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:

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

- Where $\mathscr{R}$ 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 $X = (x_1, x_2, \ldots, x_M)$ 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 $A = (\alpha_1, \ldots, \alpha_m)$:

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

K-SVD

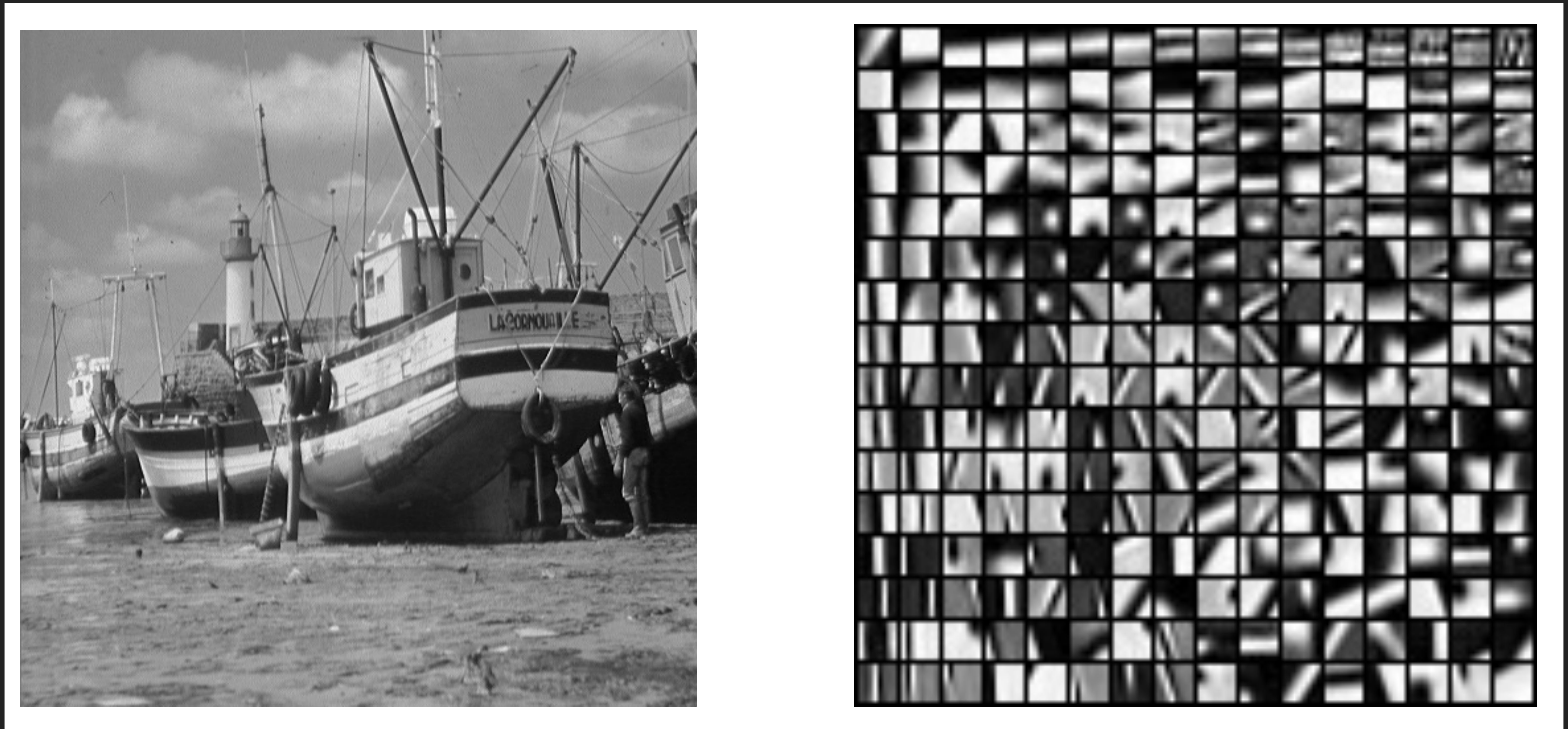
- [Elad and Aharon, 2006]
- Idea: alternate minimization between A and Φ
- Initialize Φ
- Sparse coding for each x_m: $\alpha_m = \underset{\alpha}{\operatorname{argmin}} \frac{1}{2}\|x_m - \Phi\alpha_m\|^2$ s.t. $\|\alpha\|_0 \leq s$
- Dictionary update : $\min \frac{1}{2}\|X - \Phi A\|^2$
- Can be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

DICTIONARY LEARNING: ALTERNATING MINIMIZATION

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

CONVOLUTIVE 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 $\{d_k\}_k$:

