Locating Salient Items in Large Data Collections With Compressive Linear Measurements

Jarvis Haupt
Dept. of Electrical and Computer Engineering
University of Minnesota, Minneapolis MN
Email: jdhaupt@umn.edu

***Abstract*—Recent advances in compressive sensing (CS) have established that high-dimensional signals that possess sparse representations in some basis or dictionary can be accurately recovered from relatively few linear measurements. As a result, CS strategies have been proposed and developed in a number of application domains where sensing resource efficiency is of primary importance. This paper examines a class of compressive anomaly detection tasks, where the aim is to identify the locations of a nominally small number of outliers in a large collection of data (which may be scalar or multivariate) using a small number of observations of the form of linear combinations of subsets of the data. We introduce a generalized notion of sparsity termed here as saliency, and establish that a novel sensing and inference technique called Compressive Saliency Sensing (CSS), comprised of a randomized linear sensing strategy and associated computationally efficient inference procedure based on techniques from group testing, can accurately identify the locations of k outliers in a collection of n items from only $m = O(k \log n)$ linear measurements. We describe several inference tasks to which our approach is suited, including "traditional" k-sparse support recovery problems; identification of k outliers in the "simple" signal model of Donoho and Tanner, characterized by nominally binary vectors having k entries strictly in $(0, 1)$; and identification of vectors that are outliers from a common (low-dimensional) linear subspace.**

I. Introduction

Let $\mathbf{X} = \{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ denote a collection of n individual data elements, where for a field $\mathbb{F}$ and integer $p \geq 1$ we have that $\mathbf{X}_j \in \mathbb{F}^p$ for all j; for concreteness, in what follows we take $\mathbb{F} = \mathbb{R}$. We suppose that a small number of the n data elements of $\mathbf{X}$ are "outliers," in the sense that they exhibit characteristics that differ from those of the bulk of the data (in a specific manner to be described below). Our overall aim is to identify these data outliers, using only a small number of measurements, in the form of linear combinations of elements of $\mathbf{X}$. Our approach is motivated by, and can be viewed as an extension of, compressive sensing (CS) techniques [1], [2], which leverage sparsity for inference tasks performed on undersampled data.

We introduce a generalized notion of sparsity that we refer to as *saliency* – a description chosen to embody the notion that the data outliers are deemed so merely by virtue of their deviation from the characteristics exhibited by the bulk of the data – which may be formalized as follows. Suppose that most of the data elements can be described as elements of a set $\mathcal{Z}$, which we refer to as the "common set." The outliers of interest, then, are elements of $\mathbf{X}$ not belonging to $\mathcal{Z}$. In an analogous manner to the notion of signal support in traditional sparse data models, we define the *salient support* of $\mathbf{X}$ in terms of the set $\mathcal{Z}$, as

$$\mathrm{salsupp}_{\mathcal{Z}}(\mathbf{X}) \triangleq \{\ j\ :\ \mathbf{X}_j \notin \mathcal{Z}\ \}, \tag{1}$$

and we say that $\mathbf{X}$ is *k-salient* with respect to the set $\mathcal{Z}$ when $|\mathrm{salsupp}_{\mathcal{Z}}(\mathbf{X})| = k$. In what follows, we will often simply refer to $\mathbf{X}$ as k-salient when the set $\mathcal{Z}$ is clear from context, and in this case we write the salient support of $\mathbf{X}$ simply as $\mathrm{salsupp}(\mathbf{X})$.

J.H. graciously acknowledges the support of NSF Grant CCF-1217751.

Informally, we may interpret $\mathcal{Z}$ as a (possibly infinite and uncountable) set of possible "uninteresting" elements to which a large number of the n individual elements of $\mathbf{X}$ belong. Traditional sparsity models, for example, may be described in terms of a (trivial) common set $\mathcal{Z} = \{0\}$, and in this case nonzero elements of $\mathbf{X}$ are (by contrast) the "interesting" elements. In this sense, our definition of saliency necessarily encompasses the traditional notion of sparsity as a special case, but also provides a more generalized notion of "data sparsity."

