Tensor-Based Sparsity Order Estimation for Big Data Applications

Kefei Liu*, Florian Roemer†, João Paulo C. L. da Costa‡,
Jie Xiong§, Yi-Sheng Yan§, Wen-Qin Wang§ and Giovanni Del Galdo†¶
*Indiana University School of Medicine, Indianapolis, IN, USA
†Ilmenau University of Technology, Institute for Information Technology, Germany
‡Department of Electrical Engineering, University of Brasília, Brasília, Brazil
§School of Communication Engineering, University of Electronic Science and Technology of China, ChengDu, China
¶Fraunhofer Institute for Integrated Circuits IIS, Ilmenau, Germany

***Abstract*—In Big Data Processing we typically face very large data sets that are highly structured. To save the computation and storage cost, it is desirable to extract the essence of the data from a reduced number of observations. One example of such a structural constraint is sparsity. If the data possesses a sparse representation in a suitable domain, it can be recovered from a small number of linear projections into a low-dimensional space. In this case, the degree of sparsity, referred to as sparsity order, is of high interest. It has recently been shown that if the measurement matrix obey certain structural constraints, one can estimate the sparsity order directly from the compressed data. The rich structure of the measurement matrix allows to rearrange the multiple-snapshot measurement vectors into a fourth-order tensor with rank equal to the desired sparsity order. In this paper, we exploit the multilinear structure of the data for accurate sparsity order estimation with improved identifiability. We discuss the choice of the parameters, i.e., the block size, block offset, and number of blocks, to maximize the sparsity order that can be inferred from a certain number of observations, and compare state-of-the-art order selection algorithms for sparsity order estimation under the chosen parameter settings. By performing an extensive campaign of simulations, we show that the discriminant function based method and the random matrix theory algorithm outperform other approaches in small and large snapshot-number scenarios, respectively.**

***Index Terms*—Compressed sensing, sparsity order, order selection, tensor decomposition**

I. Introduction

The availability of low-cost massive digital storage devices in connection with the ever-growing availability of significant computational power, sensing devices and high-speed communication links has led to an unprecedented need of processing very large sets of data efficiently. Applications include health care, government and public sector, natural resource management, commerce, social networking and many more [1]. The tremendous size of the data calls for new paradigms in signal processing as classical tools like PCA are inapplicable due to their complexity [1], [2].

Despite its size, the big data sets are typically highly structured [2]. It is therefore highly desirable to extract the essence of the data (its structure) from a reduced-dimensional view (also referred to as a sketch [3]) of the data. One example of such a structural constraint is sparsity, i.e., the assumption that the data possesses a sparse representation in a (known) basis. Leveraging the results from the compressed sensing field [4], [5], we know that in this case, the structure of the data can be inferred from a relatively small number of linear projections of the data, provided that the degree of sparsity (the sparsity order) is low.

To this end, the knowledge of the sparsity order appears to be an important factor as it allows to judge the complexity of the data. On the one hand, this information can be used to tune the number of projections that need to be taken, on the other hand, it can assist a subsequent sparse recovery stage as well.

In [6], we have shown that the sparsity order estimation (SOE) problem can be transferred to the problem of estimating the rank of an observation matrix constructed from rearranging the low-dimensional observation vector into a matrix form. As shown in [6] this is possible if and only if the sensing matrix is equal to the Khatri-Rao product of two full rank matrices. Moreover, when the columns of the observation matrix overlap, one of the two factors needs to be a Vandermonde matrix. The structure of the Vandermonde sensing matrix allows to rearrange the observation vector into a third-order tensor in the single measurement vector case or a fourth-order tensor in the multiple measurement vector case, respectively. In both cases, the rank of the tensor is equal to the desired sparsity order.

In this paper, we exploit the rich multilinear structure of the observed data to improve the identifiability and detection performance. We discuss the choice of the parameters, i.e., the block size, block offset, and number of blocks, to maximize the sparsity order that can be inferred from a certain number of observations, and compare state-of-the-art model order selection algorithms for sparsity order estimation under such parameter settings. Simulation results show that in the presence of independent and identically distributed (i.i.d.) Gaussian white noise, the discriminant function based method (DFBM) in [7] presents the overall best performance in the small snapshot-number scenario and the random matrix theory (RMT) algorithm in [8] performs the best in the large snapshot-number scenario.

