Inflating Compressed Samaples: A Joint Source-Channel Coding Approach for noise-Resistant Compressed Sensing

A. Hesam Mohseni, Massoud Babaie-Zadeh, Christian Jutten

INFLATING COMPRESSED SAMPLES: A JOINT SOURCE-CHANNEL CODING APPROACH FOR NOISE-RESISTANT COMPRESSED SENSING

*A. HesamMohseni, M. Babaie-Zadeh**

Department of Electrical Engineering, Sharif
University of Technology, Tehran, Iran
hesammohseni@gmail.com
mbzadeh@sharif.edu

C. Jutten

GIPSA-LAB
Grenoble, France
christian.jutten@inpg.fr

ABSTRACT

Recently, a lot of research has been done on compressed sensing, capturing compressible signals using random linear projections to a space of radically lower dimension than the ambient dimension of the signal. The main impetus of this is that the radically dimension-lowering linear projection step can be done totally in analog hardware, in some cases even in constant time, to avoid the bottleneck in sensing and quantization steps where a large number of samples need to be sensed and quantized in short order, mandating the use of a large number of fast expensive sensors and A/D converters. Reconstruction algorithms from these projections have been found that come within distortion levels comparable to the state of the art in lossy compression algorithms. This paper considers a variation on compressed sensing that makes it resistant to spiky noise. This is achieved by an analog real-field error-correction coding step. It results in a small asymptotic overhead in the number of samples, but makes exact reconstruction under spiky measurement noise, one type of which is the salt and pepper noise in imaging devices, possible. Simulations are performed that corroborate our claim and in fact substantially improve reconstruction under unreliable sensing characteristics and are stable even under small perturbations with Gaussian noise.

Index Terms— Compressed Sensing, Joint Source-channel coding, Sparse solution problems, Natural Images, Error-Correcting codes

1. INTRODUCTION

Compressed sensing has generated a lot of research activity in recent years. This is chiefly motivated by the desire to build cheaper imaging devices exploiting the way digital images are typically stored and manipulated, a lossy compressed representation containing only the most important details, and the inefficiency of current image sensor architectures that require a number of light-intensity measurements proportional to the pixel size of the image, rather than its inherent information content or the desired level of detail.

The main breakthrough came from the works of [1, 2] showing that the minimum ℓ_0-norm solution to most underdetermined linear systems coincides with their minimum ℓ_1-norm solution, efficiently solvable by virtue of ℓ_1-norm convexity. Wakin et al. in [3] demonstrated the real-world applicability of this concept by making

a so-called *single-pixel camera* using a Digital Micromirror Device (DMD) chip to produce a series of linear combinations of light intensities from across the desired scene and reconstructing it by finding the minimum ℓ_1-norm solution of the resulting linear system in wavelet domain.

Interest in this and other applications has resulted in a number of fast mathematical programming and iterative algorithms for finding sparse solutions to underdetermined linear systems (see [4, 5, 6] and references therein). In this paper we propose and analyze an enhanced method of compressed sensing that while preserving the hardware implementability property of [3] provides for a tunable measure of immunity to spike noise in image sensor.

The basic idea of this method is to make use, in addition to the compressibility of natural images, of spiky nature of noise introduced into measurements by precoding the compressed samples using a suitable encoding matrix, allowing the circuitry on the digital side to do an additional denoising step in post processing and therefore compensate to some degree for the unsatisfactory noise characteristics. This is specially relevant to image acquisition as one of the common forms of noise corrupting images is the so called salt and pepper noise that has a spiky nature and typically is a result of defective pixels in image sensors [7]. In ordinary imaging architecture, these errors are hard to correct for in hardware because the functionality of a pixel is dependent on its location [7], but our setup promises an approach that can sidestep this issue completely.

The outline of the paper is as follows. Section 2 expounds on some background knowledge, introducing our notation and a step by step description of the proposed setup. We then clarify the relationships of our method both to coding theory, specially real and complex-field linear codes, and to coded aperture methods of image acquisition used in astronomical imaging. The steps involved will be explained in complete detail, drawing attention to some important points. Section 3 reports on the numerical results obtained for our experimental setup. Various parameters for describing the problem are defined and the performance of algorithm under their changes is investigated with a series of plots showing how our algorithm compares with ordinary compressed sensing under the presence of spiky noise. Finally, section 4 concludes the discussion and shows direction for future research.

