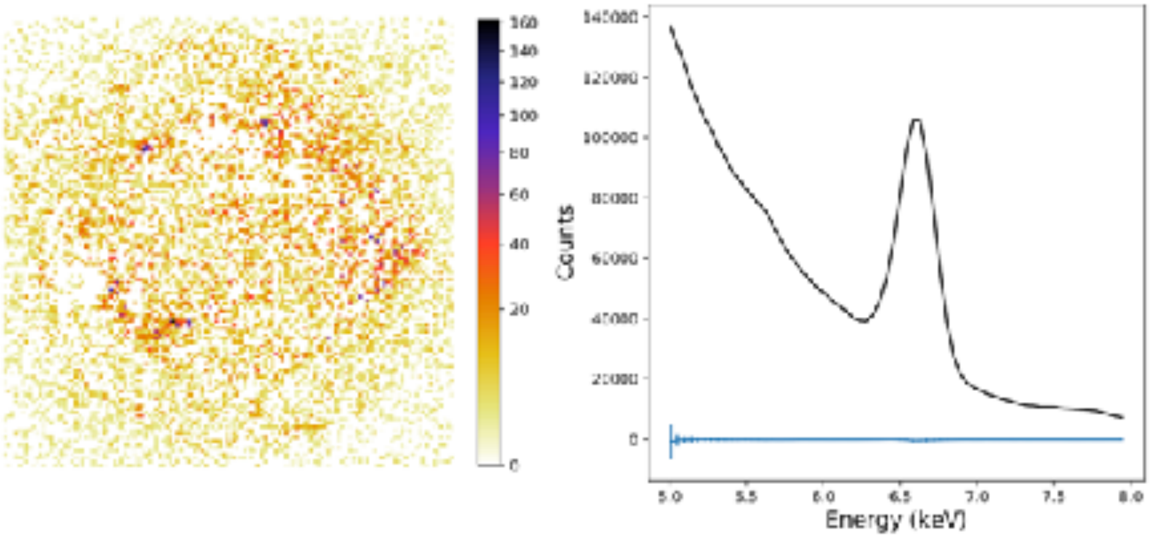
Application to the Chandra data

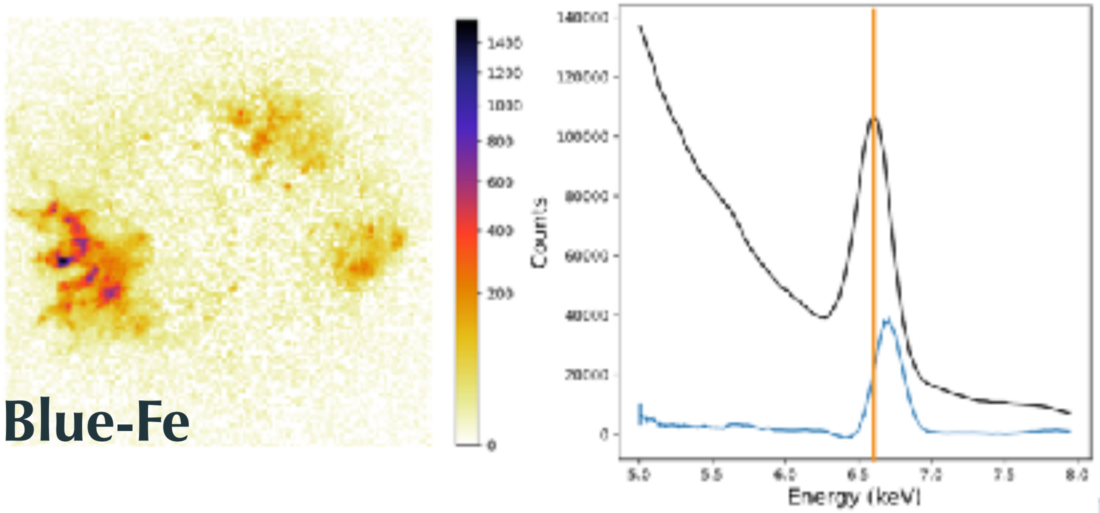




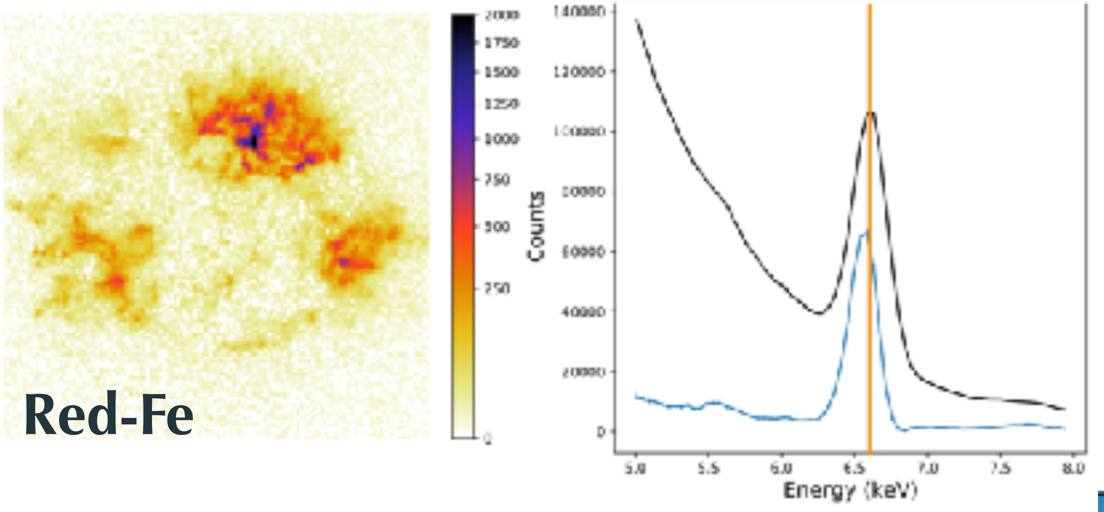


Synchrotron



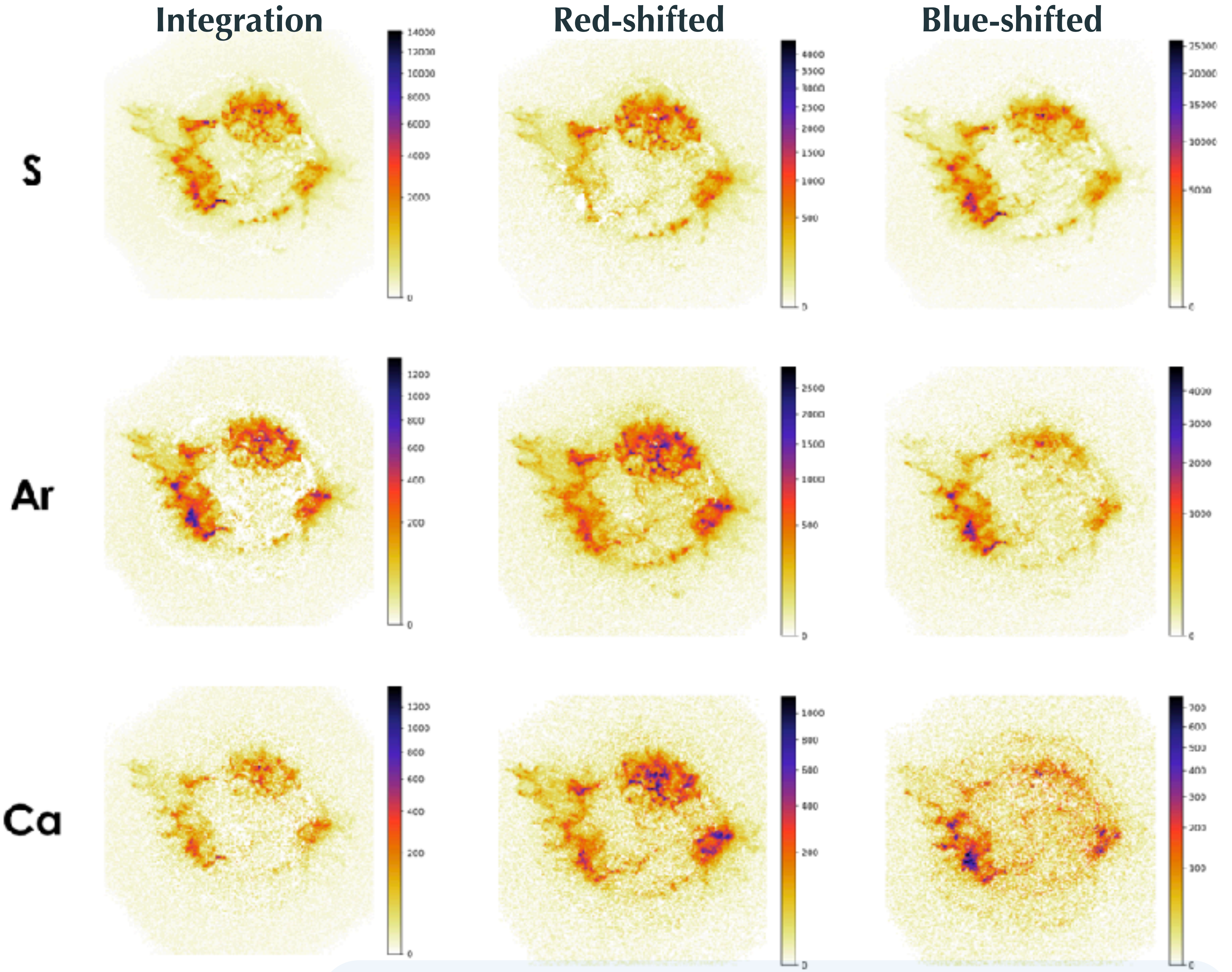


Blue-Fe



Red-Fe

Application to the Chandra data



Picquenot et al, A&A, 2019.

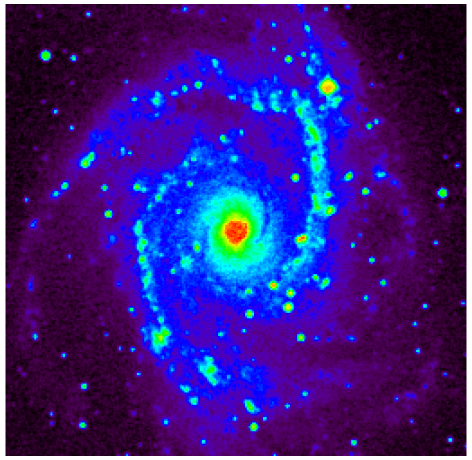
Blindly estimates red/blue-shifted atomic components !