The remainder of the paper is organized as follows. Section II presents the data model for sparsity order estimation,

ISBN 978-0-9928626-7-1

and shows how to rearrange the measurement vectors at multiple snapshots into a fourth-order tensor with rank equal to the desired sparsity order. Section III discusses how to exploit the multilinear structure in the measured data to improve the identifiability and performance of sparsity order estimation. Simulations results are presented in Section IV to assess the sparsity order estimation performance of state-of-the-art algorithms. Finally, conclusions are drawn in Section V.

II. Data Model

Let us assume that we observe one or multiple snapshots of a data vector $\boldsymbol{x}(t) \in \mathbb{C}^{N\times 1}$ (a signal or an image), $t = 1, 2, \ldots, T$, where N is the number of samples or pixels and can be very big. We assume that $\boldsymbol{x}(t)$ has a sparse representation in a known basis $\boldsymbol{A} \in \mathbb{C}^{N\times N}$ which is assumed unitary without loss of generality. Therefore, we can write $\boldsymbol{x}(t) = \boldsymbol{A}\cdot\boldsymbol{s}(t)$, where $\boldsymbol{s}(t) \in \mathbb{C}^{N\times 1}$ is K-sparse with $K \ll N$. In the multiple snapshot case, we assume that the sparsity pattern is constant over the T snapshots, i.e., the support of $\boldsymbol{s}(t)$ does not change over T. In practice, this is often a reasonable assumption for some not too large observation window where the signal is stationary. Note that if the support pattern does change within T snapshots, it is possible to accommodate this change by defining an effective support set $\mathcal{S} = \bigcup_{t=1,2,\ldots,T} \operatorname{supp}\{\boldsymbol{s}(t)\}$ as long as $K = |\mathcal{S}|$ is still small enough in the sense that $K \ll N$.

Since the dimension N is very large, we do not want to carry out any operations on the vector $\boldsymbol{x}(t)$ directly. Instead, we assume that all we have available is a set of $M \ll N$ linear measurements of $\boldsymbol{x}(t)$, which can be described via

$$\boldsymbol{y}(t) = \boldsymbol{\Psi}\cdot\boldsymbol{x}(t), \tag{1}$$

where $\boldsymbol{\Psi} \in \mathbb{C}^{M\times N}$ is the measurement kernel. As explained in [6], for the purpose of sparsity order estimation, we can segment each single observation vector $\boldsymbol{y}(t)$ into B overlapping smaller vectors of size m, and obtain an observation matrix $\boldsymbol{Y}(t) \in \mathbb{C}^{m\times B}$. Specifically, the b-th column of $\boldsymbol{Y}(t)$ contains samples $(b-1)\cdot p+1$ up to $(b-1)\cdot p+m$, where p is the block advance and we have $B = \frac{M-m}{p} + 1$. We can then show that for a K-sparse signal we have $\operatorname{rank}\{\boldsymbol{Y}(t)\} = K$ if and only if we choose $\boldsymbol{\Psi} = (\boldsymbol{C} \diamond \boldsymbol{\Phi}_0)\cdot\boldsymbol{A}^{\mathrm{H}}$ where $\diamond$ represents the Khatri-Rao product (columnwise Kronecker product) of two matrices, and $\boldsymbol{\Phi}_0 \in \mathbb{C}^{p\times N}$ and $\boldsymbol{C} \in \mathbb{C}^{\frac{M}{p}\times N}$ both have a Kruskal-rank[1] of at least K. In addition, if $p < m$, which means that adjacent blocks are overlapping, $\boldsymbol{C}$ needs to be a Vandermonde matrix.