2. BACKGROUND AND NOTATIONAL CONVENTIONS

The classical compressed sensing scenario is as follows:

Given a natural signal $\boldsymbol{x} \in \mathbb{R}^N$ and an orthonormal basis $\boldsymbol{\Phi} \in \mathbb{R}_{N \times N}$ that is sparsifying for $\boldsymbol{x}$, i.e. $\boldsymbol{x} = \boldsymbol{\Phi}\boldsymbol{\theta}$ and $\boldsymbol{\theta}$ is semi-sparse with at most S significant coefficients. We multiply $\boldsymbol{x}$ by a random

matrix $D \in \mathbb{R}_{n \times N}$ where $n \ll N$ to obtain $y = Dx = D\Phi\theta \in \mathbb{R}^n$. We recover an approximation $\hat{x}$ to x by finding the sparsest θ' for which $D\Phi\theta' = y$, calling it $\hat{\theta}$ and calculating $\hat{x}$ as $\Phi\hat{\theta}$.

Almost all recent theoretical advances on the above problem amount to proving some variation of these results: the ℓ_1-norm minimizer obeying the constraints of the above problem coincides with the sparsest solution (subject to some conditions on matrix D and S). Also, subject to stronger conditions on D and S, even if θ is not exactly S-sparse, the distance between ℓ_1-norm minimizer and its best S-term approximation is proportional to the energy in the non-sparse tail of θ [2]. Furthermore, if the observations are corrupted with an error vector e whose length is upper bounded by ϵ, the distance between the ℓ_1-norm minimizer θ' subject to the inequality constraint $\left\| y - D\Phi\theta' \right\|_2 \leq \epsilon$ and the best S-term approximation to θ is increased at most proportional to ϵ compared to the noiseless case. These sufficient conditions on D and S are true with high probability when D is selected from various random matrix ensembles and S is sufficiently small [8]. we adopt the language of [6] and call the equality constrained problem and its inequality constrained counterpart Basis Pursuit (BP) and Basis Pursuit Denoise (BPDN), respectively.

Another corollary of these results that has attracted theoretical [9], but not practical interest is this: Assume $H \in \mathbb{R}_{k \times K}$ is a matrix satisfying the same kind of conditions as D for d-sparse signals (in our work, k and K are generally much smaller than n and N respectively with K being of the same order of magnitude as n but strictly larger). Assume also that $G \in \mathbb{R}_{K \times n}$ is a matrix that is left-annihilated by H ($HG = 0$) and defines an injective mapping from $\mathbb{R}^n$ to $\mathbb{R}^K$ (has full column rank). This G is the coding matrix of a linear code of dimension n and length K that can correct up to d gross errors and H is its corresponding parity check matrix. The decoding process is identical to that of binary linear codes. We first make a message-independent syndrome s by multiplying H with the error-infested message r, then find the sparsest error pattern consistent with it by solving the system $He = Hr = H(c + e) = H(Gy + e)$ for its sparsest solution e (here, y and r denote the uncoded and coded message respectively and e is the error vector). Having e at hand allows for finding the coded message c by simply calculating $c = r - e$. Injectivity of G assures that having c is equivalent to having y (for example by calculating $G^\dagger y$ where $G^\dagger$ denotes the pseudoinverse of G). Let us clarify our matrix and vector naming scheme. D is so named because it **D**eflates (compresses) the natural signal x. H and G are conventional names, respectively for parity check and generator matrices in coding theory. c is the coded message and r is the received codeword.

Note that repetition of n and y in this discussion is deliberate and that our aim is in fact to encode the compressed samples y with G, in an approach not dissimilar to joint source-channel coding. The important difference is that in this setup both source and channel coding are to be done in analog hardware. This means complex and nonlinear signal-dependent compression algorithms are not available and we must rely on theoretical guarantees of compressed sensing to ensure a reasonable compression ratio.

A high-level description of the steps involved is as follows:

1. Multiply the signal with a known sparsifying basis with a suitable *wide* random matrix $D \in \mathbb{R}_{n \times N}$ to obtain $y = Dx$ (analog hardware).

2. Multiply y by a *tall* encoding matrix $G \in \mathbb{R}_{K \times n}$ to obtain $c = Gy$ (analog hardware).

3. Sense and quantize c to obtain $r = c + e$ where e has a small number of large components (noise spikes) interspersed by small values of thermal noise (Analog/Digital interface).

4. Multiply r by the *wide* $H \in \mathbb{R}_{k \times K}$ to obtain the syndrome and use $He = Hr = s$ to solve for the sparsest e, call it $\hat{e}$ (digital hardware).

