Robust 1-Bit Compressive Sensing via Gradient Support Pursuit

Bahmani, S.; Boufounos, P.T.; Raj, B.

Abstract

This paper studies a formulation of 1-bit Compressed Sensing (CS) problem based on the maximum likelihood estimation framework. In oder to solve the problem we apply the recently proposed Gradient Support Pursuit algorithm, with a minor modification. Assuming the proposed objective function has a Stable Restricted Hessian, the algorithm is shown to accurately solve the 1-bit CS problem. Furthermore, the algorithm is compared to the state-of- the-art 1-bit CS algorithms through numerical simulations. The results suggest that the proposed method is robust to noise and at mid to low input SNR regime it achieves the best reconstruction SNR vs. execution time trade-off.

Cornell University Library

Robust 1-bit Compressive Sensing via Gradient Support Pursuit

Sohail Bahmani, *Student, IEEE*, Petros T. Boufounos, *Member, IEEE*, Bhiksha Raj

Abstract

This paper studies a formulation of 1-bit Compressed Sensing (CS) problem based on the maximum likelihood estimation framework. In oder to solve the problem we apply the recently proposed Gradient Support Pursuit algorithm, with a minor modification. Assuming the proposed objective function has a Stable Restricted Hessian, the algorithm is shown to accurately solve the 1-bit CS problem. Furthermore, the algorithm is compared to the state-of-the-art 1-bit CS algorithms through numerical simulations. The results suggest that the proposed method is robust to noise and at mid to low input SNR regime it achieves the best reconstruction SNR vs. execution time trade-off.

Index Terms

1-bit compressed sensing, sparsity, quantization

I. INTRODUCTION

Quantization is an indispensable part of digital signal processing and digital communications systems. To incorporate Compressed Sensing (CS) methods in these systems, it is thus necessary to analyze and evaluate them considering the effect of measurement quantization. In the recent years there is a growing interest in quantized CS in the literature [1]–[6], particularly the extreme case of quantization to a single bit, known as 1-bit Compressed Sensing [7], where only the sign of the linear measurements are recorded. The advantage of this acquisition scheme is that it can be implemented using simple hardware that is

S.B. is with Department of Electrical and Computer Engineering, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA, 15213 e-mail: `sbahmani@cmu.edu`

P.T.B. is with Mitsubishi Electric Research Laboratory, 201 Broadway, Cambridge, MA 02139 e-mail: `petrosb@merl.com`

B.R. is with Language Technologies Institute, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA, 15213 e-mail: `bhiksha@cs.cmu.edu`

not expensive and can operate at very high sampling rates. The formulation of this problem is also very similar to the sparse logistic regression, very useful in machine learning applications [8].

As in standard CS, the algorithms proposed for the 1-bit CS problem can be categorized into convex methods and non-convex greedy methods. In [7] an algorithm is proposed for 1-bit CS that induces sparsity through the ℓ_1-norm while penalizes incosistency with the 1-bit sign measurements via a convex regularization term. In a noise-free scenario, the 1-bit measurements do not convey any information about the length of the signal. Therefore, the algorithm of [7], as well as other 1-bit CS algorithms, aim at accurate estimation of the normalized signal. Requiring the 1-bit CS estimate to lie on the surface of the unit-ball imposes a non-convex constraint in methods that perform an (approximate) optimization, even those that use the convex ℓ_1-norm to induce sparsity. Among greedy 1-bit CS algorithms, an algorithm called Matching Sign Pursuit (MSP) is proposed in [9] based on the CoSaMP algorithm [10]. This algorithm is empirically shown to perform better than the standard CoSaMP algorithm for estimation of the normalized sparse signal. In [11] the Restricted-Step Shrinkage (RSS) algorithm is proposed for 1-bit CS problems, improving the performance of ℓ_1-based algorithms. This algorithm, which is similar to *trust-region* algorithms in non-convex optimization, is shown to converge to a stationary point of the objective function regardless of the initialization. More recently, [12] derived a lower bound on the best achievable reconstruction error of any 1-bit CS algorithm in noise-free scenarios. Furthermore, using the notion of "binary stable embeddings", they have shown that Gaussian measurement matrices can be used for 1-bit CS problems both in noisy and noise-free regime. The Binary Iterative Hard Thresholding (BIHT) algorithm is also proposed in [12] and shown to have favorable performance compared to the RSS and MSP algorithms through numerical simulations. For robust 1-bit CS in presence of noise, [13] also proposed the Adaptive Outlier Pursuit (AOP) algorithm. In each iteration of the AOP , first the sparse signal is estimated similar to BIHT with the difference that the potentially corrupted measurements are excluded. Then with the new signal estimate fixed, the algorithm updates the list of likely corrupted measurements. The AOP is shown to improve on performance of BIHT through numerical simulations. [14] proposed a linear program to solve the 1-bit CS problems in a noise-free scenario. The algorithm is proved to provide accurate solutions, albeit using a sub-optimal number of measurements. Furthermore, in [8] a convex program is proposed that is robust to noise in 1-bit measurements and achieves the optimal number of measurements and in [15] it is shown that non-gaussian matrices can be used for acquisition unders certain conditions on the acquired signal..

