Compressed Sensing and Linear Codes over Real Numbers

Fan Zhang and Henry D. Pfister
Department of Electrical and Computer Engineering
Texas A&M University
{fanzhang,hpfister}@tamu.edu

***Abstract*—Compressed sensing (CS) is a relatively new area of signal processing and statistics that focuses on signal reconstruction from a small number of linear (e.g., dot product) measurements. In this paper, we analyze CS using tools from coding theory because CS can also be viewed as syndrome-based source coding of sparse vectors using linear codes over real numbers. While coding theory does not typically deal with codes over real numbers, there is actually a very close relationship between CS and error-correcting codes over large discrete alphabets. This connection leads naturally to new reconstruction methods and analysis. In some cases, the resulting methods provably require many fewer measurements than previous approaches.**

I. Introduction

Compressed sensing (CS) is a framework for signal sensing and compression that finds it origin in the sparse representation of signals [4], [6], [3], [5], [2]. The basic idea is that any n-dimensional signal, with a k-sparse representation in some basis, can be reconstructed from m fixed linear dot-product measurements where $m = O(k \ln n)$. In particular, m random linear projections are often sufficient for reconstruction of the signal. In many applications, the two relevant performance measures are the oversampling ratio m/k required for good performance and the computational complexity of reconstruction.

Optimal decoding (in terms of oversampling ratio) requires a combinatorial search that is known to be NP-Hard [3]. Practical reconstruction algorithms tend to either be based on linear programming (e.g., basis pursuit (BP) [4]) or low-complexity iterative algorithms (e.g., Orthogonal Matching Pursuit (OMP) [19]). A wide range of algorithms allow one to trade-off the oversampling ratio for reconstruction complexity.

Representing an n-dimensional signal using $m < n$ measurements is largely an exercise in compression. Therefore, there is a close connection between CS and information theory. In particular, signal compression via random linear projections (sometimes called *syndrome source coding*) dates back to the 1970s and is usually associated with Slepian-Wolf coding [18], [21], [1].

In this paper, we explore the connection between CS, syndrome source coding, and error-correcting codes. Recent progress in capacity-achieving codes allows us to propose and analyze sampling/reconstruction algorithms based on sparse-graph codes and message-passing decoding [12]. The algorithms have linear complexity in n and offer nearly optimal oversampling ratios. Some challenges remain, however, for their application in some CS systems.

There are a few significant differences between coding theory and CS. The first is that coding theory typically uses discrete alphabets (see [20] for an exception to this) while CS deals with signals over the real numbers. Fortunately, some codes designed for large discrete alphabets (e.g., the q-ary symmetric channel) can be adapted to the real numbers. The second is that CS is typically interested in the case where the signal is very sparse. From a syndrome source coding perspective, this implies that code rate is very high. The third is distinction between reconstruction of all sufficiently sparse signals (called *uniform reconstruction*) versus almost-all sufficiently sparse signals (called *non-uniform reconstruction*). Though coding theory and CS deal with both types, CS tends to focus more on uniform reconstruction while coding theory tends to focus more on non-uniform reconstruction.

II. Background

A. Problem Statement

Compressed sensing (CS) originated with the observation that many signal processing systems sample a large amount of data (e.g., at the Nyquist rate), perform a linear transform (e.g., to a wavelet basis), and then throw all the small coefficients away. If the locations of the transform coefficients with large magnitude are known in advance, then one could just sample those values directly and reduce the complexity enormously. Compressed sensing is field that emerged when researchers realized that random sampling could be used (with a small penalty) to achieve the same result without prior knowledge of where the coefficients of large magnitude are located.

