

# L1-Norm Principal-Component Analysis of Complex Data

Nicholas Tsagkarakis , Panos P. Markopoulos , *Member, IEEE*, George Sklivanitis , *Member, IEEE*, and Dimitris A. Pados , *Senior Member, IEEE*

**Abstract**—L1-norm Principal-Component Analysis (L1-PCA) of real-valued data has attracted significant research interest over the past decade. L1-PCA of complex-valued data remains to date unexplored despite the many possible applications (in communication systems, for example). In this paper, we establish theoretical and algorithmic foundations of L1-PCA of complex-valued data matrices. Specifically, we first show that, in contrast to the real-valued case for which an optimal polynomial-cost algorithm was recently reported by Markopoulos, Karystinos, and Pados, complex L1-PCA is formally NP-hard. Then, casting complex L1-PCA as a unimodular optimization problem, we present the first two suboptimal algorithms in the literature for its solution. Extensive experimental studies included in this paper illustrate the sturdy resistance of complex L1-PCA against faulty measurements/outliers in the processed data.

**Index Terms**—Data analytics, dimensionality reduction, erroneous data, faulty measurements, L1-norm, machine learning, principal-component analysis, outlier resistance.

## I. INTRODUCTION

FOR more than a century, Principal-Component Analysis (PCA) has been a core operation in data/signal processing [2], [3]. Conceptually, PCA can be viewed as the pursuit of a coordinate system (defined by the principal components) that reveals underlying linear trends of a data matrix. In its conventional form, the new coordinate system is calculated such that it preserves the energy content of the data matrix

to the maximum possible extent. Conventionally, the energy content of a data point is expressed by means of its L2-norm, i.e., the Euclidean distance of the data point from the center of the coordinate system. Thus, for any complex data matrix  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbb{C}^{D \times N}$ , PCA searches for the size- $K$  ( $1 \leq K < r = \text{rank}(\mathbf{X})$ ) orthonormal basis (or  $K$ -dimensional coordinate system) that solves

$$\mathbf{Q}_{L2}^{\text{opt}} = \underset{\mathbf{Q} \in \mathbb{C}^{D \times K}; \mathbf{Q}^H \mathbf{Q} = \mathbf{I}_K}{\text{argmax}} \|\mathbf{Q}^H \mathbf{X}\|_2 \quad (1)$$

where, for any  $\mathbf{A} \in \mathbb{C}^{m \times n}$ , the L2-norm<sup>1</sup> is defined as  $\|\mathbf{A}\|_2 = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{i,j}|^2}$ ,  $|\cdot|$  denotes the magnitude of a complex number (coinciding with the absolute value of a real number),  $\mathbf{I}_K$  is the size- $K$  identity matrix, and  $(\cdot)^H$  denotes the Hermitian conjugate of a matrix or vector.

The popularity of L2-PCA can be partially credited to the computational simplicity by which the solution to (1) can be obtained. The solution  $\mathbf{Q}_{L2}^{\text{opt}}$  is formed by the  $K$  dominant singular vectors of  $\mathbf{X}$  and is, thus, obtained by means of Singular-Value Decomposition (SVD) of  $\mathbf{X}$  with quadratic complexity in the number of data samples  $N$  [5]. Moreover, L2-PCA is a scalable operation in the sense that  $\mathbf{Q}_{L2D \times (K+1)}^{\text{opt}}$  is  $\mathbf{Q}_{L2D \times K}^{\text{opt}}$  augmented by a new orthogonal column to be found from  $\mathbf{Q}_{L2D \times K}^{\text{opt}}$  and  $\mathbf{X}$ . In addition, there are several algorithms that can efficiently update the solution to (1) as new data points become available [6]. Finally, by the Projection Theorem [5] it is easy to show that the maximum-L2-norm-projection problem in (1) is equivalent to the familiar minimum L2-norm error problem

$$\underset{\substack{\mathbf{Q} \in \mathbb{C}^{D \times K}; \mathbf{Q}^H \mathbf{Q} = \mathbf{I}_K \\ \mathbf{Z} \in \mathbb{C}^{K \times N}}}{\text{minimize}} \|\mathbf{X} - \mathbf{QZ}\|_2. \quad (2)$$

On the downside, conventional L2-PCA seeking to maximize the L2-norm of the projected data points in (1) is well-known to be overly sensitive to faulty measurements/outliers in the processed matrix. Such outliers may leak into the data matrix due to a number of different causes, including sensing/hardware malfunctions, external interference, and errors in data storage or transcription. Regardless of the cause, outliers are described/understood as unexpected, erroneous values that lie far from the nominal data subspace and are not to be seen under normal system operation. Since its original inception [2], engineers and mathematicians have been trying to robustify

Manuscript received August 1, 2017; revised December 13, 2017 and March 3, 2018; accepted March 10, 2018. Date of publication April 27, 2018; date of current version May 10, 2018. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Adel Belouchrani. This work was supported in part by the National Science Foundation under Grant ECCS-1462341 and in part by the Office of the Vice President for Research, Rochester Institute of Technology. This paper was presented in part at the 16th IEEE International Workshop on Signal Processing Advances in Wireless Communications, KTH Royal Institute of Technology, Stockholm, Sweden, June 2015. (*Corresponding author: Dimitris A. Pados.*)

N. Tsagkarakis was with the Department of Electrical Engineering, University at Buffalo, The State University of New York, Buffalo, NY 14260 USA. He is now with Ericsson AB, 417 56 Göteborg, Sweden (e-mail: nikolaos.tsagkarakis@ericsson.com).

P. P. Markopoulos is with the Department of Electrical and Microelectronic Engineering, Rochester Institute of Technology, Rochester, NY 14623 USA (e-mail: panos@rit.edu).

G. Sklivanitis and D. A. Pados are with the Department of Computer and Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL 33431 USA (e-mail: gsklivanitis@fau.edu; dpados@fau.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2018.2821641

<sup>1</sup> The L2-norm is also known as Frobenius or Euclidean norm [4], [5].

L2-PCA against outliers. Popular outlier-resistant versions of PCA are weighted PCA (WPCA) [7], [8], influence-function PCA [9], robust PCA [10], and most recently direct L1-norm PCA [20], [21].

From an algebraic viewpoint, of all robust versions of PCA, L1-PCA is arguably the most straightforward approach. Mathematically, L1-PCA of a *real-valued* data matrix  $\mathbf{X} \in \mathbb{R}^{D \times N}$  is formulated as

$$\underset{\mathbf{Q} \in \mathbb{R}^{D \times K}; \mathbf{Q}^T \mathbf{Q} = \mathbf{I}_K}{\text{maximize}} \quad \|\mathbf{Q}^T \mathbf{X}\|_1 \quad (3)$$

where  $\|\cdot\|_1$  is the L1-norm operator, such that for any  $\mathbf{A} \in \mathbb{C}^{m \times n}$ ,  $\|\mathbf{A}\|_1 = \sum_{i=1}^m \sum_{j=1}^n |a_{i,j}|$ . By not placing squared emphasis on the magnitude of each point (as L2-PCA does), L1-PCA is far more resistant to outlying, peripheral data points. Importantly, recent studies have shown that when the processed data are not outlier corrupted the solutions of L1-PCA and L2-PCA describe a nearly identical subspace.

Due to its outlier resistance, L1-PCA of real-valued data matrices has attracted extensive documented research interest since the mid-twentieth century [11], [12] and even more so in the past decade. In [13]–[15], the authors focused on L1-PCA in the form of residual-error minimization. Minimum-error L1-PCA has to date no known solution for its general multiple-component case. Alternatively, researchers considered projection-maximization L1-PCA [16]–[19] where data projection is measured by means of the L1-norm. Projection-maximization L1-PCA is the focal point of this present work and is henceforth simply referred to as “L1-PCA.” In [20], [21], Markopoulos, Karystinos, and Pados provided the first exact solvers for L1-PCA of real valued data. In [22]–[24], state-of-the-art approximate/efficient algorithms with favorable complexity/performance trade-off were presented. A low-cost/high-performance L1-PCA/SVD hybrid method was presented in [25]. In [26], L1-PCA was employed for outlier identification and excision. In [27]–[33], L1-PCA was used for image fusion, robust face recognition, and video surveillance. In [34], L1-PCA was used for the design of reduced-rank filters for sensor-array processing. In [35], L1-PCA was used for the design of robust nearest-subspace classifiers for radar-based indoor motion recognition. Finally, L1-PCA-based methods for direction-of-arrival (DoA) estimation were presented in [1] and [36].

In the real-valued L1-PCA literature, it was shown that real-valued L1-PCA can be converted into a combinatorial problem over antipodal binary variables ( $\pm 1$ ) solvable with intrinsic complexity, polynomial in the data record size,  $\mathcal{O}(N^{\text{rank}(\mathbf{X})K-K+1})$  [20]. L1-PCA for complex-data processing remains to date largely unexplored. Similar to (3), complex L1-PCA is formulated as

$$\underset{\mathbf{Q} \in \mathbb{C}^{D \times K}; \mathbf{Q}^H \mathbf{Q} = \mathbf{I}_K}{\text{argmax}} \quad \|\mathbf{Q}^H \mathbf{X}\|_1. \quad (4)$$

Interestingly, in contrast to real-valued L1-PCA, complex L1-PCA in (4) has no obvious connection to a combinatorial problem. Moreover, no finite-step algorithm (exponential or other) has ever been reported for optimally solving (4). Yet, as a robust analogous to complex L2-PCA, complex L1-PCA in (4) can be traced to many important applications that involve

complex-valued measurements, e.g., in the fields communications, radar processing, or general signal processing in complex-domain transformations (such as Fourier) of real-valued data.

Our contributions in this paper are summarized as follows.

- 1) We prove that (4) can be cast as an optimization problem over the set of unimodular matrices.<sup>2</sup>
- 2) We provide the first two fast algorithms to solve (4) sub-optimally.
- 3) We offer numerical and experimental studies that evaluate the performance of the complex L1-PCA algorithms.

Importantly, our numerical studies illustrate that complex L1-PCA exhibits sturdy resistance against outliers, while it performs similarly to standard L2-PCA when the processed data are outlier-free.

The rest of the paper is organized as follows. Section II offers as brief overview of technical preliminaries and notation. Section III is devoted to the presentation of our theoretical findings and the derivation of the proposed algorithms. Section IV holds our numerical studies. Finally, some concluding remarks are drawn in Section V.

## II. PRELIMINARIES AND NOTATION

### A. Complex Sign

Every complex number  $a$  can be written as the product of its magnitude and a complex exponential; i.e.,  $a = |a|\text{sgn}(a)$  where  $\text{sgn}(a) \triangleq e^{j\text{Arg}(a)}$ . The argument  $\text{Arg}(a)$  is the angle that  $a$  creates with the axis of the real numbers; i.e.,  $\text{Arg}(a)$  satisfies  $\text{Arg}(a) = \tan^{-1} \left( \frac{\Im\{a\}}{\Re\{a\}} \right)$ . Alternatively, we can refer to the sign of a complex number  $a$  as the uniquely defined complex number that has unit modulus and argument equal to the argument of  $a$ . We can, then, equivalently define  $\text{sgn}(a) \triangleq \frac{a}{|a|}$ . The sign of any complex number belongs to the unitary complex circle

$$U \triangleq \{a \in \mathbb{C} : |a| = 1\}. \quad (5)$$

For every  $a \in \mathbb{C}$ , with  $|a| > 0$ ,  $\text{sgn}(a)$  is the point in  $U$  that lies nearest to  $a$  in the magnitude sense. That is,  $\text{sgn}(a) = \underset{b \in U}{\text{argmin}} |b - a|$  or, by elementary algebraic manipulations,

$$\text{sgn}(a) = \underset{b \in U}{\text{argmax}} \Re\{b^* a\}. \quad (6)$$

The maximum value in (6) is the magnitude of  $a$ .

