Subspace Outliers Detection by Signal Subspace Matching

Mati Wax
Email: matiwax@gmail.com

Amir Adler
Email: adleram@braude.ac.il

***Abstract*—We present a novel and computationally efficient solution to the problem of subspace outlier detection that does not assume knowledge of the number of outliers nor exact knowledge of the dimension of the inliers subspace. The solution is based on a powerful representation of the inliers subspace, referred to as soft projection, and on a novel goodness-of-fit metric, referred to as signal subspace matching (SSM). Experimental results, demonstrating the performance of the SSM solution, are included.**

***Index Terms*—Subspace outlier detection, coherence metric, signal subspace matching.**

I. Introduction

Many types of signals, images and text are being modeled as vectors in *a lower-dimensional subspace*, referred to as *signal subspace*. Indeed, this model is common in face recognition [1], emitter localization [2], object recognition [3], radar [4], EEG [5], hyperspectral images [6], and text classification [7], to name a few key areas. Unfortunately, the data characterizing these low-dimensional subspaces is in many cases contaminated by *outliers*. These outliers can be either *unstructured*, i.e., random vectors lying outside the inliers' subspace, or *structured*, i.e., generated by a *different* low-dimensional subspace. As the solutions based on principal-component-analysis (PCA) are very *sensitive* to these outliers, [8], [9], coping with these outliers has been a central problem in data analysis in the last decade. For a review of this work, see [10]–[12].

Most of the work on this subject was aimed at *direct subspace recovery*. The proposed solutions were based either on L_1-norm minimization, which is known to be more robust to outliers than the L_2 norm [13]–[17], or on decomposing the data into a sum of a low-rank matrix and a sparse matrix, with the columns of the sparse matrix representing the outliers [18]–[30]. Yet, these solutions require solving optimization problems involving a large number of iterations, each with a high computational load. More critically, they all *require knowledge of the dimension* of the low-rank subspace, with some requiring also knowledge of the number of outliers, which are both typically *unknown*. Also, the algorithms based on sparse outlier models can only handle a relatively small number of outliers. A different solution, aimed at the *direct detection of the outliers*, was presented in [31]. This solution, inspired by [32] and referred to as Coherence Pursuit (CoP), is based on using a *"coherence"* metric measuring the coherence of each vector with all the other vectors. The vectors are sorted in descending coherence order, and the outliers are declared as those vectors with the lowest coherence score. This solution was shown to be computationally simple, handle both unstructured and structured outliers, and have equal or better performance than the other more computationally complex solutions. Yet, the solution suffers from two drawbacks. First, though the coherence metric is a good metric for the "similarity" of the vectors to each other, it *does not truly capture the nature of the underlying subspace*, especially of high-dimensional subspaces. Second, and more critically, the solution *assumes knowledge* of either the number of outliers or the dimension of the inliers subspace, which is typically *unknown* in practice.

In this paper, we present a radically different solution that does not assume knowledge of the number of outliers nor exact knowledge of the dimension of the inliers subspace. The solution is based on a powerful representation of the inliers subspace, referred to as soft projection, that does not require explicit determination of the inliers subspace, and on a novel goodness-of-fit metric, referred to as signal subspace matching (SSM), that measures the distance between the given vectors and the inliers subspace.

II. Problem Formulation

Due to limited space, and since "unstructured" outliers pose a significantly less challenging problem, we confine our discussion to "structured" outliers only.

Suppose that we are given a total of N $P \times 1$ vectors $\{\mathbf{y}_i\}_{i=1}^{N}$, with N_I vectors $\{\mathbf{y}_i\}_{i=1}^{N_I}$ referred to as *inliers*, and $N_O = N - N_I$ vectors $\{\mathbf{y}_i\}_{i=N_I+1}^{N}$ referred to as *outliers*. Suppose that the inliers are generated by the following low-rank model:

$$\mathbf{y}_i = \mathbf{A}_I \mathbf{s}_i^I + \mathbf{n}_i^I, \qquad i = 1, ..., N_I, \tag{1}$$

where $\mathbf{A}_I$ is the $P \times Q_I$ matrix characterizing the inliers subspace, $\mathbf{s}_i^I$ is a $Q_I \times 1$ coefficients-vector characterizing the i-th inlier, and $\mathbf{n}_i^I$ is the $P \times 1$ vector of the i-th inliers noise.

Suppose further that the outliers are generated by the following low-rank model:

$$\mathbf{y}_i = \mathbf{A}_O \mathbf{s}_i^O + \mathbf{n}_i^O, \qquad i = N_I + 1, ..., N. \tag{2}$$

where $\mathbf{A}_O$ is the $P \times Q_O$ matrix characterizing the outliers subspace, $\mathbf{s}_i^O$ is a $Q_O \times 1$ coefficients vector characterizing i-th outlier, and $\mathbf{n}_i^O$ is the $P \times 1$ vector of the i-th outliers noise.

