HAL
open science

L1-PCA signal subspace identification for non-sphered data under the ICA model

Rubén Martín-Clemente, Vicente Zarzoso

To cite this version:

Rubén Martín-Clemente, Vicente Zarzoso. L1-PCA signal subspace identification for non-sphered data under the ICA model. Proc. CAMSAP-2017, 7th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, Dec 2017, Curaçao, Netherlands Antilles. hal-01719134

HAL Id: hal-01719134

https://hal.science/hal-01719134v1

Submitted on 1 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

in: Proc. CAMSAP, 7th IEEE Intl. Workshop on Computational Advances in Multi-Sensor Adaptive Processing, Curaçao, Dutch Antilles, Dec. 10-13, 2017.

L1-PCA Signal Subspace Identification for Non-sphered Data Under the ICA Model

Rubén Martín-Clemente
Department of Signal Theory and Communications
University of Seville, Spain
ruben@us.es

Vicente Zarzoso
Université Nice Côte d'Azur, CNRS
I3S Laboratory, Sophia Antipolis, France
zarzoso@i3s.unice.fr

***Abstract*—Principal component analysis (PCA) is an ubiquitous data compression and feature extraction technique in signal processing and machine learning. As compared with the classical L2-norm PCA, its L1-norm version offers increased robustness to outliers that are usually present in faulty data. Recently, L1-PCA was shown to perform source recovery when the observed data follow an independent component analysis (ICA) model. However, proof of this result requires the data to be sphered, i.e., to be preprocessed to constrain their covariance matrix to be the identity. The present contribution extends this result by relaxing the sphering assumption and allowing the data to have arbitrary covariance matrix. We prove that L1-PCA is indeed able to identify the mixing matrix columns associated with the strongest independent sources, thus performing signal subspace identification with improved robustness to outliers. Numerical experiments illustrate and confirm the theoretical findings.**

I. Introduction

Principal Component Analysis (PCA) is probably the most popular multivariate data analysis technique [4], as it constitutes the basis for a variety of dimensionality reduction and noise filtering approaches. Multivariate data can be considered as vectors in a high-dimensional vector space, say $\mathbb{R}^p$. The primary goal of PCA, as defined originally by Pearson, is to find the direction of the line that best fits the data in the p-dimensional space. In the traditional approach, this goal is obtained by maximizing the sum of the squares of the data projections onto the sought direction. Actually, several best fitting directions, or principal axes, can be computed by solving repeatedly the above optimization problem under the constraint that the nth principal axis has to be orthogonal to the previous $(n-1)$ axes. The subspace $V \subset \mathbb{R}^p$ spanned by the orthogonal basis vectors pointing in the main principal directions is called the signal subspace, another fundamental concept in multivariate signal processing.

It is intuitive that the principal axes reveal much of the structure of the data. In the context of the processing of faulty data, PCA is mostly useful when the data of interest is concentrated around only a few principal axes, so that the dimension of the signal subspace V is much lower than p, whereas the noise is distributed isotropically through all $\mathbb{R}^p$. In this case, a significant reduction of the noise, while preserving the information content of the original data, can be obtained by projecting the faulty data onto the signal subspace.

Recently, L1-norm PCA has been proposed in an attempt to increase robustness to outliers of classical PCA. Experimentally, the L1-PCA approach has proven to be effective for restoration of faulty data in, e.g., the reconstruction of occluded images [6], [10], [11], as well as in pattern recognition and dimensionality reduction [1], [6]–[10], [13], [14], [16]–[18]. Although fast yet suboptimal algorithms for L1-PCA exist [5], [6], one of the most attractive features of this approach is that it can be also performed by optimal algorithms with guaranteed global convergence, as recently shown in [10].

The link between L1-PCA and independent component analysis (ICA), another popular multivariate data processing technique [3], [15], is carried out in [12], thus opening the possibility to perform ICA using globally optimal algorithms for L1-PCA. L1-PCA is shown to extract the independent sources under the ICA model in the case where the data are sphered, i.e., they have been whitened to present an identity covariance matrix. The present contribution extends this result by relaxing the assumption that the data be sphered, which defines a more general framework. A theoretical analysis supported by experimental results shows that L1-PCA is able to identify the dominant signal subspace under the ICA model, thus enabling source recovery via beamforming in a further processing stage.

