

SPARSITY ORDER ESTIMATION FOR SINGLE SNAPSHOT COMPRESSED SENSING

F. Römer¹, A. Lavrenko¹, G. Del Galdo¹, T. Hotz², O. Arikan³, and R. S. Thomä¹

Technische Universität Ilmenau,
¹ Institute for Information Technology
² Institute for Mathematics
P.O. Box 10 05 65, 98684 Ilmenau, Germany

³ Bilkent University
Electrical and Electronics Eng. Dep.
TR-06800 Bilkent, Ankara, Turkey

Abstract — In this paper we discuss the estimation of the sparsity order for a Compressed Sensing scenario where only a single snapshot is available. We demonstrate that a specific design of the sensing matrix based on Khatri-Rao products enables us to transform this problem into the estimation of a matrix rank in the presence of additive noise. Thereby, we can apply existing model order selection algorithms to determine the sparsity order. The matrix is a rearranged version of the observation vector which can be constructed by concatenating a series of non-overlapping or overlapping blocks of the original observation vector. In both cases, a Khatri-Rao structured measurement matrix is required with the main difference that in the latter case, one of the factors must be a Vandermonde matrix. We discuss the choice of the parameters and show that an increasing amount of block overlap improves the sparsity order estimation but it increases the coherence of the sensing matrix. We also explain briefly that the proposed measurement matrix design introduces certain multilinear structures into the observations which enables us to apply tensor-based signal processing, e.g., for enhanced denoising or improved sparsity order estimation.

1. INTRODUCTION

Compressed Sensing (CS) is a novel paradigm in sampling theory that allows to acquire signals at sampling rates significantly below the Nyquist rate without any loss of information, provided that the signals possess a sparse representation in some basis. A vast amount of theoretical results is available showing under which conditions the recovery of the signal can be achieved efficiently, i.e., by solving convex optimization problems [1]. Obviously, the sparsity order, i.e., the number of non-zero coefficients in the sparsity-providing basis, has a tremendous impact on the recovery stage, in particular it determines how many measurements are required for successful recovery [2].

However, there exist a large number of applications where the sparsity order is not known beforehand and may even vary with time. In such cases, it would be desirable if we could estimate the sparsity order before we run the recovery algorithm using an estimator that is significantly less complex than the reconstruction itself. Moreover, this would even allow us to adapt our reconstruction strategy, i.e., to choose a recovery algorithm whose performance and complexity is best suited to the current sparsity order. Also, if we find the sparsity order too large to expect successful CS recovery, we can provide a feedback to the measurement stage to perform more measurements provided that the application allows for it. Thereby, the measurement effort can be adapted to the complexity of the current signal/scene.

In [3] we have discussed sparsity order estimation in the spe-

cial case where the scene (a) can be measured multiple times, (b) is “stationary” such that the sparsity pattern does not change during these measurements, and (c) provides linearly independent measurements, e.g., by observing modulated signals. In this case, the compressed observations represent a linear mixture superimposed by additive noise so that the sparsity order is equal to the effective rank of the observation matrix and model order selection techniques can be applied for its estimation as shown in [3].

In this paper we extend this work to the more challenging case where either only a single snapshot is available or the scenario is completely static so that observing it multiple times with the same measurement matrix does not provide linearly independent observations. We develop a measurement matrix design that recovers linear independence and thus allows to estimate the sparsity order from the effective rank of a matrix constructed by concatenating blocks of the observed vector along its columns. We discuss the choice of the parameters for both the case of non-overlapping blocks (in which case the measurement matrix needs to be Khatri-Rao structured) as well as overlapping blocks (in which case one of the Khatri-Rao factors needs to be a Vandermonde matrix). This paper is structured as follows: in Section 2 we analytically derive the required structure of the measurement matrix for rank recovery both for the non-overlapping as well as the overlapping case. In Section 3 we analyze the coherence of the Khatri-Rao structured measurement matrix in order to show the effect of the required structure on the recovery performance. Section 4 contains a discussion on the choice of the parameters as well as some notes on the links between the proposed design and concepts from multilinear (tensor) algebra. Numerical results are presented in Section 5 before concluding in Section 6.

2. PROPOSED DESIGN

Consider CS scenario of the following form

$$\mathbf{y} = \Phi \cdot \mathbf{A} \cdot \mathbf{s} + \mathbf{n}, \quad (1)$$

where $\mathbf{y} \in \mathbb{C}^{M \times 1}$ represents a vector of compressed observations, $\mathbf{s} \in \mathbb{C}^{N \times 1}$ is the K -sparse coefficient vector (i.e., it contains exactly K non-zero elements), $\mathbf{A} \in \mathbb{C}^{N \times N}$ is the sparsity-providing basis, \mathbf{n} contains the additive measurement noise, and $\Phi \in \mathbb{C}^{M \times N}$ is the measurement matrix. Moreover, we require the following assumptions:

- (A1) The measurement matrix Φ can be designed freely.
- (A2) The basis \mathbf{A} is an $N \times N$ identity matrix.
- (A3) The sparsity order K satisfies $K \leq K_{\max}$.

