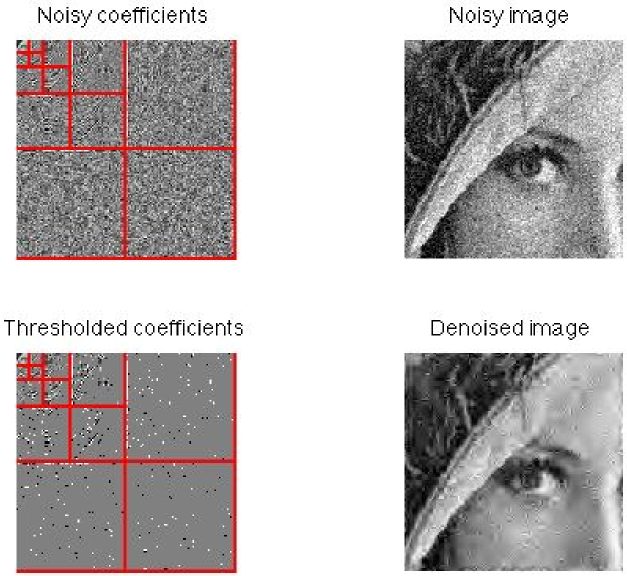
Dictionary learning

Louis MARTIN

Why do we want sparsity ?

- Compression
- Denoising

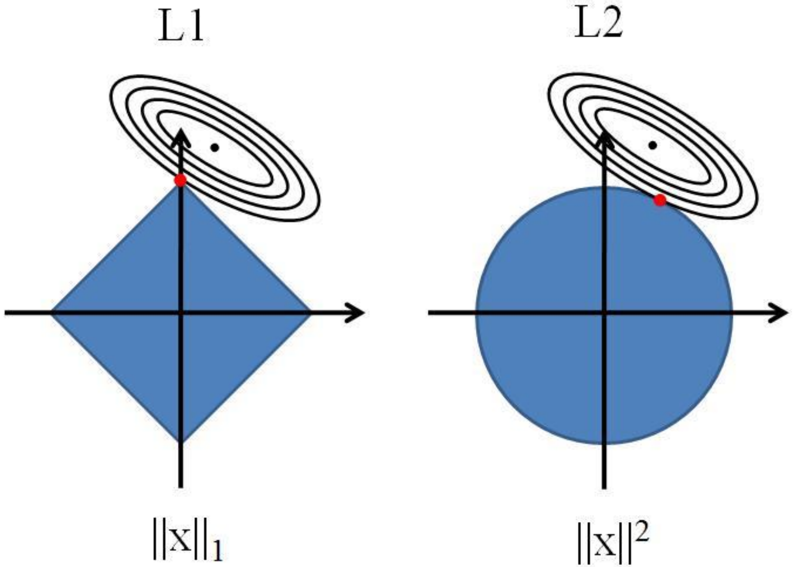


From G. Peyré

Why do we want sparsity ?

- Inpainting

How to measure sparsity ?

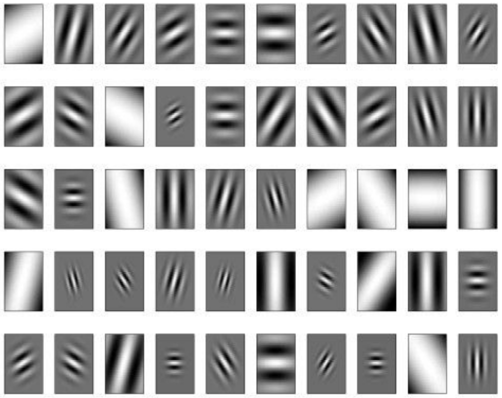


Predefined basis or learned dictionary ?

Predefined:

- Fourier
- Wavelet

Universal & standard

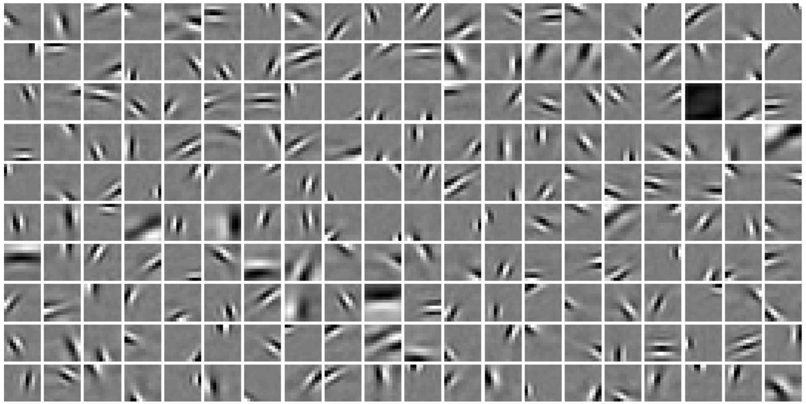


Learned:

- Dictionary

Adapted to the data

Overcomplete - Highly correlated



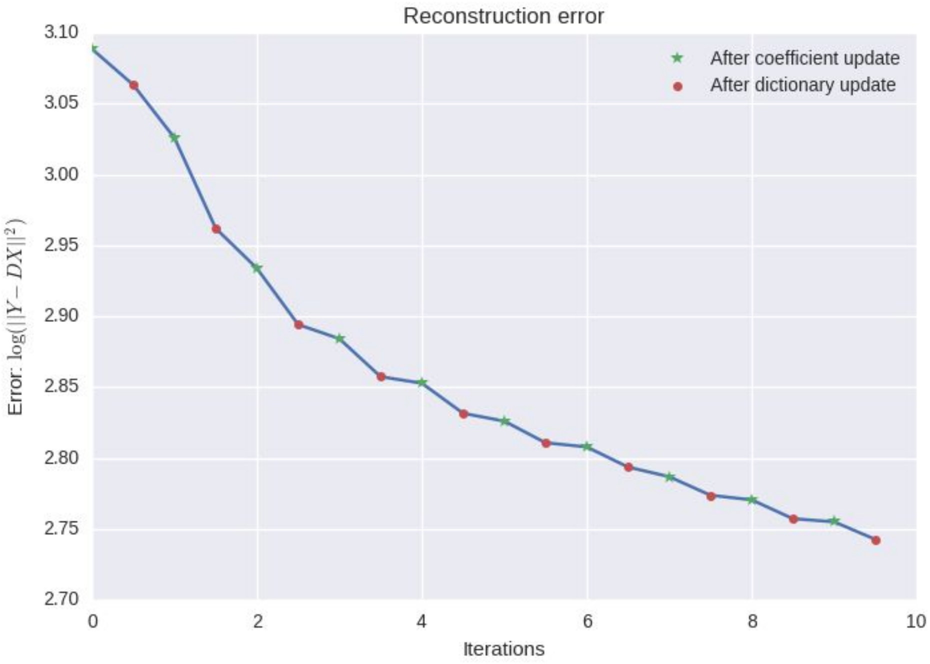
Dictionary learning

$$\min_{D \in \mathcal{D}, X \in \chi_k} E(X, D) = \min_{D \in \mathcal{D}, X \in \chi_k} \|Y - DX\|_F^2$$

Often not jointly solvable !

Two step approach:

- Sparse coding
- Dictionary update



Three algorithms

- **Online matrix factorization**
  - Adapted for large datasets
  - Column by column update

$$
\begin{aligned}
D_t &= \arg\min_{D\in\mathcal{D}} \frac{1}{t}(\frac{1}{2}\|Y_t - DX_t\|_F^2 + \lambda\|X_t\|_{1,1}) \\
&= \arg\min_{D\in\mathcal{D}} \frac{1}{t}(\frac{1}{2}Tr(D^T DX_t X_t^T) - Tr(D^T Y_t X_t^T)) \\
&= \arg\min_{D\in\mathcal{D}} \frac{1}{t}(\frac{1}{2}Tr(D^T DA_t) - Tr(D^T B_t))
\end{aligned}
$$

We can decompose A_t and B_t as follows:

$$A_t = \sum_{i=1}^{t} x_i x_i^T \text{ and } B_t = \sum_{i=1}^{t} y_i x_i^T$$

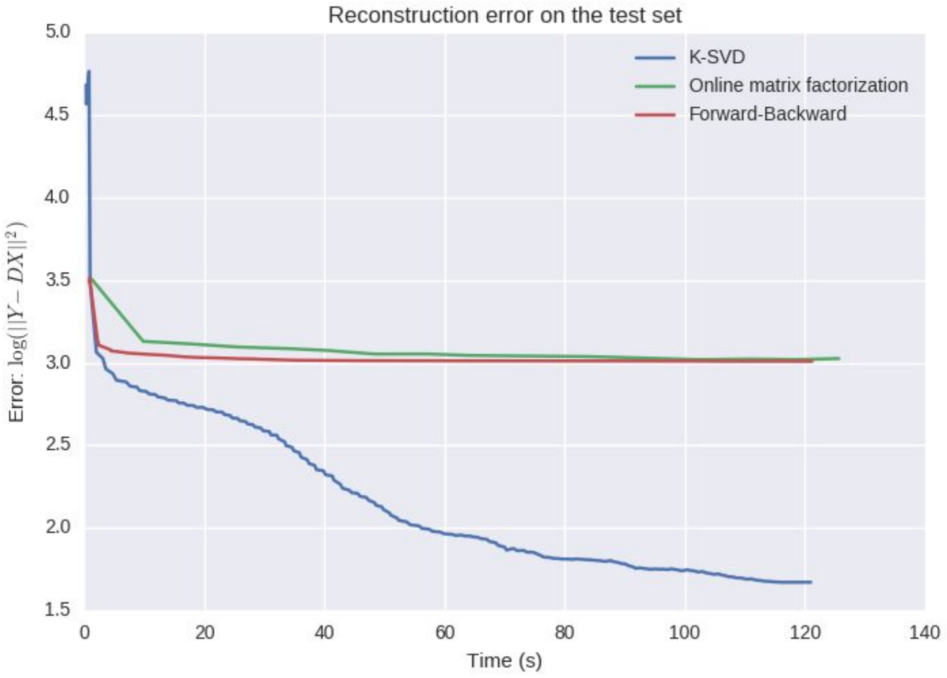
- **K-SVD**
  - Inspired from k-means
  - Fast selective updates

$$
\begin{aligned}
E(X, D) &= \|Y - DX\|_F^2 \\
&= \|Y - \sum_{j=1}^{p} d_j x^j\| \\
&= \|(Y - \sum_{j\neq k} d_j x^j) - d_k x^k\|
\end{aligned}
$$

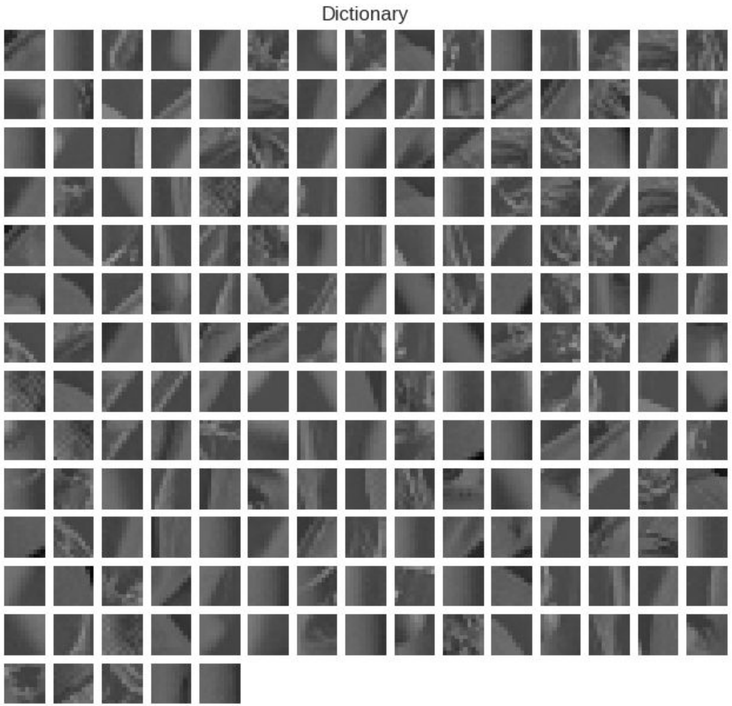
- **Forward-Backward**
  - Proximal method
  - Intuitive projected gradient descent

$$
\begin{aligned}
x^* \in \arg\min_x F(x) + G(x) &\iff 0 \in \nabla F(x^*) + \partial G(x^*) \\
&\iff (x^* - \gamma\nabla F(x^*)) \in x^* + \gamma\partial G(x^*) \\
&\iff x^* = Prox_{\gamma G}(x^* - \gamma\nabla F(x^*))
\end{aligned}
$$

Three algorithms



Application to natural image



Can be applied to denoising and inpainting.

Many oriented edges

Thank you for listening !