AI4SIP

Matthieu Kowalski (Univ Paris-Saclay)

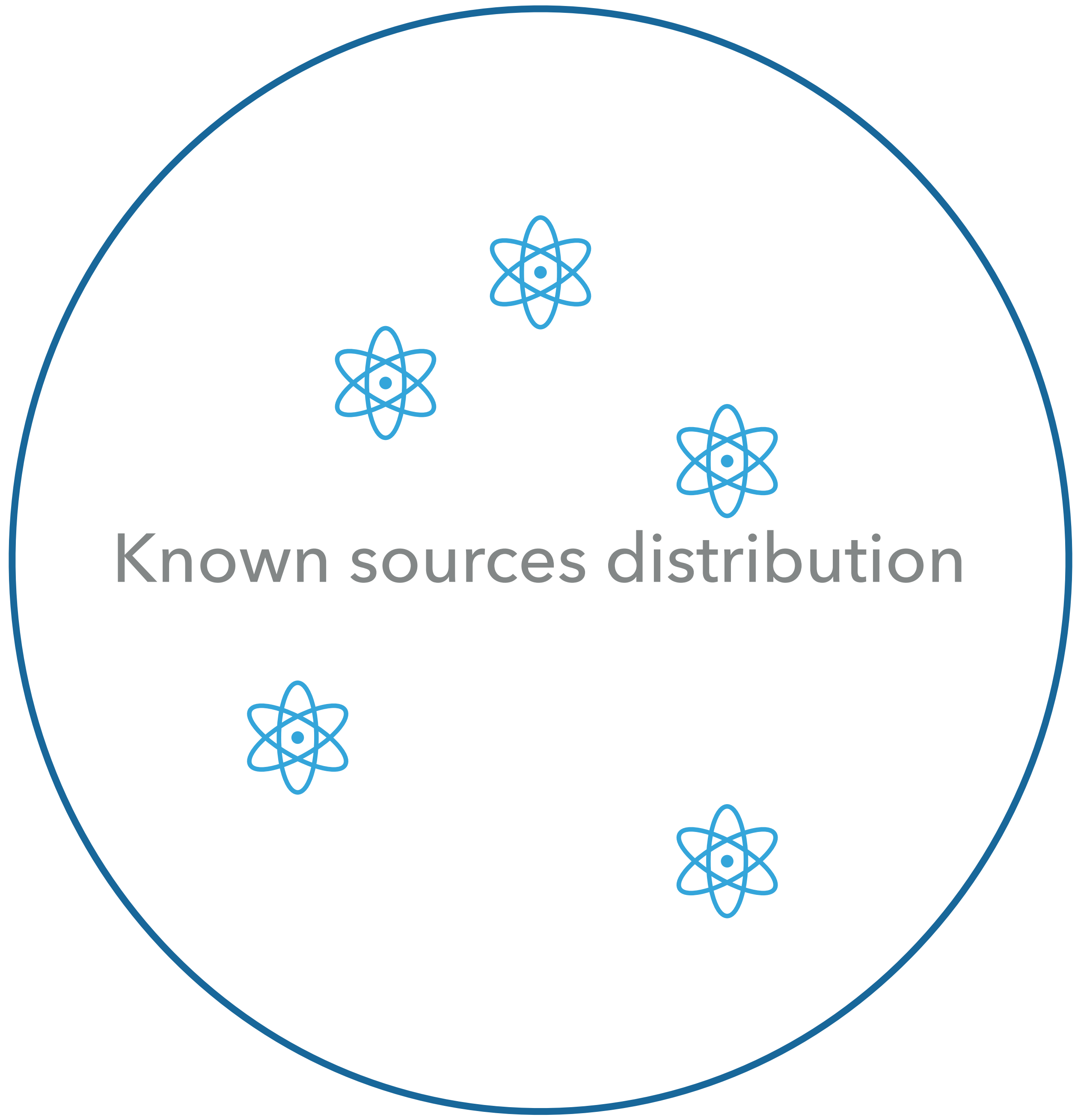
INVERSE PROBLEMS: FROM SPARSE TIME-FREQUENCY SYNTHESIS TO DICTIONARY LEARNING

OUTLINE

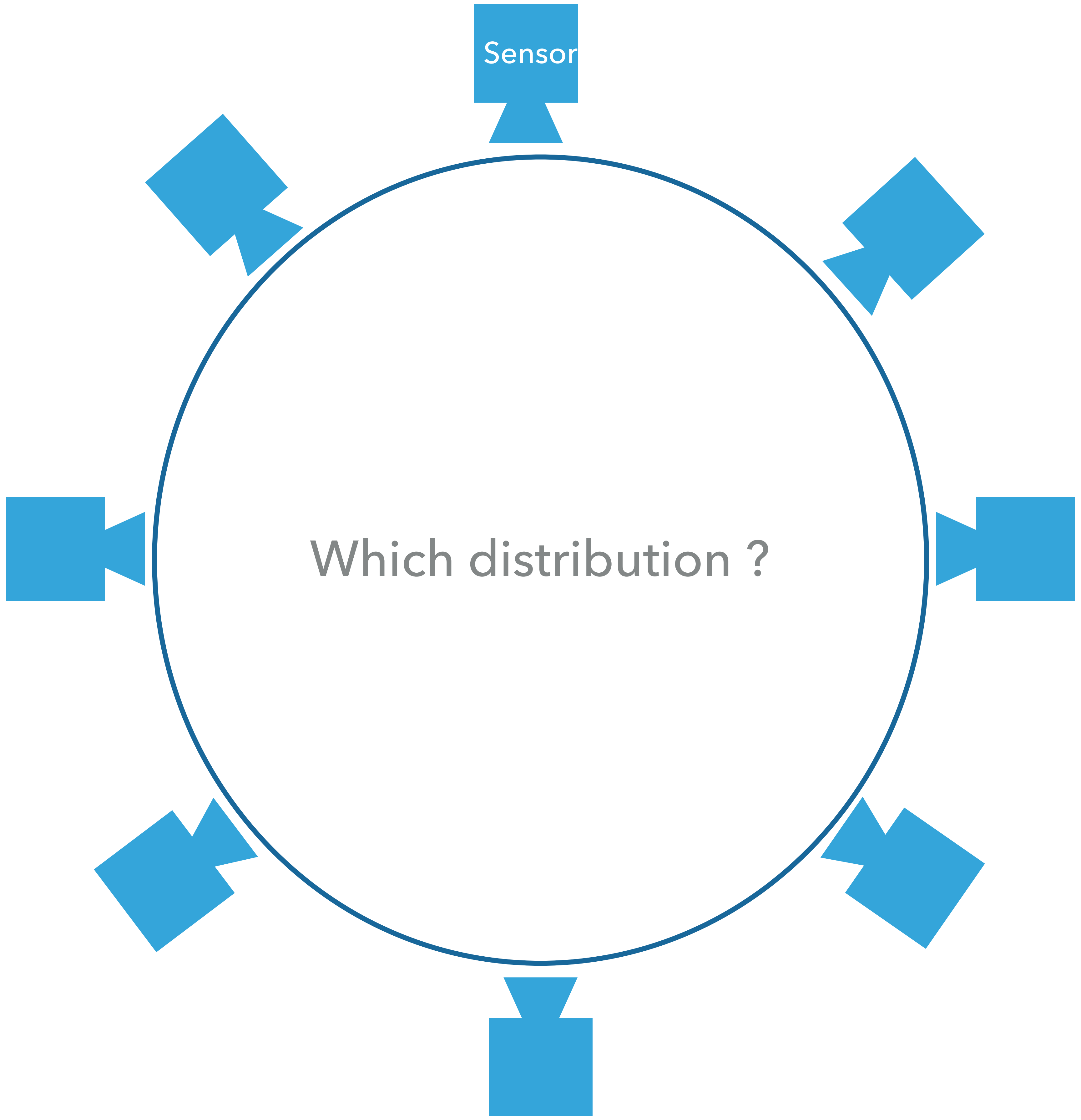
- Introduction – state of the art
  - Linear Inverse Problems
  - Time-Frequency dictionary
  - Sparse Coding and inverse problems
- Iterative Shrinkage/Thresholding and thresholding rules (illustration on audio declipping)
- Dictionary Learning
  - From alternate minimization to Deep Network
  - Non negative Matrix factorization

INTRODUCTION: LINEAR INVERSE PROBLEMS

DIRECT PROBLEM: DEFINITION A CONTRARIO



Direct Problem



Inverse Problem

DIRECT PROBLEM

- Let $x \in \mathbb{R}^N$ be a signal
- Let $A \in \mathbb{R}^{MN}$ be a sensing matrix
- We observe $y \in \mathbb{R}^M$ such that

$$y = Ax + e$$

- With $e \in \mathbb{R}^M$ some error (measure, noise etc.)
- When $M < N$ the problem is said "under-determined": infinity of solutions

DIRECT PROBLEM: INVERSION

- We observe

$$y = Ax + e$$

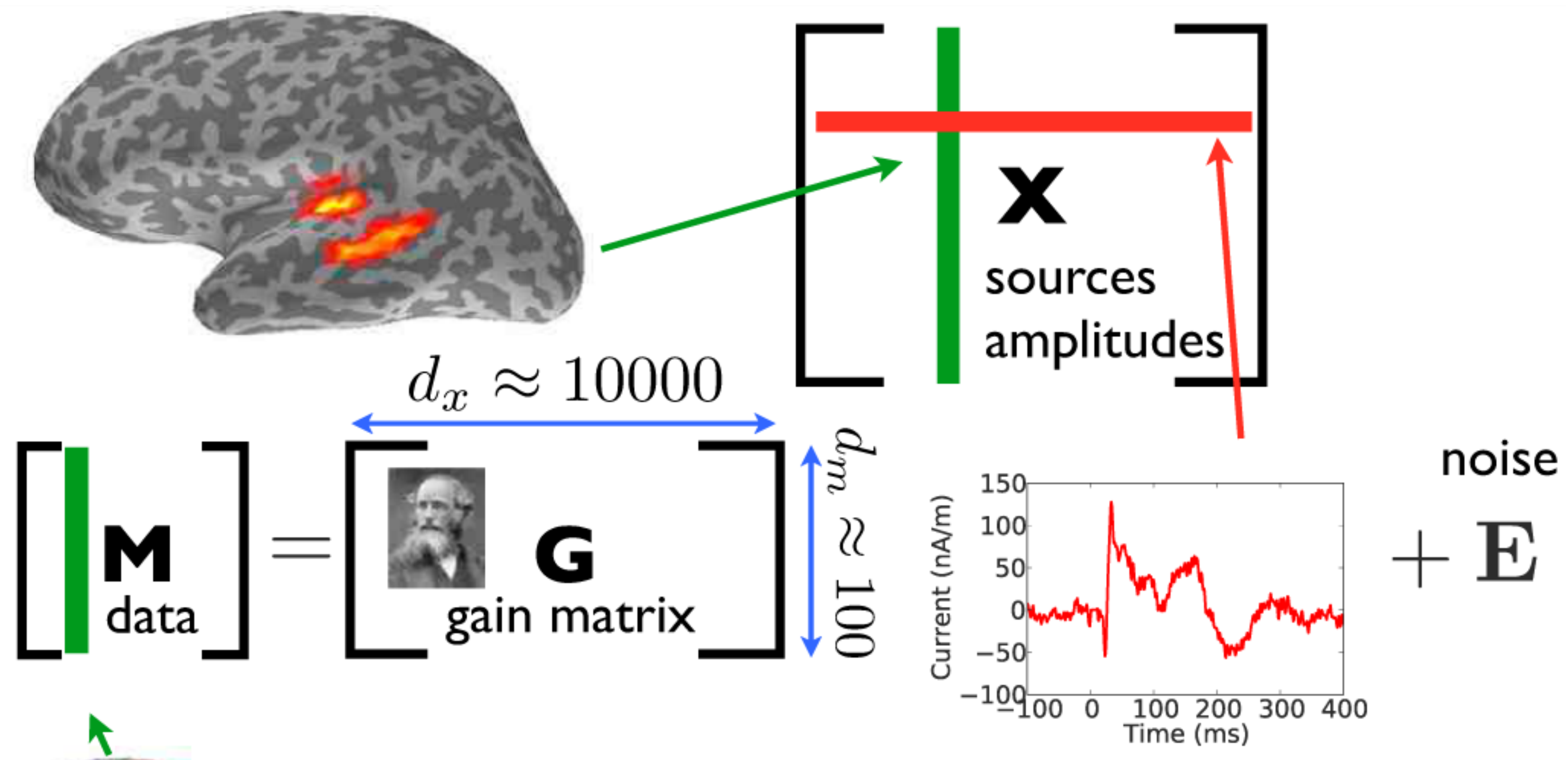
- Inversion by optimization:

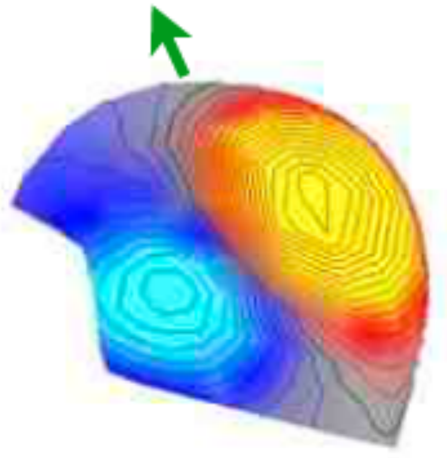
$$x = \underset{x}{\operatorname{argmin}} \; \mathscr{L}(x) + \mathscr{R}(x)$$

With

- $\mathscr{L}$: the loss or "data fit" term or "prior" on the noise
- $\mathscr{R}$: regularization or "prior" term on the sources

EXAMPLE: M/EEG INVERSE PROBLEM





THM: Following **Maxwell's equations** each source adds its contribution **linearly**

Minimum Norm Estimate (MNE – 1994):

$$X = \underset{X}{\operatorname{argmin}} \frac{1}{2}\|M - GX\|^2 + \frac{\lambda}{2}\|X\|^2$$

Minimum Current Estimate (MCE – 1999):

$$X = \underset{X}{\operatorname{argmin}} \frac{1}{2}\|M - GX\|^2 + \lambda\|X\|_1$$

Minimum mixed-norm Estimate (MxNE [2012]):

$$X = \underset{X}{\operatorname{argmin}} \frac{1}{2}\|M - GX\|^2 + \lambda\|X\|_{21}$$