ISBN: 978-9-4645-9360-0

We make the following further assumptions regarding the inliers and outliers:

A1: The number of outliers N_O is *unknown*.

A2: The subspace-dimensions Q_I and Q_O are *unknown* and obey $Q_I < P$ and $Q_O < P$.

A3: The matrices $\mathbf{A}_I$ and $\mathbf{A}_O$ are *unknown* and full-column-rank, i.e., $\mathrm{rank}[\mathbf{A}_I] = Q_I$ and $\mathrm{rank}[\mathbf{A}_O] = Q_O$.

A4: The coefficient matrices are *full-row-rank*, i.e., $\mathrm{rank}[\mathbf{s}_1^I ...., \mathbf{s}_{N_I}^I] = Q_I$ and $\mathrm{rank}[\mathbf{s}_{(N_I+1)}^O ...., \mathbf{s}_N^O] = Q_O$.

A5: The noise vectors $\{\mathbf{n}_i^I\}$ and $\{\mathbf{n}_i^O\}$ are zero-mean and independent of $\{\mathbf{s}_i^I\}$ and $\{\mathbf{s}_i^O\}$.

Let $\mathbf{Y}$ denote the $P \times N$ matrix of the given data,

$$\mathbf{Y} = [\mathbf{Y}_I, \mathbf{Y}_O], \tag{3}$$

where $\mathbf{Y}_I = [\mathbf{y}_1, ..., \mathbf{y}_{N_I}]$ denotes the $P \times N_I$ matrix of the inliers, and $\mathbf{Y}_O = [\mathbf{y}_{N_I+1}, ..., \mathbf{y}_N]$ denotes the $P \times N_O$ matrix of the outliers.

We can now state the subspace outliers detection problem as follows: *given the vector set* $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^N$, *detect the inliers vectors* $\{\mathbf{y}_i\}_{i=1}^{N_I}$.

III. The Soft Projection

Ignoring the noise and the outliers, and denoting by $\bar{\mathbf{y}}_i$ the noise-free $\mathbf{y}_i$, from (1), we have

$$\bar{\mathbf{y}}_i = \mathbf{A}_I \mathbf{s}_i^I, \quad i = 1, ..., N_I, \tag{4}$$

which implies, by A2-A4, that $\langle \bar{\mathbf{Y}}_I \rangle = \langle \mathbf{A}_I \rangle$, where $\bar{\mathbf{Y}}_I = [\bar{\mathbf{y}}_1, ..., \bar{\mathbf{y}}_N]$ and $\langle \bullet \rangle$ denotes the column span of the bracketed matrix. That is, $\langle \mathbf{A}_I \rangle$ and $\langle \bar{\mathbf{Y}}_I \rangle$ span the *same* Q_I-dimensional subspace, referred to as the *inliers signal subspace*. Denoting by $\mathbf{P}_{\mathbf{A}_I}$ and $\tilde{\mathbf{P}}_{\bar{\mathbf{Y}}_I}$ the projection matrices on $\langle \mathbf{A}_I \rangle$ and $\langle \bar{\mathbf{Y}}_I \rangle$, i.e.,

$$\mathbf{P}_{\mathbf{A}_I} = \mathbf{A}_I (\mathbf{A}_I^H \mathbf{A}_I)^{-1} \mathbf{A}_I^H, \tag{5}$$

and

$$\mathbf{P}_{\bar{\mathbf{Y}}_I} = \bar{\mathbf{Y}}_I (\bar{\mathbf{Y}}_I^H \bar{\mathbf{Y}}_I)^{-1} \bar{\mathbf{Y}}_I^H, \tag{6}$$

where $(\bullet)^H$ is the Hermitian operator, it follows from (4) that

$$\mathbf{P}_{\bar{\mathbf{Y}}_I} = \mathbf{P}_{\mathbf{A}_I}. \tag{7}$$

Note that for $N_I > P$, the $N_I \times N_I$ matrix $\bar{\mathbf{Y}}_I^H \bar{\mathbf{Y}}_I$ is rank-deficient and hence *singular*. To solve the singularity problem, and more critically, to make the projection matrix a *good estimate* of $\mathbf{P}_{\mathbf{A}_I}$ in the *presence of noise*, we use the *soft projection*, which was introduced in [33] [34] and is given by

$$\tilde{\mathbf{P}}_{\mathbf{Y}} = \mathbf{Y}(\mathbf{Y}^H \mathbf{Y} + \delta \mathbf{I}_N)^{-1} \mathbf{Y}^H, \tag{8}$$

where $\mathbf{I}_N$ is the $N \times N$ identity matrix and δ is the diagonal loading factor given by