II. Problem Formulation

Let us suppose that we observe N samples $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ of a p-dimensional zero-mean random vector $\mathbf{x}$, which can be arranged into the $(p \times N)$ matrix

$$\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N].$$

Let $\mathbf{w} \in \mathbb{R}^p$, $\|\mathbf{w}\|_2 = 1$, define a projection direction. As mentioned above, the goal of classical PCA is to find the random variable $y = \mathbf{w}^{\mathrm{T}}\mathbf{x}$ with the largest empirical variance. The first principal component vector solves the following problem:

$$\max_{\|\mathbf{w}\|_2=1} \sum_{n=1}^{N} (\mathbf{w}^{\mathrm{T}}\mathbf{x}_n)^2$$

or, in matrix form:

$$\max_{\|\mathbf{w}\|_2=1} \left\|\mathbf{w}^{\mathrm{T}}\mathbf{X}\right\|_2. \tag{1}$$

Though mathematically appealing, the L2-norm is rather sensitive to impulsive noise or outliers since squaring the

projections overemphasizes the effects of far-off data points, which is a serious drawback.

To increase the robustness of PCA when processing faulty data, we can try to eliminate the outliers from the data or, alternatively, replace the L2-norm cost function by a more robust criterion. The following L1-norm based variant of traditional PCA is proposed in [6]:

$$\max_{\|\mathbf{w}\|_2=1} \left\|\mathbf{w}^{\mathrm{T}}\mathbf{X}\right\|_1 \tag{2}$$

where $\|\mathbf{y}\|_1 = \sum_{n=1}^{N}|y_i|$ represents the sum of the absolute entries of vector $\mathbf{y} = \mathbf{w}^{\mathrm{T}}\mathbf{X}$. The above two PCA criteria are generally not equivalent, even when the data contain no outliers. A key problem is to determine under which conditions the solutions of (1) and (2) are the same, or at least similar, in the absence of outliers. When the data are sphered, i.e., the covariance matrix $\mathbf{R_x} = \frac{1}{N}\mathbf{X}\mathbf{X}^{\mathrm{T}}$ equals the identity matrix, it follows that $\left\|\mathbf{w}^{\mathrm{T}}\mathbf{X}\right\|_2 = \|\mathbf{w}\|_2$ and therefore no principal axis can be defined or identified according to (1). This is no surprising as sphering is often performed by classical PCA, so that the information that the L2 criterion can extract from sphered data is somewhat exhausted. By contrast, L1-PCA can yield interesting directions even for sphered data. A generative model that can provide useful insights is the case where each data point $\mathbf{x}_i$ is made up of the linear superposition

$$\mathbf{x}_i = \sum_{j=1}^{q} \mathbf{a}_j\, s_{ij} \tag{3}$$

where s_{ij} represents the degree to which $\mathbf{a}_j \in \mathbb{R}^p$ is present in observation $\mathbf{x}_i$ and q is the actual dimension of the data. The assumption that coefficients s_{ij} are mutually independent for different values of j is physically plausible in many real-world problems, and gives rise to the independent component analysis (ICA) model [3]. The solutions of L1-PCA under the ICA model when $p = q$ and the data are sphered are analyzed in depth in [12], showing that projections associated with independent components are locally stable stationary points of the L1-PCA criterion.

The remaining of this paper explores the extension of this result to the more general case of non-sphered data, where $\mathbf{R_x} \neq \mathbf{I}$, and proves that L1-PCA yields principal directions in a similar sense to L2-PCA but with the benefit of increased robustness to outliers.

III. Theoretical Analysis

We conduct a preliminary study for the case $q = 2$, implying that the data points can be described by only two independent components. This case is simple enough to allow a theoretical analysis, while still retaining the important features of the problem. Results are stated without proof for lack of space.

A. General Analysis for $q = 2$

We start by adopting a distributional viewpoint so that, according to model (3), the observed data vector $\mathbf{x}$ can be expressed as

$$\mathbf{x} = \mathbf{a}_1 s_1 + \mathbf{a}_2 s_2 = \mathbf{A}\mathbf{s} \tag{4}$$

with $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2]$ and $\mathbf{s} = [s_1, s_2]^{\mathrm{T}}$. Components s_1 and s_2 represent zero-mean, mutually statistically independent random variables, which are also assumed to have unit variance. If this is not the case, the power of the variables can always be incorporated into the length of the vectors $\mathbf{a}_i$ without loss of generality.

