

# Online Dictionary Learning for Sparse Coding

Julien Mairal<sup>1</sup> Francis Bach<sup>1</sup>  
Jean Ponce<sup>2</sup> Guillermo Sapiro<sup>3</sup>

<sup>1</sup>INRIA–Willow project <sup>2</sup>Ecole Normale Supérieure  
<sup>3</sup>University of Minnesota

ICML, Montréal, June 2008

## What this talk is about

- Learning efficiently dictionaries (basis set) for sparse coding.
- Solving a large-scale matrix factorization problem.
- Making some large-scale image processing problems tractable.
- Proposing an algorithm which extends to NMF, sparse PCA, . . .

1 The Dictionary Learning Problem

2 Online Dictionary Learning

3 Extensions

1 The Dictionary Learning Problem

2 Online Dictionary Learning

3 Extensions

# The Dictionary Learning Problem



$$\underbrace{\mathbf{y}}_{\text{measurements}} = \underbrace{\mathbf{x}_{orig}}_{\text{original image}} + \underbrace{\mathbf{w}}_{\text{noise}}$$

# The Dictionary Learning Problem

[Elad & Aharon ('06)]

## Solving the denoising problem

- Extract all overlapping  $8 \times 8$  patches  $\mathbf{x}_i$ .
- Solve a matrix factorization problem:

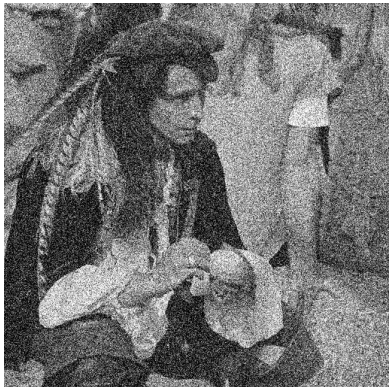
$$\min_{\alpha_i, \mathbf{D} \in \mathcal{C}} \sum_{i=1}^n \underbrace{\frac{1}{2} \|\mathbf{x}_i - \mathbf{D}\alpha_i\|_2^2}_{\text{reconstruction}} + \underbrace{\lambda \|\alpha_i\|_1}_{\text{sparsity}},$$

with  $n > 100,000$

- Average the reconstruction of each patch.

# The Dictionary Learning Problem

[Mairal, Bach, Ponce, Sapiro & Zisserman ('09)]



Denoising result

# The Dictionary Learning Problem

[Mairal, Sapiro & Elad ('08)]

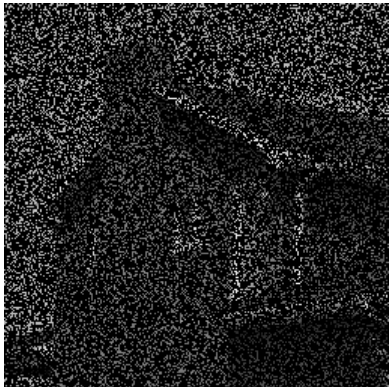
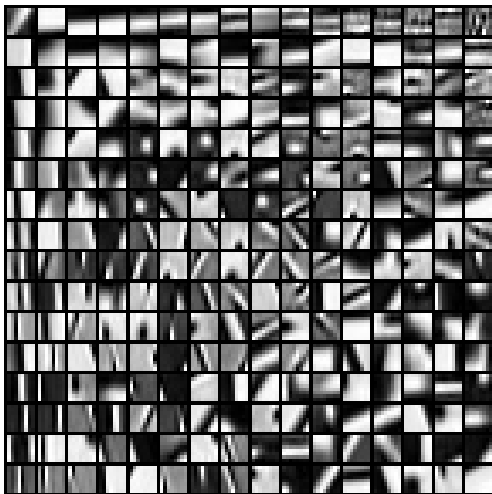


Image completion example



# The Dictionary Learning Problem

What does  $\mathbf{D}$  look like?



