

Représentations parcimonieuses pour le traitement d'image et la vision artificielle

Julien Mairal

Department of Statistics, University of California, Berkeley
INRIA - projet Willow, Paris



Journées ORASIS, Praz-sur-Arly
7 Juin 2011



Remerciements



Francis
Bach



Jean
Ponce



Rodolphe
Jenatton



Guillaume
Obozinski



Guillermo
Sapiro



Miki
Elad



Andrew
Zisserman



Florent
Couzinié-Devy



Y-Lan
Boureau



Willow
people

What is this lecture about?

- **Why sparsity, what for and how?**
- **Signal and image processing:** Restoration, reconstruction.
- **Machine learning:** Selecting relevant features.
- **Computer vision:** Modelling image patches and image descriptors.
- **Optimization:** Solving challenging problems.

- 1 Image Processing Applications
- 2 Sparse Linear Models and Dictionary Learning
- 3 Computer Vision Applications

1 Image Processing Applications

- Image Denoising
- Inpainting, Demosaicking
- Video Processing
- Other Applications

2 Sparse Linear Models and Dictionary Learning

3 Computer Vision Applications

The Image Denoising Problem



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

Sparse representations for image restoration

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

Energy minimization problem - MAP estimation

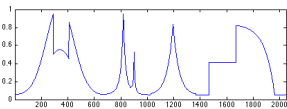
$$E(\mathbf{x}) = \underbrace{\frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2}_{\text{relation to measurements}} + \underbrace{Pr(\mathbf{x})}_{\text{image model (-log prior)}}$$

Some classical priors

- Smoothness $\lambda \|\mathcal{L}\mathbf{x}\|_2^2$
- Total variation $\lambda \|\nabla\mathbf{x}\|_1^2$
- MRF priors
- ...

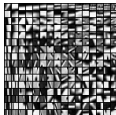
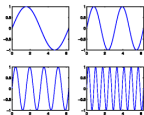
What is a Sparse Linear Model?

Let \mathbf{x} in \mathbb{R}^m be a signal.



Let $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_p] \in \mathbb{R}^{m \times p}$ be a set of normalized “basis vectors”.

We call it **dictionary**.



\mathbf{D} is “adapted” to \mathbf{y} if it can represent it with a few basis vectors—that is, there exists a **sparse vector** α in \mathbb{R}^p such that $\mathbf{y} \approx \mathbf{D}\alpha$. We call α the **sparse code**.

$$\underbrace{\begin{pmatrix} \mathbf{y} \end{pmatrix}}_{\mathbf{y} \in \mathbb{R}^m} \approx \underbrace{\begin{pmatrix} \mathbf{d}_1 & \mathbf{d}_2 & \dots & \mathbf{d}_p \end{pmatrix}}_{\mathbf{D} \in \mathbb{R}^{m \times p}} \underbrace{\begin{pmatrix} \alpha[1] \\ \alpha[2] \\ \vdots \\ \alpha[p] \end{pmatrix}}_{\alpha \in \mathbb{R}^p, \text{ sparse}}$$

First Important Idea

Why Sparsity?

A dictionary can be good for representing a class of signals, but not for representing white Gaussian noise.

The Sparse Decomposition Problem

$$\min_{\alpha \in \mathbb{R}^p} \underbrace{\frac{1}{2} \|\mathbf{y} - \mathbf{D}\alpha\|_2^2}_{\text{data fitting term}} + \underbrace{\lambda\psi(\alpha)}_{\text{sparsity-inducing regularization}}$$

ψ induces sparsity in α . It can be

- the ℓ_0 “pseudo-norm”. $\|\alpha\|_0 \triangleq \#\{i \text{ s.t. } \alpha[i] \neq 0\}$ (NP-hard)
- the ℓ_1 norm. $\|\alpha\|_1 \triangleq \sum_{i=1}^p |\alpha[i]|$ (convex),
- ...

This is a **selection** problem. When ψ is the ℓ_1 -norm, the problem is called Lasso [Tibshirani, 1996] or basis pursuit [Chen et al., 1999]

Sparse representations for image restoration

Designed dictionaries

[Haar, 1910], [Zweig, Morlet, Grossman ~70s], [Meyer, Mallat, Daubechies, Coifman, Donoho, Candes ~80s-today]... (see [Mallat, 1999])

Wavelets, Curvelets, Wedgelets, Bandlets, ... lets

Learned dictionaries of patches

[Olshausen and Field, 1997], [Engan et al., 1999], [Lewicki and Sejnowski, 2000], [Aharon et al., 2006], [Roth and Black, 2005], [Lee et al., 2007]

$$\min_{\alpha_i, \mathbf{D} \in \mathcal{D}} \sum_i \underbrace{\frac{1}{2} \|\mathbf{y}_i - \mathbf{D}\alpha_i\|_2^2}_{\text{reconstruction}} + \underbrace{\lambda \psi(\alpha_i)}_{\text{sparsity}}$$

- $\psi(\alpha) = \|\alpha\|_0$ (“ ℓ_0 pseudo-norm”)
- $\psi(\alpha) = \|\alpha\|_1$ (ℓ_1 norm)

Sparse representations for image restoration

Solving the denoising problem

[Elad and Aharon, 2006]

- Extract all overlapping 8×8 patches \mathbf{y}_i .
- Solve a matrix factorization problem:

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

with $n > 100,000$

- Average the reconstruction of each patch.

Sparse representations for image restoration

K-SVD: [Elad and Aharon, 2006]

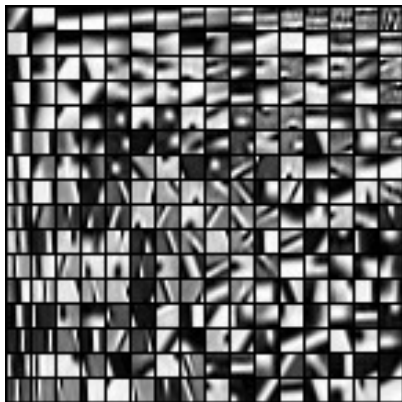
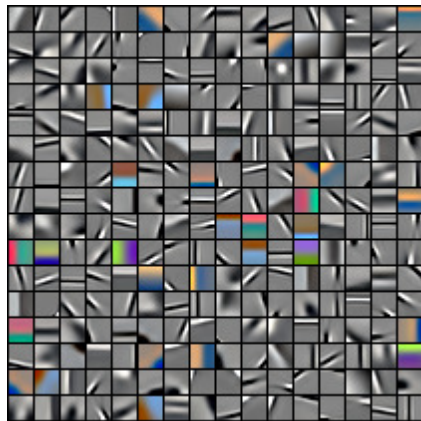
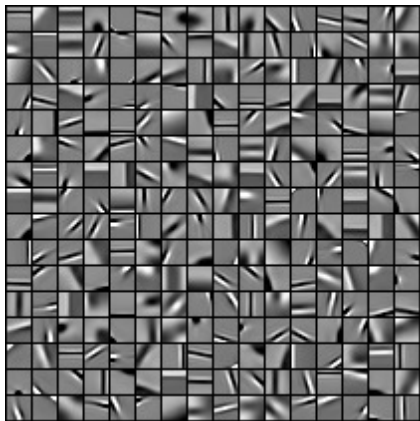


Figure: Dictionary trained on a noisy version of the image boat.

