Combinatorial Algorithms for Compressed Sensing

Graham Cormode
Bell Laboratories
cormode@bell-labs.com

S. Muthukrishnan
Rutgers University
muthu@cs.rutgers.edu

Abstract— In sparse approximation theory, the fundamental problem is to reconstruct a signal $\mathbf{A} \in \mathbb{R}^n$ from linear measurements $\langle \mathbf{A}, \psi_i \rangle$ with respect to a dictionary of ψ_i's. Recently, there is focus on the novel direction of *Compressed Sensing* [1] where the reconstruction can be done with very few—$O(k \log n)$—linear measurements over a modified dictionary if the signal is *compressible*, that is, its information is concentrated in k coefficients with the original dictionary. In particular, these results [1], [2], [3] prove that there exists a single $O(k \log n) \times n$ measurement matrix such that any such 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 and because of its potential applications.

In this paper, we address outstanding open problems in Compressed Sensing. Our main result is an explicit construction of a non-adaptive measurement matrix and the corresponding reconstruction algorithm so that with a number of measurements polynomial in $k, \log n, 1/\varepsilon$, we can reconstruct compressible signals. This is the first known polynomial time explicit construction of any such measurement matrix. In addition, our result improves the error guarantee from $O(1)$ to $1 + \varepsilon$ and improves the reconstruction time from $\text{poly}(n)$ to $\text{poly}(k \log n)$.

Our second result is a randomized construction of $O(k \, \text{polylog}(n))$ measurements that work for each signal with high probability and gives per-instance approximation guarantees rather than over the class of all signals. Previous work on Compressed Sensing does not provide such per-instance approximation guarantees; our result improves the best known number of measurements known from prior work in other areas including Learning Theory [4], [5], Streaming algorithms [6], [7], [8] and Complexity Theory [9] for this case.

Our approach is combinatorial. In particular, we use two parallel sets of group tests, one to filter and the other to certify and estimate; the resulting algorithms are quite simple to implement.

I. Introduction

We study a modern twist to a fundamental problem in sparse approximation theory, called *Compressed Sensing* which has been proposed recently in the Mathematics community.

II. Sparse Approximation Theory Background.

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 $i = j$.[1] 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.[2] From now on (for convenience of reference only), we reorder the vectors in the dictionary so $|\theta_1| \geq |\theta_2| \geq \ldots \geq |\theta_n|$.

In the area of sparse approximation theory [10], 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, $\mathbf{R}_{\text{opt}}^k$, therefore takes k coefficients with the largest $|\theta_i|$'s. The error then is $\|\mathbf{A} - \mathbf{R}_{\text{opt}}^k\|_2^2 = \sum_{i=k+1}^n \theta_i^2$. This is the error in representing the signal $\mathbf{A}$ in a compressed form using k coefficients from Ψ.

In any application (say audio signal processing), one has a "class" of input signals ($\mathbf{A}$'s) (e.g., sinusoidal waveforms comprising the audio signal), one chooses an appropriate dictionary Ψ (say discrete Fourier) 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. In the past, e.g., in audio applications, researchers focused on the α-*exponentially decaying case* where the coefficients decay faster than any polynomial. That is, for some α, $|\theta_i| = O(2^{-\alpha i})$, for all i. More recently, there is focus on the p-*Compressible case*. Specifically the coefficients have a power-law decay: for some $p \in (0, 1)$, and for all i, $|\theta_i| = O(i^{-1/p})$. Consequently, $\|\mathbf{A} - \mathbf{R}_{\text{opt}}^k\|_2^2 \leq C_p k^{1-2/p}$ 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 $\mathbf{R}_{\text{opt}}^k = \mathbf{A}$.

Study of sparse approximation problems involves the art of identifying suitable Ψ so signals from an application are compressible, and studying their mathematical properties. This is a mature area of Mathematics with many applications to signal processing, communication theory and compression [10].