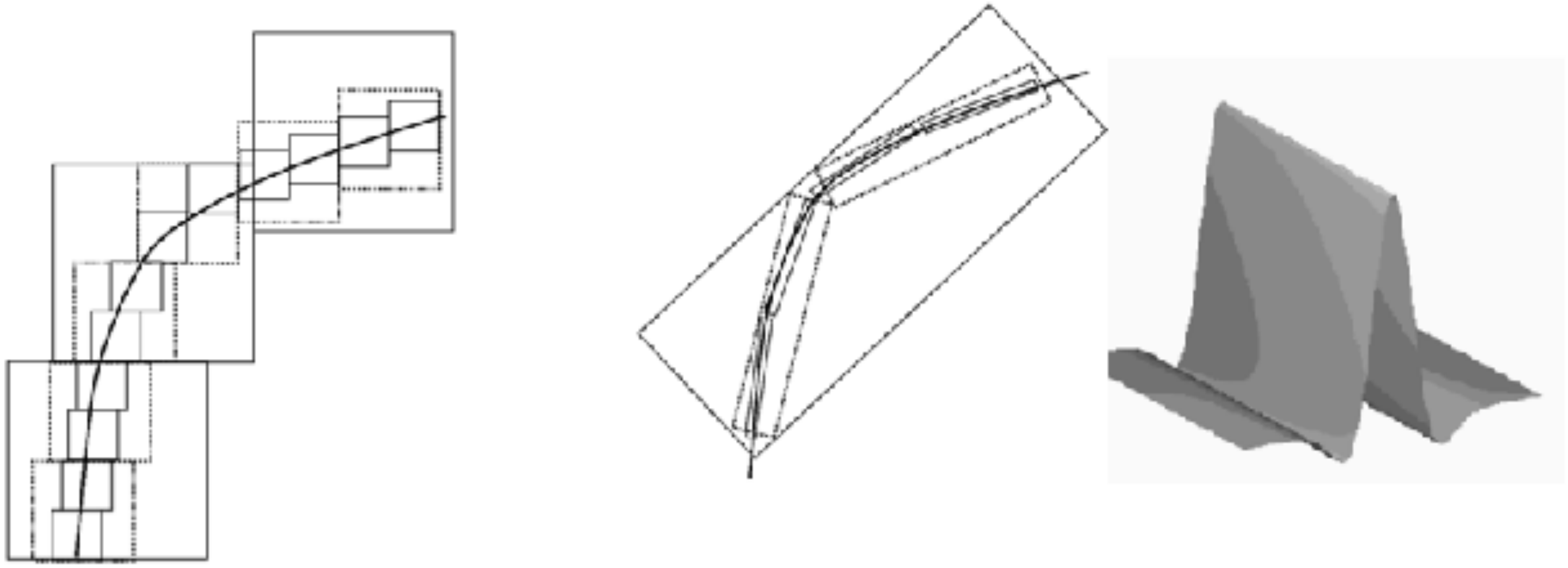
Beyond sparse modelling

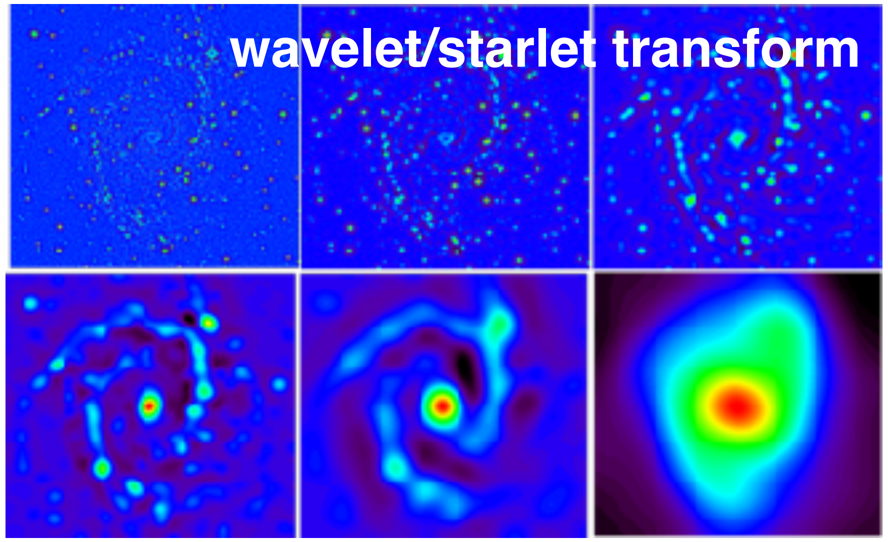
Sparse modelling

All along these courses, we have explored how signals representations can be built to obtain sparse representations.

All these representations are built on certain generic morphological/geometrical specificities of the signal to be modeled.







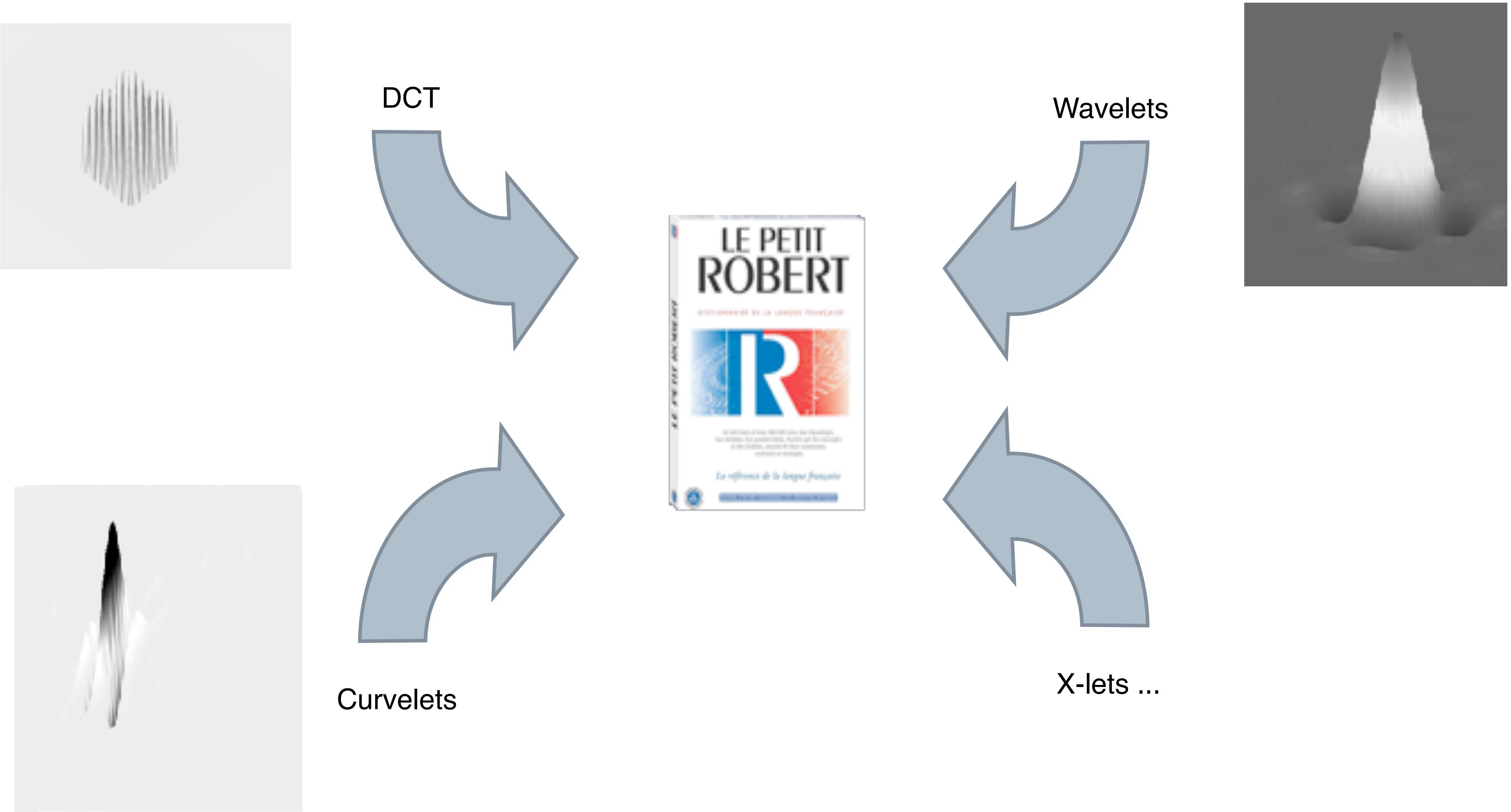
ridgelets, curvelets, contourlets, bandelettes, etc.

Main advantage: they are adapted to the content of a very large span of “natural” images, they come with fast transforms.

Main drawback: they are not specifically adapted to the content of individual signal/images/… which might typical of specific data/applications, etc.

Sparse modelling

All these signal representations can be combined to sparsely represent more complex images that combines morphologies of various nature; see Morphological Component Analysis (MCA), etc.



