OPTIMAL SPARSE L1-NORM PRINCIPAL-COMPONENT ANALYSIS*

Shubham Chamadia and Dimitris A. Pados

Dept. of Electrical Engineering, The State Univ. of New York at Buffalo, Buffalo, NY 14260

E-mail: {shubhamc, pados}@buffalo.edu

ABSTRACT

We present an algorithm that computes exactly (optimally) the S-sparse ($1 \leq S < D$) maximum-L_1-norm-projection principal component of a real-valued data matrix $\mathbf{X} \in \mathbb{R}^{D \times N}$ that contains N samples of dimension D. For fixed sample support N, the optimal L_1-sparse algorithm has linear complexity in data dimension, $\mathcal{O}(D)$. For fixed dimension D (thus, fixed sparsity S), the optimal L_1-sparse algorithm has polynomial complexity in sample support, $\mathcal{O}(N^S)$. Numerical studies included in this paper illustrate the theoretical developments and demonstrate the remarkable robustness to faulty data/measurements of the calculated sparse-L_1 principal components.

Index Terms— Faulty measurements, feature extraction, L_1-norm, machine learning, outlier resistance, principal component analysis, robust data processing, sparsity.

1. INTRODUCTION

Principal-component analysis (PCA) has long been a workhorse in the fields of machine learning and data signal processing. Conventional PCs describe the directions/subspaces over which the maximum variance of the data is captured [1] and are easily evaluated by standard (L_2-norm based) singular-value decomposition of the data matrix or, equivalently, eigen-value decomposition of the data autocorrelation matrix.

Nevertheless, in several applications not all data coordinates/dimensions are equally important. We may prefer, instead, to extract meaningful physical interpretation from few -but undetermined yet- coordinates [2]. Coordinate-based preference of data processing motivates the introduction of the concept of sparsity over the designed principal components. Recently, sparse PCA (SPCA) became a topic of very active research [2]-[15].

To enforce sparsity, an L_0-norm constraint is introduced in the modeling of principal components [5]. Regretfully, due to this additional constraint, SPCA becomes an $\mathcal{NP}$-hard problem [6]. As a consequence, a plethora of approximate SPCA solutions have appeared in the literature. Arguably, simplest among them is thresholding where the coordinates having absolute value smaller than a certain threshold are forced to zero [7]. LASSO-based solutions [8], semidefinite programming relaxation [9], rank-d approximation [10], power methods (i.e. Gpower [11], Tpower [12]), and expectation maximization [13] offer a variety of attractive approximate solutions. Under certain data matrix conditions, [14], [15] produce exact, optimal sparse solutions.

Research in the literature so far involves maximizing the variance (L_2-norm based processing) of the data along the principal direction under the given sparsity constraint. It is widely known, however, that conventional L_2-based PCA is sensitive to the presence of outliers (faulty measurements) or heavy-tailed noise in the data matrix[1] [16]-[21]. Lack of robustness and suboptimality of the available solvers are factors limiting the use and effectiveness of sparse L_2-PCA in practice [22].

In this paper, for the first time in the literature, we present an algorithm for the optimal computation of the sparse L_1-norm principal component of any data matrix $\mathbf{X} \in \mathbb{R}^{D \times N}$. Specifically, for fixed sample size N, the algorithm has linear computational complexity in the data dimension, $\mathcal{O}(D)$. Under fixed dimension D (implying fixed sparsity $S<D$), the algorithm has polynomial complexity in sample support, $\mathcal{O}(N^S)$. In the following section, we present the algorithm in complete detail for direct implementation.

2. OPTIMAL SPARSE L_1-PRINCIPAL COMPONENT COMPUTATION

The computation of the $S(<D)$-sparse maximum L_1-norm-projection principal component of a data matrix $\mathbf{X} \in \mathbb{R}^{D \times N}$ can be mathematically formulated as

$$\mathbf{q}_{L_1}^S = \underset{\substack{\mathbf{q} \in \mathbb{R}^D,\ \|\mathbf{q}\|=1, \\ \|\mathbf{q}\|_0 = S}}{\arg\max} \ \|\mathbf{X}^T \mathbf{q}\|_1. \tag{1}$$

2.1. Existing sub-optimal approach

An approximate solver of (1) has appeared in the literature [22] that relaxes the non-convex L_0-constraint in (1) to a convex L_1-constraint, i.e. $\|\mathbf{q}\|_0 \rightarrow \|\mathbf{q}\|_1$ (a practice often referred to as convexification [23]). This transforms the original sparse problem in (1) to the relaxed version

*The work was supported in part by the National Science Foundation under Grant ECSS-1462341.

