Ada-JSR: SAMPLE EFFICIENT ADAPTIVE JOINT SUPPORT RECOVERY FROM EXTREMELY COMPRESSED MEASUREMENT VECTORS

Sina Shahsavari, Pulak Sarangi, Mehmet Can Hücümenoğlu, and Piya Pal

Dept. of Electrical and Computer Engineering,
University of California, San Diego, USA
E-mail: sshahsav@eng.ucsd.edu, psarangi@eng.ucsd.edu, mhucumen@eng.ucsd.edu, pipal@eng.ucsd.edu

ABSTRACT

This paper considers the problem of recovering the joint support (of size K) of a set of unknown sparse vectors in $\mathbb{R}^d$, each of which can be sensed using a different measurement matrix. Such models have wide applicability ranging from communication to multi-task learning. We develop an adaptive strategy called Adaptive Joint Support Recovery (Ada-JSR) that enables exact support recovery in the extreme compression regime with only $m = 1$ measurement per unknown vector while requiring a total complexity of no more than $K\lceil\log_2(d)\rceil$ measurements. Unlike existing support recovery techniques which require suitable assumptions on the correlation structure or distribution of the unknown signals in order to operate in the regime $m < K$, we show that the flexibility of adaptive measurement design alone allows us to operate in this extreme compression regime, without the need for imposing any correlation or sub-Gaussian priors. [1]

Index Terms— Joint Support Recovery, Multiple Measurement Vector, Adaptive Sensing, mmWave Channel Subspace Estimation, Multi-Task Learning.

1. INTRODUCTION

The problem of joint support recovery in Multiple Measurement Vector (MMV) models is extensively studied with a variety of applications from source localization, sparse linear regression, sparse spectrum sensing, to multi-task learning, group testing [1, 2]. The goal is to identify the common support of unknown sparse vectors from compressed measurements.

The most general setup of the problem concerns with recovering signals $\{\mathbf{x}_i\}_{i=1}^n \in \mathbb{C}^d$ that share a common support $\mathcal{S} \subset [d]$, $|\mathcal{S}| = K$, from their compressed linear sketches $\mathbf{y}_i = \mathbf{\Phi}_i\mathbf{x}_i,\ i \in [n]$, acquired using sensing matrices $\mathbf{\Phi}_i \in \mathbb{C}^{m\times d}$. When $n = 1$, the model reduces to the so-called Single Measurement Vector (SMV) model, where information-theoretic results show that $m = O(K\log(d/K))$ measurements are necessary and sufficient to recover a deterministic vector $\mathbf{x}$ [3–5]. In the MMV setting, recovering the $\mathbf{x}_i$'s requires $m > K$ measurements, since when $m < K$, the non-zero values of $\mathbf{x}_i$ cannot be uniquely identified even with the knowledge of the true support $\mathcal{S}$ [6, 7]. Notice that in the MMV setting, we allow $\mathbf{x}_i$'s to vary with i, as a result of which two kinds of sample complexities emerge: (i) the dimension of the compressed sketch (m), which is henceforth referred to as *Measurements Per Vector (MPV)*, and (ii) the number of measurement vectors (n). In practical settings, these two types of sample complexities may denote different physical quantities (such as space and time). The total number of measurements mn is called the *Total Sample Complexity (TSC)*. Understanding various trade-offs between MPV and TSC in the MMV model becomes an important consideration, which does not arise in SMV models.

In a large number of applications, it is the common support $\mathcal{S}$ of $\mathbf{x}_i$'s that is of interest, not the recovery of individual signals $\mathbf{x}_i$. Some examples include channel subspace estimation/channel path estimation in mmWave communication, source localization, spectrum sensing, group testing, and so forth [1, 2, 8–10]. For SMV, $m > K$ measurements are also known to be necessary for support recovery [5, 11–13]. However, for MMV models, somewhat surprisingly, the MPV can be reduced to $m < K$, under suitable assumptions on $\mathbf{x}_i$'s and yet support recovery can be guaranteed (although the recovery of $\mathbf{x}_i$'s is no longer possible).