A. Problem Statement

Our problem of interest here amounts to a generalized support recovery task – we aim to estimate the salient support $\mathrm{salsupp}_{\mathcal{Z}}(\mathbf{X})$ of $\mathbf{X}$ from compressive measurements. To that end, we suppose that a total of m observations of $\mathbf{X}$ may be obtained, each of which is a linear combination of the individual elements $\{\mathbf{X}_j\}_{j=1}^n$, where the scalar coefficients associated with each linear combination are all elements of some specified coefficient set $\mathcal{C} \subseteq \mathbb{R}$. Formally, we obtain observations $\mathbf{Y}_i$ of the form

$$\mathbf{Y}_i = \sum_{j=1}^{n} A_{i,j}\ \mathbf{X}_j, \quad \text{for } \ i = 1, 2, \ldots, m, \tag{2}$$

where $A_{i,j} \in \mathcal{C}$ for all $i = 1, \ldots, m$ and $j = 1, \ldots, n$. We seek an estimate $\widehat{\mathcal{S}} = \widehat{\mathcal{S}}_{\mathcal{Z}}\left(\{\mathbf{Y}_i\}_{i=1}^m, \{A_{i,j}\}_{(i,j)\in\{1,\ldots,m\}\times\{1,\ldots,n\}}\right)$ that is an accurate estimate of the true salient support $\mathrm{salsupp}(\mathbf{X})$.

B. Assumptions

We introduce two assumptions on the common set $\mathcal{Z}$ and the coefficient set $\mathcal{C}$ that, when satisfied, provide sufficient conditions under which our inference approach described in the following section will succeed in identifying the salient support of data $\mathbf{X}$ from compressive measurements. Specifically, we will be interested here in settings where the pair $\{\mathcal{Z}, \mathcal{C}\}$ satisfy the following two assumptions:

Assumption A 1 (Restricted Closure Under Addition). *For any finite integer $L \in \mathbb{N}$, if $c_\ell \in \mathcal{C}$ and $\mathbf{X}_\ell \in \mathcal{Z}$ for all $\ell = 1, 2, \ldots, L$, then $\sum_{\ell=1}^{L} c_\ell \mathbf{X}_\ell \in \mathcal{Z}$.*

Assumption A 2 (Single Outlier Identifiability). *For any finite integer $L \in \mathbb{N}$, if $c_\ell \in \mathcal{C}$ and $z_\ell \in \mathcal{Z}$ for all $\ell = 1, 2, \ldots, L-1$, but $z_L \notin \mathcal{Z}$ and $c_L \in \mathcal{C} \setminus \{0\}$, then $\sum_{\ell=1}^{L-1} c_\ell \mathbf{X}_\ell + c_L \mathbf{X}_L \notin \mathcal{Z}$.*

In words, assumption **A1** specifies that the set $\mathcal{Z}$ is closed under certain (but *not necessarily all*) linear combinations – precisely we require closure only under linear combinations where the coefficients of each element in the combination are elements of $\mathcal{C}$. Assumption **A2** specifies that any linear combination comprised of elements of $\mathcal{Z}$ (with coefficients in $\mathcal{C}$) that also contains exactly one element that is *not in* $\mathcal{Z}$ (scaled by a nonzero coefficient from $\mathcal{C}$), will not belong to $\mathcal{Z}$. For the sake of illustration, we identify below several examples of pairs $\{\mathcal{Z}, \mathcal{C}\}$ that satisfy the assumptions **A1-A2** above, and describe each in terms of an associated salient support recovery task.

978-1-4673-3146-3/13/$31.00 ©2013IEEE

1) Sparse Support Recovery: As noted above, the conventional sparsity model may be described in terms of the common set $\mathcal{Z} = \{0\}$. Clearly, assumption **A1** holds here, for example, when the coefficient set $\mathcal{C} = \mathbb{R}$ (many others choices of $\mathcal{C}$ are also valid for this particular $\mathcal{Z}$). Further, in this case assumption **A2** also holds, since any linear combination comprised of a number of "zeros" and one "nonzero" element (with nonzero coefficient) will itself be nonzero. In this example, the salient support recovery task is traditional support recovery for k-sparse vectors.

2) Identifying Outliers in "Simple" Signals: In the context of a body of work aimed at identifying necessary and sufficient conditions for sparse recovery using convex programming methods, Donoho and Tanner introduced in [3] the notion of k-simple signals, which are nominally binary vectors having $n-k$ entries that are elements of the set $\{0, 1\}$ and k entries that are strictly in $(0, 1)$. Here, assumptions **A1-A2** hold if we choose $\mathcal{Z} = \mathbb{Z}$, the set of all real integers, and let $\mathcal{C} = \{0, 1\}$ (again, other choices of $\mathcal{C}$ are possible, including $\mathcal{C} = \mathbb{Z}$). The salient support recovery task here amounts to identifying the locations of the k non-binary elements in a k-simple signal.