INVERSE PROBLEM

$$x = \underset{x}{\text{argmin}} \;\; \mathscr{L}(x) + \mathscr{R}(x)$$

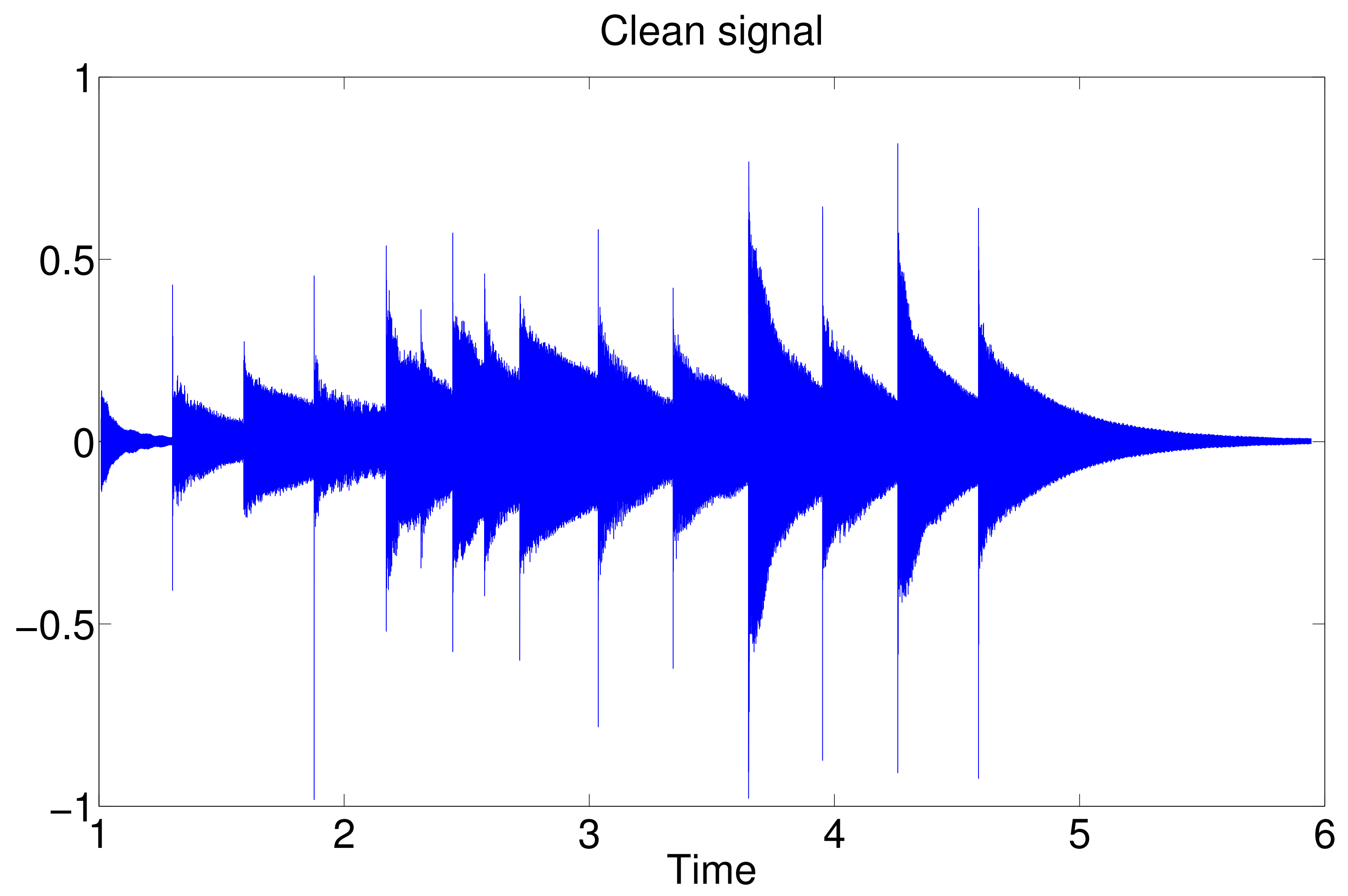
With

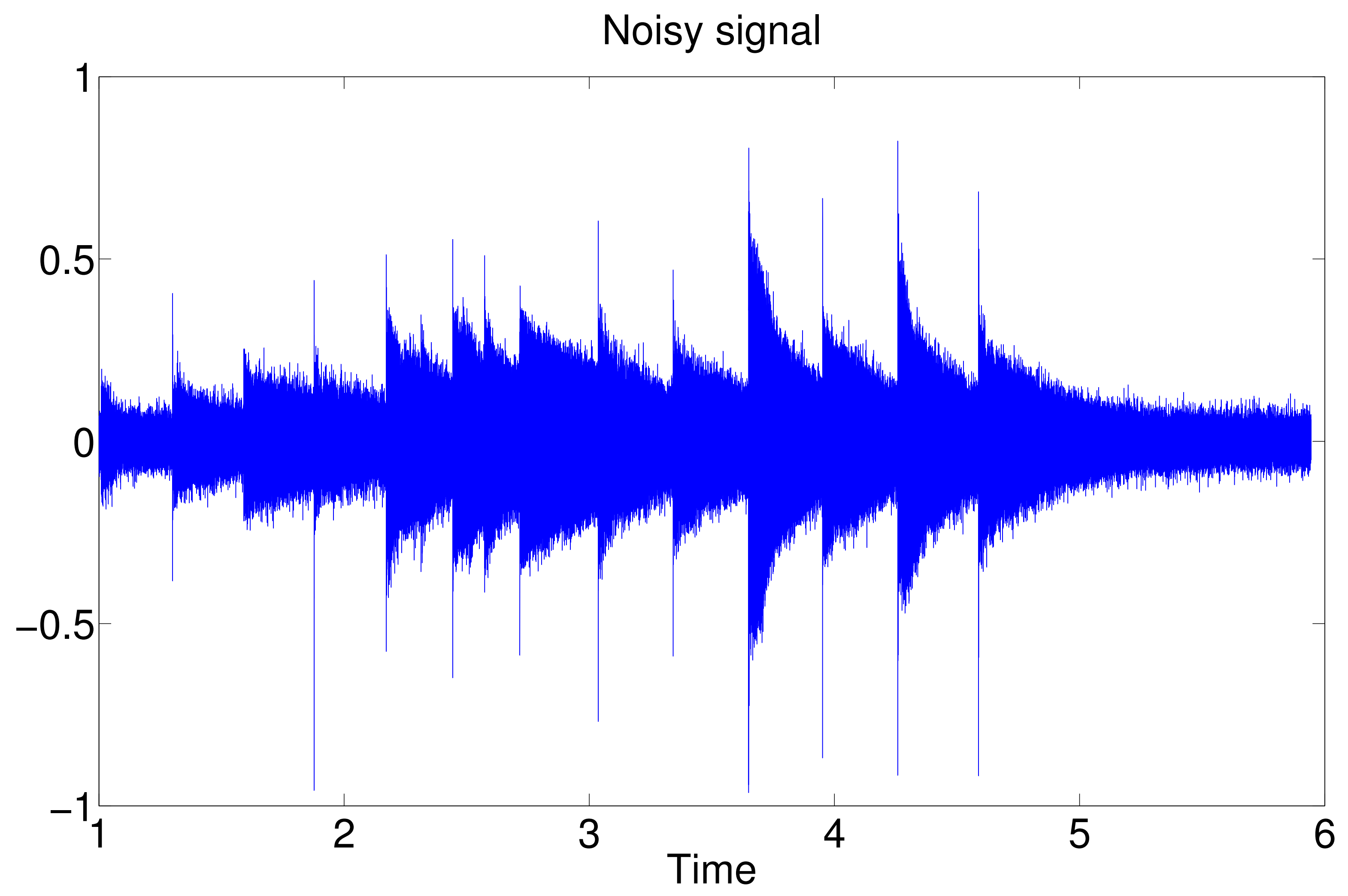
- $\mathscr{L}$: the loss or "data fit" term or "prior" on the noise
- $\mathscr{R}$: regularization or "prior" term on the sources

How to choose $\mathscr{L}$ and $\mathscr{R}$?

INTRODUCTION: TIME-FREQUENCY DICTIONARY

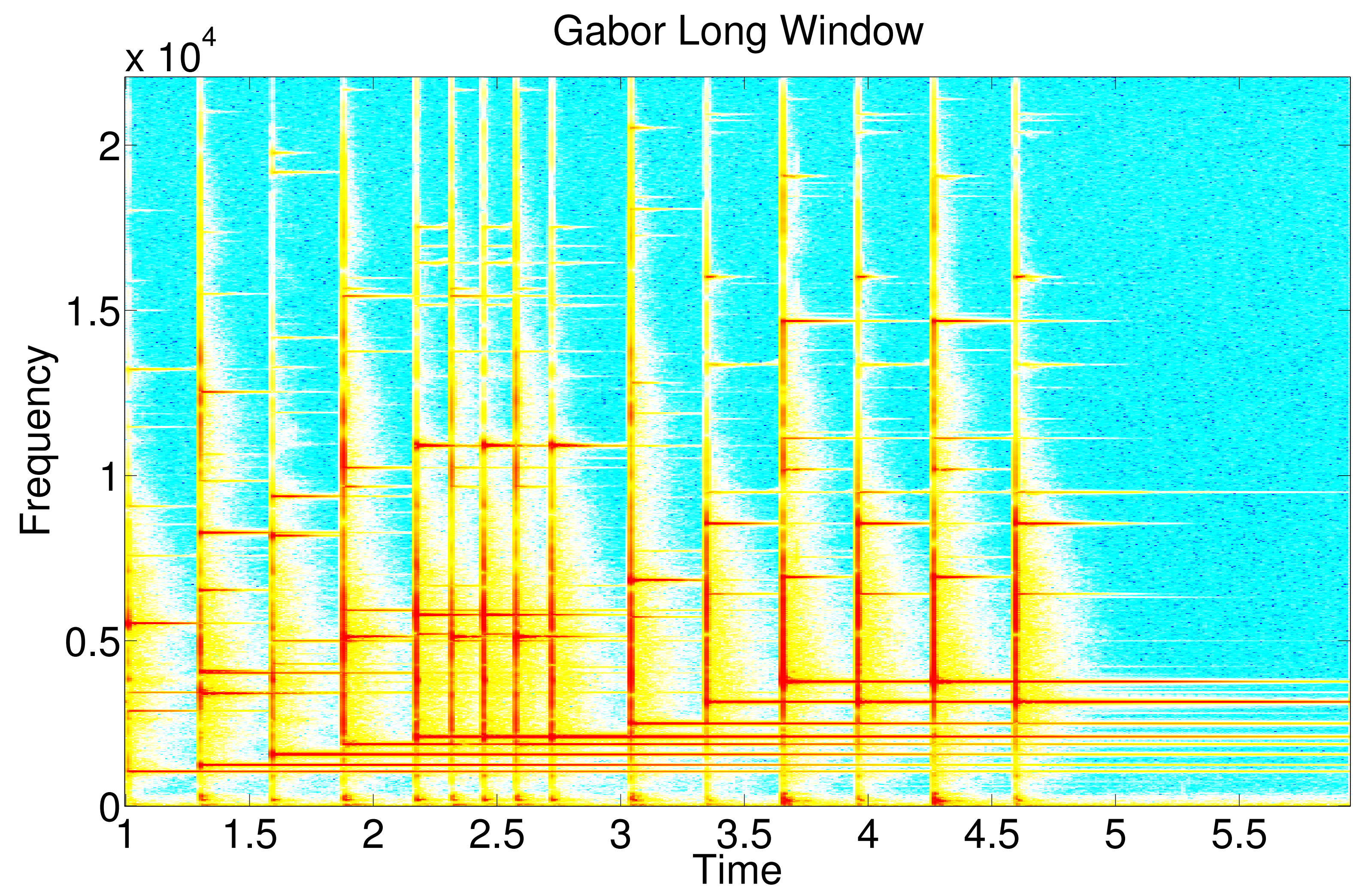
A TOY SIGNAL: THE GLOCKENSPIEL

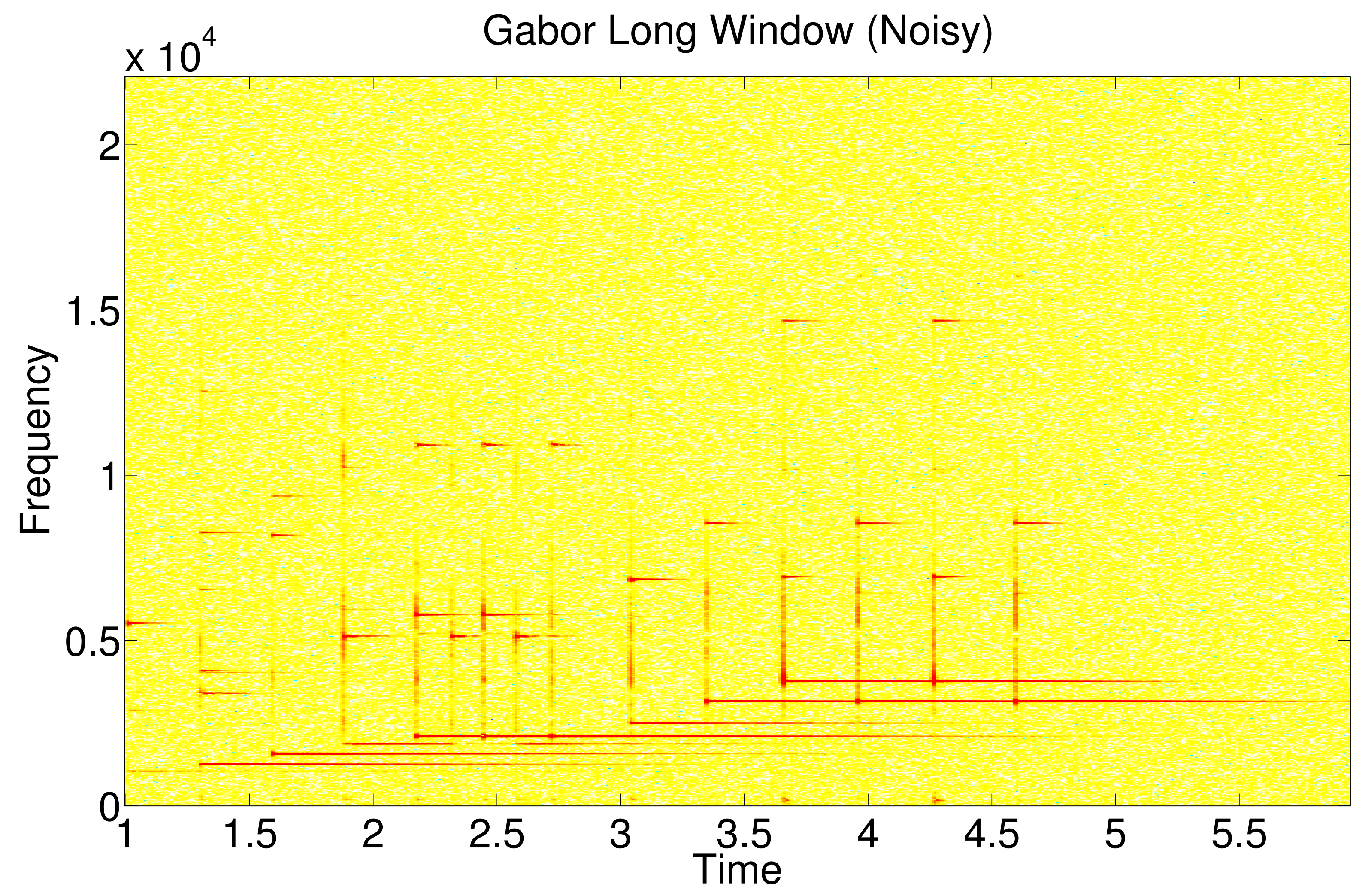




- well localized ``percussive'' transients
- clear harmonics/tonal layer
- Sampling rate: 44100Hz
- Length: 6s (2^{18}samples)
- White gaussian noise : input SNR = 10db

VARIOUS TIME-FREQUENCY REPRESENTATIONS -- LONG WINDOW

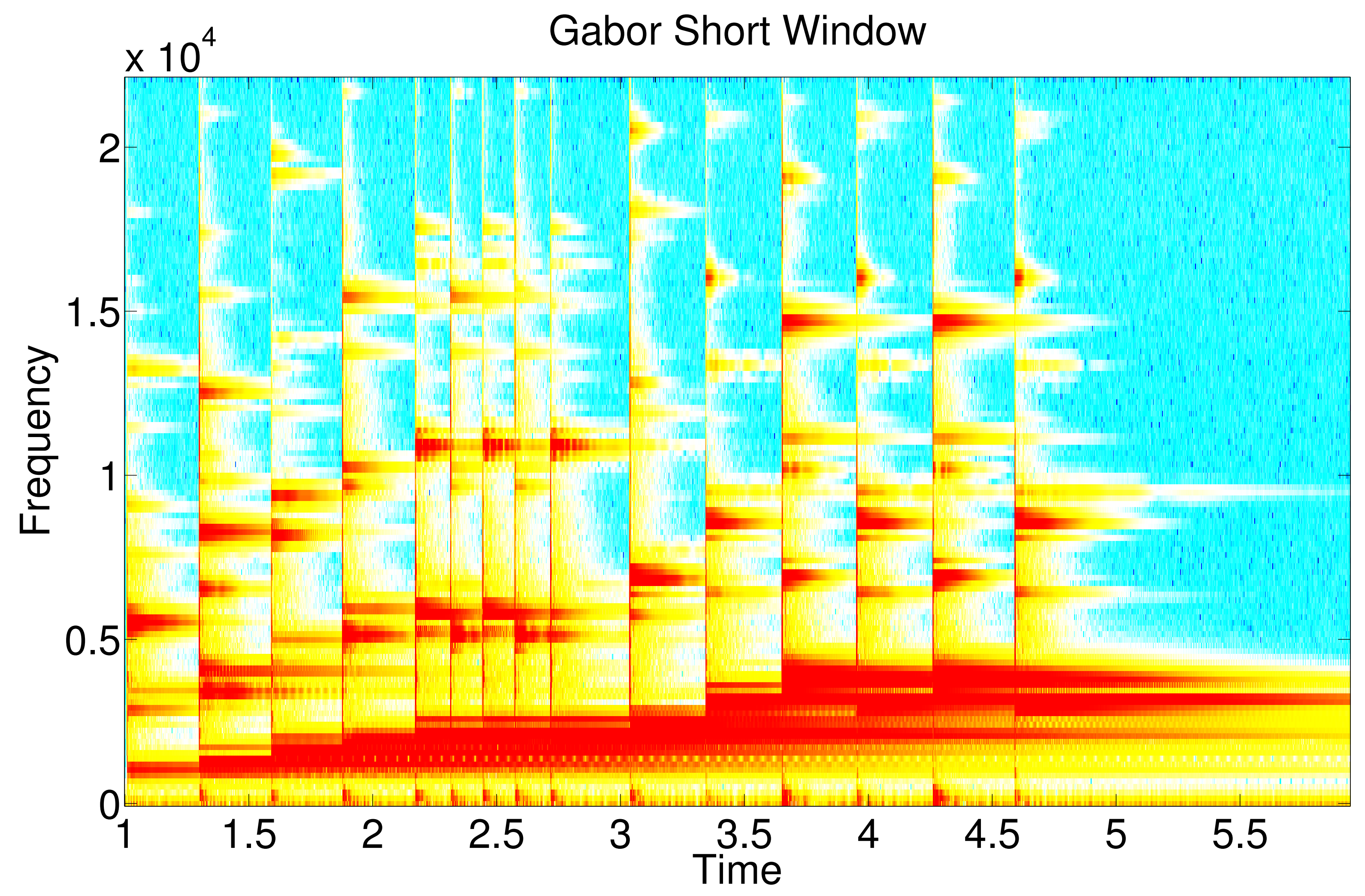


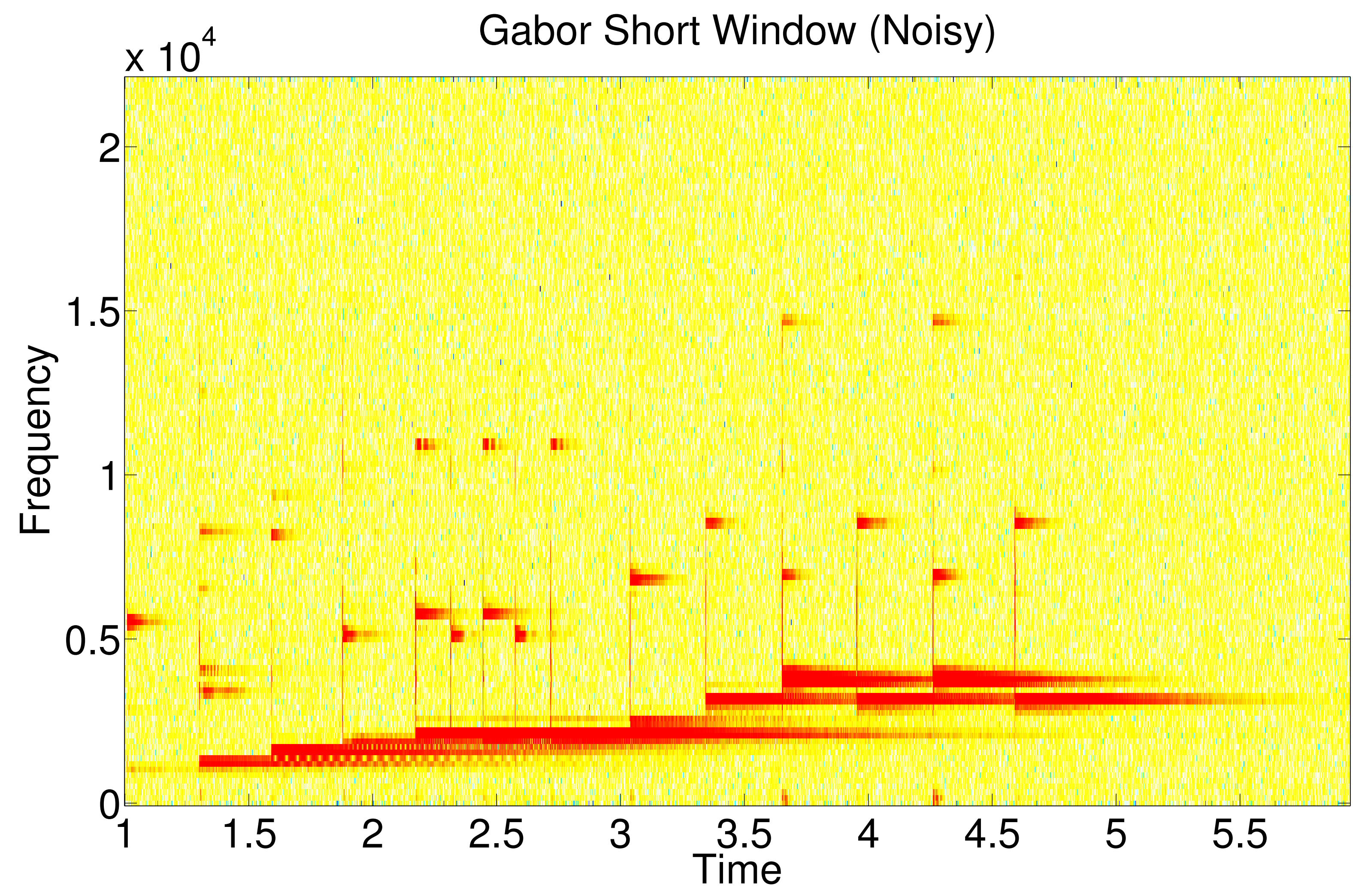


- Let $w[t] \in \mathbb{R}_+^L$ a smooth window
- Let $\varphi_{n,\nu}[t] = w[t - an]e^{-i2\pi\frac{\nu}{bL}t}$ a time-frequency atom
- $X[n,\nu] = \langle x[t], \varphi_{n,\nu}[t] \rangle$