$$\delta = \alpha \mathrm{tr}(\mathbf{Y}\mathbf{Y}^H), \tag{9}$$

with α denoting a small scalar, and $\mathrm{tr}(\bullet)$ denoting the trace of the bracketed matrix. As shown in [33] [34], α can be selected in two ways. In applications such as image recognition problems, wherein the pixels of each image are stacked into a vector with more than 1000 elements, the value of α is set to a small *data-independent* value

$$\alpha \in [10^{-3}, ..., 10^{-7}]. \tag{10}$$

In other applications, such as sensor array problems, wherein there is a need to cope with large variability in the impinging signals, α is set to a *data-dependent* value, given by

$$\alpha = \frac{1}{\frac{1}{\sqrt{P}} \sigma_\lambda}, \tag{11}$$

where σ_λ is the sample-standard-deviation of the eigenvalues of the sample-covariance-matrix $\hat{\mathbf{R}} = \frac{1}{N} \mathbf{Y}\mathbf{Y}^H$, given by

$$\sigma_\lambda = \sqrt{\frac{1}{P} \sum_{i=1}^{P} (\lambda_i - \bar{\lambda})^2} = \frac{1}{\sqrt{P}} ||(\hat{\mathbf{R}} - \bar{\lambda} \mathbf{I}_P)||_F, \tag{12}$$

with λ_i denoting the i-th eigenvalues of $\hat{\mathbf{R}}$, and $\bar{\lambda}$ denoting the sample-average of $\{\lambda_i\}_{i=1}^P$, given by

$$\bar{\lambda} = \frac{1}{P} \sum_{i=1}^{P} \lambda_i = \frac{1}{P} \mathrm{tr}(\hat{\mathbf{R}}). \tag{13}$$

If α is selected accordingly, it was demonstrated in [33] [34] that if there are no outliers, $\tilde{\mathbf{P}}_{\bar{\mathbf{Y}}}$ is a good approximation to the projection matrix on the signal subspace, i.e.,

$$\tilde{\mathbf{P}}_{\mathbf{Y}} \approx \mathbf{P}_{\mathbf{A}_I}. \tag{14}$$

Note that $\tilde{\mathbf{P}}_{\mathbf{Y}}$ does not require specifying explicitly the dimension of the underlying subspace – it is determined *implicitly* by the data. This is in contrast to PCA, wherein the dimension must be specified *explicitly* by the user, giving rise to the well-known and challenging *order-selection* problem. Note also that $\tilde{\mathbf{P}}_{\mathbf{Y}}$ is not a *proper* projection matrix. Indeed, because of the diagonal loading, its eigenvalues are *not necessarily* zero or one, in contrast to a proper projection matrix.

The form of the soft projection given by (8) is computationally complex when N is large, as it requires the inversion of a $N \times N$ matrix. In the case that $N > P$, a computationally simpler form can be obtained by using the matrix inversion lemma. Indeed, we readily get

$$(\delta \mathbf{I}_P + \mathbf{Y}\mathbf{Y}^H)^{-1} = \delta^{-1}(\mathbf{I}_P - \mathbf{Y}(\delta \mathbf{I}_N + \mathbf{Y}^H \mathbf{Y})^{-1} \mathbf{Y}^H), \tag{15}$$

which can be rewritten as

$$\tilde{\mathbf{P}}_{\mathbf{Y}} = \mathbf{I}_P - \delta(\mathbf{Y}\mathbf{Y}^H + \delta \mathbf{I}_P)^{-1}, \tag{16}$$

requiring only the inversion of a $P \times P$ matrix.

IV. Solution Ingredients

A. The Signal Subspace Matching (SSM) Score

To introduce the SSM score, we first extract a small set of vectors that capture as best as possible the inliers subspace. To this end, we use the ”coherence” metric [31], measuring the ”coherence” of each vector in a set with the other vectors. More specifically, the CoP score of $\mathbf{y}_i$ is given by the accumulated squared cosines of the angles between $\mathbf{y}_i$ and all the other vectors:

$$\text{CoP}(\mathbf{y}_i) = \sum_{k=1,k\neq i}^{N} |(\mathbf{y}_i/|\mathbf{y}_i|)^H(\mathbf{y}_k/|\mathbf{y}_k|)|^2, \quad i=1,...,N. \tag{17}$$

We compute the CoP scores of each vector $\mathbf{y}_i$, sort the values $\{\text{CoP}(\mathbf{y}_i)\}_{i=1}^N$ in *descending order*, and select the Q vectors $\mathbf{y}_i$ yielding the *highest* CoP score. Q is a parameter that should obey $Q \geq Q_I$ so as to ensure that the selected set of vectors spans the inliers subspace. Yet, Q should not be too large, so as to ensure that the selected set includes *only inliers*. Thus, a loose upper bound on the dimension should do. The selected Q vectors are then used to form the $P \times Q$ matrix $\tilde{\mathbf{Y}}$ and its corresponding soft projection,

$$\tilde{\mathbf{P}}_{\tilde{\mathbf{Y}}} = \tilde{\mathbf{Y}}(\tilde{\mathbf{Y}}^H\tilde{\mathbf{Y}} + \tilde{\delta}\mathbf{I}_Q)^{-1}\tilde{\mathbf{Y}}^H, \tag{18}$$

where

$$\tilde{\delta} = \tilde{\alpha}\text{tr}(\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^H). \tag{19}$$

