UNIVERSITY OF THE PHILIPPINES

COMPRESSIVE SENSING:

APPLICATIONS FROM 1-D TO N-D

By

KENNETH V. DOMINGO

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 a signal 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 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

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 into something interpretable or useful. In the endeavor of keeping up with the inflow of information, 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 chip, consequently limiting how powerful and how fast electronic systems can become (barring a paradigm shift in the fundamental design of semiconductors). The second is the Nyquist-Shannon sampling theorem (NST), which limits the range of frequencies a recording device can successfully capture. 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

7

and still follows the same principle: there should be at least twice the number of pixels codimensional with 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 values such that the loss of information is virtually imperceptible. Thus, the process of acquiring massive amounts of data followed by compression is extremely wasteful. CS aims to directly acquire the parts of the signal that would otherwise survive this compression stage in the classical sampling scheme.

Consider a signal $\mathbf{x} \in \mathbb{R}^n$; this notaton indicates that $\mathbf{x}$ is real-valued (but can easily be extended to complex-valued signals), one-dimensional vector of size n. The process of acquisition or sensing this signal can be modeled as a linear system, where the physical signal properties we wish to capture are transformed into digital values by applying a linear transformation

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

or in the literature of signal processing [26], by correlating them with a waveform basis

$$y_k = \langle \mathbf{x} | \mathbf{a}_k \rangle, \quad k \in \mathbb{N} \le n \tag{2.2}$$

In conventional sampling, $\mathbf{a}_k$ are Dirac basis vectors which turn $\mathbf{y}$ into a vector containing samples of $\mathbf{x}$ in the temporal or spatial domain; if $\mathbf{a}_k$ are Fourier basis vectors (i.e., sinusoids), then $\mathbf{y}$ is a vector of Fourier coefficients. If the signal

8

has been sampled sufficiently in the sense that the number of measurements m is equal to the dimension n of the signal, then $\mathbf{A}$ is a square matrix, and the original signal $\mathbf{x}$ can be reconstructed from the information vector $\mathbf{y}$ by inversion of (2.1). However, the process of recovering $\mathbf{x} \in \mathbb{R}^n$ from $\mathbf{y} \in \mathbb{R}^m$ becomes ill-posed when we consider the undersampled case ($m \ll n$), as the sensing matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$—whose row vectors are denoted as $\mathbf{a}_m$—causes the system to become underdetermined: there exist infinitely many candidate solutions $\hat{\mathbf{x}}$ which satisfy (2.1). To add to this, we also consider the possibility that the measurements are not perfect, and are contaminated with noise. How then do we recover a signal from measurements which are incomplete and most likely inaccurate? The answer lies in enforcing constraints based on models of natural signals, as well as constraints based on optimization techniques.

2.1   Sparsity

Most natural signals, especially those with some underlying periodicity, can be represented sparsely when expressed in the appropriate basis. This process of "sparsifying" can be expressed as

$$f = \langle \mathbf{x} | \boldsymbol{\psi}(k) \rangle \tag{2.3}$$

Similar to (2.2), this involves correlating the signal with the appropriate basis function to yield a representation in the sparse domain. Image information, for example, are commonly expressed in the DCT domain by

$$f_k = \sum_{n=0}^{N-1} x_n \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2}\right)k\right], \quad 0 \le k < N \tag{2.4}$$

9

and its corresponding inverse is

$$x_k = \frac{1}{2}f_0 + \sum_{n=1}^{N-1} f_n \cos\left[\frac{\pi}{N}\left(k + \frac{1}{2}\right)n\right], \quad 0 \le k < N \tag{2.5}$$

