Some Options for $L_1$-subspace Signal Processing∗

Panos P. Markopoulos
Electrical Engineering Dept.
State University of New York at Buffalo
Buffalo, NY 14260
Email: pmarkopo@buffalo.edu

George N. Karystinos
Electronic and Computer Engineering Dept.
Technical University of Crete
Chania, 73100, Greece
Email: karystinos@telecom.tuc.gr

Dimitris A. Pados†
Electrical Engineering Dept.
State University of New York at Buffalo
Buffalo, NY 14260
Email: pados@buffalo.edu

Abstract—We describe ways to define and calculate $L_1$-norm signal subspaces which are less sensitive to outlying data than $L_2$-calculated subspaces. We focus on the computation of the $L_1$ maximum-projection principal component of a data matrix containing $N$ signal samples of dimension $D$ and conclude that the general problem is formally NP-hard in asymptotically large $N, D$. We prove, however, that the case of engineering interest of fixed dimension $D$ and asymptotically large sample support $N$ is not and we present an optimal algorithm of complexity $O(N D^2)$.

I. INTRODUCTION

Subspace signal processing theory and practice rely, conventionally, on the familiar $L_2$-norm based singular-value decomposition (SVD) of the data matrix. The SVD solution traces its origin to the fundamental problem of $L_2$-norm low-rank matrix approximation, which is equivalent to the problem of maximum $L_2$-norm orthonormal data projection with as many projection (“principal”) components as the desired low-rank value [1]. Practitioners have long observed, however, that $L_2$-norm principal component analysis (PCA) is sensitive to the presence of outlier values in the data matrix, that is, values that are away from the nominal distribution data, appear only few times in the data matrix, and are not to appear again under normal system operation upon design. This paper makes a case for $L_1$-subspace signal processing. Interestingly, in contrast to $L_2$, subspace decomposition under the $L_1$ error minimization criterion and the $L_1$ projection maximization criterion are not the same. A line of recent research pursues calculation of $L_1$ principal components under error minimization [2] or projection maximization [3], [4]. No algorithm has appeared so far with guaranteed convergence to the criterion-optimal subspace and no upper bounds are known on the expended computational effort.

In this present work, given any data matrix $X \in \mathbb{R}^{D \times N}$ of $N$ signal samples of dimension $D$, we show that the general problem of finding the maximum $L_1$-projection principal component of $X$ is formally NP-hard for asymptotically large $N, D$. We prove, however, that the case of engineering interest of fixed given dimension $D$ is not NP-hard. In particular, for the case where $N < D$, we present in explicit form an algorithm to find the optimal component with computational cost $2^N$. For the case where the sample support exceeds the data dimension ($N \geq D$) –which is arguably of more interest in signal processing applications– we present an algorithm that computes the $L_1$-optimal principal component with complexity $O(\text{rank}(X))$, and asymptotically $N \leq D$. We generalize the effort to the problem of calculating multiple $L_1$ components (necessarily a joint computational problem) and present an explicit optimal algorithm for multi-component subspace design in the form of nuclear-norm maximization.

II. PROBLEM STATEMENT

Consider $N$ real-valued measurements $x_1, x_2, \ldots, x_N$ of dimension $D$ that form the $D \times N$ data matrix

$$X = [x_1 \ x_2 \ \ldots \ x_N].$$  \hspace{1cm} (1)

We are interested in describing (approximating) the data matrix $X$ by a rank-$K$ product $RS^T$ where $R \in \mathbb{R}^{D \times K}, S \in \mathbb{R}^{N \times K}$, $K \leq D$, in the form of Problem $P_1^{L_2}$ defined below,

$$P_1^{L_2} : \ (R_{L_2}, S_{L_2}) = \arg\min_{R \in \mathbb{R}^{D \times K}, \ S \in \mathbb{R}^{N \times K}} \|X - RS^T\|_2$$  \hspace{1cm} (2)

where $\|A\|_2 = \sqrt{\sum_{i,j}|A_{i,j}|^2}$ is the $L_2$ matrix norm (Frobenius) of matrix $A$ with elements $A_{i,j}$. By the Projection Theorem [1], $S = X^T R$ for any fixed $R$, $R^T R = I_K$. Hence, we obtain the equivalent problem

$$P_2^{L_2} : \ R_{L_2} = \arg\min_{R \in \mathbb{R}^{D \times K}} \|X - RR^T X\|_2$$  \hspace{1cm} (3)

frequently referred to as left-side $K$-SVD. Since $\|A\|_2^2 = \text{tr}(A^T A)$ where tr(·) denotes the trace of a matrix, $P_2^{L_2}$ is also equivalent to $L_2$ projection (energy) maximization,

$$P_3^{L_2} : \ R_{L_2} = \arg\max_{R \in \mathbb{R}^{D \times K}} \|X^T R\|_2.$$  \hspace{1cm} (4)