3) Finding Vector Outliers from a Linear Subspace: Consider a collection of linearly independent vectors $\mathbf{u}_i \in \mathbb{R}^p$, $i = 1, \ldots, d$. Collectively, the vectors span a d-dimensional linear subspace of $\mathbb{R}^p$; for shorthand, let us denote $\mathcal{U} \triangleq \text{span}\,(\mathbf{u}_1, \ldots, \mathbf{u}_d) = \{\mathbf{v} \in \mathbb{R}^p : \mathbf{v} = \sum_{\ell=1}^{d} \alpha_\ell \mathbf{u}_\ell, \alpha_\ell \in \mathbb{R}\ \forall\ \ell\}$. In this case, assumptions **A1-A2** hold for the choice $\mathcal{Z} = \mathcal{U}$ and $\mathcal{C} = \mathbb{R}$, provided that $\mathcal{U} \neq \mathbb{R}^p$ (i.e., the subspace $\mathcal{U}$ must be a proper subspace of $\mathbb{R}^p$). This follows from the fact that linear subspaces are closed under (all) linear combinations, while sums containing one vector having a component outside of the d-dimensional subspace will itself be outside of the subspace. Here, the salient support recovery task corresponds to identifying the k (vector) outliers from the subspace $\mathcal{U}$.

C. Our Contribution

Our main contribution here is to establish that, for data models characterized by pairs $\{\mathcal{Z}, \mathcal{C}\}$ satisfying assumptions **A1-A2** above, a computationally-efficient sensing and inference approach called "Compressive Saliency Sensing" (CSS) is a provably accurate method for salient support identification from compressive measurements. In the following section we describe the CSS procedure, and we state and prove our main theoretical result quantifying its performance. A few conclusions are briefly discussed in Section III.

D. Relation to Existing Works

The idea of using sparse measurement strategies for sparse inference is related to sketching notions that are, by now, well studied in the computer science literature – see, for example, [4], [5]. A number of recent works have examined connections between sparse sampling, group testing, and sketching ideas, and sparse signal recovery and compressive sensing tasks including, for example, [6]–[9]. The essential focus of our effort here amounts to the analysis of a noisy group testing task, and we note that previous efforts have examined various aspects of noisy group testing problems [10], and proposed and analyzed efficient procedures – see, e.g., [11], [12].

We also note the recent work [13] which examines a related problem to what we examine here – that of identifying the salient elements in a set of variables, features, or covariates – and establishes quantitatively similar results to ours for certain sparse recovery tasks, under a model where data are assumed to be random (iid) quantities and saliency is quantified by conditional independence conditions.

E. A Note on Notation

In what follows we employ a "MATLAB-inspired" notation to denote row and column vectors of a given matrix; namely, for a matrix $\mathbf{M}$, we denote by $\mathbf{M}_{i,:}$ its i-th row, and by $\mathbf{M}_{:,j}$ its j-th column. We define the support of a vector to be the set of locations at which the vector takes nonzero values; e.g., $\text{supp}(\mathbf{M}_{:,j}) \triangleq \{i : M_{i,j} \neq 0\}$.

II. Compressive Saliency Sensing

The essential idea underlying our "Compressive Saliency Sensing" (CSS) approach is to "map" the salient support inference problem to a group testing proxy task, as follows. We associate to the compressive observations $\mathbf{Y} = \{\mathbf{Y}_1, \ldots, \mathbf{Y}_m\}$ a binary vector $\mathbf{y} \in \{0, 1\}^m$, whose i-th element takes the value 1 when $\mathbf{Y}_i \notin \mathcal{Z}$ and is zero otherwise, for $i = 1, 2, \ldots, m$; that is, we let $\mathbf{y}_i = \mathbf{1}_{\{\mathbf{Y}_i \notin \mathcal{Z}\}}$, where $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function of its argument. Likewise, to the collection of coefficients $\{A_{i,j}\}$ we associate a binary matrix $\mathbf{M} \in \{0, 1\}^m$ having elements $M_{i,j} = \mathbf{1}_{\{A_{i,j} \neq 0\}}$ for $i = 1, \ldots, m$ and $j = 1, \ldots, n$. Now, if we interpret $\mathbf{x} \in \{0, 1\}^n$ as a vector whose j-th element is 1 if and only if $\mathbf{X}_j$ is in the salient support of $\mathbf{X}$ – that is, $\mathbf{x}_j = \mathbf{1}_{\{\mathbf{X}_j \notin \mathcal{Z}\}}$ – then our task of identifying the salient support of $\mathbf{X}$ reduces to the task of identifying the support of the (binary) vector $\mathbf{x}$ from the binary observations $\mathbf{y}$ and matrix $\mathbf{M}$. The overall procedure is depicted as Algorithm 1.

