Compressed Sensing using Channel Code Matrices
Project Report

Abhinav Gupta, Paarth Jain
190050003, 190050076

May 10, 2021

Contents

1 Introduction

Firstly in section 2 we will start with defining the problems of channel coding as well as for compressed sensing. Thereafter in section , we will establish mathematical connection between the channel coding problem and compressed sensing. We will try to see how parity-check matrix for the binary symmetric channel (BSC) which corrects k-bit flip errors under CC-LPD (will be defined later), can be used to recover all k-sparse error signals under *basis pursuit* (referred as CS-LPD in [1]).
In the last section, we will implement one such binary measurement matrix and see how using binary matrices we can speed up the computations with almost same reconstruction quality as in random gaussian random measurement matrices.
This report is majoritily based on the papers [1] and [2].

2 Defining problems

2.1 Compressed Sensing

We will start with familiar compressed sensing problem. We have 2 formulations,

1. **CS-OPT:**

$$\begin{aligned} \text{minimise} \quad & ||e'||_0 \\ \text{subject to} \quad & H_{CS} \cdot e' = s \end{aligned}$$

   where H_{CS} is the $m \times n$ measurement matrix and s is the measured vector with m entries and we are estimating e which is a vector with n entries, which is sparse in the elementary basis. This is same as $P0$ problem from the class lectures.

2. **CS-LPD:**

$$\begin{aligned} \text{minimize} \quad & ||e'||_1 \\ \text{subject to} \quad & H_{CS} \cdot e' = s \end{aligned}$$

   All symbols are of same meaning as in CS-OPT. This is l_1 relaxation of CS-OPT, covered as *basis pursuit* in the course lectures.

In class we discussed the RIP condition of order k for matrix H_{CS}, which establishes that the both the above will give the same solutions for $k-$sparse vectors e with high probability.
Here we will use another characterisation, i.e. nullspace characterisation for the measurement matrix H_{CS}. We first define the nullspace property.

Definition 1. *Let $k \in \mathbb{Z}_{>0}$ and let $C \in \mathbb{R}_{>0}$. We say that $m \times n$ H_{CS} measurement matrix has the **nullspace property** $NSP_{\mathbb{R}}^{\leq}(k, C)$ and write $H_{CS} \in NSP_{\mathbb{R}}^{\leq}(k, C)$ if for all subset S of $\{1, \ldots, n\}$ such that $|S| \leq k$ and for all vector v in nullspace of H_{CS} we have*

$$||v_S||_1 \leq ||v_{\bar{S}}||_1 \tag{1}$$

Similarly, we define strict nullspace property $NSP_{\mathbb{R}}^{<}(k, C)$, when we remove equality from equation 1.

Now, we state a well-established theorem of compressed sensing literature as follows.

Theorem 1. *Let H_{CS} be a measurement matrix. Furthermore, we have measurement $s = H_{CS} \cdot e$ where $||e||_0 \leq k$. Then the estimate $\hat{e}$ produced by CS-LPD will be equal to the estimate produced by CS-OPT if $H_C S \in NSP_{\mathbb{R}}^{<}(k, C)$.*

Not only this, if the matrix H_{CS} satisfies the nullspace property then we can also recover signals that are not exactly k-sparse upto l_1/l_1 approximation guarantees.

Theorem 2. *Let H_{CS} be a measurement matrix and $C > 1$ a constant. Furthermore, we have measurement $s = H_{CS} \cdot e$. Then, for any set $S \subseteq \{1, \ldots, n\}$ with $|S| \leq k$, the estimate $\hat{e}$ produced by CS-LPD will satisfy*

$$||e - \hat{e}||_1 \leq 2 \cdot \frac{C+1}{C-1} \cdot ||e_{\bar{S}}||_1$$

if $H_C S \in NSP_{\mathbb{R}}^{\leq}(k, C)$ i.e. $\hat{e}$ is as good as the best k approximation of e times some constant factor.

2.2 Channel Coding

Now we define the channel coding problem over a binary symmetric channel (BSC) i.e. the probability p of bit 1 flipping to 0 is same as bit 0 flipping to 1. For that we need to define some terms first.