Note that with the above choice of $\boldsymbol{\Psi}$, the observation matrix $\boldsymbol{Y}(t) \in \mathbb{C}^{m\times B}$ can be written as

$$\boldsymbol{Y}(t) = \left(\boldsymbol{C}_{\frac{m}{p}} \diamond \boldsymbol{\Phi}_0\right)\cdot\operatorname{diag}\{\boldsymbol{s}(t)\}\cdot\boldsymbol{C}_B^{\mathrm{T}}, \tag{2}$$

where $\boldsymbol{C}_q$ denotes the first q rows of the (Vandermonde) matrix $\boldsymbol{C}$. It is clear from (2) that instead of a measurement matrix $\boldsymbol{Y}(t)$ we can define a measurement tensor $\boldsymbol{\mathcal{Y}}$ of size $p \times \frac{m}{p} \times B \times T$ which is given by

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{I}}_{4,N} \times_1 \boldsymbol{\Phi}_0 \times_2 \boldsymbol{C}_{\frac{m}{p}} \times_3 \boldsymbol{C}_B \times_4 \boldsymbol{S}^{\mathrm{T}} \tag{3}$$

where $\boldsymbol{\mathcal{I}}_{4,N}$ is the four-way $N \times N \times N \times N$ identity tensor and $\boldsymbol{S} \in \mathbb{C}^{N\times T}$ collects the complex-valued coefficients $\boldsymbol{s}(t)$ at all time instants. The link between (2) and (3) is given by $\boldsymbol{Y}(t) = \left[\boldsymbol{\mathcal{Y}}_{(:,:,:,t)}\right]_{(3)}^{\mathrm{T}}$, where $[\cdot]_{(n)}$ denotes the n-mode matrix unfolding, cf. [9], and $\boldsymbol{\mathcal{Y}}_{(:,:,:,t)}$ represents the t-th 4-mode slice (of size $p \times \frac{m}{p} \times B$) of the four-way tensor $\boldsymbol{\mathcal{Y}}$. Since $\boldsymbol{S}$ is K-row-sparse, we can reformulate (3) into

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{I}}_{4,K} \times_1 \boldsymbol{\Phi}_{0,K} \times_2 \boldsymbol{C}_{\frac{m}{p},K} \times_3 \boldsymbol{C}_{B,K} \times_4 \boldsymbol{S}_K^{\mathrm{T}}, \tag{4}$$

where the additional index K means that the matrix is limited to the K columns corresponding to the active support set (i.e., the non-zero rows of $\boldsymbol{S}$). As evident from (4), $\boldsymbol{\mathcal{Y}}$ has a rank of K. Compared with the matrix representation in (2), the tensor representation in (4) manifests the multilinear structure in the measured data which allows to increase the maximum number of detectable signals (i.e., detection capability) [10] as well as improve the order estimation accuracy.

So far, the data model was formulated in the absence of noise. In practice, we may have to account for an additive noise term, either due to noisy measurements or to signals that are not exactly but only approximately K-sparse (e.g., compressible signals) [11]. Therefore, our goal is to infer the sparsity order K from an observed noisy version of (4).

III. Sparsity Order Estimation Exploiting the Tensor Structure

Tensor-based approaches to multilinear rank estimation such as the core consistency diagnostic [12], difference in fit [13] and numerical convex hull [14], require computationally expensive CANDECOMP/PARAFAC decomposition [15] which factorizes a tensor into a sum of component rank-one tensors. A computationally more attractive way is to rearrange the elements of the tensor data into a matrix with the unfolding operation [9], and then use the eigenspectrum associated with the unfolded matrix for detecting the number of signals. Under independent and identically distributed (i.i.d.) Gaussian assumption of the data samples, the eigenvalues that are associated with the noise-only space will be similar in magnitude and much smaller than the signal-bearing eigenvalues. The number of signals can therefore be estimated either by testing the equality of the smallest eigenvalues using information theoretic criterion, e.g., the minimum description length (MDL) and Akaike information criterion (AIC) [16], or by analyzing the properties of the noise-only eigenvalues based on the RMT to detect the gap between the signal-bearing and noise-only eigenvalues [8], [17].

A. Choice of Unfolded Matrix of $\boldsymbol{\mathcal{Y}}$ and Size Parameters to Improve Identifiability

The unfolding, also known as matricization, of a tensor is the operation for transforming a tensor into a matrix. The relationship between the rank of an unfolded matrix of a tensor

[1]The Kruskal-rank of a matrix $\boldsymbol{A}$ is defined as the largest integer k such that any set of k columns of $\boldsymbol{A}$ are linearly independent. Obviously, if a matrix $\boldsymbol{A}$ has full Kruskal-rank, it also has full rank.