Let $x \in \mathbb{R}^n$ be the signal vector, $\Phi \in \mathbb{R}^{m \times n}$ be the $m \times n$ measurement matrix, and $s = \Phi x$ be the m linear observations of x. Given the observation s, the valid set of signal vectors is

$$V(s) = \{x \in \mathbb{R}^n | \Phi x = s\}. \tag{1}$$

Since there is clearly no unique inverse when $m < n$, we can only choose the "best" solution based on prior knowledge of the signal. For example, if the signal components are i.i.d. zero mean Gaussian random variables, then the maximum-likelihood (ML) solution is the least-square solution

$$\hat{x} = \arg \min_{x \in V(s)} \|x\|_2 = \Phi^T \left(\Phi \Phi^T\right)^{-1} s, \tag{2}$$

where $\left(\|x\|_p\right)^p = \sum_{i=1}^n |x_i|^p$ for $p \in (0,\infty)$. If instead, we know that the signal x is strictly sparse (e.g., entries are non-zero with probability $p < 1/2$), then the ML solution (if unique) satisfies

$$\hat{x} = \arg\min_{x\in V(s)} \|x\|_H , \tag{3}$$

where $\|x\|_H = \lim_{p\to 0} \|x\|_p^p$ is the Hamming weight of x. In both cases, the purpose of the prior distribution for x is simply to provide a partial ordering of the set $V(s)$ so that one of the "best" elements can be chosen. In general, if the signal components are drawn i.i.d. according to $f_X(x) = \left[2\Gamma\left(\frac{\alpha+1}{\alpha}\right)\right]^{-1} e^{-|x|^\alpha}$, then the ML solution (if unique) satisfies

$$\hat{x} = \arg\min_{x\in V(s)} \|x\|_\alpha . \tag{4}$$

B. Connections with Coding Theory

Compressed sensing is closely related to a number of other well-established research areas such as signal processing, statistics, approximation theory, and information theory. In approximation theory, the signal is modeled as an element in a set $\mathcal{S} \subset \mathbb{R}^n$ and one seeks an m-dimensional linear projection that minimizes the maximum approximation error with respect to some metric [5], [11]. In this case, good upper/lower bounds have been derived for many interesting cases. Information theory and coding theory have also been use to get interesting results for CS [8], [16], [15], [22]. However, we believe this connection still has many interesting results to offer.

Coding theory typically deals with error correction in a message transmission setting. For example, a codeword $c \in \mathbb{F}^n$ is chosen from the codebook $\mathcal{C}$ (over the field $\mathbb{F}$) and transmitted through a channel and received as $\hat{c}$ where $\hat{c} = c + e$ and $e \in \mathbb{F}^n$ is the additive noise. With the received word $\hat{c}$ and the knowledge of the codebook $\mathcal{C}$, the decoder tries to estimate c (or equivalently e). For linear codes, the codebook $\mathcal{C}$ can be described compactly in terms of its parity-check matrix $H \in \mathbb{F}^{m\times n}$ as $\mathcal{C} = \{c \in \mathbb{F}^n | Hc = 0\}$.

A syndrome decoder works by first calculating the syndrome $s = H\hat{c} = H(c + e) = He$ and then finding the most likely error patten $\hat{e}$ given the syndrome [10]. We note that, in contrast to standard coding theory, we assume that s, c, e, x are column vectors. If the probability of an error vector e is strictly decreasing w.r.t. some weight function, then the syndrome decoder is equivalent to the minimum distance decoder w.r.t. that weight function. For the q-ary symmetric channel, the error probability is strictly decreasing w.r.t. Hamming weight and the syndrome decoder is a minimum Hamming distance decoder. In this case, error correction is only successful if e is the sparsest vector that satisfies $s = He$.

A syndrome source coding system uses a linear code with parity-check matrix $H \in \mathbb{F}^{m\times n}$ to compress a vector $x \in \mathbb{F}^n$ by computing its syndrome $s = Hx$ [1]. Given a prior distribution on x, the decoder reconstructs x by maximizing $\Pr(x|s)$. The basic idea is that the vector x is treated as an error vector for the transmission of some codeword c that satisfies $Hc = 0$. Since successful error correction relies on the fact that the error vector is sparse, we find that reconstruction is only successful if x is the sparsest vector that satisfies $s = Hx$.

C. Low-Density Parity Check Codes

Low-density parity-check (LDPC) codes are linear codes with a sparse parity-check matrix H. If the codes are defined over $\mathbb{F}_q$, the entries of H are elements from $\mathbb{F}_q$ and the parity-check equations are also computed in $\mathbb{F}_q$. Message-passing decoding works on the natural bipartite graph associated with the parity-check matrix [9]. The decoding complexity is linear in the block length and linear in the number of iterations.