Our approach is simple in principle, though its analysis requires treatment of subtle issues that arise when mapping the problem to the binary proxy task. First, it is evident that a significant number of the elements in the collection of coefficients $\{A_{i,j}\}$ for $i = 1, \ldots, m$ and $j = 1, \ldots, n$ should be equal to zero for this approach; if not, the associated matrix $\mathbf{M}$ in the group testing proxy task would contain mostly 1's, and group testing strategies applied in this case would be uninformative. A second (more subtle) point concerns the construction of the proxy vector $\mathbf{y}$. Even when the coefficients $\{A_{i,j}\}$ are chosen to yield an appropriately sparse matrix $\mathbf{M}$, we still need to ensure that the individual elements of $\mathbf{y}$ are (mostly) accurate. Informally speaking, we must be able to identify, with some level of certainty and from the observed data itself, when a linear combination of the data elements $\{\mathbf{X}_j\}_{j=1}^n$ does or does not contain elements of $\mathbf{X}$ not belonging to the common set $\mathcal{Z}$. For this, we invoke assumptions **A1-A2**, which together imply that the proxy task can be treated as a *noisy* group testing problem, albeit with a non-standard form of noise, from the perspective of previous group testing investigations.

A. Theoretical Guarantees

Our main result quantifies the performance of the CSS procedure described in Algorithm 1.

Theorem 1. *Let $\{\mathcal{Z}, \mathcal{C}\}$ satisfy assumptions **A1-A2**, and let $\mathbf{X}$ be k-salient with respect to $\mathcal{Z}$. Choose $\delta \in (0, 1)$, $c, k' > 0$ and $\alpha \in (0, 1/5]$ as fixed parameters. Collect $m = \lceil ck' \log(n/\delta) \rceil$ randomized measurements of $\mathbf{X}$ via the CSS procedure of Algorithm 1, and form the support estimate $\widehat{\mathcal{S}}$, using threshold $\tau = m\alpha(1-2\alpha)/k'$. If k' is an upper bound for the true sparsity level (i.e., $k' \geq k$) and the factor c satisfies $c \geq (2/\alpha^2) \cdot \max\{k'/k, 1/\alpha\}$, then the salient support estimate $\widehat{\mathcal{S}}$ satisfies $Pr\left(\widehat{\mathcal{S}} \neq \text{salsupp}(\mathbf{X})\right) \leq 2\delta$, where the probability is with respect to the randomness of the sensing strategy.*

Note that when the parameter k' does not overestimate the true saliency level k by too much, so that $k' \leq \beta k$ for some constant $\beta \geq 1$, it follows that the salient support of $\mathbf{X}$ can be accurately

Algorithm 1 Compressive Saliency Sensing

Input:

Common set: $\mathcal{Z}$; Coefficient set: $\mathcal{C}$
Error tolerance parameter: $\delta \in (0,1)$
Saliency level parameter: $k' \in \mathbb{N}$
Oversampling factor: $c > 0$
Threshold parameter: $\alpha \in (0, 1/5]$

Initialize:

Sensing matrix sparsity parameter: $q = \alpha/k'$
Number of measurements: $m = \lceil ck' \log(n/\delta) \rceil$
Threshold: $\tau = m\alpha(1-2\alpha)/k'$;

Collect Compressive Observations:

Set $A_{i,j} = \epsilon_{i,j} c_{i,j}$ where $c_{i,j} \in \mathcal{C}$ and $\epsilon_{i,j} \overset{\text{iid}}{\sim} \text{Bernoulli}(q)$, for $i = 1, \ldots, m$ and $j = 1, \ldots, n$
Collect randomized linear observations $\mathbf{Y}_i = \sum_{j=1}^n A_{i,j} \mathbf{X}_j$, for $i = 1, \ldots, m$

Form Binary Proxies:

Compute $\mathbf{M} \in \{0,1\}^{m \times n}$, where $M_{i,j} = \mathbf{1}_{\{A_{i,j} \neq 0\}}$, for $i = 1, \ldots, m$ and $j = 1, \ldots, n$, and
$\mathbf{y} \in \{0,1\}^m$ with $y_i = \mathbf{1}_{\{\mathbf{Y}_i \notin \mathcal{Z}\}}$ for $i = 1, \ldots, m$

Perform Support Estimation:

Let $\widehat{\mathcal{S}}_\tau = \{j \in \{1, \ldots, n\} : |\text{supp}(\mathbf{M}_{:,j}) \setminus \text{supp}(\mathbf{y})| \leq \tau\}$

identified from only $m = O(k \log n)$ compressive measurements. This is intuitively pleasing, as in this sense our result recovers essentially the same sample complexity identified for a host of related *sparse* inference tasks in CS (see, e.g., [1], [2]). The essential contribution here is in the fact that our result applies also to a broader class of outlier identification problems characterized by pairs $\{\mathcal{Z}, \mathcal{C}\}$ satisfying assumptions **A1-A2**, as described above.[1]

B. Proof of Main Result

By DeMorgan's Laws and the union bound, we have that

$$\Pr\left(\widehat{\mathcal{S}} \neq \text{salsupp}(\mathbf{X})\right) \leq \sum_{j \in \text{salsupp}(\mathbf{X})} \Pr\left(|\text{supp}(\mathbf{M}_{:,j}) \setminus \text{supp}(\mathbf{y})| > \tau\right) + \sum_{j \notin \text{salsupp}(\mathbf{X})} \Pr\left(|\text{supp}(\mathbf{M}_{:,j}) \setminus \text{supp}(\mathbf{y})| \leq \tau\right). \quad (3)$$

Our proof proceeds by considering each of the two sums on the right-hand side of (3) in turn. We note that our analysis here is related to, and in some sense motivated by, the approach employed in a related effort [11] that examined noisy group testing problem under a different error model than we consider here.