where the cosine term corresponds to $\psi(k)$ in (2.3). We can express (2.4) more conveniently as $\mathbf{f} = \mathbf{\Psi x}$, where $\mathbf{\Psi} \in \mathbb{R}^{n \times n}$ is the sparsifying matrix. Figure 2.1 shows this sparsifying process in action: given a test image, taking its DCT and zooming into the first 20,000 coefficients show that most of the signal energy is concentrated in just a few of the coefficients close to the DC term and decay rapidly. All the other coefficients, when compared to the k highest coefficients, are practically zero; such a signal is referred to as k-sparse. The compressed image resulting from discarding all but the first 25,000 coefficients and performing the inverse transform shows that any difference from the original image is virtually imperceptible. A similar concept is used in JPEG compression, wherein an image is divided into 8×8 blocks, and in each block, a certain number of DCT coefficients are discarded depending on the desired quality factor Q [27].
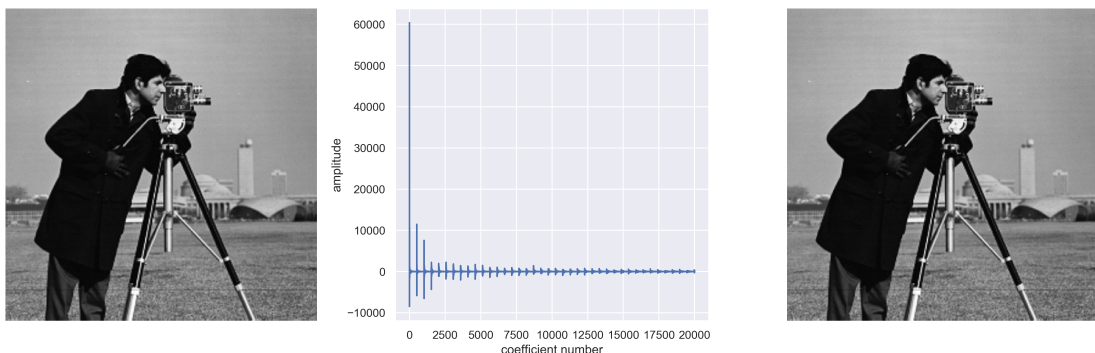

Figure 2.1: Original 512×512, 8-bit image (left), and its first 20,000 DCT coefficients (middle). Most of the signal energy is concentrated in just a few of the coefficients that are close to the origin. By discarding all but the first 25,000 coefficients and performing the inverse transform, the resulting image (right) is perceptually no different from the original.

10

2.2 Incoherence

Suppose we have two matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ which are involved in the sensing of a signal. As before, $\boldsymbol{\Psi}$ is the sparsifying matrix which converts the signal into a sparse representation, and $\boldsymbol{\Phi}$ is the actual sensing matrix. The coherence between these two bases is expressed as

$$\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi}) = \sqrt{n} \max_{0 \leq i,j < n} |\langle \boldsymbol{\varphi}_i | \boldsymbol{\psi}_j \rangle| \tag{2.6}$$

In other words, the coherence is the measure of the largest correlation between the column vectors of $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$. In compressive sensing, we are interested in low-coherence basis pairs (i.e., basis pairs for which $\mu \to 1$). For example, in the classical sampling scheme, $\boldsymbol{\Phi}$ is the Dirac basis $\boldsymbol{\varphi}_k(t) = \delta(t - k)$, and $\boldsymbol{\Psi}$ is the DCT basis (2.4). This basis pair in particular is also called a time-frequency pair and achieves maximum incoherence ($\mu = 1$) regardless of the number of dimensions [16]. Additionally, any orthonormal basis $\boldsymbol{\Phi}$ containing independent identically distributed (i.i.d.) entries are also largely incoherent with a fixed basis $\boldsymbol{\Psi}$ [26]. The consequence of this is that CS performs most efficiently when sensing with incoherent and random systems.

2.3 Reconstruction strategies

Another measure of the sparsity of a signal its ℓ_0 pseudo-norm, denoted $\|\mathbf{x}\|_0$, which simply counts the number of non-zero coefficients of $\mathbf{x}$. As such, the goal of the reconstruction stage in CS is to find the sparsest representation of the vector $\mathbf{x}$ in terms of the sensing matrix $\boldsymbol{\Phi}$ by solving the combinatorial optimization problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \mathbf{y} = \mathbf{\Phi}\mathbf{x} \tag{2.7}$$

which, as the name implies, requires one to enumerate all possible k-element combinations of the columns of $\mathbf{\Phi}$, and determining the smallest combination which approximates the signal the closest. However, this process quickly becomes intractable even for a modestly-sized signal. This requirement is therefore relaxed by instead minimizing the ℓ_1 norm

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{y} = \mathbf{\Phi}\mathbf{x} \tag{2.8}$$

where the ℓ_1 norm is defined as

$$\|\mathbf{x}\|_1 = \sum_{i=0}^{N-1} |x_i| \tag{2.9}$$

and is commonly called the taxicab or Manhattan distance. Aside from ℓ_1 minimization, other approximation methods include greedy approaches (e.g., basis pursuit), and convex programming.

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.

[27] CCITT Study Group VIII and Joint Photographics Expert Group, *T.81 – Digital compression and coding of continuous-tone still images – Requirements and guidelines* (1982).