arXiv:1410.4593v1 [math.ST] 16 Oct 2014

Adaptive Compressed Sensing for Estimation of Structured Sparse Sets

Rui M. Castro and Ervin Tánczos*

October 20, 2014

Abstract

This paper investigates the problem of estimating the support of structured signals via adaptive compressive sensing. We examine several classes of structured support sets, and characterize the fundamental limits of accurately estimating such sets through compressive measurements, while simultaneously providing adaptive support recovery protocols that perform near optimally for these classes. We show that by adaptively designing the sensing matrix we can attain significant performance gains over non-adaptive protocols. These gains arise from the fact that adaptive sensing can: (i) better mitigate the effects of noise, and (ii) better capitalize on the structure of the support sets.

1 Introduction

Compressive sensing provides an efficient way to estimate signals that have a sparse representation in some basis or frame [11], [14], [10], [12], [21]. If the measurements can be chosen adaptively it is possible to achieve performance gains in the sense that weaker signals can be estimated more accurately than in the non-adaptive setting [13],[16]. Furthermore, in some situations the signal of interest may have additional structure that can be exploited. For instance, in gene expression studies the signals of interest are supported on a submatrix of the gene-expression matrix, and are not arbitrary sparse signals. In network monitoring anomalous behavior may "radiate" from infected nodes creating star-shaped patterns in the network graph. The natural question that arises is if further performance gains can be realized using this structural information when estimating signals using compressive measurements? Furthermore, can adaptively and sequentially designing the sensing actions provide further performance gains over non-adaptive schemes? The answer to both questions is essentially affirmative, and this work quantifies such gains in a general way.

Related work. The current work is built on a number of recent contributions on detection and estimation of sparse signals using compressive sensing. Considering general sparse signals without structure [9] and [13] provide theoretical performance limits of adaptive compressive sensing, whereas [15], [17] and [16] provide efficient near optimal procedures

*The authors are with the Department of Mathematics, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands (email `e.t.tanczos@tue.nl` and `rmcastro@tue.nl`). This work was partially supported by NWO Grant 613.001.114.

for estimation. Considering the problem of detection [5] provides both theoretical limits and optimal procedures both in the non-adaptive and adaptive sensing settings.

The problem of estimating structured sparse signals was examined in the past in several different settings. In the normal means model [4], [1] and [8] consider estimating various structured signals in the non-adaptive framework. In [19] the authors examine the same problem when measurements are collected adaptively. Similar problems have been investigated in the compressive sensing setting as well. In [7] and [18] the authors consider recovering tree-structured signals, whereas [6] investigates the problem of finding block-structured activations in a signal matrix considering both non-adaptive and adaptive measurements.

Contributions. In this work we further investigate the problem of recovering the support of structured sparse signals using adaptive compressive measurements. Our focus is on the performance gains one can achieve when adaptively designing the sensing matrix compared to the situation where the sensing matrix is constructed non-adaptively. The classes of structured support sets under consideration in this paper are

- **s-sets:** any subset of $\{1, \dots, n\}$ with size s
- **s-intervals:** sets consisting of s consecutive elements of $\{1, \dots, n\}$
- **unions of s-intervals:** unions of k disjoint s-intervals
- **s-stars:** any star of size s in a complete graph (where the edges of the graph are identified with $\{1, \dots, n\}$)
- **unions of s-stars:** unions of k disjoint s-stars
- **s-submatrices:** any submatrix of a given size $s_r \times s_c$ of an $n_r \times n_c$ matrix

We analyze the fundamental limits of recovering support sets of the above classes under non-adaptive and adaptive sensing paradigms. This is done by showing both upper and lower performance bounds. Furthermore, we provide adaptive sensing protocols with near optimal performance to show the tightness of the lower bounds, and to illustrate how adaptive compressed sensing can capitalize on the structure of the support sets in the estimation. Finally, we provide procedures that next to being near optimal also perform estimation using only a small number of measurements and are thus feasible from a practical point of view.

Table 1 summarizes our results, showing necessary and sufficient conditions for the signal magnitude at which accurate support estimation is possible in the various scenarios. It also highlights two different facets of the gains of adaptive sensing over non-adaptive sensing. First, note that the necessary conditions of non-adaptive sensing include a $\sqrt{\log n}$ factor for each of the classes under consideration. This factor is replaced by the logarithm of the sparsity when considering adaptive sensing, and this is due to the fact that adaptive strategies are better able to mitigate the effects of noise. Second, for certain classes adaptive sensing can gain greater leverage from the structure of the support sets compared to non-adaptive sensing. This phenomenon is best visible considering the class of s-stars, where estimators using non-adaptive sensing gain practically nothing from the structural information whereas adaptive sensing benefits greatly from it. Note that the necessary and

Table 1: Summary of scaling laws for the signal magnitude.

	Non-Adaptive Sensing (necessary)	Adaptive Sensing (necessary)	Adaptive Sensing (sufficient)
s-sets	$\mu \sim \sqrt{\frac{n}{m} \log n}$	$\mu \sim \sqrt{\frac{n}{m} \log s}$	$\mu \sim \sqrt{\frac{n}{m} \log s}$
unions of k disjoint s-intervals	$\mu \sim \frac{1}{s}\sqrt{\frac{n}{m} \log \frac{n}{ks}}$	$\mu \sim \frac{1}{s}\sqrt{\frac{n}{m} \log ks}$	$\mu \sim \frac{1}{s}\sqrt{\frac{n}{m} \log ks}$
unions of k disjoint s-stars	$\mu \sim \sqrt{\frac{n}{m} \log \frac{\sqrt{n}}{ks}}$	$\mu \sim \frac{1}{s}\sqrt{\frac{n}{m} \log ks}$	$\mu \sim \frac{1}{s}\sqrt{\frac{n}{m} \log ks}$
$\sqrt{s} \times \sqrt{s}$ submatrices of an $\sqrt{n} \times \sqrt{n}$ matrix	$\mu \sim \sqrt{\frac{n}{\sqrt{s}m} \log \frac{n}{s}}$	$\mu \sim \frac{1}{s}\sqrt{\frac{n}{m} \log s}$	$\mu \sim \sqrt{\frac{n}{sm} \log s}$

Scaling laws for the signal magnitude μ (constants omitted) which are necessary/sufficient for $\max_{S \in \mathcal{C}} \mathbb{E}(\widehat{S} \triangle S) \to 0$ as $n \to \infty$, where $\mathcal{C}$ denotes the corresponding class of support sets. All the results assume sparsity, for exact conditions see relevant propositions of Section 3.1.

sufficient conditions for the class of submatrices using adaptive sensing do not match, and a full characterization of the problem in that case remains open. We also remark at this point that the results derived in this paper are non-asymptotic in nature and also account for the constant factors in the scaling laws. The asymptotic presentation in Table 1 merely makes it easier to highlight the main contributions of the work.

The paper is structured as follows. Section 2 describes the problem setting in detail. In Section 3 we provide adaptive sensing procedures for structured support recovery and analyze the theoretical limits of the problem, both under non-adaptive and adaptive sensing paradigms. In this section we only make a restriction to the sensing power available, but not on the number of projective measurements we are allowed to make. In Section 4 we further restrict the number of measurements. Finally we provide some concluding remarks in Section 5.

2 Problem Setting

In this work we consider the following statistical model. Let $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ be a vector of the form

$$x_i = \begin{cases} \mu & , \text{if } i \in S \\ 0 & , \text{if } i \notin S \end{cases} , \tag{1}$$

where $\mu > 0$ and S is an unknown element of a class of sets denoted by $\mathcal{C}$. We refer to $\boldsymbol{x}$ as the *signal* and to S as the *support* or *significant/active components* of the signal. The set S is our main object of interest. The signal model (1) may seem overly restrictive at first because of the fact that each non-zero entry has the same value μ. However, our results can be generalized to signals with active components of arbitrary magnitudes and signs, in which case the value μ would play the role of the minimal absolute value of the non-zero components. For sake of simplicity we do not discuss this extension here, but refer the reader to [1], [5], [16] for details on how this can be done.

We are allowed to collect multiple measurements of the form

$$Y_j =< A_j, \boldsymbol{x} > +W_j,\ j = 1, 2, \dots\ , \tag{2}$$

where j indexes the jth measurement. Thus each measurement is the inner product of the signal $\boldsymbol{x}$ with the vector $A_j \in \mathbb{R}^n$, contaminated by Gaussian noise. The noise terms $W_j \sim \mathcal{N}(0, 1)$ are independent and identically distributed (i.i.d.) standard normal random variables, also independent of $\{A_i\}_{i=1}^{j}$. Under the adaptive sensing paradigm A_j are allowed to be functions of the past observations $\{Y_i, A_i\}_{i=1}^{j-1}$. This model is only interesting if one poses some constraint on the total amount of sensing energy available. Let A denote the matrix whose jth row is A_j. We require

$$\sup_{S \in \mathcal{C}} \mathbb{E}_S \left(\|A\|_F^2 \right) \ = \ \mathbb{E}_S \left(\sum_j \|A_j\|^2 \right) \ \leq m\ , \tag{3}$$

where $\| \cdot \|_F$ denotes the Frobenius norm, m is our total energy budget, and $\mathbb{E}_S$ denotes the expectation with respect to the joint distribution of $\{A_j, Y_j\}_{j=1,2,\dots}$ when $S \in \mathcal{C}$ is the support set.

2.1 Inference Goals

This work aims at characterizing the difficulty of recovering structured sparse signal supports with adaptive compressive sensing. We are interested in settings where the class $\mathcal{C}$ contains sets with some sort of structure, for instance the active components of $\boldsymbol{x}$ are consecutive. For the unstructured case, that is, when $\mathcal{C}$ contains every set of a given cardinality, there already exists a lower bound in [13], and a procedure that achieves this lower bound in [16]. The main goal of this work is to provide results for the problem of recovering structured sparse sets.

We are interested in two aspects of adaptive compressive sensing. First, given n, m, ε and $\mathcal{C}$ the aim is to characterize the minimal signal strength μ for which S can be reliably estimated, namely, to ensure that for a given $\varepsilon > 0$,

$$\max_{S \in \mathcal{C}} \mathbb{E}_S(|\widehat{S} \triangle S|) \leq \varepsilon\ , \tag{4}$$

where $\widehat{S} \triangle S$ is the symmetric set difference. Furthermore, we aim to construct an adaptive sensing strategy which produces an estimate $\widehat{S}$ of the unknown support set S that is able to achieve above lower bounds. Although the setting above makes sense whenever $\varepsilon \in [0, |S|]$, the problem is only interesting when ε is small. Hence we will take ε as an element of $[0, 1]$. Second, given n, m, μ, ε and $\mathcal{C}$ we wish to characterize the minimal number of samples needed to ensure (4). Considering the unstructured case, we know that non-adaptive procedures need at least $O(s \log \frac{n}{s})$ measurements [2] and that this bound is achievable [12] (these results apply when the signal strength μ is close to the threshold of estimability). On the other hand, to the best knowledge of the authors, an exact characterization of the sample complexity for adaptive procedures is not yet available, though there has been work done on the topic [3]. In that work the authors present a result that states the sample complexity of the problem scales essentially as s. However, it is not clear if that bound it tight, as there are no procedures able to attain that lower bound. In Section 5 we provide a bit more insight on this question.

In what follows we use the notation $\mathbf{1}$ to denote both the usual indicator function (e.g., $\mathbf{1}\{i \in S\}$ takes the value 1 if $i \in S$ and zero otherwise), and to denote binary vectors with support S. For instance $\mathbf{1}_S$ denotes an element of $\{0,1\}^n$ for which the entries in S have value 1 and all the other entries have value 0. Furthermore, let $\mathbb{P}_S$ denote the joint distribution of $\{A_j, Y_j\}_{1,2,\ldots}$ when $S \in \mathcal{C}$ is the support set, and $\mathbb{E}_S$ denote the expectation with respect to $\mathbb{P}_S$.

3 Signal strength

We now examine the minimal signal strength required to reliably recover structured support sets. In this setup we are allowed to make potentially infinite amount of measurements of the form (2). This might not be feasible from a practical standpoint, however, it serves as a good starting point to understand the limitations of adaptive compressive sensing.

3.1 Procedures

It is instructive to briefly consider a simple support recovery algorithm for the unstructured case. When the support set can be any set of a given cardinality and there is no restriction on the number of samples we are allowed to take the situation becomes similar to that of [19], where the authors consider coordinate-wise observations. A simple procedure in this case is to perform a Sequential Likelihood Ratio Test (SLRT) for each coordinate separately. More precisely for every coordinate $i = 1, \ldots, n$ collect observations of the form

$$Y_{i,j} = ax_i + W_j =< a\mathbf{1}_{\{i\}}, \boldsymbol{x} > +W_j \ , j = 1, \ldots, N_i \ ,$$

with some fixed $a > 0$, where we recall that $\mathbf{1}_{\{i\}}$ is a singleton vector. The number of observations N_i is random and is given by

$$N_i = \min \left\{ n \in \mathbb{N} : \sum_{j=1}^{n} \log \frac{\mathrm{d}\mathbb{P}_1(Y_j)}{\mathrm{d}\mathbb{P}_0(Y_j)} \notin (l, u) \right\} \ ,$$

where $\mathbb{P}_0$ ($\mathbb{P}_1$) is the distribution of the observations when component i is non-active (active), and $l < 0 < u$ are the lower and upper stopping boundaries of the SLRT. Then our estimator $\widehat{S}$ will be the collection of components i for which the log-likelihood process above hits the upper stopping boundary u. Considering the test of component $\boldsymbol{x}_i$ we have the following.