DCT

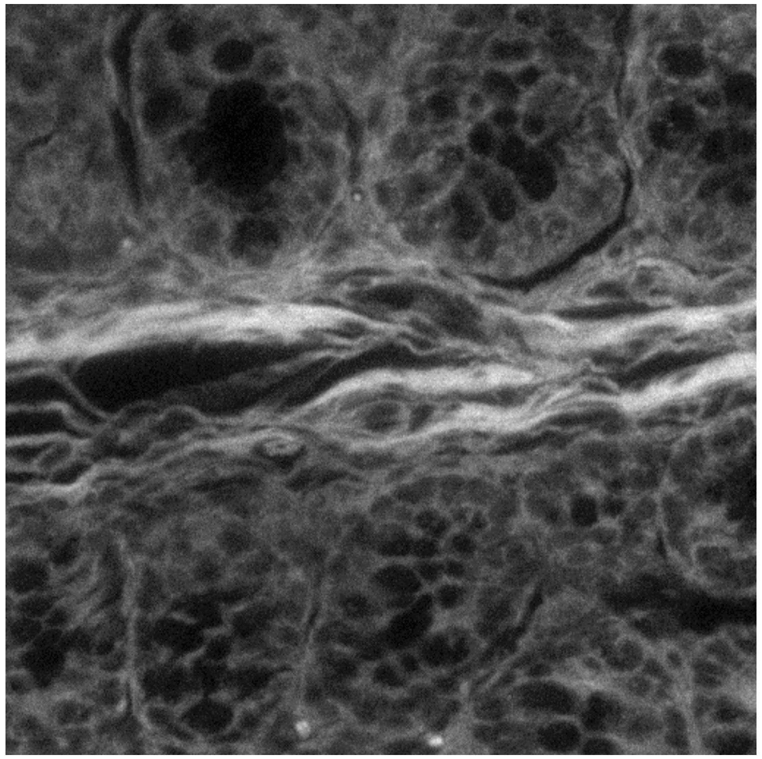
Wavelets

Curvelets

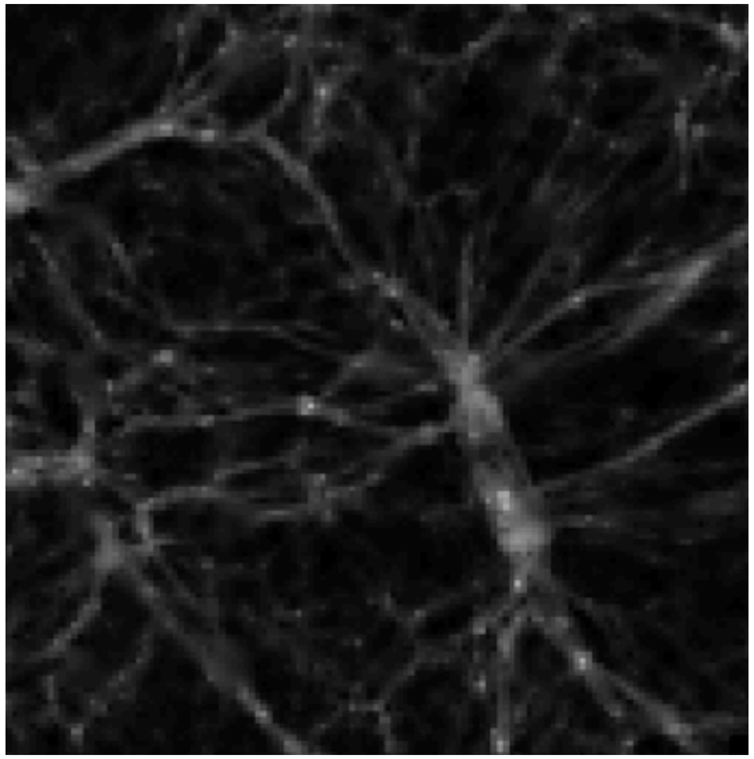
X-lets ...

Dictionary learning

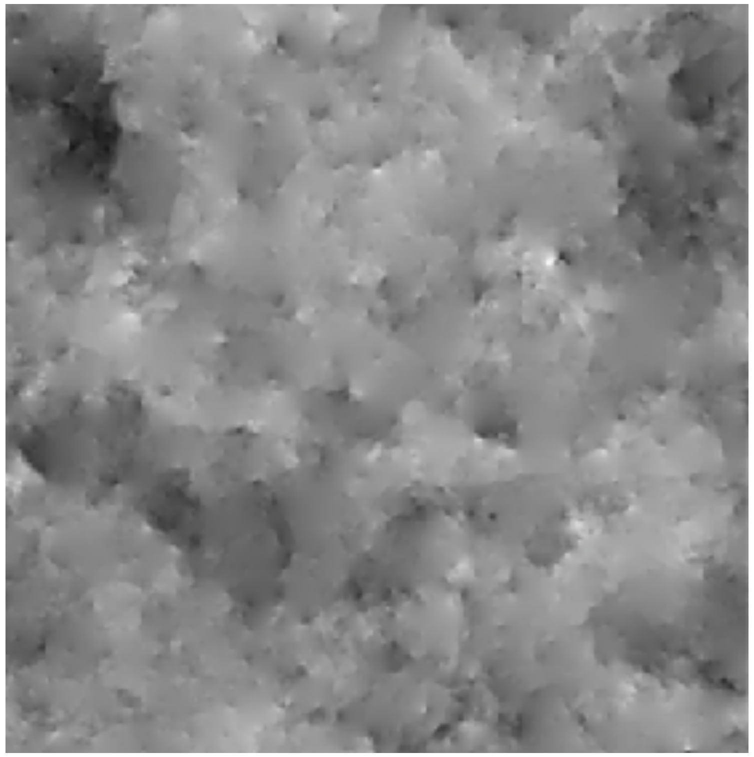
What about this very specific signals ?



Tissue observed through a confocal microscope



Simulation of the Cosmic Web (galaxy distribution)



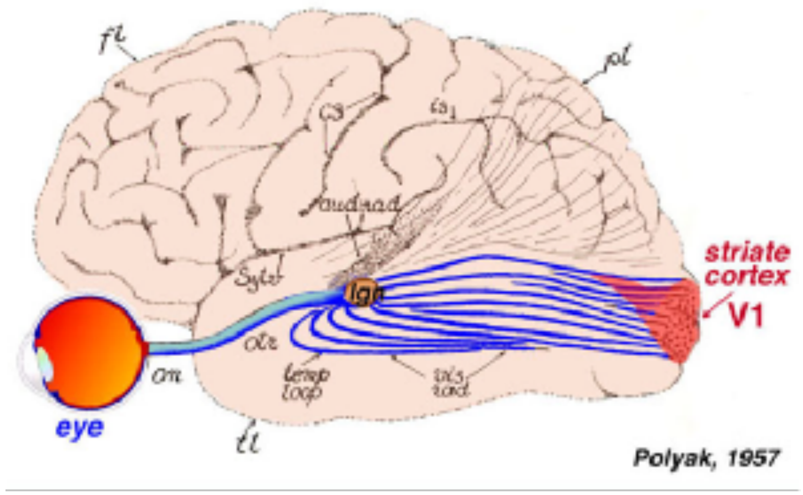
Simulation of cosmic strings

Ideally, one would like to learn a dictionary/sparse representation that is adapted to the specific morphological component of a single image or a class of images.

Dictionary learning

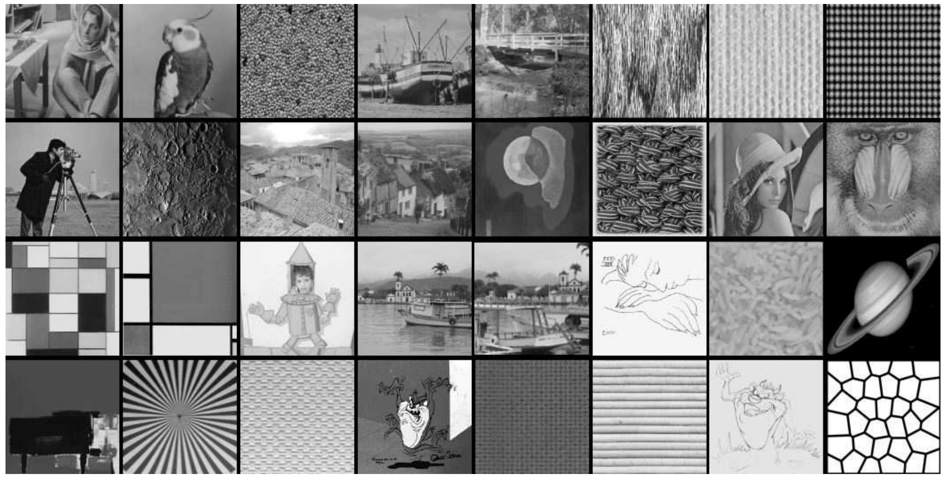
The question of learning a “relevant” representation for natural images first was first raised in the field of neuroscience by Olshausen & Fields.

Their work focused on understanding the kind of patterns in natural images the primary visual cortex (V1) is sensitive to.



Polyak, 1957

***Learning* experiment:**



each image is divide into small patches and stored in some matrix X

$$\min_{\mathbf{\Phi},\mathbf{S}} \|\mathbf{X} - \mathbf{S}\mathbf{\Phi}\|_F^2 + \lambda \sum_i f(S_i)$$

Signal representation

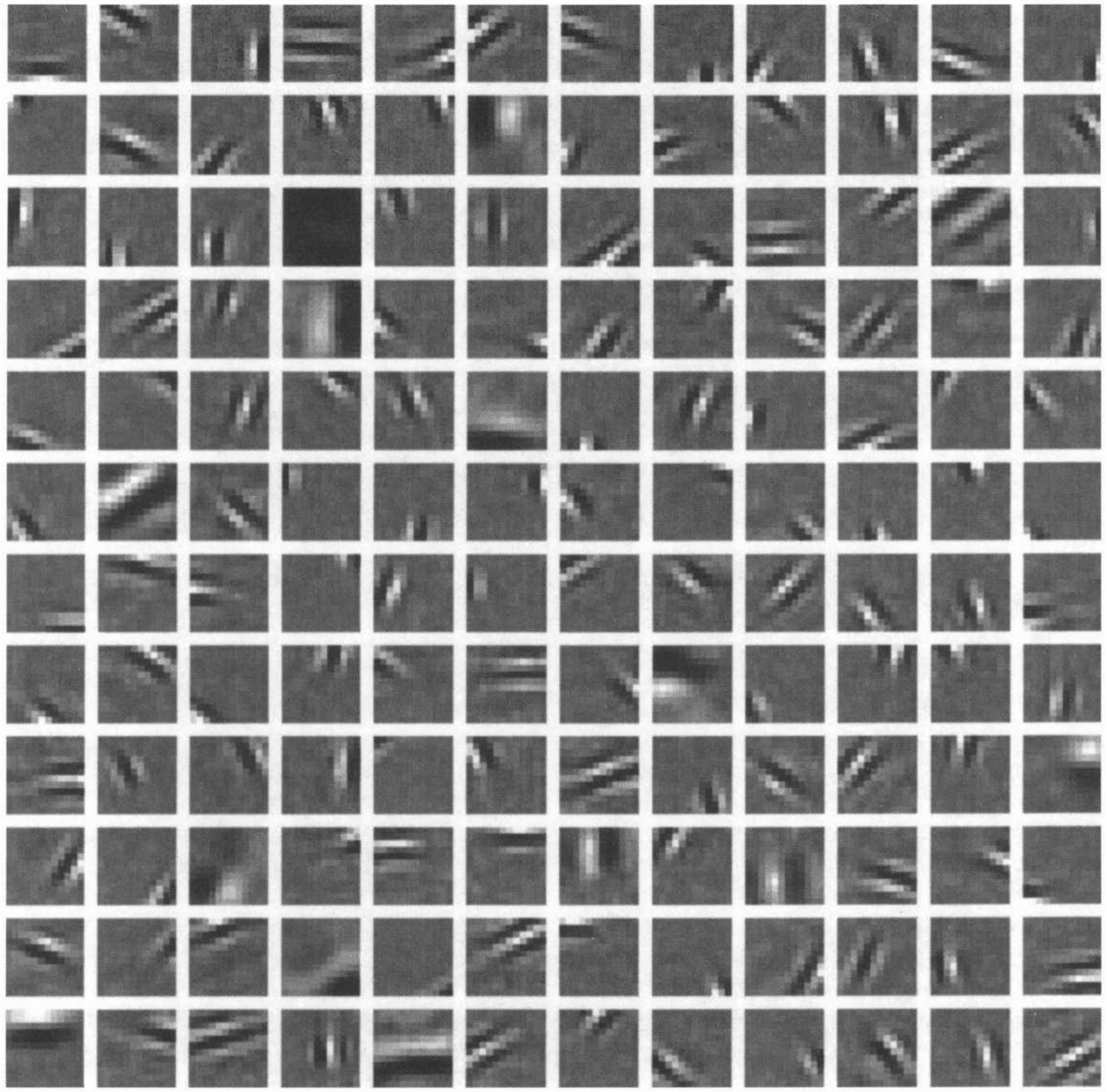
Expansion coefficient of each data patch

