Determining the number of signals correlated across multiple data sets for small sample support

Yang Song[1], Tanuj Hasija[1], Peter J. Schreier[1], and David Ramírez[2,3]
[1] Signal and System Theory Group, Universität Paderborn, Germany
[2] Signal Processing Group, University Carlos III of Madrid, Spain
[3] Gregorio Marañón Health Research Institute, Spain
Email: {yang.song,tanuj.hasija,peter.schreier}@sst.upb.de, david.ramirez@uc3m.es

***Abstract*—This paper presents a detection scheme for determining the number of signals that are correlated across multiple data sets when the sample size is small compared to the dimensions of the data sets. To accommodate the sample-poor regime, we decouple the problem into several independent two-channel order-estimation problems that may be solved separately by a combination of principal component analysis (PCA) and canonical correlation analysis (CCA). Since the signals that are correlated across all data sets must be a subset of the signals that are correlated between any pair of data sets, we keep only the correlated signals for each pair of data sets. Then, a criterion inspired by a traditional information-theoretic criterion is applied to estimate the number of signals correlated across all data sets. The performance of the proposed scheme is verified by simulations.**

***Index Terms*—Canonical correlation analysis, model-order selection, multiple data fields, principle component analysis, small sample support.**

I. Introduction

The analysis of association between more than two data sets plays an important role in many applications, e.g. in array signal processing, biomedicine, and climate science. A key problem is to determine how many signals are common or correlated across all data sets. In this paper, we consider this model-order selection problem for a small number of samples, which may even be smaller than the dimensions of each data set.

In the literature, model-order selection for more than two data sets has not yet received significant attention. Among the very few published approaches are [1]–[3]. These three papers assume that all signals that are correlated between any two data sets are also correlated between all remaining data sets. This assumption is not unrealistic in some applications. For instance, in array signal processing, the same signals may be received by multiple spatially separated arrays. However, in many other applications, this assumption may be too restrictive. For instance, in brain imaging, the signals correlated between two neurological modalities may be uncorrelated with the other modalities [4], [5]. Moreover, among these three prior works only [3] is able to work in the sample-poor regime.

In this paper, we present a detection scheme that (i) allows signals correlated between two data sets to be uncorrelated with the remaining data sets, thereby relaxing the assumption made in [1]–[3]; and (ii) is capable of handling small sample-support. To achieve these two goals, we propose a scheme consisting of three steps:

1) A "max-min" detector, which was proposed recently in [6], [7], is applied to determine the number of correlated signals in all pairs of data sets.
2) For each pair of data sets, only the correlated signals are estimated and kept.
3) A criterion based on the traditional information-theoretic criterion for two data sets [8] is applied to identify the number of signals that are correlated across all data sets.

II. Problem Formulation

We observe M independent and identically distributed (i.i.d.) sample vectors $\mathbf{x}_m^{(\ell)} \in \mathbb{C}^{n_\ell}, \ell = 1,\ldots,L$, $m = 1,\ldots,M$, that are drawn from the L-channel measurement model

$$\mathbf{x}^{(\ell)} = \mathbf{A}^{(\ell)}\mathbf{s}^{(\ell)} + \mathbf{n}^{(\ell)}.$$

The ℓth signal vector $\mathbf{s}^{(\ell)} \in \mathbb{C}^{Q_\ell}$ contains Q_ℓ independent Gaussian random variables $s_q^{(\ell)}$, with zero mean and fixed but unknown standard deviation $\sigma_q^{(\ell)}, q = 1,\ldots,Q_\ell$. The matrices $\mathbf{A}^{(\ell)} \in \mathbb{C}^{n_\ell \times Q_\ell}, \ell = 1,\ldots,L$, as well as the dimensions Q_ℓ are fixed but unknown. Without loss of generality, $\mathbf{A}^{(\ell)}$ may be assumed to have full column-rank. The noise $\mathbf{n}^{(\ell)} \in \mathbb{C}^{n_\ell}, \ell = 1,\ldots,L$, is independent of the signals and independent across all L data sets, zero-mean Gaussian with *unknown* covariance matrices $\mathbf{R}_{n^{(\ell)}n^{(\ell)}}$. The cross-covariance matrix between two signal vectors $\mathbf{s}^{(i)} \in \mathbb{C}^{Q_i}$ and $\mathbf{s}^{(j)} \in \mathbb{C}^{Q_j}$ is (assuming here w.l.o.g. that $Q_i \leq Q_j$)

