

Explicit Matrices for Sparse Approximation

Amin Khajehnejad, Arash Saber Tehrani, Alexandros G. Dimakis, Babak Hassibi

Abstract—We show that girth can be used to certify that sparse compressed sensing matrices have good sparse approximation guarantees. This allows us to present the first deterministic measurement matrix constructions that have an optimal number of measurements for ℓ_1/ℓ_1 approximation. Our techniques are coding theoretic and rely on a recent connection of compressed sensing to LP relaxations for channel decoding.

I. INTRODUCTION

Assume we observe m linear measurements of an unknown vector $\mathbf{e} \in \mathbb{R}^n$:

$$\mathbf{H} \cdot \mathbf{e} = \mathbf{s},$$

where \mathbf{H} is a real-valued matrix of size $m \times n$, called the measurement matrix. When $m < n$ this is an underdetermined system of linear equations and one fundamental compressed sensing problem involves recovering \mathbf{e} assuming that it is also k -sparse, *i.e.* it has k or less non-zero entries. The sparse approximation problem goes beyond exactly sparse vectors and requires the recovery of a k -sparse vector $\hat{\mathbf{e}}$ that is close to \mathbf{e} , even if \mathbf{e} is not exactly k -sparse itself. Recent breakthrough results [11]–[13] showed that it is possible to construct measurement matrices with $m = O(k \log(n/k))$ rows that recover k -sparse signals exactly in polynomial time. This scaling is also optimal, as discussed in [2], [3]. These results rely on randomized matrix constructions and establish that the optimal number of measurements will be sufficient with high probability over the choice of the matrix and/or the signal. Unfortunately the required properties RIP [11], Nullspace [14], [19] and high expansion (expansion quality $\epsilon < 1/6$) have no known ways to be deterministically constructed or efficiently checked. There are several explicit constructions of measurement matrices (*e.g.* [4], [7]) which, however, require a slightly sub-optimal number of measurements (m growing super-linearly as a function of n for $k = p \cdot n$). In this paper we focus in the *linear sparsity* regime where k is a fraction of n and optimal number of measurements will also be a fraction of n . The explicit construction of measurement matrices with an optimal number of rows is a well-known open problem in compressed sensing theory (see *e.g.* [2] and references therein). A closely related issue is that of checking or certifying in polynomial time that a given candidate matrix has good recovery guarantees.

Our Contributions: Consider a sparse matrix \mathbf{H} in $\{0, 1\}^{m \times n}$ that has d_c ones per row and d_v ones per column. If the bipartite graph corresponding to \mathbf{H} has $\Omega(\log n)$ girth, then for $k = p \cdot n$ and an optimal number of measurements $m = c_2 \cdot n$, we show that \mathbf{H} offers ℓ_1/ℓ_1 sparse approximation under basis pursuit decoding. Our technical requirement of girth, unlike expansion or RIP, is easy to check and several deterministic constructions of matrices with $m = c \cdot n$ and $\Omega(\log n)$ exist, starting with the early construction in Gallager’s thesis [16], and the progressive edge-growth Tanner graphs of [17].

Our result is a weak bound, also known as a ‘for-every signal’ guarantee [2]. This means that we have a fixed deterministic matrix and show the ℓ_1/ℓ_1 sparse approximation guarantee with high probability over the support of the signal. To the best of our knowledge, this is the first deterministic construction of matrices with an optimal number of measurements and the strong bound (‘for-all signals’) equivalent remains open.

Our techniques are coding-theoretic and rely on recent developments that connect the channel decoding LP relaxation by Feldman *et al.* [20] to compressed sensing [8]. We rely on a primal-based density evolution technique initiated by Koetter and Vontobel [18] and analytically strengthened in the breakthrough paper of Arora *et al.* [9] that established the best known finite-length threshold results for LDPC codes under LP decoding.

To show our ℓ_1/ℓ_1 sparse approximation result, we first need to extend [8] for non-sparse signals. Specifically, the first step in our analysis (Theorem 3) is to show that ℓ_1/ℓ_1 sparse approximation corresponds to a channel decoding problem for a *perturbed symmetric channel*. The second component is to extend the Arora *et al.* argument for this perturbed symmetric channel, an analysis achieved in Theorem 2. This is performed by showing how a similar tree recursion allows us to establish robustness of the fundamental cone (FCP), *i.e.* every pseudocodeword in the fundamental cone [18] has a non-negative pseudoweight even if the flipped bit likelihoods are multiplied by a factor larger than one. The FCP condition shows that the matrix \mathbf{H} , taken as an LDPC code can tolerate a constant fraction of errors for this perturbed symmetric channel under LP decoding.

We note that even though our analysis involves a rigorous density evolution argument, our decoder is always the basis pursuit linear relaxation which is substantially different from the related work on message-passing algorithms for compressed sensing [15], [28].

Amin Khajehnejad and Babak Hassibi are with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA, USA. email: {amin,hassibi}@caltech.edu

A. Saber Tehrani and A. G. Dimakis are with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA, USA. email: {saber, dimakis}@usc.edu

II. BACKGROUND

In this section we provide a brief background for the major topics we will discuss. We begin by introducing the noiseless compressed sensing problem and the basis pursuit. Then, we mention the channel coding problem and its relaxation.

A. Compressed Sensing Preliminaries

The simplest noiseless compressed sensing (CS) problem for exactly sparse signals consists of recovering the sparsest real vector \mathbf{e}' of a given length n , from a set of m real-valued measurements \mathbf{s} , given by $\mathbf{H} \cdot \mathbf{e}' = \mathbf{s}$; namely

$$\begin{aligned} \text{CS-OPT:} \quad & \text{minimize} && \|\mathbf{e}'\|_0 \\ & \text{subject to} && \mathbf{H} \cdot \mathbf{e}' = \mathbf{s}. \end{aligned}$$

Since ℓ_0 minimization is NP-hard, one can relax **CS-OPT** by replacing the ℓ_0 norm with ℓ_1 , specifically

$$\begin{aligned} \text{CS-LPD:} \quad & \text{minimize} && \|\mathbf{e}'\|_1 \\ & \text{subject to} && \mathbf{H} \cdot \mathbf{e}' = \mathbf{s}. \end{aligned}$$

This LP relaxation is also known as basis pursuit. A central question in compressed sensing is under what conditions the solution given by **CS-LPD** equals (or is very close to, specially in the case of approximately sparse signals) the solution given by **CS-OPT**, *i.e.*, the LP relaxation is tight. There has been a substantial amount of work in this area, see *e.g.* [2], [11]–[14], [19].