5. Calculate $\hat{c} = r - \hat{e}$ (digital hardware).

6. Calculate $\hat{x}$ by first finding the sparsest solution $\hat{\theta}$ to $\hat{c} = GD\Phi\theta$ and from it $\hat{x} = \Phi\hat{\theta}$ (digital hardware).

Some notes are in order:

Steps 1 and 2 can and should in fact be combined in a joint source-channel coding $K \times N$ operator $J \triangleq GD$, because having c, we don't need y by injectivity of G. Calculating y on the analog side is not needed and on the digital side we only need to know the decomposition of J into GD for which H is a left-annihilator of G. We construct H by drawing its entries from a suitable random matrix family. Any n linearly independent vectors from the null space of H can form the columns of G. In our simulations, we use the first n columns from the output of the MATLAB function `null(H)` that calculates an orthonormal basis in matrix form for the kernel of its input parameter.

That θ is not exactly sparse and e has a large number of small components of a thermal noise nature will not lead to numerical blowup, as their finding involves solving inverse *sparse approximation problems*, which are stable as discussed earlier. For typical values of problem parameters, this implies that the support of desired vector (location of its significant coefficients) and the associated coefficient values are accurately identified and calculated.

The $K \times N$ matrix $J = GD$ is rank-deficient as its rank cannot be larger than $n < K$. Note that the row nullity of J (number of rows minus the row rank) is at least $K - n$ (with hight probability row nullity is equal to $K - n$), this measures the redundancy in a system of equations with coefficient matrix J. Note that for an injective G and its annihilator H to exist we should have $K \geq k + n$ by a simple dimensional argument, therefore $K - n \geq k$. Combined with the fact that for a random H to recover d-sparse signals, k should be on the order of $d \log (K/d)$ [8], this reaffirms k as a measure of the amount of redundancy added for gross error correction.

Our description of the setup elaborated on the similarities to real-field coding when discussing the choice of notation, we also note one difference here: typically in real-field coding the encoding matrix is chosen to satisfy some desirable properties or to be similar with some coding scheme over finite fields, e.g. real field DFT codes [10], these design constraints most often manifest themselves in terms of restrictions on the structure of coding matrix. The method proposed here, to the contrary, constrains the design of parity check matrix (by restricting it to be suitable for compressed sensing) and finds some coding matrix among many that is annihilated by that specific parity check matrix. This has some similarities to the way sparse graph based codes like Low Density Parity Check (LDPC) Codes [11] are constructed and perhaps interestingly, these constructions also rely on randomness to obtain asymptotically optimal properties.

There are also similarities with coded aperture methods in astronomical imaging that use a masking pattern between the scene (region of the sky we are interested in studying) and a sensor array to detect a number of photon sources, each point source of photons projects the mask pattern onto the sensor array consistent with its own spacial location and therefore if the number of sensors is large enough we are able to localize and detect the intensity of all sources in the region of interest. Of particular interest to this work are some coded aperture reconstruction methods that try to exploit sparsity in
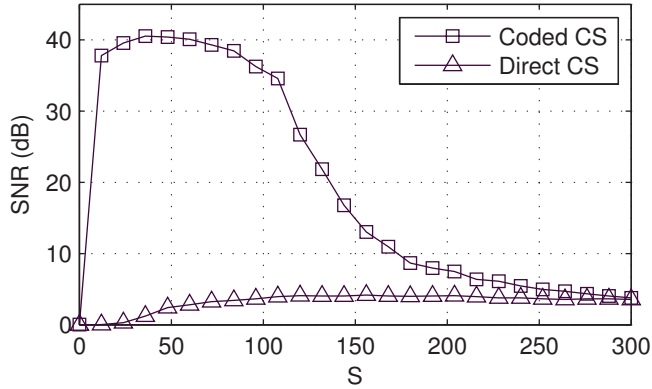
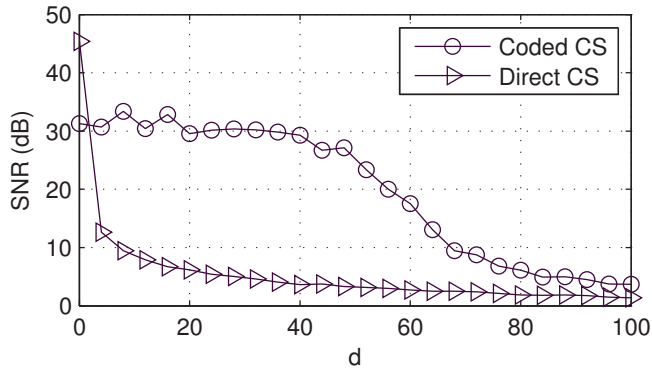**Fig. 1**. SNR versus S ($N = 1000$, $n = 400$, $K = 600$, $k = 200$ and $d = 40$).

