Sequentially Designed Compressed Sensing

Jarvis Haupt[1], Richard Baraniuk[2], Rui Castro[3], and Robert Nowak[4],

[1] Dept. of Electrical and Computer Engineering; University of Minnesota; Minneapolis, MN 55455
[2] Dept. of Electrical and Computer Engineering; Rice University; Houston, TX 77005
[3] Dept. of Mathematics, Eindhoven University of Technology; Eindhoven, The Netherlands
[4] Dept. of Electrical and Computer Engineering; University of Wisconsin; Madison, WI 53706

***Abstract*—A sequential adaptive compressed sensing procedure for signal support recovery is proposed and analyzed. The procedure is based on the principle of *distilled sensing*, and makes used of sparse sensing matrices to perform sketching observations able to quickly identify irrelevant signal components. It is shown that adaptive compressed sensing enables recovery of weaker sparse signals than those that can be recovered using traditional non-adaptive compressed sensing approaches.**

***Index Terms*—Adaptive sensing, compressed sensing, support recovery**

I. Introduction

A sparse vector can be uniquely identified from a relatively small number of random measurements. Specifically, consider $\mathbf{x} \in \mathbb{R}^n$ with $k < n$ non-zero elements and suppose we measure $\mathbf{y} = \mathbf{A}\mathbf{x}$, where $\mathbf{A} \in \mathbb{R}^{m \times n}$, with $m < n$. The dimension of $\mathbf{x}$ is greater than the number of measurements, but because $\mathbf{x}$ is sparse it can be perfectly recovered from $\mathbf{y}$ if $\mathbf{A}$ is designed carefully. The theory of Compressed Sensing shows that if the elements of $\mathbf{A}$ are generated by i.i.d. zero mean Gaussian random variables, for example, then $\mathbf{x}$ is can be recovered perfectly if $m = O(k \log n)$; see, eg., [1], [2].

Compressed Sensing has a variety of potential applications, but in many cases the measurements may be corrupted by noise. Instead of measuring $\mathbf{y} = \mathbf{A}\mathbf{x}$, we instead measure $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$, where $\mathbf{w} \in \mathbb{R}^m$ is a noise or error vector. In general, perfect recovery is no longer possible in the noisy case. Two objectives to consider are:

support recovery	–	infer the locations of non-zero elements
signal estimation	–	estimate $\mathbf{x}$

Clearly both objectives are challenging if $\mathbf{x}$ is weak relative to $\mathbf{w}$. More precisely, assume the noise is i.i.d. zero mean Gaussian with variance σ^2 and let us impose the constraint

$$\mathbb{E}\|\mathbf{A}\|_F^2 \leq M, \tag{1}$$

where $\|\cdot\|_F$ is the Frobenius matrix norm. The constraint simply limits the total amount of sensing energy – for example, it is satisfied if $\mathbf{A}$ has $m = M$ unit norm rows, though in general the number of measurements m and the total measurement budget M need not be equal. A constraint on the available sensing energy is natural in any conceivable practical implementation, and from a mathematical perspective the issue of noise is irrelevant without it. Under this constraint, the probability of correct support recovery and quality of the estimation are limited by the relative strength of $\mathbf{x}$ compared to $\mathbf{w}$. Note that if the correct support is identified, then estimating the signal reduces to the classical problem of estimating an arbitrary vector in $\mathbb{R}^k$. Therefore, we focus on support recovery in this paper.

The sensing matrix $\mathbf{A}$ can be designed non-adaptively or adaptively. In a non-adaptive design, $\mathbf{A}$ is designed a priori before any measurements are acquired. The i.i.d. Gaussian design mentioned above is an example of this. *Adaptive designs, in contrast, are sequential.* The design of each row of $\mathbf{A}$ can depend on the measurements obtained from previous rows. Adaptive designs are more flexible and may be able to focus sensing toward the non-zero elements of $\mathbf{x}$, improving the SNR of the measurements. In fact, it is known that adaptive designs can offer significant advantages over non-adaptive designs when noise is present in the measurements.

