On a development of sparse PCA method for face recognition problem

Loc TRAN
CHArt Laboratory
EPHE - Paris-Saclay University
Paris, France
tran0398@umn.edu

Bich NGO
University of Science HCMC, Vietnam
Vietnam National University, HCMC, Vietnam
1711064@student.hcmus.edu.vn

Tuan TRAN
CHArt Laboratory
EPHE - Paris-Saclay University
Paris, France
anh-tuan.tran@etu.ephe.psl.eu

Lam PHAM
University of Information Technology, HCMC, Vietnam
Vietnam National University, HCMC, Vietnam
16520645@gm.uit.edu.vn

An MAI*
Department of Computer Science & Engineering
International University, Ho Chi Minh city, Vietnam
Vietnam National University, Ho Chi Minh city, Vietnam
mhban@hcmiu.edu.vn

***Abstract*—Face recognition is a very significant branch of application in the pattern recognition area. It has multiple applications in the military and finance, to name a few. In reality, sometimes we have to deal with sparse representation of the facial data. In this paper, to deal with sparsity, two advanced versions of Sparse PCA are developed, including Proximal Gradient Sparse PCA (PG Sparse PCA) and Fast Iterative Shrinkage-Thresholding Algorithm Sparse PCA (FISTA Sparse PCA). Then they will be respectively applied to solve the face recognition problem by considering their combination with nearest-neighbor method and with the kernel ridge regression method. Experimental results illustrate that the accuracies of PG Sparse PCA and FISTA Sparse PCA are equivalent in the combination with nearest-neighbor, while PG Sparse PCA perform better than FISTA Sparse PCA in the case of using kernel ridge regression. However, we recognize that the computing process of FISTA Sparse PCA, on average, is always faster than the PG sparse PCA version due to the use of a fast proximal gradient version.**

***Index Terms*—sparse PCA, kernel ridge regression, face recognition, nearest-neighbor, proximal gradient, FISTA**

I. Introduction and related works

Face recognition is one type of bio-metric security systems. It has a lot of applications in finance and military. In the years 1990, computer scientists have employed the Eigenface [1] to recognize faces. The Eigenface method utilized the Principle Component Analysis (PCA) which is one of the dimensional reduction methods [2], [3] to reduce the dimensions of the data in order to remove noise and redundant features in the data. This will lead to the low time complexity of the face recognition system. The PCA method has been utilized in a lot of applications such as speech recognition [4], [5]. Then, the Eigenface used the nearest-neighbor method [6] to recognize faces.

To recognize faces, the machine learning approach [7]–[12] can also be employed such as the Support Vector Machine (SVM) method. From 2000 to 2010, the Support Vector Machine learning method can be considered the state of the art machine learning method. Note that the kernel ridge regression method can be viewed as the simplest form of the Support Vector Machine learning method. In this paper, in addition to the nearest-neighbor method, we will also consider the kernel ridge regression method in the combination of two proposed Sparse PCA versions to recognize faces for the illustration.

It should be noted that the original PCA has two major weaknesses which are the nonexistence of sparsity of the loading vectors and each principle component is the linear combination of all features, that may not reflect all relations of the features. To overcome these weaknesses and to present sparsity, many methods have been suggested such as [13]–[16]. In our work, we will introduce two new approaches solving sparse PCA using the Proximal Gradient method [17]–[19], and the Fast Iterative Shrinkage-Thresholding Algorithm method (i.e. FISTA method) [17]–[19]. Finally, we will try to combine these Sparse PCA dimensional reduction methods with the nearest-neighbor method (and with the kernel ridge regression method) to solve the face recognition problem.

We will organize the paper as follows: In section I, we shortly introduce our proposed discussion in the paper and review several notable related works. Then, section II will present the detailed version of the sparse PCA algorithm using the proximal gradient method in detail. Section III continues to present the detailed version of the sparse PCA algorithm using the FISTA method in detail. In section IV, we will apply the combination of these advanced Sparse PCA algorithms with the nearest-neighbor algorithm and with the kernel ridge regression algorithm to recognize the faces in the dataset available from [17]. Finally, section V will conclude this paper and discuss the future directions of researches of this face recognition problem.

