

The Research of Quick Dictionary Learning Algorithm under the Framework of Compressed Sensing

Wenchun Yu and Fei Fang

College of Computer science, NeiJiang Normal University, NeiJiang Sichuan
 641110, china

Engineering and Technology College, NeiJiang Normal University, NeiJiang
 Sichuan 641110, china

ywclmxx@163.com(Wenchun Yu)

Abstract

Signal sparse matrix structure, the degree of relationship between signal sparse representation, which affect application of compression perception to the effect of recovery reconstruction for signal. In order to solve this problem, a variety of dictionary learning algorithm such as KSVD, OLM (Online dictionary learning method) should be put forward. These algorithms used overlapping image blocks to build a dictionary, produced a large number of sparse coefficients, resulting in a fitting and calculation too slowly, and cannot ensure convergence. Based on this, it designed a fast dictionary learning algorithm based on proximal gradient. Algorithm based on the analysis of proximal gradient multiple, on the basis of convex optimization problem, applied to the dictionary learning involved in solving optimization, reduce the complexity of each iteration, reduces the iterative overhead, at the same time to ensure the convergence. Experiments on synthetic data show that the proposed algorithm dictionary learning speed, the time is short, and obtain a better dictionary.

Keywords: The dictionary to learn; Sparse representation; The proximal gradient; Global convergence.

1. Introduction

Compressed Sensing (Compressed Sensing) is a kind of using signals of compressibility or sparse on refactoring technology for signals. Compressed sensing subverts the traditional signal sampling method, the signal sparse representation to ensure that the original signal of the main structure, can be less to recover the data sampling to the signal reconstruction, has developed into a new kind of data sampling technique; It goes without saying that its advantage is to reduce the data sampling rate, direct access to the sparse representation of signal, greatly reduce the data access time and storage space. Figure 1 shows the theory process of compression perception. Compressed sensing includes three aspects: signal sparse representation, the design of the observation matrix, signal reconstruction.

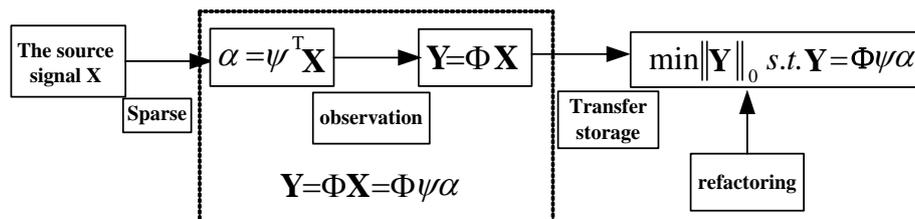


Figure 1. The Theory Process of Compression Perception

Reference [1] presents a sparse representation of mathematical definition: signal $x \in R^n$ under the orthogonal basis $(\psi_i : i = 1, \dots, n)$ of transform coefficient vector is $\theta_i = \langle \psi, x_i \rangle$, If for $0 < p < 2$ and $K > 0$, Coefficient vector meet: $\|\theta_p\| = (\sum_i |\theta_i|^p)^{1/p} \leq K$, it shows the coefficient vector is sparse in a certain sense. TERENCE Tao gives another definition: For a given signal $x \in R^n$, The base $D \in R^{n \times K}$, If meet $x = Dy$ (or $x \approx Dy$), and in middle of $y \in R^K$ only a small part of value other than zero, then we call signal x in the base D is sparse (or approximate sparse).

Assume that the signal x can be sparse representation in the base D , as long as receive signals x under the given base D of linear measurement value $b = x + \xi$ (ξ shows noise), can restore signal x by the sparse representation. The problem is equivalent to optimal solution [2]

$$\min_y \|y\|_0, \quad s.t. \|Dy - b\|_2^2 \leq \varepsilon, \quad (1)$$

Among them, $\|\cdot\|_0$ shows ℓ_0 norm, it shows the number of non zero elements in the vector; ε shows control error. To solve the y by formula (1), then it can obtain the original signal x by $x = Dy$.