[1] The L_2-norm (squared value) calculators magnify the impact of erroneous entries compared to L_1-norm (absolute value).

$$\tilde{\mathbf{q}}_{L_1}^{S} = \underset{\substack{\mathbf{q}\in\mathbb{R}^D,\ \|\mathbf{q}\|=1,\\ \|\mathbf{q}\|_1=S}}{\arg\max} \|\mathbf{X}^T\mathbf{q}\|_1. \tag{2}$$

An approximate iterative solution to (2) is then pursued by initializing the solver arbitrarily to a unit vector $\tilde{\mathbf{q}}_{L_1}^{S^{(0)}} \in \mathbb{R}^D$ and continuing by

$$\begin{aligned}\mathbf{b}^{(i+1)} &= \operatorname{sgn}\left(\mathbf{X}^T\tilde{\mathbf{q}}_{L_1}^{S^{(i)}}\right),\\ \tilde{\mathbf{q}}_{L_1}^{S^{(i+1)}} &= \frac{\Delta\left(\mathbf{X}\mathbf{b}^{(i+1)},\ S\right)}{\left\|\Delta\left(\mathbf{X}\mathbf{b}^{(i+1)},\ S\right)\right\|},\ i=0,\ 1,\ 2,\ \cdots,\end{aligned} \tag{3}$$

until convergence [22]. Here, $\operatorname{sgn}(\cdot)$ represents conventional sign extraction and $\Delta(\cdot, S)$ is the function that preserves and returns only the S largest absolute-value entries of the input vector with absolute values reduced by the $(S{+}1)$-largest absolute value of the input vector. Certainly, the iterative greedy algorithm in [22] described by (2) and (3) does not guarantee an optimal L_1-sparse solution to (1) and frequently exhibits heavy performance loss in the optimization metric.

2.2. Proposed optimal sparse L_1-principal component

The core idea supporting our algorithm is to translate the involved L_0-norm and L_1-norm in the optimization problem (1) to an equivalent tractable function as discussed below.

We observe that at any instant, only the S nonzero active entries of $\mathbf{q}$ (and the corresponding S rows of $\mathbf{X}$) in (1) participate in the maximization problem. This shrinks the coordinate search space of $\mathbf{q}$ from D (data dimension) to S (sparsity) and modifies (1) to the combinatorial search problem

$$\max_{\substack{\mathbf{q}\in\mathbb{R}^D,\ \|\mathbf{q}\|=1,\\ \|\mathbf{q}\|_0=S}} \|\mathbf{X}^T\mathbf{q}\|_1 = \max_{\substack{\mathcal{I}\subseteq\mathcal{D},\\ |\mathcal{I}|=S}} \max_{\substack{\mathbf{q}\in\mathbb{R}^S,\\ \|\mathbf{q}\|=1}} \|\mathbf{X}_{\mathcal{I},:}^T\mathbf{q}\|_1 \tag{4}$$

where $\mathcal{D} \triangleq \{1, 2, \cdots, D\}$. Further, the L_1-norm maximization in (4) can be expressed as an equivalent binary maximization problem [19]

$$\begin{aligned}\max_{\substack{\mathcal{I}\subseteq\mathcal{D},\\ |\mathcal{I}|=S}} \max_{\substack{\mathbf{q}\in\mathbb{R}^S,\\ \|\mathbf{q}\|=1}} \|\mathbf{X}_{\mathcal{I},:}^T\mathbf{q}\|_1 &= \max_{\substack{\mathcal{I}\subseteq\mathcal{D},\\ |\mathcal{I}|=S}} \max_{\substack{\mathbf{q}\in\mathbb{R}^S,\\ \|\mathbf{q}\|=1}} \max_{\mathbf{b}\in\{\pm1\}^N} \mathbf{b}^T\mathbf{X}_{\mathcal{I},:}^T\mathbf{q}\\ &= \max_{\mathbf{b}\in\{\pm1\}^N} \max_{\substack{\mathcal{I}\subseteq\mathcal{D},\\ |\mathcal{I}|=S}} \max_{\substack{\mathbf{q}\in\mathbb{R}^S,\\ \|\mathbf{q}\|=1}} \mathbf{b}^T\mathbf{X}_{\mathcal{I},:}^T\mathbf{q}.\end{aligned} \tag{5}$$