One sufficient way to certify that a given measurement matrix is “good” is through the well-known restricted isometry property (RIP), which guarantees that the LP relaxation will be tight for all k -sparse vectors \mathbf{e} and further the recovery will be robust to approximate sparsity [11], [12]. However, RIP condition is not a complete characterization of the LP relaxation of “good” measurement matrices (see, *e.g.*, [21]). In this paper we rely on the nullspace characterization (see, *e.g.*, [22], [23]) instead, that gives a necessary and sufficient condition for a matrix to be “good”.

Definition 1: Let $\mathcal{S} \subset \{1, \dots, n\}$ and let $C_R \geq 0$. We say that \mathbf{H} has the nullspace property $NSP_{\mathbb{R}}^{\leq}(\mathcal{S}, C_R)$, and write $\mathbf{H} \in NSP_{\mathbb{R}}^{\leq}(\mathcal{S}, C_R)$, if

$$C_R \cdot \|\nu_{\mathcal{S}}\|_1 \leq \|\nu_{\overline{\mathcal{S}}}\|_1, \text{ for all } \nu \in \mathcal{N}(\mathbf{H}).$$

We say that \mathbf{H} has the strict nullspace property $NSP_{\mathbb{R}}^{<}(\mathcal{S}, C_R)$ and write $\mathbf{H} \in NSP_{\mathbb{R}}^{<}(\mathcal{S}, C_R)$, if

$$C_R \cdot \|\nu_{\mathcal{S}}\|_1 < \|\nu_{\overline{\mathcal{S}}}\|_1, \text{ for all } \nu \in \mathcal{N}(\mathbf{H}) \setminus \{\mathbf{0}\}.$$

The next performance metric (see, *e.g.*, [1], [6]) for CS involves recovering approximations to signals that are not exactly k -sparse. This is the main performance metric we use in this paper:

Definition 2: An ℓ_1/ℓ_1 approximation guarantee means that the Basis Pursuit LP relaxation outputs an estimate $\hat{\mathbf{e}}$ that is within a constant factor from the best k -sparse approximation for \mathbf{e} , *i.e.*,

$$\|\mathbf{e} - \hat{\mathbf{e}}\|_1 \leq 2 \frac{C_R + 1}{C_R - 1} \cdot \min_{\mathbf{e}' \in \Sigma_{\mathbb{R}^n}^{(k)}} \|\mathbf{e} - \mathbf{e}'\|_1. \quad (1)$$

where $\Sigma_{\mathbb{R}^n}^{(k)} = \{\mathbf{w} \in \mathbb{R}^n \mid |\text{supp}(\mathbf{w})| \leq k\}$. The nullspace condition is a necessary and sufficient for a measurement matrix to yield ℓ_1/ℓ_1 approximation guarantees [8], [27].

B. Channel Coding Preliminaries

A linear binary code \mathcal{C} of length n is defined by a $m \times n$ parity-check matrix \mathbf{H}_{CC} , *i.e.* $\mathcal{C} \triangleq \{\mathbf{x} \in \mathbb{F}_2^n \mid \mathbf{H}_{\text{CC}} \cdot \mathbf{x} = \mathbf{0}\}$. We define the set of codeword indices $\mathcal{I} \triangleq \mathcal{I}(\mathbf{H}_{\text{CC}}) \triangleq \{1, \dots, n\}$, the set of check indices $\mathcal{J} \triangleq \mathcal{J}(\mathbf{H}_{\text{CC}}) \triangleq \{1, \dots, m\}$, the set of check indices that involves the i -th codeword position $\mathcal{J}_i \triangleq \mathcal{J}_i(\mathbf{H}_{\text{CC}}) \triangleq \{j \in \mathcal{J} \mid [\mathbf{H}_{\text{CC}}]_{j,i} = 1\}$, and the set of codeword positions that are involved in the j -th check $\mathcal{I}_j \triangleq \mathcal{I}_j(\mathbf{H}_{\text{CC}}) \triangleq \{i \in \mathcal{I} \mid [\mathbf{H}_{\text{CC}}]_{j,i} = 1\}$. For a regular code, the degrees of the codeword adjacency $|\mathcal{J}_i|$ and the size of check equations $|\mathcal{I}_j|$ are denoted by d_v and d_c respectively. If a codeword $\mathbf{x} \in \mathcal{C}$ is transmitted through a channel and an output sequence \mathbf{y} is received, then one can potentially decode \mathbf{x} by solving for the maximum likelihood codeword in \mathcal{C} , namely

$$\begin{aligned} \text{CC-MLD:} \quad & \text{minimize} && \lambda^T \mathbf{x}' \\ & \text{subject to} && \mathbf{x}' \in \text{conv}(\mathcal{C}), \end{aligned}$$

where λ is the likelihood vector defined by $\lambda_i = \log\left(\frac{\mathbb{P}(y_i|x_i=0)}{\mathbb{P}(y_i|x_i=1)}\right)$, and $\text{conv}(\mathcal{C})$ is the convex hull of all codewords of \mathcal{C} in \mathbb{R}^n . **CC-MLD** solves the ML decoding problem by the virtue of the fact that the objective $\lambda^T \mathbf{x}$ is minimized on a corner point of $\text{conv}(\mathcal{C})$, which is a codeword. **CC-MLD** is NP-hard and therefore an efficient description of the exact codeword polytope is very unlikely to exist.