The definition and properties of the sign can be generalized to vectors and matrices. We define the sign of a matrix  $\mathbf{A} \in \mathbb{C}^{m \times n}$  as the matrix that contains the signs of the individual entries of  $\mathbf{A}$ . That is, we define

$$\text{sgn}(\mathbf{A}) \triangleq \begin{bmatrix} \text{sgn}(a_{1,1}) & \dots & \text{sgn}(a_{1,n}) \\ \vdots & \ddots & \vdots \\ \text{sgn}(a_{m,1}) & \dots & \text{sgn}(a_{m,n}) \end{bmatrix}. \quad (7)$$

In accordance to (6), the sign of  $\mathbf{A}$  can be expressed as the solution to the maximization problem

$$\text{sgn}(\mathbf{A}) = \underset{\mathbf{B} \in U^{n \times m}}{\text{argmax}} \Re\{\text{Tr}\{\mathbf{B}\mathbf{A}\}\}. \quad (8)$$

<sup>2</sup>In this work, we call a matrix unimodular if every entry takes values on the unitary complex circle.

Moreover, the optimal objective value of (8) is the L1-norm of  $\mathbf{A}$ ; that is,

$$\|\mathbf{A}\|_1 = \max_{\mathbf{B} \in U^{n \times m}} \Re\{\text{Tr}\{\mathbf{B}\mathbf{A}\}\} = \text{Tr}\left\{\text{sgn}(\mathbf{A})^H \mathbf{A}\right\}. \quad (9)$$

Finally, by the above definitions it holds that the sign of the product of two complex numbers is equal to the product of the individual signs. In addition, the sign of a number is  $+1$  if and only if the number is real and positive and  $-1$  if and only if the number is real and negative.

### B. Nuclear Norm

Consider any matrix  $\mathbf{A} \in \mathbb{C}^{m \times n}$  with  $m > n$  without loss of generality. Let  $\mathbf{A}$  admit SVD  $\mathbf{A} = \mathbf{U}\text{Diag}(\boldsymbol{\sigma})\mathbf{V}^H$ , where  $\mathbf{U}^H\mathbf{U} = \mathbf{V}^H\mathbf{V} = \mathbf{I}_n$  and  $\boldsymbol{\sigma} \in \mathbb{R}_+^{n \times 1}$  contains the singular values of  $\mathbf{A}$  in descending order (i.e.,  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$ ).<sup>3</sup> The nuclear norm of  $\mathbf{A}$  is then defined as the summation of the singular values of  $\mathbf{A}$ ,

$$\|\mathbf{A}\|_* \triangleq \sum_{i=1}^r \sigma_i. \quad (10)$$

It holds that  $\|\mathbf{A}\|_* = \|\mathbf{A}^H\|_*$ . The nuclear norm can be expressed in several different ways. For example, in connection to the *Orthogonal Procrustes Theorem* [5], [37],

$$\|\mathbf{A}\|_* = \max_{\mathbf{Q} \in \mathbb{C}^{m \times n}; \mathbf{Q}^H \mathbf{Q} = \mathbf{I}_n} \Re\{\text{Tr}\{\mathbf{Q}^H \mathbf{A}\}\}. \quad (11)$$

Moreover, denoting by  $\text{unt}(\mathbf{A})$  the  $m \times n$  unitary matrix that maximizes (11) and assuming that  $\mathbf{A}$  has full column rank (i.e.,  $\text{rank}(\mathbf{A}) = n$ ), it holds that<sup>4</sup>

$$\mathbf{A} = \text{unt}(\mathbf{A}) (\mathbf{A}^H \mathbf{A})^{\frac{1}{2}}, \quad (12)$$

which is known as the *polar decomposition* of  $\mathbf{A}$  [5], [38]. Finally,  $\text{unt}(\mathbf{A})$  can be calculated by SVD of  $\mathbf{A}$  as

$$\text{unt}(\mathbf{A}) = \mathbf{U}\mathbf{V}^H. \quad (13)$$

Using the above preliminaries and notation, in the following section we present our new developments on complex L1-PCA.

<sup>3</sup>By  $\text{Diag}(\boldsymbol{\sigma})$  we denote the square diagonal matrix with the elements of vector  $\boldsymbol{\sigma}$  on its main diagonal.

<sup>4</sup>For any positive semi-definite matrix  $\mathbf{Z} \in \mathbb{C}^{n \times n}$ ,  $\mathbf{Z}^{\frac{1}{2}}$  is defined such that  $\mathbf{Z} = \mathbf{Z}^{\frac{1}{2}} \mathbf{Z}^{\frac{1}{2}}$ ; that is, if  $\mathbf{Z}$  admits SVD  $\mathbf{Z} = \mathbf{Q}\boldsymbol{\Lambda}^2\mathbf{Q}^H$ , then  $\mathbf{Z}^{\frac{1}{2}} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^H$ . Also, for any  $\mathbf{A} \in \mathbb{C}^{m \times n}$ ,  $\|\mathbf{A}\|_* = \text{Tr}\{(\mathbf{A}^H \mathbf{A})^{\frac{1}{2}}\}$ .

## III. COMPLEX L1-PCA

### A. Problem Connection to Unimodular Optimization

In view of (8) and (11) we can rewrite the complex L1-PCA problem in (4) as

$$\max_{\mathbf{Q} \in \mathbb{C}^{D \times K}; \mathbf{Q}^H \mathbf{Q} = \mathbf{I}_K} \|\mathbf{Q}^H \mathbf{X}\|_1 \quad (14)$$

$$= \max_{\mathbf{Q} \in \mathbb{C}^{D \times K}; \mathbf{Q}^H \mathbf{Q} = \mathbf{I}_K} \text{Tr}\left\{\text{sgn}(\mathbf{Q}^H \mathbf{X})^H \mathbf{Q}^H \mathbf{X}\right\} \quad (15)$$

$$\stackrel{(8)}{=} \max_{\substack{\mathbf{Q} \in \mathbb{C}^{D \times K}; \mathbf{Q}^H \mathbf{Q} = \mathbf{I}_K \\ \mathbf{B} \in U^{N \times K}}} \Re\{\text{Tr}\{\mathbf{B}\mathbf{Q}^H \mathbf{X}\}\} \quad (16)$$

$$\stackrel{(11)}{=} \max_{\mathbf{B} \in U^{N \times K}} \|\mathbf{X}\mathbf{B}\|_*. \quad (17)$$

That is, we establish in (17) that complex L1-PCA is directly connected to a maximization problem over the set of  $N \times K$  unimodular matrices. Interestingly, Markopoulos *et al.* [20], [22] proved an analogous result for the real case. Specifically, [20] reformulated real-valued L1-PCA to a nuclear-norm maximization over the set of  $N \times K$   $\{\pm 1\}$ -valued matrices,  $\{\pm 1\}^{N \times K}$ . Considering that  $\{\pm 1\}$  is in fact the intersection of  $U$  with the axis of real numbers, we realize that the binary-nuclear-norm maximization to which real-valued L1-PCA corresponds (see [20]) constitutes a relaxation of the unimodular-nuclear-norm maximization in (17). Due to the finite size of  $\{\pm 1\}^{N \times K}$ , finite-step algorithms can be devised for the solution of real-valued L1-PCA. Regretfully, since  $U^{N \times K}$  is uncountably infinite, this is not the case for complex L1-PCA.

Even though unimodular-nuclear-norm maximization in (17) cannot be solved by evaluating the objective function at every feasible point, there are necessary optimality conditions that we can use to devise efficient algorithms for solving (17) at least locally. The following proposition introduces the first of these optimality conditions.

**Proposition 1:** Let  $(\mathbf{Q}_{L1}^{\text{opt}}, \mathbf{B}^{\text{opt}})$  be an optimal solution pair for (16). Then,

$$\mathbf{Q}_{L1}^{\text{opt}} = \text{unt}(\mathbf{X}\mathbf{B}^{\text{opt}}) \quad \text{and} \quad \mathbf{B}^{\text{opt}} = \text{sgn}(\mathbf{X}^H \mathbf{Q}_{L1}^{\text{opt}}). \quad (18)$$

Moreover,  $\mathbf{Q}_{L1}^{\text{opt}}$  is a solution to (4) and  $\mathbf{B}^{\text{opt}}$  is a solution to (17). ■

Proposition 1 derives directly from (8), (11), and the fact that both (4) and (17) are equivalent to (16). Most importantly, Proposition 1 establishes that, if  $\mathbf{B}^{\text{opt}}$  is a solution to (17), then  $\mathbf{Q}_{L1}^{\text{opt}} = \text{unt}(\mathbf{X}\mathbf{B}^{\text{opt}})$  is a solution to L1-PCA in (4). Thus, one can focus on solving (17) and then use its solution to derive the L1-PCs by means of simple SVD (see definition of  $\text{unt}(\cdot)$  in (13)). In addition, the two equations in (18) can be combined to form a new pair of necessary optimality conditions on the individual problems (4) and (17). The new optimality conditions are presented in Corollary 1 below.

**Corollary 1:** Let  $\mathbf{Q}_{L1}^{\text{opt}}$  be a solution to (4); then,  $\mathbf{Q}_{L1}^{\text{opt}} = \text{unt}(\mathbf{X}\text{sgn}(\mathbf{X}^H \mathbf{Q}_{L1}^{\text{opt}}))$ . Let  $\mathbf{B}^{\text{opt}}$  be a solution to (17); then,  $\mathbf{B}^{\text{opt}} = \text{sgn}(\mathbf{X}^H \text{unt}(\mathbf{X}\mathbf{B}^{\text{opt}}))$ . ■

### B. Complex L1-PCA When $\text{Rank}(\mathbf{X}) < D$

Consider  $\mathbf{X} \in \mathbb{C}^{D \times N}$  with  $\text{rank}(\mathbf{X}) = r < D$ .  $\mathbf{X}$  admits thin SVD<sup>5</sup>  $\mathbf{X} = \mathbf{U}_x \mathbf{S}_x \mathbf{V}_x^H$  where  $\mathbf{U}_x^H \mathbf{U}_x = \mathbf{V}_x^H \mathbf{V}_x = \mathbf{I}_r$  and  $\mathbf{S}_x$  is the  $r \times r$  diagonal matrix that contains the non-zero singular values of  $\mathbf{X}$ . To obtain the  $K \leq r$  L1-PCs of  $\mathbf{X}$ ,  $\mathbf{Q}_{L1}^{\text{opt}}$ , we can work in two steps: (i) Obtain the solution  $\mathbf{B}^{\text{opt}}$  to (17) and (ii) perform SVD on  $\mathbf{X}\mathbf{B}^{\text{opt}}$  and return  $\mathbf{Q}_{L1}^{\text{opt}} = \text{unt}(\mathbf{X}\mathbf{B}^{\text{opt}})$ . We focus for a moment on the first step. We observe that

$$\|\mathbf{X}\mathbf{B}\|_* = \|\mathbf{U}_x \mathbf{S}_x \mathbf{V}_x^H \mathbf{B}\|_* \quad (19)$$

$$= \text{Tr} \left\{ \left( (\mathbf{U}_x \mathbf{S}_x \mathbf{V}_x^H \mathbf{B})^H (\mathbf{U}_x \mathbf{S}_x \mathbf{V}_x^H \mathbf{B}) \right)^{\frac{1}{2}} \right\} \quad (20)$$

$$= \text{Tr} \left\{ \left( \mathbf{B}^H \mathbf{V}_x^H \mathbf{S}_x^H \mathbf{S}_x \mathbf{V}_x^H \mathbf{B} \right)^{\frac{1}{2}} \right\} \quad (21)$$

$$= \|\mathbf{S}_x \mathbf{V}_x^H \mathbf{B}\|_* = \|\mathbf{Y}\mathbf{B}\|_* \quad (22)$$

where  $\mathbf{Y} \triangleq \mathbf{S}_x \mathbf{V}_x^H \in \mathbb{C}^{r \times N}$ . Then,  $\mathbf{B}^{\text{opt}}$  maximizes both (19) (by definition) and (22) (by equivalence). Notice also that  $\text{unt}(\mathbf{X}\mathbf{B}^{\text{opt}}) = \mathbf{U}_x \text{unt}(\mathbf{Y}\mathbf{B}^{\text{opt}})$ . Proposition 2 below follows.