# The Dictionary Learning Problem

$$\min_{\substack{\boldsymbol{\alpha} \in \mathbb{R}^{k \times n} \\ \mathbf{D} \in \mathcal{C}}} \sum_{i=1}^n \frac{1}{2} \|\mathbf{x}_i - \mathbf{D}\boldsymbol{\alpha}_i\|_2^2 + \lambda \|\boldsymbol{\alpha}_i\|_1$$

$$\mathcal{C} \triangleq \{\mathbf{D} \in \mathbb{R}^{m \times k} \text{ s.t. } \forall j = 1, \dots, k, \|\mathbf{d}_j\|_2 \leq 1\}.$$

- Classical optimization alternates between  $\mathbf{D}$  and  $\boldsymbol{\alpha}$ .
- Good results, but **very slow!**

1 The Dictionary Learning Problem

2 Online Dictionary Learning

3 Extensions

## Classical formulation of dictionary learning

$$\min_{\mathbf{D} \in \mathcal{C}} f_n(\mathbf{D}) = \min_{\mathbf{D} \in \mathcal{C}} \frac{1}{n} \sum_{i=1}^n l(\mathbf{x}_i, \mathbf{D}),$$

where

$$l(\mathbf{x}, \mathbf{D}) \triangleq \min_{\boldsymbol{\alpha} \in \mathbb{R}^k} \frac{1}{2} \|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1.$$

Which formulation are we interested in?

$$\min_{\mathbf{D} \in \mathcal{C}} \left[ f(\mathbf{D}) \triangleq \mathbb{E}_{\mathbf{x}} [l(\mathbf{x}, \mathbf{D})] \approx \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n l(\mathbf{x}_i, \mathbf{D}) \right]$$

## Online learning can

- handle potentially infinite datasets,
- adapt to dynamic training sets,
- be dramatically faster than batch algorithms [Bottou & Bousquet ('08)].

# Online Dictionary Learning

## Proposed approach

- 1: **for**  $t=1, \dots, T$  **do**
- 2:     Draw  $\mathbf{x}_t$
- 3:     Sparse Coding

$$\boldsymbol{\alpha}_t \leftarrow \arg \min_{\boldsymbol{\alpha} \in \mathbb{R}^k} \frac{1}{2} \|\mathbf{x}_t - \mathbf{D}_{t-1} \boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1,$$

- 4:     Dictionary Learning

$$\mathbf{D}_t \leftarrow \arg \min_{\mathbf{D} \in \mathcal{C}} \frac{1}{t} \sum_{i=1}^t \left( \frac{1}{2} \|\mathbf{x}_i - \mathbf{D} \boldsymbol{\alpha}_i\|_2^2 + \lambda \|\boldsymbol{\alpha}_i\|_1 \right),$$

- 5: **end for**

# Online Dictionary Learning

## Proposed approach

### Implementation details

- Use LARS for the sparse coding step,
- Use a block-coordinate approach for the dictionary update, with warm restart,
- Use a mini-batch.



# Online Dictionary Learning

## Proposed approach

### Which guarantees do we have?

Under a few reasonable assumptions,

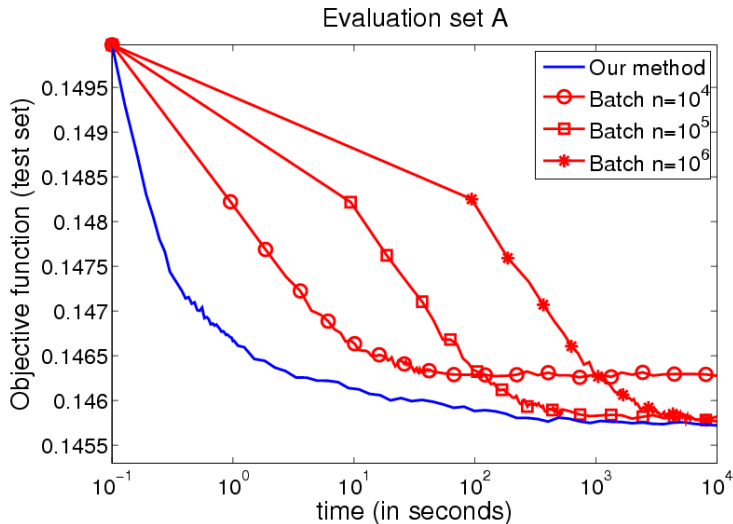
- we build a surrogate function  $\hat{f}_t$  of the expected cost  $f$  verifying

$$\lim_{t \rightarrow +\infty} \hat{f}_t(\mathbf{D}_t) - f(\mathbf{D}_t) = 0,$$

- $\mathbf{D}_t$  is asymptotically close to a stationary point.