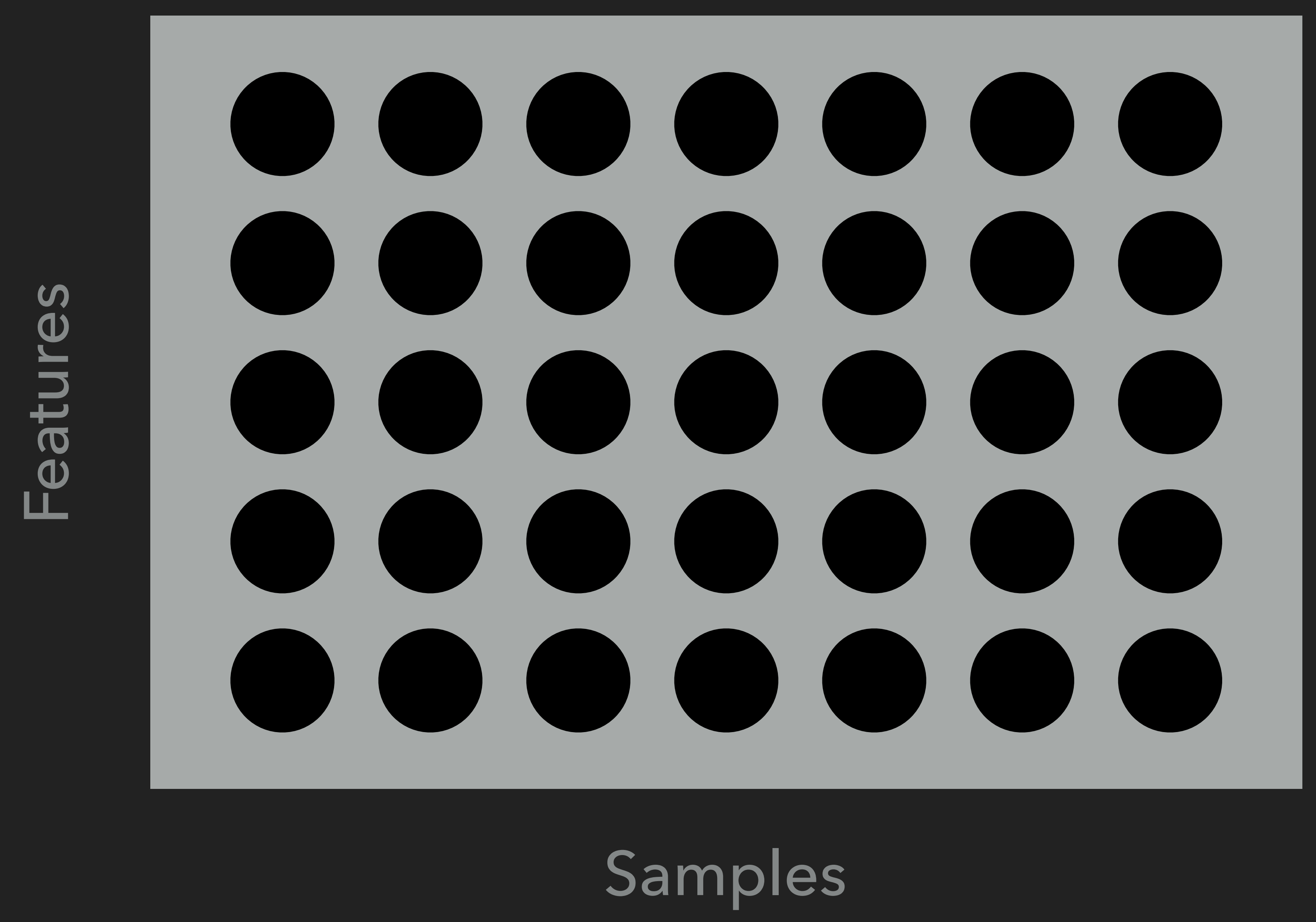
$$\min_{d_k, z_k^m} \sum_m \frac{1}{2} \|x^m - \sum_k d_k * z_k^m\|^2 + \lambda \|z_k\|_1$$