MALLAT puts forward using super complete dictionary as the base signal to represent [3] in 1993, and to ensure signal of the sparse representation, it opens the first time of the sparse representation. The study found that [4], signal after sparse representation is more sparse, the higher the accuracy of the signal reconstruction, and sparse representation can according to their structure characteristics of the signal adaptive to choose the appropriate complete dictionary, thus for the purpose of signal sparse representation is to find an adaptive dictionary enable signal to the most sparse expression. So the core problem of sparse representation is to choose an optimal dictionary and suitable sparse decomposition algorithm.

The dictionary can be divided into two categories: analysis of dictionary and learning of dictionary. A common dictionary has wavelet dictionary [4], a complete DCT dictionary and CURVELETS, etc. Using analysis dictionary for sparse representation, easy to realize, but the signal representation of a single and do not have adaptability; by contrast, the learning dictionary can better adapt to the different image data, adaptive ability strong [1,5,6], commonly used method of learning dictionary are: the KSVD algorithm [7] proposed by MICHAEL ELAD in 2006, the OLM algorithm [8] proposed by MAIRAL in 2010.

KSVD and OLM core thought is alternating iterative optimization, there are two major steps: signal sparse representation – sparse representation under fixed dictionary; the dictionary update – under the fixed sparse coefficient of the dictionary update. This kind of dictionary learning algorithm using overlapping image block building the dictionary for sparse representation, alternating iterative update optimization method is applied at the same time, the computational complexity is high, and cannot ensure the convergence of the algorithm.

To solve above problems, this paper presents a dictionary learning algorithm based on accelerating the proximal gradient. In each iteration, the algorithm uses the combined optimization, dictionary atomic updates and sparse representation model at the same time, namely in the optimization to solve of the sparse representation does not restrict the dictionary update. Compared with other algorithms, this algorithm greatly reduces the every step of the iteration complexity, computing faster, at the same time to ensure the global convergence.

2. The Dictionary Learning

In this section, we will introduce two mentioned in the article of the classic KSVD and OLM dictionary learning algorithm.

Assume that the training data set $\mathbf{X} \in R^{n \times p}$, matrix in each column is a signal vector; dictionary $\mathbf{D} \in R^{n \times K}$.KSVD by solving the following optimization problem to achieve a dictionary to learn [7,9]:

$$\begin{aligned} \min_{\mathbf{D}, \mathbf{X}} & \|\mathbf{D}\mathbf{Y} - \mathbf{X}\|_F^2, \\ \text{s.t.} & \|d_i\|_2 = 1, i = 1, \dots, K; \\ & \|y_j\|_0 \leq s, j = 1, \dots, p \end{aligned} \quad (2)$$

Among them, $\|\square\|_2$ shows ℓ_2 norm, s shows the parameter of sparse degree, d_i shows the i column of the dictionary.

Formula (2) can be solved by KSVD algorithm alternating update Y and D, the iteration process major includes two steps: first of all, under the current dictionary for sparse decomposition in middle of X's signal, a lot of algorithms to realize such a sparse representation, such as the OMP, BP algorithm; second, update the atoms in the dictionary, its core is SVD algorithm. Among them, the optimization solution involves the SVD calculation of matrix, a large amount of calculation, largely limits the computation speed; when the image size and the number of iterations is larger, the calculation amount and great consumption of calculation, the using time is too long, the practical value is limited by larger, at the same time, the algorithm is not ensure the convergence.

The another typical dictionary learning method is OLM (Online dictionary learning) [8,10], it by solving the following optimization problem to implement the dictionary to learn:

$$\begin{aligned} \min_{\mathbf{D}, \mathbf{Y}} & \frac{1}{2} \|\mathbf{D}\mathbf{Y} - \mathbf{X}\|_F^2 + \lambda \|\mathbf{Y}\|_1 \\ \text{s.t.} & \|d_i\|_2 \leq 1, i = 1, \dots, K \end{aligned} \quad (3)$$

Among them, $\|\mathbf{Y}\|_1 = \sum_{i,j} |y_{ij}|$, as a $\|\square\|_0$ convex relaxation; λ shows regular parameters, to balance the parameter fitting and sparse degree. In the process of OLM optimization solving, alternately update Y and D:

Sparse representation. Fixed dictionary D, randomly selected the signal vector from matrix X, combined into small signal block, and then calculate the coefficient YS of sparse representation on the dictionary D (among them, the S shows the selection of sample block index);

Update the dictionary. To solve the following optimization problem, update the dictionary D:

$$\min_{\mathbf{D}} \left\{ \|\mathbf{D}\mathbf{Y}_s - \mathbf{X}_s\|_F^2, \|d_i\|_2 \leq 1, \forall i \right\}$$

Among them, \mathbf{X}_s shows matrix X sub-matrices, its elements shows the matrix X of all contains index S refer to elements. The algorithm efficiency depends on the distribution of the training sample, when the sample is with the distribution of its run faster than the KSVD algorithm, but is still difficult to guarantee its convergence; Each iteration involves the matrix block operation, large amount of calculation [10].

Of course, there are many other dictionary learning algorithm [11] and the sparse representation model of more complex [12,13,14], here is no longer here.

Fast solving method of this paper is dedicated to formula (3) research, the basic idea is introduced to accelerate the proximal gradient, uses the combined optimization strategy,

in each iteration, update the dictionary and the coefficient matrix, accelerates the calculation, at the same time to ensure the convergence.

3. Fast Dictionary to be Learn based on the Proximal Gradient

As we know, through the study of dictionary, the dictionary is more suitable for natural images, can shows signal better. There are many ways to study dictionary such as MOD [15], KSVD, OLM, *etc.* In this paper, we adopt a new method to implement the dictionary learning, namely the application of the proximal gradient algorithm to speed up the solving formula (3) optimization problem.

3.1. Accelerate the Proximal Gradient Method Based on Block

For multiple convex optimization problems, Y. XU and W. YIN put forward a kind of based on BPG (Block proximal gradient) solving method [16]; S. MA puts forward a kind of based on APG (Alternating Proximal gradient) solving method [17]. In this paper, we will combine the two methods to solve type such as formula (3) double convex optimization problem:

$$\min_{x,y} f(x, y) + r_x(x) + r_y(y), \quad (4)$$

Among them, function f differentiable, and the variable x 、 y , when one of them is fixed, function f to the another variable for convex function. r_x 、 r_y shows the value expansion of convex function.

BPG method of the k step iteration, x 、 y update the formula is:

$$x^k = \arg \min_x (\nabla_x f(x^k, y^{k-1}), x - x^k) + \frac{L_x^k}{2} \|x - x^k\|_2^2 + r_x(x), \quad (5a)$$

$$y^k = \arg \min_y (\nabla_y f(x^k, y^k), y - y^k) + \frac{L_y^k}{2} \|y - y^k\|_2^2 + r_y(y) \quad (5b)$$

Among them, L_x^k shows function $\nabla_x f(x, y^{k-1})$ about x of LIPSCHITZ constant, L_y^k shows function $\nabla_y f(x^k, y)$ about y of LIPSCHITZ constant. And the corresponding extrapolation point to meet:

$$x^k = x^{k-1} + \omega_x^k (x^{k-1} - x^{k-2}),$$

$$y^k = y^{k-1} + \omega_y^k (y^{k-1} - y^{k-2})$$

Among them, $\omega_x^k \geq 0$, $\omega_y^k \geq 0$.

In some boundary conditions meet, the literature [16] confirmed the algorithm convergence of neutrons sequence. If further satisfy the KL (KURDYKA-LOJASIEWICZ) [18] nature, then the formula (5) generated sequence $\{x_k, y_k\}$ global convergence of a stable point in the formula (4).

3.2. The Dictionary to Learn

By solving Equation (3), learn dictionary from the data sample X .