ISBN 978-0-9928626-7-1 © EURASIP 2017

and the tensor rank is investigated in Theorem 4.2 of [18]. The results show that under mild conditions the rank of a tensor is equal to its mode rank with probability 1. Since the sparsity order is unknown, for safety we should employ the unfolded matrix that allows us to be able to detect the maximum sparsity order. For the eigenvalue-based model order selection schemes [8], [16], [17], this often means that we choose the unfolded matrix of $\boldsymbol{\mathcal{Y}}$ with maximize size[2].

However, from the construction of $\boldsymbol{\mathcal{Y}}$ in Section II, the largest rank that a unfolded matrix of $\boldsymbol{\mathcal{Y}}$ is actually able to accommodate may be less than its size. For example, the maximum column rank of $[\boldsymbol{\mathcal{Y}}]_{(4)}$ is M instead of mB. The unfolded matrices of the fourth-order tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{C}^{p\times\frac{m}{p}\times B\times T}$ in Eq. (3), their sizes, and the largest rank that they are able to accommodate, are summarized in Table I.

TABLE I
THE UNFOLDED MATRICES OF $\boldsymbol{\mathcal{Y}} \in \mathbb{C}^{p\times\frac{m}{p}\times B\times T}$ IN EQ. (3), THEIR SIZES AND THE LARGEST RANK THAT THEY ARE ABLE TO ACCOMMODATE.

Unfolding of $\boldsymbol{\mathcal{Y}}$		Size	Maximum Rank
1-mode unfolding:	$[\boldsymbol{\mathcal{Y}}]_{(1)}$	$p \times \frac{mBT}{p}$	$\min\left\{p, \frac{MT}{p}\right\}$
2-mode unfolding:	$[\boldsymbol{\mathcal{Y}}]_{(2)}$	$\frac{m}{p} \times pBT$	$\min\left\{\frac{m}{p}, pBT\right\}$
3-mode unfolding:	$[\boldsymbol{\mathcal{Y}}]_{(3)}$	$B \times mT$	$\min\{B, mT\}$
4-mode unfolding:	$[\boldsymbol{\mathcal{Y}}]_{(4)}$	$T \times mB$	$\min\{T, M\}$
(1,2)-mode unfolding:	$[\boldsymbol{\mathcal{Y}}]_{(1,2)}$	$m \times BT$	$\min\{m, BT\}$
(1,3)-mode unfolding:	$[\boldsymbol{\mathcal{Y}}]_{(1,3)}$	$pB \times \frac{mT}{p}$	$\min\left\{pB, \frac{mT}{p}\right\}$
(1,4)-mode unfolding:	$[\boldsymbol{\mathcal{Y}}]_{(1,4)}$	$pT \times \frac{mB}{p}$	$\min\left\{pT, \frac{M}{p}\right\}$

In Table I, we have chosen only one unfolded matrix from each pair of mutually-transposed unfolded matrices of $\boldsymbol{\mathcal{Y}}$, since they share a common set of non-zero eigenvalues. In the first four unfolded matrices, each column is formed by one mode of the tensor while in the last three matrices each column is formed by two modes of the tensor [19], [20]. Note that the maximum column rank of $[\boldsymbol{\mathcal{Y}}]_{(1)}$, $[\boldsymbol{\mathcal{Y}}]_{(4)}$ and $[\boldsymbol{\mathcal{Y}}]_{(1,4)}$ are $\frac{MT}{p}$, M and $\frac{M}{p}$, instead of $\frac{mBT}{p}$, mB and $\frac{mB}{p}$, respectively.

We find the block length m, block advance p and number of blocks B to maximize the rank that the unfolded matrices of $\boldsymbol{\mathcal{Y}}$ can accommodate:

$$\begin{aligned} \underset{p,m,B}{\text{maximize}} \quad & f(p,m,B) \\ \text{subject to} \quad & (B-1)p+m=M, \\ & 1\le p<m\le M, 1\le B\le M, \\ & p,m,B \text{ are integers and } p \text{ divides } m, \end{aligned} \tag{5}$$

where $f(p,m,B)$ is defined at the top of next page.