Note that as long as the basis \mathbf{A} is invertible and known when designing the measurement kernel Φ , (A2) holds without loss of generality since for $\mathbf{A} \neq \mathbf{I}_N$ we can replace Φ by $\tilde{\Phi} = \Phi \cdot \mathbf{A}^{-1}$ and achieve the same result. Regarding (A3), the maximum allowable sparsity order K_{\max} depends on the amount of block overlap. As we show in Section 4, it is given by $K_{\max} = \sqrt{M} - 1$ for non-overlapping blocks, it grows with larger overlap, and it eventually reaches $K_{\max} = \frac{M}{2} - 1$ for maximum overlap.

For simplicity, let us consider the noise-free case $\mathbf{n} = \mathbf{0}$. In order to recover the sparsity order K from \mathbf{y} we would like to break \mathbf{y} into smaller blocks $\mathbf{y}_b \in \mathbb{C}^{m \times 1}$, $b = 1, 2, \dots, B$ and define a matrix $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_B] \in \mathbb{C}^{m \times B}$ such that $\text{rank}\{\mathbf{Y}\} = K$. From (1) it is clear that $\mathbf{y}_b = \Phi_b \cdot \mathbf{s}$ where $\Phi_b \in \mathbb{C}^{m \times N}$ contains the m rows that correspond to the b -th block of \mathbf{Y} . This sparks the question which condition Φ must fulfill such that for any K -sparse vector \mathbf{s} , we have $\text{rank}\{\mathbf{Y}\} = K$. For the case of non-overlapping blocks, this question is answered by the following Theorem.

Theorem 1. For B non-overlapping blocks of size $m = \frac{M}{B}$, any $K \leq \min(B, m)$, and any K -sparse \mathbf{s} , we have $\text{rank}\{\mathbf{Y}\} = K$ if and only if $\Phi = \mathbf{C} \diamond \Phi_0$, where $\mathbf{C} \in \mathbb{C}^{B \times N}$ and $\Phi_0 \in \mathbb{C}^{m \times N}$, the Kruskal-rank of \mathbf{C} and Φ_0 is $\geq K$, and \diamond denotes the column-wise Kronecker (Khatri-Rao) product.

Proof: The ‘‘only-if’’ part becomes evident by considering $K = 1$. In this case, we have $\mathbf{y}_b = \Phi_b \cdot \mathbf{s} = \varphi_{b,n} \cdot s_n$, where s_n is the value of the single non-zero in \mathbf{s} at position n and $\varphi_{b,n}$ is the n -th column of Φ_b . For \mathbf{Y} we then have $\mathbf{Y} = [\varphi_{1,n}, \dots, \varphi_{B,n}] \cdot s_n$ which is rank-one only if all columns $\varphi_{1,n}$ are scaled version of one common non-zero vector $\varphi_{0,n}$, i.e., $\varphi_{b,n} = c_{b,n} \cdot \varphi_{0,n}$. Stacking $\varphi_{b,n}$ back into Φ we obtain $\varphi_n = c_n \otimes \varphi_{0,n}$ and therefore $\Phi = \mathbf{C} \diamond \Phi_0$.

For the ‘‘if’’ part, consider $\mathbf{y} = \Phi \cdot \mathbf{s} = (\mathbf{C} \diamond \Phi_0) \cdot \mathbf{s}$. Its reshaped version \mathbf{Y} can be expressed as $\mathbf{Y} = \Phi_0 \cdot \text{diag}\{\mathbf{s}\} \cdot \mathbf{C}^T = \Phi_{0,K} \cdot \text{diag}\{\mathbf{s}_K\} \cdot \mathbf{C}_K^T$, where $\Phi_{0,K} \in \mathbb{C}^{m \times K}$, $\mathbf{C}_K \in \mathbb{C}^{B \times K}$, and \mathbf{s}_K contain only the K columns/values corresponding to the non-zero entries in \mathbf{s} . Since $\Phi_{0,K}$ and \mathbf{C}_K have full column rank K and provide a rank factorization of \mathbf{Y} , we have $\text{rank}\{\mathbf{Y}\} = K$. Obviously, this can only be fulfilled if $K \leq B$ and $K \leq m$. \square

Note that the theorem requires the sparsity order to satisfy $K \leq \min(m, B)$. Since $m = \frac{M}{B}$, to maximize this upper bound it is best to choose $m = B = \sqrt{M}$ if M is a square number.