Note that, if $K < D$ and we possess the solution $R_{L_2}^{(K)}$ for $K$ singular/eigen-vectors in (2), (3), (4), then the solution for rank $K + 1$ is derived readily by $R_{L_2}^{(K+1)} = [R_{L_2}^{(K)} \ r_{L_2}^{(K+1)}]$ with

$$r_{L_2}^{(K+1)} = \arg\max_{r \in \mathbb{R}^D, \ \|r\|_2 = 1} \|X^T (I_D - R_{L_2}^{(K)} R_{L_2}^{(K)}) r\|_2.$$  \hspace{1cm} (5)

This is known as the PCA scalability property.

By minimizing the sum of squared errors, $L_2$ principal component calculation becomes sensitive to extreme error value occurrences caused by the presence of outlier measurements.
in the data matrix. Motivated by this observed drawback of $L_2$ subspace signal processing, in this work we study and pursue subspace-decomposition approaches that are based on the $L_1$ norm, $\|A\|_1 = \sum_{i,j} |A_{i,j}|$. We may “translate” the three equivalent $L_2$ optimization problems (2), (3), (4) to new problems that utilize the $L_1$ norm as follows,

$$\mathcal{P}_{L_1}^1 : (R_{L_1}, S_{L_1}) = \arg \min_{R \in \mathbb{R}^{D \times K}, R^T R = I_K} \|X - RS^T\|_1, \tag{5}$$

$$\mathcal{P}_{L_1}^2 : R_{L_1} = \arg \min_{R \in \mathbb{R}^{D \times K}, R^T R = I_K} \|X - RR^T X\|_1, \tag{6}$$

$$\mathcal{P}_{L_1}^3 : R_{L_1} = \arg \max_{R \in \mathbb{R}^{D \times K}, R^T R = I_K} \|X^T R\|_1. \tag{7}$$

A few comments appear useful at this point: (i) Under the $L_1$ norm, the three optimization problems $\mathcal{P}_{L_1}^1$, $\mathcal{P}_{L_1}^2$, and $\mathcal{P}_{L_1}^3$ are no longer equivalent. (ii) Under $L_1$, the PCA scalability property does not hold (due to loss of the Projection Theorem). (iii) Even for reduction to a single dimension (rank $K = 1$ approximation), the three problems are difficult to solve.

In this present work, we focus exclusively on $\mathcal{P}_{L_1}^3$.

III. THE $L_1$-NORM PRINCIPAL COMPONENT

In this section, we concentrate on the calculation of the $L_1$-maximum-projection component of a data matrix $X \in \mathbb{R}^{D \times N}$ (Problem $\mathcal{P}_{L_1}^3$ in (7), $K = 1$). First, we show that the problem is in general NP-hard and review briefly suboptimal techniques from the literature. Then, we prove that, if the data dimension $D$ is fixed, the principal $L_1$-norm component of $X$ is in fact computable in polynomial time and present a calculation algorithm with complexity $O\left(\frac{N^{\text{rank}(X)}}{\text{rank}(X)}\right)$, rank($X$) $\leq D$.

A. The Hardness of the Problem and an Exhaustive-search Algorithm Over the Binary Field

In Proposition 1 below, we present a fundamental property of Problem $\mathcal{P}_{L_1}^3$, $K = 1$, that will lead us to an efficient solution. The proof is omitted due to lack of space and can be found in [6].

