Bayesian Compressed Sensing with Heterogeneous Side Information

Evangelos Zimos∗, João F. C. Mota†, Miguel R. D. Rodrigues†, and Nikos Deligiannis∗

∗Vrije Universiteit Brussel - iMinds
Dept. Electronics and Informatics
Pleinlaan 2, 1050
Brussels, Belgium
ezimos@etro.vub.ac.be
ndeligia@etro.vub.ac.be

†University College London
Dept. Electronic and Electrical Engineering
Torrington Place
London WC1E 7JE, UK
j.mota@ucl.ac.uk
m.rodrigues@ucl.ac.uk

Abstract

The classical compressed sensing (CS) paradigm can be modified so as to leverage a signal correlated to the signal of interest, called side information, which is assumed to be provided a priori at the decoder in order to aid reconstruction. In this work, we propose a novel CS reconstruction method based on belief propagation principles, which manages to exploit side information generated from a diverse (or heterogeneous) data source by using the statistical model of copula functions. Through simulations, we demonstrate that the proposed method yields significant reduction in the mean-squared error of the reconstructed signal as compared to state-of-the-art methods in classical compressed sensing and compressed sensing with side information.

Introduction

Compressed sensing (CS) is a data acquisition paradigm that emerged as a new research area in signal processing during the last decade. The main contribution of the CS theory is that a plethora of signals, namely sparse, approximately sparse or compressible signals, can be recovered using fewer measurements than dictated by the Shannon-Nyquist theorem. As many signals are compressible, CS finds applications in several domains, such as magnetic resonance imaging (MRI) [1], tomographic imaging [2], as well as storage system architectures and wireless sensor networks [3].

In the classical CS framework, the signal of interest $\underline{x} \in \mathbb{R}^N$ can be written in the form $\underline{x} = \mathbf{\Psi}\underline{s}$, where $\underline{s}$ is its K-sparse representation (i.e., $\|\underline{s}\|_0 = K$) and $\mathbf{\Psi} \in \mathbb{R}^{N \times N_0}$ is an orthonormal or overcomplete basis, called dictionary. Let $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ be another matrix, called sensing (or encoding) matrix, such that the measurement matrix $\mathbf{A} = \mathbf{\Phi}\mathbf{\Psi}$ satisfies either the mutual coherence property [4], the Restricted Isometry Property [5] or the Null Space Property [6]. Then, the CS theory states that $\underline{x}$ can be recovered using the measurement matrix $\mathbf{A}$ and $M \ll N$ linear random measurements $\underline{y} = \mathbf{\Phi}\underline{x} = \mathbf{A}\underline{s}$. If the number of measurements is sufficiently large, then $\underline{s}$ is the unique minimizer of the following optimization problem, known as Basis Pursuit [7]:

$$\hat{\underline{s}} = \arg\min_{\underline{s}} \|\underline{s}\|_1 \quad \text{s.t.} \quad \underline{y} = \mathbf{A}\underline{s}.$$

Instead of assuming that $\underline{s}$ is strictly sparse (i.e., $\|\underline{s}\|_0 = K$), several works [8] (including this) focus on approximately K-sparse signals, i.e., signals that have only K large coefficients while the rest are small but not necessarily zero.

Motivated by many real-life applications such as medical imaging [9], hyperspectral imaging [10] and compressive video sensing [11], the classical CS framework can be modified so as to leverage a signal correlated to the signal of interest, called side information (SI), which is assumed to be provided *a priori* at the decoder in order to aid reconstruction. In the theory of CS with SI, the decoder aims at recovering the signal $\underline{x}$ based on the measurements $\underline{y}$, the measurement matrix $\mathbf{A}$ and a SI vector, say $\underline{w}$, which is correlated with $\underline{s}$. There have been a few approaches that attempt to exploit SI in CS. The CS problem where the support of the signal of interest is partly known was studied in [12]. Although sufficient conditions for exact reconstruction and error bounds were derived, the performance of the algorithm was not fully characterized. A different scheme that lacks theoretical analysis was proposed in [13, 14], where the authors made the assumption that the prediction error between the SI and the signal is sparser than the signal itself. Hence, the estimation error was recovered instead of the sparse signal. The problem of CS with prior information was studied in [15–17], where, through bounds and geometrical interpretations, it was shown that ℓ_1-ℓ_1 minimization improves the performance of CS if the SI is of good enough quality.

