On the Performance of Adaptive Sensing for Sparse Signal Inference

Rui M. Castro
Eindhoven University of Technology
The Netherlands
Email: rmcastro@tue.nl

Abstract—In this short paper we survey recent results characterizing the fundamental draws and limitations of adaptive sensing for sparse signal inference. We consider two different adaptive sensing paradigms, based either on single-entry or linear measurements. Signal magnitude requirements for reliable inference are shown for two different inference goals, namely signal detection and signal support estimation.

II. SINGLE-ENTRY MEASUREMENTS

This sensing model was first proposed in [15]. Measurements are of the form

\[ Y_k = x_{A_k} + \Gamma_k^{-1/2}W_k, \quad k = 1, 2, \ldots, \]

where \( A_k, \Gamma_k \) are taken to be functions of \( \{ Y_i, A_i, \Gamma_i \}_{i=1}^{k-1} \), and \( W_k \) are standard normal random variables, independent of \( \{ Y_i \}_{i=1}^{k-1} \) and also independent of \( \{ A_i, \Gamma_i \}_{i=1}^{k-1} \). In words, each measurement corresponds to a single signal entry corrupted with additive Gaussian noise, and the choice of entry and noise level can be controlled. However, there is a total sensing budget constraint that must be satisfied, namely

\[ \sum_{k=1}^{\infty} \Gamma_k \leq m, \]

where \( m > 0 \). In the above model \( A_k \) should be viewed as the sensing action taken at time \( k \), and \( \Gamma_k \) is the precision of the corresponding measurement. We have control over both quantities. Informally stated, measurements are collected sequentially, and for each measurement we can choose which entry of \( x \) to observe, and what is the precision (i.e. accuracy) of the measurement. We are allowed to collect as many measurements as desired provided the cumulative precision used satisfies the budget (1). Note that in this model we are allowed to collect an infinite (but countable) number of measurements, provided the precision \( \Gamma_k \) converges to zero as \( k \) grows. Although this might seem strange at first, it is not entirely unreasonable in practice - in many sensing modalities the precision is directly proportional to the amount of time necessary to collect a measurement, and therefore (1) can be viewed simply as a time constraint. This is the case in various imaging modalities (e.g. in astronomy) where long exposure times are used to reduce the noise level, which is inversely proportional to the exposure time. It is important to note that there are also settings where the actual number of measurements is limited, and there is little control on the precision level. In that case (1) might represent a constrain on the total number of measurements, provided \( \Gamma_k \) is not a function of \( k \). The results in the latter setting are similar to the ones presented in the current paper, especially when studying asymptotics (when both \( n \) and \( m \) grow).

It is important to note that we can consider both deterministic sequential designs or random sequential designs. In
the latter we allow the choices $A_k$ and $\Gamma_k$ to incorporate extraneous randomness, which is not explicitly described in the model. The collection of conditional distributions of $A_k, \Gamma_k$ given \( \{Y_i, A_i, \Gamma_i\}_{i=1}^n \) for all $k$ is referred to as the sensing strategy. Note that, within the sensing paradigm above we can also consider non-adaptive sensing, meaning that the choice of sensing actions and corresponding precision is made before collecting any data. Formally this means that \( \{A_k, \Gamma_k\}_{k \in \mathbb{N}} \) is statistically independent from \( \{Y_k\}_{k \in \mathbb{N}} \). Note that a non-adaptive design can still be random.

The case $m = n$ is of particular interest, allowing a direct comparison between adaptive and non-adaptive sensing methodologies. When $m = n$ we allow on average one unit of precision per each of the signal entries. So, if there is no reason to give preference to any particular entry of $x$, the natural optimal non-adaptive sensing strategy should simply measure each entry of $x$ exactly once, with precision one. This corresponds to the well studied normal means model.

For simplicity of presentation we consider only signals of the form

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

where $\mu > 0$ is called the signal amplitude. This restriction is also considered in [16], [17] in the non-adaptive sensing context and does not substantially hinder the generality of the results presented in this manuscript.

As stated before we consider two different inference problems: (i) signal detection and (ii) support estimation. For the detection problem (i) the goal is to determine if a signal is present or absent. We formulate the problem as a binary hypothesis testing, and test a simple null hypothesis against a

$$H_0 : S = \emptyset,$$

and the alternative hypothesis $H_1$ is $S \subset \mathcal{C}$, where $\mathcal{C}$ is some class of non-empty subsets of \( \{1, \ldots, n\} \).