$$\mathbf{R}_{s^{(i)}s^{(j)}} = \left[\mathbf{diag}\left(\rho_1^{(i,j)}\sigma_1^{(i)}\sigma_1^{(j)},\ldots,\rho_{Q_i}^{(i,j)}\sigma_{Q_i}^{(i)}\sigma_{Q_i}^{(j)}\right), \mathbf{0}_{Q_i\times(Q_j-Q_i)}\right]$$

where $\rho_q^{(i,j)}$ is the unknown correlation coefficient between $s_q^{(i)}$ and $s_q^{(j)}$. There are $d^{(i,j)}$ (which is a priori unknown) nonzero correlation coefficients, hence $d^{(i,j)}$ correlated components between $\mathbf{s}^{(i)}$ and $\mathbf{s}^{(j)}$. In this paper, we are interested in determining the number of components, d, that are correlated across *all* L data sets (or channels). This number d is less than or equal to the number of signals correlated between pairs of data sets. Figure 1 shows an example involving three data sets, where there are two correlated components between any pair of data sets, but only $d = 1$ component (shown in red) correlated across all three data sets.

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

The prior works [1]–[3] on model-order selection for multiple data sets assumed that $d = d^{(i,j)}$, $\forall i, j \in \{1, \ldots, L\}, i \neq j$. In this paper, we relax this assumption and allow cases such as the one shown in Fig. 1, where components correlated between one pair of data sets are uncorrelated with the remaining data sets. However, we still have to make the simplifying assumption that components correlated between one pair of data sets are either (i) correlated with all other remaining data sets or (ii) uncorrelated with all other remaining data sets. Hence, in the example shown in Fig. 1 we would not allow an additional nonzero correlation coefficient between $s_2^{(2)}$ and $s_2^{(3)}$ if $s_2^{(3)}$ and $s_2^{(1)}$ remain uncorrelated.

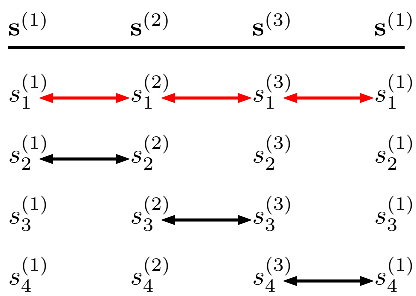

Fig. 1: Example for the correlation structure between three data sets. The arrow "↔" indicates that two components are correlated. The red arrow indicates the components that are correlated across all three data sets.

III. Proposed Detection Scheme

The d signals that are correlated across all data sets must be a subset of those signals that are correlated between pairs of data sets. In the following procedure, we first determine the number of correlated signals in all pairs of data sets (Step 1), and then extract only the correlated signals from the data sets (Step 2). Finally, in Step 3, we use an information-theoretic criterion to determine d.

A. Step 1: determine the number of correlated signals, $d^{(i,j)}$, between each pair of data sets $\left(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}\right), i, j \in \{1, \ldots, L\}, i \neq j$