- Hann window analysis
- Length: 48 ms (2048 samples)
- Hop size: 12ms (512 samples)
- Well adapted to tonal
- Transients are badly localized

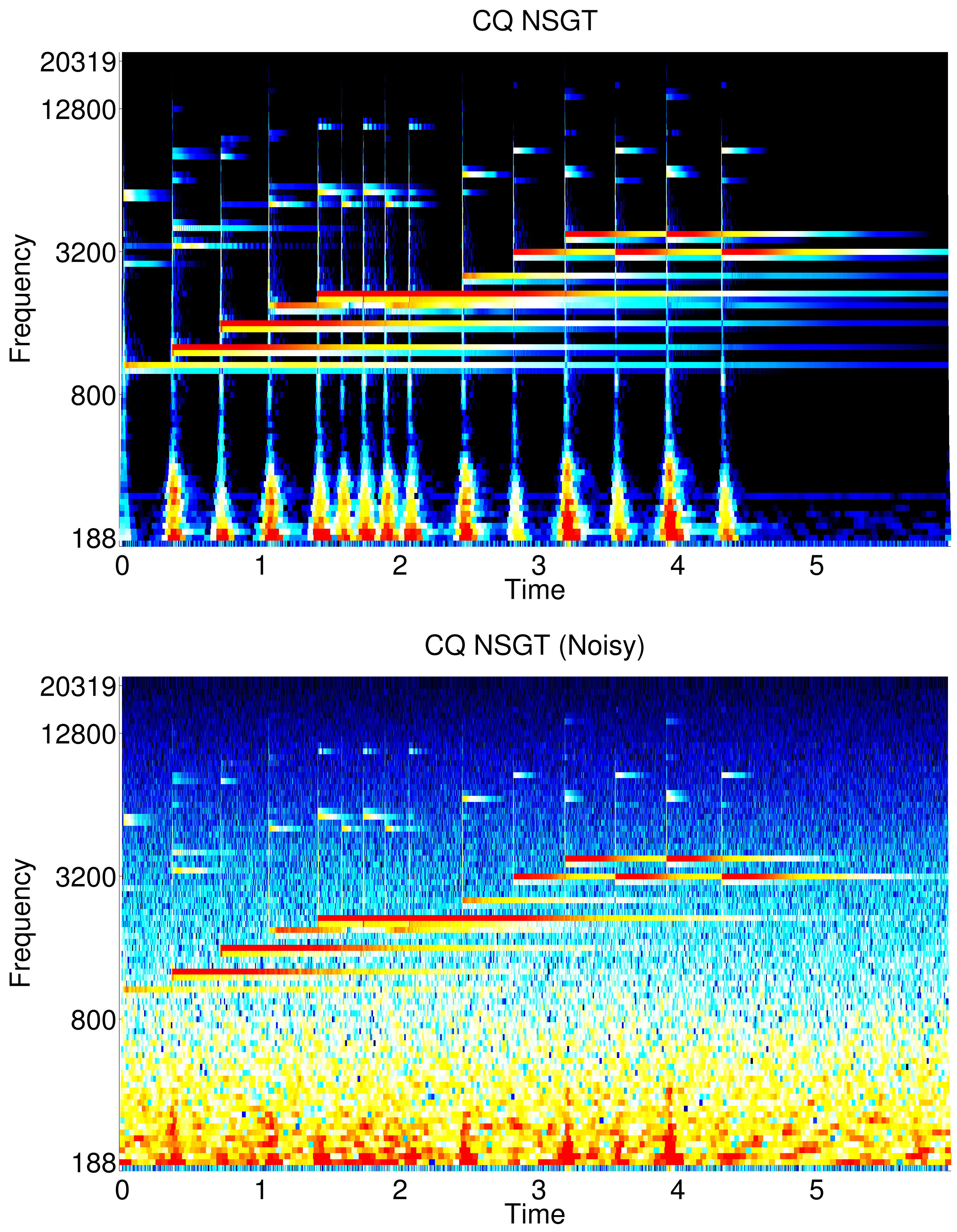
VARIOUS TIME-FREQUENCY REPRESENTATIONS -- SHORT WINDOW





- Hann window analysis
- Length: 6 ms (256 samples)
- Hop size: 1 ms (64 samples)
- Well adapted to transients
- Tonals are badly localized

VARIOUS TIME-FREQUENCY REPRESENTATIONS -- CONSTANT Q



Nice compromise but:

- less harmonics than the long window analysis
- transients are not as well localized as with the short window

INTRODUCTION: SPARSE CODING

SPARSE CODING

- Let $x \in \mathbb{R}^N$ be a signal
- Let $\Phi \in \mathbb{R}^{NK}$ be a dictionary (wavelets, time-frequency…) with $K \geq N$
- Sparse coding:

$$x = \Phi\alpha$$

  - $\alpha \in \mathbb{R}^K$ are the **synthesis** coefficients, supposed to be **sparse**
  - Remark: when $K > N$ there is an **infinity** of solution for α

ANALYSIS VS SYNTHESIS COEFFICIENTS

$$x = \Phi\alpha$$

- $\alpha \in \mathbb{R}^K$ are the **synthesis** coefficients
- **Analysis** coefficients: $\langle x, \varphi_k \rangle$, φ_k being one column of Φ
- Analysis coefficients = synthesis coefficients, if Φ is **orthogonal**
- Analysis coefficients ⊂ synthesis coefficients

WHICH DICTIONARY ?

- Usual dictionaries:
  - Time-frequency
  - Wavelets
  - *-lets
  - Union of dictionaries
- Can we learn the dictionary ? *(Second part of the talk)*

SPARSITY: SYNTHESIS APPROACH

- Hypothesis: $x = \Phi\alpha$ admits a sparse representation α in Φ
- Ideal solution: $\hat{\alpha} = \underset{\alpha}{\text{argmin}} \frac{1}{2}\|y - \Phi\alpha\|_2^2 + \lambda\|\alpha\|_0$
- Problem: very hard to solve in a finite time. We relax the ℓ_0 constraint into ℓ_1
- Lasso [Tibshirani 96] or Basis Pursuit Denoising [Chen et al. 98]:

$$\hat{\alpha} = \underset{\alpha}{\text{argmin}} \frac{1}{2}\|y - \Phi\alpha\|_2^2 + \lambda\|\alpha\|_1$$

INVERSE PROBLEM WITH SPARSE CODING

- Let $x \in \mathbb{R}^N$ be a signal, $A \in \mathbb{R}^{MN}$ be a sensing matrix, and $\Phi \in \mathbb{R}^{NK}$ a dictionary
- We observe $y \in \mathbb{R}^M$ such that

$$y = Ax + e = A\Phi\alpha + e$$

- With $e \in \mathbb{R}^M$ some error (measure, noise etc.)
- Inversion:

$$\alpha = \underset{\alpha}{\operatorname{argmin}} \frac{1}{2}\|y - A\Phi\alpha\|^2 + \lambda\|\alpha\|_1$$

ITERATIVE SHRINKAGE/THRESHOLDING ALGORITHM

- Lasso: $\hat{\alpha} = \operatorname{argmin}_{\alpha} \frac{1}{2}\|y - A\Phi\alpha\|_2^2 + \lambda\|\alpha\|_1$
- ISTA [Combettes & Wajs 05], [Daubechies & al 04], [Figuereido & Nowak 03] :

$$\alpha^{(t+1)} = \mathcal{S}_{\lambda/L}\left(\alpha^{(t)} + \frac{1}{L}\Phi^* A^* (y - A\Phi\alpha^{(t)})\right)$$

With:

- $\mathcal{S}_\lambda(z) = \operatorname{argmin}_{\alpha} \frac{1}{2}\|z - \alpha\|^2 + \lambda\|\alpha\|_1$ the "proximity operator" of ℓ_1
- $L = \|A\Phi\|^2 \leq \|A\|^2\|\Phi\|^2$
- Fast version: [Nesterov 07] [Beck & Teboulle 09] [Dossal & Chambolle 15]

ITERATIVE SHRINKAGE/THRESHOLDING ALGORITHM

- Let $\mathscr{R}$ be any measure of (structured) sparsity:

$$\hat{\alpha} = \underset{\alpha}{\text{argmin}} \frac{1}{2}\|y - A\Phi\alpha\|_2^2 + \lambda\mathscr{R}(\alpha)$$

- ISTA:

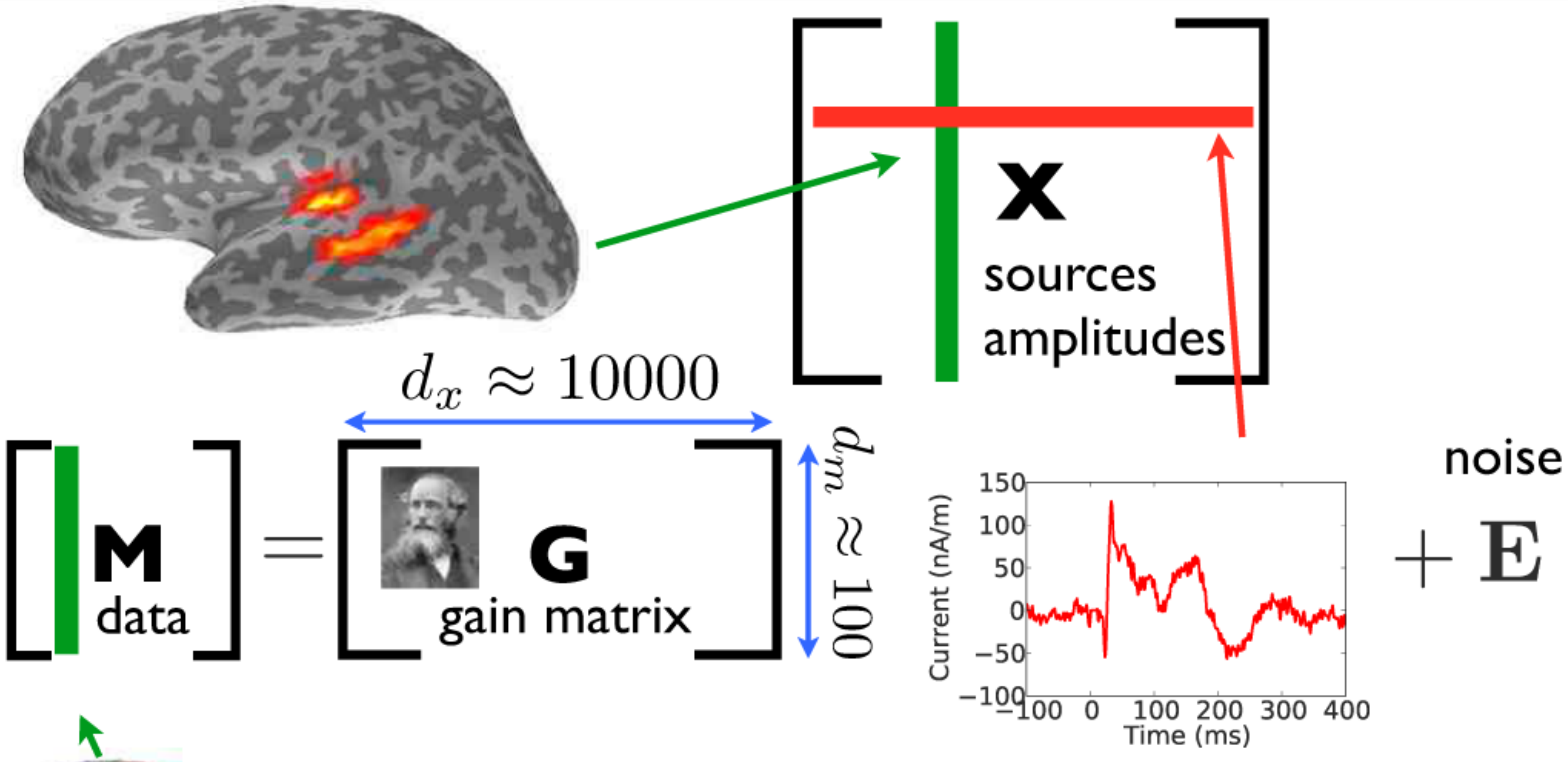
$$\alpha^{(t+1)} = \text{prox}_{\frac{\lambda}{L}\mathscr{R}}\left(\alpha^{(t)} + \frac{1}{L}\Phi^* A^*(y - A\Phi\alpha^{(t)})\right)$$

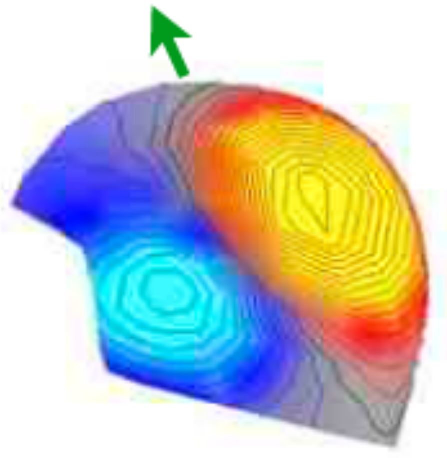
With:

$$\text{prox}_{\gamma\mathscr{R}}(z) = \underset{\alpha}{\text{argmin}} \frac{1}{2}\|z - \alpha\|^2 + \gamma\mathscr{R}(\alpha)$$

EXAMPLES

M/EEG INVERSE PROBLEM





THM: Following **Maxwell's equations** each source adds its contribution **linearly**

Minimum Norm Estimate (MNE – 1994):

$$X = \underset{X}{\text{argmin}} \frac{1}{2}\|M - GX\|^2 + \frac{\lambda}{2}\|X\|^2$$

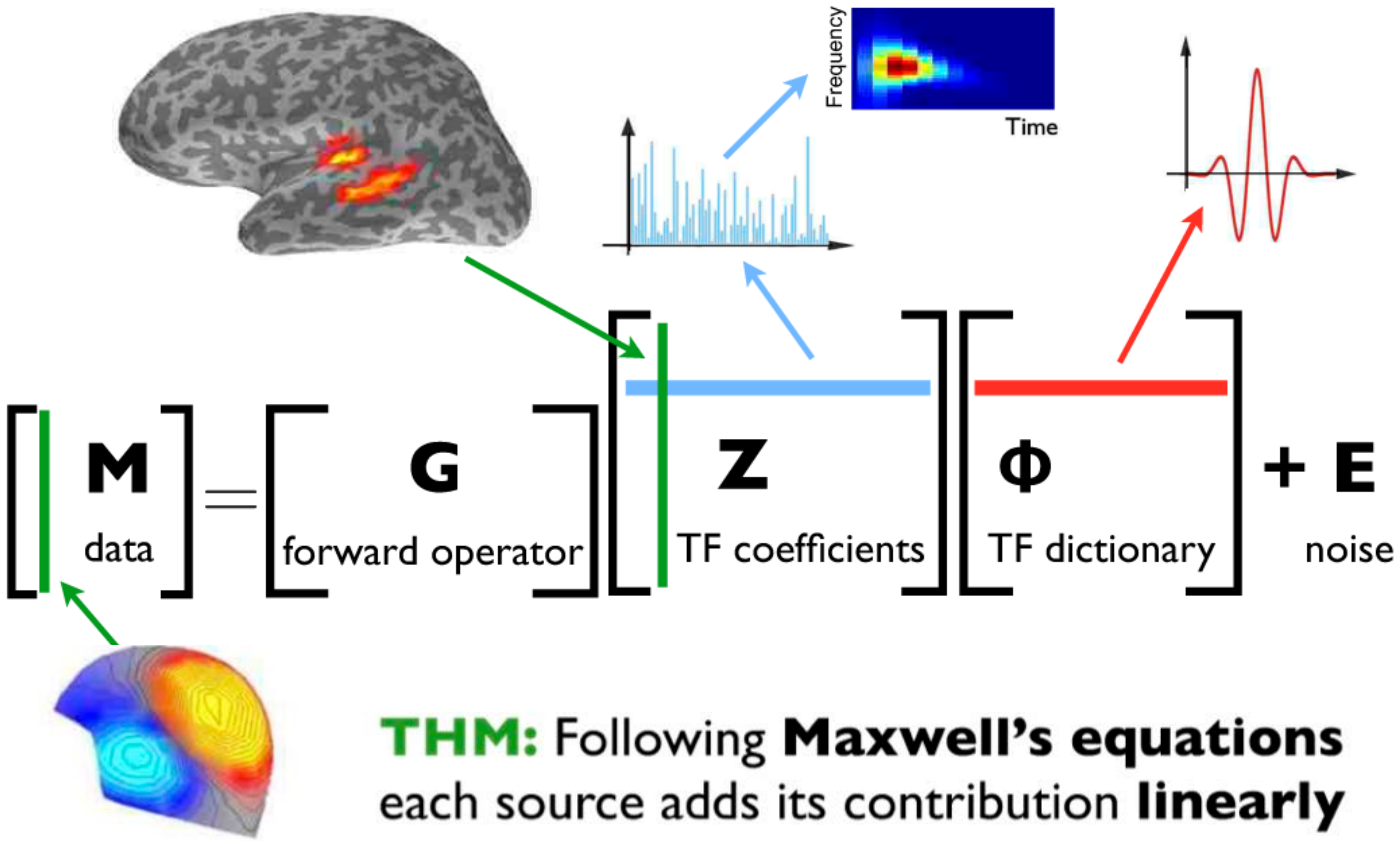
Minimum Current Estimate (MCE – 1999):

$$X = \underset{X}{\text{argmin}} \frac{1}{2}\|M - GX\|^2 + \lambda\|X\|_1$$

Minimum mixed-norm Estimate (MxNE [2012]):