Lemma 1. *Set $l = \log \frac{\beta}{1-\alpha}$ and $u = \log \frac{1-\beta}{\alpha}$, and let the type I and type II error probabilities of the SLRT described above be α_a and β_a. Then $\alpha_a \to \alpha$ and $\beta_a \to \beta$ as $a \to 0$. Furthermore*

$$a^2 \mathbb{E}_0(N_i) \leq \frac{2}{\mu^2} \left(\alpha \log \frac{\alpha}{1-\beta} + (1-\alpha) \log \frac{1-\alpha}{\beta} \right) \leq \frac{2}{\mu^2} \log \frac{1}{\beta}$$

and

$$a^2\mathbb{E}_1(N_i) \le \frac{2}{\mu^2}\left(\beta \log \frac{\beta}{1-\alpha} + (1-\beta)\log\frac{1-\beta}{\alpha}\right) \le \frac{2}{\mu^2}\log\frac{1}{\alpha}$$

as $a \to 0$.

Proof. The proof goes the same way as that of Proposition 1 in [19]. □

Using the previous result we can immediately analyze the procedure above. Set $\alpha = \varepsilon/2n$ and $\beta = \varepsilon/2s$ in the proposition above, and choose a to be small. Hence α_a and β_a will be close to the nominal error probabilities α and β and we ensure (4). Then using the other part of Lemma 1 we can upper bound the expected energy used by the tests. Summing this over all the tests and using (3) we arrive at the following.

Proposition 1. *Testing each component* $\boldsymbol{x}_i,\ i = 1,\ldots,n$ *as described above yields an estimator satisfying* (3) *and* (4) *whenever*

$$\mu \ge \sqrt{\frac{2n}{m}\log\frac{2s}{\varepsilon} + \frac{2s}{m}\log\frac{2n}{\varepsilon}}\ .$$

When the support is sparse, the first term dominates the bound above. This coincides with the lower bound of [13] showing that the simple procedure above is near optimal.

Remark 1. *Note that the lower bound presented in [13] is valid for a slightly broader class than the s-sets, namely one also has to include* $(s-1)-$ *sets into the class. However, the procedure outlined above works without any modifications for this broadened class as well, and so the result of Proposition 1 holds for this larger class. A similar comment applies to all the procedures presented later on: the procedures are presented for classes of a given sparsity for sake of clarity, but the analysis shows that they also work for classes containing sets of slightly different sparsity. This is important to note as because of technical reasons the lower bounds of Section 3.2 can only deal with such classes.*

The procedures for recovering structured support sets will be very similar in nature, but slightly modified to take advantage of the structural information. In particular we know from [5] that it is possible to detect the presence of weak signals using compressive sensing. In order to take advantage of this property our procedures consist of two phases: a *search phase* and a *refinement phase*. The aim of the search phase is to find the approximate location of the signal using a detection type method, namely by identifying a subset of components $\boldsymbol{P} \subset \{1,\ldots,n\}$ which is small, on one hand, but contains the true support with high probability. More precisely $|\boldsymbol{P}| \ll n$ and $S \subset \boldsymbol{P}$ with high probability. Once this is done we can focus our attention exclusively on $\boldsymbol{P}$ in the refinement phase and estimate the support in the same manner as in the unstructured case.

3.1.1 Unions of s-intervals

Consider the class of sets that are unions of k disjoint intervals of length s. Formally,

$$\mathcal{C} = \left\{ S \subset \{1,\ldots,n\} :\ S = \bigcup_{i=1}^{k} S_i\ ,\ S_i = \{l_i,\ldots,l_i+s-1\},\ S_i \cap S_j = \emptyset\ \forall i \ne j \right\}\ .$$

Our procedure for estimating S is as follows. Split the index set $\{1,\ldots,n\}$ into consecutive bins of length $s/2$ denoted by $\boldsymbol{P}^{(1)},\ldots,\boldsymbol{P}^{(2n/s)}$. We suppose $2n$ is divisible by s, as it makes the presentation less cluttered. The procedure can be easily modified in case the previous condition is not met. Of these bins at least k (and at most $2k$) are contained entirely in S. In the search phase we aim to find the approximate location of the support by finding k such bins. To do this we test the following hypotheses

$$H_0^{(i)}:\ \boldsymbol{P}^{(i)}\cap S=\emptyset \qquad \text{versus} \qquad H_1^{(i)}:\ \boldsymbol{P}^{(i)}\subset S \qquad i=1,\ldots,2n/s\ .$$

We use a SLRT to decide between $H_0^{(i)}$ and $H_1^{(i)}$ for each $i=1,\ldots,n$, all with the same type I and type II error probabilities α and β. The choices of α and β and the exact way of carrying out the tests will be described later. As an output of the search phase, we define the set $\boldsymbol{P}$ based on the tests above. Since some $\boldsymbol{P}^{(i)}$ may only partially intersect the support S we set $\boldsymbol{P}$ to be the union of those bins $\boldsymbol{P}^{(i)}$ for which either $H_1^{(i-1)}, H_1^{(i)}$ or $H_1^{(i+1)}$ was accepted. This way we ensure $\mathbb{P}_S(S\not\subseteq\boldsymbol{P})\leq 2k\beta$. We also wish to ensure that $\boldsymbol{P}$ is small, and to do so we must to choose α appropriately. Once this is done we can move on to the search phase and find the support within $\boldsymbol{P}$. We can do this in a very crude way and use a similar procedure as in the unstructured case with type I and II error probabilities α',β'. The sensing energy used in this phase will be negligible due to $\boldsymbol{P}$ being small. Finally the estimator $\widehat{S}$ will be the collection of components that were deemed active at the end of the refinement phase. Note that the procedure above might be improved in several ways, for instance, by using a more sophisticated method in the refinement phase. Nevertheless, for sake of simplicity we consider the procedure outlined above.

We now choose $\alpha,\beta,\alpha',\beta'$ to ensure the estimator satisfies (4). We write

$$\begin{aligned}\mathbb{E}_S\left(|\widehat{S}\triangle S|\right) &\leq \mathbb{E}_S\left(|\widehat{S}\triangle S|\ \middle|S\not\subseteq\boldsymbol{P}\right)\mathbb{P}_S(S\not\subseteq\boldsymbol{P})+\mathbb{E}_S\left(|\widehat{S}\triangle S|\ \middle|S\subseteq\boldsymbol{P}\right)\\ &\leq \mathbb{E}_S\left(|S\setminus\boldsymbol{P}|+\sum_{i\in\boldsymbol{P}:\ i\notin S}\alpha'+\sum_{i\in\boldsymbol{P}:\ i\in S}\beta'\ \middle|\ S\not\subseteq\boldsymbol{P}\right)2k\beta\\ &\quad+n\alpha'+ks\beta'\ .\end{aligned}$$

Hence choosing $\alpha'=\varepsilon/4n,\beta'=\varepsilon/4ks$ and $\beta=\varepsilon/8k^2s^2$ ensures (4). Note that α does not influence the probability of error. However, it will influence the size of $\boldsymbol{P}$, and hence the total sensing energy required by the procedure. Note also that the choices above are very conservative and can be improved. Nonetheless these simple choices lead to near optimal performance.

To perform the ith test of the search phase we collect measurements using projection vectors of the form $a\mathbf{1}_{\boldsymbol{P}^{(i)}}$ with some $a>0$ and perform a SLRT with stopping boundaries $l<0<u$. Let $\mathbb{E}_0$ and $\mathbb{E}_1$ denote the expectation when $H_0^{(i)}$ or $H_1^{(i)}$ is true respectively. Similarly to the unstructured case we now have the following.

Lemma 2. *Set $l=\log\frac{\beta}{1-\alpha}$ and $u=\log\frac{1-\beta}{\alpha}$, and let the type I and type II error probabilities of the SLRT described above be α_a and β_a. Then $\alpha_a\to\alpha$ and $\beta_a\to\beta$ as $a\to 0$. Furthermore*

$$a^2\mathbb{E}_0(N_i)\leq\frac{2}{(s/2)^2\mu^2}\left(\alpha\log\frac{\alpha}{1-\beta}+(1-\alpha)\log\frac{1-\alpha}{\beta}\right)\leq\frac{2}{(s/2)^2\mu^2}\log\frac{1}{\beta}$$

and

$$a^2\mathbb{E}_1(N_i) \le \frac{2}{(s/2)^2\mu^2}\left(\beta\log\frac{\beta}{1-\alpha} + (1-\beta)\log\frac{1-\beta}{\alpha}\right) \le \frac{2}{(s/2)^2\mu^2}\log\frac{1}{\alpha}$$

as $a \to 0$.

Using this we can upper bound the amount of sensing energy used for the test of $\boldsymbol{P}^{(i)}$ under $H_0^{(i)}$ and $H_1^{(i)}$. However, now it is possible that neither $H_0^{(i)}$ nor $H_1^{(i)}$ is true for a given bin $\boldsymbol{P}^{(i)}$. Considering a test where neither of them is true we can still carry out the the same calculations as in Lemma 1 and thus upper bound the expected sensing energy used for the test.

Lemma 3. *Set* $l = \log\frac{\beta}{1-\alpha}$ *and* $u = \log\frac{1-\beta}{\alpha}$*, and let* $\widetilde{s}$ *denote the true number of signal components in* $\boldsymbol{P}^{(i)}$*. Suppose that in the setting above neither* $H_0^{(i)}$ *nor* $H_1^{(i)}$ *is true, that is* $0 < \widetilde{s} < s/2$*. Furthermore suppose* $\widetilde{s} \ne s/4$*. Then as* $a \to 0$ *we have*

$$a^2\mathbb{E}_{\widetilde{s}}(N) \le \frac{2}{s\mu^2}\log\max\left\{\frac{1-\alpha}{\beta}, \frac{1-\beta}{\alpha}\right\} \le \frac{2}{s\mu^2}\log\frac{1}{\min\{\alpha,\beta\}}\ ,$$

where $\mathbb{E}_{\widetilde{s}}$ *denotes the expectation when the number of signal components in* $\boldsymbol{P}^{(i)}$ *is* $\widetilde{s}$*.*

Proof. Fix $i \in \{1,\ldots,n\}$. The log-likelihood ratio for an observation Y is again

$$z = \log\frac{\mathrm{d}\mathbb{P}_1(Y)}{\mathrm{d}\mathbb{P}_0(Y)} = \frac{as\mu Y}{2} - \frac{a^2s^2\mu^2}{8}\ .$$

Suppose first that $s/4 < \widetilde{s} < s/2$. Note that now the drift of the log-likelihood ratio process is positive. Now $z_1 \sim N\left((\widetilde{s}-\frac{s}{4})\frac{a^2s\mu^2}{2}, \frac{a^2s^2\mu^2}{4}\right)$. From normality we still have $\mathbb{E}(z_1|z_1 \ge 0) \ge \mathbb{E}(z_1 - c|z_1 \ge c),\ \forall c > 0$. Combining this with Wald's identity we get

$$\mathbb{E}(N)\mathbb{E}(z_1) = \mathbb{E}(\bar{z}_N) \le u + \mathbb{E}(z_1|z_1 \ge 0)\ .$$

Denoting $\xi \sim N(0,1)$ we also have

$$\begin{aligned}\mathbb{E}(z_1|z_1 \ge 0) &\le 2\mathbb{E}(z_1\mathbf{1}\{z_1 \ge 0\})\\ &\le (\widetilde{s}-\frac{s}{4})\frac{a^2s\mu^2}{2} + 2\mathbb{E}\left(\frac{as\mu}{2}\xi\mathbf{1}\{\xi \ge -(\widetilde{s}-\frac{s}{4})\mu\}\right)\\ &\le as\mu\left((\widetilde{s}-\frac{s}{4})\frac{a\mu}{2}+1\right)\ .\end{aligned}$$

Plugging this in, and using that $\mathbb{E}(z_1) \ge \frac{a^2s\mu^2}{2}$ we get

$$a^2\mathbb{E}(N) \le \frac{2}{s\mu^2}u + \frac{2a}{\mu}\left((\widetilde{s}-\frac{s}{4})\frac{a\mu}{2}+1\right)\ .$$

Hence in the limit $a \to 0$ we get

$$a^2\mathbb{E}(N) \le \frac{2}{s\mu^2}\log\frac{1-\beta}{\alpha} \le \frac{2}{s\mu^2}\log\frac{1}{\alpha}\ .$$

We can treat the case $0 < \widetilde{s} < s/4$ in a similar fashion. ∎

Remark 2. *When $\widetilde{s} = s/4$ the argument of the proof breaks down, because of ties when s is divisible by 4. However this is only a technical issue that can be simply circumvented by choosing the bins to be of size $s/2 - 1$, for instance.*

Hence, given α and β we can upper bound the total expected sensing energy used in the search phase. By Lemma 1 we can upper bound the expected sensing energy in the refinement phase given α', β' and $|\boldsymbol{P}|$.

Now we are ready upper bound the expected sensing energy used by the procedure. Note that we have

$$\mathbb{E}_S(|\boldsymbol{P}|) \leq 3ks + \frac{3s}{2} \sum_{i:\ \boldsymbol{P}^{(i)} \not\subseteq S} \alpha \ .$$

Thus choosing $\alpha = \varepsilon/6n$ we have $\mathbb{E}_S(|\boldsymbol{P}|) \leq 3ks + \varepsilon/2 \leq 4ks$. Note that this could again be improved, but this choice will ease the following discussion.

By denoting the part of the sensing matrix A corresponding to the search and refinement phases by A_{search} and $A_{refinement}$ respectively, we have

$$\begin{aligned} \mathbb{E}_S(\|A\|_F) &\leq \mathbb{E}_S(\|A_{search}\|_F) + \mathbb{E}_S\left(\mathbb{E}_S(\|A_{refinement}\|_F \big| |\boldsymbol{P}|)\right) \\ &\leq \frac{16n}{s^2\mu^2} \log\frac{2ks}{\varepsilon} + \frac{4k}{s\mu^2}\log\frac{6n}{\varepsilon} + \frac{2k}{\mu^2}\log\frac{6n}{\varepsilon} \\ &+ \frac{16ks}{\mu^2}\log\frac{4n}{\varepsilon} \ . \end{aligned} \tag{5}$$

When $|S| \ll n$ the first term dominates the bound above. Using this and combining the above with (3) we arrive at the following.