In this paper, we formulate the 1-bit CS problem as an sparsity-constrained optimization. As described in Section II, the objective function is obtained by adjusting the loss function that arises in the Maximum

Likelihood Estimation (MLE) formulation of the problem. To solve this optimization problem in Section III we propose a slightly modified version of the Gradient Support Pursuit (GraSP) algorithm proposed in [16]. In Section IV, the algorithm is shown to yield an approximate solution provided that the objective function satisfies certain sufficient conditions. Furthemore, in Section V, we compare the performance of our algorithm with the BIHT algorithm and the non-convex variant of the 1-bit CS solver introduced in [8] through numerical simulations. As an aside, we show that the non-convex solver described in [8] has an explicit solution (see Appendix B). Finally, we discuss and conclude in Section VI.

Notation: In the remainder of the paper we denote the positive part of a real number x by $(x)_+$. For a positive integer k, the set $\{1, 2, \ldots, k\}$ is denoted by $[k]$. The indicator function is denoted by $\mathbb{1}(\cdot)$. Vectors and matrices are denoted by boldface characters. With the exception of $\mathcal{N}$ which we use to denote the normal distribution as in $\mathcal{N}(0, 1)$, calligraphic letters are reserved for denoting sets. The support set (i.e., the set of non-zero coordinates) of a vector $\mathbf{x}$ is denoted by $\mathrm{supp}(\mathbf{x})$. The best k-term approximation of a vector $\mathbf{v}$ is denoted by $\mathbf{v}_k$. Depending on the context $\mathbf{v}_{[k]}$ may also be a k-dimensional vector denoting the restriction of $\mathbf{v}$ to its k largest entries in magnitudes (i.e., truncated best k-term approximation of $\mathbf{v}$). Restriction of an n-dimensional vector $\boldsymbol{v}$ to its entries corresponding to an index set $\mathcal{I} \subseteq [n]$ is denoted by $\mathbf{v}|_{\mathcal{I}}$. Similarly $\mathbf{A}_{\mathcal{I}}$ denotes the restriction of a matrix $\mathbf{A}$ to the columns enumerated by $\mathcal{I}$. Restriction of the identity matrix to the columns enumerated by $\mathcal{I}$ is particularly denoted by $\mathbf{P}_{\mathcal{I}}$. The operator norm of $\mathbf{A}$ with respect to the ℓ_2-norm is denoted by $\|A\|$. For square matrices $\mathbf{A}$ and $\mathbf{B}$ we write $\mathbf{B} \preccurlyeq \mathbf{A}$ to state that $\mathbf{A} - \mathbf{B}$ is positive semidefinite. We use $\nabla f(\cdot)$ and $\nabla^2 f(\cdot)$ to denote the gradient and the Hessian of a twice continuously differentiable function $f : \mathbb{R}^n \mapsto \mathbb{R}$. For an index set $\mathcal{I} \subset [n]$, the restriction of the gradient to the entries selected by $\mathcal{I}$ and the restriction of the Hessian to the entries selected by $\mathcal{I} \times \mathcal{I}$ are denoted by $\nabla_{\mathcal{I}} f(\cdot)$ and $\nabla^2_{\mathcal{I}} f(\cdot)$, respectively. Finally, numerical superscripts within parentheses denote the iteration index.

II. PROBLEM FORMULATION

We cast the 1-bit CS problem in the framework of statistical parametric estimation which is also considered in [4]. In 1-bit CS, binary measurements $y \in \{\pm 1\}$ of a signal $\mathbf{x}^\star \in \mathbb{R}^n$ are collected based on the model

$$y = \mathrm{sgn}\left(\langle \mathbf{a}, \mathbf{x}^\star \rangle + e\right), \tag{1}$$

where $\mathbf{a}$ is a measurement vector and e denotes an additive noise with distribution $\mathcal{N}(0,\sigma^2)$. It is straightforward to show the conditional likelihood of y given $\mathbf{a}$ and signal $\mathbf{x}$ can be written as

$$\Pr\left\{y \mid \mathbf{a}; \mathbf{x}\right\} = \Phi\left(y \frac{\langle \mathbf{a}, \mathbf{x}\rangle}{\sigma}\right),$$

with $\Phi\left(\cdot\right)$ denoting the standard normal cumulative distribution function (CDF). Then, for measurement pairs $\{(\mathbf{a}_i, y_i)\}_{i=1}^m$ the MLE loss function is given by

$$f_{\mathrm{MLE}}\left(\mathbf{x}\right) := -\frac{1}{m}\sum_{i=1}^m \log\left(\Phi\left(y_i \frac{\langle \mathbf{a}_i, \mathbf{x}\rangle}{\sigma}\right)\right).$$

The solution should exploit the sparsity of the solution and sparsely minimize the function above. Note, however, that at high Signal-to-Noise Ratio (SNR) this function does lend itself easily to optimization. To observe this behavior, rewrite f_{MLE} as

$$f_{\mathrm{MLE}}\left(\mathbf{x}\right) = \frac{1}{m}\sum_{i=1}^m g_\eta\left(y_i \left\langle \mathbf{a}_i, \frac{\mathbf{x}}{\|\mathbf{x}\|_2}\right\rangle\right),$$

where $\eta := \frac{\|\mathbf{x}\|_2}{\sigma}$ is the SNR and $g_\omega\left(t\right) := -\log\Phi\left(\omega t\right)$ for all $\omega \geq 0$. As $\eta \to +\infty$ the function $g_\eta\left(t\right)$ tends to

$$g_\infty\left(t\right) := \begin{cases} 0 & t > 0 \\ \log 2 & t = 0 \\ +\infty & t < 0 \end{cases}.$$

Therefore, as the SNR increases to infinity $f_{\mathrm{MLE}}\left(\mathbf{x}\right)$ tends to a sum of discontinuous constant functions do not uniquely identify the solution and are difficult to handle in practice. Whether the noise level is too low or the signal too strong relative to the noise, in a high SNR scenario the measurement vectors are likely to become linearly separable with respect to the corresponding binary measurements. In these cases, the minimizer of f_{MLE} would not be unique, resulting to significant ambiguity about the true amplitude of the signal. This is essentially the same problem as the amplitude ambiguity demontrated in the original formulation in [7].

