Tạp chí Khoa học Giáo dục Kỹ thuật, số 11
Đại học Sư phạm Kỹ thuật Thành phố Hồ Chí Minh

NON-NEGATIVE SPARSE PRINCIPAL COMPONENT ANALYSIS

Thanh D. X. Duong
Ton Duc Thang University, Vietnam
{thanhddx@tut.edu.vn}

ABSTRACT

With applications throughout science and engineering, sparse principal component analysis considers the problem of maximizing the variance explained by a particular linear combination of the input variables where the number of nonzero coefficients is constrained. In this paper, we consider nonnegative sparse principal component analysis in which the coefficients in the combination are required to be non-negative. We improve our recent results by applying an appropriate re-weighted algorithms. Numerical results illustrate the improvement.

Key words: *principal component analysis, semi-definite relaxation, semi-definite programming, l_1 -minimization, iterative reweighting, bisection algorithm*

1. INTRODUCTION

Principal component analysis (PCA) is a popular technique used to reduce multidimensional data sets to lower dimensions for analysis with applications throughout science and engineering, see [17]. This reduction is achieved by transforming to a new set of variables, the principal components, which are uncorrelated and ordered so that the first few retain most of the variation present in all of the original variables. PCA was first introduced by Pearson in [22], and developed independently by Hotelling in [12]. It can be performed via a singular value decomposition of the data matrix.

However, PCA has drawbacks since the principal components are usually linear combinations of all variables and the loadings are typically non-zero. This makes it often difficult to be applied in many applications where the principal components would be convenient if these components contained very few nonzero loadings. Besides the application to trajectory-based air traffic management [7], some other applications are financial asset trading strategies in which fewer non-zero loadings imply fewer transaction costs and gene expression data analysis where the sparsity is necessary for finding focalized local patterns hidden in the data, see [2]. Hence, it is desirable to study sparse principal components explaining most of the variance present in the data. To achieve this, it is necessary to sacrifice some of the explained variance and the orthogonality of the principal components. There are some approach to sparse PCA. Rotation techniques in [15] can be consider the first approach. In [25], the author studied simple principal components by restricting the loadings to take values from a small set of allowable integers such as 0, 1 and -1. Simple thresholding techniques [4] was an ad hoc way to deal with the problem, where the loadings with small absolute value are thresholded to zero. SCoTLASS [16] and SLRA [27, 28] were introduced to get modified principal components with possible zero loadings. ESPCA [20] used discrete spectral formulation based on variational eigenvalue bounds and an effective greedy strategy to give provably optimal solutions via branch-and-bound search. For very large problems, SPCA [29] was proposed via a regression type optimization problem and DSPCA [6] via relaxing a hard cardinality constraint with a convex approximation.

But sparsity is still not enough for some applications where the nonnegativity property of the loadings are required. In particular, non-negative loadings increase efficiency of risk reduction in large portfolios, see [13], and is required due to the robustness of biological systems in [2]. Using matrix factorization approach, NSPCA in [26] studied PCA with both nonnegativity and sparsity properties. This method depends on two parameters - the first one is a balancing parameter between reconstruction and orthogonality and the second controls the amount of additional sparsity required. However, there is no algorithm designed for finding the suitable parameters.

Recently, we propose a direct approach improving the sparsity of the nonnegative principal components - also called NSPCA in [8], by directly incorporating a sparsity criterion in the PCA problem formulation. Next, the problem is relaxed to be a semi-definite program (SDP), which can be solved efficiently in polynomial time via interior-point methods [23, 24]. Then, we add a postprocessing technique, since the outputs of all available approaches do not satisfied sparsity constraint - i.e. if we hope to find a principal component with less than k non-zero entries, the output often contains more than k non-zero entries by using re-weighted l1 minimization in [3, 5]. In this way, we assume that the input data is appropriate to find out a good non-negative principal component. Hence, we remove the negative entries of the principal component obtained via re-weighted l1 minimization iteration, and normalize it. In this paper, we want to improve the removing naturally by setting the infinity weights for the negative entries at each re-weighted l1 minimization iteration. This paper is organized as follows. The next section is the main results, where we presents our NSPCA method by applying relaxing technique and improving re-weighted l1 minimization technique. Section 3 is the numerical experiment on both artificial data and real-life data.