Build a function:

$$\ell(\mathbf{D}, \mathbf{Y}) = \frac{1}{2} \|\mathbf{D}\mathbf{Y} - \mathbf{X}\|_F^2$$

Obviously, the above formula shows the fidelity term of formula (3); taking the formula (5) into the formula (3), immediately available \mathbf{D} , \mathbf{Y} updates formula:

$$\mathbf{D}^k = \arg \min_{\mathbf{D} \in D} \left\langle \nabla_{\mathbf{D}} \ell(\mathbf{D}^k, \mathbf{Y}^{k-1}), \mathbf{D} - \mathbf{D}^k \right\rangle + \frac{L_d^k}{2} \|\mathbf{D} - \mathbf{D}^k\|_F^2 \tag{6a}$$

$$\mathbf{Y}^k = \arg \min_{\mathbf{Y}} (\nabla_{\mathbf{Y}} \ell(\mathbf{D}^k, \mathbf{Y}^k), \mathbf{Y} - \mathbf{Y}^k) + \frac{L_y^k}{2} \|\mathbf{Y} - \mathbf{Y}^k\|_F^2 + \lambda \|\mathbf{Y}\|_1, \tag{6b}$$

Among them: $D = \{\mathbf{D} : \|d_i\|_2 \leq 1, i = 1, \dots, K\}$, 且 $\mathbf{D} = \mathbf{D}^{k-1} + \omega_d^k (\mathbf{D}^{k-1} - \mathbf{D}^{k-2})$,

$$\mathbf{Y}^k = \mathbf{Y}^{k-1} + \omega_y^k (\mathbf{Y}^{k-1} - \mathbf{Y}^{k-2}).$$

Among them, $\omega_d^k, \omega_y^k \leq 1$, L_d^k 、 L_y^k show function respectively $\nabla_{\mathbf{D}} \ell(\mathbf{D}^k, \mathbf{Y}^{k-1})$ to \mathbf{D} of LIPSCHITZ constant and function $\nabla_{\mathbf{Y}} \ell(\mathbf{D}^k, \mathbf{Y}^k)$ to \mathbf{Y} of LIPSCHITZ constant.

Formula (6) update formula can be rewritten as:

$$\mathbf{D}^k = p_D(\mathbf{D}^k - \frac{1}{L_d^k} \nabla_{\mathbf{D}} \ell(\mathbf{D}^k, \mathbf{Y}^{k-1})) \tag{7a}$$

$$\mathbf{Y}^k = S_{\lambda/L_y^k}(\mathbf{Y}^k - \frac{1}{L_y^k} \nabla_{\mathbf{Y}} \ell(\mathbf{D}^k, \mathbf{Y}^k)), \tag{7b}$$

Among them, $p_D(\square)$ shows in the D projection, it is defined as:

$$(p_D(\mathbf{D}))_i = \frac{d_i}{\max(1, \|d_i\|_2)}, i = 1, \dots, k$$

$S_\tau(\square)$ shows the soft threshold operator, it is defined as follows:

$$S_\tau(\mathbf{Y})_{ij} = \text{sign}(y_{ij}) \square \max(|y_{ij}| - \tau, 0), \forall i, j.$$

Due to the $\nabla_{\mathbf{D}} \ell(\mathbf{D}, \mathbf{Y}) = (\mathbf{D}\mathbf{Y} - \mathbf{X})\mathbf{Y}^T$, So there are:

$$\begin{aligned} \left\| \nabla_{\mathbf{D}} \ell(\mathbf{D}, \mathbf{Y}) - \nabla_{\mathbf{D}} \ell(\mathbf{D}, \mathbf{Y}) \right\|_F &= \left\| (\mathbf{D} - \mathbf{D})\mathbf{Y}\mathbf{Y}^T \right\|_F \\ &\leq \left\| \mathbf{Y}\mathbf{Y}^T \right\|_F \left\| \mathbf{D} - \mathbf{D} \right\|_F, \forall \mathbf{D}, \mathbf{D} \end{aligned}$$

Among them, $\|\square\|$ shows matrix norm; as a result, $\|\mathbf{Y}\mathbf{Y}^T\|$ shows function $\nabla_{\mathbf{D}} \ell(\mathbf{D}, \mathbf{Y})$ to \mathbf{D} of LIPSCHITZ constant.