With $\tilde{\mathbf{P}}_{\tilde{\mathbf{Y}}}$ serving as an estimate of the projection matrix onto the inliers subspace, we next introduce the SSM score. To this end, let $\mathbf{P}_{\mathbf{y}_i}$ denote the projection matrix on $\mathbf{y}_i$,

$$\mathbf{P}_{\mathbf{y}_i} = \mathbf{y}_i(\mathbf{y}_i^H\mathbf{y}_i)^{-1}\mathbf{y}_i^H, \tag{20}$$

and let the distance between $\tilde{\mathbf{P}}_{\tilde{\mathbf{Y}}}$ and $\mathbf{P}_{\mathbf{y}_i}$ be measured by the following SSM metric, inspired by [33] [34]:

$$\text{SSM}(\tilde{\mathbf{P}}_{\tilde{\mathbf{Y}}}, \mathbf{P}_{\mathbf{y}_i}) = ||\tilde{\mathbf{P}}_{\tilde{\mathbf{Y}}} - \mathbf{P}_{\mathbf{y}_i}||_F^2. \tag{21}$$

Using the fact that $\text{tr}(\mathbf{P}_{\mathbf{y}_i}^2) = \text{tr}(\mathbf{P}_{\mathbf{y}_i}) = 1$, this metric can be rewritten as

$$\text{SSM}(\tilde{\mathbf{P}}_{\tilde{\mathbf{Y}}}, \mathbf{P}_{\mathbf{y}_i}) = \text{tr}(\tilde{\mathbf{P}}_{\tilde{\mathbf{Y}}}^2) + 1 - 2|\tilde{\mathbf{P}}_{\tilde{\mathbf{Y}}}(\mathbf{y}_i/|\mathbf{y}_i|)|^2, \tag{22}$$

where $|\bullet|$ denotes the Euclidean norm. Ignoring the first two terms, since they are independent of $\mathbf{y}_i$, the SSM score is defined as

$$\text{SSM}(\mathbf{y}_i) = |\tilde{\mathbf{P}}_{\tilde{\mathbf{Y}}}(\mathbf{y}_i/|\mathbf{y}_i|)|^2, \quad i=1,...,N. \tag{23}$$

The SSM score has a very intuitive interpretation: it measures the squared norm of the projection of the unit vector $\mathbf{y}_i/|\mathbf{y}_i|$ onto the inliers subspace. Thus, *the higher* is $\text{SSM}(\mathbf{y}_i)$, *the more likely is* $\mathbf{y}_i$ *an inlier*.

B. The SSM error Criterion

As the SSM score measures the likelihood of being an inlier, *sorting* the SSM scores $\{\text{SSM}(\mathbf{y}_i)\}_{i=1}^N$ in descending order yields an *ordered list* wherein the order represents the likelihood of being an inlier. The outliers are supposedly at the "end" of this ordered list. The question is how to determine the "border" between inliers and outliers.

To this end, let $\mathbf{y}_{(t)}$ denote the t-th vector in the *ordered list*, let $\mathbf{Y}_t$ denote the $P \times t$ matrix constructed from the first t vectors in the ordered list,

$$\mathbf{Y}_t = [\mathbf{y}_{(1)}, ..., \mathbf{y}_{(t)}], \tag{24}$$

and let $\tilde{\mathbf{P}}_{\mathbf{Y}_t}$ denote the corresponding soft projection matrix,

$$\tilde{\mathbf{P}}_{\mathbf{Y}_t} = \mathbf{Y}_t(\mathbf{Y}_t^H\mathbf{Y}_t + \delta_t\mathbf{I}_t)^{-1}\mathbf{Y}_t^H, \tag{25}$$

where

$$\delta_t = \alpha_t\text{tr}(\mathbf{Y}_t\mathbf{Y}_t^H). \tag{26}$$

Note that as t increases, $\mathbf{Y}_t$ includes more and more inliers, and as a result $\tilde{\mathbf{P}}_{\mathbf{Y}_t}$ captures better and better the inliers subspace. Yet, at some value of t, outliers start entering $\mathbf{Y}_t$ and corrupt it, and consequently $\tilde{\mathbf{P}}_{\mathbf{Y}_t}$ starts deteriorating. Recalling that $\tilde{\mathbf{P}}_{\tilde{\mathbf{Y}}}$ characterizes the inliers subspace, we propose to detect the "border" between inliers and outliers by evaluating the distance between $\tilde{\mathbf{P}}_{\tilde{\mathbf{Y}}}$ and $\tilde{\mathbf{P}}_{\mathbf{Y}_t}$, given by