---

[1] Examples of other basis are *discrete Fourier* where $\psi_{i,j} = \frac{1}{\sqrt{n}} \exp(-2\pi\sqrt{-1}ij/n)$; and *Haar wavelet* where every ψ_i is a scaled and shifted copy of the same step like function. By applying an appropriate rotation to the basis and signal vectors, our problem can be thought of in the standard basis only.

[2] We refer to θ_i where $\mathbf{A}$ is implicitly clear.

III. COMPRESSED SENSING.

Recently, Donoho posed a fundamental question [1]: 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 two years, the following result has emerged:

Theorem 1: [1], [2], [3] 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 $\|\mathbf{A} - \mathbf{R}\|_2^2 = O(k^{1-2/p})$.

There are several important points to note. First, 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. Second, even if the signal consisted of precisely k nonzero coefficients $\theta_{i_1}, \ldots, \theta_{i_k}$, one needs k measurements $\langle \mathbf{A}, \psi_{i_j} \rangle$ for $j \in [1,k]$; hence, the set V of measurements is only a $\log(n/k)$ factor larger than the naive lower bound of measurements needed. Third, the proof shows existence of V by showing that a random set of V vectors will satisfy the theorem with nonzero probability. The proof immediately gives a *Monte Carlo* randomized algorithm by using such a random V.

This result has generated much interest, and a sequence of papers have improved different aspects of the result [1], [11], [2], [3]; found interesting applications including MR imaging [12] wireless communication [3] and generated implementations [13]; found mathematical applications to coding and information theory [14]; and extended the results to noisy and distributed settings [15]. The interest arises for two main reasons. First, there is deep mathematics underlying the results, with interpretations in terms of high dimensional geometry [3], uncertainty principles [2], and linear algebra [1]. Second, there are serious applications—for example, in going from analog to digital representation of the signals, existing hardware chips can execute measurements $\langle \mathbf{A}, v_i \rangle$ extremely efficiently, so performing $O(k \log(n/k))$ measurements is significantly more efficient than measuring each component of the signal (hence "compressed sensing"). The results have inspired a number of recent workshops, meetings and talks [12], [16], [17].

IV. OUTSTANDING PROBLEMS AND OUR RESULTS.

There are several outstanding questions in Compressed Sensing.

The most fundamental issue is to explicitly construct the non-adaptive measurement set of vectors V (or equivalently, a *transformation matrix* T in which $T[i,j] = v_i[j]$) in the theorem. The existing results first show that if T satisfies certain conditions, the theorem holds; then 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 [1], and

testing that each such submatrix is an isometry, behaving like an orthonormal system [2]. No explicit construction is known to produce T's with these properties. Instead, algorithms for Compressed Sensing choose a random T, and assume that the conditions are met. Thus, these are *Monte Carlo* algorithms, with some probability of failure. This is a serious drawback for Compressed Sensing applications motivated by hardware implementations which will sense many, many signals over time. So it is highly desirable that there be an explicit construction of T suitable for Compressed Sensing. A natural approach is to take a random T and test whether it satisfies the necessary conditions. However, this is much too expensive, taking time at least $\Omega(n^{k \log n})$.

There are several other outstanding questions. For example, the time to obtain a representation from the measurements is significantly *superlinear* in n (it typically involves solving a Linear Program [1], [2], [3]). For large signals, this cost is overly burdensome. Since we make a small number of measurements, it is much better to find algorithms with running time polynomial in the number of measurements and hence, sublinear in n. Lastly, the guarantee given by the above theorem is not relative to the best possible for the given signal (i.e., *per-instance*), but to the *worst case* over the whole class of p-compressible signals. Clearly per-instance error guarantees (equivalently, true approximation algorithms) are preferable.

We address these questions and present the first known explicit algorithms for Compressed Sensing. Our approach is combinatorial, and yields a number of technical improvements such as sublinear time reconstruction, and tolerance to error. Our main results are twofold.