To clarify the exposition that follows, we will find it useful to first introduce the quantity $\mathbf{y}^*$, which we interpret as an ideal or "error-free" version of the binary vector $\mathbf{y}$. Formally, for $\mathbf{x} \in \{0,1\}^n$ having elements $x_j = \mathbf{1}_{\{\mathbf{X}_j \notin \mathcal{Z}\}}$, which are nonzero only at locations $j \in \text{salsupp}(\mathbf{X})$, and $\mathbf{M} \in \{0,1\}^{m \times n}$, we introduce the vector $\mathbf{y}^* \in \{0,1\}^m$ whose elements are given by $y_i^* = \bigvee_{j=1}^n M_{i,j} \wedge x_j$, for $i = 1, 2, \ldots, m$, where the symbols $\wedge$ and $\vee$ denote, respectively, the Boolean "and" and "or" operators. In words, an element y_i^* of $\mathbf{y}^*$ is equal to 1 if and only if $M_{i,j} = 1$ for at least one $j \in \text{salsupp}(\mathbf{X})$. Note that this definition is equivalent to the statement

$$\text{supp}(\mathbf{y}^*) = \bigcup_{j \in \text{salsupp}(\mathbf{X})} \text{supp}(\mathbf{M}_{:,j}). \quad (4)$$

Now, assumption **A1** ensures that the (potentially noisy) observation vector $\mathbf{y}$ does not contain "false positives," but assumption **A2** is by itself not strong enough to rule out "false negatives." Together, these imply $\text{supp}(\mathbf{y}) \subseteq \text{supp}(\mathbf{y}^*)$. With this, we proceed with the proof.

1) Part 1: Support error bounds for indices $j \in \text{salsupp}(\mathbf{X})$: We first consider indices in the true salient support of $\mathbf{X}$, and seek to obtain an upper bound for the first term on the right-hand side of (3). To that end, we first fix any $j \in \text{salsupp}(\mathbf{X})$ and establish an upper bound on the quantity $\Pr\left(|\text{supp}(\mathbf{M}_{:,j}) \setminus \text{supp}(\mathbf{y})| > \tau\right)$.

For shorthand, and to clarify the exposition, let us define $\epsilon_j \triangleq |\text{supp}(\mathbf{M}_{:,j}) \setminus \text{supp}(\mathbf{y})|$. Now, note that for $\mathbf{y}^*$ as above, we have (essentially by construction) that $|\text{supp}(\mathbf{M}_{:,j}) \setminus \text{supp}(\mathbf{y}^*)| = 0$. This implies that the (random) quantity ϵ_j defined above quantifies the number of 1's in $\mathbf{y}^*$ at locations $i \in \text{supp}(\mathbf{M}_{:,j})$ that are erroneously "flipped" to 0 in $\mathbf{y}$; that is, $\epsilon_j = |\{i \in \text{supp}(\mathbf{M}_{:,j}) : y_i \neq y_i^*\}|$. Note that assumption **A2** guarantees only that $y_i = 1$ when $|\text{supp}(\mathbf{M}_{i,:}) \cap \text{salsupp}(\mathbf{X})| = 1$. It follows that a *necessary* condition for any element y_i, $i \in \text{supp}(\mathbf{M}_{:,j})$, to be erroneous is that the support of the corresponding row $\mathbf{M}_{i,:}$ intersect $\text{salsupp}(\mathbf{X})$ at least twice. Further, since we restrict our attention to $i \in \text{supp}(\mathbf{M}_{:,j})$ we know that $\text{supp}(\mathbf{M}_{i,:})$ must intersect $\text{salsupp}(\mathbf{X})$ at least once – at location j; if not, then $i \notin \text{supp}(\mathbf{M}_{:,j})$. Thus, we have that $\{i \in \text{supp}(\mathbf{M}_{:,j}) : y_i \neq y_i^*\} \subseteq \{i \in \text{supp}(\mathbf{M}_{:,j}) : |\text{supp}(\mathbf{M}_{i,:}) \setminus \{j\} \cap \text{salsupp}(\mathbf{X}) \setminus \{j\}| \geq 1\}$, implying that we can bound the number of errors using the inequality $\epsilon_j \leq \epsilon_j' \triangleq \sum_{i \in \text{supp}(\mathbf{M}_{:,j})} \mathbf{1}_{\left\{\left|\text{supp}(\mathbf{M}_{i,:}) \setminus \{j\} \cap \text{salsupp}(\mathbf{X}) \setminus \{j\}\right| \geq 1\right\}}$. Overall, our approach will be to obtain a bound on ϵ_j by establishing a bound on the probability that ϵ_j' exceeds the threshold τ.

