EXPLICIT MEASUREMENTS WITH ALMOST OPTIMAL THRESHOLDS FOR COMPRESSED SENSING

Farzad Parvaresh * *and Babak Hassibi* †

* Center for Mathematics of Information † Department of Electrical Engineering
California Institute of Technology,
1200 East California, Pasadena, CA, 91125, USA.
E-mail: {fparvaresh, hassibi}@caltech.edu

ABSTRACT

We consider the deterministic construction of a measurement matrix and a recovery method for signals that are *block* sparse. A signal that has dimension $N = nd$, which consists of n blocks of size d, is called (s, d)-block sparse if only s blocks out of n are nonzero. We construct an explicit linear mapping Φ that maps the (s, d)-block sparse signal to a measurement vector of dimension M, where $s \cdot d < N\left(1 - \left(1 - \frac{M}{N}\right)^{\frac{d}{d+1}}\right) - o(1)$. We show that if the (s, d)-block sparse signal is chosen uniformly at random then the signal can almost surely be reconstructed from the measurement vector in $O(N^3)$ computations.

***Index Terms*—** Convex optimization, sparse signals, Reed-Solomon codes, decoding algorithms, compressed sensing.

1. INTRODUCTION

Consider the set of signals of dimension N with at most s nonzero element over $\mathbb{C}^N$. This set of signals spans the union of $\binom{N}{s}$ subspaces of dimension s over $\mathbb{C}^N$. If we project these subspaces to a random subspace of dimension $s+1$, then with high probability we get a one to one mapping between the projected sparse signals and the original ones. The recent results of Candés, Donoho, Romberg, and Tao [2, 3, 5], applied to applications such as tomography and digital photography, have revealed the power of random sampling. Recently, many other applications for compressed sensing have been developed in areas such as data mining, DNA microarrays, and A/D converters.

Let $\Phi_{M,N}$ denotes the linear measurement matrix, so that the samples or the measurements of a sparse signal $x \in \mathbb{C}^N$ become $y = \Phi \cdot x \in \mathbb{C}^M, M \geqslant s + 1$. To reconstruct the signal x from the measurement vector y one needs to solve the underdetermined linear system of equations $\Phi x = y$, for a given y, under the condition that x is a s-sparse signal. This can be represented as the following optimization problem:

$$\min_x \|x\|_0 \text{ subject to } \Phi x = y \tag{1}$$

Here the ℓ_0 norm or the Hamming norm is the number of nonzero elements of x.

A naive exhaustive search checks all the possible $\binom{N}{s}$ nonzero coordinates for the signal x to find the minimum and that takes an exponential time in N. However, one may try to solve (1) by relaxing the ℓ_0 norm to ℓ_1 norm.

$$\min_x \|x\|_1 \text{ subject to } \Phi x = y \tag{2}$$

Assume δ is equal to M/N and ρ is s/M and the measurement matrix Φ is chosen uniformly at random from the set of linear projections from $\mathbb{C}^N$ to $\mathbb{C}^M$. Donoho and Tanner [5, 11] determined the region (δ, ρ) for which the ℓ_1 optimization and ℓ_0 coincide. They compute two different types of "strong", $\rho_S(\delta)$, and "weak", $\rho_W(\delta)$, threshold functions. The strong threshold function ensures that ℓ_1 and ℓ_0 are equivalent for $s < M\,\rho_S(M/N)$ with overwhelming probability in the uniform selection of measurement matrix Φ. For the weak threshold the equivalence between (1) and (2) holds for *most* signals x when $s < M\,\rho_W(M/N)$ with overwhelming probability in uniform selection of Φ. (cf. Figure 1)

How much do we pay by relaxing the ℓ_0 optimization to ℓ_1? Let's define $\rho_S^{\mathsf{opt}}(\delta)$ and $\rho_W^{\mathsf{opt}}(\delta)$ to be the supremum of all the threshold functions over all the linear measurements. We know that $\rho_S^{\mathsf{opt}} = 1/2, \rho_W^{\mathsf{opt}} = 1$. There is a large gap between the storing and week thresholds $\rho_{S(W)}(\delta)$ and $\rho_{S(W)}^{\mathsf{opt}}(\delta)$.

How do we choose the measurement matrix Φ? In most of the literature in compressed sampling the measurement matrix is an instance of a class of random matrices. Then, with overwhelming probability, Φ satisfies certain reconstruction

This work was supported in parts by the National Science Foundation under grants no. CCR-0133818 and CCR-0326554, by the David and Lucille Packard Foundation, and by Caltechs Lee Center for Advanced Networking.