A. Deterministic Algorithms

We present a deterministic algorithm that in time polynomial in k and n constructs a non-adaptive transformation matrix T of number of rows polynomial in $k \log n$, and present an associated reconstruction algorithm in the spirit of Theorem 1. More specifically, our algorithm outputs a representation $\mathbf{R}$ for a compressible signal $\mathbf{A}$ such that $\|\mathbf{R} - \mathbf{A}\|_2^2 < \|\mathbf{R}_{\text{opt}}^k - \mathbf{A}\|_2^2 + \varepsilon \|\mathbf{C}_{\text{opt}}^k\|_2$. Here, $\|\mathbf{C}_{\text{opt}}^k\|_2$ denotes the optimal error over the whole class of signals considered. This is the first explicit construction known for this problem in polynomial time.

In addition, this result leads to the following improvements: (a) the reconstruction time is subquadratic in the number of measurements (and hence sublinear in n), (b) the overall error is optimal up to $1 + \varepsilon$ of the worst case error $\|\mathbf{C}_{\text{opt}}^k\|_2$ for p-compressible signals, improving the $O(1)$ approximation factor in prior results, and (c) the approach applies to other cases of compressible signals with tighter bounds. For the exponentially decaying and k-sparse family, the size of T is only $O(k^2 \operatorname{polylog}(n))$. The algorithms are simple and easy to implement, without linear programming and without running into precision-issues inherent in the choice of Gaussian random T in prior methods. The formal statement of our results are as follows:

Theorem 2: We can construct a set of measurements for a signal $\mathbf{A}$ in time polynomial in k and n and return a $\mathbf{R}$ for $\mathbf{A}$ of at most k coefficients $\hat{\theta}$ under Ψ such that $\|\hat{\theta} - \theta\|_2^2 = \|\mathbf{R} - \mathbf{A}\|_2^2 < \|\mathbf{R}_{\text{opt}}^k - \mathbf{A}\|_2^2 + \varepsilon \|\mathbf{C}_{\text{opt}}^k\|_2^2$, and (i) if $p < \frac{1}{2}$, then the number of measurements is $O((k\varepsilon^p)^{4/(1-2p)} \log^4 n)$ and the time to produce the coefficients from the measurements is $O((k\varepsilon^p)^{6/(1-2p)} \log^6 n)$. (ii) if the p-compressible case is tight, i.e. $|\theta_i| = \Theta(i^{-1/p})$ then the number of measurements is $O((k\varepsilon^p)^{4/(1-p^2)} \log^4 n)$ and the time to find coefficients is $O((k\varepsilon^p)^{6/(1-p)^2} \log^6 n)$.

Theorem 3: We can construct a set of $O(k^2 \operatorname{polylog}(n))$ measurements in time polynomial in k and n. For any α-exponentially decaying signal or k-sparse signal, $\mathbf{A}$, from these measurements of $\mathbf{A}$, we can return a representation $\mathbf{R}$ for $\mathbf{A}$ of at most k coefficients $\hat{\theta}$ under Ψ such that $\|\hat{\theta} - \theta\|_2^2 = \|\mathbf{R} - \mathbf{A}\|_2^2 < \|\mathbf{R}_{\text{opt}}^k - \mathbf{A}\|_2^2 + \varepsilon \|\mathbf{C}_{\text{opt}}^k\|_2^2$. The time required to produce the coefficients from the measurements is $O(k^2 \operatorname{polylog}(n))$

B. Randomized Algorithms

We address the issue of obtaining per-instance guarantees for each signal. We present a randomized algorithm that on any given $\mathbf{A}$, produces a T with $O(\frac{k}{\epsilon^2} \operatorname{polylog}(n))$ rows such that in time linear in $O(k \operatorname{polylog}(n))$, we can reconstruct a $\mathbf{R}$ with $\|\mathbf{A} - \mathbf{R}\|_2^2 \leq (1 + \varepsilon)\|\mathbf{A} - \mathbf{R}_{\text{opt}}^k\|_2^2$, with probability at least $1 - \frac{1}{n^{O(1)}}$.