**Proposition 2:** Consider  $\mathbf{X} \in \mathbb{C}^{D \times N}$  with  $\text{rank}(\mathbf{X}) = r \leq \min\{D, N\}$ , admitting thin SVD  $\mathbf{X} = \mathbf{U}_x \mathbf{S}_x \mathbf{V}_x^H$  (i.e.,  $\mathbf{S}_x$  is  $r \times r$ ). Define  $\mathbf{Y} = \mathbf{S}_x \mathbf{V}_x^H$  and  $\tilde{\mathbf{Q}}_{L1}^{\text{opt}} = \arg\max_{\{\mathbf{Q} \in \mathbb{C}^{r \times K}; \mathbf{Q}^H \mathbf{Q} = \mathbf{I}_K\}} \|\mathbf{Q}^H \mathbf{Y}\|_1$ . Then,

$$\mathbf{Q}_{L1}^{\text{opt}} = \mathbf{U}_x \tilde{\mathbf{Q}}_{L1}^{\text{opt}} \quad (23)$$

is a solution to the L1-PCA problem in (4). Moreover,  $\|(\mathbf{Q}_{L1}^{\text{opt}})^H \mathbf{X}\|_1 = \|\mathbf{X}\mathbf{B}^{\text{opt}}\|_* = \|\mathbf{Y}\mathbf{B}^{\text{opt}}\|_* = \|(\tilde{\mathbf{Q}}_{L1}^{\text{opt}})^H \mathbf{Y}\|_1$ . ■

Proposition 2 shows that the L1-PCs of a rank- $r$   $D \times N$  matrix can always be obtained through the L1-PCA of a rank- $r$   $r \times N$  matrix. Therefore, Proposition 2 steers our algorithmic focus to problems where  $\mathbf{X}$  has full row-rank (i.e.,  $r = \text{rank}(\mathbf{X}) = D$ ).

### C. The Single-Component Case and L1-PCA Hardness

In its simplest non-trivial form, complex L1-PCA is the search for a single ( $K = 1$ ) component  $\mathbf{q} \in \mathbb{C}^{D \times 1}$ ,  $\|\mathbf{q}\|_2 = 1$ , that maximizes  $\|\mathbf{q}^H \mathbf{X}\|_1$ . According to our general developments for the multi-component ( $K \geq 1$ ) case, the pursuit of a single L1-PC can also be rewritten as unimodular nuclear-norm maximization. That is,

$$\max_{\mathbf{q} \in \mathbb{C}^{D \times 1}; \|\mathbf{q}\|_2=1} \|\mathbf{q}^H \mathbf{X}\|_1 = \max_{\mathbf{b} \in U^{N \times 1}} \|\mathbf{X}\mathbf{b}\|_* \quad (24)$$

$$= \max_{\mathbf{b} \in U^{N \times 1}} \|\mathbf{X}\mathbf{b}\|_2 \quad (25)$$

$$= \sqrt{\max_{\mathbf{b} \in U^{N \times 1}} \mathbf{b}^H \mathbf{X}^H \mathbf{X} \mathbf{b}}. \quad (26)$$

Equation (26) derives from the fact that any complex vector  $\mathbf{a} \in \mathbb{C}^m$  admits SVD  $\mathbf{a} = \mathbf{u}\sigma$  with  $\mathbf{u} = \mathbf{a}/\|\mathbf{a}\|_2^{-1}$  and  $\sigma = \|\mathbf{a}\|_2 = \|\mathbf{a}\|_1$  (trivially, the dominant right-hand singular vector is 1). We observe that under the square root in (26), the objective function is a quadratic polynomial and the optimization problem

is unimodular quadratic maximization (UQM), which has been well-studied in the literature [39]. By Proposition 1, the L1-PC  $\mathbf{q}_{L1}^{\text{opt}}$  that solves (24) is

$$\mathbf{q}_{L1}^{\text{opt}} = \text{unt}(\mathbf{X}\mathbf{b}^{\text{opt}}) = \mathbf{X}\mathbf{b}^{\text{opt}} \|\mathbf{X}\mathbf{b}^{\text{opt}}\|_2^{-1} \quad (27)$$

where  $\mathbf{b}^{\text{opt}}$  is the solution to (26). Also,  $\mathbf{b}^{\text{opt}} = \text{sgn}(\mathbf{X}^H \mathbf{q}_{L1}^{\text{opt}})$ .

Interesting applications of (24)–(27) include the design of maximum-signal-to-interference-plus-noise-ratio (max-SINR) phased arrays and the design of unimodular codes [39]. For the real data case, UQM in (26) takes the form of binary quadratic maximization [40], [41] as proven in [20].

Certainly, the necessary optimality condition presented in Corollary 1 for (17) also applies to (26). Specifically, if  $\mathbf{b}^{\text{opt}}$  is a solution to (26), then

$$\mathbf{b}^{\text{opt}} = \text{sgn}(\mathbf{X}^H \text{unt}(\mathbf{X}\mathbf{b}^{\text{opt}})) = \text{sgn}(\mathbf{X}^H \mathbf{X}\mathbf{b}^{\text{opt}}). \quad (28)$$

However, in contrast to what has been shown for the real case [16], [17], (28) is not a sufficient condition for local optimality. The reason is that  $\mathbf{b} = \text{sgn}(\mathbf{X}^H \mathbf{X}\mathbf{b})$  is also satisfied by “saddle” points of  $\|\mathbf{X}\mathbf{b}\|_2$ . Below, we provide a stronger optimality condition than (28) that is necessary and sufficient for local optimality.

For compactness in notation, we begin by defining

$$\boldsymbol{\omega}(\mathbf{b}) \triangleq \mathbf{b}^* \odot (\mathbf{X}^H \mathbf{X}\mathbf{b}) \in \mathbb{C}^{N \times 1}, \quad (29)$$

for any  $\mathbf{b} \in U^N$  where  $(\cdot)^*$  performs complex-conjugation and  $\odot$  is the element-wise product (Hadamard) operator. Even though the entries of  $\boldsymbol{\omega}(\mathbf{b})$  are complex, their summation  $\sum_{n=1}^N \omega_n$  is real and positive and equal to the quadratic form  $\mathbf{b}^H \mathbf{X}^H \mathbf{X} \mathbf{b}$ . The following proposition presents a necessary condition for optimality for the UQM problem of (26) that is satisfied by all local maximizers, but not by saddle points (or minimizers).

**Proposition 3:** A unimodular vector  $\mathbf{b}$  is a local maximizer to (26) if and only if

$$\omega_n(\mathbf{b}) \in \mathbb{R} \quad \text{and} \quad \omega_n(\mathbf{b}) \geq \|\mathbf{x}_n\|_2^2 \quad \forall n \in \{1, \dots, N\}. \quad (30)$$

*Proof:* For any  $\mathbf{b} \in U^{N \times 1}$ , there exists an angle vector  $\boldsymbol{\phi} \in [0, 2\pi)^{N \times 1}$  such that  $\mathbf{b} = e^{j\boldsymbol{\phi}} \triangleq [e^{j\phi_1}, \dots, e^{j\phi_N}]^T$ . The quadratic in the UQM problem of (26) can be then rewritten as  $(e^{j\boldsymbol{\phi}})^H \mathbf{X}^H \mathbf{X} (e^{j\boldsymbol{\phi}})$ , which is a function twice continuously differentiable in the angle vector  $\boldsymbol{\phi}$ ; the corresponding first and second derivatives are

$$\mathbf{g}(\mathbf{b}) = 2\Im\{\boldsymbol{\omega}(\mathbf{b})\} \quad (31)$$

and

$$\mathbf{H}(\mathbf{b}) = 2\Re\{\text{Diag}(\mathbf{b})^H \mathbf{X}^H \mathbf{X} \text{Diag}(\mathbf{b}) - \text{Diag}(\boldsymbol{\omega}(\mathbf{b}))\}, \quad (32)$$

respectively. Any local maximizer of (26) will null the gradient and render the Hessian in (32) negative semidefinite. Below, we prove the two directions of the statement of Proposition 3.

<sup>5</sup>In “thin SVD” a matrix is written only in terms of its singular vectors that correspond to non-zero singular values.



If  $\mathbf{b}$  is a local maximizer of (26), then  $\mathbf{H}(\mathbf{b}) \preceq 0$ . Therefore, for every  $n \in \{1, 2, \dots, N\}$ ,

$$\mathbf{e}_n^T \mathbf{H}(\mathbf{b}) \mathbf{e}_n \leq 0 \Leftrightarrow 2\Re\{b_n^* \mathbf{x}_n^H \mathbf{x}_n b_n - \omega_n(\mathbf{b})\} \leq 0 \quad (33)$$

$$\Leftrightarrow \omega_n(\mathbf{b}) \geq \|\mathbf{x}_n\|_2^2 \quad (34)$$

where  $\mathbf{e}_n$  the  $n$ -th column of  $\mathbf{I}_N$ .

Consider  $\mathbf{b}$  such that all entries of  $\omega(\mathbf{b})$  satisfy (30). Then, for every  $\mathbf{z} \in \mathbb{C}^N$ ,

$$\mathbf{z}^T \mathbf{H}(\mathbf{b}) \mathbf{z} = 2\mathbf{z}^T (\text{Diag}(\mathbf{b})^H \mathbf{X}^H \mathbf{X} \text{Diag}(\mathbf{b}) - \text{Diag}(\omega(\mathbf{b}))) \mathbf{z} \quad (35)$$

$$= 2 \left\| \sum_{n=1}^N z_n b_n \mathbf{x}_n \right\|_2^2 - 2 \sum_{n=1}^N z_n^2 \omega_n(\mathbf{b}) \quad (36)$$

$$\leq 2 \sum_{n=1}^N \|z_n b_n \mathbf{x}_n\|_2^2 - 2 \sum_{n=1}^N z_n^2 \omega_n(\mathbf{b}) \quad (37)$$

$$= 2 \sum_{n=1}^N z_n^2 (\|\mathbf{x}_n\|_2^2 - \omega_n(\mathbf{b})) \leq 0 \quad (38)$$

which implies that the Hessian at  $\mathbf{b}$ ,  $\mathbf{H}(\mathbf{b})$ , is negative semidefinite. This, in turn, implies that  $\hat{\mathbf{b}}$  is a local maximizer of (26). ■

The following statement is a direct corollary of Proposition 3.

*Corollary 2:* A unimodular vector  $\mathbf{b}$  is a local maximizer to (26) if and only if

$$\mathbf{b} = \text{sgn}(\mathbf{A}_d \mathbf{b}) \quad (39)$$

where  $\mathbf{A}_d \triangleq \mathbf{X}^H \mathbf{X} - \text{Diag}([\|\mathbf{x}_1\|_2^2, \dots, \|\mathbf{x}_N\|_2^2]^T)$ .

*Proof:* For every  $n \in \{1, 2, \dots, N\}$ ,

$$\omega_n(\mathbf{b}) \geq \|\mathbf{x}_n\|_2^2 \Leftrightarrow \omega_n(\mathbf{b}) - \|\mathbf{x}_n\|_2^2 \geq 0 \quad (40)$$

$$\Leftrightarrow \text{sgn}(\omega_n(\mathbf{b}) - \|\mathbf{x}_n\|_2^2) = 1 \quad (41)$$

$$\Leftrightarrow b_n = \text{sgn} \left( \sum_{m \neq n} \mathbf{x}_n^H \mathbf{x}_m b_m \right) \quad (42)$$

which, by the definition of  $\mathbf{A}_d$ , implies (39). ■

The quantitative difference between conditions (28) and (39) lies in the corresponding  $\omega(\cdot)$  variables. On the one hand, condition (28) guarantees that  $\omega_n(\mathbf{b}^{\text{opt}})$  is positive; on the other hand, condition (39) guarantees that, for every  $n \in \{1, 2, \dots, N\}$ ,  $\omega_n(\mathbf{b}^{\text{opt}})$  is not only positive but also greater than or equal to  $\|\mathbf{x}_n\|_2^2$ . Hence, (39) is a stronger condition than (28) ((39) implies (28), but not vice versa). For example, saddle points in  $U^N$  could satisfy the mild condition in (28) but not the necessary-and-sufficient local optimality condition in (39). Proposition 3 and Corollary 2 brought us a step closer to solving (26) efficiently.

