Dictionary Learning - Algorithms

EN.580.709 - Fall 2019

In search for sparse representations

$$\min_{\boldsymbol{\gamma}} \|\mathbf{y} - \mathbf{D}\boldsymbol{\gamma}\|_2^2 \text{ s.t. } \|\boldsymbol{\gamma}\|_0 \leq k$$

... but for what *dictionary* $\mathbf{D}$?

In search for sparse representations

$$\min_{\boldsymbol{\gamma}} \|\mathbf{y} - \mathbf{D}\boldsymbol{\gamma}\|_2^2 \text{ s.t. } \|\boldsymbol{\gamma}\|_0 \leq k$$

... but for what *dictionary* $\mathbf{D}$?

Dictionary Learning

Given a distribution of signals, $\mathcal{D}$, find a dictionary $\mathbf{D}$ so that

$$\min_{\mathbf{D}} \mathop{\mathbb{E}}_{\mathbf{y}\sim\mathcal{D}} \|\mathbf{y} - \mathbf{D}\boldsymbol{\gamma}^*\|_2^2,$$

$$\boldsymbol{\gamma}^* = \arg\min_{\boldsymbol{\gamma}} \|\mathbf{y} - \mathbf{D}\boldsymbol{\gamma}\|_2^2 \text{ s.t. } \|\boldsymbol{\gamma}\|_0 \leq k$$

In search for sparse representations

$$\min_{\boldsymbol{\gamma}} \|\mathbf{y} - \mathbf{D}\boldsymbol{\gamma}\|_2^2 \text{ s.t. } \|\boldsymbol{\gamma}\|_0 \leq k$$

... but for what *dictionary* $\mathbf{D}$?

Dictionary Learning

Given a distribution of signals, $\mathcal{D}$, find a dictionary $\mathbf{D}$ so that

$$\min_{\mathbf{D}} \mathop{\mathbb{E}}_{\mathbf{y}\sim\mathcal{D}} \|\mathbf{y} - \mathbf{D}\boldsymbol{\gamma}^*\|_2^2,$$

$$\boldsymbol{\gamma}^* = \arg\min_{\boldsymbol{\gamma}} \|\mathbf{y} - \mathbf{D}\boldsymbol{\gamma}\|_2^2 \text{ s.t. } \|\boldsymbol{\gamma}\|_0 \leq k$$

Given a set of training samples, $\{\mathbf{y}_i\}_{i=1}^N = \mathbf{Y} \in \mathbb{R}^{n\times N}$,

$$\min_{\mathbf{D},\boldsymbol{\Gamma}} \|\mathbf{Y} - \mathbf{D}\boldsymbol{\Gamma}\|_2^2 \quad \text{s.t.} \quad \begin{cases} \|\boldsymbol{\gamma}_i\|_0 \leq k, \ \forall i \\ \|\mathbf{d}_j\|_2 = 1, \ \forall j \end{cases}$$

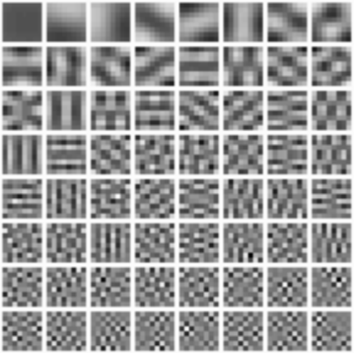
Physiological understanding of neurons in visual cortex

- **spatially localized** – each neuron was sensitive only to light in a particular region of the image
- **band-pass** – adding high-frequency components had a negligible effect on the response
- **oriented** – rotating images with sharp edges produced responses only when the edge was within some range of angles

Physiological understanding of neurons in visual cortex

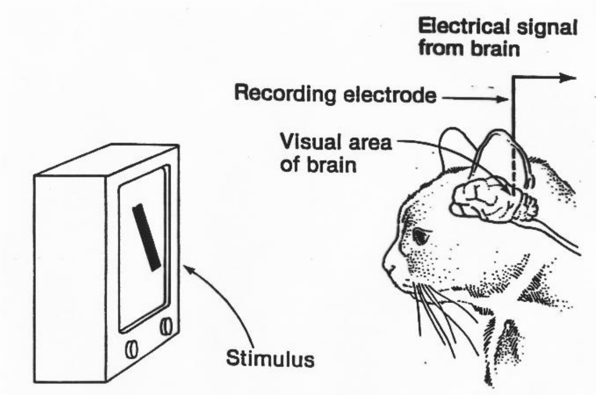
- **spatially localized** – each neuron was sensitive only to light in a particular region of the image
- **band-pass** – adding high-frequency components had a negligible effect on the response
- **oriented** – rotating images with sharp edges produced responses only when the edge was within some range of angles

What representations are being computed by these neurons?



Physiological understanding of neurons in visual cortex

Hubel and Weisel's Experiment



Video of Demo

Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1?

BRUNO A. OLSHAUSEN,‡ DAVID J. FIELD†

Received 16 July 1996; in revised form 24 December 1996

"We conjecture that natural images have "sparse structure" that is, any given image can be represented in terms of a small number of descriptors out of a large set "

Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1?

BRUNO A. OLSHAUSEN,‡ DAVID J. FIELD†

Received 16 July 1996; in revised form 24 December 1996

"We conjecture that natural images have "sparse structure" that is, any given image can be represented in terms of a small number of descriptors out of a large set "

$$\min_{\boldsymbol{\gamma}_i, \mathbf{D}} \sum_i \|\mathbf{y}_i - \mathbf{D}\boldsymbol{\gamma}_i\|_2^2 + \phi(\boldsymbol{\gamma}_i)$$

$$\phi(x) = \sum \log(1 + x^2)$$

Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1?

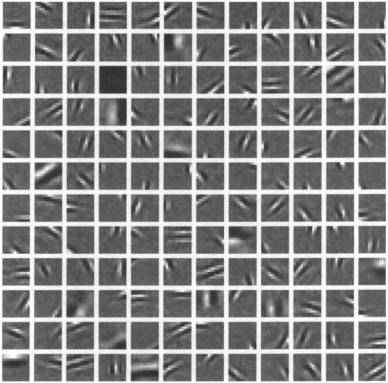
BRUNO A. OLSHAUSEN,‡ DAVID J. FIELD†

Received 16 July 1996; in revised form 24 December 1996

"We conjecture that natural images have "sparse structure" that is, any given image can be represented in terms of a small number of descriptors out of a large set "

$$\min_{\boldsymbol{\gamma}_i, \mathbf{D}} \sum_i \|\mathbf{y}_i - \mathbf{D}\boldsymbol{\gamma}_i\|_2^2 + \phi(\boldsymbol{\gamma}_i)$$

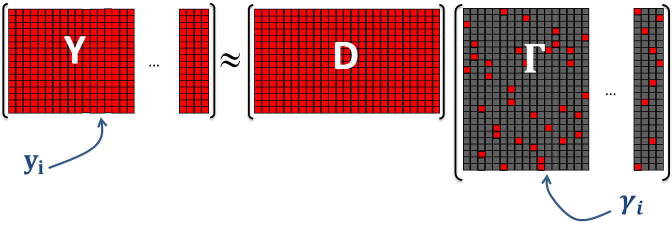
$$\phi(x) = \sum \log(1 + x^2)$$



Dictionary Learning

Given a set of training samples, $\{\mathbf{y}_i\}_{i=1}^{N} = \mathbf{Y} \in \mathbb{R}^{n \times N}$,

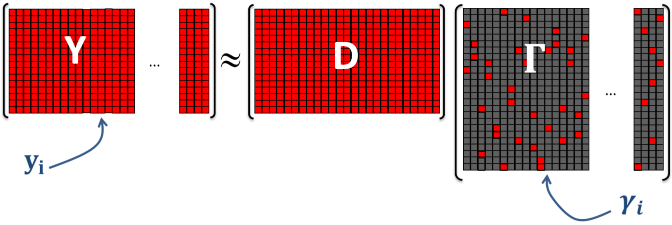
$$\min_{\mathbf{D},\boldsymbol{\Gamma}} \|\mathbf{Y} - \mathbf{D}\boldsymbol{\Gamma}\|_2^2 \quad \text{s.t.} \quad \begin{cases} \|\boldsymbol{\gamma}_i\|_0 \leq k, \ \forall i \\ \|\mathbf{d}_j\|_2 = 1, \ \forall j \end{cases}$$



Dictionary Learning

Given a set of training samples, $\{\mathbf{y}_i\}_{i=1}^N = \mathbf{Y} \in \mathbb{R}^{n\times N}$,

$$\min_{\mathbf{D},\mathbf{\Gamma}} \|\mathbf{Y} - \mathbf{D}\mathbf{\Gamma}\|_2^2 \quad \text{s.t.} \quad \begin{cases} \|\boldsymbol{\gamma}_i\|_0 \le k, \ \forall i \\ \|\mathbf{d}_j\|_2 = 1, \ \forall j \end{cases}$$



Alternating Minimization

1. Fix $\mathbf{D}$, and solve $\min_{\mathbf{\Gamma}} \|\mathbf{Y} - \mathbf{D}\mathbf{\Gamma}\|_2^2$ s.t. $\|\mathbf{\Gamma}_i\|_0 \le k \ \forall \ i$
2. Fix $\mathbf{\Gamma}$, and solve $\min_{\mathbf{D}} \|\mathbf{Y} - \mathbf{D}\mathbf{\Gamma}\|_2^2$ s.t. $\|\mathbf{d}_j\|_2 = 1 \ \forall \ j$

Method of Optimal Directions (MOD) [Engan, Rao, Kreutz-Delgado '99]

$$\min_{\mathbf{D},\mathbf{\Gamma}} \|\mathbf{Y} - \mathbf{D}\mathbf{\Gamma}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{\Gamma}_i\|_0 \leq k \;\forall\; i$$

1. Fix $\mathbf{D}^k$, and solve $\min_{\mathbf{\Gamma}} \|\mathbf{Y} - \mathbf{D}^k\mathbf{\Gamma}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{\Gamma}_i\|_0 \leq k \;\forall\; i$ e.g., w/OMP