Notice crucially that this second result does not produce a T that works for all p-compressible signals, merely that on any given signal $\mathbf{A}$, we can produce a good $\mathbf{R}$ with high probability. In this regime, which is quite different from the regime in earlier papers on Compressed Sensing where a fixed T works for all p-compressible signals, many results in the Computer Science literature apply, in particular, from learning theory [4], [5], streaming algorithms [7], [6] and complexity theory [9]. Some of these results do not completely translate to our scenario: the learning theory approaches assume that the signal can be probed in the light of the results of prior measurements (this is similar to adaptive group testing). Other results can be thought of as producing a T with $O(k^{2+O(1)} \operatorname{polylog}(n))$ rows which is improved by our result here. An exception is the result in [18] which works by sampling (that is, finding $\langle \mathbf{A}, v_i \rangle$ where $v_{i,j} = 1$ for some j and is 0 elsewhere) for the Fourier basis, but can be thought of as solving our problem using $O(k \operatorname{polylog}(1/\varepsilon, \log n, \log \|\mathbf{A}\|))$ measurements. Our result improves [18] in the term $\operatorname{polylog}(\|\mathbf{A}\|_2)$ which governs the number of iterations in [18]. Finally, we extend to the case when the measurements are noisy—an important practical concern articulated in [19]—and obtain novel results that give per-instance approximation results. Formally, we show:

Theorem 4: We can construct a dictionary $\Psi' = T\Psi$ of $O(\frac{ck \log^3 n}{\varepsilon^2})$ vectors, in time $O(cn^2 \log n)$. For any signal $\mathbf{A}$, given the measurements $\Psi' \mathbf{A}$, we can find a representation $\mathbf{R}$ of $\mathbf{A}$ under Ψ such that with probability at least $1 - \frac{1}{n^c}$ $\|\mathbf{R} - \mathbf{A}\|_2^2 \leq (1 + \varepsilon)\|\mathbf{R}_{\text{opt}}^k - \mathbf{A}\|_2^2$. The reconstruction process takes time $O(\frac{c^2 k \log^3 n}{\varepsilon^2})$.

Theorem 5: We can construct a dictionary $\Psi' = T\Psi$ of $O(\frac{ck \log n}{\varepsilon^2})$ vectors, in time $O(cn^2 \log n)$. For any signal $\mathbf{A}$, given the measurements $\Psi' \mathbf{A}$, we can find a representation $\mathbf{R}$ of $\mathbf{A}$ under Ψ such that with probability at least $1 - \frac{1}{n^c}$ $\|\mathbf{R} - \mathbf{A}\|_2^2 \leq (1 + \varepsilon)\|\mathbf{R}_{\text{opt}}^k - \mathbf{A}\|_2^2$. The reconstruction process takes time $O(cn \log n)$.

Theorem 6: 1. If a fraction $\rho = O(1)$ of the measurements are chosen at random to be corrupted in an arbitrary fashion, we can still recover a representation $\mathbf{R}$ with error $\|\mathbf{R} - \mathbf{A}\|_2^2 \leq (1 + \varepsilon)\|\mathbf{R}_{\text{opt}}^k - \mathbf{A}\|_2^2$ in time $O(cn \log n)$.
2. If only a $\rho = O(\log^{-1} n)$ fraction of the measurements are corrupted we can recover a representation $\mathbf{R}$ with error $\|\mathbf{R} - \mathbf{A}\|_2^2 \leq (1 + \varepsilon)\|\mathbf{R}_{\text{opt}}^k - \mathbf{A}\|_2^2$ in time $O(\frac{kc^2 \log n}{\varepsilon^2})$.

V. TECHNICAL OVERVIEW.

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 collections of $O((k \log n)^2)$ nonadaptive 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. In general, there is no direct connection between $\langle \mathbf{A}, \chi_{S_i} \rangle$ and presence or absence of the largest k coefficients in S_i when the signal is from the p-compressible class. This is also the reason that prior work on this problem has delved into the linear-algebraic and geometric structure of the problem.