Proposition 2. *Consider the class of k disjoint s-intervals and suppose $\frac{n}{\log 6n} \geq ks^3$. Then the above estimator satisfies* (3) *and* (4) *whenever*

$$\mu \geq \sqrt{\frac{64n}{s^2 m}\log\frac{2ks}{\varepsilon}} \ .$$

Remark 3. *Note that in the entire analysis we made crude choices for the error probabilities and upper bounds to make the discussion and the formulas as smooth as possible. This results in the constants in the above expression to be much larger than necessary. However, our main interest is the scaling of the bounds in terms of the parameters k, s, n, m and ε. The constants can be improved by performing the same analysis with a little more attention to detail. This remark also applies for all the procedures considered later on.*

Remark 4. *The condition on the sparsity in the proposition is needed to ensure that the term corresponding to the search phase in* (5) *becomes dominant. By performing the refinement phase in a more sophisticated way one can relax that condition. For instance using k binary searches to find the left endpoint of the intervals the sparsity condition becomes $\frac{n}{\log 6n} \geq ks^2 \log s$. We expect this to be essentially the best condition one can hope for, as the lower bounds of Section 3.2 show that the first term in* (5) *is unavoidable.*

The bound of Proposition 2 matches the lower bound in Section 3.2, hence in this sparsity regime the procedure above is optimal apart from constants.

3.1.2 Unions of s-stars

Let the components of $\boldsymbol{x}$ be in one-to-one correspondence to edges of a complete graph $G = (V, E)$. Let $e_i \in E$ denote the edge corresponding to component $\boldsymbol{x}_i$, and for a vertex $v \in V$ and edge $e \in E$ let $v \in e$ denote that e is incident with v. We call a support set $S \subset \{1, \ldots, n\}$ an s-star if $|S| = s$ and $\exists v \in V: \ \forall i \in S: \ v \in e_i$. Let $\mathcal{C}$ be the class of unions of k disjoint s-stars. In what follows we use the notation $|V| = p$.

The procedure for support estimation is very similar to that presented for s-intervals. We introduce the procedure when $k = 1$, but the idea can be carried through for larger k. Consider the subsets $\boldsymbol{P}^{(i)}, \ i = 1, \ldots, p$, defined as follows:

$$\boldsymbol{P}^{(i)} = \{j \in E : \ v_i \in e_j\} \ ,$$

that is $\boldsymbol{P}^{(i)}$ contains all the components whose corresponding edges lie on the vertex v_i. These subsets are not a partition of $\{1, \ldots, n\}$ as they are not disjoint. Nonetheless we know that

$$|\boldsymbol{P}^{(i)} \cap S| \in \{0, 1, s\} \qquad \forall i = 1, \ldots, p \ .$$

We can use this to find the approximate location of S. Thus in the search phase we test the hypotheses

$$H_0^{(i)} : \ |\boldsymbol{P}^{(i)} \cap S| = 1 \qquad \text{versus} \qquad H_1^{(i)} : \ |\boldsymbol{P}^{(i)} \cap S| = s \qquad i = 1, \ldots, p \ .$$

In words we test whether vertex v_i is the center of the star or not for $i = 1, \ldots, p$. Note that when vertex v_i is not the center of the star we have $|\boldsymbol{P}^{(i)} \cap S| \in \{0, 1\}$. By specifying $H_0^{(i)}$ as above we ensure that if $|\boldsymbol{P}^{(i)} \cap S| = 0$ both the probability of error and the expected number of steps of the SLRT will be smaller than if $|\boldsymbol{P}^{(i)} \cap S| = 1$, due to the monotonicity of the likelihood ratio.

Again we use independent SLRTs for the tests with common type I and type II error probabilities α, β, where the details will be covered later. Using these tests we can define $\boldsymbol{P}$, the output of the search phase, as the union of those $\boldsymbol{P}^{(i)}$ for which $H_1^{(i)}$ is accepted. With the appropriate choices for α and β we can ensure that with high probability $S \subset \boldsymbol{P}$ and that $|\boldsymbol{P}|$ is small. In fact we would like to accept exactly one $H_1^{(i)}$. Again the right choice for β will ensure $\mathbb{P}_S(S \not\subseteq \boldsymbol{P})$ is small whereas the right choice of α ensures that $|\boldsymbol{P}|$ is small with high probability. In the subsequent refinement phase we estimate S within $\boldsymbol{P}$. We do this using the same procedure as in the unstructured case with error probabilities α', β'. Finally the estimator $\widehat{S}$ will be the collection of those components which were deemed active in the refinement phase.

Now we choose the error probabilities for the tests such that we can ensure (4) for our procedure. We have

$$\begin{aligned}
\mathbb{E}_S\left(|\widehat{S} \triangle S|\right) &\leq \mathbb{E}_S\left(|\widehat{S} \triangle S| \ \middle| S \not\subseteq \boldsymbol{P}\right) \mathbb{P}_S(S \not\subseteq \boldsymbol{P}) + \mathbb{E}_S\left(|\widehat{S} \triangle S| \ \middle| S \subseteq \boldsymbol{P}\right) \\
&\leq \mathbb{E}_S\left(|S \setminus \boldsymbol{P}| + \sum_{i \in \boldsymbol{P}: \ i \notin S} \alpha' + \sum_{i \in \boldsymbol{P}: \ i \in S} \beta' \ \middle| \ S \not\subseteq \boldsymbol{P} \right) \beta \\
&\quad + n\alpha' + s\beta' \ .
\end{aligned}$$

Thus the choices $\beta = \varepsilon/4s$ and $\alpha' = \varepsilon/4n, \beta' = \varepsilon/4s$ suffice. As noted before, the choice of α will influence the size of $\boldsymbol{P}$ and will be discussed later.

To test $H_0^{(i)}$ versus $H_1^{(i)}$ we collect observations using the sensing vector $a\mathbf{1}_{\boldsymbol{P}^{(i)}}$ and perform a SLRT such as the one in Lemma 2. When there is no active component in $\boldsymbol{P}^{(i)}$ the drift of the likelihood-ratio process is smaller than if there was one active component by monotonicity of the likelihood ratio. This results in the test terminating sooner in expectation than it would under $H_0^{(i)}$ and the probability of accepting $H_1^{(i)}$ is also smaller than the type I error probability α.

We continue by upper bounding the expected sensing energy used by the procedure. Again we have results similar to Lemma 2 for the tests carried out in the search phase, and we can use Lemma 1 to bound the energy used in the refinement phase. Hence given $\alpha, \beta, \alpha', \beta'$ and $\boldsymbol{P}$ we can bound the total energy used by the procedure. Also note that

$$\mathbb{E}_S(|\boldsymbol{P}|) \leq p + p \sum_{i:\ \boldsymbol{P}^{(i)} \not\subseteq S} \alpha \ ,$$

thus choosing $\alpha = \varepsilon/2n$ ensures $\mathbb{E}_S(|\boldsymbol{P}|) \leq 2p$. As in the case of s-intervals, these choices are crude but ease the discussion and do not affect the overall performance scaling.

Using the notation A_{search} and $A_{refinement}$ as before we get

$$\begin{aligned}\mathbb{E}_S(\|A\|_F) &\leq \mathbb{E}_S(\|A_{search}\|_F) + \mathbb{E}_S\left(\mathbb{E}_S(\|A_{refinement}\|_F \big| |\boldsymbol{P}|)\right) \\ &\leq \frac{2p(p-1)}{(s-1)^2\mu^2} \log\frac{4s}{\varepsilon} + \frac{2p}{(s-1)^2\mu^2}\log\frac{4n}{\varepsilon} \\ &+ \frac{4p}{\mu^2}\log\frac{4n}{\varepsilon} \ .\end{aligned}$$

When $s \ll n$ the first term dominates the bound. Combining this with (3) we get the following.

Proposition 3. *Consider the class of s-stars and suppose $\frac{\sqrt{n}}{\log 4n} \geq s^2$. Then the above estimator satisfies* (3) *and* (4) *whenever*

$$\mu \geq \sqrt{\frac{12n}{(s-1)^2 m}\log\frac{4s}{\varepsilon}} \ .$$

In Section 3.2 we show that the bound of Proposition 3 is near optimal in this sparsity regime. Also we show there that the sparsity assumption in the proposition above is needed and is not an artifact of our method.

When $k > 1$ (S consists of two or more s-stars) similar arguments hold. When $k \ll s$ it is possible to modify the procedure such that the search phase aims to find the center of the k stars. The modifications include setting $H_0(i):\ |\boldsymbol{P}^{(i)} \cap S| = k$, and slightly changing $\alpha, \beta, \alpha', \beta'$ to account for the fact that there are more than one stars. For instance choosing α, α' to be the same as before and setting $\beta = \beta' = \varepsilon/4ks$ we get the following.

Proposition 4. *Consider the class of k disjoint s-stars and suppose $k < s$ and $\frac{\sqrt{n}}{\log 4n} \geq k(s-k)^2$. Then the modified estimator satisfies* (3) *and* (4) *whenever*

$$\mu \geq \sqrt{\frac{12n}{(s-k)^2 m}\log\frac{4sk}{\varepsilon}} \ .$$

We see is Section 3.2 that the bound above is near the optimal one when k is much smaller than s.

3.1.3 s_r, s_c-submatrices

Let the components of $\boldsymbol{x}$ be in one-to-one correspondence to elements of a matrix M with n_r rows and n_c columns (and let $n = n_r \times n_c$). We call a set $S \subset \{1, \ldots, n\}$ an s_r, s_c-submatrix if the elements $m_i \in M$ corresponding to the components $i \in S$ form an $s_r \times s_c$ submatrix in M. Let $\mathcal{C}$ be the class of all s_r, s_c-submatrices in $\boldsymbol{x}$. Suppose without loss of generality that $s_r \geq s_c$ and recall that the number of non-zero components of $\boldsymbol{x}$ is simply $s = s_r \times s_c$.

One way to estimate S in this case is to first find active columns in the search phase and then focus on one or more active columns in the refinement phase to find the active rows. Let $\boldsymbol{c}^{(i)}$ denote the ith column of $\boldsymbol{x}$, $i = 1, \ldots, n_c$. In order to find the active columns we need to decide between

$$H_0^{c(i)} : \ |\boldsymbol{c}^{(i)} \cap S| = 0 \qquad \text{versus} \qquad H_1^{c(i)} : \ |\boldsymbol{c}^{(i)} \cap S| = s_r \qquad i = 1, \ldots, n_c \ .$$

To do this we perform independent SLRTs with type I and type II error probabilities α and β respectively for every $i = 1, \ldots, n_c$. At the end of the search phase we return $\boldsymbol{P}$, which is the union of columns $c^{(i)}$ for which $H_1^{c(i)}$ was accepted. Choosing α, β appropriately ensures that with high probability $\boldsymbol{P}$ contains all the active columns and only those. In the refinement phase we test if row j of $\boldsymbol{P}$ is active or not using a similar method as above, with error probabilities α', β' for every $j = 1, \ldots, n_r$. In particular the tests are formulated as

$$H_0^{r(j)} : \ |(\boldsymbol{r}^{(j)} \cap \boldsymbol{P}) \cap S| = 0 \qquad \text{versus} \qquad H_1^{r(j)} : \ |(\boldsymbol{r}^{(j)} \cap \boldsymbol{P}) \cap S| = s_c \qquad j = 1, \ldots, n_r \ ,$$

where $\boldsymbol{r}^{(j)}$ denotes the jth row of $\boldsymbol{x}$, $j = 1, \ldots, n_r$. Finally our estimate $\widehat{S}$ are those elements that are in a row and column that were both deemed active.

Now we choose the error probabilities $\alpha, \beta, \alpha', \beta'$. Now we simply have

$$\mathbb{E}_S(|\widehat{S} \triangle S|) \leq n\alpha + s\beta + n\alpha' + s\beta' \ ,$$

as every type I error in the search phase can result in at most n_r errors in $\widehat{S}$ and there can be at most n_c type I errors in the search phase, whereas a type II error can produce at most s_r errors in the end and there are s_c possibilities to make such an error. A similar argument holds for tests in the refinement phase. Hence the choices $\alpha = \alpha' = \varepsilon/4n$ and $\beta = \beta' = \varepsilon/4s$ ensure (4).

We move on to bounding the expected energy used by the procedure. To test the ith hypothesis in the search phase we collect measurements using sensing vector $a\mathbf{1}_{\boldsymbol{c}^{(i)}}$ for all $i = 1, \ldots, n_c$ and perform a SLRT similar to that described in the previous cases. To perform the jth SLRT of the refinement phase we collect measurements of the form $a\mathbf{1}_{r^{(j)} \cap \boldsymbol{P}}$. For these tests we have results identical to Lemmas 2 and 3. Also for the number of columns in $\boldsymbol{P}$ denoted by $\widetilde{n}_c$ we have

$$\mathbb{E}_S(\widetilde{n}_c) \leq s_c + n_c\alpha \leq 2s_c \ .$$

Putting everything together yields

$$\begin{aligned}\mathbb{E}_S(\|A\|_F) &\le \mathbb{E}_S(\|A_{search}\|_F) + \mathbb{E}_S\left(\mathbb{E}_S(\|A_{refinement}\|_F \big| |\boldsymbol{P}|)\right) \\ &\le \frac{2n}{s_r^2\mu^2}\log\frac{4s}{\varepsilon} + \frac{2n_r s_c}{s_r^2\mu^2}\log\frac{4n}{\varepsilon} \\ &+ \frac{4n_r}{s_c\mu^2}\log\frac{4n}{\varepsilon}\ .\end{aligned}$$

When $s \ll n$ the first term dominates the bound above. Combining this with (3) yields the following.