**Proposition 1:** For any data matrix $X \in \mathbb{R}^{D \times N}$, the solution to $\mathcal{P}_{L_1}^3 : r_{L_1} = \arg \max_{r \in \mathbb{R}^{D}, \|r\|_1 = 1} \|X^T r\|_1$ is given by

$$r_{L_1} = \frac{Xb_{\text{opt}}}{\|Xb_{\text{opt}}\|_2}, \tag{8}$$

where

$$b_{\text{opt}} = \arg \max_{b \in \{\pm 1\}^N} \|Xb\|_2 = \arg \max_{b \in \{\pm 1\}^N} b^T X^T Xb. \tag{9}$$

In addition, $\|X^T r_{L_1}\|_1 = \|Xb_{\text{opt}}\|_2$. \hfill $\Box$

The straightforward approach to solve (9) is an exhaustive search among all $2^N$ binary vectors of length $N$. Proposition 2 below declares that, indeed, in its general form $\mathcal{P}_{L_1}^3$, $K = 1$, is NP-hard for jointly asymptotically large $N,D$. The proof can be found in [6].

**Proposition 2:** Computation of the $L_1$ principal component of $X \in \mathbb{R}^{D \times N}$ by maximum $L_1$-norm projection (Problem $\mathcal{P}_{L_1}^3$, $K = 1$) is NP-hard in jointly asymptotic $N,D$. \hfill $\Box$

B. Existing Approaches in Literature

There has been a growing documented effort to calculate subspace components by $L_1$ projection maximization [3], [4]. For $K = 1$, both algorithms in [3], [4] are identical and can be described by the simple single iteration

$$b_{(i+1)} = \text{sgn}(X^T X b_{(i)}), \quad i = 1, 2, \ldots, \tag{10}$$

for the computation of $b_{\text{opt}}$ in (9). Equation (10), however, does not guarantee convergence to the $L_1$-optimal component solution (convergence to one of the many local maxima may be observed). In the following section, we present for the first time in the literature an optimal algorithm to calculate the $L_1$ principal component of a data matrix with complexity polynomial in the sample support $N$ when the data dimension $D$ is fixed.

C. Computation of the $L_1$ Principal Component in Polynomial Time

In the following, we show that, if $D$ is fixed, then computation of $r_{L_1}$ is no longer NP-hard (in $N$). We state our result in the form of Proposition 3 below.

**Proposition 3:** For any fixed data dimension $D$, computation of the $L_1$ principal component of $X \in \mathbb{R}^{D \times N}$ has complexity $O\left(\frac{N^{\text{rank}(X)}}{\text{rank}(X)}\right)$, rank($X$) $\leq D$. \hfill $\Box$

By Proposition 2, computation of the $L_1$ principal component of $X$ is equivalent to computation of $b_{\text{opt}}$ in (9). To prove Proposition 3, we will then prove that $b_{\text{opt}}$ can be computed with complexity $O\left(\frac{N^{\text{rank}(X)}}{\text{rank}(X)}\right)$. We begin our developments by defining

$$d \triangleq \text{rank}(X) \leq D. \tag{11}$$

Then, $X^T X$ has also rank $d$ and can be decomposed by

$$X^T X = QQ^T, \quad Q_{N \times d} = [q_1, q_2, \ldots, q_d], \quad q_i^T q_i = 0, \quad i \neq j, \tag{12}$$

where $q_1, q_2, \ldots, q_d$ are the $d$ eigenvalue-weighted eigenvectors of $X^T X$ with nonzero eigenvalue. By (9),

$$b_{\text{opt}} = \arg \max_{b \in \{\pm 1\}^N} b^T QQ^T b = \arg \max_{b \in \{\pm 1\}^N} \|Q^T b\|_2. \tag{13}$$

For the case $N < D$, the optimal binary vector $b_{\text{opt}}$ can be obtained directly from (13) by an exhaustive search among all $2^N$ binary vectors $b \in \{\pm 1\}^N$. Therefore, we can design the $L_1$-optimal principal component $r_{L_1}$ with computational cost $2^N < 2^D = O(1)$. For the case where the sample support exceeds the data dimension ($N \geq D$) - which is arguably of higher interest in signal processing applications - we find it useful in terms of both theory and practice to present our developments separately for data rank $d = 1, d = 2$, and $2 < d \leq D$.

1) Case $d = 1$: If the data matrix has rank $d = 1$, then $Q = q_1$ and (13) becomes

$$b_{\text{opt}} = \arg \max_{b \in \{\pm 1\}^N} |q_1^T b| = \text{sgn}(q_1). \tag{14}$$
By (8), the $L_1$-optimal principal component is
\[ r_{L_1} = \frac{X \text{sgn}(q_1)}{\|X \text{sgn}(q_1)\|_2} \] (15)
designed with complexity $O(N)$. It is of notable practical importance to observe at this point that even when $X$ is not of true rank one, (15) presents us with a quality, trivially calculated approach of the $L_1$ principal component of $X$. Calculate the $L_2$ principal component $q_1$ of the $N \times N$ matrix $X^T X$, quantize to $\text{sgn}(q_1)$, and project and normalize to obtain $r_{L_1} \simeq X \text{sgn}(q_1)/\|X \text{sgn}(q_1)\|_2$.