In a series of past works, we have shown that for the standard MMV model when the measurement matrix is fixed for all i, (i.e. $\mathbf{\Phi}_i = \mathbf{\Phi}, \forall i$) certain correlation priors on $\mathbf{x}_i$'s can be exploited to attain a significant reduction in MPV ($m \ll K$). In such correlation-aware scenarios introduced in our earlier works [14–17], the $\mathbf{x}_i$'s are modeled as random sparse vectors whose non-zero elements are statistically uncorrelated. Using suitably-designed deterministic measurement matrix $\mathbf{\Phi}$, the support can be identified with $m = \Omega(\sqrt{K})$ measurements, provided that n is larger than a threshold that depends on the specific choice of $\mathbf{\Phi}$. In a recent work [18], it has been shown that *when the measurement matrices are allowed to change for every* $i \in [n]$, then an MPV of $m \geqslant \log^2 K$ is achievable with a TSC of $mn = \Theta((K^2/m)\log(K(d-K)))$, provided non-zero

[1]This work was supported in part by the Office of Naval Research under Grants ONR N00014-19-1-2227, and DE-SC0022165, and in part by the National Science Foundation under Grant NSF CAREER ECCS 1700506, and NSF 2124929.

elements of $\{\mathbf{x}_i\}_{i=1}^n$ are drawn i.i.d from a sub-Gaussian distribution and $\mathbf{\Phi}_i$'s are also i.i.d sub-Gaussian measurement matrices.

Our Contributions: In this paper, we study if the flexibility of changing $\mathbf{\Phi}_i$ alone can enable us to achieve an extreme reduction/compression per measurement vector, (in particular an MPV of $m = 1$) with a TSC of $mn = K\lceil \log_2(d) \rceil$, *without the need for imposing any correlation or sub-Gaussian priors on* $\mathbf{x}_i$*'s.* In Sec. 3, we propose an adaptive strategy for joint support recovery, called Ada-JSR, that sequentially and adaptively designs $\mathbf{\Phi}_i$'s. In our setting $\mathbf{x}_i$'s can be sparse over any basis $\mathbf{A}$, which is very relevant for applications such as mmWave channel sensing, and optical imaging, where $\mathbf{A}$ typically has a Fourier structure. Our simulations show that Ada-JSR outperforms state-of-the-art non-adaptive techniques in this extreme compressive regime.

2. PROBLEM SETUP AND BACKGROUND

Consider n unknown vectors $\mathbf{x}_i \in \mathbb{C}^d, 1 \leqslant i \leqslant n$, each of which admits a sparse representation over a known basis $\mathbf{A} \in \mathbb{C}^{d\times d}$, satisfying $\mathbf{x}_i = \mathbf{A}\mathbf{z}_i, \quad i = 1, 2\cdots, n$. The coefficients $\mathbf{z}_i$'s share a common support $\mathcal{S} \subset [d]$, $|\mathcal{S}| = K$, $(K < d)$. We acquire n compressive linear sketches (or measurement vectors) $\mathbf{y}_i \in \mathbb{C}^m, i \in [n]$ of the unknown signals, using sensing matrices $\mathbf{\Phi}_i \in \mathbb{C}^{m\times d}$ as follows,

$$\mathbf{y}_i = \mathbf{\Phi}_i \mathbf{x}_i, \quad i \in [n] \tag{1}$$

The goal is to determine the support set $\mathcal{S}$ given $\{\mathbf{y}_i\}_{i=1}^n$. Notice that since rank$(\mathbf{A}_{\mathcal{S}'}) = K$ for any $\mathcal{S}' \subset [d]$ with $|\mathcal{S}'| = K$, the problem of joint support recovery is equivalent to recovering the true subspace $\mathcal{R}(\mathbf{A}_{\mathcal{S}})$ from $\binom{N}{K}$ possible choices, each of dimension K.