To ensure the probability of exactly recovering the support is close to 1, the non-zero elements of $\mathbf{x}$ must be significantly large relative to the noise level. For non-adaptive designs, the amplitude of the smallest non-zero element of $\mathbf{x}$ must exceed a constant times $\sqrt{\frac{n}{M}\sigma^2 \log n}$; see, eg., [3]–[5]. The factor of M/n is the sensing energy per dimension and $\sqrt{\log n}$ is needed to ensure that the signal is larger than the largest noise contribution. As we show in this paper, adaptive designs succeed under the weaker requirement that the amplitude of the smallest non-zero element of $\mathbf{x}$ exceeds a constant times $\sqrt{\frac{n}{M}\sigma^2(\log k + \log\log_2 \log n)}$. It is worth emphasizing that this improvement can lead to dramatic gains in practice, since in many applications of interest $n > 10^6$ and $k \ll n$.

II. Main Result

Our main contribution is a sequential procedure for the design of compressed measurements, referred to as Sequential Compressed Sensing (SCS), and stated in Algorithm 1 in the next section. The algorithm consists of two stages. The first stage involves $\log_2 \log n$ steps. Each step uses several sparse compressed sensing matrices (consisting of about of $k \log k \log\log n$ measurements in total), which are used to remove half of the zero components (in expectation) at each step, while guaranteeing that all the non-zero components are retained (with large probability). Thus, the expected number of components remaining after the first stage is bounded by $n/\log n + k$, and the k non-zero components are guaranteed to be contained in this set. The second stage faces a lower dimensional problem (approximately $n/\log n$ instead of n) and this makes it possible to reliably remove all remaining zero components using only $k \log n$ additional measurements. The key advantage of this sequential algorithm is that the support can be recovered exactly at much lower SNRs compared to traditional (non-sequential, non-adaptive) compressed sensing methods. Our main result is the following theorem.

Theorem 1. *The SCS procedure described in Algorithm 1 uses $O(k \max\{\log k \log\log n, \log n\})$ measurements. If the minimum non-zero value in $\mathbf{x}$ is greater than a constant times*

$$\sqrt{\frac{n}{M}\sigma^2(\log k + \log\log_2 \log n)}\ ,$$

then Algorithm 1 exactly recovers the support of $\mathbf{x}$ with probability at least $1 - \delta$, for any choice of $\delta > 0$ used in the algorithm. Furthermore, for sufficiently large n and $k \leq n/\log_2 \log n$, the expected sensing energy used by the algorithm is at most M.

It is worth noting that when $\mathbf{x}$ is sufficiently sparse, i.e., when $k = O(n^{\beta/\log\log n})$ for some fixed $\beta > 0$, the total number of measurements satisfies $O(k \log n)$, so this procedure can be compared fairly with classical compressed sensing in these regimes.

III. Sequential Compressed Sensing Algorithm

The sequential compressed sensing procedure we propose here is based on the sequential sensing-and-refinement approach proposed in our earlier work on *distilled sensing* (DS) [6]. The essential idea of DS is to acquire the unknown sparse vector using a multi-step acquisition process, where each step involves measurement and refinement. The refinements sequentially reduce the dimension of the search space, in order to preferentially place more sensing energy on the locations where the signal components are most likely to be present, while not "wasting" sensing energy on locations where signal components are likely absent. Our initial DS work considered a direct (i.e., non-compressive) measurement model, identified conditions on the minimum signal component amplitude sufficient to ensure that DS successfully recovers all but an asymptotically vanishing fraction of signal support, and showed that DS succeeds at recovering weaker signals than comparable non-adaptive approaches. Here we extend these ideas to compressive sensing, and seek to identify conditions sufficient to enable a stronger *exact* support recovery result.

This SCS procedure proposed here (depicted in Algorithm 1) also improves on our earlier compressive distilled sensing (CDS) work [7], which employed a different measurement and estimation process (based on random Gaussian sensing matrices). The recovery results we obtained in [7] were valid only for sparse signals with dynamic range (the ratio of the largest to smallest nonzero component amplitudes) on the order of a constant. Our current approach removes this dynamic range constraint.