By the Cauchy-Schwartz inequality, for any given binary vector $\mathbf{b}$ and support set $\mathcal{I}$ in (5), the optimal $\mathbf{q} \in \mathbb{R}^S$ is

$$\mathbf{q}(\mathbf{b}, \mathcal{I}) = \frac{\mathbf{X}_{\mathcal{I},:}\mathbf{b}}{\|\mathbf{X}_{\mathcal{I},:}\mathbf{b}\|}. \tag{6}$$

Substituting the optimal $\mathbf{q}(\mathbf{b}, \mathcal{I})$ in (5) enables us to conceive our original sparse L_1-norm maximization problem (1) as an equivalent binary quadratic combinatorial maximization problem

$$\max_{\substack{\mathbf{q}\in\mathbb{R}^D,\ \|\mathbf{q}\|=1,\\ \|\mathbf{q}\|_0=S}} \|\mathbf{X}^T\mathbf{q}\|_1 = \max_{\mathbf{b}\in\{\pm1\}^N} \max_{\substack{\mathcal{I}\subseteq\mathcal{D},\\ |\mathcal{I}|=S}} \|\mathbf{X}_{\mathcal{I},:}\mathbf{b}\|. \tag{7}$$

Depending now on the relative value of the sample support N and data dimension D, we split our analysis and algorithmic developments into two explicit cases.

Case 1 ($N < D$)
Proposition 1 For fixed sample support N and asymptotically large D, optimal calculation of the S-sparse L_1-principal component $\mathbf{q}_{L_1}^S$ of a data matrix $\mathbf{X} \in \mathbb{R}^{D\times N}$ has computational complexity linear in D, $\mathcal{O}(D)$. ■

Below we prove Proposition 1 and present an implementation algorithm. We begin by introducing the function $\Psi(\mathbf{v}, S)$ which outputs the index set of the $S(<D)$ largest squared-value (or absolute-value) entries of its input vector $\mathbf{v} \in \mathbb{R}^D$,

$$\Psi(\mathbf{v}, S) = \underset{\substack{\mathcal{I}\subseteq\mathcal{D},\\ |\mathcal{I}|=S}}{\arg\max} \sum_{i\in\mathcal{I}} (\mathbf{v}_i)^2 = \underset{\substack{\mathcal{I}\subseteq\mathcal{D},\\ |\mathcal{I}|=S}}{\arg\max} \sum_{i\in\mathcal{I}} |\mathbf{v}_i|. \tag{8}$$

For any given binary vector $\mathbf{b} \in \{\pm1\}^N$ in (7), the optimal index set $\mathcal{I}_{\text{opt}}(\mathbf{b})$ is simply

$$\mathcal{I}_{\text{opt}}(\mathbf{b}) = \Psi(\mathbf{X}\mathbf{b}, S) \tag{9}$$

produced with computational cost $\mathcal{O}(D)$. For fixed sample support N, we can search among all possible 2^N binary vectors with cost $\mathcal{O}(2^N)$ to obtain an optimal binary solution

$$\mathbf{b}_{\text{opt}} = \underset{\mathbf{b}\in\{\pm1\}^N}{\arg\max} \|\mathbf{X}_{\mathcal{I}_{\text{opt}}(\mathbf{b}),:}\mathbf{b}\|. \tag{10}$$

Finally, the S-nonzero active coordinates of the optimal S-sparse L_1-principal component $\mathbf{q}_{L_1}^S$ in (1) are given by $\mathcal{I}_{\text{opt}}(\mathbf{b}_{\text{opt}})$ and have value

$$\frac{\mathbf{X}_{\mathcal{I}_{\text{opt}}(\mathbf{b}_{\text{opt}}),:}\mathbf{b}_{\text{opt}}}{\|\mathbf{X}_{\mathcal{I}_{\text{opt}}(\mathbf{b}_{\text{opt}}),:}\mathbf{b}_{\text{opt}}\|} \in \mathbb{R}^S. \tag{11}$$

The overall computational cost of the algorithm is $\mathcal{O}(2^N D)$, which for fixed N and asymptotically large D establishes Proposition 1.