When a joint statistical characterization between the signal of interest and the SI signal is available, Bayesian inference can provide another CS recovery approach. A Bayesian CS framework that was not considering SI was proposed in [18], where relevance vector machines were used for signal estimation. Furthermore, the authors in [8] performed asymptotically optimal Bayesian inference using LDPC-like codes for signal reconstruction. Regarding schemes that incorporate SI, in [19] the authors developed an SI-aided approximate message passing (SI-AMP) algorithm, where the asymptotic prediction performance and noise-sensitivity analysis of the scheme were derived. Furthermore, the work in [8] has been extended to distributed compressed video sensing setups [20] and multi-view image acquisition [21], where Gaussian models were used to describe the correlation between the sparse signal and the SI in a belief-propagation context. All these works assume that both the signal of interest and the SI are jointly sparse and they describe their underlying statistical dependence using (*i*) models with common sparse supports [22] or sparse common components [22, 23], (*ii*) simple additive models [20, 21], or (*iii*) joint Gaussian Mixture Models (GMM) [24]. These models assume that the signal of interest and the SI are produced from *homogeneous* sources, namely, sources that produce signals of the same type and similar statistical descriptions are used to model them. However, the SI available in many applications [25] is *heterogeneous* and comes from different modalities. For instance, an MRI image can be reconstructed using as an aid a CT or a PET scan [26].

Contributions

This work answers to the following questions:

- How can one port SI from a heterogeneous correlated source into a CS recon-

struction framework?

- How can we leverage the potential correlation among the *sparse* (or *approximately sparse*) signal of interest and a *non-sparse* correlated SI signal in a CS framework?

We propose a novel CS reconstruction method that exploits a SI signal generated from diverse (or *heterogeneous*) sources. Our method uses copula functions [27, 28] to capture the correlation between the different sources, which can have different statistical characterizations. For example, the entries of $\underline{x}$ can be modelled as Laplacian and the entries of $\underline{w}$ as Poisson, Gaussian, or also Laplacian. This implies that SI can aid the reconstruction even if it is not sparse. Through experimentation, we show that the proposed method outperforms the $\ell_1 - \ell_1$ algorithm in [15] offering mean-squared-error (MSE) reductions of up to 32.73% compared to the case where $\underline{x}$ and $\underline{w}$ are both Laplace distributed. Also, we show that the proposed method can efficiently exploit diverse SI, modelled using Poisson or Gaussian prior statistics, and provide gains of up to 45.93% and 49.69%, respectively, compared to the classical CS scenario [8].

Statistical Modelling

We assume that the signal-of-interest vector $\underline{s} = [s(1), \ldots, s(i), \ldots, s(N)]^T$ is an approximately K-sparse signal. Also, the vector $\underline{w} = [w(1), \ldots, w(i), \ldots, w(N)]^T$ is assumed to be the SI that is correlated with $\underline{s}$. To provide a joint statistical description, we assume that the vectors $\underline{s}$ and $\underline{w}$ are samples (or, realizations) of the random vectors $\boldsymbol{S} = [S_1, \ldots, S_i, \ldots S_N]^T$ and $\boldsymbol{W} = [W_1, \ldots, W_i, \ldots W_N]^T$, respectively. Moreover, we consider that $\underline{s}$ and $\underline{w}$ are drawn *i.i.d.* from their joint probability density function (pdf) $f_{\boldsymbol{S},\boldsymbol{W}}(\boldsymbol{s}, \boldsymbol{w}) = \prod_{i=1}^{N} f_{S_i}(s_i) f_{W_i|S_i}(w_i|s_i)$, where $f_{S_i}(s_i)$ and $f_{W_i|S_i}(w_i|s_i)$ denote the marginal pdf of S_i and the conditional pdf of W_i given S_i, respectively.

The most common distributions used in the literature describing approximately sparse signals are: (a) the zero-mean Laplace distribution $\mathcal{L}(0, b_{s_i})$, where b_{s_i} is the scaling parameter, and (b) the zero-mean Gaussian Mixture Model (GMM) $\mathcal{G}_M(\beta_{s_i}, \sigma_{0,i}, \sigma_{1,i})$ [8], with $\beta_{s_i}, \sigma_{0,i}, \sigma_{1,i}$, being the mixture weight, the low and the high standard deviation, respectively, and (c) the zero-mean generalized Pareto distribution (GPD) $\mathcal{G}_P(q_{s_i}, \sigma_{s_i})$ [29], with order q_{s_i} and shape parameter σ_{s_i}.