Note that the case $k = 1$ is somewhat trivial as $\text{salsupp}(\mathbf{X})$ contains only a single index – say j – and in this case, we have necessarily that $\epsilon_j = 0$. To address the more general case $k \geq 2$, we recall that the elements of $\mathbf{M}$ are iid Bernoulli(q) random variables, which implies that conditioned on $\text{supp}(\mathbf{M}_{:,j})$ the quantity ϵ_j' defined above is conditionally Binomial($|\text{supp}(\mathbf{M}_{:,j})|, \gamma$), with $\gamma = 1 - (1-q)^{k-1}$. [2] Further, we have by construction of $\mathbf{M}$ that $|\text{supp}(\mathbf{M}_{:,j})|$ is Binomial(m, q) distributed. Overall, then, we have that the random quantity ϵ_j' is (unconditionally) Binomial($m, q\gamma$) distributed. Further, since $\{\epsilon_j > \tau\} \subseteq \{\epsilon_j' > \tau\} \subseteq \{\epsilon_j' \geq \tau\}$, we have that $\Pr(\epsilon_j > \tau) \leq \Pr(\epsilon_j' \geq \tau)$. Thus, by the Chernoff bound (see, e.g., [14, Theorem 2.3]) we have for any $\lambda \geq 0$ and $\tau \geq (1+\lambda) mq\gamma$, that $\Pr(\epsilon_j > \tau) \leq \exp\left(-\frac{\lambda^2 mq\gamma}{2+\lambda}\right)$, implying, in particular, that for $\tau \geq 3mq\gamma$, we have $\Pr(\epsilon_j > \tau) \leq \exp(-mq\gamma)$.

Now, in order to use this bound here, we need to ensure that $\tau \geq 3mq\gamma$, where γ is as above. From the initializations of Algorithm 1 we have that $q = \alpha/k'$ for $\alpha \in (0, 1/5]$ and $\tau = m\alpha(1-2\alpha)/k' = mq(1-2\alpha)$. Now, since $q \in (0,1)$ we have that $(1-q)^{k-1} \geq (1-q)^k \geq 1 - kq$ where the last inequality follows from Bernoulli's Inequality. Thus, $\gamma = 1-(1-q)^{k-1} \leq kq = \alpha(k/k')$. It follows that

[1] In practice, k' may be informed by domain knowledge or could be selected simply as a conservative upper bound. Note also that the parameter α also shows up the sufficient condition on c (which can be viewed as an "oversampling factor"). In practice, one would aim to make c as small as possible, which here motivates that α be taken as large as possible. That α not exceed $1/5$ is a condition of our analysis approach which, on account of several bounding steps, may not illuminate the best possible constants.

[2] Here, the specification of γ follows directly from the fact that given $\text{supp}(\mathbf{M}_{:,j})$ the quantity $|\text{supp}(\mathbf{M}_{i,:}) \setminus \{j\} \cap \text{salsupp}(\mathbf{X}) \setminus \{j\}|$ is conditionally Binomial($k-1, q$) distributed.

a sufficient condition to ensure that $\tau \geq 3mq\gamma$ is that $mq(1-2\alpha) \geq 3mq\alpha(k/k')$, or $k'/k \geq 3\alpha/(1-2\alpha)$. Since $3\alpha/(1-2\alpha) \leq 1$ by choice of α, this condition is satisfied whenever $k' \geq k$, which was a condition of the Theorem.

Now, let $\Delta_1 \triangleq \sum_{j \in \text{salsupp}(\mathbf{X})} \Pr\left(|\text{supp}(\mathbf{M}_{:,j}) \setminus \text{supp}(\mathbf{y})| > \tau\right)$ and consider the first term on the right-hand side of (3). Under the conditions specified in Algorithm 1, and when $k' \geq k$ we have that $\Delta_1 = 0$ when $k = 1$, while for $k \geq 2$, we have by the union bound that $\Delta_1 \leq k \ \exp\left(-mq\gamma\right)$. This implies, in particular, that for any $\delta \in (0,1)$ and $k \geq 1$ we have $\Delta_1 \leq \delta$ provided that $m \geq \left(\frac{1}{\alpha\gamma}\right) \ k' \log\left(\frac{k}{\delta}\right)$. Now, it is easy to show[3] that $\gamma \geq \alpha k/(2k')$ when $k \geq 2$, so that overall $\Delta_1 \leq \delta$ whenever $m \geq \left(\frac{2k'}{\alpha^2 k}\right) \ k' \log\left(\frac{k}{\delta}\right)$.