To avoid the problems mentioned above we consider a modified loss function

$$f_0\left(\mathbf{x}\right) := -\frac{1}{m}\sum_{i=1}^m \log\left(\Phi\left(y_i \langle \mathbf{a}_i, \mathbf{x}\rangle\right)\right), \tag{2}$$

while we merely use an alternative formulation of (1) given by

$$y = \mathrm{sgn}\left(\eta \langle \mathbf{a}, \mathbf{x}^\star\rangle + e\right),$$

in which $\eta > 0$ denotes the true SNR, $\mathbf{x}^\star$ is assumed to be unit-norm, and $e \sim \mathcal{N}(0, 1)$. The aim is accurate estimation of the unit-norm signal $\mathbf{x}^\star$ which is assumed to be s-sparse. Disregarding computational complexity, the candidate estimator would be

$$\arg \min_{\mathbf{x}} \ f_0(\mathbf{x}) \quad \text{s.t. } \|\mathbf{x}\|_0 \leq s \text{ and } \|\mathbf{x}\|_2 \leq 1. \tag{3}$$

However, finding the exact solution (3) may be computationally intractable, thereby we merely focus on approximate solutions to this optimization problem.

III. ALGORITHM

In this section we introduce a modified version of the GraSP algorithm, outlined in Algorithm 1, for estimation of bounded sparse signals associated with a cost function. While in this paper the main goal is to study the 1-bit CS problem and in particular the objective function described by (2), we state performance guarantees of Algorithm 1 in more general terms. As in GraSP, in each iteration first the $2s$ coordinates at which the gradient of the cost function at the iterate $\mathbf{x}^{(t)}$ has the largest magnitudes are identified. These coordinates, denoted by $\mathcal{Z}$, are then merged with the support set of $\mathbf{x}^{(t)}$ to obtain the set $\mathcal{T}$ in the second step of the iteration. Then, as expressed in line 3 of Algorithm 1, a crude estimate $\mathbf{b}$ is computed by minimizing the cost function over vectors of length no more than r whose supports are subsets of $\mathcal{T}$. Note that this minimization would be a convex program and therefore tractable, provided that the sufficient conditions proposed in Section IV hold. In the final step of the iteration (i.e., line 4) the crude estimate is pruned to its best s-term approximation to obtain the next iterate $\mathbf{x}^{(t+1)}$. By definition we have $\|\mathbf{b}\|_2 \leq r$, thus the new iterate remains in the feasible set (i.e., $\|\mathbf{x}^{(t+1)}\|_2 \leq r$).

IV. ACCURACY GUARANTEES

In order to provide accuracy guarantees for Algorithm 1, we rely on the notion of SRH described in [16] with a slight modification in its definition. The original definition of SRH basically characterizes the cost functions that have bounded curvature over sparse canonical subspaces, possibly at locations arbitrarily far from the origin. However, we only require the bounded curvature condition to hold at locations that are within a sphere around the origin. More precisely, we redefine the SRH as follows.

Definition 1 (Stable Restricted Hessian)**.** Suppose that $f : \mathbb{R}^n \mapsto \mathbb{R}$ is a twice continuously differentiable function and let $k < n$ be a positive integer. Furthermore, let $\alpha_k(\mathbf{x})$ and $\beta_k(\mathbf{x})$ be in turn the largest and smallest real numbers such that

$$\beta_k(\mathbf{x}) \|\mathbf{\Delta}\|_2^2 \leq \mathbf{\Delta}^{\mathrm{T}} \nabla^2 f(\mathbf{x}) \mathbf{\Delta} \leq \alpha_k(\mathbf{x}) \|\mathbf{\Delta}\|_2^2, \tag{4}$$

Algorithm 1: GraSP with Bounded Thresholding

$\qquad\qquad s \qquad$ desired sparsity level

input : $\ r \qquad$ radius of the feasible set

$\qquad\qquad f(\cdot) \quad$ the cost function

$t \longleftarrow 0$

$\mathbf{x}^{(t)} \longleftarrow \mathbf{0}$

repeat

$\quad$ 1 $\quad \mathcal{Z} \longleftarrow \mathrm{supp}\left(\left[\nabla f\left(\mathbf{x}^{(t)}\right)\right]_{2s}\right)$

$\quad$ 2 $\quad \mathcal{T} \longleftarrow \mathrm{supp}\left(\mathbf{x}^{(t)}\right) \cup \mathcal{Z}$

$\quad$ 3 $\quad \mathbf{b} \longleftarrow \arg\min_{\mathbf{x}} f\left(\mathbf{x}\right) \quad$ s.t. $\ \mathbf{x}|_{\mathcal{T}^c} = \mathbf{0}$ and $\|\mathbf{x}\|_2 \leq r$