Notation. In this paper, we denote the set of symmetric matrices of size n by $\mathbf{S}^n$, the vector of ones by $\mathbf{1}$, the cardinality (number of non-zero elements) of a vector x by **Card** (x), and the number of non-zero coefficients in a matrix X by **Card** (X). For $X \in S^n$, the notation $X \geq 0$ means that X is positive semi-definite, and $\|x\|_2$ is the 2-Euclide norm for $x \in \mathbf{C}^n$.

2. MAIN RESULTS

In this section, we derive an SDP relaxation for the problem of maximizing the variance explained by a non-negative vector while constraining its cardinality via a improved re-weighted l_1 technique. Then, we apply the problem to decompose a data matrix into non-negative sparse factors.

2.1 Semi-definite Relaxation

Let $A \in \mathbf{S}^n$ be a covariance matrix, i.e. $A \geq 0$, and k be an integer with $1 \leq k \leq n$. We consider the problem of maximizing the variance of a non-negative vector $x \in \mathbf{R}^n$ while constraining its cardinality:

$$\begin{aligned} \text{maximize} \quad & x^T Ax, \\ \text{subject to} \quad & \|x\|_2 = 1, \\ & \mathbf{Card}(x) \le k, \\ & x \ge 0. \end{aligned} \tag{1}$$

It is remarkable that problem (1) is a NP-hard problem. And to relax it to be a convex problem, we do the same procedures as in [8] by setting $X = xx^T \ge 0$. Using the lifting procedure for semi-definite relaxation, truncation technique and l_1 norm constraint technique, see for examples [1, 3, 11, 6, 18, 19], we get a relaxation of (1) as follows:

$$\begin{aligned} \text{maximize} \quad & \mathbf{Tr}(AX), \\ \text{subject to} \quad & \mathbf{Tr}(X) = 1, \\ & \mathbf{1}^T X \mathbf{1} \le k, \\ & X \ge 0, \\ & X \, ?0. \end{aligned} \tag{2}$$

This means, we will solve the semi-definite problem (2) to get solution X, and an approximation solution of (1) is the non-negative parts of the dominant eigenvector of X.

2.2 Cardinality Constraint Refinement

Let x_* be the approximation solution of (1). It is clear that x_* does not satisfying cardinality constraint $\mathbf{Card}(x) \le k$ in general. Hence, we consider the following cardinality constraint refinement problem:

$$\begin{aligned} \text{minimize} \quad & \mathbf{Card}(x), \\ \text{subject to} \quad & \|x\|_2 = 1, \\ & x^T Ax \ge c_*, \\ & x \ge 0, \end{aligned} \tag{3}$$

where $x_* := x_*^T Ax_*$.

By the same arguments as in the last subsection and setting, $X = xx^T \ge 0$, we can relax the problem (3) as follows:

$$\begin{aligned} \text{minimize} \quad & \mathbf{Card}(X), \\ \text{subject to} \quad & \mathbf{Tr}(X) = 1, \\ & \mathbf{Tr}(AX) \ge c_*, \\ & X \ge 0, \\ & X \, ?0. \end{aligned} \tag{4}$$

It is noticeable that the condition $X \ge 0$ implies the nonnegative of its dominant eigenvector provided **rank**(X). Hence, by the re-weighted l_1 minimization technique, see [11], we consider the following relaxation of the problems (4):

$$\begin{aligned} \text{minimize} \quad & \mathbf{Tr} W^T X, \\ \text{subject to} \quad & \mathbf{Tr}(X) = 1, \\ & \mathbf{Tr}(AX) \geq c_*, \\ & X \geq 0, \\ & X \geq 0, \end{aligned} \tag{5}$$

where $W > 0$ is positive rank one weight matrix with an improved iterative algorithm that tries to warranty the condition **rank**$(X) = 1$:

1. Set the iteration count m to zero and solve problem (2) to get the approximate solution $x(0)$.
2. Update the weights: for each $I = 1,\ldots, n$, set
$$w_i^{(m+1)} := \frac{1}{x_i^{(m)} + \varepsilon}, \tag{6}$$
and $W := ww^T$ - a rank one matrix.
3. Solve the weighted l_1 minimization problem (5) to get the approximate solution $x^{(m)}$.
4. Terminate when **Card**$(x^{(m)}) \leq$ k or m attains a specified maximum number of iterations m_{max}. Otherwise, increment m and go to step 2.

The sparse decomposition for the next components can be proceeded as in [8] with a remark that the number of principle component should not over than *rank(A)*.

3. NUMERICAL EXPERIMENTS

Since the benefit of NSPCA has been presented in [8], we only compare the effectiveness of the proposed approach with the NSPCA in [8].

3.1 Artificial Data

To compare the result with that of existing algorithms, we consider the simulation example proposed by [29]. In this example, three hidden factors are first created

$$V_1 \sim N(0{,}290),\ V_2 \sim N(0{,}300),\ V_3 = 0{,}3V_1 + 0{,}925V_2 + \varepsilon,\ \varepsilon \sim N(0{,}1)$$

where V_1, V_2 and ε are independent. Then 10 observed variables are generated as the follows

$$\begin{array}{lll} X_i = V_1 + \varepsilon_i^1, & \varepsilon_i^1 \sim N(0, 1) & i = 1, 2, 3, 4, \\ X_i = V_2 + \varepsilon_i^2, & \varepsilon_i^2 \sim N(0, 1) & i = 5, 6, 7, 8, \\ X_i = V_3 + \varepsilon_i^3, & \varepsilon_i^1 \sim N(0, 1) & i = 9, 10, \\ \varepsilon_i^j \text{ are independent}, & j = 1,2,3, & i = 1,\ldots, 10. \end{array}$$

To avoid the simulation randomness, the exact covariance matrix which is an infinity amount of data generated from the above model is used to compute principal components using the different approaches. The variance of the three underlying factors is nearly the same (290, 300 and 283.8, respectively). Since the first two are associated with four variables while the last one is associated with only two variables, V_1 and V_2 are almost equally important, and they are both significantly more important than V_3. In [29], the first two principal components explain 99.6% of the total variance. Hence, we shall choose the sparsity constraint is $k = 4$ when using only the variables X_1, X_2, X_3 and X_4 to recover the factor V_1, and only X_5, X_6, X_7 and X_8 for the second sparse principal component to recover V_2. And the results in [8] show that the post processing of cardinality refinement is not necessary. Hence, we compare the results in the case $k = 5$ given in Table 1.

Table 1. *The first two principal components with k=5*

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	**Explained variance**
NSPCA in [8], PC_1	0	0	0	0	.47	.47	.47	.47	0	.38	49.7%
NSPCA in [8], PC_2	.5	.5	.5	.5	0	0	0	0	0	0	39.5%
New method, PC_1	0	0	0	0	.48	.48	.48	.48	0	.28	50.1%
New method, PC_2	.49	.49	.49	.49	0	0	0	0	0.19	0	40.3%

With $\varepsilon = 10^{-2}$, the results in Table 1 show that the principal components of the new method explain more variance than NSPCA. Moreover, the first principal component of NSPCA is obtained after 27 sparsity constraint refinement iterations, but one of the new method is after 9 iterations.

3.2 Pit Props Data

The pit props data (consisting of 180 observations and 13 measured variables) was introduced in [14] and is another benchmark example used to test SPCA. In [8], the best results is obtained with sparsity constraint 5, 2, 3, 1, 1, and 1 as in Table 2.