Method of Optimal Directions (MOD) [Engan, Rao, Kreutz-Delgado '99]

$$\min_{\mathbf{D},\boldsymbol{\Gamma}} \|\mathbf{Y} - \mathbf{D}\boldsymbol{\Gamma}\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{\Gamma}_i\|_0 \leq k \ \forall \ i$$

1. Fix $\mathbf{D}^k$, and solve $\min_{\boldsymbol{\Gamma}} \|\mathbf{Y} - \mathbf{D}^k\boldsymbol{\Gamma}\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{\Gamma}_i\|_0 \leq k \ \forall \ i$ e.g., w/OMP
2. Fix $\boldsymbol{\Gamma}$,

$$\min_{\mathbf{D}} \|\mathbf{Y} - \mathbf{D}\boldsymbol{\Gamma}\|_2^2$$

Method of Optimal Directions (MOD) [Engan, Rao, Kreutz-Delgado '99]

$$\min_{\mathbf{D},\mathbf{\Gamma}} \|\mathbf{Y} - \mathbf{D}\mathbf{\Gamma}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{\Gamma}_i\|_0 \le k \ \forall \ i$$

1. Fix $\mathbf{D}^k$, and solve $\min_{\mathbf{\Gamma}} \|\mathbf{Y} - \mathbf{D}^k\mathbf{\Gamma}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{\Gamma}_i\|_0 \le k \ \forall \ i$ e.g., w/OMP
2. Fix $\mathbf{\Gamma}$,

$$\min_{\mathbf{D}} \|\mathbf{Y} - \mathbf{D}\mathbf{\Gamma}\|_2^2$$

Method of Optimal Directions (MOD) [Engan, Rao, Kreutz-Delgado '99]

$$\min_{\mathbf{D},\mathbf{\Gamma}} \|\mathbf{Y} - \mathbf{D}\mathbf{\Gamma}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{\Gamma}_i\|_0 \leq k \ \forall \ i$$

1. Fix $\mathbf{D}^k$, and solve $\min_{\mathbf{\Gamma}} \|\mathbf{Y} - \mathbf{D}^k\mathbf{\Gamma}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{\Gamma}_i\|_0 \leq k \ \forall \ i$ e.g., w/OMP
2. Fix $\mathbf{\Gamma}$,

$$\min_{\mathbf{D}} \|\mathbf{Y} - \mathbf{D}\mathbf{\Gamma}\|_2^2$$

$$(\mathbf{D}\mathbf{\Gamma} - \mathbf{Y})\mathbf{\Gamma}^T = 0$$

$$\mathbf{D}^{k+1} = \mathbf{Y}\mathbf{\Gamma}^T(\mathbf{\Gamma}\mathbf{\Gamma}^T)^{-1}$$

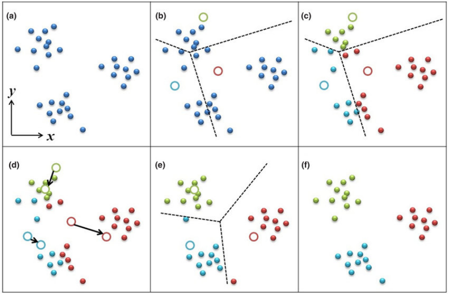
K-SVD [Aharon, Bruckstein, Elad, '06]

K-SVD [Aharon, Bruckstein, Elad, '06]

Recall K-means

Given observations $\{\mathbf{y}_i\}_{i=1}^{N}$, partition them in k sets $S = \{S_1, \ldots, S_k\}$ so as to

$$\min_{S} \sum_{i=1}^{k} \sum_{\mathbf{y}_i \in S_j} \|\mathbf{y}_i - \mathbf{d}_j\|_2^2$$

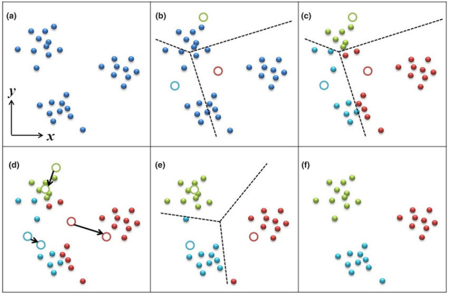


K-SVD [Aharon, Bruckstein, Elad, '06]

Recall K-means

Given observations $\{\mathbf{y}_i\}_{i=1}^N$, partition them in k sets $S = \{S_1, \ldots, S_k\}$ so as to

$$\min_S \sum_{i=1}^k \sum_{\mathbf{y}_i \in S_j} \|\mathbf{y}_i - \mathbf{d}_j\|_2^2 = \min_{\mathbf{D},\mathbf{C}} \|\mathbf{Y} - \mathbf{DC}\|_2^2 \qquad \mathbf{C}_{i,j} = 1 \text{ if } \mathbf{y}_j \in S_i$$

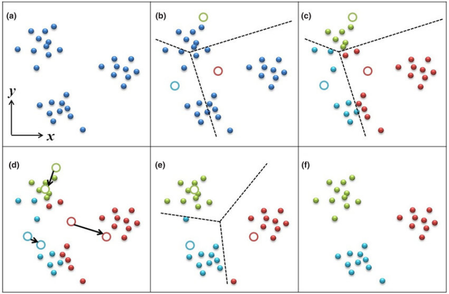


K-SVD [Aharon, Bruckstein, Elad, '06]

Recall K-means

Given observations $\{\mathbf{y}_i\}_{i=1}^N$, partition them in k sets $S = \{S_1, \ldots, S_k\}$ so as to

$$\min_S \sum_{i=1}^k \sum_{\mathbf{y}_i \in S_j} \|\mathbf{y}_i - \mathbf{d}_j\|_2^2 = \min_{\mathbf{D},\mathbf{C}} \|\mathbf{Y} - \mathbf{DC}\|_2^2 \qquad \mathbf{C}_{i,j} = 1 \text{ if } \mathbf{y}_j \in S_i$$



1. Assignment: each observation is assigned to the cluster with its closest mean

$$\mathbf{C}_{i,j} = 1 \text{ if}$$

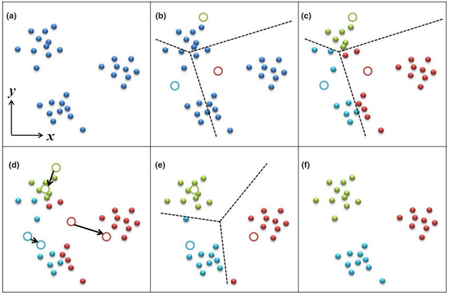
$$\|\mathbf{y}_j - \mathbf{d}_i\|_2^2 \leq \|\mathbf{y}_j - \mathbf{d}_k\|_2^2 \; \forall \; k \neq i$$

K-SVD [Aharon, Bruckstein, Elad, '06]

Recall K-means

Given observations $\{\mathbf{y}_i\}_{i=1}^{N}$, partition them in k sets $S = \{S_1, \dots, S_k\}$ so as to

$$\min_{S} \sum_{i=1}^{k} \sum_{\mathbf{y}_i \in S_j} \|\mathbf{y}_i - \mathbf{d}_j\|_2^2 = \min_{\mathbf{D},\mathbf{C}} \|\mathbf{Y} - \mathbf{DC}\|_2^2 \qquad \mathbf{C}_{i,j} = 1 \text{ if } \mathbf{y}_j \in S_i$$



1. Assignment: each observation is assigned to the cluster with its closest mean

$$\mathbf{C}_{i,j} = 1 \text{ if}$$

$$\|\mathbf{y}_j - \mathbf{d}_i\|_2^2 \le \|\mathbf{y}_j - \mathbf{d}_k\|_2^2 \ \forall\ k \neq i$$

2. Update: calculate new centroids

$$\mathbf{d}_j = \frac{1}{|S_j|} \sum_{\mathbf{y}_i \in S_j} \mathbf{y}_i$$

K-SVD [Aharon, Bruckstein, Elad, '06]

Back to Dictionary Learning

K-SVD [Aharon, Bruckstein, Elad, '06]

Back to Dictionary Learning

1. Assignment/Sparse Coding

$$\mathbf{\Gamma} \leftarrow \min_{\mathbf{\Gamma}} \|\mathbf{Y} - \mathbf{D}\mathbf{\Gamma}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{\Gamma}_i\|_0 \leq k \ \forall\ i \ \ \text{(OMP)}$$

K-SVD [Aharon, Bruckstein, Elad, '06]

Back to Dictionary Learning

1. Assignment/Sparse Coding

$$\mathbf{\Gamma} \leftarrow \min_{\mathbf{\Gamma}} \|\mathbf{Y} - \mathbf{D}\mathbf{\Gamma}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{\Gamma}_i\|_0 \leq k \; \forall \; i \;\; \text{(OMP)}$$

2. Optimize centroids/atoms sequentially, keeping others fixed

$$\mathbf{d}_j = \arg\min_{\mathbf{d}_j} \|\mathbf{Y} - \sum_i \mathbf{d}_i \mathbf{\Gamma}_i^T\|_2^2$$

K-SVD [Aharon, Bruckstein, Elad, '06]

Back to Dictionary Learning

1. Assignment/Sparse Coding

$$\mathbf{\Gamma} \leftarrow \min_{\mathbf{\Gamma}} \|\mathbf{Y} - \mathbf{D}\mathbf{\Gamma}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{\Gamma}_i\|_0 \leq k \;\forall\; i \;\; \text{(OMP)}$$

2. Optimize centroids/atoms sequentially, keeping others fixed

$$\begin{aligned}\mathbf{d}_j &= \arg\min_{\mathbf{d}_j} \|\mathbf{Y} - \sum_i \mathbf{d}_i \mathbf{\Gamma}_i^T\|_2^2 \\ &= \arg\min_{\mathbf{d}_j} \|\mathbf{Y} - \sum_{i \neq j} \mathbf{d}_i \mathbf{\Gamma}_i^T - \mathbf{d}_j \mathbf{\Gamma}_j^T\|_2^2\end{aligned}$$

K-SVD [Aharon, Bruckstein, Elad, '06]