2) Case $d = 2$: If $d = 2$, then $Q = [q_1 q_2]$ and (13) becomes
\[ b_{opt} = \arg \max_{b \in \{\pm 1\}^N} \left\{ (q_1^T b)^2 + (q_2^T b)^2 \right\} . \] (16)

The binary optimization problem (16) was seen and solved for the first time in [7] by the auxiliary-angle method [8] with complexity $O(N \log N)$. Due to lack of space, we omit the specifics of the Case $d = 2$ and move directly to the general case $2 \leq d \leq D$.

3) Case $2 \leq d \leq D$: If $d \geq 2$, we design the $L_1$-optimal principal component of $X$ with complexity $O(N^d)$ by considering the multiple-auxiliary-angle approach that was presented in [9] as a generalization of the work in [7].

Consider a unit vector $c \in \mathbb{R}^d$. By Cauchy-Schwarz, for any $a \in \mathbb{R}^d$,
\[ a^T c \leq \|a\|_2 \|c\|_2 = \|a\|_2 \] (17)
with equality if and only if $c$ is codirectional with $a$. Then,
\[ e \in \mathbb{R}^d, \|e\|_1 = 1 : a^T c = \|a\|_2 . \] (18)

By (18), the optimization problem in (13) becomes
\[ \max_{b \in \{\pm 1\}^N} \|Q^T b\|_2 = \max_{b \in \{\pm 1\}^N} \max_{e \in \mathbb{R}^d, \|e\|_1 = 1} b^T Qc = \max_{e \in \mathbb{R}^d, \|e\|_1 = 1} b^T Qc . \] (19)

For every $c \in \mathbb{R}^d$, inner maximization in (19) is solved by the binary vector
\[ b(c) = \text{sgn}(Qc) . \] (20)

which is obtained with complexity $O(N)$. Then, by (19), the solution to the original problem in (13) is met if we collect all binary vectors $b(c)$ returned as $c$ scans the unit-radius $d$-dimensional hypersphere. That is, $b_{opt}$ in (13) is in $^{2}$
\[ S = \bigcup_{e \in \mathbb{R}^d, \|e\|_1 = 1, c \geq 0} b(c) . \] (21)

Two fundamental questions for the computational problem under consideration are what the size (cardinality) of set $S$ is and how much computational effort is expended to form $S$.

The candidate vector set $S$ has cardinality $|S| = \sum_{g=0}^{d-1} \binom{N-1}{g} = O(N^{d-1})$ and it suffices to solve
\[ Q_{T,c} = 0 \] (22)
for every $I \subset \{1, 2, \ldots, N\}$, $|I| = d - 1$ (i.e., $Q_{T,c}$ contains any $d - 1$ rows of $Q$). The solution to (22) is the unit vector in the null space of the $(d - 1) \times d$ matrix $Q_{T,c}$. Then, the binary vectors $b$ of interest are obtained by
\[ \text{sgn}(Qc) \] (23)
with complexity $O(N)$. Note that (23) presents ambiguity regarding the sign of the intersecting $d - 1$ hypersurfaces (zero values). A straightforward way to resolve the ambiguity is to consider all $d^{d-1}$ sign combinations for the $d - 1$ zero value positions. Since complexity $O(N)$ is required to solve (23) for each subset of $d - 1$ rows of $Q$, the overall complexity of the construction of $S$ is $O(N^d)$ for any given matrix $Q_{N \times d}$. Our complete, new algorithm for the computation of the $L_1$-optimal principal component of a rank-$d$ matrix $X \in \mathbb{R}^{d \times N}$ that has complexity $O(N^d)$ is presented in detail in Fig. 1.

IV. MULTIPLE $L_1$-NORM PRINCIPAL COMPONENTS

In this section, we switch our interest to the joint design of $K > 1$ principal $L_1$ components of a $D \times N$ matrix $X$.