Case 2 ($D < N$)
Proposition 2 For fixed data dimension D and asymptotically large sample size N, computation of the optimal S-sparse L_1-principal component $\mathbf{q}_{L_1}^S$ of $\mathbf{X} \in \mathbb{R}^{D\times N}$ has polynomial-time complexity in N, $\mathcal{O}(N^S)$. ■

We begin the proof of Proposition 2 by exchanging (finite search set) the order of the maximizations in (7). There are $\binom{D}{S}$ possible index support sets $\mathcal{I}$ to consider with implementation cost $\mathcal{O}(D^S)$. For each one of them,

$$\mathbf{b}_{\text{opt}}(\mathcal{I}) = \underset{\mathbf{b}\in\{\pm1\}^N}{\arg\max} \|\mathbf{X}_{\mathcal{I},:}\mathbf{b}\|,\ \ \mathcal{I} = \mathcal{I}_1,\ \mathcal{I}_2, \cdots,\ \mathcal{I}_{\binom{D}{S}}. \tag{12}$$

Optimal S-Sparse L_1-Principal Component Algorithm

Input: $\mathbf{X}_{D\times N}$ data matrix, $S(<D)$ sparsity, $\mathbf{q}_{L_1}^S = \mathbf{0}_D$

1: **if** $N < D$
   $\forall\, \mathbf{b}^{(i)} \in \{\pm 1\}^{N\times 1}$, evaluate
   $[\text{val}^{(i)} \;\; \mathcal{I}^{(i)}] \leftarrow \Psi(\mathbf{X}\mathbf{b}^{(i)}, S)$
2: **else**
   $\forall\, \mathcal{I}^{(i)} \subseteq \mathcal{D}$ and $|\mathcal{I}| = S$, evaluate
   $\mathbf{b}^{(i)} = \arg\max_{\mathbf{b}\in\{\pm 1\}^N} \|\mathbf{X}_{\mathcal{I}^{(i)},:}\mathbf{b}\|$ by [19], $\text{val}^{(i)} = \|\mathbf{X}_{\mathcal{I}^{(i)},:}\mathbf{b}^{(i)}\|$
3: **end if**
4: opt $\leftarrow \arg\max_i \text{val}^{(i)}$

Output: $\mathbf{q}_{L_1}^S(\mathcal{I}^{(\text{opt})}) = \mathbf{X}_{\mathcal{I}^{(\text{opt})},:}\,\mathbf{b}^{(\text{opt})} / \|\mathbf{X}_{\mathcal{I}^{(\text{opt})},:}\,\mathbf{b}^{(\text{opt})}\|$

Function: $\Psi(\mathbf{v}, S)$

Input: $\mathbf{v} \in \mathbb{R}^D, S$

1: $\mathcal{I} = \arg\max_{\mathcal{I}\subseteq\mathcal{D},\, |\mathcal{I}|=S} \sum_{i\in\mathcal{I}} |\mathbf{v}_i|$, $\text{val} = \sum_{i\in\mathcal{I}} (\mathbf{v}_i)^2$

Output: $[\text{val} \;\; \mathcal{I}]$

Fig. 1. Computation of the optimal S-sparse L_1-principal component of data matrix $\mathbf{X} \in \mathbb{R}^{D\times N}$.

For any given $\mathcal{I}$, we can solve (12) with the algorithm[2] developed in [19], which computes the optimal binary vector $\mathbf{b}_{\text{opt}}$ in (12) with cost $\mathcal{O}(N^{rank(\mathbf{X}_{\mathcal{I},:})}) \leq \mathcal{O}(N^S)$. Once we extract the optimal pair $(\mathbf{b}_{\text{opt}},\, \mathcal{I}_{\text{opt}})$, similar to Case 1, we design the optimal S-sparse L_1-principal vector $\mathbf{q}_{L_1}^S$ by (11). The overall computational cost is $\mathcal{O}(D^S N^S)$. For fixed D (implying fixed S) and asymptotically large N, the algorithm is executed in polynomial time with respect to N, i.e. $\mathcal{O}(N^S)$, which establishes Proposition 2.

Special Case of Nonnegative Data: $\mathbf{X} \in \mathbb{R}_+^{D\times N}$
We conclude this section with a note on the special case of all-positive (negative) data, which is of significant engineering importance. When $\mathbf{X} \in \mathbb{R}_+^{D\times N}$, the optimal binary binary vector in (7) simplifies to $\mathbf{b}_{\text{opt}} = \mathbf{1}^N$. Thus, the optimal index set is $\mathcal{I}_{\text{opt}} = \Psi(\mathbf{X}\mathbf{1}^N, S)$, which is the indices of the S data rows with highest row-wise mean absolute-value. The computational complexity in this special case is only $\mathcal{O}(D)$.

