EC601 Project 1: Low-density parity-check matrices for compressive sensing applications case study

Steven Hastings

1 Introduction

The high bandwidth of modern communications technologies require very high sampling rates in order to satisfy the Nyquist criterion. Consequently high speed ADCs and large storage capacities are needed. The high power consumption and high costs of such systems however has led to the search for ways to improve performance. With the advent of compressive sensing (CS), new approaches are being discovered that allow for signal reconstruction at sub-Nyquist rates. There are two main areas of research in CS: algorithms for reconstructions e.g., basis pursuit and orthogonal matching pursuit (OMP), and finding measurement matrices that satisfy or with high probability satisfy the restricted isometry property (RIP). In this case study, a special type of measurement matrix is examined. In coding theory, there are a class of codes known as low-density parity-check (LDPC) codes. For one type of LDPC construction that we will look at, the parity check matrices thereof so happen to meet the RIP criteria with high probability, allowing for sparse signal reconstruction under CS.

2 Compressive Sensing

Compressive sensing involves non-uniform random sampling of a signal with something called an Analog to Information Converter (AIC) and reconstructing the original signal using optimization techniques. The signal is assumed to be sparse. A signal x of size N with K non-zero coefficients and $N-K$ zero coefficients in some transform basis such that $K \ll N$ is said to be sparse, or K-sparse. Sparse signals require fewer samples to reconstruct than dense signals do. The system can be represented as

$$y = \Phi x = \Phi \Psi s = \Theta s$$

where Φ is the $M \times N$ measurement matrix, y is the $M \times 1$ measured values, Ψ is the $N \times N$ representation matrix, and s is the N coefficients. When written in succinct form, the matrix Θ is the $M \times N$ reconstruction matrix. If the basis is linearly dependent, it is called a frame. The basis and frame are often called the dictionary, the elements of which are called atoms. The aim is to find s given y and Θ. Then x can be easily computed with an inverse transform.

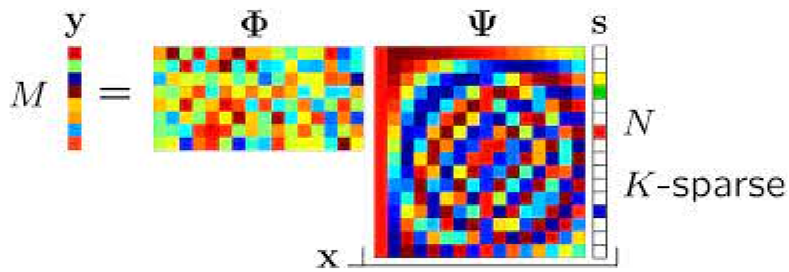

Fig. 1. The system
Source: http://users.isr.ist.utl.pt/~aguiar/CS_notes.pdf

In most cases, the reconstruction of the signal cannot be absolutely guaranteed, but under certain conditions the signal can be reconstructed with very high probability. The first condition that must be met is that the matrices have the restricted isometry property (RIP) or have it with high probability, which simply stated, means that any set of columns in the matrix are approximately orthogonal, or that the transform $y = \Theta s$ is unitary. The RIP condition is met if Φ and Ψ are incoherent (uncorrelated). The second condition is that $cK \log(N/K))$ measurements be made, where c is a constant depending on coherence between Φ and Ψ.
Because the system has fewer equations than unknowns, it is an underdetermined system and has infinite solutions. We are looking for the solution that is most sparse. We may minimize the ℓ_0-norm of s wrt. the system to obtain the sparsest solution.

The ℓ_0-norm–rigorously speaking it is not a norm, so it is often referred to as a pseudonorm–or sparse norm counts the number of non-zero elements.

$$||s||_0 = \lim_{p \to 0} \sum_k |s_k|^p$$

However, for $0 < p < 1$, finding the global minimum is intractible as it is an NP-hard problem, meaning there is no known algorithm that can solve the optimization problem in polynomial time. The ℓ_1-norm approximates the ℓ_0-norm in many cases as can be seen geometrically in Figure 2. It is also known as the Taxicab norm and the Manhattan norm because it measures the distance a taxicab would have to drive to reach its destination in certain parts of Manhattan (where the streets form a gridlike structure).

$$||s||_1 = \sum_k |s_k|$$

By the triangle inequality, it can be shown that the ℓ_1-norm is a convex function, and we will therefore need to employ convex optimization methods in order to minimize it. But it is possible to convert the problem to a linear programming problem which is more computationally efficient.