This work is related to other recent efforts in adaptive compressive sensing. In [8] a measurement model without the additive noise term was examined, and an adaptive sensing procedure that recovers k-sparse signals (or approximate nearly sparse signals to within a constant factor of the best k-term approximation) using only $O(k \log \log n)$ measurements was proposed. The authors of [9] considered a noisy measurement model much like the one we adopt here, and in the context of establishing fundamental lower bounds for ℓ_2 signal estimation error associated with any (adaptive or non-adaptive) procedure, obtained a lower bound on the expected number of support errors associated with any sensing and testing procedure. In contrast, here we seek stronger guarantees on the probability of exact support recovery. Other related efforts include [10], which examined adaptive sensing techniques for exact support recovery using *non-compressive* measurements; [11], which identified sufficient conditions for support recovery of structured (tree-)sparse signals via adaptive compressive sensing; and [12], which examined an adaptive compressive sensing technique under a model where additive noise enters the signal itself, prior to each measurement.

A. Structured Random Measurement Matrices

Let $\mathbf{x} = [x_1 \ \ldots \ x_n]^T$. For the remainder of the paper assume $x_i \geq 0$, and let $\mathcal{S} = \{i : x_i \neq 0\}$. This non-negativity assumption can be easily removed, as discussed in [6], but makes the entire presentation much simpler. At each stage of our procedure, we utilize structured random matrices with r rows and n columns, where each column of the matrix has a single nonzero entry whose location is chosen uniformly at random with replacement from the set $\{1, 2, \ldots, r\}$. The amplitude of each nonzero entry is selected at random from the set $\{-\alpha, +\alpha\}$, where $\alpha > 0$ is a parameter that can be specified (amplitudes and locations are mutually independent). For shorthand, we denote random matrices $\mathbf{A} \in \mathbb{R}^{r \times n}$ generated in this fashion as being drawn from a distribution denoted by $\mathcal{A}(r, n, \alpha)$; in particular, we will write $\mathbf{A} \sim \mathcal{A}(r, n, \alpha)$ to denote an independent realization of such a structured random matrix. Similar sensing matrix constructions have been examined in the sketching literature; see, eg., [13], [14].

Algorithm 1: Sequential Compressed Sensing.

```
Input Parameters:
  $n, k, M$ and confidence parameter $\delta > 0$;
Internal Algorithm Parameters:
  $s = \lceil \log_2 \log n \rceil$ (number of refinement steps)
  $\tau_1 = \left\lceil \log \frac{2}{\delta} + \log k + \log \log_2 \log n \right\rceil$
      (number of sensing matrices per step in Stage 1)
  $\tau_2 = \left\lceil 32 \max \left\{ 9 \log \frac{8}{\delta} + 9 \log k \, , \, \log \frac{8}{\delta} + \log(n-k) \right\} \right\rceil$
      (number of sensing matrices in Stage 2)
  $r = 6k$ (number of rows in each sensing matrix)
  $\alpha = \sqrt{\frac{M}{6n\tau_1}}$ (magnitude of non-zero entries)
  $\mathcal{I}_1 = \{1, 2, \ldots, n\}$ (initial index set)

Stage 1: (Refinement)
  for $j = 1$ to $s$ do
    for $t = 1$ to $\tau_1$ do
      Generate matrix: $\mathbf{A}_t \sim \mathcal{A}(r, n, \alpha)$;
      Mask: Set $\ell$-th column of $\mathbf{A}_t$ to zero $\forall \ell \notin \mathcal{I}_j$;
      Collect measurements: $\mathbf{y}_t = \mathbf{A}_t \mathbf{x} + \mathbf{w}_t$;
      Backproject: $\widetilde{\mathbf{x}}_t = \mathbf{A}_t^T \mathbf{y}_t$;
    end
    Compute proxy $\widehat{\mathbf{x}}$ with entries $\widehat{x}_i = \sum_{t=1}^{\tau_1} \mathrm{sgn}(\widetilde{x}_{i,t})$;
    Refine search space: $\mathcal{I}_{j+1} = \{i \in \mathcal{I}_j \ : \ \widehat{x}_i > 0\}$;
  end
Stage 2: (Estimation)
  for $t = 1$ to $\tau_2$ do
    Generate matrix: $\mathbf{A}_t \sim \mathcal{A}(r, n, \alpha)$;
    Mask: Set $\ell$-th column of $\mathbf{A}_t$ to zero $\forall \ell \notin \mathcal{I}_{s+1}$;
    Collect measurements: $\mathbf{y}_t = \mathbf{A}_t \mathbf{x} + \mathbf{w}_t$;
    Backproject: $\widetilde{\mathbf{x}}_t = \mathbf{A}_t^T \mathbf{y}_t$;
  end
  Compute proxy $\widehat{\mathbf{x}}$ with entries $\widehat{x}_i = \sum_{t=1}^{\tau_2} \mathrm{sgn}(\widetilde{x}_{i,t})$;
  Final Support Estimate: $\widehat{\mathcal{S}} = \{i \in \mathcal{I}_j \ : \ \widehat{x}_i > 1/4\}$
```