The complete optimal algorithm is given in pseudo code in Fig. 1.

3. EXPERIMENTAL STUDIES

In this section, we carry out experimental studies to illustrate the developed optimal L_1-sparse PCA algorithm and compare its performance/robustness against the state of the art popular sparse schemes. It is rather interesting to note that due to the data record size of the presented experiments, it is computationally infeasible to evaluate the optimal sparse L_2-norm principal component [15]. Nonetheless, the optimal sparse L_1-norm principal component can be easily calculated by the developed algorithm.

[2]Due to lack of space, we refrain from discussing the algorithm of [19] in detail.

Experiment 1 - On-off signal detection

We consider a uniform linear array of 14 antennas, each recording 100 input observations. All input observations contain additive white Gaussian noise (AWGN) of zero-mean, unit-variance, $\mathcal{N}(0, 1)$. Thirty (30) out of the 100 observations contain also an active signal randomly drawn from a Gaussian distribution with mean $+4$ or -4 and variance 2, $\mathcal{N}(\pm 4, 2)$. We are interested in classifying signal presence/absence by means of principal-component analysis of the observation data $\mathbf{X} \in \mathbb{R}^{100\times 14}$.

To make the experiment more challenging and test algorithmic robustness to faulty measurements, we select any two (out of the 14) antennas and contaminate any 5 of their observations by adding additive white Gaussian disturbance of high variance, $\mathcal{N}(0, 15)$. We denote this corrupted version of the data set by $\mathbf{X}^{\text{CRPT}}$. In Fig. 2, we plot the normalized explained variance (NEV)

$$\text{NEV} \triangleq \|\mathbf{X}^T \mathbf{q}_{(\text{CRPT})}^S\|^2 / \|\mathbf{X}^T \mathbf{q}\|^2 \qquad (13)$$

where $\mathbf{q}_{(\text{CRPT})}^S$ is the S-sparse principal component evaluated over the available corrupted data set $\mathbf{X}^{\text{CRPT}}$ by four different methods: (a) The proposed optimal L_1-sparse PCA algorithm, (b) the RSPCA method [22], (c) The Tpower method [12], and (d) the EM method [13]; $\mathbf{q}$ is the standard (non-sparse) L_2-principal component of the *original clean data matrix* $\mathbf{X}$. All algorithms successfully saturate around cardinality 30, indicating the number of significant signals present in the data set. However, the proposed optimal L_1-sparse PC scheme is greatly superior in capturing the active signal subspace (highest explained variance) as compared to the other schemes.

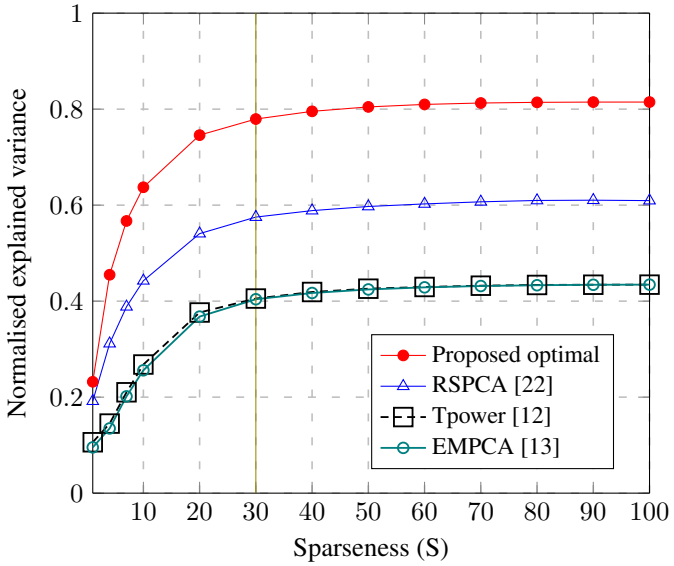

Fig. 2. Normalized explained variance versus principal-component sparseness.

(a) (b) (c)

(d) (e) (f) (g)

Fig. 3. (a) Original image; (b) noisy image; (c) "salt-and-pepper" disturbed image; (d) RSPCA [22], (e) Tpower [12], (f) EMPCA [13], (g) proposed optimal L_1-sparse restored image.

