UNIVERSITY OF THE PHILIPPINES

COMPRESSIVE SENSING: APPLICATIONS FROM 1-D TO N-D

By

KENNETH V. DOMINGO

A Baccalaureate Thesis Submitted to the

National Institute of Physics

College of Science

University of the Philippines

Diliman, Quezon City

In Partial Fulfillment of the Requirements

For the Degree of

B.S. Applied Physics (Instrumentation)

April 2020

Acknowledgments

Abstract

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The trend of both curiosity and profit-driven human development has caused a surge in the amount of openly accessible raw data. More often than not, the data is generated much faster than it can be processed. In the endeavor of keeping up with information inflow, there are two major factors that significantly hinder our progress. First, Moore's law implicitly sets a physical limit to the number of transistors that can be placed on a fixed-size chip (barring a paradigm change in the design and manufacture of semiconductors), consequentially limiting how powerful and how fast electronic systems can become. The second is the Nyquist-Shannon sampling theorem (NST), which limits the range of frequencies a recording device can successfully capture. This study explores the use of compressive sensing (CS)—an emergent sampling theorem that allows reconstruction of signals from much fewer samples than required by the NST by taking advantage of some inherent sparsity or periodicity in a signal—as a viable method for compression, encryption, and/or enhancement. In this framework, the computational burden of encoding/decoding is shifted from the sampling device to the device performing

reconstruction, decompression, and other modes of post-processing. As such, there exist many ways to reconstruct a signal from compressive measurements. CS has found its applications in simple audio signals containing pure tones [1, 2] and speech [3–5], images [6–8], and audioless grayscale videos [9, 10]. The formulation of a sensing matrix in CS requires a basis conforming to some uniform uncertainty principle, and the most common starting point would be a partial discrete cosine transform (DCT). Recent studies, however, have shown that learned bases perform much better on more complex signals [11–13], i.e., those that would be typically encountered in practical situations. The learning algorithms associated in the construction of these bases range from classical iterative methods, which have long been used in optimization problems, to the more contemporary machine learning methods.

1.1 Related literature

In 2006, Candès, Romberg, Tao [14], and Donoho [15] asked the question,

> With the recent breakthroughs in lossy compression technologies, we now know that most of the data we acquire can be thrown away with minimal perceptual loss. Why bother to acquire all the data when we can just directly measure the part that will not be thrown away?

which was eventually answered in many different ways by the same people, ultimately birthing the field which we now know as compressive sensing. The methods in CS apply concepts from time-frequency uncertainty principles [16] and sparse representations, which were studied rigorously by Donoho and Elad [17]. CS can be viewed as a strategic undersampling method: the signal is sampled

at random locations in the real domain, and the ratio of the indices where it is sampled to the signal size can be associated with some quasi-frequency which may or may not be lower than the Nyquist rate. The reason for this indifference towards the Nyquist rate is that the compressive samples are not equitemporal and thus, cannot "violate" the Nyquist theorem. Therefore, CS should not be considered a "replacement" or "rival" to the NST, but rather, a completely distinct model of signal acquisition.

[18] demonstrated the use of deterministic chaos filters to acquire samples instead of random distributions. Sampling using a Gaussian-Logistic map was applied to acoustic signals in [1]. Normally, a deterministic chaotic function will need one or more initialization values as a "seed", and the sequence of numbers produced by different combinations of initial values rapidly diverge from each other. This phenomenon led to investigating the use of compressive sensing as an encryption algorithm. Simultaneous compression and encryption was achieved by [6], and it was found that perturbations on the order of 10^{-15} to the initial values are sufficient to make an encrypted image remain as such. Their image compression-encryption model via compressive sensing was shown to have a key space on the order of 10^{34}, making it extremely resistant to brute force and other types of attacks. This was extended in [7] to utilize higher-dimensional chaotic systems, subsequently expanding the key space to the order of 10^{83}. In the methods above, sampling was performed in the signal domain (i.e., time domain for audio, spatial domain for images), and the reconstruction was performed in the frequency domain. [2] proposed a method to perform both sampling and reconstruction in the time domain using differential evolution.

Audio signals, compared to images, are much more densely packed with

information. Whereas image size is dependent on the resolution and bit depth of the imaging device, audio size scales proportionally with time. The accepted frequency range of human hearing is from 20 Hz to 20 kHz, so by the NST, a sampling frequency of at least 40 kHz is needed to preserve the fidelity of recorded audio. Any meaningful audio recording, especially those containing speech, will certainly have a duration of a few seconds up to a few hours, so one cannot straightforwardly apply methodologies used for images or recordings with relatively static frequencies, as the first challenge this would pose for electronic systems is insufficient memory to process the entire signal all at once. Low [3, 4] circumvented this problem by transforming the signal to the modulation domain, i.e., the signal's spectrogram, essentially raising a one-dimensional signal to N-dimensions, where the value of N is dependent on the desired spectrogram resolution, number of subbands, and percent overlap between adjacent bands. In such signals, recordings were also observed to have a noise floor, which could be easily removed by an inherent denoising property of CS [19].