The standard approach for determining the number of correlated signals between two zero-mean random vectors $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(j)}$ uses canonical correlation analysis (CCA) [9]. In order to estimate the canonical correlations, we collect M i.i.d. sample pairs $(\mathbf{x}_m^{(i)}, \mathbf{x}_m^{(j)}), m = 1, \ldots, M$, and arrange them in data matrices $\mathbf{X}^{(i)} = [\mathbf{x}_1^{(i)}, \ldots, \mathbf{x}_M^{(i)}]$ and $\mathbf{X}^{(j)} = [\mathbf{x}_1^{(j)}, \ldots, \mathbf{x}_M^{(j)}]$. From these we compute the sample covariance matrices $\hat{\mathbf{R}}_{i,i} = \mathbf{X}^{(i)}[\mathbf{X}^{(i)}]^H/M$, $\hat{\mathbf{R}}_{j,j} = \mathbf{X}^{(j)}[\mathbf{X}^{(j)}]^H/M$, and $\hat{\mathbf{R}}_{i,j} = \mathbf{X}^{(i)}[\mathbf{X}^{(j)}]^H/M$. The sample canonical correlations $1 \geq \hat{k}_1^{(i,j)} \geq \cdots \geq \hat{k}_{\min(n_i,n_j)}^{(i,j)} \geq 0$ are the singular values of the sample coherence matrix $\hat{\mathbf{R}}_{i,i}^{-1/2}\hat{\mathbf{R}}_{i,j}\hat{\mathbf{R}}_{j,j}^{-H/2}$ [10].

Based on the sample canonical correlations, there are two main methods for determining $d^{(i,j)}$: hypothesis tests [11] and information theoretic criteria (ITC) [8]. However, these traditional methods are inapplicable in a small sample-size scenario. This is because the sample canonical correlations can be extremely misleading as they are substantially overestimated [12]. Indeed, if $M < n_i + n_j$ then $n_i + n_j - M$ sample canonical correlations are identically one regardless of the two-channel model that generates the data samples.

To avoid this, a rank-reduction preprocessing may be applied before subjecting the two data sets to CCA. The most commonly used preprocessing is PCA. The PCA proprocessing for $\mathbf{X}^{(i)}$ and $\mathbf{X}^{(j)}$ proceeds as follows: We first determine the singular value decompositions (SVDs) of $\mathbf{X}^{(i)} = \mathbf{U}_i\mathbf{\Sigma}_i\mathbf{V}_i^H$ and $\mathbf{X}^{(j)} = \mathbf{U}_j\mathbf{\Sigma}_j\mathbf{V}_j^H$. Then the reduced-rank PCA descriptions of $\mathbf{X}^{(i)}$ and $\mathbf{X}^{(j)}$ are

$$\bar{\mathbf{X}}^{(i)} = \mathbf{U}_i(:, 1:r_i)\mathbf{\Sigma}_i(1:r_i, 1:r_i)\mathbf{V}_i^H(:, 1:r_i), \quad (1)$$

$$\bar{\mathbf{X}}^{(j)} = \mathbf{U}_j(:, 1:r_j)\mathbf{\Sigma}_j(1:r_j, 1:r_j)\mathbf{V}_j^H(:, 1:r_j), \quad (2)$$

where $\mathbf{U}_i(:, 1:r_i)$ and $\mathbf{V}_i(:, 1:r_i)$ consist of the r_i columns associated with the largest r_i singular values in $\mathbf{\Sigma}_i$, and $\mathbf{\Sigma}_i(1:r_i, 1:r_i)$ is a submatrix of $\mathbf{\Sigma}_i$ containing the first r_i rows and the first r_i columns of $\mathbf{\Sigma}_i$. The other matrices are defined analogously. Now let $\hat{\bar{\mathbf{R}}}_{i,i} = \bar{\mathbf{X}}^{(i)}[\bar{\mathbf{X}}^{(i)}]^H/M$, $\hat{\bar{\mathbf{R}}}_{j,j} = \bar{\mathbf{X}}^{(j)}[\bar{\mathbf{X}}^{(j)}]^H/M$, and $\hat{\bar{\mathbf{R}}}_{i,j} = \bar{\mathbf{X}}^{(i)}[\bar{\mathbf{X}}^{(j)}]^H/M$ be the sample covariance matrices computed from the reduced-rank PCA descriptions. The corresponding estimated canonical correlations $\hat{k}_n^{(i,j)}(r_i, r_j)$, which depend on the PCA ranks, are the singular values of the reduced-rank sample coherence matrix $\hat{\bar{\mathbf{R}}}_{i,i}^{-1/2}\ \hat{\bar{\mathbf{R}}}_{i,j}\ \hat{\bar{\mathbf{R}}}_{j,j}^{-H/2}$. Here, the pseudoinverse may be necessary because $\hat{\bar{\mathbf{R}}}_{i,i}$ and $\hat{\bar{\mathbf{R}}}_{j,j}$ may be rank deficient.