Next, we discuss the hardness of UQM and, therefore, complex L1-PCA ( $K = 1$ ). In [39] and references therein, it is formally shown that UQM in (26) is an  $\mathcal{NP}$ -hard problem. Accordingly, the equivalent complex L1-PCA problem for  $K = 1$  is also  $\mathcal{NP}$ -hard. In addition, the nuclear-norm maximization in

(17) and the equivalent L1-PCA in (4) are also  $\mathcal{NP}$ -hard, since (17) is a generalization of (26) for  $K \geq 1$ . In conclusion, in contrast to the real-field case of [20], complex L1-PCA remains  $\mathcal{NP}$ -hard in the sample size  $N$  even for fixed dimension  $D$ .

#### D. Proposed Algorithms for Complex L1-PCA

Based on the theoretical analysis above, we present two algorithms for complex L1-PCA. Both algorithms are iterative and guaranteed to converge. With proper initialization, both algorithms can return upon convergence the global optimal solution of (17). Our first algorithm relies on (28) and can be applied to general  $K \geq 1$  cases. Our second algorithm relies on the stronger condition (39) and is applicable only to the  $K = 1$  case.

1) *Algorithm 1:* For any given data matrix  $\mathbf{X} \in \mathbb{C}^{D \times N}$  and number of sought after L1-PCs  $K < \text{rank}(\mathbf{X})$ , the algorithm initializes at an arbitrary unimodular matrix  $\mathbf{B}^{(0)} \in U^{N \times K}$ . Then, in view of the optimality condition in (28), the algorithm performs the iteration

$$\mathbf{B}^{(i)} = \text{sgn} \left( \mathbf{X}^H \text{unt} \left( \mathbf{X} \mathbf{B}^{(i-1)} \right) \right), \quad i = 1, 2, \dots \quad (43)$$

until the objective value in (17) converges. That is, practically, the algorithm terminates at the first iteration  $t$  that satisfies

$$\|\mathbf{X} \mathbf{B}^{(t)}\|_* - \|\mathbf{X} \mathbf{B}^{(t-1)}\|_* \leq \delta \quad (44)$$

for some arbitrarily low convergence threshold  $\delta \geq 0$ . Then, the algorithm returns  $\mathbf{B}^{(t)}$  as (approximate) solution to (17) and, in accordance to (18),  $\hat{\mathbf{Q}}_{L1} = \text{unt}(\mathbf{X} \mathbf{B}^{(t)})$  as (approximate) solution to the L1-PCA problem in (4). Below we provide a proof of convergence for the iteration in (43), for any initialization  $\mathbf{B}^{(0)}$ .

Convergence of (43) is guaranteed because the sequence  $\{\|\mathbf{X} \mathbf{B}^{(i)}\|_*\}_{i=0}^\infty$  is (i) upper bounded by  $\|\mathbf{X} \mathbf{B}^{\text{opt}}\|_*$  and (ii) monotonically increasing. That is, for every  $i$ ,  $\|\mathbf{X} \mathbf{B}^{(i-1)}\|_* \leq \|\mathbf{X} \mathbf{B}^{(i)}\|_* \leq \|\mathbf{X} \mathbf{B}^{\text{opt}}\|_*$ . The monotonicity of the sequence can be proven as follows.

$$\|\mathbf{X} \mathbf{B}^{(i)}\|_* \quad (45)$$

$$= \max_{\mathbf{Q} \in \mathbb{C}^{D \times K}; \mathbf{Q}^H \mathbf{Q} = \mathbf{I}_K} \Re \left\{ \text{Tr} \left\{ \mathbf{Q}^H \mathbf{X} \mathbf{B}^{(i)} \right\} \right\} \quad (46)$$

$$\geq \Re \left\{ \text{Tr} \left\{ \text{unt} \left( \mathbf{X} \mathbf{B}^{(i-1)} \right) \right\}^H \mathbf{X} \mathbf{B}^{(i)} \right\} \quad (47)$$

$$= \Re \left\{ \text{Tr} \left\{ \text{unt} \left( \mathbf{X} \mathbf{B}^{(i-1)} \right) \right\}^H \mathbf{X} \text{sgn} \left( \mathbf{X}^H \text{unt} \left( \mathbf{X} \mathbf{B}^{(i-1)} \right) \right) \right\} \quad (48)$$

$$= \max_{\mathbf{B} \in U^{N \times K}} \Re \left\{ \text{Tr} \left\{ \text{unt} \left( \mathbf{X} \mathbf{B}^{(i-1)} \right) \right\}^H \mathbf{X} \mathbf{B}' \right\} \quad (49)$$

$$\geq \Re \left\{ \text{Tr} \left\{ \text{unt} \left( \mathbf{X} \mathbf{B}^{(i-1)} \right) \right\}^H \mathbf{X} \mathbf{B}^{(i-1)} \right\} \quad (50)$$

$$= \|\mathbf{X} \mathbf{B}^{(i-1)}\|_*. \quad (51)$$

The inequality in (47) holds because we have substituted  $\mathbf{Q}$  with a point in the feasibility set of the maximization in (46),

**Algorithm 1: Iterative complex L1-PCA for general  $K \geq 1$  (mild condition).**


---

**Input:**  $\mathbf{X} \in \mathbb{C}^{D \times N}$ ;  $K < \text{rank}(\mathbf{X})$ ; init.  $\mathbf{B}$ ;  $\delta > 0$

- 1:  $\alpha \leftarrow \|\mathbf{XB}\|_*$
- 3: while true
- 4:    $\mathbf{B} \leftarrow \text{sgn}(\mathbf{X}^H \text{unt}(\mathbf{XB}))$
- 5:   if  $\|\mathbf{XB}\|_* - \alpha > \delta$ ,  $\alpha \leftarrow \|\mathbf{XB}\|_*$
- 6:   else, break
- 7:    $[\mathbf{U}, \mathbf{S}_{K \times K}, \mathbf{V}] \leftarrow \text{svd}(\mathbf{XB})$
- 8:    $\mathbf{Q} \leftarrow \mathbf{UV}^H$

**Output:**  $\hat{\mathbf{Q}}_{L1} \leftarrow \mathbf{Q}$  and  $\hat{\mathbf{B}} \leftarrow \mathbf{B}$

---

Fig. 1. Proposed algorithm for L1-PCA of  $\mathbf{X} \in \mathbb{C}^{D \times N}$  (general  $K \geq 1$  case). The algorithm relies on the mild optimality condition in (28).

but not necessarily the maximizer. Similarly, the inequality in (50) holds because we have substituted  $\mathbf{B}$  with a point in the feasibility set of the maximization in (49), but not necessarily the maximizer.

A detailed presentation of Algorithm 1 is offered in pseudocode in Fig. 1.

2) *Algorithm 2* ( $K = 1$ ): Our second algorithm relies on the strong optimality condition of (39). Specifically, given  $\mathbf{X} \in \mathbb{C}^{D \times N}$  and an initialization point  $\mathbf{b} \in U^N$ , Algorithm 2 iterates

$$\mathbf{b}^{(i)} = \text{sgn}(\mathbf{A}_d \mathbf{b}^{(i-1)}), \quad i = 1, 2, \dots, \quad (52)$$

until  $\|\mathbf{A}_d \mathbf{b}^{(t)}\|_1 - \|\mathbf{A}_d \mathbf{b}^{(t-1)}\|_1 \leq \delta$  for some arbitrary small threshold  $\delta$  and converging iteration index  $t$ . Then the algorithm returns  $\mathbf{b}^{(t)}$  as (approximate) solution to (26) and, in accordance to Proposition 1 and (27),  $\hat{\mathbf{q}}_{L1} = \mathbf{X} \mathbf{b}^{(t)} \|\mathbf{X} \mathbf{b}^{(t)}\|_2^{-1}$  as approximate solution to (24). The iteration in (52) will converge because the sequence  $\{\|\mathbf{A}_d \mathbf{b}^{(i)}\|_1\}_{i=1}^\infty$  is (i) upper bounded by  $\|\mathbf{A}_d \mathbf{b}^{\text{opt}}\|_1$  and (ii) increases monotonically since

$$\|\mathbf{A}_d \mathbf{b}^{(i)}\|_1 = \max_{\mathbf{b} \in U^{N \times 1}} \Re \left\{ \mathbf{b}^H \mathbf{A}_d \mathbf{b}^{(i)} \right\} \quad (53)$$

$$\geq \Re \left\{ \mathbf{b}^{(i-1)H} \mathbf{A}_d \mathbf{b}^{(i)} \right\} \quad (54)$$

$$= \Re \left\{ \mathbf{b}^{(i-1)H} \mathbf{A}_d \text{sgn}(\mathbf{A}_d \mathbf{b}^{(i-1)}) \right\} \quad (55)$$

$$= \|\mathbf{A}_d \mathbf{b}^{(i-1)}\|_1. \quad (56)$$

A detailed presentation of Algorithm 2 is offered in pseudocode in Fig. 2.

#### IV. NUMERICAL STUDIES

##### A. Convergence Illustrations

The convergence of Algorithm 1 was formally proven in the previous section. At this point, to visualize the convergence, we fix  $D = 10$ ,  $N = 100$  and generate  $\mathbf{X} \in \mathbb{C}^{D \times N}$  with entries drawn independently from  $\mathcal{CN}(0, 1)$  (zero-mean, unit-variance, complex Gaussian distribution). Then, we run Algorithm 1 on  $\mathbf{X}$  for  $K = 5$  (initialized at arbitrary  $\mathbf{B}^{(0)}$ ) and plot in Fig. 3

**Algorithm 2: Iterative complex L1-PCA for  $K = 1$  (strong condition).**


---

**Input:**  $\mathbf{X} \in \mathbb{C}^{D \times N}$ ; init.  $\mathbf{b}$ ;  $\delta > 0$

- 1:  $\mathbf{A}_d = \mathbf{X}^H \mathbf{X} - \text{Diag}(\|\mathbf{x}_1\|_2^2, \dots, \|\mathbf{x}_N\|_2^2)^T$ ,  $\alpha \leftarrow \|\mathbf{A}_d \mathbf{b}\|_1$
- 3: while true
- 4:    $\mathbf{b} \leftarrow \text{sgn}(\mathbf{A}_d \mathbf{b})$
- 5:   if  $\|\mathbf{A}_d \mathbf{b}\|_1 - \alpha > \delta$ ,  $\alpha \leftarrow \|\mathbf{A}_d \mathbf{b}\|_1$
- 6:   else, break
- 7:    $\mathbf{q} \leftarrow \mathbf{X} \mathbf{b} \|\mathbf{X} \mathbf{b}\|_2^{-1}$

**Output:**  $\hat{\mathbf{q}}_{L1} \leftarrow \mathbf{q}$  and  $\hat{\mathbf{b}} \leftarrow \mathbf{b}$

---

Fig. 2. Proposed algorithm for calculation of the prime ( $K = 1$ ) L1-PC of  $\mathbf{X} \in \mathbb{C}^{D \times N}$ . The algorithm relies on the strong optimality condition in (39).

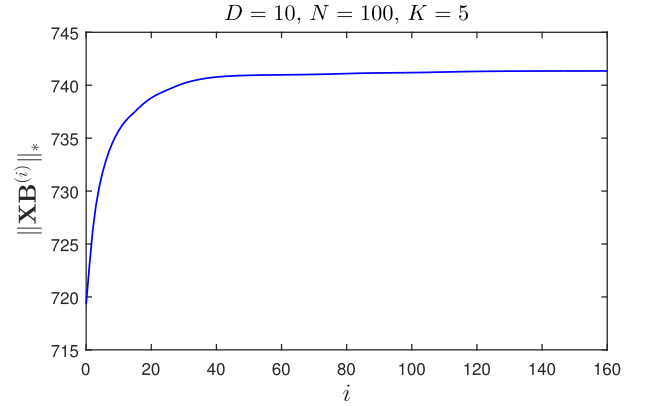


Fig. 3. The objective metric of (17) versus iteration index  $i$  for Algorithm 1 (single realization;  $D = 10$ ,  $N = 100$ ,  $K = 5$ ).