Given a projection unit-norm vector $\mathbf{w}$, the projected data are defined by $y = \mathbf{w}^{\mathrm{T}}\mathbf{x} = g_1\, s_1 + g_2\, s_2$, where $g_i = \mathbf{w}^{\mathrm{T}}\mathbf{a}_i$. Denoting by a_{ji} the jth entry of $\mathbf{a}_i$, we can also write

$$g_i = \sum_{j=1}^{p} w_j\, a_{ji}.$$

Observe that, assuming ergodicity conditions, the L1-norm criterion in (2) becomes (up to an irrelevant scale factor) proportional to $\mathrm{E}\{|y|\}$ for large enough sample size:

$$\|\mathbf{y}\|_1 \underset{N\to+\infty}{\longrightarrow} N\mathrm{E}\{|y|\}$$

where $\mathbf{y} = \mathbf{w}^{\mathrm{T}}\mathbf{X}$. For this reason, instead of (2), let us consider here the equivalent problem

$$\max_{\mathbf{w}} \mathrm{E}\{|y|\} \quad \text{subject to } \|\mathbf{w}\|_2 = 1$$

whose Lagrangian is given by

$$\mathcal{L}(\mathbf{w}, \lambda) = \frac{1}{2}\mathrm{E}\{|y|\} + \lambda(\|\mathbf{w}\|_2^2 - 1) \tag{5}$$

where λ is the Lagrange multiplier. Differentiating with respect to the components of $\mathbf{w}$ by the chain rule, and using some formulas in [12], it can be proven that

$$\frac{\partial \mathcal{L}}{\partial w_n} = \sum_{i=1}^{2} a_{ni}\mathcal{G}_i + 2\lambda w_n$$

where

$$\mathcal{G}_i = -\int_{-\infty}^{\infty} s f_i(s) F_j\left(-\frac{g_i}{g_j}s\right) \mathrm{d}s \tag{6}$$

with $i, j \in \{1, 2\}$, $i \neq j$,. Symbols $f_i(\cdot)$ and $F_j(\cdot)$ denote the probability density function and the distribution function, respectively, of s_j. This equation can be rewritten in matrix form as

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \mathbf{A}\mathbf{b} + 2\lambda\mathbf{w}$$

where $\mathbf{b} = [\mathcal{G}_1, \mathcal{G}_2]^{\mathrm{T}}$. Finally, the stationary points of the criterion must verify:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \mathbf{0} \Rightarrow \mathbf{A}\mathbf{b} = -2\lambda\mathbf{w}. \tag{7}$$

This equation is difficult to solve because $\mathbf{b}$ depends nonlinearly on $\mathbf{w}$ through $\mathbf{g}$ in (6). Hence, rather than searching for a general solution, we will consider a simple yet significant particular problem for now.

B. Particular Case: Data Clustered Around 1D Subspace

Now let us focus on the important situation where the data appear approximately clustered around a one-dimensional subspace, a straight line. Without loss of generality, we assume that this line is in the direction of $\mathbf{a}_1$, the case of $\mathbf{a}_2$ being treated in a totally analogous manner. A necessary condition for the observed data clustering is that $\|\mathbf{a}_1\|_2 \gg \|\mathbf{a}_2\|_2$ under the unit-variance source assumption. Vector $\mathbf{a}_1$ defines a principal axis in the sense of the dominant signal subspace, while the term $\mathbf{a}_2\, s_2$ is interpreted here as a small amount of added noise defining a noise subspace. We are going to study whether L1-PCA is able to identify the signal subspace. This is tantamount to verifying whether

$$\mathbf{w} \approx \pm \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|_2} \tag{8}$$

is a solution to (7). Since, by definition,

$$\mathbf{g} = \mathbf{w}^{\mathrm{T}}\mathbf{A} \approx \pm\Big[1, \mathbf{a}_1^{\mathrm{T}}\mathbf{a}_2/\|\mathbf{a}_1\|_2\Big]$$

we can assume that $\frac{g_2}{g_1} \simeq 0$. Then, the following Maclaurin-based approximation of (6) holds for $i = 2$:

$$\mathcal{G}_2 \approx -\int_{-\infty}^{\infty} s f_2(s) \left[F_1(0) - s\, f_1(0)\, \frac{g_2}{g_1}\right] \mathrm{d}s = -f_1(0)\, \frac{g_2}{g_1}$$

where we have exploited the assumption that s_2 has zero mean and unit variance. Clearly, $\frac{g_2}{g_1} \simeq 0$ implies $\mathcal{G}_2 \simeq 0$. Substituting $\mathcal{G}_2 = 0$ in (7), we readily arrive at (8): this proves that $\mathbf{a}_1$ is a stationary point of the L1-norm criterion. We know check its stability.