$$\text{SSM}(\tilde{\mathbf{P}}_{\mathbf{Y}_t}, \tilde{\mathbf{P}}_{\tilde{\mathbf{Y}}}) = ||\tilde{\mathbf{P}}_{\mathbf{Y}_t} - \tilde{\mathbf{P}}_{\tilde{\mathbf{Y}}}||_F^2, \tag{27}$$

and searching for the value of t yielding its minimal value,

$$\hat{t} = \text{argmin}_t ||\tilde{\mathbf{P}}_{\mathbf{Y}_t} - \tilde{\mathbf{P}}_{\tilde{\mathbf{Y}}}||_F^2. \tag{28}$$

The inliers are then given by:

$$\hat{\mathbf{Y}}_I = [\mathbf{y}_{(1)}, ..., \mathbf{y}_{(\hat{t})}]. \tag{29}$$

C. Recursive Expression for $\tilde{\mathbf{P}}_{\mathbf{Y}_t}$

The evaluation of $\text{SSM}(\tilde{\mathbf{P}}_{\mathbf{Y}_t}, \tilde{\mathbf{P}}_{\tilde{\mathbf{Y}}})$ requires the computation of $\tilde{\mathbf{P}}_{\mathbf{Y}_t}$ for every value of t. This, in turn, by (16), requires matrix inversion for every t, which is computationally expensive. To alleviate this computational burden we apply the matrix inversion lemma to (16), with some straightforward manipulations, resulting in the following recursive algorithm:

$$\tilde{\mathbf{P}}_{\mathbf{Y}_t} = \tilde{\mathbf{P}}_{\mathbf{Y}_{t-1}} + \frac{1}{\gamma_t\delta_{t-1}}\tilde{\mathbf{y}}_{(t)}\tilde{\mathbf{y}}_{(t)}^H \tag{30}$$

where

$$\tilde{\mathbf{y}}_{(t)} = (\mathbf{I}_P - \tilde{\mathbf{P}}_{\mathbf{Y}_t})\mathbf{y}_{(t)}, \tag{31}$$

and

$$\gamma_t = 1 + \frac{1}{\delta_{t-1}}\mathbf{y}_{(t)}^H\tilde{\mathbf{y}}_{(t)}. \tag{32}$$

This is an intuitively pleasing *rank-one* update of $\tilde{\mathbf{P}}_{\mathbf{Y}_t}$. Indeed, the updating vector, $\tilde{\mathbf{y}}_{(t)}$, is the projection of $\mathbf{y}_{(t)}$ on $\mathbf{I}_P - \tilde{\mathbf{P}}_{\mathbf{Y}_{t-1}}$, the orthogonal complement of $\tilde{\mathbf{P}}_{\mathbf{Y}_{t-1}}$, with $\gamma_t\delta_{t-1}$ serving as a normalization scalar.

To numerically stabilize (30), we apply a forgetting factor β, with $\beta < 1$, resulting in the following recursion:

$$\tilde{\mathbf{P}}_{\mathbf{Y}_t} = \beta\tilde{\mathbf{P}}_{\mathbf{Y}_{t-1}} + \frac{1}{\gamma_t\delta_{t-1}}\tilde{\mathbf{y}}_{(t)}\tilde{\mathbf{y}}_{(t)}^H. \tag{33}$$

V. Experimental Results

We next present experimental results for a sensor array problem, comparing the performance of the SSM solution to the CoP solution [31]. The CoP solution was selected since it is based on direct outliers detection, and as such enables straightforward performance comparison, without resorting to the estimation of the inliers subspace

The performance is compared by two error metrics:

$$CER1 = \frac{\text{number of inliers classified as outliers}}{N_I}, \tag{34}$$

TABLE I
INLIERS AND OUTLIERS SUBSPACES

$\mathbf{A}_I$	$[\mathbf{a}(10), \mathbf{a}(20), \mathbf{a}(30), \mathbf{a}(40), \mathbf{a}(50), \mathbf{a}(60), \mathbf{a}(70), \mathbf{a}(80)]$
$\mathbf{A}_O$	$[\mathbf{a}(130), \mathbf{a}(140)]$
$\mathbf{A}_O$	$[\mathbf{a}(130), \mathbf{a}(140), \mathbf{a}(150), \mathbf{a}(160), \mathbf{a}(170), \mathbf{a}(180)]$

and

$$CER2 = \frac{\text{number of outliers classified as inliers}}{N_O}. \qquad (35)$$

