Some Algorithmic Problems and Results in Compressed Sensing

S. Muthukrishnan
Google Inc.
muthu@google.com

***Abstract*— In Compressed Sensing [9], we consider a signal $\mathbf{A} \in \mathbb{R}^n$ that is *compressible* with respect to some dictionary of ψ_i's, that is, its information is concentrated in k coefficients $\langle \mathbf{A}, \psi_i \rangle$. The goal is to reconstruct such signals using only a few *measurements* $\langle \mathbf{A}, v_i \rangle$, for carefully chosen v_i's which depend on $\{\psi_i\}$.**

Known results [9], [3], [21] prove that there exists a single $O(k \log n) \times n$ measurement matrix such that any compressible signal can be reconstructed from these measurements, with error at most $O(1)$ times the worst case error for the class of such signals. Compressed sensing has generated tremendous excitement both because of the sophisticated underlying mathematics including Linear Algebra [9], Geometry [21] and Uncertainty Principle [3], and because of its potential applications to signal processing, communication theory and compression.

In this paper, we focus on algorithmic aspects of compressed sensing and present new problems and results. For example, we

1) **Present a simple, deterministic explicit construction of $\text{poly}(k, \log n)$ linear measurements that suffice for compressed sensing.**[1]
2) **Introduce *functional compressed sensing*, that is, extend compressed sensing of a signal to that of functions of the signal, for various functions.**
3) **Extend compressed sensing to a distributed, continuous model of computing.**

All the results above are obtained by simple combinatorial or number-theoretic ideas.

I. Introduction to Compressed Sensing

Dictionary. The *dictionary* Ψ denotes an orthonormal basis for $\mathbb{R}^n$, i.e. Ψ is a set of n real-valued vectors ψ_i each of dimension n and $\psi_i \perp \psi_j$. The *standard basis* is the traditional coordinate system for n dimensions, namely, for $i = 1, \ldots, n$, the vector $\psi_i = [\psi_{i,j}]$ where $\psi_{i,j} = 1$ iff $i = j$. By applying an appropriate rotation to any basis and signal, our discussion henceforth can be restricted to the standard basis Ψ only.

Signal Transformation. A *signal* vector $\mathbf{A}$ in $\mathbb{R}^n$ is transformed by this dictionary into a vector of *coefficients* $\theta(\mathbf{A})$ formed by inner products between $\mathbf{A}$ and vectors from Ψ. That is, $\theta_i(\mathbf{A}) = \langle \mathbf{A}, \psi_i \rangle$ and $\mathbf{A} = \sum_i \theta_i(\mathbf{A})\psi_i$ by the orthonormality of Ψ. We refer to θ_i where $\mathbf{A}$ is implicitly clear. From now on, we reorder the vectors in the dictionary so $|\theta_1| \geq |\theta_2| \geq \ldots \geq |\theta_n|$.

Sparse Representation. In the area of sparse approximation theory [8], one seeks representations of $\mathbf{A}$ that are *sparse*, i.e., use few coefficients. Formally, $\mathbf{R} = \sum_{i \in K} \theta_i \psi_i$, for some set K of coefficients, $|K| = k \ll n$. Clearly, $\mathbf{R}(\mathbf{A})$ cannot exactly equal the signal $\mathbf{A}$ for all signals. The error is typically taken as $\|\mathbf{R} - \mathbf{A}\|_2^2 = \sum_i (\mathbf{R}_i - \mathbf{A}_i)^2$. By the classical Parseval's equality, this is equivalently $\|\theta(\mathbf{A}) - \theta(\mathbf{R})\|_2^2$. The optimal k representation of $\mathbf{A}$ under Ψ, $\mathbf{R}^k_{\text{opt}}$, therefore takes k coefficients with the largest $|\theta_i|$'s. The error then is $\|\mathbf{A} - \mathbf{R}^k_{\text{opt}}\|_2^2 = \sum_{i=k+1}^n \theta_i^2$. This is the error in representing the signal $\mathbf{A}$ in a compressed form using k coefficients from Ψ.

Compressible Signals. In any application, one has a "class" of input signals; hence, one chooses an appropriate dictionary so that most of the signals are "compressible" using that dictionary, and represents the signal using the adequate number ($k \ll n$) of coefficients ($\theta_1, \ldots, \theta_k$). There are different notions of a signal being compressible in a dictionary, most typically, the *p-Compressible case*. Specifically the coefficients have a power-law decay: for some $p \in (0, 1)$, and for all i,

[1] Throughout, $\text{poly}()$ denotes a multinomial in its arguments.

