

Online Dictionary Learning and Image Inpainting



Jia-Wei Liao, Kuok-Tong Ng

Department of Applied Mathematics
National Chiao Tung University

December 30, 2020

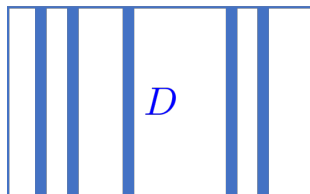
Outline

- 1 Sparse representation / Sparse coding
- 2 Sparse dictionary learning
- 3 Image inpainting

Sparse representation / Sparse coding

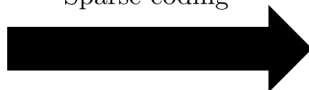
Problem

$$\min \|z\|_0 \text{ s.t. } Dz = x$$


 $m \times k$

 $m \times 1$

Sparse coding


 $k \times 1$

Sparse representation / Sparse coding

SR problem

$$z^* = \arg \min_{z \in \mathbb{R}^k} \left(\frac{1}{2} \|x - Dz\|_2^2 + \lambda \|z\|_0 \right)$$

where

- x is signal vector in \mathbb{R}^m
- D is dictionary matrix in $\mathbb{R}^{m \times k}$
- z is sparse coefficient vector in \mathbb{R}^k
- $\lambda > 0$ is a penalty parameter

Sparse representation / Sparse coding

It is inefficient to compute $\|z\|_0$ directly when n is large. In practice, we will use the ℓ^1 norm instead of the ℓ^0 norm.

SR problem

$$z^* = \arg \min_{z \in \mathbb{R}^k} \left(\frac{1}{2} \|x - Dz\|_2^2 + \lambda \|z\|_1 \right), \lambda > 0.$$

Sparse dictionary learning

SDL problem

$$\min_{D, z_i} \left(\frac{1}{2} \sum_{i=1}^n \|x_i - Dz_i\|_2^2 + \lambda \sum_{i=1}^n \|z_i\|_1 \right)$$

subject to $\|d_j\|_2 \leq 1, \forall 1 \leq j \leq k.$

where

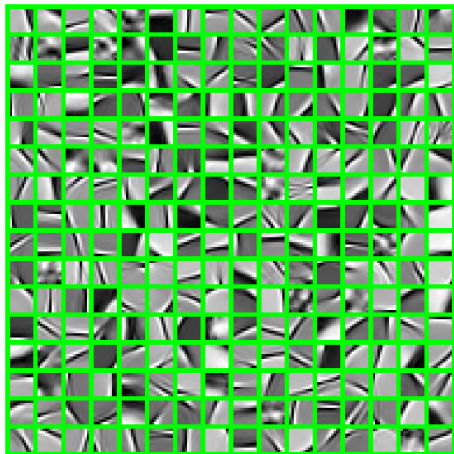
- x_i is a given signals in \mathbb{R}^m
- $D = [d_1, d_2, \dots, d_k]$ is dictionary matrix in $\mathbb{R}^{m \times k}$
- z is sparse coefficient vector in \mathbb{R}^k
- $\lambda > 0$ is a penalty parameter

Training set



$$\begin{array}{ccc}
 x_1 & \cdots & x_n \\
 \begin{array}{|c|} \hline \\ \hline \end{array} & & \begin{array}{|c|} \hline \\ \hline \end{array} \\
 100 \times 1 & & 100 \times 1
 \end{array}$$

Dictionary



Online dictionary learning

ODL method

Given an initial dictionary $D^{(0)}$, for $t = 1, 2, \dots, T$, we solve the following two sub-problems alternately:

$$x^{(t)} = \text{Random drawing from } \{x_1, \dots, x_n\}$$

$$z^{(t)} = \arg \min_{z \in \mathbb{R}^k} \left(\frac{1}{2} \|x^{(t)} - D^{(t-1)} z\|_2^2 + \lambda \|z\|_1 \right)$$

$$D^{(t)} = \arg \min_{D \in \mathcal{C}} \frac{1}{t} \sum_{i=1}^t \left(\frac{1}{2} \|x^{(i)} - D z^{(i)}\|_2^2 + \lambda \|z^{(i)}\|_1 \right)$$

$$\text{where } \mathcal{C} = \{D \in \mathbb{R}^{m \times k} \mid d_j^\top d_j \leq 1, \forall 1 \leq j \leq k\}$$

We iterate until convergence is achieved.