We now move to the case where we let blocks overlap. To this end, we divide $\mathbf{y} \in \mathbb{C}^{M \times 1}$ into B blocks of m samples with an offset of p samples from block to block. In other words, the b -th block $\mathbf{y}_b \in \mathbb{C}^{m \times 1}$ contains samples $(b-1) \cdot p + 1$ up to $(b-1) \cdot p + m$. To cover all M samples with B blocks we therefore obtain the condition $(B-1) \cdot p + m = M$ which implies that the number of blocks is given by $B = \frac{M-m}{p} + 1$ and that $M-m$ must be divisible by p . The case $p = m$ is the one where the blocks do not overlap. For the overlapping case $1 \leq p < m$ the following theorem provides the required structure for Φ to allow to obtain K from the rank of \mathbf{Y} :

Theorem 2. For B overlapping blocks of size m (with an offset of p samples between blocks), any $K \leq \min(B, m)$, and any K -sparse \mathbf{s} , we have $\text{rank}\{\mathbf{Y}\} = K$ if and only if the matrix Φ can be constructed by taking the first M rows of $\mathbf{C} \diamond \Phi_0$, where $\Phi_0 \in \mathbb{C}^{p \times N}$, $\mathbf{C} \in \mathbb{C}^{\lceil \frac{M}{p} \rceil \times N}$, \mathbf{C} is a Vandermonde matrix, Φ and \mathbf{C} have a Kruskal-rank $\geq K$, and $\lceil \cdot \rceil$ denotes the operation of rounding to the next larger integer number.

Proof: The proof proceeds in a similar fashion to the proof of Theorem 1. We begin by considering $K = 1$ to prove the ‘‘only-if’’ part. As before we obtain $\mathbf{Y} = [\varphi_{1,n}, \dots, \varphi_{B,n}] \cdot s_n$, where

$\varphi_{b,n}$ is the b -th block of the n -th column of Φ . To obtain the desired property that $\text{rank}\{\mathbf{Y}\} = K = 1$, we therefore require that all $\varphi_{b,n}$ are linearly dependent, i.e., $\varphi_{b,n} = c_{b,n} \cdot \varphi_{b-1,n}$. For non-overlapping blocks, all the $c_{b,n}$ are different, which immediately yields the Khatri-Rao structure with an arbitrary matrix \mathbf{C} . We now show that for overlapping blocks, there must be much more structure in the coefficients \mathbf{C} .

To this end, let us consider three consecutive columns $b-1$, b , and $b+1$. We have the two scaling conditions

$$\varphi_{b,n} = c_{b,n} \cdot \varphi_{b-1,n} \quad (2)$$

$$\varphi_{b+1,n} = c_{b+1,n} \cdot \varphi_{b,n} \quad (3)$$

which must be valid for all elements in the vectors. However, since the blocks overlap by $m-p$ samples, they have common elements. In particular

$$[\varphi_{b-1,n}]_{(k)} = [\varphi_{b,n}]_{(k-p)} \quad (4)$$

$$[\varphi_{b,n}]_{(k)} = [\varphi_{b+1,n}]_{(k-p)} \quad (5)$$

for all $k = p+1, p+2, \dots, m$. Inserting (2) and (3) into (5) we obtain

$$c_{b,n} \cdot [\varphi_{b-1,n}]_{(k)} = c_{b+1,n} \cdot [\varphi_{b,n}]_{(k-p)}. \quad (6)$$

However, since (4) must be true for all $k = p+1, p+2, \dots, m$ this implies that $c_{b+1,n} = c_{b,n}$. Since b is arbitrary, the same argument can be applied to show that $c_{b,n} = c_n \forall b = 1, 2, \dots, B$. Therefore, we have

$$\varphi_{b,n} = c_n \cdot \varphi_{b-1,n} = c_n^2 \cdot \varphi_{b-2,n} = \dots = c_n^{b-1} \cdot \varphi_{1,n}. \quad (7)$$

Due to the overlap, to satisfy (7) we can choose only the first p elements of $\varphi_{1,n}$ as well as the constant $c_n \in \mathbb{C}$ freely. The remaining elements of φ_n are filled by replicating these p elements, scaled by c_n, c_n^2 , and so on. This leads to the Vandermonde structure of \mathbf{C} .

For the ‘‘if’’ part, it is easy to see that for every K we can write \mathbf{Y} as $\tilde{\Phi}_0 \cdot \text{diag}\{\mathbf{s}\} \cdot \mathbf{C}_B^T$, where $\tilde{\Phi}_0$ contains the first m rows of $\mathbf{C} \lceil \frac{m}{p} \rceil \diamond \Phi_0$, and $\mathbf{C} \lceil \frac{m}{p} \rceil$ and \mathbf{C}_B represent the first $\lceil \frac{m}{p} \rceil$ and the first B rows of \mathbf{C} , respectively. Therefore, we can write \mathbf{Y} as $\tilde{\Phi}_{0,K} \cdot \text{diag}\{\mathbf{s}_K\} \cdot \mathbf{C}_{B,K}^T$, where $\tilde{\Phi}_{0,K} \in \mathbb{C}^{m \times K}$, $\mathbf{C}_{B,K} \in \mathbb{C}^{B \times K}$, and \mathbf{s}_K contain the K columns/values corresponding to the nonzero entries in \mathbf{s} . Since \mathbf{C} and Φ_0 have full Kruskal-rank K , the same holds true for $\tilde{\Phi}_0$ and \mathbf{C}_B [4] and thus we have $\text{rank}\{\mathbf{Y}\} = K$. \square

Note that Theorem 2 implies that in the special case $p = 1$ (maximal overlap), the entire sensing matrix Φ must be a Vandermonde matrix. Vandermonde structured measurement matrices have been proposed in the CS context before [5, 6] and rank recovery for Vandermonde mixtures has been studied in the context of harmonic retrieval [7]. In fact, for the case $p = 1$ the mapping from \mathbf{y} to \mathbf{Y} is known as spatial smoothing [8] in the harmonic retrieval context where it is applied as a preprocessing step for subspace-based estimators in order to decorrelate coherent signals.