*Corresponding author.

II. The proximal gradient method

Suppose that we would like to solve the unconstrained optimization problem like the following

$$\min_z f(z). \tag{1}$$

One of the simplest methods solving the above unconstrained optimization problem is the gradient method. This gradient method solve the above optimization as follows

$$z_0 \in R^n, z_{k+1} = z_k - t_{k+1}\nabla f(z_k), \tag{2}$$

where t_{k+1} is the suitable step size.

The idea of this gradient method can also be applied to the l_1-regularization problem like the following.

Suppose that we are given a function f that can be decomposed into two functions as follows

$$f(x) = g(x) + \lambda h(x). \tag{3}$$

For this decomposable function f, we want to find the solution of the following optimization problem

$$\min_z f(z) = g(z) + \lambda h(z). \tag{4}$$

Note that function g is differentiable. So, the solution of this optimization is

$$\begin{aligned} x^+ &= \operatorname*{argmin}_z g(x) + \nabla g(x)^T(z-x) + \frac{1}{2t}||z-x||_2^2 + \lambda h(z) \\ &= \operatorname*{argmin}_z \frac{1}{2t}||z-(x-t\nabla g(x))||_2^2 + \lambda h(z) \\ &= \operatorname*{prox}_{h,\lambda t}(x - t\nabla g(x)). \end{aligned} \tag{5}$$

Here, we should pay attention to a point that is $\nabla^2 g(x)$ is replaced by $1/tI$.

This method can also be called the ISTA method. In the next section, we will present the FISTA method which is the extended version of the ISTA method and is a lot faster than the ISTA method.

In our sparse PCA problem, $g(x) = -x^T D^T D x$ and $h(x) = ||x||_1$. So, we know that $\nabla g(x) = -2D^T Dx$. Hence the solution of the sparse PCA problem is

$$\begin{aligned} x^+ &= \operatorname*{prox}_{h,\lambda t}(x + 2tD^T Dx) \\ &= \operatorname*{prox}_{h,\lambda t}((I + 2tD^T D)x). \end{aligned} \tag{6}$$

Last but not least, in detail, the proximal operator is defined like the following

$$\begin{aligned} \operatorname*{prox}_{h,\lambda t}(x - t\nabla g(x))_i = (|x + 2tD^T Dx|_i - \lambda t)_+ \\ \cdot sign((x + 2tD^T Dx)_i). \end{aligned} \tag{7}$$

We repeat the above formula until it converges to the final solution.

Next, we will give the algorithm of one pass of the proximal gradient method (i.e. the ISTA method).

Algorithm 1 One pass of proximal gradient method

1: Start with x_k
2: Compute $x_{k+1} = \operatorname{prox}_{h,\lambda t}((I + 2tD^T D)x_k)$
3: x_{k+1} will be served as the input for the next pass of the ISTA method

III. The FISTA method

In this section, we will describe the modified version of the proximal gradient method (i.e. FISTA method) used to solve the sparse PCA problem.

First, randomly choose *t_old_old, t_old, x_old_old, x_old*. Then, we need to compute

$$y = x_{\text{old}} + \frac{t_{\text{old\_old}} - 1}{t_{\text{old}}}(x_{\text{old}} - x_{\text{old\_old}}). \tag{8}$$

Next, we need to compute the solution of the following formula

$$x^+ = \operatorname*{prox}_{h,\lambda t}(y - t\nabla g(y)). \tag{9}$$

We then obtain $t = \frac{1 + \sqrt{1 + 4t_{\text{old}}^2}}{2}$.

Finally, we set $t\_old = t$, $t\_old\_old = t\_old$, $x\_old = x^+$, $x\_old\_old = x\_old$.

We repeat the above procedure until it converges to the final solution.