The well-known channel decoding LP relaxation is:

$$\begin{aligned} \text{CC-LPD:} \quad & \text{minimize} && \lambda^T \mathbf{x}' \\ & \text{subject to} && \mathbf{x}' \in \mathcal{P}(\mathbf{H}_{\text{CC}}), \end{aligned}$$

where $\mathcal{P} = \mathcal{P}(\mathbf{H}_{\text{CC}})$ is known as the fundamental polytope [18], [20]. The fundamental polytope is compactly described as follows: If h_j^T is the j -th row of \mathbf{H}_{CC} , then

$$\mathcal{P} = \bigcap_{1 \leq j \leq m} \text{conv}(\mathcal{C}_j) \quad (2)$$

Where $\mathcal{C}_j = \{x \in \mathbb{F}_2^n \mid h_j^T x = 0 \text{ mod } 2\}$. Due to the symmetries of the fundamental polytope [20] we can focus on the cone around the all-zeros codeword without loss of generality. Given the parity check matrix \mathbf{H}_{CC} , its fundamental cone $\mathcal{K}(\mathbf{H}_{\text{CC}})$ is defined as the smallest cone in \mathbb{R}^n that encompasses $\mathcal{P}(\mathbf{H}_{\text{CC}})$, and can be formally defined as follows.

Definition 3: The fundamental cone $\mathcal{K} \triangleq \mathcal{K}(\mathbf{H}_{\text{CC}})$ of \mathbf{H}_{CC} is the set of all vectors $\mathbf{w} \in \mathbb{R}^n$ that satisfy

$$\begin{aligned} w_i &\geq 0, && \text{for all } i \in \mathcal{I}, \\ w_i &\leq \sum_{i' \in \mathcal{I}_j} w_{i'}, && \text{for all } i \in \mathcal{I}_j, j \in \mathcal{J}. \end{aligned} \quad (3)$$

$$w_i \leq \sum_{i' \in \mathcal{I}_j} w_{i'}, \quad \text{for all } i \in \mathcal{I}_j, j \in \mathcal{J}. \quad (4)$$

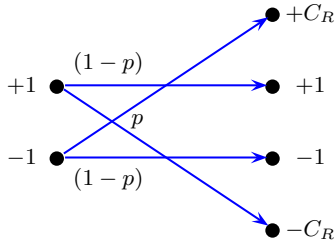


Fig. 1. Perturbed symmetric channel model.

Given the fundamental cone of a code \mathcal{C} , we define the following property.

Definition 4: Let $S \subset \{1, 2, \dots, n\}$ and $C_R \geq 1$ be fixed. A code \mathcal{C} with parity check matrix $\mathbf{H}_{\mathcal{C}\mathcal{C}}$ is said to have the fundamental cone property $\text{FCP}(\mathcal{S}, C_R)$ if for every $\mathbf{w} \in \mathcal{K}(\mathbf{H}_{\mathcal{C}\mathcal{C}})$ the following holds:

$$C_R \|\mathbf{w}_S\|_1 < \|\mathbf{w}_{\bar{S}}\|_1. \quad (5)$$

We introduce the perturbed symmetric channel (PSC) shown in Fig. 1, where each bit $+1$ (-1)¹ is flipped into $-C_R$ ($+C_R$) with probability p and remains unchanged with probability $(1-p)$. For $C_R = 1$, the perturbed symmetric channel is the same as the binary symmetric channel (BSC). Note that the data can be received error free through the perturbed channel. Through solving **CC-LPD**, however, it is apparent that the recovery probability of the perturbed channel with $C_R > 1$ is less than BSC.

III. ℓ_1/ℓ_1 GUARANTEE FOR $\Omega(\log n)$ GIRTH MATRICES

We start by stating the main Theorem of this paper.

Theorem 1: Let \mathbf{H} be an $m \times n$ 0-1 measurement matrix with girth $g = \Omega(\log(n))$, and column and row densities d_v and d_c respectively. There exists a fraction $p^*(d_c, d_v)$, solely a function of d_c and d_v , so that for every $0 \leq p \leq p^*(d_c, d_v)$, there exists some $C_R > 1$ such that \mathbf{H} provides ℓ_1/ℓ_1 guarantee (with parameter C_R) for the **CS-LPD** and for a randomly chosen support set S of size $p \cdot n$, with probability higher than $1 - \mathcal{O}(e^{-\alpha n})$ for some $\alpha > 0$.

The function $p^*(d_c, d_v)$ and C_R are deterministic functions computed from a recursion on a tree (see theorem 2). For example, for $d_c = 6, d_v = 3$ and we numerically obtain $p^* \geq 0.045$, and therefore we can recover almost all $k = 0.045n$ -sparse signals. To the best of our knowledge, this is the best provable recovery threshold for sparse compressed sensing measurement matrices. Furthermore, for $p = 0.04$ we obtain $C_R \geq 1.08$, which means that for compressible signals, ℓ_1 minimization gives ℓ_1/ℓ_1 error bound, given by the constant $2 \frac{C_R+1}{C_R-1} = 52$, i.e. the ℓ_1 norm of the recovery error is bounded by a factor 52 of the ℓ_1 norm of the smallest $(1 - 0.04)n$ coefficient of the signal (see (1)).

¹Note that the bits 0,1 are mapped to +1,-1, respectively.

Proof: To prove this Theorem, we take the steps shown in Fig. 2. We import the best performance guarantees for LDPC code \mathcal{C} from LP decoding [9] through the bridge established in [8]. First through Theorem 2, we show that the code \mathcal{C} has the fundamental cone property $\text{FCP}(\mathcal{S}, C_R)$ with probability at least $1 - \mathcal{O}(e^{-\alpha n})$. Next, through Lemma 2, we demonstrate that \mathcal{C} corrects the error configuration \mathcal{S} at the output of the perturbed symmetric channel. Finally, by Theorem 3, we establish a connection between the properties of $\mathbf{H}_{\mathcal{C}\mathcal{C}}$ as parity check matrix (i.e. FCP condition) and its null space properties as a measurement matrix in compressed sensing. Consequently, we show that the solution of **CS-LPD** satisfies the ℓ_1/ℓ_1 condition. ■

Note that, if the girth is $g = \Omega(\log \log(n))$, then $\mathbf{H}_{\mathcal{C}\mathcal{C}}$ has ℓ_1/ℓ_1 guarantee with probability higher than $1 - \mathcal{O}(1/n)$. Theorem 1 is our main result and we will build the blocks necessary to complete the proof in the remaining of this paper.