In our past and recent works [14–16], we have shown that it is possible to recover $\mathcal{S}$ with a MPV of $m = \Omega(\sqrt{K})$, provided we utilize certain correlation priors on $\mathbf{x}_i$'s, and $\mathbf{\Phi}$ is suitably chosen (non-random designs). Very recently, the idea of operating with a MPV of $m < K$ has been extended to the case where $\mathbf{\Phi}_i$'s can be independently varied across measurement vectors (leading to the so-called Generalized version of the MMV model). In this setting, an MPV of $m \geqslant \log^2(K)$ was shown to be achievable with a TSC of $\Theta(K^2/m \log(K(d-K)))$, provided certain sub-Gaussian priors on the unknown $\mathbf{x}_i$'s are exploited, and $\mathbf{\Phi}_i$'s are chosen independently.

In this work, we ask the question "*Does the flexibility of changing* $\mathbf{\Phi}_i$ *per measurement vector enable us to achieve an extreme compression in MPV with* $m = 1$*, without imposing any correlation and/or sub-Gaussian priors on* $\mathbf{x}_i$*'s*?" We provide an affirmative answer to this question by proposing an adaptive method called Ada-JSR, which achieves an MPV of $m = 1$ and TSC of $mn = K\lceil \log_2(d) \rceil$ (under some mild conditions on $\mathbf{x}_i$'s) via a sequential (and adaptive) design of structured measurement matrices $\mathbf{\Phi}_i$, that are cognizant of the structure of the basis $\mathbf{A}$.

3. ADAPTIVE JOINT SUPPORT RECOVERY (Ada-JSR)

The proposed Adaptive Joint Support Recovery (Ada-JSR) strategy is inspired by techniques employed in adaptive compressed sensing [19–21], which were previously developed for the SMV model ($n = 1$). For the generalized MMV problem that is of interest to us, we show that adaptive techniques can also yield significant improvements in the MPV (enabling $m = 1$), and outperform non adaptive i.i.d designs of sensing matrices used in [18].

3.1. Adaptive measurement design

The Ada-JSR strategy is summarized in Table 1. Here we briefly explain the main idea. Each sensing vector $\boldsymbol{\phi}_i$ is associated with an ordered set of indices $\mathcal{D}_i \subseteq [d]$ (here each i corresponds to a tuple (j, k) in Table 1). Let $\mathcal{D} = \{u_1, u_2, \cdots, u_p\} \subseteq [d]$ be an ordered set such that $u_1 < u_2 < \cdots < u_p$. We define a function $G(.)$ as follows,

$$G(\mathcal{D}) = \{u_1, u_2, \cdots, u_{\lceil \frac{p}{2} \rceil}\}$$

Note that the output $G(\mathcal{D})$ is also an ordered subset of $[d]$.

Given a vector $\boldsymbol{\alpha} \in \mathbb{R}^d$ (which is provided as an input to the strategy), for each i, we generate $\boldsymbol{\phi}_i \in \mathbb{C}^d$ via generating a *structured vector* $\mathbf{w}_i$ as follows,

$$\boldsymbol{\phi}_i = (\mathbf{A}^{-H})\mathbf{w}_i, \quad \text{where} \quad \begin{cases} [\mathbf{w}_i]_{[d]/G(\mathcal{D}_i)} = \mathbf{0}, \\ [\mathbf{w}_i]_{G(\mathcal{D}_i)} = [\boldsymbol{\alpha}]_{G(\mathcal{D}_i)} \end{cases} \tag{2}$$

Using this $\boldsymbol{\phi}_i$, a measurement y_i is acquired as $y_i = \boldsymbol{\phi}_i^H \mathbf{x}_i$. If $y_i \neq 0$, we assign $\mathcal{D}_{i+1} = G(\mathcal{D}_i)$. Else, we assign $\mathcal{D}_{i+1} = \mathcal{D}_i / G(\mathcal{D}_i)$, and repeat.