Experiment 2 - Sparse-L_1 image fusion

We consider 15 identical copies of the gray scale 256×256 *Lenna* image (Fig. 3(a)). Each copy is corrupted by zero-mean AWGN of variance σ^2=100 (Fig. 3(b)). Then, each of the noisy images is partitioned into sixteen square patches of dimension 64×64. Eight randomly chosen (out of sixteen) such patches are overwritten by "salt and pepper noise" as in Fig. 3(c).

At the processing stage, we possess the fifteen damaged (corrupted) image copies and assume no prior information regarding the corruption process. We pursue patch-wise restoration of the original image as follows. We divide each image into $P=\frac{256\times256}{32\times32}=64$ squared patches, each of dimension 32×32. We form the data matrix $\mathbf{X}_p$ that collects the vectorized p^{th}-patch from each of the corrupted images,

$$\mathbf{X}^p = [\mathbf{i}_1^p,\ \mathbf{i}_2^p, \ldots, \mathbf{i}_{15}^p]_{(1024\times15)}\,, \quad p = 1,\ 2, \cdots, P.$$

Next, we evaluate the optimal S-sparse L_1-principal component $\mathbf{q}^p$ of $\mathbf{X}_p$ and calculate the corresponding image representation/reliability factor r_n^p [24] as

$$r_n^p = \|\mathbf{i}_n^p - \mathbf{q}^p \mathbf{q}^{pT} \mathbf{i}_n^p\|^{-2}, \quad n = 1,\ 2, \cdots, 15, \qquad (14)$$

which captures the "closeness" of each individual corrupted image to the overall sparse-L_1-PC representation. Upon normalization of the reliability factors to $w_n^p = r_n^p / \sum_n r_n^p$, we restore each patch by

$$\hat{\mathbf{i}}^p = \sum_{n=1}^{15} w_n^p \mathbf{i}_n^p.$$

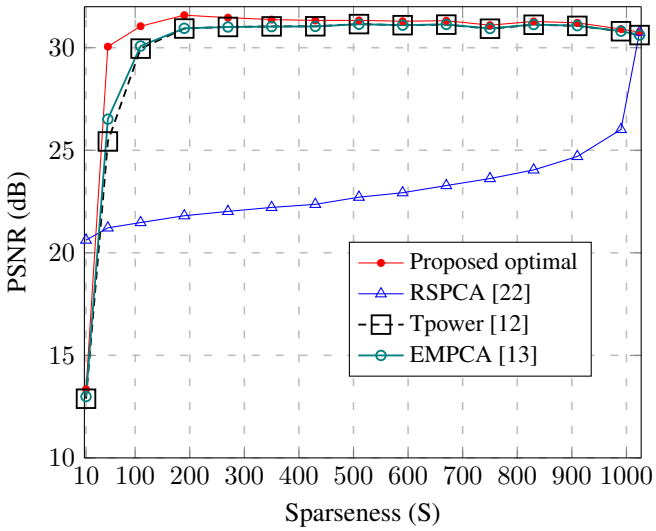

Fig. 4. PSNR of restored image versus enforced sparsity.

In Fig. 4, we plot the PSNR (in dB) of the restored image (compared to the clean original) with varying sparseness of the principal component. Alongside the proposed optimal sparse-L_1 PC calculator, we consider also RSPCA [22], Tpower [12], and EMPCA [13]. All algorithms (except RSCPA [22]) saturate at the sparseness of about 100. Fig. 3(d), (e), (f), (g) show the actual visual instance of the restored fused image by the four algorithms evaluated at sparsity S=100. The proposed optimal L_1-sparse scheme (Fig. 3(g)) offers the superior representation of the original image at most affordable computational cost[3].

[3] Since $\mathbf{X}^p \in \mathbb{R}_+$, p=1, 2, $\cdots$, P, the special case of Section 2.2 applies.

4. REFERENCES

[1] I. T. Jolliffe, Principal Component Analysis, 2nd ed., New York, NY: Springer Series in Statistics, 2002.

[2] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, pp. 6-18, Jan. 2006.

[3] Y. Zhang and L. E. Ghaoui, "Large-scale sparse principal component analysis with application to text data," in *Adv. Neural Inf. Process. Syst.*, vol. 24, pp. 532-539, Dec. 2011.