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

MATRIX FACTORIZATION

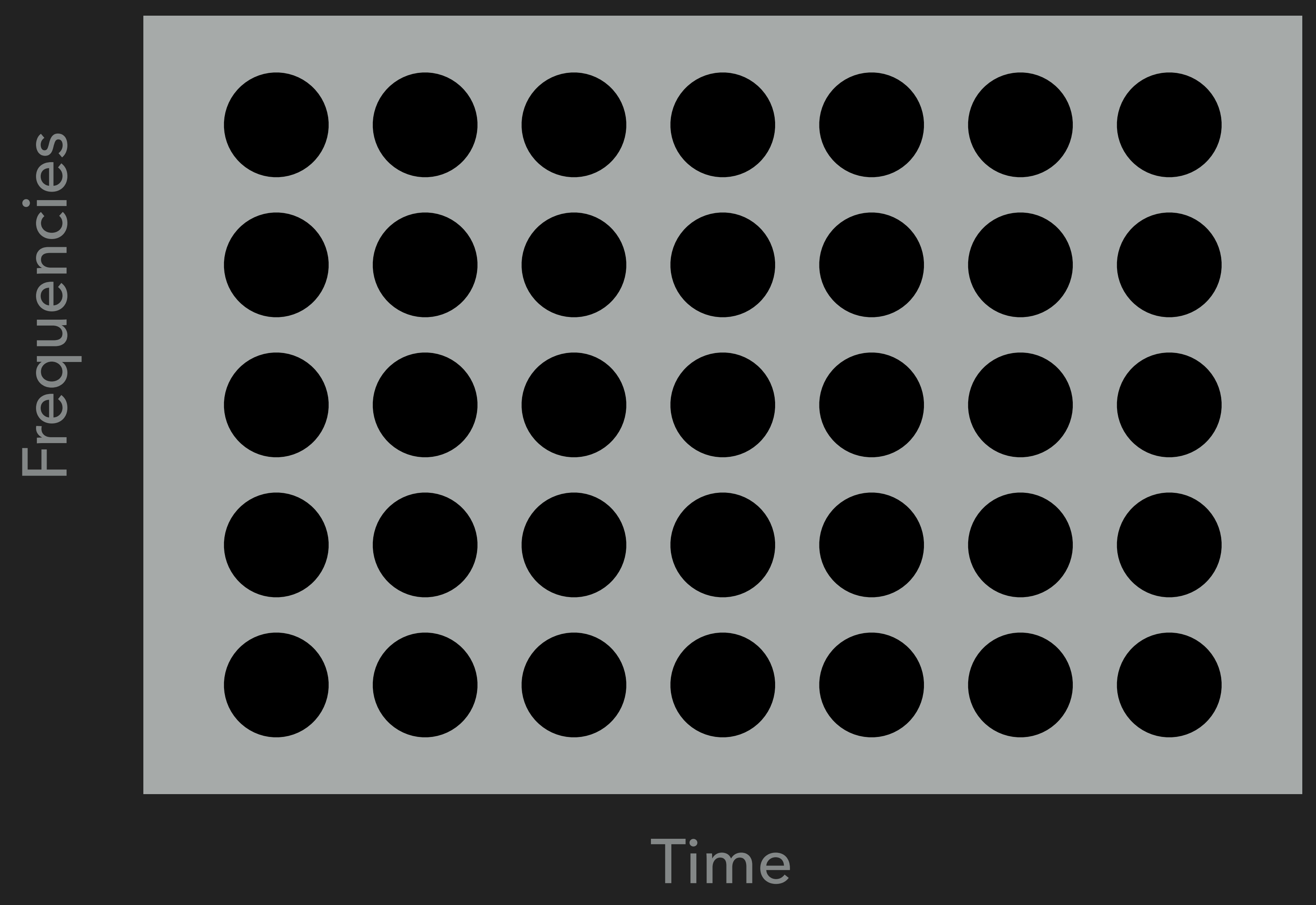
MATRIX FACTORIZATION

- Data are often available under a matrix form:



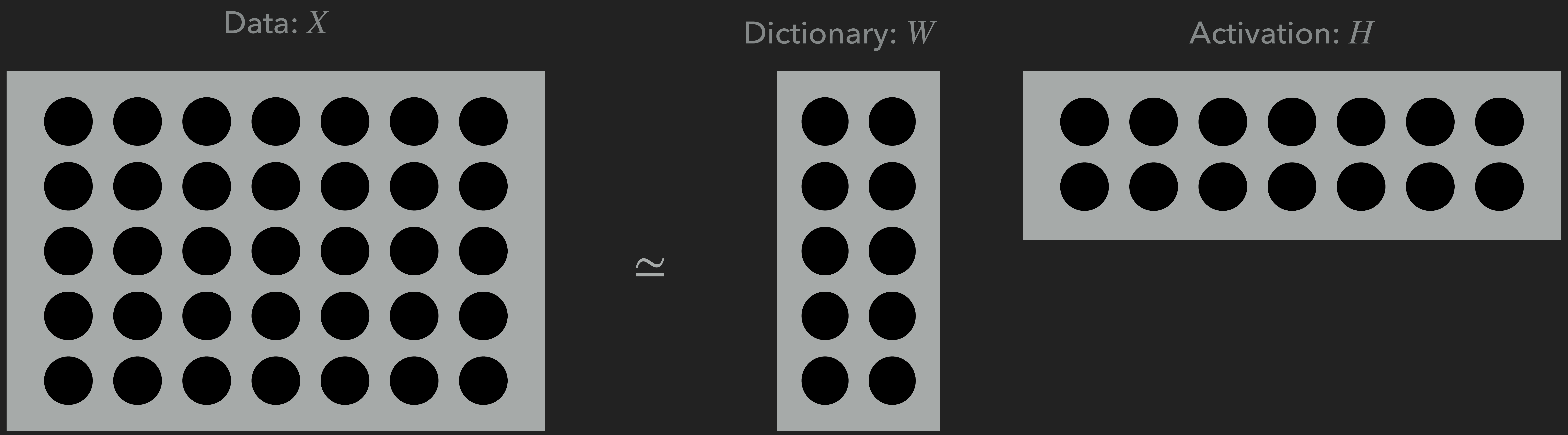
MATRIX FACTORIZATION

- Exemple: time-frequency representation



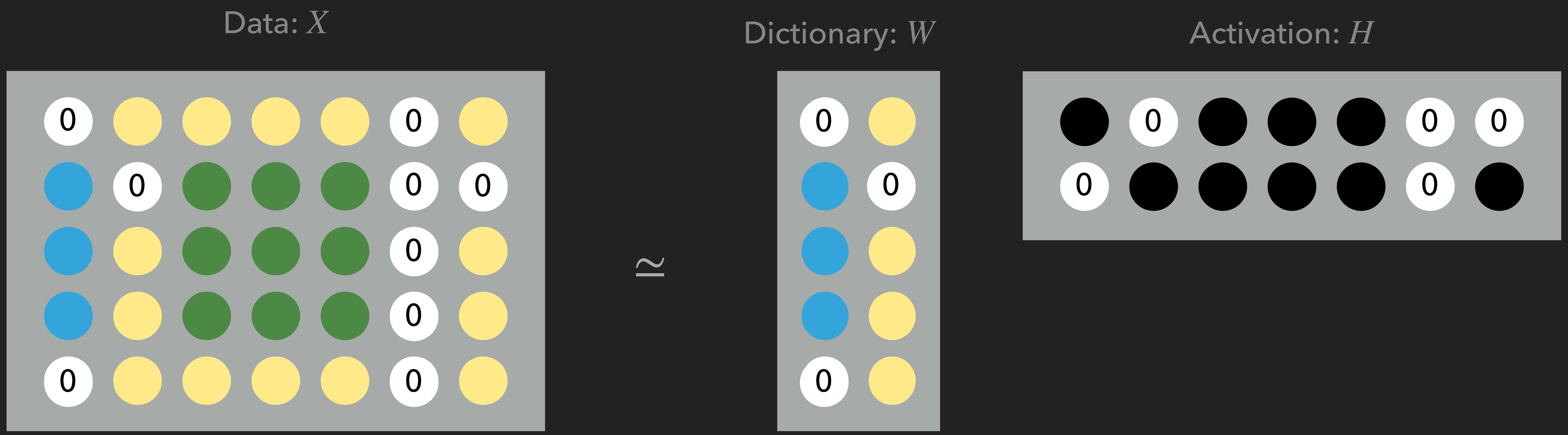
MATRIX FACTORIZATION

- Let $X \in \mathbb{R}^{MN}$. We 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

