Dictionary Learning and Application to Image Processing

Julie Digne

LIRIS - Équipe GeoMod - CNRS
Outline

1. Introduction
2. Sparse Coding
3. Dictionary Learning
Sparse processing of signals

Signal Processing aims to decompose complex signals using elementary functions which are then easier to manipulate.

\[ x(t) = \sum_{i=-\infty}^{\infty} a_i \varphi_i(t) \]

Image: Pier Luigi Dragotti
Sparse processing of signals

Image: Pier Luigi Dragotti
Sparse processing of signals

Between two representations of a signal pick the ones with the higher number of zero coefficients.
Patch-based approaches for images and surfaces

- Texture synthesis [Efros 99], Non local means [Buadès et al. 2005].

Compressive sensing theory [Candès et al. 2006]
There exists spaces, in which the signals would be sparsely represented, that are especially well suited for processing the signals.

- Sparse regularization for image analysis, inpainting... [Elad et al. 2006], [Mairal 2009] The K-SVD algorithm
Outline

1 Introduction

2 Sparse Coding

3 Dictionary Learning
A brief reminder on norms

**Norm definition**

Let $E$ be a vector space over a subfield $K$, a norm on $E$ is an application with nonnegative values $\|\cdot\| : E \to \mathbb{R}$ such that for all $\alpha \in K$ and $u, v \in E$:

- $\|\alpha v\| = |\alpha| \|v\|$ (positive homogeneity)
- $\|u + v\| \leq \|u\| + \|v\|$ (subadditivity)
- $\|u\| = 0_K \iff u = 0_E$ (separation)
A brief reminder on norms

**Norm definition**

Let $E$ be a vector space over a subfield $K$, a norm on $E$ is an application with nonnegative values $||| : E \to \mathbb{R}$ such that for all $\alpha \in K$ and $u, v \in E$:

- $\|\alpha v\| = |\alpha|\|v\|$ (positive homogeneity)
- $\|u + v\| \leq \|u\| + \|v\|$ (subadditivity)
- $\|u\| = 0_K \Leftrightarrow u = 0_E$ (separation)

The $\ell^2$ norm is also called the euclidean norm. Let $x$ be a vector in $\mathbb{R}^n$ with coordinates $(x_1, \ldots, x_n)$ in the canonical basis, the $\ell^2$ norm writes:

$$\|x\|_2 = \sqrt{x \cdot x^T} = \left(\sum_{i=1}^n x_i^2\right)^{\frac{1}{2}}$$
Norm Examples on vectors of $\mathbb{R}^n$

- $\ell^1$ Norm (Manhattan)

$$\|x\|_1 = \left( \sum_{i=1}^{n} |x_i| \right)$$
Norm Examples on vectors of $\mathbb{R}^n$

- $\ell^1$ Norm (Manhattan)

$$\|x\|_1 = \left( \sum_{i=1}^{n} |x_i| \right)$$

- $\ell^3$

$$\|x\|_3 = \left( \sum_{i=1}^{n} |x_i|^3 \right)^{\frac{1}{3}}$$
Norm Examples on vectors of $\mathbb{R}^n$

- $\ell^1$ Norm (Manhattan)

$$\|x\|_1 = \left(\sum_{i=1}^{n} |x_i|\right)$$

- $\ell^3$

$$\|x\|_3 = \left(\sum_{i=1}^{n} |x_i|^3\right)^{\frac{1}{3}}$$

- $\ell^{2.1}$

$$\|x\|_{2.1} = \left(\sum_{i=1}^{n} x_i^{2.1}\right)^{\frac{1}{2.1}}$$
Norm Examples on vectors of $\mathbb{R}^n$

- $\ell^1$ Norm (Manhattan)

$$
\|x\|_1 = (\sum_{i=1}^{n} |x_i|)
$$

- $\ell^3$

$$
\|x\|_3 = \left(\sum_{i=1}^{n} |x_i|^3\right)^{\frac{1}{3}}
$$

- $\ell^{2.1}$

$$
\|x\|_{2.1} = \left(\sum_{i=1}^{n} x_i^{2.1}\right)^{\frac{1}{2.1}}
$$

- $\ell^p$ pour $p \geq 1$

$$
\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}
$$
Norm Examples on vectors of $\mathbb{R}^n$