Back to Dictionary Learning

1. Assignment/Sparse Coding

$$\mathbf{\Gamma} \leftarrow \min_{\mathbf{\Gamma}} \|\mathbf{Y} - \mathbf{D}\mathbf{\Gamma}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{\Gamma}_i\|_0 \leq k \;\forall\; i \;\; \text{(OMP)}$$

2. Optimize centroids/atoms sequentially, keeping others fixed

$$\begin{aligned}
\mathbf{d}_j &= \arg\min_{\mathbf{d}_j} \|\mathbf{Y} - \sum_i \mathbf{d}_i \mathbf{\Gamma}_i^T\|_2^2 \\
&= \arg\min_{\mathbf{d}_j} \|\mathbf{Y} - \sum_{i \neq j} \mathbf{d}_i \mathbf{\Gamma}_i^T - \mathbf{d}_j \mathbf{\Gamma}_j^T\|_2^2 \\
&= \arg\min_{\mathbf{d}_j} \|\mathbf{E}_j - \mathbf{d}_j \mathbf{\Gamma}_j^T\|_2^2
\end{aligned}$$

K-SVD [Aharon, Bruckstein, Elad, '06]

Back to Dictionary Learning

1. Assignment/Sparse Coding

$$\boldsymbol{\Gamma} \leftarrow \min_{\boldsymbol{\Gamma}} \|\mathbf{Y} - \mathbf{D}\boldsymbol{\Gamma}\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{\Gamma}_i\|_0 \leq k \; \forall \; i \;\; \text{(OMP)}$$

2. Optimize centroids/atoms sequentially, keeping others fixed

$$\begin{aligned}
\mathbf{d}_j &= \arg\min_{\mathbf{d}_j} \|\mathbf{Y} - \sum_i \mathbf{d}_i \boldsymbol{\Gamma}_i^T\|_2^2 \\
&= \arg\min_{\mathbf{d}_j} \|\mathbf{Y} - \sum_{i \neq j} \mathbf{d}_i \boldsymbol{\Gamma}_i^T - \mathbf{d}_j \boldsymbol{\Gamma}_j^T\|_2^2 \\
&= \arg\min_{\mathbf{d}_j} \|\mathbf{E}_j - \mathbf{d}_j \boldsymbol{\Gamma}_j^T\|_2^2 \\
&= \arg\min_{\mathbf{d}_j} \|\tilde{\mathbf{E}}_j - \mathbf{d}_j \tilde{\boldsymbol{\Gamma}}_j^T\|_2^2, \quad \tilde{\mathbf{E}}_j : n \times \|\boldsymbol{\Gamma}_j^T\|_0
\end{aligned}$$

K-SVD [Aharon, Bruckstein, Elad, '06]

Back to Dictionary Learning

1. Assignment/Sparse Coding

$$\mathbf{\Gamma} \leftarrow \min_{\mathbf{\Gamma}} \|\mathbf{Y} - \mathbf{D}\mathbf{\Gamma}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{\Gamma}_i\|_0 \leq k \ \forall \ i \ \ \text{(OMP)}$$

2. Optimize centroids/atoms sequentially, keeping others fixed

$$\begin{aligned}
\mathbf{d}_j &= \arg\min_{\mathbf{d}_j} \|\mathbf{Y} - \sum_i \mathbf{d}_i \mathbf{\Gamma}_i^T\|_2^2 \\
&= \arg\min_{\mathbf{d}_j} \|\mathbf{Y} - \sum_{i \neq j} \mathbf{d}_i \mathbf{\Gamma}_i^T - \mathbf{d}_j \mathbf{\Gamma}_j^T\|_2^2 \\
&= \arg\min_{\mathbf{d}_j} \|\mathbf{E}_j - \mathbf{d}_j \mathbf{\Gamma}_j^T\|_2^2 \\
&= \arg\min_{\mathbf{d}_j} \|\tilde{\mathbf{E}}_j - \mathbf{d}_j \tilde{\mathbf{\Gamma}}_j^T\|_2^2, \quad \tilde{\mathbf{E}}_j : n \times \|\mathbf{\Gamma}_j^T\|_0 \\
(\mathbf{d}_j, \tilde{\mathbf{\Gamma}}_j) &= (\mathbf{u}_1, \sigma_1 \mathbf{v}_1); \qquad \tilde{\mathbf{E}}_j = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T
\end{aligned}$$

K-SVD [Aharon, Bruckstein, Elad, '06]

- Prone to get stuck in local minima $\Rightarrow$ Extra regularization techniques

K-SVD [Aharon, Bruckstein, Elad, '06]

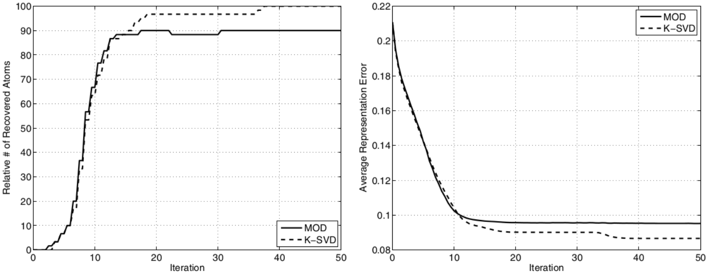
- Prone to get stuck in local minima $\Rightarrow$ Extra regularization techniques
- **unused atoms**: replace with better guesses

K-SVD [Aharon, Bruckstein, Elad, '06]

- Prone to get stuck in local minima $\Rightarrow$ Extra regularization techniques
- **unused atoms**: replace with better guesses
- **too similar atoms**: replace $\mathbf{d}_j$ if $\exists i : \; |\mathbf{d}_j^T \mathbf{d}_i| \geq \tau$, say, $\tau = 0.99$.

K-SVD [Aharon, Bruckstein, Elad, '06]

- Prone to get stuck in local minima $\Rightarrow$ Extra regularization techniques
- **unused atoms**: replace with better guesses
- **too similar atoms**: replace $\mathbf{d}_j$ if $\exists i : \ |\mathbf{d}_j^T \mathbf{d}_i| \geq \tau$, say, $\tau = 0.99$.

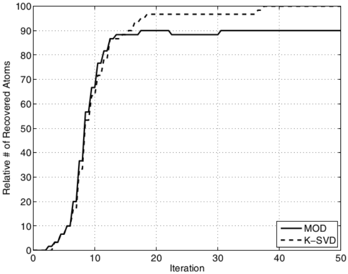


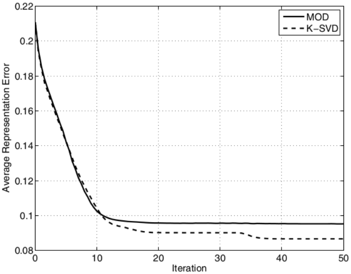
Uniqueness

Given $\mathbf{Y} = \mathbf{D}\boldsymbol{\Gamma}$, if (i) $\|\boldsymbol{\gamma}_i\|_0 = k <$ spark$(\mathbf{D})/2$,

K-SVD [Aharon, Bruckstein, Elad, '06]

- Prone to get stuck in local minima $\Rightarrow$ Extra regularization techniques
- **unused atoms**: replace with better guesses
- **too similar atoms**: replace $\mathbf{d}_j$ if $\exists i: \ |\mathbf{d}_j^T \mathbf{d}_i| \geq \tau$, say, $\tau = 0.99$.



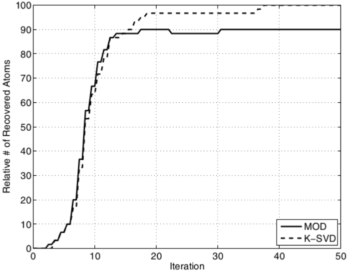


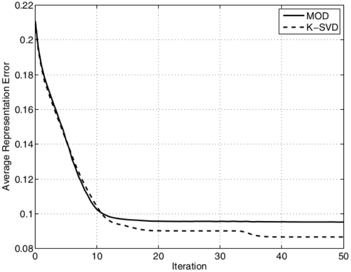
Uniqueness

Given $\mathbf{Y} = \mathbf{D\Gamma}$, if (i) $\|\boldsymbol{\gamma}_i\|_0 = k <$ spark$(\mathbf{D})/2$, (ii) at least $(k+1)$ signals $\mathbf{y}_i$ from every possibly combination of atoms,

K-SVD [Aharon, Bruckstein, Elad, '06]

- Prone to get stuck in local minima $\Rightarrow$ Extra regularization techniques
- **unused atoms**: replace with better guesses
- **too similar atoms**: replace $\mathbf{d}_j$ if $\exists i: |\mathbf{d}_j^T \mathbf{d}_i| \geq \tau$, say, $\tau = 0.99$.



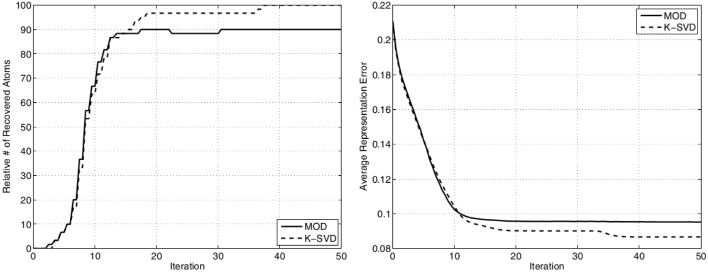


Uniqueness

Given $\mathbf{Y} = \mathbf{D}\mathbf{\Gamma}$, if (i) $\|\boldsymbol{\gamma}_i\|_0 = k <$ spark$(\mathbf{D})/2$, (ii) at least $(k+1)$ signals $\mathbf{y}_i$ from every possibly combination of atoms, (iii) "Non-degeneracy" of atoms,

K-SVD [Aharon, Bruckstein, Elad, '06]

- Prone to get stuck in local minima $\Rightarrow$ Extra regularization techniques
- **unused atoms**: replace with better guesses
- **too similar atoms**: replace $\mathbf{d}_j$ if $\exists i: \ |\mathbf{d}_j^T \mathbf{d}_i| \geq \tau$, say, $\tau = 0.99$.