# Online Dictionary Learning

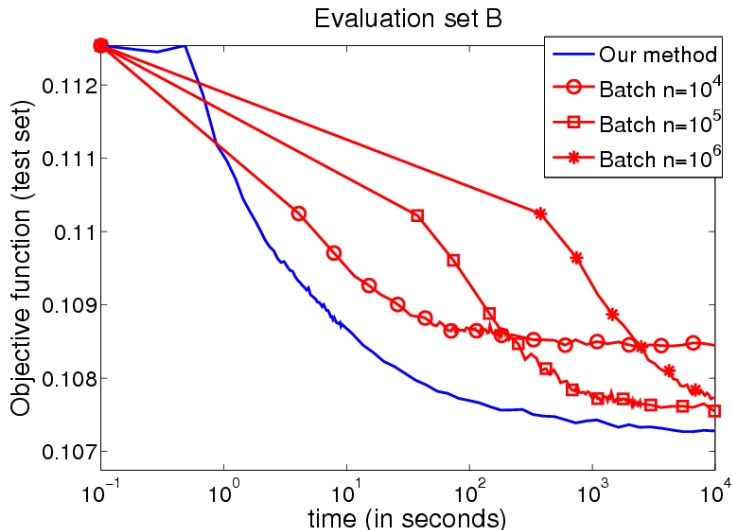
Experimental results, batch vs online



$$m = 8 \times 8, k = 256$$

# Online Dictionary Learning

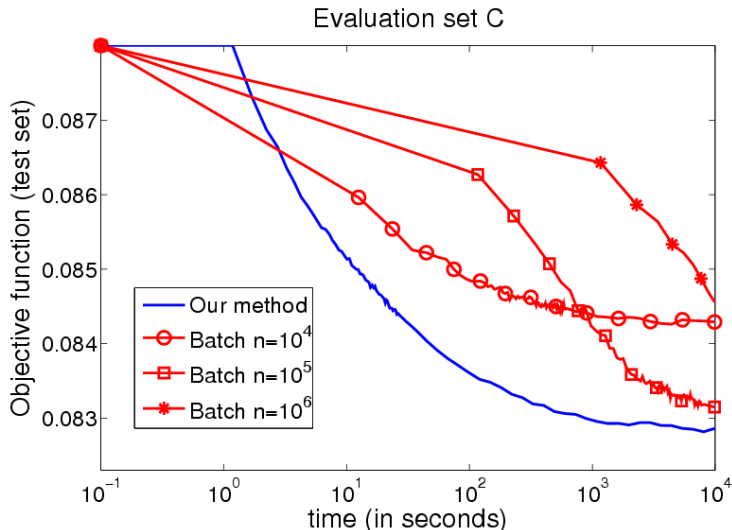
Experimental results, batch vs online



$$m = 12 \times 12 \times 3, k = 512$$

# Online Dictionary Learning

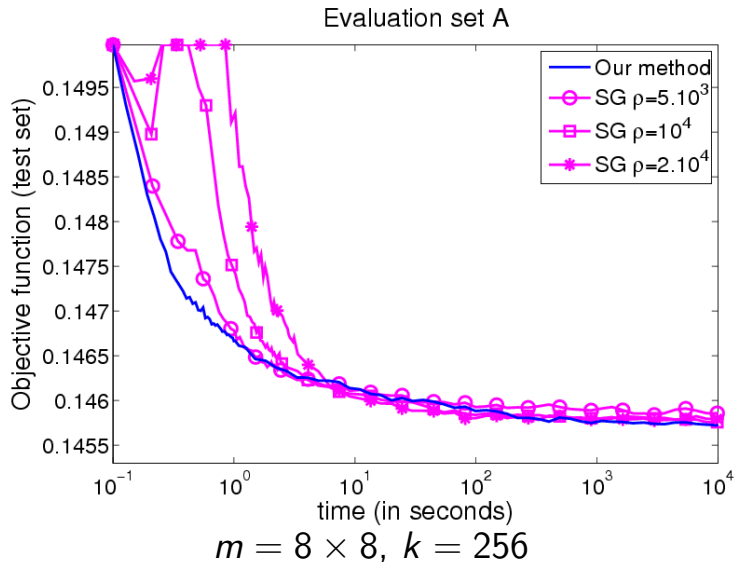
Experimental results, batch vs online



$$m = 16 \times 16, k = 1024$$

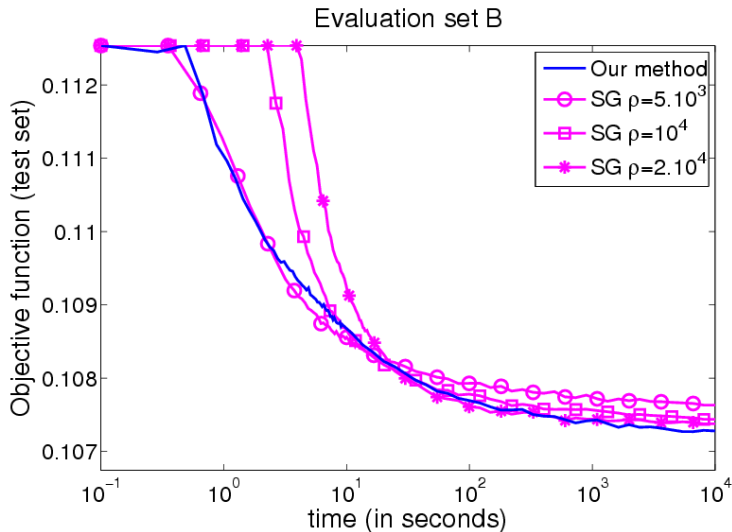
# Online Dictionary Learning

Experimental results, ODL vs SGD



# Online Dictionary Learning

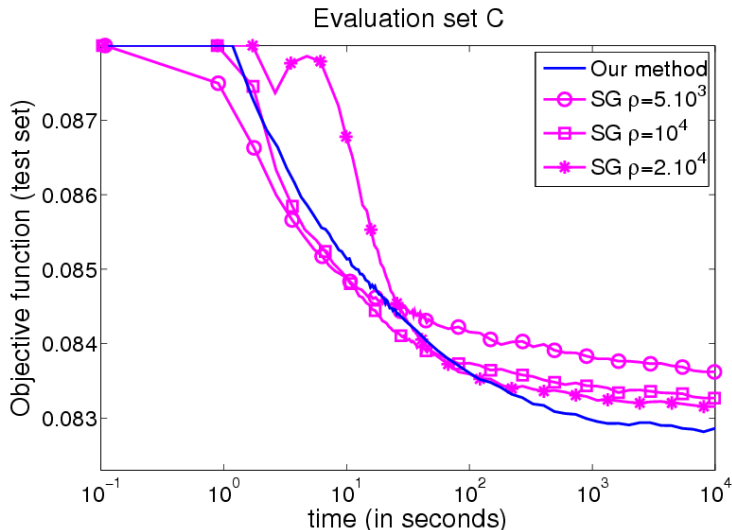
Experimental results, ODL vs SGD



$$m = 12 \times 12 \times 3, k = 512$$

# Online Dictionary Learning


Experimental results, ODL vs SGD



$$m = 16 \times 16, k = 1024$$

# Online Dictionary Learning

## Inpainting a 12-Mpixel photograph



THE SALINAS VALLEY is in Northern California. It is a long narrow scale 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 load 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 eyes. The memory of odors is very rich.

