

AI4SIP

Matthieu Kowalski (Univ Paris-Saclay)

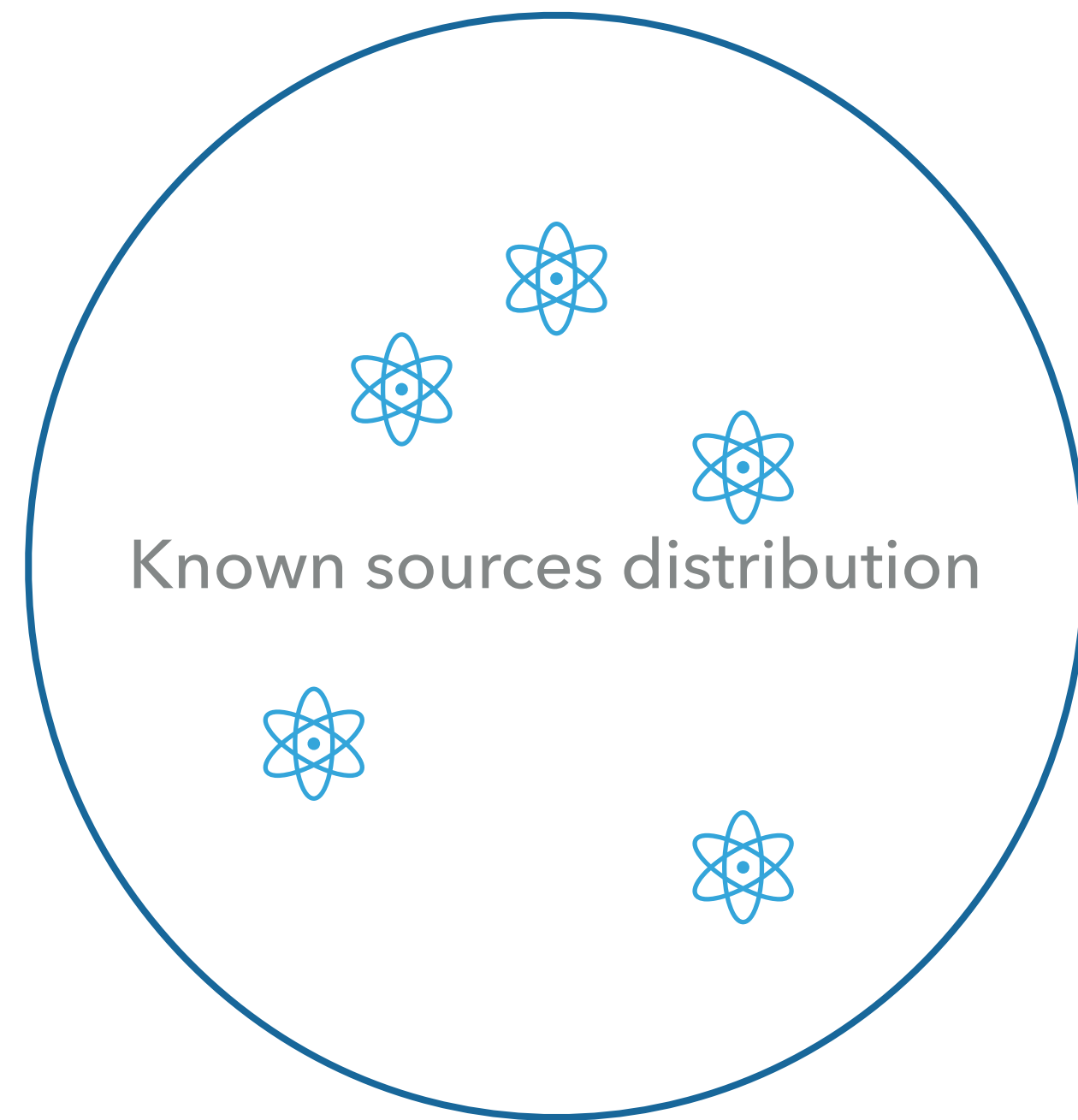
INVERSE PROBLEMS: FROM SPARSE TIME- FREQUENCY SYNTHESIS TO DICTIONARY LEARNING

OUTLINE

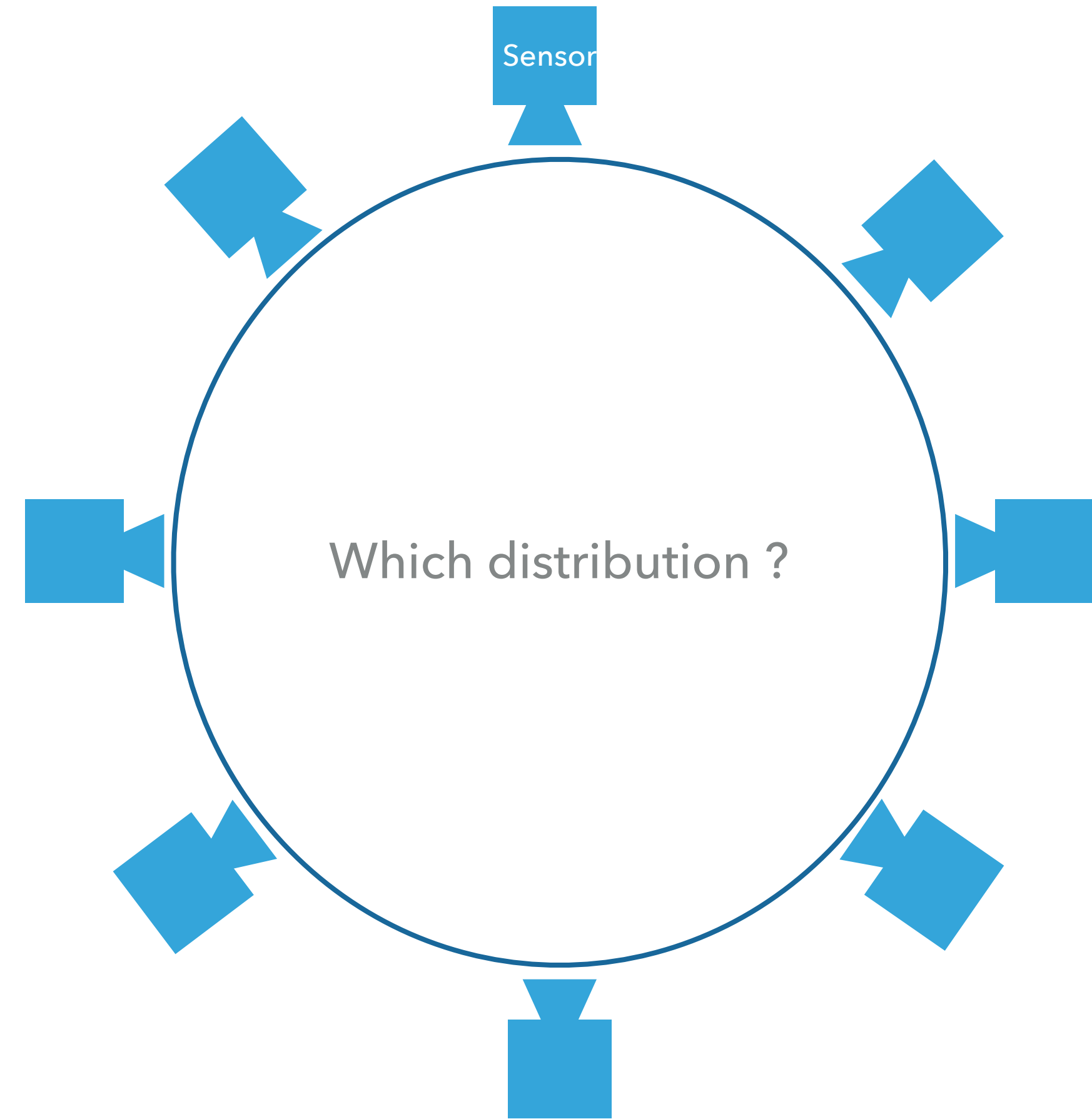
- ▶ Introduction – state of the art
 - ▶ Linear Inverse Problems
 - ▶ Time-Frequency dictionary
 - ▶ Sparse Coding and inverse problems
- ▶ Iterative Shrinkage/Thresholding and thresholding rules (illustration on audio declipping)
- ▶ Dictionary Learning
 - ▶ From alternate minimization to Deep Network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

$$y = Ax + e$$

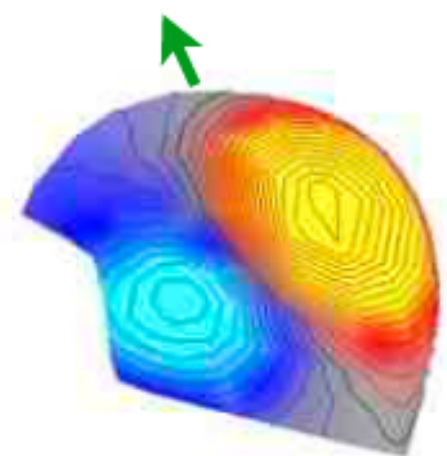
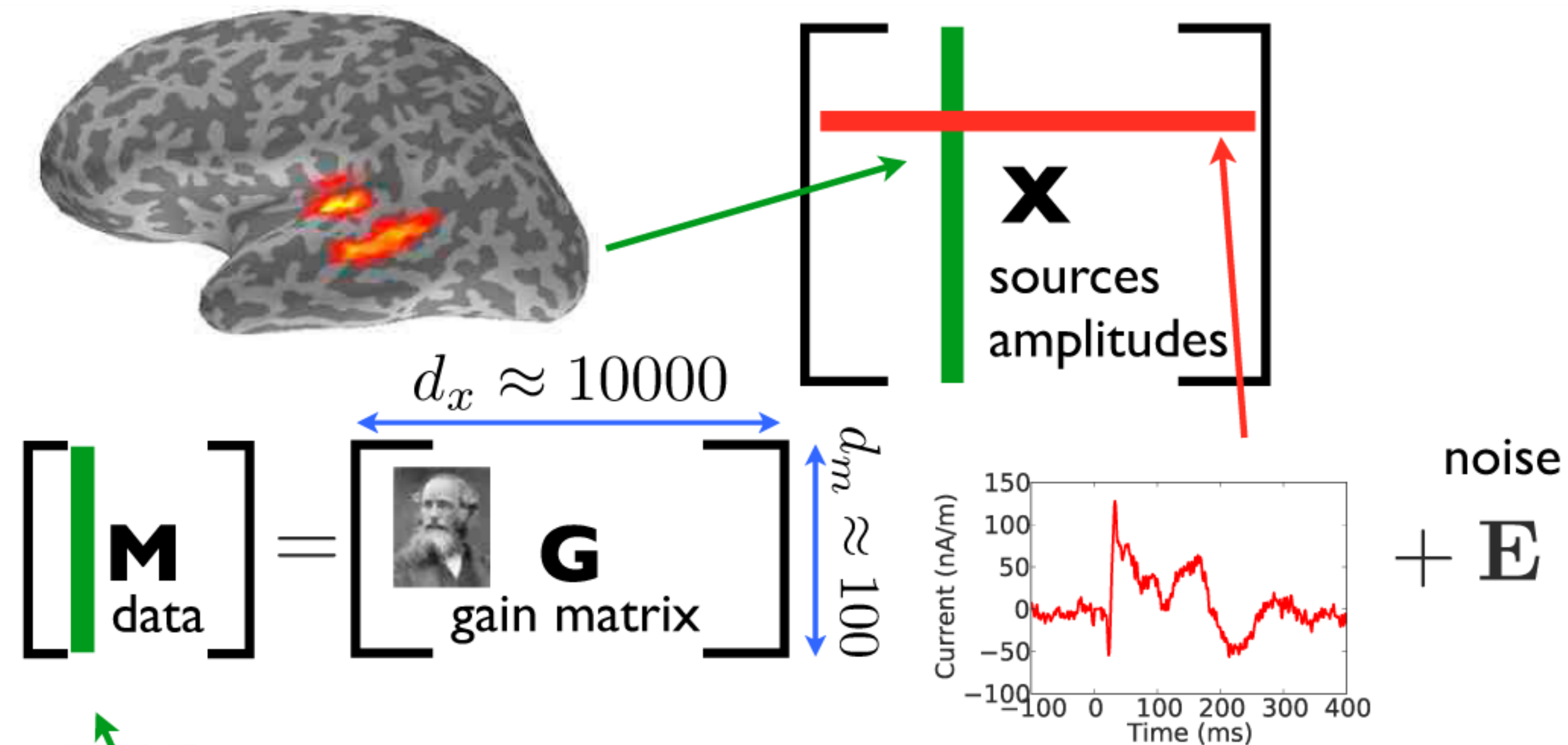
- ▶ Inversion by optimization:

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

With

- ▶ \mathcal{L} : the loss or "data fit" term or "prior" on the noise
- ▶ \mathcal{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):

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

Minimum Current Estimate (MCE – 1999):

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

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

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

INVERSE PROBLEM

$$x = \operatorname{argmin}_x \mathcal{L}(x) + \mathcal{R}(x)$$

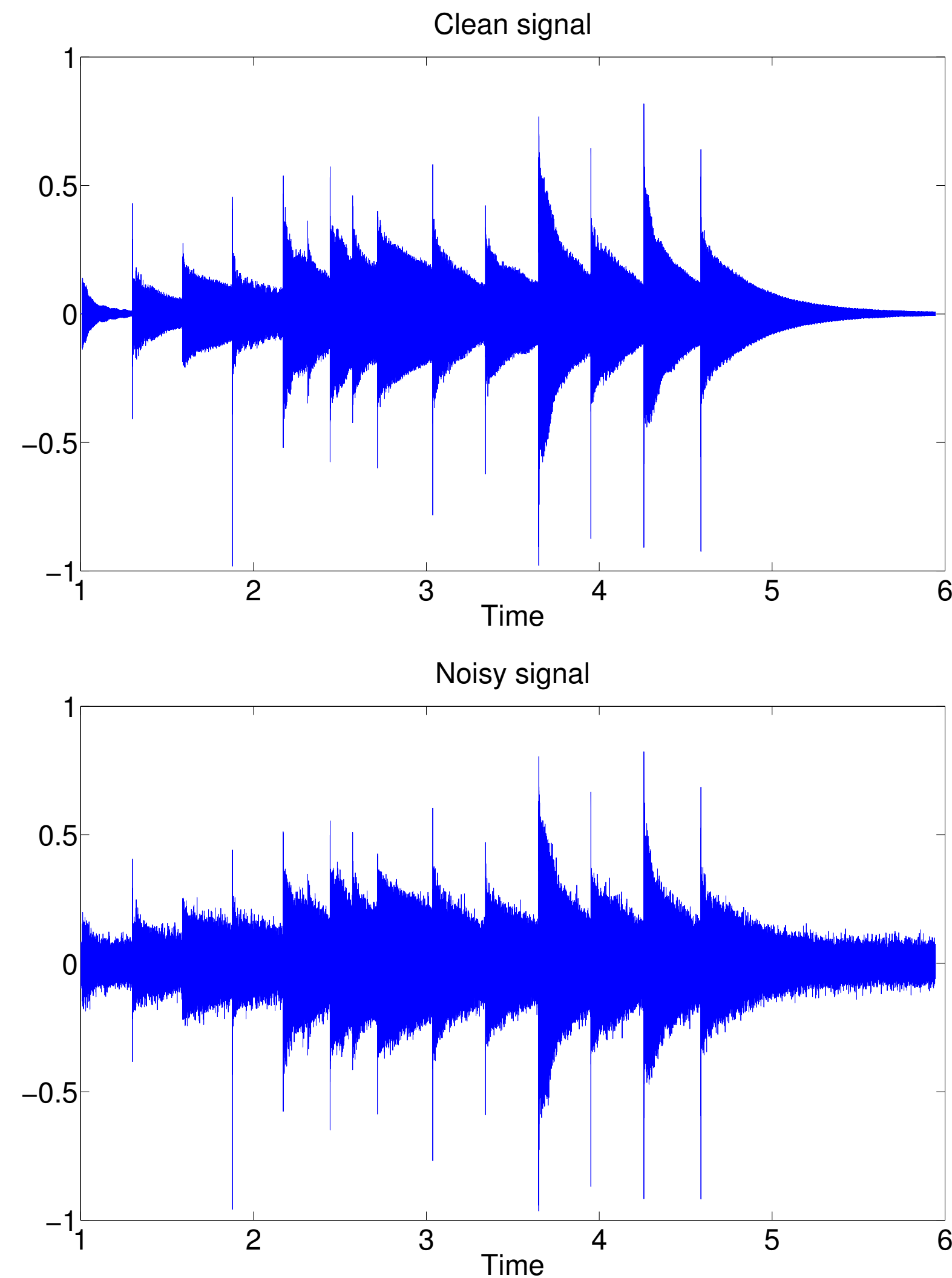
With

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

How to choose \mathcal{L} and \mathcal{R} ?

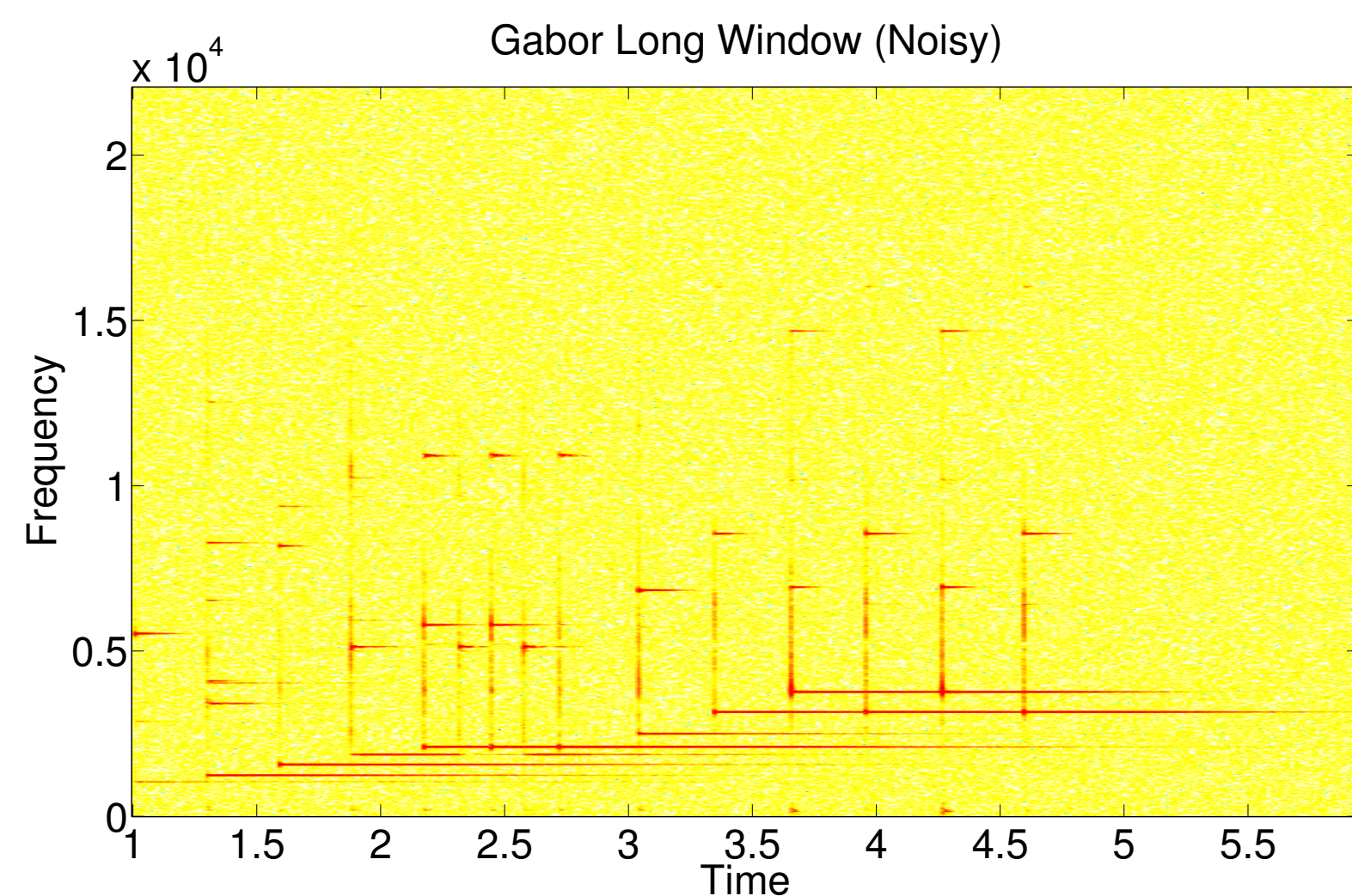
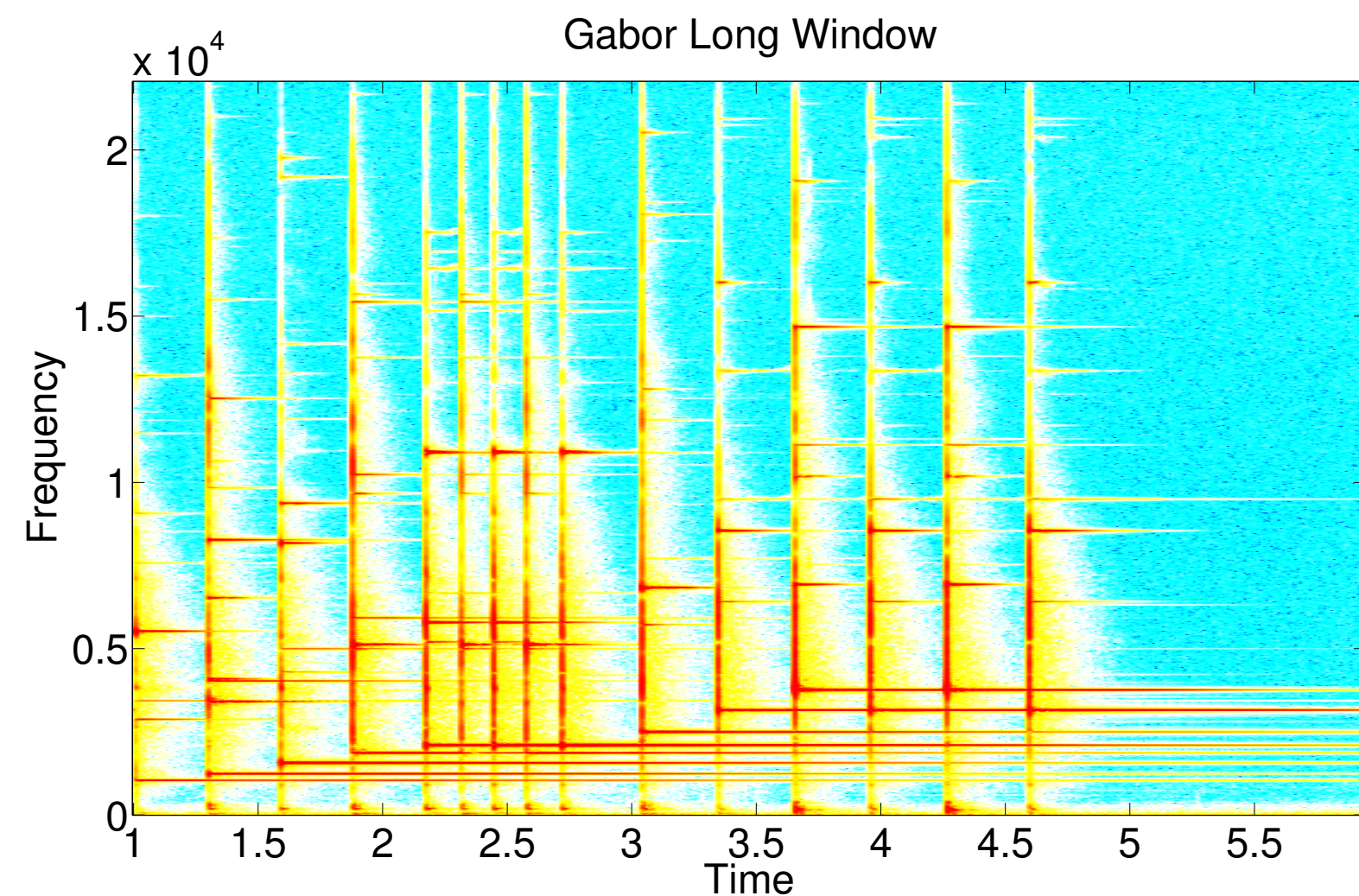
INTRODUCTION: TIME-FREQUENCY DICTIONARY