Our approach here is combinatorial. Our first results show that one can focus attention on some $k' > k$ coefficients, in order to meet our error guarantees. Then, we show that separating the k' coefficients using group testing methods serves as a filter and subsequently, using a different set of group tests serves to certify and estimate the largest k coefficients in magnitude. This use of two parallel sets of group tests is novel. For the second set of results, combinatorial group testing has been applied previously in Learning Theory [4], [5], Streaming Algorithms [6], [7], [8] and Complexity Theory [9]. Here, our contribution is to adapt the approach from our first set of results and provide a tighter analysis of the error in terms of $\|\mathbf{R}_{\text{opt}}^k - \mathbf{A}\|_2$ rather than in terms of $\|\mathbf{A}\|_2$ as is more typical.

Note. Preliminary versions of these results have appeared as technical reports [20], which are superseded by the results here.

VI. Concluding Remarks

We present a simple combinatorial approach of two sets of group tests with different separation properties that yields the first known polynomial time explicit construction of a non-adaptive transformation matrix and a reconstruction algorithm for the Compressed Sensing problem. The polynomial dependency is large, but we emphasize that no other construction with polynomial creation time is known, and the cost may be improved in future work. Our approach yields other results including sublinear reconstruction, improved approximation in error and others. Given the excitement about Compressed Sensing in the Applied Mathematics community, we expect many new results soon. The main open problem is to reduce the number of measurements used by explicit algorithms: our result here gives a cost polynomial in k, which is not close to the linear factor k in the existential results of [2], [1], [3]. For the case of k-sparse signals, (which have no more than k nonzero coefficients) Indyk recently developed a set of measurements, near linear in k in number (but has other superlogarithmic factors in n) [21]. Another outstanding question is to tease apart other properties of Compressed Sensing results—such as their ability to measure in one basis and reconstruct in another—and study their algorithmics.

Acknowledgments

References

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

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

[3] 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.

[4] E. Kushilevitz and Y. Mansour, "Learning decision trees using the fourier spectrum," *SIAM Journal on Computing*, vol. 22, no. 6, pp. 1331–1348, 1993.

[5] Y. Mansour, "Randomized interpoloation and approximation of sparse polynomials," *SIAM Journal of Computing*, vol. 24, no. 2, 1995.

[6] A. Gilbert, S. Guha, P. Indyk, Y. Kotidis, S. Muthukrishnan, and M. Strauss, "Fast, small-space algorithms for approximate histogram maintenance," in *Proceedings of the ACM Symposium on Theory of Computing*, 2002, pp. 389–398.

[7] A. Gilbert, S. Guha, P. Indyk, S. Muthukrishnan, and M. Strauss, "Near-optimal sparse Fourier representation via sampling," in *Proceedings of the ACM Symposium on Theory of Computing*, 2002.

[8] G. Cormode and S. Muthukrishnan, "What's hot and what's not: Tracking most frequent items dynamically," in *Proceedings of ACM Principles of Database Systems*, 2003, pp. 296–306.

[9] A. Akavia, S. Goldwasser, and S. Safra, "Proving hard-core predicates by list decoding," in *IEEE Conference on Foundations of Computer Science*, 2003, pp. 146–157.

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

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

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

[13] "Compressed sensing website," http://www.dsp.ece.rice.edu/CS/.

[14] E. Candès, M. Rudelson, T. Tao, and R. Vershynin, "Error correction via linear programming," in *IEEE Conference on Foundations of Computer Science*, 2005.

[15] 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, unpublished Manuscript.

[16] "IEEE International Conference on Acoustics, Speech, and Signal Processing," 2005, http://www.icassp2005.com/.

[17] "Integration of Sensing and Processing," 2005, http://www.ima.umn.edu/2005-2006/W12.5-9.05/.

[18] A. Gilbert, S. Muthukrishnan, and M. Strauss, "Improved time bounds for near-optimal sparse Fourier representations," in *SPIE Conference on Wavelets*, 2005.

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

[20] G. Cormode and S. Muthukrishnan, "Towards an algorithmic theory of compressed sensing," Center for Discrete Mathematics and Computer Science (DIMACS), Tech. Rep. 2005-25, 2005.

[21] P. Indyk, "Personal communication," 2005.