$$X = \underset{X}{\text{argmin}} \frac{1}{2}\|M - GX\|^2 + \lambda\|X\|_{21}$$

M/EEG INVERSE PROBLEM + TF SPARSE CODING

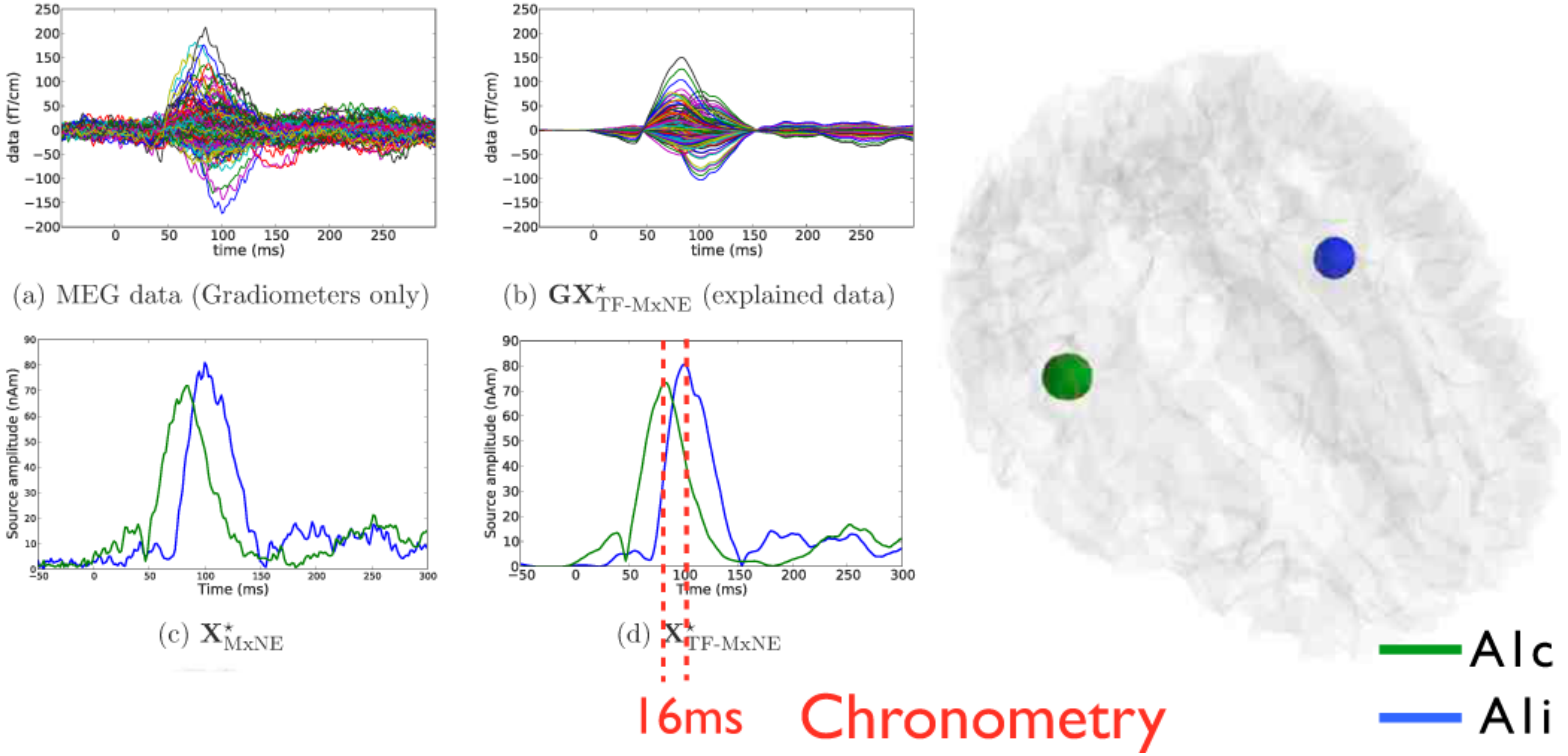


$$Z = \underset{Z}{\text{argmin}} \frac{1}{2}\|M - GZ\Phi^*\|^2 + \mathcal{R}(Z)$$

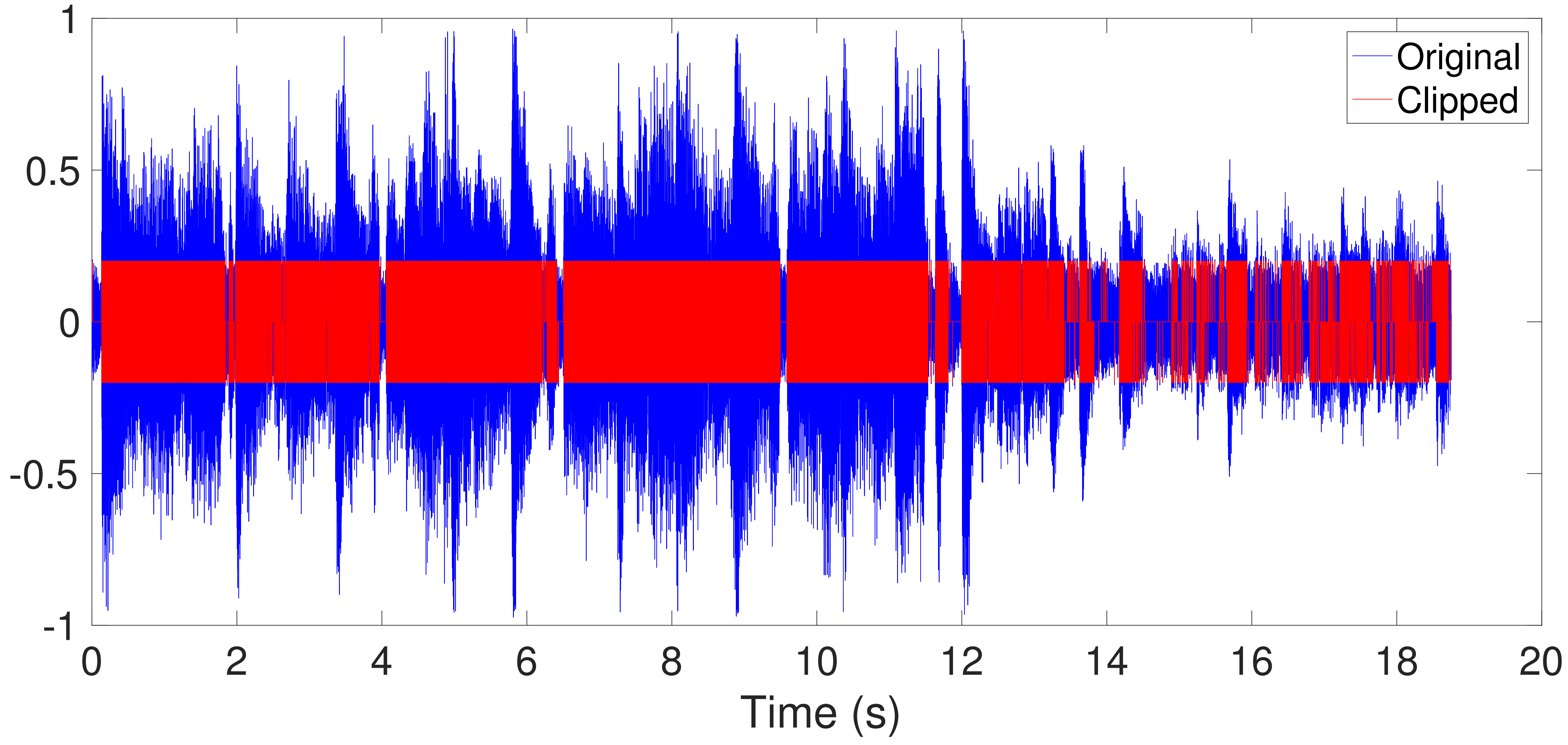
[MK, Gramfort & al 13] :

$$\mathcal{R}(Z) = \gamma\|Z\|_{2,1} + (1-\gamma)\|Z\|_1 \ , \quad 0 \leq \gamma \leq 1$$

M/EEG INVERSE PROBLEM: AUDITORY DATA



AUDIO DECLIPPING

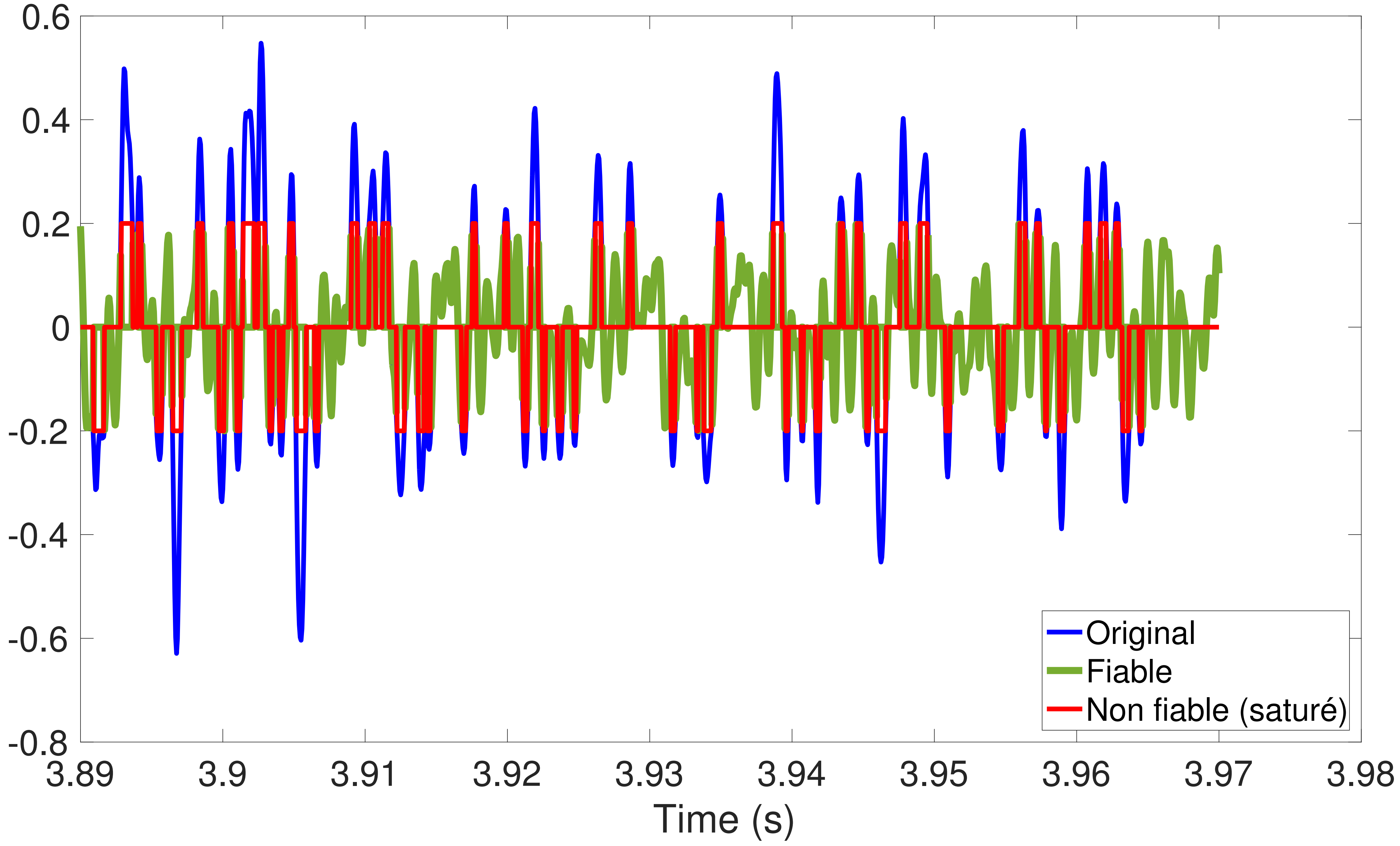


- How to recover the original (blue) signal from the clipped (red) signal ?

(SNR = 4dB)

[A. Adler, V. Emiya & al 2011]

AUDIO DECLIPPING

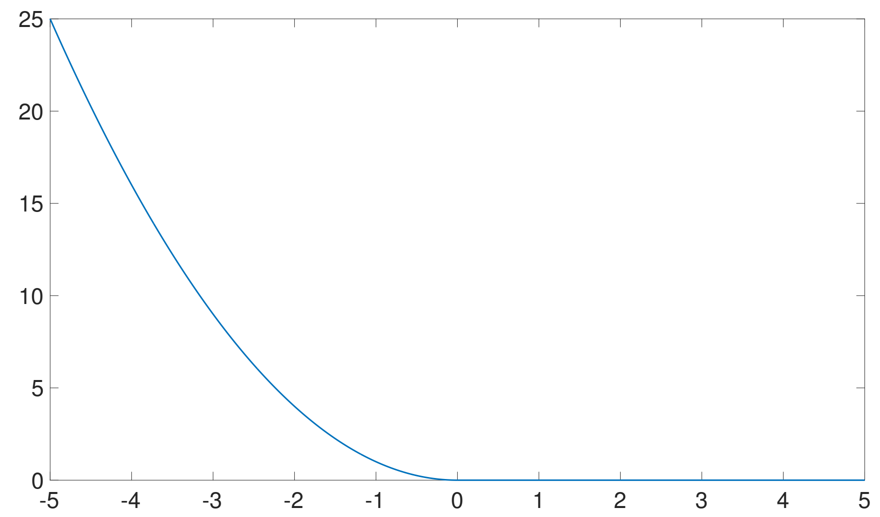


- Reliable samples (green): $y^r = M^r x = M^r \Phi \alpha$
- Loss on reliable samples (green):

$$\frac{1}{2}\|y^r - M^r \Phi \alpha\|_2^2$$

- Clipped samples (red): $y^m = \text{sgn}(M^m x)\theta^{clip}$
- Loss on clipped samples (red):

$$[\theta^{clip} - M^m \Phi \alpha]_+^2$$



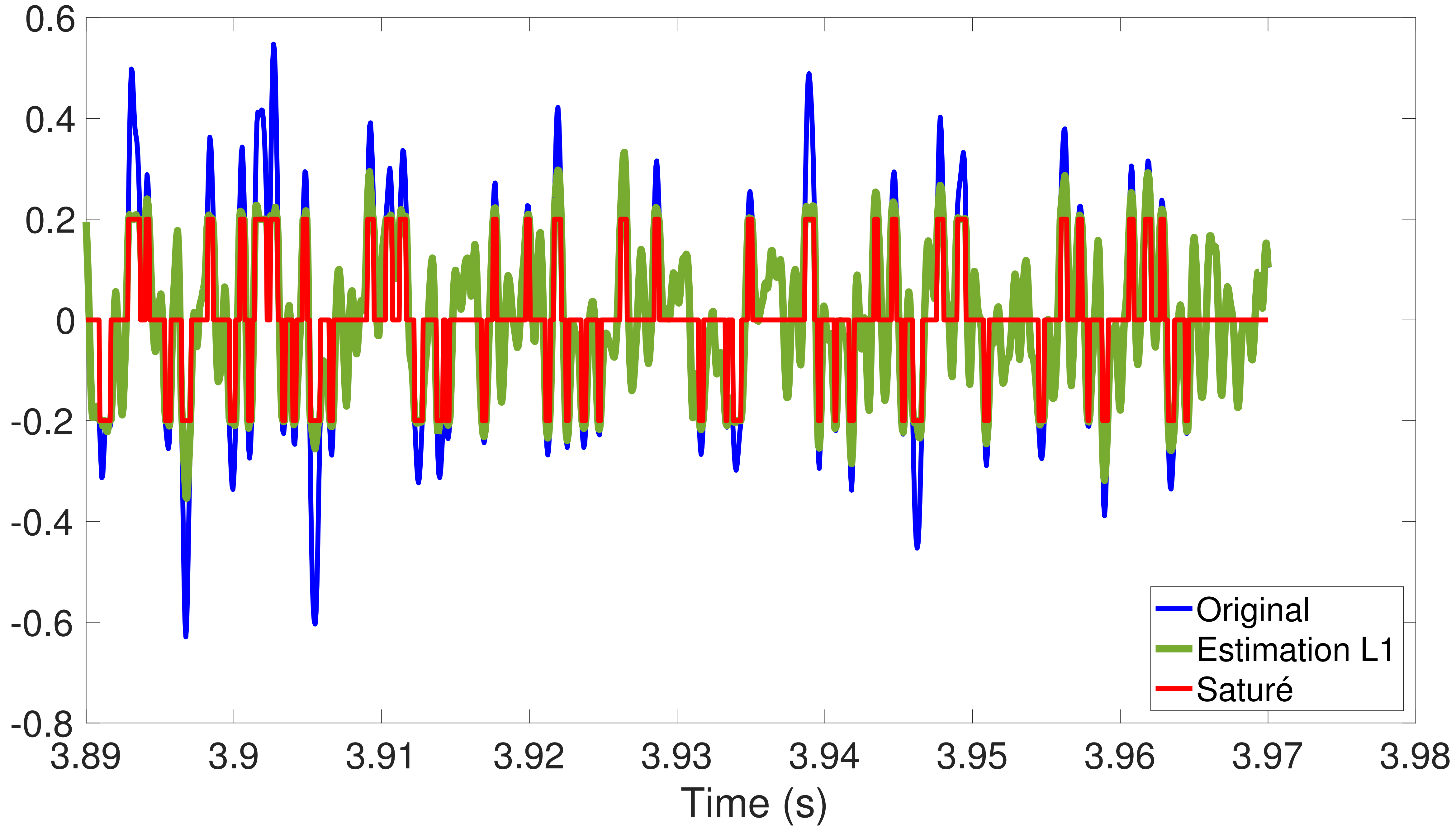
AUDIO DECLIPPING

- Declipping :

$$\hat{\alpha} = \underset{\alpha}{\text{argmin}} \frac{1}{2}\|y^r - M^r\Phi\alpha\|_2^2 + \frac{1}{2}[\theta^{clip} - M^m\Phi\alpha]_+^2 + \lambda\|\alpha\|_1$$

- Algorithm: ISTA
- Results: SNR = 4.5 dB

How to improve the estimation ?



INVERSE PROBLEM WITH SPARSE CODING

$$\alpha = \underset{\alpha}{\text{argmin}} \frac{1}{2}\|y - A\Phi\alpha\|^2 + \mathcal{R}(\alpha)$$

- Which choice for $\mathcal{R}$?
  - (Structured) sparsity:
    - lasso ℓ_1 [Tibshirani 96] [Chen et al. 98]
    - Group-lasso [Yuan, Lin 06] and mixed norms ℓ_{21}, ℓ_{12} [MK 09]
    - composite norms $\ell_{21} + \ell_1$ [Jenatton et al. 11]
- Which algorithm ?
  - Iterative Shrinkage/Thresholding (IST)
- Can we play on the Shrinkage/Thresholding function instead of $\mathcal{R}$?