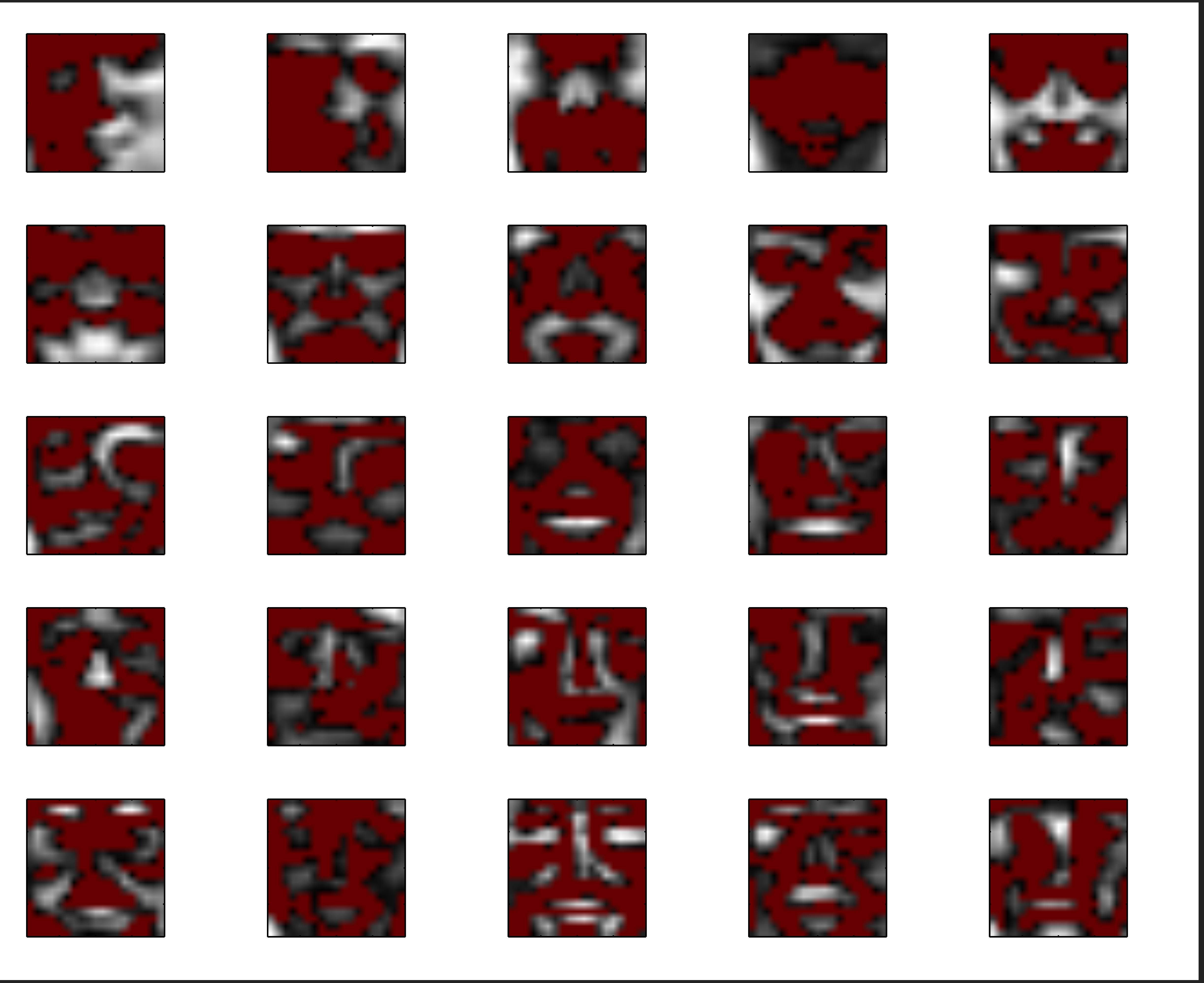
- $X \in \mathbb{R}_+^{MN}$ a matrix with non negative entries
- $X = WH$ with $W \in \mathbb{R}_+^{MK}$ and $H \in \mathbb{R}_+^{KN}$
- W is a matrix of non negative features (exemples: spectrum)
- H is a mature of non negative coefficients
- The Features can only add each other (no soustraction)
-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