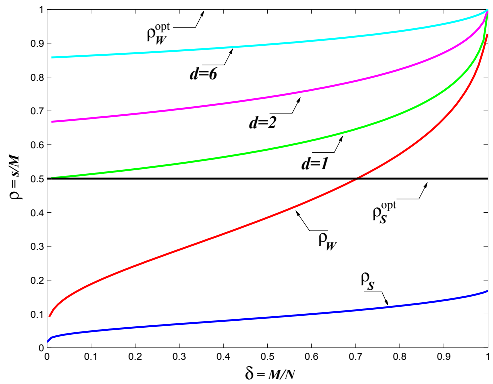

Fig. 1. *Thresholds for the recovery of sparse signals. ρ_S is the strong threshold and ρ_W is the weak threshold for linear program. The rest are thresholds for recovery of (s, d)-block sparse signals. Using our proposed method the improvement over ρ_W is clear. As d grows, our threshold approaches ρ_W^{opt}.*

properties. However, there is no efficient method for *verifying* that a given matrix has these properties [12]. Recently, a line of research on compressed sensing has been devoted to the *explicit* construction of the measurement matrix, however, the threshold functions of these explicit constructions are usually worst than those of ℓ_1 optimization.

1.1. Contributions

A connection between compressed sensing and Reed-Solomon codes over the complex filed is already implicit in various works in the literature, often under different names such as annihilator filters and recovery of a measure from its moments [1, 7]. In this paper, we make this connection explicit by choosing the measurement vector to be essentially the syndrome of the code. The sparse signal can then be recovered by any well known decoding algorithm such as Berlekamp-Massey for RS codes.

However, recently, there have been many remarkable breakthroughs in list-decoding of Reed-Solomon codes such as Sudan [10] and Guruswami-Sudan [6] algorithms; To the best of our knowledge we are not aware of any research on using these classes of algorithms for compressed sampling. In a Reed-Solomon code of length N and dimension K, the Berlekamp-Massey decoder only needs the syndrome vector of dimension $N-K$ to find the *error locater* polynomial. The dimension of the syndrome vector is smaller than the dimension of the received word; that is equivalent to having a measurement vector with smaller dimension than the sparse signal in compressed sampling. However, in the list-decoding algorithms the whole received word is being used, and not the syndromes, to perform the decoding. One contribution of ours is to show that one can construct a "received word" out of the syndrome vector to perform the list-decoding algorithm for compressed sensing applications.

One of the crucial steps of all the Sudan-type list-decoding algorithms is to factor a bivariate polynomial over the underlying field. This factorization can be done efficiently over finite fields but we are not aware of any efficient algorithm for factoring a bivariate polynomial over the complex field. Instead of list-decoding, we propose to use Coppersmith and Sudan [4] decoding. This algorithm is probabilistic and decodes with probability of $1-O(N^c/q)$ where, c is a constant, q is the size of the finite field, and N is the length of the code; considering that errors are generated uniformly at random in $\mathbb{F}_q$. However, we show that over the complex field, the Coppersmith-Sudan algorithm will *almost surely* recover the random sparse signal. In computer science, Reed-Solomon codes have mostly been used at rates that approaches zero and the authors in [4] basically give decoding bounds that are suitable for these rates. We use more advance tools from algebra, such as working with the Gröbner basis of certain ideal of polynomials and we improve the decoding bound of [4]. The new bound shows improvement compare to conventional decoding algorithms for all rates in $[0, 1]$.

In addition, the Coppersmith-Sudan algorithm can be used to recover curves in three and more dimensions. That is tantamount to the possibility of recovering ***block*** sparse signals with a small number of measurements. Consider a signal of dimension N which consists of n blocks of size $d = N/n$. We say the signal is (s, d)-block sparse if only s blocks out of n is nonzero. We show that using syndrome measurements one can almost surely recover an (s, d)-block sparse signal from M measurements efficiently if $s \cdot d < N\left(1 - \left(1 - \frac{M}{N}\right)^{\frac{d}{d+1}}\right) - o(1)$ (Check Figure 1 for the plot of thresholds).