Due to the large size of video information as a consequence of its high dimensionality, it is possible, but impractical, to apply image CS techniques on an entire frame-by-frame basis. Correlations between adjacent frames are utilized instead, and can be obtained using dictionary learning [11] or principal components analysis [9]. For the same reason, the application of CS to grayscale videos presuppose the use of machine learning methods. Iliadis [20, 21] came up with two different deep neural network architectures whose inputs and outputs are patches derived from grayscale videos.

4

1.2 Novelty

Previous studies worked with either pure audio or pure image/image sequence signals as the target for compressive sensing. This study works on applying previously established compressive sensing methods to N-dimensional signals, such as spectrograms and full-color video coupled with audio. N will be bound by the type of signals being worked with, as well as the computational power of the decoding device. In particular, large values of N are useful in encryption, where a signal is first raised to a high dimension, the sensing matrix is derived also from a high dimension, and the result is usually cast back to dimension one or two.

Chapter 2

Preliminaries

2.1 Compressed sensing

Consider a discrete time signal $\mathbf{x} \in \mathbb{R}^N$. Consider also an orthonormal basis $\mathbf{\Psi} \in \mathbb{R}^{N \times N}$ whose column vectors are expressed as $\boldsymbol{\psi}_i$. The signal $\mathbf{x}$ can be expressed as

$$\mathbf{x} = \langle \mathbf{\Psi} | \boldsymbol{\alpha} \rangle = \sum_{i=1}^{N} \alpha_i^{\top} \boldsymbol{\psi}_i \tag{2.1}$$

where α_i are the coefficients of $\mathbf{x}$ in the $\mathbf{\Psi}$ domain. For practical signal processing applications, a sampled signal is usually of lower dimension than its original representation. We define the compressed vector $\mathbf{y} \in \mathbb{R}^M$, $M < N$ which is obtained by

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{2.2}$$

where $\mathbf{A} \in \mathbb{R}^{M \times N}$ ($M \leq N$) is the sensing matrix. According to the NST, a periodic signal—which may be composed of a linear superposition of sinusoids with

different parameters—has a characteristic bandwidth f_B, defined by its highest frequency component. The signal can be successfully reconstructed by a sampling device if the signal is sampled at a fixed rate with frequency f_s, which is at least twice f_B; that is $f_s \geq 2f_B$, where $2f_B$ is known as the Nyquist rate [22]. Under the NST, the original signal $\mathbf{x}$ in (2.2) can be recovered by inversion, least squares approximation, or more commonly, sinc interpolation. More often than not, the inversion of (2.2) is ill-posed because it is an underdetermined system. However, recovery is still possible provided that $\mathbf{x}$ is sparse or can be represented sparsely in some domain, and $\mathbf{A}$ satisfies one of the following properties [23]:

- **Low mutual coherence**: The mutual coherence of a matrix $\mathbf{A}$ is defined as

$$\mu(\mathbf{A}) = \max_{i \neq j} |\langle \mathbf{a}_i | \mathbf{a}_j \rangle| \tag{2.3}$$

- **Restricted isometry property** (RIP): Matrix $\mathbf{A}$ satisfies the RIP with sparsity parameter k and restricted isometry constant δ if for all k-sparse vectors $\mathbf{x}$,

$$(1 - \delta)\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 - \delta)\|\mathbf{x}\|_2^2 \tag{2.4}$$

The sparsity parameter k entails that a vector has, at most, k nonzero coefficients, in which case the vector is said to be k-sparse [14].

Once these have been satisfied, $\mathbf{x}$ can be recovered exactly from the compressive measurements $\mathbf{y}$ by solving the combinatorial minimization problem

$$\min \|\mathbf{x}\|_0 \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{y} \tag{2.5}$$

where $\|\mathbf{x}\|_0$ denotes the ℓ_0 pseudo-norm of $\mathbf{x}$, which extracts the number of nonzero

coefficients. However, this problem is computationally intractable even for a small signal. A more tractable solution is to reduce this to a convex problem

$$\min \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{Ax} = \mathbf{y} \tag{2.6}$$

where $\|\mathbf{x}\|_1 \equiv \sum_i |x_i|$ denotes the ℓ_1 norm of $\mathbf{x}$. For a sufficiently sparse signal, the solutions to (2.5) and (2.6) are identical [24]. A plethora of algorithms exist dedicated to solving (2.6). The scope of this study utilizes the following algorithms used commonly in the literature:

- **Least absolute shrinkage and selection operator** (LASSO): recasts the ℓ_1 minimization problem as an ℓ_1-regularized least squares problem:

$$\min_{\mathbf{x}} \frac{1}{2N} \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \alpha \|\mathbf{x}\|_1 \tag{2.7}$$

  where α is the regularization hyperparameter. Optimization is performed via coordinate descent [25].

2.2 Associative memory networks

Chapter 3

Two-dimensional compressive sensing (2DCS)

Chapter 4

One-dimensional compressive sensing (1DCS)

Chapter 5

Generalization to N-dimensions

Chapter 6

Conclusions

Bibliography

[1] M. R. Mathew and B. Premanand, Sub-Nyquist Sampling of Acoustic Signals Based on Chaotic Compressed Sensing, *Procedia Technol.* **24**, 941 (2016), ISSN 22120173.