B. Sensing and Refinement

Each measurement step in our SCS procedure will employ a *collection* of independent structured random sensing matrices generated as described above. We describe the measurements corresponding to each sensing matrix in the collection by $\mathbf{y}_t = \mathbf{A}_t \mathbf{x} + \mathbf{w}_t$, $t = 1, 2, \ldots, \tau$, where the $\mathbf{A}_t$ are matrices constructed as described above, but with some columns "masked" to zero. Here, t indexes the elements of the collection, and there are a total of τ such elements. The aim of the refinement step is to quickly identify locations of $\mathbf{x}$ which are no longer deemed worthy of further investigation, and to adapt the subsequent sensing matrices so that no sensing energy is expended measuring those locations.

Measurements obtained in each stage are used to form a set of *estimates* of the form $\widetilde{\mathbf{x}}_t = \mathbf{A}_t^T \mathbf{y}_t$ for $t = 1, \ldots, \tau$. The estimates are combined to form a *signal proxy* $\widehat{\mathbf{x}}$ with entries $\widehat{\mathbf{x}}_i = \sum_{t=1}^{\tau} \mathrm{sgn}(\widetilde{\mathbf{x}}_{i,t})$. Here the notation $\widetilde{\mathbf{x}}_{i,t}$ is shorthand for the i-th entry of the vector $\widetilde{\mathbf{x}}_t$, and $\operatorname{sgn}(\cdot)$ denotes the signum function which takes the value -1 $(+1)$ when the argument is less than (greater than) zero, and is zero when the argument is zero.

In the initial measurement step all locations of $\mathbf{x}$ are observed (equivalently, the sensing matrices have nonzero entries in all columns). Each refinement step then identifies a set of locations at which measurements will be obtained in the next measurement step, simply by identifying the entries of the current proxy $\widehat{\mathbf{x}}$ that exceed zero. The rationale is, as we will show below, that the signs of the entries of the proxy are very likely to match the corresponding signs of the unknown $\mathbf{x}$ at locations in the support set, and the signs of the entries of the proxy are equally likely off of the signal support. Since we have assumed that the nonzero entries of $\mathbf{x}$ are all non-negative, identifying the set of locations at which the proxy is non-negative is therefore likely to result in a refined set containing *all* of the locations of the true signal components, and about *half* of the locations off of the support. This process is iterated several times, resulting in a sequence of refinements of the search space, which then hone in on the true signal support.

IV. Proof of Main Result

The proof proceeds in a constructive way, and provides a justification for the internal parameter choices, namely s, r, α, τ_1 and τ_2. Note that the s must be an integer, and τ_1 and τ_2 must be odd integers. For simplicity, in the analysis below we don't consider these constraints, as this doesn't qualitatively change the results. Let $\mathcal{S} = \{i : x_i \neq 0\}$ and $\mathcal{S}^c$ denote the true support set and its complement, respectively. Let $x_{\min} = \min_{i \in \mathcal{S}} |x_i|$. The following lemma characterizes the proxy obtained at each step.