Uniqueness

Given $\mathbf{Y} = \mathbf{D}\mathbf{\Gamma}$, if (i) $\|\boldsymbol{\gamma}_i\|_0 = k <$ spark$(\mathbf{D})/2$, (ii) at least $(k+1)$ signals $\mathbf{y}_i$ from every possibly combination of atoms, (iii) "Non-degeneracy" of atoms, then the factorization $\mathbf{Y} = \mathbf{D}\mathbf{\Gamma}$ is unique (up to permutations/scaling).

K-SVD [Aharon, Bruckstein, Elad, '06]

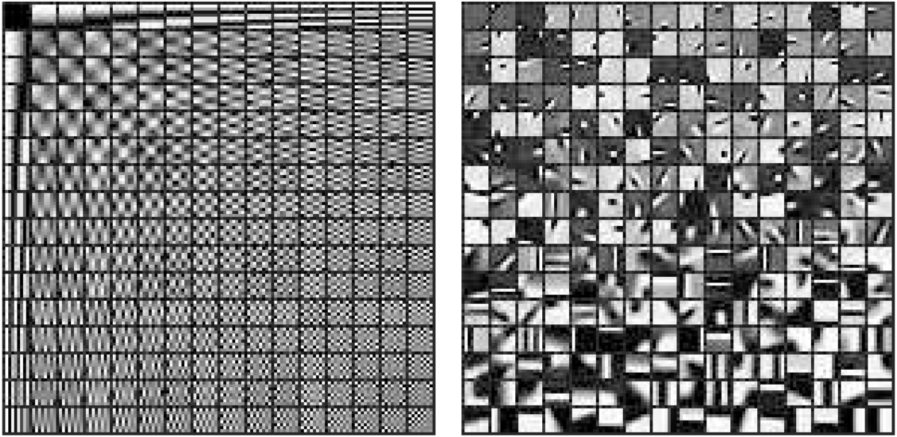


Fig. 2. Left: Overcomplete DCT dictionary. Right: Globally trained dictionary.

Stochastic Optimization

$$\min_{\mathbf{D}\in\mathcal{C}} \frac{1}{n}\sum_{1}^{n} L(\mathbf{y}_i, \mathbf{D})$$

where

$$L(\mathbf{y}_i, \mathbf{D}) \triangleq \min_{\boldsymbol{\gamma}} \frac{1}{2}\|\mathbf{y}_i - \mathbf{D}\boldsymbol{\gamma}\|_2^2 + \lambda\|\boldsymbol{\gamma}\|_1$$

Stochastic Optimization

$$\min_{\mathbf{D}\in\mathcal{C}} \frac{1}{n}\sum_{1}^{n} L(\mathbf{y}_i, \mathbf{D})$$

where

$$L(\mathbf{y}_i, \mathbf{D}) \triangleq \min_{\boldsymbol{\gamma}} \frac{1}{2}\|\mathbf{y}_i - \mathbf{D}\boldsymbol{\gamma}\|_2^2 + \lambda\|\boldsymbol{\gamma}\|_1$$

Stochastic (projected) Gradient Descent

1: Init $\mathbf{D} = \mathbf{D}_0$
2: **for** $t = 1$ to T **do**
3: Sparse Coding:

$$\boldsymbol{\gamma}_t = \arg\min_{\boldsymbol{\gamma}} \frac{1}{2}\|\mathbf{y}_t - \mathbf{D}\boldsymbol{\gamma}\|_2^2 + \lambda\|\boldsymbol{\gamma}\|_1$$

Stochastic Optimization

$$\min_{\mathbf{D}\in\mathcal{C}} \frac{1}{n}\sum_{1}^{n} L(\mathbf{y}_i, \mathbf{D})$$

where

$$L(\mathbf{y}_i, \mathbf{D}) \triangleq \min_{\boldsymbol{\gamma}} \frac{1}{2}\|\mathbf{y}_i - \mathbf{D}\boldsymbol{\gamma}\|_2^2 + \lambda\|\boldsymbol{\gamma}\|_1$$

Stochastic (projected) Gradient Descent

Init $\mathbf{D} = \mathbf{D}_0$

for $t = 1$ to T **do**

Sparse Coding:

$$\boldsymbol{\gamma}_t = \arg\min_{\boldsymbol{\gamma}} \frac{1}{2}\|\mathbf{y}_t - \mathbf{D}\boldsymbol{\gamma}\|_2^2 + \lambda\|\boldsymbol{\gamma}\|_1$$

Dictionary Update:

$$\mathbf{D} \leftarrow \Pi_{\mathcal{C}}[\mathbf{D} - \eta^t(\mathbf{D}\boldsymbol{\gamma}_t - \mathbf{y})\boldsymbol{\gamma}_t^T]$$

end for

Online Dictionary Learning [Mairal, Bach, Ponce, Sapiro, '09]

$$\min_{\mathbf{D}\in\mathcal{C}} \; f(\mathbf{D}) = \mathbb{E}_{\mathbf{y}} \; [L(\mathbf{y},\mathbf{D})] \qquad \text{via} \qquad \min_{\mathbf{D}\in\mathcal{C}} \; f_n(\mathbf{D}) = \frac{1}{n}\sum_{i=1}^{n} L(\mathbf{y}_i,\mathbf{D})$$

where

$$L(\mathbf{y},\mathbf{D}) \triangleq \min_{\boldsymbol{\gamma}} \frac{1}{2}\|\mathbf{y}-\mathbf{D}\boldsymbol{\gamma}\|_2^2 + \lambda\|\boldsymbol{\gamma}\|_1$$

Online Dictionary Learning [Mairal, Bach, Ponce, Sapiro, '09]

$$\min_{\mathbf{D}\in\mathcal{C}} \ f(\mathbf{D}) = \mathbb{E}_{\mathbf{y}} \left[L(\mathbf{y},\mathbf{D})\right] \qquad \text{via} \qquad \min_{\mathbf{D}\in\mathcal{C}} \ f_n(\mathbf{D}) = \frac{1}{n}\sum_{i=1}^{n} L(\mathbf{y}_i,\mathbf{D})$$

where

$$L(\mathbf{y},\mathbf{D}) \triangleq \min_{\boldsymbol{\gamma}} \frac{1}{2}\|\mathbf{y}-\mathbf{D}\boldsymbol{\gamma}\|_2^2 + \lambda\|\boldsymbol{\gamma}\|_1$$

1: Init $\mathbf{D}_0$
2: **for** $t = 1$ to T **do**
3: Sparse Coding:

$$\boldsymbol{\gamma}_t = \arg\min_{\boldsymbol{\gamma}} \frac{1}{2}\|\mathbf{y}_t-\mathbf{D}_t\boldsymbol{\gamma}\|_2^2 + \lambda\|\boldsymbol{\gamma}\|_1$$

Online Dictionary Learning [Mairal, Bach, Ponce, Sapiro, '09]

$$\min_{\mathbf{D}\in\mathcal{C}} \; f(\mathbf{D}) = \mathbb{E}_{\mathbf{y}}\left[L(\mathbf{y},\mathbf{D})\right] \qquad \text{via} \qquad \min_{\mathbf{D}\in\mathcal{C}} \; f_n(\mathbf{D}) = \frac{1}{n}\sum_{i=1}^{n} L(\mathbf{y}_i,\mathbf{D})$$

where

$$L(\mathbf{y},\mathbf{D}) \triangleq \min_{\boldsymbol{\gamma}} \frac{1}{2}\|\mathbf{y}-\mathbf{D}\boldsymbol{\gamma}\|_2^2 + \lambda\|\boldsymbol{\gamma}\|_1$$

1: Init $\mathbf{D}_0$
2: **for** $t = 1$ to T **do**
3: Sparse Coding:

$$\boldsymbol{\gamma}_t = \arg\min_{\boldsymbol{\gamma}} \frac{1}{2}\|\mathbf{y}_t-\mathbf{D}_t\boldsymbol{\gamma}\|_2^2 + \lambda\|\boldsymbol{\gamma}\|_1$$

4: Dictionary Update:

$$\mathbf{D}_{t+1} = \arg\min_{\mathbf{D}\in\mathcal{C}} \tilde{f}_t(\mathbf{D}) = \frac{1}{t}\sum_{i=1}^{t} \frac{1}{2}\|\mathbf{y}_i-\mathbf{D}\boldsymbol{\gamma}_t\|_2^2 + \lambda\|\boldsymbol{\gamma}_t\|_1 \geq f_n(\mathbf{D})$$

5: **end for**

Online Dictionary Learning [Mairal, Bach, Ponce, Sapiro, '09]

$$\mathbf{D}_t = \arg\min_{\mathbf{D}\in\mathcal{C}} \sum_{i=1}^{t} \frac{1}{2}\|\mathbf{y}_i - \mathbf{D}\boldsymbol{\gamma}_t\|_2^2 = \arg\min_{\mathbf{d}} \frac{1}{2}\|\mathbf{Y}_t - \mathbf{D}\boldsymbol{\Gamma}_t\|_2^2$$

Online Dictionary Learning [Mairal, Bach, Ponce, Sapiro, '09]

$$\mathbf{D}_t = \arg\min_{\mathbf{D}\in\mathcal{C}} \sum_{i=1}^{t} \frac{1}{2}\|\mathbf{y}_i - \mathbf{D}\boldsymbol{\gamma}_t\|_2^2 = \arg\min_{\mathbf{d}} \frac{1}{2}\|\mathbf{Y}_t - \mathbf{D}\boldsymbol{\Gamma}_t\|_2^2$$

Block Coordinate Descent

Letting $\mathbf{B} = \mathbf{Y}\boldsymbol{\Gamma}^T$, $\mathbf{C} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^T$

$$\mathbf{d}_j = \arg\min_{\mathbf{d}\in\mathcal{C}} \frac{1}{2}\|\mathbf{Y}_t - \mathbf{D}\boldsymbol{\Gamma}_t + \mathbf{d}_j\boldsymbol{\Gamma}_j^T - \mathbf{d}\boldsymbol{\Gamma}_j^T\|_2^2$$