It is also important to note that the proposed design can also be applied to the case where \mathbf{y}_b corresponds to subsequent CS measurements in a setting where we can change the measurement matrix from one measurement to the next. In this case it recovers the required rank for scenarios where the scene is static and would hence not provide linearly independent observations by itself. In such a setting we could adapt the number of measurements to the complexity of the scene, e.g., by recovering the scene from an initial set of B_1 observations if the sparsity test suggests that K is small enough for successful recovery, otherwise continuing to observe until a sufficient number of observations have been collected.

3. COHERENCE ANALYSIS

In this section we analyze the proposed sensing matrix design in terms of the coherence of the measurement matrices Φ that are obtainable with the structure derived in Theorem 1 and 2.

For simplicity, let us assume that M is divisible by p which allows us to write $\Phi = C \diamond \Phi_0$ for both cases, the only difference being that in the case of overlap C must be Vandermonde whereas without overlap, C can be arbitrary. The mutual coherence of Φ is defined as

$$\mu(\Phi) = \max_{n_1 \neq n_2 \in [1, 2, \dots, N]} \left| \frac{\varphi_{n_1}^H \cdot \varphi_{n_2}}{\|\varphi_{n_1}\|_2 \cdot \|\varphi_{n_2}\|_2} \right|. \quad (8)$$

Using the column-wise Kronecker structure and the fact that $(\mathbf{a} \otimes \mathbf{b})^T (\mathbf{c} \otimes \mathbf{d}) = \mathbf{a}^T \mathbf{c} \cdot \mathbf{b}^T \mathbf{d}$ as well as $\|\mathbf{a} \otimes \mathbf{b}\|_2 = \|\mathbf{a}\|_2 \cdot \|\mathbf{b}\|_2$ we obtain

$$\begin{aligned} \mu(\Phi) &= \max_{n_1 \neq n_2 \in [1, 2, \dots, N]} \left| \frac{\mathbf{c}_{n_1}^H \cdot \mathbf{c}_{n_2}}{\|\mathbf{c}_{n_1}\|_2 \cdot \|\mathbf{c}_{n_2}\|_2} \right| \cdot \left| \frac{\varphi_{0, n_1}^H \cdot \varphi_{0, n_2}}{\|\varphi_{0, n_1}\|_2 \cdot \|\varphi_{0, n_1}\|_2} \right| \\ &= \max_{n_1 \neq n_2 \in [1, 2, \dots, N]} \underbrace{\mu_{n_1, n_2}(C)}_{\mu_{n_1, n_2}(C)} \cdot \underbrace{\mu_{n_1, n_2}(\Phi_0)}_{\mu_{n_1, n_2}(\Phi_0)} \\ &\leq \mu(C) \cdot \mu(\Phi_0) \end{aligned} \quad (9)$$

$$\leq \mu(C) \cdot \mu(\Phi_0) \quad (10)$$

Note that (9) is reminiscent of an argument made in [9] (Lemma 1). There, the authors had considered a Kronecker product structure for the sensing matrix which leads to the product bound (10) being tight. However, since we consider a Khatri-Rao product, this is not the case here (by permuting the columns of the factors appropriately, a lower value can be achieved). To further analyze (9) we note that for an arbitrary $P \times Q$ matrix, its coherence is bounded by $\mu \geq \sqrt{\frac{Q-P}{P(Q-1)}}$, a bound known as the Welch bound [10]. Moreover, this bound is achieved only if all pairs of columns have the same magnitude inner product, which is then equal to this lower bound (such matrices are called equiangular tight frames (ETF) [11]). In the case of no overlap ($p = m = \sqrt{M}$), C and Φ_0 can be designed freely. Therefore, if C and Φ_0 achieved the Welch bound, we would have

$$\mu(\Phi) = \sqrt{\frac{N-m}{m(N-1)}} \cdot \sqrt{\frac{N-\frac{M}{m}}{\frac{M}{m}(N-1)}} = \frac{N-m}{m(N-1)}. \quad (11)$$

However, the Welch bound is not achievable for all matrix dimensions P and Q . In particular, it is known that ETFs do not exist for $P < Q^2$, a condition which is satisfied for both C and Φ_0 for $M = m^2$. Empirically, this leads to a higher coherence than what (11) predicts. Note that the prediction may be improved using bounds that are tighter than the Welch bound for $P < Q^2$ (e.g., [12]).