Proposed Joint Statistical Distribution Using Copula Functions

Prior art [20–24] focuses on statistical models that exploit the homogeneous nature of the correlated data, meaning that the signal of interest and the SI are produced by sources of the same type. However, when we deal with heterogeneous SI, these assumptions may be inaccurate, due to, for example, variations in signal dimensionality across diverse modalities. Copula functions [27, 28] allow signals produced from diverse data sources to have arbitrary marginal distributions, while merging them into a joint multivariate pdf.

Let $F_{S_i}(s_i)$ and $F_{W_i}(w_i)$ be the marginal cumulative distribution functions (cdfs) that describe the signal coefficients and the SI symbols, respectively. By applying

the probability integral transform on S_i and W_i, they are both transformed into the uniformly-distributed random variables $U_{S_i} = F_{S_i}(s_i)$ and $U_{W_i} = F_{W_i}(w_i)$, respectively. Therefore, regardless of the marginal distribution of S_i and W_i, the transformed variables U_{S_i} and U_{W_i} always follow the uniform distribution. According to Sklar's theorem [27, 28], if $F_{S_i,W_i}(s_i, w_i)$ is the 2-dimensional joint distribution of S_i and W_i, there exists a unique 2-dimensional copula function $C : [0,1]^2 \rightarrow [0,1]$ such that

$$F_{S_i,W_i}(s_i, w_i) = C[F_{S_i}(s_i), F_{W_i}(w_i)]. \tag{1}$$

To derive the corresponding pdf, we differentiate the expression in (1) with respect to $u_{s_i} = F_{S_i}(x_i)$ and $u_{w_i} = F_{W_i}(w_i)$. Then, the joint pdf is given by

$$f_{S_i,W_i}(s_i, w_i) = c\left[F_{S_i}(s_i), F_{W_i}(w_i)\right] f_{S_i}(s_i) f_{W_i}(w_i), \tag{2}$$

where $c\left[F_{S_i}(s_i), F_{W_i}(w_i)\right]$ is the bivariate copula density. Given the marginal pdfs of the random variables, an appropriate copula function that best captures the dependencies among the prediction errors should be selected.

Several copula families exist [28], such as the elliptical (e.g., Gaussian and student t- copulas) and the Archimedean (e.g., Clayton, Frank and Gumbel families). In this work, we aim at demonstrating the capability of copula functions to couple heterogeneous SI in CS. For that reason, we use the simplest copula that belongs to the elliptical family, namely the Gaussian copula. However, depending on the application, a particular copula function that fits the data should be selected.

The multivariate normal copula function [28] is defined as

$$C_g(\mathbf{u}_i) = \Phi_{R_g}(\Phi^{-1}(u_{s_i}), \Phi^{-1}(u_{w_i})), \tag{3}$$

where Φ_{Rg} denotes the standard bivariate normal distribution with correlation matrix $\boldsymbol{R}_g$, Φ^{-1} is the inverse function of the standard univariate normal distribution, and the vector $\mathbf{u}_i = [u_{s_i}, u_{w_i}]$ contains the cdf values. The normal copula density [30] is given by

$$c_g(\boldsymbol{\xi}_i) = |\boldsymbol{R}_g|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\boldsymbol{\xi}_i\left(\boldsymbol{R}_g^{-1} - \mathbf{I}\right)\boldsymbol{\xi}_i^{\mathrm{T}}\right], \tag{4}$$

where $\boldsymbol{\xi}_i = [\Phi^{-1}(u_{s_i}), \Phi^{-1}(u_{w_i})]$ and $\mathbf{I}$ is the 2×2 identity matrix.