$\quad$ 4 $\quad \mathbf{x}^{(t+1)} \longleftarrow \mathbf{b}_s$

$\quad$ 5 $\quad t \longleftarrow t+1$

until *halting condition holds*

return $\mathbf{x}^{(t)}$

---

holds for all $\boldsymbol{\Delta}$ and $\mathbf{x}$ that obey $|\mathrm{supp}\left(\boldsymbol{\Delta}\right) \cup \mathrm{supp}\left(\mathbf{x}\right)| \leq k$ and $\|\mathbf{x}\|_2 \leq r$. Then f is said to have an Stable Restricted Hessian of order k with constant $\mu_k \geq 1$ in a sphere of radius $r > 0$, or for brevity (μ_k, r)-SRH, if $1 \leq \alpha_k\left(\mathbf{x}\right) / \beta_k\left(\mathbf{x}\right) \leq \mu_k$ for all k-sparse $\mathbf{x}$ with $\|\mathbf{x}\|_2 \leq r$.

Theorem 1. *Let $\overline{\mathbf{x}}$ be a vector such that $\|\overline{\mathbf{x}}\|_0 \leq s$ and $\|\overline{\mathbf{x}}\|_2 \leq r$. If the cost function $f\left(\mathbf{x}\right)$ have (μ_{4s}, r)-SRH corresponding to the curvature bounds $\alpha_{4s}\left(\mathbf{x}\right)$ and $\beta_{4s}\left(\mathbf{x}\right)$ in (4), then iterates of Algorithm 1 obey*

$$\left\|\mathbf{x}^{(t+1)} - \overline{\mathbf{x}}\right\|_2 \leq \left(\mu_{4s}^2 - \mu_{4s}\right) \left\|\mathbf{x}^{(t)} - \overline{\mathbf{x}}\right\|_2 + 2\left(\mu_{4s} + 1\right)\epsilon,$$

where ϵ obeys $\left\|\left[\nabla f\left(\overline{\mathbf{x}}\right)\right]_{3s}\right\|_2 \leq \epsilon\,\beta_{4s}\left(\mathbf{x}\right)$ for all $\mathbf{x}$ with $\|\mathbf{x}\|_0 \leq 4s$ and $\|\mathbf{x}\|_2 \leq r$.

The immediate implication of this theorem is that if the 1-bit CS loss $f_0\left(\mathbf{x}\right)$ has $(\mu_{4s}, 1)$-SRH with $\mu_{4s} \leq \frac{1+\sqrt{3}}{2}$ then we have $\left\|\mathbf{x}^{(t)} - \mathbf{x}^\star\right\|_2 \leq 2^{-t}\|\mathbf{x}^\star\|_2 + 2\left(3 + \sqrt{3}\right)\epsilon$.

Proof of Theorem 1 is almost identical to the proof of Theorem 1 in [16]. For brevity we will provide a proof sketch in Appendix A and elaborate only on the more distinct parts of the proof and borrow the remaining parts from [16].

V. SIMULATIONS

In our simulations using synthetic data we considered signals of dimensionality $n = 1000$ that are s-sparse with $s = 10, 20,$ or 30. The non-zero entries of the signal constitute a vector randomly drawn from the surface of the unit Euclidean ball in $\mathbb{R}^s$. The $m \times n$ measurement matrix has iid standard Gaussian entries with m varying between 100 and 2000 in steps of size 100. We also considered three different noise variances σ^2 corresponding to input SNR $\eta = 20$dB, 10dB, and 0dB. Figures 1–5 illustrate the average performance of the considered algorithm over 200 trials versus the sampling ratio (i.e., m/n). In these figures, the results of Algorithm 1 considering f_0 and f_{MLE} as the objective function are demarcated by GraSP and GrasP-η, respectively. Furthermore, the results corresponding to BIHT algorithm with one-sided ℓ_1 and ℓ_2 objective functions are indicated by BIHT and BIHT-ℓ_2, respectively. We also considered the ℓ_0-constrained optimization proposed by [8] which we refer to as PV-ℓ_0. While [8] mostly focused on studying the convex relaxation of this method using ℓ_1-norm, as shown in Appendix B the solution to PV-ℓ_0 can be derived explicitly in terms of the one-bit measurements, the measurement matrix, and the sparsity level. We do not evaluate the convex solver proposed in [8] because we did not have access to an efficient implementation of this method. Furthermore, this convex solver is expected to be inferior to PV-ℓ_0 in terms of accuracy because it operates on a feasible set with larger *mean width* [8, Theorem 1.1]. With the exception of the non-iterative PV-ℓ_0, the other four algorithms considered in our simulations are iterative; they are configured to halt when they produce an estimate whose 1-bit measurements and the real 1-bit measurements have a Hamming distance smaller than an η-dependent threshold.

Figure 1 illustrates performance of the considered algorithms in terms of the angular error between the normalized estimate $\widehat{\mathbf{x}}$ and the true signal $\mathbf{x}^\star$ defined as $\mathrm{AE}\left(\widehat{\mathbf{x}}\right) := \frac{1}{\pi} \cos^{-1} \langle \widehat{\mathbf{x}}, \mathbf{x}^\star \rangle$. As can be seen from the figure, with higher input SNR (i.e., η) and less sparse target signals the algorithms incur larger angular error. While there is no significant difference in performance of GaSP, GraSP-η, and BIHT-ℓ_2 for the examined values of η and s, the BIHT algorithm appears to be sensitive to η. At $\eta = 20$dB and low sampling ratios BIHT outperforms the other methods by a noticeable margin. However, for more noisy measurements BIHT loses its advantage and at $\eta = 0$dB it performs even poorer than the PV-ℓ_0. PV-ℓ_0 never outperforms the two variants of GraSP or the BIHT-ℓ_2, but the gap between their achieved angular error decreases as the measurements become more noisy.