Fig. 2. SNR versus d ($N = 1000$, $n = 400$, $K = 600$, $k = 200$ and $S = 120$).

Fig. 3. Reconstruction with coded CS ($N = 25$, $n = 10$, $K = 15$, $k = 5$, $S = 5$ and $d = 1$).

Fig. 4. Reconstruction with direct CS ($N = 25$, $n = 10$, $K = 15$, $k = 5$, $S = 5$ and $d = 1$).

detector function imperfections (spiky errors) in addition to scene sparsity (e.g. [12]). There are important differences though, the unusual design of coded aperture systems is at least as much affected by the difficulty of building lenses for wavelengths of interest as by the sparsity of the desired signal and the detector failure pattern, if not more.

3. SIMULATION RESULTS

This section reports on numerical results obtained from the proposed scheme. We select S indices at random from $\{1 \cdots n\}$ to be from $\mathcal{N}(0, \sigma_{\text{on}}^2)$ (representing active signal coefficients) and the rest from $\mathcal{N}(0, \sigma_{\text{off}}^2)$ ($\sigma_{\text{off}}^2 \ll \sigma_{\text{on}}^2$). The same method is used to generate the spiky noise where d elements of $\{1 \cdots K\}$ are chosen at random to be from $\mathcal{N}(0, \sigma_{\text{spike}}^2)$ (representing noise spike locations) and the rest to be from $\mathcal{N}(0, \sigma_{\text{therm}}^2)$ ($\sigma_{\text{therm}}^2 \ll \sigma_{\text{spike}}^2$). $\boldsymbol{D}$ and $\boldsymbol{H}$ are selected from the scaled Gaussian ensembles $\mathcal{N}_{n \times N}/\sqrt{n}$ and $\mathcal{N}_{k \times K}/\sqrt{k}$ respectively (scaled i.i.d standard normal entries).

To compare our results with ordinary compressed sensing, each test on a joint source-channel coding matrix $\boldsymbol{J} \in \mathbb{R}^{K \times N}$ is accompanied by a test on another random matrix $\boldsymbol{A} \in \mathbb{R}^{K \times N}$ drawn from $\mathcal{N}_{K \times N}/\sqrt{K}$, which we term the direct compressed sensing matrix (direct CS). Our own inflation approach is termed coded compressed
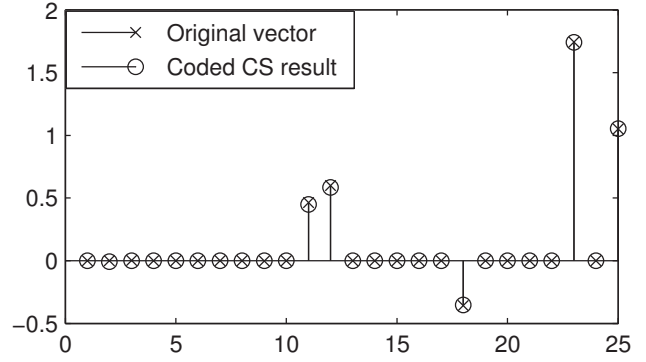
sensing (coded CS). The same spiky noise is added to both measurements. Reconstruction is based on the SPGL1 algorithm [6] as it seems to be the fastest implementation of Basis Pursuit for systems whose answer is not completely sparse. The reported SNRs were averaged over 50 trials of each experiment to smooth out random deviations.

In coded CS, error reconstruction from syndrome is done using the noiseless BP. BPDN is used in reconstructing the signal from error-subtracted samples as the error estimation step is not guaranteed to be completely accurate. BPDN is also used in direct compressed sensing setup. The ℓ_2-norm constraint value in these BPDN problems was chosen empirically to be 1.3 times the expected norm of the relevant error. To apply this for error-subtracted samples in the inflated setup, we needed to estimate the residual part of noise not recovered by the first BP, for which we used an empirical estimate of BP performance in recovering semi-sparse vectors.