During reconstruction, we use the conditional distribution $f_{S_i|W_i}(s_i|w_i)$ that, based on the Bayesian rule, can be written as

$$f_{S_i|W_i}(s_i|w_i) = \frac{f_{S_i,W_i}(s_i, w_i)}{f_{W_i}(w_i)} = c_g\left[F_{S_i}(s_i), F_{W_i}(w_i)\right] f_{S_i}(s_i). \tag{5}$$

Signal Encoding

To accelerate both the encoding and decoding procedures, we use a sparse sensing matrix $\boldsymbol{\Phi}$ to measure the signal. We build the matrix $\boldsymbol{\Phi}$ based on regular low-density parity-check (LDPC) codes [31] since, as their name dictates, they provide sparse sensing matrices that reduce the sensing complexity and storage requirements.
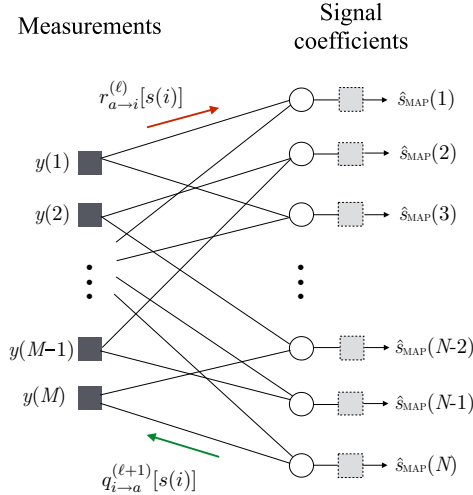
Figure 1: Tanner graph that visualises the message passing algorithm between the variable and the factor nodes.

Slightly deviating from the classical LDPC matrix construction, we consider that the non-zero entries of the matrix can be $\{-1, +1\}$[1] (Rademacher matrices) because negative values improve the system performance. Furthermore, we consider that the *row weight* of $\boldsymbol{\Phi}$ is constant to ensure that each row has the same number of non-zero entries, say L. Also, the *column weight* is assumed to be constant and equal to R. The parameters L and R define the sparsity of the sensing matrix, and can be chosen based on signal properties such as measurement noise or sparsity. Since the signal of interest[2] $\underline{x}$ is not sparse but has an approximately sparse representation $\underline{s}$ on the basis $\boldsymbol{\Psi}$, for example, a wavelet basis or a Discrete Cosine Transform (DCT) matrix, the sensing matrix $\boldsymbol{\Phi}$ is right-multiplied by the dictionary $\boldsymbol{\Psi}$, resulting in the measurement matrix $\mathbf{A}$. Due to the sparsity of $\boldsymbol{\Phi}$, a measurement symbol $y(a), a = 1, 2, ..., M$ is simply a mixture of sums and differences of a small subset of signal coefficients, enabling a fast encoding procedure via matrix-vector multiplication $\underline{y} = \boldsymbol{\Phi}\underline{x}$.

Proposed Signal Recovery via Copula-based Belief Propagation

Optimal reconstruction using SI can be performed, for example, via the maximum a posteriori (MAP) or the minimum mean-squared-error (MMSE) criteria. The *posterior* pdf takes the following form:

$$f_{\boldsymbol{S}|\boldsymbol{Y},\boldsymbol{W}}(\mathbf{s}|\mathbf{y},\mathbf{w}) \propto \prod_{a=1}^{M} f_{Y_a|\boldsymbol{S}}(y_a|\mathbf{s}) \prod_{i=1}^{N} f_{S_i|W_i}(s_i|w_i), \tag{6}$$

since the Markov chain $\boldsymbol{W} \to \boldsymbol{S} \to \boldsymbol{Y}$ holds. Given that the calculation of the pdf in (6) is not always feasible, loopy belief propagation provides an alternative

---

[1]The entries of classical LDPC parity-check matrices are 0 or 1.

[2]We refer to both $\underline{x}$ and $\underline{s}$ as "signal of interest" since knowledge of one implies knowledge of the other: they are related by a simple matrix-vector multiplication: $\underline{x} = \boldsymbol{\Psi}\underline{s}$.

estimation technique, where the estimate of each symbol, say $\hat{s}_i$, is calculated based on its conditional pdf $f_{S_i|\mathbf{Y},W_i}(s_i|\mathbf{y} = \underline{y}, w_i = w(i))$.

