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

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 Nyquist-Shannon sampling theorem (NST)—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, or 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 stable frequencies, such as pure tones [1, 2], and dynamic frequencies, such as speech [3–5], images [6–8], and grayscale videos [9, 10]. The formulation of a sensing matrix in CS requires a basis conforming to some uniform uncertainty principle, and most common starting points would be partial discrete cosine transforms (DCT) or partial discrete wavelet transforms (DWT). 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 real life situations. The

1

learning algorithms associated with 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 2004, 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?

This 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 size of the signal can be associated with some quasi-frequency which may be lower than the Nyquist rate.

Linh-Trung et al. [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 the initial values were sensitive to perturbations on the order of 10^{-15}. 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 a higher-dimensional variant of the Lorenz attractor, subsequently expanding the key space to the order of 10^{83}. In the methods above, sampling was performed in the signal domain (i.e., temporal domain for audio, spatial domain for images), and the reconstruction was performed in the frequency domain with a DCT or similar basis. [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 images are not naturally bandlimited and rather, are dependent on the spatial resolution and bit depth of the imaging device, audio size scales proportionally with time and takes on a wider range of values. 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 ensure that an audio sample is recorded correctly. 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 subbands. In such signals, recordings with an observed noise floor could be easily be denoised, which is an inherent 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. This idea was utilized in [22] who modified the architecture into a residual network containing several convolutional layers. The original design was targeted towards image reconstruction, but could easily be extended to videos.

In the same vein, neural network methods could also be used in CS of speech. Advances in natural language processing were primarily made using recurrent neural networks (RNN). In [23], a speech signal was first modeled by their proposed RNN architecture based on a noise-constrained least squares estimate, and final recovery is done via Kalman filtering. A new simple recurrent unit (SRU) network was created in [24] which maps the relation between noisy and clean speech recordings for speech enhancement.

1.2 Novelty

This study aims to provide a generalization for applying CS techniques to signals of arbitrary dimensions. Previous studies worked exclusively with either audio or image sequences as the target for CS, and due to the computational demands, the focus of most of the research in the field has been to optimize the computational complexity for real-time applications, and improve signal reconstruction quality. In the establishment of CS methods, two different general frameworks to compressively sample signals arise, namely, one-dimensional CS (1DCS) and two-dimensional CS (2DCS). It is shown that an N-dimensional signal can be decomposed into factors of one-dimensional and two-dimensional signals, and can be processed using methods appropriate for each type of signal. Furthermore, it is shown that N is bound not only by the type of signals being worked with, but also the computational power of the decoding/decompressing device. In particular, large values of N are useful in encryption, where a signal is first raised to a high dimension in a certain basis, the sensing matrix is derived from another high-dimensional basis, and the result is cast back to either one or two dimensions to yield the encrypted message.

1.3 Thesis overview

The next chapter establishes the relevant mathematical concepts and notation to be used throughout this study, algorithms used in signal reconstruction, and appropriate metrics per type of signal. Chapters 3–5 respectively focus on two-dimensional CS, one-dimensional CS, and N-dimensional CS. The reason behind the ordering of Chapters 3 & 4 will become apparent as the usage of spectrograms are introduced. Each of these chapters are self-contained methodologies, results,

and discussions to emphasize that the methods can work independently of each other, save for the generalization to N-dimensions. The study is concluded and recommendations for future studies are presented in Chapter 6.

Chapter 2

Preliminaries

2.1   Compressive sensing

Nearly all of today's signal acquisition technologies were made possible thanks to
the Nyquist-Shannon sampling theorem (NST). This states that given that you
know a signal's highest frequency component f_B, sampling it at a rate f_S that is
at least twice this frequency is sufficient to capture all of the pertinent information
regarding that signal; that is $f_S \geq 2f_B$, where f_B is known as the Nyquist rate [25],
and also as the signal bandwidth. For signals that are not naturally bandlimited,
such as images, the ability reproduce a signal is dependent on the device's resolution
and still follows the same principle: there should be at least twice the number of
pixels in the same dimension as the image's highest spatial frequency. For practical
day-to-day use, the NST covers everything you need. However, issues arise when
bandwidth and storage are at a premium. Typically, after sensing a signal, not all
of the raw data is stored. Rather, this data is converted to a compressed format
by systematically discarding the imperceptible information. Thus, the process of
acquiring massive amounts of data followed by compression is extremely wasteful

[26], and so the goal of CS is to directly acquire the information which survives the compression stage.

Consider a real-valued signal $\mathbf{x} \in \mathbb{R}^{N \times 1}$. The process of acquisition or sensing this signal can be modeled as a linear system, where the physical signal values are transformed into values that can be understood by some sensing device by applying a linear transformation as

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{2.1}$$

where $\mathbf{A} \in \mathbb{R}^{M \times N}$ is the transformation commonly known as a sensing matrix. In the NST case, $\mathbf{A}$ is a square matrix ($M = N$), and one could easily reconstruct $\mathbf{x}$ by inversion. However, we are interested in the case when the signal is undersampled, i.e., the number of measurements is less than the size of the signal ($M \ll N$).

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] H. Yao, F. Dai, S. Zhang, Y. Zhang, Q. Tian, and C. Xu, DR2-Net: Deep Residual Reconstruction Network for image compressive sensing, *Neurocomputing* **359**, 483 (2019), ISSN 18728286.

[23] Y. Xia and J. Wang, Low-dimensional recurrent neural network-based Kalman filter for speech enhancement, *Neural Networks* **67**, 131 (2015), ISSN 18792782.

[24] X. Cui, Z. Chen, and F. Yin, Speech enhancement based on simple recurrent unit network, *Appl. Acoust.* **157**, 107019 (2020), ISSN 1872910X.

[25] C. E. Shannon, Communication in the presence of noise, *Proceedings of the*

Institute of Radio Engineers **37**, 10 (1949).

[26] E. J. Candes and M. B. Wakin, An introduction to compressive sampling: A sensing/sampling paradigm that goes against the common knowledge in data acquisition, *IEEE Signal Process. Mag.* **25**, 21 (2008), ISSN 10535888.