Lemma 1. *The following are true about the entries of the proxy $\widehat{x}_j$ formed in the j-th step:*

1) $\Pr(\widehat{x}_{j,i} > 0) = \Pr(\widehat{x}_{i,j} < 0) = 1/2$ *for any $i \in \mathcal{S}_j^c$, and*
2) *for any $i \in \mathcal{S}_j$,*

$$\Pr(\widehat{x}_{j,i} < 0) \;=\; \Pr(\widehat{x}_{i,j} \leq 0) \leq \exp\left(-\frac{\tau(1-2p)^2}{4}\right)$$

provided $p < 1/2$, where

$$p \triangleq \frac{1}{2}\exp\left(-\frac{\alpha^2 x_{\min}^2}{2\sigma^2}\right) + \left(\frac{k-1}{r}\right) . \tag{2}$$

To ensure the lemma is applicable we need to guarantee condition (2), which involves both r and α. The proposed choice of r guarantees that the second term on the r.h.s. of (2) is less than $1/6$. Further, if

$$x_{\min} \geq \frac{\sqrt{2\sigma^2 \log 3}}{\alpha} = \Omega\left(\sqrt{\frac{n}{M}\sigma^2(\log k + \log\log_2 \log n)}\right) ,$$

the first term in the r.h.s. of (2) is also bounded by $1/6$, so we conclude that $p < 1/3 < 1/2$.

The next step is to ensure that, after Stage 1, we have not lost any signal components. This means we must control the probability of the event $\{\mathcal{S} \nsubseteq \mathcal{I}_s\}$. This is easily done via a union bound

$$\begin{aligned} P(\mathcal{S} \nsubseteq \mathcal{I}_s) &= P(\cup_{i\in\mathcal{S}} \cup_{j=1}^{s} \{\widehat{x}_{i,j} < 0\}) \\ &\leq ks \exp\left(-\frac{\tau_1(1-2p)^2}{4}\right), \end{aligned}$$

using the result of Lemma 1. Given $p < 1/3$ and our choice of τ_1 we conclude, after some trivial algebra, that $P(\mathcal{S} \nsubseteq \mathcal{I}_s) \leq \delta/2$. This means that, with probability $1 - \delta/2$, all the signal components are still considered in Stage 2. Denote this event as Ω_1.

Now we show that Stage 2 exactly recovers the support set. We use of the following lemma from [15], adapted to our context.

Lemma 2. *Suppose that p in (2) satisfies $p < 1/2$. Choose the final threshold value T to satisfy $T < 1 - 2p$ and let $\delta > 0$ be arbitrary. Conditioned on the event Ω_1, the choice*

$$\tau_2 \geq \max\left\{\frac{2}{(T-(1-2p))^2}\log\left(\frac{8k}{\delta}\right), \frac{2}{T^2}\log\left(\frac{8(n-k)}{\delta}\right)\right\}$$

where τ_2 is an odd integer is sufficient to ensure that $\widehat{\mathcal{S}} = \mathcal{S}$ with probability at least $1 - \delta/2$.

Since we already showed that $p < 1/3$, thresholding at $T = 1/4$ satisfies the condition of the lemma. Furthermore the choice τ_2 we took follows immediately from the lemma statement. Finally, since $P(\Omega_1) \geq 1 - \delta/2$, and that $P(\widehat{\mathcal{S}} = \mathcal{S} | \Omega_1) \geq 1 - \delta/2$ we conclude that the support set estimator $\widehat{\mathcal{S}}$ generated by Algorithm 1 is equal to $\mathcal{S}$ with probability at least $1 - \delta$.