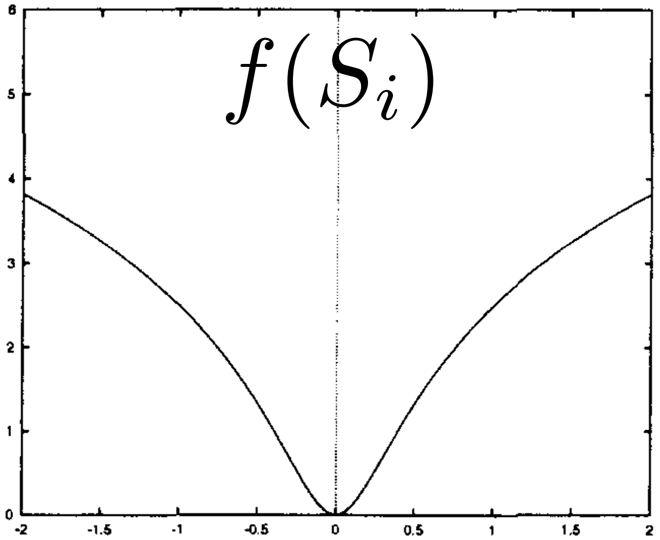
Olshausen & Fields, Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1 ?, 1997

Dictionary learning

In their experiment, they chose the penalization f so as to promote sparseness since it is believed that the primary visual cortex compresses information into a few significant features.

$$\min_{\mathbf{\Phi},\mathbf{S}} \|\mathbf{X} - \mathbf{S\Phi}\|_F^2 + \lambda \sum_i f(S_i)$$



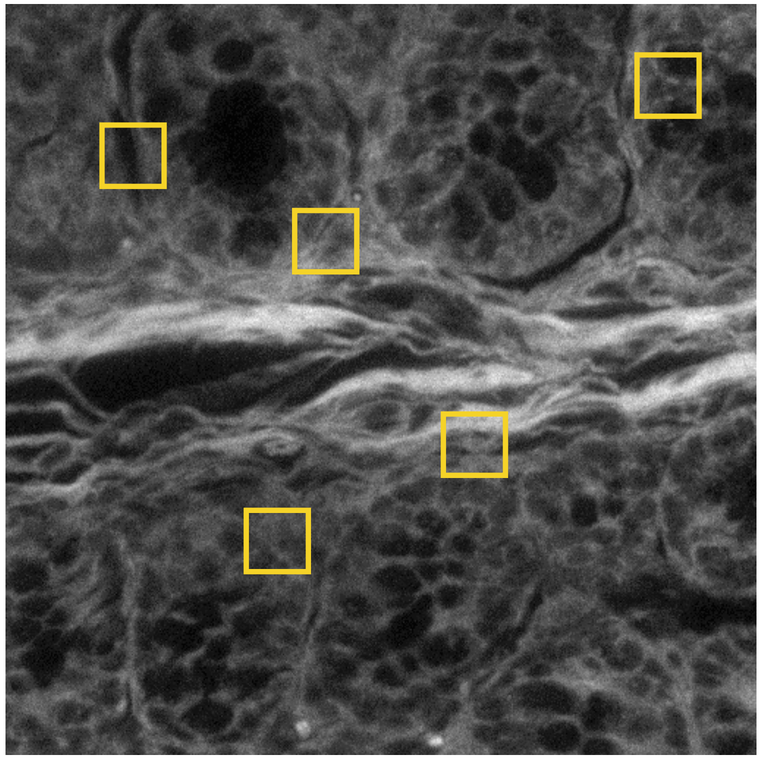


They found that the learnt dictionary contains oriented/elongated/oscillatory features which are reminiscent of Gabor-wavelets.

Dictionary learning

In the setting of imaging processing, the use of overcomplete representations has long been advocated as a mean to provide sparser and therefore more efficient signal representations.

Learning sparse overcomplete signal representation has been introduced by Elad and Aharon in 2006.



Images are modeled in a patch-based representation

$$\min_{\mathbf{\Phi},\alpha} \sum_j \mu_j \|\alpha_j\|_{\ell_0} + \|x_j - \mathbf{\Phi}\alpha_j\|_2^2$$

Expansion coefficient of each data patch

Signal representation

Elad & Aharon, Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries , 2006

Dictionary learning

The K-SVD algorithm can be described as follows:

- **Initialization, Φ is set to an overcomplete DCT**
- **Sparse coding:**

  Decompose each data patch in the current dictionary:

$$\min_{\alpha_j} \|\alpha_j\|_{\ell_0} \text{ s.t. } \|x_j - \mathbf{\Phi}\alpha_j\|_2 \leq \epsilon$$

- **Dictionary update:**

  For each atom (element of the dictionary), build the residual:

$$\mathbf{R}_k = \mathbf{X} - \sum_{k' \neq k} \sum_j \alpha_j[k']\phi_{k'}$$

part of the data "explained" the atoms except atom #k

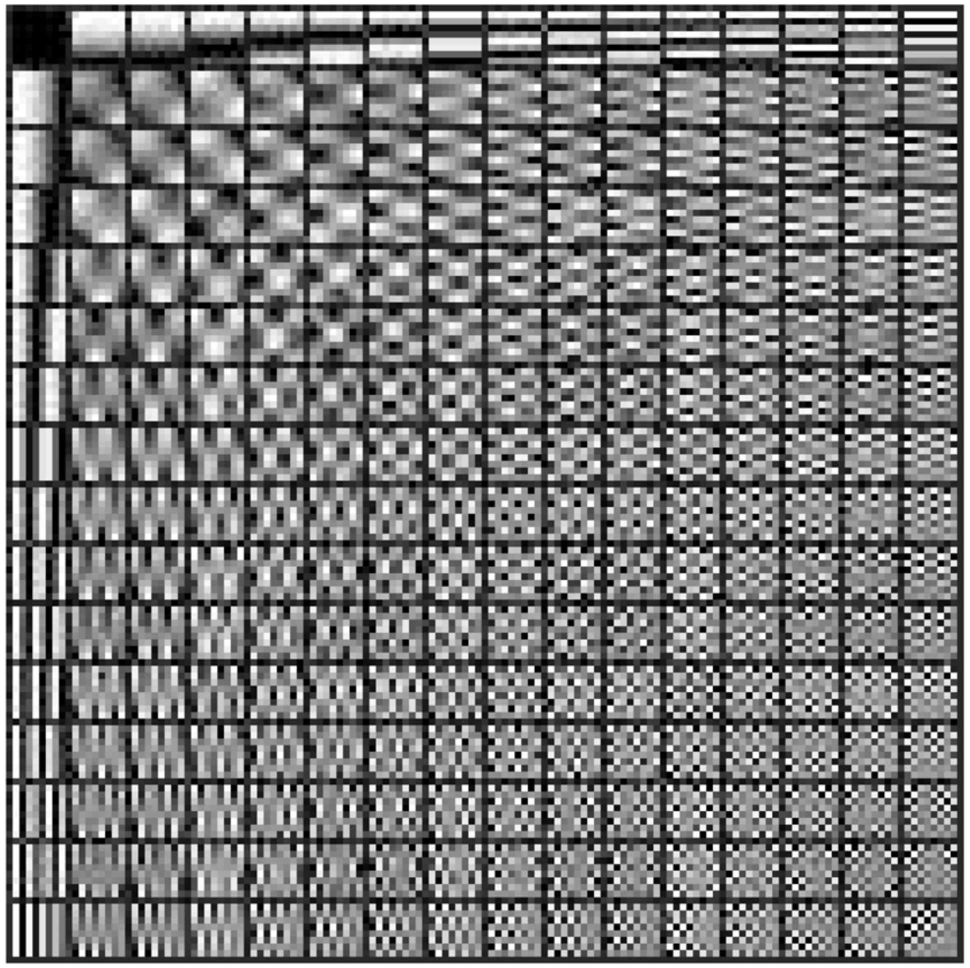
The atom and the corresponding decomposing are updated via SVD:

$$\mathbf{R}_k = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k \longrightarrow \begin{array}{l} \phi_k = \mathbf{U}_k^1 \\ \alpha^k = \mathbf{V}_{k,1} \end{array}$$