A. Existing Approaches in Literature

For the case $K > 1$, [3] proposed to design the first $L_1$ principal component $r_{L_1}$ by the coupled iteration (10) (which does not guarantee optimality) and then project the data onto the subspace that is orthogonal to $r_{L_1}$, design the $L_1$ principal component of the projected data by the same coupled iteration, and continue similarly. To avoid the above suboptimal greedy approach, [4] presented an iterative algorithm for the computation of $r_{L_1}$ altogether (that is the joint computation of the $K$ principal $L_1$ components), which does not guarantee convergence to the $L_1$-optimal subspace.

B. Exact Computation of Multiple $L_1$ Principal Components

For any $D \times K$ matrix $A$,
\[ \max_{R \in \mathbb{R}^{D \times K}, R^T R = I_K} \text{tr} (R^T A) = \|A\|_* \] (24)
where $\|A\|_*$ denotes the nuclear norm (i.e., the sum of the singular values) of $A$. Maximization in (24) is achieved by $R = UV^T$ where $U \Sigma V^T$ is the “compact” SVD of $A$, $U$ and $V$ are $D \times d$ and $K \times d$, respectively, matrices with $U^T U = V^T V = I_d$, $\Sigma$ is a nonsingular diagonal $d \times d$ matrix, and $M$ is full-rank.

2If $Q_{T,c}$ is full-rank, then its null space has rank 1 and $c$ is uniquely determined (within a sign ambiguity which is resolved by $c \geq 0$). If, instead, $Q_{T,c}$ is rank-deficient, then the intersection of the $d - 1$ hypersurfaces (i.e., the solution of (22)) is a $p$-manifold (with $p \geq 1$ in the $(d - 1)$-dimensional space and does not generate new binary vectors of interest. Hence, linearly independent combinations of $d - 1$ rows of $Q$ are ignored.

3The algorithm of Fig. 1 uses an alternative way of resolving the sign ambiguities at the intersections of hypersurfaces which was developed in [9] and led to the direct construction of a set $S$ of size $\sum_{g=0}^{d-1} \binom{N-1}{g} = O(N^{d-1})$ with complexity $O(N^d)$.
The Optimal $L_1$-Principal-Component Algorithm

**Input:** $X_{D \times N}$ data matrix

$$(U_{N \times d}, B_{d \times d}, V_{d \times N}) \leftarrow \text{svd}(X^T)$$

$Q_{N \times d} \leftarrow U \Sigma$

$B \leftarrow \text{compute_candidates}(Q)$

$m_{opt} \leftarrow \max_{m \in \{1, \ldots, N\}} \sum_{i=1}^{m} \sigma_i$

$b_{opt} \leftarrow B_{:, m_{opt}}$

**Output:** $r_{L_1} \leftarrow X b_{opt} / \| X b_{opt} \|_2$

**Function compute_candidates**

**Input:** $Q_{N \times m}$

if $m > 0$, $i \leftarrow 0$

for $I \subseteq \{1, \ldots, N\}$ s.t. $|I| = m - 1$, $i \leftarrow i + 1$

$Q_{(m-1) \times m} \leftarrow Q_{(m-1) \times (m-1)}$

$c_{m \times 1} \leftarrow \text{null}(Q)$

$c_{m \times 1} \leftarrow \text{sgn}(c_m)c$

$B_{i} \leftarrow \text{sgn}(Qc)$

for $j = 1 \colon m - 1$

$c_{(m-1) \times 1} \leftarrow \text{null}(Q_{j \times (m-1)})$

$c_{(m-1) \times 1} \leftarrow \text{sgn}(c_{m-1})c$

$B_{j} \leftarrow \text{sgn}(Q_{j \times (m-1)})$

else $m = 2$.

for $i = 1 \colon N$

$c_{2 \times 1} \leftarrow \text{null}(Q_{i \times 2})$

$c_{2 \times 1} \leftarrow \text{sgn}(c_{2})c$

$B_{i} \leftarrow \text{sgn}(Q_{i \times 2})$

else $B \leftarrow \text{sgn}(Q)$

**Output:** $B$

Fig. 1. The optimal $O(N^3)$ algorithm for the computation of the maximum $L_1$-projection component of a rank-$d$ data matrix $X_{D \times N}$ of $N$ samples of dimension $D$.