A TOY SIGNAL: THE GLOCKENSPIEL



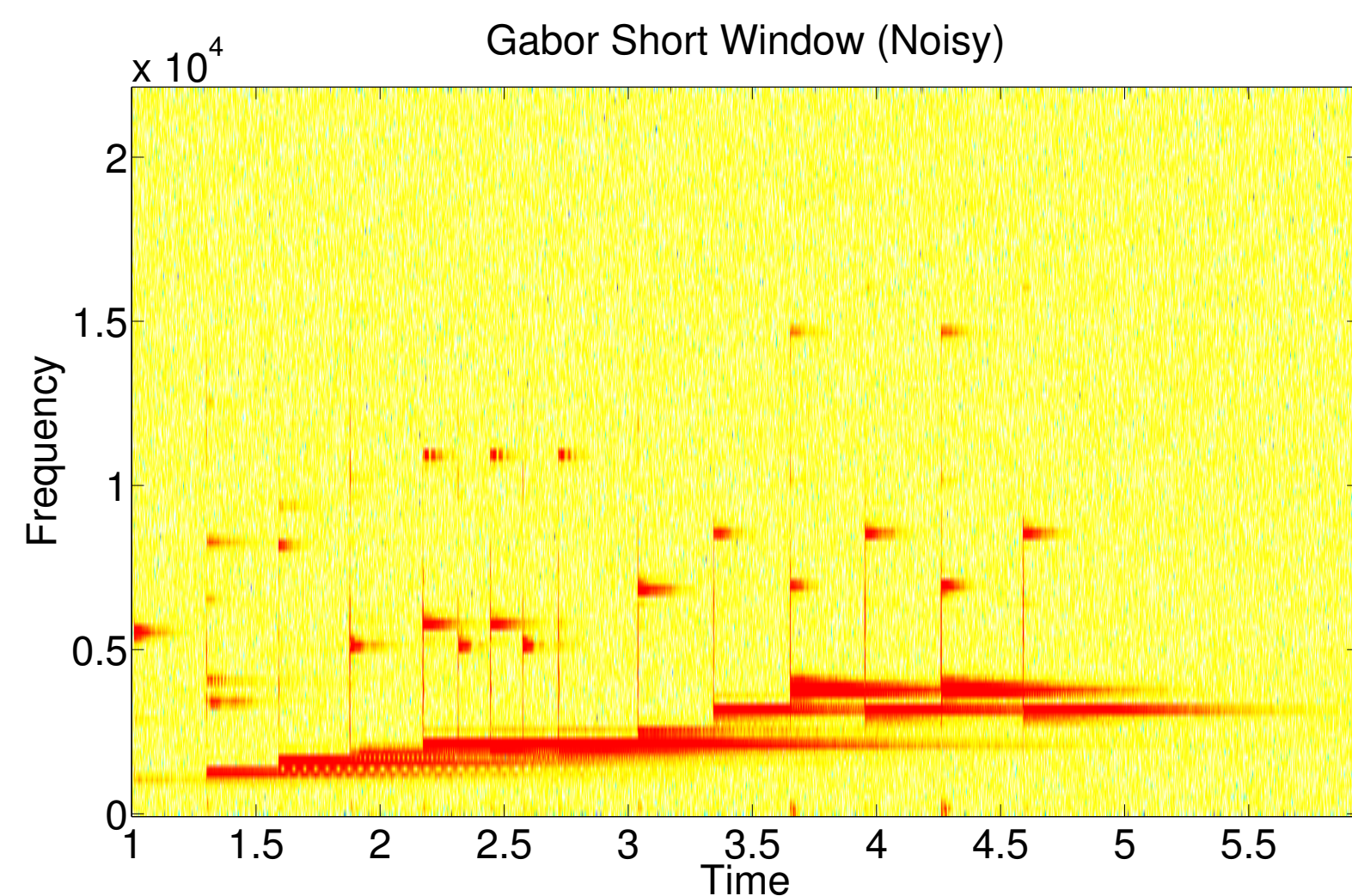
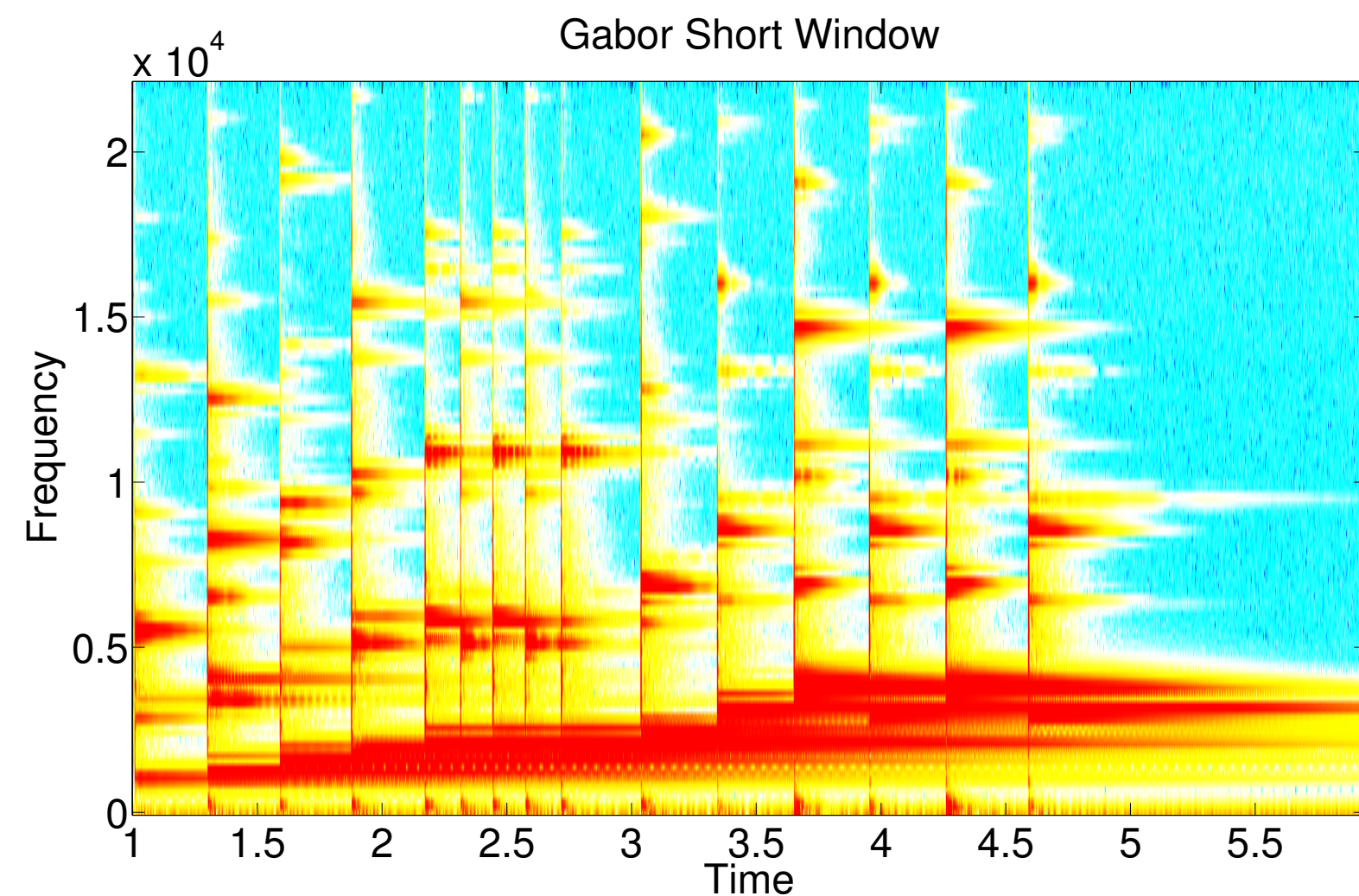
- ▶ well localized "percussive" transients
- ▶ clear harmonics/tonal layer
- ▶ Sampling rate: 44100Hz
- ▶ Length: 6s (2^{18} samples)
- ▶ White gaussian noise : input SNR = 10db

VARIOUS TIME-FREQUENCY REPRESENTATIONS -- LONG WINDOW



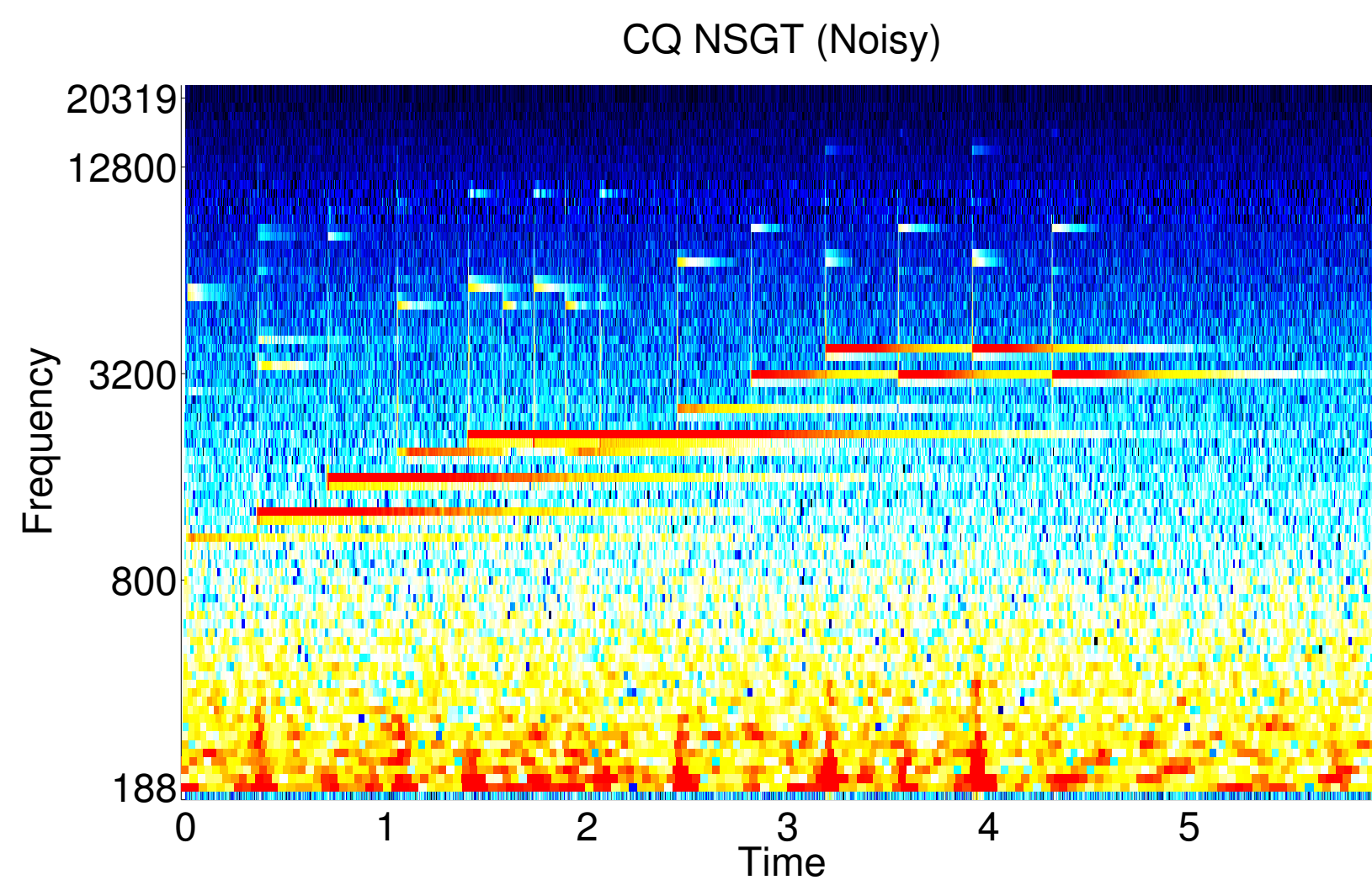
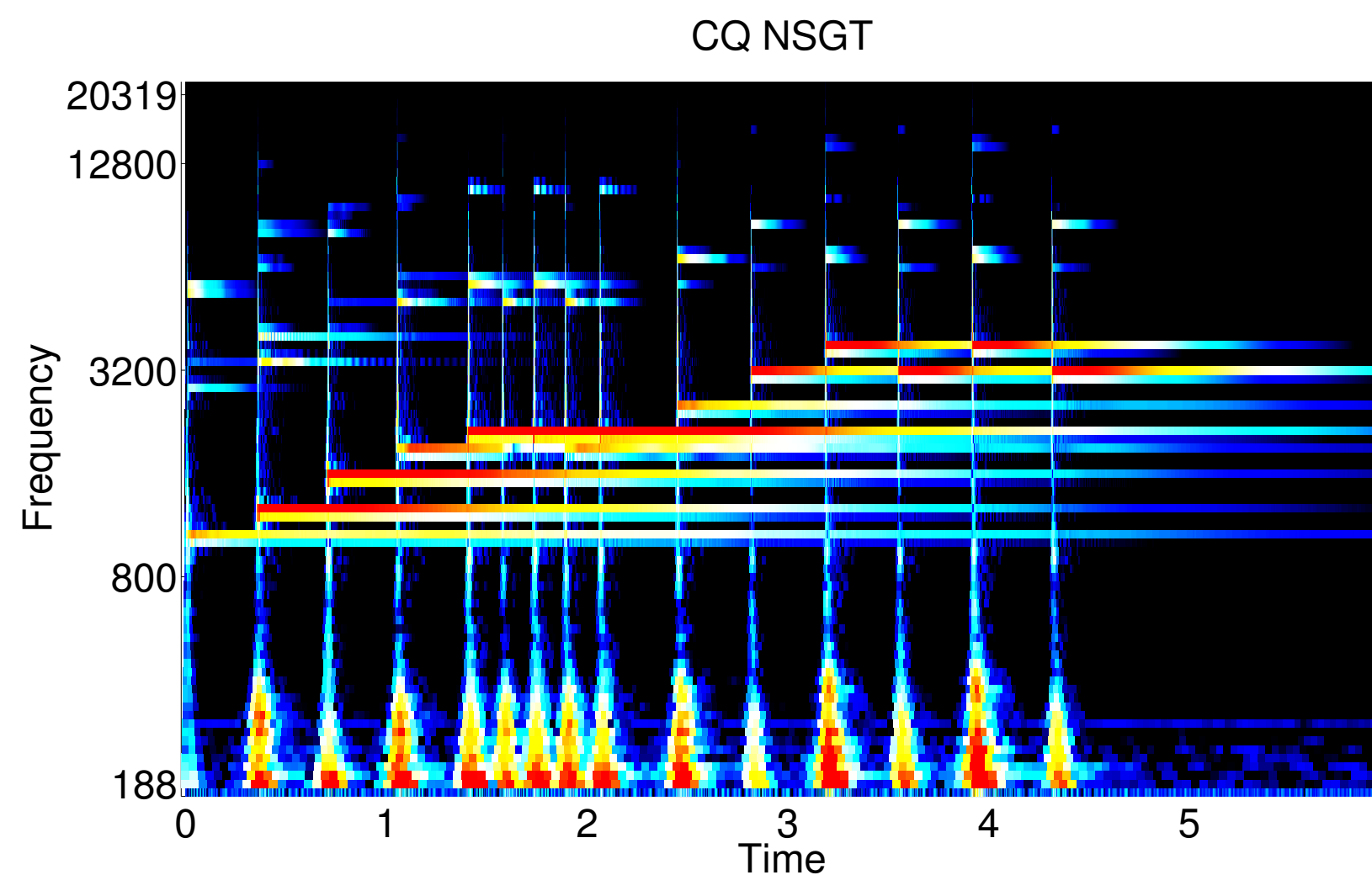
- ▶ Let $w[t] \in \mathbb{R}_+^L$ a smooth window
- ▶ Let $\varphi_{n,\nu}[t] = w[t - an]e^{-i2\pi\frac{\nu}{bL}t}$ a time-frequency atom
- ▶ $X[n, \nu] = \langle x[t], \varphi_{n,\nu}[t] \rangle$
- ▶ Hann window analysis
- ▶ Length: 48 ms (2048 samples)
- ▶ Hop size: 12ms (512 samples)
- ▶ Well adapted 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 \subset 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: $\hat{\alpha} = \operatorname{argmin}_{\alpha} \frac{1}{2} \|y - \Phi\alpha\|_2^2 + \lambda \|\alpha\|_0$
- ▶ 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]:

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

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:

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

ITERATIVE SHRINKAGE/THRESHOLDING ALGORITHM

▶ Lasso: $\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \frac{1}{2} \|y - A\Phi\alpha\|_2^2 + \lambda \|\alpha\|_1$

▶ 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:

▶ $\mathcal{S}_{\lambda}(z) = \underset{\alpha}{\operatorname{argmin}} \frac{1}{2} \|z - \alpha\|^2 + \lambda \|\alpha\|_1$ 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 \mathcal{R} be any measure of (structured) sparsity:

$$\hat{\alpha} = \operatorname{argmin}_{\alpha} \frac{1}{2} \|y - A\Phi\alpha\|_2^2 + \lambda\mathcal{R}(\alpha)$$

- ▶ ISTA:

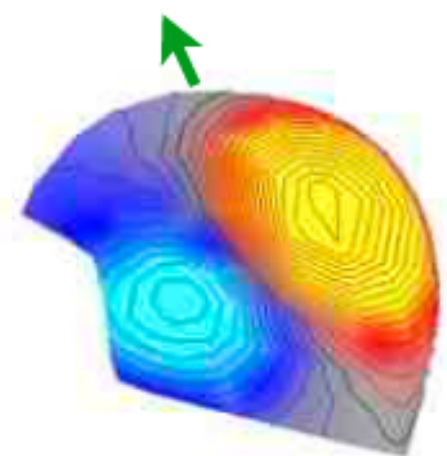
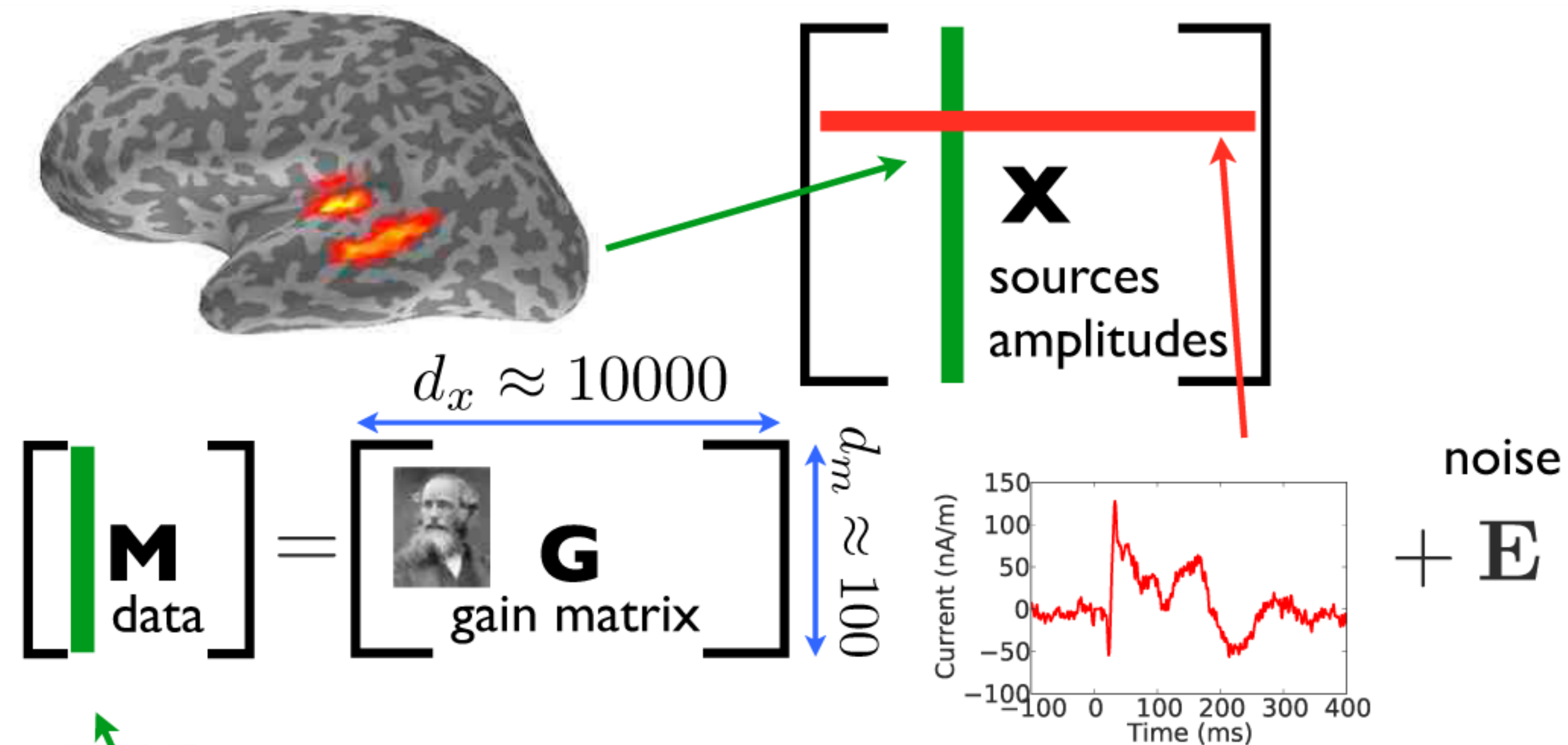
$$\alpha^{(t+1)} = \operatorname{prox}_{\frac{\lambda}{L}\mathcal{R}} \left(\alpha^{(t)} + \frac{1}{L}\Phi^*A^*(y - A\Phi\alpha^{(t)}) \right)$$

With:

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

EXAMPLES

M/EEG INVERSE PROBLEM



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

Minimum Norm Estimate (MNE – 1994):

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

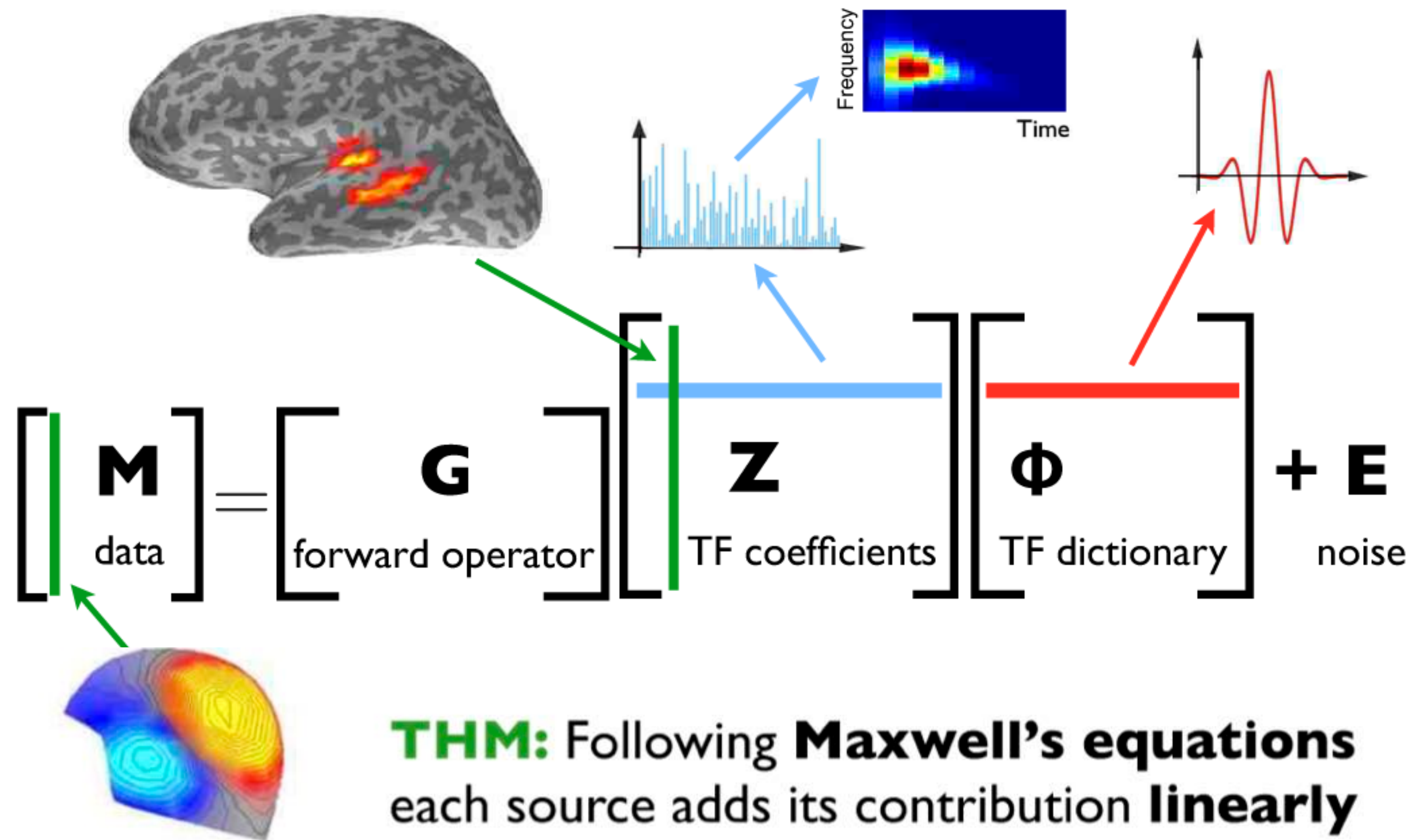
Minimum Current Estimate (MCE – 1999):

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

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

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

M/EEG INVERSE PROBLEM + TF SPARSE CODING

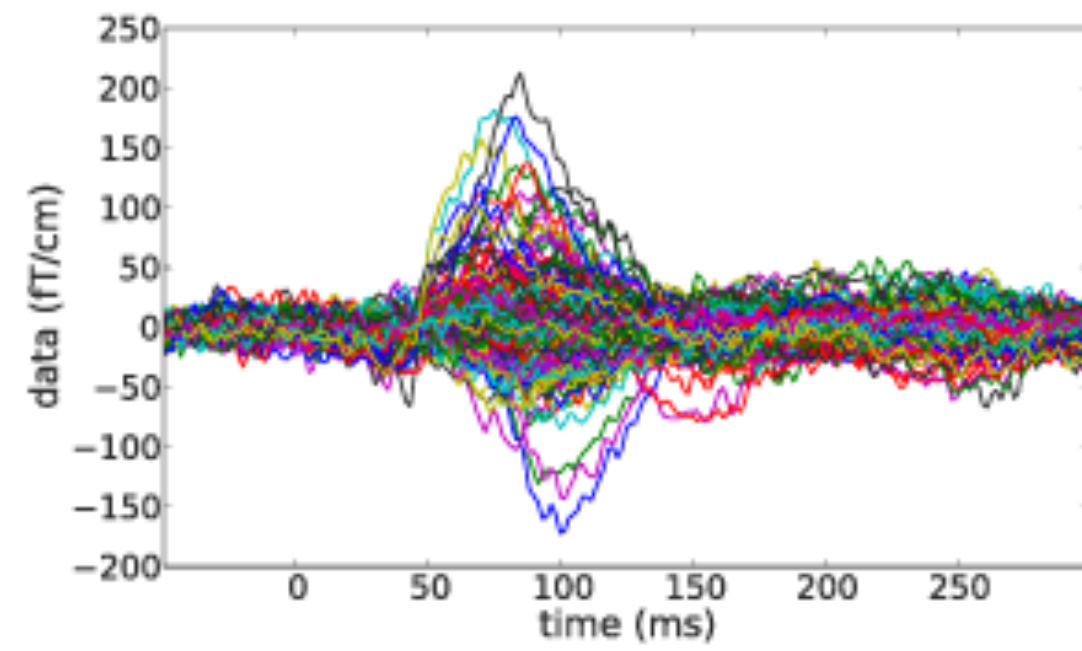


$$Z = \operatorname{argmin}_Z \frac{1}{2} \|M - GZ\Phi^*\|^2 + \mathcal{R}(Z)$$

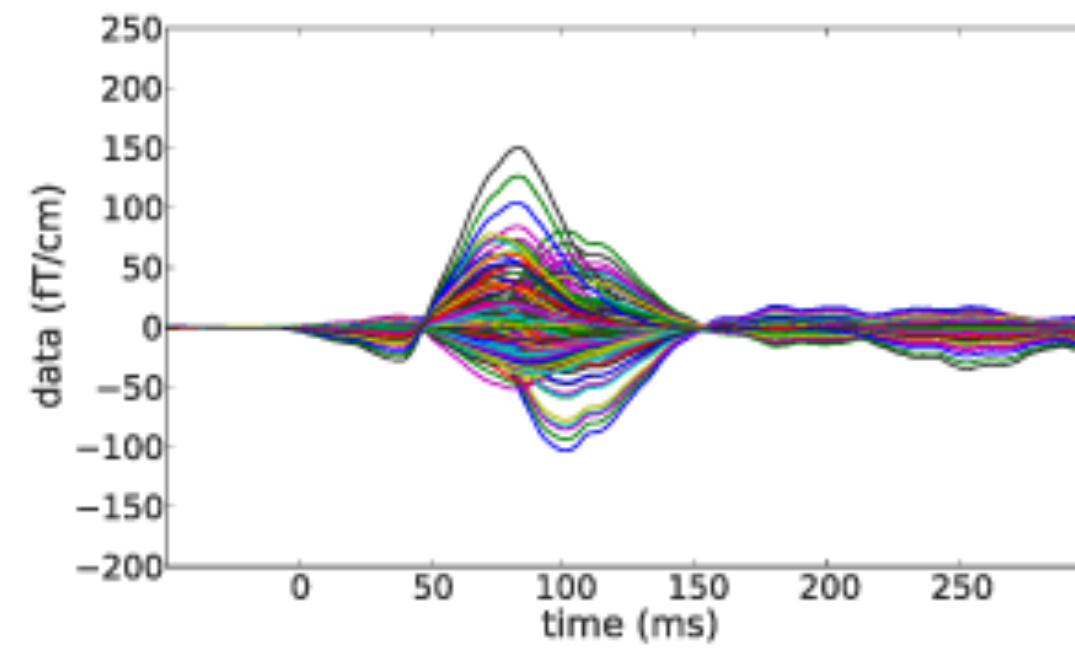
[MK, Gramfort & al 13]:

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

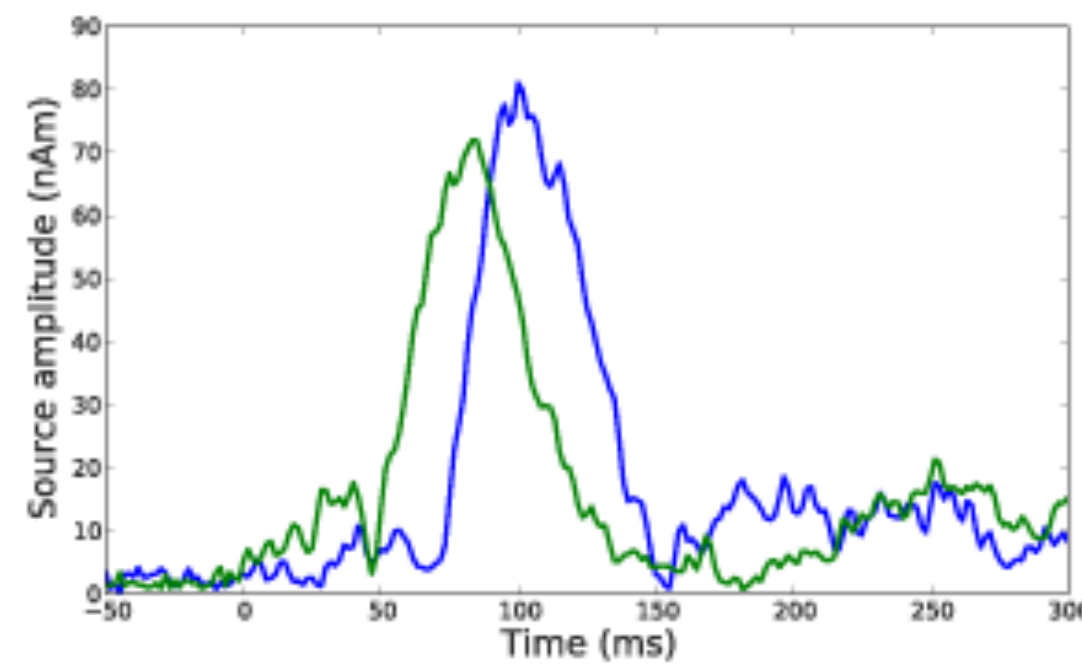
M/EEG INVERSE PROBLEM: AUDITORY DATA



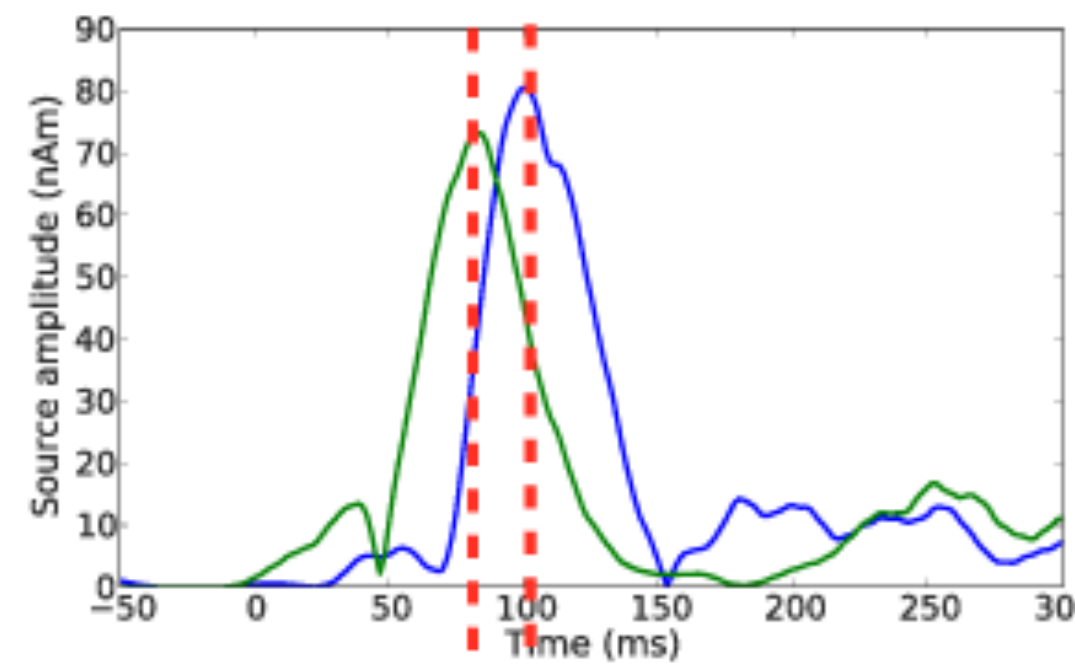
(a) MEG data (Gradiometers only)



(b) $GX_{TF-MxNE}^*$ (explained data)