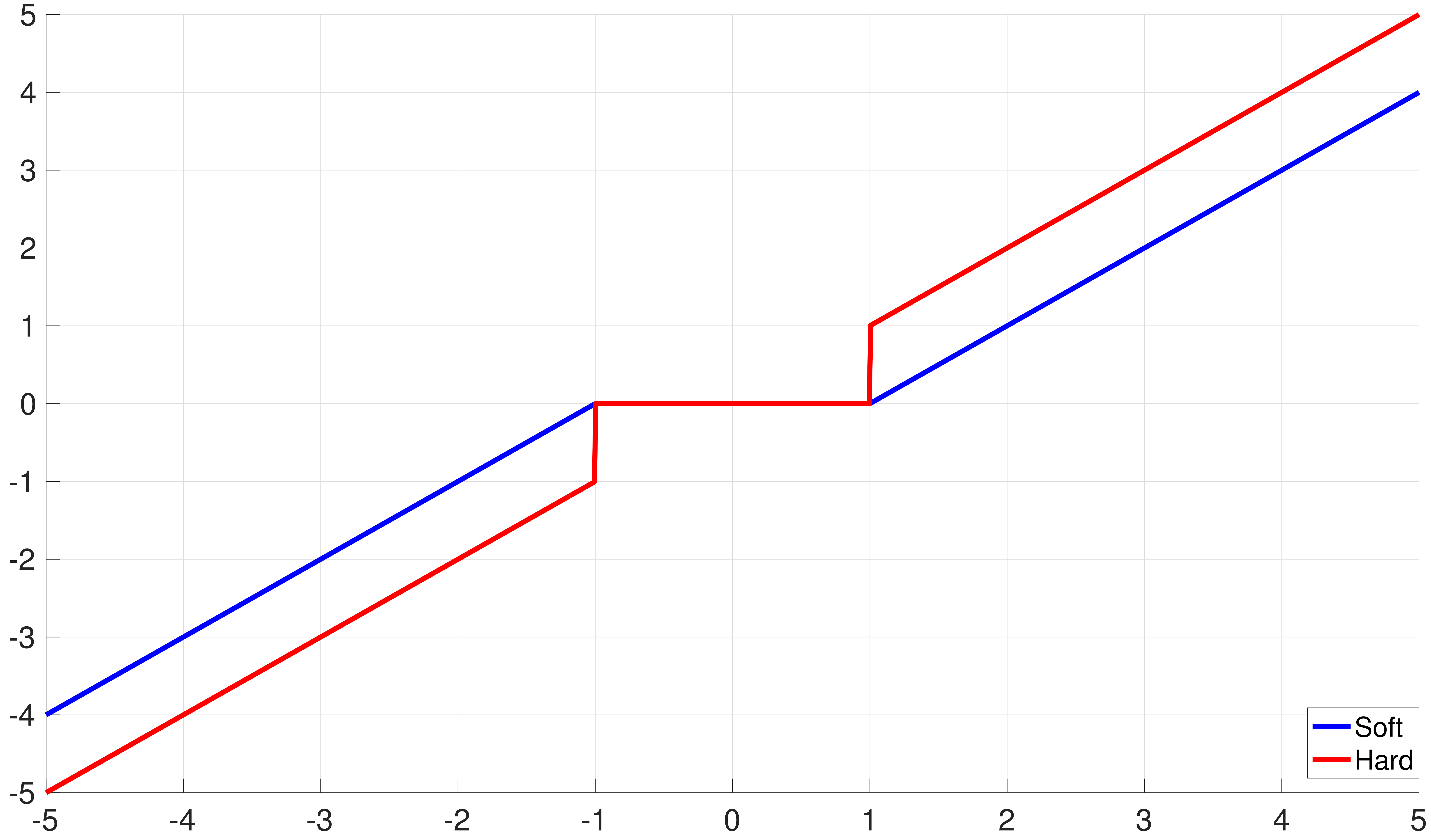
ISTA AND SHRINKAGE RULES

SHRINKAGE RULES

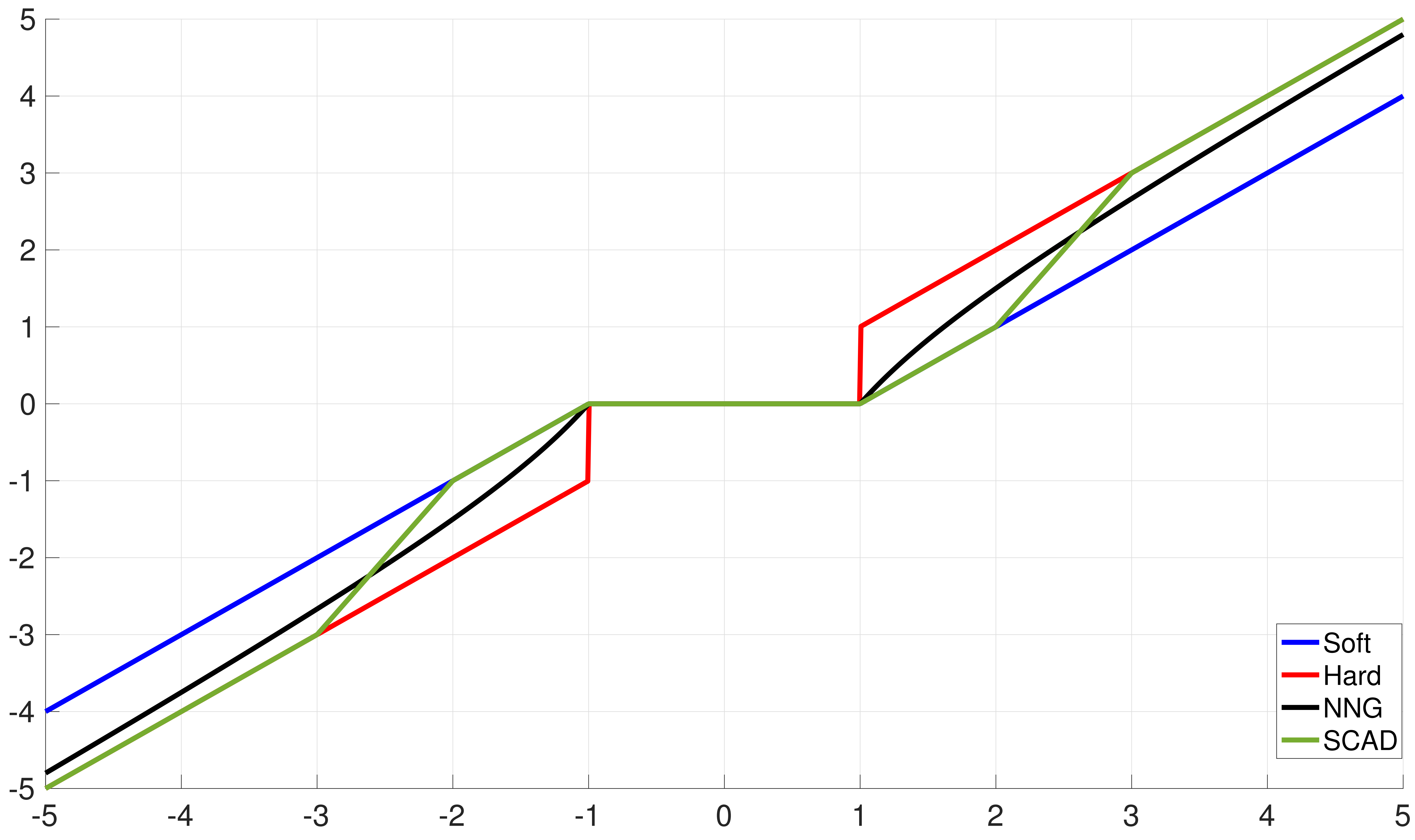
Definition [Antoniadis 07]

- $\mathbb{S}(\,.\,;\lambda)$ is odd – with $\mathbb{S}_+(\,.\,;\lambda)$ its restriction on $\mathbb{R}_+$
- $\mathbb{S}(\,.\,;\lambda)$ is a shrinkage function: $0 \leq \mathbb{S}_+(t;\lambda) \leq t,\ \forall t \in \mathbb{R}_+$.
- $\mathbb{S}_+(\,.\,;\lambda)$ is non decreasing and $\lim_{t \to +\infty} \mathbb{S}(t;\lambda) = +\infty$

SHRINKAGE RULES EXEMPLES



SHRINKAGE RULES EXEMPLES



SHRINKAGE RULES PROPERTIES

- **Definition**

A function f is semi-convex iff it exists c such that $x \mapsto f(x) + \frac{c}{2}\|x\|^2$ is convex

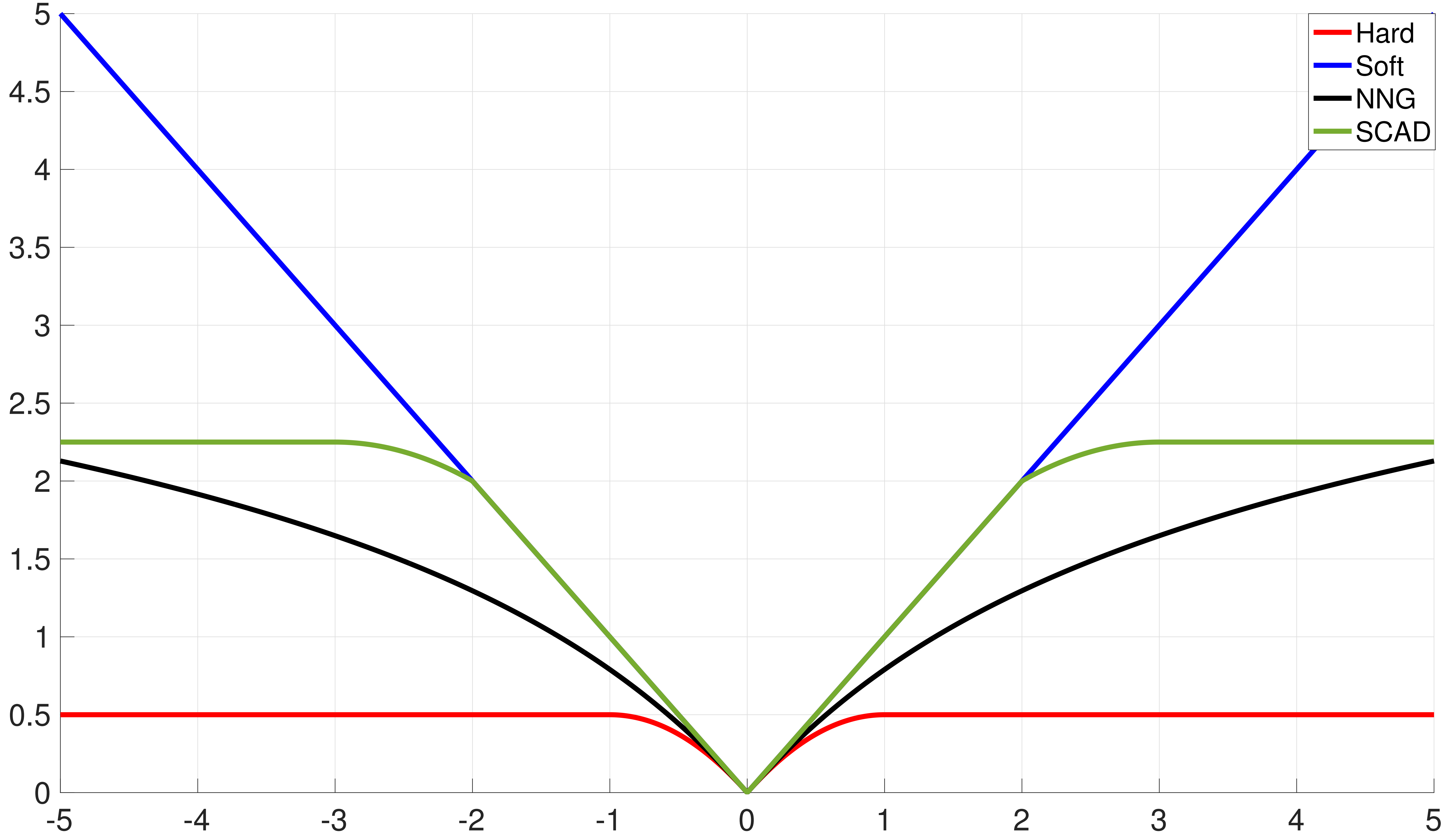
- **Theorem 1** [MK 14]

We can associate a semi-convex penalty $P(\,.\,;\lambda)$ with $c \leq 1$ to any thresholding rule

- **Theorem 2** [MK 14]

(Relaxed) ISTA converges with any thresholding rules

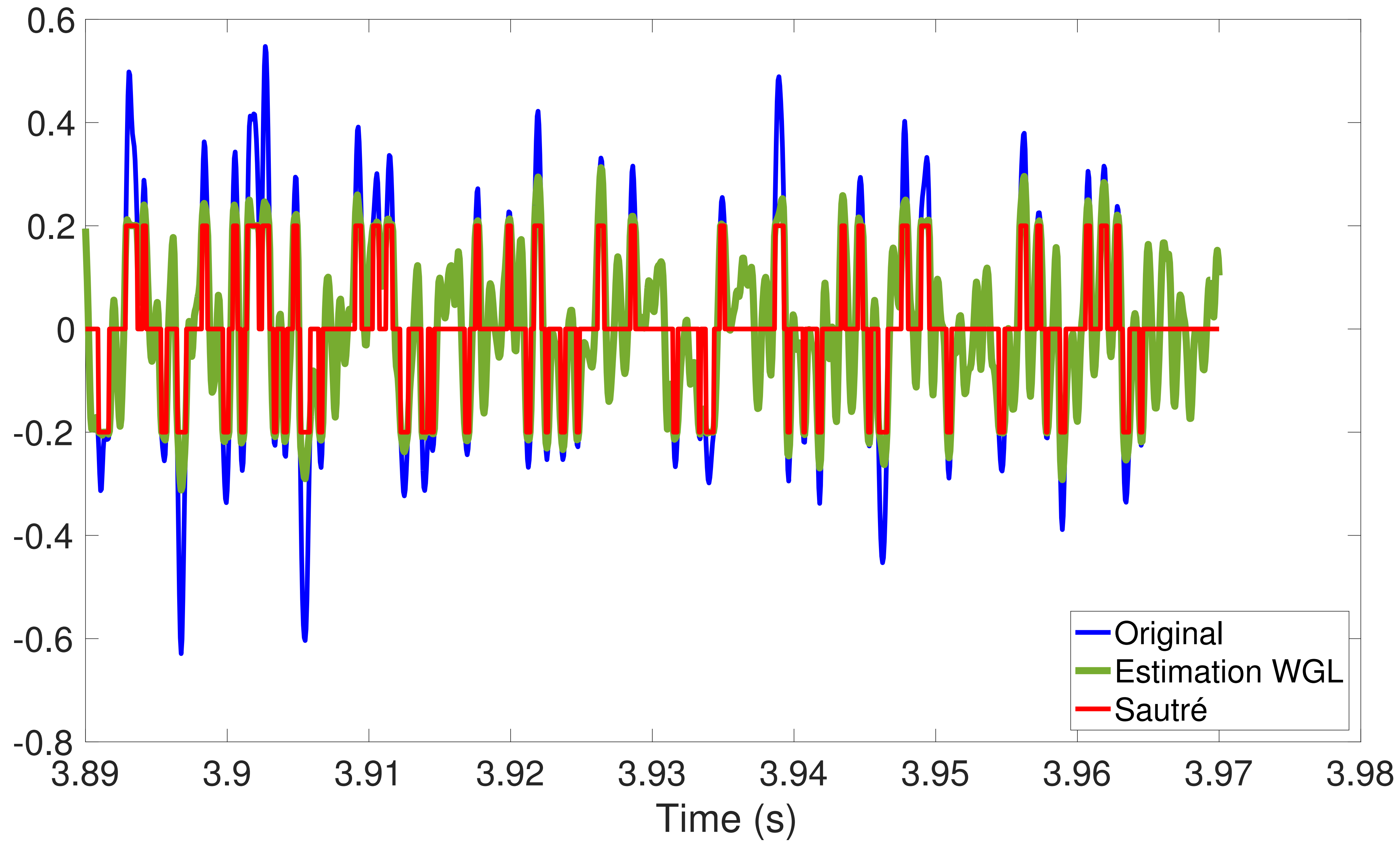
SHRINKAGE RULES EXEMPLES



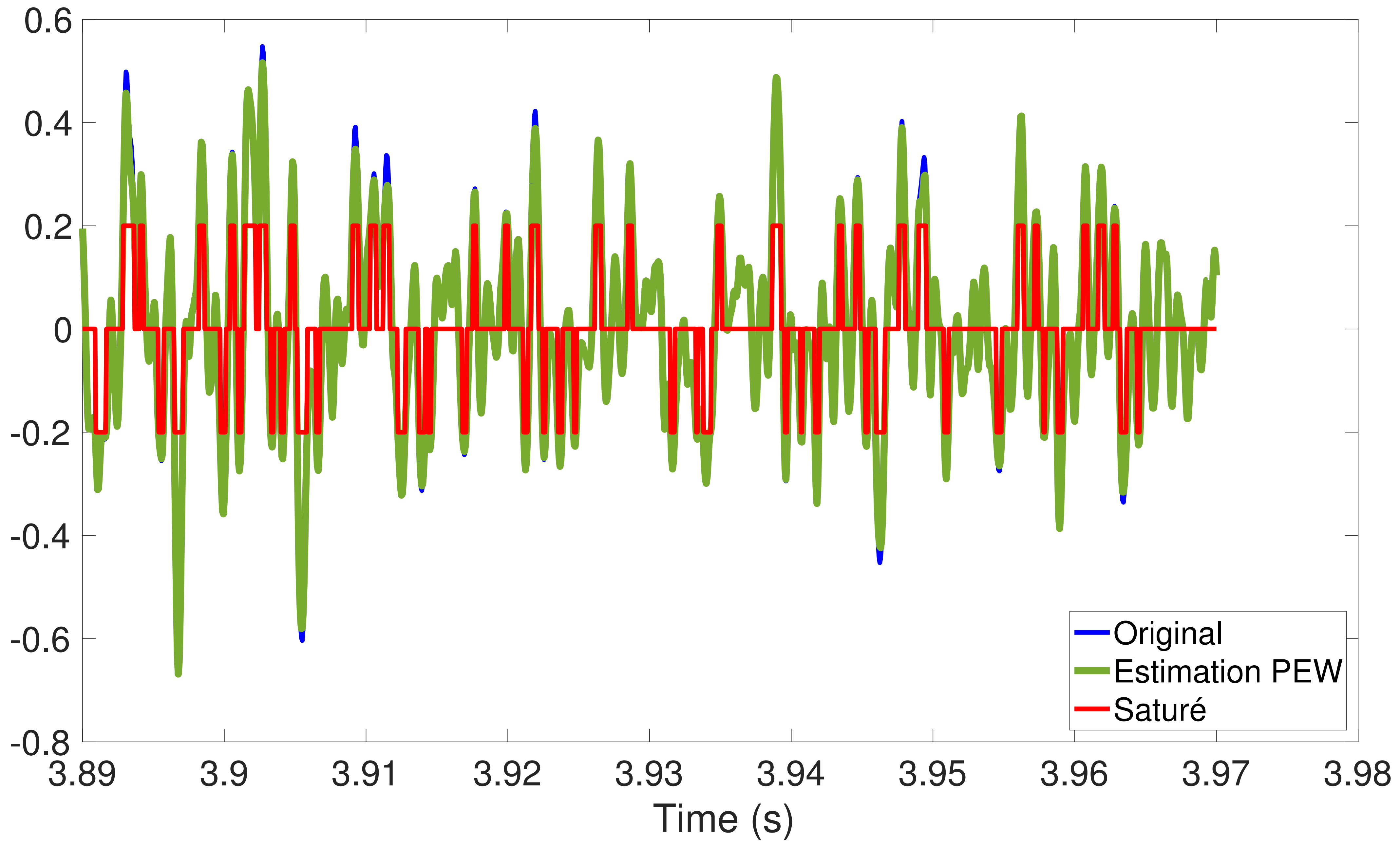
EXTENSION: STRUCTURED/SOCIAL SPARSITY [MK & AL 13]