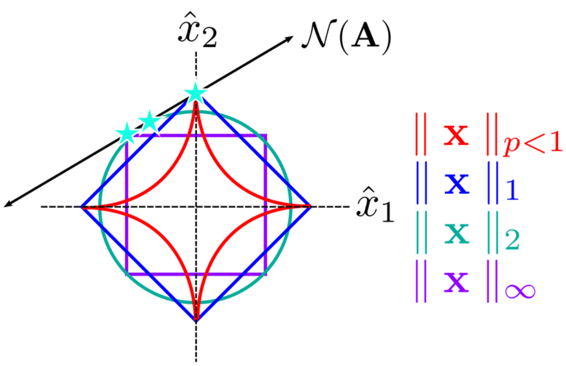

Fig. 2. p-norms
Source: https://www.researchgate.net/figure/Spheres-of-the-L-p-norm-and-their-intersections-marked-by-stars-with-the-null-space_fig3_325593608

Minimizing the ℓ_1-norm is known in the literature as Basis Pursuit (BP).

$$\hat{s} = \underset{s:y=\Theta s}{\operatorname{argmin}} ||s||_1$$

In the presence of noise, the system is expressed as $y = \Theta s + \eta$. The energy of the noise is bounded by a constant ϵ: $||\eta||_2 = ||y - \Theta s||_2 \leq \epsilon$. The optimization problem can then be written as follows

$$\hat{s} = \underset{s:||y-\Theta s||_2 \leq \epsilon}{\operatorname{argmin}} ||s||_1$$

Other methods such as Orthogonal Matching Pursuit (OMP) use a greedy heuristic algorithm to solve the system. OMP is faster and has a simpler implementation than BP, but requires more measurements than BP.

In this case study, our focus is the measurement matrix rather than reconstruction algorithms. The Gaussian, and Bernoulli random matrices as well as Fourier matrices have been shown to satisfy RIP with high probability. Gaussian matrices are the most popular in the literature. They have very good orthogonality but have a relatively high computational load because of their randomness and density. A deterministic matrix with high orthogonality and an efficient implementation would be ideal. Such a class of matrices exist. The parity check matrix of the low-density parity-check (LDPC) code used in error correction has these properties.

3 Coding Theory

LDPC codes are a type of error-correcting code often used in communications, specifically Ethernet, digital TV, satellite communications, as well as in SSD hard drives. LDPC codes are linear block codes that are subject to sparsity and orthogonality constraints for purposes of efficiency, which can be exploited by compressive sensing algorithms. The LDPC code is defined by the null space of the LDPC parity check matrix. The codes are represented using something called a Tanner graph. Tanner graphs are bipartite graphs--graphs that can be divided into two disjoint sets of vertices where none of the vertices in the same set are adjacent to each other. One set corresponds to data bits, and the other set parity check bits. The eponymous Gallager code, the original LDPC code, used a pseudo-random number generator (PRNG) to generate the graph. Gallager showed that randomized relationships between the data and parity bits allowed error correction. Later, structured constructions (non-random) were discovered. One such construction is the quasi-cyclic LDPC (QC-LDPC) code, which are used in 5G nr communications. The QC-LDPC parity check matrix consists of circulant submatrices (Figure 3). A permutation or shift matrix determines how much to shift each row of a given submatrix. The permutation matrix is generated using a progressive edge-growth (PEG) algorithm that aims to maximize the girth of the parity check matrix. Girth–the length of the smallest cycle–corresponds to the degree of orthogonality in a matrix representation. A QC-LDPC matrix with a large girth should meet the requirements of the RIP with high probability.

$$\begin{array}{ccc|ccc|ccc|ccc}
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
\hline
0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0
\end{array}$$

Fig. 3. Example of an LDPC parity check matrix
Source: https://www.researchgate.net/figure/Parity-check-matrix-for-QC-LDPC-codes_fig5_224149189

4 Methodology

The goal is to construct LDPC matrices with a large girth g. However, as the the size of the matrix and number of non-zero elements increases, the cycle lengths tend to decrease. So for a desired degree of column sparsity d, we try to find the largest g possible. The parity check matrix is constructed using PEG algorithms for different compression ratios and sizes.

