

DICTIONARY LEARNING

---

# SPARSE REPRESENTATIONS

# DICTIONARY LEARNING AND SPARSE CODING

## SPARSE CODING

- ▶ Let  $x \in \mathbb{R}^N$  be a signal
- ▶ Let  $\Phi \in \mathbb{R}^{MN}$  be a dictionary
- ▶ Sparse coding:

$$\min \frac{1}{2} \|x - \Phi\alpha\|_2^2 + \lambda \mathcal{R}(\alpha)$$

- ▶ Where  $\mathcal{R}$  is a measure of sparsity ( $\|\alpha\|_0$ ,  $\|\alpha\|_1$  etc.)

## WHICH DICTIONARY ?

- ▶ Usual dictionary:
  - ▶ Time-frequency
  - ▶ Wavelets
  - ▶ \*-lets
  - ▶ Union of dictionaries
- ▶ Can we learn the dictionary ?

## DICTIONARY LEARNING

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

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

## K-SVD

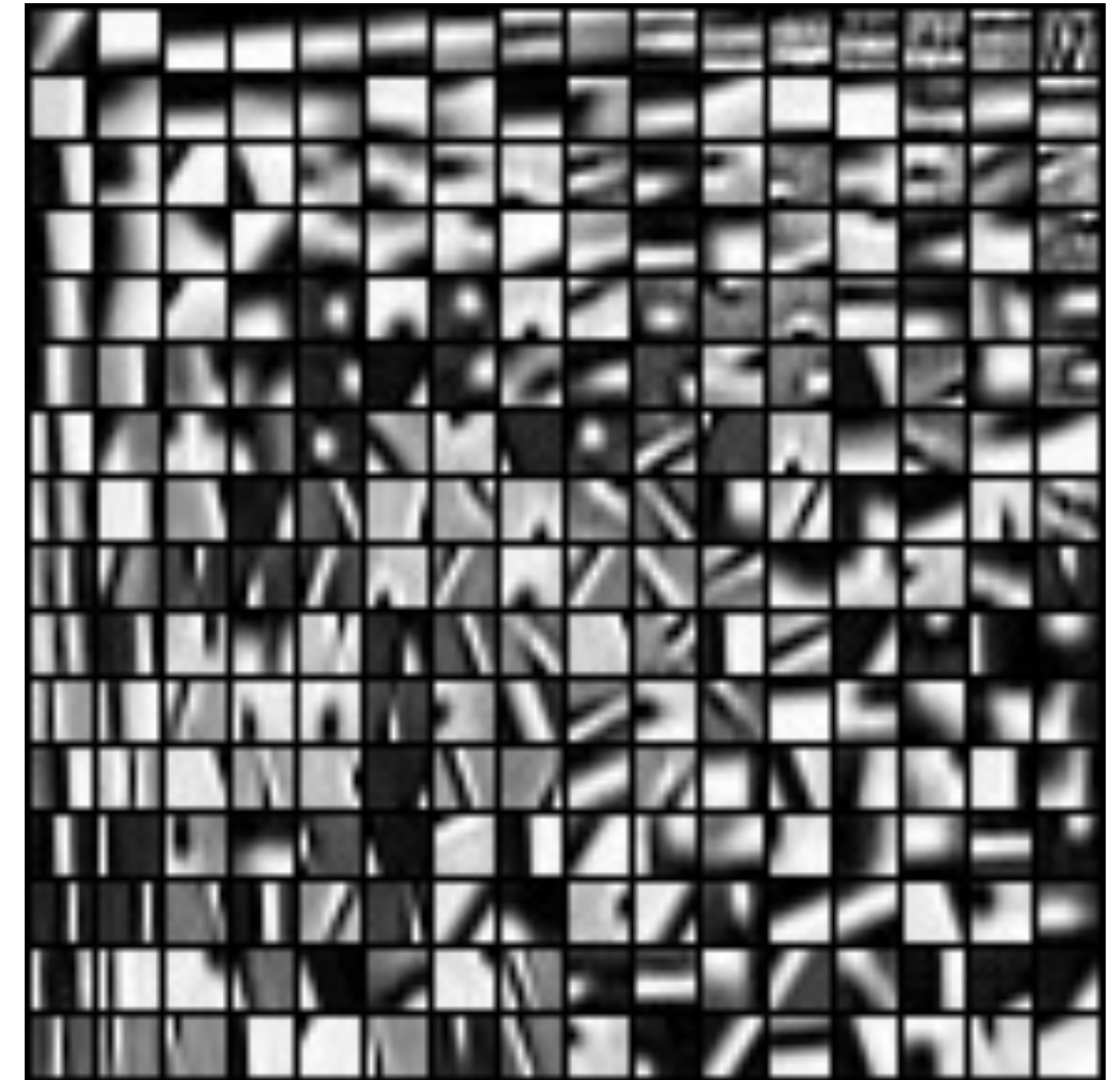
- ▶ [Elad and Aharon, 2006]
- ▶ Idea: alternate minimization between  $A$  and  $\Phi$
- ▶ Initialize  $\Phi$
- ▶ Sparse coding for each  $x_m$ :  $\alpha_m = \operatorname{argmin}_{\alpha} \frac{1}{2} \|x_m - \Phi\alpha_m\|^2$  s.t.  $\|\alpha\|_0 \leq s$
- ▶ Dictionary update :  $\min \frac{1}{2} \|X - \Phi A\|^2$
- ▶ Can be done column by column using SVD

## DICTIONARY LEARNING: ALTERNATING MINIMIZATION

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

## 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...



# CONVOLUTIONAL SPARSE CODING

## CONVOLUTIONAL SPARSE CODING

- ▶ This model assumes that the signal  $x \in \mathbb{R}^N$  can be decomposed as

$$x[t] = \sum_k (d_k * z_k)[t]$$

Where

- ▶  $d_k$  are "local filters"
- ▶  $z_k$  are the corresponding sparse representation

## CONVOLUTIONAL SPARSE CODING

- ▶ For each  $x^m[t]$ , assumes the CSC model:

$$x^m[t] = \sum_k (d_k * z_k^m)[t]$$

- ▶ Learning the dictionary  $\{d_k\}_k$ :