We consider coded data transmission over a memoryless channel with input, output alphabet $\mathbb{F}_2 = \{0, 1\}$ and channel law $P_{Y|X}(y|x)$ (gives us the bit flipping probability). The coding scheme will codes represented by set $\mathcal{C}_{CC}$ of block length n and dimension k.

1. Let $G_{CC} \in \mathbb{F}_2^{n\times k}$ be the *generator matrix* for $\mathcal{C}_{CC}$. We convert k-dimensional information v to coded vector $x = G_{CC} \cdot v(\mod 2)$ i.e. $\mathcal{C}_{CC} = \{G_{CC} \cdot v(\mod 2) | v \in \mathbb{F}_2^k\}$.
2. $H_{CC} \in \mathbb{F}_2^{m\times n}$ is the *parity-check* matrix such that $\mathcal{C}_{CC} = \{x \in \mathbb{F}_2^n | H_{CC} \cdot x = 0(\mod 2)$.
3. Let $y \in \mathbb{F}_2^n$ be the received vector, we define for each $i \in 1, \ldots, n$ the log-likelihood ratio $\lambda_i = \log \frac{P_{Y|X}(y_i|0)}{P_{Y|X}(y_i|1)}$. This is log of ratio of probabilities we will receive y_i given the bit x_i was 0 and 1 respectively.

We will use I to denote the set $\{1, \ldots, n\}$ and J to denote the set $\{1, \ldots, m\}$ (column and row indices of parity-matrix repectively).
Now, we use maximum-likelihood formulation to get the approximation $\hat{\boldsymbol{x}}$ original codeword given the received codeword $\boldsymbol{y}$ i.e.

$$\hat{\boldsymbol{x}}(\boldsymbol{y}) = \arg \max_{\boldsymbol{x'} \in \mathcal{C}_{CC}} P_{Y|X|}(\boldsymbol{y}|\boldsymbol{x'})$$

where $P_{Y|X|}(\boldsymbol{y}|\boldsymbol{x'}) = \prod_{i\in I} P_{Y|X}(y_i|x'_i)$, we call this formulation to be **CC-MLD** analogue for CC-OPT in compressed sensing.
Now we can maximise the log of the likelihood term which gives us

$$\log P_{Y|X}(y_i|x'_i) = -\lambda_i x'_i + \log P_{Y|X}(y_i|0) \qquad (\text{as } x'_i \in \{0, 1\})$$

Now as the second term is independent of $\boldsymbol{x'}$, we reformulate CC-MLD as

$$\begin{aligned} \text{minimise} \quad & \langle \boldsymbol{\lambda}, \boldsymbol{x'} \rangle \\ \text{subject to} \quad & \boldsymbol{x'} \in \mathcal{C}_{CC} \end{aligned}$$

However, the cost function is linear, and linear function attains its minimum at the extremal points of a convex set, this is essentially equivalent to

$$\begin{aligned} \text{minimise} \quad & \langle \boldsymbol{\lambda}, \boldsymbol{x'} \rangle \\ \text{subject to} \quad & \boldsymbol{x'} \in \mathit{conv}(\mathcal{C}_{CC}) \end{aligned}$$

where conv(S) denotes the convex hull of the set of S of points in $\mathbb{R}^n$. We call this formulation **CC-MLD1**, solving this takes up exponential time in the block length n.

Therefore, we formulate the following relaxation **CC-LPD**

$$\begin{aligned} \text{minimize} \quad & \langle \boldsymbol{\lambda}, \boldsymbol{x'} \rangle \\ \text{subject to} \quad & \boldsymbol{x'} \in \mathcal{P}(H_C C) \end{aligned}$$

where relaxed set $\mathcal{P}(H_{CC} \supseteq \mathit{conv}(C_{CC})$ defined as follows

Definition 2. *For every $j \in J$, let h_j^T be the j^{th} row of H_{CC} and let*

$$\mathcal{C}_{CC,j} = \{\boldsymbol{x} \in \mathbb{F}_2^n | \langle h_j, \boldsymbol{x} \rangle = 0(\mod 2)\}$$

Then, the fundamental polytope $\mathcal{P}$ of H_{CC} is defined to be the set

$$\mathcal{P} = \bigcap_{j\in J} \mathit{conv}(\mathcal{C}_{CC,j})$$

Vectors in $\mathcal{P}$ are called *pseudocodewords*. Now, we are interested in knowing when CC-LPD gives results close to the original problem CC-MLD2.

