Compression-based Acquisition of Structured Signals

Sajjad Beygi
University of Southern California

Shirin Jalali
Nokia - Bell Labs

Arian Maleki
Columbia University

Urbashi Mitra
University of Southern California

I. Introduction

The main goal of many compressed sensing (CS) algorithms is to recover an unknown high-dimensional target signal $\mathbf{x} \in \mathbb{R}^n$ from its noisy undersampled linear measurements $\mathbf{y} = A\mathbf{x} + \mathbf{z}$, where $A \in \mathbb{R}^{m\times n}$ and $\mathbf{z} \in \mathbb{R}^m$ denote the measurement matrix and the measurement noise, respectively. Despite the improvement that CS algorithms offer compared to conventional sensing methods [1]–[3], their main shortcoming is the discrepancy between the (true) complex structures present in many signals, such as natural images and videos, versus the relatively simple structures that are employed by classical algorithms, such as total variation minimization [4]–[8]. Bridging this gap can potentially lead to considerable gain in the performance of recovery algorithms in terms of their required number of measurements or reconstruction quality.

In this paper, we present a novel, complimentary approach to design generic recovery algorithms that exploit complex structures. Consider a class of signals $\mathcal{Q} \subset \mathbb{R}^n$, e.g. the class of natural images or videos, and suppose that there exists an efficient compression code for signals in $\mathcal{Q}$, e.g. JPEG2000 or MPEG4, respectively. Existence of such a compression code for a set $\mathcal{Q}$ suggests that there exists a shared *structure* between signals in this set. Such structures are typically much more complex than those exploited by CS algorithms. Therefore, the required number of measurements of a CS recovery algorithm that instead of simple structures employs such complex ones is reduced by a great factor.

This leads us to the following question: Can we efficiently use an existing compression code to build an efficient CS recovery algorithm? In other words, can we build an efficient CS recovery algorithm that takes advantage of signals' complex structures via using efficient compression codes? In response to this question, we propose a novel algorithm, namely, compression-based gradient descent (C-GD), which is an iterative algorithm that employs a compression code to solve a CS problem. Our empirical and theoretical results demonstrate the effectiveness of C-GD algorithm. C-GD can be considered as an efficient method to approximate the solution of the highly non-convex compressible signal pursuit algorithm proposed in [9].

II. Compression-based Gradient Descent

Consider a compact set $\mathcal{Q} \subset \mathbb{R}^n$ and a rate-r lossy compression code for set $\mathcal{Q}$, described by the encoding and decoding mapping pair (f_r, g_r). Given measurements $\mathbf{y} = A\mathbf{x} + \mathbf{z}$, where $\mathbf{x} \in \mathcal{Q}$ and $A \in \mathbb{R}^{m\times n}$, C-GD, iteratively, employs the following update rule:

$$\mathbf{x}^{k+1} \leftarrow g_r\left(f_r\left(\mathbf{x}^k + \alpha_k\, A^T(\mathbf{y} - A\mathbf{x}^k)\right)\right), \tag{1}$$

where $\alpha_k > 0$ is a step-size that ensures the convergence of the algorithm. The intuition of the algorithm is simple; by computing $\mathbf{x}^k + \alpha_k\, A^T(\mathbf{y} - A\mathbf{x}^k)$, C-GD moves its estimate toward the subspace $\{\mathbf{x} \in \mathbb{R}^n \;:\; \mathbf{y} = A\mathbf{x}\}$. Then, performing $g_r(f_r(\cdot))$ maps the obtained estimate on a discretized version (codebook) of $\mathcal{Q}$, which contains the set of compressible signals. We refer to this algorithm as *compression-based gradient descent* (C-GD), and prove that, given enough number of measurements, it always converges, even in the presence of measurement noise. For example, Theorem 1 shows that, with high probability, C-GD converges to a good estimate of $\mathbf{x}$, if m is a constant factor times nr, where r denotes the per-symbol rate of the compression code. After Theorem 1 we give a concrete example that clarifies the required number of measurements of C-GD.