We need to determine not only the number of correlated signals $d^{(i,j)}$, but also the ranks r_i and r_j for the PCA preprocessing step. For this, the so-called "max-min" detector was recently proposed in [6]. As far as we know, this is the only technique capable of handling the combined PCA-CCA setup in the sample-poor regime. The "max-min" detector is a hypothesis test based on the reduced-rank Bartlett-Lawley statistic, which in our case is

$$C(r_i, r_j, s) = -2\left(M - s - \frac{r_i + r_j + 1}{2} + \sum_{n=1}^{s} \hat{k}_n^{-2}(r_i, r_j)\right) \times \ln \prod_{n=s+1}^{\min(r_i, r_j)} \left(1 - \hat{k}_n^2(r_i, r_j)\right).$$

As long as the PCA ranks r_i and r_j are small compared to the number of samples M—which is the case when r_i and r_j are smaller than $r_{\max} = \min(n_i, n_j, \lfloor M/3 \rfloor)$—this statistic is approximately χ^2-distributed with $2(r_i - s)(r_j - s)$ degrees of freedom. The "max-min" detector now chooses

$$\hat{d}^{(i,j)} = \max_{r_i, r_j} \min_{s} \left\{s : C(r_i, r_j, s) < T(r_i, r_j, s)\right\}, \quad (3)$$

where $r_i, r_j \in \{1, \ldots, r_{\max}\}$ and $s \in \{0, \ldots, \min(r_i, r_j) - 1\}$. The r_i and r_j that lead to $\hat{d}^{(i,j)}$ are chosen to be the PCA ranks. The min-operator chooses the smallest s such that the statistic $C(r_i, r_j, s)$ falls below the threshold $T(r_i, r_j, s)$, which is selected to ensure a given probability of false alarm. If there is no such s, then the min-operator chooses $s = \min(r_i, r_j)$. A more detailed discussion and motivation for this approach can be found in [7].

B. Step 2: keep the correlated signals for each pair of data sets $\left(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}\right), i, j \in \{1, \ldots, L\}, i \neq j$

The canonical vectors for a given pair of data sets $(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$ are obtained as

$$\mathbf{w}_i^{(i,j)} = \left(\mathbf{F}^{(i,j)}(:, 1:d^{(i,j)})\right)^H \mathbf{R}_{i,i}^{-1/2} \mathbf{x}^{(i)},$$
$$\mathbf{w}_j^{(i,j)} = \left(\mathbf{G}^{(i,j)}(:, 1:d^{(i,j)})\right)^H \mathbf{R}_{j,j}^{-1/2} \mathbf{x}^{(j)},$$

where $\mathbf{F}^{(i,j)}(:, 1:d^{(i,j)})$ and $\mathbf{G}^{(i,j)}(:, 1:d^{(i,j)})$ contain the first $d^{(i,j)}$ left and right singular vectors, respectively, of the coherence matrix $\mathbf{R}_{i,i}^{-1/2}\mathbf{R}_{i,j}\mathbf{R}_{j,j}^{-H/2}$. It is well known for CCA-based blind source separation [13] that the canonical vectors $\mathbf{w}_i^{(i,j)}$ and $\mathbf{w}_j^{(i,j)}$ extract the $d^{(i,j)}$ signal components that are correlated between $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(j)}$ (up to some scaling factors).