Mini-batch

To simplify the formulation of the SDL problem, we define

$$X^{(t)} = \begin{bmatrix} x_1^{(t)} & x_2^{(t)} & \cdots & x_b^{(t)} \end{bmatrix} \in \mathbb{R}^{m \times b}$$

$$Z^{(t)} = \begin{bmatrix} z_1^{(t)} & z_2^{(t)} & \cdots & z_b^{(t)} \end{bmatrix} \in \mathbb{R}^{k \times b}$$

where b is the **batch size**.

Hence the problem in (ODL) can be rewritten as follows:

$$Z^{(t)} = \arg \min_{Z \in \mathbb{R}^{k \times b}} \left(\frac{1}{2} \|X^{(t)} - D^{(t-1)} Z\|_F^2 + \lambda \|Z\|_{1,1} \right) \quad (1)$$

$$D^{(t)} = \arg \min_{D \in \mathcal{C}} \frac{1}{t} \sum_{i=1}^t \left(\frac{1}{2} \|X^{(i)} - D Z^{(i)}\|_F^2 + \lambda \|Z^{(i)}\|_{1,1} \right) \quad (2)$$

where $\|Z\|_{1,1} := \sum_{j=1}^b \|z_j\|_1$.

Solving sparse coding by ADMM

To solve

$$Z^{(t)} = \arg \min_{Z \in \mathbb{R}^{k \times b}} \left(\frac{1}{2} \|X - DZ\|_F^2 + \lambda \|Z\|_{1,1} \right)$$

We add an auxiliary variable Y and a dual variable U , we define

$$f(Z) = \frac{1}{2} \|X - DZ\|_F^2, \quad g(Y) = \lambda \|Y\|_{1,1}, \quad Z = Y.$$

Then the ADMM for solving (1) is given by

- ① $Z^{(i)} = \arg \min_Z \left(\frac{1}{2} \|X - DZ\|_F^2 + \frac{\rho}{2} \|Z - Y^{(i-1)} + U^{(i-1)}\|_F^2 \right)$
- ② $Y^{(i)} = \arg \min_Y \left(\lambda \|Y\|_{1,1} + \frac{\rho}{2} \|Z^{(i)} - Y + U^{(i-1)}\|_F^2 \right)$
- ③ $U^{(i)} = U^{(i-1)} + Z^{(i)} - Y^{(i)}$

Solving Z-subproblem

Define

$$F(Z) = \frac{1}{2} \|X - DZ\|_F^2 + \frac{\rho}{2} \|Z - Y^{(i-1)} + U^{(i-1)}\|_F^2$$

Then

$$\begin{aligned} \nabla F(Z) &= -D^\top (X - DZ) + \rho I (Z - Y^{(i-1)} + U^{(i-1)}) \\ &= (D^\top D + \rho I) Z - (D^\top X + \rho(Y^{(i-1)} - U^{(i-1)})) \end{aligned}$$

Letting $\nabla F(Z) = 0$, we have

$$(D^\top D + \rho I) Z = (D^\top X + \rho(Y^{(i-1)} - U^{(i-1)}))$$

Therefore, we obtain the solution

$$Z^{(i)} = (D^\top D + \rho I)^{-1} (D^\top X + \rho(Y^{(i-1)} - U^{(i-1)}))$$

Solving Y-subproblem

Using the component-wise soft-thresholding function, the solution of Y-subproblem has the closed form:

$$Y^{(i)} = \mathcal{S}_{\lambda/\rho} \left(Z^{(i)} + U^{(i-1)} \right)$$

where

$$\mathcal{S}_{\lambda/\rho}(V) = \text{sign}(V) \odot \max \left(0, |V| - \frac{\lambda}{\rho} \right)$$

with $\text{sign}(V)$ and $|V|$ are element-wisely applied to the matrix V and \odot is the Hadamard product.