Sparse representations for image restoration

Grayscale vs color image patches

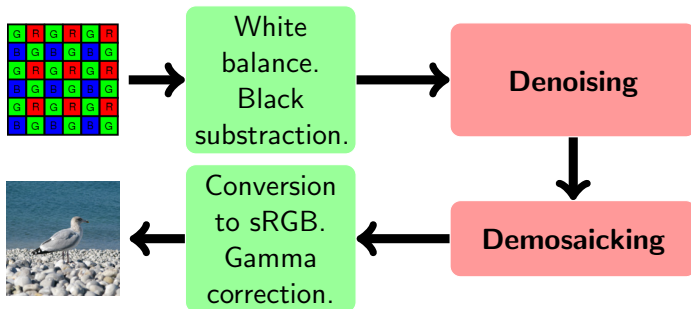


Sparse representations for image restoration

Inpainting, Demosaicking

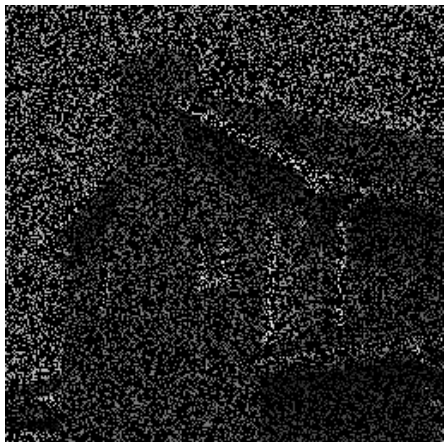
$$\min_{\mathbf{D} \in \mathcal{D}, \alpha} \sum_i \frac{1}{2} \|\beta_i(\mathbf{y}_i - \mathbf{D}\alpha_i)\|_2^2 + \lambda_i \psi(\alpha_i)$$

RAW Image Processing



Sparse representations for image restoration

[Mairal, Sapiro, and Elad, 2008d]



Sparse representations for image restoration

Inpainting, [Mairal, Elad, and Sapiro, 2008b]



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-

Sparse representations for image restoration

Inpainting, [Mairal, Elad, and Sapiro, 2008b]



Sparse representations for video restoration

Key ideas for video processing

[Protter and Elad, 2009]

- Using a 3D dictionary.
- Processing of many frames at the same time.
- Dictionary propagation.

Sparse representations for image restoration

Inpainting, [Mairal, Sapiro, and Elad, 2008d]

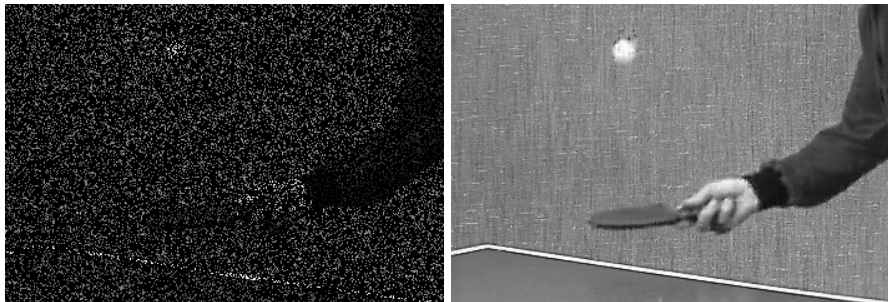


Figure: Inpainting results.

Sparse representations for image restoration

Inpainting, [Mairal, Sapiro, and Elad, 2008d]

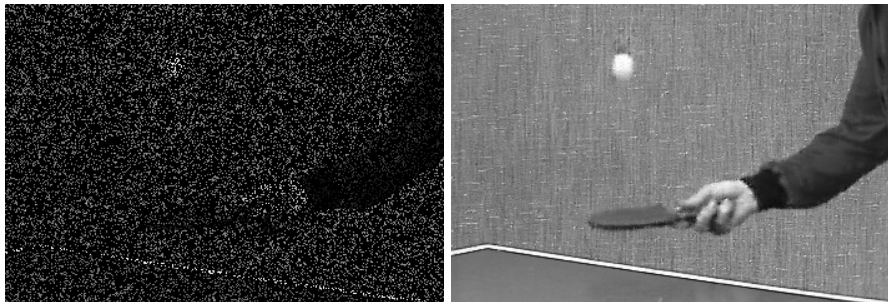


Figure: Inpainting results.

Sparse representations for image restoration

Inpainting, [Mairal, Sapiro, and Elad, 2008d]

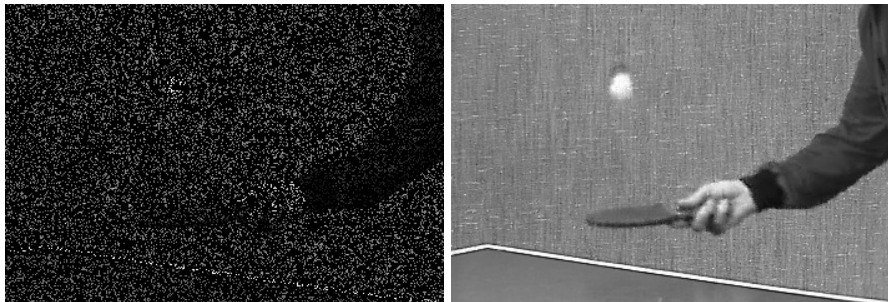


Figure: Inpainting results.

Sparse representations for image restoration

Inpainting, [Mairal, Sapiro, and Elad, 2008d]

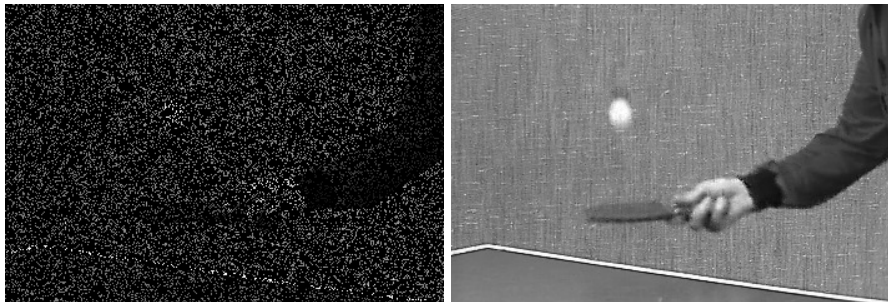


Figure: Inpainting results.

Sparse representations for image restoration

Inpainting, [Mairal, Sapiro, and Elad, 2008d]

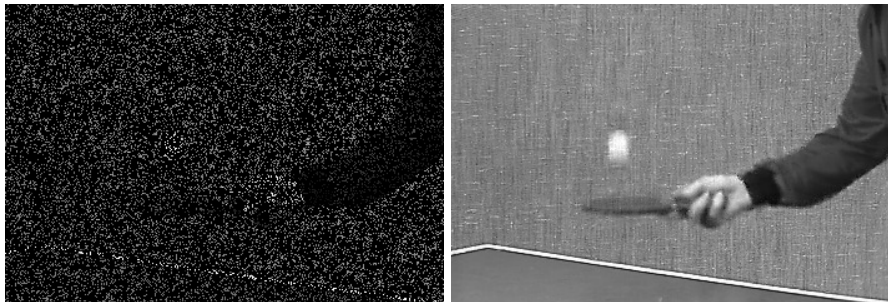


Figure: Inpainting results.