Proposition 5. *Consider the class of s_r, s_c-submatrices and suppose $\frac{n_c}{\log 4n} \ge \frac{s_r^2}{s_c}$. Then the estimator above satisfies* (3) *and* (4) *whenever*

$$\mu \ge \sqrt{\frac{6n}{s_r^2 m}\log\frac{4s}{\varepsilon}}\ .$$

Note that the condition on the sparsity in the proposition above is not very strict. Consider square submatrices within square matrices so that we have $n_r = n_c = \sqrt{n}$ and $s_r = s_c = \sqrt{s}$. Then the condition becomes $\frac{\sqrt{n}}{\log 4n} > \sqrt{s}$, which would be automatically fulfilled if there was no logarithmic term on the left. We see in Section 3.2 that in some sparsity regimes the bound above matches the lower bounds we derive, thus in those regimes this procedure is near optimal. However, in what follows we slightly modify the procedure above to have better performance for submatrices that are more sparse than the ones required in the proposition above. This combined with the results of Section 3.2 shows that the best performance we can hope for depends on the sparsity in a non-trivial manner in the case of submatrices.

Note that, for the refinement phase of the above procedure, it does suffice to find a single active column in the search phase, as accurately estimating components within *any* active column will yield all the active rows and similarly estimating components within *any* active row yields us the active columns. This motivates the following modification of the above procedure: return a single active column in the search phase, then focus on that column to find the active rows and finally focus on one active row to find the active columns. To do this we retain most of the algorithm choices done in the earlier approach, but choose a different α and β.

Ideally we would like to accept $H_1^{c(i)}$ for exactly one active column, so our choices for α, β will be made accordingly. We begin the refinement phase by randomly choosing a column from the ones that were deemed active and locate the active components within that column, using the same procedure as in the unstructured case. This gives us the active rows. Finally we choose a row deemed active, and find all the active components within that row to find the active columns. Throughout the refinement phase we set type I and type II error probabilities to be α', β'. With the right choices for the error probabilities, this procedure outperforms the previous one in certain sparsity regimes.

First we need to choose the error probabilities for the tests. We can write

$$\begin{aligned}\mathbb{E}_S(|\widehat{S}\triangle S|) &\le 2s\mathbb{P}_S(\boldsymbol{P}=\emptyset) + \left(2n\alpha' + 2s\right)\mathbb{P}_S(\exists \boldsymbol{c}^{(i)} \subset \boldsymbol{P}:\ \boldsymbol{c}^{(i)} \cap S = \emptyset) + \left(2n\alpha' + 2s\beta'\right) \\ &\le 2s\beta^{s_c} + (2n\alpha' + 2s)n_c\alpha + (2n\alpha' + 2s\beta')\ .\end{aligned}$$

Thus the conservative choices $\alpha = \varepsilon/16n^2, \beta = \sqrt[s_c]{\varepsilon/8s}, \alpha' = \varepsilon/8n, \beta' = \varepsilon/8s$ ensure (4).

Now we can move on to calculate the expected sensing energy used by the procedure. The same way as before we have

$$\begin{aligned}\mathbb{E}_S(\|A\|_F) &\leq \mathbb{E}_S(\|A_{search}\|_F) + \mathbb{E}_S(\|A_{refinement}\|_F)\\ &\leq \frac{2n}{s_c s_r^2 \mu^2}\log\frac{8s}{\varepsilon} + \frac{4n_r s_c}{s_c s_r^2 \mu^2}\log\frac{4n}{\varepsilon}\\ &+ \frac{4\max\{n_r, n_c\}}{\mu^2}\log\frac{4n}{\varepsilon}\ .\end{aligned}$$

Combining the above with (3) and using that when $s \ll n$ the first term dominates we arrive to the following.

Proposition 6. *Consider the class of s_r, s_c-submatrices and suppose $\frac{\min\{n_r,n_c\}}{2\log 4n} \geq s_c s_r^2$. Then the estimator above satisfies* (3) *and* (4) *whenever*

$$\mu \geq \sqrt{\frac{6n}{s_c s_r^2 m}\log\frac{8s}{\varepsilon}}\ .$$

The condition on the sparsity in the proposition above is stronger than that in Proposition 5. On the other hand the bound for μ is smaller. This shows that in sparser regimes it is indeed possible to outperform the procedure of Proposition 5, hinting that the sparsity regime non-trivially influences the best possible performance of adaptive support recovery procedures in the case of sub-matrices. For instance considering square matrices when $n_r = n_c = \sqrt{n}$ and $s_r = s_c = \sqrt{s}$, the condition above reads $\frac{\sqrt{n}}{2\log 4n} > \sqrt{s^3}$ which is slightly stronger than that of Proposition 5.

3.2 Lower bounds

We turn our attention to the fundamental limits of recovering the support of structured sparse signals using compressive measurements. We consider both the non-adaptive sensing and adaptive sensing settings.

3.2.1 Non-Adaptive Sensing

First we consider the non-adaptive compressive sensing setting. Comparing these lower bounds with the performance bounds of the previous section illustrates the gains adaptivity provides in the various cases. We do not make any claim on whether these lower bounds are tight or not, as these serve mostly for comparison of adaptive sensing to non-adaptive sensing.

In the non-adaptive sensing setting we need to define sensing actions before any measurements are taken. That means the sensing matrix A is specified prior to taking any observations. This does not exclude the possibility that A is random, but it has to be generated before any observations are made.

All the bounds presented here are based on Proposition 2.3 in in [20], which states

Lemma 4 (Proposition 2.3 of [20])**.** *Let $\mathbb{P}_0, \ldots, \mathbb{P}_M$ be probability measures on $(\mathcal{X}, \mathcal{A})$ and let $\Psi : \mathcal{X} \to \{0, \ldots, M\}$ be any $\mathcal{A}$-measurable function. If*

$$\frac{1}{M}\sum_{j=1}^{M} D(\mathbb{P}_j \| \mathbb{P}_0) \leq a$$

then

$$\max_{j=0,\ldots,M} \mathbb{P}_j(\Psi \neq j) \geq \sup_{0<\tau<1} \left(\frac{\tau M}{1+\tau M} \left(1 + \frac{a + \sqrt{a/2}}{\log \tau} \right) \right) .$$

We can use this to get lower bounds in the following way. Let $\mathbb{P}_0, \ldots, \mathbb{P}_M$ be the probability measures induced by sampling $\boldsymbol{x}$ with sensing matrix A, when the support set is $S_0, \ldots, S_M$ respectively, where $S_i \in \mathcal{C}$. Now note that

$$\begin{aligned}
D(\mathbb{P}_j \| \mathbb{P}_0) &= \mathbb{E}_0 \left(\sum_k \log \frac{\mathrm{d}\mathbb{P}_0(Y_k|A_k)}{\mathrm{d}\mathbb{P}_j(Y_k|A_k)} \right) \\
&= \sum_k \mathbb{E} \left(\mathbb{E}_0 \left(-\frac{1}{2} \left((Y_k - \mu < A_k, \mathbf{1}_{S_0} >)^2 - (Y_k - \mu < A_k, \mathbf{1}_{S_j} >)^2 \right| A \right) \right) \right) \\
&= \sum_k \mathbb{E} \left(\mathbb{E}_0 \left(\frac{1}{2} \left(\mu^2 (< A_k, \mathbf{1}_{S_j} >^2 - < A_k, \mathbf{1}_{S_0} >^2) - 2\mu Y_k < A_k, \mathbf{1}_{S_j} - \mathbf{1}_{S_0} > \right) \right| A \right) \right) \\
&= \frac{\mu^2}{2} \mathbb{E} \left(\sum_k \left(< A_k, \mathbf{1}_{S_j} >^2 + < A_k, \mathbf{1}_{S_0} >^2 - 2 < A_k, \mathbf{1}_{S_j} >< A_k, \mathbf{1}_{S_0} > \right) \right) \\
&= \frac{\mu^2}{2} \mathbb{E} \left(\sum_k \left(\sum_{i \in S_0 \triangle S_j} A_{k,i} \right)^2 \right) \\
&\leq \frac{\mu^2}{2} \mathbb{E} \left(\sum_k |S_0 \triangle S_j| \sum_{i \in S_0 \triangle S_j} A_{k,i}^2 \right) \\
&= \frac{\mu^2}{2} |S_0 \triangle S_j| \sum_{i \in S_0 \triangle S_j} a_i^2 \ , \qquad (6)
\end{aligned}$$

where $A_{k,j}$ is the (k, j)th element of the sensing matrix A, a_i^2 denotes $\mathbb{E}(\sum_k A_{k,i}^2)$, and in the second to last step we use Jensen's inequality.

Now consider the right side of Lemma 4 and set $\tau = 1/M$. To make the bound more transparent suppose $1 \leq (1 - 2\varepsilon) \log M$, which is essentially always satisfied if M is large enough and $\varepsilon \in (0, 1/2)$. This way we arrive to the inequality

$$2a \geq (1 - 2\varepsilon) \log M \ . \qquad (7)$$

Choosing the sets $S_0, \ldots, S_M$ and using inequality (6) to bound the average KL distance, we can use the above inequality to get lower bounds for μ. These choices will be specific to the classes we are considering.

Remark 5. *In the following statements we require n to be divisible by s. This condition is merely technical, and can be easily dropped at the expense of a cumbersome presentation.*

Proposition 7 (s-sets)**.** *Let $\mathcal{C}$ be the class of s-sets and suppose n/s is an integer. If there is a non-adaptive estimator $\widehat{S}$ that satisfies* (3) *and $\mathbb{P}_S(\widehat{S} \neq S) \leq \varepsilon \ \forall S \in \mathcal{C}$ then*

$$\mu \geq \sqrt{(1-2\varepsilon)\frac{n}{4m}\log(n-s)} \ .$$

Proof. Let $S_0 \in \mathcal{C}$ be arbitrary. Partition $\{1,\ldots,n\}$ into s bins of equal size denoted by $\boldsymbol{P}^{(1)},\ldots,\boldsymbol{P}^{(s)}$ such that each bin contains exactly one element of S_0. Let $s_i = S_0 \cap \boldsymbol{P}^{(i)}$, $i = 1,\ldots,s$. Now consider the sets $S_1,\ldots,S_M$ that we get by modifying exactly one element of S_0 in the following way: pick one element of S_0 denoted by s_i and swap it with some other element in $\boldsymbol{P}^{(i)}$ thus changing the position of the active component within $\boldsymbol{P}^{(i)}$. We can generate $M = n-s$ sets in the previous manner. From (6) we have that

$$\frac{1}{M}\sum_{j=1}^{M} D(\mathbb{P}_j\|\mathbb{P}_0) \leq \frac{1}{M}\mu^2 \sum_{j=1}^{M}\sum_{i\in S_0 \triangle S_j} a_i^2 = \frac{1}{n-s}\mu^2 \left(\sum_{i=1}^{n} a_i^2 + \frac{n-2s}{s}\sum_{i\in S_0} a_i^2\right) \ .$$

Now note that by the total energy constraint (3) we have

$$\sum_{i=1}^{n} a_i^2 \leq m \ .$$

Also note that given A we can always choose S_0 to be the one that is the most difficult to distinguish from the other sets $S_1,\ldots,S_M$. That is we have to solve

$$\max_{A:\ \|A\|_F \leq m} \ \min_{S\in\mathcal{C}} \sum_{i\in S_0} a_i^2 \ .$$

This implies $\sum_{i\in S_0} a_i^2 \leq sm/n$. Combining what we have yields

$$\frac{1}{M}\sum_{j=1}^{M} D(\mathbb{P}_j\|\mathbb{P}_0) \leq \frac{1}{n-s}\left(1 + \frac{n-2s}{n}\right) m\mu^2 \leq \frac{2m}{n}\mu^2 \ .$$

Using this with (7) concludes the proof. □

Proposition 8 (Unions of s-intervals)**.** *Let $\mathcal{C}$ be the class of unions of k disjoint s-intervals and suppose n/s is an integer. If there is a non-adaptive estimator $\widehat{S}$ that satisfies* (3) *and $\mathbb{P}_S(\widehat{S} \neq S) \leq \varepsilon \ \forall S \in \mathcal{C}$ then*

$$\mu \geq \sqrt{(1-2\varepsilon)\frac{n-(k-1)s}{4s^2 m}\log(\frac{n}{s}-k)} \ .$$

Proof. Partition $\{1,\dots,n\}$ into consecutive intervals of size s denoted by $S^{(1)},\dots,S^{(n/s)}$. Now consider the subclass whose elements are unions of the first $k-1$ intervals $S^{(1)},\dots,S^{(k-1)}$ and some other interval $S^{(i)}$. Formally, $\mathcal{C}' = \{S \in \mathcal{C} : \ S = S^{(i)} \cup \left(\bigcup_{j=1}^{k-1} S^{(j)}\right),\ i = k,\dots,n/s\}$. This way we effectively reduced this problem to finding one interval in a slightly smaller vector. Let $S_0 \in \mathcal{C}'$ be arbitrary and let $S_1,\dots,S_M$ be all the other elements of $\mathcal{C}'$, so $M = n/s - k$. Let $\widetilde{S}_0 = S_0 \setminus \cup_{j=1}^{k-1} S^{(j)}$. From (6) we have

$$\frac{1}{M}\sum_{j=1}^{M} D(\mathbb{P}_j \| \mathbb{P}_0) \leq s\mu^2 \frac{1}{M}\sum_{j=1}^{M}\sum_{i\in S_0 \triangle S_j} a_i^2 = \frac{s^2\mu^2}{n-ks}\left(\sum_{i=(k-1)s+1}^{n} a_i^2 + \frac{n-(k+1)s}{s}\sum_{i\in\widetilde{S}_0} a_i^2\right) .$$

Again, from (3) and the fact that we can choose $S_0 \in \mathcal{C}'$ after the sensing strategy has been determined we have

$$\frac{1}{M}\sum_{j=1}^{M} D(\mathbb{P}_j \| \mathbb{P}_0) \leq \frac{1}{n-ks}\left(1 + \frac{n-(k+1)s}{n-(k-1)s}\right) s^2 m\mu^2 \leq \frac{2s^2 m}{n-(k-1)s}\mu^2 .$$

Using this with (7) concludes the proof. □

Proposition 9 (Unions of s-stars)**.** *Let $\mathcal{C}$ be the class of s-stars and suppose p/s is an integer. If there is a non-adaptive estimator $\widehat{S}$ that satisfies* (3) *and $\mathbb{P}_S(\widehat{S} \neq S) \leq \varepsilon\ \forall S \in \mathcal{C}$ then*

$$\mu \geq \sqrt{(1-2\varepsilon)\frac{n}{2m}\log(\sqrt{2n} - s - 1)} .$$

Proof. Consider the $p-1$ edges of the complete graph of p vertices which share a common vertex j. Denote this set of edges by E_j. The s-stars whose center is vertex j form a class of s-sets on E_j. So we can do the same construction on this set of edges as in Proposition 7 to get

$$\frac{1}{M}\sum_{j=1}^{M} D(\mathbb{P}_j \| \mathbb{P}_0) \leq \frac{1}{p-1-s}\mu^2\left(\sum_{i\in E_j} a_i^2 + \frac{p-1-2s}{s}\sum_{i\in S_0} a_i^2\right) .$$

Now note that we can choose any star to be S_0 which implies $\sum_{i\in S_0} a_i^2 \leq sm/n$ and $\sum_{i\in E_j} a_i^2 \leq (p-1)m/n$ yielding

$$\frac{1}{M}\sum_{j=1}^{M} D(\mathbb{P}_j \| \mathbb{P}_0) \leq \frac{2m}{n}\mu^2 .$$

The statement now follows from (7) and that $p > \sqrt{2n}$. □

Considering unions of k disjoint s-stars we can get a similar lower bound by considering a subclass where $k-1$ of the s-stars are fixed and only one can change, reducing the problem to finding one s-star.

