Compressive sensing: Applications from 1-D to *N*-D

KENNETH V. DOMINGO
kdomingo@nip.upd.edu.ph
Video and Image Processing Group
Instrumentation Physics Laboratory, National Institute of Physics
University of the Philippines Diliman, Quezon City
Adviser: Maricor N. Soriano, Ph.D.

Introduction

Objectives

- To generalize **compressive sensing (CS)** to signals of arbitrary dimensions, for applications such as compression, encryption, and enhancement.
- To evaluate the reconstruction quality of compressively sampled signals using perceptually accurate metrics.
- To lay out the groundwork for similar applications of CS on signals containing a combination of audio and images (i.e., videos).

What is compressive sensing?

CS models signal treatment as an underdetermined linear system [1] which can be solved by imposing certain constraints. It exploits sparse representation of signals in some domain (typically the frequency domain) in order to recover them from much fewer samples than required by the Nyquist-Shannon sampling theorem (NST).

Why compressive sensing?

Because of the fewer samples required, CS is not as stringent in its acquisition speed and storage requirements, in contrast with conventional sampling techniques. Due to the lower rate, CS-based systems can potentially attain higher signal-to-noise ratio (SNR) [2] and dynamic range [3].

Image CS (2DCS)

Enhancement

A 1600 × 976 px image (Fig. 1) was divided into a 16 × 16 grid of 100 × 61 px patches. Each patch was compressively sampled with a 40% compression ratio, and was reconstructed by convex optimization [4]. The reconstructed image makes it apparent that the original image's background has some bias, and has been removed. For a perceptually accurate metric of the reconstruction quality, the Structural Similarity Index (SSIM) was used. SSIM values of 0.8 and above indicate acceptable quality.

Encryption

The sensing matrix is encoded by a deterministic chaotic function, whose behavior is controlled by some initial value parameter(s). These parameters act as encryption key sets, and one key set is required for each signal dimension. Simultaneous compression & encryption is achieved in Fig. 2, with a compression ratio of 75%. Figure 3 shows the MSE curves evaluated from attempted reconstruction for tiny perturbations in the initial values Δx_0 (in this example, a function with two initial value parameters was used). Reconstruction is only possible if the decoder uses the exact same key set used in the encryption stage [6].