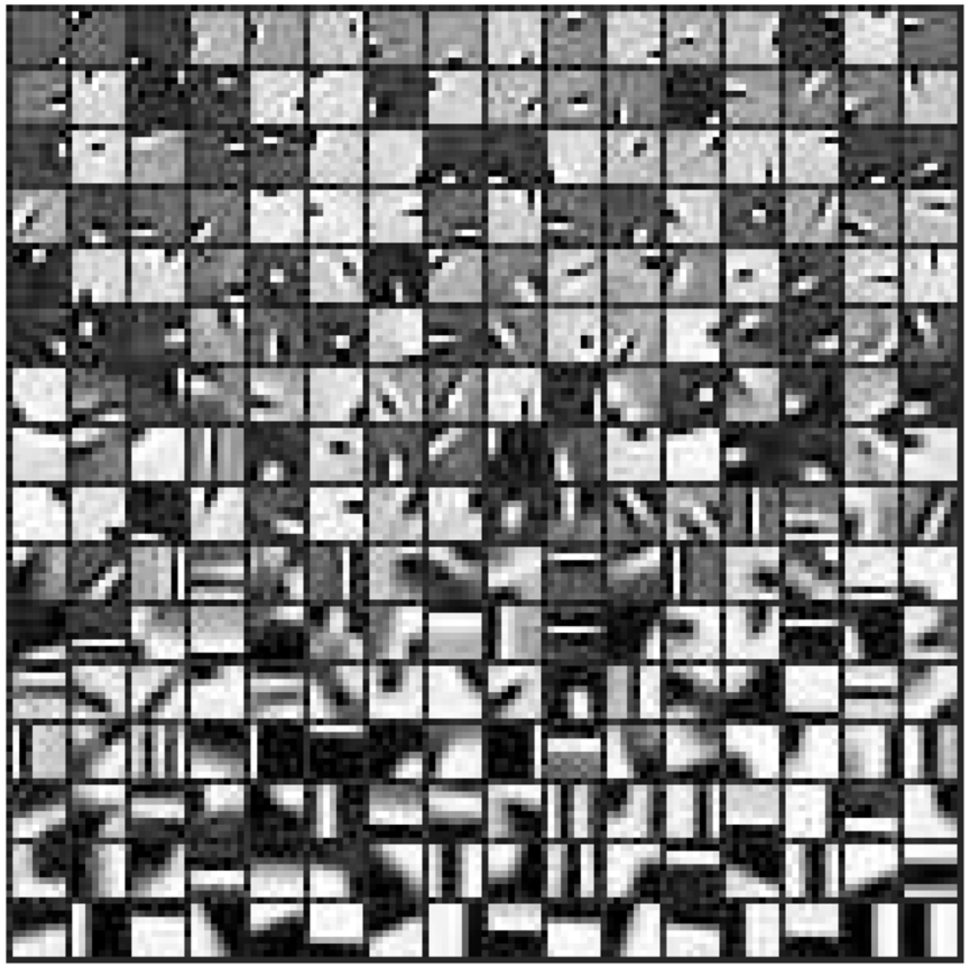
take the principal eigenvectors

Dictionary learning

Sample images



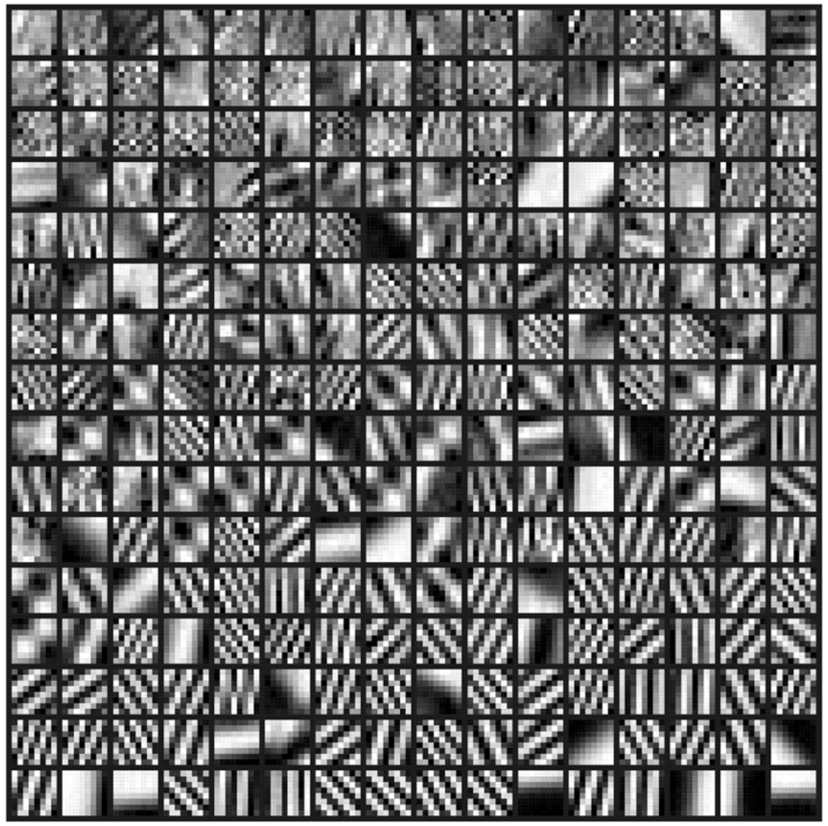
Overcomplete DCT dictionary



Learnt dictionary with K-SVD

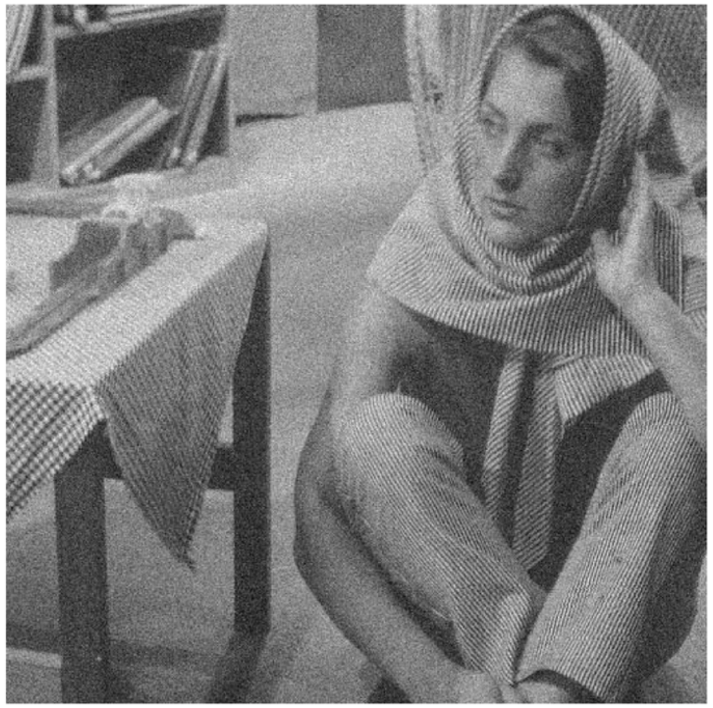
Elad & Aharon, Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries , 2006

Dictionary learning



Created Adaptive Dictionary

Original Image



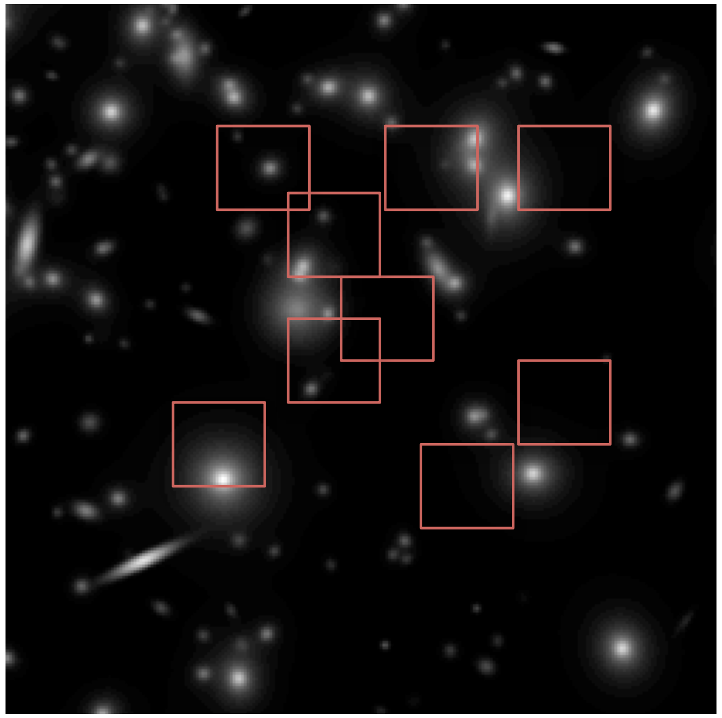
Noisy Image (22.1307 dB, σ=20)

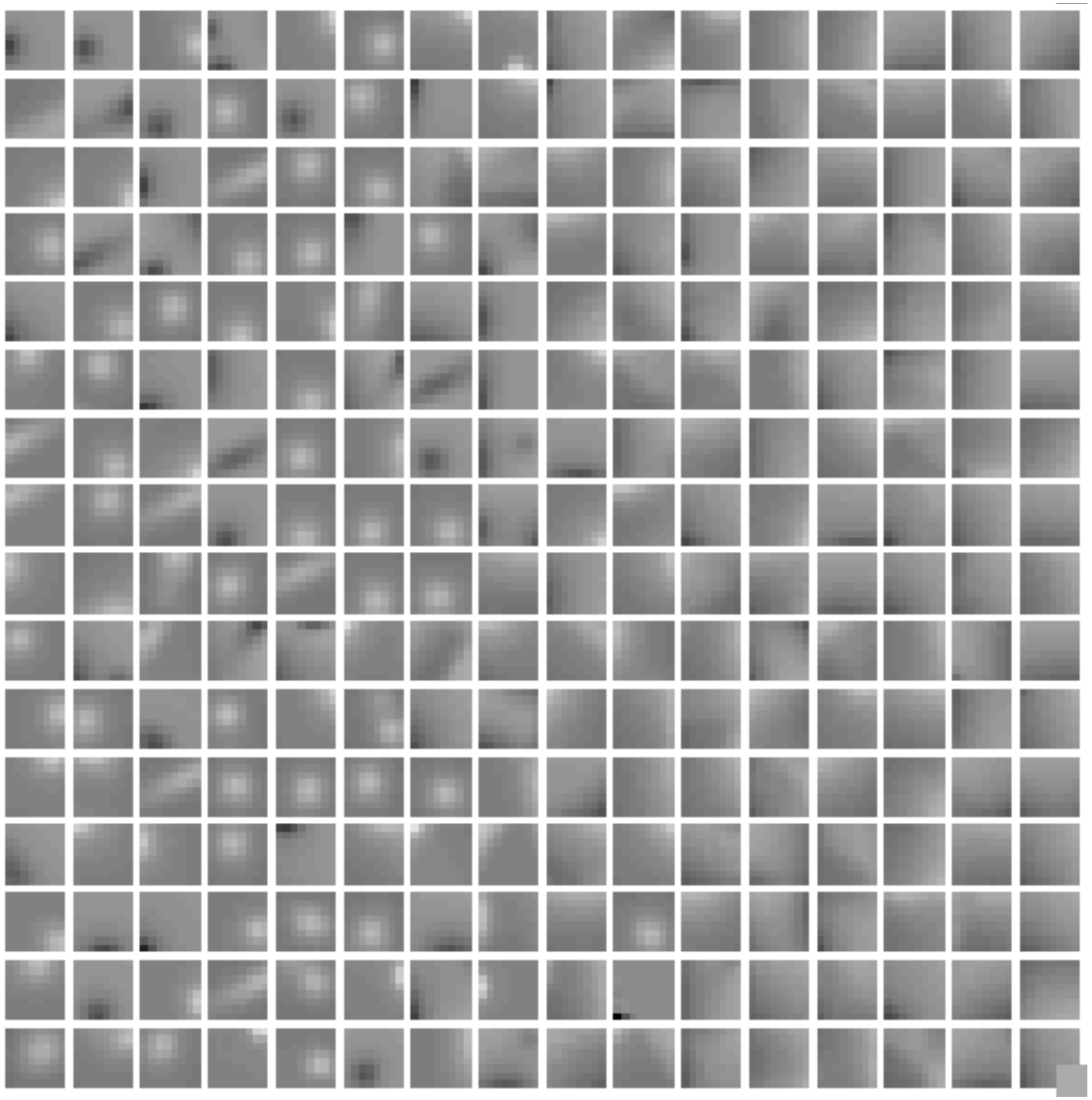
Denoised Image Using
Global Trained Dictionary (28.8528 dB)