In total, there are $\tilde{L} = \frac{1}{2}L(L-1)$ possible distinct pairings between $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(j)}$ for $i, j \in \{1, \ldots, L\}, i \neq j$. Let us arrange these pairs as follows:

$$\mathbf{w}_{\text{left}}^{(\ell)} = \begin{cases} \mathbf{w}_1^{(1,\ell+1)} & \ell = 1, \ldots, L-1 \\ \mathbf{w}_2^{(2,\ell-L+3)} & \ell = L, \ldots, 2L-3 \\ \vdots & \vdots \\ \mathbf{w}_{L-1}^{(L-1,L)} & \ell = \tilde{L} \end{cases}$$
$$\mathbf{w}_{\text{right}}^{(\ell)} = \begin{cases} \mathbf{w}_{\ell+1}^{(1,\ell+1)} & \ell = 1, \ldots, L-1 \\ \mathbf{w}_{\ell-L+3}^{(2,\ell-L+3)} & \ell = L, \ldots, 2L-3 \\ \vdots & \vdots \\ \mathbf{w}_{L}^{(L-1,L)} & \ell = \tilde{L} \end{cases}$$

For the "left" data set, the subscript on $\mathbf{w}_i$ matches the *left* index of its superscript $^{(i,\cdot)}$, and for the "right" data set, the subscript matches the right index of its superscript. This means that for a given ℓ, $\mathbf{w}_{\text{left}}^{(\ell)}$ and $\mathbf{w}_{\text{right}}^{(\ell)}$ represent the "left" and "right" canonical vectors corresponding to the same pair of data sets.

Since we work with samples, we apply the procedure outlined above to the reduced-rank data matrices $\bar{\mathbf{X}}^{(i)}$ and $\bar{\mathbf{X}}^{(j)}$ obtained in (1) and (2), using the reduced-rank sample covariance matrices $\hat{\bar{\mathbf{R}}}_{i,i}, \hat{\bar{\mathbf{R}}}_{i,j}, \hat{\bar{\mathbf{R}}}_{j,j}$. This yields sample canonical matrices $\mathbf{W}_{\text{left}}^{(\ell)}$ and $\mathbf{W}_{\text{right}}^{(\ell)}$.

C. Step 3: estimate the number of signals, d, correlated across all data sets

Because of the assumptions on correlation structure made in the last paragraph of Section 2, the ranks of the cross-covariance matrices $E\left\{\mathbf{w}_{\text{left}}^{(i)}\left(\mathbf{w}_{\text{right}}^{(j)}\right)^H\right\}$, for $i = 1, \ldots, L-1$ and $j = i+1, \ldots, L$, are all equal to d. By restricting the index of $\mathbf{w}_{\text{right}}^{(j)}$ to be greater than the index of $\mathbf{w}_{\text{left}}^{(i)}$, we avoid using two canonical vectors obtained from the same data set, e.g. $\mathbf{w}_2^{(2,3)} (= \mathbf{w}_{\text{left}}^{(L)})$ and $\mathbf{w}_2^{(1,2)} (= \mathbf{w}_{\text{right}}^{(1)})$.

For each pair of data sets $(\mathbf{w}_{\text{left}}^{(i)}, \mathbf{w}_{\text{right}}^{(j)})$, $i < j$, the information-theoretic criterion for two data sets [8], which is a function of the assumed number of correlated signals, s, is (on average) minimized at $s = d$. This means that the sum of these information-theoretic criteria over all distinct pairs of data sets is also minimized at $s = d$. This motivates the following selection rule:

$$\hat{d} = \arg \min_{s=0,\ldots,p_{\min}} \sum_{i=1}^{\tilde{L}-1} \sum_{j=i+1}^{\tilde{L}} \left[M \ln \prod_{n=1}^{s} \left(1 - \left(\hat{\gamma}_n^{(i,j)}\right)^2\right) + \ln(M) s \left(p^{(i)} + p^{(j)} - s\right) \right], \quad (4)$$

where $\hat{\gamma}_n^{(i,j)}$ denotes the nth largest sample canonical correlation computed from $\mathbf{W}_{\text{left}}^{(i)}$ and $\mathbf{W}_{\text{right}}^{(j)}$, $p_{\min} = \min(p^{(1)}, \ldots, p^{(\tilde{L})})$, with $p^{(i)}$ and $p^{(j)}$ representing the dimensions of $\mathbf{W}_{\text{left}}^{(i)}$ and $\mathbf{W}_{\text{right}}^{(j)}$, respectively. That is, $p^{(i)}$ and $p^{(j)}$ are given by the number of correlated signals in the corresponding pairs of data sets. The first term in the sum in (4) is the log-likelihood function depending on the assumed number of correlated signals, s, and the second term is the penalty term, which depends on the degrees of freedom of the model and number of samples.