Sparse representations for image restoration

Color video denoising, [Mairal, Sapiro, and Elad, 2008d]

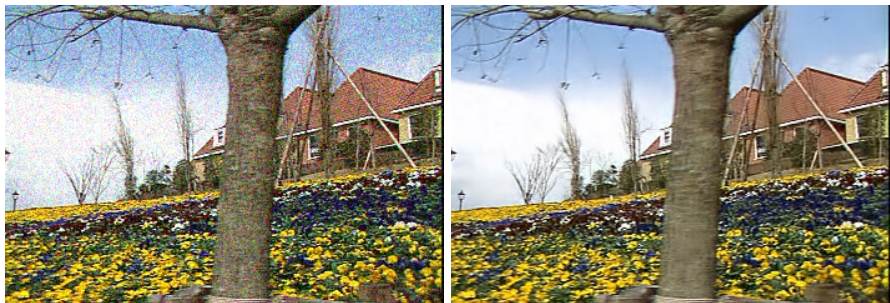


Figure: Denoising results. $\sigma = 25$

Sparse representations for image restoration

Color video denoising, [Mairal, Sapiro, and Elad, 2008d]



Figure: Denoising results. $\sigma = 25$

Sparse representations for image restoration

Color video denoising, [Mairal, Sapiro, and Elad, 2008d]



Figure: Denoising results. $\sigma = 25$

Sparse representations for image restoration

Color video denoising, [Mairal, Sapiro, and Elad, 2008d]



Figure: Denoising results. $\sigma = 25$

Sparse representations for image restoration

Color video denoising, [Mairal, Sapiro, and Elad, 2008d]



Figure: Denoising results. $\sigma = 25$

Digital Zooming

Couzinie-Devy, 2010, Original



Digital Zooming

Couzinie-Devy, 2010, Bicubic



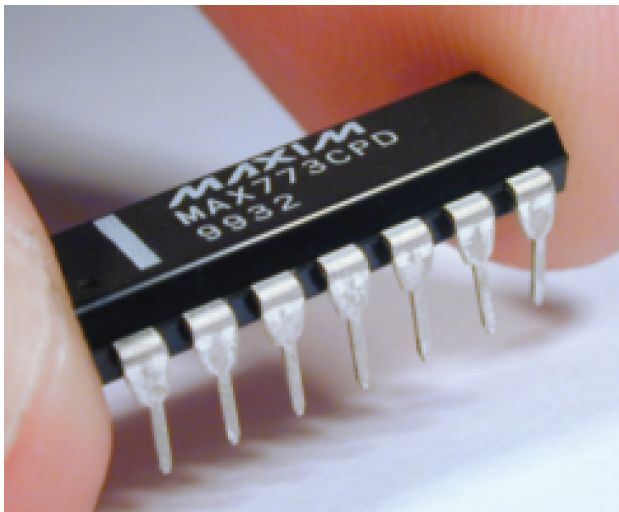
Digital Zooming

Couzinie-Devy, 2010, Proposed method



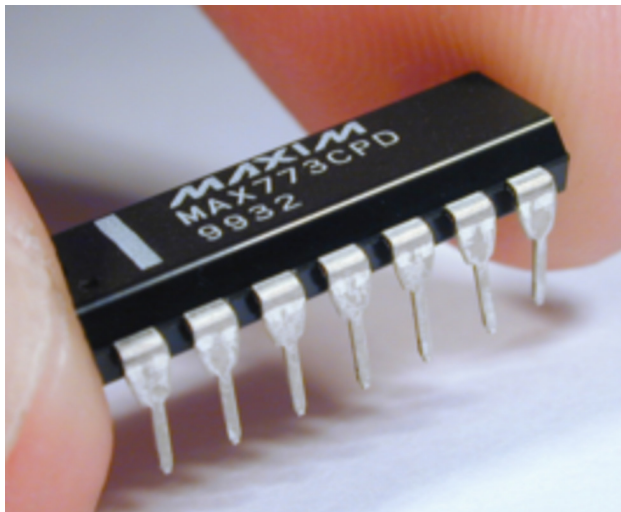
Digital Zooming

Couzinie-Devy, 2010, Original



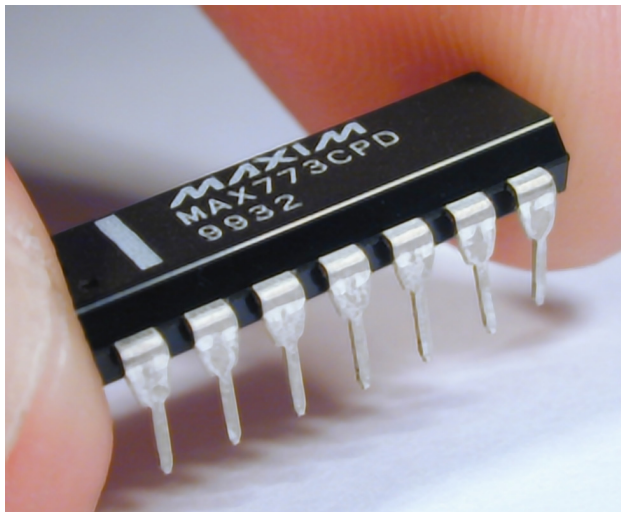
Digital Zooming

Couzinie-Devy, 2010, Bicubic



Digital Zooming

Couzinie-Devy, 2010, Proposed approach



Inverse half-toning

Original



Inverse half-toning

Reconstructed image



Inverse half-toning

Original



Inverse half-toning

Reconstructed image



Inverse half-toning

Original



Copyright © 1987 by AcademySoft-ELORG. Macintosh version © 1988 by Sphere, Inc.

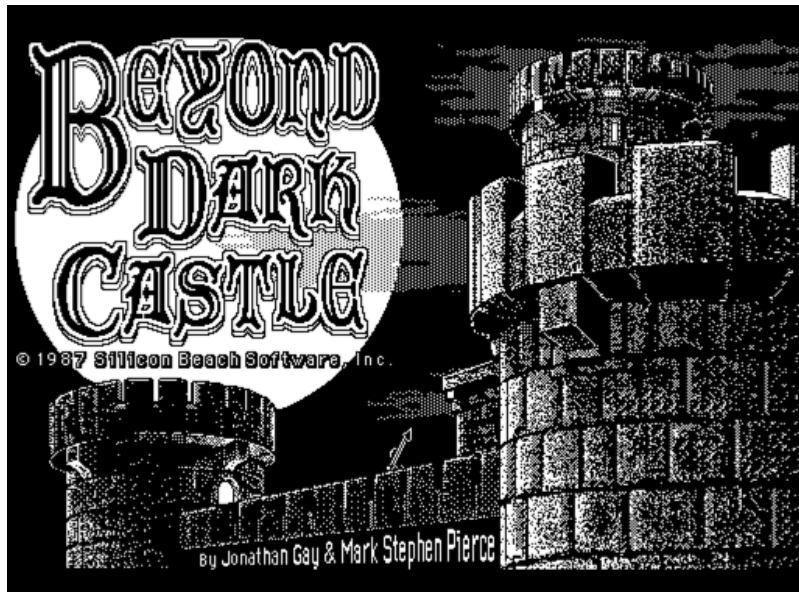
Inverse half-toning

Reconstructed image



Inverse half-toning

Original



Inverse half-toning

Reconstructed image



Inverse half-toning

Original



Inverse half-toning

Reconstructed image



Important messages

- Patch-based approaches are achieving state-of-the-art results for many image processing task.
- Dictionary Learning adapts to the data you want to restore.
- Dictionary Learning is well adapted to data that admit sparse representation. **Sparsity is for sparse data only.**

Next topics

- A bit of machine learning.
- Why does the ℓ_1 -norm induce sparsity?
- Some properties of the Lasso.
- Links between dictionary learning and matrix factorization techniques.
- A simple algorithm for learning dictionaries.