By solving (5), the (1,2)-mode unfolding is optimal, with

$$p^*=1, m^*=\frac{T}{T+1}(M+1), B^*=\frac{1}{T+1}(M+1),$$
$$f(p^*,m^*,B^*)=\min\{m^*,B^*T\}=\frac{T}{T+1}(M+1) \tag{7}$$

[2]For a measured data matrix of size Q by L, where Q and L represent the spatial dimension and sample size, respectively, the number of non-zero eigenvalues available for use is $\min\{Q,L\}$. The eigenvalue-based order selection algorithms assume that at least one non-zero eigenvalue is a pure-noise eigenvalue. Therefore, the number of signals can not exceed $\min\{Q,L\}-1$.

when $T \le M$, and with

$$m^*=M, B^*=1,\ f(p^*,m^*,B^*)=\min\{m^*,B^*T\}=M \tag{8}$$

when $T > M$.

The proof is omitted due to space limit.

B. Sparsity Order Estimation

In (4), the factor matrices $\boldsymbol{C}_{\frac{m}{p},K}$ and $\boldsymbol{C}_{B,K}$ of $\boldsymbol{\mathcal{Y}}$ have a Vandermonder structure, and the noise component matrix contains repeated entries in the case of overlapping blocks. This violates the assumptions of the stationarity of the signal and/or statistical independence of noise underlying the state-of-the-art eigenvalue-based algorithms [8], [16], [17].

Therefore, we turn our attention to a heuristic discriminant function based method [7] for sparsity order estimation due to its computational simpleness and good empirical performance. The DFBM does not rely on assumptions of stationary signals or independent noise. Instead, it considers estimation of the model order K as a classification problem where the eigenvalues are divided into two classes, one is formed of signal-bearing eigenvalues and the other is formed of pure-noise eigenvalues. The boundary of the two classes, i.e., the location of the smallest signal-bearing eigenvalue, K, is found using the following cost function:

$$\mathrm{DFBM}(k)=g_1(k)-g_2(k), \tag{9}$$

where $g_1(k)$ and $g_2(k)$ are the discriminant functions associated to the signal and noise subspaces, respectively. The $g_1(k)$ is defined such that it is much larger in the signal subspace than in the noise subspace while $g_2(k)$ is much larger in the noise subspace than in the signal subspace.

Denote the sample eigenvalues associated with the chosen unfolded matrix of $\boldsymbol{\mathcal{Y}}$ as $\ell_1 \ge \ell_2 \ge \cdots \ge \ell_Q > 0$, where Q is the number of non-zero eigenvalues. Based on the assumption that the variation of the ordered signal-bearing eigenvalues and of the ordered pure-noise eigenvalues are both linear, the discriminant functions $g_1(k)$ and $g_2(k)$ are derived in [7]:

$$g_1(k)=\frac{\ell_k}{\sum_{i=2}^{Q}\ell_i}, \quad k=2,3,\ldots,Q \tag{10}$$

and

$$g_2(k)=\frac{\xi_k}{\sum_{i=1}^{Q-1}\xi_i}, \quad k=1,2,\ldots,Q-1 \tag{11}$$

where $\xi_k=1-\alpha\frac{\ell_k-\mu_k}{\mu_k}$ with

$$\mu_k=\frac{\sum_{i=k+1}^{Q}\ell_i}{Q-k}, \quad \alpha=\frac{1}{\max_k \frac{\ell_k-\mu_k}{\mu_k}}.$$

The $g_1(k)$ is simply the eigenvalues normalized to sum to 1. It can be interpreted as the probability mass function defined on the last $(Q-1)$ eigenvalues. The ξ_k in $g_2(k)$ measures the relative variation between the k-th eigenvalue and its expected

ISBN 978-0-9928626-7-1 © EURASIP 2017

$$f(p,m,B) = \max\left\{\min\left\{p, \frac{MT}{p}\right\}, \min\left\{\frac{m}{p}, pBT\right\}, \min\{B, mT\}, \min\{T, M\}, \min\{m, BT\}, \min\left\{pB, \frac{mT}{p}\right\}, \min\left\{pT, \frac{M}{p}\right\}\right\}. \quad (6)$$

value under the null hypothesis (noise only). The smaller the variation is, the larger ξ_k is. The $g_2(k)$ is the scaled version of ξ_k and sums to 1 to imitate the probability mass function defined on the first $(Q-1)$ eigenvalues.