$|\theta_i| = O(i^{-1/p})$. Consequently,

$$\|\mathbf{A} - \mathbf{R}^k_{\text{opt}}\|_2^2 \leq \int_{k+1}^{n} O((i^{-1/p})^2) \leq C_p k^{1-2/p},$$

for some constant C_p. A simplification of these models is the k-*support case*, where the signal has at most k non-zero coefficients, so $\mathbf{R}^k_{\text{opt}} = \mathbf{A}$.

Compressed Sensing. Recently, Donoho posed a fundamental question [9]: Since most of the information in the signal is contained in only a few coefficients and the rest of the signal is not needed for the applications, can one directly determine (acquire) only the relevant coefficients without reading (measuring) each of the coefficients? In a series of papers over the past few years, the following result has emerged.

Theorem 1. *[9], [3], [21] There exists a non-adaptive set V of $O(k \log(n/k))$ vectors in $\mathbb{R}^n$ which can be constructed once and for all from the standard basis. Then, for fixed $p \in (0,1)$ and any p-compressible signal $\mathbf{A}$ in the standard basis, given only* measurements $\langle \mathbf{A}, v_i \rangle$, $v_i \in V$, *a representation $\mathbf{R}$ can be determined in time polynomial in n such that $\|\mathbf{A} - \mathbf{R}\|_2^2 = O(k^{1-2/p})$.*

Since the worst case error for a p-compressible signal is $C_p k^{1-2/p}$, the representation above is *optimal, up to constant factors* for the class of all p-compressible signals, for a fixed p. The result above shows existence of V by proving that a random set of V vectors will satisfy the theorem with nonzero probability. The number of measurements is only $\log(n/k)$ more than the trivial lower bound of k measurements one needs even in the k-support special case.

This result has generated much interest, and a sequence of papers have improved different aspects of the result [9], [25], [3], [21], [6], [14]; found interesting applications including MR imaging [23] wireless communication [21] and generated implementations [17]; found mathematical applications to coding and information theory [2]; and extended the results to noisy and distributed settings [17].

In this paper, we study algorithmic aspects, presenting algorithmic problems as well as solving some, in Compresses Sensing and its variants.

II. Explicit Construction of Measurements

Let T, the *transformation matrix* denote the non-adaptive measurement set of vectors V, that is, $T[i,j] = v_i[j]$. Existing results show that if T satisfies certain conditions, then the theorem 1 holds; additionally they show that T chosen from an appropriate random distribution suffices. The necessary conditions are quite involved, such as computing the eigenvalues of every $O(k \log n)$ square submatrix of T [9], and testing that each such submatrix is an isometry, behaving like an orthonormal system [3]. No explicit construction is known to produce T's with these necessary properties. Instead, algorithms for Compressed Sensing choose a random T directly, and hence are *Monte Carlo* algorithms, with some probability of failure. One can produce Las Vegas algorithms if one could take a random T and test whether it satisfies the necessary conditions. However, this is expensive, taking time at least $\Omega(n^{k \log n})$.

The intuitive way to think about these problems is to consider combinatorial group testing problems. We have a set $U = [n]$ of items and a set D of *distinguished* items, $|D| \leq k$. We identify the items in D by performing group tests on subsets $S_i \subseteq U$ whose output is 1 or 0, revealing whether that subset contains one or more distinguished items, that is $|S_i \cap D| \geq 1$. There exist a *fixed* collection of $O((k \log n)^2)$ tests which identify each of the distinguished items precisely.

There is a strong connection between this problem and Compressed Sensing. We can treat θ_i's as items and the largest (in magnitude) k as the members of D. Each test set S_i can be written as its characteristic vector χ_{S_i} of n dimensions. A difficulty arises in interpreting the outcome of $\langle \mathbf{A}, \chi_{S_i} \rangle$. The discussion so far has been entirely combinatorial, but the outcome of this linear-algebraic operation of inner product must be interpreted as a binary outcome to apply standard combinatorial group testing methods.