2) Part 2: Support error bounds for indices $j \notin \text{salsupp}(\mathbf{X})$: We now consider indices in the complement of the true salient support of $\mathbf{X}$, and seek to obtain an upper bound for the second term on the right-hand side of (3). To that end, we first fix any $j \notin \text{salsupp}(\mathbf{X})$ and establish an upper bound on the quantity $\Pr\left(|\text{supp}(\mathbf{M}_{:,j}) \setminus \text{supp}(\mathbf{y})| \leq \tau\right)$.

Note that since $\text{supp}(\mathbf{y}) \subseteq \text{supp}(\mathbf{y}^*)$, where $\mathbf{y}^*$ is the ideal or "error-free" vector defined above, we have that $|\text{supp}(\mathbf{M}_{:,j}) \setminus \text{supp}(\mathbf{y}^*)| \leq |\text{supp}(\mathbf{M}_{:,j}) \setminus \text{supp}(\mathbf{y})|$. The implication is that for any τ, we have $\{|\text{supp}(\mathbf{M}_{:,j}) \setminus \text{supp}(\mathbf{y}^*)| > \tau\} \subseteq \{|\text{supp}(\mathbf{M}_{:,j}) \setminus \text{supp}(\mathbf{y})| > \tau\}$. Now, it follows from this inclusion that $\Pr\left(|\text{supp}(\mathbf{M}_{:,j}) \setminus \text{supp}(\mathbf{y})| \leq \tau\right) \leq \Pr\left(|\text{supp}(\mathbf{M}_{:,j}) \setminus \text{supp}(\mathbf{y}^*)| \leq \tau\right)$. Our proof proceeds by deriving an upper bound for the right-hand side of this inequality.

Now, using (4), we have that $|\text{supp}(\mathbf{M}_{:,j}) \setminus \text{supp}(\mathbf{y}^*)| = |\text{supp}(\mathbf{M}_{:,j}) \setminus \bigcup_{\ell \in \text{salsupp}(\mathbf{X})} \text{supp}(\mathbf{M}_{:,\ell})| = |\{ \ i \ : \ \{M_{i,j} = 1\} \cap \{\bigcap_{\ell \in \text{salsupp}(\mathbf{X})} \ \{M_{i,\ell} = 0\} \ \}\}|$. Thus, it follows that $|\text{supp}(\mathbf{M}_{:,j}) \setminus \text{supp}(\mathbf{y}^*)|$ is Binomial$(m, q(1-q)^k)$ distributed. Again we employ the Chernoff bound, which implies here that for $\lambda \in [0,1]$, $\Pr\left(|\text{supp}(\mathbf{M}_{:,j}) \setminus \text{supp}(\mathbf{y}^*)| \leq (1-\lambda)mq(1-q)^k\right) \leq \exp\left(-\lambda^2 mq(1-q)^k/2\right)$. Letting $\tau = (1-\lambda)mq(1-q)^k$, and simplifying, we obtain that for any $\tau \leq mq(1-q)^k$,

$$\Pr\left(|\text{supp}(\mathbf{M}_{:,j}) \setminus \text{supp}(\mathbf{y}^*)| \leq \tau\right) \leq \exp\left(-\frac{\left(mq(1-q)^k - \tau\right)^2}{2mq(1-q)^k}\right). \quad (5)$$

In order to use the bound (5) here, we need to ensure that for τ, m, and q as specified in Algorithm 1, the condition $\tau \leq mq(1-q)^k$ is satisfied. To that end, note that by Bernoulli's Inequality it is sufficient to ensure that $\tau \leq mq(1-kq)$. Using the fact that $q = \alpha/k'$ and $\tau = m\alpha(1-2\alpha)/k' = mq(1-2\alpha)$ we see that this condition holds whenever $k'/k \geq 1/2$, which holds under the condition of the Theorem that $k' \geq k$. Further, for this choice of τ we can simplify the Chernoff bound to obtain that

$$\begin{aligned}\Pr\left(|\text{supp}(\mathbf{M}_{:,j}) \setminus \text{supp}(\mathbf{y}^*)| \leq \tau\right) &\leq \exp\left(-\frac{mq\left(1-kq-1+2\alpha\right)^2}{2(1-q)^k}\right) \\ &\leq \exp\left(-\frac{\alpha^2 mq\left(2-k/k'\right)^2}{2(1-q)^k}\right) \leq \exp\left(-\alpha^2 mq/2\right), \quad (6)\end{aligned}$$