We consider a uniform circular array (UCA) with $P = 100$ elements, spaced half-wavelength apart. The inliers and outliers are generated according to (1) and (2), using Q_I-dimensional and Q_O-dimensional subspaces, respectively, specified by the $P \times Q_I$ and $P \times Q_O$ matrices $\mathbf{A}_I = [\mathbf{a}(\theta_1), ..., \mathbf{a}(\theta_{Q_I})]$ and $\mathbf{A}_O = [\mathbf{a}(\theta_1), ..., \mathbf{a}(\theta_{Q_O})]$, where the $P \times 1$ vector $\mathbf{a}(\theta)$ denotes the "steering vector" of the array towards the bracketed direction-of-arrival (in degrees). The coefficient $\{\mathbf{s}_i\}$ are generated as white Gaussian random vectors with covariance matrix $\sigma_s^2 \mathbf{I}_P$, and the noise $\{\mathbf{n}_i\}$ as white Gaussian vectors with covariance matrix $\sigma_n^2 \mathbf{I}_P$. The signal-to-noise ratio (SNR) is defined as $10 \log_{10} \frac{\sigma_s^2}{\sigma_n^2}$.

In all the experiments, the value of the parameter Q was set to $Q = 12$, and the value of β was set to $\beta = 0.999$.

The performance is evaluated by presenting the classification errors (34) and (35) as a function of the number of outliers for an SNR of 15dB, with inliers and outliers having the same SNR. The results are obtained by averaging 20 runs.

Experiment 1 presents the results for $N_I = 100$ inliers generated by an 8-dimensional subspace $\mathbf{A}_I$, and a variable number of outliers generated by a 2-dimensional subspace $\mathbf{A}_O$. The subspaces are specified in Table I. The results are presented in Fig. 1. Note the clear superiority of the SSM over the CoP in coping with a larger number of outliers. Indeed, while the CoP classification errors start rising at 20 outliers, the SSM classification errors start rising only at 40 outliers.

As evident from the results of the CoP solution, a relatively small number of 2-dimensional outliers is sufficient to deteriorate its performance. Though the SSM uses the CoP coherence metric for the construction of $\tilde{\mathbf{P}}_{\tilde{\mathbf{Y}}}$, its performance is affected to a much less extent, as is evident from the results.

To shed more light on the SSM solution in this experiment, we examine the behavior of the SSM error (27). The results of 20 superimposed Monte Carlo runs for $N_O = 30$ are presented in Fig. 2, with each run having a different color. Note that almost all runs have a minimum at $t = 100$, implying that almost all inliers are classified correctly.

Experiment 3 presents the results for $N_I = 100$ inliers generated by an 8-dimensional subspace $\mathbf{A}_I$, and a varying number of outliers generated by a 6-dimensional subspace $\mathbf{A}_O$. The subspaces are specified in Table I. The results are presented in Fig. 3. Note again the clear superiority of the SSM over the CoP in coping with a larger number of outliers. Indeed, while the CoP classification errors start rising at 40 outliers, the SSM starts rising only at 70 outliers.

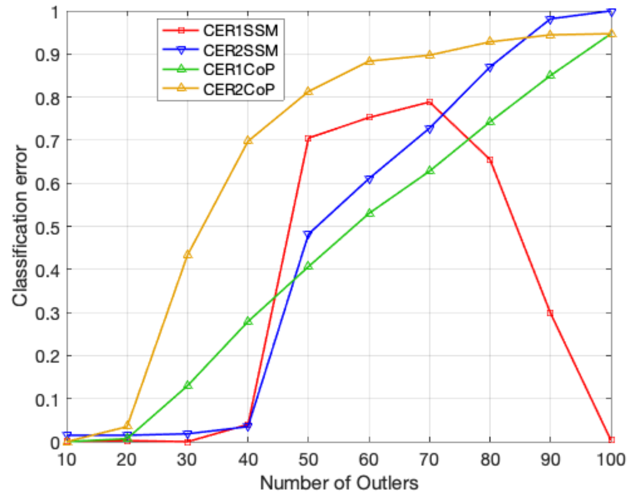

Fig. 1. Structured outliers: $N_I = 100$ inliers generated by an 8-dimensional subspace $\mathbf{A}_I$, and a varying number of outliers generated by 2-dimensional subspace $\mathbf{A}_O$. SNR=15dB.