Now as the channel is binary-symmetric channel, we can, without loss of generality, assume that the all-zero codeword was transmitted. Now, after this the success probability of CC-LDP is the probability that the all-zero codeword yields the lowest cost-function value when compared to all nonzero vectors in the fundamental

polytype. Further, as the cost function is linear this is same as saying zero vector has the minimum cost function when compared to all the nonzero vectors in the conic hull of the fundamental polytype denoted by $\mathcal{K} = conic(\mathcal{P}) = conic(H_{CC})$.
A derived definition for fundamental cone $\mathcal{K}$ of H_{CC} is set of all vectors $\boldsymbol{w} \in \mathbb{R}^n$ that satisfy

$$w_i \geq 0 \quad \forall i \in I \tag{2}$$

$$w_i \leq \sum_{i' \in I_j \backslash i} w'_i \quad \forall j \in J \text{ and all } i \in I_j \tag{3}$$

where I_j is the set of indices in the j^{th} row where elements of H_{CC} are nonzero.

Now we will setup equivalence between CC-LPD and CC-MLD.

Lemma 1. *Let H_{CC} be a parity-check matrix of a code $\mathcal{C}_{CC}$ and let $S \subseteq I$ be the set of coordinate indices that are flipped by a BSC with a nonzero cross-over probability. Ig H_{CC} is such that*

$$||\omega_S||_1 < ||\omega_{\bar{S}}||_1$$

for all $\omega \in \mathcal{K}\backslash 0$, then the $CC - LPD$ decision equals the codeword that was sent.

The inequality here is very similar to the one present in the nullspace property. We will use this to connect the two problems later in the section 3.

2.3 Pseudoweights

The influence of the channel on the vectors in the fundamental cone is represented by the pseudoweights of the particular vector.

Let $\boldsymbol{\omega}$ be a nonzero vector in $\mathbb{R}^N_{\geq 0}$ with $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)$, we define ω' be the vector with the same components as $\boldsymbol{\omega}$ but arranged in a non-increasing order. Now let

$$f(\zeta) = \omega'_i \quad (i - 1 < \zeta \leq i, 0 < \zeta \leq n)$$

$$F(\zeta) = \int_0^{\zeta} f(\zeta')d\zeta'$$

$$e = F^{-1}\left(F(n)/2\right) = F^{-1}(||\boldsymbol{\omega}||_1/2)$$

With this, the BSC pseudoweight $w_p^{BSC}(\boldsymbol{\omega})$ of $\boldsymbol{\omega}$ is defined to be $2e$.
Also, another way to define is by

$$w_p^{BSC}(\boldsymbol{\omega}) = \begin{cases} 2e & \text{if } ||\boldsymbol{\omega'}_{1,\ldots,e}||_1 = ||\boldsymbol{\omega'}_{e+1,\ldots,n}||_1 \\ 2e-1 & \text{if } ||\boldsymbol{\omega'}_{1,\ldots,e}||_1 > ||\boldsymbol{\omega'}_{e+1,\ldots,n}||_1 \end{cases} \tag{4}$$

where $\boldsymbol{\omega'}$ is sorted array as in the last definition and e is the smallest integer such that $||\boldsymbol{\omega'}_{1,\ldots,e}||_1 \geq ||\boldsymbol{\omega'}_{e+1,\ldots,n}||_1$.
Also, we define

$$w_p^{BSC,\min}(H_C C) = \min_{\boldsymbol{\omega} \in \mathcal{K}\backslash \mathbf{0}} w_p^{BSC}(\boldsymbol{\omega})$$

Now the following lemma establishes a connection between BSC pseudoweights and the condition in Lemma 1.