I remember that the Gabilan Mountains to the east of the valley were light gray 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 west. 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, so 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 pl...



# Online Dictionary Learning

Inpainting a 12-Mpixel photograph



# Online Dictionary Learning

Inpainting a 12-Mpixel photograph



# Online Dictionary Learning

Inpainting a 12-Mpixel photograph



1 The Dictionary Learning Problem

2 Online Dictionary Learning

3 Extensions

# Extension to NMF and sparse PCA

## NMF extension

$$\min_{\substack{\boldsymbol{\alpha} \in \mathbb{R}^{k \times n} \\ \mathbf{D} \in \mathcal{C}}} \sum_{i=1}^n \frac{1}{2} \|\mathbf{x}_i - \mathbf{D}\boldsymbol{\alpha}_i\|_2^2 \quad \text{s.t.} \quad \boldsymbol{\alpha}_i \geq 0, \quad \mathbf{D} \geq 0.$$

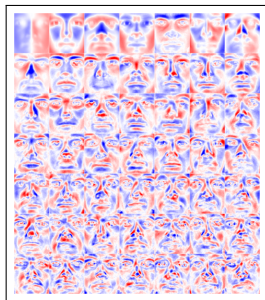
## SPCA extension

$$\min_{\substack{\boldsymbol{\alpha} \in \mathbb{R}^{k \times n} \\ \mathbf{D} \in \mathcal{C}'}} \sum_{i=1}^n \frac{1}{2} \|\mathbf{x}_i - \mathbf{D}\boldsymbol{\alpha}_i\|_2^2 + \lambda \|\boldsymbol{\alpha}_1\|_1$$