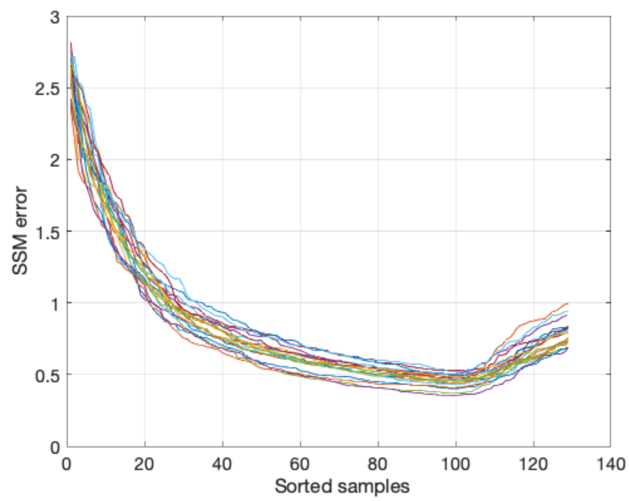

Fig. 2. Structured outliers: $N_I = 100$ inliers generated by an 8-dimensional subspace $\mathbf{A}_I$, and outliers generated by 2-dimensional subspace $\mathbf{A}_O$. SNR=15dB. 20 superimposed runs of the SSM error as a function of the sorted samples. $N_O = 30$.

Note that the performance of the CoP and SSM in Experiment 3 is better than their performance in Experiment 1. This can be attributed to the fact that the CoP metric is significantly more robust to outliers from a 6-dimensional subspace than to outliers from a 2-dimensional subspace, reflecting the *inherent* larger sensitivity of the CoP to low-dimensional outliers, being based on an angles-between-vectors metric.

Note also that in experiments 1 and 3 the SSM classification error CER1SSM falls to zero at a large number of outliers. This happens because when the number of outliers is relatively large $\tilde{\mathbf{P}}_{\mathbf{Y}_t}$ is composed of *both inliers and outliers* and hence all the vectors are classified as inliers.

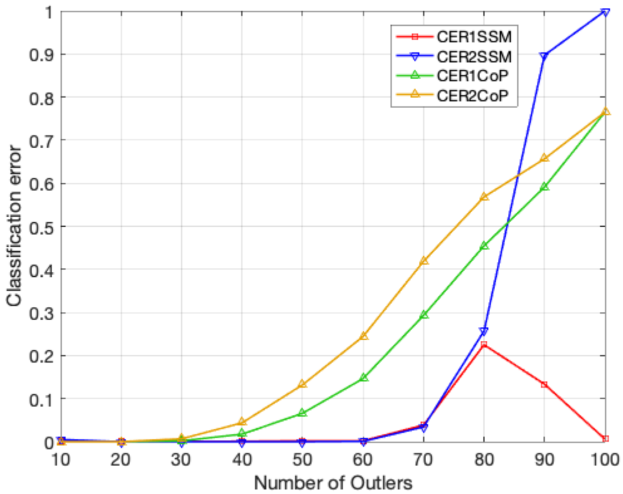

Fig. 3. Structured outliers: $N_I = 100$ inliers generated by an 8-dimensional subspace $\mathbf{A}_I$, and a varying number of outliers generated by a 6-dimensional subspace $\mathbf{A}_O$. SNR=15dB.

VI. Conclusions

We have presented a computationally efficient solution for subspace outlier detection that does not assume knowledge of the number of outliers nor exact knowledge of the dimension of the inliers subspace – a loose upper bound on the subspace dimension suffices. The solution was compared to the CoP solution in the case of structured outliers and shown to largely outperform it, notwithstanding that the CoP solution *assumes knowledge* of the number of outliers while the SSM solution *estimates* it from the data.

References

[1] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[2] E. Kupershtein, M. Wax, and I. Cohen, "Single-site emitter localization via multipath fingerprinting," *IEEE Transactions on Signal Processing*, vol. 61, no. 1, pp. 10–21, 2013.

[3] Y. Ohkawa and K. Fukui, "Kernel mutual subspace method and its application for object recognition," in *IEICE Trans. Inf. Syst.*, June 2012, vol. E95-D, pp. 1616–1627.

[4] P. Lopez-Rodriguez, D. Escot-Bocanegra, R. Fernandez-Redo and I. Bravo, "Non-cooperative identification of civil aircraft using a generalised mutual subspace method," *IET Radar, Sonar Navigation*, 2016.

[5] R. Ashari and C. Anderson, "EEG subspace analysis and classification using principal angles for brain-computer interfaces," in *2014 IEEE Symposium on Computational Intelligence in Brain Computer Interfaces (CIBCI)*, 2014, pp. 57–63.

[6] R. Aghaee and M. Mokhtarzade, "Classification of hyperspectral images using subspace projection feature space," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 9, pp. 1803–1807, 2015.

[7] E. K. Shimomoto1, F. Portet, and K. Fukui, "Text classification based on the word subspace representation," *Pattern Analysis and Applications*, 2021.

[8] M. A. Fischler and R. C. Bolles,, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," in *Readings in Comput. vision. Elsevier*, 1987, p. 726–740.

[9] F. De la Torre and M. J. Black, "A framework for robust subspace learning," *Int. J. Comput. Vis.*, vol. 54, no. 1, pp. 117–142, 2003.