Our design of $\mathbf{w}_i$ is motivated by the goal to determine if $G(\mathcal{D}_i) \bigcap \mathcal{S} = \varnothing$. Under some mild conditions on $\mathbf{x}_i$'s (as stated in Theorem 1), the measurement y_i acquired using the designed $\mathbf{w}_i$, serves as an answer to this query:

$$\begin{aligned} &\text{if} \quad y_i \neq 0, \quad \text{then} \quad G(\mathcal{D}_i) \textstyle\bigcap \mathcal{S} \neq \varnothing \\ &\text{if} \quad y_i = 0, \quad \text{then} \quad G(\mathcal{D}_i) \textstyle\bigcap \mathcal{S} = \varnothing \end{aligned} \tag{3}$$

Theorem 1. *Let* $\mathcal{S} \subset [d]$*,* $|\mathcal{S}| = K$*, and* $\{\mathbf{x}_i\}_{i=1}^n$ *be* n *unknown vectors with* $\mathbf{x}_i = \mathbf{A}\mathbf{z}_i$*, supp*$(\mathbf{z}_i) = \mathcal{S},\ \forall i \in [n]$*. Then Ada-JSR satisfies the followings:*
i) Ada-JSR terminates in no more than $K\tilde{d}$ *iterations, where* $\tilde{d} = \lceil \log_2 d \rceil$
ii) If $[\mathbf{z}_i]_l > 0$ *for all* $l \in \mathcal{S},\ i \in [n]$*, and* $\boldsymbol{\alpha} \in \mathbb{R}^d$ *is a positive vector, then the output* $\widehat{\mathcal{S}}_K$ *of Ada-JSR satisfies*

$$\widehat{\mathcal{S}}_K = \mathcal{S}$$

Sketch of proof. We provide a sketch of proof for (i) and (ii).
i) Note that at any stage $k < K$, $|\widehat{\mathcal{S}}_k| < K$, therefore the set $\mathcal{D}_0^k$ is non-empty. Also note that for $j \geqslant 1$,

$$|\mathcal{D}_j^k| = \begin{cases} \lceil \frac{|\mathcal{D}_{j-1}^k|}{2} \rceil, & \text{If} \ \mathcal{D}_j^k = \widetilde{\mathcal{D}}_{j-1}^k \\ \lfloor \frac{|\mathcal{D}_{j-1}^k|}{2} \rfloor, & \text{If} \ \mathcal{D}_j^k = \mathcal{D}_{j-1}^k / \widetilde{\mathcal{D}}_{j-1}^k \end{cases} \tag{4}$$

Table 1: Adaptive Joint Support Recovery (Ada-JSR)[2]

Input: $\mathbf{A}, \boldsymbol{\alpha}, d, K$
$k \leftarrow 0, \quad i \leftarrow 1, \quad \widehat{\mathcal{S}}_0 \leftarrow \varnothing$
while $|\widehat{\mathcal{S}}_k| < K$ **do**
  `# Beginning of Stage k:`
  $j \leftarrow 0$
  `# Removing the detected indices from the set of candidates`
  $\mathcal{D}_0^k \leftarrow [d]/\widehat{\mathcal{S}}_k$
  while $|\mathcal{D}_j^k| > 1$ **do**
    `# Beginning of Iteration j:`
    `# Measurement design`
    $\widetilde{\mathcal{D}}_j^k \leftarrow G(\mathcal{D}_j^k)$
    $[\mathbf{w}_j^k]_{[d]/\widetilde{\mathcal{D}}_j^k} \leftarrow 0,\ [\mathbf{w}_j^k]_{\widetilde{\mathcal{D}}_j^k} \leftarrow [\boldsymbol{\alpha}]_{\widetilde{\mathcal{D}}_j^k}$
    $\boldsymbol{\phi}_j^k \leftarrow (\mathbf{A}^{-H})\mathbf{w}_j^k$
    $y_i \leftarrow (\boldsymbol{\phi}_j^k)^H \mathbf{x}_i$
    `# Decision-making step`
    if $|y_i| > 0$ **then**
      $\mathcal{D}_{j+1}^k \leftarrow \widetilde{\mathcal{D}}_j^k$
    else
      $\mathcal{D}_{j+1}^k \leftarrow \mathcal{D}_j^k/\widetilde{\mathcal{D}}_j^k$
    $i \leftarrow i+1, j \leftarrow j+1$
  $\widehat{\mathcal{S}}_{k+1} \leftarrow \widehat{\mathcal{S}}_k \bigcup \mathcal{D}_j^k$
  $k \leftarrow k+1$
return $\widehat{\mathcal{S}}_k$