In the proposed system design, the decoding of the signal coefficients in $\underline{s}$ is done by employing loopy belief propagation [31], i.e., a soft-decision decoding method that is based on the sum-product algorithm. We assume that the decoder observes a $M \times 1$ measurement vector $\underline{y} = [y(1), \ldots, y(a), \ldots, y(M)]^T$, which is generated during the encoding procedure. Also, we assume $\underline{w}$ to be a SI vector, which is produced from a heterogeneous data source, correlated with the signal of interest s. The SI vector is provided *a priori* to the decoder. A graphical representation of the belief-propagation decoding procedure can be given from a Tanner graph, a factor graph that captures the statistical dependencies between the variables (see Fig. 1). The Tanner graph is a bipartite graph, meaning that the nodes are separated into two distinctive sets: (a) the factor nodes $y(a), a = 1, \ldots, M$, which represent the symbols of the measurement vector (dark-grey squares), and (b) the variable nodes $s(i), i = 1, \ldots, N$, which represent the signal coefficients (white circles). Also, there exists another type of nodes, namely the assistant nodes (light-gray squares), which store the soft-decision information (i.e., beliefs) of each variable node. The edges of the Tanner graph only connect nodes of different types. An edge occurs when there is a non-zero element in the sparse matrix $\mathbf{\Phi}$ and it is associated with a negative or a positive sign.

At the ℓ-th iteration, a variable node $s(i)$ sends a message $q_{i \to a}^{(\ell)}[s(i)]$ to each neighboring factor node $y(a)$, and a factor node $y(a)$ sends a message $r_{a \to i}^{(\ell)}[s(i)]$ to each neighbor $s(i)$ (see Fig. 1). The messages convey encoded beliefs about the most probable value of a node [8]. In our design, the encoded beliefs are the sampled conditional pdfs $f_{S_i|\mathbf{Y},W_i}(s_i|\mathbf{y} = \underline{y}, w_i = w(i))$ that correspond to the signal coefficients. The update rules of the messages are given by

$$r_{a \to i}^{(\ell)}[s(i)] = \frac{1}{C_{a \to i}^{(\ell)}} \sum_{\mathbf{s}_a \setminus i} \left(f_a(y_a|\mathbf{s}_a) \prod_{i' \in \mathcal{N}(a) \setminus i} q_{i \to a}^{(\ell)}[s(i')] \right), \tag{7}$$

$$q_{i \to a}^{(\ell+1)}[s(i)] = \frac{1}{C_{i \to a}^{(\ell+1)}} q_{i \to a}^{(0)}[s(i)] \prod_{a' \in \mathcal{M}(i) \setminus a} r_{a' - i}^{(\ell)}[s(i)], \tag{8}$$

where $q_{i \to a}^{(0)}[s(i)] = f_{S_i|W_i}(s_i|w_i = w(i))$ is the initial message sent from the variable node $s(i)$ to the neighboring measurement node $y(a)$. In (7), $\mathcal{N}(a)$ denotes the set of the neighbors of the measurement node $y(a)$, $\mathbf{s}_a$ the variables in $\mathbf{s}$ that are neighbors to the factor node $y(a)$, and $\mathbf{s}_a \setminus i$ the set of variables in vector $\mathbf{s}_a$ with the variable s_i excluded. The function $f_a(\cdot)$ is the factor function[3] that applies the constraints imposed by the neighbors of the factor node $y(a)$ via the measurement matrix $\mathbf{A}$. Moreover, $\mathcal{M}(i)$ is the set of the neighbors of the variable node $s(i)$, while the normalization factors $C_{a \to i}^{(\ell)}$ and $C_{i \to a}^{(\ell+1)}$ guarantee that $\sum_{s(i)} r_{a \to i}^{(\ell)}[s(i)] = 1$ and $\sum_{s(i)} q_{i \to a}^{(\ell+1)}[s(i)] = 1$, respectively.

The message passing takes place for a maximum number of iterations, say Λ. At that point, we calculate the final MAP estimate $\hat{s}_{\mathrm{MAP}}(i)$, based on the stored

---

[3]In this model $f_a(y_a|\mathbf{s}_a) = \delta(y_a - \sum_{i \in \mathcal{N}(a)} A_{a,i} s_i)$, where $\delta(\cdot)$ denotes a Dirac delta function.

soft-decision information at the assistant node of $s(i)$, as

$$\hat{s}_{\text{MAP}}(i) = \arg \max_{s_i} q_{i \to a}^{(0)}[s(i)] \prod_{i \in \mathcal{N}(i)} r_{a \to i}^{(\Lambda)}[s(i)]. \tag{9}$$

The messages sent between neighboring nodes are vectors containing p pdf samples. At the coefficient nodes, the multiplication of messages, as it is described by (8), corresponds to element-wise multiplication between vectors. At the measurement nodes, the message update is performed by (7), which is done in the frequency domain via the Fast Fourier Transform (FFT).