- $\ell^1$ Norm (Manhattan)

$$\|x\|_1 = \left(\sum_{i=1}^{n} |x_i|\right)$$

- $\ell^3$

$$\|x\|_3 = \left(\sum_{i=1}^{n} |x_i|^3\right)^{\frac{1}{3}}$$

- $\ell^{2.1}$

$$\|x\|_{2.1} = \left(\sum_{i=1}^{n} x_i^{2.1}\right)^{\frac{1}{2.1}}$$

- $\ell^p$ pour $p \geq 1$

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}$$

- $\ell^\infty$

$$\|x\|_\infty = \max_{i=1}^{n} |x_i|$$

**Exercice:** Prove that $\ell^\infty$ is indeed a norm?
The ball of radius 1 for norms $\ell^p$ with $p \geq 2$
The ball of radius 1 for norms $\ell^p$ with $p \geq 2$
The ball of radius 1 for norms $\ell^p$ with $p \geq 2$
The ball of radius 1 for norms $\ell^p$ with $p \geq 2$
The ball of radius 1 for norms $\ell^p$ with $p \leq 2$
The ball of radius 1 for norms $\ell^p$ with $p \leq 2$
The ball of radius 1 for norms $\ell^p$ with $p \leq 2$
The ball of radius 1 for norms $\ell^p$ with $p \leq 2$
The ball of radius 1 with norms and quasi-norms $\ell^p$
Norm and sparsity

**Sparsity definition**

A vector $x \in \mathbb{R}^N$ is said to be $s$-sparse if at most $s$ of its entries are non zero, i.e.

$$\text{card } \text{support}(x) \leq s$$

where $\text{support}(x) = \{i | x_i \neq 0\}$.

We note $\|x\|_0 = \text{card } \text{support}(x)$ and call it $\ell^0$. 
Norm and sparsity

Sparsity definition

A vector $x \in \mathbb{R}^N$ is said to be $s$-sparse if at most $s$ of its entries are non-zero, i.e.

$$\text{card } \text{support}(x) \leq s$$

where $\text{support}(x) = \{i | x_i \neq 0\}$.

We note $\|x\|_0 = \text{card } \text{support}(x)$ and call it $\ell^0$.

Is $\ell^0$ a norm?
Sparse Coding with the $\ell^0$ norm

**Problem statement**

Let $A \in \mathbb{R}^{m \times n}$ and $x$ a $s$-sparse vector in $\mathbb{R}^n$. Let $y \in \mathbb{R}^m$ such that $y = Ax$. Assume only $y$ and $A$ are known and we want to recover $x$. If $m < n$, the system is underdetermined.
Sparse Coding with the $\ell^0$ norm

**Problem statement**

Let $A \in \mathbb{R}^{m \times n}$ and $x$ a $s$-sparse vector in $\mathbb{R}^n$ Let $y \in \mathbb{R}^m$ such that $y = Ax$. Assume only $y$ and $A$ are known and we want to recover $x$. If $m < n$, the system is underdetermined.

**Sparsity hypothesis**

Identifying the solution $x$ under the $s$-sparsity hypothesis is easier.
Sparse Coding with the $\ell^0$ norm

**Optimization problem**

Given a measurement matrix $A \in \mathbb{R}^{m \times n}$ and $y$ a vector in $\mathbb{R}^n$, under the $s$-sparse assumption, the vector $x$ can be reconstructed as the solution of:

$$
\begin{align*}
\text{Minimize} & \quad \|x\|_0 \\
\text{s.t.} & \quad y = Ax
\end{align*}
$$

($P_0$)
Sparse Coding with the $\ell^0$ norm

**Optimization problem**

Given a measurement matrix $A \in \mathbb{R}^{m \times n}$ and $y$ a vector in $\mathbb{R}^n$, under the $s$-sparse assumption, the vector $x$ can be reconstructed as the solution of:

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \|x\|_0$$

$$\text{s.t. } y = Ax$$

\[ (P_0) \]

- This Optimization is a nonconvex optimization problem
Problem \((P_0)\) is a NP-hard problem

- Reformulate the problem as

\[
\begin{align*}
\text{Minimize} & \|x\|_0 \\
\text{s.t.} & \|y - Ax\|_2 \leq \eta
\end{align*}
\\quad (P_{0,\eta})
\]

Theorem

Problem \((P_{0,\eta})\) is a NP-hard problem

- NP-hardness: all problems for which a solving algorithm could be turned in polynomial time into a solving algorithm for any NP-problem.
- Proof: demonstrate that using Problem \((P_{0,\eta})\) one can solve for the exact cover 3-set problem.
- Reminder: Given a collection \(S\) of 3-subsets of a set \(X\), an exact cover of \(X\) is a subcollection \(S_{sub}\) of \(S\) such that the intersection of two distinct elements of \(S_{sub}\) is empty and the union of all elements of \(S_{sub}\) cover \(X\).
Sparse decomposition algorithm

- Efficient greedy algorithms have been proposed to find an approximate solution.
Matching Pursuit