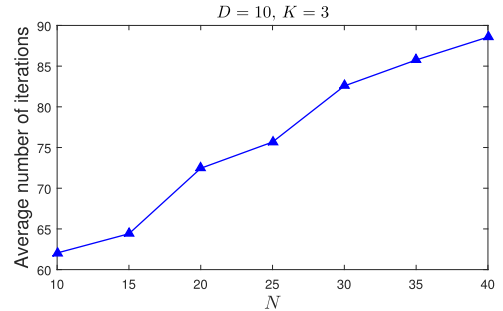


Fig. 4. Average number of iterations needed for the convergence of Algorithm 1 versus number of data points  $N$  (averages over 1000 realizations).

$\|\mathbf{XB}^{(i)}\|_*$  versus the iteration index  $i$ .<sup>6</sup> We observe that, indeed, the objective nuclear-norm maximization metric increases monotonically.

Next, we examine the number of iterations needed for Algorithm 1 to converge as the problem size parameters  $D$ ,  $N$ , and  $K$  take different values. First, we set  $D = 10$ ,  $K = 3$  and vary  $N = 10, 15, \dots, 40$ . We draw again the entries of  $\mathbf{X}$  independently from  $\mathcal{CN}(0, 1)$  and plot in Fig. 4 the average number of iterations needed for Algorithm 1 to converge (aver-

<sup>6</sup>All numerical studies in Section IV, were carried out in Matlab 2015b(R) on an Intel(R) Core(TM) i7-5600U CPU @ 2.60GHz.

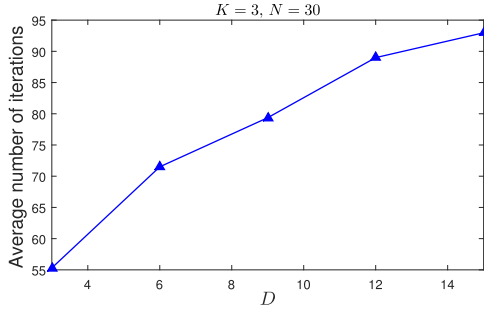


Fig. 5. Average number of iterations needed for the convergence of Algorithm 1 versus size of data point vectors  $D$  (averages over 1000 realizations).

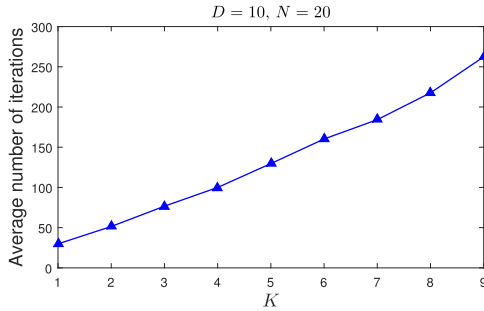


Fig. 6. Average number of iterations needed for the convergence of Algorithm 1 versus target dimensionality  $K$  (averages over 1000 realizations).

aging is carried out over 1000 independent realizations of  $\mathbf{X}$ .) We observe that, as expected, the average number of iterations increases along  $N$ . However, importantly, the increase appears to be sub-linear in  $N$ . In Fig. 5, we fix  $K = 3$ ,  $N = 30$  and plot the average number of iterations needed for convergence versus  $D$ . In Fig. 6, we fix  $D = 10$ ,  $N = 20$  and plot the average number of iterations versus  $K$ . We observe that the number of iterations increases sub-linearly along  $D$  and rather linearly along  $K$ .

### B. Subspace Comparisons

In this experiment, we investigate and compare the outlier resistance of L2-PCA and L1-PCA. We consider the data matrix  $\mathbf{X}$  of (58), as shown at the bottom of this page consisting of  $N = 10$  data points of size  $D = 5$ . Next, the first of the  $N = 10$

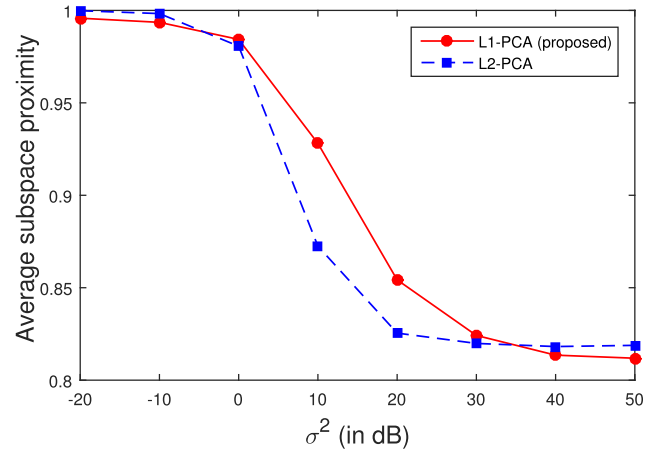


Fig. 7. Average subspace proximity versus corruption variance  $\sigma^2$ , for L2-PCA and L1-PCA (averages over 1000 realizations;  $D = 5$ ,  $N = 10$ ,  $K = 2$ ).

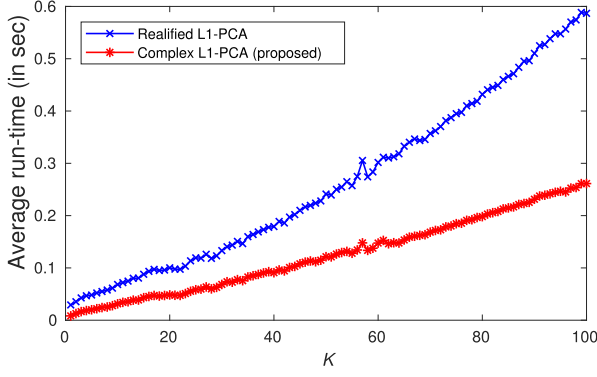
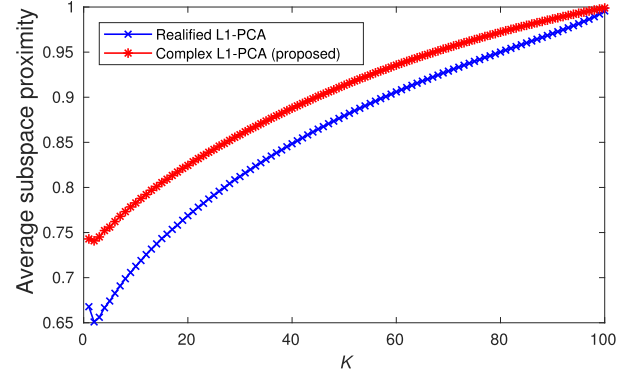
measurements in  $\mathbf{X}$  is additively corrupted by a random point  $\mathbf{c}$  drawn from  $\mathcal{CN}(\mathbf{0}_D, \sigma^2 \mathbf{I}_D)$ . Instead of  $\mathbf{X}$ , what is now available to us is its corrupted counterpart  $\mathbf{X}_{cor} = [\mathbf{x}_1 + \mathbf{c}, \mathbf{x}_2, \dots, \mathbf{x}_N]$ . We calculate  $K = 3$  L2-PCs of  $\mathbf{X}_{cor}$ ,  $\mathbf{Q}_{L2}$ , and  $K = 3$  L1-PCs of  $\mathbf{X}_{cor}$ ,  $\mathbf{Q}_{L1}$ . To quantify the corruption-resistance of the two subspaces, we measure their subspace proximity (SP) to the  $K = 3$  L2-PCA subspace  $\mathbf{Q}_{true}$  of the original clear data matrix  $\mathbf{X}$  by

$$SP(\mathbf{Q}, \mathbf{Q}_{true}) = \frac{1}{\sqrt{K}} \|\mathbf{Q}_{true}^H \mathbf{Q}\|_2 \in [0, 1], \quad (57)$$

for  $\mathbf{Q} = \mathbf{Q}_{L2}$  and  $\mathbf{Q} = \mathbf{Q}_{L1}$ .

In Fig. 7, we plot average SP (over 1000 independent corruption realizations) for L2-PCA and L1-PCA versus the corruption variance  $\sigma^2$ . We observe that for weak corruption of variance  $\sigma^2 < 0$  dB, L1-PCA and L2-PCA exhibit similar behavior with SP value near 1. We also notice that for very strong corruption of variance  $\sigma^2 > 35$  dB, L1-PCA and L2-PCA are similarly “misled” to SP of about 0.82. Interestingly, for all intermediate values of  $\sigma^2$ , L1-PCA exhibits significantly superior performance in calculating the nominal L2-PCA subspace. For example, for  $\sigma^2 = 10$  dB, L1-PCA is at 93% of SP, while L2-PCA is at 87% SP.

$$\mathbf{X} = \begin{bmatrix} -0.3003 - i 1.0117 & 0.4618 + i 0.0705 & -0.3924 - i 0.1602 & -0.5327 + i 0.1129 & 1.9368 - i 0.5685 \\ -0.3886 - i 0.6530 & -0.6204 - i 0.3556 & 0.7040 - i 1.3574 & 0.3315 + i 0.9675 & -1.5390 - i 0.8711 \\ -0.5961 + i 0.1708 & 0.6005 - i 1.8511 & -0.5541 - i 0.6086 & -0.4701 + i 0.3234 & 1.0896 + i 1.3071 \\ -0.0893 + i 0.1863 & -0.6031 + i 0.3869 & -0.7038 + i 0.0123 & 1.0782 + i 1.4440 & 0.9593 - i 0.9096 \\ -0.1678 + i 1.7097 & 0.5883 - i 0.7234 & -0.5185 - i 0.2924 & -0.3291 - i 1.7799 & -1.1252 - i 0.5569 \\ 0.2485 + i 0.6433 & -1.3913 - i 1.7947 & 0.1189 + i 0.1334 & 0.0509 - i 0.1326 & -1.2163 + i 0.4921 \\ 0.5302 - i 0.1632 & -0.9533 - i 0.3757 & 1.4074 - i 1.2147 & -0.4419 + i 0.8734 & -0.8092 - i 0.6724 \\ 0.0428 + i 0.6675 & -1.1010 + i 0.6750 & 0.6385 - i 0.7620 & 0.4554 + i 0.5840 & -0.7863 + i 1.2148 \\ -1.3608 + i 0.5011 & 1.0467 - i 0.1282 & 0.5043 + i 0.1808 & 0.2366 - i 0.8010 & 0.0459 - i 0.3441 \\ 0.5409 - i 0.7822 & 0.0075 - i 1.5285 & 1.4829 + i 0.9075 & -0.5216 - i 0.0030 & 0.8504 + i 0.8860 \end{bmatrix}^T. \quad (58)$$

Fig. 8. Run-time versus dimensionality  $K$  (averages over 100 experiments.)Fig. 9. Average proximity to nominal subspace versus dimensionality  $K$  (averages over 100 experiments.)