The solution $\hat{K}$ is obtained as k which yields the last positive value of (9).

IV. Simulation Results

We evaluate the performance of the following approaches for sparsity order estimation: AIC, MDL [16], efficient detection criterion (EDC) [21], RMT [8], empirical eigenvalue-threshold test (EET) [11] and DFBM [7].

The signals are generated according to model (4). The entries of a non-Vandermonde factor matrix are i.i.d. drawn from zero-mean circularly symmetric complex Gaussian (ZMCSCG) distribution with variance 1. The Vandermonde factor matrices are extracted from an $N \times N$ discrete Fourier transform matrix. The i.i.d. ZMCSCG noise with variance σ^2 is added to (1) and is rearranged in matrix form as described in Section II. The signal-to-noise ratio (SNR) is defined as $\text{SNR} = \frac{\sum_{t=1}^{T}\|\boldsymbol{y}(t)\|^2}{MT\sigma^2}$. The noise variance σ^2 is scaled to obtain different SNRs. For each SNR, 100 independent Monte Carlo runs have been conducted. The performance measure is the probability of correct detection (PoD), i.e. $\Pr\left(\hat{K} = K\right)$, averaged over signal and noise realizations.

The dimension of the original and observed data are respectively $N = 10^5$ and $M = 1024$. A small snapshot-number scenario $T = 2$ and a large snapshot-number scenario $T = 200$ are considered. The block length m, block advance p and number of blocks B are set according to (7) and (8).

A. Small Snapshot-Number Scenario

Fig. 1 presents the PoD versus SNR at $T = 2$, where the size parameters are described in the figure caption. The AIC, MDL and EDC totally fail due to severe overestimation of the sparsity order. The RMT also fails since it overestimates the sparsity order for a false alarm rate of 0.01. The ETT performs the best when SNR< 30 dB. However, it can only achieve a PoD around 75% in high SNR regimes due to overestimation of the sparsity order. By contrast, the DFBM method consistently improves in PoD with increasing SNR, and achieves a PoD of 100% in high SNR regimes.

From Fig. 2, we see at an SNR of 45 dB the DFBM is able to accurately detect the sparsity order up to 100 (although it fails for $K = 1$ due to overestimation of the sparsity order) while the ETT can only detect up to 10 signals. Note that for a higher sparsity order, although the DFBM and ETT fail to detect the exact sparsity order, their estimated sparsity order are relatively very close to the true one (not shown here due to space limit).

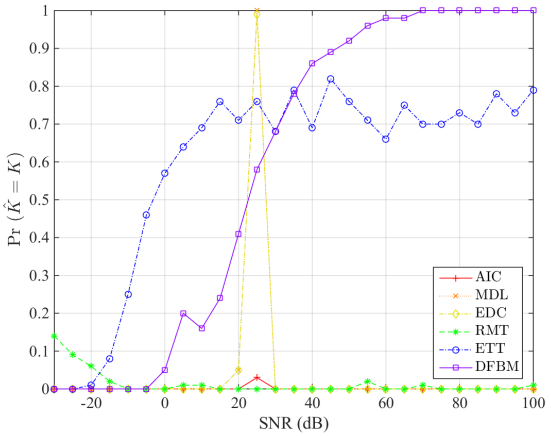

Fig. 1. PoD versus SNR, $N = 10^5$, $M = 1024$, $T = 2$. Block length: $m = 683$, block advance: $p = 1$, number of blocks: $B = 342$. True sparsity order: $K = 5$