Matching Pursuit Algorithm [Mallat & Zhang 1993]

- Set $k = 0$, $\alpha = 0_{\mathbb{R}^n}$
- While $k < s$ and $\|x - D\alpha\| > 0$ do:
  - Select index $j$ maximizing $|D_j^T \cdot (x - D\alpha)|$
  - Update coefficients $\alpha(j) = \alpha(j) + D_j^T \cdot (x - D\alpha)$
Matching Pursuit

Matching Pursuit Algorithm [Mallat & Zhang 1993]

- Set $k = 0$, $\alpha = 0_{\mathbb{R}^n}$
- While $k < s$ and $\|x - D\alpha\| > 0$ do:
  - Select index $j$ maximizing $|D_j^T \cdot (x - D\alpha)|$
  - Update coefficients $\alpha(j) = \alpha(j) + D_j^T \cdot (x - D\alpha)$

- At each step the algorithm finds the atom that best represents the residual $r = x - D\alpha$
## Matching Pursuit

### Matching Pursuit Algorithm [Mallat & Zhang 1993]

- Set $k = 0$, $\alpha = 0_{\mathbb{R}^n}$
- While $k < s$ and $\|x - D\alpha\| > 0$ do:
  - Select index $j$ maximizing $|D_j^T \cdot (x - D\alpha)|$
  - Update coefficients $\alpha(j) = \alpha(j) + D_j^T \cdot (x - D\alpha)$

At each step the algorithm finds the atom that best represents the residual $r = x - D\alpha$

The residual monotonically decreases until the residual is orthogonal to all $D_j$
Matching Pursuit

### Matching Pursuit Algorithm [Mallat & Zhang 1993]

- Set $k = 0$, $\alpha = 0_{\mathbb{R}^n}$
- While $k < s$ and $\|x - D\alpha\| > 0$ do:
  - Select index $j$ maximizing $|D_j^T \cdot (x - D\alpha)|$
  - Update coefficients $\alpha(j) = \alpha(j) + D_j^T \cdot (x - D\alpha)$

- At each step the algorithm finds the atom that best represents the residual $r = x - D\alpha$
- The residual monotonically decreases until the residual is orthogonal to all $D_j$
- How does the sparsity behave?
Matching Pursuit

Matching Pursuit Algorithm [Mallat & Zhang 1993]

- Set $k = 0$, $\alpha = 0_{\mathbb{R}^n}$
- While $k < s$ and $\|x - D\alpha\| > 0$ do:
  - Select index $j$ maximizing $|D_j^T \cdot (x - D\alpha)|$
  - Update coefficients $\alpha(j) = \alpha(j) + D_j^T \cdot (x - D\alpha)$

- At each step the algorithm finds the atom that best represents the residual $r = x - D\alpha$
- The residual monotonically decreases until the residual is orthogonal to all $D_j$
- How does the sparsity behave? nondecreasing
Orthogonal Matching Pursuit

- **Goal:** The sparsity should increase at each step.
Orthogonal Matching Pursuit

- **Goal:** The sparsity should increase at each step.

- **How?**

---

Orthogonal Matching Pursuit (OMP)

Set \( k = 0, \alpha = 0 \) \( R_n, \Gamma = \emptyset \)

While \( k < s \) and \( \| x - D\alpha \| > 0 \) do:

- Select index \( j_{\text{max}} \) maximizing \( |D^Tj \cdot (x - D\alpha)| \)
- Update the active set \( \Gamma = \Gamma \cup \{j\} \)
- Recompute \( \alpha_{\Gamma} \) minimizing \( x - D_{\Gamma}\alpha_{\Gamma} \)
- Set \( \alpha_{\bar{\Gamma}} = 0 \)

Remark: \( D_\Gamma, \alpha_\Gamma \): matrix (resp. vector) composed of the columns (resp. elements) of \( D \) (resp. \( \alpha \)) whose indices are in \( \Gamma \).
Orthogonal Matching Pursuit

- Goal: The sparsity should increase at each step.
- How? Render the residual orthogonal to all selected atoms.

Orthogonal Matching Pursuit (OMP)

\[
\text{Set } k = 0, \alpha = 0, R_n, \Gamma = \emptyset
\]

While \( k < s \) and \( \| x - D \alpha \| > 0 \) do:

- Select index \( j \) maximizing \( |D^T j \cdot (x - D \alpha)| \)
- Update the active set \( \Gamma = \Gamma \cup \{ j \} \)
- Recompute \( \alpha_\Gamma \) minimizing \( x - D_\Gamma \alpha_\Gamma \)
- Set \( \alpha_{\bar{\Gamma}} = 0 \)

Remark:

- \( D_\Gamma, \alpha_\Gamma \): matrix (resp. vector) composed of the columns (resp. elements) of \( D \) (resp. \( \alpha \)) whose indices are in \( \Gamma \).
Orthogonal Matching Pursuit

- Goal: The sparsity should increase at each step.
- How? Render the residual orthogonal to all selected atoms.
Orthogonal Matching Pursuit

- Goal: The sparsity should increase at each step.