Hence for some $j_k\,(j_k \leqslant \widetilde{d})$, we will have $|\mathcal{D}_{j_k}^k| = 1$ and inner iterations for kth stage will end after j_k iterations. Furthermore, it can be verified that $\mathcal{D}_j^k \subseteq \mathcal{D}_{j-1}^k, \quad$ for all $j \geqslant 1, \quad k < K$. Since $\mathcal{D}_0^k \bigcap \widehat{\mathcal{S}}_k = \varnothing \implies \mathcal{D}_{j_k}^k \bigcap \widehat{\mathcal{S}}_k = \varnothing$, and therefore each stage finds a new index from $[d]$,

$$|\widehat{\mathcal{S}}_{k+1}| = |\widehat{\mathcal{S}}_k| + 1, \quad \text{for all } k < K \tag{5}$$

Hence, the outer loop will terminates after K stages, and Ada-JSR terminates after a total of $\sum_{k=0}^{K-1} j_k \leqslant K\widetilde{d}$ iterations.

(ii) This statement can be proved via induction by showing that $|\mathcal{S} \bigcap \mathcal{D}_j^k| > 0$ for all $j \geqslant 0,\ k < K$ by using the fact that $[\mathbf{z}_i]_{\mathcal{S}}, \boldsymbol{\alpha}$ are positive. In particular, it holds that $|\mathcal{S} \bigcap \mathcal{D}_{j_k}^k| > 0$ and we have already shown that $|\mathcal{D}_{j_k}^k| = 1$. This would imply that $\mathcal{D}_{j_k}^k \subseteq \mathcal{S}$ and $\mathcal{D}_{j_k}^k$ is a singleton with an element of $\mathcal{S}$. Therefore, $\widehat{\mathcal{S}}_k \subseteq \mathcal{S}$ for all $k < K$. This together with (5) and the fact that $|\mathcal{S}| = K$, implies that $\widehat{\mathcal{S}}_K = \mathcal{S}$. $\square$

Remark. The positivity assumptions on $[\mathbf{z}_i]_{\mathcal{S}}$ and $\boldsymbol{\alpha}$ in Theorem 1 are due to technical considerations, and in principle can be relaxed. We will tackle the more general case in a future work. As illustrated in our simulations, Ada-JSR is able to recover the support even when $[\mathbf{z}_i]_{\mathcal{S}}$ and $\boldsymbol{\alpha}$ violates the positivity assumptions.

Remark. The focus of this paper was to establish a TSC of order $O(K \log d)$ for exact support recovery. In the future, it will be interesting to extend the problem to the case of noisy measurements,

$$y_i = \boldsymbol{\phi}_i^H \mathbf{x}_i + e_i, \quad i \in [n] \tag{6}$$

where e_i denotes the additive measurement noise. If the noise statistics are unknown and only an upper bound on the noise magnitude is available, our method can be suitably modified to recover the support from such noisy measurements. In particular, a simple modification involves altering the decision-making step by comparing the $|y_i|$ against a threshold τ instead of 0.

4. APPLICATION OF Ada-JSR IN mmWAVE COMMUNICATION

In this section, we briefly demonstrate how Ada-JSR can be applied to the important problem of Channel Subspace Estimation in Hybrid mmWave communication systems [8, 9]. According to the geometrical channel model, the SIMO uplink channel between a single-antenna mobile station and the multi-antenna base station is given by [8, 9]:

$$\boldsymbol{h}_t = \mathbf{A}\mathbf{g}_t, \quad t = 1, 2, \cdots, n$$

Here $\mathbf{g}_t \in \mathbb{C}^d$ represents the channel gain, and $\mathbf{A} = [\mathbf{a}(\theta_1), \cdots, \mathbf{a}(\theta_d)] \in \mathbb{C}^{d \times d}$ is a matrix of array steering vectors associated with quantized Angles of Arrivals (AoAs), $[\mathbf{a}(\theta)]_l = e^{j\pi l \sin(\theta)}, l = 1, \cdots, d$. Notice that we assumed the grid size to be same as the number of antennas. Since in mmWave communication, it is common to deploy a large number of antennas in a massive MIMO configuration, this assumption does not significantly limit the grid size/resolution. Due to the sparse nature of the mmWave channel, the gains $\mathbf{g}_t$'s are typically assumed to be K-sparse where K represents the total number of channel paths. In low mobility scenarios, although the gains are changing, their support $\mathcal{S}$ remains unchanged [9]. Therefore, the low-dimensional channel subspace can be estimated by estimating the support set $\mathcal{S}$. We consider a practical scenario where the hardware architecture is equipped with only a single Radio Frequency (RF) chain. The signal at the output of the RF chain is given by:

$$\boldsymbol{z}_t = \boldsymbol{w}_t^H \boldsymbol{A}\boldsymbol{g}_t, \quad t \in [n] \tag{7}$$