The only previously known algorithm to explicitly construct the measurements is in [6]. In this section, we give an alternative explicit construction using number-theoretic properties. Our results, like those of [3], [9], [6], are stated with respect to the error due to the worst case over all signals in the class, which we denote $\|\mathbf{C}^k_{\text{opt}}\|_2^2 = O(k^{1-2/p})$. For any signal that is p-compressible with fixed p and C_p it follows that $\|\mathbf{R}^k_{\text{opt}} - \mathbf{A}\|_2 \leq \|\mathbf{C}^k_{\text{opt}}\|_2$. We focus only on $p < \frac{1}{2}$; the results can be extended to $0 < p < 1$ easily.

We make use of the Hamming code matrix H_n, which is the $\lceil 1 + \log_2 n \rceil$ matrix whose ith column is 1 followed by the binary representation of i. We combine matrices together to get larger matrices by (a) concatenating the rows of N to M and get a matrix denoted $M \bigcup N$, or (b) a Tensor product-like operation we denote $\otimes$, defined as follows: Given matrices V and W of dimension $v \times n$ and $w \times n$ respectively, define the matrix $(V \otimes W)$ of dimension $vw \times n$ as $(V \otimes W)_{iv+l,j} = V_{i,j} W_{l,j}$.

Transform Definition. Our solution is to construct a group testing procedure of small number of groups. Let

$$k < p_1 < p_2 < .. < p_x$$

be the sequence of consecutive primes larger than k and with the largest index $x = O(k \log_k N)$. For each p_i, form groups

$$0 \bmod p_i, \dots, p_i - 1 \bmod p_i,$$

called p_i*-groups*. Each $j \in [1, n]$ belongs to one p_i-group for each i. We call this the *prime grouping* of size k and denote it $\mathcal{P}(k)$, and denote its size by $|\mathcal{P}(k)|$. By the Prime Number Theorem, it follows that $|\mathcal{P}(k)|$ is $\text{poly}(k, \log n)$. The required primes can be enumerated efficiently and deterministically. Prime grouping was first proposed in [19] for data stream computing, and has since been extended [12].

We define our transform matrix as follows. Let k_α be function of k, ε, $\log n$ to be defined later, and $k_\beta = |\mathcal{P}(k_\alpha)|$. Write T_1 as the matrix formed by the concatenation of χ_{S_i} for all S_i in $\mathcal{P}(k_\alpha)$. Similarly, write T_2 as characteristic matrix of $\mathcal{P}(k_\beta)$. Our transform matrix T is $(T_1 \otimes H) \bigcup T_2$. ∎

Reconstruction Algorithm. For each set of $\lceil 1 + \log n \rceil$ measurements due to $S_i \otimes H$, we recover $x_0 \dots x_{\lceil \log n \rceil} = ((\chi_{S_i} \otimes H)\mathbf{A}$, and decode an identifier j_i as

$$\sum_{b=1}^{\log n} \frac{2^{b-1}(|x_b| - \min\{|x_b|, |x_0 - x_b|\})}{\max\{|x_b|, |x_0 - x_b|\} - \min\{|x_b|, |x_0 - x_b|\}}.$$

This generates a set of coefficients $J = \{j_1, j_2 \dots\}$. We then use the measurements due to T_2 to estimate each coefficient in J: for each $j \in J$, we set $\hat{\theta}_j = \chi_{R_i}\mathbf{A}$ for $J \cap R_i = \{j\}$. The properties of $\mathcal{R}$ ensure that there will be at least one such R_i, and if there is more than one, then we can pick one arbitrarily. Our output is the set of k pairs $(j, \hat{\theta}_j)$ with the k largest values of $|\hat{\theta}_j|$. ∎

Theorem 2. *Let $p < 1/2$. Our construction above gives a set of $\text{poly}(k, \log n, 1/\varepsilon)$ measurements for a compressible signal $\mathbf{A}$ in time polynomial in the set size and returns a $\mathbf{R}$ for $\mathbf{A}$ of at most k coefficients $\hat{\theta}$ in time polynomial in the number of measurements such that*

$$\|\hat{\theta} - \theta\|_2^2 = \|\mathbf{R} - \mathbf{A}\|_2^2 < \|\mathbf{R}^k_{\text{opt}} - \mathbf{A}\|_2^2 + \varepsilon \|\mathbf{C}^k_{\text{opt}}\|_2^2.$$

Proof. Let K denote the set of the k_α largest coefficients.