- How? Render the residual orthogonal to all selected atoms.

Orthogonal Matching Pursuit (OMP)

- Set $k = 0$, $\alpha = 0_{\mathbb{R}^n}$, $\Gamma = \emptyset$

- While $k < s$ and $\|x - D\alpha\| > 0$ do:
  - Select index $j$ maximizing $|D_j^T \cdot (x - D\alpha)|$
  - Update the active set $\Gamma = \Gamma \cup \{j\}$
  - Recompute $\alpha_\Gamma$ minimizing $x - D_\Gamma \alpha_\Gamma$
  - Set $\alpha_\complement = 0$

Remark: $D_\Gamma$, $\alpha_\Gamma$: matrix (resp. vector) composed of the columns (resp. elements) of $D$ (resp. $\alpha$) whose indices are in $\Gamma$. 
What can we prove about OMP?

- The index selection is guided by finding the one that makes the error decrease most.

Tropp and Gilbert (2007): OMP is able to reliably recover a sparse vector from random measurements.
What can we prove about OMP?

- The index selection is guided by finding the one that makes the error decrease most.
- What about the case where a vector is a linear combination of 3 vectors.
What can we prove about OMP?

- The index selection is guided by finding the one that makes the error decrease most.
- What about the case where a vector is a linear combination of 3 vectors.
- Tropp and Gilbert (2007): OMP is able to reliably recover a sparse vector from random measurements.
What can we prove about OMP?

- The index selection is guided by finding the one that makes the error decrease most.
- What about the case where a vector is a linear combination of 3 vectors.
- Tropp and Gilbert (2007): OMP is able to reliably recover a sparse vector from random measurements.
- OMP is slower than MP
How can we make OMP faster?

Which step is computationally intensive?

- Computing the best index means computing $D^T_\Gamma (x - D_\Gamma \alpha_\Gamma)$ and taking the index of the smallest coefficient.
How can we make OMP faster?

Which step is computationally intensive?

- Computing the best index means computing $D_f^T(x - D_\Gamma \alpha_\Gamma)$ and taking the index of the smallest coefficient.
- Then we compute $\alpha_{\Gamma'}$ as $\alpha$ minimizing $\|x - D_{\Gamma'}\alpha\|^2$. 

Closed form solution:

$$\alpha_{\Gamma'} = \left(D_f^T D_\Gamma\right)^{-1} D_f^T x$$

Making OMP faster

Invert quickly $D_f^T D_\Gamma$ knowing the inverse of $D_f^T D_\Gamma$.
How can we make OMP faster?

Which step is computationally intensive?

- Computing the best index means computing $D_T^T(x - D_T \alpha_T)$ and taking the index of the smallest coefficient.
- Then we compute $\alpha_{\Gamma'}$ as $\alpha$ minimizing $\|x - D_{\Gamma'} \alpha\|^2$.
- Closed form solution:

$$\alpha_{\Gamma'} = (D_{\Gamma'}^T D_{\Gamma'})^{-1} D_{\Gamma'} x$$
How can we make OMP faster?

Which step is computationally intensive?

- Computing the best index means computing $D_{\Gamma}^T(x - D_{\Gamma}\alpha_{\Gamma})$ and taking the index of the smallest coefficient.
- Then we compute $\alpha_{\Gamma'}$ as $\alpha$ minimizing $\|x - D_{\Gamma'}\alpha\|^2$.
- Closed form solution:

$$\alpha_{\Gamma'} = (D_{\Gamma'}^T D_{\Gamma'})^{-1} D_{\Gamma'} x$$

Making OMP faster

Invert quickly $D_{\Gamma'}^T D_{\Gamma'}$ knowing the inverse of $D_{\Gamma}^T D_{\Gamma}$. 