### C. Complex Versus Realified L1-PCA

We have proposed a method for solving the  $\mathcal{NP}$ -hard complex L1-PCA problem in (4). To waive the burden of  $\mathcal{NP}$ -hardness, one might consider to “realify” the problem by calculating optimally [20]

$$\mathbf{Q}_{\mathbb{R}}^{\text{opt}} = \underset{\mathbf{Q} \in \mathbb{R}^{2D \times 2K}; \mathbf{Q}^T \mathbf{Q} = \mathbf{I}_{2K}}{\text{argmax}} \|\mathbf{Q}^T \bar{\mathbf{X}}\|_1 \quad (59)$$

where  $\bar{\mathbf{X}}$  is the realified version of matrix  $\mathbf{X}$  defined by [36]

$$\bar{\mathbf{X}} \triangleq \begin{bmatrix} \Re\{\mathbf{X}\} & -\Im\{\mathbf{X}\} \\ \Im\{\mathbf{X}\} & \Re\{\mathbf{X}\} \end{bmatrix}. \quad (60)$$

Then, we can “complexify” [42], [43]  $\mathbf{Q}_{\mathbb{R}^{2D \times 2K}}^{\text{opt}}$  back to  $\mathbf{Q}_{L1 \times K}^{\text{realification}}$  by

$$\mathbf{Q}_{L1 \times K}^{\text{realification}} \triangleq [\mathbf{Q}_{\mathbb{R}}^{\text{opt}}]_{1:D, 1:2:2K} + j[\mathbf{Q}_{\mathbb{R}}^{\text{opt}}]_{D+1:2D, 1:2:2K} \quad (61)$$

to approximate the direct complex solution to (4) by Algorithm 1,  $\mathbf{Q}_{L1}$ .

To compare  $\mathbf{Q}_{L1}^{\text{realification}}$  against the proposed direct  $\mathbf{Q}_{L1}$  solution, we create an arbitrary nominal unitary matrix  $\mathbf{Q}_{\text{nom}} \in \mathbb{C}^{D \times K}$  that describes a  $K$ -dimensional subspace of the entire ( $D = 101$ )-dimensional space. We draw  $N = 1000$  random points  $\mathbf{x}_n$ ,  $n = 1, 2, \dots, 1000$ , according to  $\mathbf{x}_n \sim \mathcal{CN}(\mathbf{0}_{101 \times 1}, \mathbf{Q}_{\text{nom}} \mathbf{Q}_{\text{nom}}^H)$  and contaminate every column of  $\mathbf{X}_{D \times N} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$  by additive white Gaussian noise drawn from  $\mathcal{CN}(\mathbf{0}_{D \times 1}, \mathbf{I}_D)$ . In addition, 5% of the data set is also corrupted by noise from  $\mathcal{CN}(\mathbf{0}_{D \times 1}, 100\mathbf{I}_D)$ .

First, in Fig. 8, we compare the runtime required for realified versus direct L1-PCA of our data set as a function of  $K$ . We observe that the realified method is significantly slower than the direct (proposed) complex L1-PCA method. Next, in Fig. 9, we plot the average subspace proximity (calculated by means of (57)) of the two counterparts to the nominal subspace  $\mathbf{Q}_{\text{nom}}$  versus  $K$ . We observe that, across the board, direct complex L1-PCA by Algorithm 1 outperforms the optimally calculated by [20] realified alternative.

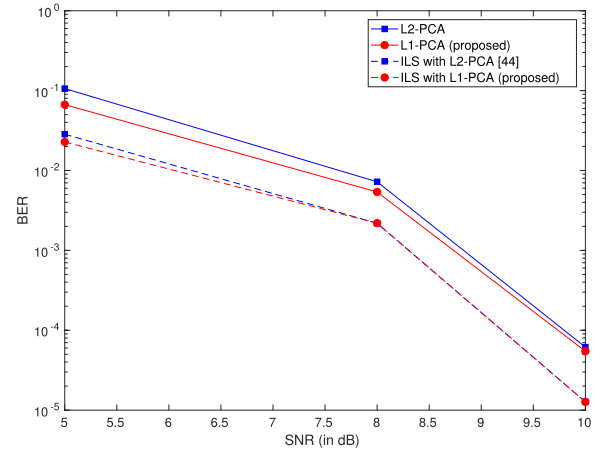


Fig. 10. BER versus estimated receive SNR (in dB) for the SDR-based channel estimation and symbol detection experiment.

### D. Radio Experiment: Channel Estimation and Symbol Detection by L1-PCA

We implement a single-input single-output (SISO) spread-spectrum wireless link with two Universal Software Radio Peripherals (USRP N210) that operate at  $f_c = 2.49$  GHz in an indoor laboratory multipath environment. We consider frame-based wireless transmissions of  $N = 1024$  binary antipodal information symbols using unknown complex-valued channel-processed spread-spectrum signatures of length  $L = 8$  chips. Channel-processed complex signatures at the receiver are estimated by complex L1 and L2-PCA on the  $L \times N$  complex received signal matrix. The channel-processed signature estimates are used for matched-filter (MF) symbol detection. Additionally, we implement and evaluate an iterative least-squares (ILS) type algorithm [44], [45] that conducts joint channel estimation and symbol detection, using L2-PCA (conventional) and L1-PCA (proposed) initialization.

We evaluate the L1 and conventional L2-PCA method in terms of bit-error rate (BER) performance for different transmit power and estimated receive signal-to-noise ratio (SNR) values in the fixed indoor radio setup. In Fig. 10, we plot their BER performance as a function of the receive SNR. We observe



that, most interestingly, L1-PCA performs somewhat better than L2-PCA both for MF and ILS. This may be explained by the presence in our data set of common measurement irregularities, such as occasional multipath reflections, phase noise, and clock shifts.

### E. Autonomous Channelization

Next, we present one more application example for complex L1-PCA from the field of wireless communications. We consider an existing system of  $K$  single-antenna primary sources using unknown complex-valued spread-spectrum signatures of length  $L$  chips. The signatures of the  $K$  sources are linearly independent (possibly orthogonal) so that they do not interfere with each other, spanning a  $K$ -dimensional subspace of  $\mathbb{C}^L$ . We assume now that  $L - K$  new secondary sources wish to access the same frequency spectrum using length- $L$  spread-spectrum signatures. Of course, the secondary sources should not interfere with the primary sources (or with themselves). Therefore, the secondary users will estimate the subspace spanned by the primary signatures and self-design signatures in its orthogonal complement.

1) *Training Phase:* We consider a collection of  $N$  snapshots from primary transmissions in the presence of additive white Gaussian noise (AWGN). To make the problem more challenging, while the snapshots are collected an unexpected strong interference source is also sporadically active. That is, the  $n$ -th recorded snapshot vector (after frequency down-conversion and pulse-matched filtering) has the form

$$\mathbf{x}(n) = \underbrace{\sum_{k=1}^K y_k(n) \mathbf{s}_k}_{\text{nominal}} + \underbrace{\mathbf{n}(n) + \gamma(n) \mathbf{i}(n)}_{\text{unexpected corruption}} \in \mathbb{C}^{L \times 1}, \quad (62)$$

$n = 1, 2, \dots, N$ , where  $y_k(n)$  accounts for the  $n$ -th amplitude-scaled information symbol transmitted by the  $k$ -th primary source over a flat-fading channel with  $E\{|y_k(n)|^2\} = 1$ ;  $\mathbf{s}_k \in \mathbb{C}^{L \times 1}$  is the signature of the  $k$ -th primary source with  $\|\mathbf{s}_k\|_2 = 1$  and  $\mathbf{s}_k^H \mathbf{s}_l = 0$  for  $k \neq l$ ;  $\mathbf{n}(n)$  is additive white Gaussian noise (AWGN), drawn from  $\mathcal{CN}(\mathbf{0}_{L \times 1}, \frac{1}{L} \mathbf{I}_L)$ ;  $\mathbf{i}(n)$  accounts for unexpected sporadic interference drawn from  $\mathcal{CN}(\mathbf{0}_{L \times 1}, \frac{100}{L} \mathbf{I}_L)$ ; and  $\{\gamma(n)\}_{n=1}^N$  are independent and identically distributed (i.i.d.)  $\{0, 1\}$ -Bernoulli( $\epsilon$ ) variables that indicate interference activity. That is, each snapshot is corrupted by an unexpected interference signal with probability  $\epsilon$ . According to the chosen values of symbol and noise variance, the primary users operate at signal-to-noise ratio of 0 dB.

In the training phase, the recorded snapshots are organized in the complex data matrix form  $\mathbf{X} \triangleq [\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(N)] \in \mathbb{C}^{L \times N}$ . From  $\mathbf{X}$ , we can estimate the  $K$ -dimensional primary-source transmission subspace  $\mathcal{S} \triangleq \text{span}(\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_K)$  as the span of  $\mathbf{Q}_{L2} = \arg\max_{\mathbf{Q} \in \mathbb{C}^{L \times K}; \mathbf{Q}^H \mathbf{Q} = \mathbf{I}_K} \|\mathbf{X}^H \mathbf{Q}\|_2$ . In view of the presented developments in this paper, we can also estimate  $\mathcal{S}$  by the span of  $\mathbf{Q}_{L1} = \arg\max_{\mathbf{Q} \in \mathbb{C}^{L \times K}; \mathbf{Q}^H \mathbf{Q} = \mathbf{I}_K} \|\mathbf{X}^H \mathbf{Q}\|_1$  with Algorithm 1. After  $\mathcal{S}$  is estimated, we select  $L - K$  orthogonal secondary signatures from the orthogonal complement of  $\text{span}(\mathbf{Q}_{L2})$  or

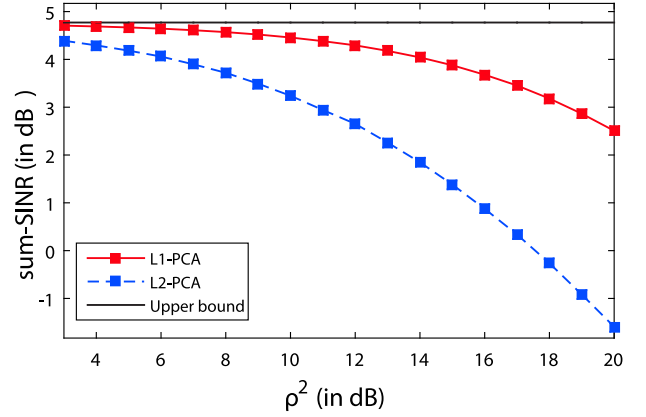


Fig. 11. Post-filtering sum-SINR versus transmission energy of secondary sources  $\rho^2$  (10000 realizations;  $L = 8$ ,  $K = 3$ ,  $N = 200$ ;  $\epsilon = 1.2\%$ ).

$\text{span}(\mathbf{Q}_{L1})$ . The secondary users employ these signatures and conduct transmissions concurrently with the primary sources.

2) *Operation Phase:* To evaluate the ability of L2/L1-PCA to identify the actual primary-source subspace (and thus enable secondary user operations that do not interfere with the primary sources), we measure and plot the post-filtering sum signal-to-interference-plus-noise ratio (SINR) of the primary users. If  $\mathbf{s}_k$ ,  $\|\mathbf{s}_k\|_2 = 1$ ,  $k = 1, 2, \dots, K$ , denote the signatures of the primary users and  $\mathbf{s}_l'$ ,  $\|\mathbf{s}_l'\|_2 = 1$ ,  $l = 1, 2, \dots, L - K$ , denote the self-designed signatures of the secondary users, the sum SINR of the primary users is  $\text{sum-SINR} = \sum_{k=1}^K \text{SINR}_k$ , where

$$\begin{aligned} \text{SINR}_k & \triangleq \frac{E\left\{|\mathbf{s}_k^H \mathbf{s}_k y_k(n)|^2\right\}}{E\left\{|\mathbf{s}_k^H \left(\sum_{m \neq k} \mathbf{s}_m y_m(n) + \sum_{l=1}^{L-K} \mathbf{s}_l' y_l'(n) + \mathbf{n}(n)\right)|^2\right\}} \\ & = \frac{1}{1 + \rho^2 \left|\sum_{l=1}^{L-K} \mathbf{s}_k^H \mathbf{s}_l'\right|^2} \end{aligned} \quad (63)$$

where  $\rho = E\{|y_l'(n)|^2\}$ ,  $l = 1, 2, \dots, L - K$ . Sum-SINR is a decreasing function of the transmission energy of the secondary sources,  $\rho^2$ . If  $\mathcal{S}$  were perfectly estimated and  $\{\mathbf{s}_l'\}_{l=1}^{L-K}$  were all designed in its orthogonal complement, then  $\mathbf{s}_k^H \mathbf{s}_l' = 0$  for all  $k, l$  and  $\text{SINR}_k$  would take its maximum value 1 (0 dB) independently of  $\rho$ .