C. Local Stability Analysis

Let $\mathbf{L}_{\mathbf{w}}$ be the Hessian matrix associated to (5), $\mathbf{L}_{\mathbf{w}} = \mathbf{F}_{\mathbf{w}} + \lambda \mathbf{H}_{\mathbf{w}}$, where

$$\mathbf{H}_{\mathbf{w}} = \left[\frac{\partial^2}{\partial w_i \partial w_j} \sum_{n=1}^{p} w_n^2\right]_{i,j} \qquad i,j = 1,\ldots,p$$

λ is the Lagrange multiplier and $\mathbf{F}_{\mathbf{w}}$ is the Hessian of $\frac{1}{2}\mathrm{E}\{|y|\}$,

$$\mathbf{F}_{\mathbf{w}} \stackrel{\text{def}}{=} \left[\frac{\partial^2}{\partial w_i \partial w_j} \frac{1}{2}\mathrm{E}\{|y|\}\right]_{i,j} \qquad i,j = 1,\ldots,p.$$

Assuming the solution (8), it can be shown that

$$\mathbf{L}_{\mathbf{w}} = \frac{f_1(0)}{\|\mathbf{a}_1\|_2}\mathbf{a}_2\mathbf{a}_2^{\mathrm{T}} - \|\mathbf{a}_1\|_2 \left(\int_0^{\infty} s f_1(s)\, \mathrm{d}s\right)\mathbf{I}.$$

Let $\mathbf{v}$ be any vector orthogonal to $\mathbf{w}$. Geometry shows that

$$\mathbf{v}^{\mathrm{T}}\mathbf{a}_2 = \mp\|\mathbf{v}\|_2\|\mathbf{a}_2\|_2 \sin(\angle \mathbf{a}_1\mathbf{a}_2)$$

where $\angle \mathbf{a}_1\mathbf{a}_2$ is the angle between $\mathbf{a}_1$ and $\mathbf{a}_2$. Then,

$$\frac{\mathbf{v}^{\mathrm{T}}\mathbf{L}_{\mathbf{w}}\mathbf{v}}{\|\mathbf{v}\|_2^2} = \frac{\|\mathbf{a}_2\|_2^2}{\|\mathbf{a}_1\|_2}\sin^2(\angle \mathbf{a}_1\mathbf{a}_2) f_1(0) - \|\mathbf{a}_1\|_2 \int_0^{\infty} s f_1(s)\, \mathrm{d}s.$$

When $\|\mathbf{a}_1\|_2$ is greater enough than $\|\mathbf{a}_2\|_2$, it follows that $\mathbf{v}^{\mathrm{T}}\mathbf{L}_{\mathbf{w}}\mathbf{v} < 0$. According to [2], *this ensures that* (8) *maximizes the criterion*. Otherwise, if $f_1(0) \gg \int_0^{\infty} s f_1(s)\, \mathrm{d}s$, as in a supergaussian distribution [12], then $\mathbf{v}^{\mathrm{T}}\mathbf{L}_{\mathbf{w}}\mathbf{v}$ may be positive: in this case, we have to minimize, not maximize, the L1-norm to recover $\mathbf{a}_1$.

According to this theoretical analysis, L1-PCA determines the signal subspace in a way similar to L2-PCA. This result, combined with the fact that L1-PCA is more robust against outliers than L2-PCA, provides additional support for the use of L1-PCA in the analysis of faulty data.

IV. Numerical Experiments

Some computer experiments are carried out to support the validity and generality of the theoretical development of the previous section. For simplicity we will assume that $p = q$, i.e., the observations have the same dimensionality as the underlying independent components generating the data.