Sparse Coding
Update of the inverse of $D^T D$ when appending a column $d$

- $u_1 \leftarrow D^T d$
- $u_2 \leftarrow (D^T D)^{-1} u_1$
- $u_3 \leftarrow d u_2$
- $A \leftarrow (X^t X)^{-1} + d u_2^T u_2$
- $s \leftarrow \frac{1}{d^T d - u_1^T u_2}$
- Updated inverse: $(A \quad -u_3 \quad -u_3^T s)$
An application of OMP: synthesizing terrains based on examples [Guérin et al. 2016]

A terrain is seen as a set of blended patches

Terrain model $\mathcal{T}$
An application of OMP: synthesizing terrains based on examples [Guérin et al. 2016]

- Build a dictionary by decomposing a real-world elevation map into patches
- Decompose patches to synthesize on it
Various possible applications in the terrain setting

Sparse Coding
Terrain Amplification

Creation of a multi-resolution dictionary
Terrain Amplification

Result

video
Another viewpoint on sparsity: the $\ell^1$ norm

- Using the $\ell^0$ or $\ell^p, p < 1$ yields nonconvex problems
- What about the $\ell^1$ norm?

$\ell^1$ norm (blue), $\ell^2$ norm (red)
Sparse Coding with the $\ell^1$ norm

**Problem**

Find $\alpha$ minimizing $\frac{1}{2} \|x - D\alpha\|_2^2 + \lambda \|\alpha\|_1$ this is called the LASSO problem
Sparse Coding with the $\ell^1$ norm

**Problem**

Find $\alpha$ minimizing $\frac{1}{2} \|x - D\alpha\|^2_2 + \lambda \|\alpha\|_1$ this is called the LASSO problem

**Coordinate descent algorithm**

- Select a coordinate index $j$ at random
Sparse Coding with the $\ell^1$ norm

**Problem**

Find $\alpha$ minimizing $\frac{1}{2} \| x - D\alpha \|_2^2 + \lambda \| \alpha \|_1$ this is called the LASSO problem

**Coordinate descent algorithm**

- Select a coordinate index $j$ at random
- update $\alpha[j]$ as the minimizer of $\| x - \sum_{l \neq j} \alpha[l]d_l - \alpha d_j \|_2^2 + \lambda |\alpha|$
Sparse Coding with the $\ell^1$ norm

**Problem**

Find $\alpha$ minimizing $\frac{1}{2}\|x - D\alpha\|_2^2 + \lambda\|\alpha\|_1$ this is called the LASSO problem

**Coordinate descent algorithm**

- Select a coordinate index $j$ at random
- update $\alpha[j]$ as the minimizer of $\|x - \sum_{l \neq j} \alpha[l]d_l - \alpha d_j\|_2^2 + \lambda|\alpha|$
- $\alpha[j] \leftarrow S_\lambda(\alpha[j] + \frac{d_j^T(x - D\alpha)}{\|d_j\|_2^2})$
Sparse Coding with the $\ell^1$ norm

**Problem**

Find $\alpha$ minimizing $\frac{1}{2} \| x - D\alpha \|_2^2 + \lambda \| \alpha \|_1$ this is called the LASSO problem

**Coordinate descent algorithm**

- Select a coordinate index $j$ at random
- update $\alpha[j]$ as the minimizer of $\| x - \sum_{l \neq j} \alpha[l] d_l - \alpha d_j \|_2^2 + \lambda |\alpha|$
- $\alpha[j] \leftarrow S_\lambda(\alpha[j] + \frac{d_j^T(x - D\alpha)}{\|d_j\|_2^2})$

- $S_\lambda$ is the soft-thresholding operator $S_\lambda(u) = \text{sign}(u) \max(|u| - \lambda, 0)$
What is the dictionary $D$?

**Problem**

In the context of Image Processing and Synthesis, we only have access to a set of signals for which we want to build a dictionary.
What is the dictionary $D$?

**Problem**

In the context of Image Processing and Synthesis, we only have access to a set of signals for which we want to build a dictionary.

**Dictionary Learning Problem**

Given a set of signals $x_i$ for $i = 1 \cdots N$ in $\mathbb{R}^n$ we want to build a matrix $D \in \mathbb{R}^{n \times m}$ and coefficients $\alpha_i \in \mathbb{R}^m$ for $i = 1 \cdots n$ solving:

$$\text{Minimize} \quad \sum_{i=1}^{N} \|x_i - D \cdot \alpha_i\|_2^2$$

$$\text{subject to} \quad \|D_i\|_2 \leq 1, \quad \forall i = 1 \cdots N, \alpha_i \in \mathbb{R}^m,$$

$$\forall i = 1 \cdots N, \|\alpha_i\|_0 = s$$

($P_{D,\alpha,0}$)
Dictionary learning problems

- Still a nonconvex problem
- Common approach: alternate minimization
  - Fix the dictionary $D$ and compute the sparse decomposition $\alpha$
  - Fix the sparse decomposition $\alpha$ and compute $D$
Method of Optimal Directions (MOD)

- First introduced by Engan et al. [1999]
Method of Optimal Directions (MOD)

- First introduced by Engan et al. [1999]
- Step 1: Compute the sparse codes?
Method of Optimal Directions (MOD)