NMF EXAMPLES

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

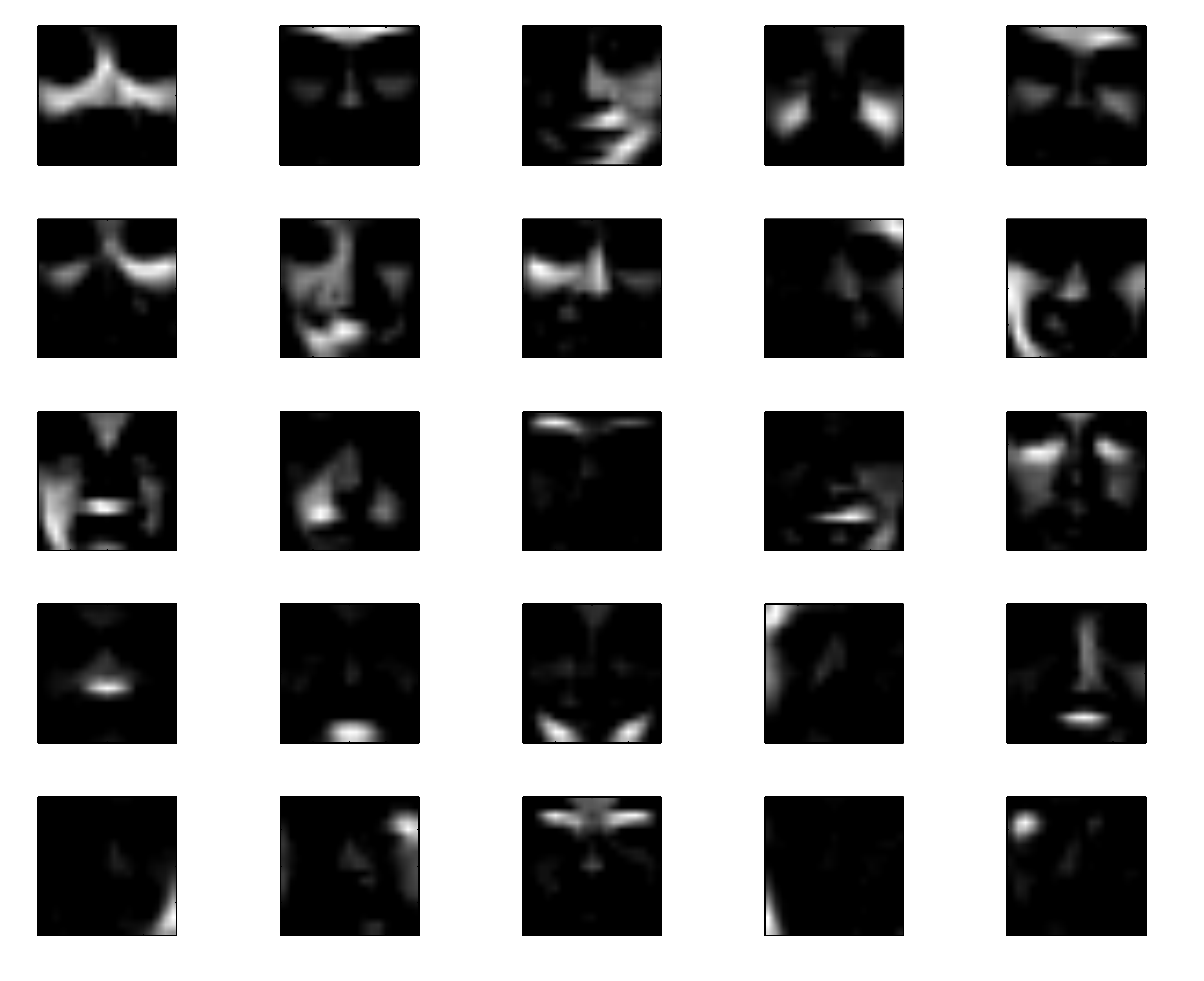
NMF EXAMPLES

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



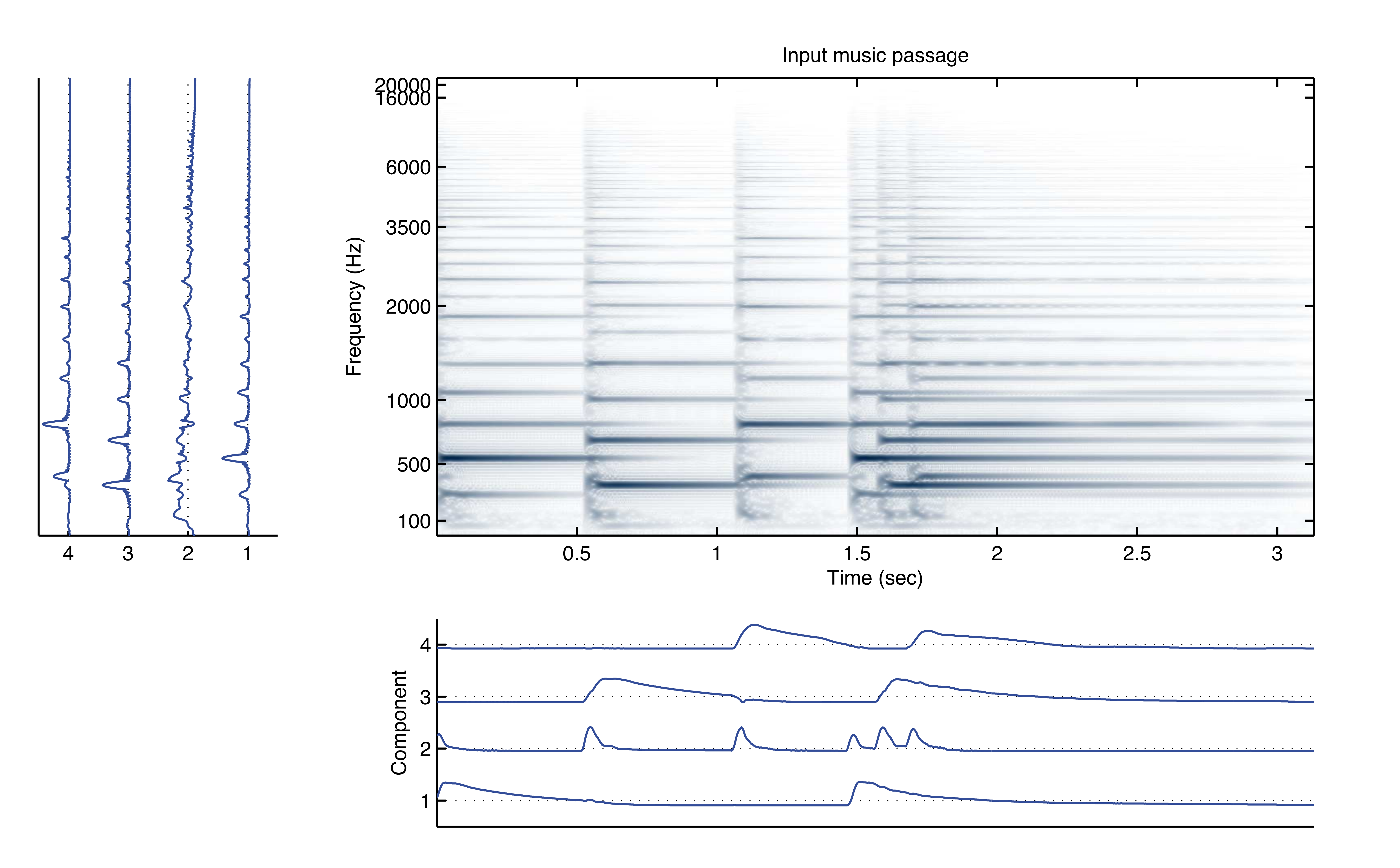