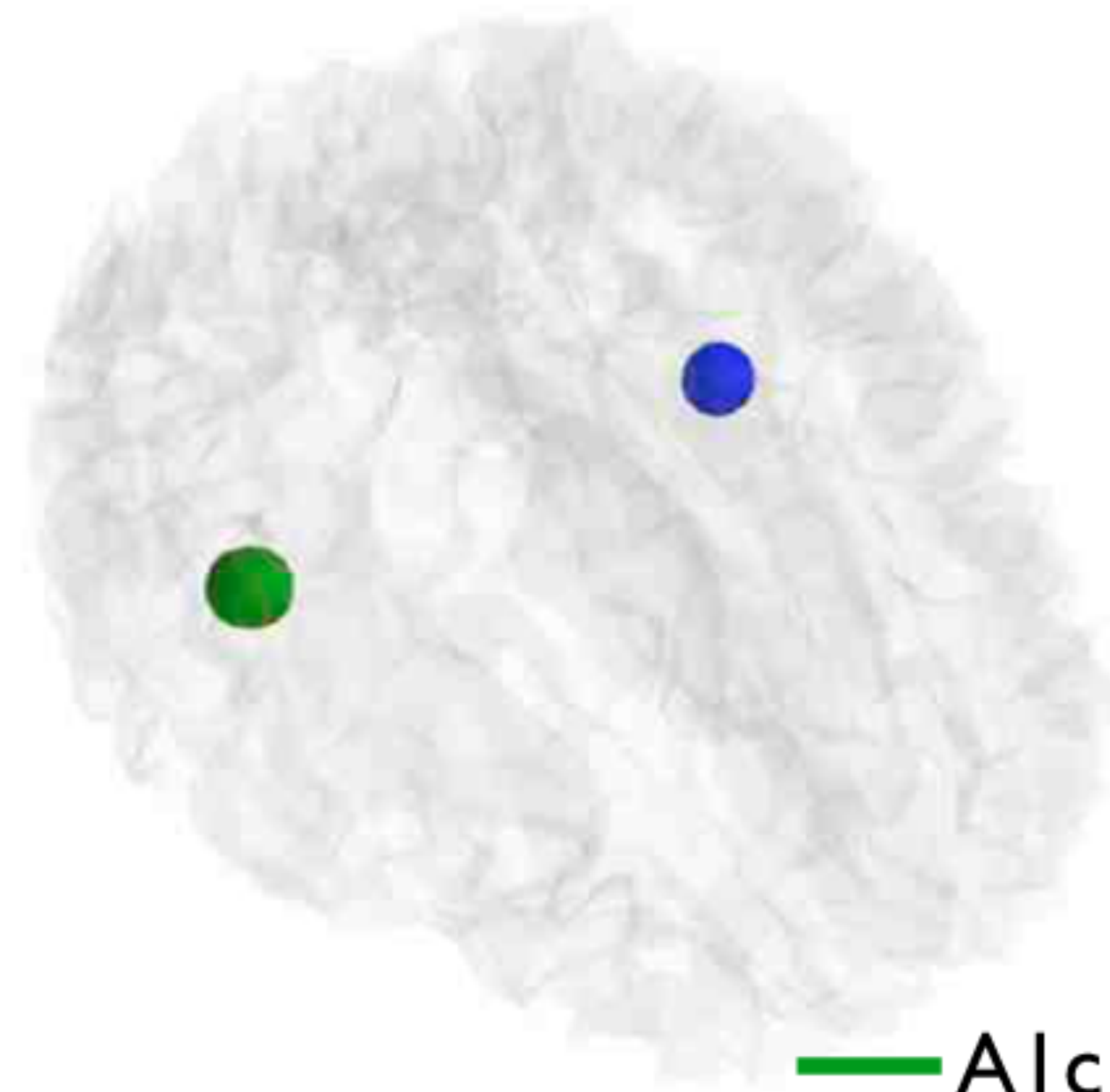


(c) X_{MxNE}^*



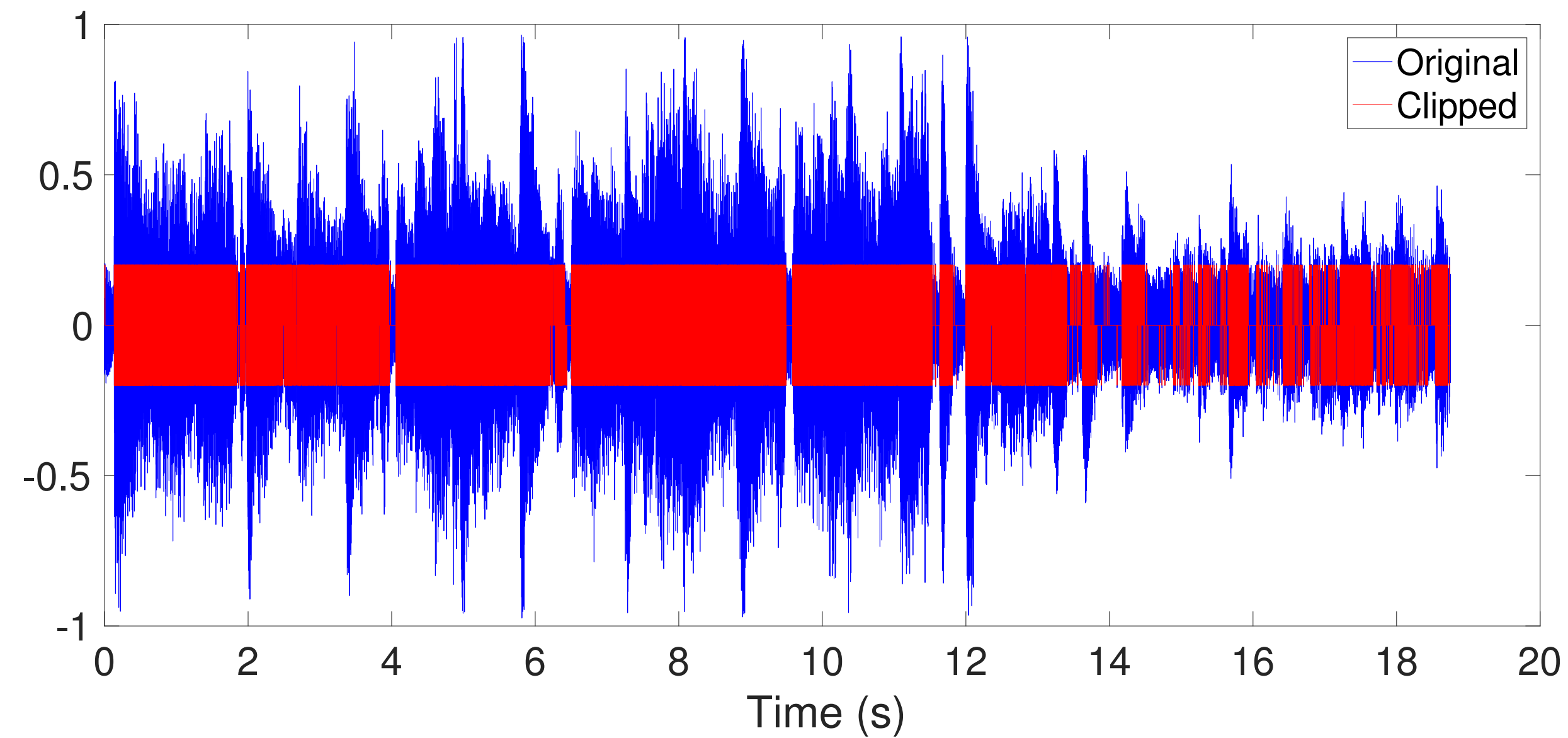
(d) $X_{TF-MxNE}^*$

16ms Chronometry



— Alc
— Ali

AUDIO DECLIPPING

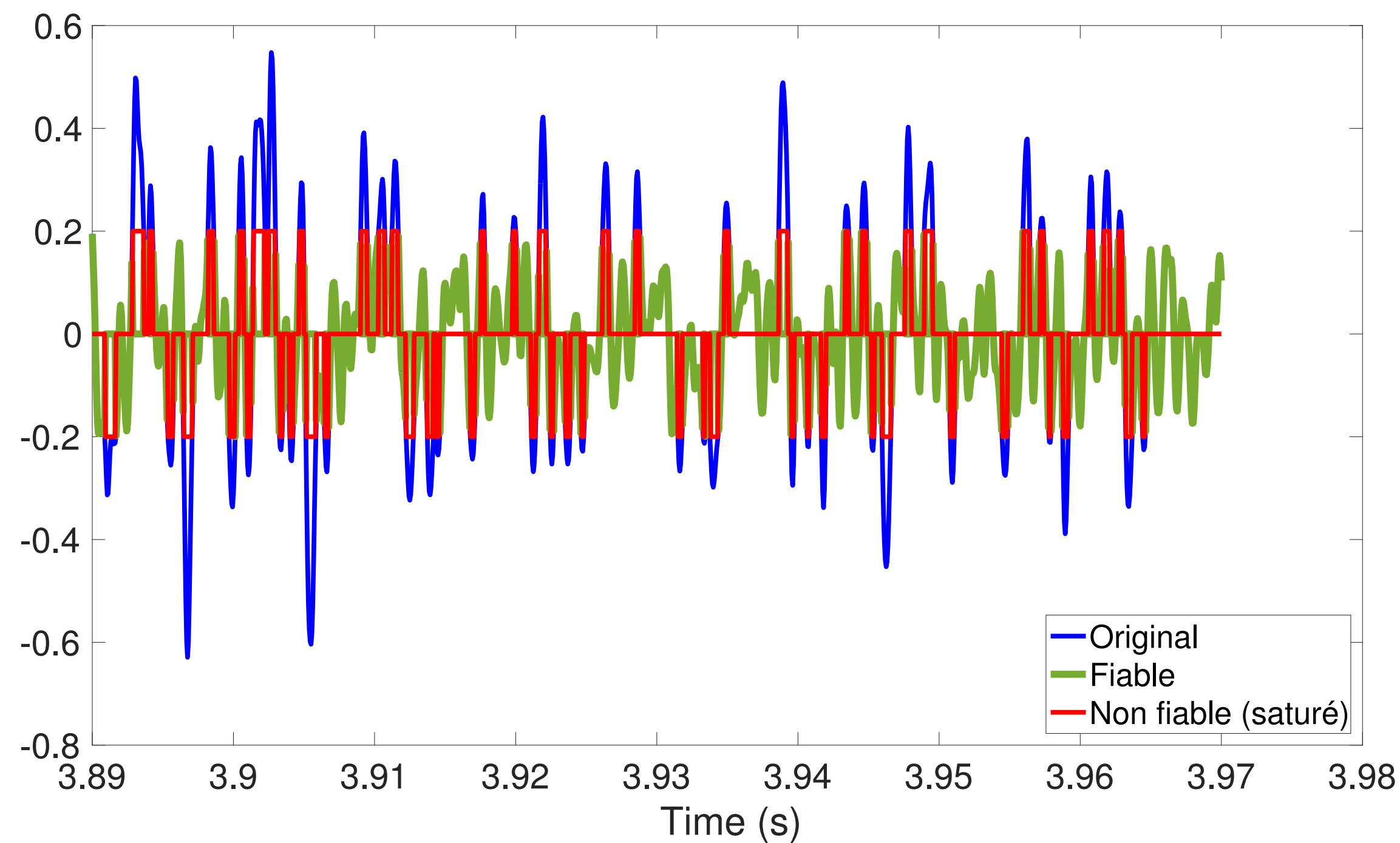


- ▶ How to recover the original (blue) signal from the clipped (red) signal ?

(SNR = 4dB)

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

AUDIO DECLIPPING



▶ Reliable samples (green): $y^r = M^r x = M^r \Phi \alpha$

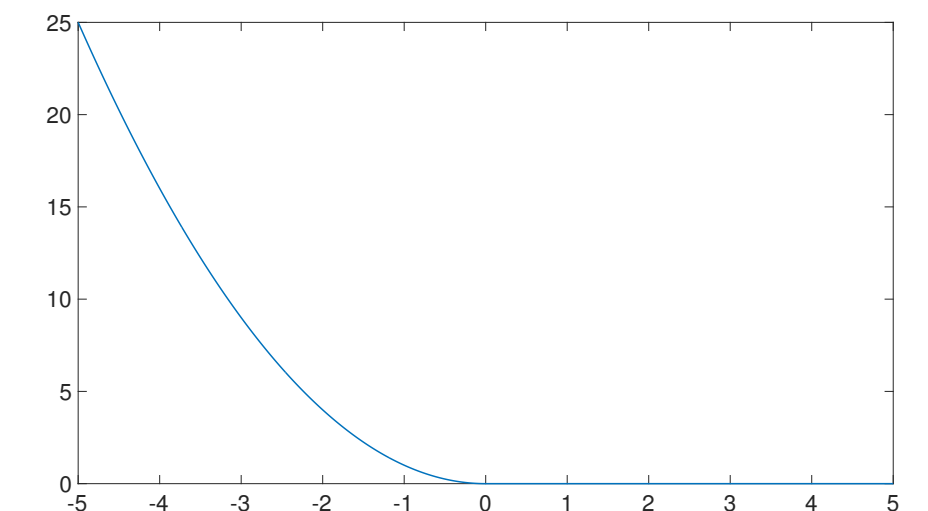
▶ Loss on reliable samples (green):

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

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

▶ Loss on clipped samples (red):

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



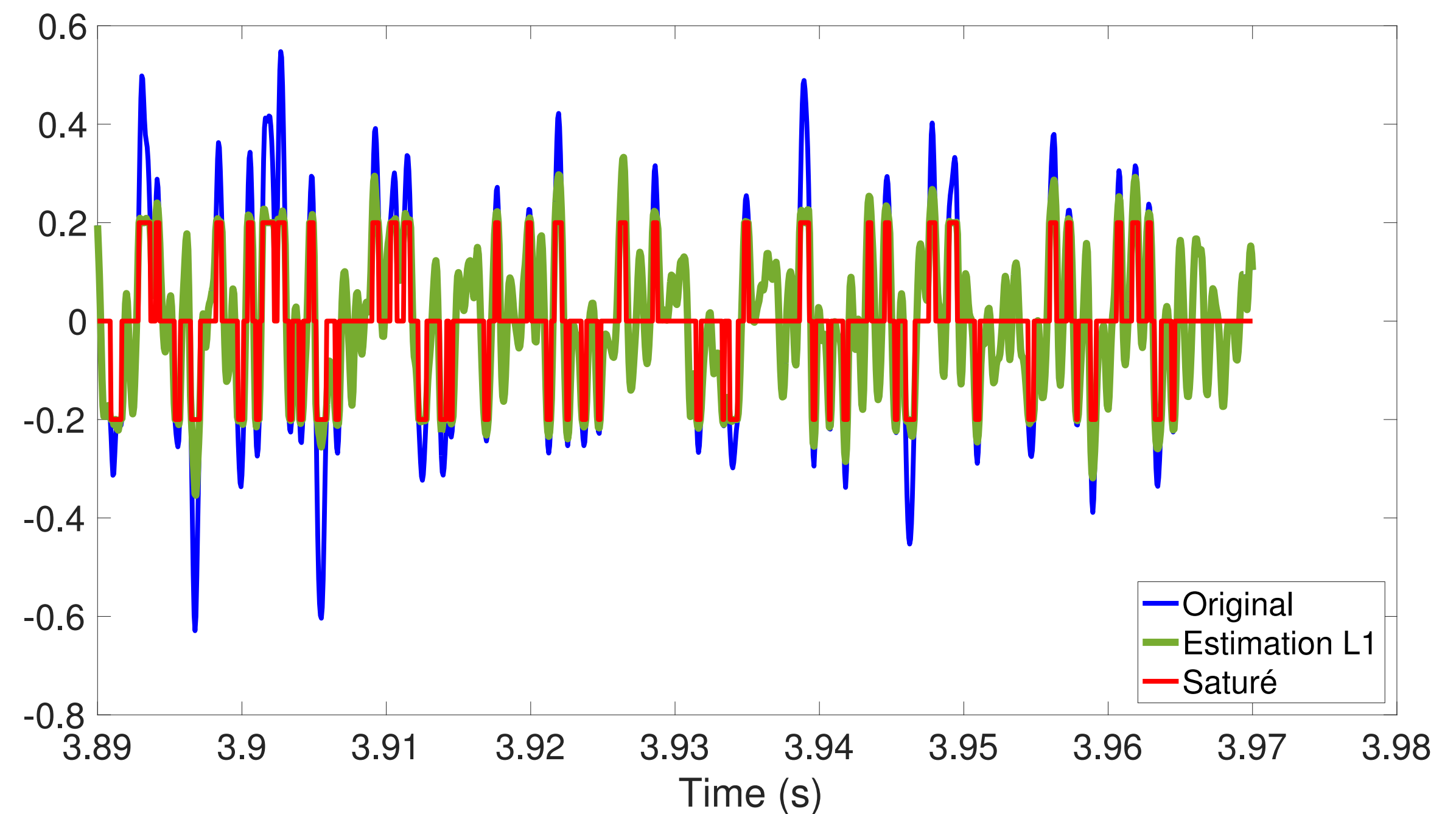
AUDIO DECLIPPING

- ▶ Declipping :

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

- ▶ Algorithm: ISTA
- ▶ Results: SNR = 4.5 dB

How to improve the estimation ?



INVERSE PROBLEM WITH SPARSE CODING

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

- ▶ 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} ?

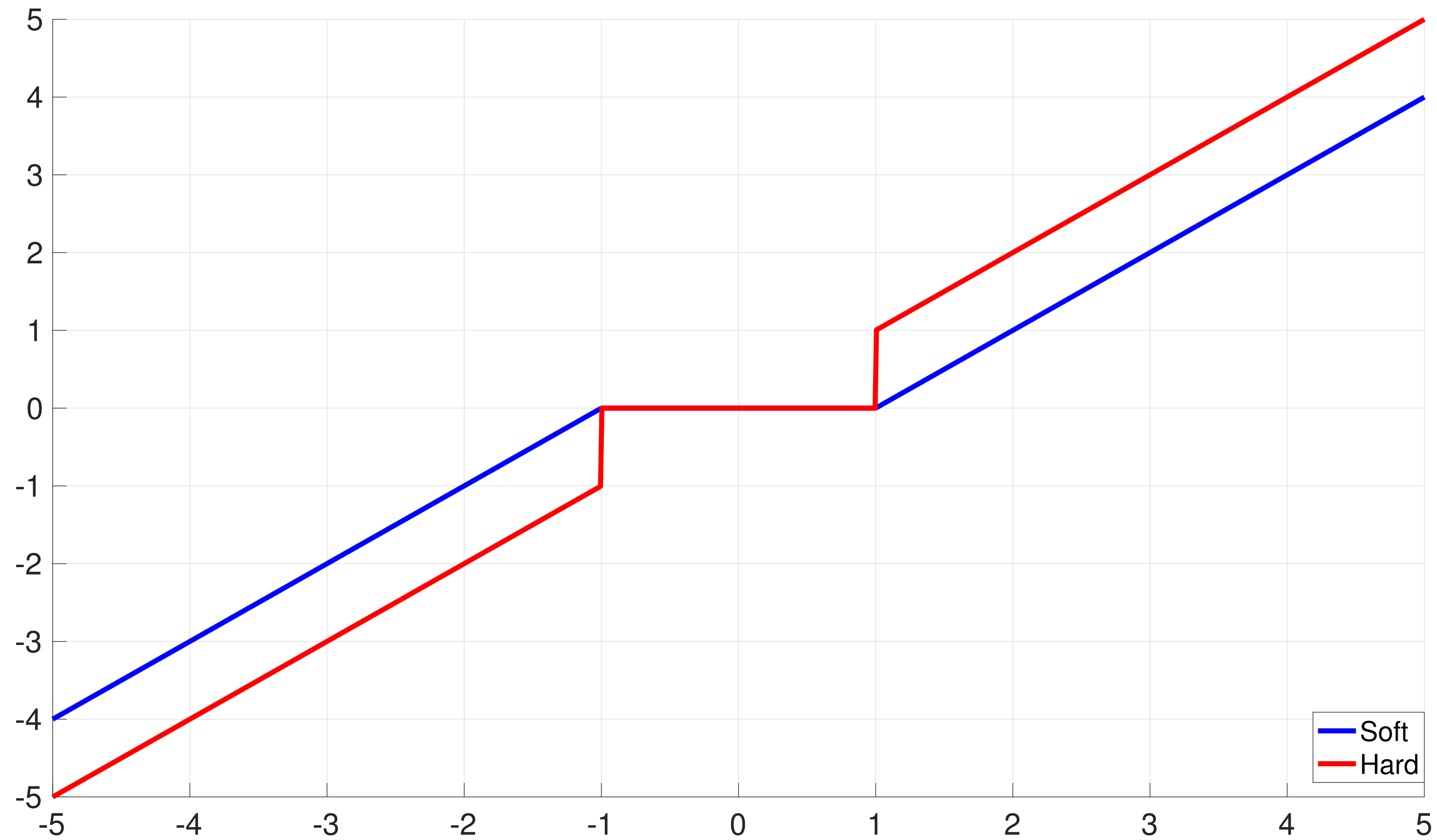
ISTA AND SHRINKAGE RULES

SHRINKAGE RULES

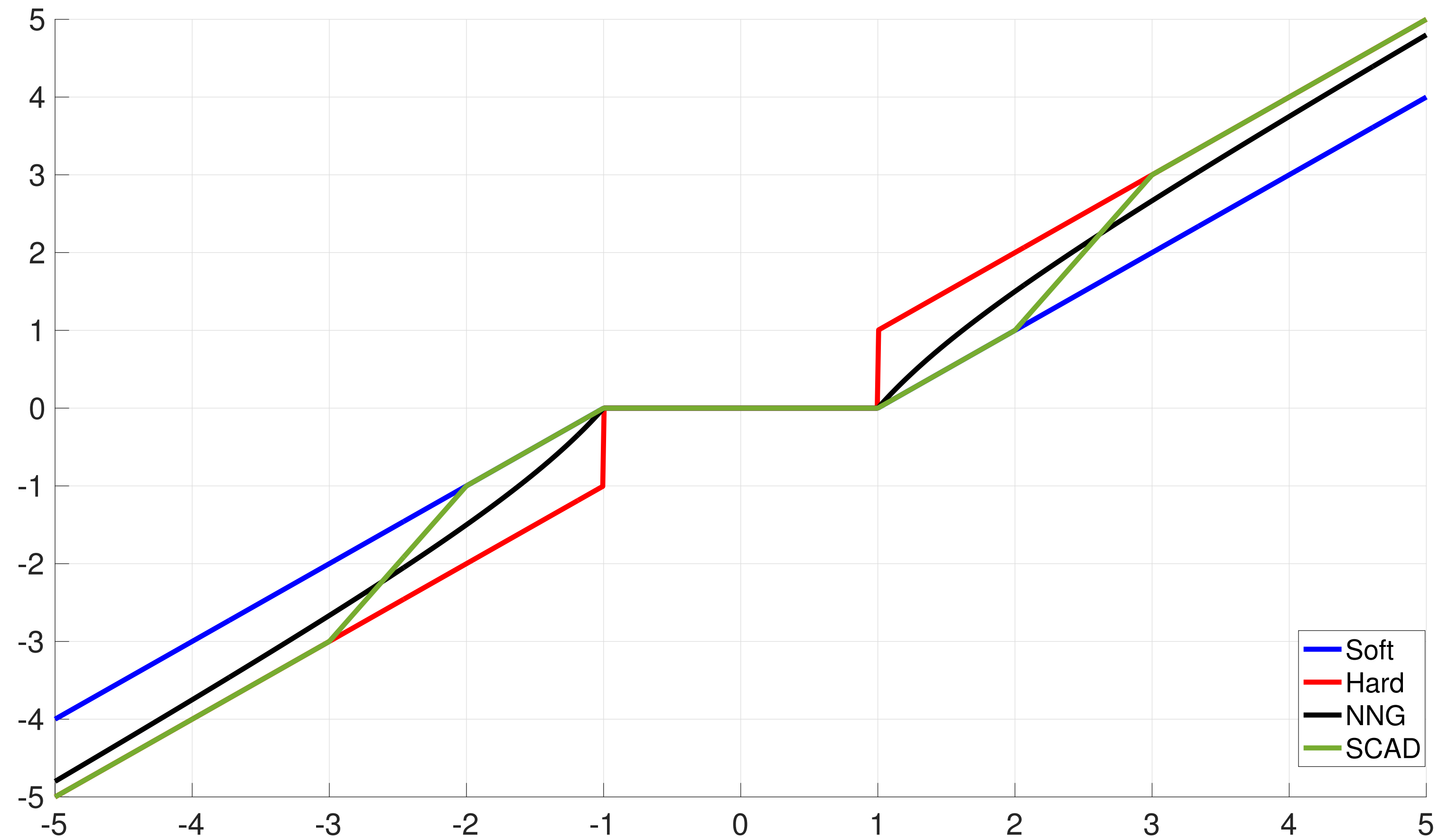
Definition [Antoniadis 07]