- Lasso: $\hat{\alpha}_k = \alpha_k \left(1 - \frac{\lambda}{|\alpha_k|}\right)^+$
- NonNegative Garotte / Empirical Wiener: $\hat{\alpha}_k = \alpha_k \left(1 - \frac{\lambda^2}{|\alpha_k|^2}\right)^+$
- Windowed group-Lasso: $\hat{\alpha}_k = \alpha_k \left(1 - \frac{\lambda}{\|\alpha_{\mathcal{N}(k)}\|_2}\right)^+$
- Persistent Empirical Wiener: $\hat{\alpha}_k = \alpha_k \left(1 - \frac{\lambda}{\|\alpha_{\mathcal{N}(k)}\|_2^2}\right)^+$

RESULTS ON DECLIPPING



WGL (SNR 4.9 dB)



PEW (SNR 15 dB)

PARTIAL CONCLUSION

- Take-Home message
  - Sparsity using a dictionary
  - Play on the thresholding rules
  - Flexible structure in a neighborhood

DICTIONARY LEARNING

DICTIONARY LEARNING: GENERAL FORMULATION

- Let $X = (x_1, x_2, \ldots, x_M)$ a set of M signals in $\mathbb{R}^N$
- We seek a dictionary $\Phi \in \mathbb{R}^{NK}$, $K > N$, such that each signal x_m admits a sparse representation
- $x_m = \Phi \alpha_m$, with $\alpha_m \in \mathbb{R}^K$, and α_m is $s-$sparse
- General formulation:

$$\min_{\Phi, \alpha_1, \ldots, \alpha_m} \sum_m \|x_m - \Phi \alpha_m\|^2 \quad s.t. \quad \|\alpha_m\|_0 \leq s \ \forall m$$

- With $A = (\alpha_1, \ldots, \alpha_m)$:

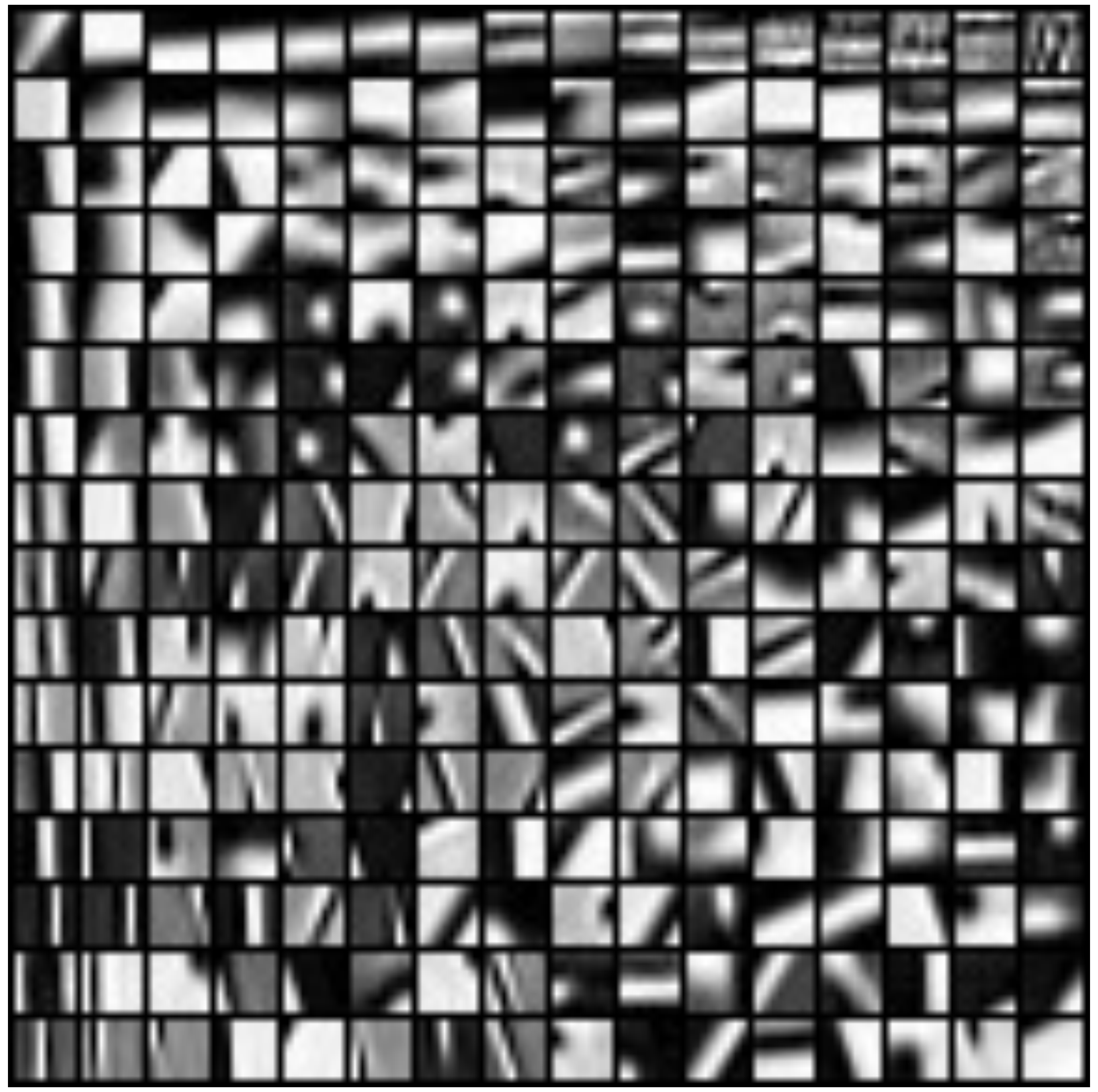
$$\min_{\Phi, A} \|X - \Phi A\|^2 \quad s.t. \quad \|\alpha_m\|_0 \leq s \ \forall m$$

DICTIONARY LEARNING: ALTERNATING MINIMIZATION

- Idea: alternate minimization between A and Φ
- Initialize Φ
- Sparse coding for each x_m: $\alpha_m = \underset{\alpha}{\operatorname{argmin}} \frac{1}{2}\|x_m - \Phi\alpha\|^2 + \lambda\|\alpha\|_1$
- Dictionary update : $\Phi = \Phi + \frac{1}{\|A\|^2}(X - \Phi A)A^*$
- Can be implemented "on line"
- Online dictionary learning [Mairal, Bach, Ponce, Sapiro 2009], K-SVD [Elad and Aharon, 2006]

EXAMPLES

- K-SVD [Elad and Aharon, 2006]
- Dictionary learned on a noisy version of the boat image



EXAMPLES

- Online dictionary learning [Mairal, Bach, Ponce, Sapiro 2009]

DICTIONARY LEARNING AS BI-LEVEL OPTIMIZATION

Let : $F(\alpha, \Phi) = \frac{1}{2}\|y - \Phi\alpha\|^2 + \lambda\|\alpha\|_1$

Dictionary learning as bi-level optimization:

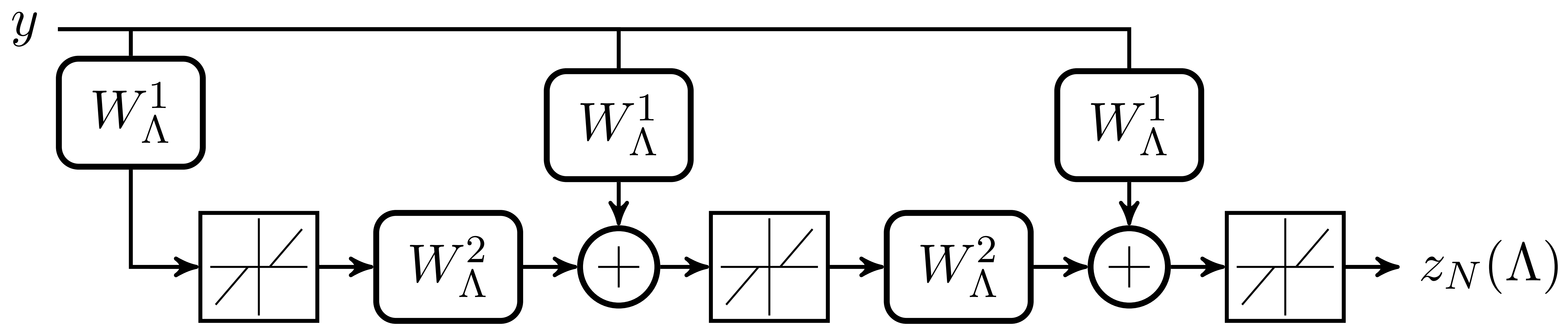
$$\min_{\Phi \in \mathscr{C}} F(\alpha^*(\Phi), \Phi) \quad \text{s.t.} \quad \alpha^*(\Phi) = \underset{\alpha}{\text{argmin}}\, F(\alpha, \Phi)$$

Where $\alpha^* = \text{ISTA}(x, \Phi) = \lim_{t \to +\infty} \mathcal{S}_{\lambda/L}\left(\alpha^{(t)} + \frac{1}{L}\Phi^*(y - \Phi\alpha^{(t)})\right)$

Idea: unroll N iterations of ISTA like a deep neural network with N layers

LEARNING ISTA [GREGOR AND LE CUN 10]

DDL: back-propagation through the algorithm



KKT conditions of the Lasso and recurrent neural network:

$$W_\Phi^1 = \frac{1}{L}\Phi^T$$

$$W_\Phi^2 = \left(I - \frac{1}{L}\Phi^T\Phi\right)$$

DICTIONARY LEARNING THROUGH DEEP NETWORK

Let : $F(\alpha, \Phi) = \frac{1}{2}\|y - A\Phi\alpha\|^2 + \lambda\|\alpha\|_1$

Dictionary learning as bi-level optimization:

$$\min_{\Phi \in \mathscr{C}} F(\alpha^*(\Phi), \Lambda) \quad \text{s.t.} \quad \alpha^*(\Lambda) = \operatorname{argmin}_{\alpha \in \mathbb{R}^L} F(\alpha, \Phi)$$

Where $\alpha^* = \text{ISTA}(x, \Phi) = \lim_{t \to +\infty} \mathcal{S}_{\lambda/L}\left(\alpha^{(t)} + \frac{1}{L}\Phi^* A^*(y - A\Phi\alpha^{(t)})\right)$

1. Compute $\alpha^N(\Phi) = \text{ISTA}^N(x, \Phi)$

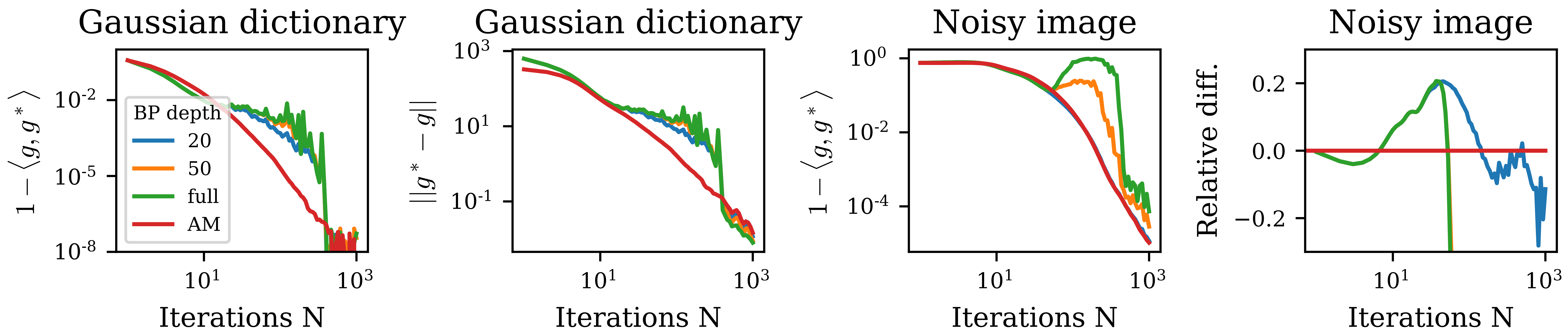
2. Compute the gradient of $F(\alpha^N(\Phi), \Phi)$ using autodifferentiation

DDL VS AM [B. MALÉZIEUX ET AL., ICLR'22]

$$\min_{\Phi \in \mathscr{C}} G_N(\Phi) = F(\alpha^{\mathbf{N}}(\Phi), \Phi)$$

- Ideal Gradient: $g^* = \nabla_\Phi G(\Phi) = \nabla_2 F(\alpha^*(\Phi), \Phi)$
- AM: $g_N^{AM}(\Phi) = \nabla_2 F(\alpha^N(\Phi), \Phi)$ Linear convergence
- DDL: $g_N^{DDL}(\Phi) \in \nabla_2 F(\alpha^N(\Phi), \Phi) + J_N \left(\partial_1 F(\alpha^N(\Phi), \Phi\right)$ Quadratic convergence on the support
- $\|J_N - J^*\| \leq A_N + B_N$, where $A_N \to 0$ and B_N an error term

DDL : NUMERICAL EXPERIMENTS [B. MALÉZIEUX ET AL., ICLR'22]

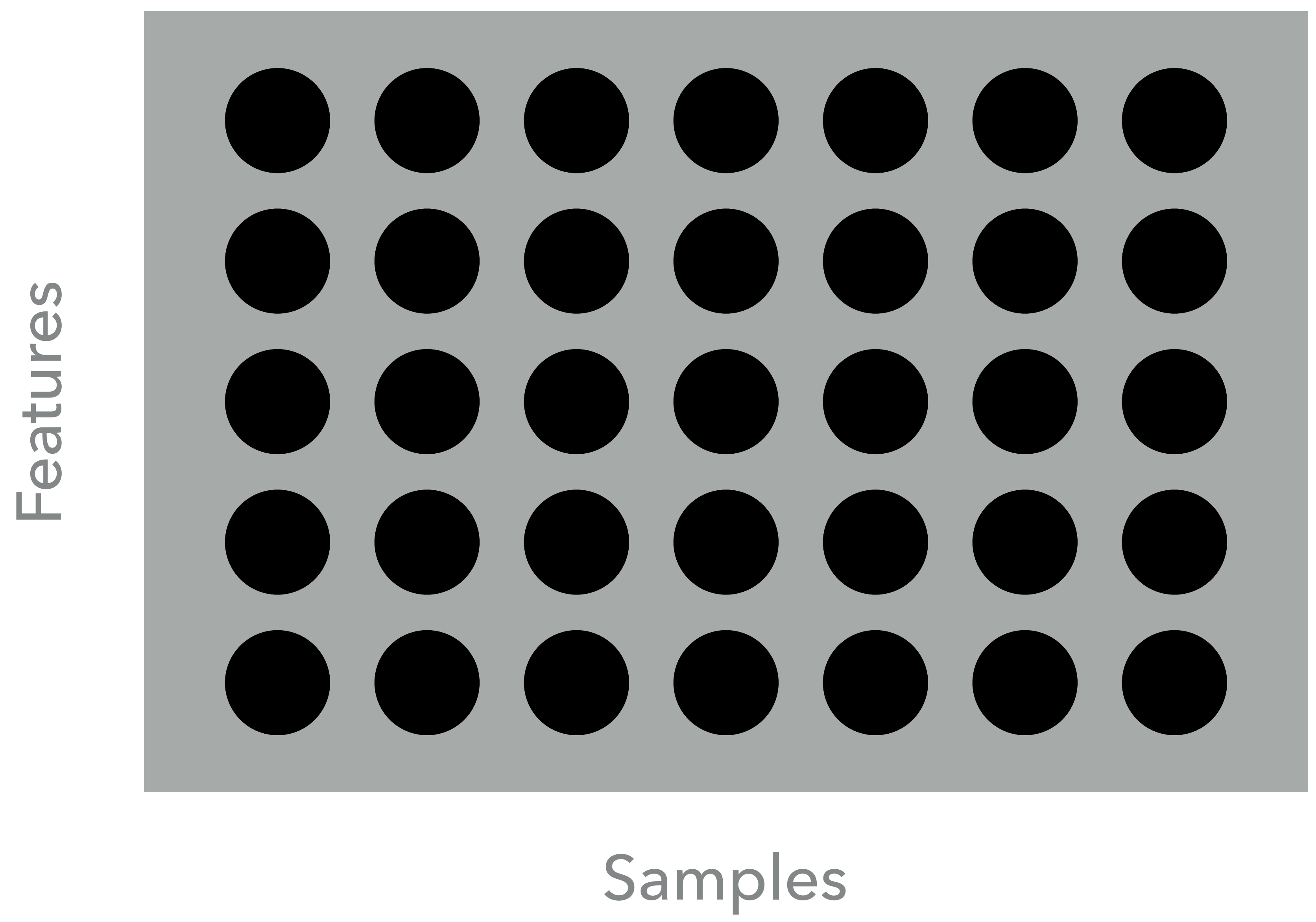


- First iterations: stable behavior
- Too many iterations: accumulations of errors !
- On the support: convergence to g^*