2. THE MEASUREMENT MATRIX

Let $\mathbb{C}$ denote the complex field. We use $\mathbb{C}[X_1, X_2, \cdots, X_c]$ to denote the rings of polynomials over $\mathbb{C}$ in several variables. Reed-Solomon codes are obtained by evaluation of certain function in $\mathbb{C}[X]$ in a set of points $\mathcal{D} = \{\omega_0, \omega_1, \cdots, \omega_{N-1}\}$ in $\mathbb{C}$. Throughout this work we choose $\omega_i = \lambda^i$ for $i = 0, 1, \cdots, N-1$, where λ is the N-th root of unity. [1] A Reed-Solomon code $\textsf{\textit{RS}}(N, K)$ of length N and dimension K over the complex field is defined as follows:

$$\textsf{\textit{RS}}(N, K) \stackrel{\text{def}}{=} \{(f(\omega_0), f(\omega_1), \cdots, f(\omega_{N-1})) : \\ f(X) \in \mathbb{C}[X], \deg f < K\}$$

Notice that $\textsf{\textit{RS}}(N, K)$ is a subspace of dimension K in $\mathbb{C}^N$. Define $\textsf{\textit{Synd}}(N, K)$ to be the orthogonal space to

[1] In the general case ω_i's can be any set of different numbers in $\mathbb{C}$. We choose them to be on the unit circle with equal distance so that the measurement becomes the inverse Fourier transform.

$\mathit{RS}(N, K)$.

$$\mathit{Synd}(N, K) \stackrel{\text{def}}{=} \{(v_1, v_2, \cdots, v_N) \in \mathbb{C}^N : \langle v, c\rangle = 0, \text{ for all } c \in \mathit{RS}(N, K)\}$$

We call $\mathit{Synd}(N, K)$ the *syndrome* or the *measurement* space.

Definition 1 *Set $K = N - M$ and consider the corresponding linear projection Φ from $\mathbb{C}^N$ to $\mathit{Synd}(N, K)$. We define the M-dimensional measurement of x to be*

$$y = \Phi \cdot x. \tag{3}$$

Lemma 1 *Assume that the evaluation points of the RS code are the consecutive powers of the N-th root of unity, i.e. $\omega_i = \lambda^i$ for $i = 0, 1, \cdots, N-1$, then the measurement vector y in (3) is the inverse Fourier transform of x at frequencies $\omega_K, \omega_{K+1}, \cdots, \omega_{N-1}$.*

Lemma 2 *Any vector $x \in \mathbb{C}^N$ can be written uniquely as a summation of vectors $r \in \mathit{Synd}(N, K)$ and $c \in \mathit{RS}(N, K)$:*

$$\begin{array}{ll} y = \Phi x & \in \mathbb{C}^{N-K} \\ r = \Phi^\dagger y & \in \mathit{Synd}(N, K) \\ c = x - r & \in \mathit{RS}(N, K) \end{array} \tag{4}$$

where $\Phi^\dagger$ is the conjugate transpose of Φ.

For a given measurement vector y of the s-sparse signal x we construct the "received vector" $r = \Phi^\dagger y$. Now, from Lemma 2, we know that $r = x + c$ for some $c \in \mathit{RS}(N, K)$. That means, r is simply a RS codeword c that has been corrupted at s positions. Thus, for example, if we use the Berlekamp-Massey algorithm to decode r, as far as the number of corrupted coordinates is smaller than half the minimum distance of the code $s < (N - K)/2 = M/2$, the decoder outputs the codeword c and sparse signal x. In the next section, we explore the possibility of using other RS decoding algorithms for compressed sensing.

3. RECOVERY FROM THE MEASUREMENTS

Now that we have established a connection between the recovery of the sparse signal x from the measurement vector $y = \Phi x$ and the RS decoding of $r = \Phi^\dagger y$, we can use other advanced decoding algorithms such as the list-decoding algorithm of Guruswami-Sudan [6] for recovery. The bottleneck of the Guruswami-Sudan algorithm over complex fields is the factorization part. We are not aware of any efficient factorization algorithm over the complex field. Considering the fact that there are many efficient algorithms to factor univariate polynomials over the complex field, one can use the Ruth-Ruckenstein [9] algorithm. However, the algorithm, in principle, is sensitive to numerical inaccuracies.

Another elegant decoding algorithm with bounds comparable to the list-decoding algorithm of GS was introduced by Coppersmith and Sudan [4]. Their algorithm does not rely on tools such as the factoring of multivariate polynomials. Basically, given a received word, they construct a matrix $\mathbf{A}$ such that the right kernel of $\mathbf{A}$ with high probability consists of vectors with support that is entirely on the "non-error" coordinates of the received vector.

We show that, over the complex field their algorithm *almost surly* recovers the codeword if the sparse signal is chosen uniformly at random over $\mathbb{C}^N$, we further improve the bounds [4] and show that the performance of the algorithm is comparable to the list-decoding algorithm of GS at all rates in $[0, 1]$.