A. Extension of CC-LPD

The work of Arora et. al [9] provides the *best existing* guarantees for the performance of LP decoding. For a fixed code rate, the probabilistic analysis of [9] yields a (weak) threshold p^* for the probability of bit flip, below which **CC-LPD** can recover the output of a binary bit flipping channel, with high probability, provided that the code has a girth at least doubly logarithmic in the size of the code. The resulting threshold is significantly tighter than the previously achieved bounds, namely those obtained in [24], which were established for expander graph based codes. To the best of our knowledge, the thresholds of [9] are the highest proved thresholds for LP decoding, and as mentioned, hold for codes that retain large girths, rather than expander codes.

Building upon the work of Arora *et al.*, we further prove that codes with optimal girth maintain the FCP property in a weak notion. In other words, we prove that if the probability of bit flip p is sufficiently small, there is a finite $C_R > 1$ such that with high probability, the output of the perturbed symmetric channel in Figure 1 with C_R can be recovered by **CC-LPD**. The parameter C_R surely depends on p . We compute an upper bound on the achievable robustness factor C_R as a function of p , which approaches 1 as p approaches the recoverable threshold p^* of LP decoding based on the analysis of [9]. Using the connection between the **CS-LPD** and **CC-LPD**, this result will allow us to explicitly find the relationship between the ℓ_1/ℓ_1 robustness factor C_R and the fraction p , when the results are translated to the context of compressed sensing. In order to state the main theorem of this section, first we define η to be a random variable that takes the value $-C_R$ with probability p and value 1 with probability $1-p$, i.e. the output of the PSC when 0 is transmitted. Also, let the sequences of random variables $X_i, Y_i, i \geq 0$ be defined in the following way:

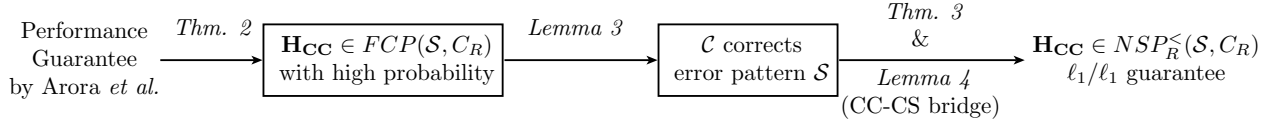


Fig. 2. The procedure of importing the performance guarantee from LP decoding into compressed sensing.

$$\begin{aligned}
 Y_0 &= \eta, \\
 X_i &= \min\{Y_i^{(1)}, \dots, Y_i^{(d_c-1)}\} \forall i > 0, \\
 Y_i &= 2^i \eta + X_{i-1}^{(1)} + \dots + X_{i-1}^{(d_v-1)} \forall i > 0, \quad (6)
 \end{aligned}$$

where $X^{(j)}$ s are independent copies of a random variable X .

Theorem 2: Let \mathcal{C} be a regular (d_v, d_c) -LDPC code with girth equal to g , and $0 \leq p \leq 1/2$ be the probability of bit flip, and S be the random set of flipped bits. If for some $j \in \mathbb{N}$,

$$c = \gamma^{1/(d_v-2)} \min_{t \geq 0} \mathbb{E} e^{-tX_j} < 1$$

where $\gamma = (d_c - 1) \frac{C_R + 1}{C_R} (\frac{C_R p}{1-p})^{1/(C_R+1)} (1-p) < 1$, then with probability at least $1 - O(n)c^{d_v(d_v-1)^{T-1}}$ the code \mathcal{C} has the FCP(\mathcal{S}, C_R), where T is any integer with $j \leq T < g/4$.

Note that the value c can be derived for specific values of p , d_v , d_c , and C_R . The proof of the above theorem falls along the same lines as the arguments of [9]. The bottom line is the existence of a certificate in the primal LP problem for the success of LP decoding, which can be extended to the case of the perturbed channel. In order to define the certificate, we first bring the following definitions from [9]. In the sequel, we denote the bipartite graph that corresponds to \mathbf{H} by \mathcal{G} .

Definition 5: A tree \mathcal{T} of height $2T$ is called a skinny subtree of \mathcal{G} , if it is rooted at some variable node v_{i_0} , for every variable node v in \mathcal{T} all the neighboring check nodes of v in \mathcal{G} are also present in \mathcal{T} , and for every check node c in \mathcal{T} exactly two neighboring variable nodes of c in \mathcal{G} are present in \mathcal{T} .

Definition 6: Let $\mathbf{w} \in [0, 1]^T$ be a fixed vector. A vector $\beta^{(\mathbf{w})}$ is called a minimal T -local deviation, if there is a skinny subtree of \mathcal{G} of height $2T$, say \mathcal{T} , so that for every variable node v_i $1 \leq i \leq n$,

$$\beta_i^{(\mathbf{w})} = \begin{cases} w_{h(i)} & \text{if } v_i \in \mathcal{T} \setminus \{v_{i_0}\} \\ 0 & \text{otherwise} \end{cases},$$

where $h_i = \frac{1}{2}d(v_{i_0}, v_i)$.

The key to the derivation of a certificate for LP decoding is the following lemma:

Lemma 1 (Lemma 1 of [9]): For any vector $\mathbf{z} \in \mathcal{P}$, and any positive vector $\mathbf{w} \in [0, 1]^T$, there exists a distribution on the minimal T -local deviations $\beta^{(\mathbf{w})}$, such that

$$\mathbb{E} \beta^{(\mathbf{w})} = \alpha \mathbf{z},$$

where $0 < \alpha \leq 1$.

Lemma 1 has the following interpretation. If a linear property holds for all minimal T -local deviations ($f(\beta^{(\mathbf{w})}) \geq 0$, where $f(\cdot)$ is a linear function), then it also holds for all pseudocodewords ($f(\mathbf{z}) \geq 0 \forall \mathbf{z} \in \mathcal{P}$). Interestingly enough, the success of LP decoding over the perturbed symmetric channel of Figure 1 for a given set of bit flips \mathcal{S} has a linear certificate, namely FCP(\mathcal{S}, C_R). In other words, if we define:

$$f_C^{(\mathcal{S})}(x) = \sum_{i \in \mathcal{S}} x_i - C_R \sum_{i \in \mathcal{S}^c} x_i, \quad (7)$$

then LP decoder is successful, if and only if $f_C^{(\mathcal{S})}(z) \geq 0$ for every pseudocodeword $z \in \mathcal{P}$. Therefore, according to Lemma 1, it suffices that the condition be true for all T -local deviations, which is equivalent to FCP(\mathcal{S}, C_R).