IV. Simulation Results

In this section, the performance of our proposed scheme is examined. We consider a setting with five complex-valued data sets, each of which has 9 signals with variances $[7, 8, 9, 8, 6, 7, 3, 2, 1]$. The noise in each data set is generated independently with unit variance and is either white or colored. In the colored case, the auto-covariance matrix has elements $\left[\mathbf{R}_{n^{(\ell)}n^{(\ell)}}\right]_{i,j} = 0.4^{|i-j|}, \ell \in \{1, \ldots, 5\}, i, j \in \{1, \ldots, n_\ell\}$. The mixing matrices $\mathbf{A}^{(\ell)}, \ell \in \{1, \ldots, 5\}$ are randomly generated unitary matrices. The probability of false alarm for the "max-min" detector in Step 1 has been set to 0.005.

We first consider an example with varying sample sizes and fixed dimensions $n_1 = 30, n_2 = 40, n_3 = 50, n_4 = 60, n_5 = 70$. The signals between the five data sets are correlated with correlation coefficients

$$\boldsymbol{\rho}^{(1,2)} = [0.80, 0.90, 0.86, 0.79, 0.94, 0.73, 0, \ldots, 0],$$
$$\boldsymbol{\rho}^{(1,3)} = [0.78, 0.85, 0.92, 0.85, 0.80, 0.78, 0, \ldots, 0],$$
$$\boldsymbol{\rho}^{(1,4)} = [0.83, 0.85, 0.78, 0.80, 0.80, 0.82, 0, \ldots, 0],$$
$$\boldsymbol{\rho}^{(1,5)} = [0.78, 0.91, 0.75, 0.90, 0.90, 0.80, 0, \ldots, 0],$$
$$\boldsymbol{\rho}^{(2,3)} = [0.76, 0.88, 0.79, 0.75, 0.87, 0.89, 0, \ldots, 0],$$
$$\boldsymbol{\rho}^{(2,4)} = [0.86, 0.88, 0.92, 0.83, 0.81, 0.80, 0, \ldots, 0],$$
$$\boldsymbol{\rho}^{(2,5)} = [0.81, 0.92, 0.76, 0.93, 0.91, 0.86, 0, \ldots, 0],$$
$$\boldsymbol{\rho}^{(3,4)} = [0.84, 0.85, 0.79, 0.77, 0.91, 0.82, 0, \ldots, 0],$$
$$\boldsymbol{\rho}^{(3,5)} = [0.76, 0.84, 0.82, 0.73, 0.81, 0.90, 0, \ldots, 0],$$
$$\boldsymbol{\rho}^{(4,5)} = [0.80, 0.90, 0.89, 0.73, 0.89, 0.84, 0, \ldots, 0].$$

Figure 2 shows simulation results for this scenario for white and colored noise and for different number of samples M,

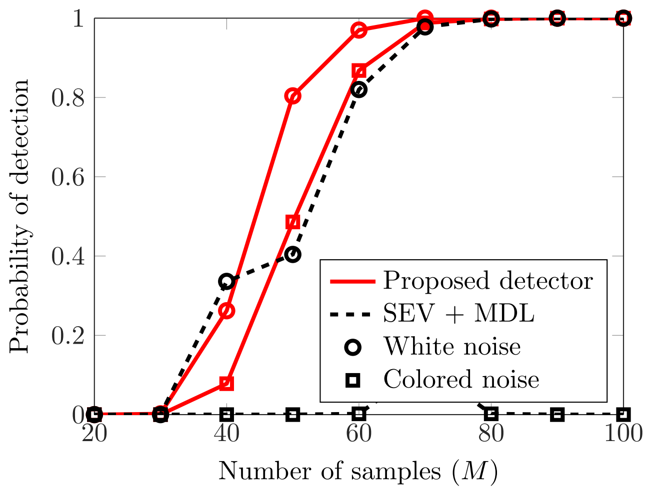

Fig. 2: Detection performance for different sample sizes for both white and colored noise. The number of correlated signals is identical for each pair of data sets.