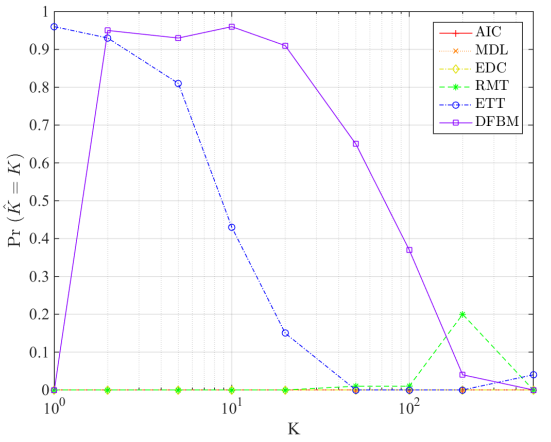

Fig. 2. PoD versus sparsity order K at SNR=45 dB. The size parameters are the same as in Fig. 1

B. Large Snapshot-Number Scenario

In Fig. 3, we increase the number of snapshots to $T = 200$. The AIC, MDL and EDC again totally fail due to severe overestimation of the sparsity order. The RMT shows the best performance, and is able to detect up to 500 signals, as shown in Fig. 4. Moreover, for a higher sparsity order, although the RMT fails to detect the exact sparsity order, its estimated sparsity order is centered around the true one (not shown here). Similar observations are obtained for the DFBM and ETT algorithms.

ISBN 978-0-9928626-7-1 © EURASIP 2017

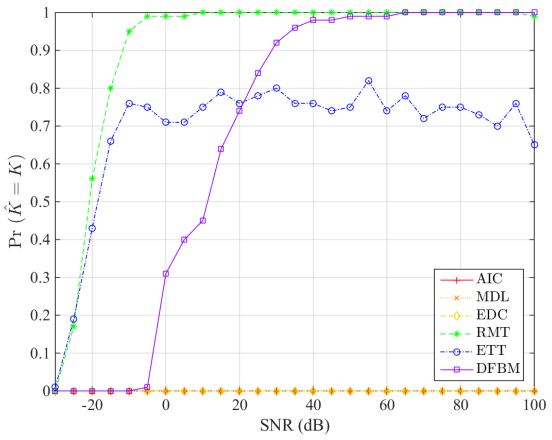

Fig. 3. PoD versus SNR, $N = 10^5$, $M = 1024$, $T = 200$. Block length: $m = 1019$, block advance: $p = 1$, number of blocks: $B = 6$. True sparsity order: $K = 5$

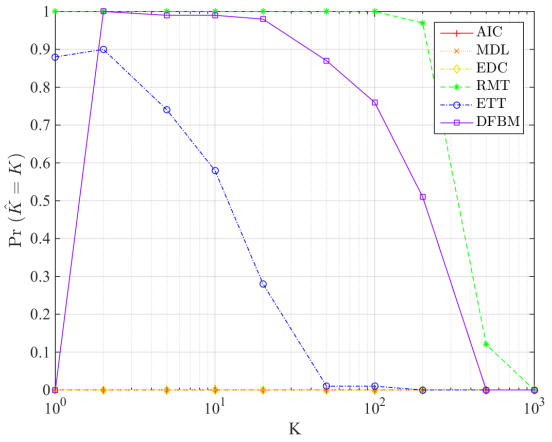

Fig. 4. PoD versus sparsity order K at SNR=45 dB. The size parameters are the same as in Fig. 3

V. Conclusion

The sparsity structure in data is a crucial information to allow a drastic reduction on data storage and communication in big data applications. The authors in [6] have recently shown that if the data possesses a sparse representation in a suitable domain, one can estimate the sparsity order directly from the compressed data obtained by using a Khatri-Rao structured measurement matrix for compressed sensing. The structure of the measurement matrix allows to rearrange the multiple-snapshot observation vectors into a fourth-order tensor. In this paper, we show how to employ the rich multilinear structure in the observed data to increase the identifiability (i.e, the largest sparsity order that can be inferred from a limited number of observations) and to improve the performance of sparsity order estimation.

Acknowledgment

This work was partially supported by the Carl-Zeiss Foundation under the postdoctoral scholarship project "EMBiCoS". The authors would also like to thank the Fundação de Apoio à Pesquisa do Distrito Federal (FAPDF) (call no. 01/2017), the DPI/DPG/UnB and Fundação de Empreendimentos Científicos e Tecnológicos (Finatec) (call no. 01/2017), and the CNPq (Grant no. 303905/ 2014-0) for their financial supports.