Online Dictionary Learning [Mairal, Bach, Ponce, Sapiro, '09]

$$\mathbf{D}_t = \arg\min_{\mathbf{D}\in\mathcal{C}} \sum_{i=1}^{t} \frac{1}{2}\|\mathbf{y}_i - \mathbf{D}\boldsymbol{\gamma}_t\|_2^2 = \arg\min_{\mathbf{d}} \frac{1}{2}\|\mathbf{Y}_t - \mathbf{D}\boldsymbol{\Gamma}_t\|_2^2$$

Block Coordinate Descent

Letting $\mathbf{B} = \mathbf{Y}\boldsymbol{\Gamma}^T$, $\mathbf{C} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^T$

$$\mathbf{d}_j = \arg\min_{\mathbf{d}\in\mathcal{C}} \frac{1}{2}\|\mathbf{Y}_t - \mathbf{D}\boldsymbol{\Gamma}_t + \mathbf{d}_j\boldsymbol{\Gamma}_j^T - \mathbf{d}\boldsymbol{\Gamma}_j^T\|_2^2$$

$$\vdots$$

$$\mathbf{d}_j = \arg\min_{\mathbf{d}\in\mathcal{C}} \frac{1}{2}\|\frac{1}{\mathbf{C}[j,j]}(\mathbf{b}_j - \mathbf{D}\mathbf{c}_j) + \mathbf{d}_j - \mathbf{d}\|_2^2$$

Online Dictionary Learning [Mairal, Bach, Ponce, Sapiro, '09]

$$\mathbf{D}_t = \arg\min_{\mathbf{D}\in\mathcal{C}} \sum_{i=1}^{t} \frac{1}{2}\|\mathbf{y}_i - \mathbf{D}\boldsymbol{\gamma}_t\|_2^2 = \arg\min_{\mathbf{d}} \frac{1}{2}\|\mathbf{Y}_t - \mathbf{D}\boldsymbol{\Gamma}_t\|_2^2$$

Block Coordinate Descent

Letting $\mathbf{B} = \mathbf{Y}\boldsymbol{\Gamma}^T$, $\mathbf{C} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^T$

$$\mathbf{d}_j = \arg\min_{\mathbf{d}\in\mathcal{C}} \frac{1}{2}\|\mathbf{Y}_t - \mathbf{D}\boldsymbol{\Gamma}_t + \mathbf{d}_j\boldsymbol{\Gamma}_j^T - \mathbf{d}\boldsymbol{\Gamma}_j^T\|_2^2$$

$$\vdots$$

$$\mathbf{d}_j = \arg\min_{\mathbf{d}\in\mathcal{C}} \frac{1}{2}\|\frac{1}{\mathbf{C}[j,j]}(\mathbf{b}_j - \mathbf{D}\mathbf{c}_j) + \mathbf{d}_j - \mathbf{d}\|_2^2$$

$$\mathbf{d}_j = \Pi_{\mathcal{C}}\left[\frac{1}{\mathbf{C}[j,j]}(\mathbf{b}_j - \mathbf{D}\mathbf{c}_j) + \mathbf{d}_j\right]$$

Online Dictionary Learning [Mairal, Bach, Ponce, Sapiro, '09]

$$\min_{\mathbf{D}\in\mathcal{C}} \ \mathbb{E}_{\mathbf{y}} \left[L(\mathbf{y}_i, \mathbf{D})\right]$$

Init $\mathbf{D} = \mathbf{D}_0$, $\mathbf{B} = \mathbf{0}, \mathbf{C} = \mathbf{0}$

for $t = 1$ to T **do**

  Sparse Coding: $\boldsymbol{\gamma}_t = \arg\min_{\boldsymbol{\gamma}} \frac{1}{2}\|\mathbf{y}_t - \mathbf{D}\boldsymbol{\gamma}\|_2^2 + \lambda\|\boldsymbol{\gamma}\|_1$

  Update memory variables

$$\mathbf{B} \leftarrow \mathbf{B} + \mathbf{y}_t\boldsymbol{\gamma}_t^T \ ; \quad \mathbf{C} \leftarrow \mathbf{C} + \boldsymbol{\gamma}_t\boldsymbol{\gamma}_t^T$$

  Dictionary Update:

$$\mathbf{d}_j = \frac{1}{\mathbf{C}[j,j]}(\mathbf{b}_j - \mathbf{D}\mathbf{c}_j) + \mathbf{d}_j$$

$$\mathbf{d}_j = \frac{1}{\|\mathbf{d}_j\|_2}\mathbf{d}_j$$

end for

Online Dictionary Learning [Mairal, Bach, Ponce, Sapiro, '09]

Details

- Extensions to mini-batches
- Pruning unused atoms too

Online Dictionary Learning [Mairal, Bach, Ponce, Sapiro, '09]

Details

- Extensions to mini-batches
- Pruning unused atoms too

Convergence Guarantees

This proposes to minimize the quadratic surrogate $\tilde{f}_n(\mathbf{D}) \geq f_n(\mathbf{D})$
Under mild assumptions,

- $\tilde{f}_n(\mathbf{D}_t)$ converges a.s.
- $f_n(\mathbf{D}_t) - \tilde{f}_n(\mathbf{D}_t) \to 0$ a.s.
- $f_n(\mathbf{D}_t)$ converges a.s.

Online Dictionary Learning [Mairal, Bach, Ponce, Sapiro, '09]

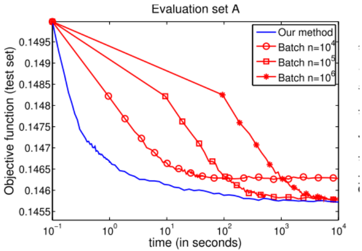
Details

- Extensions to mini-batches
- Pruning unused atoms too

Convergence Guarantees

This proposes to minimize the quadratic surrogate $\tilde{f}_n(\mathbf{D}) \geq f_n(\mathbf{D})$
Under mild assumptions,

- $\tilde{f}_n(\mathbf{D}_t)$ converges a.s.
- $f_n(\mathbf{D}_t) - \tilde{f}_n(\mathbf{D}_t) \to 0$ a.s.
- $f_n(\mathbf{D}_t)$ converges a.s.



Online Dictionary Learning [Mairal, Bach, Ponce, Sapiro, '09]

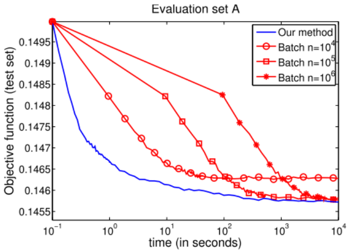
Details

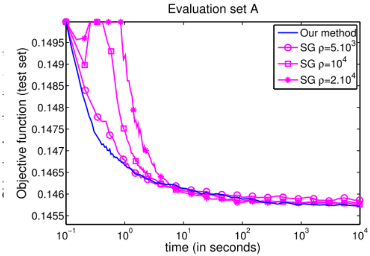
- Extensions to mini-batches
- Pruning unused atoms too

Convergence Guarantees

This proposes to minimize the quadratic surrogate $\tilde{f}_n(\mathbf{D}) \geq f_n(\mathbf{D})$
Under mild assumptions,

- $\tilde{f}_n(\mathbf{D}_t)$ converges a.s.
- $f_n(\mathbf{D}_t) - \tilde{f}_n(\mathbf{D}_t) \to 0$ a.s.
- $f_n(\mathbf{D}_t)$ converges a.s.





Applications to Inverse Problems in Computer Vision

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{x}\|_2^2 + P(\mathbf{x})$$

Applications to Inverse Problems in Computer Vision

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{x}\|_2^2 + P(\mathbf{x})$$

- Global Representation Prior:

$$\min_{\boldsymbol{\gamma}} \|\mathbf{y} - \mathbf{W}\boldsymbol{\gamma}\|_2^2 + \lambda\|\boldsymbol{\gamma}\|_0$$

Applications to Inverse Problems in Computer Vision

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{x}\|_2^2 + P(\mathbf{x})$$

- Global Representation Prior:

$$\min_{\boldsymbol{\gamma}} \|\mathbf{y} - \mathbf{W}\boldsymbol{\gamma}\|_2^2 + \lambda\|\boldsymbol{\gamma}\|_0$$

Noisy Image: $\sigma = 20$, PSNR = 22.11 dB

$\hat{\mathbf{x}} = \mathbf{W}\mathcal{H}_\lambda(\mathbf{W}^T\mathbf{y})$ PSNR = 28.36 dB

From Global to Local Priors

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 + \lambda \sum_i p(\mathbf{R}_i\mathbf{x})$$

From Global to Local Priors

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \sum_i p(\mathbf{R}_i \mathbf{x})$$

$$\min_{\mathbf{x}, \{\boldsymbol{\gamma}_i\}} \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 + \frac{\lambda}{2} \sum_i \|\mathbf{R}_i \mathbf{x} - \mathbf{D}\boldsymbol{\gamma}_i\|_2^2 + \lambda \|\boldsymbol{\gamma}_i\|_0$$

From Global to Local Priors

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 + \lambda \sum_i p(\mathbf{R}_i \mathbf{x})$$

$$\min_{\mathbf{x},\{\boldsymbol{\gamma}_i\}} \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 + \frac{\lambda}{2} \sum_i \|\mathbf{R}_i \mathbf{x} - \mathbf{D}\boldsymbol{\gamma}_i\|_2^2 + \lambda \|\boldsymbol{\gamma}_i\|_0$$

Alternating Minimization:

- fix $\mathbf{x} = \mathbf{y}$, and

$$\min_{\{\boldsymbol{\gamma}_i\}} \frac{\lambda}{2} \sum_i \|\mathbf{R}_i \mathbf{x} - \mathbf{D}\boldsymbol{\gamma}_i\|_2^2 + \lambda \|\boldsymbol{\gamma}_i\|_0$$

$$\Rightarrow \hat{\boldsymbol{\gamma}}_i = \arg\min_{\boldsymbol{\gamma}_i} \|\boldsymbol{\gamma}_i\|_0 \ \text{ s.t. } \ \|\mathbf{R}_i \mathbf{x} - \mathbf{D}\boldsymbol{\gamma}_i\|_2^2 \le n\sigma^2 c, \ \forall i$$