Solving sparse coding by ADMM

Therefore, the iterative scheme can be posed as follows:

$$\textcircled{1} \quad Z^{(i)} = (D^T D + \rho I)^{-1} \left(D^T X + \rho(Y^{(i-1)} - U^{(i-1)}) \right)$$

$$\textcircled{2} \quad Y^{(i)} = \mathcal{S}_{\lambda/\rho} \left(Z^{(i)} + U^{(i-1)} \right)$$

$$\textcircled{3} \quad U^{(i)} = U^{(i-1)} + Z^{(i)} - Y^{(i)}$$

Update dictionary

The problem (2) in (ODL) is

$$D^{(t)} = \arg \min_{D \in \mathcal{C}} \frac{1}{t} \sum_{i=1}^t \left(\frac{1}{2} \|X^{(i)} - DZ^{(i)}\|_F^2 + \lambda \|Z^{(i)}\|_{1,1} \right)$$

Note that

$$\begin{aligned} & \|X^{(i)} - DZ^{(i)}\|_F^2 \\ &= \text{tr} \left[(X^{(i)} - DZ^{(i)})^\top (X^{(i)} - DZ^{(i)}) \right] \\ &= \text{tr}(X^{(i)\top} X^{(i)}) + \text{tr}(Z^{(i)\top} D^\top D Z^{(i)}) - 2\text{tr}(Z^{(i)\top} D^\top X^{(i)}) \end{aligned}$$

Then we have

$$D^{(t)} = \arg \min_{D \in \mathcal{C}} \frac{1}{t} \sum_{i=1}^t \left(\frac{1}{2} \text{tr}(Z^{(i)\top} D^\top D Z^{(i)}) - \text{tr}(Z^{(i)\top} D^\top X^{(i)}) \right)$$

Update dictionary

Let $A^{(t)} = \sum_{i=1}^t Z^{(i)} Z^{(i)\top}$ and $B^{(t)} = \sum_{i=1}^t X^{(i)} Z^{(i)\top}$.

We can show that

- $\sum_{i=1}^t \text{tr}(Z^{(i)\top} D^\top D Z^{(i)}) = \text{tr}(D^\top D A^{(t)})$
- $\sum_{i=1}^t \text{tr}(Z^{(i)\top} D^\top X^{(i)}) = \text{tr}(D^\top B^{(t)})$

Therefore, the problem can be rewritten to

$$D^{(t)} = \arg \min_{D \in \mathcal{C}} \frac{1}{t} \left(\frac{1}{2} \text{tr}(D^\top D A^{(t)}) - \text{tr}(D^\top B^{(t)}) \right)$$

Update dictionary: coordinate descent

To solve

$$D^{(t)} = \arg \min_{D \in \mathcal{C}} \frac{1}{t} \left(\frac{1}{2} \text{tr}(D^\top D A^{(t)}) - \text{tr}(D^\top B^{(t)}) \right)$$

First, let

$$F(D) = \frac{1}{2} \text{tr}(D^\top D A^{(t)}) - \text{tr}(D^\top B^{(t)})$$

Then we can get

$$\nabla_{d_i} F(D) = D a_i - b_i$$

where a_i, b_i are the column vector of $A^{(t)}$ and $B^{(t)}$.

Update dictionary: coordinate descent

Applying coordinate decent method, d_i can update by

$$d_i^{(t)} = d_i^{(t-1)} - \frac{1}{A_{i,i}}(Da_i - b_i).$$

Since we hope $d_i \leq 1, \forall 1 \leq i \leq k$,

$$d_i^{(t)} \leftarrow \frac{d_i^{(t)}}{\max(\|d_i\|_2, 1)}$$

Algorithm

Algorithm Sparse Coding

- 1: **Require:** $D, X, Y^{(0)}, U^{(0)}, \lambda, \rho$
 - 2: **for** $i = 1$ to maxstep **do**
 - 3: $Z^{(i)} = (D^T D + \rho I)^{-1} (D^T X + \rho(Y^{(i-1)} - U^{(i-1)}))$
 - 4: $Y^{(i)} = \text{sign}(Z^{(i)} + U^{(i-1)}) \odot \max(0, |Z^{(i)} + U^{(i-1)}| - \frac{\lambda}{\rho})$
 - 5: $U^{(i)} = U^{(i-1)} + Z^{(i)} - Y^{(i)}$
 - 6: **end for**
 - 7: **return** Z
-