Notations and definitions. Given Δ, let $\mathcal{M}_{K,\Delta}$ be the set of monomials X^aY^b with $a + (K-1)b \leqslant \Delta$. For a positive integer p, let $\mathcal{S}_p = \{(d, e) : d + e < p\}$. Given $(d, e) \in \mathcal{S}_p$, let $\mathbf{f}^{[d,e]}_{\alpha,\beta;\mathcal{M}}$ be the vector in $\mathbb{C}^{|\mathcal{M}|}$ whose coordinates are indexed by monomials M in $\mathcal{M}$ and whose Mth coordinate is $\frac{\partial^{d+e}}{\partial X^d \partial Y^e} X^aY^b|_{(\alpha,\beta)}$ if $M = X^aY^b$.

Algorithm 1 Coppersmith-Sudan decoding algorithm

Input: Received vector $r \in \mathbb{C}^N$, multiplicity p, and codeword dimension K.
Output: Codeword $c \in \mathbb{C}^N$ or **FAIL.**

1: **Parameters:** Set Δ sufficiently large such that $|\mathcal{M}_{K,\Delta}| \geqslant N \cdot |\mathcal{S}_p|$.
2: **Step 1:** Let $\mathbf{A}$ be the matrix whose columns are indexed by pairs $(i, (d, e))$ with $i \in \{0, 1, \cdots, N-1\}$ and $(d, e) \in \mathcal{S}_p$ where the $(i, (e, d))$th column is $\mathbf{f}^{[d,e]}_{\omega_i, r_i;\mathcal{M}}$. Let $\mathbf{b}$ be a non-zero vector such that $\mathbf{A} \cdot \mathbf{b} = \mathbf{0}$.
3: **Step 2:** Let J be the set of all indices $i \in \{0, 1, \cdots, N-1\}$ such that there exists a tuple $(d, e) \in \mathcal{S}_p$ for which the $(i, (d, e))$th coordinate of $\mathbf{b}$ is nonzero.
4: **if** there exists a polynomial $f(X)$ with $\deg f(X) < K$ such that $f(\omega_i) = r_i$ for every $i \in J$ **then**
5:    **return** $c = (f(\omega_0), f(\omega_1), \cdots, f(\omega_{N-1}))$.
6: **else**
7:    **return FAIL.**
8: **end if**

4. ANALYSIS OF THE ALGORITHM

Due to lack of space we omit the proofs. For details, the reader is referred to [8]. Let I denote the set of non-error positions of r and $t = |I|$. Let $f(X)$ be the corresponding polynomial of the RS codeword c. We prove the following Lemmas. First, the matrix $\mathbf{A}$ does have a rank less than $N \cdot |\mathcal{S}_p|$ and thus a vector $\mathbf{b}$ as required in **Step 1** does exist. Second, with high probability the subset J found in **Step 2** is a subset of I. Third, the size of J is at least K and so there is at most **one** polynomial $f(X)$ of degree less than K that interpolates

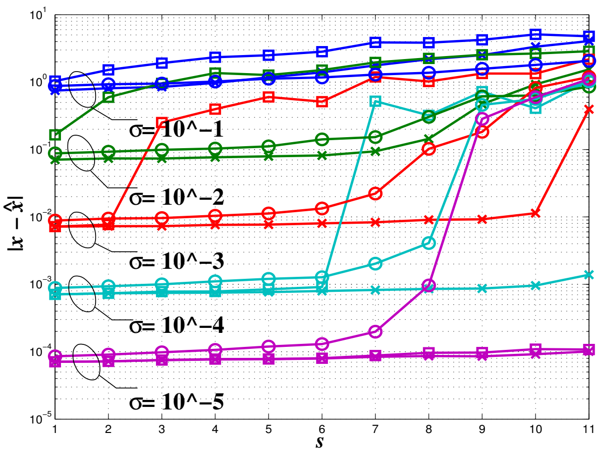

Fig. 2. *Simulation results at different noise levels.* $N = 50$, $M = 25$, $\square$ *is for the Berlekamp-Massey,* $\times$ *is for the Coppersmith-Sudan,* $\circ$ *is for the Lasso.*

through points of J.

Let $\mathbf{B}$ be the $|\mathcal{M}| \times (|\mathcal{S}_p| \cdot t)$ matrix consisting of those columns of $\mathbf{A}$ that correspond to $i \in I$.