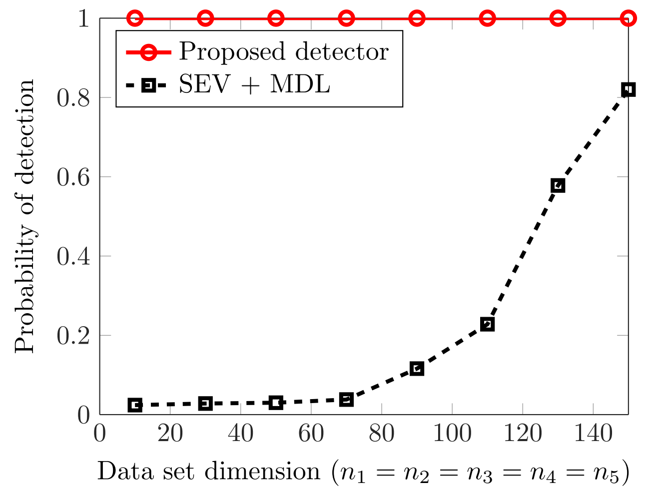

Fig. 3: Detection performance for different dimensions and white noise. The number of correlated signals is *not* identical for each pair of data sets.

based on 500 independent Monte Carlo trials. This figure also shows results for a competing approach, which is based on the minimum description length (MDL) information-theoretic criterion proposed in [2] for multiple data sets. Because [2] is not designed to handle small sample support, we used PCA preprocessing steps whose dimensions were determined by the sample eigenvalue-based (SEV) technique [14]. The correlation structure in our example satisfies the assumption made in [2] that the number of correlated signals is identical for each pair of data sets. Indeed, for white noise, our proposed technique and the combination of SEV + MDL perform comparably. However, for colored noise, the combination of SEV + MDL does not work at all because SEV is not able to separate the signal from the noise subspace. Our technique, on the other hand, still performs well.

We now consider the same system for white noise, but with varying dimensions and fixed sample size $M = 100$. The signals are correlated with correlation coefficients

$$\begin{aligned}
\boldsymbol{\rho}^{(1,2)} &= [0.80, 0.90, 0.86, 0.79, 0.00, 0.00, 0, \ldots, 0],\\
\boldsymbol{\rho}^{(1,3)} &= [0.78, 0.85, 0.92, 0.00, 0.00, 0.00, 0, \ldots, 0],\\
\boldsymbol{\rho}^{(1,4)} &= [0.83, 0.85, 0.78, 0.00, 0.00, 0.00, 0, \ldots, 0],\\
\boldsymbol{\rho}^{(1,5)} &= [0.78, 0.91, 0.75, 0.00, 0.00, 0.00, 0, \ldots, 0],\\
\boldsymbol{\rho}^{(2,3)} &= [0.76, 0.88, 0.79, 0.00, 0.87, 0.00, 0, \ldots, 0],\\
\boldsymbol{\rho}^{(2,4)} &= [0.86, 0.88, 0.92, 0.00, 0.00, 0.00, 0, \ldots, 0],\\
\boldsymbol{\rho}^{(2,5)} &= [0.81, 0.92, 0.76, 0.00, 0.00, 0.00, 0, \ldots, 0],\\
\boldsymbol{\rho}^{(3,4)} &= [0.84, 0.85, 0.79, 0.00, 0.00, 0.00, 0, \ldots, 0],\\
\boldsymbol{\rho}^{(3,5)} &= [0.76, 0.84, 0.82, 0.00, 0.00, 0.90, 0, \ldots, 0],\\
\boldsymbol{\rho}^{(4,5)} &= [0.80, 0.90, 0.89, 0.00, 0.00, 0.00, 0, \ldots, 0].
\end{aligned}$$

In this setup, only the first three signals in each data set are correlated across all data sets, thus the assumption made in [2] is violated. Indeed, Fig. 3 shows that the competing SEV + MDL approach does not work for small data set dimensions, whereas our proposed technique performs very well for all dimensions.