Here $\boldsymbol{w}_t$ denotes the beamforming vector used at the t-th time slot. The proposed Ada-JSR strategy can be applied to (7) where the beamforming vectors $\boldsymbol{w}_t$ are designed adaptively to aid channel subspace estimation with low training overhead of at most $K\lceil \log_2(d) \rceil$ training time slots. Note that Ada-JSR departs from the adaptive strategy proposed in [22], where the beamformers are selected from a pre-designed hierarchical codebook requiring a total $O(K^2)$ training time slots with 1 RF chain. Furthermore, the resulting beamformers generated

[2]The implementation of AdaJSR is available here

by Ada-JSR are computed "on-the-fly" which can lead to non-hierarchical multi-lobe beamformers. The implementation of such multi-lobe beamformers under RF hardware constraints would be an interesting direction for future work.

5. NUMERICAL RESULTS

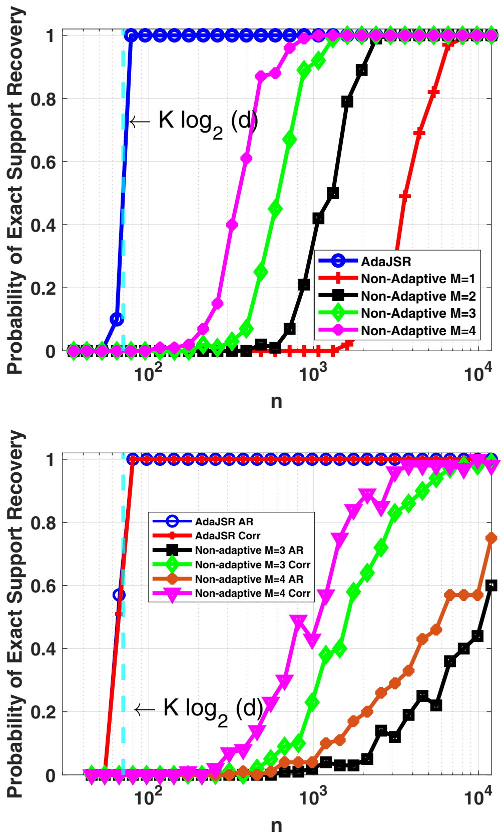

Fig. 1: Probability of exact support recovery of Ada-JSR and the non-adaptive method in [18] when (top) signals $\mathbf{z}_i$'s are drawn from i.i.d Gaussian distributions, and (bottom) signals $\mathbf{z}_i$'s obey correlation priors [Corr: Correlated signals, AR: Autoregressive model]. In both figures, $\alpha_i \overset{i.i.d}{\sim} \mathcal{N}(0,1), i = 1, 2, \cdots, d^3$.

In the first set of simulations, in order to compare the performance of Ada-JSR with the joint support recovery technique in [18], we assume $\mathbf{A} = \mathbf{I}_d$. In Fig. 1 (top) we show the probability of exact support recovery of Ada-JSR for $K = 10$, and $d = 100$, as a function of the n when $[\mathbf{z}_i]_{\mathcal{S}} \overset{i.i.d}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_K)$. We compare Ada-JSR against the non-adaptive method proposed in [18] with plots for different values of $m\,(m \geqslant 1)$ overlaid. As can be seen, the Ada-JSR only requires at most $K\lceil\log_2(d)\rceil$ total measurements, whereas the non-adaptive method needs a TSC of order $\Theta((K^2/m)\log(d(K-d)))$. to achieve exact support recovery, even with $m > 1$. Next, we assume the unknown signal $\mathbf{x}_i$'s possess an additional correlation structure besides the sparsity. In particular, we consider two types of signals which are encountered in many real-world applications: 1)$\{\mathbf{z}_i\}_{i=1}^n$ are generated according to a vector Autoregressive process, 2) Spatially correlated signals (non-zero correlation between the entries of $[\mathbf{z}_i]_{\mathcal{S}}$). In Fig. 1 (bottom), we show that such additional assumptions do not affect the sample complexity of Ada-JSR. However, the performance of the non-adaptive methods from [18] significantly degrades, since $\mathbf{z}_i$'s generated as above violate the assumptions in [18].