Figures 1 and 2 show the performance of the algorithm versus S and d ($N = 1000$, $n = 400$, $K = 600$, $k = 200$, $\sigma_{\text{spike}}^2/\sigma_{\text{therm}}^2 = \sigma_{\text{on}}^2/\sigma_{\text{off}}^2 = 1000$). Note how in fig. 1 we have zero SNR for the first data point ($S = 0$). For $S \leq 100$, coded CS performance is quite impressive and after that it decreases smoothly for larger S, asymptotically reaching direct CS but always performing better. Note that the performance of direct CS shows some improvement when S increases! This might be related to the less sparse signals being some-

what better at masking the effects of spiky noise. Figure 2 shows the effect of changing d on performance. Note the extreme sensitivity of direct CS to spiky noise (a common form of noise in image sensors) by comparing its performance for $d = 0$ and $d = 4$ (first and second data points). Coded CS outperforms direct CS as expected, except for the degenerate case of $d = 0$. Its performance smoothly decreases for larger d's, again asymptotically reaching direct CS but never becoming worse.

Figures 3 and 4 compare performance of coded and direct CS on an actual low-dimensional problem instance ($N = 25$, $n = 10$, $K = 15$, $k = 10$, $S = 5$, $d = 1$, $\sigma^2_{\text{spike}}/\sigma^2_{\text{therm}} = \sigma^2_{\text{on}}/\sigma^2_{\text{off}} = 1000$). We see that our coded CS approach has correctly identified the location and magnitude of active coefficients whereas direct CS has failed to detect three of the five active components correctly under only a single noise spike.

We repeated above experiments using SL0 algorithm [5] and the results have been essentially the same. We don't reproduce these results here for reasons of space. SL0 may be advantageous from a computational standpoint. Unlike SPGL1 it requires computation or storage of measurement matrix pseudoinverse but typically converges in fewer iterations and requires fewer matrix vector products.

4. CONCLUSION

An extension to the classical compressed sensing scenario was proposed and connections with related ideas in coding and astronomical imaging were considered. Simulations showed that the proposed idea is promising when spiky errors are present in the measurement process, even when they are not the only source of error and that numerical stability is not an issue in reconstruction for moderate values of parameters. We are also experimenting with partial Fourier transform matrices for sensing and parity check as they are more amenable to analog hardware implementation. Also, the approach to redundant measurements proposed here may be of use for reducing other types of noise. Our primary results show that even for completely dense noise, coded compressed sensing provides a $3dB$ improvement in SNR. This is not as impressive as the results for sparse noise, but may be useful as it can be implemented in analog optical domain and used to *learn* the thermal noise of sensors and other electronic components.

5. REFERENCES

[1] D. L Donoho, "For most large underdetermined systems of equations, the minimal l1-norm near-solution approximates the sparsest near-solution," *Communications on Pure and Applied Mathematics*, vol. 59, pp. 907–934, 2004.

[2] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, pp. 1207–1223, 2006.

[3] M.B. Wakin, J.N. Laska, M.F. Duarte, D. Baron, S. Sarvothamand, D. Takhar, K.F. Kelly, and R.G. Baraniuk, "An architecture for compressive imaging," in *Proceedings of IEEE International Conference on Image Processing (ICIP)*, 2006, pp. 1273–1276.

[4] D. Needell and J.A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *accepted for publication in Applied and Computational Harmonic Analysis*, 2008, available as eprint arXiv: 0803.2392.

[5] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed ℓ^0 norm," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 289–301, January 2009.

[6] E. van den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *To appear in SIAM Journal on Scientific Computing*, 2008, available at `http://www.cs.ubc.ca/~mpf/downloads/BergFriedlander08.pdf`.

[7] K. N. Salama and A. Al-Yamani, "Analysis of self-correcting active pixel sensors," in *Image Quality and System Performance II. Proceedings of SPIE*, 2005, vol. 5668, pp. 262–269.

[8] R. Baraniuk, M. Davenport, R. Devore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, 2008.

[9] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, pp. 4203–4215, 2005.

[10] F. Marvasti, M. Hasan, M. Echhart, and S. Talebi, "Efficient algorithms for burst error recovery using FFT and other transform kernels," *IEEE Transactions on Signal Processing*, vol. 47, pp. 1065–1075, 1999.

[11] A. Shokrollahi, "LDPC codes: An introduction," in *Proceedings of Workshop on Coding, Cryptography and Combinatorics*, 2003.

[12] G. K. Skinner and M. R. Nottingham, "Analysis of data from coded-mask telescopes by maximum likelihood." *Nuclear Instruments and Methods in Physics Research*, vol. 333, pp. 540–547, Sept. 1993.