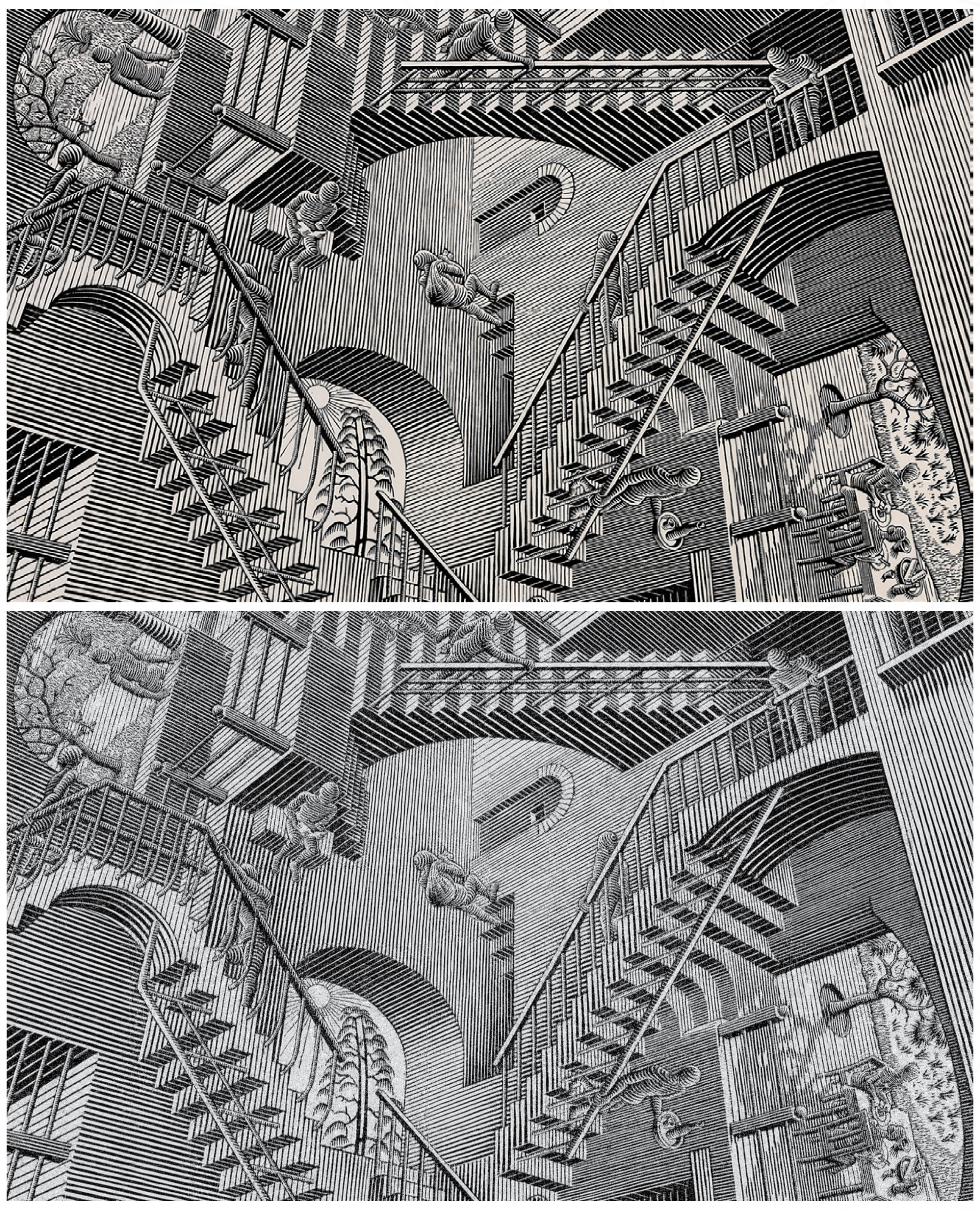

Figure 1: *Relativity*, by M.C. Escher (top), and the same image compressively sampled and reconstructed using 40% random samples from the original (bottom). SSIM = 0.88

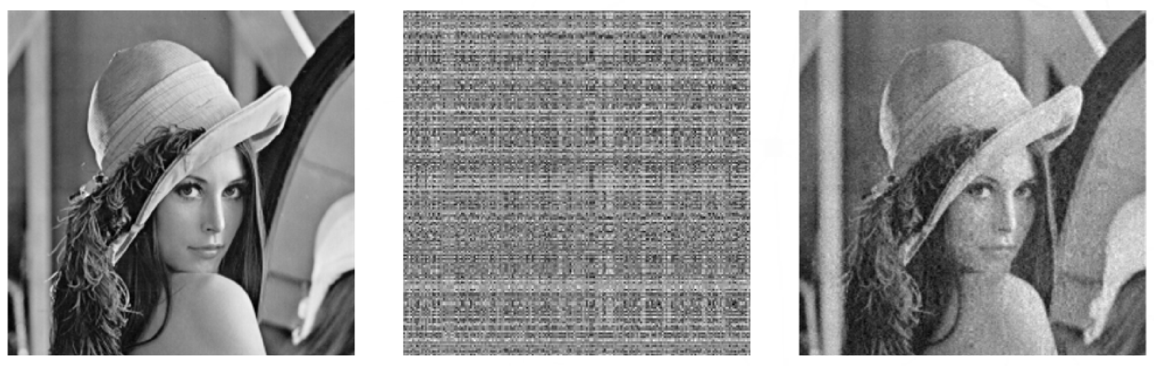

Figure 2: Simultaneous compression & encryption achieved with CS: original 256 × 256 image (left), encrypted 192 × 192 representation (middle), and decrypted image (right). SSIM = 0.82

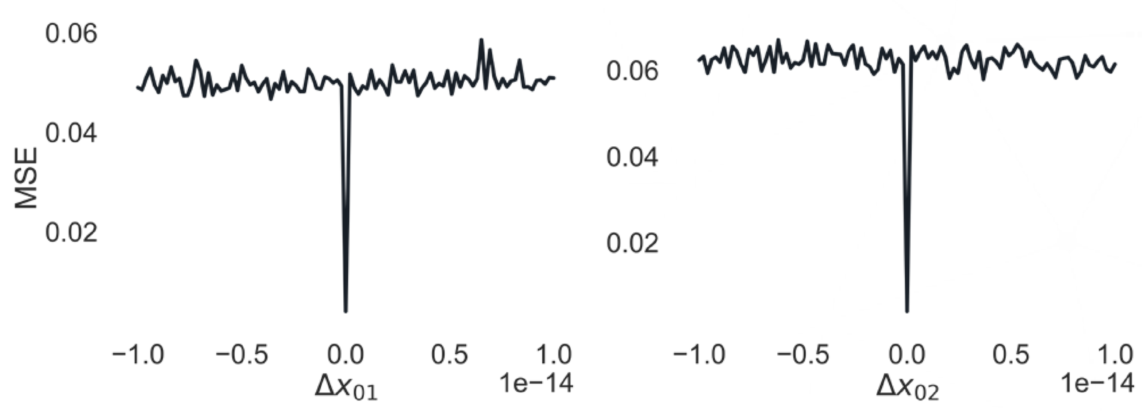

Figure 3: MSE curves resulting from evaluation of reconstruction error for tiny perturbations in the initial values Δx_{01} & Δx_{02}. Reconstruction is only possible for $\Delta x = 0$, i.e., the exact encryption keys.