Now that we ensured the algorithm recovers the exact signal support with high probability, we will quantify the number of measurements it requires and the total sensing budget. First, the total number of measurements is easily computed as follows. The algorithm makes $rs\tau_1$ measurements in Stage 1 and $r\tau_2$ measurements in the Stage 2. Therefore, the total number of measurements is $m = rs\tau_1 + r\tau_2 \;\leq\; C\,k\left((\log k + \log\log_2 \log n + \log \delta^{-1})\log_2 \log n + \log n\right)$ where $C > 0$ is a constant (e.g., $C = 1728$ will suffice). Note that $m = O\left(k \max\{\log k \log\log n\, ,\, \log n\}\right)$. Next consider the expected total sensing energy used by the algorithm. To this end it is important to characterize the expected problem size in Stage 2. We first characterize the problem size after each step in Stage 1, given by $|\mathcal{I}_j|$. Making use of the first part of Lemma 1,

$$\begin{aligned} \mathbb{E}|\mathcal{I}_j| &= \mathbb{E}|\mathcal{I}_j \cap \mathcal{S}^c| + \mathbb{E}|\mathcal{I}_j \cap \mathcal{S}| \\ &\leq \mathbb{E}|\mathcal{I}_j \cap \mathcal{S}^c| + k \\ &= \sum_{i\in\mathcal{S}^c} \Pr\left(\bigcap_{\ell=1}^{j-1} \widehat{x}_{i,\ell} > 0\right) \quad + k \\ &\leq (n-k)\max_{i\in\mathcal{S}^c} \Pr\left(\bigcap_{\ell=1}^{j-1} \widehat{x}_{i,\ell} > 0\right) \quad + k, \end{aligned}$$

from which it follows that $\mathbb{E}|\mathcal{I}_j| \leq \frac{n-k}{2^{j-1}} + k$. Using this result we can easily evaluate the total sensing energy. Note that, for $j = 1, \ldots, s$ the sensing matrix at each step satisfies

$$\mathbb{E}\|\mathbf{A}_t\|_F^2 \;=\; \alpha^2 \mathbb{E}|\mathcal{I}_j| \;\leq\; \alpha^2 \frac{n-k}{2^{j-1}} \quad + k \;.$$

This means the expected total sensing energy used in Stage 1, which we will denote by E_1, is bounded by

$$E_1 \;\leq\; \alpha^2 \tau_1 \;(2(n-k) + sk) \;\leq\; \frac{M}{3} + \frac{Mk\log_2 \log n}{6n},$$

where the last inequality follows by inserting the specified values of α and τ_1. It follows that if $k \leq n/\log_2 \log n$, then $E_1 \leq M/2$. For Stage 2, note that each sensing matrix satisfies

$$\mathbb{E}\|\mathbf{A}_t\|_F^2 \;=\; \alpha^2 \mathbb{E}|\mathcal{I}_{s+1}| \;\leq\; \alpha^2\left(\frac{n-k}{2^s} + k\right) \;.$$

Therefore the total expected energy used in Stage 2, denoted by E_2, is bounded by τ_2 times the quantity above. That is,

$$\begin{aligned} E_2 &\leq \tau_2 \alpha^2\left(\frac{n-k}{2^s} + k\right) \\ &\leq 33\alpha^2 \log n\left(\frac{n-k}{\log n} + k\right) \\ &\leq 33\frac{M}{6n\tau_1}\,(n - k + k\log n) \\ &\leq M/2 \;, \end{aligned}$$

provided n is sufficiently large and $k \leq n/\log_2 \log n$. In the second inequality we made use of the fact that $\tau_2 < 33 \log n$ for large enough n. Taken together, these imply the budget constraint is satisfied.

V. Acknowledgements

This work was supported in part by grant DARPA/ONR N66001-10-1-4090.

Appendix

A. Proof of Lemma 1

For ease of notation, we drop the subscripts j on the sets $\mathcal{S}_j$ and $\mathcal{S}_j^c$ here, with the implicit understanding that we're considering the refinement in a particular step. Further, to better describe the structured sparse sensing matrices we employ in our procedure, it is useful to introduce the following random variables: take $(j,t) \in \{1,\ldots,n\} \times \{1,\ldots,\tau\}$ and let $Z_{j,t}$ be i.i.d. Rademacher random variables and $H_{j,t}$ be i.i.d. discrete uniform random variables over $\{1,\ldots,r\}$, also independent from the Rademacher random variables. The jth column of matrix A_t has a single non-zero element in row $H_{j,t}$, taking the value $\alpha Z_{j,t}$, where $\alpha > 0$. Formally, $[A_t]_{ij} = \alpha Z_{j,t} \mathbf{1}\{i = H_{j,t}\}$.