The ensemble of LDPC codes that we use in this paper was introduced in [13], [14] and is defined by the edge degree distribution (d.d.) functions $\lambda(x) = \sum_{k\geq 2} \lambda_k x^{k-1}$ and $\rho(x) = \sum_{k\geq 2} \rho_k x^{k-1}$. In [13], the authors also design a class of capacity-achieving codes for the binary erasure channel (BEC) by carefully choosing the variable-node and check-node degree distributions.

D. The q-ary Symmetric Channel and Verification Decoding

The q-SC channel is defined by

$$\Pr(y|x) = \begin{cases} 1-\delta & \text{if } y = x \\ \delta/(q-1) & \text{otherwise} \end{cases},$$

where $x, y \in \mathbb{F}_q$ are the input and output of the channel respectively. Verification decoding is a message-passing decoding algorithm which marks messages as *verified* when they are very likely to be correct. For example, if two independent observations through a q-SC give the same value, then the probability that value is not correct is roughly $\delta^2/(q-1)$. From this, we find that verifying symbol nodes based on two independent matching observations gives a probability of false verification (FV) that approaches zero as q goes to infinity.

In [12], two decoding algorithms based on verification were proposed. The first algorithm, dubbed LM1 in this paper, is very similar to the peeling decoder for the erasure channel from [13]. The algorithm works by sequentially verifying code symbols and removing all edges attached to verified symbols. To get started, any parity-check which sums to zero is used to verify all adjacent code symbols with value "0". It proceeds by using parity-checks with degree-1 to verify another code symbol and remove all its attached edges. The second algorithm, dubbed LM2 in this paper, adds one more rule. If any two parity-checks predict the same value for a code symbol, then that symbol is verified (with that value) and all its edges are removed. Decoding succeeds if all code symbols become verified.

Since verification is based on the idea that two independent error are unlikely to match when q is large, it is easily generalized to $\mathbb{F} = \mathbb{R}$. In this case, two real numbers chosen independently from continuous distributions will almost surely be distinct.

III. Main Results

A. Compressed Sensing and Coding

In this paper, we show that the basic CS problem can be interpreted as a syndrome source coding problem and solved using standard coding techniques generalized to $\mathbb{F} = \mathbb{R}$. In particular, we will modify coding techniques designed for the q-ary symmetric channel with large q [17], [12], [23]. In fact, the reconstruction algorithm proposed in [16] for CS is almost identical to the decoding algorithm proposed in [12] for the q-ary symmetric channel. These approaches currently work only for *strictly sparse signals* (i.e., many entries are exactly zero). Generalizations to *approximately sparse signals* (i.e, $\|x\|_p \le R$ for some $0 < p < 2$) are currently underway.

In our approach, the measurement matrix Φ is chosen to be the parity-check matrix of a low-density parity-check (LDPC) code over $\mathbb{R}$. Signal reconstruction is based on an iterative message-passing decoder for this code. In particular, verification decoding is used to mark parts of the signal which are known correctly with high probability [17], [12], [23]. The message-passing decoder is analyzed using density evolution (DE) to track the average fraction of verified messages as decoding progresses. If this fraction converges to 1, then decoding is successful with high probability. We note that this DE analysis can only establish non-uniform reconstruction.

Consider a CS system for a sequence of strictly sparse signals of increasing dimension n with $k_n = \lfloor \delta n \rfloor$ non-zero elements. For each n, the measurement matrix Φ is chosen randomly from the ensemble of LDPC matrices parametrized by $(\lambda(x), \rho(x))$. In this case, there is a threshold δ^* such that, for all $\delta < \delta^*$, iterative decoding will recover the original signal with high probability as $n \to \infty$. The number of measurements produced by this system is the same as the number of rows in Φ and is given by

$$m_n = \frac{\int_0^1 \rho(x)dx}{\int_0^1 \lambda(x)dx} n.$$

In this case, the required oversampling ratio is given by $\frac{m_n}{\delta^* n}$.