where the first inequality follows from Bernoulli's Inequality applied to the term $(1-q)^k$, the second inequality from the specification of q (and some simplification), and the third inequality follows since $(1-q)^k \leq 1$ and $k'/k \geq 1$. Now, we turn our attention back to the second term on the right-hand side of (3). Let us denote $\Delta_2 \triangleq \sum_{j \notin \text{salsupp}(\mathbf{X})} \Pr\left(|\text{supp}(\mathbf{M}_{:,j}) \setminus \text{supp}(\mathbf{y})| \leq \tau\right)$. By the union bound, we have that $\Delta_2 \leq (n-k)\exp\left(-\alpha^2 mq/2\right)$, which implies that for any $\delta \in (0,1)$ we have $\Delta_2 \leq \delta$ provided that $m \geq \left(\frac{2}{\alpha^3}\right) \ k' \log\left(\frac{n-k}{\delta}\right)$.

3) Putting the Results Together: Overall, we have established that for the parameter specifications of Algorithm 1, and when $k' \geq k$, we have that $\Pr\left(\widehat{\mathcal{S}} \neq \text{salsupp}(\mathbf{X})\right) \leq \Delta_1 + \Delta_2 \leq 2\delta$ whenever $m \geq \max\left\{ \ \left(\frac{2k'}{\alpha^2 k}\right) \ k' \log\left(\frac{k}{\delta}\right) \ , \ \left(\frac{2}{\alpha^3}\right) \ k' \log\left(\frac{n-k}{\delta}\right) \ \right\}$. The condition $m \geq \max\left\{\frac{2k'}{\alpha^2 k}, \frac{2}{\alpha^3}\right\} \ k' \log\left(\frac{n}{\delta}\right)$ suffices, as claimed.

III. Conclusions

We have shown that a simple inference approach based on group testing ideas provably succeeds at identifying the locations of k outliers in a large data collection, with high probability, given only $m = O(k \log n)$ linear combinations of the data elements. The inference procedure itself is computationally efficient, requiring at most $O(mn) = O(kn \log n)$ operations, and requires only the storage and processing of binary data structures. Due to space limitations, we defer in-depth numerical investigations to future work.

[3] Write $k' = \beta k$ for some $\beta \geq 1$. It is easy to see the that bound holds for $k = 2$; that it holds for $k \geq 3$ follows from a simple monotonicity argument.

References

[1] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[2] E. Candes, "Compressive sampling," in *Proceedings of the International Congress of Mathematics*, Madrid, Spain, 2006, pp. 1433–1452.

[3] D. L. Donoho and J. Tanner, "Counting the faces of randomly-projected hypercubes and orthants, with applications," *Discrete & computational geometry*, vol. 43, no. 3, pp. 522–541, 2010.

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

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

[6] G. Cormode and S. Muthukrishnan, "Combinatorial algorithms for compressed sensing," in *Structural Information and Communication Complexity*, pp. 280–294. Springer, 2006.

[7] A. C. Gilbert, M. A. Iwen, and M. J. Strauss, "Group testing and sparse signal recovery," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, 2008, pp. 1059–1063.

[8] P. Indyk, "Explicit constructions for compressed sensing of sparse signals," in *Proc. ACM-SIAM Symposium on Discrete Algorithms*, 2008, pp. 30–33.

[9] W. Wang, M. J. Wainwright, and K. Ramchandran, "Information-theoretic limits on sparse signal recovery: Dense versus sparse measurement matrices," *IEEE Trans. Inform. Theory*, vol. 56, no. 6, pp. 2967–2979, 2010.

[10] G. K. Atia and V. Saligrama, "Boolean compressed sensing and noisy group testing," *IEEE Trans Inform Theory*, vol. 58, no. 3, pp. 1880–1901, 2012.

[11] M. Cheraghchi, A. Hormati, A. Karbasi, and M. Vetterli, "Group testing with probabilistic tests: Theory, design and application," *IEEE Trans Inform Theory*, vol. 57, no. 10, pp. 7057–7067, 2011.

[12] C. L. Chan, S. Jaggi, V. Saligrama, and S. Agnihotri, "Non-adaptive group testing: Explicit bounds and novel algorithms," in *Proc. IEEE Intl Symposium on Inform Theory*, 2012, pp. 1837–1841.

[13] C. Aksoylar, G. Atia, and V. Saligrama, "Sparse signal processing with linear and non-linear observations: A unified Shannon-theoretic approach," *Submitted*, 2012, online at arxiv.org/abs/1304.0682.

[14] C. McDiarmid, "Concentration," in *Probabilistic methods for algorithmic discrete mathematics*, pp. 195–248. Springer, 1998.