Denoised Image Using
Adaptive Dictionary (30.8295 dB)

Elad & Aharon, Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries , 2006

Dictionary learning





Dictionary learning

From sparse modelling to learning-based representations

Representation learning and 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

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

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

Representation learning and 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

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

Representation learning and plug&play methods

- Virtually any denoiser architecture can be used

- U-Net quite popular for image denoising

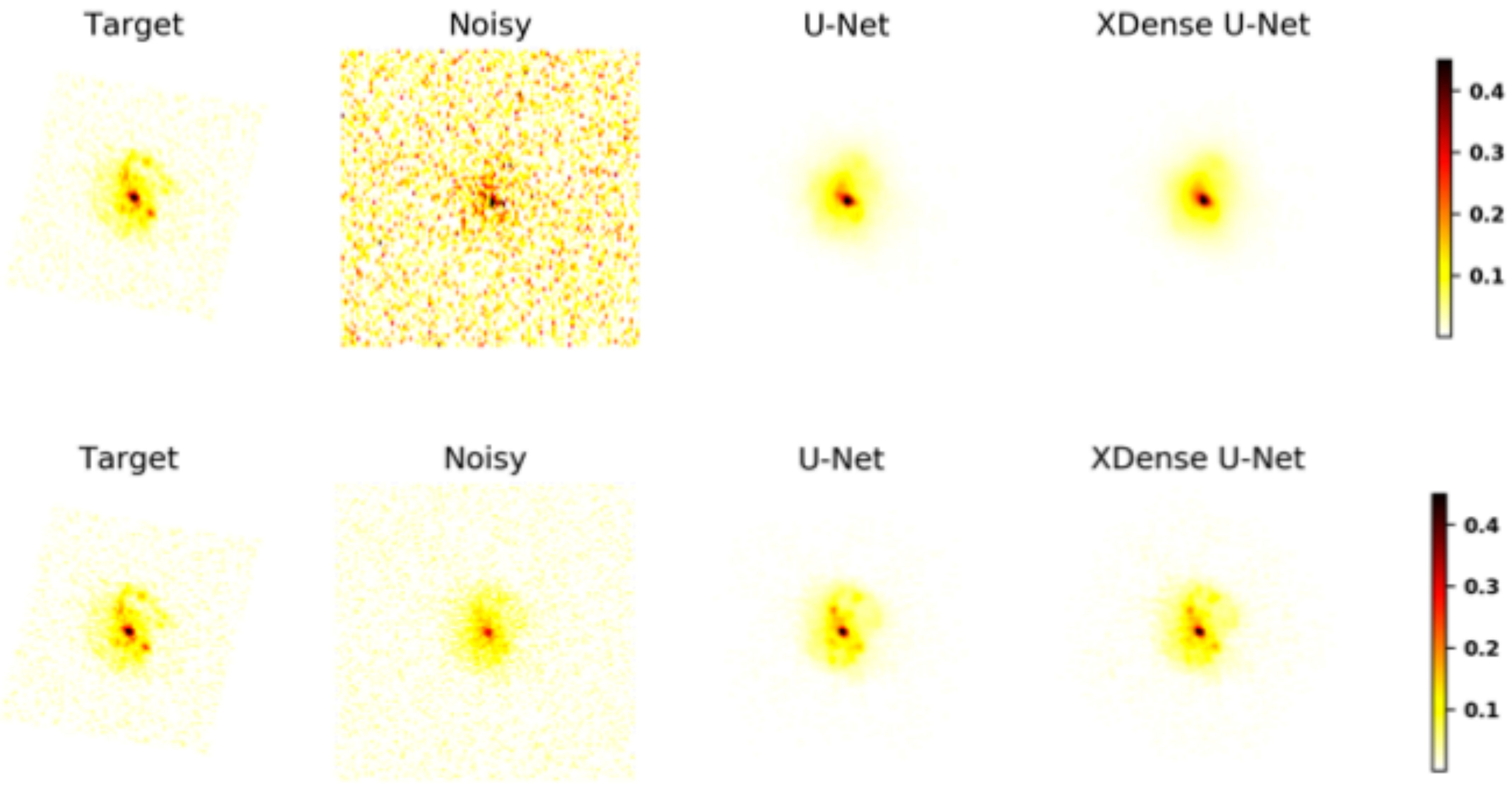
(Though being highly over-parameterized)

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

Sureau et al. 2019

Representation learning and plug&play methods

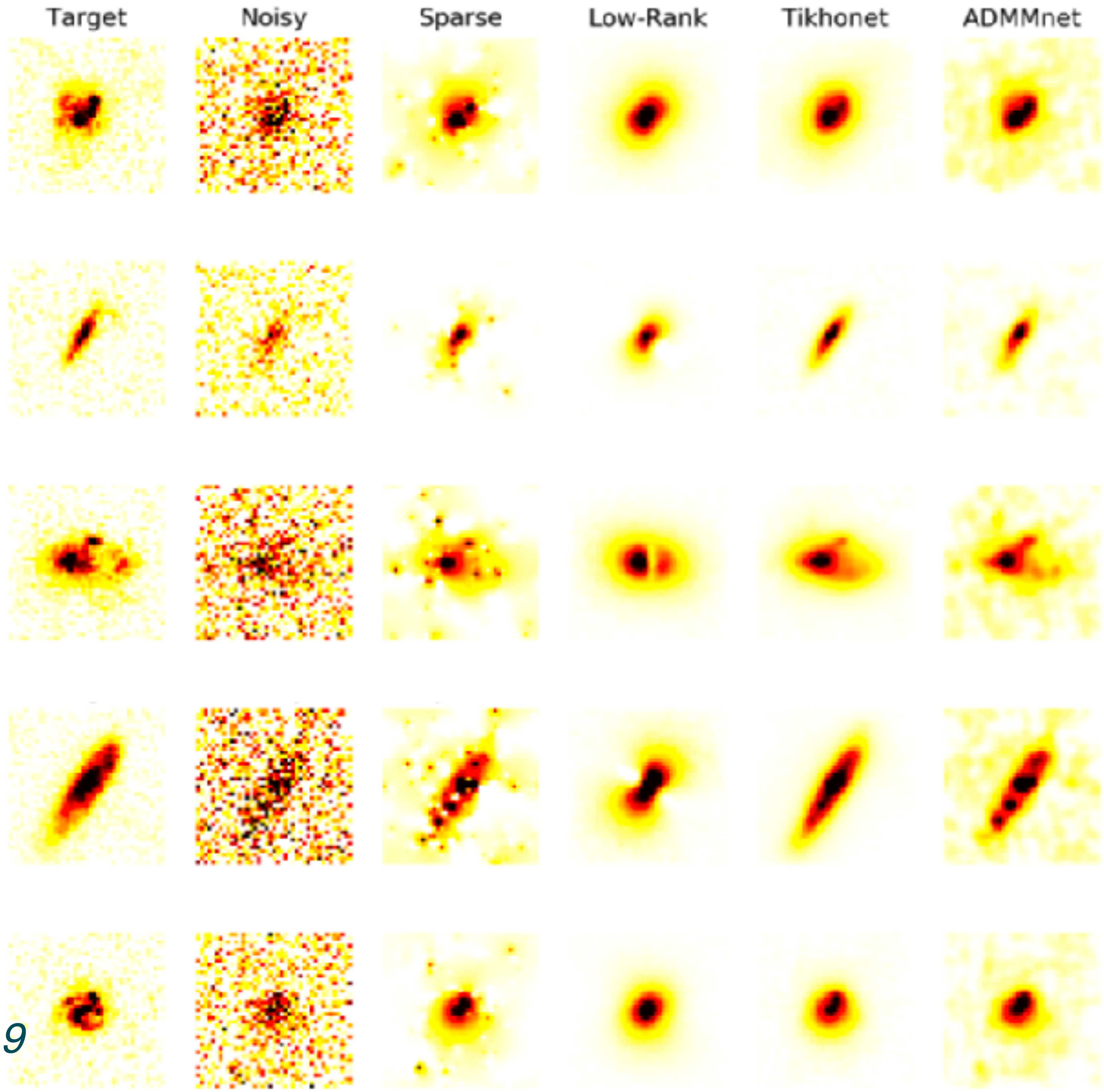
U-Net architectures for galaxy image denoising