In this paper, by numerical tests, we set the following LIPSCHITZ constants and extrapolation method involved in the weight:

$$L_d^k = \left\| \mathbf{Y}^{k-1} (\mathbf{Y}^{k-1})^T \right\|, \quad L_y^k = \left\| (\mathbf{D}^k)^T \mathbf{D}^k \right\| \tag{8}$$

$$\omega_d^k = 0.99999 \min(\omega^k, \sqrt{L_d^{k-1} / L_d^k}) \tag{9a}$$

$$\omega_y^k = 0.99999 \min(\omega^k, \sqrt{L_y^{k-1} / L_y^k}) \tag{9b}$$

Among them, $t_0 = 1, \omega^k = \frac{t_{k-1} - 1}{t_k}, t_k = \frac{1}{2} \left(1 + \sqrt{1 + 4t_{k-1}^2} \right)$

The iterative joint method design and weight ω^k applications have used in other algorithms, such as FISTA [19] algorithm; in this paper application of proximal gradient to solve formula (3) optimization problem, have played an important role in accelerating operation. The optimal solution of another key is to ensure that the whole objective function increment. In literature [16, 17] demonstrates the objective function under the condition of not increasing ensures that the algorithm is globally convergent, it reflects the objective function of non increase to the algorithm importance of robustness and fast convergence. So, when $F(\mathbf{D}^k, \mathbf{Y}^k) > F(\mathbf{D}^{k-1}, \mathbf{Y}^{k-1})$, we set up by $\omega_d^k = \omega_y^k = 0$ to repeat the k step iteration, to ensure $F(\mathbf{D}^k, \mathbf{Y}^k)$ is bigger than $F(\mathbf{D}^{k-1}, \mathbf{Y}^{k-1})$.

Among them, formula (3) the optimization problem of objective function is:

$$F(\mathbf{D}, \mathbf{Y}) = \frac{1}{2} \|\mathbf{D}\mathbf{Y} - \mathbf{X}\|_F^2 + \lambda \|\mathbf{Y}\|_1$$

The algorithm of this paper uses proximal gradient method in optimization solving, in the process of update Y and D at the same time, the difference in the KSVD and OLM uses minimization method alternating update D and Y. Keeping on D and Y sub-problems have closed solution, reduce the complexity of the algorithm every step of iteration; the extrapolation is used to design at the same time, further speed up the convergence, reduced the number of iterations, greatly shorten the operation time.

Algorithm : fast dictionary to learn based on the proximal gradient

Data : training sample X, $\lambda > 0$,

Initial point $(\mathbf{D}^{-1}, \mathbf{Y}^{-1}) = (\mathbf{D}^0, \mathbf{Y}^0)$

when $k = 1, 2, \dots$,

According to the formula (8), (9) are calculated respectively L_d^k, ω_d^k :

To calculate $\mathbf{D}^k = \mathbf{D}^{k-1} + \omega_d^k (\mathbf{D}^{k-1} - \mathbf{D}^{k-2})$, and by the formula (7 a) to calculate \mathbf{D}^k ;

According to the formula (8), (9) are calculated respectively L_y^k, ω_y^k :

To calculate $\mathbf{Y}^k = \mathbf{Y}^{k-1} + \omega_y^k (\mathbf{Y}^{k-1} - \mathbf{Y}^{k-2})$, and by the formula (7b) to calculate \mathbf{Y}^k ;

If $F(\mathbf{D}^k, \mathbf{Y}^k) > F(\mathbf{D}^{k-1}, \mathbf{Y}^{k-1})$, then make $\mathbf{D}^k = \mathbf{D}^{k-1}, \mathbf{Y}^k = \mathbf{Y}^{k-1}$, and according to the formula (7a) (7b) update again $\mathbf{D}^k, \mathbf{Y}^k$;

If meet the stop condition (*), then stop operation, and the output $(\mathbf{D}^k, \mathbf{Y}^k)$.