From Global to Local Priors

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 + \lambda \sum_i p(\mathbf{R}_i\mathbf{x})$$

$$\min_{\mathbf{x},\{\boldsymbol{\gamma}_i\}} \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 + \frac{\lambda}{2} \sum_i \|\mathbf{R}_i\mathbf{x}-\mathbf{D}\boldsymbol{\gamma}_i\|_2^2 + \lambda\|\boldsymbol{\gamma}_i\|_0$$

Alternating Minimization:

- fix $\mathbf{x}=\mathbf{y}$, and

$$\min_{\{\boldsymbol{\gamma}_i\}} \frac{\lambda}{2} \sum_i \|\mathbf{R}_i\mathbf{x}-\mathbf{D}\boldsymbol{\gamma}_i\|_2^2 + \lambda\|\boldsymbol{\gamma}_i\|_0$$

$$\Rightarrow \hat{\boldsymbol{\gamma}}_i = \arg\min_{\boldsymbol{\gamma}_i} \|\boldsymbol{\gamma}_i\|_0 \ \ \text{s.t.} \ \ \|\mathbf{R}_i\mathbf{x}-\mathbf{D}\boldsymbol{\gamma}_i\|_2^2 \leq n\sigma^2 c, \ \forall i$$

- fix $\boldsymbol{\gamma}_i$ and

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 + \frac{\lambda}{2} \sum_i \|\mathbf{R}_i\mathbf{x}-\mathbf{D}\boldsymbol{\gamma}_i\|_2^2$$

From Global to Local Priors

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 + \lambda \sum_i p(\mathbf{R}_i\mathbf{x})$$

$$\min_{\mathbf{x},\{\boldsymbol{\gamma}_i\}} \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 + \frac{\lambda}{2} \sum_i \|\mathbf{R}_i\mathbf{x}-\mathbf{D}\boldsymbol{\gamma}_i\|_2^2 + \lambda\|\boldsymbol{\gamma}_i\|_0$$

Alternating Minimization:

- fix $\mathbf{x}=\mathbf{y}$, and

$$\min_{\{\boldsymbol{\gamma}_i\}} \frac{\lambda}{2} \sum_i \|\mathbf{R}_i\mathbf{x}-\mathbf{D}\boldsymbol{\gamma}_i\|_2^2 + \lambda\|\boldsymbol{\gamma}_i\|_0$$

$$\Rightarrow \hat{\boldsymbol{\gamma}}_i = \arg\min_{\boldsymbol{\gamma}_i} \|\boldsymbol{\gamma}_i\|_0 \ \text{ s.t. } \ \|\mathbf{R}_i\mathbf{x}-\mathbf{D}\boldsymbol{\gamma}_i\|_2^2 \leq n\sigma^2 c,\ \forall i$$

- fix $\boldsymbol{\gamma}_i$ and

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 + \frac{\lambda}{2} \sum_i \|\mathbf{R}_i\mathbf{x}-\mathbf{D}\boldsymbol{\gamma}_i\|_2^2$$

$$\Rightarrow \hat{\mathbf{x}} = \left(\lambda\mathbf{I} + \sum_i \mathbf{R}_i^T\mathbf{R}_i\right)^{-1} \left(\mathbf{y} + \sum_i \mathbf{R}_i^T\mathbf{D}\hat{\boldsymbol{\gamma}}_i\right)$$

Global - PSNR = 28.36 dB

Global - PSNR = 28.36 dB

Local - PSNR = 30.01 dB

From Global to Adaptive Local Priors

$$\min_{\mathbf{x},\{\boldsymbol{\gamma}_i\},\mathbf{D}} \ \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 + \frac{\lambda}{2}\sum_i \|\mathbf{R}_i\mathbf{x} - \mathbf{D}\boldsymbol{\gamma}_i\|_2^2 + \lambda\|\boldsymbol{\gamma}_i\|_0$$

From Global to Adaptive Local Priors

$$\min_{\mathbf{x},\{\boldsymbol{\gamma}_i\},\mathbf{D}} \; \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 + \frac{\lambda}{2}\sum_i \|\mathbf{R}_i\mathbf{x} - \mathbf{D}\boldsymbol{\gamma}_i\|_2^2 + \lambda\|\boldsymbol{\gamma}_i\|_0$$

- fix $\mathbf{x}$ and

$$\min_{\mathbf{x},\mathbf{D}} \sum_i \|\boldsymbol{\gamma}_i\|_0 \;\; \text{s.t.} \;\; \|\mathbf{R}_i\mathbf{x} - \mathbf{D}\boldsymbol{\gamma}_i\|_2^2 \le n\sigma^2 c$$

From Global to Adaptive Local Priors

$$\min_{\mathbf{x},\{\boldsymbol{\gamma}_i\},\mathbf{D}} \ \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 + \frac{\lambda}{2}\sum_i \|\mathbf{R}_i\mathbf{x} - \mathbf{D}\boldsymbol{\gamma}_i\|_2^2 + \lambda\|\boldsymbol{\gamma}_i\|_0$$

- fix $\mathbf{x}$ and

$$\min_{\mathbf{x},\mathbf{D}} \sum_i \|\boldsymbol{\gamma}_i\|_0 \ \text{ s.t. } \ \|\mathbf{R}_i\mathbf{x} - \mathbf{D}\boldsymbol{\gamma}_i\|_2^2 \le n\sigma^2 c$$

$$\text{iterate}\begin{cases} \hat{\boldsymbol{\gamma}}_i^{k+1} = \arg\min_{\boldsymbol{\gamma}_i} \|\boldsymbol{\gamma}_i\|_0 \ \text{ s.t. } \ \|\mathbf{R}_i\mathbf{x} - \mathbf{D}^k\boldsymbol{\gamma}_i\|_2^2 \le n\sigma^2 c, \ \forall i \\ \mathbf{D}^{k+1} = \arg\min_{\mathbf{D}} \|\mathbf{R}_i\mathbf{x} - \mathbf{D}\boldsymbol{\gamma}_i^{k+1}\|_2^2 \end{cases}$$

From Global to Adaptive Local Priors

$$\min_{\mathbf{x},\{\boldsymbol{\gamma}_i\},\mathbf{D}} \ \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 + \frac{\lambda}{2}\sum_i \|\mathbf{R}_i\mathbf{x} - \mathbf{D}\boldsymbol{\gamma}_i\|_2^2 + \lambda\|\boldsymbol{\gamma}_i\|_0$$

- fix $\mathbf{x}$ and

$$\min_{\mathbf{x},\mathbf{D}} \sum_i \|\boldsymbol{\gamma}_i\|_0 \ \ \text{s.t.} \ \ \|\mathbf{R}_i\mathbf{x} - \mathbf{D}\boldsymbol{\gamma}_i\|_2^2 \le n\sigma^2 c$$

$$\text{iterate}\begin{cases} \hat{\boldsymbol{\gamma}}_i^{k+1} = \arg\min_{\boldsymbol{\gamma}_i} \|\boldsymbol{\gamma}_i\|_0 \ \ \text{s.t.} \ \ \|\mathbf{R}_i\mathbf{x} - \mathbf{D}^k\boldsymbol{\gamma}_i\|_2^2 \le n\sigma^2 c, \ \forall i \\ \mathbf{D}^{k+1} = \arg\min_{\mathbf{D}} \|\mathbf{R}_i\mathbf{x} - \mathbf{D}\boldsymbol{\gamma}_i^{k+1}\|_2^2 \end{cases}$$

- Solve for $\mathbf{x}$,

$$\Rightarrow \hat{\mathbf{x}} = \left(\lambda\mathbf{I} + \sum_i \mathbf{R}_i^T\mathbf{R}_i\right)^{-1}\left(\mathbf{y} + \sum_i \mathbf{R}_i^T\mathbf{D}\hat{\boldsymbol{\gamma}}_i\right)$$

Local DCT - PSNR = 30.01 dB

Local DCT - PSNR = 30.01 dB

Local K-SVD - PSNR = 30.96 dB

[Aharon, Bruckstein, Elad, '06]

Beyond K-SVD denoising: EPLL

$$\min_{\mathbf{x},\{\boldsymbol{\gamma}_i\},\mathbf{D}} \ \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 + \frac{\lambda}{2}\sum_i \|\mathbf{R}_i\mathbf{x}-\mathbf{D}\boldsymbol{\gamma}_i\|_2^2 + \lambda\|\boldsymbol{\gamma}_i\|_0$$

- fix $\mathbf{x}$ and

$$\text{iterate}\begin{cases} \hat{\gamma}_i^{k+1} = \arg\min_{\boldsymbol{\gamma}_i} \|\boldsymbol{\gamma}_i\|_0 \ \text{ s.t. } \ \|\mathbf{R}_i\mathbf{x}-\mathbf{D}^k\boldsymbol{\gamma}_i\|_2^2 \le n\sigma^2 c, \ \forall i \\ \mathbf{D}^{k+1} = \arg\min_{\mathbf{D}} \|\mathbf{R}_i\mathbf{x}-\mathbf{D}\boldsymbol{\gamma}_i^{k+1}\|_2^2 \end{cases}$$

- Solve for $\mathbf{x}$,

$$\Rightarrow \hat{\mathbf{x}} = \left(\lambda\mathbf{I} + \sum_i \mathbf{R}_i^T\mathbf{R}_i\right)^{-1}\left(\mathbf{y} + \sum_i \mathbf{R}_i^T\mathbf{D}\hat{\boldsymbol{\gamma}}_i\right)$$

Beyond K-SVD denoising: EPLL

$$\min_{\mathbf{x},\{\boldsymbol{\gamma}_i\},\mathbf{D}} \ \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 + \frac{\lambda}{2}\sum_i \|\mathbf{R}_i\mathbf{x}-\mathbf{D}\boldsymbol{\gamma}_i\|_2^2 + \lambda\|\boldsymbol{\gamma}_i\|_0$$

- fix $\mathbf{x}$ and

$$\text{iterate}\begin{cases} \hat{\gamma}_i^{k+1} = \arg\min_{\boldsymbol{\gamma}_i} \|\boldsymbol{\gamma}_i\|_0 \ \ \text{s.t.} \ \ \|\mathbf{R}_i\mathbf{x}-\mathbf{D}^k\boldsymbol{\gamma}_i\|_2^2 \leq n\sigma^2 c, \ \forall i \\ \mathbf{D}^{k+1} = \arg\min_{\mathbf{D}} \|\mathbf{R}_i\mathbf{x}-\mathbf{D}\boldsymbol{\gamma}_i^{k+1}\|_2^2 \end{cases}$$

- Solve for $\mathbf{x}$,

$$\Rightarrow \hat{\mathbf{x}} = \left(\lambda\mathbf{I} + \sum_i \mathbf{R}_i^T\mathbf{R}_i\right)^{-1}\left(\mathbf{y} + \sum_i \mathbf{R}_i^T\mathbf{D}\hat{\gamma}_i\right)$$

Keep iterating! – but how much noise is there in $\mathbf{R}_i\hat{\mathbf{x}}$?