The claim is, each coefficient in K will be isolated in at least one of the p_i-groups in $\mathcal{P}(k_\alpha)$, away from the others in K. This is true because any such coefficient δ can collide with a different member δ' of K in at most $\log_{k_\alpha} n$ p_j-groups with different j's. Otherwise, the difference $|\delta - \delta'| < n$ would be divisible by $\log_{k_\alpha} n + 1$ different primes $> k_\alpha$ which is a contradiction.

Consider $j \leq k_\alpha$. We know that there is some set $S_i \in \mathcal{P}$ so that $K \cap S_i = \{j\}$. Consider the vector of measurements involving this set, $x = (\chi_{S_i} \otimes H)\mathbf{A}$. It is simple to see that the identity j will be recovered provided j is the majority item in this set, i.e., $|\theta_j| > \sum_{l \neq j,\, l \in S_i} |\theta_l|$ [5]. This can be at most $\sum_{l > k_\alpha} |\theta_l| \leq c k_\alpha^{1-1/p}$, for a suitable constant c. Hence, provided $|\theta_j| > c k_\alpha^{1-1/p}$, $j \in J$.

We now show that $\forall j \in J : \ |\hat{\theta}_j - \theta_j| \leq \frac{\varepsilon}{5\sqrt{k}} \|\mathbf{C}^k_{\text{opt}}\|_2$. Each $S_i \in \mathcal{P}(k_\alpha)$ generates at most one $j \in J$. Since $k_\beta = |\mathcal{P}(k_\alpha)|$, and we use $\mathcal{P}(k_\beta)$ for estimation, we can guarantee for each $j \in J$ there is at least one $S_i \in \mathcal{P}k_(\beta)$ such that $J \cap S_i = \{j\}$. We can choose the estimate of θ_j as any measurement of θ_j that avoids all other members of J. Consider the error in the estimation of θ_j. We have $|\hat{\theta}_j - \theta_j| \leq \sum_{l \notin J} |\theta_l|$, and $l \notin J$ implies either $l \leq k_\alpha$ and $\theta_j^2 \leq c k_\alpha^{2-2/p}$ or $l > k_\alpha$.

Hence,

$$
\begin{aligned}
|\hat{\theta}_j - \theta_j| &\leq \sum_{l<k_\alpha, l\notin J} |\theta_l| + \sum_{l>k_\alpha} |\theta_l| \\
&\leq c(k_\alpha - 1)k_\alpha^{1-1/p} + ck_\alpha^{1-1/p} \\
&\leq ck_\alpha^{2-1/p}.
\end{aligned}
$$

By our choice of k_α and c, we need to insure that $ck_\alpha^{2-1/p} \leq \frac{\varepsilon}{5\sqrt{k}}\|\mathbf{C}_{\text{opt}}^k\|_2$. Then, $ck_\alpha^{2-1/p} \leq \frac{\varepsilon}{5\sqrt{k}}ck^{1/2-1/p}$, that is, $k_\alpha \leq C(k\varepsilon^{-p})^{1/(1-2p)}$, for suitable constant C.

It is now a technical detail to show that given $\hat{\theta}(\mathbf{A}) = \{\hat{\theta}_i(\mathbf{A})\}$ such that $(\hat{\theta}_i - \theta_i)^2 \leq \frac{\varepsilon^2}{25k}\|\mathbf{C}_{\text{opt}}^k\|_2^2$ if $\theta_i^2 \geq \frac{\varepsilon^2}{25k}\|\mathbf{C}_{\text{opt}}^k\|_2^2$, picking the k largest coefficients from $\hat{\theta}(\mathbf{A})$ gives an error $\|\mathbf{R}_{\text{opt}}^k - \mathbf{A}\|_2^2 + \varepsilon\|\mathbf{C}_{\text{opt}}^k\|_2^2$ k-term representation of $\mathbf{A}$ [6]. ∎

We have not fully optimized the various polynomial factors, but still, the polynomial we will obtain after optimizing will be worse than $k \log n/k$, which is the existential bound.

Problem 1. *Is there an explicit construction of a measurement matrix of smaller size, with an efficient reconstruction algorithm?*

Even for the k-support case, the best known explicit construction is of size $k^2 \log^2 n$ [6], or kn^ϵ [16] which is still off from the $k \log n/k$ existential bound.