For simplicity of presentation we assume that all the sets in $\mathcal{C}$ have the same cardinality $s$. A test procedure based on the (adaptive) measurements is described by a binary test function $\hat{\phi} (\{A_i, \Gamma_i, Y_i\}_{i=1}^n) \in \{0, 1\}$, and a natural way to measure the performance of such a test function is the worst case risk

$$R(\hat{\phi}) = \mathbb{P}_S (\hat{\phi} = 0) + \max_{S \in \mathcal{C}} \mathbb{P}_S (\hat{\phi} = 1),$$

where $\mathbb{P}_S$ denotes the joint probability distribution of $\{A_i, \Gamma_i, Y_i\}_{i=1}^n$ for a given support set $S$. Characterizing the relation between $R(\hat{\phi})$, $n$, $m$, $\mu$, and $\mathcal{C}$ is our main objective.

The goal of the estimation problem (ii) is (statistically) more ambitious, as we seek to actually identify the support set $S$. An estimation procedure is a function $\hat{S}$ mapping $\{A_i, \Gamma_i, Y_i\}_{i=1}^n$ to a subset of $\{1, \ldots, n\}$. There are several ways to measure "closeness" between $\hat{S}$ and the true support set $S$, for instance the worst case probability of making any errors

$$\max_{S \in \mathcal{C}} \mathbb{P}_S [\hat{S} \neq S].$$

A somewhat more stringent metric is the worst case expected number of errors $\max_{S \in \mathcal{C}} \mathbb{E}_S [E_S [\hat{S} \Delta S]$, and clearly $\mathbb{P}_S [\hat{S} \neq S] \leq \mathbb{E}_S [\hat{S} \Delta S]$. We will focus mainly on the first metric in this manuscript, but remark that the two metrics are essentially equivalent in several cases.

A. Single-entry Measurements: Results

In this section we present the fundamental tradeoffs for the inference problems presented above. Clearly these results bear some dependency on the class of sets $\mathcal{C}$.

**Definition II.1** (symmetric class). Let $S$ be a random set, drawn uniformly at random from $\mathcal{C}$. If for all $i \in \{1, \ldots, n\}$ we have $\mathbb{P}(i \in S) = s/n$ the class $\mathcal{C}$ is said to be symmetric.

In words, in a symmetric class of sets there is no reason to give a priori preference to any individual entry. Many classes $\mathcal{C}$ of interest satisfy this mild symmetry, for instance all the classes in [16]. Of particular interest is the maximal class of all the subsets of $\{1, \ldots, n\}$ with cardinality $s$, which corresponds to lack of structure in the sparsity pattern $S$. If the class $\mathcal{C}$ is smaller then we say the sparsity patterns $S$ have structure. An example of a structured class is presented later.

**Theorem II.1** ([18]). Let $\mathcal{C}$ be a symmetric class, and let $\hat{\Phi}$ be an arbitrary adaptive sensing testing procedure. For any $0 < \epsilon < 1$, if $R(\hat{\Phi}) \leq \epsilon$ then necessarily

$$\mu \geq \sqrt{\frac{2n}{sm} \log \frac{1}{2\epsilon}}.$$

As argued before, the case $m = n$ is of particular interest, as it allows for comparison between adaptive and non-adaptive sensing performance: in that case the above bound is of the order $\sqrt{2/s}$. It is remarkable that the extrinsic signal dimension $n$ plays no role in this bound, and only the intrinsic signal dimension $s$ is relevant. This is in stark contrast to what is known for the same problem if one restricts to the classical setting of non-adaptive sensing, as in [19], [20], [17]. For instance, for the class of all subsets with cardinality $s$ the non-adaptive sensing lower bound is of the order $\sqrt{\log(n/s^2)}$ if $s < o(\sqrt{n})$. Therefore signals need to be much stronger in order to be reliably detected when using non-adaptive sensing.

The above adaptive sensing lower bound is valid for any symmetric class, and in particular for the maximal class of all subsets $S$ with cardinality $s$. For this class there is a adaptive sensing methodology able to nearly achieve the lower bound.

**Proposition II.1** ([18]). Let $s_n \to \log \log \log n$ and consider the class $\mathcal{C}$ of all subsets with cardinality $s_n$. Furthermore let $\mu > \sqrt{\frac{s_n \log \log \log n}{m}}$. There is an adaptive sensing testing strategy for which

$$R(\hat{\Phi}) \to 0,$$

as $n \to \infty$.

The mentioned procedure is based on the idea of distilled sensing [15], but it does require some simple modifications to attain the desired bound (see [18]). Note that the order of the bound matches the one of the lower bound up to a factor $\log \log \log n$. It is conjectured that this is an artifact of the specific procedure, however, there are currently no known procedures able to tighten this gap. Perhaps more noteworthy
is the fact that extra structure in the class $C$ is not helpful in the adaptive sensing detection scenario! This is quite different than in the non-adaptive sensing case, where the structure of the set $C$ can play a very prominent role as well documented in [16], [21], [22], for instance.

The estimation problem exhibit similar trends, but structure of the set $C$ can give important cues on the design of adaptive sensing methodologies. We focus first on the unstructured case where $C$ is the class of all subsets of $\{1, \ldots, n\}$ with cardinality $s$.