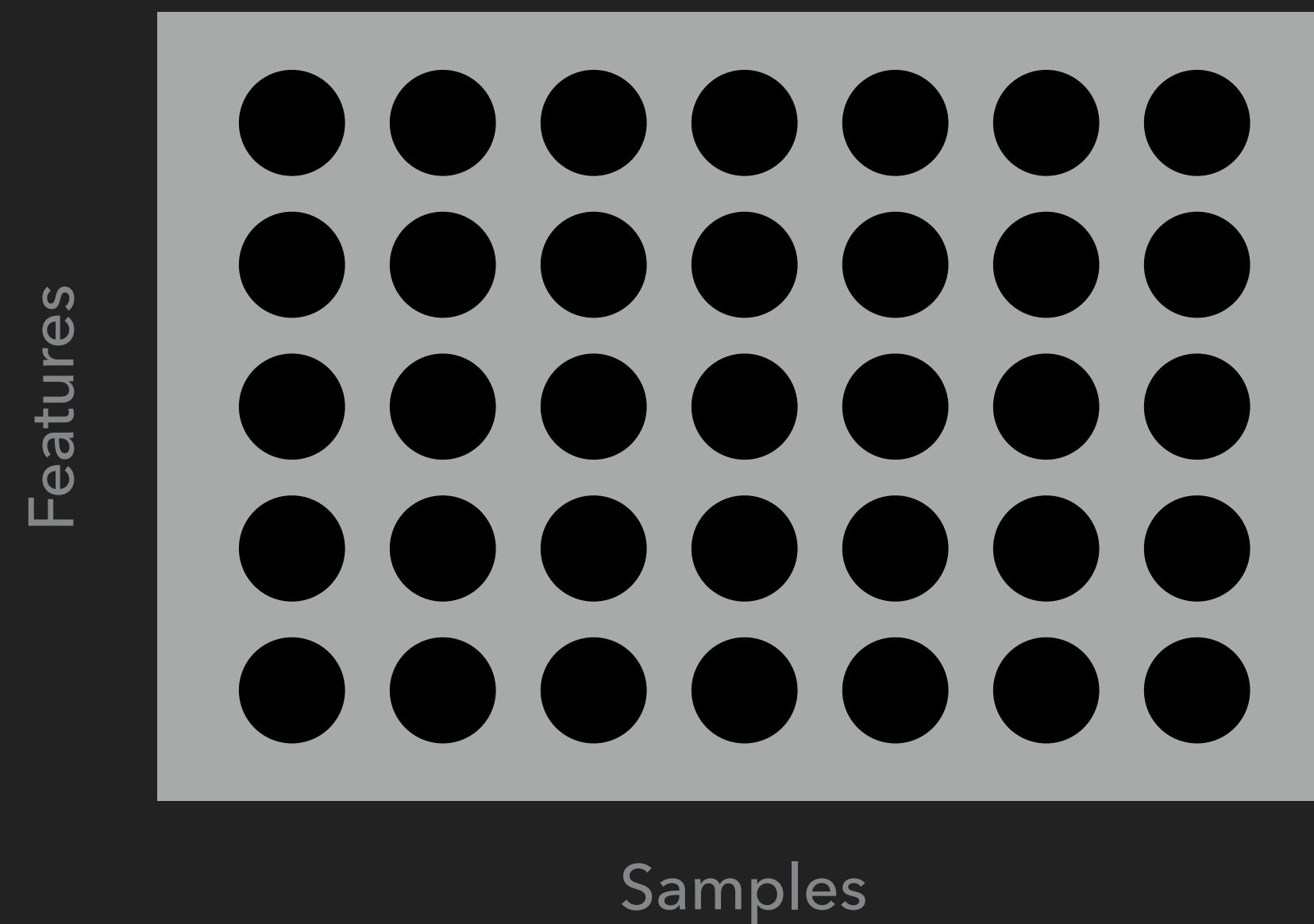
$$\min_{d_k, z_k^m} \sum_m \frac{1}{2} \|x^m - \sum_k d_k * z_k^m\|^2 + \lambda \|z_k\|_1$$

- ▶ [Bristow, Eriksson & Lucey 2013]  
[Heide, Heidrich, & Wetzstein 2015]  
[Pappyan, Romano, Sulam & Elad 2017]