Problem 2. (Subsignal compression) *Say we are given a signal* $\mathbf{A}$ *for preprocessing. Thereafter, each query is a range* $[i, j]$*, and the desired result is the best* k*-term representation for* $\mathbf{A}[i, j]$ *for a suitable subdictionary of size* $j - 1 + 1$*. What is the best tradeoff between the preprocessing time and space versus the query time one can achieve?*

The problem above is in the style of a data structure problem. Naturally one can precompute the correct answer for all ranges and the query becomes trivial; on the other extreme, one can do no preprocessing and solve the problem as needed when the query arrives taking time $O(j - i)$. The interesting tradeoffs will be between these two extremes.

Problem 3. (Universal Decoding) *Using random linear measurements for compressed sensing has certain universal decoding property: the signal can be measured in one basis and reconstructed in a different basis, if the two bases were mutually incoherent. Is there a combinatorial characterization of the other basis in which the reconstruction works for the measurements in this paper based on prime groups?*

One of the alluring aspects of compressed sensing is its universal decoding ability. The problem above is an attempt at quantifying to what extent the measurements based on prime groups are useful for universal decoding.

III. Distributed, Continuous Compressed Sensing

Our motivation is a network of distributed sensors each of which measures a portion of the underlying signal. The system as a whole has to continuously track the signal, such as in monitoring applications.

We formalize the problem as follows. There are ℓ clients $c_1, \ldots, c_\ell$. At any time t, each client has the signal $\mathbf{A}_i(t)$ obtained from a series of updates up until time t. They all communicate with a central server S. S needs to be able to maintain enough information at any time t to approximate $\mathbf{A}(t) = \sum_i \mathbf{A}_i(t)$ (in applications, one may need the average of the signals to be maintained [20], but that problem is equivalent to the sum problem). The problem is to minimize the number of bits communicated altogether, over the course of updates, at each time t.

This problem may be thought of as being related to a classical result in communication theory [24]. The Slepian-Wolf theorem shows a bound on the bits needed for two distributed sources to communicate their correlated data without loss to a central server. Here, we focus on lossy transmission since we seek k coefficient representation of the $\mathbf{A}(t)$ only. Moreover, we seek bounds for any given instance of the signals for each t, not in asymptotics for probabilistic inputs.

If each client sents each of its updates to S every time instant, then S can easily update the grouping solution from above with only $\text{poly}(\ell, \log n)$ memory; at any time t, the method can accurately retrieve the ℓ coefficient approximation to the signal $\mathbf{A}(t)$ in time $\text{poly}(\ell, \log n)$. But the communication will be $O(t\ell)$ bits. Our goal is to improve upon this "naive" solution.

In what follows, we present a simple, efficient, distributed, continuous solution. This solution uses the p_i groups described earlier, and will be in the style of Compressed Sensing, that is, the entire method will be deterministic, with a fixed set of measurements monitored by each of the sources. There is a growing body of work in database research on continuously monitoring various functions on distributed signals (see survey [4]), but they tend to be probabilistic (i.e., succeeding only with some probability and need common random seeds across the clients), and work typically when the values they track are monotonic. In our case, as $\mathbf{A}_i(t)$ changes, its coeficients change not necessarily in a monotonic way. The methods surveyed in [4] such as maintaining L_2 sketches, quantiles and other functions do not give the k coefficient approximation we seek here.

Our approach is as follows. Each client c_i maintains its prime groups, Then, client c_i maintains $\lambda_i[j](t) = (((T_1 \otimes H) \bigcup T_2)\mathbf{A}_i)[j](t)$, from the previous section, over time t. We make two claims.

- It suffices to maintain $|(((T_1 \otimes H) \bigcup T_2)\mathbf{A})[j]|$'s accurate upto $\pm\delta\frac{\varepsilon}{\sqrt{k}}\|\mathbf{C}^k_{\text{opt}}\|_2$ for the algorithm in the previous section, for some suitable constant fraction δ.
- There is an algorithm to maintain $|((T_1 \otimes H) \bigcup T_2)\mathbf{A})[j]|$'s accurate upto $\pm\gamma$ using $O(1)$ bits each of $O(\frac{\sum_t |\lambda_i[j](t)-\lambda_i[j](t-1)|}{\gamma/\ell})$ communications per client.