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

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

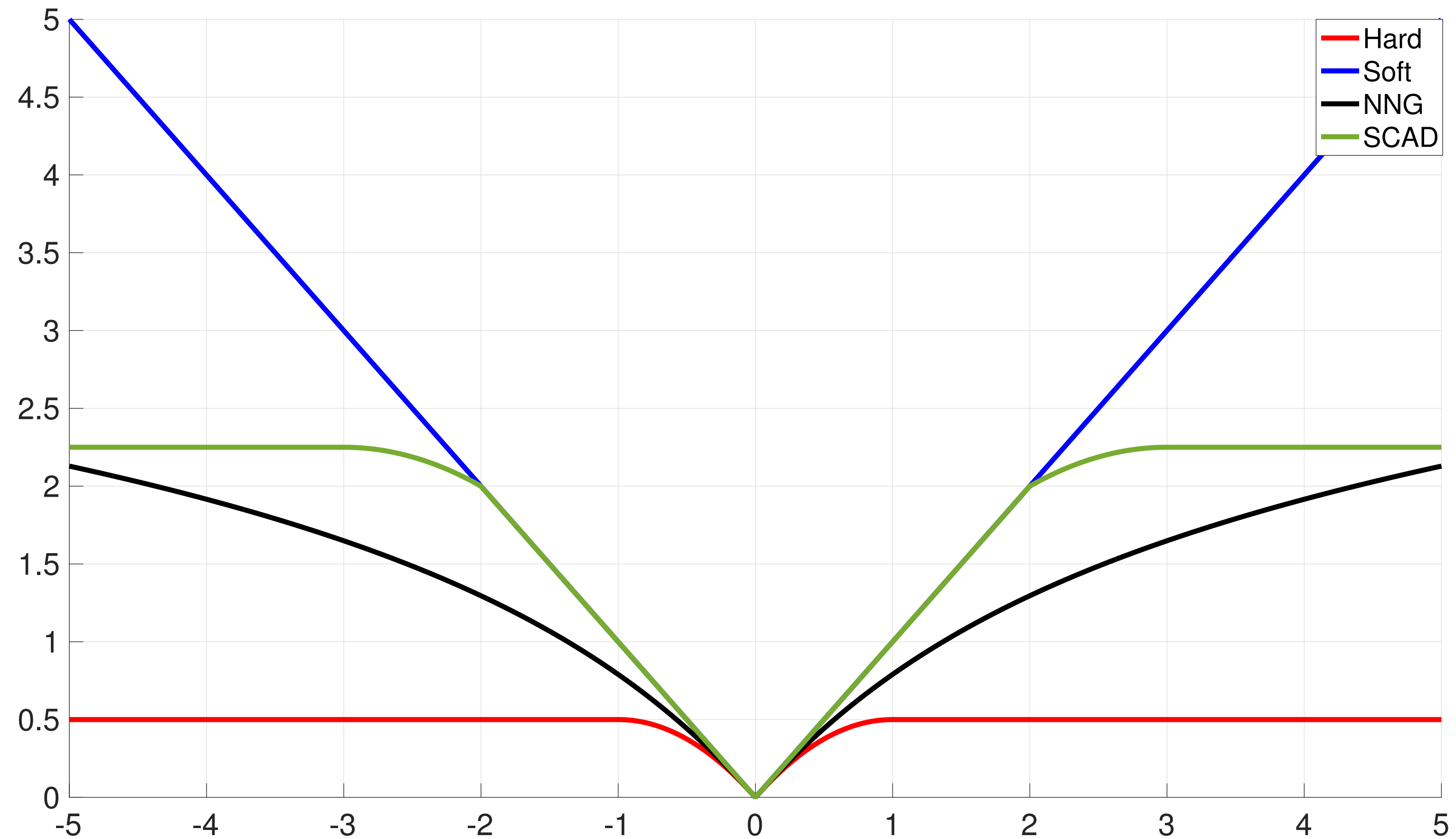
▶ Theorem 1 [MK 14]

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

▶ Theorem 2 [MK 14]

(Relaxed) ISTA converges with any thresholding rules

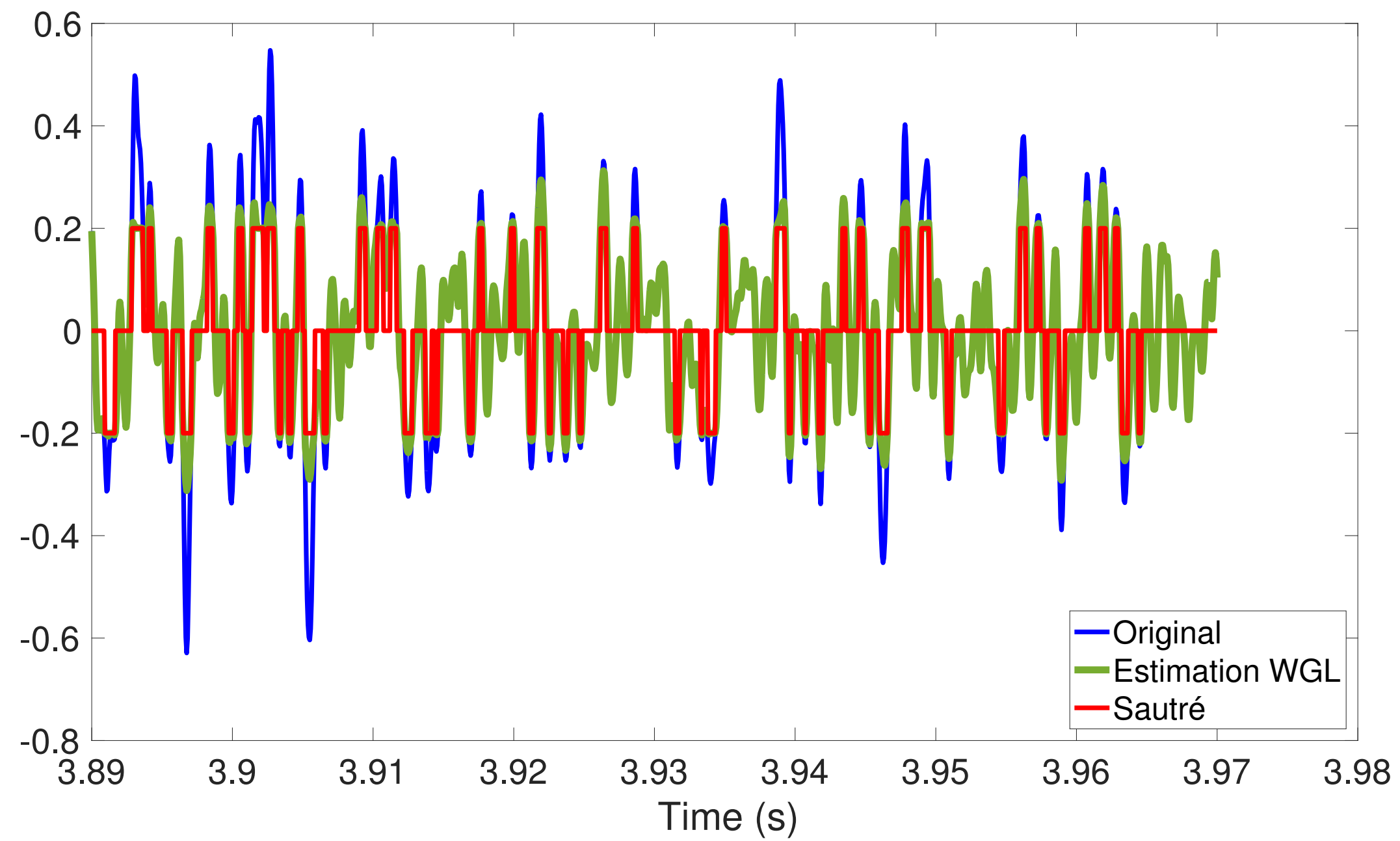
SHRINKAGE RULES EXEMPLES



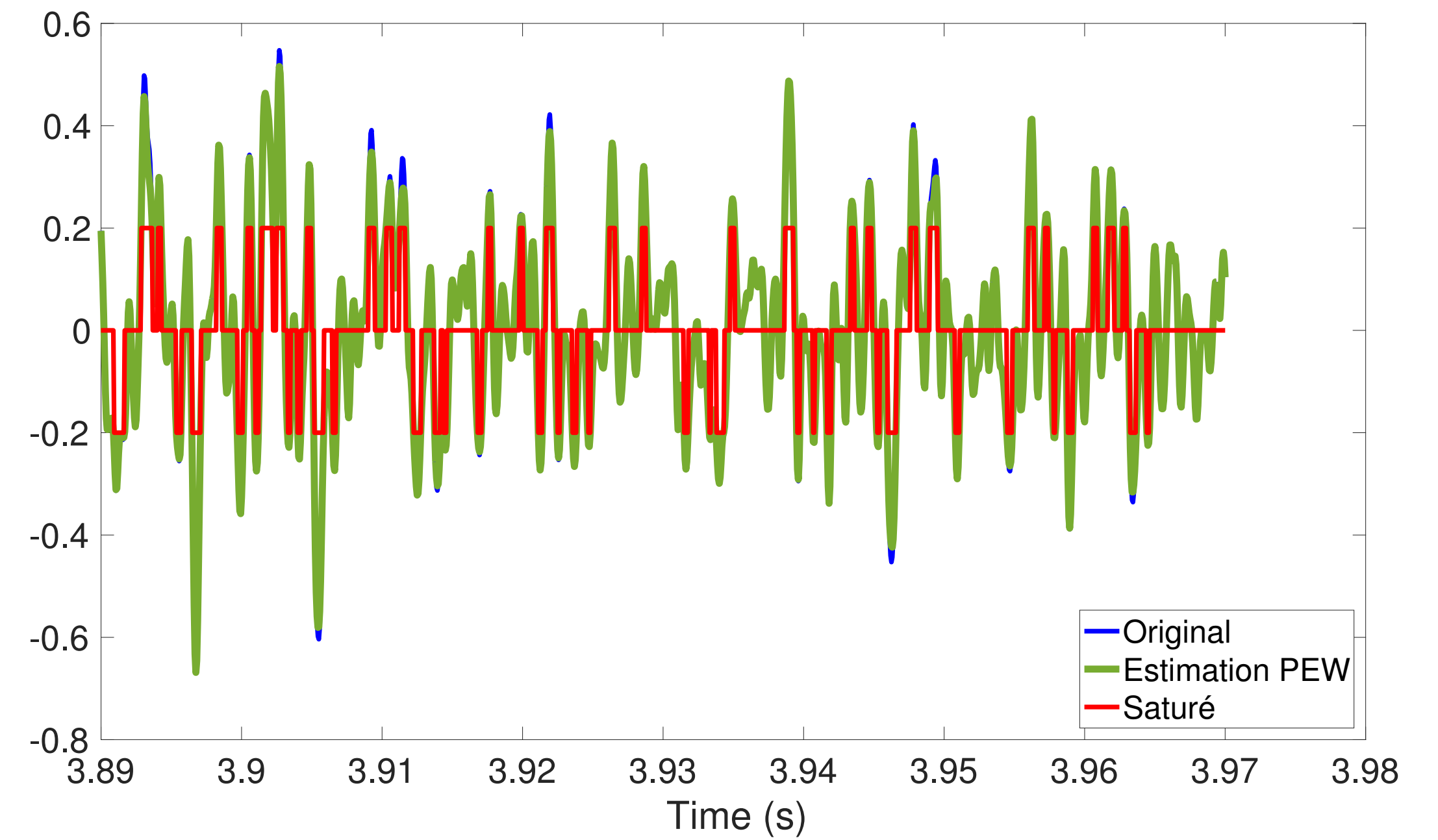
EXTENSION: STRUCTURED/SOCIAL SPARSITY [MK & AL 13]

- ▶ 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 $X = (x_1, x_2, \dots, x_M)$ 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:

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

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

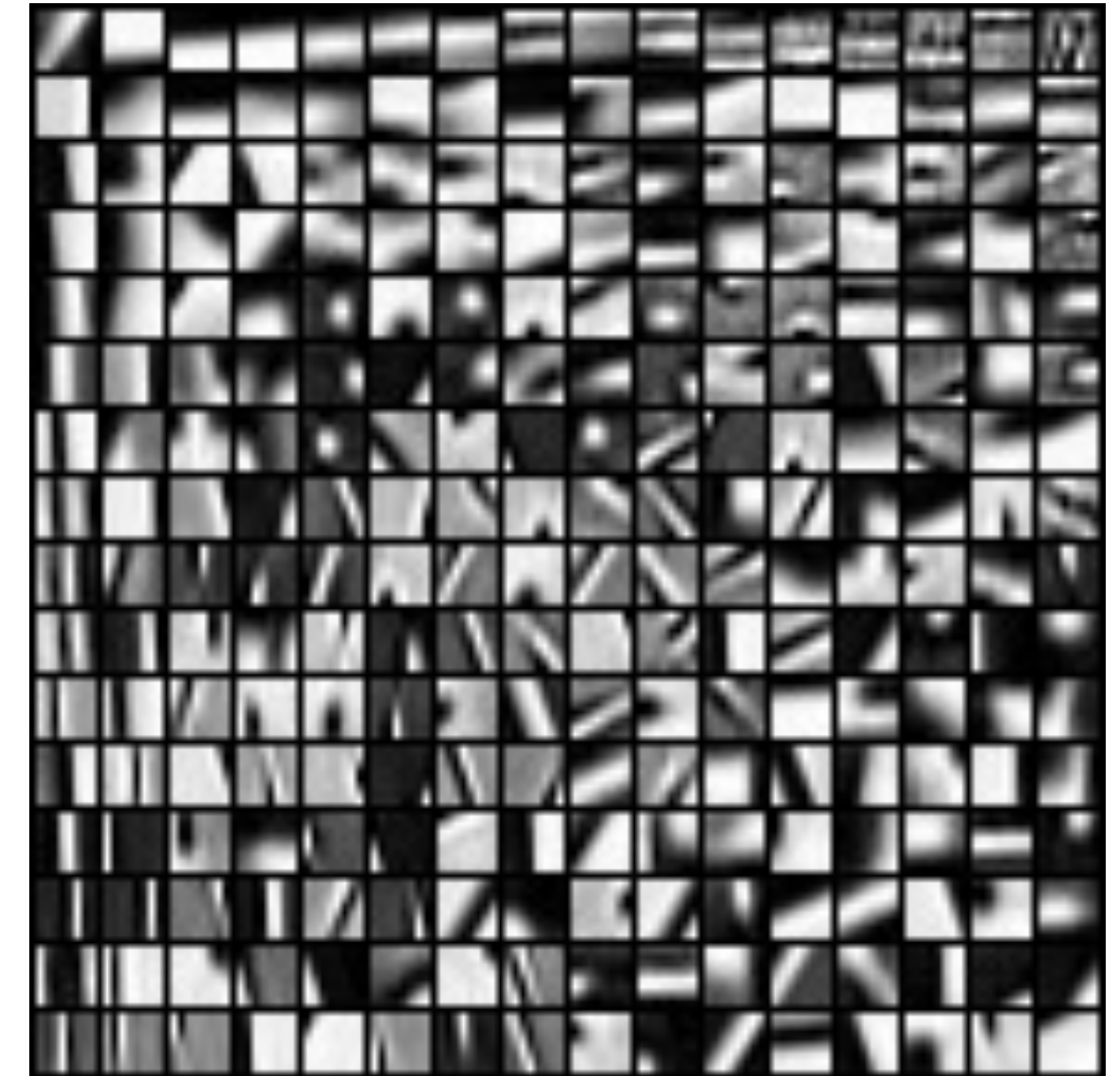
$$\min_{\Phi, A} \|X - \Phi A\|^2 \quad s.t. \quad \|\alpha_m\|_0 \leq s \quad \forall m$$

DICTIONARY LEARNING: ALTERNATING MINIMIZATION

- ▶ Idea: alternate minimization between A and Φ
- ▶ Initialize Φ
- ▶ Sparse coding for each x_m : $\alpha_m = \operatorname{argmin}_{\alpha} \frac{1}{2} \|x_m - \Phi\alpha\|^2 + \lambda \|\alpha\|_1$
- ▶ 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]

THE SALINAS VALLEY is in Northern California. It is a long narrow swale between two ranges of mountains, and the Salinas River winds and twists up the center until it falls at last into Monterey Bay.

I remember my childhood names for grasses and secret flowers. I remember where a toad may live and what time the birds awaken in the summer-and what trees and seasons smelled like-how people looked and walked and smelled even. The memory of odors is very rich.

I remember that the Gabilan Mountains to the east of the valley were light gay mountains full of sun and loveliness and a kind of invitation, so that you wanted to climb into their warm foothills almost as you want to climb into the lap of a beloved mother. They were beckoning mountains with a brown grass love. The Santa Lucias stood up against the sky to the west and kept the valley from the open sea, and they were dark and brooding-unfriendly and dangerous. I always found in myself a dread of west and a love of east. Where I ever got such an idea I cannot say, unless it could be that the morning came over the peaks of the Gabilans and the night drifted back from the ridges of the Santa Lucias. It may be that the birth and death of the day had some part in my feeling about the two ranges of mountains.

From both sides of the valley little streams slipped out of the hill canyons and fell into the bed of the Salinas River. In the winter of wet years the streams ran full-freshet, and they swelled the river until sometimes it raged and boiled, bank full, and then it was a destroyer. The river tore the edges of the farm lands and washed whole acres down; it toppled barns and houses into itself, to go floating and bobbing away. It trapped cows and pigs and sheep and drowned them in its muddy brown water and carried them to the sea. Then when the late spring came, the river drew in from its edges and the sand banks appeared. And in the summer the river didn't run at all above ground. Some pools would be left in the deep swirl places under a high bank. The tules and grasses grew back, and willows straightened up with the flood debris in their upper branches. The Salinas was only a part-time river. The summer sun drove it underground. It was not a fine river at all, but it was the only one we had and so we boasted about it-how dangerous it was in a wet winter and how dry it was in a dry summer. You can boast about anything if it's all you have. Maybe the less you have, the more you are required to boast.

The floor of the Salinas Valley, between the ranges and below the foothills, is level because this valley used to be the bottom of a hundred-mile inlet from the sea. The river mouth at Moss Landing was centuries ago the entrance to this long inland water. Once, fifty miles down the valley, my father bored a well. The drill came up first with topsoil and then with gravel and then with white sea sand full of shells and even pi...