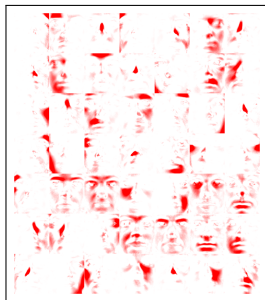
$$\mathcal{C}' \triangleq \{\mathbf{D} \in \mathbb{R}^{m \times k} \quad \text{s.t.} \quad \forall j \quad \|\mathbf{d}_j\|_2^2 + \gamma \|\mathbf{d}_j\|_1 \leq 1\}.$$

# Extension to NMF and sparse PCA

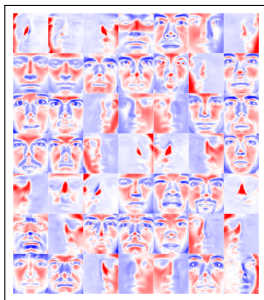
Faces: Extended Yale Database B



(a) PCA



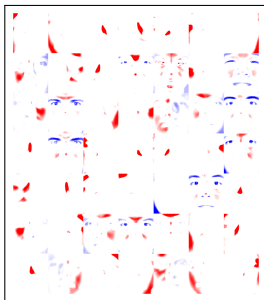
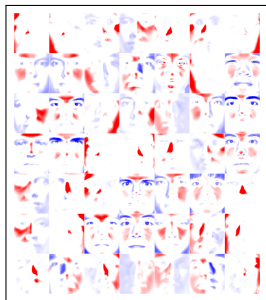
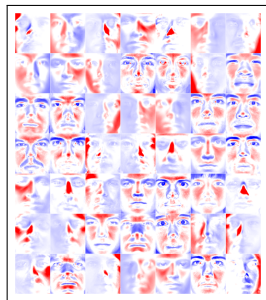
(b) NNMF



(c) DL

# Extension to NMF and sparse PCA

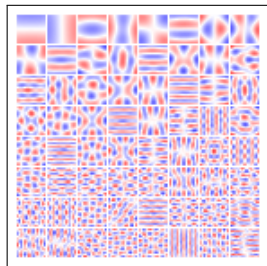
Faces: Extended Yale Database B



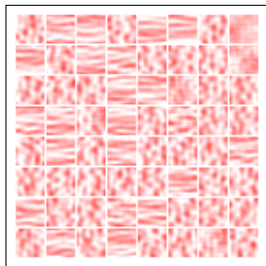
(d) SPCA,  $\tau = 70\%$  (e) SPCA,  $\tau = 30\%$  (f) SPCA,  $\tau = 10\%$

# Extension to NMF and sparse PCA

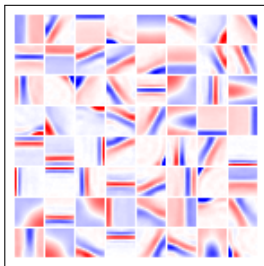
## Natural Patches



(a) PCA



(b) NNMF

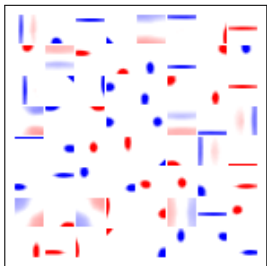
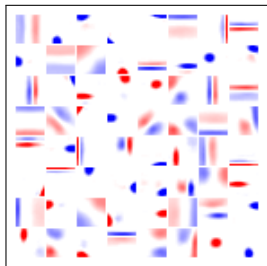
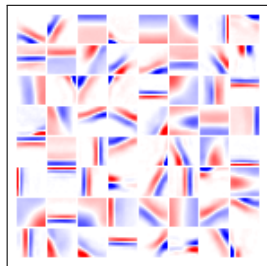


(c) DL



# Extension to NMF and sparse PCA

## Natural Patches



(d) SPCA,  $\tau = 70\%$  (e) SPCA,  $\tau = 30\%$  (f) SPCA,  $\tau = 10\%$

## Take-home message

- Online techniques are adapted to the dictionary learning problem.
- Our method makes some large-scale image processing tasks tractable—...
- ...— and extends to various matrix factorization problems.