Proposition 10 (s-submatrices)**.** *Let $\mathcal{C}$ be the class of s-submatrices of a fixed size $s_c \times s_r$, and suppose both n_c/s_c and n_r/s_r are integers. If there is a non-adaptive estimator $\widehat{S}$ that satisfies* (3) *and $\mathbb{P}_S(\widehat{S} \neq S) \leq \varepsilon \ \forall S \in \mathcal{C}$ then*

$$\mu \geq \sqrt{(1-2\varepsilon)\frac{n}{4m}\max\left\{\frac{1}{s_r}\frac{n_c - s_c}{n_c}, \frac{1}{s_c}\frac{n_r - s_r}{n_r}\right\}\log\left(\max\{n_r - s_r, n_c - s_c\}\right)} \ .$$

Proof. Let $S_0 \in \mathcal{C}$ be arbitrary. Denote the indexes of the rows of S_0 by $r_1, \ldots, r_{s_r}$, and let $S_0^{(j)}$ denote the jth row of S_0. Consider a partition of the indexes $\{1, \ldots, n_r\}$ into $\boldsymbol{r}^{(1)}, \ldots, \boldsymbol{r}^{(s_r)}$ such that all of the are of the same size and $\boldsymbol{r}^{(j)}$ contains exactly one active row indexed by r_j for every $j = 1, \ldots, r_{s_r}$.

Now let $S_1, \ldots, S_M$ be elements of $\mathcal{C}$ that we get by replacing exactly one row index of S_0 such that if we modify r_j, then the new row index is in $\boldsymbol{r}^{(j)}$. There are $n_r - s_r$ such submatrices. The same way as for the s-sets we get

$$\frac{1}{M}\sum_{j=1}^{M} D(\mathbb{P}_j \| \mathbb{P}_0) \leq \frac{1}{n_r - s_r}\mu^2 \left(\sum_{(i,l):\ l \in C_0} a_{(i,l)}^2 + \frac{n_r - 2s_r}{s_r} \sum_{(i,l) \in S_0} a_{(i,l)}^2 \right) \ .$$

Again, the fact that we can choose an arbitrary $S_0 \in \mathcal{C}$ after the sensing strategy has been fixed results in the upper bound

$$\frac{1}{M}\sum_{j=1}^{M} D(\mathbb{P}_j \| \mathbb{P}_0) \leq \frac{2s_c m}{n}\frac{n_r}{n_r - s_r}\mu^2 \ .$$

Plugging this into (7) and rearranging gives a lower bound. Repeating the same arguments for columns concludes the proof. □

3.2.2 Adaptive Sensing

Here we provide lower bounds considering the adaptive sensing framework. Comparing these bounds with the performance bounds of Section 3.1 shows the near optimality of the procedures presented there.

s-sets

Adaptive sensing lower bounds for unstructured classes were proved in [13]. There lower bounds are derived by slightly broadening the class, which we state here for convenience. Note that the fact that the following lower bound is valid for a slightly larger class than the class of s-sets does not cause a problem, see Remarks 1 and 6. Let $\mathcal{C}_s$ denote the class of s-sets. We have the following.

Proposition 11. *Let $\mathcal{C} = \mathcal{C}_s \cup \mathcal{C}_{s-1}$, and suppose there exists an estimator $\widehat{S}$ that satisfies* (3) *and* (4)*. Then we have*

$$\mu \geq \sqrt{\frac{2(n-s+1)}{m}\left(\log\frac{s}{2\varepsilon} + \log\frac{n-s+1}{n+1}\right)} \ .$$

Remark 6. *Note that the bound above holds for estimators for sets with sparsity s or $s-1$. The procedure presented in Section 3.1 works for this class of sets without any modifications. Later on for the structured classes we rely on the proposition above to derive lower bounds, hence a similar comment applies in those cases as well.*

s-intervals and unions of s-intervals

For s-intervals we have multiple ways of deriving lower bounds, just as in the case of coordinate wise sampling studied in [19]. First we consider $\mathbb{P}_S(\widehat{S} \neq S)$ as the error metric. This is more forgiving than $\mathbb{E}_S(|\widehat{S} \triangle S|)$, hence lower bounds with the former metric in mind apply as lower bounds with the latter metric as well (since $\mathbb{E}_S(|\widehat{S} \triangle S|) \geq \mathbb{P}_S(\widehat{S} \neq S)$). The following result is analogous to the lower bound in [6], and the proof is included here for the sake of clarity.

Proposition 12. *Let $\mathcal{C}$ be the class of s-intervals and suppose there is an estimator $\widehat{S}$ satisfying* (3) *and $\max\limits_{S \in \mathcal{C}} \mathbb{P}_S(\widehat{S} \neq S) \leq \varepsilon$. Furthermore suppse n/s is an integer. Then*

$$\mu \geq (1-\varepsilon)\sqrt{\frac{n}{2s^2 m}} \ .$$

Proof. Consider the subclass of consecutive disjoint s-intervals

$$\{\{1, \ldots, s\}, \{s+1, \ldots, 2s\}, \ldots \{n-s+1, \ldots, n\}\} \ .$$

Partition this subclass into two subclasses of equal size denoted by $\mathcal{C}_1$ and $\mathcal{C}_2$. Let π_i denote the uniform distribution on the subclass $\mathcal{C}_i$ for $i = 1, 2$, and consider the two hypotheses $H_i : \ S \sim \pi_i, \ i = 1, 2$. If there exists an estimator $\widehat{S}$ satisfying (3), then there exists a test function $\Phi : D \to \{1, 2\}$ such that $\mathbb{P}_1(\Phi(D) = 2) + \mathbb{P}_2(\Phi(D) = 1) \leq \varepsilon$, where $\mathbb{P}_i$ denotes the distribution of $D = \{Y_j, A_j\}_{j=1,2,\ldots}$ when H_i is true, $i = 1, 2$. Let $\mathbb{P}_0$ denote the distribution of D when in fact $S = \emptyset$. We have

$$\begin{aligned}
\varepsilon &\geq \mathbb{P}_1(\Phi(D) = 2) + \mathbb{P}_2(\Phi(D) = 1) \geq 1 - TV(\mathbb{P}_1, \mathbb{P}_2) \\
&\geq 1 - (TV(\mathbb{P}_0, \mathbb{P}_1) + TV(\mathbb{P}_0, \mathbb{P}_2)) = 1 - 2TV(\mathbb{P}_0, \mathbb{P}_1) \\
&\geq 1 - \sqrt{2KL(\mathbb{P}_0, \mathbb{P}_1)} \ ,
\end{aligned}$$

where $TV(.,.)$ denotes the total variation distance and $KL(.,.)$ denotes the Kullback-Leibler divergence of two distributions. Now the goal is to upper bound $KL(\mathbb{P}_0, \mathbb{P}_1)$. Let Y denote the observations $Y_1, Y_2, \ldots$, and let $\mathbb{P}_S$ denote the distribution of Y for a fixed support S.

We have

$$
\begin{aligned}
KL(\mathbb{P}_0,\mathbb{P}_1) &= \mathbb{E}_0\left(\log\frac{\mathrm{d}\mathbb{P}_0(D)}{\mathrm{d}\mathbb{P}_1(D)}\right) = \mathbb{E}_0\left(\log\frac{\mathrm{d}\mathbb{P}_0(Y)}{\mathrm{d}\mathbb{P}_1(Y)}\right)\\
&= -\mathbb{E}_0\left(\log\frac{\mathrm{d}\mathbb{P}_1(Y)}{\mathrm{d}\mathbb{P}_0(Y)}\right) = -\mathbb{E}_0\left(\log\frac{\mathbb{E}_{S\sim\pi_1}\left(\mathrm{d}\mathbb{P}_S(Y)\right)}{\mathrm{d}\mathbb{P}_0(Y)}\right)\\
&\le -\mathbb{E}_0\left(\mathbb{E}_{S\sim\pi_1}\left(\log\frac{\mathrm{d}\mathbb{P}_S(Y)}{\mathrm{d}\mathbb{P}_0(Y)}\right)\right)\\
&= -\mathbb{E}_0\left(\mathbb{E}_{S\sim\pi_1}\left(-\frac{1}{2}\sum_{j=1}^{\infty}\left((Y_j-\mu < A_j,\mathbf{1}_S >)^2 - Y_j^2\right)\right)\right)\\
&= \frac{1}{2}\mathbb{E}_0\left(\mathbb{E}_{S\sim\pi_1}\left(\sum_{j=1}^{\infty}\left(\mu^2 < A_j,\mathbf{1}_S >^2 - 2\mu < A_j,\mathbf{1}_S > Y_j\right)\right)\right)\\
&= \frac{\mu^2}{2}\mathbb{E}_0\left(\mathbb{E}_{S\sim\pi_1}\left(\sum_{j=1}^{\infty} A_j^T\mathbf{1}_S\mathbf{1}_S^T A_j\right)\right)\\
&= \frac{\mu^2}{2}\mathbb{E}_0\left(\sum_{j=1}^{\infty} A_j^T\mathbb{E}_{S\sim\pi_1}\left(\mathbf{1}_S\mathbf{1}_S^T\right) A_j\right) .
\end{aligned}
$$

Now $\mathbb{E}_{S\sim\pi_1}\left(\mathbf{1}_S\mathbf{1}_S^T\right) = \frac{2s}{n}I'$ where $I'\in\mathbb{R}^{n\times n}$ is block diagonal with $n/2s$ blocks of size $s\times s$ consisting of all ones, and the rest of the matrix consists of zeros. Thus we can continue as

$$
\begin{aligned}
KL(\mathbb{P}_0,\mathbb{P}_1) &\le \frac{\mu^2}{2}\mathbb{E}_0\left(\sum_{j=1}^{\infty} A_j^T\mathbb{E}_{S\sim\pi_1}\left(\mathbf{1}_S\mathbf{1}_S^T\right) A_j\right)\\
&= \mu^2\frac{s}{n}\mathbb{E}_0\left(\sum_{j=1}^{\infty} A_j^T I' A_j\right) = \mu^2\frac{s}{n}\mathbb{E}_0\left(\sum_{j=1}^{\infty} < A_j, I'A_j >\right)\\
&\le \mu^2\frac{s}{n}\mathbb{E}_0\left(\sum_{j=1}^{\infty} |< A_j, I'A_j >|\right)\\
&\le \mu^2\frac{s}{n}\mathbb{E}_0\left(\sum_{j=1}^{\infty} \|A_j\|_2\|I'A_j\|_2\right)\\
&\le \mu^2\frac{s}{n}\mathbb{E}_0\left(\sum_{j=1}^{\infty} \|A_j\|_2^2\|I'\|_2\right)\\
&\le \mu^2\frac{ms^2}{n} ,
\end{aligned}
$$

where the last step follows from $\|I'\|_2 \le s$ and (3). Thus we arrive at the inequality

$$
\varepsilon \ge 1 - \sqrt{2\mu^2\frac{ms^2}{n}} ,
$$

from which the statement follows. □

In the previous bound the dependence on ϵ is clearly loose. When considering the Hamming distance as the error metric, we can also get lower bounds by slightly broadening the class. We cover this by considering the case of unions of k disjoint s-intervals, which as a special case contains the class of s-intervals when $k = 1$. We broaden this class by adding unions of $k - 1$ disjoint s-intervals as well.

Proposition 13. *Let $\mathcal{C}$ be the class of unions of k or $k-1$ disjoint s-intervals with $k > 0$ fixed, and suppose n/s is an integer. Suppose there is an estimator satisfying* (3) *and* $\max_{S\in\mathcal{C}} \mathbb{E}_S\big(d(\widehat{S}, S)\big) \le \varepsilon$. *Then*

$$\mu \ge \sqrt{\frac{2\big(n - s(k-1)\big)}{s^2 m}\left(\log\frac{ks}{8\varepsilon} + \log\frac{n - s(k-1)}{n+s}\right)} \ .$$

Proof. Partition $\{1, \dots, n\}$ into consecutive disjoint s-intervals denoted by $S^{(1)}, \dots, S^{(n/s)}$, that is $S^{(d)} = \{(d-1)s + 1, \dots, ds\}$, and consider the subclass $\mathcal{C}'$ of $\mathcal{C}$ consisting of all the sets in $\mathcal{C}$ that can be written in the form $\cup\, S^{(d)}$. This subclass is similar to a general sparse class of sparsity k or $k-1$ with the intervals S_d playing the role of the components. This is exactly what we wish to formalize, and then use Proposition 11.