Theorem 1. *Let δ denote the supremum distortion of code (f_r, g_r) over set $\mathcal{Q}$. Let $A \in \mathbb{R}^{m\times n}$ be a random Gaussian measurement matrix with i.i.d $\mathcal{N}(0, \sigma_a^2/n)$ entries, and $\mathbf{z} \in \mathbb{R}^m$ be an i.i.d. $\mathcal{N}(0, \sigma_z^2)$ Gaussian noise vector. Let $\alpha = \frac{1}{m\sigma_a^2}$ and $\tilde{\mathbf{x}} = g_r(f_r(\mathbf{x}))$. Then, given $\epsilon > 0$ and $m \geq 80nr(1+\epsilon)$, with a probability larger than $1 - 2^{-2\epsilon nr+1}$, for $k = 0, 1, 2, \ldots$,*

$$\frac{1}{\sqrt{n}}\|\mathbf{x}^{k+1} - \tilde{\mathbf{x}}\|_2 \leq \frac{0.9}{\sqrt{n}}\|\mathbf{x}^k - \tilde{\mathbf{x}}\|_2 + 2\left(2 + \sqrt{\frac{n}{m}}\right)^2 \frac{\delta}{\sqrt{n}} + \frac{\sigma_z}{\sigma_a}\sqrt{\frac{8(1+\epsilon)nr}{m}}.$$

Note that $\lim_{k\to\infty} \frac{1}{\sqrt{n}}\|\mathbf{x}^{k+1} - \tilde{\mathbf{x}}\|_2 \leq 20\left(2 + \sqrt{\frac{n}{m}}\right)^2 \frac{\delta}{\sqrt{n}} + 10\frac{\sigma_z}{\sigma_a}\sqrt{\frac{8(1+\epsilon)nr}{m}}$. If we assume that the distortion of the code is low, then the only remaining error will be due to the noise. However, small values of δ often correspond to large values of r. Hence, the smaller the first error term is, the larger the second error term will be.

To connect this result with more classical results in CS, we consider a very simple example. Let $\mathcal{Q} \subset \mathbb{R}^n$ denote the set of k-sparse signals that lie inside the Euclidean unit ball. A simple compression code for $\mathbf{x} \in \mathcal{Q}$ consists of describing i) the sparsity pattern of $\mathbf{x}$ using $\approx \log_2 \binom{n}{k}$ bits, and ii) the quantized value of the non-zero elements of $\mathbf{x}$. To achieve supremum distortion δ, this code requires a rate r, where

$$r \leq \frac{1}{n}\log_2\binom{n}{k} + \frac{k}{n}\log_2\left(\frac{\sqrt{k}}{\delta}\right) + c\frac{k}{n},$$

Let $\delta = \frac{1}{n}$. Then, from Theorem 1, by using this compression code inside C-GD algorithm, if the number of measurements satisfies $m > 80k\log n$, with high probability, at each step of C-GD,

$$\frac{1}{\sqrt{n}}\|\mathbf{x}^{k+1} - \tilde{\mathbf{x}}\|_2 \leq \frac{0.9}{\sqrt{n}}\|\mathbf{x}^k - \tilde{\mathbf{x}}\|_2 + 2\left(2 + \sqrt{\frac{n}{m}}\right)^2 \frac{1}{n\sqrt{n}} + \frac{\sigma_z}{\sigma_a}\sqrt{\frac{8(1+\epsilon)k\log n}{m}}.$$

Note that both the number of measurements and the sensitivity to the Gaussian noise can be compared with the similar results in [10] and are nearly optimal.

III. Numerical Results

In our numerical results, we study CS of test images for both i.i.d. Gaussian and random partial Fourier measurement matrices. We consider two different noise levels, $\mathrm{SNR} = 10\mathrm{dB}$ and $30\mathrm{dB}$, and two different sampling rates. Tables I and II report the peak signal-to-noise ratios (PSNR) of recoveries based on i) state-of the art compressive imaging algorithms and ii) C-GD employing JPEG2000.

TABLE I: The PSNR performance of C-GD algorithm that employs JPEG2000 compression code is compared with the state-of-the-art NLR-CS algorithm [11]. 512×512 test images are sampled by a random partial-DCT measurement matrix and distorted by Gaussian noise with various SNR values.