In the overlapping case, the matrix C must be Vandermonde. Among the class of $P \times Q$ Vandermonde matrices V , the frame obtained by considering the top P rows of a $Q \times Q$ DFT matrix achieves the lowest coherence (cf. [13], which also shows that for prime Q these frames have maximal spark). Its coherence is given by considering the inner product between two adjacent columns which can easily be shown to be

$$\mu(V) = \left| \frac{\mathbf{v}_n^H \cdot \mathbf{v}_{n+1}}{\|\mathbf{v}_n\|_2 \cdot \|\mathbf{v}_{n+1}\|_2} \right| = \frac{\sin\left(\pi \cdot \frac{P}{Q}\right)}{P \cdot \sin\left(\frac{\pi}{Q}\right)}. \quad (12)$$

Using a similar reasoning as in the non-overlapping case, an optimally designed matrix $\Phi_0 \in \mathbb{C}^{p \times N}$ together with the Vandermonde

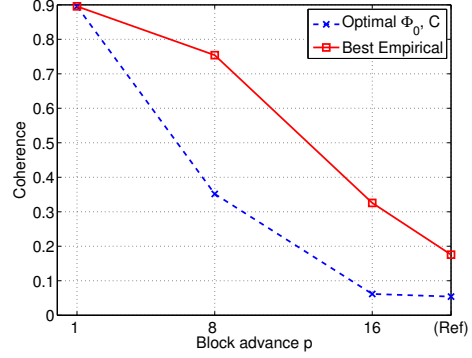


Fig. 1. Coherence of the matrix Φ for different amount of block advance p : $p = 1$ corresponds a the Vandermonde matrix, $p = 8$ to overlapping blocks, $p = 16$ to non-overlapping blocks, and “(Ref)” to an unstructured matrix Φ .

matrix $C \in \mathbb{C}^{\frac{M}{p} \times N}$ achieves a coherence given by

$$\mu(\Phi) = \frac{\sin\left(\pi \cdot \frac{M}{p \cdot N}\right)}{\frac{M}{p} \cdot \sin\left(\frac{\pi}{N}\right)} \cdot \sqrt{\frac{N-p}{p(N-1)}}. \quad (13)$$

Figure 1 demonstrates the dependence of the achievable coherence on the block overlap. We consider a scenario where $M = 256$ and $N = 1000$. We compare the case of maximal block overlap ($p = 1$), no overlap ($p = \sqrt{M} = 16$), the intermediate case $p = 8$ and as a reference the case where Φ is unstructured matrix (which does not allow any sparsity order estimation). For each case, we display the value of the coherence that is predicted by the Welch bound (according to (11) and (13)) as well as the coherence achieved by drawing the matrices randomly (from a complex Gaussian distribution), choosing the best among 200000 trials. As expected, the coherence we achieve practically is higher than the Welch bound predicts. However, both the theoretical and the empirical results show the same trend, namely, the larger the block overlap (i.e., the smaller p), the higher the coherence.

4. DISCUSSION

4.1. Choice of the parameters

The structure of Φ that is provided by Theorem 1 and Theorem 2 provides an exhaustive answer to the question which sensing strategies allow to estimate the sparsity order from the matrix rank of a rearranged version of the single observation vector \mathbf{y} . As we have shown there are essentially two parameters we can choose: the block length m and the block advance p . In this section, we discuss the implications of these parameters on the design of Φ as well as the sparsity order estimation step.

Let us begin with the block length m . Note that the design of the sensing matrix Φ does not depend on m . Therefore, once a suitable value of p has been selected, the block length m can be chosen without affecting the sensing or the sparse recovery stage. It determines the dimensions of the matrix \mathbf{Y} which is important for the rank estimation step that is used to find the sparsity order. More specifically, \mathbf{Y} is of size $m \times B$, where $B = \frac{M-m}{p} + 1$. To maximize the size of \mathbf{Y} ,

we can choose p such that $m \approx B$, which leads to $m \approx \frac{M+p}{p+1}$. Therefore, we propose to select m as the integer value closest to $\frac{M+p}{p+1}$ such that $M - m$ divides p .

The second parameter we can adjust is the block advance p , which controls the amount of overlap between adjacent blocks (equal to $m - p$ samples). The smaller p is chosen, the more we reuse elements of \mathbf{y} which results in a larger overall matrix \mathbf{Y} . This has a positive effect on the sparsity order estimation step and it allows to estimate larger values of K since $K \leq K_{\max} = \min(m, B) - 1$ where $B = \frac{M-m}{p} + 1$. For maximum overlap ($p = 1$), this bound is maximized and becomes $K_{\max} = \frac{M}{2} - 1$ while in the case of no overlap ($p = m = \sqrt{M}$) we have $K_{\max} = \sqrt{M} - 1$. On the other hand, a larger overlap leads to a reduced flexibility in the sensing matrix design since a growing part of Φ has to obey the Vandermonde scaling law shown in Theorem 2. As shown in Section 3, this has a negative impact on the coherence of Φ . Therefore, there is a fundamental trade-off between the performance of the sparsity order estimation and the performance of the sparse recovery step. However, note that we consider a system where the measurement matrix Φ can be adapted at will. This allows to switch between measurement matrices designed for the two different purposes: a ‘‘probing’’ matrix which is optimized for the sparsity order estimation step (using a small value of p , e.g., $p = 1$), and a measurement matrix, which is optimized to the recovery stage (using a larger value of p , e.g., $p = m$).