Numerical results

In this section, we compare the performance of the proposed method with the state-of-the-art algorithms in CS with SI [15] and classical CS [8]. In our simulations, we consider $\boldsymbol{\Psi} = \boldsymbol{I}_{N \times N}$ and, hence, $\mathbf{A} = \boldsymbol{\Phi}$. The vector length of the signals is $N = 1000$ and the length of each message vector (which contains the pdf samples) is set to $p = 243$ in order to provide fast FFT calculations and satisfactory pdf descriptions. Moreover, the column weight of the sensing matrix $\boldsymbol{\Phi}$ has a value of $L = 20$, which offers a good balance between the reconstruction gain and the number of measurements.

The signal of interest is assumed to take continuous values drawn *i.i.d.* from the zero-mean Laplace distribution $\mathcal{L}(0, b_{s_i})$ with scaling parameter $b_{s_i} = 2$. To demonstrate the efficiency of the proposed algorithm when dealing both with homogeneous and heterogeneous SI, we assume that the SI signal has one of the following distributions: (a) Laplace distribution with $b_{w_i} = 2.5$, (b) Poisson distribution $\mathcal{P}(\lambda_{s_i})$ with shape parameter $\lambda_{w_i} = 8$, and (c) Gaussian distribution $\mathcal{N}(\mu_{w_i}, \sigma_{w_i})$ with mean $\mu_{w_i} = 10$ and standard deviation $\sigma_{w_i} = 5$. The first model implies an approximately sparse SI signal, whereas the second and the third imply that the SI signal is discrete and continuous non-sparse, respectively. In order to have a fair comparison with the state of the art, we perform experiments using synthetic data generated from bivariate copulas other than the Gaussian. Specifically, we generate data from the Clayton and Frank copula families [28], which are parameterized by α_c and α_f, respectively. For the proposed model, the joint statistical description of the sources is modeled using a bivariate Gaussian copula function, where the correlation parameter ρ is fitted on the generated data via maximum likelihood estimation.

Firstly, we compare the performance of the proposed algorithm against that of the ADMM-based $\ell_1 - \ell_1$ algorithm presented in [15]. The SI signal is assumed to follow a Laplace distribution in order to fairly compare the two schemes. In Fig. 2 we plot the results for different values of α_c and α_f, namely, $\alpha_c = 1, 5, 15$ for the Clayton copula, and $\alpha_f = 4, 8, 20$ for the Frank copula. The choice of the values of α_c and α_f was done so as to provide results in the weak, moderate and strong correlation regime, respectively. We observe that both the proposed algorithm and the $\ell_1 - \ell_1$ method manage to efficiently exploit the SI and systematically improve the reconstruction compared to the classical CS method [8]. As expected, the performance improvements are increasing with the amount of correlation between the source and the SI.
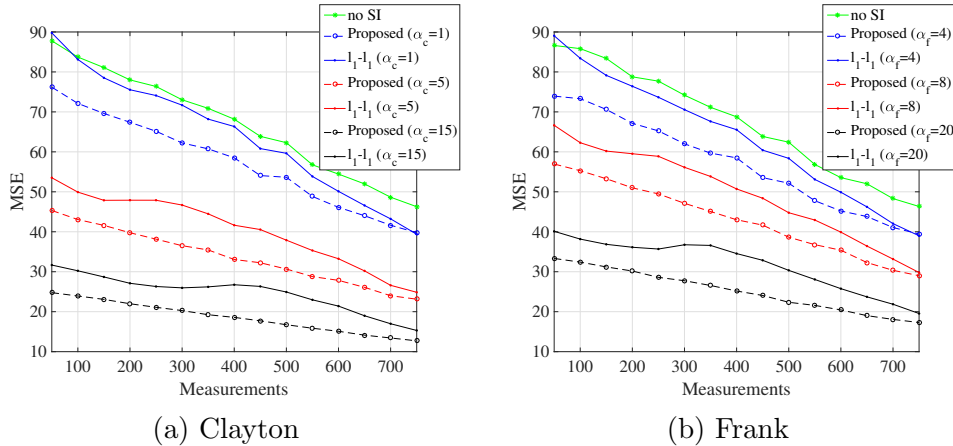
Figure 2: Decoding error expressed in MSE for a varying number of measurements M and different parameters. Note that $\alpha_c = 1$ (and $\alpha_f = 4$) imply weak correlation, $\alpha_c = 5$ (and $\alpha_f = 8$) moderate correlation, and $\alpha_c = 15$ (and $\alpha_f = 20$) strong correlation. The SI signal is assumed to follow a Laplace distribution and the data are generated by (a) Clayton, and (b) Frank copula.