A. Case $q = 2$

Let $\mathbf{a}_1 = \frac{\alpha}{\sqrt{2}}[1,1]^{\mathrm{T}}$ and $\mathbf{a}_2 = \frac{1}{2}[\sqrt{3},-1]^{\mathrm{T}}$. These vectors make an angle of $45°$ and $-30°$, respectively, with the horizontal axis so that $\mathbf{R}_{\mathbf{x}} = \mathbf{A}\mathbf{A}^{\mathrm{T}} \neq \mathbf{I}$. Assuming that $\mathbf{w} \approx \pm\frac{\mathbf{a}_1}{\|\mathbf{a}_1\|_2}$ as in (8), it follows that $\|\mathbf{a}_1\|_2 \gg \|\mathbf{a}_2\|_2$, as required, when $\alpha \to \infty$. We generate $N = 200$ samples of the random variable $\mathbf{x}$ using model (4), where s_1 and s_2 are drawn from independent zero-mean unit-variance uniform distributions. L1-PCA is then applied on the data using the iterative algorithm presented in [6]. First, Figure 1 (top) shows vector $\mathbf{a}_1$ for $\alpha = 1.5$ (solid-line arrow), the vector $\mathbf{w}_{L1}$ obtained by L1-PCA (bold solid line), the first principal component $\mathbf{w}_{L2}$ generated by traditional L2-PCA (dashed line) and the scatter plot of the observations of $\mathbf{x}$. In the figure, vectors $\mathbf{a}_1, \mathbf{w}_{L1}$ and $\mathbf{w}_{L2}$ have been shifted and scaled differently for clarity. In this particular experiment, observe that L1-PCA estimates $\mathbf{a}_1$, the vector pointing in the main direction of the data, better than L2-PCA: the angle between $\mathbf{w}_{L1}$ and $\mathbf{a}_1$ equals $3.2°$, whereas the angle between $\mathbf{w}_{L2}$ and $\mathbf{a}_1$ is $6.3°$. Next, we repeat the experiment after adding several outliers to the same data. Results are shown in Figure 1 (bottom). As expected, the estimation is negatively affected by the presence of outliers: the angle between $\mathbf{w}_{L1}$ and $\mathbf{a}_1$ now becomes $3.6°$. Yet L1-PCA proves more robust than L2-PCA, since the angle between $\mathbf{w}_{L2}$ and $\mathbf{a}_1$ is now equal to $12.5°$.

Figure 2 (left) shows the angle of vector $\mathbf{w}_{L1}$ determined by L1-PCA with the horizontal axis as a function of α after convergence of the L1-PCA algorithm. As α increases, this angle tends to $45°$ so that $\mathbf{w}_{L1}$ is parallel to $\mathbf{a}_1$ as expected. On the other hand, if α tends to zero, the angle approaches $-30°$ and $\mathbf{w}_{L1}$ coincides with $\mathbf{a}_2$, which is also consistent with our analysis as $\mathbf{a}_2$ now becomes dominant. Finally, when $\alpha = 1$, so that $\mathbf{a}_1$ and $\mathbf{a}_2$ have the same length, $\mathbf{w}_{L1}$ points in the direction of the sum of both vectors. In Figure 2 (right), as in a previous experiment, data contain 5% of Gaussian outliers and therefore the figure also compares the robustness of L1-PCA and L2-PCA.

B. Case $q > 2$

We now generate $N = 500$ independent samples of $\mathbf{x} = \sum_{j=1}^{7} \mathbf{a}_j\, s_j$, $\mathbf{x} \in \mathbb{R}^7$. The coordinates of vectors $\mathbf{a}_j$ are