DICTIONARY LEARNING AS BI-LEVEL OPTIMIZATION

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

Dictionary learning as bi-level optimization:

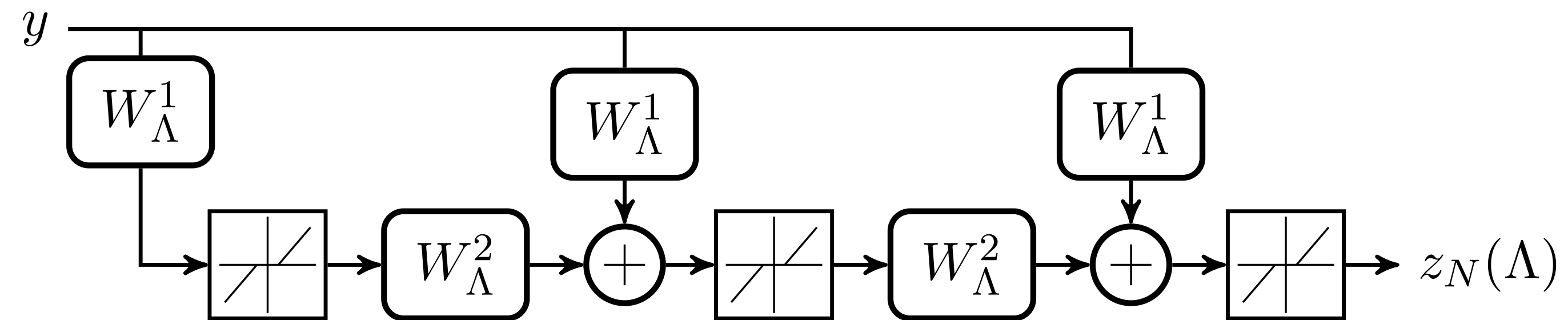
$$\min_{\Phi \in \mathcal{C}} F(\alpha^*(\Phi), \Phi) \quad \text{s.t.} \quad \alpha^*(\Phi) = \underset{\alpha}{\operatorname{argmin}} F(\alpha, \Phi)$$

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

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:

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

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

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

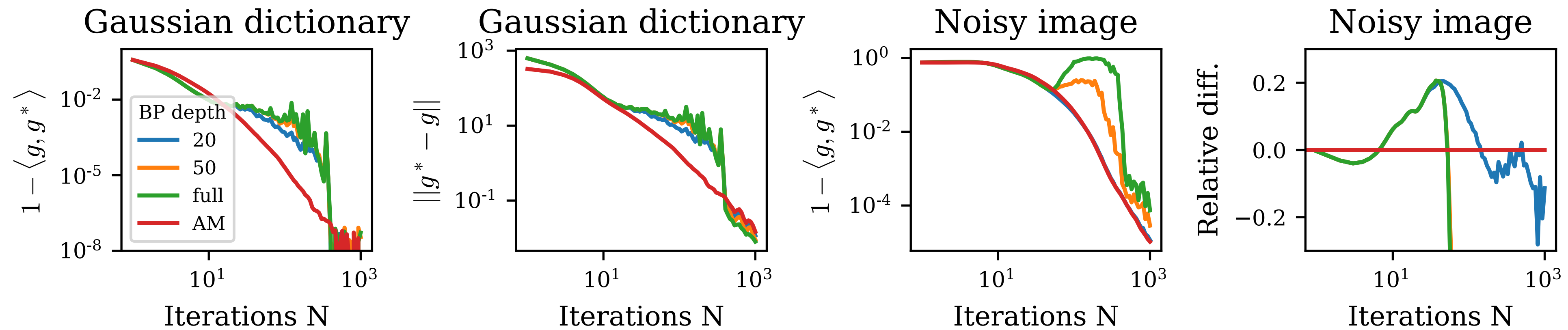
2. Compute the gradient of $F(\alpha^N(\Phi), \Phi)$ using autodifferentiation

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

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

- ▶ 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: $g_N^{DDL}(\Phi) \in \nabla_2 F(\alpha^N(\Phi), \Phi) + J_N(\partial_1 F(\alpha^N(\Phi), \Phi))$ Quadratic convergence on the support
- ▶ $\|J_N - J^*\| \leq A_N + B_N$, where $A_N \rightarrow 0$ and B_N an error term

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

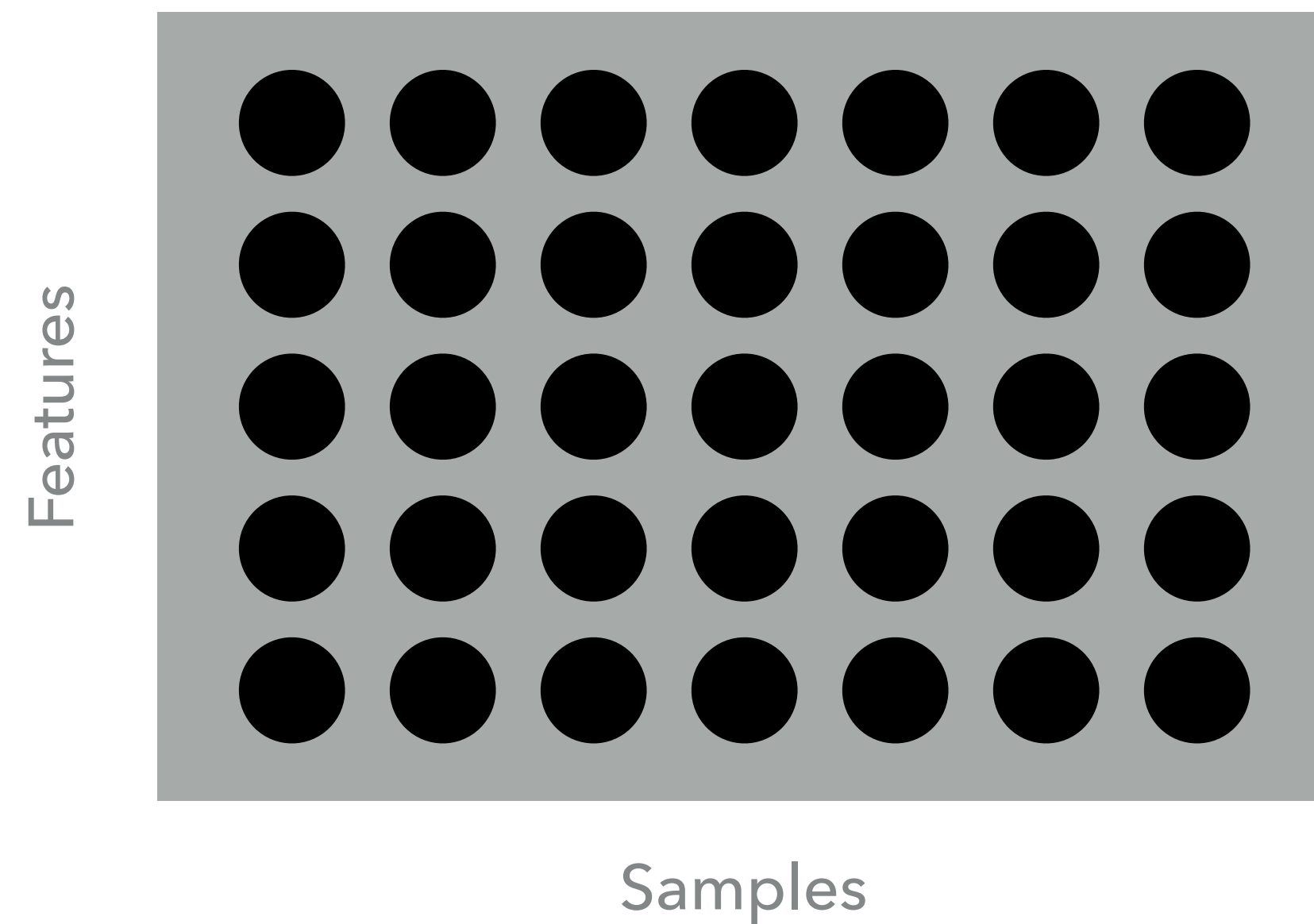


- ▶ First iterations: stable behavior
- ▶ Too many iterations: accumulations of errors !
- ▶ On the support: convergence to g^*

DICTIONARY LEARNING AND MATRIX FACTORIZATION

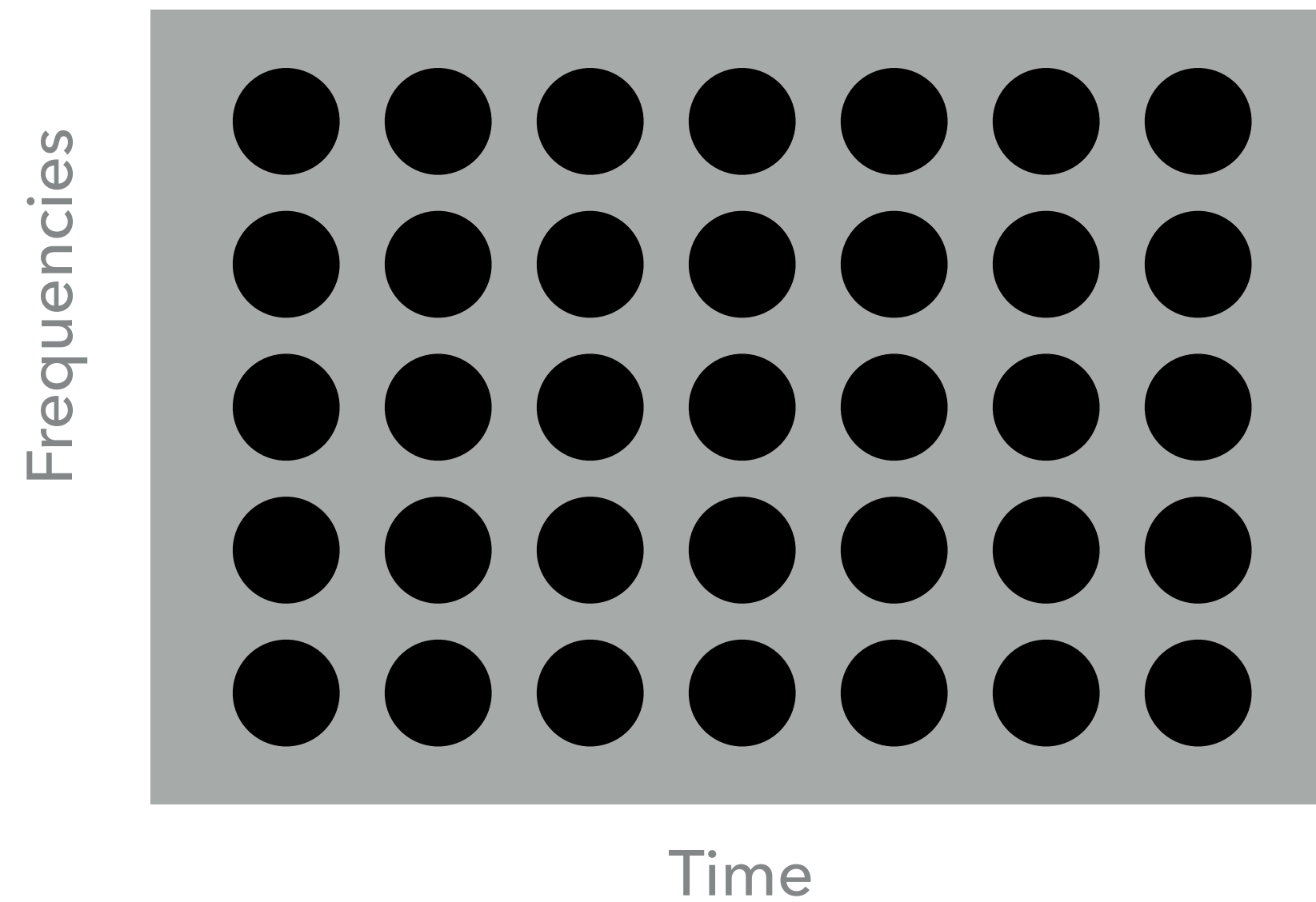
MATRIX FACTORIZATION

- ▶ Data are often available under a matrix form:



MATRIX FACTORIZATION

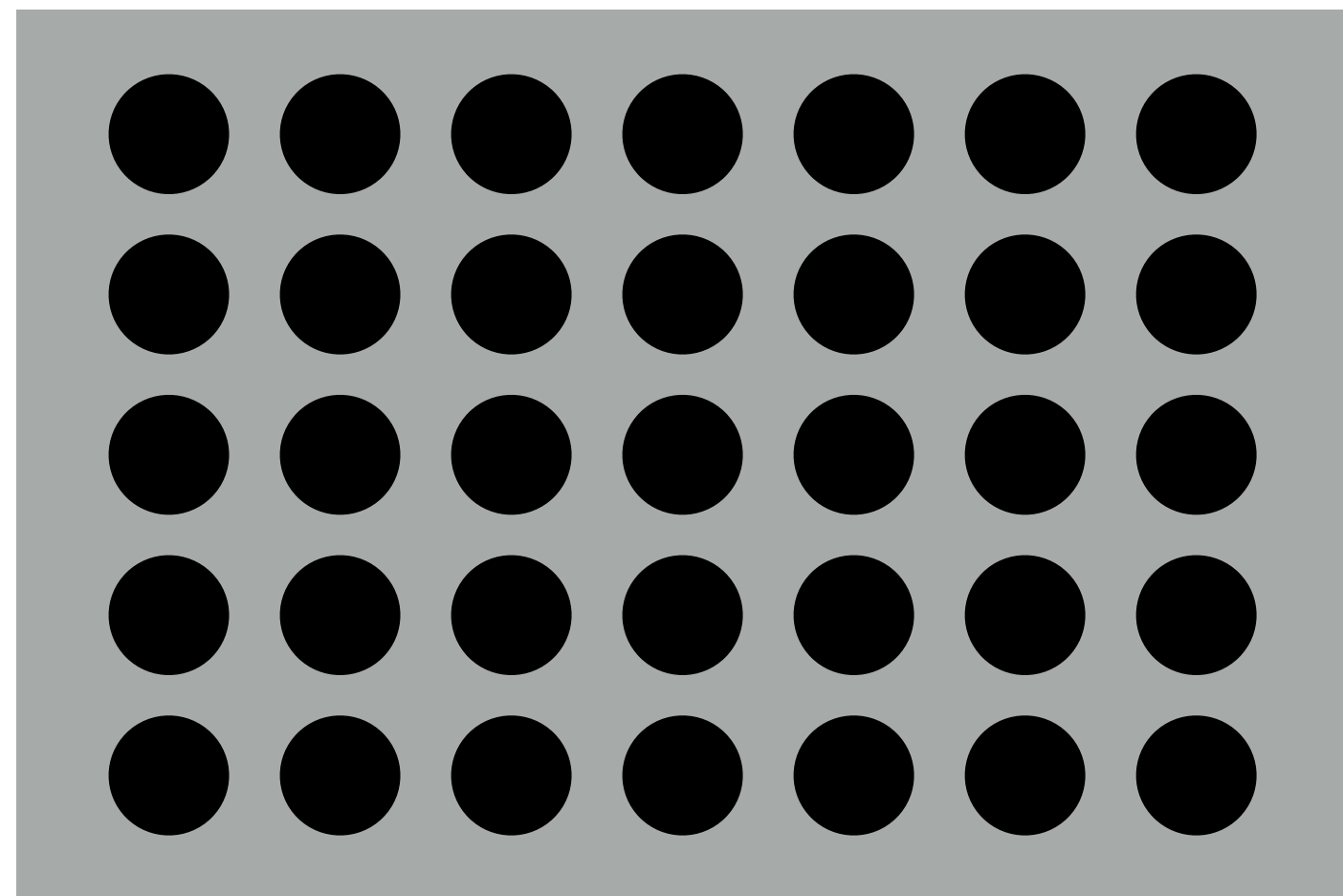
- ▶ Example: time-frequency representation



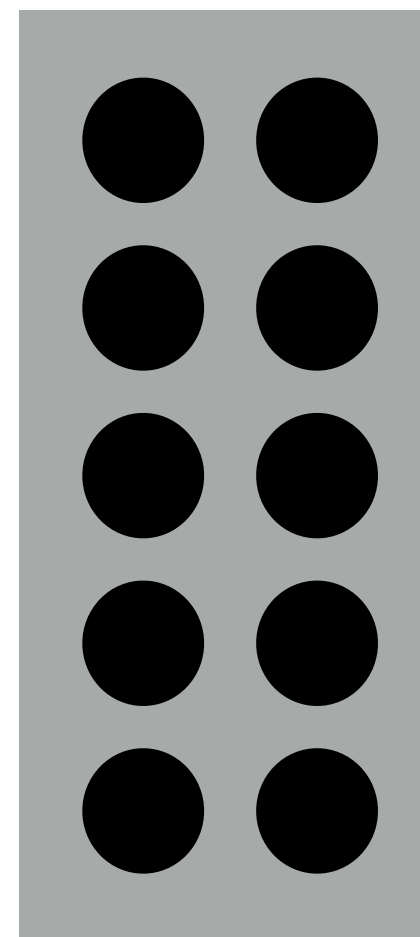
MATRIX FACTORIZATION

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

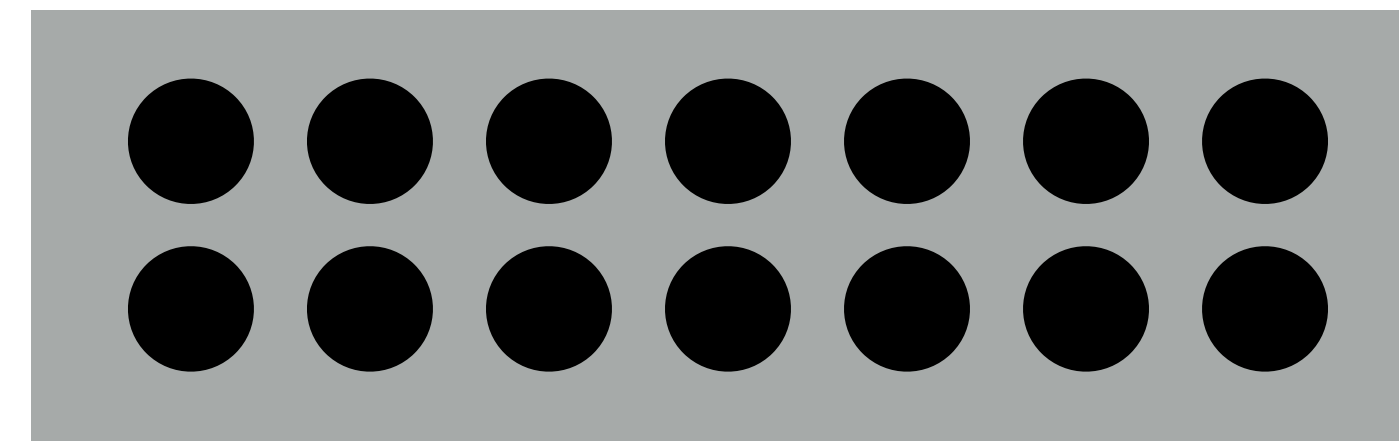
Data: X



Dictionary: W



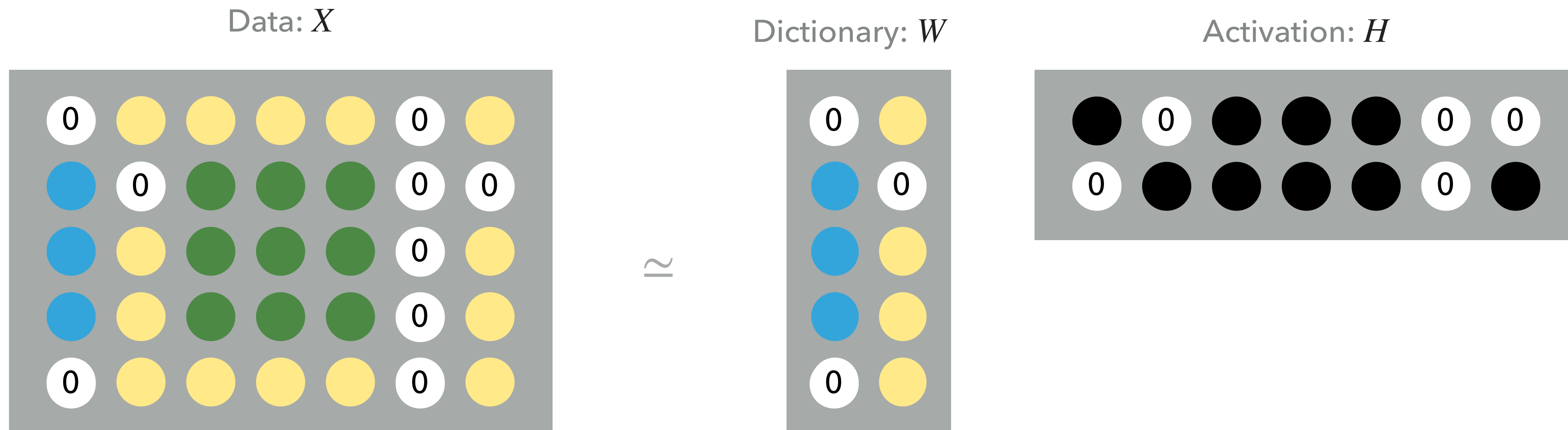
Activation: H



\mathbb{R}

MATRIX FACTORIZATION

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



MATRIX FACTORIZATION: EXAMPLES

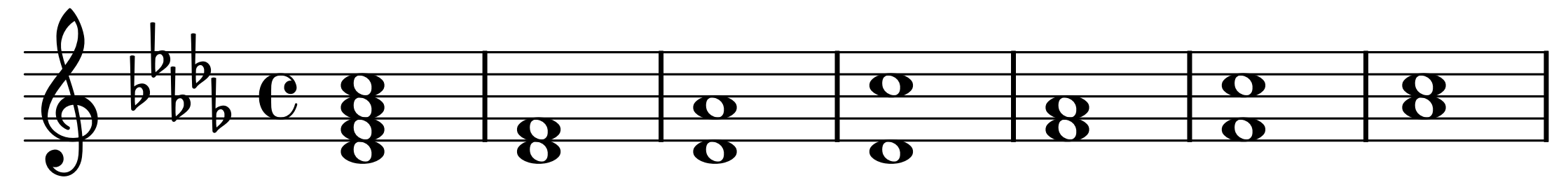
- ▶ PCA : $X \in \mathbb{R}^{NN} = UDU^T$, U being the eigen vectors (and is orthogonal) and D the eigen values
- ▶ SVD: $X \in \mathbb{R}^{MN} = UDV^T$, D being the singular values, U and V are orthogonal
- ▶ ICA
- ▶ etc.