3.3. Convergence Analysis

In this paper, the formula (3) optimization problem is equivalent to:

$$\min_{\mathbf{D}, \mathbf{Y}} \frac{1}{2} \|\mathbf{D}\mathbf{Y} - \mathbf{X}\|_F^2 + \lambda \|\mathbf{Y}\|_1 + \delta_D(\mathbf{D}) \quad (10)$$

Among them, $\delta_D(\square)$ shows about D of indicator function. According to the literature [16], The objective function of formula (10) to meet KL (KURDYKA-LOJASIEWICZ) nature, and sequence $\{\mathbf{D}^k\}$ shows bounded set D; due to the objective function of the formula (10) is not increased, while $\lambda > 0$, so sequence $\{\mathbf{Y}^k\}$ bounded. So, $(\mathbf{D}^k, \mathbf{Y}^k)$

shows limited set of points, and it determined by the formula (8) the LIPSCHITZ constant also must be bounded. In other words, L_d^k 、 L_y^k consensus is greater than zero, the sequence $\{\mathbf{D}^k\}$ and $\{\mathbf{Y}^k\}$ consistently deviating from the starting point. Therefore, according to the theorem 2.8 in literature [16], we are easy to have a conclusion:

Assume that $(\mathbf{D}^k, \mathbf{Y}^k)$ shows according to the algorithm generated sequence, if sequence $\{\mathbf{D}^k\}$ and $\{\mathbf{Y}^k\}$ consistently deviating from the starting point. then $(\mathbf{D}^k, \mathbf{Y}^k)$ will converge to the formula (10), or a stable point in the formula (3).

4. Numerical Experiments

This paper sets up two groups of contrast experiment, which are respectively with the classic KSVD, OLM contrast experiments on synthetic data. We chose to use KSVD and OLM comparison, because these two kinds of algorithm is widely used, but also because of the effects of two kinds of algorithms are fully validated. Test laboratory index for: the average computation speed, dictionary learning efficiency.

Dictionary learning efficiency, this article is defined as follows. From the original dictionary \mathbf{D} recover each atom d meet:

$$\max_{1 \leq i \leq K} \frac{|d^T d_i|}{\|d\|_2 \|d_i\|_2} \geq 0.99$$

We say that dictionary update as successful. Among them, d_i shows predict the dictionary \mathbf{D} of the i column. This paper uses a dictionary update rate to refer to a dictionary learning efficiency.

Termination conditions of algorithm in this paper are as follows:

$$\frac{|F(\mathbf{D}^k, \mathbf{Y}^k) - F(\mathbf{D}^{k+1}, \mathbf{Y}^{k+1})|}{1 + F(\mathbf{D}^k, \mathbf{Y}^k)} \leq 10^{-4} \quad (*)$$

Combined with the literature [7, 10], we could begin experiment by the following steps:

The first step, generate experimental data needed. First of all, to generate a dictionary $\mathbf{D} \in R^{n \times K}$ (namely random (n, K)), and normalized for each column; and then, in the n d space generate p signal, composition the training sample \mathbf{X} ($\mathbf{X} \in R^{n \times p}$). Among them, each sample is evenly randomly selected in the dictionary \mathbf{D} of r column of atoms was formed by linear combination, and combination coefficient is Gaussian random generated.

The second step, on the above generated data, running KSVD, OLM and algorithm (FDL_PG) for dictionary learning. Among them, the essence of the KSVD dictionary to learn shows formula (2) optimize the solving process, OLM and the essence of this paper algorithm is the formula (3) optimization solving process.

In the formula (2) we assume that the sparse degree $s = r$, we set $\lambda = 0.5/\sqrt{n}$ in the formula (3). In this paper experiment, setting $n=36$.

The third step, we take three groups of different data to the experiment of (K, p), and in each group (K, p), we test five groups of data under different sparseness degree r, among them, the range of r for {4, 6, 8, 10, 12}. Experiment we run 100 times independently in each group, and then statistical computing the average run time (T) and recovery accuracy (R).