# MATRIX FACTORIZATION

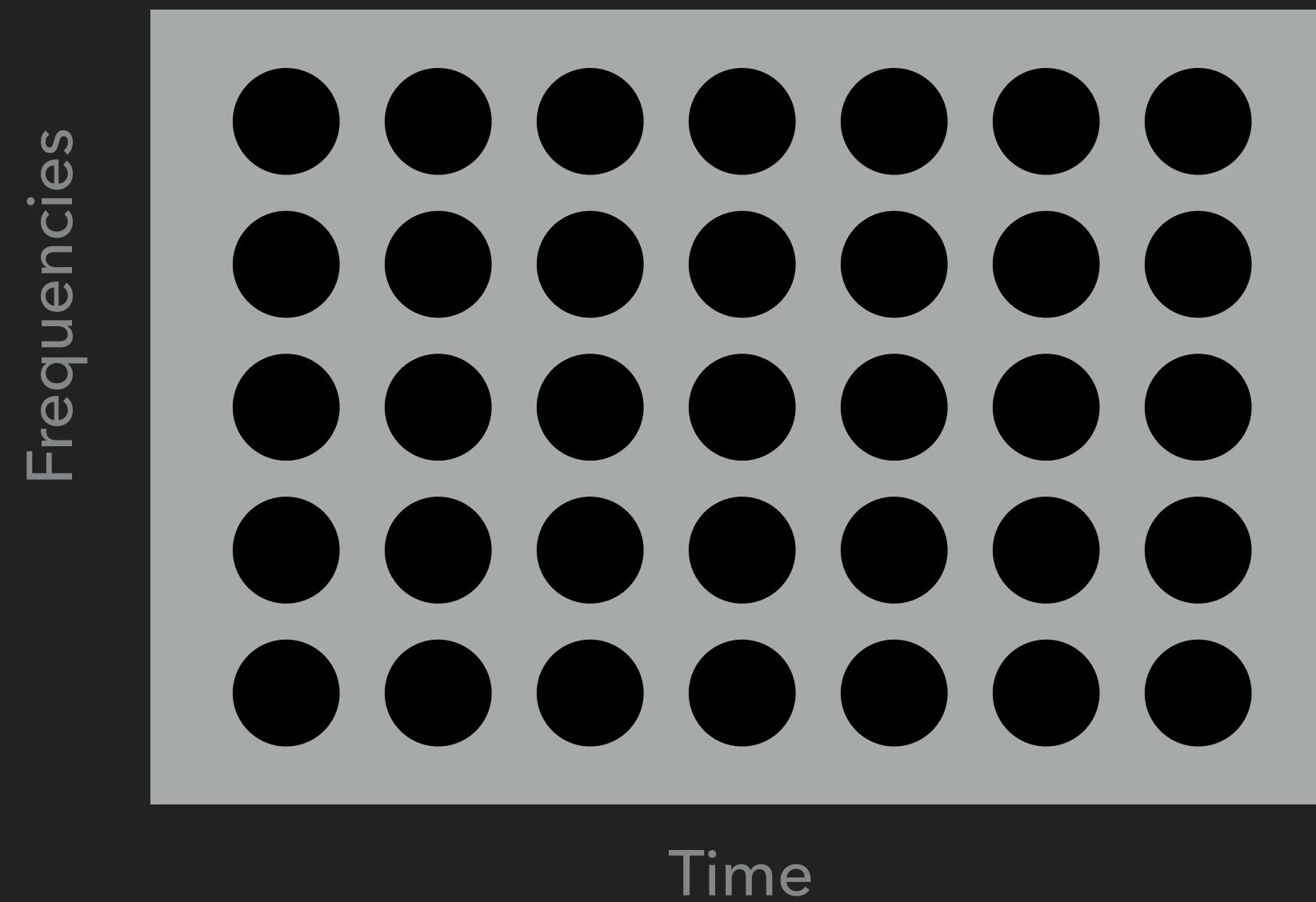
# MATRIX FACTORIZATION

- ▶ Data are often available under a matrix form:



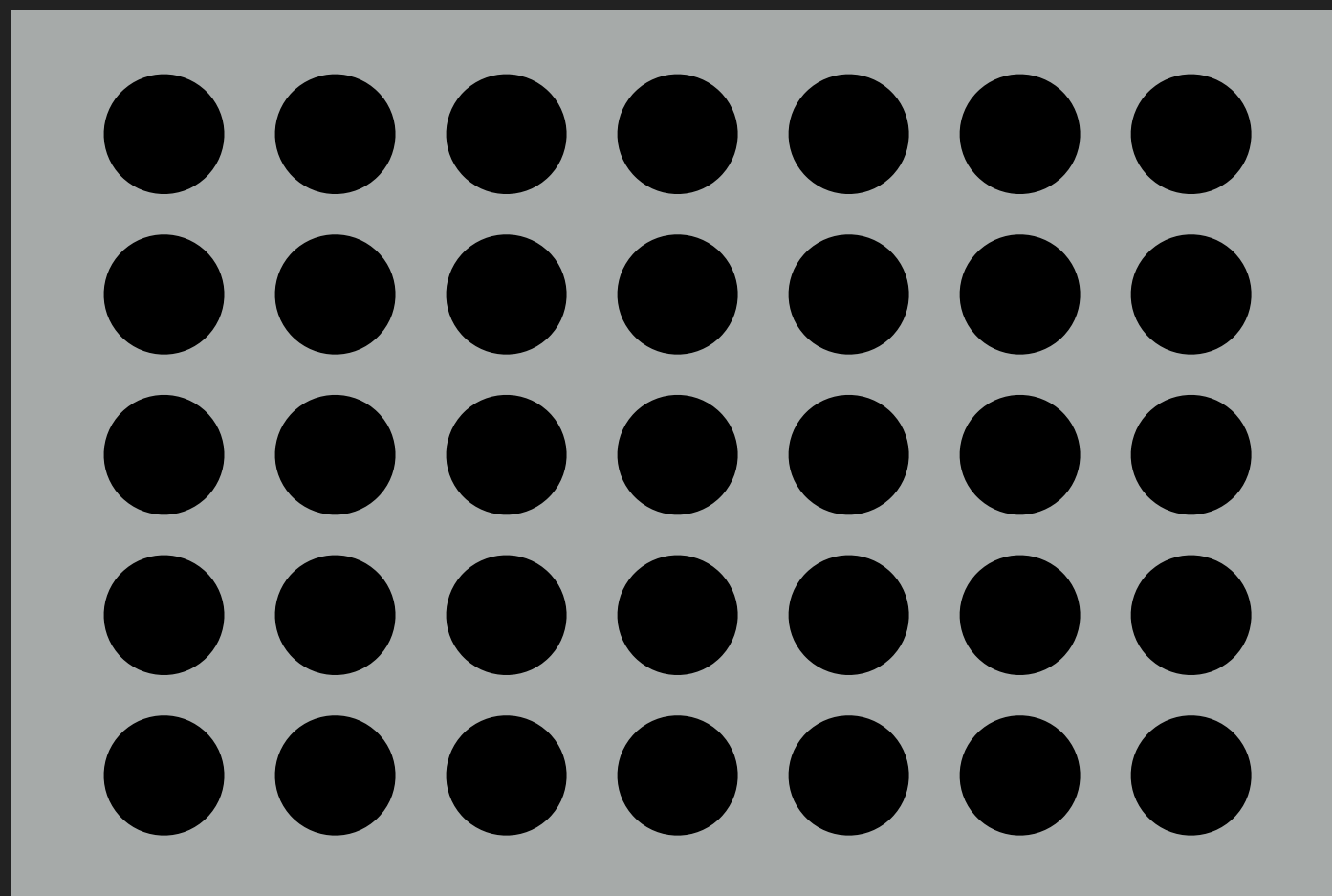
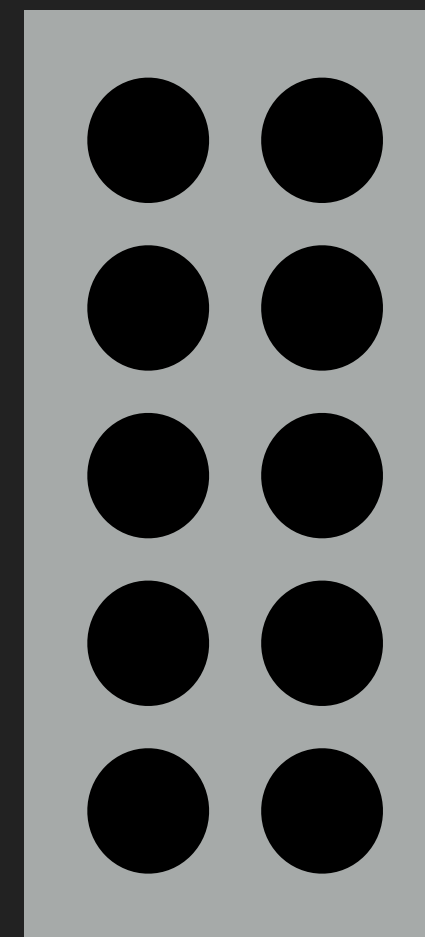
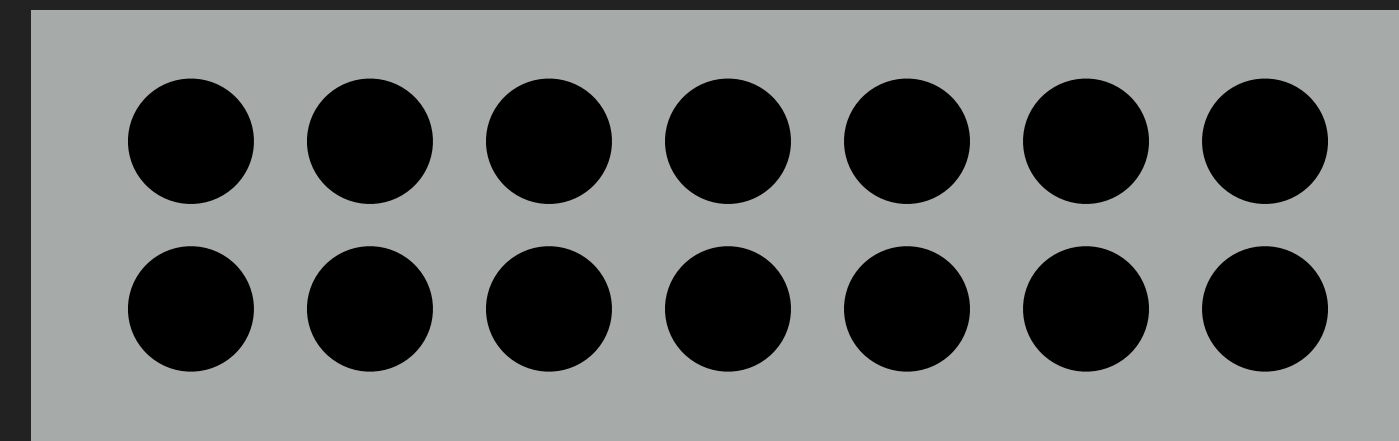
# MATRIX FACTORIZATION