- 1 Image Processing Applications
- 2 Sparse Linear Models and Dictionary Learning
 - The machine learning point of view
 - Why does the ℓ_1 -norm induce sparsity?
 - Dictionary Learning and Matrix Factorization
- 3 Computer Vision Applications

Sparse Linear Model: Machine Learning Point of View

Let $(y^i, \mathbf{x}^i)_{i=1}^n$ be a training set, where the vectors \mathbf{x}^i are in \mathbb{R}^p and are called features. The scalars y^i are in

- $\{-1, +1\}$ for **binary** classification problems.
- $\{1, \dots, N\}$ for **multiclass** classification problems.
- \mathbb{R} for **regression** problems.

In a linear model, one assumes a relation $y \approx \mathbf{w}^\top \mathbf{x}$ (or $y \approx \text{sign}(\mathbf{w}^\top \mathbf{x})$), and solves

$$\min_{\mathbf{w} \in \mathbb{R}^p} \underbrace{\frac{1}{n} \sum_{i=1}^n \ell(y^i, \mathbf{w}^\top \mathbf{x}^i)}_{\text{data-fitting}} + \underbrace{\lambda \psi(\mathbf{w})}_{\text{regularization}} .$$

Sparse Linear Models: Machine Learning Point of View

A few examples:

Ridge regression:
$$\min_{\mathbf{w} \in \mathbb{R}^p} \frac{1}{2n} \sum_{i=1}^n (y^i - \mathbf{w}^\top \mathbf{x}^i)^2 + \lambda \|\mathbf{w}\|_2^2.$$

Linear SVM:
$$\min_{\mathbf{w} \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^n \max(0, 1 - y^i \mathbf{w}^\top \mathbf{x}^i) + \lambda \|\mathbf{w}\|_2^2.$$

Logistic regression:
$$\min_{\mathbf{w} \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^n \log \left(1 + e^{-y^i \mathbf{w}^\top \mathbf{x}^i} \right) + \lambda \|\mathbf{w}\|_2^2.$$

The squared ℓ_2 -norm induces **smoothness** in \mathbf{w} . When one knows in advance that \mathbf{w} should be sparse, one should use a **sparsity-inducing** regularization such as the ℓ_1 -norm. [Chen et al., 1999, Tibshirani, 1996]

The purpose of the regularization is to add **additional a-priori knowledge** in the regularization.

Sparse Linear Models: the Lasso

- Signal processing: \mathbf{D} is a dictionary in $\mathbb{R}^{n \times p}$,

$$\min_{\alpha \in \mathbb{R}^p} \frac{1}{2} \|\mathbf{y} - \mathbf{D}\alpha\|_2^2 + \lambda \|\alpha\|_1.$$

- Machine Learning:

$$\min_{\mathbf{w} \in \mathbb{R}^p} \frac{1}{2n} \sum_{i=1}^n (y^i - \mathbf{x}^{i\top} \mathbf{w})^2 + \lambda \|\mathbf{w}\|_1 = \min_{\mathbf{w} \in \mathbb{R}^p} \frac{1}{2n} \|\mathbf{y} - \mathbf{X}^\top \mathbf{w}\|_2^2 + \lambda \|\mathbf{w}\|_1,$$

with $\mathbf{X} \triangleq [\mathbf{x}^1, \dots, \mathbf{x}^n]$, and $\mathbf{y} \triangleq [y^1, \dots, y^n]^\top$.

Useful tool in signal processing, machine learning, statistics, ... as long as one wishes to **select** features.

Why does the ℓ_1 -norm induce sparsity?

Exemple: quadratic problem in 1D

$$\min_{\alpha \in \mathbb{R}} \frac{1}{2}(y - \alpha)^2 + \lambda|\alpha|$$

Piecewise quadratic function with a kink at zero.

Derivative at 0_+ : $g_+ = -y + \lambda$ and 0_- : $g_- = -y - \lambda$.

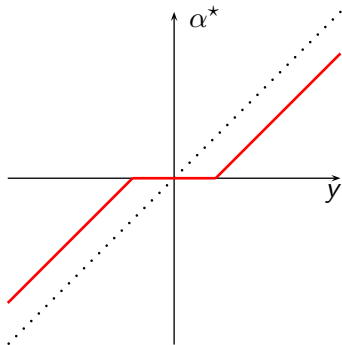
Optimality conditions. α is optimal iff:

- $|\alpha| > 0$ and $(y - \alpha) + \lambda \text{sign}(\alpha) = 0$
- $\alpha = 0$ and $g_+ \geq 0$ and $g_- \leq 0$

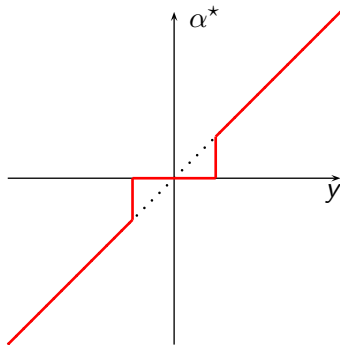
The solution is a **soft-thresholding**:

$$\alpha^* = \text{sign}(y)(|y| - \lambda)^+.$$

Why does the ℓ_1 -norm induce sparsity?



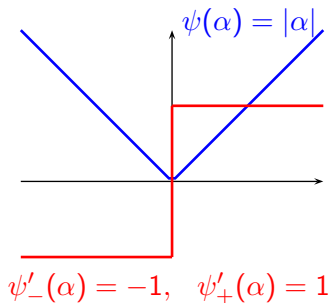
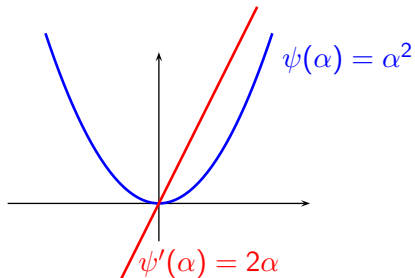
(a) soft-thresholding operator



(b) hard-thresholding operator

Why does the ℓ_1 -norm induce sparsity?

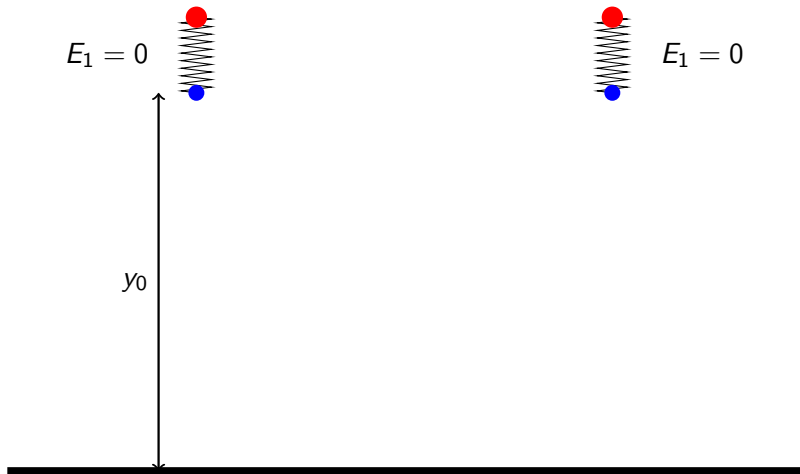
Analysis of the norms in 1D



The gradient of the ℓ_2 -norm vanishes when α get close to 0. On its differentiable part, the norm of the gradient of the ℓ_1 -norm is constant.