To describe the observation process we introduce the "noise" variables $W_{i,t}$, which are i.i.d. normal random variables with zero mean and variance $\sigma^2 = 1$. These are also independent from the other random variables introduced so far. The projective observations are τ vectors of the form

$$Y_t = A_t x + W_t \quad t \in \{1,\ldots,\tau\} \ .$$

The entries of each vector can also be written as

$$Y_{i,t} = \sum_{j:H_{j,t}=i} \alpha Z_{j,t} x_j + W_{i,t} = \sum_{j \in \mathcal{S}:H_{j,t}=i} \alpha Z_{j,t} x_j + W_{i,t} \ .$$

Following the proposed algorithm we compute the proxy vectors, which after simple algebraic manipulation can be shown to be

$$\tilde{x}_{i,t} = \alpha^2 x_i + \alpha^2 Z_{i,t} \sum_{\ell \in \mathcal{S}:\ell \neq i} Z_{\ell,t} x_\ell \mathbf{1}\{H_{\ell,t} = H_{i,t}\} + \alpha Z_{i,t} W_{H_{i,t},t} \ . \tag{3}$$

Finally, in our algorithm we take $\hat{x}_i = \sum_{t=1}^{\tau} \operatorname{sgn}(\tilde{x}_{i,t})$, and these are the main quantities we want to characterize. We consider two cases.

Case 1): Let $i \notin \mathcal{S}$. Note that, since $x_i = 0$ we have

$$\tilde{x}_{i,t} = \alpha Z_{i,t} \left(W_{H_{i,t},t} + \sum_{\ell \in \mathcal{S}:\ell \neq i} \alpha Z_{\ell,t} x_\ell \mathbf{1}\{H_{\ell,t} = H_{i,t}\} \right) \ .$$

Note this is the product of two independent random variables, since the terms in the summation depend only on $Z_{\ell,t}$ with $\ell \neq i$. Furthermore one easily can verify that each term in this product is a symmetric random variable around zero (this can be easily done using a conditioning argument). In addition, note that with probability one each $\tilde{x}_{i,t}$ is different than zero, so the $\operatorname{sgn}(\tilde{x}_{i,t}) = 0$ with zero probability. These two facts together immediately show that $\operatorname{sgn}(\tilde{x}_{i,t})$ are Rademacher random variables. Now clearly $\hat{x}_i = \sum_{t=1}^{\tau} \operatorname{sgn}(\tilde{x}_{i,t})$ are also symmetric zero mean random variables. Since τ is assumed odd it is immediate that $\hat{x}_i$ cannot take the value zero, and therefore the first result of the lemma follows.

Case 2): Let $i \in \mathcal{S}$. Recall equation (3). Notice that the middle term in that equation is the one that complicates the analysis. However, this term is actually zero with significant probability. Define the event $\Omega_{i,t} = \{\forall \ell \in \mathcal{S}, \quad H_{\ell,t} \neq H_{i,t}\}$. It is elementary to show that

$$P(\Omega_{i,t}) = \left(1 - \frac{1}{r}\right)^{|\mathcal{S}|-1} \geq 1 - \frac{|\mathcal{S}|-1}{r} \ .$$