In our numerical simulation, we set  $N = 200$ ,  $L = 8$ , and  $K = 3$ . In Fig. 11 we plot the average value of sum-SINR (calculated over 10000 independent experiments) versus  $\rho^2$  for snapshot-corruption probability  $\epsilon = 1.2\%$ . As a benchmark, we also plot the horizontal line of  $10 \log_{10}(K)$  dB, which corresponds to the sum-SINR if  $\mathcal{S}$  was accurately estimated (i.e.,  $\mathbf{s}_k^H \mathbf{s}_l' = 0$  for all  $k, l$ ). We observe that the L1-PCA-based design exhibits significantly superior performance compared to L2-PCA.

### F. Direction-of-Arrival Estimation

Direction-of-Arrival (DoA) estimation is a core operation in many applications, such as wireless node localization and network topology estimation. Super-resolution DoA estimation relies, traditionally, on L2-PCA of collected data as, e.g., in Multiple-Signal Classification (MUSIC) [46].

In this numerical study, we consider a receiver equipped with a uniform linear array (ULA) of  $D$  antenna elements that receives signals from  $K$  sources of interest located at angles  $\Theta = \{\theta_1, \theta_2, \dots, \theta_K\}$  with respect to the broadside. The inter-element spacing of the array is fixed at half the wavelength of the incoming signals and the array response vector for a signal that arrives at angle  $\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$  is

$$\mathbf{s}(\phi) \triangleq [1, e^{-j\pi \sin(\phi)}, \dots, e^{-j(D-1)\pi \sin(\phi)}]^T \in \mathbb{C}^{D \times 1}. \quad (64)$$

As in the signal model of (62), we assume that apart from the  $K$  sources of interest, there are potentially up to  $J$  unexpected, sporadically interfering sources (jammers) impinging on the ULA from angles  $\Theta' = \{\theta'_1, \theta'_2, \dots, \theta'_J\}$ . The  $n$ th collected data snapshot at the receiver is of the form

$$\mathbf{x}(n) = \underbrace{\sum_{k=1}^K y_k(n) \mathbf{s}(\theta_k)}_{\text{nominal}} + \underbrace{\sum_{j=1}^J \gamma_{n,j} y'_j(n) \mathbf{s}(\theta'_j)}_{\text{unexpected jamming}}, \quad (65)$$

$n = 1, 2, \dots, N$ . In (65),  $y_k(n)$  and  $y'_j(n)$  account for compound amplitudes of the  $k$ -th source and  $j$ -th jammer, respectively;  $\gamma_{n,j}$  is a  $\{0, 1\}$ -Bernoulli( $\epsilon$ ) activity indicator for jammer  $j$  at snapshot  $n$ ; and  $\mathbf{n}(n)$  is AWGN component drawn from  $\mathcal{CN}(\mathbf{0}_D, \sigma^2 \mathbf{I}_D)$ .

The popular MUSIC DoA estimation method collects all  $N$  snapshots in  $\mathbf{X} = [\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(N)]$  and calculates the source-signal subspace by the span of  $\mathbf{Q}_{L2} = \arg\max_{\{\mathbf{Q} \in \mathbb{C}^{D \times K}; \mathbf{Q}^H \mathbf{Q} = \mathbf{I}_K\}} \|\mathbf{X}^H \mathbf{Q}\|_2$ . DoA estimation is carried out by identifying the  $K$  highest peaks of the, so called, MUSIC spectrum

$$P(\phi; \mathbf{Q}_{L2}) = \|(\mathbf{I}_D - \mathbf{Q}_{L2} \mathbf{Q}_{L2}^H) \mathbf{s}(\phi)\|_2^{-1}. \quad (66)$$

In this experiment, we will replace L2-PCA by L1-PCA and carry out DoA estimation by identifying the  $K$  highest peaks of the L1-PCA spectrum  $P(\phi; \mathbf{Q}_{L1})$  where  $\mathbf{Q}_{L1} = \arg\max_{\{\mathbf{Q} \in \mathbb{C}^{D \times K}; \mathbf{Q}^H \mathbf{Q} = \mathbf{I}_K\}} \|\mathbf{X}^H \mathbf{Q}\|_1$ . We set  $D = 12$ ,  $K = 4$ ,  $J = 3$ , source DoAs  $\Theta = \{-40^\circ, -21^\circ, -7^\circ, 60^\circ\}$ , and jammer DoAs  $\Theta' = \{0^\circ, 20^\circ, 80^\circ\}$ . The number of snapshots available for DoA estimation,  $N$ , varies from 10 to 100. The  $K$  sources of interest operate at SNR of 0 dB and active jammers operate at SNR of 15 dB. We conduct 10000 independent DoA estimation experiments and evaluate the average DoA estimation performance of the L2-PCA-based and L1-PCA-based methods by means of the standard root-mean-squared error (RMSE) defined as

$$\text{RMSE} \triangleq \sqrt{\frac{1}{10000} \sum_{m=1}^{10000} \sum_{k=1}^K |\theta_k - \hat{\theta}_k(m)|^2} \quad (67)$$

where  $\hat{\theta}_k(m)$  is the estimate of  $\theta_k$  in the  $m$ -th experiment.

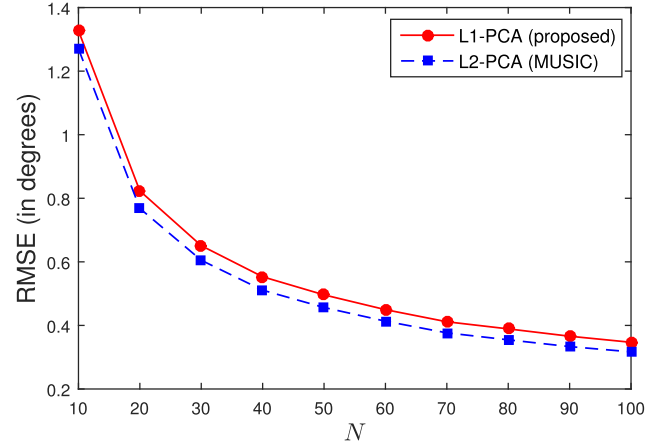


Fig. 12. RMSE versus data record size  $N$  in nominal operation ( $\epsilon = 0$ ); 10000 realizations;  $D = 12$ ,  $\Theta = \{-40^\circ, -21^\circ, -7^\circ, 60^\circ\}$ ,  $\Theta' = \{0^\circ, 20^\circ, 80^\circ\}$ .

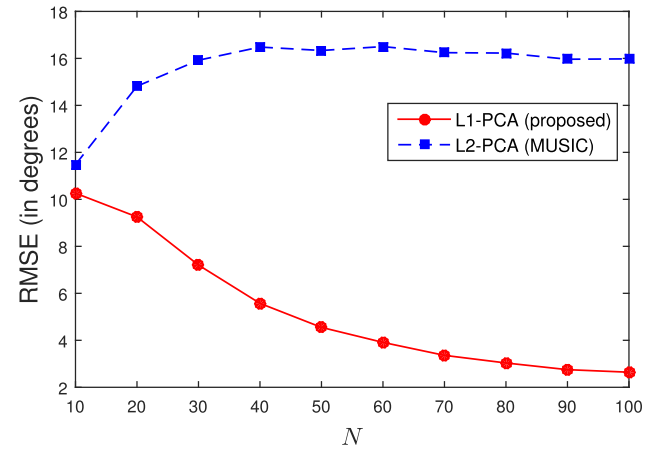


Fig. 13. RMSE versus data record size  $N$  in the presence of jammers ( $\epsilon = 2\%$ ); 10000 realizations;  $D = 12$ ,  $\Theta = \{-40^\circ, -21^\circ, -7^\circ, 60^\circ\}$ ,  $\Theta' = \{0^\circ, 20^\circ, 80^\circ\}$ .


In Fig. 12, we set  $\epsilon = 0$  (i.e., no unexpected jamming) and plot RMSE for L2-PCA and L1-PCA versus the number of snapshots,  $N$ . We observe that L2-PCA and L1-PCA have almost identical RMSE performance and of course improve as  $N$  increases. In Fig. 13, we set  $\epsilon$  to 2%. The performance of L2-PCA (standard MUSIC [46]) changes astoundingly and deteriorates as the sample-support  $N$  increases, converging to a plateau of  $\text{RMSE} = 16^\circ$ . Interestingly, L1-PCA resists the directional corruption by the jammers and as  $N$  increases attains decreasing RMSE approaching  $2^\circ$ .

### V. CONCLUSIONS

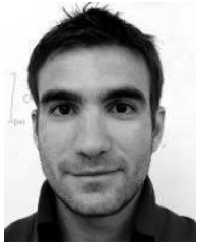
We showed that, in contrast to the real-valued case, complex L1-PCA is formally  $\mathcal{NP}$ -hard. We established that complex L1-PCA can be cast and solved through a unimodular-nuclear-norm maximization problem. We conducted optimality analysis and provided necessary conditions for global optimality. For the case of  $K = 1$  principal component, we provided necessary and sufficient conditions for local optimality. Based on the

optimality conditions, we presented the first two sub-optimal/iterative algorithms in the literature for complex L1-PCA. Finally, we presented extensive numerical and experimental studies from the fields of data analysis and wireless communications and observed that when the processed complex data are outlier-free, L1-PCA and L2-PCA perform very similarly. When the processed data are corrupted by faulty measurements, complex L1-PCA exhibits sturdy resistance significantly robustifying PCA-driven solutions.