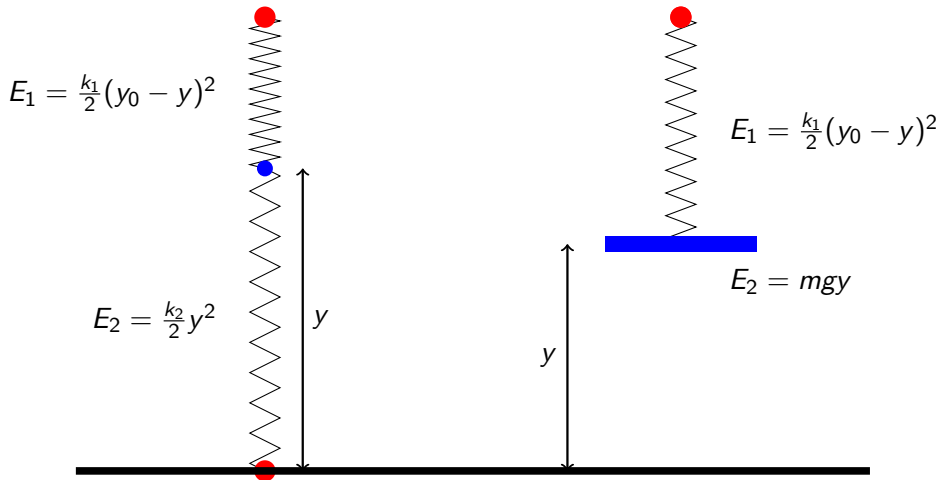
Why does the ℓ_1 -norm induce sparsity?

Physical illustration



Why does the ℓ_1 -norm induce sparsity?

Physical illustration



Why does the ℓ_1 -norm induce sparsity?

Physical illustration

$$E_1 = \frac{k_1}{2}(y_0 - y)^2$$

$$E_2 = \frac{k_2}{2}y^2$$

y

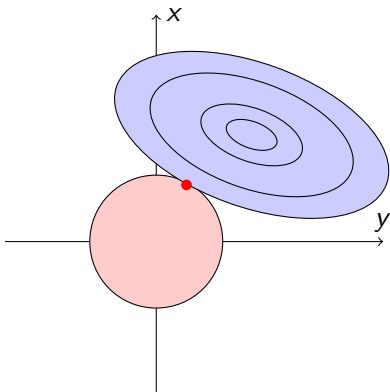
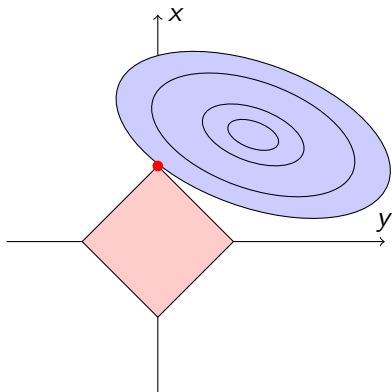
$$E_1 = \frac{k_1}{2}(y_0 - y)^2$$

$y = 0 !!$

$$E_2 = mgy$$

Why does the ℓ_1 -norm induce sparsity?

Geometric explanation



$$\min_{\alpha \in \mathbb{R}^p} \frac{1}{2} \|\mathbf{y} - \mathbf{D}\alpha\|_2^2 + \lambda \|\alpha\|_1$$
$$\min_{\alpha \in \mathbb{R}^p} \|\mathbf{y} - \mathbf{D}\alpha\|_2^2 \quad \text{s.t.} \quad \|\alpha\|_1 \leq T.$$

Important property of the Lasso

Piecewise linearity of the regularization path

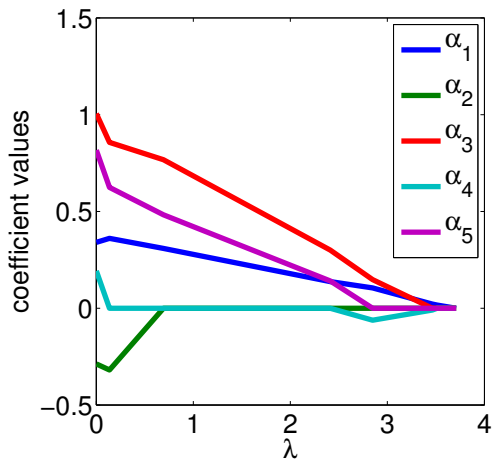


Figure: Regularization path of the Lasso

Optimization for Dictionary Learning

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

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

- Classical optimization alternates between \mathbf{D} and α .
- Good results, but **slow!**
- **Instead use online learning [Mairal et al., 2009]**

Optimization for Dictionary Learning

Inpainting a 12-Mpixel photograph

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 games 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 blown grass love. The Santa Lucia 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 flat 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...

Optimization for Dictionary Learning

Inpainting a 12-Mpixel photograph



Optimization for Dictionary Learning

Inpainting a 12-Mpixel photograph



Optimization for Dictionary Learning

Inpainting a 12-Mpixel photograph



Matrix Factorization Problems and Dictionary Learning

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

can be rewritten

$$\min_{\substack{\alpha \in \mathbb{R}^{p \times n} \\ \mathbf{D} \in \mathcal{D}}} \frac{1}{2} \|\mathbf{Y} - \mathbf{D}\alpha\|_F^2 + \lambda \|\alpha\|_1,$$

where $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n]$ and $\alpha = [\alpha_1, \dots, \alpha_n]$.

Matrix Factorization Problems and Dictionary Learning

PCA

$$\min_{\substack{\alpha \in \mathbb{R}^{p \times n} \\ \mathbf{D} \in \mathbb{R}^{m \times p}}} \frac{1}{2} \|\mathbf{Y} - \mathbf{D}\alpha\|_F^2 \quad \text{s.t.} \quad \mathbf{D}^\top \mathbf{D} = \mathbf{I} \text{ and } \alpha\alpha^\top \text{ is diagonal.}$$

$\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_p]$ are the principal components.

Matrix Factorization Problems and Dictionary Learning

Hard clustering

$$\min_{\substack{\alpha \in \{0,1\}^{p \times n} \\ \mathbf{D} \in \mathbb{R}^{m \times p}}} \frac{1}{2} \|\mathbf{Y} - \mathbf{D}\alpha\|_F^2 \quad \text{s.t.} \quad \forall i \in \{1, \dots, p\}, \sum_{j=1}^p \alpha_i[j] = 1.$$

$\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_p]$ are the centroids of the p clusters.

Matrix Factorization Problems and Dictionary Learning

Soft clustering

$$\min_{\substack{\alpha \in \mathbb{R}_+^{p \times n} \\ \mathbf{D} \in \mathbb{R}^{m \times p}}} \frac{1}{2} \|\mathbf{Y} - \mathbf{D}\alpha\|_F^2, \quad \text{s.t.} \quad \forall i \in \{1, \dots, p\}, \sum_{j=1}^p \alpha_i[j] = 1.$$

$\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_p]$ are the centroids of the p clusters.