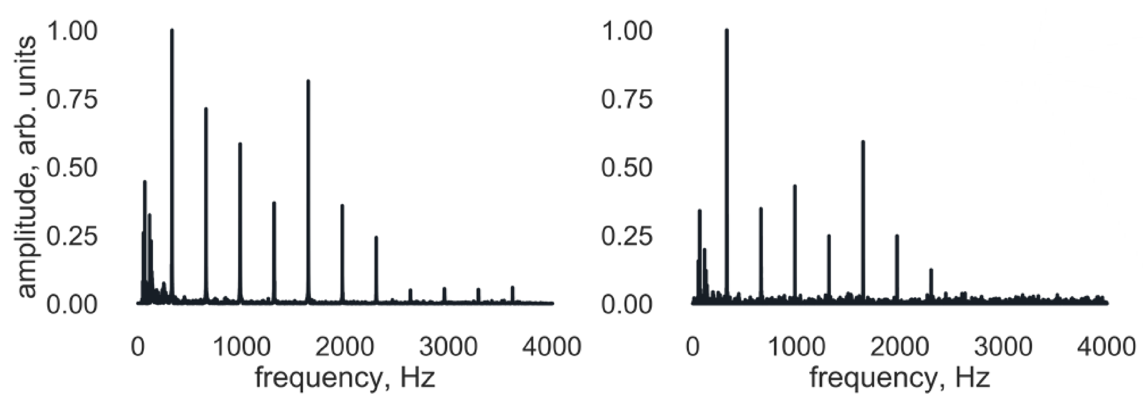

Figure 4: Base frequency (330 Hz) and harmonics of a recorded guitar tone (left). Reconstruction was able to recover the base, as well as up to the 6th harmonic (right). CS rate used was 1 kHz, corresponding to a Nyquist rate of 500 Hz.

References

[1] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, arXiv:math/0409186v1 (2006). | [2] 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. | [3] M. Davenport, J. Laska, J. Treichler, and R. Baraniuk. The pros and cons of compressive sensing for wideband signal acquisition: noise folding vs. dynamic range, arXiv:1104.4842 (2011). | [4] S. Diamond, and S. Boyd. CVXPY: A Python-embedded modeling language for convex optimization, *Journal of Machine Learning Research 17*(83), 1–5 (2016). | [5] K. Dabov, A. Foi, V. Kavotnik et al. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. on Image Processing 16*(8), 2080–2095 (2007). | [6] Y. Mo, A. Zhang, F. Zheng, N. Zhou. An image compression-encryption algorithm based on 2-D compressive sensing. *Journal of Computational Information Systems 9*(24), 10057–10064 (2013). | [7] F. Pedregosa, G. Varoquaux, A. Gramfort et al. Scikit-learn: Machine learning in Python, arXiv:1201.0490v4 (2012). | [8] J. S. Garafolo, L. F. Lamel, W. M. Fisher et al. TIMIT Acoustic-Phonetic Continuous Speech Corpus LDC93S1. Linguistic Data Consortium (1993). | [9] S. Y. Low, D. S. Pham, S. Venkatesh. Compressive speech enhancement. *Speech Communication 55*(6), 757–768 (2013).

Audio CS

Pure tones (1DCS)

A recorded guitar signal playing a single E_4 note (330 Hz) was compressively sampled with a 12.5% compression ratio in the DCT domain and was reconstructed using LASSO [7]. Figure 4 shows the frequency content of the original and reconstructed signals, and also shows that CS can recover frequencies beyond the Nyquist rate.

Speech (*N*DCS)

A speech audio file was obtained from the TIMIT Corpus [8] and was transformed into the spectrogram domain. LASSO reconstruction was performed on each time segment, and the full reconstruction was obtained by constant-overlap-and-add (COLA) method (Fig. 5). For a perceptually accurate metric of the reconstruction quality, the Perceptual Evaluation of Speech Quality (PESQ) was used [9]. A PESQ value of 3.0 and above indicates acceptable quality. Additionally, the segmental SNR (SNR_{seg}) was also used. These two metrics were evaluated for varying combinations of compression ratios and number of subbands/sampling windows (Fig. 6).