- First introduced by Engan et al. [1999]
- Step 1: Compute the sparse codes? MP, OMP, Iterative Hard Thresholding
Method of Optimal Directions (MOD)

- First introduced by Engan et al. [1999]
- Step 1: Compute the sparse codes? MP, OMP, Iterative Hard Thresholding
- Step 2: Update the dictionary
Assume all coefficients $\alpha_i$ are fixed, Problem $(P_{D,\alpha,0})$ becomes

$$\text{Minimize } \sum_{i=1}^{N} \left\| x_i - D \cdot \alpha_i \right\|_2^2$$

subject to $D \in \mathbb{R}^{n \times m}$ and $\|D_i\|_2 \leq 1$. This problem is a convex problem on a convex set. Discarding the 1-norm constraint yields a least squares objective.
MOD: Dictionary Update

- Assume all coefficients $\alpha_i$ are fixed, Problem $(P_{D,\alpha,0})$ becomes

  $$\text{Minimize} \sum_{i=1}^{N} \| x_i - D \cdot \alpha_i \|^2_2$$

  $$\text{subject to } D \in \mathbb{R}_{n \times m}, \| D_i \|_2 \leq 1$$

- This problem is a convex problem on a convex set
MOD: Dictionary Update

- Assume all coefficients $\alpha_i$ are fixed, Problem $(P_{D,\alpha,0})$ becomes

$$\text{Minimize} \sum_{i=1}^{N} \|x_i - D \cdot \alpha_i\|_2^2$$

- This problem is a convex problem on a convex set
- Discarding the 1-norm constraint yields a least squares objective
Assume all coefficients $\alpha_i$ are fixed, Problem $(P_{D,\alpha,0})$ becomes

$$\text{Minimize} \sum_{i=1}^{N} \sum_{D \in \mathbb{R}^{n \times m}, \|D_i\|_2 \leq 1} \|x_i - D \cdot \alpha_i\|_2^2$$

This problem is a convex problem on a convex set.

Discarding the 1-norm constraint yields a least squares objective.

Idea: Solve the least squares problem and project the solution onto the convex set of admissible solutions.
MOD: Dictionary Update

- Assume all coefficients $\alpha_i$ are fixed, Problem $(P_{D,\alpha,0})$ becomes

$$\min_{D \in \mathbb{R}^{n \times m}, \|D_i\|_2 \leq 1} \sum_{i=1}^{N} \|x_i - D \cdot \alpha_i\|^2_2$$

- This problem is a convex problem on a convex set
- Discarding the 1-norm constraint yields a least squares objective
- Idea: Solve the least squares problem and project the solution onto the convex set of admissible solutions

Bertsekas 1999

In general, solving the general problem and projecting the solution on the convex constraints set yields a poor solution
MOD: Dictionary Update

- Assume all coefficients $\alpha_i$ are fixed, Problem $(P_{D, \alpha, 0})$ becomes

$$\minimize_{D \in \mathbb{R}^{n \times m}, \|D_i\|_2 \leq 1} \sum_{i=1}^{N} \|x_i - D \cdot \alpha_i\|_2^2$$

- This problem is a convex problem on a convex set
- Discarding the 1-norm constraint yields a least squares objective
- Idea: Solve the least squares problem and project the solution onto the convex set of admissible solutions

Bertsekas 1999

Since the $\ell^0$ norm remains constant when a vector undergoes a nonzero rescaling, the projection is valid.
Dictionary Update

Least Squares Problem

Solve for $D$ in:

$$\text{Minimize}_{D \in \mathbb{R}^{m \times n}} \sum_{i=1}^{N} \| x_i - D \alpha_i \|_2^2$$
Dictionary Update

Least Squares Problem

Solve for $D$ in:

$$\text{Minimize} \sum_{i=1}^{N} \|x_i - D\alpha_i\|^2_2$$

- Compute the gradient + set to 0: $\sum_{i=1}^{N} (x_i - D\alpha_i)\alpha_i^T = 0$
Dictionary Update

Least Squares Problem

Solve for $D$ in:

$$\text{Minimize}_{D \in \mathbb{R}^{m \times n}} \sum_{i=1}^{N} \|x_i - D\alpha_i\|_2^2$$

- Compute the gradient + set to 0: $\sum_{i=1}^{N} (x_i - D\alpha_i)\alpha_i^T = 0$
- $D = (\sum_{i=1}^{N} x_i\alpha_i^T)(\sum_{i=1}^{N} (\alpha_i\alpha_i^T)^{-1}$
Dictionary Update