Matrix Factorization Problems and Dictionary Learning

Non-negative matrix factorization [Lee and Seung, 2001]

$$\min_{\substack{\alpha \in \mathbb{R}_+^{p \times n} \\ \mathbf{D} \in \mathbb{R}_+^{m \times p}}} \frac{1}{2} \|\mathbf{Y} - \mathbf{D}\alpha\|_F^2$$

Matrix Factorization Problems and Dictionary Learning

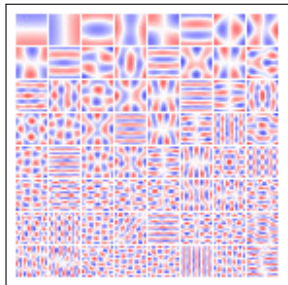
NMF+sparsity?

$$\min_{\substack{\alpha \in \mathbb{R}_+^{p \times n} \\ \mathbf{D} \in \mathbb{R}_+^{m \times p}}} \frac{1}{2} \|\mathbf{Y} - \mathbf{D}\alpha\|_F^2 + \lambda \|\alpha\|_1.$$

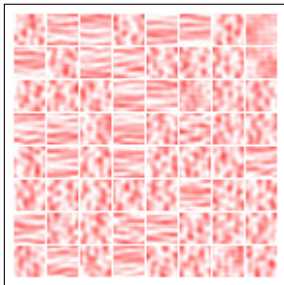
Most of these formulations can be addressed the same types of algorithms.

Matrix Factorization Problems and Dictionary Learning

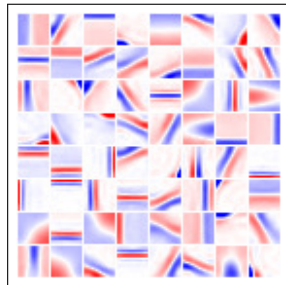
Natural Patches



(a) PCA



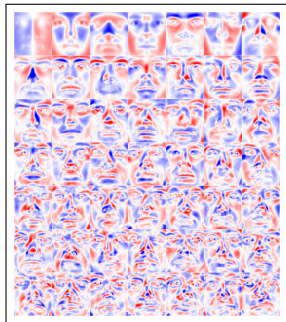
(b) NNMF



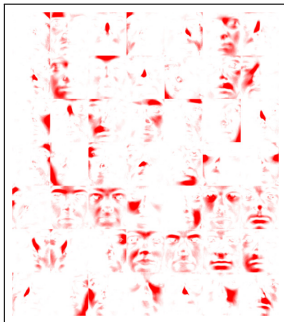
(c) DL

Matrix Factorization Problems and Dictionary Learning

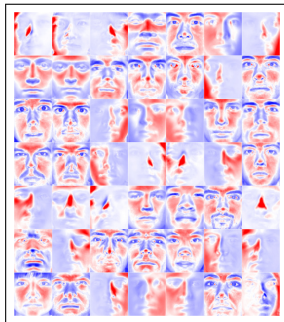
Faces



(d) PCA



(e) NNMF



(f) DL

Important messages

- The ℓ_1 -norm induces sparsity and shrinks the coefficients (soft-thresholding)
- The regularization path of the Lasso is piecewise linear.
- Learning the dictionary is simple, fast and scalable.
- Dictionary learning is related to several matrix factorization problems.

Software SPAMS is available for all of this:

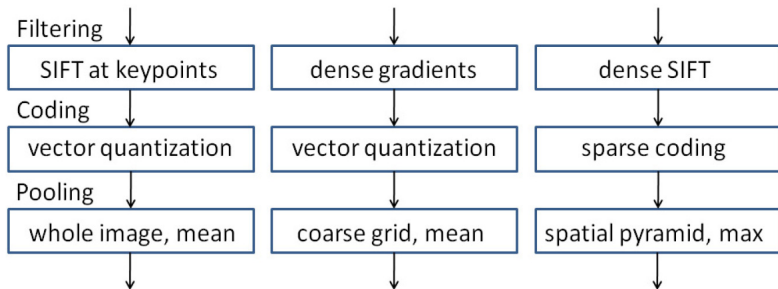
www.di.ens.fr/willow/SPAMS/.

Next topics: Computer Vision

- Intriguing results on the use of dictionary learning for bags of words.
- Modelling the local appearance of image patches.

- 1 Image Processing Applications
- 2 Sparse Linear Models and Dictionary Learning
- 3 Computer Vision Applications
 - Learning codebooks for image classification
 - Modelling the local appearance of image patches

Learning Codebooks for Image Classification



Idea

Replacing Vector Quantization by Learned Dictionaries!

- unsupervised: [Yang et al., 2009]
- supervised: [Boureau et al., 2010, Yang et al., 2010]

Learning Codebooks for Image Classification

Let an image be represented by a set of low-level descriptors \mathbf{y}_i at N locations identified with their indices $i = 1, \dots, N$.

- hard-quantization:

$$\mathbf{y}_i \approx \mathbf{D}\boldsymbol{\alpha}_i, \quad \alpha_i \in \{0, 1\}^p \quad \text{and} \quad \sum_{j=1}^p \alpha_i[j] = 1$$

- soft-quantization:

$$\alpha_i[j] = \frac{e^{-\beta \|\mathbf{y}_i - \mathbf{d}_j\|_2^2}}{\sum_{k=1}^p e^{-\beta \|\mathbf{y}_i - \mathbf{d}_k\|_2^2}}$$

- sparse coding:

$$\mathbf{y}_i \approx \mathbf{D}\boldsymbol{\alpha}_i, \quad \boldsymbol{\alpha}_i = \arg \min_{\boldsymbol{\alpha}} \frac{1}{2} \|\mathbf{y}_i - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1$$

Learning Codebooks for Image Classification

Table from Boureau et al. [2010]

Method	Caltech-101, 30 training examples		15 Scenes, 100 training examples	
	Average Pool	Max Pool	Average Pool	Max Pool
Results with basic features, SIFT extracted each 8 pixels				
Hard quantization, linear kernel	51.4 ± 0.9 [256]	64.3 ± 0.9 [256]	73.9 ± 0.9 [1024]	80.1 ± 0.6 [1024]
Hard quantization, intersection kernel	64.2 ± 1.0 [256] (1)	64.3 ± 0.9 [256]	80.8 ± 0.4 [256] (1)	80.1 ± 0.6 [1024]
Soft quantization, linear kernel	57.9 ± 1.5 [1024]	69.0 ± 0.8 [256]	75.6 ± 0.5 [1024]	81.4 ± 0.6 [1024]
Soft quantization, intersection kernel	66.1 ± 1.2 [512] (2)	70.6 ± 1.0 [1024]	81.2 ± 0.4 [1024] (2)	83.0 ± 0.7 [1024]
Sparse codes, linear kernel	61.3 ± 1.3 [1024]	71.5 ± 1.1 [1024] (3)	76.9 ± 0.6 [1024]	83.1 ± 0.6 [1024] (3)
Sparse codes, intersection kernel	70.3 ± 1.3 [1024]	71.8 ± 1.0 [1024] (4)	83.2 ± 0.4 [1024]	84.1 ± 0.5 [1024] (4)
Results with macrofeatures and denser SIFT sampling				
Hard quantization, linear kernel	55.6 ± 1.6 [256]	70.9 ± 1.0 [1024]	74.0 ± 0.5 [1024]	80.1 ± 0.5 [1024]
Hard quantization, intersection kernel	68.8 ± 1.4 [512]	70.9 ± 1.0 [1024]	81.0 ± 0.5 [1024]	80.1 ± 0.5 [1024]
Soft quantization, linear kernel	61.6 ± 1.6 [1024]	71.5 ± 1.0 [1024]	76.4 ± 0.7 [1024]	81.5 ± 0.4 [1024]
Soft quantization, intersection kernel	70.1 ± 1.3 [1024]	73.2 ± 1.0 [1024]	81.8 ± 0.4 [1024]	83.0 ± 0.4 [1024]
Sparse codes, linear kernel	65.7 ± 1.4 [1024]	75.1 ± 0.9 [1024]	78.2 ± 0.7 [1024]	83.6 ± 0.4 [1024]
Sparse codes, intersection kernel	73.7 ± 1.3 [1024]	75.7 ± 1.1 [1024]	83.5 ± 0.4 [1024]	84.3 ± 0.5 [1024]

	Unsup	Discr
Linear	83.6 ± 0.4	84.9 ± 0.3
Intersect	84.3 ± 0.5	84.7 ± 0.4

Yang et al. [2009] have won the PASCAL VOC'09 challenge using this kind of technique.