Clearly $\max_{S\in\mathcal{C}'} \mathbb{E}_S\big(d(\widehat{S}, S)\big) \le \varepsilon$. Using $\widehat{S}$ we can construct an estimator $\widetilde{S}$ which only takes values of the form $\cup\, S^{(d)}$, and has the property $\max_{S\in\mathcal{C}'} \mathbb{E}_S\big(d(\widetilde{S}, S)\big) \le 4\varepsilon$. For instance let $\widetilde{S}$ be such that for every $d = 1, \dots, n/s:\ S^{(d)} \subset \widetilde{S}$ if and only if $|\widehat{S} \cap S^{(d)}| \ge s/2$. The expected Hamming-distance for such estimators can be written as

$$\mathbb{E}_S\left(d(\widetilde{S}, S)\right) = s\sum_{d=1}^{n/s} \mathbb{P}_S\left(\mathbf{1}\{S^{(d)} \subset \widetilde{S}\} \ne \mathbf{1}\{S^{(d)} \subset S\}\right) \ .$$

The measurements $Y_j, j = 1, 2, \dots$ can be written in the following form

$$Y_j =< A_j, \boldsymbol{x} > + W_j = \mu \sum_{i\in S} a_{i,j} + W_j = s\mu \sum_{S^{(d)}\in S} \frac{1}{s} \sum_{i\in S^{(d)}} a_{i,j} + W_j \ .$$

Also from Jensen's inequality we have

$$\sum_{d=1}^{n/s}\left(\frac{1}{s}\sum_{i\in S^{(d)}} a_{i,j}\right)^2 \le \sum_{d=1}^{n/s}\frac{1}{s}\sum_{i\in S^{(d)}} a_{i,j}^2 = \frac{1}{s}\sum_{d=1}^{n/s}\sum_{i\in S^{(d)}} a_{i,j}^2 \le \frac{m}{s} \ .$$

Therefore the problem can be viewed as estimating a general sparse support set. The sparsity is either k or $k-1$, the length of the vector is n/s, the signal strength is $s\mu$, the total sensing budget is m/s and the desired accuracy in expected Hamming-distance is $4\varepsilon/s$. From Proposition 11 we have

$$s\mu \ge \sqrt{\frac{2(n/s - k + 1)}{m/s}\left(\log\frac{ks}{8\varepsilon} + \log\frac{n/s - k + 1}{n/s + 1}\right)} \ ,$$

which concludes the proof. □

s-stars and unions of s-stars

For these classes exactly the same arguments follow as were used for s-intervals and unions of s-intervals. The only thing that needs to be altered is that instead of disjoint s-intervals we use disjoint s-stars. The difference this makes is that whereas before the new problem dimension became n/s, since the entire signal vector could be covered by disjoint intervals, the same can not be said when considering s-stars.

Let $N(p, s)$ denote the number of disjoint s-stars that can be packed in a complete graph with p vertices. We can easily check that the following inequality holds (see Lemma 2 in [19])

$$N(p, s) \geq \frac{p(p-1-s)}{2s} .$$

The left hand side is approximately n/s when the signal is sparse, thus essentially the same results hold as in the case of unions of intervals. Thus the analogue of Proposition 12 for s-stars is the following.

Proposition 14. *Let $\mathcal{C}$ be the class of s-stars and suppose there is an estimator $\widehat{S}$ satisfying* (3) *and $\max_{S \in \mathcal{C}} \mathbb{P}_S(\widehat{S} \neq S) \leq \varepsilon$. Then*

$$\mu \geq (1-\varepsilon)\sqrt{\frac{N(p, s)}{2sm}} .$$

Remark 7. *When $s \ll n$ the bound above scales as $(1-\varepsilon)\sqrt{\frac{n}{s^2 m}}$.*

We also have an analogue of Proposition 13 for the case of multiple stars.

Proposition 15. *Let $\mathcal{C}$ be the class of unions of k or $k-1$ disjoint s-stars. Suppose there is an estimator satisfying* (3) *and $\max_{S \in \mathcal{C}} \mathbb{E}_S\big(d(\widehat{S}, S)\big) \leq \varepsilon$. Then*

$$\mu \geq \frac{1}{s}\sqrt{\frac{2\big(N(p, s)-k+1\big)}{m/s}\left(\log\frac{ks}{8\varepsilon} + \log\frac{N(p, s)-k+1}{N(p, s)+1}\right)} .$$

Remark 8. *When $s \ll n$ the bound above scales as $\sqrt{\frac{n}{s^2 m}\log\frac{ks}{\varepsilon}}$.*

We also present another simple lower bound that illustrates that the assumption on the sparsity in Proposition 3 requiring approximately that $s^4 \leq n$ is needed and is not only an artifact of our method.

Consider a setting where the support set is a star of size s or $s-1$. Now consider the sub-problem of estimating the support of such a star when the center of the star is given by an oracle. This is an unstructured problem on a vector of size $p-1$. Hence we can directly apply Proposition 11 to get the following result.

Proposition 16. *Let $\mathcal{C}$ be the class of stars with sparsity s and $s-1$ and suppose there is an estimator $\widehat{S}$ satisfying* (3) *and* (4)*. Then*

$$\mu \geq \sqrt{\frac{2(p-s)}{m}\left(\log\frac{s}{2\varepsilon} + \log\frac{p-s}{p}\right)} .$$

Remark 9. *When $s \ll n$ the bound above scales as $\sqrt{\frac{\sqrt{n}}{m}\log\frac{s}{\varepsilon}}$.*

Combining the results of Propositions 15 and 16 shows that considering s-stars the scaling of the signal strength needs to be at least

$$\max\left\{\frac{n}{s^2 m}\log\frac{s}{\varepsilon}, \frac{\sqrt{n}}{m}\log\frac{s}{\varepsilon}\right\} .$$

The first term in the maximum above dominates the second when $s^4 \leq n$. This shows that the performance of Proposition 3 can only be achieved in that sparsity regime.

Remark 10. *Note that the setting of the proposition above is slightly different than the one considered in Section 3.1.2. However, we present this result here merely to make a remark on the conditions in Proposition 3 and it only serves an illustrative purpose. Furthermore the procedure presented in Section 3.1.2 can be easily modified to handle classes considered in the above proposition and have similar performance guarantees to Proposition 3.*

s_r, s_c-submatrices

The case of submatrices has been studied in [6], where the authors consider block-structured activations in matrices. They provide a lower bound akin to that of Proposition 12 and a near optimal procedure. Our setting is more general as we consider arbitrary sub-matrices of a given dimension. Nonetheless the same type of lower bound holds in this case as well.

Proposition 17. *Let $\mathcal{C}$ be the class of s_r, s_c-submatrices, and for sake of simplicity assume that both n_r/s_r and n_c/s_c are integers. Suppose there is an estimator satisfying* (3) *and* $\max_{S\in\mathcal{C}} \mathbb{E}_S\big(d(\widehat{S}, S)\big) \leq \varepsilon$. *Then*

$$\mu \geq (1-\varepsilon)\sqrt{\frac{n}{2s^2 m}} .$$

Proof. Since both n_r/s_r and n_c/s_c are integers the proof goes the same way as that of Proposition 12. ☐

However, our procedures do not reach this lower bound, hence the question arises whether the lower bound above is loose or the procedures are suboptimal? We partially answer this question by presenting another simple lower bound with which we illustrate that in certain sparsity regimes the procedure of Proposition 5 is indeed optimal. Consider the class containing all $s_r \times s_c$ and $s_r \times (s_c - 1)$ submatrices, and consider the sub-problem of estimating the support when the active columns are given. This is a problem of estimating s_c or $s_c - 1$ disjoint s_r-intervals in a signal of size $s_r \cdot n_c$. Note that the procedure of Proposition 5 can handle such classes without any modifications. Now we can directly apply Proposition 13 to get the following.

Proposition 18. *Let $\mathcal{C}$ be the class containing all submatrices of size $s_r \times s_c$ and $s_r \times (s_c - 1)$. Suppose there is an estimator $\widehat{S}$ satisfying* (3) *and* (4). *Then*

$$\mu \geq \sqrt{\frac{2(n_c - s_c + 1)}{s_r m}\left(\log\frac{s}{8\varepsilon} + \log\frac{n_c - s_c + 1}{n_c + 1}\right)} .$$

When $s_r \approx n_r$ (for instance we have linear sparsity *in the rows*: $s_r = cn_r$ with some $c \in (0,1]$) the performance bound of Proposition 5 becomes essentially identical to the lower bound above. This shows that in certain regimes that procedure is optimal. Note that the condition on the number of active rows does not determine the sparsity of the signal, as there is no requirement on the number of active columns for the results to hold. Also note that by Proposition 6 in certain regimes it is possible to outperform the procedure of Proposition 5 indicating that the gains one can hope for in the case of submatrices depends on the interplay between the dimensions of the problem n_r, n_c, s_r, s_c. On a final note if we assume that the support set is such that either the active rows of active columns (but not necessary both) are consecutive then one can simply modify the procedure presented in Section 3.1.3 to even reach the lower bound of Proposition 17. The exact performance characterization of the case of submatrices with arbitrary dimensions remains an interesting open problem, which is not addressed in this article.

4 Sample complexity

In the preceding sections we presented near optimal procedures for structured support recovery using adaptive compressive sensing. Those procedures provided an insight into how one can use the structure of the support sets to achieve performance gains, but we paid no regard to the number of measurements that are collected. However an important aspect of compressive sensing is the possibility to perform estimation using only a small number of observations. Therefore we now present procedures for structured support recovery that only use a small number of observations.

4.1 Procedures

All the procedures presented here are based on an algorithm named Compressive Adaptive Sense and Search (CASS), introduced and analyzed in [16]. This procedure is designed to recover non-structured support sets. To ease presentation we briefly describe and analyze the procedure here, though for the reader is referred to [16] where this has already been done in more detail.

4.1.1 s-sets

Assume the support set is any s-sparse set. Partition $\{1, \dots, n\}$ into $2s$ bins of equal size, denoted by $\boldsymbol{I}_1^{(1)}, \dots, \boldsymbol{I}_{2s}^{(1)}$. For each of the $2s$ bins we wish to decide between

$$H_{i,0}^{(1)} : \boldsymbol{I}_i^{(1)} \cap S = \emptyset \qquad \text{versus} \qquad H_{i,1}^{(1)} : \boldsymbol{I}_i^{(1)} \cap S \neq \emptyset, \; i = 1, \dots, 2s \; .$$

Once having identified the non-empty bins, we split each of these into two bins of equal size denoted by $\boldsymbol{I}_1^{(2)}, \dots, \boldsymbol{I}_{2n_1}^{(2)}$, where n_1 denotes the number of bins deemed non empty previously, and do the same as before. We know that at most s bins can be non-empty, thus we will enforce in our procedure that $n_1 \leq s$. Hence in step j we consider bins $\boldsymbol{I}_1^{(j)}, \dots, \boldsymbol{I}_{2n_{j-1}}^{(j)}$, where $n_{j-1} \leq s$, and test the hypotheses

$$H_{i,0}^{(j)} : \boldsymbol{I}_i^{(j)} \cap S = \emptyset \qquad \text{versus} \qquad H_{i,1}^{(j)} : \boldsymbol{I}_i^{(j)} \cap S \neq \emptyset, \; i = 1, \dots, 2n_{j-1} \; .$$

When $j = \log_2 \frac{n}{2s}$ the bins consist of single components of $\boldsymbol{x}$, and the estimator of the support $\widehat{S}$ will consist of the ones deemed non-empty in this final step.

To decide between $H_{i,0}^{(j)}$ and $H_{i,1}^{(j)}$, $j = 1, \ldots, \log_2 \frac{n}{2s}; i = 1, \ldots, 2n_{j-1}$ we collect a single measurement of the form

$$Y_i^{(j)} =< a\sqrt{j}\mathbf{1}_{\boldsymbol{I}_i^{(j)}}, \boldsymbol{x} > +W_i^{(j)},\ j = 1, \ldots \log_2 \frac{n}{2s}; i = 1, \ldots, 2n_{j-1} ,$$

where $W_i^{(j)} \sim N(0,1)$ i.i.d., and $a > 0$. The parameter $a > 0$ needs to be chosen such that (3) is fulfilled. Since the length of the bins $I_i^{(j)}$ is $n/2^j s$ for every $i = 1, \ldots, 2n_{j-1}$, $n_{j-1} \leq s$ and there are $\log_2 \frac{n}{2s}$ steps we can write

$$\mathbb{E}_S\left(\|A\|_F\right) = \sum_{j=1}^{\log_2 \frac{n}{2s}} 2s\frac{n}{2^j s} j a^2 \leq na^2 \sum_{j=1}^{\infty} j 2^{-(j-1)} = 4na^2 .$$

Combining this with (3) yields $a = \sqrt{\frac{m}{4n}}$. If the bin $\boldsymbol{I}_i^{(j)}$ is non-empty then $\mathbb{E}_S(Y_i^{(j)}) \geq \mu\sqrt{\frac{jm}{4n}}$. Therefore we conclude that the bin $\boldsymbol{I}_i^{(j)}$ is empty if $Y_i^{(j)} \leq \frac{\mu}{2}\sqrt{\frac{jm}{4n}}$, otherwise we conclude the opposite. If at any step $j = 1, \ldots, \log_2 \frac{n}{2s}$ more than s bins are deemed non-empty, we select those which correspond to the s largest observations. For the method described above both the type I and type II error probabilities for the test between $H_{i,0}^{(j)}$ and $H_{i,1}^{(j)}$, $j = 1, \ldots \log_2 \frac{n}{2s}; i = 1, \ldots, 2n_{j-1}$ can be upper bounded using the Gaussian tail bound

$$\mathbb{P}(X > \eta) \leq \frac{1}{2} e^{-\eta^2/2} \tag{8}$$

by

$$\frac{1}{2} e^{-\frac{jm\mu^2}{32n}} .$$

Hence the probability of error can be bounded from above as follows

$$\mathbb{P}_S(\widehat{S} \neq S) \leq \sum_{j=1}^{\log_2 \frac{n}{2s}} s\ e^{-\frac{jm\mu^2}{32n}} .$$

Thus whenever $\mu^2 \geq \frac{32n}{m} \log \frac{2s}{\varepsilon}$ we have

$$\mathbb{P}_S(\widehat{S} \neq S) \leq \sum_{j=1}^{\log_2 \frac{n}{2s}} s\left(\frac{\varepsilon}{2s}\right)^j \leq \sum_{j=1}^{\log_2 \frac{n}{2s}} \left(\frac{\varepsilon}{2}\right)^j \leq \varepsilon .$$

When considering the expected Hamming-distance as the error metric we can use the procedure above with probability of error set to $\varepsilon/2s$. This method then yields an near-optimal estimator for the support recovery problem described in Section 2 by collecting at most $2s \log_2 \frac{n}{2s}$ measurements.