Least Squares Problem

Solve for $D$ in:

$$\begin{array}{c}
\text{Minimize} \\
D \in \mathbb{R}^{m \times n} \\
\sum_{i=1}^{N} \| x_i - D \alpha_i \|_2^2
\end{array}$$

- Compute the gradient + set to 0:
  $$\sum_{i=1}^{N} (x_i - D \alpha_i) \alpha_i^T = 0$$
- $$D = (\sum_{i=1}^{N} x_i \alpha_i^T)(\sum_{i=1}^{N} (\alpha_i \alpha_i^T)^{-1})$$
- Setting $A = (\alpha_1 \mid \alpha_2 \mid \cdots \mid \alpha_N)$, $X = (x_1 \mid x_2 \mid \cdots \mid x_N)$
  one has: $D = X A^T (A A^T)^{-1}$

Projection on the constraint set: normalizing each column of $D$ if its norm is above 1.
Dictionary Update

Least Squares Problem

Solve for $D$ in:

$$\text{Minimize}_{D \in \mathbb{R}^{m \times n}} \sum_{i=1}^{N} \| x_i - D\alpha_i \|_2^2$$

- Compute the gradient + set to 0: $\sum_{i=1}^{N} (x_i - D\alpha_i)\alpha_i^T = 0$
- $D = (\sum_{i=1}^{N} x_i\alpha_i^T) (\sum_{i=1}^{N} (\alpha_i\alpha_i^T)^{-1}$
- Setting $A = (\alpha_1 | \alpha_2 | \cdots | \alpha_N)$, $X = (x_1 | x_2 | \cdots | x_N)$
  one has: $D = X A^T (A A^T)^{-1}$
- Projection on the constraint set: normalizing each column of $D$ if its norm is above 1.
MOD algorithm for the $\ell^1$ norm

- The same principle as before can be used: alternate directions
MOD algorithm for the $\ell^1$ norm

- The same principle as before can be used: alternate directions
- Compute the sparse codes using coordinate gradient descent (or homotopy)
MOD algorithm for the $\ell^1$ norm

- The same principle as before can be used: alternate directions
- Compute the sparse codes using coordinate gradient descent (or homotopy)
- Update the dictionary by minimizing $\frac{1}{2} \sum_i \|x_i - D\alpha_i\|^2_2$
Dictionary update for MOD-L1

<table>
<thead>
<tr>
<th>Updating column $d_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find $d$ s.t. $|d|<em>2 \leq 1$ minimizing $\sum</em>{i=1}^{n} \frac{1}{2} |x_i - \sum_{l \neq j} \alpha_i(l)d_l - \alpha_i(j)d|_2^2$</td>
</tr>
</tbody>
</table>

- Derive the final update for the dictionary.
K-SVD algorithm

- Still an alternating direction minimization method
K-SVD algorithm

- Still an alternating direction minimization method
- Goal: Incorporate the sparsity constraint also in the dictionary update step
Goal

- A set of training signals \( \{y_i\}_{i=1}^N \in \mathbb{R}^n \)
- Design a dictionary \( D \in \mathbb{R}^{n \times K} \) such that there exists \( x \in \mathbb{R}^k \) such that either 
  \( y = Dx \)
  or \( y \approx Dx \) s.t. \( \|y - Dx\|_p \leq \varepsilon \)
- if \( n < K \) and \( D \) is full-ranked the solution must be constrained
  - \( \min_x \|x\|_0 \) s.t. \( y = Dx \)
  - \( \min_x \|x\|_0 \) s.t. \( \|y - Dx\|_2 \leq \varepsilon \)
- Design \( D \) in order to best fit the sparsity model imposed
An extension of K-means

- K-means search for the best possible representative enforcing that each representation uses a single atom with coefficient 1.
- K-SVD solves \( \min_{D,X} \|Y - DX\|_F^2 \) s.t. \( \forall i \|x_i\|_0 \leq T_0 \)
- An iterative approach that alternates between two steps
  - \textit{Sparse coding} of the examples based on the current dictionary
  - \textit{Update of the dictionary} so as to better fit the data
- \( Y \in \mathbb{R}^{n \times N} \): training samples, \( X \in \mathbb{R}^{K \times N} \) matrix of coefficients
Sparse Coding stage

- $D$ is fixed, compute the best representation $x_i$ of sample $y_i$
- Find $x_i$ minimizing $\|y_i - Dx_i\|_2^2$ s.t. $\|x_i\|_0 \leq T_0$
- Can be done using a pursuit algorithm (e.g. Orthogonal Matching Pursuit)
Dictionary Update stage