$d$ is the rank of $A$. This is due to the trace version of the Cauchy-Schwarz inequality [10], according to which

$$\text{tr}(R^T A) \leq \left\| U \Sigma^2 \right\|_F \left\| \Sigma^2 V^T R^T \right\|_F = \left\| \Sigma^2 \right\|_F^2 = \text{tr}(\Sigma) = \left\| A \right\|_*$

with equality if $(U \Sigma^2)^T = \Sigma^2 V^T R^T$ which is satisfied by $R = UV^T$.

To identify the optimal $L_1$ subspace for any number of components $K$, we begin by presenting a property of $P_L$ in the form of Proposition 4 below. The proof is omitted and can be found in [6].

**Proposition 4:** For any data matrix $X \in \mathbb{R}^{D \times N}$, the solution to $P_L: R_{L_1} = \arg \max_{R \in \mathbb{R}^{D \times K}, R^T R = I_K} \| R^T X \|_1$ is given by

$$R_{L_1} = UV^T$$

(26)

where $U$ and $V$ are the $D \times K$ and $N \times K$ matrices that consist of the $K$ highest-singular-value left and right, respectively, singular vectors of $X b_{opt}$ with

$$b_{opt} = \arg \max_{b \in \{\pm 1\}^{N \times K}} \| X b \|_*.$$  

(27)

In addition, $\| R_{L_1}^T X \|_1 = \| X b_{opt} \|_2$.

By Proposition 4, to find exactly the optimal $L_1$-norm projection operator $R_{L_1}$ we can perform the following steps:

1) Solve (27) to obtain $b_{opt}$.

2) Perform SVD on $X b_{opt} = U \Sigma V^T$.

3) Return $R_{L_1} = U_{:,1} V^T$.

Step 1 can be executed by an exhaustive search among all $2^{NK}$ binary matrices of size $N \times K$ followed by evaluation in the metric of interest in (27). That is, with computational cost $O(2^{NK})$ we identify the $L_1$-optimal $K$ principal components of $X$. An optimal algorithm for the computation of the $L_1$-optimal $K$ principal components of $X$ with complexity $O(NK\min(K, N-K)+1)$, rank($X$) $\leq D$, is presented in [6].

V. EXPERIMENTAL STUDIES

**Experiment 1 - Data Dimensionality Reduction**

We generate a data-set $X_{D \times N}$ of $N = 50$ two-dimensional $(D = 2)$ observation points drawn from the Gaussian distribution $N\left(0_{2 \times 1}, \begin{bmatrix} 15 & 13 \\ 13 & 26 \end{bmatrix} \right)$ as seen in Fig. 2(a). We calculate the $L_2$ (by standard SVD) and $L_1$ (by Section III.C, Case $d = 2$, complexity about $50 \log_2 50$) principal component of the data matrix $X$. Then, we assume that our data matrix is corrupted by three outlier measurements, $o_1, o_2, o_3$, shown in the bottom right corner of Fig. 2(b). We recalculate the $L_2$ and $L_1$ principal component of the corrupted data matrix $X_{\text{CRPT}} = [X, o_1, o_2, o_3]$ and notice (Fig. 2(a) versus Fig. 2(b)) how strongly the $L_2$ component responds to the outliers compared to $L_1$. To quantify the impact of the outliers, in Fig. 2(c) we generate 1000 new independent evaluation data points from $N\left(0_{2 \times 1}, \begin{bmatrix} 15 & 13 \\ 13 & 26 \end{bmatrix} \right)$ and estimate the mean square-fit-error $E\{\|x - \sigma^2 \|^2\}$ when $r = r_{L_2}(X_{\text{CRPT}})$ or $r_{L_1}(X_{\text{CRPT}})$. We find $E\{\|x - r_{L_2}(X_{\text{CRPT}})\|^2\} = 10.1296$ versus $E\{\|x - r_{L_1}(X_{\text{CRPT}})\|^2\} = 6.8387$. In contrast, when the principal component is calculated from the clean training set, $r = r_{L_2}(X)$ or $r_{L_1}(X)$, we find mean square-fit-error $6.3736$ and $4.2344$, correspondingly. We conclude that dimensionality reduction by $L_1$ principal components may loose only little in mean-square fit compared to $L_2$ when the designs are from clean training sets, but can protect significantly from outlier corrupted training.