The first is the extension of the proof from the previous section. Now we will provide the algorithm for the second claim, which is quite simple. Each client keeps a budget γ/ℓ and any time t that $|\lambda_i[j](t)|$ differs from $|\lambda_i[j](t')|$ by γ/ℓ, it updates the server with its current $\lambda_i[j](t)$, in a suitable encoding. It is easy to see that then $\lambda[j] = \sum_i \lambda_i[j]$ is maintained within accuracy $\pm\gamma$. The nontrivial part is to analyze the communication of this scheme. It depends on how the $\lambda_i[j]$'s change over time. A trivial upper bound is stated above, but tighter bounds can be stated in terms of the distribution of the $\lambda_i[j]$'s. The bound above is intuitive since $\sum_t |\lambda_i[j](t) - \lambda_i[j](t-1)|$ is the *total absolute change* $\Delta_i[j]$ in each site over time. Note that $\Delta_i[j]$ denotes the change in the counts we maintain, and not in the signal or its coefficients. This lets us conclude (we use upper bound $\|\mathbf{C}^k_{\text{opt}}\|_2$ in the algorithm):

Theorem 3. *Let $p < 1/2$. There is a fixed set of $\text{poly}(k, \log n, 1/\varepsilon)$ counts to maintain at all times for each site $1, \ldots, \ell$, independent of each other, such that for any signal $\mathbf{A}(t)$ that is p-compressible for all t, with at most $O(\frac{\ell\sqrt{k}\sum_{i,j}\Delta_i[j]}{\varepsilon\|\mathbf{C}^k_{\text{opt}}\|_2})$ sized communication altogether, at any time t, in time $\text{poly}(k, \log n, 1/\varepsilon)$, one can obtain a $\mathbf{R}$ for $\mathbf{A}$ of at most k coefficients $\hat{\theta}$ such that*

$$\|\hat{\theta} - \theta\|_2^2 = \|\mathbf{R} - \mathbf{A}\|_2^2 < \|\mathbf{R}^k_{\text{opt}} - \mathbf{A}\|_2^2 + \varepsilon\|\mathbf{C}^k_{\text{opt}}\|_2^2.$$

A few other distributed continuous algorithms are also natural including trying to maintain the top k coefficients directly, or using the central site to more carefully monitor what budgets clients use, etc. It will be worthwhile to study these methods experimentally as is typically done in references in [4].

Problem 4. (Functional Slepian-Wolf) *The basic Slepian-Wolf bound deals with two correlated sources $\mathbf{A}_1(t)$ and $\mathbf{A}_2(t)$ transmitting to a central server in close to information-bound without knowledge of each others input. Derive a similar result when the goal is not to communicate the two signals fully, but only approximate the k coefficient representation for the sum of two signals, $\mathbf{A}_1(t) + \mathbf{A}_2(t)$.*

The problem above is a combination of communication theory and network coding with that of sparse approximation. In particular, there are heuristics such as using modelling and prediction to decrease the number of bits communicated. The problem above is to consider the fundamental bounds.

Problem 5. *Design a deterministic method such as the one above for distributed continuous maintenance for estimating how many coefficients are in a given range of values.*

The problem above can be solved using what is known as inverse sampling [7], combined with the method we have described above for maintaining counts upto additive error approximations. However, such a method gives a result that will work probabilistically, and require common random bits across clients. The main focus in the problem above is to develop results in the style of compressed sensing by using a deterministic, fixed set of counts or measurements.

IV. Functional Compressed Sensing

We generalize compressed sensing. Given a function f, determine the set V of linear measurements such that for any signal $\mathbf{A} \in \mathbb{R}^n$, given $\langle \mathbf{A}, v_i \rangle$, $v_i \in V$ only, $f(\mathbf{A})$ can be computed. We call this *functional compressed sensing*. In standard compressed sensing, $f()$ is the top k coefficients $\theta(\mathbf{A})$ in some basis Ψ, and serves to compress $\mathbf{A}$ lossily. In general, other functions may be of interest.

There are two versions: f is defined on the signal $\mathbf{A}$ in one and in the other on the transformed coefficients θ. While these two versions of the functions are mathematically equivalent (using the forward and inverse transforms), the functional compressed sensing problem may differ in complexity for the two versions since the measurements are of the signal $\mathbf{A}$.

We pose a specific problem and discuss the variations and the complexities.