For a particular degree distribution $(\lambda(x), \rho(x))$ and decoding rule, the threshold δ^* can be computed with high precision using DE. For example, the ensemble of measurement matrices generated by the degree distribution $\left(x^2, x^5\right)$ (i.e., 3 non-zero elements (n.z.e.) per column and 6 n.z.e. per row) has $\delta^* \approx 0.169$ when the first decoding rule in [12] is used. This corresponds to a required oversampling ratio of $\frac{0.5}{0.169} \approx 3$. One advantage this approach has over all previous results is that the required oversampling ratio can be computed precisely.

Since CS is interested primarily in the case where δ is very small or approaches 0 as $n \to \infty$, we also consider a "scaling law" analysis of iterative decoding in the high rate regime. This allows us to derive tight bounds on the DE thresholds in limit as $\delta \to 0$. To make this precise, we consider sequences of (j, l) regular LDPC code ensembles where j is fixed and $l \to \infty$.

Theorem 1: Consider a CS system for strictly sparse signal based on (j, l) regular LDPC code ensembles with LM1 reconstruction. Non-uniform reconstruction succeeds (w.h.p as $n \to \infty$) if $\delta < \delta^* = (l-1)^{-j/(j-1)}$ for $j \ge 3$ and $l \ge 2$. This results in a required oversampling ratio of $\frac{j/l}{(l-1)^{-j/(j-1)}} < j l^{1/(j-1)} = j \delta^{*-1/j}$.

Sketch of Proof: Using a scaling law for LM1 DE as $l \to \infty$, we can show that DE converges in this case. ■

Corollary 1: For the non-uniform reconstruction with LM1, choosing $j = \left\lceil \ln \frac{1}{\delta} \right\rceil$ results in a required oversampling ratio of at most $\left\lceil \ln \frac{1}{\delta} \right\rceil e$.

Theorem 2: Consider a CS system for strictly sparse signal based on $(4, l)$ regular LDPC code ensembles with LM2 reconstruction. Reconstruction succeeds (w.h.p as $n \to \infty$) if $\delta < \delta^* = (l-1)^{-1}$ for $l \ge 2$. This results in a required oversampling ratio of $\frac{4/l}{(l-1)^{-1}} < 4$. Therefore, LM2 achieves a constant oversampling ratio as $\delta \to 0$.

Sketch of Proof: Using a scaling law for LM2 DE as $l \to \infty$, we can show that DE converges in this case. ■

Conjecture 1: Consider a CS system for strictly sparse signal based on (j, l) regular LDPC code ensembles with LM1 reconstruction. A randomly chosen measurement matrix (w.h.p as $n \to \infty$) achieves uniform reconstruction of all signals with $\delta < \delta^* = \frac{1}{2}(l-1)^{-j/(j-2)}$ for $j \ge 3$ and $l \ge 2$. This results in a required oversampling ratio of $\frac{j/l}{\frac{1}{2}(l-1)^{-j/(j-2)}} < 2 j l^{1/(j-2)} = 2 j \delta^{*-2/j}$.

Sketch of Proof: Using a scaling law for stopping set analysis of LM1 decoding as $l \to \infty$, we have shown this is the correct scaling but there are still some details to verify. ■

Corollary 2: For uniform reconstruction with LM1, choosing $j = \left\lceil \ln \frac{1}{\delta} \right\rceil$ results in a required oversampling ratio of at most $2 \left\lceil \ln \frac{1}{\delta} \right\rceil e^2$.

B. Caveats

There is one drawback to this approach that may not be obvious at first glance. The coding approach solves the CS problem for the case the signal-of-interest is sparse in the standard basis. In many cases, sampling occurs relative to one basis and the signal is sparse in another basis. The linear programming approach (i.e., basis pursuit) for signal reconstruction handles this very elegantly by adding n auxiliary variables and n linear equality constraints. On the other hand, most iterative or fast approaches cannot handle this easily.

Of course, sampling in the wrong basis can also be handled by a simple linear transform of the measurement matrix. The problem with this approach is two-fold. First, the correct basis may not be known in advance or the complexity of communicating the correct basis to the sensor may be prohibitive. Second, randomized sampling may occur naturally as part of some physical process and may not be controllable by the sensor. In both cases, LP based reconstruction works without any problem.

One possible solution is to use error-correcting codes and decoders designed for channels with memory [7]. In this case, the sampling basis can be distinct from the basis of sparsity and iterative decoding has been shown to work quite well.