Sureau et al. 2019

Representation learning and plug&play methods

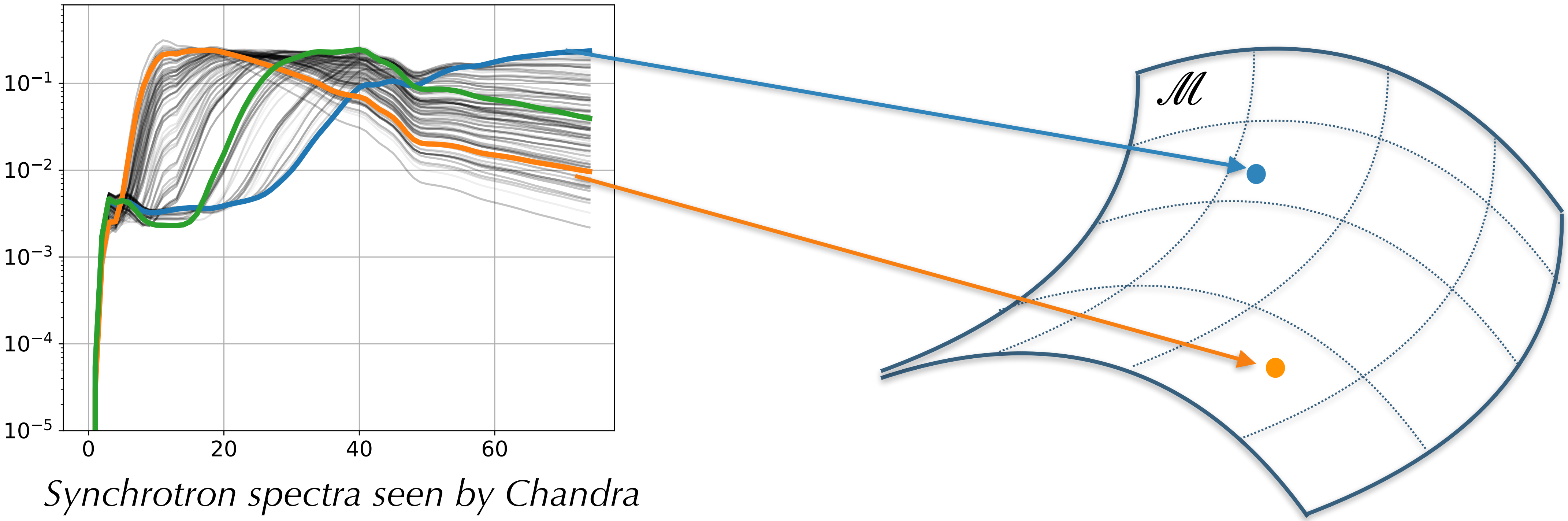
U-Net architectures for galaxy image deconvolution



Sureau et al. 2019

Learning non-linear signal representations

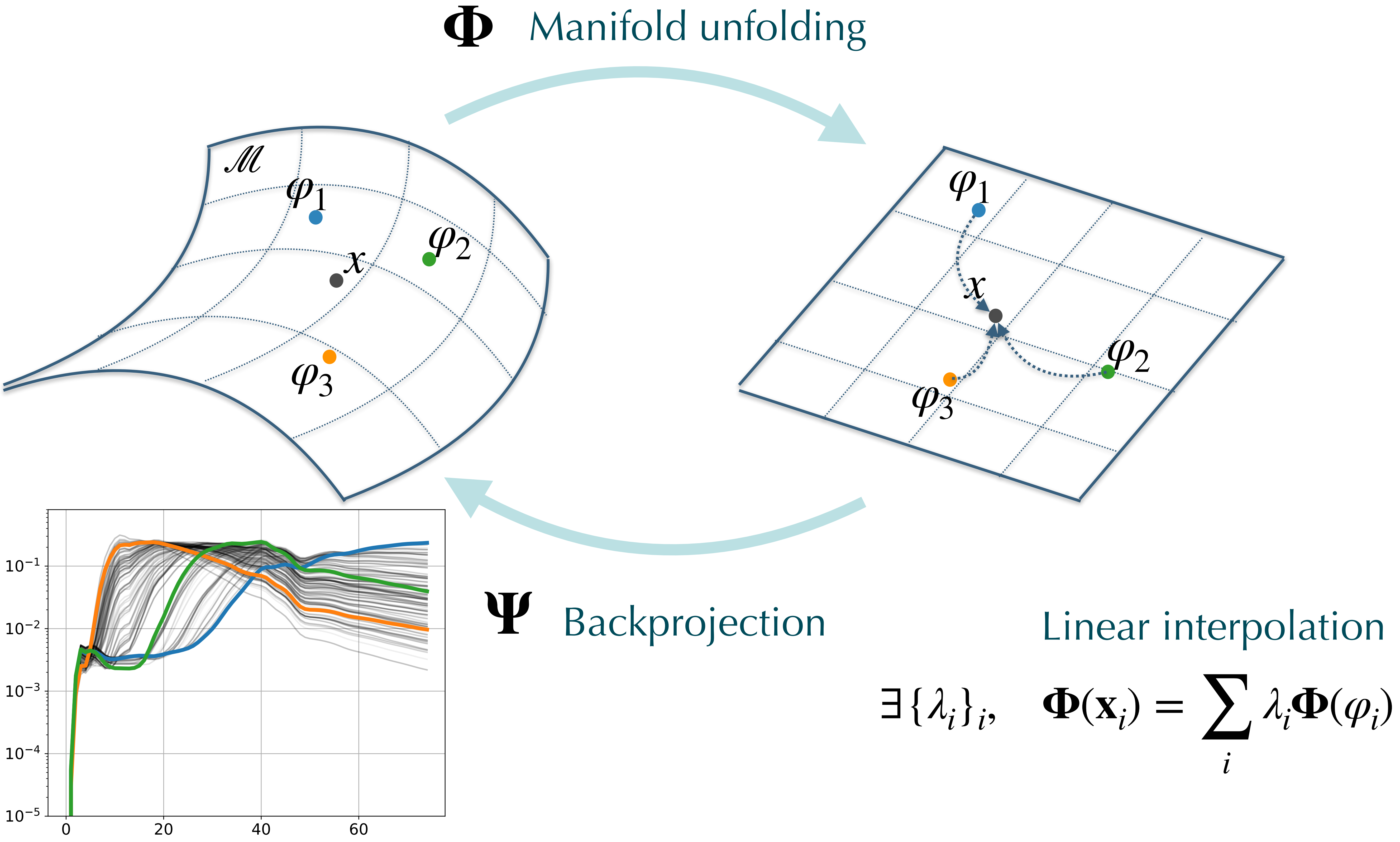
Physically relevant signals generally belong to a smooth low-dimensional manifold



Synchrotron spectra seen by Chandra

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

Learning non-linear signal representations



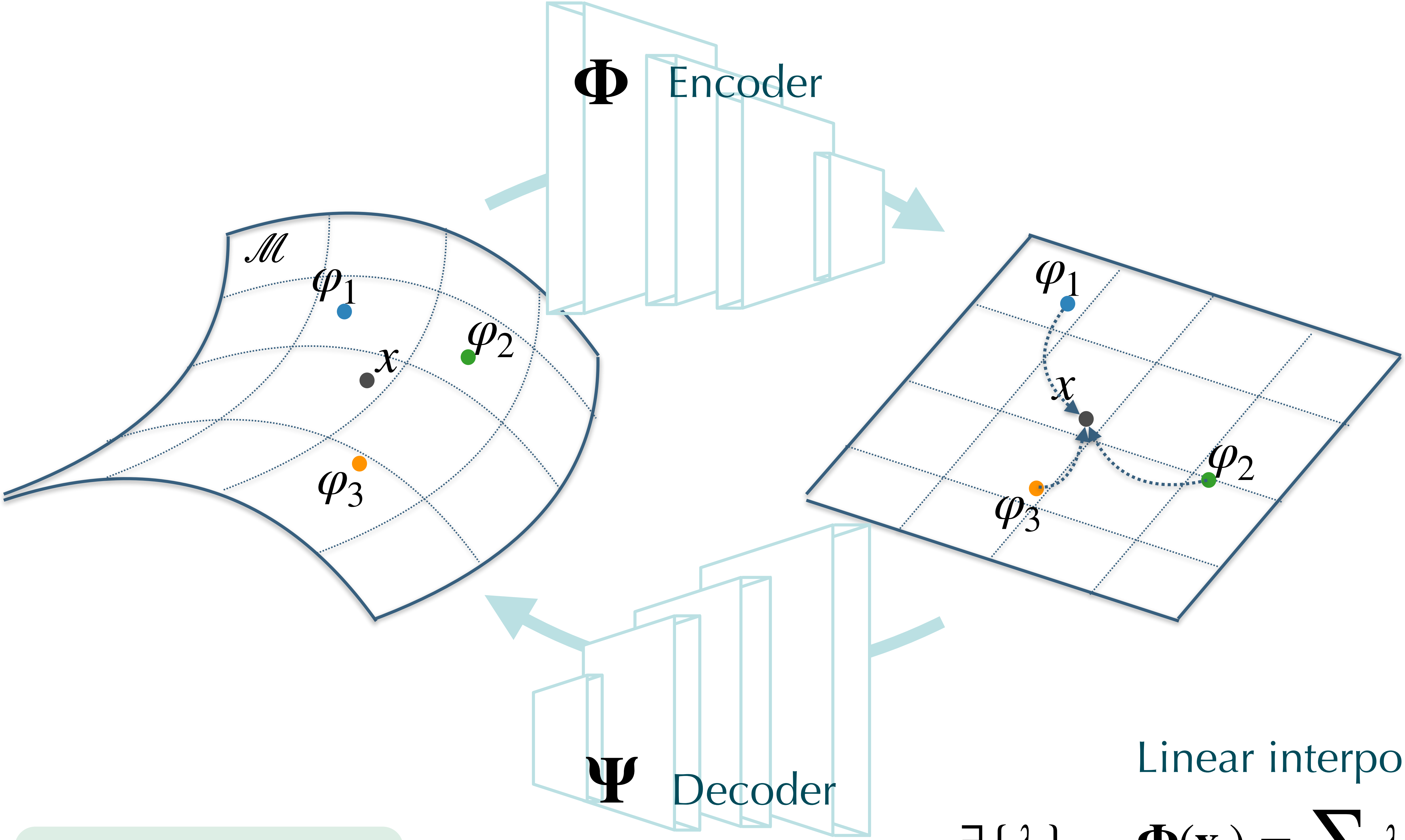
$\mathbf{\Phi}$ Manifold unfolding

$\mathbf{\Psi}$ Backprojection

Linear interpolation

$$\exists \{\lambda_i\}_i, \quad \mathbf{\Phi}(\mathbf{x}_i) = \sum_i \lambda_i \mathbf{\Phi}(\varphi_i)$$

Learning non-linear signal representations



Linear interpolation

$$\exists \{\lambda_i\}_i, \quad \mathbf{\Phi}(\mathbf{x}_i) = \sum_i \lambda_i \mathbf{\Phi}(\varphi_i)$$