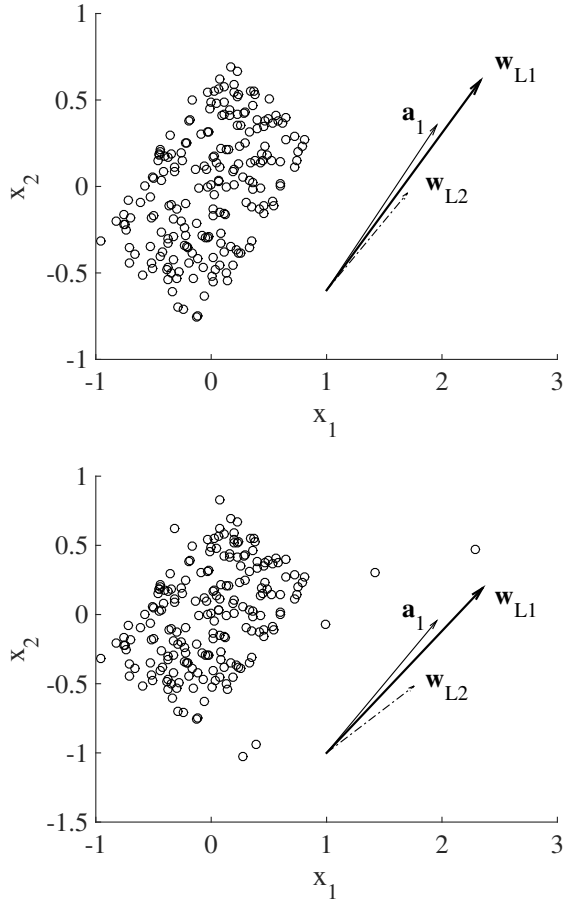

Fig. 1. Scatter plot of a two-dimensional observation ($p = 2$) composed of two independent components ($q = 2$) and $N = 200$ samples. The independent component direction $\mathbf{a}_1$ is represented by a solid line arrow. The L1-PCA solution $\mathbf{w}_{L1}$ (bold solid-line arrow) is aligned with the main axis of the data, as predicted by our analysis. For the purposes of comparison, the L2-PCA solution $\mathbf{w}_{L2}$ is also shown (dashed-line arrow). Vectors are not drawn to scale. Top plot: No outliers. Bottom plot: 5% of outliers. Outliers are drawn at random from a normalized Gaussian distribution. It is apparent in this figure that L1-PCA is more robust to outliers than L2-PCA.

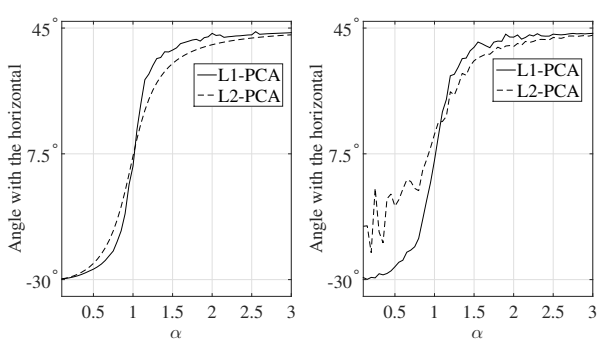

Fig. 2. Angle of L1-PCA solution $\mathbf{w}_{L1}$ with the horizontal axis as a function of the dominant axis magnitude α. As α increases, the angle tends to $45°$ and $\mathbf{w}_{L1}$ points in the direction of $\mathbf{a}_1$. As α tends to zero, $\mathbf{a}_2$ becomes dominant and $\mathbf{w}_{L1}$ points it its direction, at an angle of $-30°$. For comparison, the angle between the L2-PCA solution and the horizontal axis is also shown in the dashed line. The curves represent the average of 100 independent experiments. Left plot: no outliers. Right plot: data contain 5% of outliers.

drawn from a standardized Gaussian distribution, except for $\mathbf{a}_4$, where the variance is set to 10, so that $\mathbf{a}_4$ defines the main axis of the data. The variables s_j are zero-mean and uniformly distributed. Figure 3 shows the angle of vector $\mathbf{w}_{L1}$ recovered by L1-PCA with each of the basis vectors $\mathbf{a}_i$, simply computed as $\arccos\left(\frac{|\mathbf{w}^{\mathrm{T}}\mathbf{a}_i|}{\|\mathbf{w}\|\|\mathbf{a}_i\|}\right)$, averaged over 100 independent Monte Carlo realizations. A confidence interval is also depicted: the total length of the vertical line equals twice the standard deviation of the angle. As expected, $\mathbf{w}_{L1}$ is more aligned with $\mathbf{a}_4$ than with any other vector $\mathbf{a}_i$, $i \neq 4$. This result demonstrates that the theoretical analysis of Sec. III can be generalized to the case of data with more than two dimensions ($q > 2$). For comparison, we also show the results obtained by L2-PCA.