Problem 6. (Compressed Sensing for the Median Energy) *In version I, $f(\mathbf{A})$ is to find an i such that $\sum_{i \le j} \mathbf{A}[i]^2 \ge \frac{\sum_i \mathbf{A}[i]^2}{2}$ and $\sum_{i \le j-1} \mathbf{A}[i]^2 < \frac{\sum_i \mathbf{A}[i]^2}{2}$. In version II, $f(\mathbf{A})$ is to find an i such that $\sum_{i \le j} \theta[i]^2 \ge \frac{\sum_i \theta[i]^2}{2}$ and $\sum_{i \le j-1} \theta[i]^2 < \frac{\sum_i \theta[i]^2}{2}$.*

The median is an order statistic; extension to quantiles is standard. The idea of asking for median energy in version II is somewhat artificial since the ordering of the coefficients could be arbitrary; it may be more natural in version I where the dimensions may have a well defined ordering such as time in time series data. On the other hand, for other functions, such as estimating top k coefficients, analog of version II may be more natural since the transform, rather than the signal, is likely to be sparse. Version I can be converted to version II and vice versa if we have the time and space to perform the (inverse) transform. In some instances of compressed sensing, one can measure from the signal and reconstruct in a different basis that is unknown during the measurement [17], which is similar to solving version II.

We focus on version I. A natural algorithm is to design a procedure using random projections to estimate $\mathbf{A}[i]^2$, use the same procedure hierarchically with signals obtained by collapsing dyadic intervals, and then do a binary search to find the median. This gives:

Theorem 4. *Given a signal $\mathbf{A}$, there is a set of measurements of size $\mathrm{poly}(1/\varepsilon, \log n)$ from which we can determine j with high probability such that $\sum_i \frac{(1-\varepsilon)\mathbf{A}[i]^2}{2} \le \sum_{i \le j} \mathbf{A}[i]^2 \le \sum_i \frac{(1+\varepsilon)\mathbf{A}[i]^2}{2}$.*

The result above however does not give a compressed sensing style result where the same set of measurement vectors work for all (compressible) signals, which we leave open. If we are allowed to take measurements on $\mathbf{A}^2$, one can follow the same procedure as above using prime groups at each of the dyadic intervals, but now need a different procedure to estimate $\mathbf{A}[i]^2$ for any given i, based on the counts for the prime groups, which is presented in [12]. That gives,

Theorem 5. *There is an explicit set of vectors V of size $\mathrm{poly}(1/\varepsilon, \log n)$. Given the measurements for any $\mathbf{A}^2$ with V, we can determine j such that $\sum_i \frac{(1-\varepsilon)\mathbf{A}[i]^2}{2} \le \sum_{i \le j} \mathbf{A}[i]^2 \le \frac{(1+\varepsilon)\sum_i \mathbf{A}[i]^2}{2}$.*

It will be of interest to remove the condition that $\mathbf{A}^2$ be measured, and improve the bounds in the poly term. For example, if one considers p-compressible signals for very small p, one can perhaps use fewer prime groups.

There have been very few results, if any, on functional compressed sensing. Here are example problems to consider:

Problem 7. *Solve version I above for $f(\mathbf{A}) = ||\mathbf{A}||_0$, that is, $f(\mathbf{A})$ is the number of nonzero entries in $\mathbf{A}$. Does there exist a small set of measurements assuming the signal is p-compressible?*

Problem 8. *Solve version I above for $f(\mathbf{A})$ where f measures some notion of information content of $\mathbf{A}$, such as the kth order (empirical) entropy of the signal. Does there exist a small set of measurements assuming the signal is p-compressible?*

V. Concluding Remarks

There are a lot of fundamental bounds to resolve in Compressed Sensing (for example, with arbitrary, not necessarily orthonormal basis, dictionaries). In this paper, we have presented new directions: functional compressed sensing, and distributed continuous compressed sensing. We have also presented explicit constructions for compressed sensing. All the methods used prime groups, which may be a technique of independent interest.

Recently, compressed sensing type results have emerged for low rank matrix approximation, by sampling rows and columns [10] or by taking random projections [22]. Further extensions and improvements will be of great interest. Also, exensions to nonuniform sparse approximation problems will be of interest [18].

Finally, one may look for nonlinear measurements of the signal for compressed sensing. There are well-known methods based on measuring *polynomial* sum's of the signal components [19], [11] for simple data stream problems. It is worth exploring if (a) applications in signal processing will support such measurements, and (b) application of such measurements will get better results than linear measurements.