4.2. A link to multilinear algebra

Beyond enabling the sparsity order estimation for a single snapshot via rank estimation, the proposed Khatri-Rao design for the measurement matrix has a strong link to multilinear algebra. In particular, using a Khatri-Rao structured measurement matrix allows to rearrange the observed data in form of a tensor that has (in the noise-free case) a rank- K Canonical Polyadic Decomposition (CPD) [14]. In fact, there are multiple special cases of the proposed designs where such tensors occur, which we would like to list here. For simplicity, let us assume that the parameters p and m are chosen such that p divides M and m .

Firstly, in the case of overlapping blocks we have shown in Theorem 2 that $\mathbf{Y} = (\mathbf{C}_{\frac{m}{p}} \diamond \Phi_0) \cdot \text{diag}\{\mathbf{s}\} \cdot \mathbf{C}_B^T \in \mathbb{C}^{m \times B}$. This matrix can be reshaped into an $m \times p \times B$ tensor \mathcal{Y} which obeys

$$\begin{aligned} \mathcal{Y} &= \mathcal{I}_{3,N} \times_1 \mathbf{C}_{\frac{m}{p}} \times_2 \Phi_0 \times_3 (\mathbf{C}_B \cdot \text{diag}\{\mathbf{s}\}) \\ &= \mathcal{I}_{3,K} \times_1 \mathbf{C}_{\frac{m}{p},K} \times_2 \Phi_{0,K} \times_3 (\mathbf{C}_{B,K} \cdot \text{diag}\{\mathbf{s}_K\}). \end{aligned} \quad (14)$$

where $\mathcal{I}_{3,p}$ is the $p \times p \times p$ identity tensor. Moreover, $\mathbf{C}_{\frac{m}{p},K}$, $\Phi_{0,K}$, and $\mathbf{C}_{B,K}$ contain only the K columns corresponding to the support (i.e., the non-zero elements of \mathbf{s}). Obviously, (14) is a rank- K CPD which shows that \mathcal{Y} is rank K .

Secondly, in a case where we do have $T > 1$ linearly independent snapshots according to $\mathbf{Y} = \Phi \cdot \mathbf{X}$, where $\mathbf{Y} \in \mathbb{C}^{M \times T}$, $\mathbf{X} \in \mathbb{C}^{N \times T}$ is row-sparse and $\Phi = \mathbf{C} \diamond \Phi_0$, applying the proposed measurement matrix design allows to reshape the given $M \times T$ observation matrix \mathbf{Y} into a $B \times p \times T$ tensor \mathcal{Y} where $B = \frac{M}{p}$. Note that this corresponds to the operation we perform to \mathbf{y} for non-overlapping blocks, i.e., $m = p$. The resulting tensor can be expressed as

$$\mathcal{Y} = \mathcal{I}_{3,N} \times_1 \mathbf{C} \times_2 \Phi_0 \times_3 \mathbf{S}^T = \mathcal{I}_{3,K} \times_1 \mathbf{C}_K \times_2 \Phi_{0,K} \times_3 \mathbf{S}_K^T,$$

where, as above, \mathbf{S}_K^T contains only the columns of \mathbf{S}^T corresponding to the support (i.e., the non-zero rows of \mathbf{S}).

Thirdly, we can combine these two approaches for the case of multiple snapshots and overlapping signals. In this case we transform the $M \times T$ observations into a $m \times p \times B \times T$ tensor with loading matrices given by $\mathbf{C}_{\frac{m}{p},K}$, $\Phi_{0,K}$, $\mathbf{C}_{B,K}$, and \mathbf{S}_K^T .

Finally, we could also decompose Φ into more than two matrices, e.g., $\Phi = \mathbf{C}_1 \diamond \dots \diamond \mathbf{C}_G \diamond \Phi_0$ which can then be reshaped into a $(G + 1)$ -dimensional tensor if we have a single snapshot and into a $(G + 2)$ -dimensional tensor in the multiple snapshot case.

Exploring the potential benefit of this rich multilinear structure in our data is an aspect of future work. We see potential benefit of it in an enhanced denoising, similar to the improvement tensor-based subspace estimation schemes have brought for high-resolution parameter estimation [15]. Moreover, the tensor structure can be used to improve the sparsity order estimation step using tensor-based model order estimation (based on, e.g., [16]). Finally, in the case that the sensing matrix is (partially) unknown (e.g., in a distributed setting), a CPD of the observed tensor \mathcal{Y} could be computed to reveal it. Note that connections between CS and tensors have been discussed in slightly different contexts before, e.g., big low-rank tensors with unknown factors [17] or tensors with specially structured ‘‘block-sparsity’’ cores [18].