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 matrix of non negative coefficients
- ▶ The Features can only add each other (no soustraction)
- ▶ Facilitate the interpretability

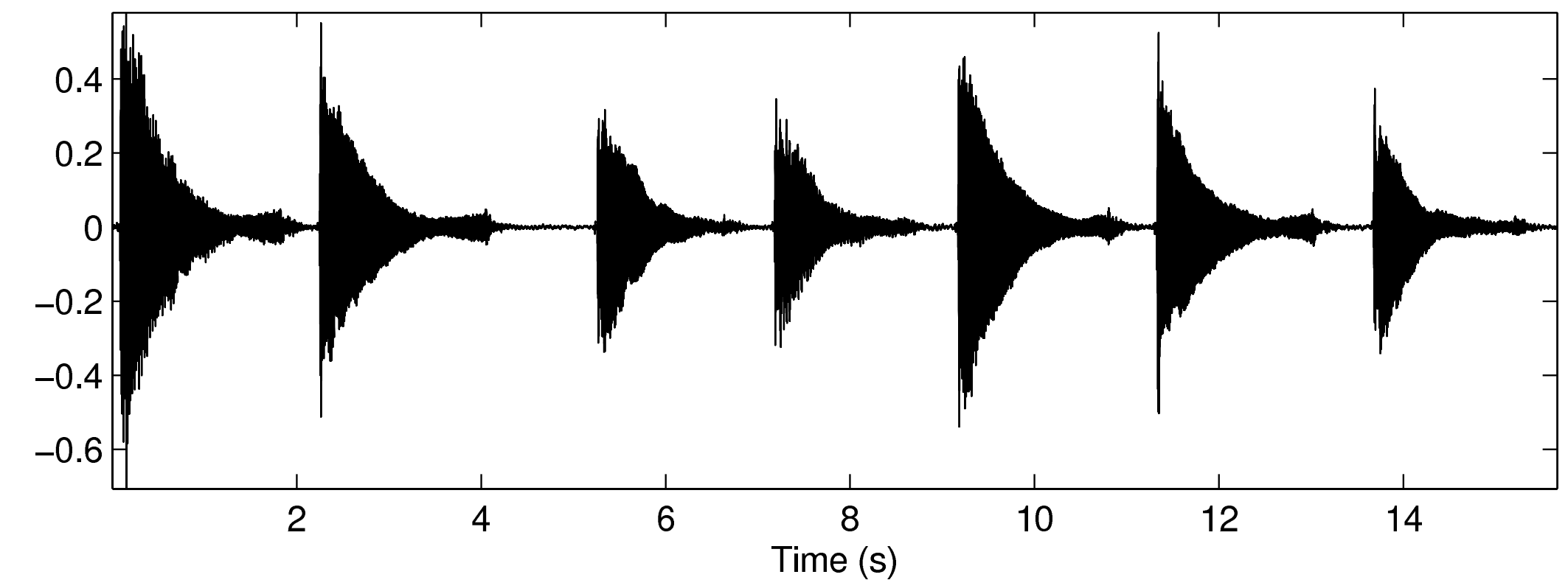
NMF EXAMPLES

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

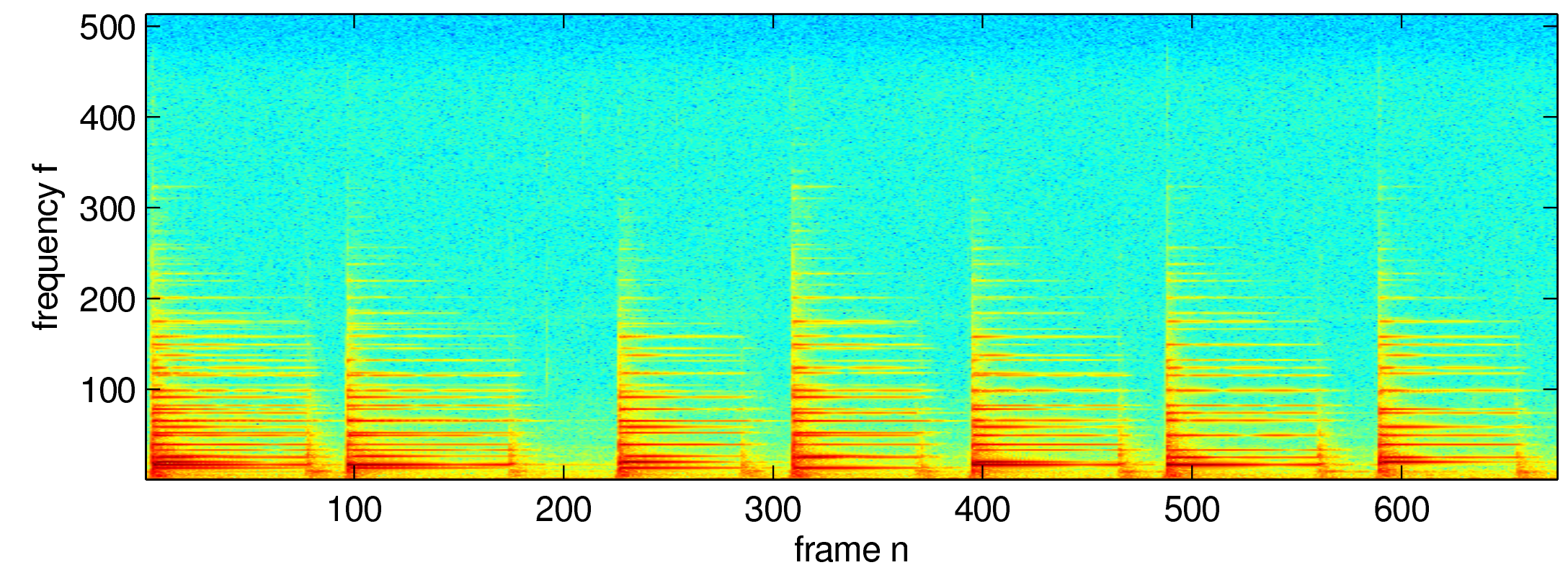


(MIDI numbers : 61, 65, 68, 72)

Signal x



Log-power spectrogram

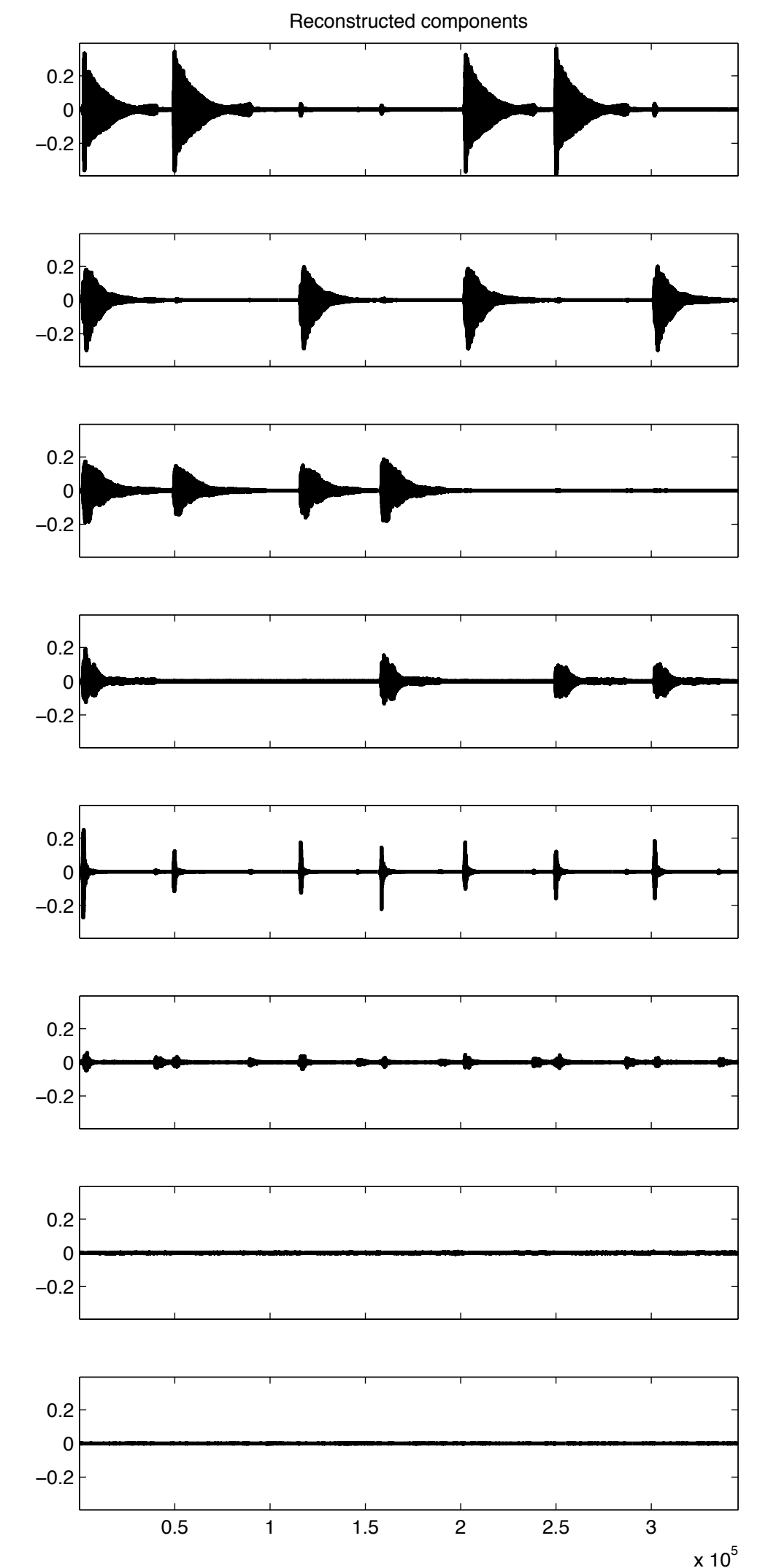
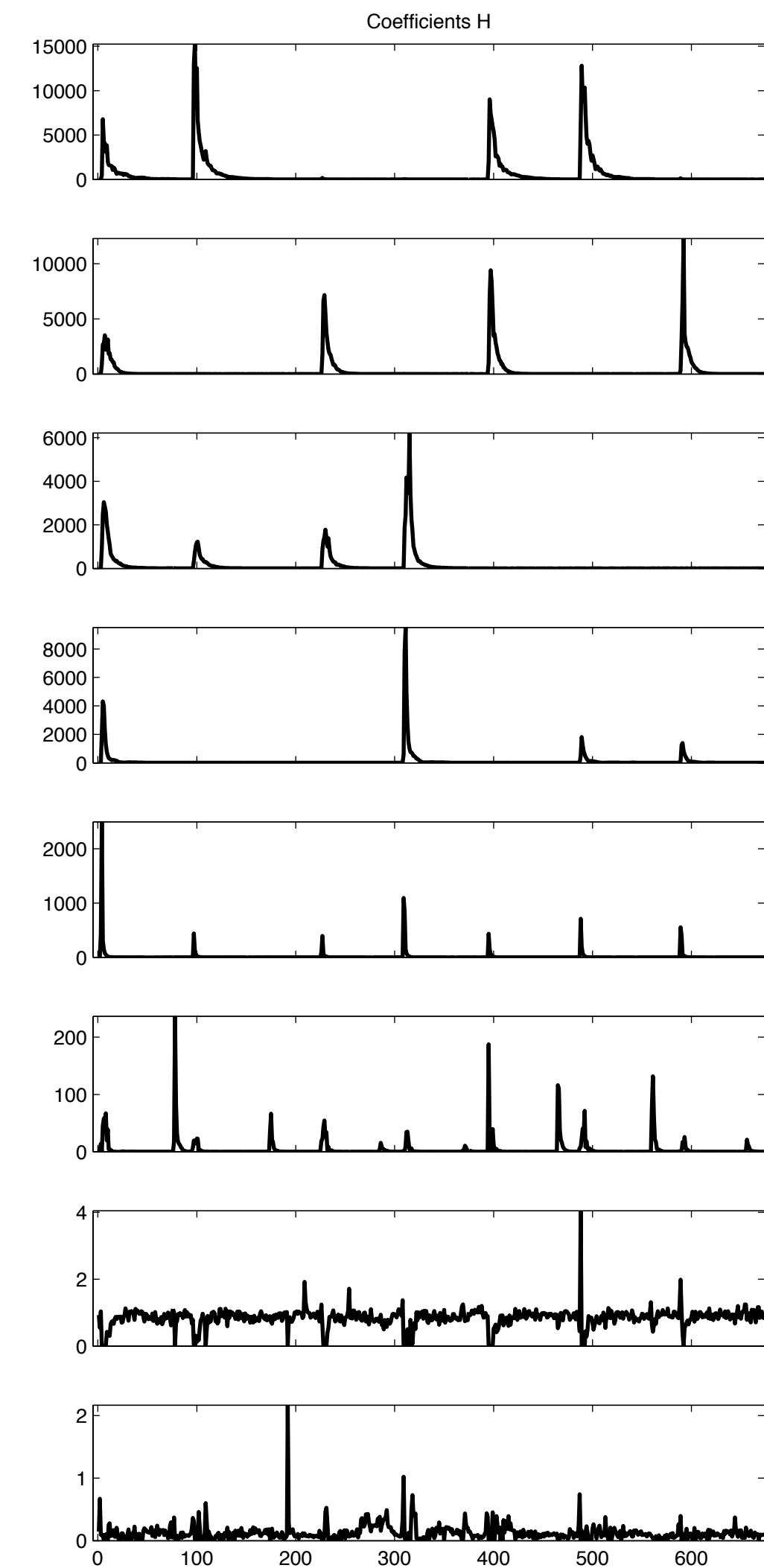
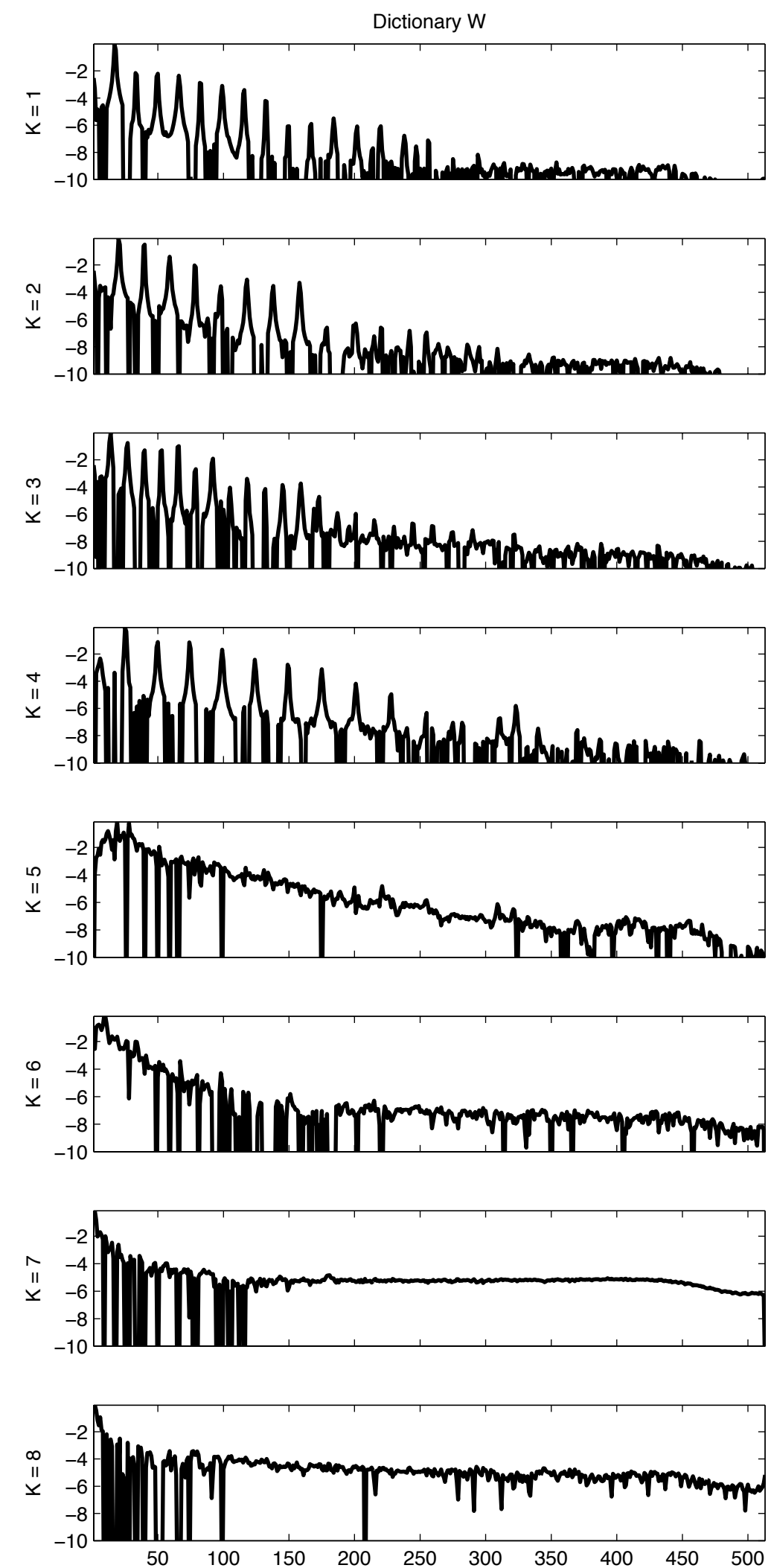


NMF EXAMPLES

▶ Piano example from [Fevotte 2009]

▶ Demo available at:

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



x 10⁵

NMF DECOMPOSITIONS

- ▶ Choose a measure of fit between X and WH :

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

- ▶ $D(X | WH) = \sum_{n,k} d(X[n, k] | \{WH\}[n, k])$, where d is a scalar cost function

- ▶ Exemple of cost functions :

- ▶ Euclidian : $D(X | WH) = \|X - WH\|^2$ [Paatero & Tapper, 1994] [Lee & Seung, 2001]

- ▶ Kullback-Leibler divergence: $d(x | y) = x \log \frac{x}{y} + (x - y)$ [Lee & Seung, 1999] [Finesso & Spreij, 2006]

- ▶ Itakura-Saito divergence: $d(x | y) = \frac{x}{y} - \log \frac{x}{y} - 1$ [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

▶ $x = \Phi\alpha + e$

▶ $\alpha \sim \mathcal{N}_c(0, \text{diag}(WH))$ 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