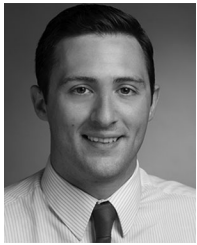
## REFERENCES

- [1] N. Tsagkarakis, P. P. Markopoulos, and D. A. Pados, "Direction finding by complex L1-principal component analysis," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun.*, Stockholm, Sweden, Jun. 2015, pp. 475–479.
- [2] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosoph. Mag.*, vol. 2, pp. 559–572, 1901.
- [3] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Physcomerika*, vol. 1, pp. 211–218, Sep. 1936.
- [4] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*. Philadelphia, PA, USA: SIAM, 2000.
- [5] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 1996.
- [6] M. Brand, "Fast low-rank modifications of the thin singular value decomposition," *Linear Algebra Appl.*, vol. 415, pp. 20–30, May 2006.
- [7] N. Srebro and T. Jaakkola, "Weighted low-rank approximations," in *Proc. Int. Conf. Mach. Learn.*, Washington DC, USA, Aug. 2003, pp. 720–727.
- [8] P. Anandan and M. Irani, "Factorization with uncertainty," *Int. J. Comput. Vis.*, vol. 49, pp. 101–116, Sep. 2002.
- [9] F. De La Torre and M. Black, "A framework for robust subspace learning," *Int. J. Comput. Vis.*, vol. 54, pp. 117–142, Aug. 2003.
- [10] G. Mateos and G. B. Giannakis, "Robust PCA as bilinear decomposition with outlier-sparsity regularization," *IEEE Trans. Signal Process.*, vol. 60, pp. 5176–5190, Oct. 2012.
- [11] R. R. Singleton, "A method for minimizing the sum of absolute values of deviations," *Ann. Math. Statist.*, vol. 11, pp. 301–310, Sep. 1940.
- [12] I. Barrodale, "L1 approximation and the analysis of data," *J. Royal Statist. Soc., Appl. Statist.*, vol. 17, pp. 51–57, 1968.
- [13]  and T. Kanade, "Robust subspace computation using L1 norm," *Int. Rep. Comput. Sci. Dept.*, Carnegie Mellon Univ., Pittsburgh, PA, USA, Rep. CMU-CS-03172, Aug. 2003.
- [14] J. P. Brooks, J. H. Dula, and E. L. Boone, "A pure L1-norm principal component analysis," *J. Comput. Statist. Data Anal.*, vol. 61, pp. 83–98, May 2013.
- [15] N. Tsagkarakis, P. P. Markopoulos, and D. A. Pados, "On the L1-norm approximation of a matrix by another of lower rank," in *Proc. IEEE Int. Conf. Mach. Learn. App.*, Anaheim, CA, USA, Dec. 2016, pp. 768–773.
- [16] N. Kwak, "Principal component analysis based on L1-norm maximization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, pp. 1672–1680, Sep. 2008.
- [17] F. Nie, H. Huang, C. Ding, D. Luo, and H. Wang, "Robust principal component analysis with non-greedy L1-norm, maximization," in *Proc. Int. Joint Conf. Artif. Intell.*, Barcelona, Spain, Jul. 2011, pp. 1433–1438.
- [18] M. McCoy and J. A. Tropp, "Two proposals for robust PCA using semidefinite programming," *Electron. J. Statist.*, vol. 5, pp. 1123–1160, Jun. 2011.
- [19] C. Ding, D. Zhou, X. He, and H. Zha, " $R_1$ -PCA: Rotational invariant L1-norm principal component analysis for robust subspace factorization," in *Proc. Int. Conf. Mach. Learn.*, Pittsburgh, PA, USA, Jun. 2006, pp. 281–288.
- [20] P. P. Markopoulos, G. N. Karystinos, and D. A. Pados, "Optimal algorithms for L1-subspace signal processing," *IEEE Trans. Signal Process.*, vol. 62, pp. 5046–5058, Oct. 2014.
- [21] P. P. Markopoulos, G. N. Karystinos, and D. A. Pados, "Some options for L1-subspace signal processing," in *Proc. Int. Symp. Wireless Commun. Syst.*, Ilmenau, Germany, Aug. 2013, pp. 622–626.
- [22] S. Kundu, P. P. Markopoulos, and D. A. Pados, "Fast computation of the L1-principal component of real-valued data," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Florence, Italy, May 2014, pp. 8028–8032.
- [23] P. P. Markopoulos, S. Kundu, S. Chamadia, and D. A. Pados, "Efficient L1-norm principal-component analysis via bit flipping," *IEEE Trans. Signal Process.*, vol. 65, pp. 4252–4264, Aug. 2017.
- [24] P. P. Markopoulos, S. Kundu, S. Chamadia, and D. A. Pados, "L1-norm principal-component analysis via bit flipping," in *Proc. IEEE Int. Conf. Mach. Learn. Appl.*, Anaheim, CA, USA, Dec. 2016, pp. 326–332.
- [25] P. P. Markopoulos, D. A. Pados, G. N. Karystinos, and M. Langberg, "L1-norm principal-component analysis in L2-norm-reduced-rank data subspaces," in *Proc. SPIE Defense Commercial Sens.*, Anaheim, CA, USA, Apr. 2017, pp. 1021104:1–1021104:10.
- [26] S. Chamadia and D. A. Pados, "Optimal sparse L1-norm Principal-Component analysis," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, New Orleans, LA, USA, Mar. 2017, pp. 2686–2690.
- [27] P. P. Markopoulos, S. Kundu, and D. A. Pados, "L1-fusion: Robust linear-time image recovery from few severely corrupted copies," in *Proc. IEEE Int. Conf. Image Process.*, Quebec City, Canada, Sep. 2015, pp. 1225–1229.
- [28] M. Johnson and A. Savakis, "Fast L1-eigenfaces for robust face recognition," in *Proc. IEEE West. New York Image Signal Process. Workshop*, Rochester, NY, USA, Nov. 2014, pp. 1–5.
- [29] F. Maritato, Y. Liu, S. Colonnese, and D. A. Pados, "Cloud-assisted individual L1-PCA face recognition using wavelet-domain compressed images," in *Proc. Eur. Workshop Visual Inf. Process.*, Marseilles, France, Oct. 2016, pp. 1–6.
- [30] F. Maritato, Y. Liu, S. Colonnese, and D. A. Pados, "Face recognition with L1-norm subspaces," in *Proc. SPIE Defense Commercial Sens.*, Baltimore, MD, USA, Apr. 2016, pp. 98570K:1–98570K:8.
- [31] M. Pierantozzi, Y. Liu, D. A. Pados, and S. Colonnese, "Video background tracking and foreground extraction via L1-subspace updates," in *Proc. SPIE Defense Commercial Sens.*, Baltimore, MD, USA, Apr. 2015, pp. 985708:1–985708:16.
- [32] Y. Liu and D. A. Pados, "Compressed-sensed-domain L1-PCA video surveillance," *IEEE Trans. Multimedia*, vol. 18, pp. 351–63, Mar. 2016.
- [33] D. G. Chachlakis, P. P. Markopoulos, R. J. Muchhala, and A. Savakis, "Visual tracking with L1-Grassmann manifold modeling," in *Proc. SPIE Defense Commercial Sens.*, Anaheim, CA, USA, Apr. 2017, pp. 1021102:1–1021102:12.
- [34] P. P. Markopoulos, "Reduced-rank filtering on L1-norm subspaces," in *Proc. IEEE Sens. Array Multichannel Signal Process. Workshop*, Rio de Janeiro, Brazil, Jul. 2016, pp. 1–5.
- [35] P. P. Markopoulos and F. Ahmad, "Indoor human motion classification by L1-norm subspaces of micro-Doppler signatures," in *Proc. IEEE Radar Conf.*, Seattle, WA, USA, May 2017, pp. 1807–1810.
- [36] P. P. Markopoulos, N. Tsagkarakis, D. A. Pados, and G. N. Karystinos, "Direction finding with L1-norm subspaces," in *Proc. SPIE Defense, Security, Sens.*, Baltimore, MD, USA, May 2014, pp. 91090J:1–91090J:11.
- [37] P. H. Schönemann, "A generalized solution of the orthogonal procrustes problem," *Psychometrika*, vol. 31, pp. 1–10, Mar. 1966.
- [38] N. J. Higham, "Computing the polar decomposition—with applications," *SIAM J. Sci. Statist. Comput.*, vol. 7, pp. 1160–1174, Oct. 1986.
- [39] M. Soltanalian and P. Stoica, "Designing unimodular codes via quadratic optimization," *IEEE Trans. Signal Process.*, vol. 62, pp. 1221–1234, Mar. 2014.
- [40] G. N. Karystinos and D. A. Pados, "Rank-2-optimal adaptive design of binary spreading codes," *IEEE Trans. Inf. Theory*, vol. 53, pp. 3075–3080, Sep. 2007.
- [41] G. N. Karystinos and A. P. Liavas, "Efficient computation of the binary vector that maximizes a rank-deficient quadratic form," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3581–3593, Jul. 2010.
- [42] P. A. Businger and G. H. Golub, "Algorithm 358: Singular value decomposition of a complex matrix  $[F_1, 4, 5]$ ," *Commun. ACM*, vol. 12, pp. 564–565, 1969.
- [43] A. Bunse-Gerstner and W. B. Gragg, "Singular value decompositions of complex symmetric matrices," *J. Comput. Appl. Math.*, vol. 21, pp. 41–54, 1988.
- [44] M. Li, S. N. Batalama, D. A. Pados, T. Melodia, M. J. Medley, and J. D. Matyjas, "Cognitive code-division links with blind primary-system identification," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 3743–3753, Nov. 2011.
- [45] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Kyoto, Japan, Mar. 2012, pp. 3249–3252.
- [46] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, pp. 276–280, Mar. 1986.





**Nicholas Tsagkarakis** received the Diploma degree in electronic and computer engineering from the Technical University of Crete, Chania, Greece, in 2011 and the Ph.D. degree in electrical engineering from The State University of New York at Buffalo, Buffalo, NY, USA, in 2017. His research interests include the areas of adaptive signal processing, wireless communications, and data analysis.



**Panos P. Markopoulos** (S'11–M'15) was born in Athens, Greece, in 1986. He received the Diploma and M.S. degrees in electronic and computer engineering from the Technical University of Crete, Chania, Greece, in 2010 and 2012, respectively, and the Ph.D. degree in electrical engineering from The State University of New York at Buffalo, Buffalo, NY, USA, in 2015. Since 2015, he has been an Assistant Professor with the Department of Electrical and Microelectronic Engineering, Rochester Institute of Technology, Rochester, NY, USA, where he directs

the Signal Processing for Data Communication and Analysis Research Group. His research interests include signal processing, communications, data analysis, and machine learning.

Dr. Markopoulos is a member of the IEEE Communications, IEEE Signal Processing, and IEEE Computer Societies. Since September 2017, he is Editor of the IEEE Wireless Communication Letters. In 2016, he received the Exemplary Reviewer distinction from the IEEE Wireless Communications Letters. In 2013, he received the Best Paper Award in Physical Layer Communications and Signal Processing of the 2013 International Symposium on Wireless Communication Systems. He serves regularly as Reviewer to the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON MULTIMEDIA, and many IEEE conferences. In 2017 and 2018, he was a co-Organizer and the co-Chairman of the IEEE International Workshop on Wireless Communications and Networking in Extreme Environments. In 2018, he was co-Organizer and co-Chairman of the IEEE GlobalSIP Symposium on Tensor Methods for Signal Processing and Machine Learning.



**George Sklivanitis** (S'11–M'18) received the Diploma degree in electronic and computer engineering from the Technical University of Crete, Chania, Greece, in 2010, and the Ph.D. degree in electrical engineering from The State University of New York at Buffalo, Buffalo, NY, USA, in 2018. He is currently a Research Assistant Professor with the Department of Computer and Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL, USA. His research interests include the areas of signal processing, software-defined wireless

communications and networking, cognitive radio, and underwater acoustic communications. In 2014, he was the first finalist and was a recipient of the 2014 Nutaq Software-Defined Radio Academic U.S. National Contest and in 2015 the 10th ACM International Conference on Underwater Networks and Systems Best Demo Award. He was also a recipient of the 2015 SUNY Buffalo Graduate Student Award for Excellence in Teaching, the 2016 SUNY Buffalo Student Entrepreneur Fellowship, and the 2017 SUNY Chancellors Award for Student Excellence.



**Dimitris A. Pados** (M'95–SM'15) was born in Athens, Greece, on October 22, 1966. He received the Diploma degree in computer science and engineering (five-year program) from the University of Patras, Patras, Greece, in 1989 and the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, VA, USA, in 1994.

From 1994 to 1997, he was an Assistant Professor with the Department of Electrical and Computer Engineering and the Center for Telecommunications Studies, University of Louisiana, Lafayette, LA, USA. From 1997 to 2017, he was with the Department of Electrical Engineering, The State University of New York at Buffalo, holding in sequence the titles of Assistant Professor, Associate Professor, Professor, and Clifford C. Furnas Chair Professor of Electrical Engineering. He was also appointed Associate Chair and Chair of the Department of Electrical Engineering. He was elected University Faculty Senator four times and served on the Faculty Senate Executive Committee for two terms. In August 2017, he joined Florida Atlantic University, Boca Raton, FL, USA, where he currently leads the university initiative on autonomous systems and is a Fellow of the I-SENSE research institute and Professor of Computer and Electrical Engineering and Computer Science. His research interests include the general areas of data and signal processing and communications theory.

Dr. Pados is a member of the IEEE Communications, IEEE Signal Processing, IEEE Information Theory, and IEEE Computational Intelligence Societies. He was an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS and the IEEE TRANSACTIONS ON NEURAL NETWORKS. He was a recipient of a number of awards for articles that he coauthored with his students including the 2001 IEEE International Conference on Telecommunications Best Paper Award, the 2003 IEEE Transactions on Neural Networks Outstanding Paper Award, the 2010 IEEE International Communications Conference Best Paper Award in Signal Processing for Communications, the 2013 International Symposium on Wireless Communication Systems Best Paper Award in Physical Layer Communications and Signal Processing, Best of IEEE GLOBECOM 2014-Top 50 Papers distinction, and the Best Paper Selection distinction in the 2016 IEEE International Conference on Multimedia Big Data. He was also a recipient of the 2009 SUNY-wide Chancellor's Award for Excellence in Teaching and the 2011 University at Buffalo Exceptional Scholar - Sustained Achievement Award.