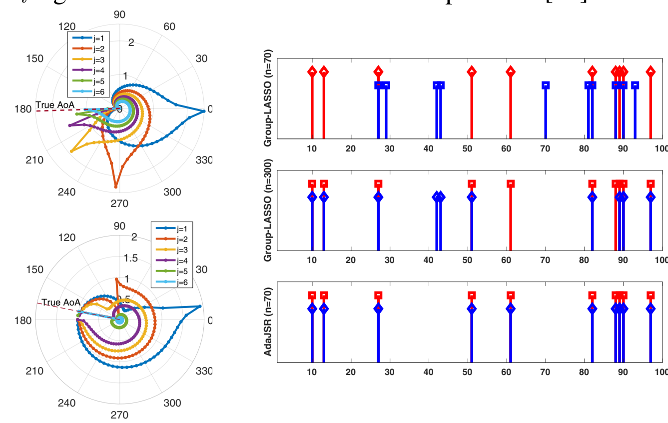

Fig. 2: Left) Measurement vectors/Beamformers generated by Ada-JSR, visualized in frequency domain. Here, $d = 100$, top: $k = 1 < K = 5$, bottom: $k = K = 5$. Right) Channel path estimation performance ($K = 10, m = 1, d = 100$): Group LASSO ($n = 70, 300$) vs Ada-JSR ($n = 70$). Here red shows true channel paths, whereas blue shows the recovery results.

Finally, we employ Ada-JSR to solve a channel subspace estimation problem as discussed in Section 4. Measurements follow the model (7), where $[\mathbf{g}_t]_{\mathcal{S}} \overset{i.i.d}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_K)$ and θ_i's are chosen on a grid of size 100 from 0 to 2π. In Fig. 2 (right) we show that Ada-JSR can effectively resolve the channel paths with 1 RF-chain and very limited temporal snapshots ($n = 70$). However, Group LASSO [23], which is a well-known non-adaptive method for joint support recovery, fails to detect the AoA's using $m = 1$ even with a TSC of 300 measurements.

In Fig. 2 (left), we also provide a frequency domain representation of the measurement vectors ϕ_j^k's which are designed by the Ada-JSR process. It is interesting to observe that they mimic the beam patterns of multi-lobe beamformers, and they look very different from the hierarchical beamformers designed by Alkhateeb et al, in [22].

6. CONCLUSION

We showed that it is possible to recover the joint support in a generalized MMV problem in the extreme compression regime of $m = 1$ measurement per vector, while maintaining a TSC of at most $K\lceil\log_2 d\rceil$ measurements, without the need for imposing any correlation priors on the unknown signals. In particular, we achieve this by proposing an adaptive sensing strategy (Ada-JSR) which designs the sensing operators sequentially and adaptively. Numerical simulations demonstrate the effectiveness of our approach, and its potential application in mmWave channel sensing with low training overhead, which can be an exciting direction for future research.

[3]Although the Theorem 1 assumes $\boldsymbol{\alpha}$ to be a positive vector, in simulations, a random choice of $\boldsymbol{\alpha}$ (in this case, from standard normal distribution) also leads to successful recovery with high probability.

References

[1] D. Yin, R. Pedarsani, Y. Chen, and K. Ramchandran, “Learning mixtures of sparse linear regressions using sparse graph codes,” *IEEE Transactions on Information Theory*, vol. 65, no. 3, pp. 1430–1451, 2018.

[2] A. Argyriou, T. Evgeniou, and M. Pontil, “Convex multi-task feature learning,” *Machine learning*, vol. 73, no. 3, pp. 243–272, 2008.