4.1.2 Unions of s-intervals

We can simply modify the CASS procedure of [16] to estimate unions of k disjoint s-intervals. Similarly to the procedure presented in Section 3.1 the one discussed here will consist of two phases, a search phase and a refinement phase. As before, in the search phase we wish to identify the approximate location of the support, that is return a subset of components $\boldsymbol{P} \subset \{1,\ldots,n\}$ such that $|\boldsymbol{P}| \ll n$ and $S \subset \boldsymbol{P}$ with high probability. Again we start by splitting $\{1,\ldots,n\}$ into consecutive bins of size $s/2$ denoted by $\boldsymbol{P}^{(1)},\ldots,\boldsymbol{P}^{(2n/s)}$. To ease the presentation we assume $2n/s$ is an integer since the case when this is not satisfied can be handled with simple modifications. The same holds for any divisibility issue that we encounter further on. Of these bins at least k will consist entirely of signal components. Roughly speaking we think of these bins as signal components of a vector of size $2n/s$, and use a CASS procedure to find them. Once that is done, we set $\boldsymbol{P}$ as the bins deemed active and their neighboring bins, and move on to the refinement phase. In the refinement phase we estimate the active components in $\boldsymbol{P}$ for instance by using another CASS procedure.

We now describe the method in full detail. Consider the binning $\boldsymbol{P}^{(1)},\ldots,\boldsymbol{P}^{(2n/s)}$ described before. Partition the bins into $4k$ groups denoted by $\boldsymbol{I}_1^{(1)},\ldots,\boldsymbol{I}_{4k}^{(1)}$. For each of these we test the hypothesis

$$H_{i,0}^{(1)}:\ \boldsymbol{I}_i^{(1)} \cap S = \emptyset \qquad \text{versus} \qquad H_{i,1}^{(1)}:\ |\boldsymbol{I}_i^{(1)} \cap S| \geq s/2,\ i=1,\ldots,4k\ .$$

The groups for which $H_{i,1}^{(1)}$ is accepted are split into two in the middle giving us the groups $\boldsymbol{I}_1^{(2)},\ldots,\boldsymbol{I}_{2n_1}^{(2)}$. We now test a similar hypotheses as before for these new groups. Since at most $3k$ groups can contain signal components, we will specifically enforce $n_1 \leq 3k$. Iterating this, in step j we have groups denoted by $\boldsymbol{I}_1^{(j)},\ldots,\boldsymbol{I}_{2n_{j-1}}^{(j)}$, where $n_{j-1} \leq 3k$, and we wish to decide between

$$H_{i,0}^{(j)}:\ \boldsymbol{I}_i^{(j)} \cap S = \emptyset \qquad \text{versus} \qquad H_{i,1}^{(j)}:\ |\boldsymbol{I}_i^{(j)} \cap S| \geq s/2,\ i=1,\ldots,2n_{j-1}\ .$$

When $j = \log_2 n/2ks$ the groups consist of single bins. The set $\boldsymbol{P}$ will consist of the ones for which $H_{i,1}^{(1)}$ is accepted in this final step and the bins adjacent to those.

To decide between $H_{i,0}^{(j)}$ and $H_{i,1}^{(j)},\ j=1,\ldots,\log_2 \frac{n}{2s}; i=1,\ldots,2n_{j-1}$ we collect a single measurement of the form

$$Y_i^{(j)} = < a\sqrt{j}\mathbf{1}_{\boldsymbol{I}_i^{(j)}}, \boldsymbol{x} > + W_i^{(j)},\ j=1,\ldots \log_2 \frac{n}{2s}; i=1,\ldots,2n_{j-1}\ ,$$

where $W_i^{(j)} \sim N(0,1)$ i.i.d., and $a > 0$. The parameter $a > 0$ needs to be chosen such that (3) is fulfilled. We will use half of our energy budget for the search phase. Since the groups $I_i^{(j)}$ contain $n/2^{j+1}k$ components for every $i=1,\ldots,2n_{j-1}$, $n_{j-1} \leq 3k$ and there are $\log_2 \frac{n}{2ks}$ steps we can write

$$\mathbb{E}_S\left(\|A_{search}\|_F\right) = \sum_{j=1}^{\log_2 \frac{n}{2ks}} 6k\frac{n}{2^{j+1}k} j a^2 = \frac{3}{2} n a^2 \sum_{j=1}^{\log_2 \frac{n}{2s}} j 2^{-(j-1)} \leq 6na^2\ .$$

Since we use at most $m/2$ energy in the search phase we get $a = \sqrt{\frac{m}{12n}}$. If group $\boldsymbol{I}_i^{(j)}$ contains a bin which is contained in S, we have $\mathbb{E}_S(Y_i^{(j)}) \geq \frac{s\mu}{2}\sqrt{\frac{jm}{12n}}$. Therefore we declare that the group contains no signal components if $Y_i^{(j)} \leq \frac{s\mu}{4}\sqrt{\frac{jm}{12n}}$, otherwise we declare the opposite. If in step $j = 1, \ldots, \log_2 \frac{n}{2ks}$ we accept $H_{i,1}^{(j)}$ for more than $3k$ groups, we choose those corresponding to the highest $3k$ observations. Considering a single test the type I and type II error probabilities can both be upper bounded using (8) by

$$\frac{1}{2} e^{-\frac{js^2 m\mu^2}{384n}} .$$

Remark 11. *Note that the constant in the above bound is large, which is due to the crude algorithm choices made previously and using the same analysis as [16]. The authors made these choices to make the presentation easy to follow, as the main focus of this work is to analyze the rates of the bounds and not the constants. The constants can potentially be improved by wiser algorithm choices and a more careful analysis.*

It is also possible that neither the null or the alternative is true, and the group contains some bins that intersect with S, but are not contained in S. However we need not pay any attention to those, as by construction $\boldsymbol{P}$ will also contain neighboring bins of those we deem non-empty. The probability of either concluding $H_{i,1}^{(j)}$ when the group $\boldsymbol{I}_i^{(j)}$ contains no signal or concluding $H_{i,0}^{(j)}$ when in fact $H_{i,1}^{(j)}$ is true can be bounded from above by

$$\sum_{j=1}^{\log_2 \frac{n}{2ks}} 3k\; e^{-\frac{js^2 m\mu^2}{384n}} .$$

Thus whenever $\mu \geq \sqrt{\frac{384n}{s^2 m} \log \frac{9k}{\varepsilon}}$ we have that

$$\mathbb{P}_S(S \not\subseteq \boldsymbol{P}) \leq \sum_{j=1}^{\log_2 \frac{n}{2ks}} 3k \left(\frac{\varepsilon}{9k}\right)^j \leq \sum_{j=1}^{\log_2 \frac{n}{2ks}} \left(\frac{\varepsilon}{3}\right)^j \leq \varepsilon/2 .$$

We also have by construction that $|\boldsymbol{P}| \leq 3ks$. Hence in the refinement phase we can measure each component in $\boldsymbol{P}$ separately, say, to produce $\widehat{S}$. We have $m/6ks$ energy for each of the components in $\boldsymbol{P}$, hence it is easy to check using (8) that the probability of making an error in the refinement phase is at most

$$\frac{3ks}{2} e^{-\frac{m\mu^2}{48ks}} .$$

Whenever $\mu \geq \sqrt{\frac{48ks}{m} \log \frac{3ks}{\varepsilon}}$ the probability above is at most $\varepsilon/2$. Thus the procedure given an estimator $\widehat{S}$ for which $\mathbb{P}_S(\widehat{S} \neq S) \leq \varepsilon$ whenever

$$\mu \geq \sqrt{\max\{\frac{384n}{s^2 m} \log \frac{9k}{\varepsilon}, \frac{48ks}{m} \log \frac{3ks}{\varepsilon}\}} .$$

When considering the expected Hamming-distance as the error metric we can use the procedure above with probability of error set to $\varepsilon/2ks$. This method then yields an near-optimal estimator for the support recovery problem described in Section 2 by collecting at most $3k\left(\log_2 \frac{n}{2ks} + s\right)$ measurements.

Proposition 19. *Consider the class of k disjoint s-intervals and suppose $n > ks^3$. Then the procedure above satisfies* (3) *and* (4) *whenever*

$$\mu \geq \sqrt{\frac{384n}{s^2 m} \log \frac{18ks}{\varepsilon}} \ .$$

Furthermore, the procedure collects at most $3k\left(\log_2 \frac{n}{2ks} + s\right)$ observations.

Remark 12. *As with Proposition 2 the condition on the sparsity is an artifact of the simple method above and can be avoided by using a more elaborate method in the refinement phase, for instance binary search.*

4.1.3 Unions of s-stars

Consider the class of k disjoint s-stars. To ease the discussion we focus on the case $k = 1$, but the idea can be applied to larger k. The procedure is very similar to the one used for unions of s-intervals, however due to the different nature of the structure we provide a detailed description of the procedure in the Appendix.

Proposition 20. *Consider the class of s-stars, and suppose $\sqrt{2n} > p \geq 2s^2$. Then the procedure described in the Appendix satisfies* (3) *and* (4) *whenever*

$$\mu \geq \sqrt{\frac{128n}{s^2 m} \log \frac{6s}{\varepsilon}} \ .$$

Furthermore, the procedure collects at most $4\log_2 \frac{p}{4} + 2s\log_2 \frac{p-1}{s} \leq 8\log_2 n + 2s\log_2 \frac{\sqrt{2n}-1}{s}$ observations.

Similar ideas can be used to treat the case of k disjoint s-stars when $k > 1$, but $k \ll s$.

4.1.4 s_r, s_c-submatrices

Consider the class of submatrices of size $s_r \times s_c$ of a matrix of size $n_r \times n_c$, and suppose $s_r \geq s_c$. The procedure we present now is very similar to the one used for unions of s-intervals, hence we only provide an outline and present performance guarantees here.

Once more we break the procedure into two phases, a search phase and a refinement phase. The aim of the search phase is to find the active columns of the signal matrix, whereas the refinement phase aims to find the active rows once the active columns are found. If we view the columns of the signal matrix as components of a vector of dimension n_c, then finding the active columns can be viewed as estimating an unstructured s_c-sparse support set. Likewise the problem of the refinement phase can be viewed as finding an s_r-set in a signal of dimension n_r. Hence we can immediately use the CASS procedure for both sub-problems with modifications similar to those used in the case of unions of s-intervals. Thus we get the following.

Proposition 21. *Consider the class of s_r, s_c-submatrices and suppose $n_r > 2s_r^2/s_c$. There exists a procedure which yields an estimator satisfying* (3) *and* (4) *whenever*

$$\mu \geq \sqrt{\frac{64n}{s_r^2 m} \log \frac{4s}{\varepsilon}} \ .$$

Furthermore the estimator takes at most $2s_c \log_2 \frac{n_c}{2s_c} + 2s_r \log_2 \frac{n_r}{2s_r}$ *measurements.*

4.2 Sample Complexity lower bounds

Necessary conditions for the sample complexity of compressive sensing have been studied both in the adaptive and non-adaptive setting in [2] and [3]. In both works sample complexity was studied for the unstructured case of s-sets. For the non-adaptive setting the authors show in Theorem 4.1 of [2] that the sample complexity can be lower bounded by an expression that scales essentially like $s \log \frac{n}{s}$. Furthermore they also show that the signal to noise ratio plays a role in the sample complexity of compressive sensing, and this phenomenon is also explicitly captured in their bound. Though the setting considered in their work is slightly different from that in the present work, Theorem 4.1 of [2] can be translated into our setting in the following manner.

Lemma 5 (Theorem 4.1 of [2])**.** *Consider the class of s-sets, and suppose there exists a non-adaptive estimator satisfying* (3) *and for which* $\frac{1}{|\mathcal{C}|}\sum_{S\in\mathcal{C}} \mathbb{P}_S(\widehat{S} \neq S)$ *is not asymptotically bounded away from zero as $n, s \to \infty$. Let $k(n,s)$ denote the number of measurements the estimator makes. Then*

$$k(n,s) \geq \frac{cs \log \frac{n}{s}}{\log\left(\mu^2 \frac{m}{n} + 1\right)} ,$$

with some constant c.

This shows that the procedure presented in the previous section for s-sets performs as well in terms of sample complexity as the best non-adaptive procedure. Furthermore, when estimating structured support sets, potentially less samples are enough to perform accurate estimation. We now briefly discuss necessary conditions on sample complexity for non-adaptive estimators for the structured classes we examined before.