Beyond K-SVD denoising: EPLL

$$\min_{\mathbf{x},\{\boldsymbol{\gamma}_i\},\mathbf{D}} \;\; \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 + \frac{\lambda}{2}\sum_i \|\mathbf{R}_i\mathbf{x} - \mathbf{D}\boldsymbol{\gamma}_i\|_2^2 + \lambda\|\boldsymbol{\gamma}_i\|_0$$

- fix $\mathbf{x}$ and

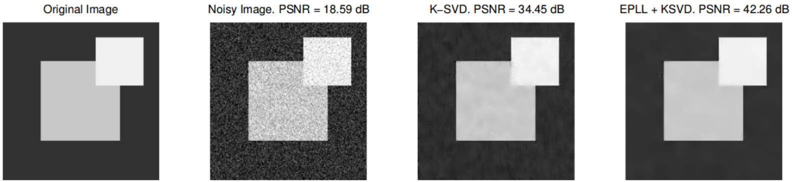
$$\text{iterate}\begin{cases} \hat{\gamma}_i^{k+1} = \arg\min_{\boldsymbol{\gamma}_i} \|\boldsymbol{\gamma}_i\|_0 \;\; \text{s.t.} \;\; \|\mathbf{R}_i\mathbf{x} - \mathbf{D}^k\boldsymbol{\gamma}_i\|_2^2 \le n\sigma^2 c, \;\forall i \\ \mathbf{D}^{k+1} = \arg\min_{\mathbf{D}} \|\mathbf{R}_i\mathbf{x} - \mathbf{D}\boldsymbol{\gamma}_i^{k+1}\|_2^2 \end{cases}$$

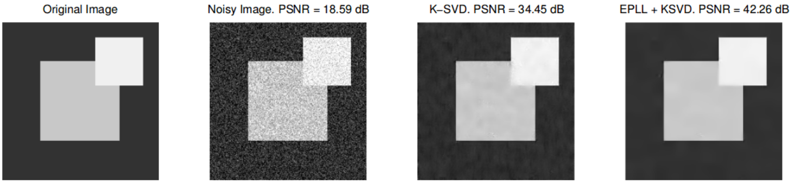
- Solve for $\mathbf{x}$,

$$\Rightarrow \hat{\mathbf{x}} = \left(\lambda\mathbf{I} + \sum_i \mathbf{R}_i^T\mathbf{R}_i\right)^{-1}\left(\mathbf{y} + \sum_i \mathbf{R}_i^T\mathbf{D}\hat{\gamma}_i\right)$$

Keep iterating! – but how much noise is there in $\mathbf{R}_i\hat{\mathbf{x}}$?

$\rightarrow$ Recall that $\|\mathbf{R}_i\hat{\mathbf{x}} - \mathbf{R}_i\mathbf{y}\|_2^2 \propto \frac{k}{n}\sigma^2$


Original Image
Noisy Image. PSNR = 18.59 dB
K–SVD. PSNR = 34.45 dB
EPLL + KSVD. PSNR = 42.26 dB



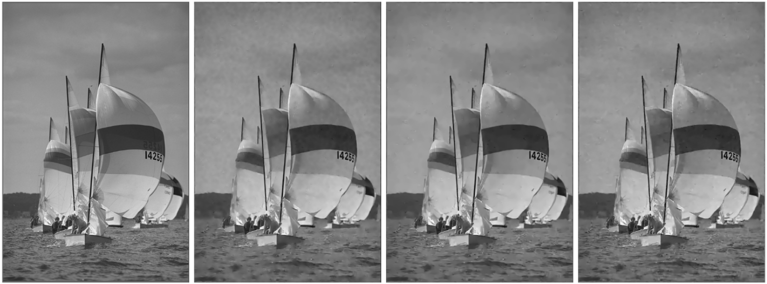


Fig. 8. Denoising results of an image from the Kodak Database, initially corrupted with additive white Gaussian nose ($\sigma = 25$). From left to right: Original Image, K-SVD (PSNR = 31.42 dB), EPLL with Sparse Prior (PSNR = 31.83 dB), and EPLL with GMM (PSNR = 31.85 dB).

[S., Elad, '15]

Inpainting

$\mathbf{y} = \mathbf{M}\mathbf{x}$

Inpainting

$$\mathbf{y} = \mathbf{M}\mathbf{x} \Rightarrow \min_{\mathbf{x},\{\boldsymbol{\gamma}_i\}} \ \frac{1}{2}\|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 + \frac{\lambda}{2}\sum_i \|\mathbf{R}_i\mathbf{M}\mathbf{x} - \mathbf{D}\boldsymbol{\gamma}_i\|_2^2 + \lambda\|\boldsymbol{\gamma}_i\|_0$$

Inpainting

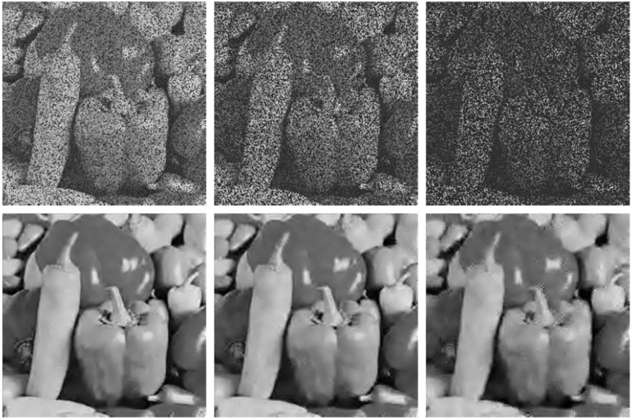
$$\mathbf{y} = \mathbf{M}\mathbf{x} \Rightarrow \min_{\mathbf{x},\{\boldsymbol{\gamma}_i\}} \ \frac{1}{2}\|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 + \frac{\lambda}{2}\sum_i \|\mathbf{R}_i\mathbf{M}\mathbf{x} - \mathbf{D}\boldsymbol{\gamma}_i\|_2^2 + \lambda\|\boldsymbol{\gamma}_i\|_0$$

$$\hat{\boldsymbol{\gamma}}_i = \arg\min_{\boldsymbol{\gamma}_i} \|\boldsymbol{\gamma}_i\|_0 \ \ \text{s.t.} \ \ \|\mathbf{R}_i\mathbf{M}\mathbf{x} - \mathbf{D}\boldsymbol{\gamma}_i\|_2^2 \leq \delta_i, \ \forall i$$

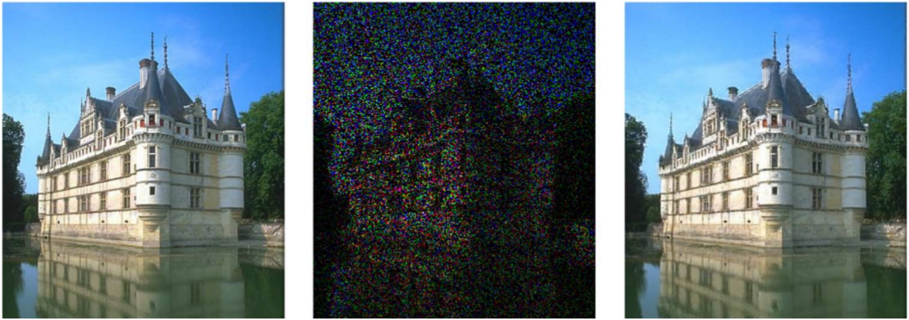
Inpainting

$$\mathbf{y} = \mathbf{M}\mathbf{x} \Rightarrow \min_{\mathbf{x},\{\boldsymbol{\gamma}_i\}} \ \frac{1}{2}\|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 + \frac{\lambda}{2}\sum_i \|\mathbf{R}_i\mathbf{M}\mathbf{x} - \mathbf{D}\boldsymbol{\gamma}_i\|_2^2 + \lambda\|\boldsymbol{\gamma}_i\|_0$$

$$\hat{\boldsymbol{\gamma}}_i = \arg\min_{\boldsymbol{\gamma}_i} \|\boldsymbol{\gamma}_i\|_0 \ \ \text{s.t.} \ \ \|\mathbf{R}_i\mathbf{M}\mathbf{x} - \mathbf{D}\boldsymbol{\gamma}_i\|_2^2 \leq \delta_i, \ \forall i$$








Since 1699, when French explorers
landed at the great bend of the
Mississippi River and celebrated
the first Mardi Gras in North America,
New Orleans has brewed a fascinating
melange of cultures. It was French,
then Spanish, then French again, then
sold to the United States. Through all
these years, and even into the 1900s,
others arrived from everywhere:
Acadians (Cajuns), Africans, indige-

Morphological Component Analysis

$$\mathbf{y} = \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{v}$$

Morphological Component Analysis

$$\mathbf{y} = \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{v} = \mathbf{D}_A \boldsymbol{\gamma}_A + \mathbf{D}_B \boldsymbol{\gamma}_B + \mathbf{v}$$