IV. Conclusion

In this paper, we discuss the connections between compressed sensing (CS) and error-correcting codes. In particular, we point out that CS is identical to syndrome source coding [1] with linear codes over the real numbers. We propose and analyze new techniques for CS by adapting the construction and analysis of good codes for the q-ary symmetric channel [17], [12], [23]. This results in linear-time algorithms for signal reconstruction that have sparsity and oversampling thresholds which can be computed precisely. We also analyze the scaling of the oversampling ratio as the signal becomes extremely sparse. One important new result is that non-uniform linear-time reconstruction of strictly sparse signals is possible with a constant oversampling ratio as the fraction of non-zero elements goes to zero.

References

[1] T. Ancheta Jr. Syndrome-source-coding and its universal generalization. *IEEE Trans. Inform. Theory*, 22(4):432–436, 1976.

[2] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.

[3] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, 2005.

[4] S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comp.*, 20(1):33–61, 1998.

[5] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[6] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Inform. Theory*, 47(7):2845–2862, Nov. 2001.

[7] C. Douillard, M. Jézéquel, C. Berrou, A. Picart, P. Didier, and A. Glavieux. Iterative correction of intersymbol interference: Turbo equalization. *Eur. Trans. Telecom.*, 6(5):507–511, Sept. – Oct. 1995.

[8] M. Duarte, S. Sarvotham, D. Baron, M. Wakin, and R. Baraniuk. Distributed compressed sensing of jointly sparse signals. In *Proc. Asilomar Conf. on Signals, Systems & Computers*, 2005.

[9] R. G. Gallager. *Low-Density Parity-Check Codes*. The M.I.T. Press, Cambridge, MA, USA, 1963.

[10] R. G. Gallager. *Information Theory and Reliable Communication*. Wiley, New York, NY, USA, 1968.

[11] B. S. Kashin and V. N. Temlyakov. A remark on compressed sensing. *Mathematical Notes*, 82(5):748–755, 2007.

[12] M. Luby and M. Mitzenmacher. Verification-based decoding for packet-based low-density parity-check codes. *IEEE Trans. Inform. Theory*, 51(1):120–127, 2005.

[13] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman. Efficient erasure correcting codes. *IEEE Trans. Inform. Theory*, 47(2):569–584, Feb. 2001.

[14] T. J. Richardson and R. L. Urbanke. The capacity of low-density parity-check codes under message-passing decoding. *IEEE Trans. Inform. Theory*, 47(2):599–618, Feb. 2001.

[15] S. Sarvotham, D. Baron, and R. G. Baraniuk. Measurements vs. bits: Compressed sensing meets information theory. In *Proc. 45th Annual Allerton Conf. on Commun., Control, and Comp.*, Monticello, IL, Sept. 2006.

[16] S. Sarvotham, D. Baron, and R. G. Baraniuk. Sudocodes–fast measurement and reconstruction of sparse signals. pages 2804–2808, July 2006.

[17] A. M. Shokrollahi and W. Wang. Low-density parity-check codes with rates very close to the capacity of the q-ary symmetric channel for large q. In *Proc. IEEE Int. Symp. Information Theory*, page 275, Chicago, IL, June 2004.

[18] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. Inform. Theory*, 19(4):471–480, July 1973.

[19] J. A. Tropp and A. C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inform. Theory*, 53(12):4655–4666, 2007.

[20] J. K. Wolf. Redundancy, the discrete Fourier transform, and impulse noise cancellation. *IEEE Trans. Commun.*, 31(3):458–461, March 1983.

[21] A. Wyner. Recent results in the Shannon theory. *IEEE Trans. Inform. Theory*, 20(1):2–10, 1974.

[22] W. Xu and B. Hassibi. Efficient compressive sensing with deterministic guarantees using expander graphs. In *Proc. IEEE Inform. Theory Workshop*, pages 414–419, Lake Tahoe, CA, Sept. 2007.

[23] F. Zhang and H. D. Pfister. List-message passing achieves capacity on the q-ary symmetric channel for large q. In *Proc. IEEE Global Telecom. Conf.*, Washington, DC, Nov. 2007.