DICTIONARY LEARNING AND MATRIX FACTORIZATION

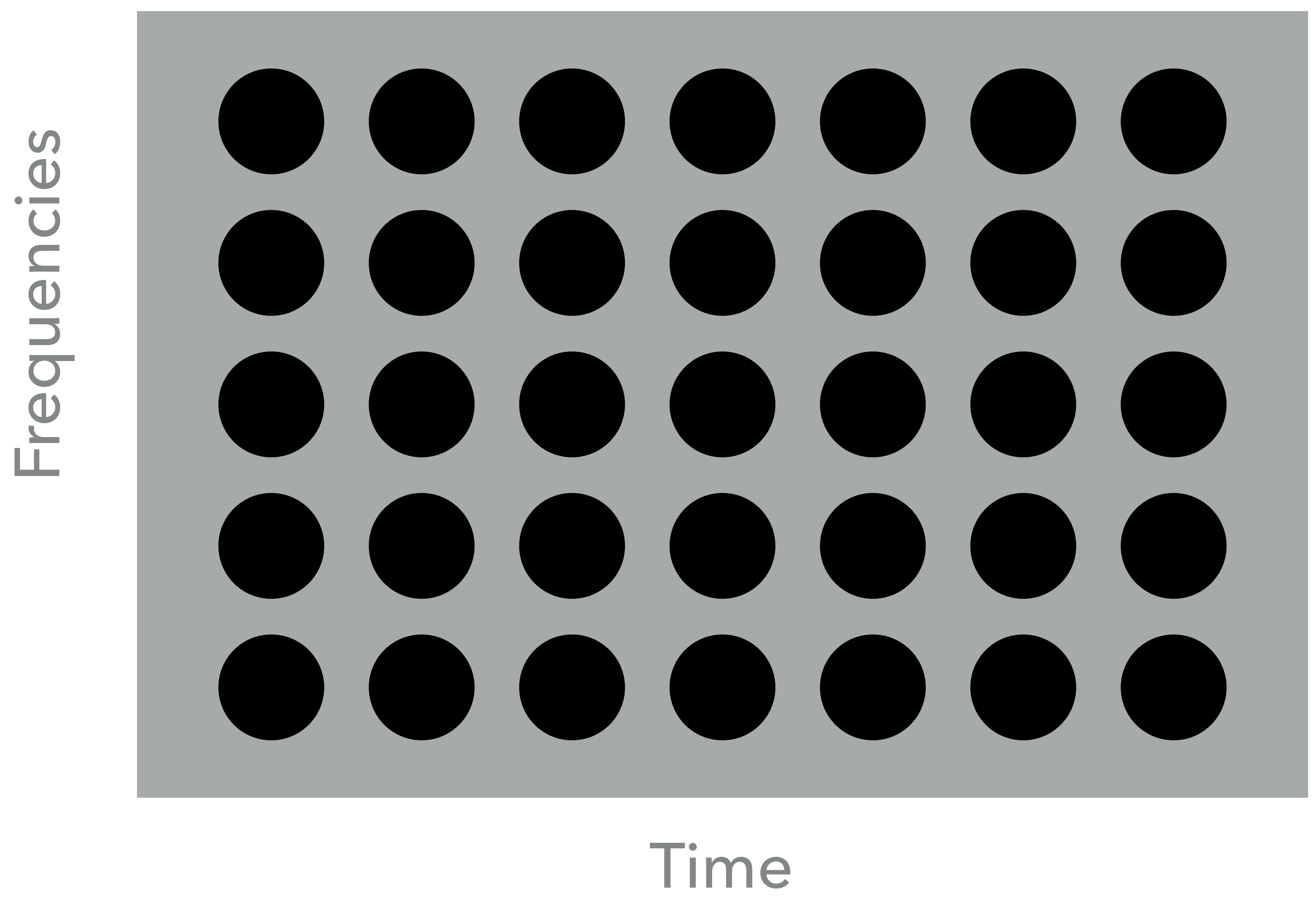
MATRIX FACTORIZATION

- Data are often available under a matrix form:



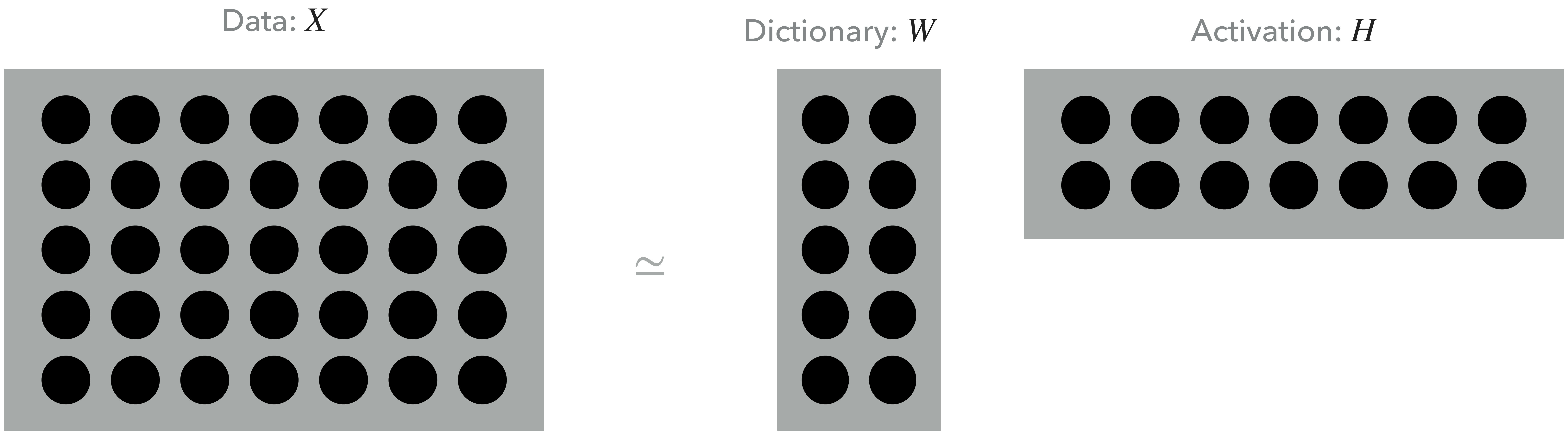
MATRIX FACTORIZATION

- Exemple: time-frequency representation



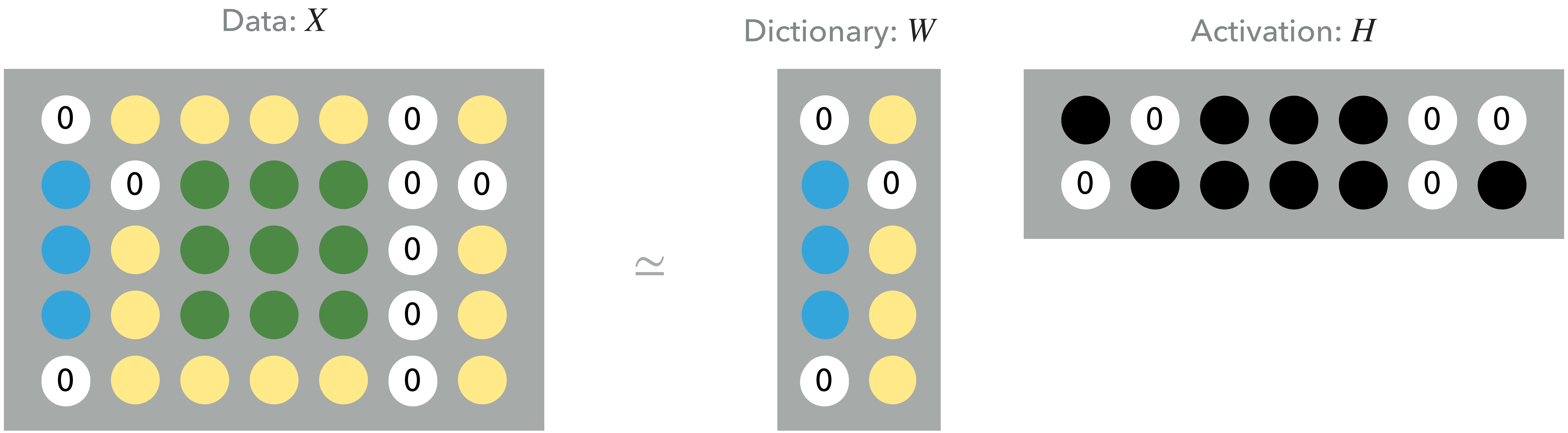
MATRIX FACTORIZATION

- Let $X \in \mathbb{R}^{MN}$. We aims to factorize $X \simeq WH$, with $W \in \mathbb{R}^{MK}$ and $H \in \mathbb{R}^{KN}$



MATRIX FACTORIZATION

- Let $X \in \mathbb{R}^{MN}$. We aims to factorize $X \simeq WH$, with $W \in \mathbb{R}^{MK}$ and $H \in \mathbb{R}^{KN}$



MATRIX FACTORIZATION: EXAMPLES

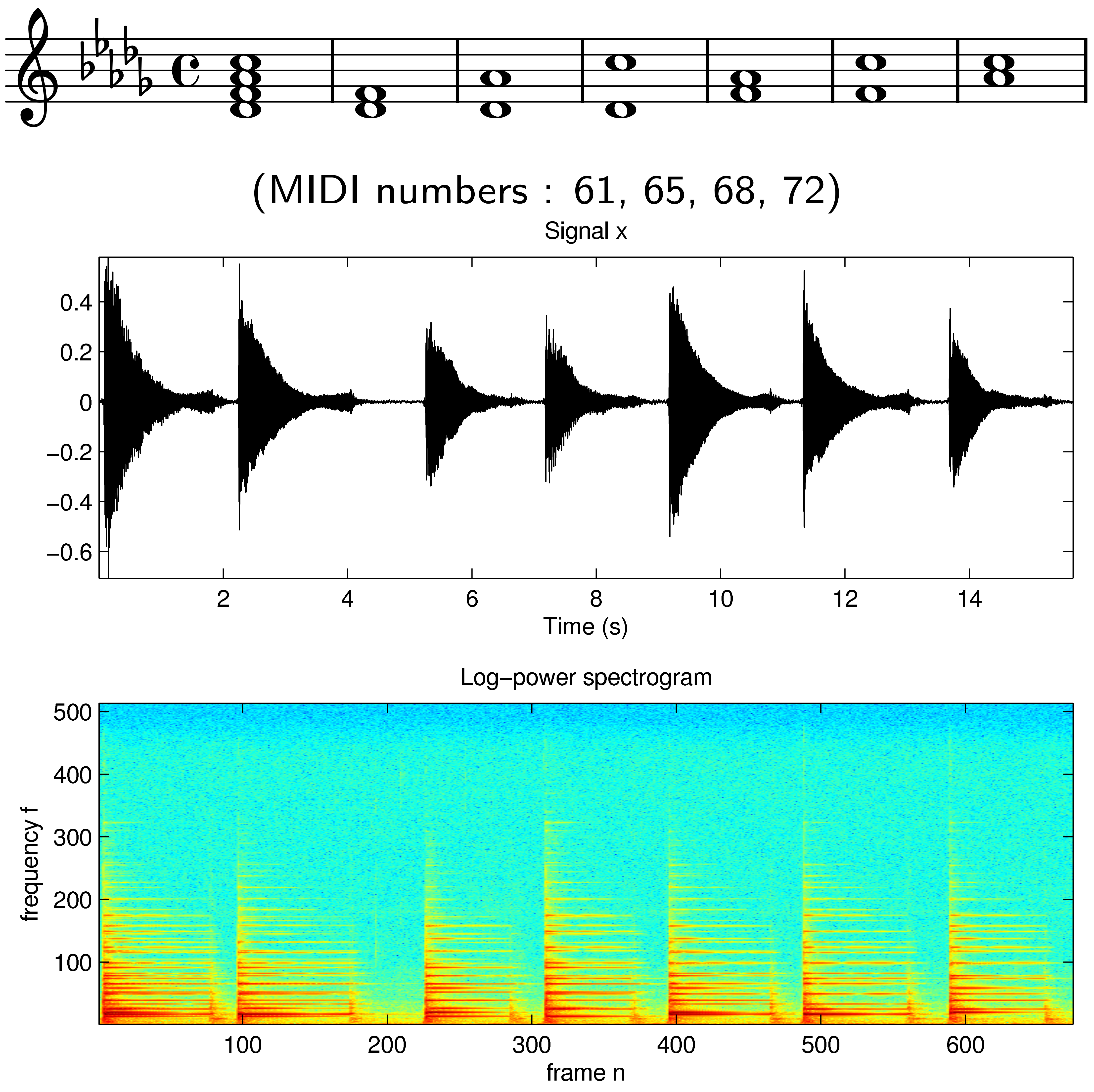
- PCA : $X \in \mathbb{R}^{NN} = UDU^T$, U being the eigen vectors (and is orthogonal) and D the eigen values
- SVD: $X \in \mathbb{R}^{MN} = UDV^T$, D being the singular values, U and V are orthogonal
- ICA
- etc.

NON NEGATIVE MATRIX FACTORIZATION

- $X \in \mathbb{R}_+^{MN}$ a matrix with non negative entries
- $X = WH$ with $W \in \mathbb{R}_+^{MK}$ and $H \in \mathbb{R}_+^{KN}$
- W is a matrix of non negative features (exemples: spectrum)
- H is a mature of non negative coefficients
- The Features can only add each other (no soustraction)
- Facilitate the interpretability

NMF EXAMPLES

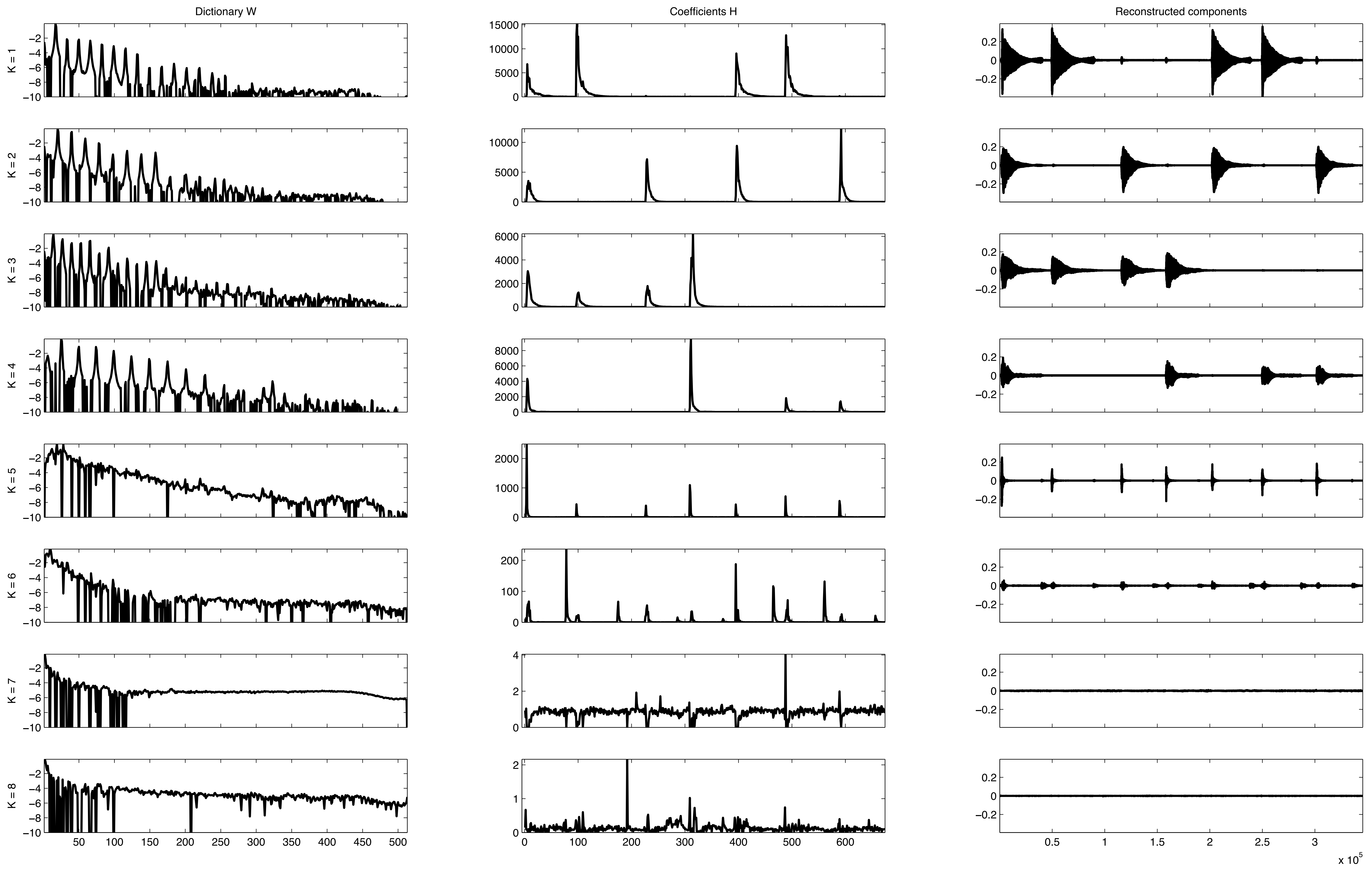
- Piano example from [Fevotte 2009]
- Demo available at:
  https://www.irit.fr/~Cedric.Fevotte/extras/machine_audition/



NMF EXAMPLES

- Piano example from [Fevotte 2009]
- Demo available at:

https://www.irit.fr/~Cedric.Fevotte/extras/machine_audition/



NMF DECOMPOSITIONS

- Choose a measure of fit between X and WH :

$$W, H = \underset{W,H \geq 0}{\operatorname{argmin}} \; D(X \mid WH)$$

- $D(X \mid WH) = \sum_{n,k} d(X[n,k] \mid \{WH\}[n,k])$, where d is a scalar cost function
- Exemple of cost functions :
  - Euclidian : $D(X \mid WH) = \|X - WH\|^2$ [Paatero & Tapper, 1994] [Lee & Seung, 2001]
  - Kullback-Leibler divergence: $d(x \mid y) = x \log \frac{x}{y} + (x - y)$ [Lee & Seung, 1999] [Finesso & Spreij, 2006]
  - Itakura-Saito divergence: $d(x \mid y) = \frac{x}{y} - \log \frac{x}{y} - 1$ [Févotte, Bertin & Durrieu, 2009]
- Optimisation by multiplicative update (keep the non negativity)

LOW-RANK TIME-FREQUENCY SYNTHESIS [MK & FEVOTTE 14,18]

Let $\Phi = \{\varphi_{f,n}\}$ a TF dictionary. The model reads

- $x = \Phi\alpha + e$
- $\alpha \sim \mathcal{N}_c(0,\mathrm{diag}(WH))$ with $W \in \mathbb{R}_+^{FK}$ and $H \in \mathbb{R}_+^{KT}$
- $e \sim \mathcal{N}(0,\lambda)$
- Remark: related to the sparse bayesian learning [Wipf 04]

LRTFS — ESTIMATION

$$C_{\mathsf{JL}}(\alpha, W, H, \lambda) = -\log p(x, \alpha \,|\, W, H, \lambda)$$
$$= \frac{1}{\lambda}\|x - \Phi\alpha\|_2^2 + D_{\mathsf{IS}}(|\alpha|^2 \,|\, WH) + \log(|\alpha|^2) + M \log \pi$$

- With

$$D_{\mathsf{IS}}(|\alpha|^2 \,|\, WH) = \sum_{fn} d_{\mathsf{IS}}(|\alpha_{fn}|^2 \,|\, [WH]_{fn}) = \sum_{fn} \frac{|\alpha_{fn}|^2}{[WH]_{fn}} - \log\left(\frac{|\alpha_{fn}|^2}{[WH]_{fn}}\right) - 1$$