Morphological Component Analysis

$$\mathbf{y} = \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{v} = \mathbf{D}_A \boldsymbol{\gamma}_A + \mathbf{D}_B \boldsymbol{\gamma}_B + \mathbf{v}$$

Then,

$$\min_{\boldsymbol{\gamma}_A, \boldsymbol{\gamma}_B, \mathbf{D}_A, \mathbf{D}_B} \lambda_1 \|\boldsymbol{\gamma}_A\|_1 + \lambda_2 \|\boldsymbol{\gamma}_B\|_1 + \frac{1}{2} \|\mathbf{y} - \mathbf{D}_A \boldsymbol{\gamma}_A - \mathbf{D}_B \boldsymbol{\gamma}_B\|_2^2$$

Morphological Component Analysis

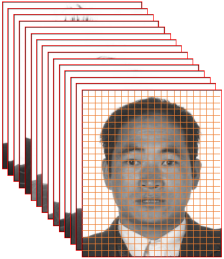
$$\mathbf{y} = \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{v} = \mathbf{D}_A \boldsymbol{\gamma}_A + \mathbf{D}_B \boldsymbol{\gamma}_B + \mathbf{v}$$

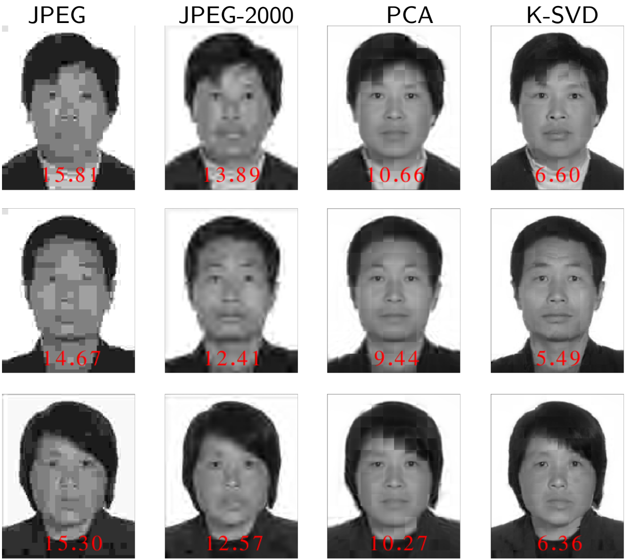
Then,

$$\min_{\boldsymbol{\gamma}_A, \boldsymbol{\gamma}_B, \mathbf{D}_A, \mathbf{D}_B} \lambda_1 \|\boldsymbol{\gamma}_A\|_1 + \lambda_2 \|\boldsymbol{\gamma}_B\|_1 + \frac{1}{2} \|\mathbf{y} - \mathbf{D}_A \boldsymbol{\gamma}_A - \mathbf{D}_B \boldsymbol{\gamma}_B\|_2^2$$

$$\Rightarrow \hat{\mathbf{x}}_A = \hat{\mathbf{D}}_A \hat{\boldsymbol{\gamma}}_A \ , \quad \hat{\mathbf{x}}_A = \hat{\mathbf{D}}_B \hat{\boldsymbol{\gamma}}_B$$

Image Compression



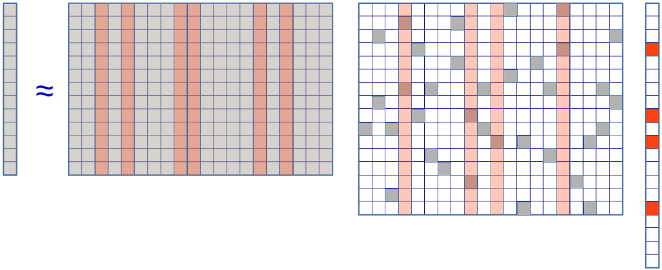


Double Sparsity and Trainlets

Double Sparsity and Trainlets

Let $\mathbf{D} = \mathbf{\Phi A}$, where $\|\mathbf{A}_i\|_0 \leq l$, $\mathbf{\Phi}$: fixed, low-complexity base dictionary

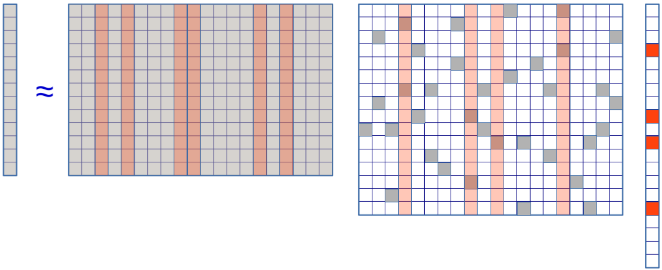
$$\min_{\mathbf{A},\mathbf{\Gamma}} \frac{1}{2}\|\mathbf{Y} - \mathbf{\Phi A \Gamma}\|_2^2 \text{ s.t. } \begin{cases} \|\boldsymbol{\gamma}_i\|_0 \leq k \ \forall i \\ \|\mathbf{a}_j\|_0 \leq l \ \forall j \end{cases}$$



Double Sparsity and Trainlets

Let $\mathbf{D} = \mathbf{\Phi}\mathbf{A}$, where $\|\mathbf{A}_i\|_0 \leq l$, $\mathbf{\Phi}$: fixed, low-complexity base dictionary

$$\min_{\mathbf{A},\mathbf{\Gamma}} \frac{1}{2}\|\mathbf{Y} - \mathbf{\Phi}\mathbf{A}\mathbf{\Gamma}\|_2^2 \ \text{ s.t. } \begin{cases} \|\boldsymbol{\gamma}_i\|_0 \leq k \ \forall i \\ \|\mathbf{a}_j\|_0 \leq l \ \forall j \end{cases}$$

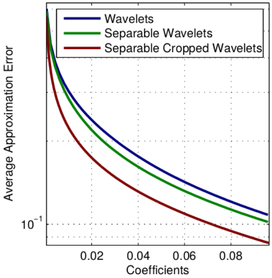


- Reduced computational complexity
- Reduced d.o.f.

[Rubinstein, Elad, '10] [S. , Elad, '16]

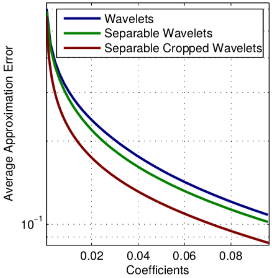
Double Sparsity and Trainlets

$\mathbf{\Phi}$: "cropped" separable wavelet atoms, for large image dimensions



Double Sparsity and Trainlets

$\mathbf{\Phi}$: "cropped" separable wavelet atoms, for large image dimensions



Sparse Coding:

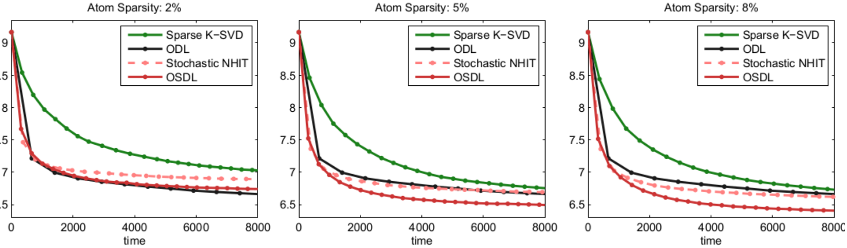
$$\min_{\mathbf{\Gamma}} \frac{1}{2}\|\mathbf{Y} - \mathbf{\Phi}\mathbf{A}\mathbf{\Gamma}\|_2^2 \text{ s.t. } \|\boldsymbol{\gamma}_i\|_0 \leq k \ \forall i$$

Dictionary Update:

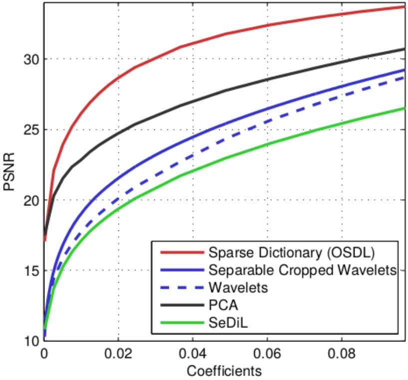
$$\min_{\mathbf{A}} \frac{1}{2}\|\mathbf{Y} - \mathbf{\Phi}\mathbf{A}\mathbf{\Gamma}\|_2^2 \text{ s.t. } \|\mathbf{a}_i\|_0 \leq l \ \forall i$$

$$\Rightarrow \mathbf{A}^{t+1} = \mathcal{H}_l \left(\mathbf{A}^t - \eta^t \mathbf{\Phi}^T (\mathbf{Y} - \mathbf{\Phi}\mathbf{A}\mathbf{\Gamma})\mathbf{\Gamma}^T \right)$$

Double Sparsity and Trainlets

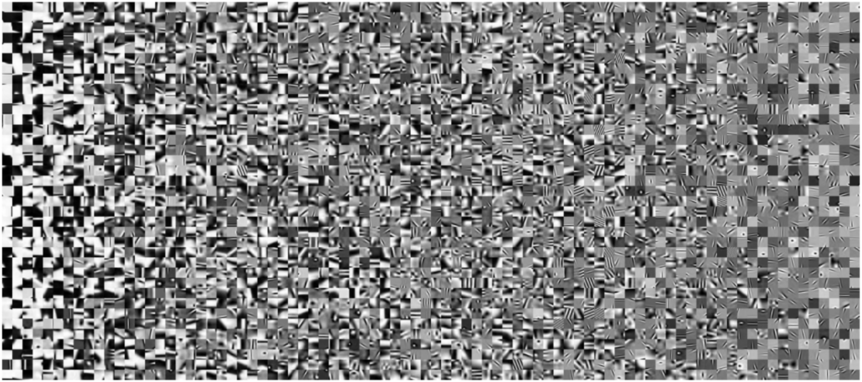


Double Sparsity and Trainlets



Double Sparsity and Trainlets

Dictionaries for *tiles* of 32×32



Double Sparsity and Trainlets