We proceed by noting that the middle term of (3) is zero under $\Omega_{i,t}$, and therefore

$$\begin{aligned} P(\tilde{x}_{i,t} < 0 | \Omega_{i,t}) &= P(\alpha^2 x_i \ + \ \alpha Z_{i,t} W_{H_{i,t},t} \ < \ 0 \mid \Omega_{i,t}) \\ &= P(\alpha x_i \ + \ W_{H_{i,t},t} \ < \ 0 \mid \Omega_{i,t}) \\ &\leq \frac{1}{2} \exp\left(-\frac{\alpha^2 x_i^2}{2\sigma^2}\right) \ , \end{aligned}$$

where the final step make use of standard Gaussian tail bounds. By removing the conditioning on $\Omega_{i,t}$ we conclude that

$$P(\tilde{x}_{i,t} < 0) \leq \frac{1}{2} \exp\left(-\frac{\alpha^2 x_i^2}{2\sigma^2}\right) \ + \ \frac{|\mathcal{S}|-1}{r} \ .$$

We now focus our attention on each individual value of $\hat{x}_i$, for a fixed but arbitrary $i \in \mathcal{S}$. By construction $\{\tilde{x}_{i,t}\}_{t=1}^{\tau}$ are i.i.d., therefore we can use Hoeffding's inequality to characterize $\hat{x}_i$. Let $p_i = P(\tilde{x}_{i,t} < 0)$ and assume for now that $p_i < 1/2$. Then

$$\begin{aligned} P(\hat{x}_i < 0) &= P(\sum_{t=1}^{\tau} \operatorname{sgn}(\tilde{x}_{i,t}) < 0) \\ &= P(\sum_{t=1}^{\tau} \operatorname{sgn}(\tilde{x}_{i,t}) - \tau(1-2p_i) < -\tau(1-2p_i)) \\ &\leq \exp\left(-\frac{\tau(1-2p_i)^2}{2\sigma^2}\right) \ . \end{aligned}$$

Finally, recall the expression (2) and note that $p_i \leq p$ for all $i \in \mathcal{S}$.

References

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

[2] E. Candes and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[3] Martin J. Wainwright, "Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (lasso)," *IEEE Trans. Inform. Theory*, vol. 55, no. 5, pp. 2183–2202, 2009.

[4] C. Genovese, J. Jin, and L. Wasserman, "Revisiting marginal regression," *Manuscript*, Nov. 2009.

[5] S. Aeron, V. Saligrama, and M. Zhao, "Information theoretic bounds for compressed sensing," *IEEE Trans. Inform. Theory*, vol. 56, no. 10, pp. 5111–5130, 2010.

[6] J. Haupt, R. Castro, and R. Nowak, "Distilled sensing: Adaptive sampling for sparse detection and estimation," *IEEE Trans. Inform. Theory*, vol. 57, no. 9, pp. 6222–6235, 2011.

[7] J. Haupt, R. Baraniuk, R. Castro, and R. Nowak, "Compressive distilled sensing: Sparse recovery using adaptivity in compressive measurements," in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, Nov 2009, pp. 1551–1555.

[8] P. Indyk, E. Price, and D.P. Woodruff, "On the power of adaptivity in sparse recovery," in *Proc. IEEE Foundations of Computer Science (FOCS)*, Oct. 2011, pp. 285–294.

[9] E. Arias-Castro, E. J. Candes, and M. Davenport, "On the fundamental limits of adaptive sensing," *Manuscript*, Nov. 2011.

[10] M. Malloy and R. Nowak, "Sequential analysis in high dimensional multiple testing and sparse recovery," in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, 2011, pp. 2661–2665.

[11] A. Soni and J. Haupt, "Efficient adaptive compressive sensing using sparse hierarchical learned dictionaries," in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, 2011, pp. 1250–1254.

[12] M. Iwen and A. Tewfik, "Adaptive group testing strategies for target detection and localization in noisy environments," *IEEE Trans. Signal Proc.*, vol. 60, no. 5, pp. 2344–2353, 2012.

[13] S. Muthukrishnan, *Data streams: Algorithms and applications*, Now Publishers, 2005.

[14] P. Indyk and A. Gilbert, "Sparse recovery using sparse matrices," *Proc. IEEE*, vol. 98, no. 6, pp. 937–947, 2010.

[15] J. Haupt and R. Baraniuk, "Robust support recovery using sparse compressive sensing matrices," in *Proc. 45th Annual Conf. on Information Sciences and Systems*, Baltimore, MD, March 2011, pp. 1–6.