		House		Barbara		Shore	
Method	Sampling ratio	SNR=10	SNR=30	SNR=10	SNR=30	SNR=10	SNR=30
JP2K-CG	10%	**17.33**	**24.40**	**16.53**	18.65	**16.40**	**24.03**
	30%	**19.56**	**35.38**	**18.82**	**26.19**	**19.18**	**35.39**
NLR	10%	11.66	24.14	12.10	**19.83**	10.83	22.20
	30%	12.6	26.84	13.32	23.05	11.76	24.98

TABLE II: The PSNR performance of C-GD algorithm that employs JPEG2000 compression code is compared with the state-of-the-art recovery algorithm BM3D-AMP [12] for Gaussian sensing matrices. 128×128 test images are sampled by a random Gaussian measurement matrix and distorted by Gaussian measurement noise with various SNR values. While BM3D-AMP often outperforms JPEG2000-GD, (i) as is clear from the table, when the noise is higher, the difference is much lower and in some cases (Shore image) JPEG2000-GD outperforms BM3D-AMP, (ii) BM3D-AMP does not work on partial-Fourier matrices that are used in many applications, (iii) by using better compression algorithms one can improve the performance of C-GD.

		House		Barbara		Shore	
Method	Sampling ratio	SNR=10	SNR=30	SNR=10	SNR=30	SNR=10	SNR=30
JP2K-CG	30%	21.86	26.52	21.42	24.23	22.80	24.15
	50%	24.82	28.67	22.78	27.13	**25.64**	29.64
BM3D-AMP	30%	**24.12**	**35.40**	**22.05**	**30.36**	**23.85**	**34.01**
	50%	**25.72**	**37.07**	**23.38**	**33.06**	24.81	**36.28**

References

[1] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly. Compressed sensing MRI. *IEEE Sig. Process. Mag.*, 25(2):72–82, 2008.

[2] R. G. Baraniuk and P. Steeghs. Compressive radar imaging. In *Proc. IEEE Radar Conf.*, pages 128–133, Apr. 2007.

[3] S. Beygi, U. Mitra, and E. G. Ström. Nested sparse approximation: Structured estimation of v2v channels using geometry-based stochastic channel model. *IEEE Trans. Signal Process.*, 63(18):4940–4955, 2015.

[4] Y. Kim, M. S. Nadar, and A. Bilgin. Compressed sensing using a gaussian scale mixtures model in wavelet domain. In *Proc. IEEE Int. Conf. Image Processing (ICIP)*, pages 3365–3368, September 2010.

[5] S. Som and P. Schniter. Compressive imaging using approximate message passing and a Markov-tree prior. *IEEE Trans. Signal Process.*, 60(7):3439–3448, 2012.

[6] G. Peyre, S. Bougleux, and L. Cohen. Non-local regularization of inverse problems. In D. Forsyth, P. Torr, and A. Zisserman, editors, *Computer Vision*, volume 5304 of *Lecture Notes in Computer Science*, pages 57–68. Springer Berlin Heidelberg, 2008.

[7] J. Zhang, S. Liu, and D. Zhao. Improved total variation based image compressive sensing recovery by nonlocal regularization. *IEEE Int. Symp. Cir. and Sys. (ISCAS)*, pages 2135–2138, March 2013.

[8] W. Dong, G. Shi, X. Li, Y. Ma, and F. Huang. Compressive sensing via nonlocal low-rank regularization. *IEEE Trans. Image Process.*, 2014.

[9] S. Jalali and A. Maleki. From compression to compressed sensing. *Appl. Comp. Harmonic Anal. (ACHA)*, 40(2):352–385, 2016.

[10] P. J. Bickel, Y. Ritov, and A. B. Tsybakov. Simultaneous analysis of LASSO and Dantzig selector. *Annals of Stat.*, pages 1705–1732, 2009.

[11] W. Dong, G. Shi, X. Li, Y. Ma, and F. Huang. Compressive sensing via nonlocal low-rank regularization. *IEEE Trans. Image Process.*, 23(8):3618–3632, 2014.

[12] C. A. Metzler, A. Maleki, and R. G. Baraniuk. From denoising to compressed sensing. *IEEE Trans. Inf. Theory*, 62(9):5117–5144, Sep. 2016.