VI. Acknowledgements

Sincere thanks to Graham Cormode for all the brainstorming. Thanks to Sumit Ganguly and Anirban Mazumdar for showing me a copy of [12].

References

[1] E. Candès, J. Romberg, and T. Tao. Stable signals recovery from incomplete and inaccurate measurements. http://www.acm.caltech.edu/~emmanuel/papers/StableRecovery.pdf, 2005.

[2] E. Candès, M. Rudelson, T. Tao, and R. Vershynin. Error correction via linear programming. In *FOCS*, 2005.

[3] E. Candès and T. Tao. Near optimal signal recovery from random projections and universal encoding strategies. http://arxiv.org/abs/math.CA/0410542, 2004.

[4] G. Cormode and M. Garofalakis. Efficient strategies for continuous distributed tracking tasks. *IEEE Data Engineering Bulletin*, 33-39, 2005.

[5] G. Cormode and S. Muthukrishnan. What's hot and what's not: Tracking most frequent items dynamically. In *ACM PODS*, pages 296–306, 2003.

[6] G. Cormode and S. Muthukrishnan. Towards an algorithmic theory of compressed sensing. DIMACS Tech Report 2005-25, 2005. Also, Proc. SIROCCO, 2006.

[7] G. Cormode, S. Muthukrishnan and I. Rozenbaum. Summarizing and mining inverse distributions on data streams via dynamic inverse sampling. VLDB 2005: 25-36.

[8] R. Devore and G. G. Lorentz. *Constructive Approximation*, volume 303. Springer Grundlehren, 1993.

[9] D. Donoho. Compressed sensing. http://www-stat.stanford.edu/~donoho/Reports/2004/CompressedSensing091604.pdf, 2004.

[10] P. Drineas, M. Mahoney and S. Muthukrishnan Polynomial time algorithm for column-row based relative error low-rank matrix approximation. RANDOM 2006 and ESA 2006.

[11] S. Ganguly and A. Mazumdar. Deterministic K-set structure. *Proc. PODS*, 2006.

[12] S. Ganguly and A. Mazumdar. CR-Precis: A deterministic data structure for update streams. *CS Archives*, cs/0609032, 2006.

[13] A. Gilbert, S. Guha, P. Indyk, Y. Kotidis, S. Muthukrishnan, and M. Strauss. Fast, small-space algorithms for approximate histogram maintenance. In *STOC*, 389–398, 2002.

[14] A. Gilbert, M. Strauss, J. Tropp, R. Vershynin. Algorithmic linear dimension reduction in the L_1 norm for sparse vectors. 44th Annual Allerton Conference on Communication, Control, and Computing, 2006.

[15] J. Haupt and R. Nowak. Signal reconstruction from noisy random projections. http://homepages.cae.wisc.edu/~jhaupt/Publications/RandomProjections05.%pdf, 2005.

[16] P. Indyk. Explicit constructions for compressed sensing problems. *Allerton Conf*, 2006.

[17] Compressed sensing website. http://www.dsp.ece.rice.edu/CS/.

[18] S. Muthukrishnan. Subquadratic algorithms for workload-aware Haar wavelet synopses. *Proc. FSTTCS*, 2005.

[19] S. Muthukrishnan. Data Streams: Algorithms and Applications. *Foundations and Trends in Theoretical Computer Science*, Vol 2, 2006.

[20] S. Papadimitriou, J. Sun and C. Faloutsos. Streaming pattern discovery in multiple time-series. *Proc. VLDB* 2005.

[21] M. Rudelson and R. Vershynin. Geometric approach to error correcting codes and reconstruction of signals. http://www.math.ucdavis.edu/~vershynin/papers/ecc.pdf, 2005. Unpublished Manuscript.

[22] T. Sarlos. Improved Approximation Algorithms for Large Matrices via Random Projections. *Proc. FOCS*, 2006.

[23] SPARS05 Signal Processing with Adaptative Sparse Structured Representations, 2005. http://www.irisa.fr/manifestations/2005/SPARS05/.

[24] D. Slepian and J.K. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. Inform. Theory*, IT-19:471-480, 1973.
2005.

[25] Y. Tsaig and D. Donoho. Extensions of compressed sensing. http://www-stat.stanford.edu/~donoho/Reports/2004/ExtCS-10-22-04.pdf, 2004.