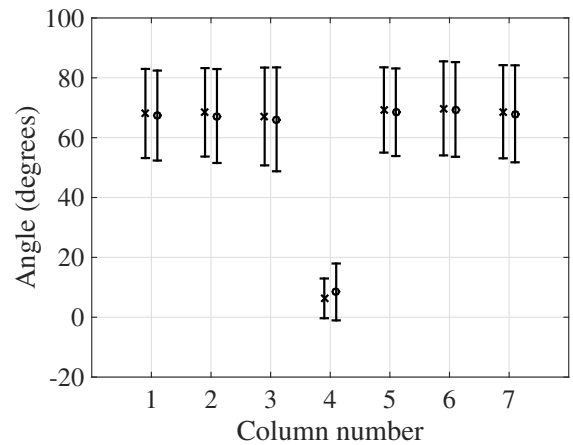

Fig. 3. Results with $q = 7$ random directions, where $\mathbf{a}_4$ has the strongest magnitude, and $N = 500$ samples. Each pair of vertical lines represents the confidence intervals of the angles formed by L1-PCA vector $\mathbf{w}_{L1}$ (left line) and the L2-PCA vector $\mathbf{w}_{L2}$ (right line) with each data direction over 100 independent Monte Carlo realizations. L1-PCA succeeds in finding the principal direction naturally defined by the dominant axis $\mathbf{a}_4$. The results are also similar to those obtained by applying L2-PCA.

V. Conclusions

This work has found a relationship between L1-PCA and the principal axes defining the dominant signal subspace when the data are not sphered. This result is consistent with the fact that the criterion seeks for projections with large amplitude. The dominant source subspace identified by L1-PCA proves more robust in the presence of outliers than classical L2-PCA. Further work should aim at the theoretical characterization of the influence of outliers and the extension of our analysis to more than two underlying dimensions.

Acknowledgments

This work is partially funded by the Spanish Ministry of Economy and Competitiveness under project TEC2014-53103-P CMANS.

V. Zarzoso is a member of the *Institut Universitaire de France*.

REFERENCES

[1] J.P. Brooks, J.H. Dulá, and E.L. Bone, "A pure L_1-norm principal component analysis", *Journal Computational Stat. Data Analysis*, vol. 61, pp. 83–98, May 2013.

[2] E.K.P. Chong, and S. H. Zak, "An Introduction to Optimization", *John Wiley & Sons, Inc.*, 1996.

[3] P. Comon, and C. Jutten, eds. "Handbook of Blind Source Separation: Independent component analysis and applications", *Academic press*, 2010.

[4] I.T. Jolliffe, "Principal Component Analysis", *Springer*, 2nd ed., 2002.

[5] S. Kundu, P. P. Markopoulos, and D. A. Pados, "Fast computation of the L1-principal component of real-valued data", in *Proceedings of 39th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014)*, Florence, Italy, May 2014.

[6] N. Kwak, "Principal component analysis based on L1-norm maximization", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 1672–1680, Sept. 2008.

[7] N. Kwak and J. Oh, "Feature extraction for one-class classification problems: enhancements to biased discriminant analysis", *Pattern Recognition*, vol. 42, pp. 17–26, Jan. 2009.

[8] X. Li, Y. Pang, and Y. Yuan, "L1-norm based 2DPCA", *IEEE Trans. on Systems, Man and Cybernetics*, vol. 40, pp. 1170–1175, August 2009.

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

[10] P. P. Markopoulos, G. N. Karystinos, and D. A. Pados, "Optimal algorithms for L1-subspace signal processing", *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 5046–5058, Oct. 2014.

[11] P. P. Markopoulos, S. Kundu, S. Chamadia, and Dimitris A. Pados, "Efficient L1-Norm Principal-Component Analysis via Bit Flipping", *IEEE Transactions on Signal Processing*, vol. 65, no. 16, pp. 4252–4264, Aug. 2017.

[12] R. Martín-Clemente, V. Zarzoso, "On the link between L1-PCA and ICA", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 3, pp. 515–527, March 2017.

[13] D. Meng, Q. Zhao, and Z. Xu, "Improve robustness of sparse PCA by L1 -norm maximization", *Pattern Recognition*, vol. 45, pp. 487–497, Jan. 2012.

[14] Y. Pang, X. Li, and Y. Yuan, "Robust tensor analysis with L1-norm", *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 20, pp. 172–178, Feb. 2010.

[15] S. Roberts, and R. Everson, eds. "Independent Component Analysis: Principles and Practice", *Cambridge University Press*, 2001.

[16] L. Yu, M. Zhang, and C. Ding, "An efficient algorithm for L1-norm principal component analysis", *Proc. IEEE ICASSP 2012*, Kyoto, Japan, pp. 1377–1380, Mar. 2012.

[17] H. Wang, Q. Tang, and W. Zheng, "L1-norm-based common spatial patterns", *IEEE Trans. on Biomedical Engineering*, vol. 59, pp. 653–662, Mar. 2012.

[18] H. Wang, "Block principal component analysis with L1-norm for image analysis", *Pattern Recognition Letters*, vol. 33, pp. 537-542, Apr. 2012.