Lemma 2. *Let H_{CC} be a parity-check matrix of a code $\mathcal{C}_{CC}$ and let $\boldsymbol{\omega} \in \mathcal{K}\backslash \mathbf{0}$. Then, for all sets $S \subseteq I$ with*

$$|S| < \frac{1}{2} \cdot w_p^{BSC}(\boldsymbol{\omega}) \text{ or with } |S| < \frac{1}{2} \cdot w_p^{BSC'}(\boldsymbol{\omega})$$

it holds that

$$||\boldsymbol{\omega_S}||_1 < ||\boldsymbol{\omega_{\bar{S}}}||_1$$

3 Bridging CC-LPD and CS-LPD

In both the problems, their is a condition on the matrix H_{CS} (or H_{CC}) so that the result matches that of CS-OPT and CC-MLD respectively.
For CS-LPD to produce the same result as CS-OPT for all k-sparse vectors, H_{CS} has to satisfy the null space property $NSP^{<}_{\mathbb{R}}(k, C = 1)$ i.e. -

$$\forall v \in \mathcal{N}(H_{CS}), ||v_S||_1 \leq ||v_{\bar{S}}||_1$$

Similarly for CC-LPD to produce the same result as CC-MLD for all k-bit flipping errors, the condition mentioned in lemma 2 is to be satisfied which is -

$$k < \frac{1}{2} \cdot w_p^{BSC}(\boldsymbol{\omega}) \text{ or with } k < \frac{1}{2} \cdot w_p^{BSC'}(\boldsymbol{\omega})$$

for any arbitrary $\omega \in \mathcal{K} \backslash \{\mathbf{0}\}$.

Now to establish the mathematical similarity between the 2 problems, we form the following relationship between the points in the nullspace of H_{CS} and points in the fundamental cone of H_{CS}.

Lemma 3. *Let H_{CS} be a zero-one measurement matrix. Then*

$$\boldsymbol{v} \in Nullsp_{\mathbb{R}}(H_{CS}) \implies |\boldsymbol{v}| \in \mathcal{K}(H_{CS})$$

where $|\boldsymbol{v}|$ is the element-wise absolute value operator applied to vector $\boldsymbol{v}$.

Proof. Let $\boldsymbol{v}$ be a vector in nullspace of H_{CS}. Now $|\boldsymbol{v}|$ will definitely satisfy equation 2 by construction. Now we prove that the vector $\omega = |\boldsymbol{v}|$ satisfy 3 also.
As $\boldsymbol{v}$ lies in the nullspace of H_{CS}, it follows that for all $j \in J$ (J being the set of indices of rows of H_{CS}), we have $\sum_{i \in I} h_{j,i} v_i = 0$. This fact can be stated in other form as

$$\forall j \in J, \sum_{i \in I_j} v_i = 0 \tag{5}$$

where I_j are the set of indices where the j^{th} row h_j is non-zero. This implies

$$\omega_i = |v_i| = \left| - \sum_{i' \in I_j \backslash i} v_{i'} \right| \leq \sum_{i' \in I_j \backslash i} |v_{i'}| = \sum_{i' \in I_j \backslash i} \omega_{i'}$$

where first inequality is just LHS to RHS transfer in the equation 5. □

This gives us the one-way result that every point in the $\mathbb{R}$-nullspace of the measurement matrix H_{CS}, we can associate a point in the fundamental cone of H_{CS}. Hence, here we can start to see the similarities in the formulations for "good" parity-check matrix H_{CS} and "good" measurement H_{CS}. Namely, if we have no low pseudoweight vectors in the fundamental cone of H_{CS} then there are no problematic points in the nullspace of H_{CS}. This statement would be more clear after we see the next lemma.
Note that Lemma 3 preserves the support of the given vector $\boldsymbol{v}$.