- ▶ Exemple: time-frequency representation



# MATRIX FACTORIZATION

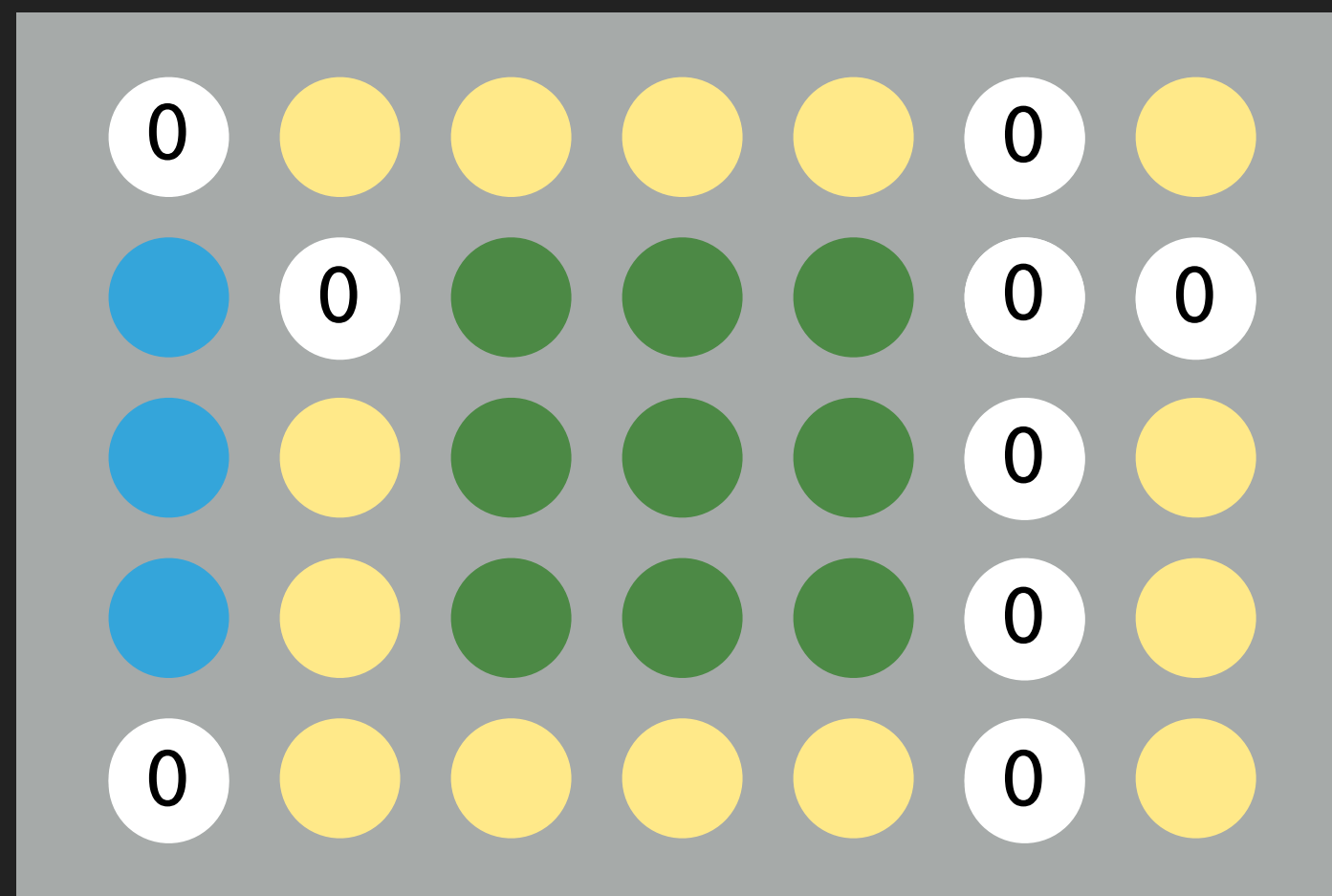
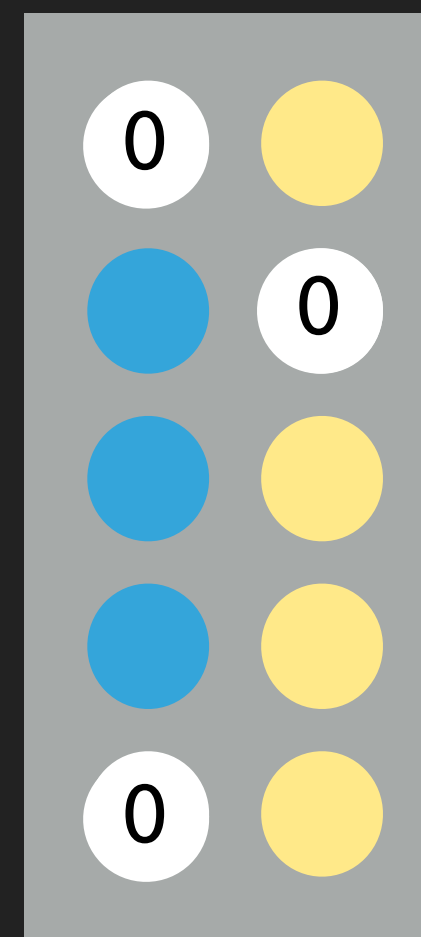
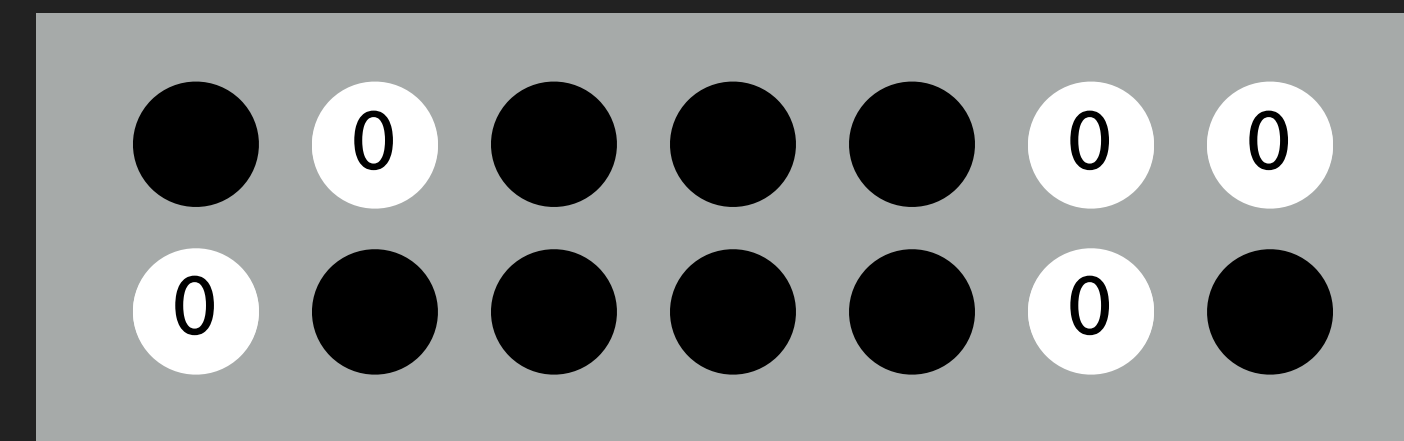
- ▶ Let  $X \in \mathbb{R}^{MN}$ . We aim to factorize  $X \simeq WH$ , with  $W \in \mathbb{R}^{MK}$  and  $H \in \mathbb{R}^{KN}$