[3] E. J. Candès, 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, 2006.

[4] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Transactions on information theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[5] M. J. Wainwright, “Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting,” *IEEE transactions on information theory*, vol. 55, no. 12, pp. 5728–5741, 2009.

[6] B. D. Rao, K. Engan, and S Cotter, “Diversity measure minimization based method for computing sparse solutions to linear inverse problems with multiple measurement vectors,” in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, IEEE, vol. 2, 2004, pp. ii–369.

[7] J. Chen and X. Huo, “Theoretical results on sparse representations of multiple-measurement vectors,” *IEEE Transactions on Signal processing*, vol. 54, no. 12, pp. 4634–4643, 2006.

[8] S. Park, J. Park, A. Yazdan, and R. W. Heath, “Exploiting spatial channel covariance for hybrid precoding in massive mimo systems,” *IEEE Transactions on Signal Processing*, vol. 65, no. 14, pp. 3818–3832, 2017.

[9] S. Haghighatshoar and G. Caire, “Low-complexity massive mimo subspace estimation and tracking from low-dimensional projections,” *IEEE Transactions on Signal Processing*, vol. 66, no. 7, pp. 1832–1844, 2018.

[10] P. Sarangi, S. Shahsavari, and P. Pal, “Robust doa and subspace estimation for hybrid channel sensing,” in *2020 54th Asilomar Conference on Signals, Systems, and Computers*, IEEE, 2020, pp. 236–240.

[11] M. Akcakaya and V. Tarokh, “Shannon-theoretic limits on noisy compressive sampling,” *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 492–504, 2009.

[12] G. Tang and A. Nehorai, “Performance analysis for sparse support recovery,” *IEEE transactions on information theory*, vol. 56, no. 3, pp. 1383–1399, 2010.

[13] Y. Jin, Y.-H. Kim, and B. D. Rao, “Limits on support recovery of sparse signals via multiple-access communication techniques,” *IEEE Transactions on Information Theory*, vol. 57, no. 12, pp. 7877–7892, 2011.

[14] P. Pal and P. Vaidyanathan, “Pushing the limits of sparse support recovery using correlation information,” *IEEE Transactions on Signal Processing*, vol. 63, no. 3, pp. 711–726, 2014.

[15] A. Koochakzadeh, H. Qiao, and P. Pal, “On fundamental limits of joint sparse support recovery using certain correlation priors,” *IEEE Transactions on Signal Processing*, vol. 66, no. 17, pp. 4612–4625, 2018.

[16] H. Qiao and P. Pal, “Guaranteed localization of more sources than sensors with finite snapshots in multiple measurement vector models using difference co-arrays,” *IEEE Transactions on Signal Processing*, vol. 67, no. 22, pp. 5715–5729, 2019.

[17] S. Shahsavari, J. Millhiser, and P. Pal, “Fundamental trade-offs in noisy super-resolution with synthetic apertures,” in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2021, pp. 4620–4624.

[18] L. Ramesh, C. R. Murthy, and H. Tyagi, “Sample-measurement tradeoff in support recovery under a subgaussian prior,” *IEEE Transactions on Information Theory*, 2021.

[19] A. Gupta, R. Nowak, and B. Recht, “Sample complexity for 1-bit compressed sensing and sparse classification,” in *2010 IEEE International Symposium on Information Theory*, IEEE, 2010, pp. 1553–1557.

[20] C. Aksoylar and V. Saligrama, “Information-theoretic bounds for adaptive sparse recovery,” in *2014 IEEE International Symposium on Information Theory*, IEEE, 2014, pp. 1311–1315.

[21] A. Koochakzadeh, P. Sarangi, and P. Pal, “A sequential approach for sparse support recovery using correlation priors,” in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, IEEE, 2019, pp. 586–590.

[22] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, “Channel estimation and hybrid precoding for millimeter wave cellular systems,” *IEEE journal of selected topics in signal processing*, vol. 8, no. 5, pp. 831–846, 2014.

[23] M. Yuan and Y. Lin, “Model selection and estimation in regression with grouped variables,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.