5. NUMERICAL RESULTS

To demonstrate the single-snapshot sparsity order estimation based on the proposed design, we perform a numerical experiment. We consider the recovery of a $K = 7$ -sparse vector \mathbf{s} of length $N = 1000$ from a single vector of observations \mathbf{y} of length $M = 256$. We compare three different strategies. Firstly, choosing $m = p = 16$ so that there is no overlap and thus Φ_0 and \mathbf{C} can be chosen freely. Secondly, setting $p = 1$ for maximum overlap, in which case Φ needs to be Vandermonde, where we set $m = 128$ to maximize the size of \mathbf{Y} . Thirdly, as an intermediate case, $p = 8$ in which case $m = 32$ leads to the largest matrix \mathbf{Y} and Φ is composed of an 8×1000 Vandermonde matrix \mathbf{C} and an arbitrary 32×1000 matrix Φ_0 . For a direct comparison, the size of the data matrix \mathbf{Y} that is used for rank estimation is 16×16 in the non-overlapping case, 128×129 in the case $p = 1$, and 32×29 for $p = 8$.

The matrices that we can choose freely (\mathbf{C} in the case of no overlap and Φ_0 in all cases) are drawn from a zero-mean circularly symmetric complex Gaussian (ZMCSCG) distribution. The K non-zeros in the vector \mathbf{s} are placed randomly and their values are given by $e^{j\varphi_k}$ where $\varphi_k \sim \mathcal{U}[0, 2\pi)$. The additive noise \mathbf{n} is also drawn from a ZMCSCG distribution with variance P_N . The SNR is defined as $\text{SNR} = 1/P_N$.

The sparsity order is estimated by applying existing model order selection criteria to the the singular values σ_i , $i = 1, 2, \dots, 16$ of the matrix $\mathbf{Y} \in \mathbb{C}^{m \times B}$. In particular, we consider Akaike’s Information Criterion (AIC) [19] as well as the Empirical Threshold Test (ETT) proposed in [3]. Figure 2 shows the estimated model order as a function of the SNR, averaged over 1500 Monte-Carlo trials. The result demonstrates that the correct sparsity order can be identified from \mathbf{y} , provided the SNR is not too low. AIC suffers from the very small sample support (which is assumed to be large in the original derivation), which is handled much better by the ETT. We also notice that a larger amount of block overlap (a smaller value p) provides a better sparsity order estimation performance. This is not surprising since the size of the matrix \mathbf{Y} used for the rank estimation step grows with increasing block overlap. We also observe that ETT tends to overestimate the model order slightly for the case of maximum overlap. However, this is not a critical issue since for the CS context, it is better to overestimate (to have some head room) than to underestimate

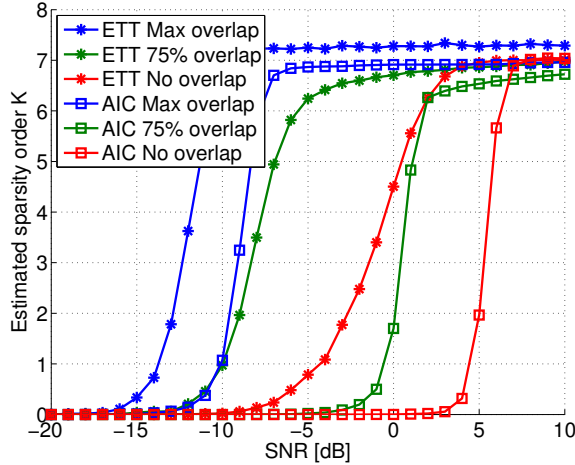


Fig. 2. Estimated sparsity order K vs. the SNR for $M = 256$ and $K = 7$. We compare three cases: $p = 1, m = 128$ (“max overlap”), $p = 8, m = 32$ (“75 % overlap”) and $p = m = 8$ (“no overlap”).

(which can lead to a fatal breakdown of the CS recovery).

Figure 3 shows the probability of correct support estimation for the same scenario, using the Orthogonal Matching Pursuit (OMP) algorithm for the sparse recovery stage. In addition to the three scenarios discussed above, the curve labeled “Reference” depicts the case where an unstructured matrix Φ is used for the measurement (which does not allow the sparsity order estimation discussed in this paper). We observe that larger overlaps lead to a degradation of the recovery performance, which is expected due to the increase in the mutual coherence (as discussed in Section 3). Interestingly, the Khatri-Rao structured matrix Φ in the non-overlapping case performs almost identically to the unstructured matrix Φ shown as a reference.

6. CONCLUSIONS

In this work we have investigated the problem of sparsity order estimation from a single data snapshot. We analytically show that a specific Khatri-Rao design of the measurement matrix is a necessary and a sufficient condition for the recovery of the linear independence in a single vector of observations. The sparsity order can then be estimated as the effective rank of a matrix constructed by concatenating blocks of this observation vector. We analyze the influence of the parameters of the proposed Khatri-Rao design on the resulting matrix coherence and numerically show the trade-off between the achievable estimation and recovery performance. Additionally, we show that the proposed design introduces certain multi-linear structures into the data. These could possibly be exploited by applying tensor-based signal processing (e.g., tensor-based denoising and model order estimation).