Algorithm 2 One pass of FISTA method

1: Start with *t_old_old, t_old, x_old_old, x_old*
2: Compute $y = x_{\text{old}} + \frac{t_{\text{old\_old}} - 1}{t_{\text{old}}}(x_{\text{old}} - x_{\text{old\_old}})$
3: Compute $x_{k+1} = \operatorname{prox}_{h,\lambda t}((I + 2tD^T D)y)$
4: Compute $t = \frac{1 + \sqrt{1 + 4t_{\text{old}}^2}}{2}$
5: $x_{k+1}, x_{\text{old}}, t, t_{\text{old}}$ will be served as the inputs for the next pass of the FISTA method

IV. Experiments and Results

In this paper, the set of 120 face samples recorded of 15 different people (8 face samples per people) is the training set. Then another set of 45 face samples of these people is the testing set. This dataset is available from [20]. Then, we will merge all rows of the face sample (i.e. the matrix) sequentially from the first row to the last row into a single big row which is the R^{1*1024} row vector. Then, to increase the sparsity for illustration, we intentionally add a number of zeros row by row. These row vectors will be used as the feature vectors of the learning models for the recognition problem, in this paper, nearest neighbor and kernel ridge regression.

Next, two versions of Sparse PCA algorithms will be applied to recognize the faces in the training set and the testing set to reduce the dimensions of the facial representation vectors. Then the nearest-neighbor method and the kernel ridge regression method will be applied to these new transformed feature vectors.

In this section, we illustrate our experiment based on two machine learning algorithms of nearest-neighbor method and kernel ridge regression method, and evaluate the results in terms of accuracy. The accuracy measure Q is given as follows:

$$Q = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

in which TP is True Positive, TN is True Negative, FP is False Positive and FN is False Negative.

All experiments were implemented in Matlab 6.5 on virtual machine. The accuracies of the above proposed methods are given in the following table I and table II and the time complexities of the computing processes of the Sparse PCA using the proximal gradient (PG) and the FISTA methods are given in the corresponding table III.

Table I: Accuracies of the PG Sparse PCA and FISTA Sparse PCA in the combination of nearest-neighbor method in the face recognition problem

Accuracy (%)	
PG Sparse PCA (d = 200) + The nearest-neighbor method	87.86
PG Sparse PCA (d = 300) + The nearest-neighbor method	87.96
PG Sparse PCA (d = 400) + The nearest-neighbor method	87.85
PG Sparse PCA (d = 500) + The nearest-neighbor method	87.96
PG Sparse PCA (d = 600) + The nearest-neighbor method	87.86
FISTA Sparse PCA (d = 200) + The nearest-neighbor method	88.15
FISTA Sparse PCA (d = 300) + The nearest-neighbor method	88.44
FISTA Sparse PCA (d = 400) + The nearest-neighbor method	88.74
FISTA Sparse PCA (d = 500) + The nearest-neighbor method	89.04
FISTA Sparse PCA (d = 600) + The nearest-neighbor method	88.74

Table II: Accuracies of the PG Sparse PCA and FISTA Sparse PCA in the combination of the kernel ridge regression method in the face recognition problem

Accuracy (%)	
PG Sparse PCA (d = 200) + The kernel ridge regression method	**95.26**
PG Sparse PCA (d = 300) + The kernel ridge regression method	**96.15**
PG Sparse PCA (d = 400) + The kernel ridge regression method	**96.44**
PG Sparse PCA (d = 500) + The kernel ridge regression method	**96.15**
PG Sparse PCA (d = 600) + The kernel ridge regression method	**96.15**
FISTA Sparse PCA (d = 200) + The kernel ridge regression method	87.26
FISTA Sparse PCA (d = 300) + The kernel ridge regression method	89.93
FISTA Sparse PCA (d = 400) + The kernel ridge regression method	90.52
FISTA Sparse PCA (d = 500) + The kernel ridge regression method	92.30
FISTA Sparse PCA (d = 600) + The kernel ridge regression method	91.11

From the above tables I, II and III, we recognize the accuracies of using PG and FISTA Sparse PCA methods are promising to deal with face recognition problem with sparse facial data. But as we choose the initial vector of the proximal gradient method (i.e. the ISTA method) and the FISTA method randomly, this will not lead to the best accuracy performance measures of these two methods in this experiments. This vectors need to be tuned carefully, but it is outside the scope of this paper since the goal is just indicate the potential application of these two Sparse PCA versions.