Table 2. *The first six principal components of NSPCA with sparsity constraint 5, 2, 3, 1, 1, and 1*

Variable	**PC_1**	**PC_3**	**PC_3**	**PC_4**	**PC_5**	**PC_6**
topdiam	0.48	0	0	0	0	0
length	0.50	0	0	0	0	0
moist	0	0.71	0	0	0	0
testsg	0	0.71	0	0	0	0
ovensg	0	0	0.47	0	0	0
ringtop	0	0	0.71	0	0	0
ringbut	0.40	0	0.52	0	0	0
bowmax	0	0	0	1.00	0	0
bowdist	0.42	0	0	0	0	0
whorls	0.43	0	0	0	0	0
clear	0	0	0	0	1.00	0
knots	0	0	0	0	0	1.00
diaknot	0	0	0	0	0	0
Number of refinement iteration	64	0	912	0	0	0
Explained variance	26.20	14.48	14.19	7.69	7.69	7.69

Table 3. *The first six principal components of the new method with sparsity constraint 5, 2, 3, 1, 1, and 1*

Variable	PC_1	PC_3	PC_3	PC_4	PC_5	PC_6
topdiam	0.57	0	0	0	0	0
length	0.58	0	0	0	0	0
moist	0	0.71	0	0	0	0
testsg	0	0.71	0	0	0	0
ovensg	0	0	0.43	0	0	0
ringtop	0	0	0.71	0	0	0
ringbut	0.26	0	0.57	0	0	0
bowmax	0	0	0	1.00	0	0
bowdist	0.37	0	0	0	0	0
whorls	0.37	0	0	0	0	0
clear	0	0	0	0	1.00	0
knots	0	0	0	0	0	1.00
diaknot	0	0	0	0	0	0
Number of refinement iteration	13	0	18	0	0	0
Explained variance	26.43	14.48	14.37	7.69	7.69	7.69

With the same effectiveness as in the artificial data, the results in Table 3 illustrates the improvement of the proposed re-weighted l_1 minimization technique.

CONCLUSIONS AND PERSPECTIVES

In this paper, we attempted to improve the Non-negative Sparse PCA method (NSPCA) in [8] to find the non-negative principal components not only explaining most of the variance present in the data but also satisfying sparsity constraints through the solving of semi-definite problems. However, the re-weighted l_1 minimization technique try to refine the sparsity of the principal component as most as possible, but do not guarantee that the principal component satisfies the desired sparsity constraint. To achieve this, it is necessary to sacrifice some of the explained variance, albeit hopefully not too much. By dropping the entries which have smallest absolute values, we obtain a truncated principal component which gives a lower bound for the explained variance. Exploiting this observation, we may get a heuristic bisection algorithm to find the optimal principle component satisfying the required sparsity with an given error in the explained variance as in [9]. An other improvement can be done by using weights directly in the constraint of problem (2) as in [10].