Furthermore, it is clear that the proposed algorithm systematically outperforms the $\ell_1 - \ell_1$ algorithm, providing MSE reductions of up to 27.41% (Frank, $\alpha_f = 20$) and 32.73% (Clayton, $\alpha_c = 15$).

Subsequently, we compare the performance of the proposed algorithm and the $\ell_1 - \ell_1$ method under the availability of a Poisson-distributed SI signal. Fig. 3(a) shows that our algorithm achieves MSE reductions of up to 20.29% (Frank, $\alpha_f = 4$) and 45.93% (Clayton, $\alpha_c = 5$) compared to classical CS. In the same scenario, the $\ell_1 - \ell_1$ method constantly underperforms classical CS due to the fact that the algorithm has been designed to perform well mainly under Laplace statistics.

Finally, as shown in Fig. 3(b), the same holds when the SI is a normally distributed non-sparse signal, where the MSE reduction of proposed algorithm reaches up to 19.80% (Frank, $\alpha_f = 4$) and 49.69% (Clayton, $\alpha_c = 5$) compared to classical CS.

Conclusion

We proposed a copula-based CS reconstruction method that builds on belief propagation and message passing principles. As shown by experimentation, the proposed scheme manages to leverage the underlying correlation between the signal of interest and the SI even when they are produced from diverse sources. Irrespective of whether the SI signal is approximately sparse, discrete or continuous non-sparse, the proposed method yields significant MSE reductions compared to the state-of-the-art works in classical CS and CS with SI. Our framework can therefore offer a new perspective to several CS scenarios where heterogenous SI can be exploited during reconstruction.
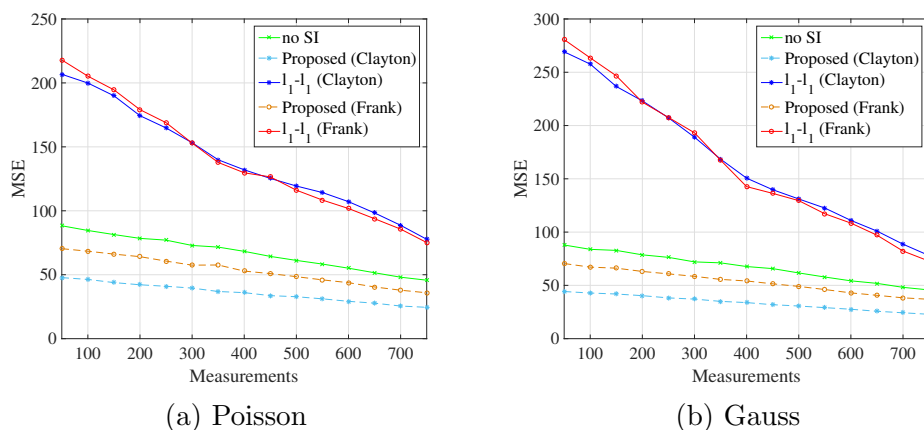
Figure 3: Decoding error expressed in MSE for a varying number of measurements M. The SI signal is assumed to follow a: (a) Poisson, and (b) Gaussian distribution.

References

[1] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 72–82, 2008.

[2] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[3] A. Talari and N. Rahnavard, "CStorage: Distributed data storage in wireless sensor networks employing compressive sensing," in *IEEE Global Com. Conf. (GLOBECOM)*, 2011, pp. 1–5.

[4] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.

[5] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[6] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[7] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[8] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 269–280, 2010.

[9] L. Weizman, Y. C. Eldar, D. B. Bashat, D. Nguyen, L.-N. Tran, P. Pirinen, M. Latva-aho, A. Mansoor, V. Patsekin, D. Scherl *et al.*, "Compressed sensing for longitudinal MRI: An adaptive-weighted approach," *arXiv:1407.2602*, 2014.

[10] X. Yuan, T.-H. Tsai, R. Zhu, P. Llull, D. Brady, and L. Carin, "Compressive hyperspectral imaging with side information," *IEEE J. Sel. Topics in Signal Process.*, vol. 9, no. 6, pp. 964–976, 2015.