The reconstruction SNR of the estimates produced by the algorithms are compared in Figure 2. The reconstruction SNR conveys the same information as the angular error as it can be calculated through
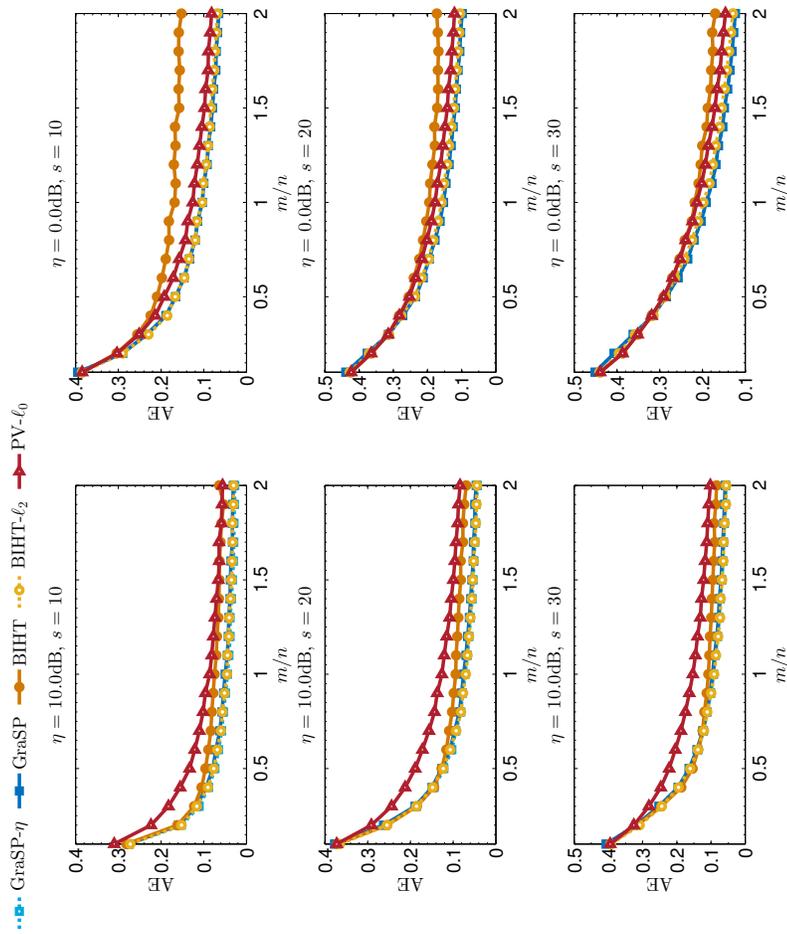
Figure 1. Angular error (AE) vs. the sampling ratio (m/n) at different values of input SNR (η) and sparsity (s)

the formula

$$\mathrm{R} - \mathrm{SNR}\left(\widehat{\mathbf{x}}\right) := -20\log_{10}\left\|\widehat{\mathbf{x}} - \mathbf{x}^\star\right\|_2$$

$$= -10\log_{10}\left(2 - 2\cos\mathrm{AE}\left(\widehat{\mathbf{x}}\right)\right).$$

However, it magnifies small differences between the algorithms that were difficult to trace using the angular error. For example, it can be seen in Figure 2 that at $\eta = 20$dB and $s = 10$, GraSP-η has an advantage (of up to 2dB) in reconstruction SNR.

Furthermore, we evaluated performance of the algorithms in terms of identifying the correct support set of the target sparse signal by are comparing their achieved False Negative Rate

$$\mathrm{FNR} = \frac{\left|\mathrm{supp}\left(\mathbf{x}^\star\right)\setminus\mathrm{supp}\left(\widehat{\mathbf{x}}\right)\right|}{\left|\mathrm{supp}\left(\mathbf{x}^\star\right)\right|}$$

and False Positive Rate

$$\mathrm{FPR} = \frac{\left|\mathrm{supp}\left(\widehat{\mathbf{x}}\right)\setminus\mathrm{supp}\left(\mathbf{x}^\star\right)\right|}{n - \left|\mathrm{supp}\left(\mathbf{x}^\star\right)\right|}.$$

Figures 3 and 4 illustrate these rates for the studied algorithms. It can be seen in Figure 3 that at $\eta = 20$dB, BIHT achieves a FNR slightly lower than that of the variants of GraSP, whereas PV-ℓ_0 and BIHT-ℓ_2 rank first and second, respectively, in the highest FNR at a distant from the other algorithms. However, as η decreases the FNR of BIHT deteriorates relative to the other algorithms while BIHT-ℓ_2 shows improved FNR. The GraSP variants exhibit better performance overall at smaller values of η especially with $s = 10$, but for $\eta = 10$dB and at low sampling ratios BIHT attains a slightly better FNR. The relative performance of the algorithms in terms of FPR, illustrated in Figure 4, is similar.