Proof of Theorem 2: We denote the set of variable nodes and check nodes by X_v and X_c respectively. For a fixed $\mathbf{w} \in [0, 1]^T$, let \mathcal{B} be the set of all minimal T -local deviations, and \mathcal{B}_i be the set of minimal T -local deviations defined over a skinny tree rooted at the variable node v_i . Also, assume S is the random set of flipped bits, when the flip probability is p . Interchangeably, we also use S to refer to the set of variable nodes corresponding to the flipped bits indices. We are interested in the probability that for all $\beta^{(\mathbf{w})} \in \mathcal{B}$, $f_C^{(\mathcal{S})}(\beta^{(\mathbf{w})}) \geq 0$. For simplicity we denote this event by $f_C^{(\mathcal{S})}(\mathcal{B}) \geq 0$. Since the bits are flipped independently and with the same probability, we have the following union bound

$$\mathbb{P} \left(f_C^{(\mathcal{S})}(\mathcal{B}) \geq 0 \right) \geq 1 - n \mathbb{P} \left(f_C^{(\mathcal{S})}(\mathcal{B}_1) \geq 0 \right). \quad (8)$$

Now consider the full tree of height $2T$, that is rooted at the node v_0 , and contains every node u in \mathcal{G} that is no more than $2T$ distant from v_0 , i.e. $d(v_0, u) \leq 2T$. We denote this tree by $B(v_0, 2T)$. To every variable node u of $B(v_0, 2T)$, we assign a label, $I(u)$, which is equal to $-C_R \omega_{h(u)}$ if $u \in S$, and is $\omega_{h(u)}$ if $u \in S^c$, where $(\omega_0, \omega_1, \dots, \omega_{2T-2}) = \mathbf{w}$. We can now see that the event $f_C^{(\mathcal{S})}(\mathcal{B}_1) \geq 0$ is equivalent to the event that for all skinny subtrees \mathcal{T} of $B(v_0, 2T)$ of height $2T$, the sum of the labels on the variable nodes of \mathcal{T} is positive. In other words, if Γ_1 is the set of all skinny trees of height $2T$ that are rooted at v_0 , then $f_C^{(\mathcal{S})}(\mathcal{B}_1) \geq 0$ is equivalent to:

$$\min_{\mathcal{T} \in \Gamma_1} \sum_{v \in \mathcal{T} \cap X_v} I(v) \geq 0. \quad (9)$$

We assign to each node u (either check or variable node) of $B(v_0, 2T)$ a random variable Z_u , which is equal to the contribution to the quantity $\min_{\mathcal{T} \in \Gamma_1} \sum_{v \in \mathcal{T} \cap X_v} I(v)$

by the offspring of the node u in the tree $B(v_0, 2T)$, and the node u itself. The value of Z_u can be determined recursively from all of its children. Furthermore, the distribution of Z_u only depends on the height of u in $B(v_0, 2T)$. Therefore, to find the distribution of Z_u , we use X_0, X_1, \dots, X_{T-1} as random variables with the same distribution as Z_u when u is a variable node (X_0 is assigned to the lowest level variable node) and likewise Y_1, Y_2, \dots, Y_{T-1} for the check nodes. It then follows that:

$$\begin{aligned} Y_0 &= \omega_0 \eta_C, \\ X_i &= \min\{Y_i^{(1)}, \dots, Y_i^{(d_c-1)}\} \quad \forall i > 0, \\ Y_i &= \omega_i \eta_C + X_{i-1}^{(1)} + \dots + X_{i-1}^{(d_v-1)} \quad \forall i > 0, \end{aligned} \quad (10)$$

where $X^{(j)}$'s are independent copies of a random variable X , and η_C is a random variable that takes the value $-C_R$ with probability p and value 1 with probability $1-p$. It follows that

$$\begin{aligned} \mathbb{P}\left(f_C^{(S)}(\mathcal{B}_1) \leq 0\right) &= \mathbb{P}\left(X_{T-1}^{(1)} + \dots + X_{T-1}^{(d_v)} \leq 0\right) \\ &\leq (\mathbb{E}(e^{-tX_{T-1}}))^{d_v}. \end{aligned} \quad (11)$$

The last inequality is by Markov inequality and is true for all $t > 0$. The rest of the proof is modifications to the proof of Lemma 8 in [9], for the Laplace transform evolution of the variables X_i 's and Y_i 's, to account for a non-unitary robustness factor C_R . By upper bounding the Laplace transform of the variables recursively it is possible to show that (see Lemma 8 of [9], the argument is completely the same for our case)

$$\begin{aligned} \mathbb{E}e^{-tX_i} &\leq (\mathbb{E}e^{-tX_j})^{(d_v-1)^{i-j}} \\ &\quad \prod_{0 \leq k \leq i-j-1} ((d_c-1)\mathbb{E}e^{-tw_{i-k}\eta})^{(d_v-1)^k}, \end{aligned} \quad (12)$$

for all $1 \leq j \leq i < T$. If we take the weight vector as $\mathbf{w} = (1, 2, \dots, 2^j, \rho, \rho, \dots, \rho)$ for some integer $1 \leq j < T$, and use equation (12) for $i = T-1$, we obtain:

$$\begin{aligned} \mathbb{E}e^{-tX_{T-1}} &\leq (\mathbb{E}e^{-tX_j})^{(d_v-1)^{T-j-1}} \\ &\quad \cdot ((d_c-1)\mathbb{E}e^{-t\rho\eta})^{\frac{(d_v-1)^{T-j-1}-1}{d_v-2}}, \end{aligned}$$