NMF EXAMPLES

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



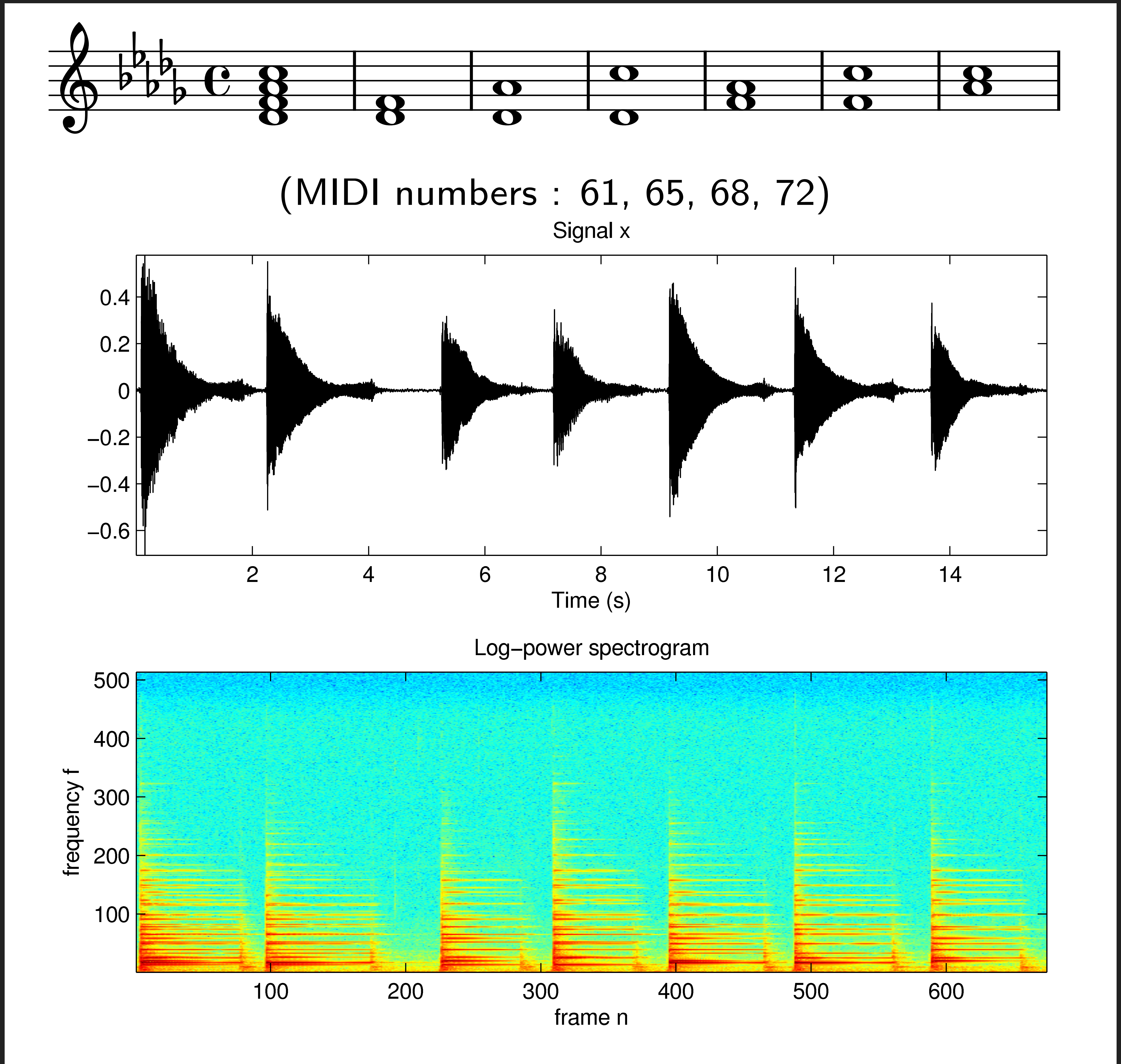
NMF EXAMPLE

- From [Smaragdis 2013]
- Spectral unmixing



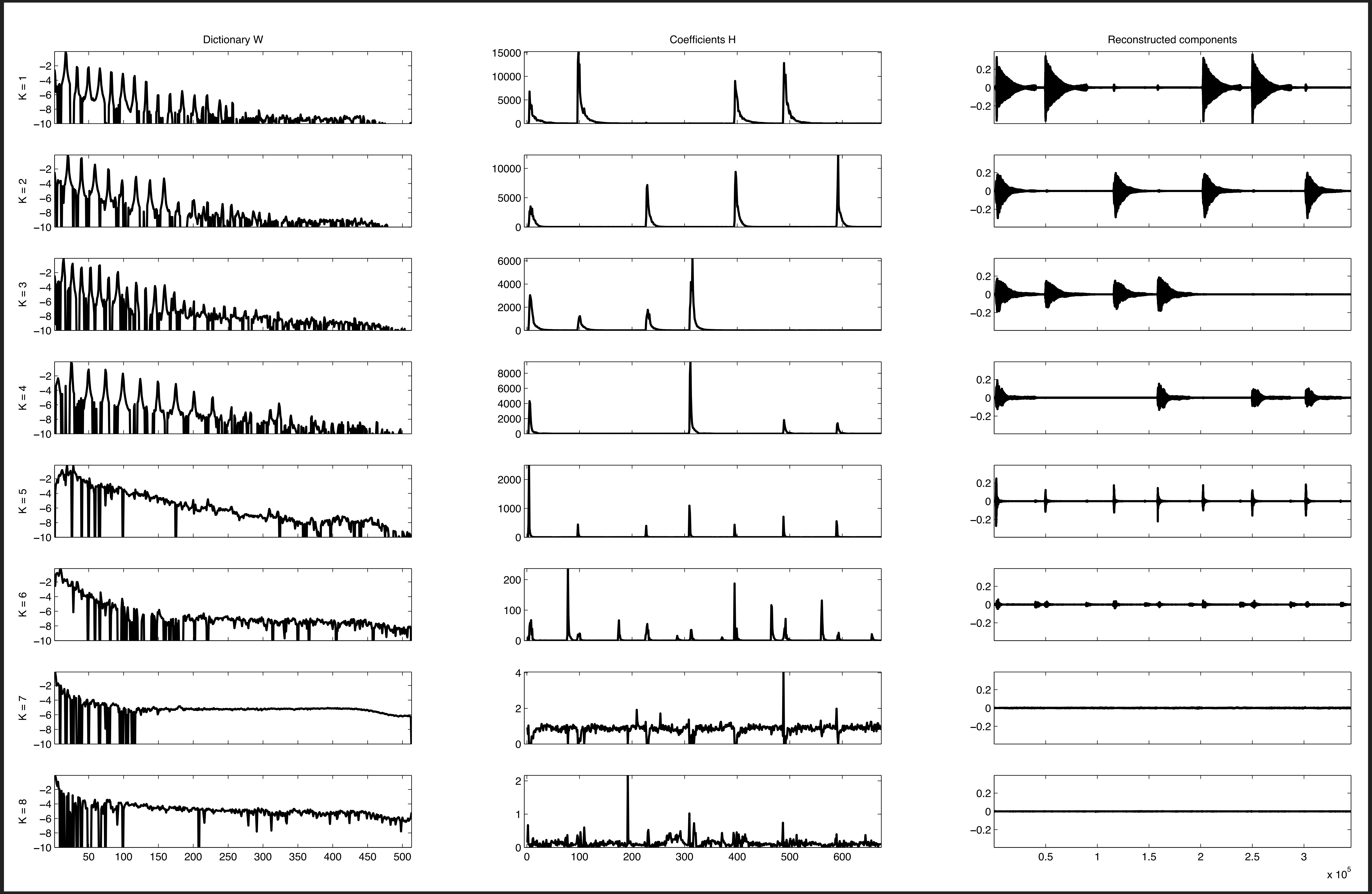
NMF EXAMPLES

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



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 :

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

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