We also compared the algorithms in terms of their average execution time (T) measured in seconds. The simulation was ran on a PC with an AMD Phenom$^{\mathrm{TM}}$II X6 2.60GHz processor and 8.00GB of RAM. The average execution time of the algorithms, all of which are implemented in MATLAB$^{\circledR}$, is illustrated in 5 in log scale. It can be observed from the figure that PV-ℓ_0 is the fastest algorithm which can be attributed to its non-iterative procedure. Furthermore, in general BIHT-ℓ_2 requires significantly longer time compared to the other algorithms. The BIHT, however, appears to be the fastest among the iterative algorithms at low sampling ratio or at large values of η. The GraSP variants generally run at similar speed, while they are faster than BIHT at low values of η and high sampling ratios.

VI. Conclusion

In this paper we revisited a formulation of the 1-bit CS problem based on the maximum likelihood estimation. Furthermore, we applied a variant of the GraSP algorithm [16] to this problem. We showed
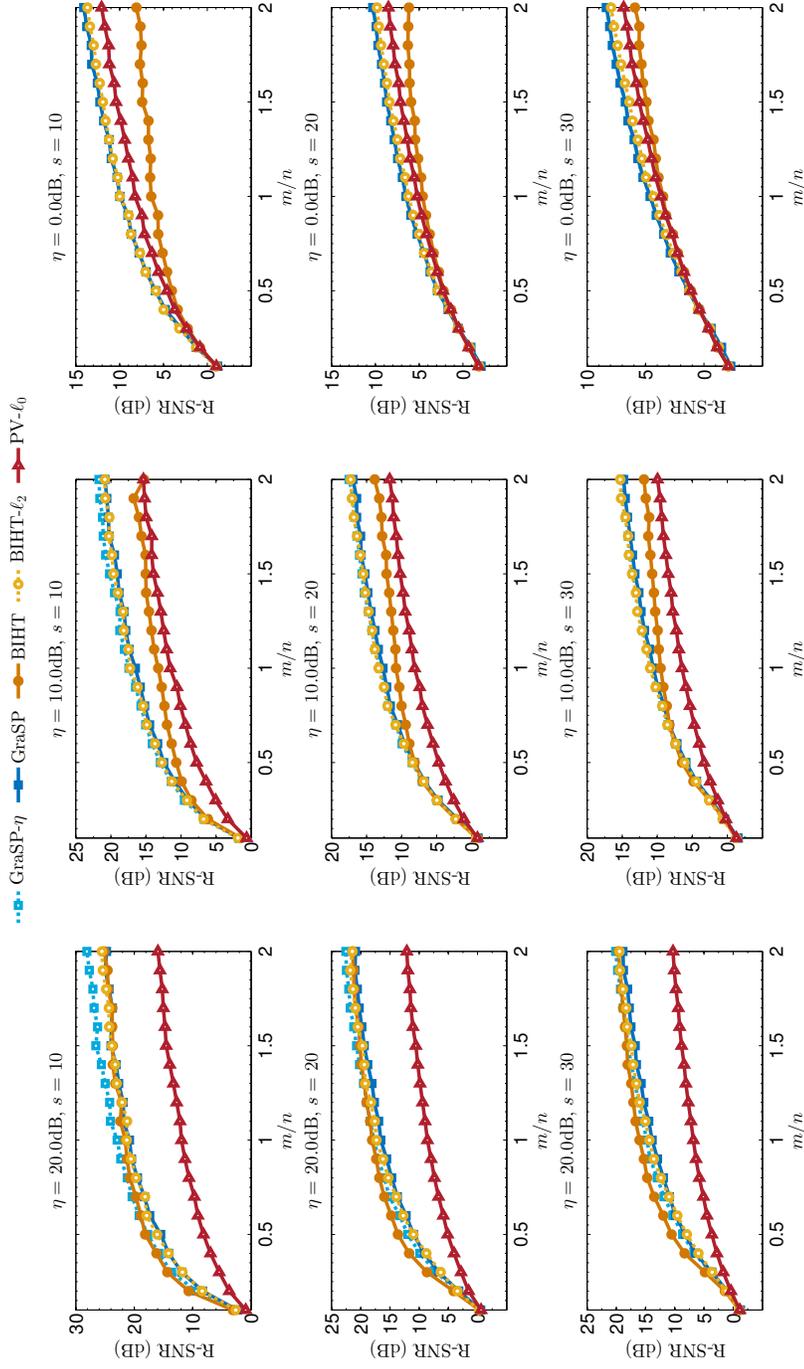
Figure 2.   Reconstruction SNR on the unit ball (R-SNR) vs. the sampling ratio (m/n) at different values of input SNR (η) and sparsity (s)
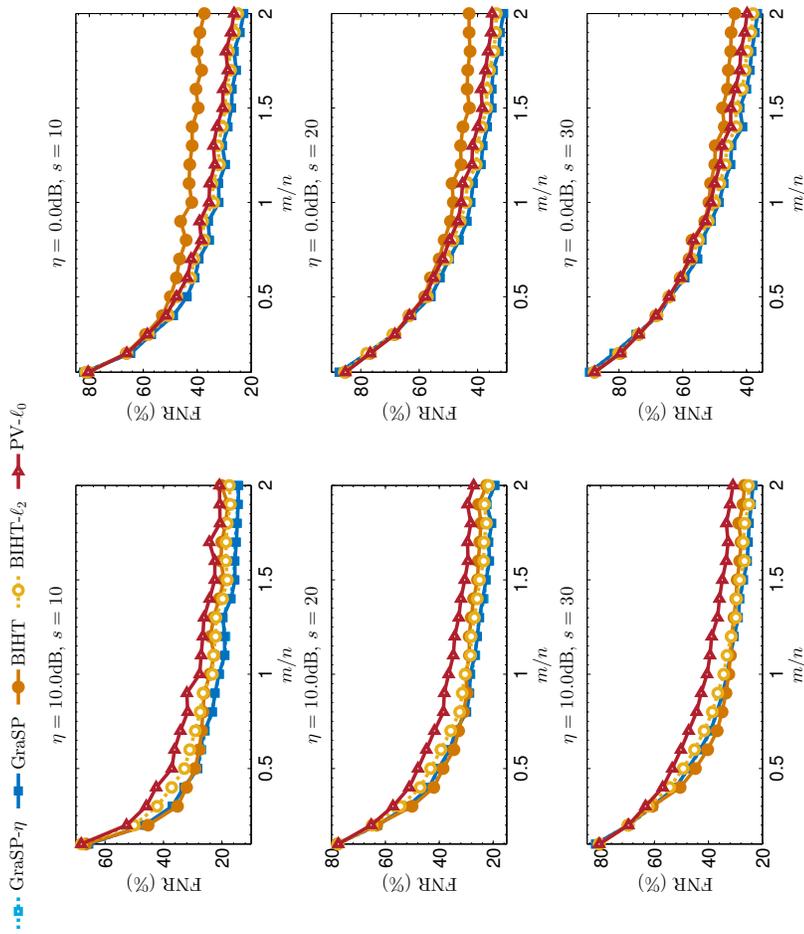
Figure 3. False Negative Rate (FNR) vs. the sampling ratio (m/n) at different values of input SNR (η) and sparsity (s)
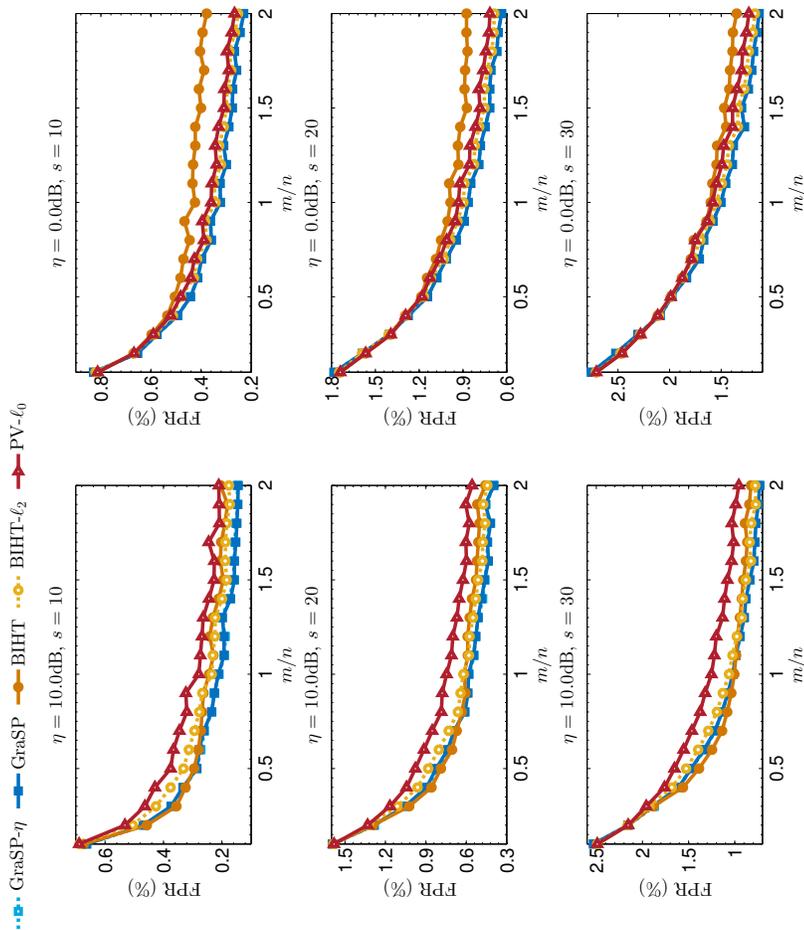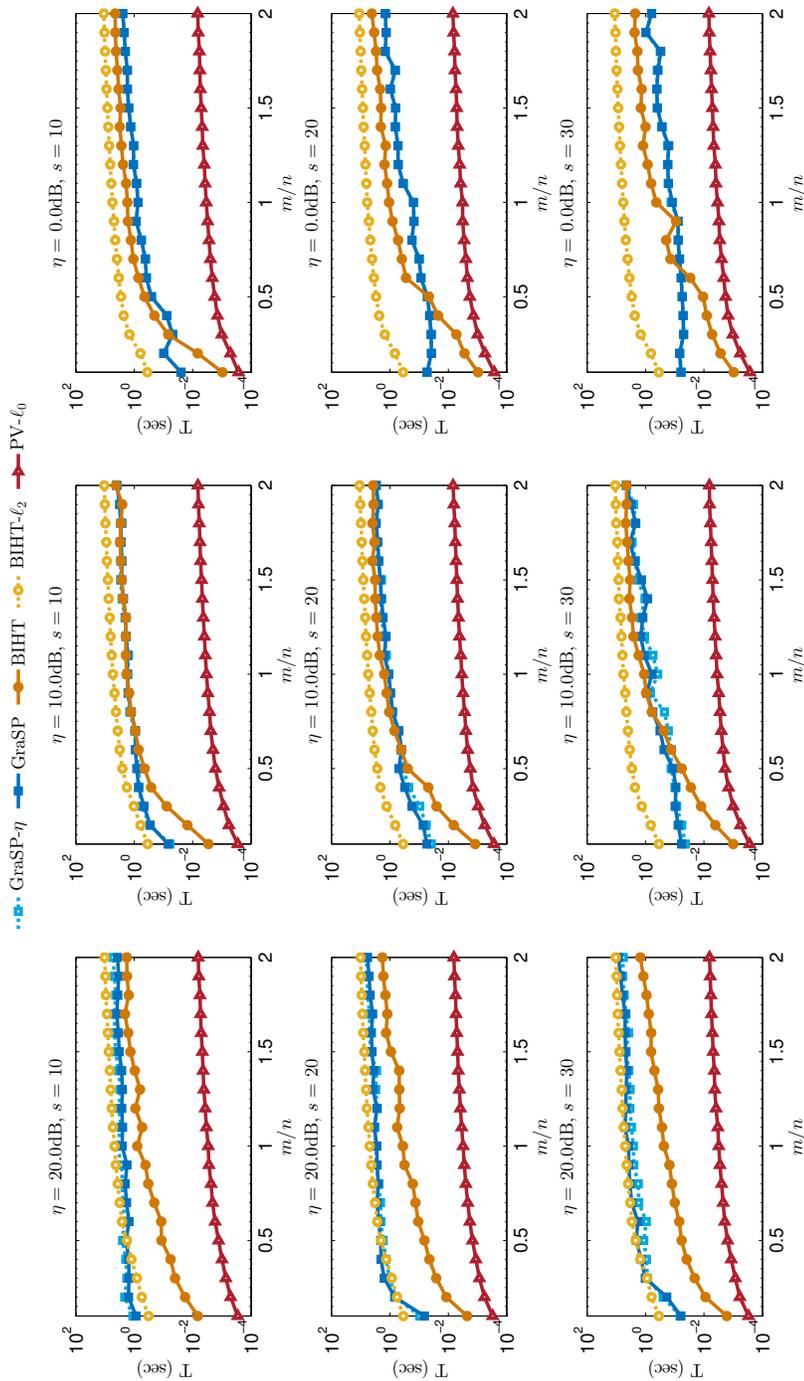
Figure 4. False Positive Rate (FPR) vs. the sampling ratio (m/n) at different values of input SNR (η) and sparsity (s)