**Experiment 2 - Direction-of-Arrival Estimation**

We consider a uniform linear antenna array of $D = 5$ elements that takes $N = 10$ snapshots of two incoming signals with angles of arrival $\theta_1 = -30^\circ$ and $\theta_2 = 50^\circ$,

$$x_n = A_1 s_{\theta_1} + A_2 s_{\theta_2} + n_n, \quad n = 1, \ldots, 10,$$

(28)

where $A_1, A_2$ are the received-signal amplitudes with array response vectors $s_{\theta_1} \text{ and } s_{\theta_2}$, correspondingly, and $n \sim C_N(0_5, \sigma^2 I_5)$ is additive white complex Gaussian noise. We assume that the signal-to-noise ratio (SNR) of the two signals is $\text{SNR}_1 = 10 \log_{10} \frac{\sigma^2}{\sigma^2} \text{dB} = 2 \text{dB}$ and $\text{SNR}_2 = 10 \log_{10} \frac{\sigma^2}{\sigma^2} \text{dB} = 3 \text{dB}$. Next, we assume that one arbitrarily selected measurement out of the ten observations $X_{5 \times 10} = $
Fig. 2. (a) Training data matrix $X_{2 \times 50}$ with its $L_1$ and $L_2$ principal components $(K = 1)$. (b) Training data matrix $X_{2 \times 50}$ corrupted by three additional outlier points in bottom right with recalculated $L_1$ and $L_2$ principal components. (c) Evaluation data set of 1000 nominal points against the outlier infected (Fig. 2(b)) $L_1$ and $L_2$ principal components.

Fig. 3. MUSIC power spectrum with $K = 2$ $L_2$ or $L_1$ calculated principal components (data set of $N = 10$ measurements with signals at $\theta_1 = -30^\circ$ and $\theta_2 = 50^\circ$ of which one measurement is additive-jammer corrupted with $\theta_J = 20^\circ$; SNR$_1 = 2$dB; SNR$_2 = 2$SNR$_J = 3$dB).

$[x_1, \ldots, x_{10}] \in \mathbb{C}^{5 \times 10}$ is corrupted by a jammer operating at angle $\theta_J = 20^\circ$ with amplitude $A_J = A_2$. We call the resulting corrupted observation set $X_{CRPT} \in \mathbb{C}^{5 \times 10}$ and create the real-valued version $X_{CRPT} = [Re\{X_{CRPT}\}^T, \text{Im}\{X_{CRPT}\}^T]^T \in \mathbb{R}^{10 \times 10}$ by $Re\{\cdot\}, \text{Im}\{\cdot\}$ part concatenation. We calculate the $K = 2$ $L_2$-principal components of $X_{CRPT}$, $R_{L_2} = [r_{L_2}^{(1)}, r_{L_2}^{(2)}] \in \mathbb{R}^{10 \times 2}$, and the $K = 2$ $L_1$-principal components of $X_{CRPT}$, $R_{L_1} = [r_{L_1}^{(1)}, r_{L_1}^{(2)}] \in \mathbb{R}^{10 \times 2}$. In Fig. 3, we plot the standard $L_2$ MUSIC spectrum [11]

$$P(\theta) \triangleq \frac{1}{s_\theta^T(l_{2D} - R_{L_2} l_{L_2}) s_\theta}, \quad \theta \in \left(\frac{\pi}{2}, \frac{3\pi}{2}\right), \quad (29)$$

where $s_\theta = [\text{Re}\{s_{\theta}\}^T, \text{Im}\{s_{\theta}\}^T]^T$, as well as what we may call “$L_1$ MUSIC spectrum” with $R_{L_1}$ in place of $R_{L_2}$. It is interesting to observe how $L_1$ MUSIC (in contrast to $L_2$ MUSIC) does not respond to the one-out-of-ten outlying jammer value in the data set and shows only the directions of the two actual nominal signals.

VI. CONCLUSIONS

We presented for the first time in the literature optimal (exact) algorithms for the calculation of the maximum-$L_1$-

projection component of data sets with complexity polynomial in the sample support size (and exponent equal to the data dimension). We generalized to multiple $L_1$-max-projection components and presented an explicit optimal $L_1$ subspace calculation algorithm in the form of matrix nuclear-norm evaluations. When $L_1$ subspaces are calculated on nominal “clean” training data, they differ little—arguably—from their $L_2$-subspace counterparts in least-squares fit. However, subspaces for data sets with possibly erroneous, “outlier” entries, $L_1$ Subspace calculation offers significant robustness/resistance to the presence of inappropriate data values.

REFERENCES