Data:  $X$ Dictionary:  $W$ Activation:  $H$  $\simeq$



# 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:  $H$ 
 $\approx$

## 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

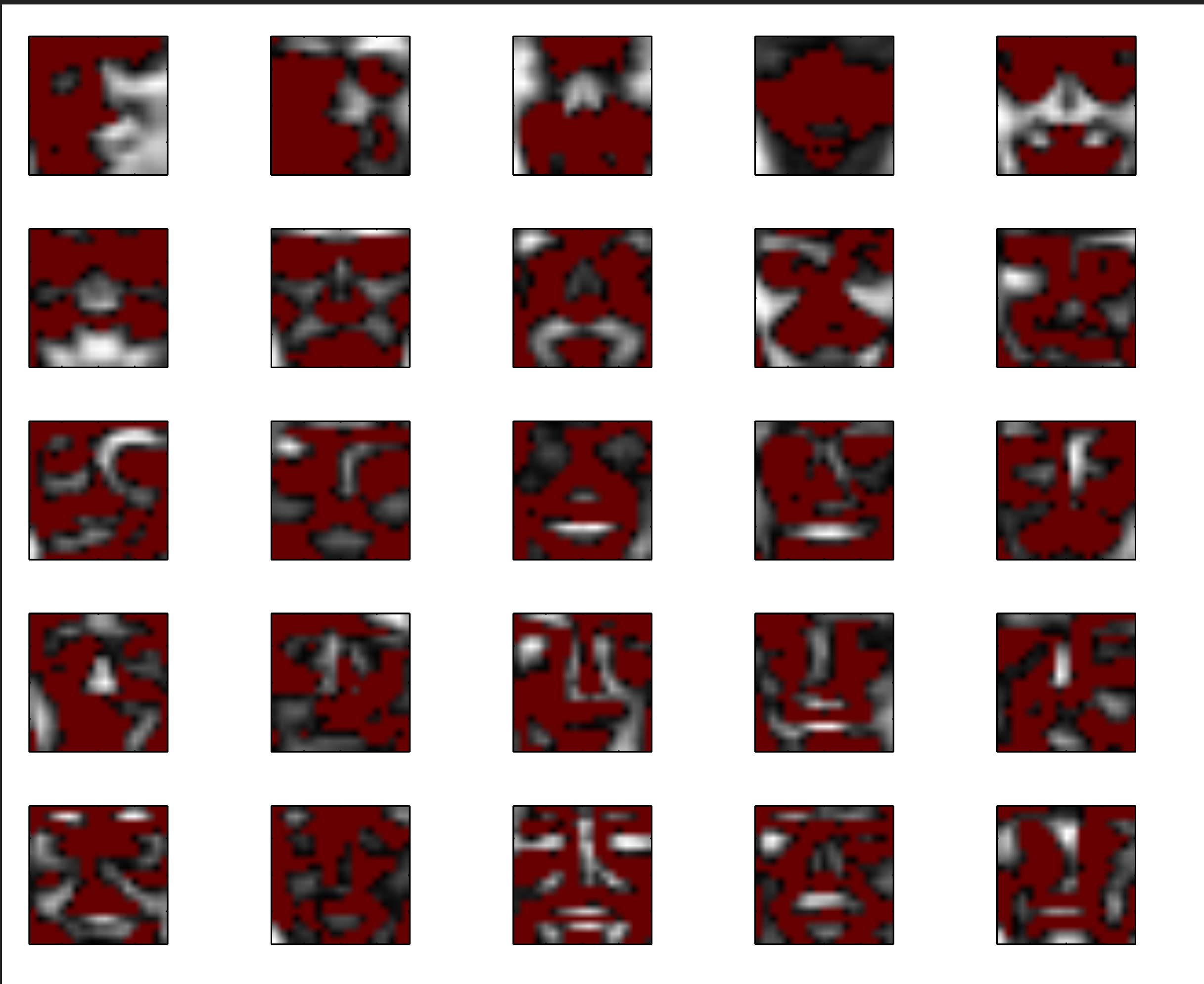
## NMF EXAMPLES

- ▶ Example from [Lee & Seung 1999]
- ▶ 49 images from MIT's CBCL data set



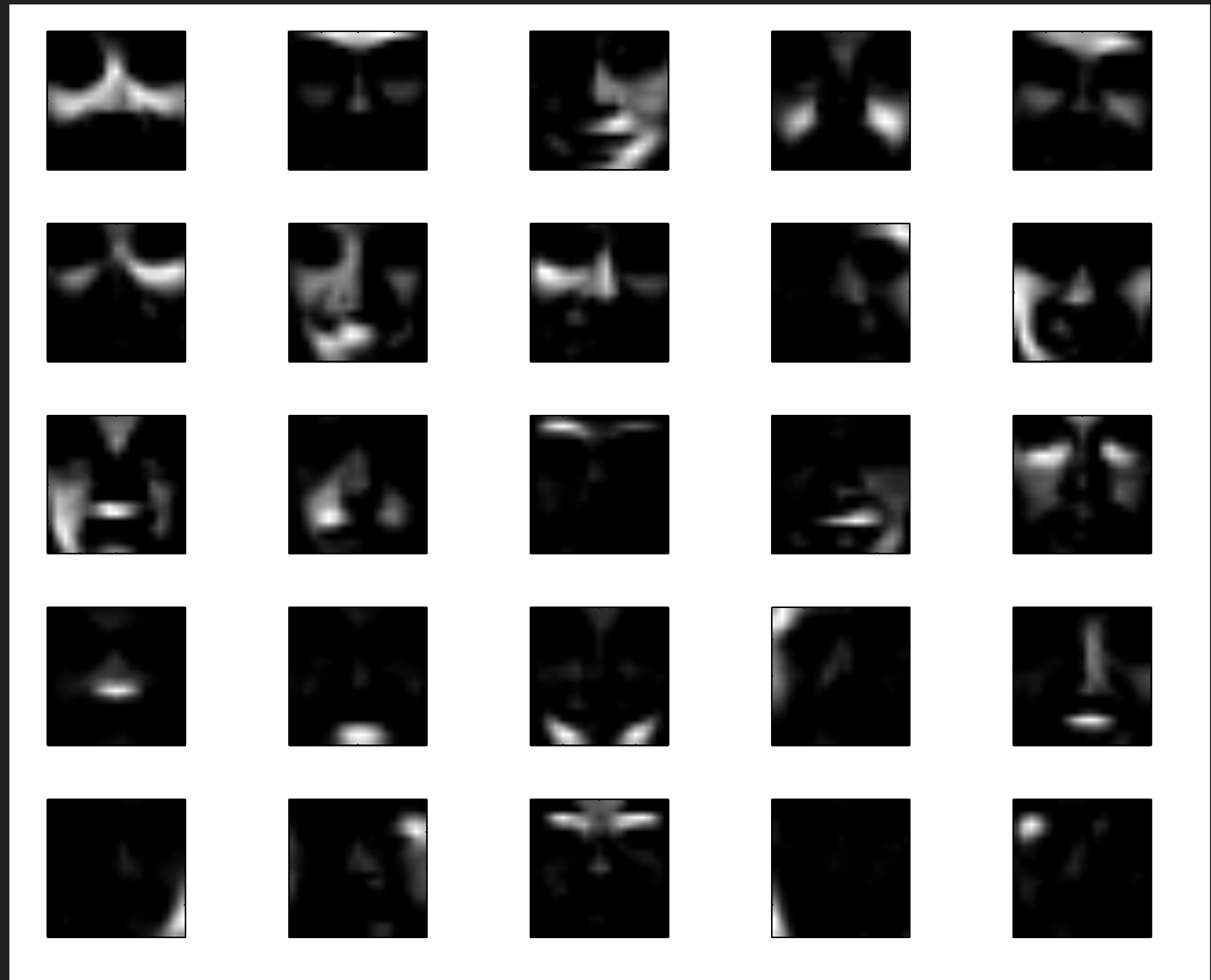
## NMF EXAMPLES

- ▶ Example from [Lee & Seung 1999]
- ▶ PCA decomposition
- ▶ Red values = negative pixel