Learning dictionaries with a discriminative cost function

Idea:

Let us consider 2 sets S_- , S_+ of signals representing 2 different classes. Each set should admit a dictionary best adapted to its reconstruction.

Classification procedure for a signal $\mathbf{y} \in \mathbb{R}^n$:

$$\min(\mathbf{R}^*(\mathbf{y}, \mathbf{D}_-), \mathbf{R}^*(\mathbf{y}, \mathbf{D}_+))$$

where

$$\mathbf{R}^*(\mathbf{y}, \mathbf{D}) = \min_{\alpha \in \mathbb{R}^p} \|\mathbf{y} - \mathbf{D}\alpha\|_2^2 \text{ s.t. } \|\alpha\|_0 \leq L.$$

“Reconstructive” training

$$\begin{cases} \min_{\mathbf{D}_-} \sum_{i \in S_-} \mathbf{R}^*(\mathbf{y}_i, \mathbf{D}_-) \\ \min_{\mathbf{D}_+} \sum_{i \in S_+} \mathbf{R}^*(\mathbf{y}_i, \mathbf{D}_+) \end{cases}$$

[Grosse et al., 2007], [Huang and Aviyente, 2006],
[Sprechmann et al., 2010] for unsupervised clustering (CVPR '10)

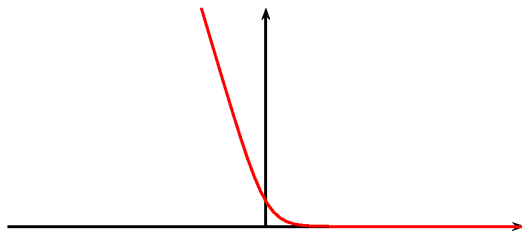
Learning dictionaries with a discriminative cost function

“Discriminative” training

[Mairal, Bach, Ponce, Sapiro, and Zisserman, 2008a]

$$\min_{\mathbf{D}_-, \mathbf{D}_+} \sum_i \mathcal{D} \left(\lambda z_i (\mathbf{R}^*(\mathbf{y}_i, \mathbf{D}_-) - \mathbf{R}^*(\mathbf{y}_i, \mathbf{D}_+)) \right),$$

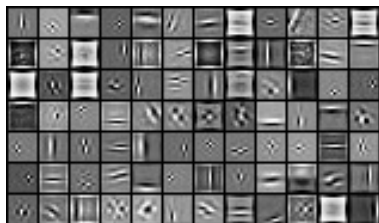
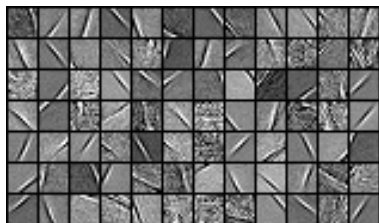
where $z_i \in \{-1, +1\}$ is the label of \mathbf{y}_i .



Logistic regression function

Learning dictionaries with a discriminative cost function

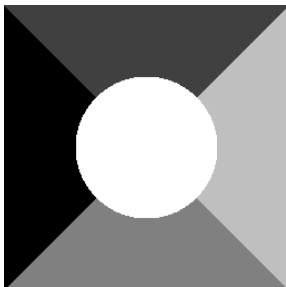
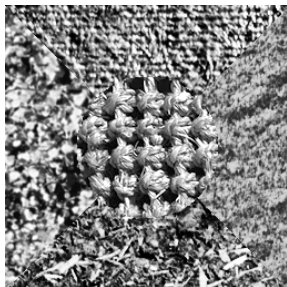
Examples of dictionaries



Top: reconstructive, Bottom: discriminative, Left: Bicycle, Right: Background.

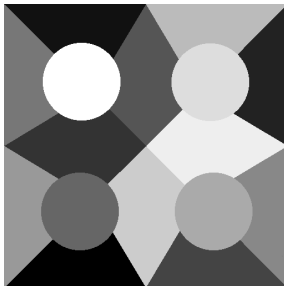
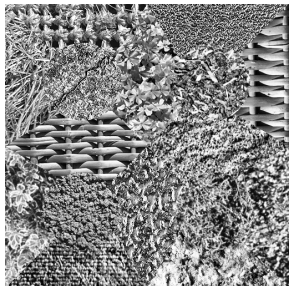
Learning dictionaries with a discriminative cost function

Texture segmentation



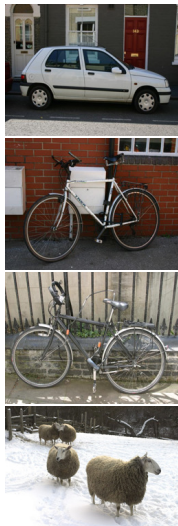
Learning dictionaries with a discriminative cost function

Texture segmentation



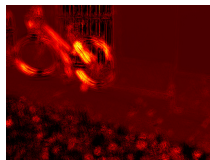
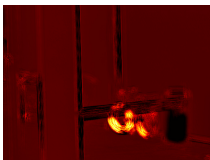
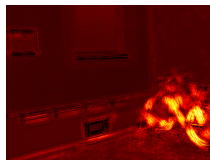
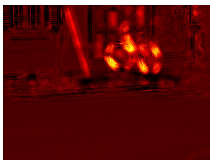
Learning dictionaries with a discriminative cost function

Pixelwise classification



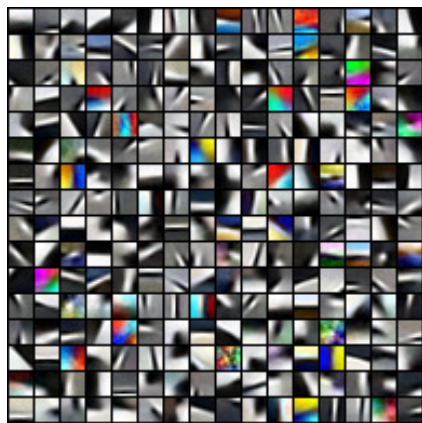
Learning dictionaries with a discriminative cost function

weakly-supervised pixel classification

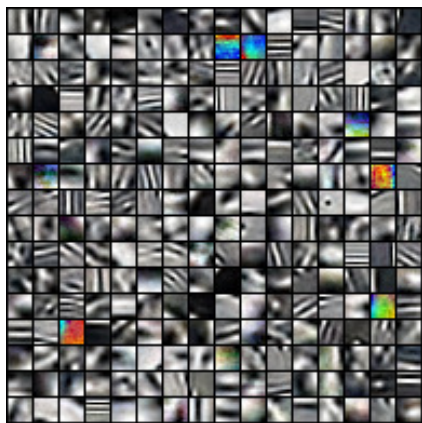


Application to edge detection and classification

[Mairal, Leordeanu, Bach, Hebert, and Ponce, 2008c]