5 Results

Using mean-square error (MSE) as one performance metric, it was found that LDPC matrices of various sizes significantly outperform Gaussian and sparse random matrices of equivalent size beyond a certain d depending on the size of the matrix. Figure 4. shows a comparison of the MSE using LDPC, Gaussian, and sparse random matrices of size 1024x4096 with sparsity $k = 200$ and $k = 300$.

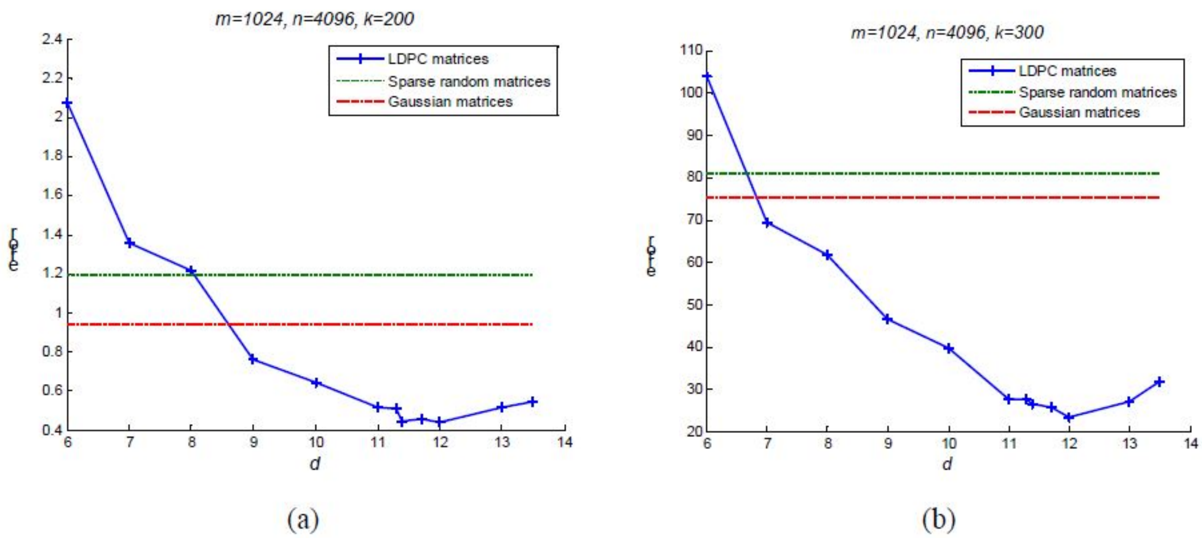

Fig. 4. MSE comparison. Source: [1]

6 Summary & Conclusions

LDPC matrices as measurement matrices have been shown to outperform Gaussian matrices in the MSE of the reconstructed signal. In addition, the non-random quasi-cyclic construction lends itself well to shift-register architecture allowing for more efficient processing in hardware than hardware for generating Gaussian random variables.

7 Future work

The next steps were I to continue on this topic in the future would be to try and reproduce the results in [1] and [2] and tune the parameters for an application of interest. One application worth exploring is pulse compression radar. Some research [4] has found that Costas coded radar pulses can be reconstructed at as much 16x below the Nyquist rate! It would be interesting to see if using LDPC rather than Gaussian measurement matrices, as were used in the study, would yield even more precise reconstructions at such low sampling rates.

References

[1] W. Lu, Kidiyo Kpalma, and J. Ronsin, "Sparse Binary Matrices of LDPC Codes for Compressed Sensing," HAL (Le Centre pour la Communication Scientifique Directe), Apr. 2012, doi: https://doi.org/10.1109/dcc.2012.60.

[2] Y. Wang, Y. Qin, and H. Ren, "Deterministic Construction of Compressed Sensing Measurement Matrix with Arbitrary Sizes via QC-LDPC and Arithmetic Sequence Sets," Electronics, vol. 12, no. 9, pp. 2063–2063, Apr. 2023, doi: https://doi.org/10.3390/electronics12092063.

[3] S. L. Brunton and J Nathan Kutz, Data-Driven Science and Engineering Machine Learning, Dynamical Systems, and Control. Cambridge University Press, 2019.

[4] A. Hanif, A. B. Mansoor, and A. S. Imran, "Sub-Nyquist sampling and detection in Costas coded pulse compression radars," EURASIP Journal on Advances in Signal Processing, vol. 2016, no. 1, Nov. 2016, doi: https://doi.org/10.1186/s13634-016-0424-1.