Consider first the case of unions of k disjoint s-intervals. Without giving a rigorous formal proof we argue that the number of samples required in the non-adaptive case must scale as $k \log \frac{n}{sk}$. Let $S_1, \ldots, S_{n/s}$ be consecutive disjoint s-intervals of $\{1, \ldots, n\}$ and let

$$\mathcal{C}' = \left\{ S \in \mathcal{C} : \; S = \bigcup_{j=1}^{k} S_{i_j}, \; i_1, \ldots, i_k \in \{1, \ldots, n/s\} \right\} ,$$

that is unions of intervals that are constructed from $S_1, \ldots, S_{n/s}$. This class roughly behaves like a class of k-sparse sets of a vector of dimension n/s, except that there is an increase in the relative sensing power arising from the fact that the building blocks of the class are s-sets instead of singletons. This results in that it is possible to detect somewhat weaker signals (see Proposition 8), but because of the weak dependence of the sample complexity bound of Lemma 5 on the signal to noise ratio, the scaling of the bound will still be dictated by the numerator.

The class of unions of k disjoint s-stars is even more simple to consider. Suppose $k = 1$, and that the center of the star is given by an oracle. The remaining problem is the estimation of an s-sparse set in a vector with dimension roughly $\sqrt{2n}$. Hence the sample complexity remains essentially the same as that of the unstructured case.

Finally for the class of s_r, s_c-submatrices, if an oracle provides the active columns, the problem reduces to the unions of intervals case.

This shows that the procedures presented in the previous section for structured support recovery perform as well in terms of sample complexity as the best non-adaptive procedures. It is plausible however that adaptive procedures might outperform non-adaptive ones in terms of sample complexity. This question was investigated in [3], where the authors provide a necessary condition for any adaptive algorithm to recover unstructured s-sets. The number of samples required is dependent on the signal to noise ratio in this case as well. Their results show that when the signal to noise ratio is near the boundary where accurate estimation is possible (see Proposition 11, and [13]) the number of samples needs to scale essentially like s. It is still an open question whether this bound is achievable or not.

Although not yet having a rigorous proof, the authors of this work conjecture that although some performance gain might be present, it is not substantial and the number of samples needs to scale essentially like $s \log \frac{n}{s}$ for adaptive estimators as well. The reason behind this conjecture is roughly the following. Consider the 1-sparse case. It can be easily seen that by taking one measurement, a fraction of the n hypotheses (namely that the signal component is at coordinate $1, \ldots, n$) remains essentially indistinguishable. Focusing the next measurement on these potential signal components, again a fraction of them will remain essentially indistinguishable. With a bit of work this line of reasoning will, in principle, provide a lower bound on the sampling complexity. However, formalizing this argument is challenging, because each projection does contain some small amount of information about these "indistinguishable" hypotheses. So one needs to show that this small amounts of information are negligible as a whole, even after collecting multiple projections. Nonetheless, the authors conjecture that because of this heuristic, a term that is logarithmic in the dimension should also be present in the sample complexity lower bounds.

5 Final remarks

In this work we examined the problem of recovering structured support sets through adaptive compressive measurements. We have seen that adaptively designing the sensing matrix it is possible to achieve performance gains over non-adaptive protocols, and that the gains can be quite dramatic for instance in the case of s-stars. We have also seen that these gains can be realized by simple and practically feasible estimation procedures.

However a complete characterization of the problem for the class of submatrices is still missing. This could prove to be an interesting area for future research considering the practical relevance of that model in gene expression studies. Furthermore it remains unclear if the sample complexity of support recovery using compressive measurements can be significantly reduced by adaptively designing the rows of the sensing matrix.

Appendix

Description of the procedure of Section 4.1.3

We begin with a search phase to find the approximate location of the support. Again we consider the subsets $\boldsymbol{P}^{(i)},\ i=1,\ldots,p$, where $\boldsymbol{P}^{(i)}$ contains all the components whose corresponding edges lie on the vertex v_i. Our goal is to find the center of the s-star. We begin by forming 4 groups $\boldsymbol{I}_1^{(1)},\ldots,\boldsymbol{I}_4^{(1)}$, where each of them is a union of $p/4$ different $\boldsymbol{P}^{(i)}$, and no subset $\boldsymbol{P}^{(i)}$ is contained in more than one group. We then take one measurement per group

$$Y_i^{(1)} = < a\mathbf{1}_{\boldsymbol{I}_i^{(1)}}, \boldsymbol{x} > + W_i^{(1)},\ i=1,\ldots,4,$$

where $W_i^{(1)}$ i.i.d. standard normals and $a>0$. Large measurements should correspond to groups containing a lot of signal components, and particularly the one containing the center of the star. However, because of the structure of the support and the fact that these groups are not disjoint, large observations may also correspond to groups not containing the center of the star. Therefore instead of performing hypothesis tests we choose the two highest observations, and consider the groups corresponding to those. Once we have these groups, we split each in half in the sense that half of the $\boldsymbol{P}^{(i)}$ in a given group will form one new group, and the other half will form another new group. This way we end up with 4 groups, again not disjoint, and do the same as before. Let the groups in step j be denoted by $\boldsymbol{I}_1^{(j)},\ldots,\boldsymbol{I}_4^{(j)}$. The measurements we collect are

$$Y_i^{(j)} = < a\sqrt{j}\mathbf{1}_{\boldsymbol{I}_i^{(j)}}, \boldsymbol{x} > + W_i^{(j)},\ j=1,\ldots,\log_2\frac{p}{4}; i=1,\ldots,4\ .$$

In the final step $j=\log_2\frac{p}{4}$ each group consists of a single $\boldsymbol{P}^{(i)}$. The output set of the search phase $\boldsymbol{P}$ will consist of the union of those two groups for which the final observation is largest.

First we specify the parameter a so as to ensure we don't use more than half of our measurement budget. Each $\boldsymbol{I}_i^{(j)}$ contains $(p-1)\frac{p}{2^{j+1}} = n/2^j$ components $i=1,\ldots,4$, and $j=1,\ldots,\log_2\frac{p}{4}$, hence

$$\mathbb{E}_S\left(\|A_{search}\|_F\right) = \sum_{j=1}^{\log_2\frac{p}{4}} \frac{n}{2^{j-2}} ja^2 \leq 8na^2\ .$$

Therefore $a=\sqrt{\frac{m}{16n}}$ ensures we use at most $m/2$ energy in the search phase.

Now we need to show that $S\subset\boldsymbol{P}$ with high probability. Without loss of generality suppose that $\boldsymbol{I}_1^{(j)},\ldots,\boldsymbol{I}_4^{(j)}$ are indexed such that the center of the star is in group $\boldsymbol{I}_1^{(j)}$, and for the number of signal components in $\boldsymbol{I}_i^{(j)}$ denoted by $N_i^{(j)}$ we have $N_i^{(j)} \geq N_{i+1}^{(j)}$. Hence $\boldsymbol{I}_1^{(j)}$ contains exactly s components, and because $\sum_{i=2}^{4} N_i^{(j)} \leq s$ we know $N_3^{(j)} \leq s/2$. Using this we conclude that in each step $j=1,\ldots,\log_2\frac{p}{4}$ the probability that $Y_1^{(j)} < \max\{Y_3^{(j)}, Y_4^{(j)}\}$ can be bounded from above with (8) by

$$2\cdot\frac{1}{2}e^{-\frac{js^2m\mu^2}{128n}}\ .$$

From this we get that whenever $\mu \geq \sqrt{\frac{128n}{s^2 m} \log \frac{3}{\varepsilon}}$ we have

$$\mathbb{P}_S(S \not\subseteq \boldsymbol{P}) \leq \sum_{j=1}^{\log_2 \frac{p}{4}} \left(\frac{\varepsilon}{3}\right)^j \leq \varepsilon/2 \ .$$

By construction we make $4 \log_2 \frac{p}{4}$ observations in this phase, and also $|\boldsymbol{P}| \leq 2(p-1)$.

In the search phase we can directly apply the CASS procedure on $\boldsymbol{P}$ to estimate the support. We know that whenever $\mu \geq \sqrt{\frac{64(p-1)}{m} \log \frac{4s}{\varepsilon}}$ the probability of error is at most $\varepsilon/2$, and we take at most $2s \log_2 \frac{p-1}{s}$ measurements. When considering $\mathbb{E}_S(|\widehat{S} \triangle S|)$ as the error metric one can set the probability of error to $\varepsilon/2s$ and use the procedure above.

Sketch proof of Proposition 21

We use half the energy for the search phase, and half for the refinement phase. In step j of the search phase the groups $\boldsymbol{I}^{(j)}$ contain $n/2^j s_c$ components and there are at most $2s_c$ components. Hence the energy used is at most

$$\sum_{j=1}^{\log_2 \frac{n_c}{2s_c}} 2s_c \frac{n}{2^j s_c} j a^2 = 4na^2 \ .$$

Thus $a = \sqrt{\frac{m}{8n}}$. This means that for the probability of error we have

$$\sum_{j=1}^{\log_2 \frac{n_c}{2s_c}} 2s_c \frac{1}{2} e^{-\frac{s_r^2 m \mu^2}{64n}} \ ,$$

so whenever $\mu \geq \sqrt{\frac{64n}{s_r^2 m} \log \frac{2s_c}{\varepsilon}}$ the probability of error is at most $\varepsilon/2$.

In the refinement phase the energy used is

$$\sum_{j=1}^{\log_2 \frac{n_r}{2s_r}} 2s_r \frac{n_r s_c}{2^j s_r} j a^2 = 4n_r s_c a^2 \ ,$$

hence $a = \sqrt{\frac{m}{8n_r s_c}}$. Therefore the probability of error is at most

$$\sum_{j=1}^{\log_2 \frac{n_r}{2s_r}} 2s_r \frac{1}{2} e^{-\frac{s_c m \mu^2}{64n_r}} \ ,$$

which means whenever $\mu \geq \sqrt{\frac{64n_r}{s_c m} \log \frac{2s_r}{\varepsilon}}$ the probability of error is at most $\varepsilon/2$.

Considering the expected Hamming-distance as the error metric, we can use the procedure above with probability of error set to $\varepsilon/2s$

References

[1] Addario-Berry, L., Broutin, N., Devroye, L., and Lugosi, G. On combinatorial testing problems. *The Annals of Statistics 38*, 5 (2010), 3063–3092.

[2] Aksoylar, C., Atia, G., and Saligrama, V. Sparse signal processing with linear and non-linear observations: A unified shannon theoretic approach. In *Information Theory Workshop (ITW), 2013 IEEE* (2013), IEEE, pp. 1–5.

[3] Aksoylar, C., and Saligrama, V. Information-theoretic bounds for adaptive sparse recovery. *arXiv preprint arXiv:1402.5731* (2014).

[4] Arias-Castro, E., Candes, E. J., Helgason, H., and Zeitouni, O. Searching for a trail of evidence in a maze. *The Annals of Statistics* (2008), 1726–1757.

[5] Arias-Castro, E., et al. Detecting a vector based on linear measurements. *Electronic Journal of Statistics 6* (2012), 547–558.

[6] Balakrishnan, S., Kolar, M., Rinaldo, A., and Singh, A. Recovering block-structured activations using compressive measurements. *arXiv preprint arXiv:1209.3431* (2012).

[7] Baraniuk, R. G., Cevher, V., Duarte, M. F., and Hegde, C. Model-based compressive sensing. *Information Theory, IEEE Transactions on 56*, 4 (2010), 1982–2001.

[8] Butucea, C., and Ingster, Y. I. Detection of a sparse submatrix of a high-dimensional noisy matrix. *arXiv preprint arXiv:1109.0898* (2011).

[9] Candes, E., Arias-Castro, E., and Davenport, M. On the fundamental limits of adaptive sensing. *arXiv preprint arXiv:1111.4646* (2013).

[10] Candes, E., and Tao, T. The dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics* (2007), 2313–2351.

[11] Candes, E. J., and Tao, T. Near-optimal signal recovery from random projections: Universal encoding strategies? *Information Theory, IEEE Transactions on 52*, 12 (2006), 5406–5425.

[12] Candès, E. J., and Wakin, M. B. An introduction to compressive sampling. *Signal Processing Magazine, IEEE 25*, 2 (2008), 21–30.

[13] Castro, R. M. Adaptive sensing performance lower bounds for sparse signal estimation and testing. *Bernoulli 20*, 4 (2014), 2217–2246.

[14] Donoho, D. L. Compressed sensing. *Information Theory, IEEE Transactions on 52*, 4 (2006), 1289–1306.

[15] Haupt, J. D., Baraniuk, R. G., Castro, R. M., and Nowak, R. D. Compressive distilled sensing: Sparse recovery using adaptivity in compressive measurements. In *Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on* (2009), IEEE, pp. 1551–1555.

[16] Malloy, M. L., and Nowak, R. D. Near-optimal adaptive compressed sensing. In *Signals, Systems and Computers (ASILOMAR), 2012 Conference Record of the Forty Sixth Asilomar Conference on* (2012), IEEE, pp. 1935–1939.

[17] Malloy, M. L., and Nowak, R. D. Near-optimal compressive binary search. *arXiv preprint arXiv:1203.1804* (2012).

[18] Soni, A., and Haupt, J. Efficient adaptive compressive sensing using sparse hierarchical learned dictionaries. In *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on* (2011), IEEE, pp. 1250–1254.

[19] Tánczos, E., and Castro, R. M. Adaptive sensing for estimation of structured sparse signals. *arXiv preprint arXiv:1311.7118* (2013).

[20] Tsybakov, A. B. *Introduction to Nonparametric Estimation*, vol. 41 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer, Berlin, 2009.

[21] Wainwright, M. J. Sharp thresholds for high-dimensional and noisy sparsity recovery using-constrained quadratic programming (lasso). *Information Theory, IEEE Transactions on 55*, 5 (2009), 2183–2202.