**Theorem II.2** ([18]). Let $C$ be the class of all subsets with cardinality $s$, and let $\hat{S}$ be an arbitrary adaptive sensing support estimator. For any $0 < \epsilon < 1$, if
\[
\max_{S \subseteq C} P_S[\hat{S} \neq S] \leq \max_{S \subseteq C} \mathbb{E}_S[|\hat{S} \Delta S|] \leq \epsilon \text{ then necessarily}
\]
\[
\mu^2 \geq \sqrt{\frac{2n}{m} \left( \log s + \log \frac{n-s}{n+1} + \log \frac{1}{2\epsilon} \right)} .
\]

Again, focusing on the case $m = n$ and assuming also the signal is sufficiently sparse (meaning $s_n = o(n)$), we see that $\mu$ needs to be on the order of $\sqrt{2\log\frac{n}{s_n}}$ to ensure the probability of making any errors goes to zero as $n$ increases. This result is again in stark contrast with what is possible with non-adaptive sensing, where the signal magnitude $\mu$ needs to be on the order of $\sqrt{2\log n}$ to ensure the probability of error goes to zero. Furthermore the above lower bound is tight, as there is a procedure that allows for exact support recovery with probability approaching 1 provided the signal magnitude $\mu$ is of the order $\sqrt{2\log\frac{n}{s_n}}$, namely $\frac{1}{\sqrt{s_n}}\sqrt{n/m}$. Although this result is somewhat similar to the detection problem was been carefully studied in [28] and the author has shown that for reliable detection it is necessary and sufficient for the signal magnitude to be of the order $\frac{1}{s_n}\sqrt{n/m}$. In the results described below we consider only the sensing budget restriction (2) and assume the number of measurements $l$ can be potentially infinite.

As linear measurements are more powerful/general than entry-wise ones, we might expect some performance improvement in both the detection and estimation inference tasks. The detection problem was been carefully studied in [28] and the author has shown that for reliable detection it is necessary and sufficient for the signal magnitude to be of the order $\frac{1}{s_n}\sqrt{n/m}$. Although this result is somewhat similar to the one in Theorem II.1 we notice that the dependency on the sparsity level $s_n$ is better, and therefore weaker signals can be detected using linear measurements. Perhaps surprisingly adaptive sensing is of no help in this scenario, and detection procedures achieving the optimal performance can be non-adaptive. Furthermore, the structure of the class $C$ does not help, provided the class is symmetric. This means that, like in the single-entry measurement case, structure is of no use for detection. However, this statement is true both for adaptive and non-adaptive sensing paradigms, meaning that the extra flexibility of adaptive sensing provides no advantage for detection using linear measurements.

For the estimation problem the story is a bit different: adaptive sensing can exhibit an advantage over non-adaptive sensing, as documented in [29], [30], [31]. Furthermore structural information about $S$ can be extremely helpful. In [18] it is shown that for the unstructured case the same lower bound as in Theorem II.2 applies in the context of linear measurements (although the proof of the result requires a few small modifications). Procedures achieving (or nearly achieving) this bound exist, namely [31], [32]. For the non-adaptive sensing paradigm information theoretical lower bounds have also been shown, namely the signal amplitude must exceed a constant times $\sqrt{\frac{2}{m}}\sigma^2\log n$, as shown for instance in [33].
The factor of $n/m$ is the sensing energy per dimension and $\sqrt{\log n}$ is needed to ensure that the signal is larger than the largest noise contribution. Therefore adaptive sensing is advantageous, especially in the typical case when the signal dimension $n$ is very large.

If the sparsity patterns exhibit some structure there are also results contrasting adaptive and non-adaptive sensing, but the story is far from complete. In [34] the authors devise an algorithm that can identify the support set $S$ with high probability when $S$ is an “interval” (see the last paragraph of Section II) provided the signal magnitude is of the order $\sqrt{(n/m)(\log s_n)/s_n}$. Furthermore they prove a lower bound of the form $\sqrt{(n/m)/s_n}$, which matches the upper bound apart from the $\sqrt{\log s_n}$ factor (which does not appear to be an artifact of the algorithm). Again, note that linear measurements are advantageous over entry-wise ones, for which signal magnitude must scale like $\sqrt{(n/m)(\log(s_n)/s_n)}$ for this problem.

IV. FINAL REMARKS

In this brief note we surveyed existing results over adaptive sensing of sparse signals. We considered both entry-wise and linear measurements and clarified in which situations can adaptive sensing yield interesting gains over non-adaptive designs. A clear picture exists for the unstructured scenario, where one assumes only that the support set $S$ is sparse. If in addition one can make structural assumptions over $S$ than it is clear that support estimation is possible for even weaker signals. With so few results available along those lines this remains an interesting avenue for future research.

ACKNOWLEDGMENT

The author would like to thank the anonymous reviewers for their helpful pointers and remarks.

REFERENCES