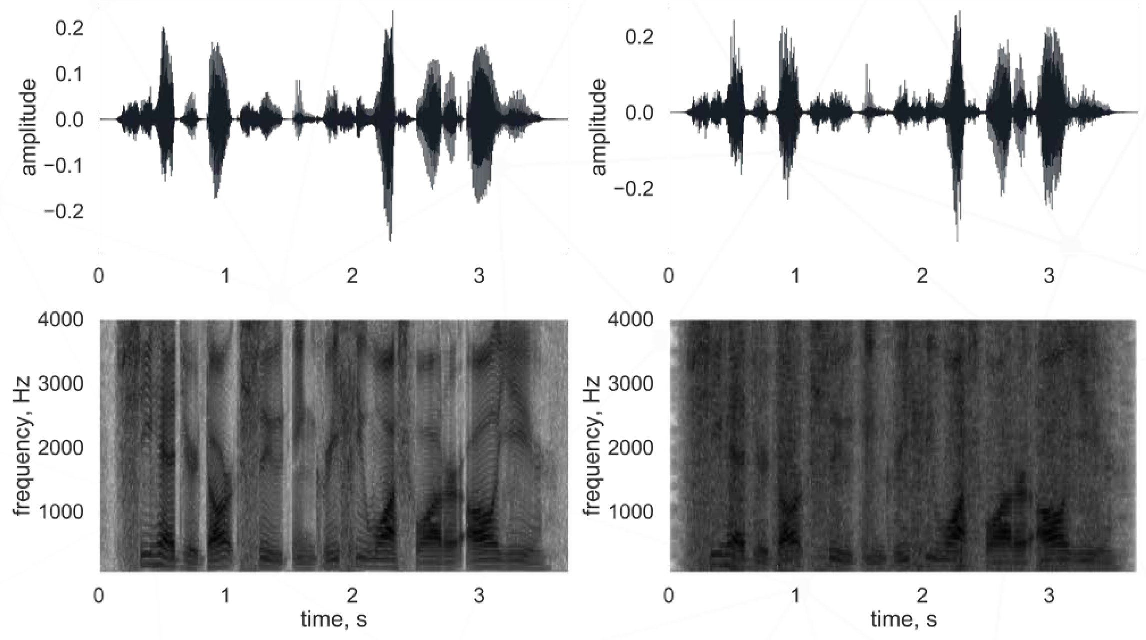

Figure 5: Test speech signal obtained from the TIMIT Corpus, with the original (left) and reconstructed (right) signals. The top row shows the temporal representation, while the bottom row shows the spectrogram representation (25ms window, 75% overlap).

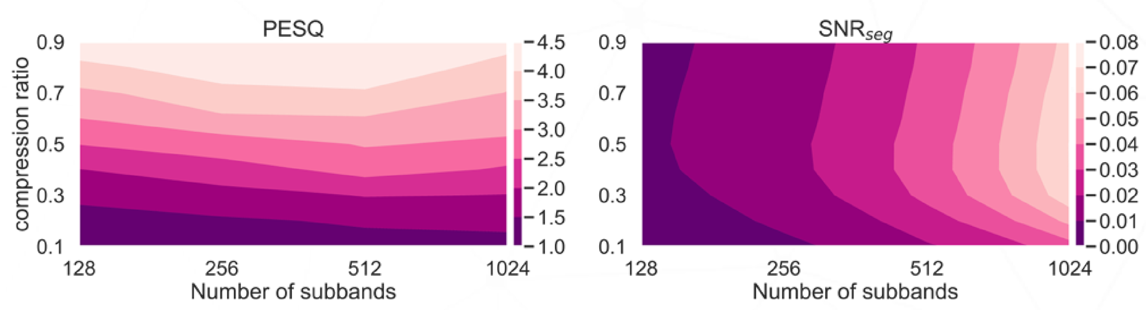

Figure 6: Most commonly used metrics in evaluating speech recording quality. PESQ (left) is sensitive to signal compression, and is more perceptually accurate. SNR_{seg} (right) is sensitive to the number of subbands.

Conclusions

I showed the potential applications of compressive sensing on arbitrary signals, including but not limited to compression, encryption, and enhancement. These techniques and algorithms can easily be extended to high-dimensional signals, such as color videos and hyperspectral images. Additionally, by error space mapping, I determined that SSIM & PESQ are appropriate, perceptually accurate metrics for image and audio signals, respectively.