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

► With

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

► And

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

LRTFS — ALGORITHM

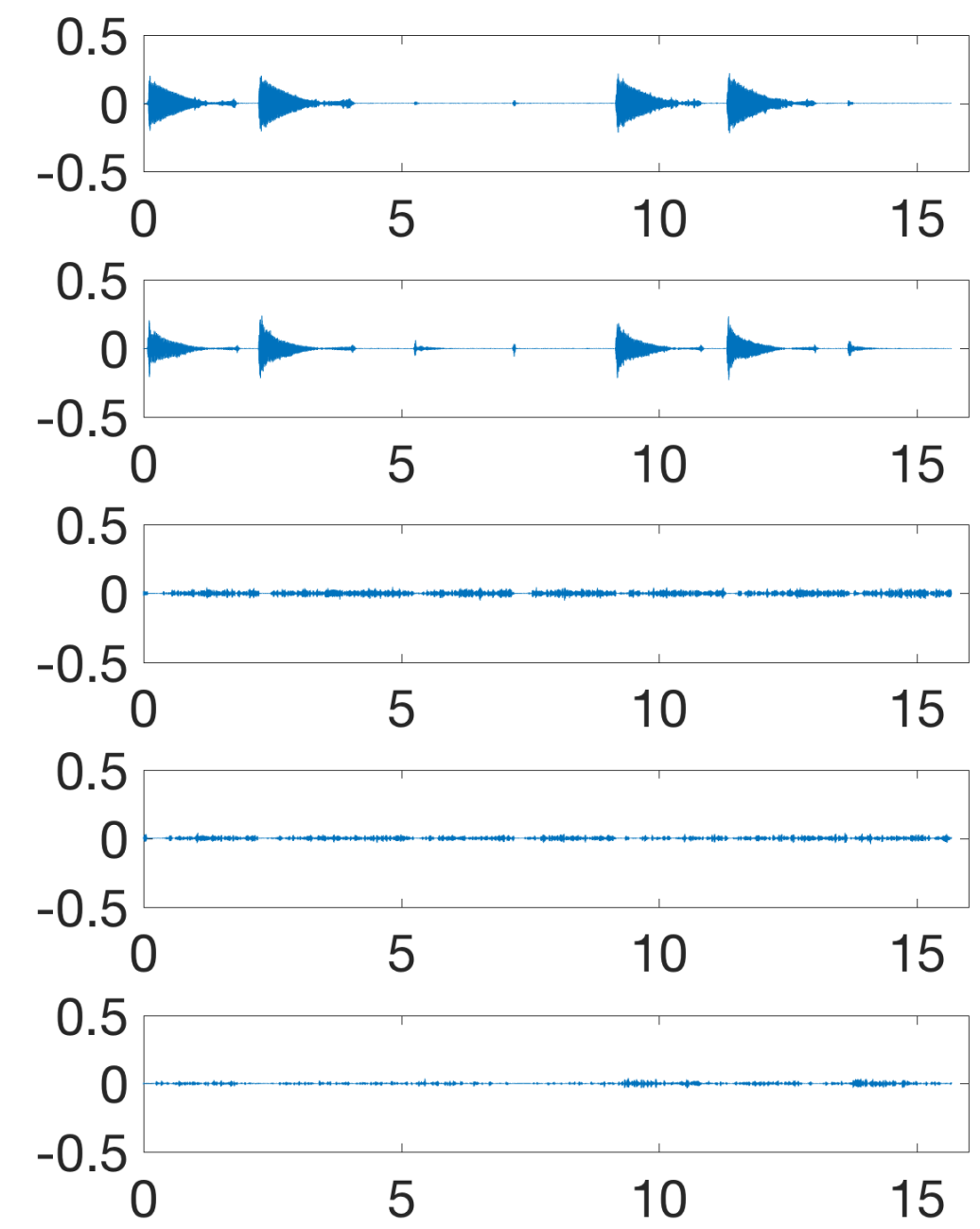
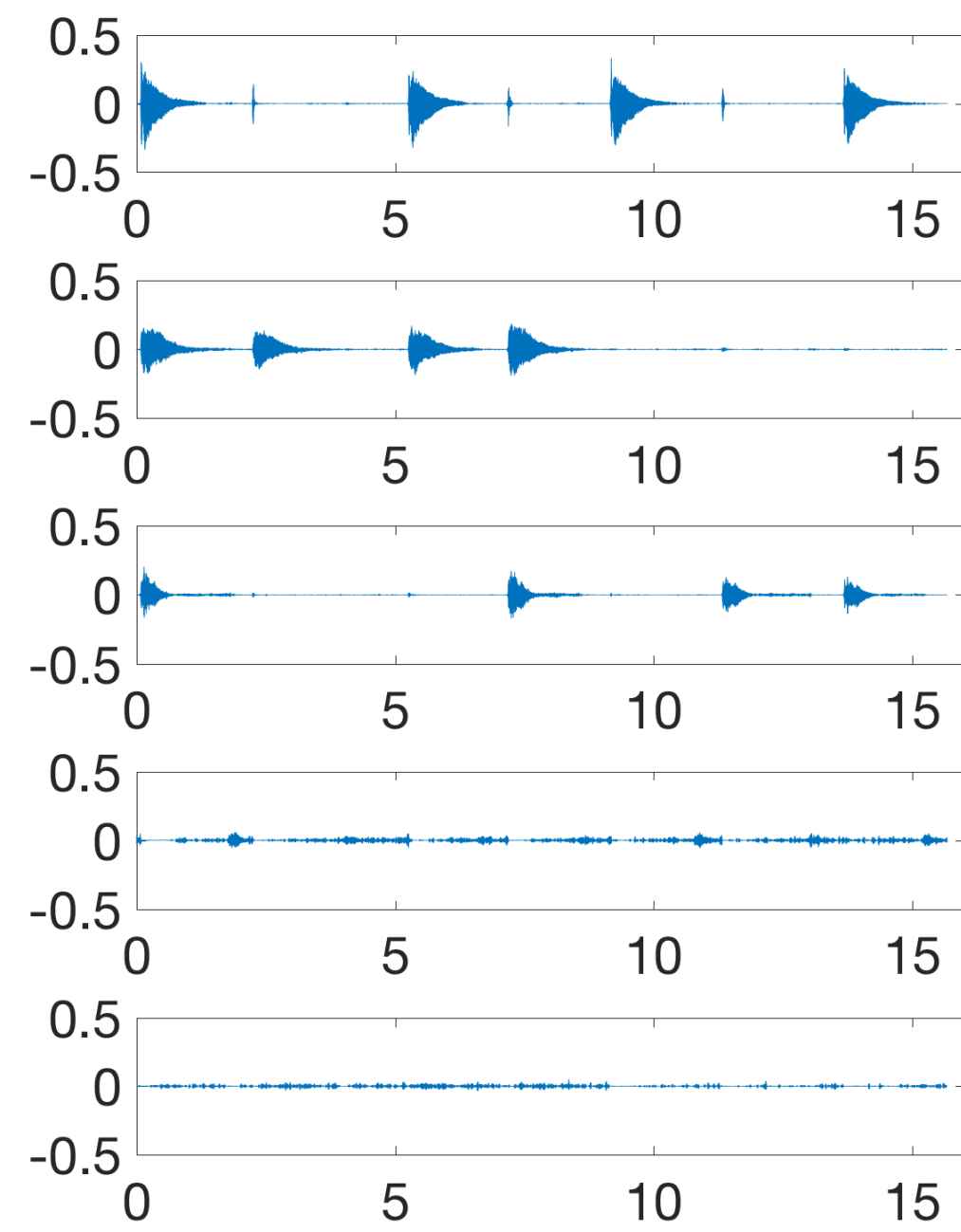
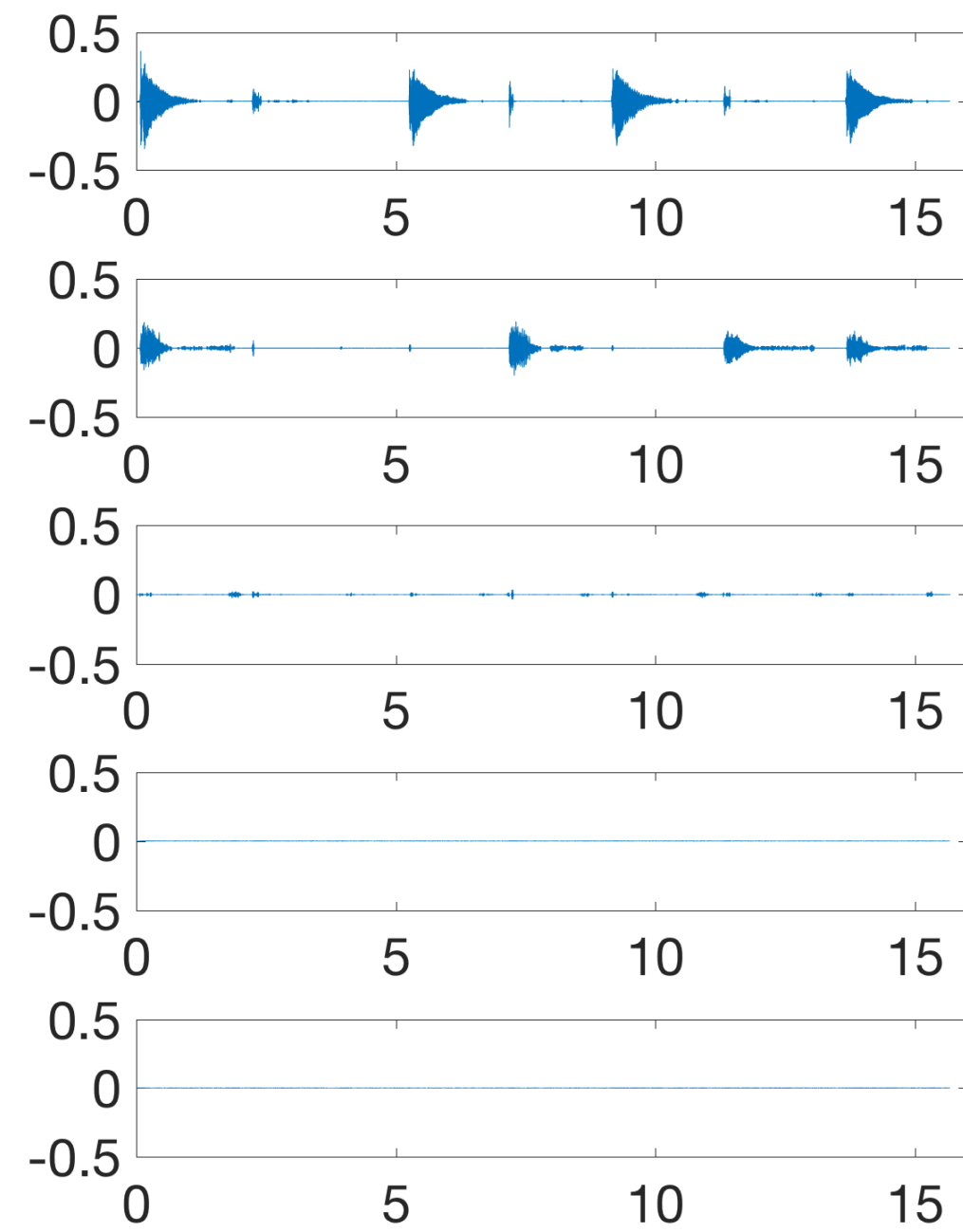
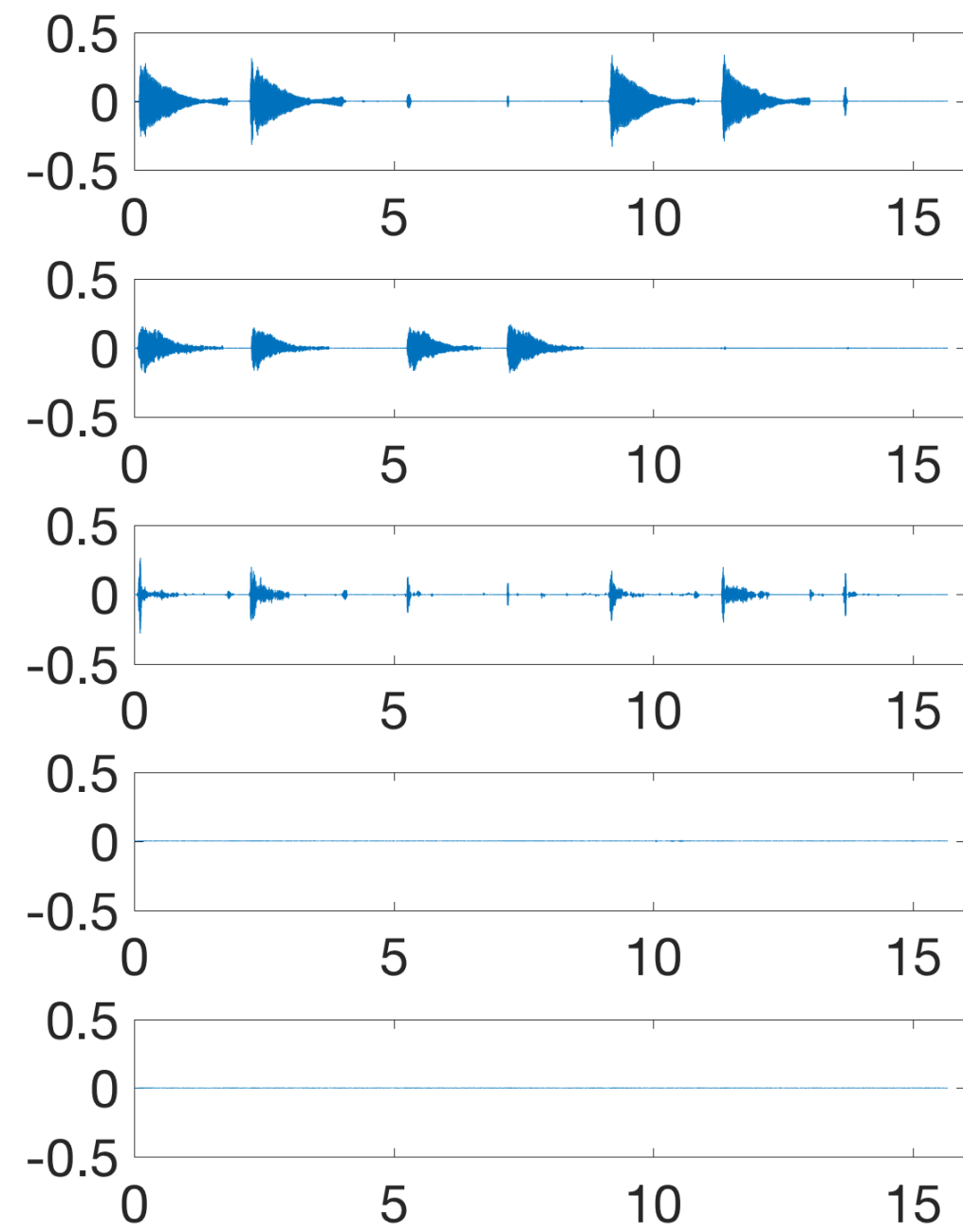
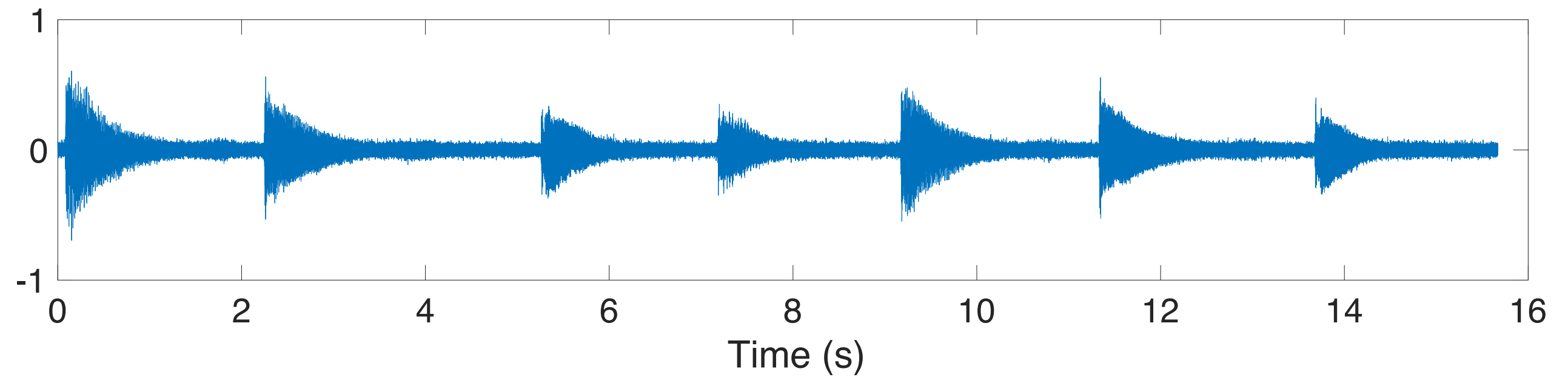
$$1. v^{(i)} = W^{(i)}H^{(i)}$$

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

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

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

EXAMPLES: PIANO + NOISE

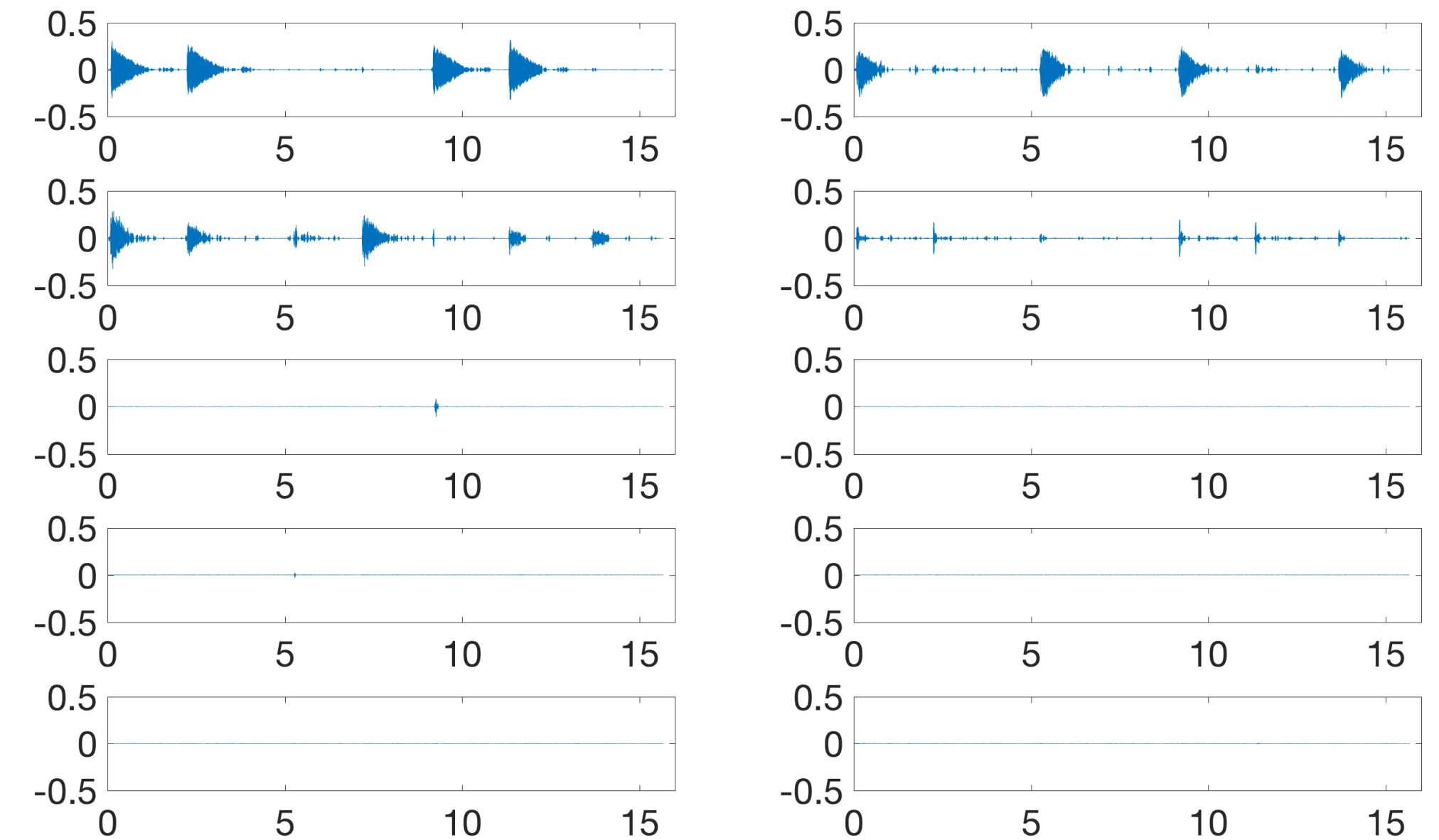
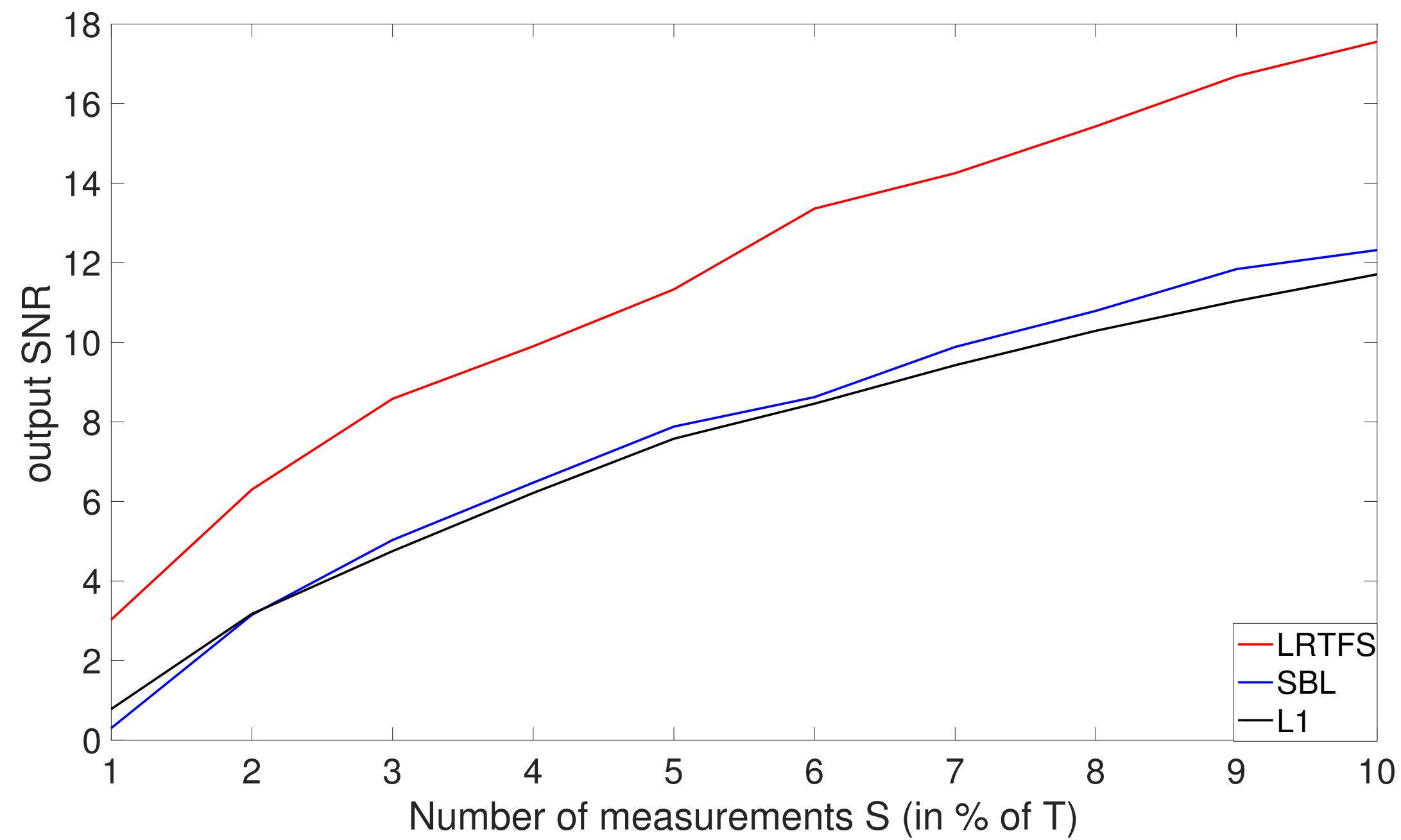


LRTFS

IS-NMF on spectrogram

EXAMPLES: PIANO AND COMPRESSED SENSING

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



CONCLUSION

- ▶ Take-home message
 - ▶ NMF can be thought 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 Convolutional 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