[10] N. Vaswani, T. Bouwmans, S. Javed and P. Narayanamurthy, "Robust subspace learning: Robust PCA, robust subspace tracking, and robust subspace recovery," *IEEE Signal Processing Magazine*, vol. 35, no. 4, pp. 32–55, 2018.

[11] N. Vaswani and P. Narayanamurthy, "Static and dynamic robust PCA and matrix completion: A review," *Proceedings of the IEEE*, vol. 106, no. 8, pp. 1359–1379, 2018.

[12] G. Lerman and T. Maunu, "An overview of robust subspace recovery," *Proceedings of the IEEE*, vol. 106, no. 8, pp. 1380–1410, 2018.

[13] C. Ding, D. Zhou, X. He, and H. Zha, "R1-PCA: Rotational invariant L1-norm principal component analysis for robust subspace factorization," in *Proc. 23rd Int. Conf. Mach. Learn.*, 2006, p. 281–288.

[14] N. Kwak, "Principal component analysis based on L1-norm maximization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 9, pp. 1672–1680, 2008.

[15] A. Y Yang, S. S Sastry, A. Ganesh, and Y. Ma, "Fast L1-minimization algorithms and an application in robust face recognition: A review," in *2010 IEEE International Conference on Image Processing*, 2010, pp. 1849–1852.

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

[17] P.P Markopoulos, S. Kundu, S. Chamadia and D.A Pados, "Efficient L1-norm principal-component analysis via bit flipping," *IEEE Transactions on Signal Processing*, vol. 65, no. 16, pp. 4252–4264, 2017.

[18] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. ACM*, vol. 58, no. 3, pp. 11–37, 2011.

[19] X. Yuan and J. Yang, "Sparse and low-rank matrix decomposition via alternating direction methods," *Pacific J. Optim.*, vol. 9, no. 1, pp. 1–11, 2009.

[20] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, "Rank sparsity incoherence for matrix decomposition," *SIAM J. Optim.*, vol. 21, no. 2, pp. 572–596, 2011.

[21] D. Hsu, S.M Kakade and T. Zhang, "Robust matrix decomposition with sparse corruptions," *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7221–7234, 2011.

[22] H. Xu, C. Caramanis and S. Sanghavi, "Robust PCA via outlier pursuit," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3047–3064, 2012.

[23] P. Filzmoser C. Croux and H. Fritz, "Robust sparse principal component analysis," *Technometrics*, vol. 55, no. 2, pp. 202–214, 2013.

[24] C. Constantine H. Xu and S. Mannor, "Outlier-robust PCA: The high-dimensional case," *IEEE Transactions on Information Theory*, vol. 59, no. 1, pp. 546–572, 2013.

[25] P. Netrapalli, U. N. Niranjan, S. Sanghavi, A. Anandkumar, and P. Jain, "Non-convex robust PCA," in *Proc. 23rd Int. Conf. Mach. Learn.*, 2014, p. 1107–1115.

[26] X. Li and J. Haupt, "Identifying outliers in large matrices via randomized adaptive compressive sampling," *IEEE Transactions on Signal Processing*, vol. 63, no. 7, pp. 1792–1807, 2015.

[27] Y. Cherapanamjeri, P. Jain, and P. Netrapalli, "Thresholding based outlier robust PCA," in *Proc. 30th Conf. Learn. Theory, Amsterdam, The Netherlands, Jul. 7–10*, 2017, p. 593–628.

[28] M. Rahmani and G. K. Atia, "Randomized robust subspace recovery and outlier detection for high dimensional data matrices," *IEEE Transactions on Signal Processing*, vol. 65, no. 6, pp. 1580–1594, 2017.

[29] M. Rahmani and G.Atia, "High dimensional low rank plus sparse matrix decomposition," *IEEE Transactions on Signal Processing*, vol. 65, no. 8, pp. 2004–2019, 2017.

[30] M. Rahmani and G. K Atia, "Low-rank matrix recovery with simultaneous presence of outliers and sparse corruption," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 6, pp. 1170–1181, 2018.

[31] M. Rahmani and G. K Atia, "Coherence pursuit: Fast, simple, and robust principal component analysis," *IEEE Transactions on Signal Processing*, vol. 65, no. 23, pp. 6260–6275, 2017.

[32] R. Heckel and H. Bölcskei, "Robust subspace clustering via thresholding," *IEEE Transactions on Information Theory*, vol. 61, no. 11, pp. 6320–6342, 2015.

[33] M. Wax, and A. Adler, "Detection of the number of signals by signal subspace matching," *IEEE Transactions on Signal Processing*, vol. 69, pp. 973–985, 2021.

[34] M. Wax and A. Adler, "Vector set classification by signal subspace matching," *IEEE Transactions on Information Theory*, vol. 69, no. 3, pp. 1853–1856, 2022.