V. Conclusions

We have presented an approach to estimate the number of signals correlated across multiple data sets when there is only small sample support available. As far as we know, no alternative approach has been suggested in the literature so far. Since this is a challenging problem, we have proposed a suboptimal heuristic technique that decomposes the original problem into several sequential problems. Nevertheless, Monte Carlo simulations show promising performance.

Acknowledgments

This research was supported by the Alfried Krupp von Bohlen und Halbach foundation under the program “Return of German scientists from abroad”, and the German Research Foundation (DFG) under grant SCHR 1384/3-1. The work of D. Ramírez has been partly supported by Ministerio de Economía of Spain under projects: COMPREHENSION (TEC2012-38883-C02-01), OTOSIS (TEC2013-41718-R), and the COMONSENS Network (TEC2015-69648-REDC), by the Ministerio de Economía of Spain jointly with the European Commission (ERDF) under project ADVENTURE (TEC2015-69868-C2-1-R), and by the Comunidad de Madrid under project CASI-CAM-CM (S2013/ICE-2845).

References

[1] M. Bhandary, “Detection of the number of signals in the presence of white noise in decentralized processing,” *IEEE Transactions on Signal Processing*, vol. 46, no. 3, pp. 800–803, 1998.

[2] Y. H. Wu, K. W. Tam, and F. Li, “Determination of number of sources with multiple arrays in correlated noise fields,” *IEEE Transactions on Signal Processing*, vol. 50, no. 6, pp. 1257–1260, June 2002.

[3] T. Hasija, Y. Song, P. J. Schreier, and D. Ramírez, "Detecting the dimension of the subspace correlated across multiple data sets in the sample poor regime," *IEEE Workshop on Statistical Signal Processing*, June 2016.

[4] T. Adali, Y. L. Schwartz, and V. D. Calhoun, "Multi-modal data fusion using source separation: Application to medical imaging," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1494–1506, August 2015.

[5] Y. L. Schwartz, Y. Song, P. J. Schreier, V. D. Calhoun, and T. Adali, "Sample-poor estimation of order and common signal subspace with application to fusion of medical imaging data," *NeuroImage*, vol. 134, pp. 486–493, July 2016.

[6] Y. Song, P. J. Schreier, and N. J. Roseveare, "Determining the number of correlated signals between two data sets using PCA-CCA when sample support is extremely small," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, April 2015.

[7] Y. Song, P. J. Schreier, D. Ramírez, and T. Hasija, "Canonical correlation analysis of high-dimensional data with very small sample support," *Signal Processing*, vol. 128, pp. 449–458, November 2016.

[8] P. Stoica, K. M. Wong, and Q. Wu, "On a nonparametric detection method for array signal processing in correlated noise fields," *IEEE Transactions on Signal Processing*, vol. 44, no. 4, pp. 1030–1032, April 1996.

[9] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3-4, pp. 321, 1936.

[10] L. L. Scharf and C. T. Mullis, "Canonical coordinates and the geometry of inference, rate, and capacity," *IEEE Transactions on Signal Processing*, vol. 48, no. 3, pp. 824–831, March 2000.

[11] D. N. Lawley, "Tests of significance in canonical analysis," *Biometrika*, vol. 46, no. 1/2, pp. 59–66, June 1959.

[12] A. Pezeshki, L. L. Scharf, M. R. Azimi-Sadjadi, and M. Lundberg, "Empirical canonical correlation analysis in subspaces," *Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 7–10, November 2004.

[13] Y. O. Li, T. Adali, W. Wang, and V. D. Calhoun, "Joint blind source separation by multiset canonical correlation analysis," *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 3918–3929, October 2009.

[14] R. R. Nadakuditi and A. Edelman, "Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 2625–2638, July 2008.