Lemma 3 (i) *If* $t > \Delta/p$, *then the matrix* $\mathbf{B}$ *has column dependency.* (ii) *There are no column dependencies in* $\mathbf{B}$ *involving fewer that* H *blocks of columns, provided* $\Delta > pH + p(p+1)/2$. (iii) *Almost surely, the matrix* $\mathbf{A}$ *has no linear dependencies involving any of the columns indexed by* $(i, (e, d))$ *where* $i \notin I$, *provided* $|\mathcal{M}_{\Delta-(2p+1)}| > N \cdot |\mathcal{S}_p|$.

Theorem 1 *For every fixed constant* d, *using the syndrome measurement matrix with the Coppersmith-Sudan decoding algorithm we can almost surely recover* (s, d)*-block sparse signals from* $M = \delta N$ *measurements if*

$$S < M\frac{1-(1-\delta)^{\frac{d}{d+1}}}{\delta} - o(1) \tag{5}$$

where $S = s \cdot d$ *is the number of nonzero elements of the sparse signal, and* N *is size of the sparse signal.*

Remark. When $d = 1$, then (5) reduces to $S < M(1 - \sqrt{1-\delta})/\delta$, which is greater than $1/2$ for all $\delta \in [0, 1]$.

5. ROBUSTNESS TO NOISE

In practice the measurement vector is usually corrupted by noise. Let $\mathcal{N}(0, \sigma)$ be a complex Gaussian random variable with zero mean and standard deviation σ. We assume that

$$y_w = \Phi \cdot x + w \tag{6}$$

where $w \in \mathcal{N}^M(0, \sigma)$. We choose x at random, i.e. the support of x is chosen uniformly at random from all the possible $\binom{N}{s}$ subsets and the values are drawn *i.i.d.* form $\mathcal{N}(0, 1)$. Due to the noise, the matrix $\mathbf{A}$ is full rank, so in Algorithm 1 we choose $\mathbf{b}$ to be the right singular vector with the smallest singular value. Figure 2 shows the median of the squared error $\|x - \hat{x}\|_2$ as a function of the sparsity s. From the Figure 2, Algorithm 1 is more robust to noise than the BM algorithm. We also compare the performance to the well known LASSO algorithm [13] which minimizes $\|x - \hat{x}\|_2$ with an ℓ_1 constraint.

6. REFERENCES

[1] M. Akçakaya and V. Tarokh, Performance Bounds on Sparse Representations Using Redundant Frames, arXiv:cs/0703045.

[2] E. Candés, J. Romberg, and T. Tao, Robust uncertainty princples: Exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. on Inform. Theory*, **52**, pp. 489–509, 2006.

[3] E. Candés and T. Tao, Near optimal singal recovery from random projections: Universal encoding strategies, *IEEE Trans. on Inform. Theory*, 2006.

[4] D. Coppersmith and M. Sudan, Reconstructing curves in three (and higher) dimensional spaces from noisy data, *Proc. 35th STOC* , pp. 136–142, San Diego, CA., June 2003.

[5] D. L. Donoho, High-dimensional centrally-symmetric polytopes with neighborliness proportional to dimension, *Disc. Comput. Geometry*, December 2005.

[6] V. Guruswami and M. Sudan, Improved decoding of Reed-Solomon and algebraic-geometric codes, *IEEE Trans. on Inform. Theory*, **45**, pp. 1755–1764, Septebmer 1999.

[7] I. Maravić and M. Vetterli, Sampling and Reconstruction of Signals With Finite Rate of Innovation in the Presence of Noise, *IEEE Trans. On Signal Proc.*, **53(8)**, pp. 2788–2805, August 2005.

[8] F. Parvaresh and B. Hassibi, Compressed sensing with almost optimal thresholds, *in preperation.*

[9] R.M. Roth and G. Ruckenstein, Efficient decoding of Reed-Solomon codes beyond half the minimum distance, *IEEE Trans. on Inform. Theory*, **46**, pp. 246–258, January 2000.

[10] M. Sudan, Decoding of Reed-Solomon codes beyond the error-correction bound, *Journal of Complexity*, **13(1)**, pp. 180–193, 1997.

[11] J. Tanner and D Dohono, Thresholds for the Recovery of Sparse Solutions via L1 Minimization, *Proc. of the Conf. on Inform. Sci. and Sys.*, March 2006.

[12] T. Tao, Open question: deterministic UUP matrices, Weblog at http://terrytao.wordpress.com/2007/07/02/open-question-deterministic-uup-matrices 2007.

[13] R. Tibshirani, Regression shrinkage and selection via the lasso, *Journal of the Royal Stat. soc. B*, **58(1)**, pp. 267–288, 1996.