[2] I. Andráš, P. Dolinský, L. Michaeli, and J. Šaliga, A time domain reconstruction method of randomly sampled frequency sparse signal, *Meas. J. Int. Meas. Confed.* **127**, 68 (2018), ISSN 02632241.

[3] S. Y. Low, D. S. Pham, and S. Venkatesh, Compressive speech enhancement, *Speech Commun.* **55**, 757 (2013), ISSN 01676393.

[4] S. Y. Low, Compressive speech enhancement in the modulation domain, *Speech Commun.* **102**, 87 (2018), ISSN 01676393.

[5] V. Abrol, P. Sharma, and A. K. Sao, Voiced/nonvoiced detection in compressively sensed speech signals, *Speech Commun.* **72**, 194 (2015), ISSN 01676393.

[6] Y. Mo, A. Zhang, F. Zheng, and N. Zhou, An image compression-encryption algorithm based on 2-D compressive sensing, *J. Comput. Inf. Syst.* **9**, 10057 (2013), ISSN 15539105.

[7] N. Zhou, S. Pan, S. Cheng, and Z. Zhou, Image compression-encryption scheme based on hyper-chaotic system and 2D compressive sensing, *Opt. Laser*

Technol. **82**, 121 (2016), ISSN 00303992.

[8] R. A. Romero, G. A. Tapang, and C. A. Saloma, in *Proceedings of the Samahang Pisika ng Pilipinas Physics Conference* (University of the Philippines Visayas, Iloilo City, Philippines, 2016), vol. 34, SPP–2016–PA–14.

[9] S. Liu, M. Gu, Q. Zhang, and B. Li, Principal component analysis algorithm in video compressed sensing, *Optik (Stuttg).* **125**, 1149 (2014), ISSN 00304026.

[10] J. Chen, K. X. Su, W. X. Wang, and C. D. Lan, Residual distributed compressive video sensing based on double side information, *Zidonghua Xuebao/Acta Autom. Sin.* **40**, 2316 (2014), ISSN 02544156.

[11] H. Liu, B. Song, H. Qin, and Z. Qiu, Dictionary learning based reconstruction for distributed compressed video sensing, *J. Vis. Commun. Image Represent.* **24**, 1232 (2013), ISSN 10473203.

[12] P. Sharma, V. Abrol, Nivedita, and A. K. Sao, Reducing footprint of unit selection based text-to-speech system using compressed sensing and sparse representation, *Comput. Speech Lang.* **52**, 191 (2018), ISSN 10958363.

[13] N. Eslahi, A. Aghagolzadeh, and S. M. H. Andargoli, Image/video compressive sensing recovery using joint adaptive sparsity measure, *Neurocomputing* **200**, 88 (2016), ISSN 18728286.

[14] E. J. Candès, J. K. Romberg, and T. Tao, Stable signal recovery from incomplete and inaccurate measurements, *Commun. Pure Appl. Math.* **59**, 1207 (2006), ISSN 00103640, arXiv:0503066v2.

[15] D. L. Donoho, Compressed sensing, *IEEE Trans. Inf. Theory* **52**, 1289 (2006), ISSN 00189448.

[16] D. Donoho and X. Huo, Uncertainty principles and ideal atomic decomposition,

IEEE Trans. Inf. Theory **47**, 2845 (2001), ISSN 00189448.

[17] D. L. Donoho and M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via L1 minimization, *Proc. Natl. Acad. Sci. U. S. A.* **100**, 2197 (2003), ISSN 00278424.

[18] N. Linh-Trung, D. Van Phong, Z. M. Hussain, H. T. Huynh, V. L. Morgan, and J. C. Gore, Compressed sensing using chaos filters, *Proc. 2008 Australas. Telecommun. Networks Appl. Conf. ATNAC 2008* 219–223 (2008).

[19] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, Image denoising by sparse 3-D transform-domain collaborative filtering, *IEEE Trans. Image Process.* **16**, 2080 (2007), ISSN 10577149, arXiv:arXiv:1011.1669v3.

[20] M. Iliadis, L. Spinoulas, and A. K. Katsaggelos, Deep fully-connected networks for video compressive sensing, *Digit. Signal Process. A Rev. J.* **72**, 9 (2018), ISSN 10512004.

[21] M. Iliadis, L. Spinoulas, and A. K. Katsaggelos, DeepBinaryMask: Learning a binary mask for video compressive sensing, *Digit. Signal Process. A Rev. J.* **96**, 102591 (2020), ISSN 10512004, arXiv:1607.03343.

[22] C. E. Shannon, Communication in the presence of noise, *Proceedings of the Institute of Radio Engineers* **37**, 10 (1949).

[23] A. Mazumdar and A. S. Rawat, Associative memory via a sparse recovery model, *Adv. Neural Inf. Process. Syst.* 2701–2709 (2015), ISSN 10495258.

[24] E. J. Candès, J. Romberg, and T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Inf. Theory* **52**, 489 (2006), ISSN 00189448, arXiv:0409186.

[25] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel,

15

M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research* **12**, 2825 (2011).