ρ and t can be chosen to jointly minimize $\mathbb{E}e^{-tX_j}$ and $\mathbb{E}e^{-t\rho\eta}$ in the above, which along with (11) results in

$$\begin{aligned} \mathbb{P}(f_C^{\mathcal{S}}(\mathcal{B}_1) \leq 0) &\leq (\mathbb{E}e^{-tX_{T-1}})^{d_v} \\ &\leq \gamma^{-d_v/(d_v-2)} \times c^{d_v(d_v-1)^{T-j-1}}, \end{aligned}$$

where $\gamma = (d_c-1)\frac{C_R+1}{C_R}(1-p)\left(\frac{C_R p}{1-p}\right)^{1/(C_R+1)}$ and $c = \gamma^{1/(d_v-2)} \min_{t \geq 0} \mathbb{E}e^{-tX_j}$. If $c < 1$, then probability of error tends to zero as stated in Theorem 2. ■

In the next Lemma, we show that if a code \mathcal{C} has the fundamental cone property $\text{FCP}(\mathcal{S}, C_R)$, then **CC-LPD** can correct the perturbed symmetric channel error configuration \mathcal{S} .

Lemma 2: Let \mathbf{H}_{CC} be a parity-check matrix of a code \mathcal{C} and let $\mathcal{S} \subset \mathcal{I}(\mathbf{H}_{\text{CC}})$ be a particular set of coordinate indices that are flipped by a perturbed channel with cross-over probability $p > 0$. If and only if \mathbf{H}_{CC} has the $\text{FCP}(\mathcal{S}, C_R)$, then the solution of **CC-LPD** equals the codeword that was sent.

Proof: Without loss of generality, we can assume that the all-zero codeword is transmitted. We begin by proving the sufficiency. Let $+1$ be the log-likelihood ratio associated with a received 0, and let $-C_R < -1$ be the log-likelihood ratio associated with a received 1. Therefore,

$$\lambda_i = \begin{cases} +1 & \text{if } i \in \mathcal{S} \\ -C_R & \text{if } i \in \bar{\mathcal{S}} \end{cases}.$$

Then it follows from the assumptions in the lemma statement that for any $\mathbf{w} \in \mathcal{K}(\mathbf{H}_{\text{CC}}) \setminus \{\mathbf{0}\}$

$$\begin{aligned} \lambda^T \mathbf{w} &= \sum_{i \in \mathcal{S}} (+1) \cdot w_i + \sum_{i \in \bar{\mathcal{S}}} (-C_R) \cdot w_i \\ &\stackrel{(a)}{=} +1 \cdot \|\mathbf{w}_{\mathcal{S}}\|_1 - C_R \cdot \|\mathbf{w}_{\bar{\mathcal{S}}}\|_1 > 0 \stackrel{(b)}{=} \lambda^T \cdot \mathbf{0}, \end{aligned}$$

where step (a) follows from the fact that $|w_i| = w_i$ for all $i \in \mathcal{I}(\mathbf{H}_{\text{CC}})$, and where step (b) follows from (5). Therefore, under **CC-LPD** the all-zero codeword has the lowest cost function value when compared to all non-zero pseudo-codewords in the fundamental cone, and therefore also compared to all non-zero pseudo-codewords in the fundamental polytope.

Note that the proof of the converse is direct and can easily be derived by taking the sufficiency proof steps, backward. ■

B. Establishing the Connection

In this section, through Lemma 3 (taken from [10]), we will establish a bridge between LP-decoding and compressed sensing. Specifically, using this bridge, we will import performance results from LP-decoding context into compressed sensing.

Lemma 3: (Lemma 6 in [10]): Let \mathbf{H}_{CC} be a zero-one measurement matrix. Then

$$\nu \in \mathcal{N}(\mathbf{H}_{\text{CC}}) \Rightarrow |\nu| \in \mathcal{K}(\mathbf{H}_{\text{CC}}).$$

Proof: Let $\mathbf{w} \triangleq |\nu|$ and to show that such a vector \mathbf{w} is indeed in the fundamental cone of \mathbf{H}_{CC} , we need to verify (3) and (4). It is apparent that \mathbf{w} satisfies (3). Therefore, let us focus on the proof that \mathbf{w} satisfies (4). Namely, from $\nu \in \mathcal{N}(\mathbf{H}_{\text{CC}})$ it follows that for all $j \in \mathcal{J}$, $\sum_{i \in \mathcal{I}} h_{j,i} \nu_i = 0$, i.e., for all $j \in \mathcal{J}$, $\sum_{i \in \mathcal{I}_j} \nu_i = 0$. This implies

$$w_i = |\nu_i| = \left| - \sum_{i' \in \mathcal{I}_j \setminus \{i\}} \nu_{i'} \right| \leq \sum_{i' \in \mathcal{I}_j \setminus \{i\}} |\nu_{i'}| = \sum_{i' \in \mathcal{I}_j \setminus \{i\}} w_{i'},$$

for all $j \in \mathcal{J}$ and all $i \in \mathcal{I}_j$, showing that \mathbf{w} indeed satisfies (4). ■

From here on, we deal with \mathbf{H}_{CC} as a zero-one measurement matrix. Note that the bridge established in [8]

and through Lemma 3 connects **CC-LPD** of the binary linear channel code and **CS-LPD** based on a zero-one measurement matrix over reals by viewing this binary parity-check matrix as a measurement matrix. This connection allows the translation of performance guarantees from one setup to the other. Using this bridge, we can show that parity-check matrices of "good" channel codes can be used as provably "good" measurement matrices under basis pursuit.

Using the bridge of Lemma 3, we show in the next theorem that the parity-check matrix \mathbf{H}_{CC} , as a measurement matrix, has the nullspace property and as a result satisfies the ℓ_1/ℓ_1 guarantee.