REFERENCES

1. Alizadeh, F.: Interior point methods in semi-definite programming with applications to combinatorial optimization. SIAM J. Optim. 5 (1995) 13-51
2. Badea, L., Tilivea, D.: Sparse factorizations of gene expression guided by binding data. In Pacific Symposium on Biocomputing (2005)
3. Boyd, S., Vandenberghe,L.: Convex optimization. Cambridge University Press. Cambridge UK (2004)
4. Cadima,J., Jolliffe, I.T.: Loadings and correlations in the interpretation of principal components. J. Appl. Statist. 22 (1995) 203-214
5. Candes, E.J., Wakin, M.B., Boyd, S.: Enhancing sparsity by re-weighted l_1 minimization. (preprint)
6. D'Aspremont, A., El Ghaoui, L., Jordan, M.I., Lanckriet, G.R.G.: A direct formulation for sparse PCA using semi-definite programming. SIAM Rev. 49 (2007) 434-448
7. Duong, V.: Dynamic models for airborne air traffic management capability: State-of-the-art analysis. (Internal report) Bretigny, France: Eurocontrol Experimental Centre (1996)
8. Duong, D.X.T, Duong, V.: Non-Negative Sparse Principal Component Analysis for Multidimensional Constrained Optimization, The Tenth Pacific Rim International Conference on Artificial Intelligence (PRICAI 08), LNAI 5351, Springer-Verlag Berlin Heidelberg, (2008) 103-114.
9. Duong, D.X.T, Duong, V.: Exact and Direct Approach for Sparse Principal Component Analysis, Vietnam Journal of Science and Technology, 46 (2008) 7-22
10. Duong, D.X.T, Duong, V.: Principal Component Analysis with Weighted Sparsity Constraint, (2009) (to appear in Applied Mathematics & Information Sciences)
11. Fazel, M., Hindi, H., Boyd, S.: A rank minimization heuristic with application to minimum order system approximation. Proceedings of the American Control Conference, Arlington, VA. Vol. 6 (2001) 4734-4739
12. Hotelling, H.: Analysis of a complex of statistical variables into principal components. J. Educ. Psychol. 24 (1933) 417-441
13. Jagannathan, R., Ma, T.: Risk reduction in large portfolios: Why imposing the wrong constraints helps. Journal of Finance 58 (2003) 1651-1684
14. Jeffers, J.: Two case studies in the application of principal components. Appl. Statist. 16 (1967) 225-236
15. Jolliffe, I.T.: Rotation of principal components: Choice of normalization constraints. J. Appl. Statist. 22 (1995) 29-35.
16. Jolliffe, I.T., Trendafilov, N.T., Uddin, M.: A modified principal component technique based on the LASSO. J.Comput. Graphical Statist. 12 (2003) 531-547
17. Jolliffe, I.T.: Principal component analysis. Springer Verlag, New York (2002)
18. Lemarechal, C., Oustry, F.: Semi-definite relaxations and lagrangian duality with application to combinatorial optimization. Rapport de recherche 3710, INRIA, France (1999)

19. Lovasz, L., Schrijver, A.: Cones of matrices and set-functions and 0-1 optimization. SIAM J. Optim. 1 (1991) 166-190
20. Moghaddam, B., Weiss, Y., Avidan, S.: Spectral Bounds for Sparse PCA: Exact & Greedy Algorithms. Advances in Neural Information Processing Systems 18 (2006) 915-922. MIT Press.
21. Nesterov, Y.: Smoothing technique and its application in semi-definite optimization. Math. Program. 110 (2007) 245-259
22. Pearson, K.: On lines and planes of closest fit to systems of points in space. Phil. Mag. 2 (1901), 559-572
23. Sturm, J.: Using SEDUMI 1.0x, a MATLAB toolbox for optimization over symmetric cones. Optim. Methods Softw. 11 (1999) 625-653
24. Toh, K.C., Todd, M.J., Tutuncu, R.H.: SDPT3 - a MA TLAB software package for semi-definite programming. Optim. Methods Softw. 11 (1999) 545-581
25. Vines, S.: Simple principal components. Appl. Statist. 49 (2000) 441-451
26. Zass, R., Shashua, A.: Non-negative Sparse PCA. Advances In Neural Information Processing Systems 19 (2007) 1561-1568
27. Zhang, Z., Zha, H., Simon, H.: Low-rank approximations with sparse factors I: Basic algorithms and error analysis. SIAM J. Matrix Anal. Appl. 23 (2002) 706-727
28. Zhang, Z., Zha, H., Simon, H.: Low-rank approximations with sparse factors II: Penalized methods with discrete Newton-like iterations. SIAM J. Matrix Anal. Appl. 25 (2004) 901-920
29. Zou, H., Hastie, T., Tibshirani, R.: Sparse principal component analysis. J. Comput. Graphical Statist. 15 (2006) 265-286