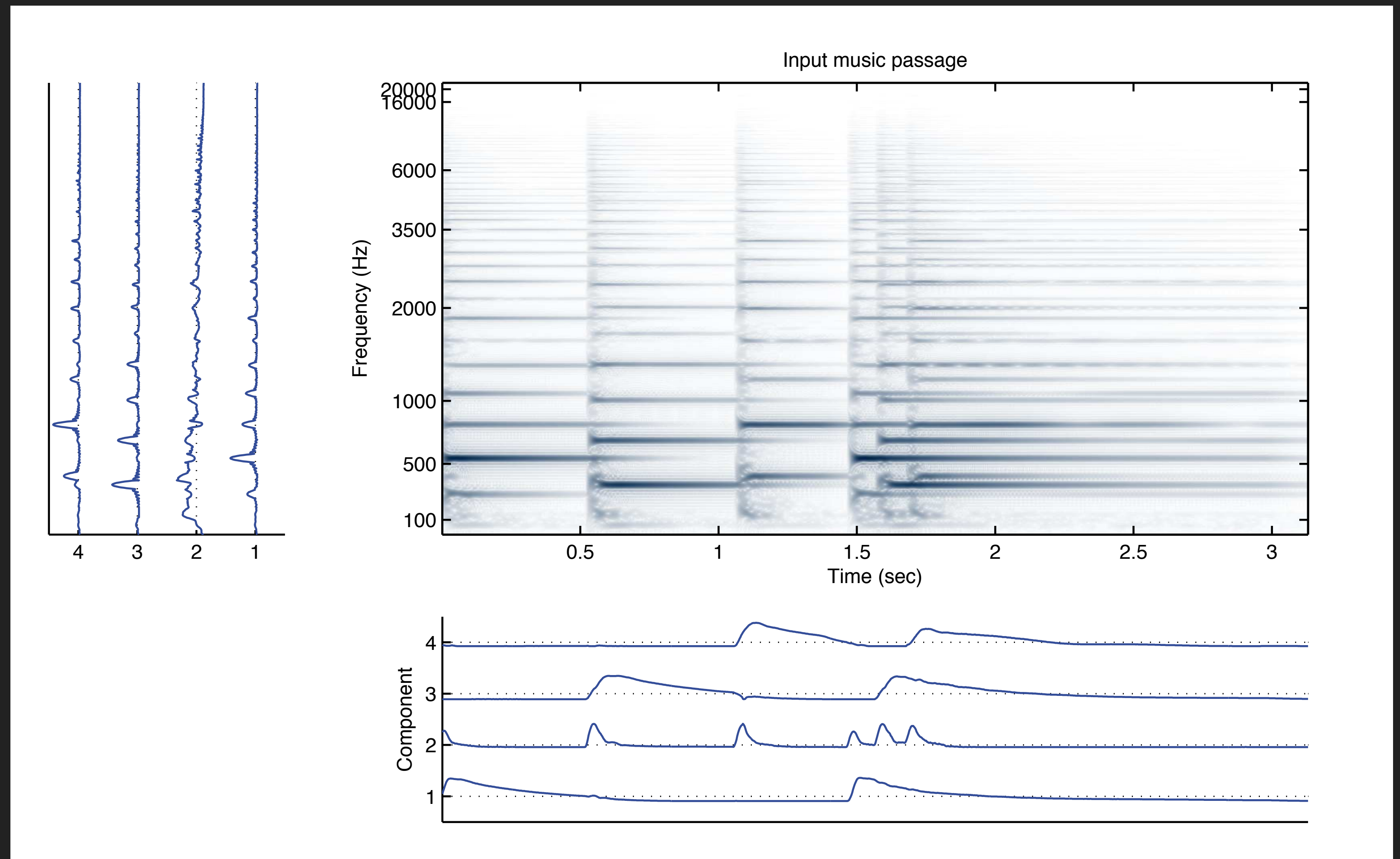
## NMF EXAMPLES

- ▶ Example from [Lee & Seung 1999]
- ▶ NMF decomposition



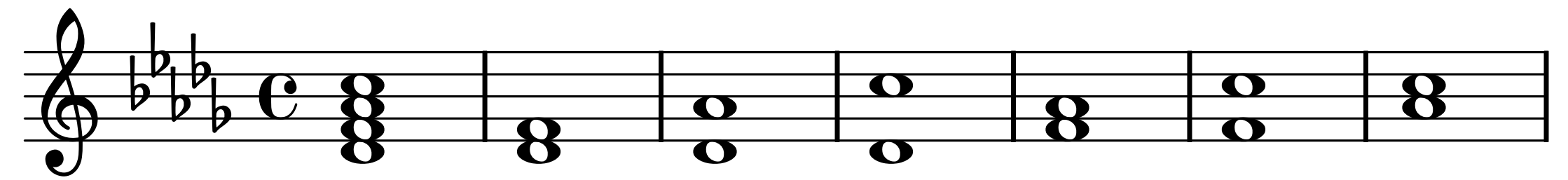
# NMF EXAMPLE

- ▶ From [Smaragdis 2013]
- ▶ Spectral unmixing



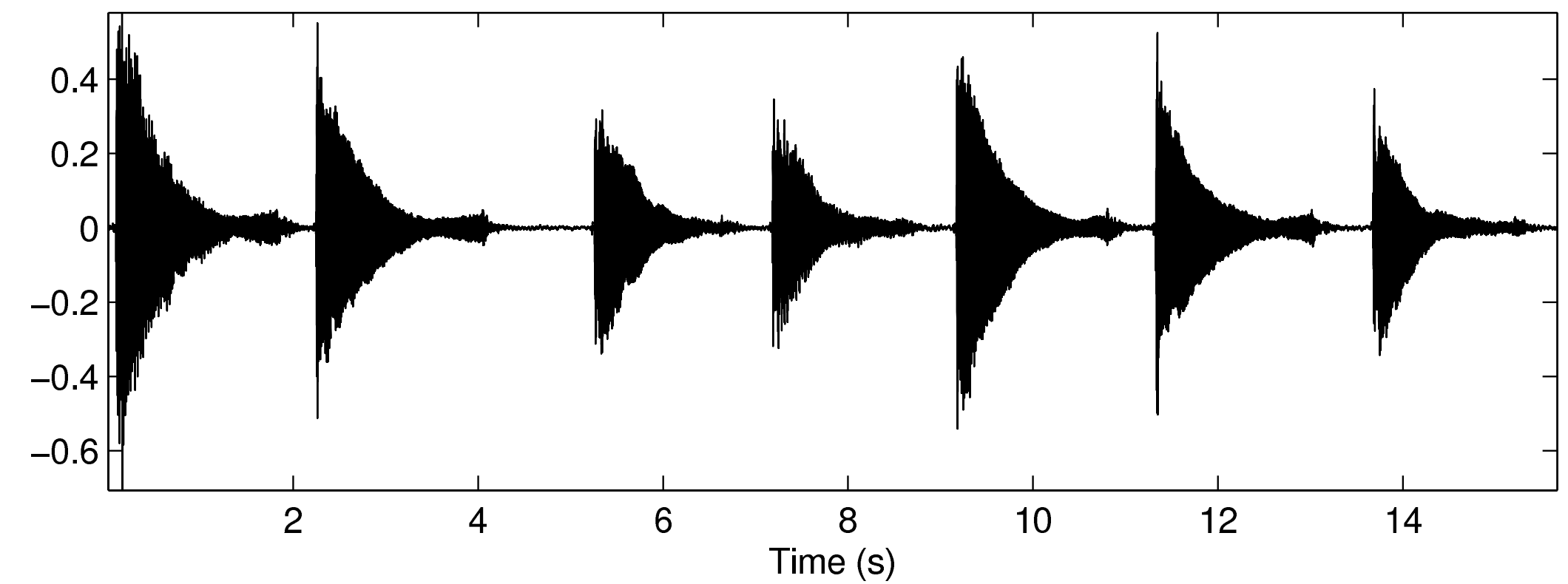
# NMF EXAMPLES

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

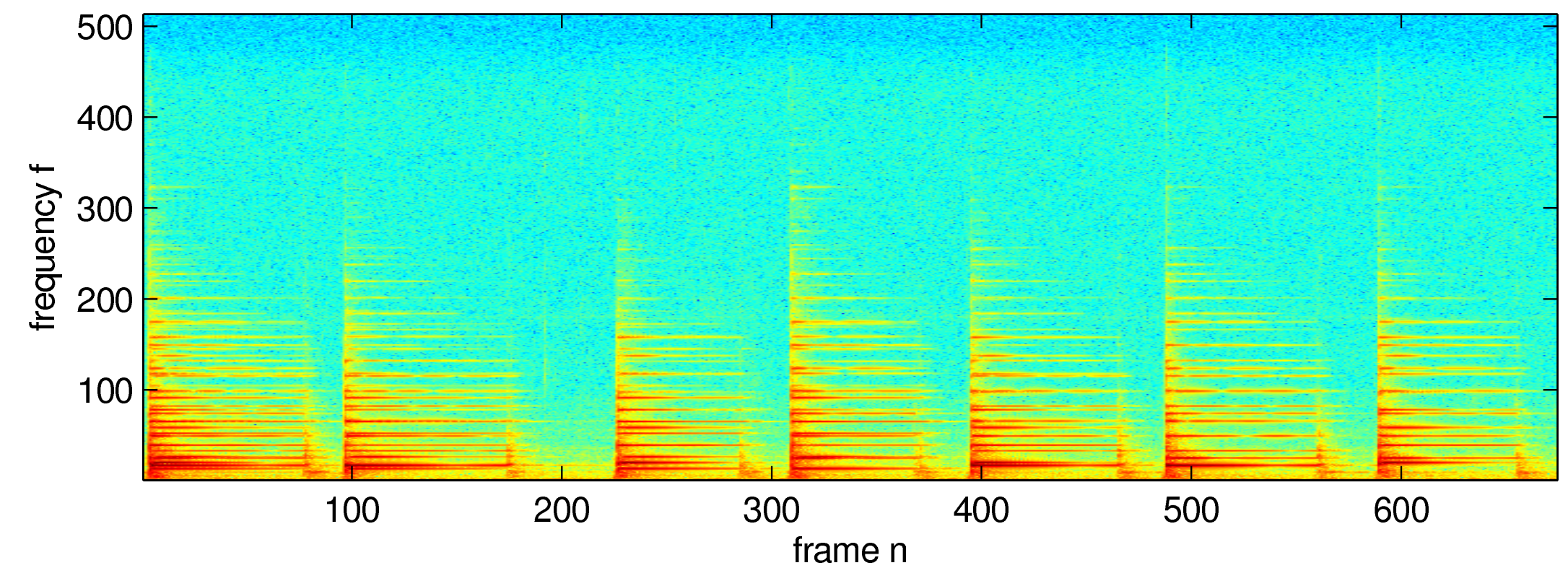


(MIDI numbers : 61, 65, 68, 72)

Signal x



Log-power spectrogram



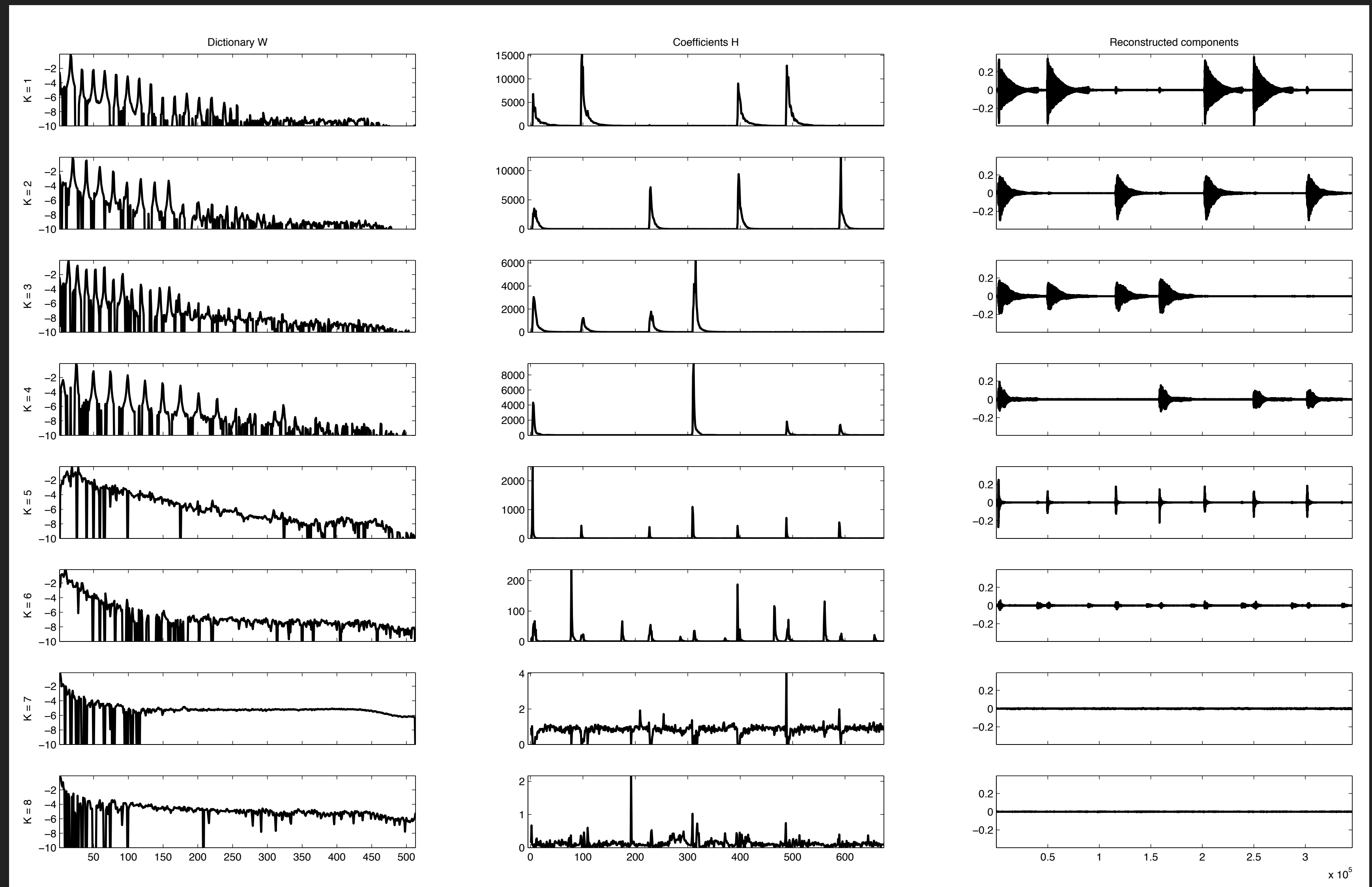


# NMF EXAMPLES

▶ Piano example from [Fevotte 2009]

▶ Demo available at:

[https://www.irit.fr/~Cedric.Fevotte/extras/machine\\_audition/](https://www.irit.fr/~Cedric.Fevotte/extras/machine_audition/)



# NMF DECOMPOSITIONS

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

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

- ▶  $D(X | WH) = \sum_{n,k} d(X[n, k] | \{WH\}[n, k])$  , 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)

## CONCLUSION

- ▶ Dictionary can be learned from the data
- ▶ Several model can be chosen (Sparse Coding, Convolutive Sparse Coding, NMF)
- ▶ One can "unroll" an algorithm to obtain a Neural Network