Learn $\mathbf{\Psi}$ and $\mathbf{\Phi}$

Bobin, Carloni-Gertosio, Bobin, Thiam, 2021

Combining with unmixing

sGMCA : a semi-blind sparse unsupervised matrix factorization method

$$\text{argmin}_{\mathbf{A},\mathbf{S}} \|\boldsymbol{\Lambda} \odot \mathbf{SW}\|_1 + \sum_{j\neq\mathcal{J}} \iota_{\mathcal{S}_m}\left(\mathbf{A}^j\right) + \sum_{j\in\mathcal{J}} \iota_{\mathcal{B}_{\phi_j}(\{\varphi^e_{\mathcal{M}_j}\})}\left(\mathbf{A}^j\right) + \frac{1}{2} \parallel \mathbf{X} - \mathbf{AS} \parallel_F^2$$

Sparsity regularization

Oblique constraint

Barycentric span constraint

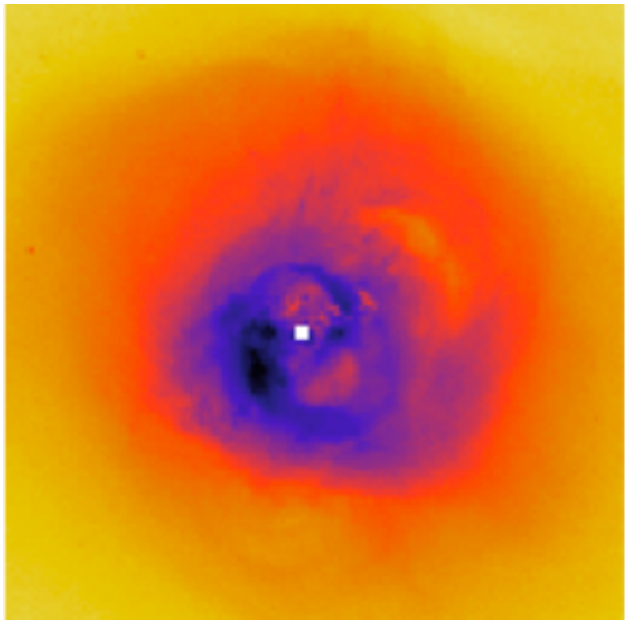
Data fidelity

- 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

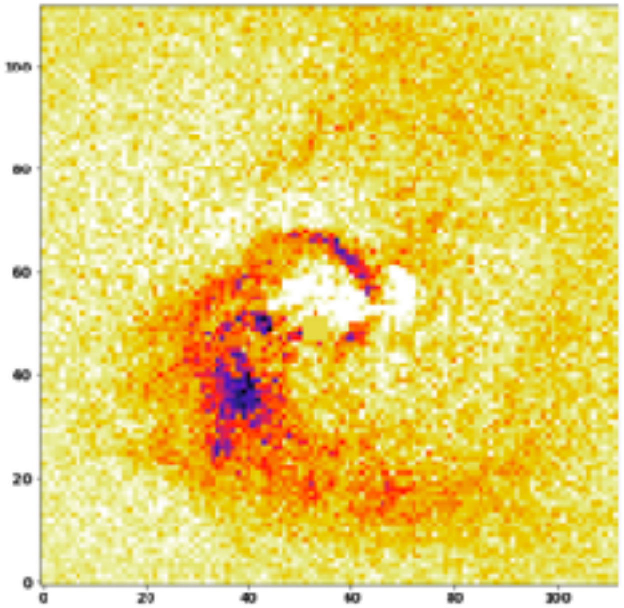
Carloni-Gertosio, Bobin, Acero, submitted to DSP, 2022

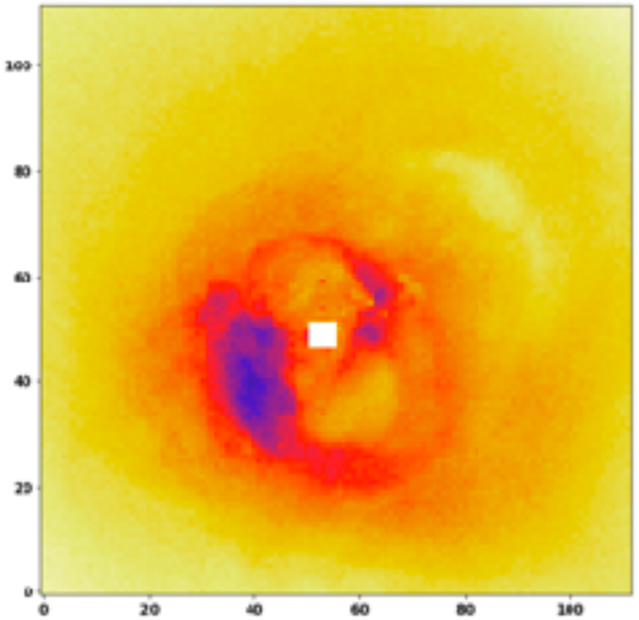
Unsupervised matrix factorisation

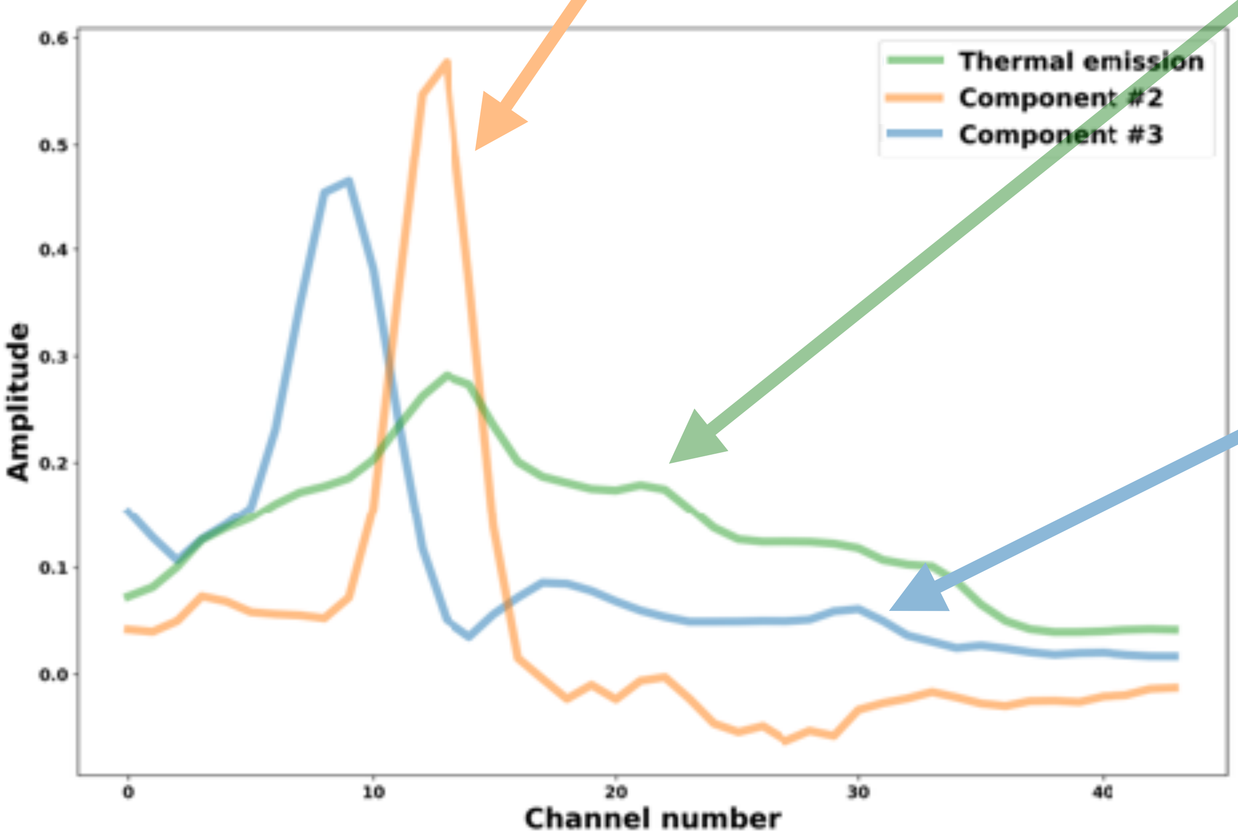
Perseus galaxy cluster



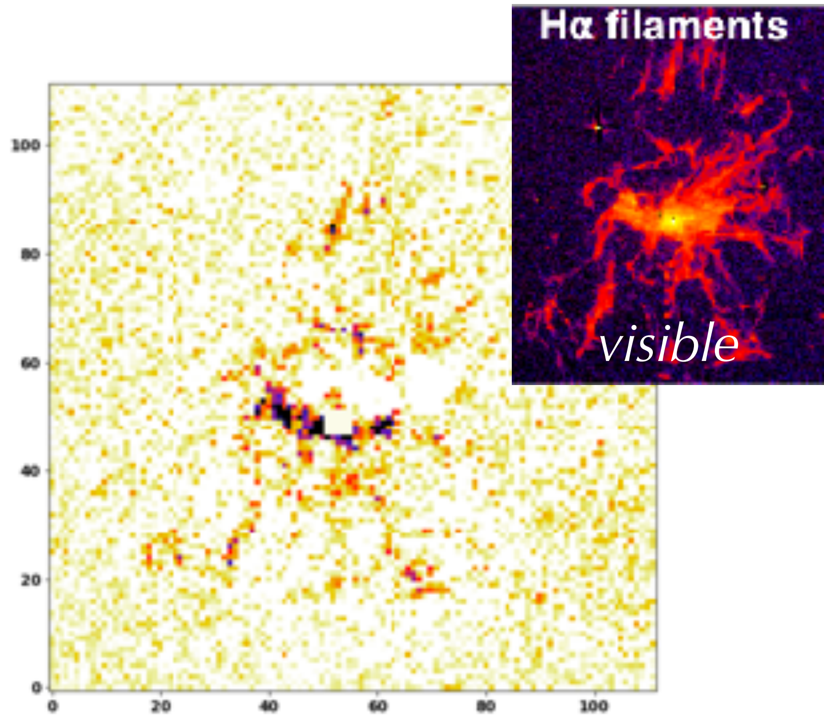
Integrated image 0.5 - 8keV







Only the thermal emission is constrained



X-ray filaments have~50-100 counts
buried under 10^4 counts
Finding features with contrast < 1%