Table III: The average time complexities of the computing processes of the PG and FISTA Sparse PCA

The time complexities (in seconds)	
PG Sparse PCA (d = 200)	2232
PG Sparse PCA (d = 300)	3117
PG Sparse PCA (d = 400)	3537
PG Sparse PCA (d = 500)	4837
PG Sparse PCA (d = 600)	5370
FISTA Sparse PCA (d = 200)	**439**
FISTA Sparse PCA (d = 300)	**806**
FISTA Sparse PCA (d = 400)	**1210**
FISTA Sparse PCA (d = 500)	**1489**
FISTA Sparse PCA (d = 600)	**1634**

Last but not least, from the experiments, the process computing the Sparse PCA algorithm using FISTA method (FISTA Sparse PCA) is a lot faster than the process computing the sparse PCA algorithm using proximal gradient method (PG Sparse PCA).

V. Conclusions

In this paper, two extended versions of the Sparse PCA method solved by the proximal gradient (PG) method and the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) method have been proposed and implemented. The experimental results show that the accuracies of these two approaches are promising to deal with face recognition problem with high sparsity level.

In the future research, we will discuss about the combination of the tensor Sparse PCA method with a lot of classification systems, for e.g. the SVM method and deep neural network method, for applying to not only the face recognition problem but also other image recognition problems such as finger print recognition problem with different levels of sparsity.

References

[1] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[2] L. H. Tran, L. H. Tran, H. Trang *et al.*, "Combinatorial and random walk hypergraph laplacian eigenmaps," *International Journal of Machine Learning and Computing*, vol. 5, no. 6, p. 462, 2015.

[3] L. Tran, A. Mai, T. Quan, and L. Tran, "Weighted un-normalized hypergraph laplacian eigenmaps for classification problems." *International Journal of Advances in Soft Computing & Its Applications*, vol. 10, no. 3, 2018.

[4] H. Trang, T. H. Loc, and H. B. H. Nam, "Proposed combination of pca and mfcc feature extraction in speech recognition system," in *2014 International Conference on Advanced Technologies for Communications (ATC 2014)*. IEEE, 2014, pp. 697–702.

[5] L. H. Tran and L. H. Tran, "The combination of sparse principle component analysis and kernel ridge regression methods applied to speech recognition problem," *Int J Adv Soft Comput Appl*, vol. 10, no. 2, 2018.

[6] Z. Zhang, "Introduction to machine learning: k-nearest neighbors," *Annals of translational medicine*, vol. 4, no. 11, 2016.

[7] B. Schölkopf, A. J. Smola, F. Bach *et al.*, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.

[8] H. Trang and L. Tran, "Kernel ridge regression method applied to speech recognition problem: A novel approach," in *2014 International Conference on Advanced Technologies for Communications (ATC 2014)*. IEEE, 2014, pp. 172–174.

[9] J. M. Zurada, *Introduction to artificial neural systems*. West St. Paul, 1992, vol. 8.

[10] L. Tran, "Application of three graph laplacian based semi-supervised learning methods to protein function prediction problem," *arXiv preprint arXiv:1211.4289*, 2012.

[11] L. Tran and L. Tran, "The un-normalized graph p-laplacian based semi-supervised learning method and speech recognition problem," *International Journal of Advances in Soft Computing & Its Applications*, vol. 9, no. 1, 2017.

[12] H. Trang and L. H. Tran, "Graph based semi-supervised learning methods applied to speech recognition problem," in *International Conference on Nature of Computation and Communication*. Springer, 2014, pp. 264–273.

[13] R. Hausman, "Constrained multivariate analysis," *Optimisation in Statistics*, pp. 137–151, 1982.

[14] S. Vines, "Simple principal components," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 49, no. 4, pp. 441–451, 2000.

[15] I. T. Jolliffe, N. T. Trendafilov, and M. Uddin, "A modified principal component technique based on the lasso," *Journal of computational and Graphical Statistics*, vol. 12, no. 3, pp. 531–547, 2003.

[16] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *Journal of computational and graphical statistics*, vol. 15, no. 2, pp. 265–286, 2006.

[17] S. Tao, D. Boley, and S. Zhang, "Local linear convergence of ista and fista on the lasso problem," *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 313–336, 2016.

[18] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[19] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 57, no. 11, pp. 1413–1457, 2004.

[20] http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html.