Lemma 4. *Let $H_{CS} \in \{0,1\}^{m \times n}$ be a CS measurement matrix and let k be a non-negative integer. Then*

$$w_p^{BSC,\min}(H_{CS}) > 2k \implies H_{CS} \in NSP_{\mathbb{R}}^{<}(k, C = 1)$$

Proof. We select any arbitrary vector $\boldsymbol{v} \in Nullsp_{\mathbb{R}}(H_{CS}) \backslash \{\mathbf{0}\}$. By lemma 3, we know that $|\boldsymbol{v}$ is a pseudocodeword in of H_{CS}. And, as $w_p^{BSC,\min}(H_{CS}) > 2k$, we know that $w_p^{BSC}(|\boldsymbol{v}|) > 2k$. Then using lemma 2, we conclude that for all sets $S \subseteq I$ with $|S| \leq k$, we must have

$$||\boldsymbol{v}_S||_1 = |||\boldsymbol{v}_S||_1 < |||\boldsymbol{v}_{\bar{S}}|||_1 = ||\boldsymbol{v}_{\bar{S}}||_1$$

Middle inequality follows from the lemma 2. And because the choice $\boldsymbol{v}$ was arbitrary, we see that $H_{CS} \in NSP_{\mathbb{R}}^{<}(k, C = 1)$. □

This results guarantees us that given a good parity-check matrix H_{CS} i.e. a matrix which can detect k-bit flip errors or in other words, have minimum pseudoweight greater than $2k$, we can use the same matrix H_{CS} as a measurement matrix that can recover $k-$sparse signals.
There are many good parity-check matrices available(or construction available) based on expander graphs with property of large minimum pseudoweights, that can be used as measurement matrices. We show one such implementation in the next section.

4 Implementation

Here we use the LDPC matrix suggested in [2]. It is a matrix $H(q,l) \in 0,1^{ql \times q^2}$, where q is prime and $l \leq q-1$. It is constructed using a permutation matrix $P \in 0,1^{q \times q}$. More precisely, it is a shifted matrix such that $P(i,i-1) = 1 \quad \forall 1 < i \leq q$ (for $i = 1$, $P(1,q) = 1$), all other entries are zero. Thus the final construction of $H(q,l)$ looks like the following-

$$H(q,l) = \begin{bmatrix} I & I & \cdots & I \\ I & P & \cdots & P^{q-1} \\ \vdots & \vdots & \ddots & \vdots \\ I & P^{l-1} & \cdots & P^{(l-1)(q-1)} \end{bmatrix}$$

If we consider the bipartite graph formed from $H(q,l)$ using its column indices as left nodes, and row indices as right nodes, and edges are formed by considering a 1 at position (i,j) as an edge from jth left node to ith right node in the graph. We can see that the graph is left-regular (i.e. it's left nodes have the same degree ($l > 1$ in this case). Also every column and row have atleast 2 ones (as q is prime and l ¿ 1), and the inner product of any two columns is atmost 1. Thus using Theorem 13, from [2], we can say that the girth (size of smallest cycle) of this matrix is 6. Hence, by Theorem 11,eq (31) from [2] gives us -

$$m \geq (k+1)n^{1/2}$$

Where, k is the sparsity of the vectors being recovered using $H(q,l)$. Now, as $m \leq n$, we have $k \leq \sqrt{n}$.
We perform the experiment for $q = 101$, or $n = 10201$, and for 2 values of k. $k = 17$(corresponding $m = 1717$) and $k = 30$($m = 3030$). The values of m for random gaussian matrix are $m = 9000$ for $k = 17$, and $m = 13800$ for $k = 30$ (These are calculated using eq (6) and the preceding theorem in [2]). And as can be seen, this also results in m for $k = 30$ becoming greater than n, so basically Gaussian matrix is useless for such high sparsity. Further the results were averaged over 10 repetitions on each (k,matrix) pair. The results of reconstruction are as follows-

	LDPC Matrix		Random Gaussian Matrix	
k	m_L	T in sec	m_G	T in sec
17	1717	0.224	9000	0.589
30	3131	1.043	13800	2.575

Table 1: Results

The time of reconstruction was calculated for equal reconstruction quality, (measured in terms of RMSE) (values can be seen by running code).

References

[1] A. G. Dimakis, R. Smarandache, and P. O. Vontobel. Ldpc codes for compressed sensing. *IEEE Transactions on Information Theory*, 58(5):3093–3114, 2012.

[2] M. Lotfi and M. Vidyasagar. Compressed sensing using binary matrices of nearly optimal dimensions. *IEEE Transactions on Signal Processing*, 68:3008–3021, 2020.