- The update will be done atom by atom.
- $\|Y - DX\|_F^2 = \|Y - \sum_{j=1}^N d_j x_j^T\|_F^2 = \|Y - \sum_{j=1, j \neq k}^N d_j x_j^T - d_k x_k^T\|_F^2$
- $E_k = Y - \sum_{j=1, j \neq k}^N d_j x_j^T$ error obtained by omitting atom $d_k$ in the decomposition
- Finally solve for:
  $$\|E_k - d_k x_k^T\|_F \text{ w.r.t. } d_k, x_k^T$$
- Solve using SVD? if so sparsity not enforced.
Trick to enforce the sparsity

- \( \omega_k = \{ i | 1 \leq i \leq K, x^k_T(i) \neq 0 \} \)
- Restrict \( E_k \) and \( x^k_T \) to \( E^R_k \) and \( x^R_k \) by selecting only the columns of indices included in the support of \( x^k_T \)
- Use SVD to decompose \( E^R_k = U \Delta V^T \)
- Set \( d_k \) to be the first column of \( U \)
- Set \( x^R_k \) to be the first column of \( V \) multiplied by \( \Delta(1, 1) \)
- the columns of \( D \) remain normalized and the support of the representations can not increase
Application to the denoising of images

- noisy input image $y$
- Build $\hat{D}$ and $(\hat{x}_i)_i$ the dictionary and representations of all patches of image $y$
- $(P_i(x))_i$ the set of all image $x$ patches.
- $\hat{D}\hat{\alpha}_i$ is the representation of patch $P_i(y)$
- Find $x$ minimizing $\lambda \|x - y\|^2_2 + \sum_i \|\hat{D}\hat{\alpha}_i - P_i(x)\|^2_2$
  
  - fidelity term
  - proximity of the reconstruction to the denoised patch
Can be tested on IPOL http://www.ipol.im/pub/algo/llm_ksvd
Learned dictionary

Dictionary learned from face patches
Train time 9.0s on 94500 patches
Learned Color dictionary
Denoising via dictionary learning
Denoising via dictionary learning
Denoising via dictionary learning

Denoising each channel separately
Denoising via dictionary learning

Denoising each channel separately (left) vs globally (right)
Comparison to NL-means

Original
Comparison to NL-means

Dictionary learning
Comparison to NL-means

NL-means
Application: Point Cloud Compression


Self-similarity for compression

[Hubo et al. 2008]
- Cluster surface patches by similarity
- Replace each patch by a word of the codebook

Compression for rendering and not precision!

Patch-based self-similarity
Local patches capture local variations, comparing them underlines the self-similarity
Two samplings of the same shape
Pipeline

Original Seeds and patches Parameterization
Patch descriptions Coefficients Dictionary
1 2 n-1 n

Dictionary Learning
Working assumptions

- **Topological condition:** Surface covered by a set of topological disks centered around seeds.

- **Sampling condition:** $R$-neighborhood of a seed containing enough points.

- **Noise level:** Noise magnitude strictly below radius $R$.

- **Seeds selection:** anchors to define local patches
Self-similarity compression

- **Seeds** selection
- Local patches represented in a comparable way
- Patches decomposed upon a dictionary found by the K-SVD algorithm
- Final data: a set of seeds with local frames, a small dictionary and the (sparse) coefficients for each patch.
Further compression

- **Seeds**: kd-tree compression [Gandoin and Devillers, 2002].
- **Local parameterization** (3 Euler angles): predictive coding
- **Dictionary**: lossless compression.
- **Coefficients**: scalar quantization (increases sparsity) followed by entropy coding.
Controlling the error

Two types of errors are introduced:

- **Resampling error**
  - Increasing the accuracy of the resampling pattern

- **Compression error**
  - Increasing the number of atoms in the dictionary
Decompression

1. Decompress
   - seed positions
   - euler angles
   - dictionary $D$
   - coefficients $X$

2. Reconstruct the patches:
   \[ P_{\text{rec}} = D \ast X \]

3. Consolidate the reconstructed point cloud in overlapping areas.
Results

Anubis (9, 9M pts) compressed to 0.96bpp; error = 0.01mm (0.003%)
St Matthew (93,5M pts) compressed to 0.83\textit{bpp}; error = 0.05cm (0.002%)
Mire (16, 1M pts) compressed to 0.73bpp; error = 0.03mm (0.011%). Screened Poisson Reconstruction [Kazdhan, 2013]
Comparison with kd-tree coding. 4.83bpp against 0.6bpp in our method.
Lovers (15, 8M pts) compressed to 0.59 bpp; error = 0.01mm (0.006%)
Breaking the working assumptions

Church (69, 9M pts) compressed to 0.76bpp; error = 1.48cm (0.005%)
rate/distortion performance compared to previous works
Recommended Readings