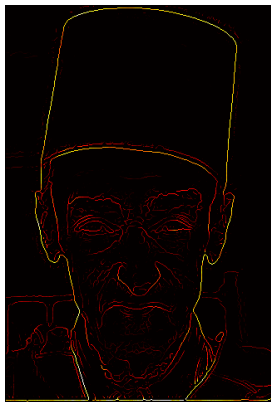
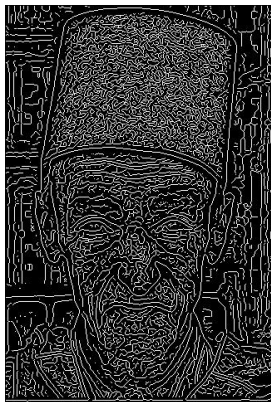
Good edges



Bad edges

Application to edge detection and classification

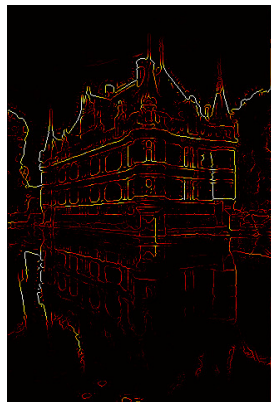
Berkeley segmentation benchmark



Raw edge detection on the right

Application to edge detection and classification

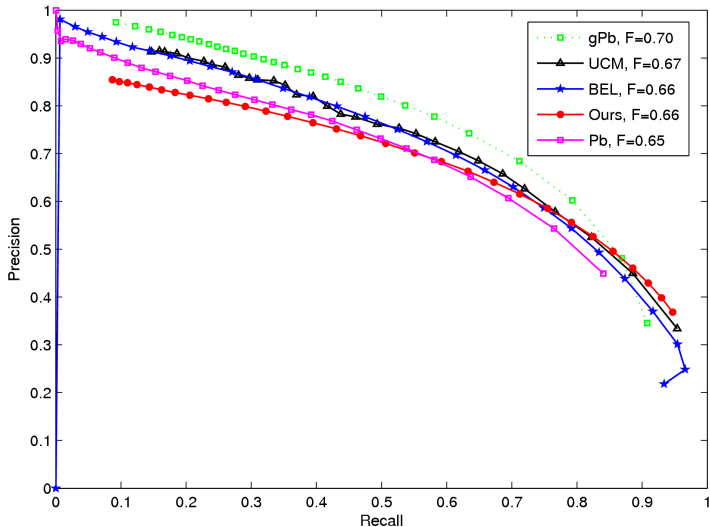
Berkeley segmentation benchmark



Raw edge detection on the right

Application to edge detection and classification

Berkeley segmentation benchmark



Application to edge detection and classification

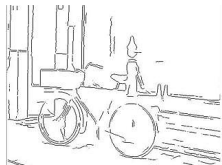
Contour-based classifier: [Leordeanu, Hebert, and Sukthankar, 2007]



Is there a bike, a motorbike, a car or a person on this image?

Application to edge detection and classification

**Input
Contours**



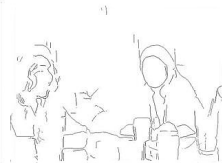
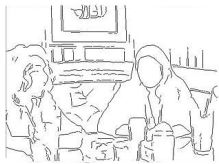
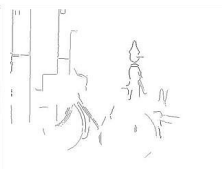
**Bike
Edge Detector**



**Bottle
Edge Detector**



**People
Edge Detector**



Application to edge detection and classification

Performance gain due to the prefiltering

Ours + [Leordeanu '07]	[Leordeanu '07]	[Winn '05]
96.8%	89.4%	76.9%

Recognition rates for the same experiment as [Winn et al., 2005] on VOC 2005.

Category	Ours+[Leordeanu '07]	[Leordeanu '07]
Aeroplane	71.9%	61.9%
Boat	67.1%	56.4%
Cat	82.6%	53.4%
Cow	68.7%	59.2%
Horse	76.0%	67%
Motorbike	80.6%	73.6%
Sheep	72.9%	58.4%
Tvmonitor	87.7%	83.8%
Average	75.9%	64.2 %

Recognition performance at equal error rate for 8 classes on a subset of images from Pascal 07.

Important messages

- Learned dictionaries are well adapted to model the local appearance of images and edges.
- They can be used to learn dictionaries of SIFT features.

References I

- M. Aharon, M. Elad, and A. M. Bruckstein. The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representations. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, November 2006.
- Y-L. Boureau, F. Bach, Y. Lecun, and J. Ponce. Learning mid-level features for recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.
- S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20:33–61, 1999.
- M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 54(12): 3736–3745, December 2006.
- K. Engan, S. O. Aase, and J. H. Husoy. Frame based signal compression using method of optimal directions (MOD). In *Proceedings of the 1999 IEEE International Symposium on Circuits Systems*, volume 4, 1999.
- R. Grosse, R. Raina, H. Kwong, and A. Y. Ng. Shift-invariant sparse coding for audio classification. In *Proceedings of the Twenty-third Conference on Uncertainty in Artificial Intelligence*, 2007.

References II

- A. Haar. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69:331–371, 1910.
- K. Huang and S. Aviyente. Sparse representation for signal classification. In *Advances in Neural Information Processing Systems*, Vancouver, Canada, December 2006.
- D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems*, 2001.
- H. Lee, A. Battle, R. Raina, and A. Y. Ng. Efficient sparse coding algorithms. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems*, volume 19, pages 801–808. MIT Press, Cambridge, MA, 2007.
- M. Leordeanu, M. Hebert, and R. Sukthankar. Beyond local appearance: Category recognition from pairwise interactions of simple features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.
- M. S. Lewicki and T. J. Sejnowski. Learning overcomplete representations. *Neural Computation*, 12(2):337–365, 2000.
- J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Discriminative learned dictionaries for local image analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008a.

References III

- J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Transactions on Image Processing*, 17(1):53–69, January 2008b.
- J. Mairal, M. Leordeanu, F. Bach, M. Hebert, and J. Ponce. Discriminative sparse image models for class-specific edge detection and image interpretation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2008c.
- J. Mairal, G. Sapiro, and M. Elad. Learning multiscale sparse representations for image and video restoration. *SIAM Multiscale Modelling and Simulation*, 7(1): 214–241, April 2008d.
- J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online dictionary learning for sparse coding. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2009.
- S. Mallat. *A Wavelet Tour of Signal Processing, Second Edition*. Academic Press, New York, September 1999.
- B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, 37:3311–3325, 1997.
- M. Protter and M. Elad. Image sequence denoising via sparse and redundant representations. *IEEE Transactions on Image Processing*, 18(1):27–36, 2009.

References IV

- S. Roth and M. J. Black. Fields of experts: A framework for learning image priors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.
- P. Sprechmann, I. Ramirez, G. Sapiro, and Y. C. Eldar. Collaborative hierarchical sparse modeling. Technical report, 2010. Preprint arXiv:1003.0400v1.
- R. Tibshirani. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B*, 58(1):267–288, 1996.
- J. Winn, A. Criminisi, and T. Minka. Object categorization by learned universal visual dictionary. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2005.
- J. Yang, K. Yu, Y. Gong, and T. Huang. Linear spatial pyramid matching using sparse coding for image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.
- J. Yang, K. Yu, , and T. Huang. Supervised translation-invariant sparse coding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.