Theorem 3: Let $\mathbf{H}_{\text{CC}} \in \{0,1\}^{m \times n}$ be a parity-check matrix of the code \mathcal{C} and let k be a non-negative integer. Further assume that code \mathcal{C} can correct the error configuration \mathcal{S} over the perturbed symmetric channel where $|\mathcal{S}| \leq k$. Additionally, assume that $\mathbf{s} = \mathbf{H}_{\text{CC}} \cdot \mathbf{e}$. Then \mathbf{H}_{CC} as a measurement matrix satisfies

$$\mathbf{H}_{\text{CC}} \in \text{NSP}_{\mathbb{R}}^{\leq}(\mathcal{S}, C_R).$$

Furthermore, the solution $\hat{\mathbf{e}}$ produced by **CS-LPD** will satisfy

$$\|\mathbf{e} - \hat{\mathbf{e}}\|_1 \leq 2 \cdot \frac{C_R + 1}{C_R - 1} \cdot \|\mathbf{e}_{\overline{\mathcal{S}}}\|_1.$$

Proof: We begin by proving the nullspace property. Since the code \mathcal{C} corrects the configuration \mathcal{S} , from Lemma 2, for any point in the fundamental cone $\mathcal{K}(\mathbf{H}_{\text{CC}})$ including \mathbf{w} and any set $\mathcal{S} \subset \{1, \dots, n\}$ with $|\mathcal{S}| \leq k$, we have

$$C_R \|\mathbf{w}_{\mathcal{S}}\|_1 < \|\mathbf{w}_{\overline{\mathcal{S}}}\|_1. \quad (13)$$

We prove by contradiction. Assume \mathbf{H}_{CC} does not have the strict nullspace property $\text{NSP}_{\mathbb{R}}^{\leq}(\mathcal{S}, C_R)$, *i.e.* there exists a point $\nu \in \mathcal{N}(\mathbf{H}_{\text{CC}})$ such that

$$C_R \|\nu_{\mathcal{S}}\|_1 \geq \|\nu_{\overline{\mathcal{S}}}\|_1.$$

Further, from Lemma 3, we know there exists $\mathbf{w} = |\nu|$ in $\mathcal{K}(\mathbf{H}_{\text{CC}})$. And

$$\begin{aligned} C_R \|\mathbf{w}_{\mathcal{S}}\|_1 &= C_R \|\nu_{\mathcal{S}}\|_1 \\ &= C_R \|\nu_{\mathcal{S}}\|_1 \\ &\geq \|\nu_{\overline{\mathcal{S}}}\|_1 \\ &= \|\nu_{\overline{\mathcal{S}}}\|_1 \\ &= \|\mathbf{w}_{\overline{\mathcal{S}}}\|_1, \end{aligned}$$

which contradicts the assumption and shows that no such a point ν exists.

We showed that \mathbf{H}_{CC} has the claimed nullspace property. Since $\mathbf{H}_{\text{CC}} \cdot \mathbf{e} = \mathbf{s}$ and $\mathbf{H}_{\text{CC}} \cdot \hat{\mathbf{e}} = \mathbf{s}$, it easily follows

that, $\nu \triangleq \mathbf{e} - \hat{\mathbf{e}}$ is in the nullspace of \mathbf{H}_{CC} . So

$$\begin{aligned} \|\mathbf{e}_{\mathcal{S}}\|_1 + \|\mathbf{e}_{\overline{\mathcal{S}}}\|_1 &= \|\mathbf{e}\|_1 \\ &\stackrel{(a)}{\geq} \|\hat{\mathbf{e}}\|_1 \\ &= \|\mathbf{e} - \nu\|_1 \\ &= \|\mathbf{e}_{\mathcal{S}} - \nu_{\mathcal{S}}\|_1 + \|\mathbf{e}_{\overline{\mathcal{S}}} - \nu_{\overline{\mathcal{S}}}\|_1 \\ &\stackrel{(b)}{\geq} \|\mathbf{e}_{\mathcal{S}}\|_1 - \|\nu_{\mathcal{S}}\|_1 + \|\nu_{\overline{\mathcal{S}}}\|_1 - \|\mathbf{e}_{\overline{\mathcal{S}}}\|_1 \\ &\stackrel{(c)}{\geq} \|\mathbf{e}_{\mathcal{S}}\|_1 - \frac{C_R - 1}{C_R + 1} \cdot \|\nu\|_1 - \|\mathbf{e}_{\overline{\mathcal{S}}}\|_1, \end{aligned} \quad (14)$$

where step (a) follows from the fact that the solution of **CS-LPD** satisfies $\|\mathbf{e}\|_1 \leq \|\hat{\mathbf{e}}\|_1$, where step (b) follows from applying the triangle inequality property of the ℓ_1 -norm twice, and where step (c) follows from

$$\begin{aligned} (C_R + 1) \cdot (-\|\nu_{\mathcal{S}}\|_1 + \|\nu_{\overline{\mathcal{S}}}\|_1) &= -C_R \cdot \|\nu_{\mathcal{S}}\|_1 - \|\nu_{\mathcal{S}}\|_1 + C_R \cdot \|\nu_{\overline{\mathcal{S}}}\|_1 + \|\nu_{\overline{\mathcal{S}}}\|_1 \\ &\stackrel{(d)}{\geq} -\|\nu_{\overline{\mathcal{S}}}\|_1 - \|\nu_{\mathcal{S}}\|_1 + C_R \cdot \|\nu_{\overline{\mathcal{S}}}\|_1 + C_R \cdot \|\nu_{\mathcal{S}}\|_1 \\ &= (C_R - 1) \cdot \|\nu_{\mathcal{S}}\|_1 + (C_R - 1) \cdot \|\nu_{\overline{\mathcal{S}}}\|_1 \\ &= (C_R - 1) \|\nu\|_1, \end{aligned}$$

where step (d) follows from applying twice the fact that $\nu \in \mathcal{N}(\mathbf{H}_{\text{CC}})$ and the assumption that $\mathbf{H}_{\text{CC}} \in \text{NSP}_{\mathbb{R}}^{\leq}(\mathcal{S}, C_R)$. Subtracting the term $\|\mathbf{e}_{\mathcal{S}}\|_1$ on both sides of (14), and solving for $\|\nu\|_1 = \|\mathbf{e} - \hat{\mathbf{e}}\|_1$ yields the promised result. \blacksquare

This Theorem was the last piece in the proof of Theorem 1, and with it, the proof is complete.