Three groups of experimental data, is shown in the table below:

Table 1. Under the First Group (K, p) Three Kinds of Algorithm Experimental Results Contrast

| | FDL_PG | KSVD | OLM |
|----------|----------------------|-------------|-------------|
| <i>r</i> | T (s) R (%) | T(s) R(%) | T(s) R(%) |
| | $(K, p) = (2n, 20n)$ | | |
| 4 | 0.676 98.75 | 28.80 98.11 | 5.635 98.65 |
| 6 | 0.808 98.47 | 32.88 91.68 | 9.893 98.26 |
| 8 | 1.052 97.84 | 34.81 0.00 | 16.43 97.71 |
| 10 | 1.454 94.90 | 38.34 0.00 | 21.50 86.53 |
| 12 | 2.212 64.31 | 41.51 0.00 | 29.45 44.38 |

Table 2. Under the Second Group (K, p) Three Kinds of Algorithm Experimental Results Contrast

| | FDL_PG | KSVD | OLM |
|----------|-----------------------|-------------|-------------|
| <i>r</i> | T(s) R(%) | T(s) R(%) | T(s) R(%) |
| | $(K, p) = (2n, 100n)$ | | |
| 4 | 2.425 99.42 | 41.61 98.54 | 6.237 99.70 |
| 6 | 2.819 98.96 | 44.53 98.81 | 11.60 99.03 |
| 8 | 3.401 98.61 | 51.90 97.29 | 19.69 99.02 |
| 10 | 4.180 98.47 | 56.59 16.94 | 25.32 99.08 |
| 12 | 5.438 96.52 | 62.98 0.00 | 33.25 98.75 |

Table 3. Under the Third Group (K, p) Three Kinds of Algorithm Experimental Results Contrast

| | FDL_PG | KSVD | OLM |
|----------|-----------------------|-------------|-------------|
| <i>r</i> | T (s) R (%) | T (s) R (%) | T(s) R(%) |
| | $(K, p) = (4n, 100n)$ | | |
| 4 | 7.080 99.03 | 71.47 98.96 | 12.51 98.95 |
| 6 | 9.382 98.82 | 77.74 97.43 | 24.49 99.18 |
| 8 | 11.76 98.02 | 83.51 87.51 | 42.53 98.97 |
| 10 | 16.36 96.18 | 89.54 0.00 | 59.70 97.45 |
| 12 | 22.86 79.83 | 94.04 0.00 | 72.20 65.26 |

The analysis of the above three groups of experimental data, we can get:

Dictionary learning effect quite under the condition of same with the KSVD contrast (namely the success rate of the dictionary updating is quite), but used time significantly shorter than KSVD algorithm in this paper, calculation speed advantage is obvious; when the sparse degree of *r* value larger (such as *r*=12) or the training sample is limited (such as *p*=20*n*), the success rate of this paper algorithm under the condition of same dictionary update significantly higher, namely the algorithm of this paper dictionary learning effect is significantly better than that of the KSVD;

Compared with OLM, observing the first group members of the data, the results are shown in table 1, can be found under the condition of same OLM dictionary update success rate is lower than the algorithm in this paper, namely the algorithm in this paper of the dictionary learning effect is better than that of OLM dictionary learning effect. After the two groups of the experiment, the experimental results as shown in table 2 and table 3, when not to consider running time, under the condition of same OLM algorithm and the algorithm of this paper dictionary update success rate similar; considering the operation time, easy to find the same conditions this algorithm with the OLM dictionary

learning effect quite, but this paper algorithm used time significantly shorter, namely its convergence speed is faster.

5. Conclusion

Dictionary learn more and more widely used in the field of image processing and so on, and the complexity of the algorithm and convergence restricts largely the dictionary to learn the application and promotion, and gradually become the focus of research. In this paper, we developed a new dictionary learning algorithm, different from other methods for solving dictionary learning problems, we quoted the proximal gradient algorithm to solve the dictionary learning involves of combination of optimization problems. Our algorithm is mainly aimed at the non convex smooth function and separable of the convex function of combination function of optimization, takes combination optimization, dictionary atomic updates and sparse representation model at the same time, reduce the complexity of each step iteration, accelerate the computation speed, at the same time to ensure the stable point of convergence, has stronger practicability, also provide a reference for sparse representation problem solving.

Acknowledgements

This work was supported by the Sichuan Education Department under Grant No.13ZA0005.