[4] S. Sharma and R. Gupta, "Improved BSP clustering algorithm for social network analysis," *Intern. Journal of Grid & Distributed Comp.*, vol. 3, pp. 67-76, Sept. 2010.

[5] J. J. Fuchs, "Recovery of exact sparse representations in the presence of bounded noise," *IEEE Trans. Inf. Theory*, vol. 51, pp. 3601-3608, Oct. 2005.

[6] B. Moghaddam, Y. Weiss, and S. Avidan, "Generalized spectral bounds for sparse LDA," in *Proc. Int. Conf. Mach. Learn.*, Pittsburgh, PA, June 2006, pp. 641-648.

[7] J. Cadima and I. Jolliffe, "Loading and correlations in the interpretation of principle compenents," *J. Appl. Stat.*, vol. 22, no. 2, pp. 203-214, 1995.

[8] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc.*, vol. 58, pp. 267-288, 1996.

[9] A. d' Aspremont, L. El Ghaoui, M. I. Jordan, and G. R. G. Lanckriet, "A direct formulation for sparse PCA using semidefinite programming," *SIAM Review*, vol. 49, no. 3, pp. 434-448, 2007.

[10] D. S. Papailiopoulos, A. G. Dimakis, and S. Korokythakis, "Sparse PCA through low-rank approximations," in *Proc. Int. Conf. Mach. Learn.*, Atlanta, GA, Jun. 2013, pp. 767-774.

[11] M. Journée, Y. Nesterov, P. Richtárik, and R. Sepulchre, "Generalized power method for sparse principal component analysis," *J. Machine Learning Research*, vol. 11, pp. 517-553, Feb. 2010.

[12] X.-T. Yuan and T. Zhang, "Truncated power method for sparse eigenvalue problems," *J. Machine Learning Research*, vol. 14, pp. 899-925, Apr. 2013.

[13] C. D. Sigg and J. M. Buhmann, "Expectation-maximization for sparse and non-negative PCA," in *Proc. Int. Conf. Mach. Learn.*, Helsinki, Finland, July 2008, pp. 960-967.

[14] A. d' Aspremont, F. Bach, and L. El Ghaoui, "Optimal solutions for sparse principal component analysis," *J. Machine Learning Research*, vol. 9, pp. 1269-1294, July 2008.

[15] M. Asteris, D. S. Papailiopoulos, and G. N. Karystinos, "Sparse principal component of a rank-deficient matrix," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Saint Petersburg, Russia, Aug. 2011, pp. 673-677.

[16] N. Kwak, "Principal component analysis based on L_1-norm maximization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, pp. 1672-1680, Sept. 2008.

[17] M. McCoy and J. A. Tropp, "Two proposals for robust PCA using semidefinite programming," *Electron. J. Stat.*, vol. 5, pp. 1123-1160, June 2011.

[18] S. Kundu, P. P. Markopoulos, and D. A. Pados, "Fast computation of the L_1-principal components of real-valued data," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, Italy, May 2014, pp. 8028-8032.

[19] P. P. Markopoulos, G. N. Karystinos, and D. A. Pados, "Optimal algorithms for L_1-subspace signal processing," *IEEE Trans. Signal Proc*, vol. 62, pp. 5046-5068, July 2014.

[20] Y. Liu and D. A. Pados, "Compressed-sensed-domain L1-PCA video surveillance," *IEEE Trans. Multimedia*, vol. 18, pp. 351-363, Mar. 2016.

[21] P. P. Markopoulos, S. Kundu, S. Chamadia, and D. A. Pados, "Efficient L_1-norm principal-component analysis via bit flipping," *IEEE Trans. Signal Proc.*, ArXiv e-print, Sep. 2016, [Online]. Available: https://arxiv.org/abs/1610.01959

[22] D. Meng, Q. Zhao, and Z. Xu, "Improve robustness of sparse PCA by L_1-norm maximization," *Pattern Recogn.*, vol. 45, pp. 487-497, Jan. 2012.

[23] J. Tropp, "Just relax: Convex programming methods for identifying sparse signals," *IEEE Trans. Inf. Theory*, vol. 51, pp. 1030-1051, Apr. 2006.

[24] P. P. Markopoulos, S. Kundu, and D. A. Pados, "L_1-fusion: Robust linear-time image recovery from few severely corrupted copies," in *Proc. IEEE Int. Conf. Image Process.*, Canada, Sep. 2015, pp. 1225-1229.