References

[1] K. Kambatla, G. Kollias, V. Kumar, and A. Grama, "Trends in big data analytics," *J. Parallel Distrib. Comput.*, vol. 74, no. 7, pp. 2561–2573, 2014.

[2] K. Slavakis, G. Giannakis, and G. Mateos, "Modeling and optimization for big data analytics : (statistical) learning tools for our era of data deluge," *IEEE Signal. Proc. Mag.*, vol. 31, no. 5, pp. 18–31, 2014.

[3] G. B. Giannakis, F. Bach, R. Cendrillon, M. Mahoney, and J. Neville, "Signal processing for big data," *IEEE Signal. Proc. Mag.*, vol. 31, no. 5, pp. 15–16, 2014.

[4] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[5] M. F. Duarte and R. G. Baraniuk, "Kronecker compressive sensing," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 494–504, 2012.

[6] F. Roemer, A. Lavrenko, G. D. Galdo, T. Hotz, O. Arikan, and R. S. Thomä, "Sparsity order estimation for single snapshot compressed sensing," in *Proc. 48th Asilomar Conference on Signals, Systems and Computers*, Nov. 2014, pp. 1220–1224.

[7] E. Radoi and A. Quinquis, "A new method for estimating the number of harmonic components in noise with application in high resolution radar," *EURASIP Journal on Applied Signal Processing*, vol. 8, pp. 1177–1188,

[8] S. Kritchman and B. Nadler, "Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3930–3941, 2009.

[9] L. de Lathauwer, B. de Moor, and J. Vanderwalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, 2000.

[10] P. Comon, J. M. F. ten Berge, L. De Lathauwer, and J. Castaing, "Generic and typical ranks of multi-way arrays," *Linear Algebra Appl.*, vol. 430, no. 11–12, pp. 2997–3007, Jun. 2009.

[11] A. Lavrenko, F. Roemer, G. D. Galdo, R. Thomä, and O. Arikan, "An empirical eigenvalue-threshold test for sparsity level estimation from compressed measurements," in *Proc. 22nd European Signal Processing Conf. (EUSIPCO)*, Sept. 2014, pp. 1761–1765.

[12] R. Bro and H. A. L. Kiers, "A new efficient method for determining the number of components in PARAFAC models," *J. Chemom.*, vol. 17, pp. 274–286, 2003.

[13] M. E. Timmerman and H. A. L. Kiers, "Three-mode principal components analysis: Choosing the numbers of components and sensitivity to local optima," *Br. J. Math. Stat. Psychol.*, vol. 53, no. 1, pp. 1–16, 2000.

[14] E. Ceulemans and H. A. L. Kiers, "Selecting among three-mode principal component models of different types and complexities: A numerical convex hull based method," *Br. J. Math. Stat. Psychol.*, vol. 59, no. 1, pp. 133–150, 2006.

[15] G. Tomasi and R. Bro, "A comparison of algorithms for fitting the PARAFAC model," *Computational Statistics & Data Analysis*, vol. 50, no. 7, pp. 1700–1734, 2006.

[16] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 33, no. 2, pp. 387–392, 1985.

[17] A. Quinlan, J.-P. Barbot, P. Larzabal, and M. Haardt, "Model order selection for short data: An exponential fitting test (EFT)," *EURASIP Journal on Applied Signal Processing*, vol. 2007, no. 1, pp. 1–11, 2007.

[18] K. Liu, J. P. C. da Costa, H. C. So, L. Huang, and J. Ye, "Detection of number of components in CANDECOMP/PARAFAC models via minimum description length," *Digital Signal Process.*, vol. 51, pp. 110–123, 2016.

[19] T. G. Kolda, *Multilinear operators for higher-order decompositions*, United States. Department of Energy, 2006.

[20] S. Ragnarsson and C. F. Van Loan, "Block tensor unfoldings," *SIAM J. Matrix Anal. Appl.*, vol. 33, no. 1, pp. 149–169, 2012.

[21] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "On detection of number of signals in presence of white noise," *Journal of Multivariate Analysis*, vol. 20, pp. 1–25, 1986.