Note that it is easy to obtain deterministic constructions of measurement matrices that are regular, have the optimal order of measurements and have $\Omega(\log n)$ girth. Therefore, a byproduct of our result is the explicit construction of measurement matrices with $m = cn$ rows which can recover almost all sparse signals with sparsity $k = p \cdot n$. These will be parity-check matrices of codes with girth $\Omega(\log n)$. To explicitly obtain such codes, we can use one of the known deterministic matrix constructions such as the progressive edge-growth (PEG) Tanner graphs [17], or the deterministic construction suggested by Gallager [16].

REFERENCES

- [1] R. Berinde, A. Gilbert, P. Indyk, H. Karloff, and M. Strauss, *Combining geometry and combinatorics: a unified approach to sparse signal recovery*, in Proc. 46th Allerton Conf. on Communications, Control, and Computing, Allerton House, Monticello, IL, USA, Sept. 23–26 2008.
- [2] A. Gilbert and P. Indyk, *A survey on Sparse Recovery Using Sparse Matrices*, Proceedings of IEEE, June 2010.
- [3] K. Do Ba, P. Indyk, E. Price and D. Woodruff, *Lower bounds for Sparse Recovery*, in Proc. ACM-SIAM Symp. Discrete Algorithms (SODA), Jan. 2010.
- [4] W. U. Bajwa, R. Calderbank, and S. Jafarpour, *Why Gabor frames? Two fundamental measures of coherence and their role in model selection*, submitted, June 2010.
- [5] M. Capalbo, O. Reingold, S. Vadhan, and A. Wigderson, *Randomness conductors and constant-degree lossless expanders*, Proc. 34th Annual ACM Symposium on Theory of Computing, 2002.

- [6] A. Cohen, W. Dahmen, and R. DeVore, *Compressed sensing and best k -term approximation*, *J. Amer. Math. Soc.*, vol. 22, pp. 211–231, July 2008.
- [7] R. A. DeVore, *Deterministic constructions of compressed sensing matrices*, *J. Complexity*, vol. 23, no. 4–6, pp. 918–925, Aug. 2007.
- [8] A. G. Dimakis, R. Smarandache, and P. O. Vontobel, *LDPC Codes for Compressed Sensing*, submitted for publication. Available online: <http://arxiv.org/abs/1012.0602>
- [9] S. Arora, C. Daskalakis, and D. Steurer, *Message-passing algorithms and improved LP decoding*, Proc. 41st Annual ACM Symp. Theory of Computing, Bethesda, MD, USA, May 31–June 2 2009.
- [10] R. Smarandache and P. O. Vontobel, *Absdet-pseudo-codewords and perm-pseudo-codewords: definitions and properties*, Proc. IEEE Int. Symp. Information Theory, Seoul, Korea, June 28–July 3 2009.
- [11] E. J. Candes and T. Tao, *Decoding by linear programming*, IEEE Trans. Inf. Theory, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
- [12] D. Donoho, *Compressed sensing*, IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [13] D. Donoho and J. Tanner, *Neighborliness of randomly-projected simplices in high dimensions*, Proc. National Academy of Sciences, 02(27), pp. 9452–9457, 2005.
- [14] D. L. Donoho and X. Huo, *Uncertainty principles and ideal atomic decomposition*, IEEE Transactions on Information Theory, 47(7), 2001.
- [15] D. L. Donoho, A. Maleki, and A. Montanari, *Message passing algorithms for compressed sensing*, Proc. Natl Acad. Sci., 2009.
- [16] R. G. Gallager, *Low-Density Parity-Check Codes*, M.I.T. Press, Cambridge, MA, 1963.
- [17] X. Y. Hu, E. Eleftheriou, and D. M. Arnold, *Regular and irregular progressive edge-growth Tanner graphs*, IEEE Trans. Inf. Theory, vol. 51, no. 1, pp. 386–398, 2005.
- [18] R. Koetter and P. O. Vontobel, *On the block error probability of LP decoding of LDPC codes*, Proc. Inaugural Workshop of the Center for Information Theory and Applications, UC San Diego, La Jolla, CA, USA, Feb. 6–10 2006.
- [19] M. Stojnic, W. Xu, and B. Hassibi, *Compressed sensing - probabilistic analysis of a null-space characterization*, IEEE International Conference on Acoustic, Speech and Signal Processing, ICASSP 2008.
- [20] J. Feldman, M. J. Wainwright, and D. R. Karger, *Using linear programming to decode binary linear codes*, IEEE Trans. Inf. Theory, vol. 51, no. 3, pp. 954–972, Mar. 2005.
- [21] J. D. Blanchard, C. Cartis, and J. Tanner, *Compressed sensing: how sharp is the restricted isometry property?*, to appear in SIAM Review, 2010, available online: <http://arxiv.org/abs/1004.5026>
- [22] W. Xu and B. Hassibi, *Compressed sensing over the Grassmann manifold: a unified analytical framework*, in Proc. 46th Allerton Conf. on Communications, Control, and Computing, Allerton House, Monticello, IL, USA, Sept. 23–26 2008.
- [23] M. Stojnic, W. Xu, and B. Hassibi, *Compressed sensing - probabilistic analysis of a null-space characterization*, in Proc. IEEE Intern. Conf. Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, Mar. 31– Apr. 4 2008, pp. 3377–3380.
- [24] C. Daskalakis, A. G. Dimakis, R. M. Karp, and M. J. Wainwright, *Probabilistic Analysis of Linear Programming Decoding*.
- [25] P. O. Vontobel and R. Koetter, *Graph-cover decoding and finite-length analysis of message-passing iterative decoding of LDPC codes*, accepted for IEEE Trans. Inform. Theory, 2007. Available online <http://www.arxiv.org/abs/cs.IT/0512078>
- [26] A. Khajehnejad, A. G. Dimakis, B. Hassibi and W. Bradley, *Iterative Reweighted LP Decoding*, Proceedings of Allerton Conference 2010.
- [27] W. Xu and B. Hassibi, *Compressed sensing over the Grassmann manifold: a unified analytical framework*, in Proc. 46th Allerton Conf. on Communications, Control, and Computing, Allerton House, Monticello, IL, USA, Sept. 23–26 2008.
- [28] F. Zhang and H. D. Pfister, *On the iterative decoding of high rate LDPC codes with applications in compressed sensing*, submitted, available online under <http://arxiv.org/abs/0903.2232>, Mar. 2009.