[11] J. F. C. Mota, N. Deligiannis, A. C. Sankaranarayanan, V. Cevher, and M. R. Rodrigues, "Dynamic sparse state estimation using $\ell_1 - \ell_1$ minimization: Adaptive-rate measurement bounds, algorithms and applications," in *IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, 2015.

[12] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Trans. Signal Process.*, vol. 58, no. 9, 2010.

[13] M. Trocan, T. Maugey, J. E. Fowler, and B. Pesquet-Popescu, "Disparity-compensated compressed-sensing reconstruction for multiview images," in *IEEE Int. Conf. Multimedia and Expo (ICME)*. IEEE, 2010, pp. 1225–1229.

[14] J. E. Fowler, S. Mun, and E. W. Tramel, "Block-based compressed sensing of images and video," *Found. and Trends Signal Process.*, vol. 4, no. 4, pp. 297–416, 2012.

[15] J. F. C. Mota, N. Deligiannis, and M. R. D. Rodrigues, "Compressed sensing with prior information: Optimal strategies, geometry, and bounds," *arXiv:1408.5250*, 2014.

[16] J. F. Mota, N. Deligiannis, and M. R. Rodrigues, "Compressed sensing with side information: Geometrical interpretation and performance bounds," in *IEEE Global Conf. Signal and Inf. Process. (GlobalSIP)*, 2014, pp. 512–516.

[17] J. F. C. Mota, L. Weizman, N. Deligiannis, Y. Eldar, and M. R. Rodrigues, "Reference-based compressed sensing: A sample complexity approach," in *IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, 2016.

[18] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, 2008.

[19] X. Wang and J. Liang, "Side information-aided compressed sensing reconstruction via approximate message passing," in *IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*. IEEE, 2014, pp. 3330–3334.

[20] Y. Liu, X. Zhu, and L. Zhang, "Noise-resilient distributed compressed video sensing using side-information-based belief propagation," in *IEEE Int. Conf. Network Infrastr. and Digital Content (IC-NIDC)*. IEEE, 2012, pp. 350–390.

[21] P. Beigi, X. Xiu, and J. Liang, "Compressive sensing based multiview image coding with belief propagation," in *Asilomar Conf. Signals, Sys., Comput.* IEEE, 2010.

[22] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *Asilomar Conf. Signals, Sys., Comput.*, 2005, pp. 1537–1541.

[23] D. Baron, M. F. Duarte, S. Sarvotham, M. B. Wakin, and R. G. Baraniuk, "An information-theoretic approach to distributed compressed sensing," in *Proc. Conf. Commun., Control, and Comput.*, 2005.

[24] F. Renna, L. Wang, X. Yuan, J. Yang, G. Reeves, R. Calderbank, L. Carin, and M. Rodrigues, "Classification and reconstruction of high-dimensional signals from low-dimensional noisy features in the presence of side information," *arXiv:1412.0614*, 2014.

[25] N. Deligiannis, E. Zimos, D. Ofrim, Y. Andreopoulos, and A. Munteanu, "Distributed joint source-channel coding with copula-function-based correlation modeling for wireless sensors measuring temperature," *IEEE Sensors J.*, vol. 15, no. 8, pp. 4496–4507, 2015.

[26] M. J. Ehrhardt, K. Thielemans, L. Pizarro, D. Atkinson, S. Ourselin, B. F. Hutton, and S. R. Arridge, "Joint reconstruction of PET-MRI by exploiting structural similarity," *Inverse Problems*, vol. 31, no. 1, p. 015001, 2015.

[27] M. Sklar, *Fonctions de répartition à n dimensions et leurs marges*. Un. Paris 8, 1959.

[28] R. B. Nelsen, *An introduction to copulas*. Springer Science, 2013, vol. 139.

[29] R. G. Baraniuk, V. Cevher, and M. B. Wakin, "Low-dimensional models for dimensionality reduction and signal recovery: A geometric perspective," *IEEE Proc.*, vol. 98, no. 6, pp. 959–971, 2010.

[30] R. T. Clemen and T. Reilly, "Correlations and copulas for decision and risk analysis," *Management Science*, vol. 45, no. 2, pp. 208–224, 1999.

[31] R. G. Gallager, "Low-density parity-check codes," *IRE Trans. on Inf. Theory*, vol. 8, no. 1, pp. 21–28, 1962.