- And

$$D_{\mathsf{IS}}(|\alpha|^2 \,|\, WH) + \log(|\alpha|^2) = \sum_{fn} \frac{|\alpha_{fn}|^2}{[WH]_{fn}} + Cst$$

LRTFS — ALGORITHM

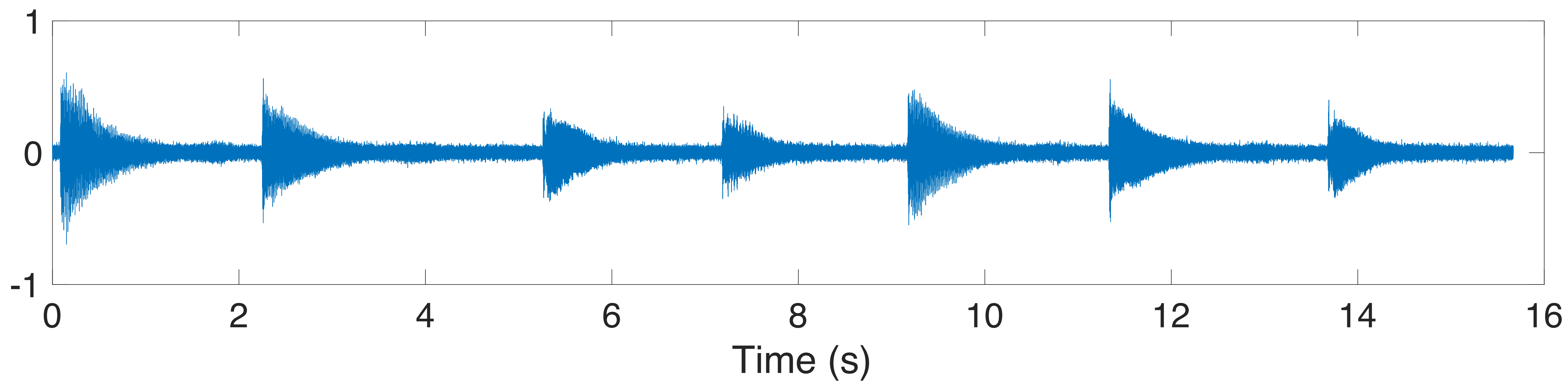
1. $v^{(i)} = W^{(i)}H^{(i)}$

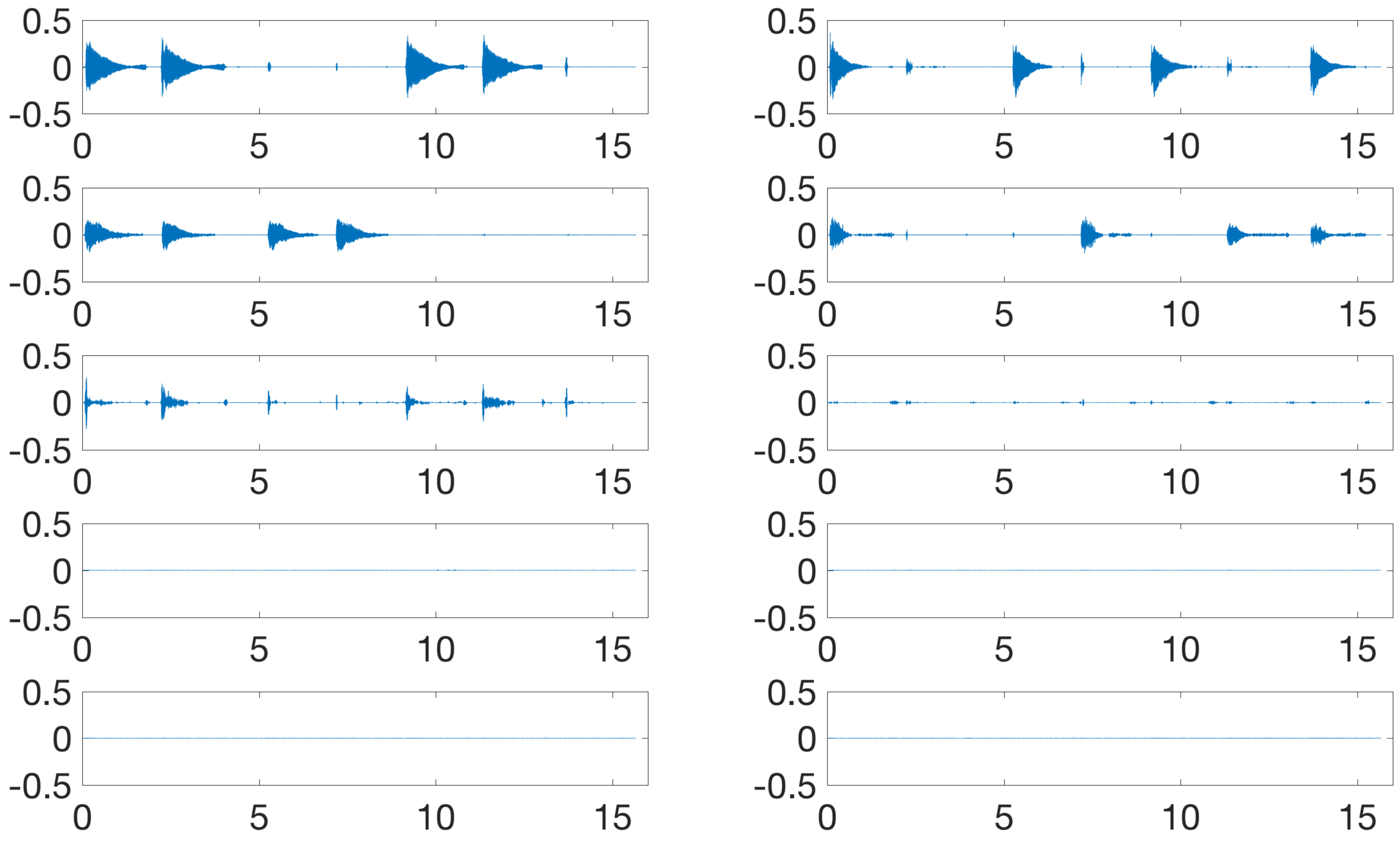
2. $z^{(i)} = \alpha^{(i)} + \frac{1}{L}\Phi^*(x - \Phi\alpha^{(i)})$

3. $\forall(f,n),\ \alpha_{fn}^{(i+1)} = \frac{v_{fn}^{(i)}}{v_{fn}^{(i)} + \lambda/L} z_{fn}^{(i)}$

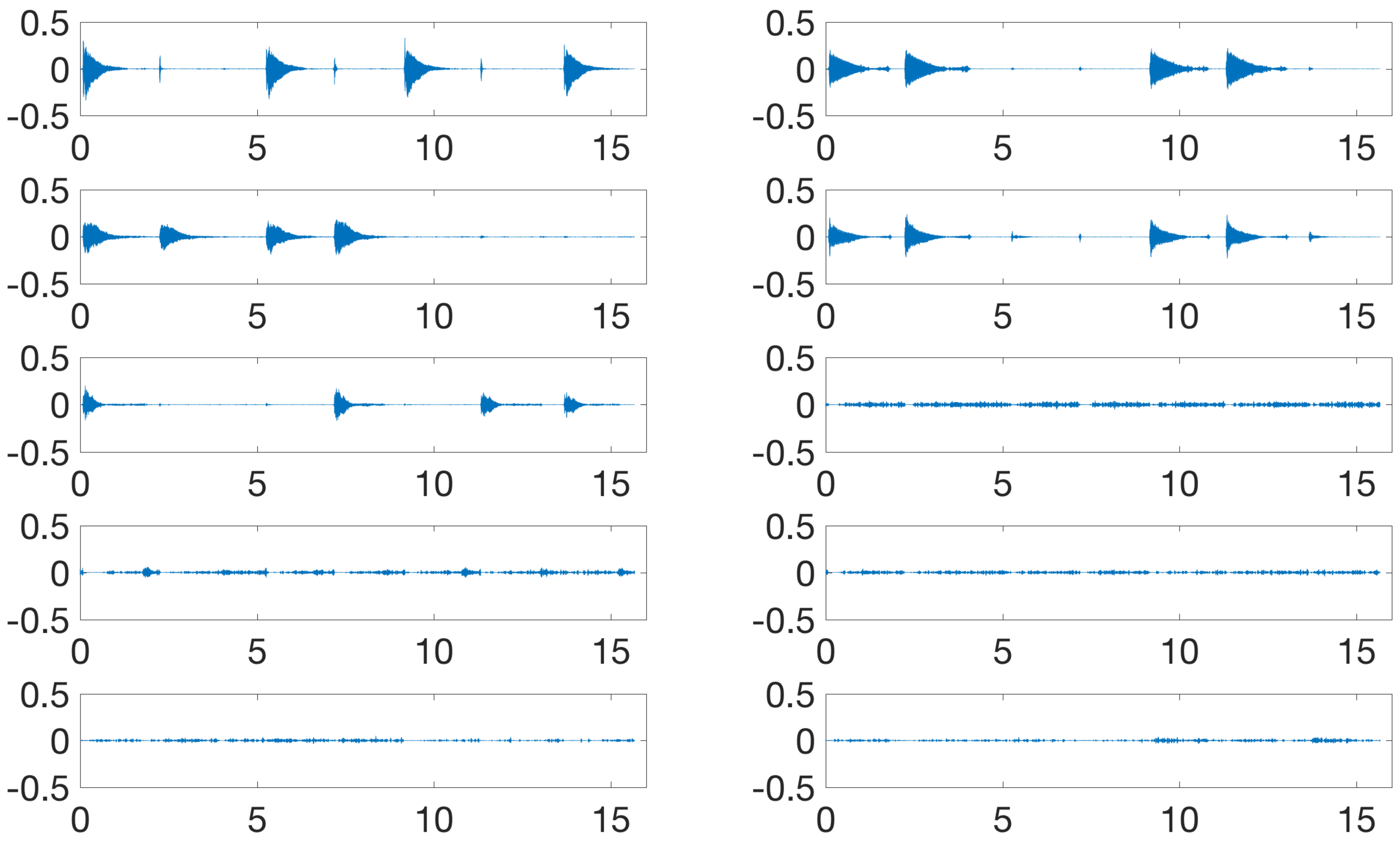
4. $(W^{(i+1)}, H^{(i+1)}) = \underset{W,H\geq 0}{\text{argmin}} \sum_{fn} D_{\text{IS}}\left(|\alpha_{fn}^{(i+1)}|^2 \,|\, [WH]_{fn} \right)$

EXEMPLES: PIANO + NOISE





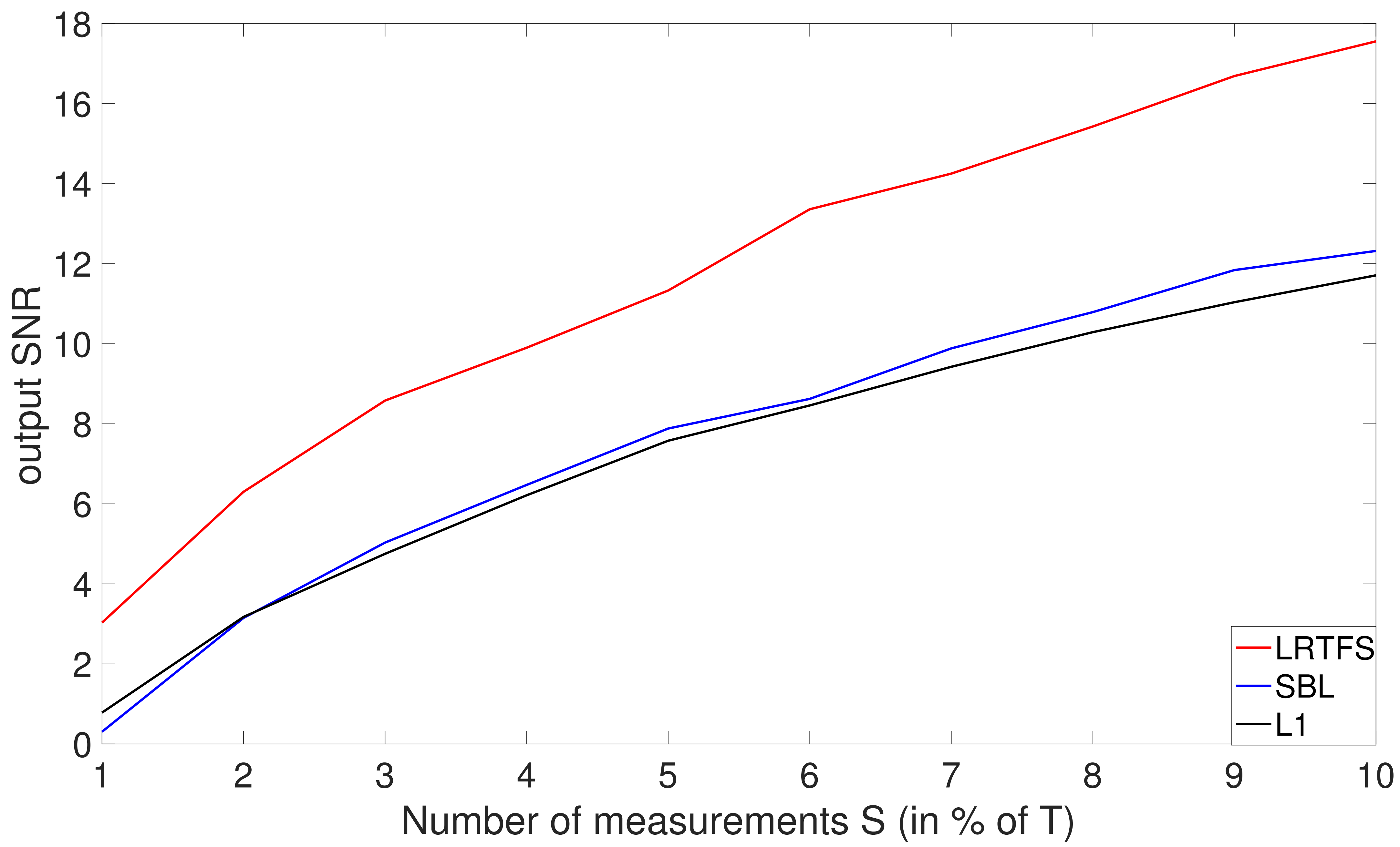
LRTFS

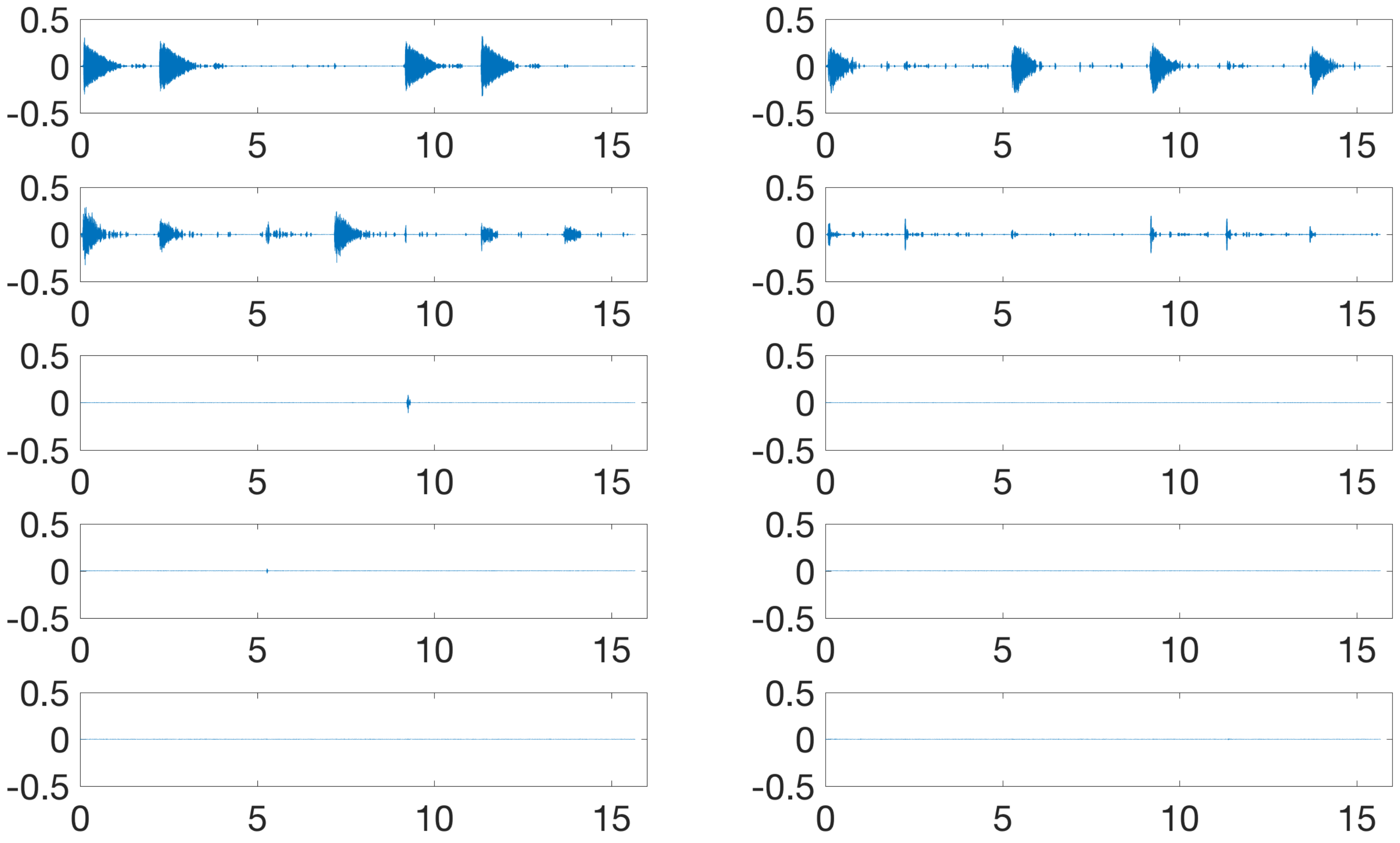


IS-NMF on spectrogram

EXAMPLES: PIANO AND COMPRESSED SENSING

- $y = A\Phi\alpha + e$ where A random





CONCLUSION

- Take-home message
  - NMF can be though as a "prior’
  - Simple IST-like algorithm
  - Scale well to learn a dictionary of frequency pattern W
- Other example of learned W in a classical NMF context for percussive+harmonic separation (C. Laroche’s thesis)
- Strong links between IS-NMF and Convolutive Sparse Coding (work in progress)

GENERAL CONCLUSION

- Sparse coding is now state of the art for inverse problem
- But the wanted solution must be sparse !
- Use of TF-dictionary for sparsifying the signal
- Learn the dictionary using DDL
- Learn frequency patterns using TF and NMF