REFERENCES

- [1] D.L. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.
- [2] E.J. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb 2006.
- [3] A. Lavrenko, F. Roemer, G. Del Galdo, R.S. Thomä, and O. Arikan, “An empirical eigenvalue-threshold test for sparsity level estimation from compressed measurements,” in *Proc. European Signal Processing Conference (EUSIPCO 2014)*, Lisbon, Portugal, Sept. 2014, submitted.

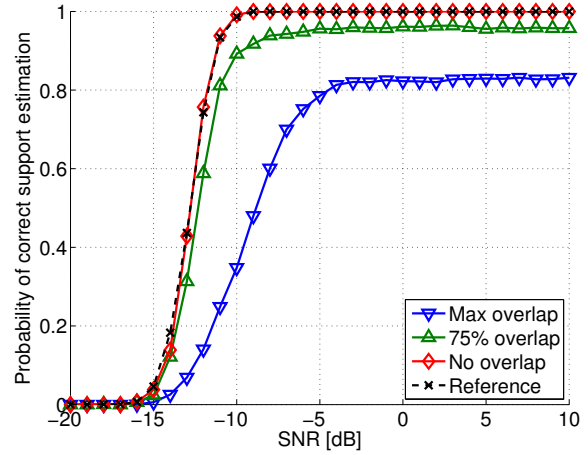


Fig. 3. Empirical probability of correct support estimation using OMP vs. the SNR for $M = 256$ and $K = 7$. We compare three cases: $p = 1, m = 128$ (“max overlap”), $p = 8, m = 32$ (“75 % overlap”) and $p = m = 8$ (“no overlap”). The curve “reference” corresponds to using an unstructured measurement matrix.

- [4] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, “Parallel factor analysis in sensor array processing,” *IEEE Transactions on Signal Processing*, vol. 48, no. 8, pp. 2377–2388, Aug. 2000.
- [5] J.-J. Fuchs, “Sparsity and uniqueness for some specific under-determined linear systems,” in *Proc. Int. Conf. Acoustics, Speech, and Sig. Proc. (ICASSP)*, Philadelphia, PA, Mar. 2005.
- [6] M. E. Dominguez-Jimenez, N. Gonzalez-Prelcic, G. Vazquez-Vilar, and R. Lopez-Valcarce, “Design of universal multicoset sampling patterns for compressed sensing of multiband sparse signals,” in *Proc. Int. Conf. Acoustics, Speech, and Sig. Proc. (ICASSP)*, Prague, Czech Republic, 2012, pp. 3337–3340.
- [7] K. Konstantinides and K. Yao, “Statistical analysis of effective singular values in matrix rank determination,” *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 36, no. 5, pp. 757–763, May 1988.
- [8] T. J. Shan, M. Wax, and T. Kailath, “On spatial smoothing for estimation of coherent signals,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 806–811, Aug. 1985.
- [9] M. F. Duarte and R. G. Baraniuk, “Kronecker compressive sensing,” *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 494–504, Feb. 2012.
- [10] T. Strohmer and R. W. Heath, Jr., “Grassmannian frames with applications to coding and communication,” *Appl. Comput. Harmon. Anal.*, vol. 14, pp. 257–275, 2003.
- [11] V. N. Malozemov and A.B. Pevnyi, “Equiangular tight frames,” *Journal of Mathematical Sciences*, vol. 157, no. 6, 2009.
- [12] P. Xia, S. Zhou, and G. B. Giannakis, “Achieving the Welch bound with difference sets,” *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1900–1907, May 2005.
- [13] B. Alexeev, J. Cahill, and D. G. Mixon, “Full spark frames,” *Journal of Fourier Analysis and Applications*, vol. 18, no. 6, pp. 1167–1194, 2012.
- [14] T. G. Kolda and B. W. Bader, “Tensor decompositions and applications,” *SIAM Review*, vol. 51, no. 3, pp. 455–500, Sept. 2009.
- [15] M. Haardt, F. Roemer, and G. Del Galdo, “Higher-order SVD based subspace estimation to improve the parameter estimation accuracy in multi-dimensional harmonic retrieval problems,” *IEEE Trans. Sig. Proc.*, vol. 56, pp. 3198–3213, July 2008.
- [16] J. P. C. L. Da Costa, F. Roemer, M. Haardt, and R. T. de Sousa Jr., “Multi-Dimensional model order selection,” *EURASIP J. on Adv. Sig. Proc.*, July 2011.
- [17] N. D. Sidiropoulos and A. Kyriklidis, “Multi-way compressed sensing for sparse low-rank tensors,” *IEEE Sig. Proc. Letters*, vol. 19, no. 11, pp. 757–760, 2012.
- [18] C. F. Caiafa and A. Cichocki, “Multidimensional compressed sensing and their applications,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 3, no. 6, pp. 355–380, Nov. 2013.
- [19] M. Wax and T. Kailath, “Detection of signals by information theoretic criteria,” *IEEE Trans. Acoustics, Speech and Sig. Proc.*, vol. 33, no. 2, pp. 387–392, 1985.