Algorithm

Algorithm Dictionary Update

- 1: **Require:** $D = [d_1, \dots, d_k]$, $A = [a_1, \dots, a_k]$, $B = [b_1, \dots, b_k]$
- 2: **repeat**
- 3: **for** $i = 1$ to k **do**
- 4: Update the i -th column:

$$u_i \leftarrow d_i - \frac{1}{A_{ii}}(Da_i - b_i)$$

$$d_i \leftarrow \frac{u_i}{\max(\|u_i\|_2, 1)}$$

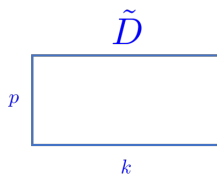
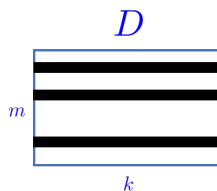
- 5: **end for**
 - 6: **until convergence**
 - 7: **return** D
-

Algorithm

Algorithm Online Dictionary Learning

- 1: **Require:** $D^{(0)}, x_1, \dots, x_n, \lambda$.
 - 2: $A^{(0)} \leftarrow 0, B^{(0)} \leftarrow 0$
 - 3: **for** $t = 1$ to maxstep **do**
 - 4: Random drawing $X^{(t)}$ from $\{x_1, \dots, x_n\}$
 - 5: **Do the Sparse Coding**
 - 6: $A^{(t)} \leftarrow A^{(t-1)} + Z^{(t)}Z^{(t)\top}$
 - 7: $B^{(t)} \leftarrow B^{(t-1)} + X^{(t)}Z^{(t)\top}$
 - 8: **Do the Dictionary Update**
 - 9: **end for**
 - 10: **return** D
-

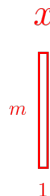
Image inpainting



Sparse coding



$$x = Dz$$



Example of image inpainting

original block



contaminated block



recovered block



- Recovery rate: 80% \sim 90%
- Sparsity: 6%

Numerical result (I) : grayscale image



Before inpainting



After inpainting

Numerical result (I) : color image

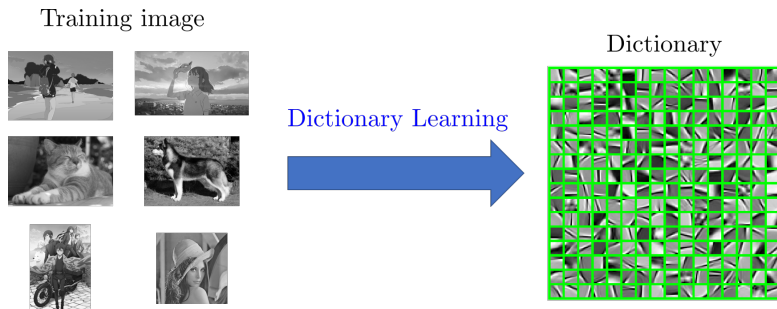


Before inpainting



After inpainting

Dictionary learning from multiple images



Numerical result (II) : grayscale image



Before inpainting



After inpainting

Numerical result (II) : color image



Before inpainting



After inpainting

Reference

- Julien Mairal, Francis Bach, Jean Ponce, Guillermo Sapiro (2009), Online Dictionary Learning for Sparse Coding.
<https://dl.acm.org/doi/10.1145/1553374.1553463>
- Xiaomin Zhang, Jinyu Xia, Jinman Zhao (2015), Image Inpainting using Online Dictionary Learning. <http://pages.cs.wisc.edu/~xiaominz/projs/CS532ImageInpainting.pdf>
- Po-I Tseng, Suh-Yuh Yang (2020), Sparse Representation and Dictionary Learning with Applications to Image Processing.
<http://www.math.ncu.edu.tw/~syyang/>

THE END

Thanks for listening!