References

- [1] D. L. Donoho, "Compressed sensing", *Information Theory, IEEE Transactions*, vol. 52, no. 4, (2006), pp. 1289-1306.
- [2] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries", *Image Processing, IEEE Transactions*, vol. 15, no. 12, (2006), pp. 3736-3745.
- [3] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries", *Signal Processing, IEEE Transactions on*, vol. 41, no. 12, (1993), pp. 3397-3415.
- [4] C. Emmanuel, R. Justin and T. Terence, "Stable signal recovery from incomplete and inaccurate measurements", *Communications on pure and applied mathematics*, vol. 59, no. 8, (2006), pp. 1207-1223.
- [5] K. K. Delgado, J. F. Murray and B. D. Rao, "Dictionary learning algorithms for sparse representation", *Neural computation*, vol. 15, no. 2, (2003), pp. 349-396.
- [6] J. Mairal, J. Ponce and G. Sapiro, "Supervised dictionary learning", *Advances in neural information processing systems*, (2009), pp. 1033-1040.
- [7] M. Elad, M. Aharon and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation", *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, (2006), pp. 4311-4322.
- [8] J. Mairal, F. Bach and J. Ponce, "Online learning for matrix factorization and sparse coding", *The Journal of Machine Learning Research*, vol. 11, (2010), pp. 19-60.
- [9] R. Ron, T. Peleg and M. Elad, "Analysis K-SVD: A dictionary-learning algorithm for the analysis sparse model", *Signal Processing, IEEE Transactions on*, vol. 61, no. 3, (2013), pp. 661-677.
- [10] J. Mairal, F. Bach and J. Ponce J, "Online dictionary learning for sparse coding", *Proceedings of the 26th Annual International Conference on Machine Learning, ACM*, (2009), pp.689-696.
- [11] I. Todic and P. Frossard, "Dictionary learning", *Signal Processing Magazine, IEEE*, vol. 28, no. 2, (2011), pp. 27-38.
- [12] A. M. Bruckstein, D. L. Donoho and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images", *SIAM review*, vol. 51, 1, (2009), pp. 34-81.
- [13] J. Mairal, F. Bach and J. Ponce, "Task-driven dictionary learning", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 4, (2012), pp. 791-804.
- [14] J. Mairal, J. Ponce and G. Sapiro, "Supervised dictionary learning", *Advances in neural information processing systems*, (2009), pp. 1033-1040.
- [15] K. Engan, S. O. Aase and J. H. Husøy, "Multi-frame compression: Theory and design", *Signal Processing*, vol. 80, no. 10, (2000), pp. 2121-2140.
- [16] Y. Xu and W. Yin, "A block coordinate descent method for multiconvex optimization with applications to nonnegative tensor factorization and completion", *RICE UNIV HOUSTON TX DEPT OF COMPUTATIONAL AND APPLIED MATHEMATICS*, (2012).

- [17] S. Ma, "Alternating proximal gradient method for convex minimization", Preprint, (2012), http://www.optimization online.org/DB_HTML/2012/09/3608.html.
- [18] J. Bolte, A. Daniilidis and A. Lewis, "The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems", SIAM Journal on Optimization, vol. 17, no. 4, (2007), pp. 1205-1223.
- [19] A. Beck and M. A. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems", SIAM Journal on Imaging Sciences, vol. 2, no. 1, (2009), pp. 183-202

Authors



Wen-Chun Yu, he was born in Sichuan Province, China, in 1974. He received the B.S. degree from the Sichuan Normal University of China (SICNUU), Chengdu, in 1997 and the M.S. degree from the Southwest Jiao tong University of China (WJITU), Chengdu, in 2009. His research interests include Wireless Local Area Network Large data processing of network and Data Mining.



Fei Fang, he was born in Sichuan Province, China, in 1974. He received the B.S. degree from the Southwest University of China (SWU), Chongqing, in 1997 and the M.S. degree from the Chongqing University of Posts and Telecommunications of China (CQUPT), Chongqing, in 2004. He received the doctor's degree from the School of Communication and Information Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, in 2014. His research interests include Wireless Local Area Network, Cognitive radio Wireless network.