13

Figure 5. Average execution time (T) vs. the sampling ratio (m/n) at different values of input SNR (η) and sparsity (s)

through numerical simulations that the proposed algorithms have robust performance in presence of noise. While at high levels of input SNR these algorithms are outperformed by a narrow margin by the competing algorithms, in low input SNR regime our algorithms show a solid performance at reasonable computational cost.

APPENDIX A

PROOFS

To prove Theorem 1 we use the following two lemmas. We omit the proofs since they can be easily adapted from Lemmas 1 and 2 in [16] using straightforward changes. It suffices to notice that

1) the proof in [16] still holds if the estimation errors are measured with respect to the true sparse minimizer or any other feasible (i.e., s-sparse) point, rather than the statistical true parameter, and

2) the iterates and the crude estimates will always remain in the sphere of radius r centered at the origin where the SRH applies.

In what follows $\int_0^1 \alpha_k (\tau \mathbf{x} + (1-\tau) \overline{\mathbf{x}}) \, \mathrm{d}\tau$ and $\int_0^1 \beta_k (\tau \mathbf{x} + (1-\tau) \overline{\mathbf{x}}) \, \mathrm{d}\tau$ are denoted by $\widetilde{\alpha}_k (\mathbf{x})$ and $\widetilde{\beta}_k (\mathbf{x})$, respectively. We also define $\widetilde{\gamma}_k (\mathbf{x}) := \widetilde{\alpha}_k (\mathbf{x}) - \widetilde{\beta}_k (\mathbf{x})$.

Lemma 1. *Let $\mathcal{Z}$ be the index set defined in Algorithm 1 and $\mathcal{R}$ denote the set* $\mathrm{supp} \left(\mathbf{x}^{(t)} - \overline{\mathbf{x}} \right)$. *Then the iterate $\mathbf{x}^{(t)}$ obeys*

$$\left\| \left(\mathbf{x}^{(t)} - \overline{\mathbf{x}} \right) \Big|_{\mathcal{Z}^c} \right\|_2 \leq \frac{\widetilde{\gamma}_{4s} \left(\mathbf{x}^{(t)} \right) + \widetilde{\gamma}_{2s} \left(\mathbf{x}^{(t)} \right)}{\widetilde{\beta}_{2s} \left(\mathbf{x}^{(t)} \right)} \left\| \mathbf{x}^{(t)} - \overline{\mathbf{x}} \right\|_2 + \frac{\left\| \nabla_{\mathcal{R} \setminus \mathcal{Z}} f \left(\overline{\mathbf{x}} \right) \right\|_2 + \left\| \nabla_{\mathcal{Z} \setminus \mathcal{R}} f \left(\overline{\mathbf{x}} \right) \right\|_2}{\widetilde{\beta}_{2s} \left(\mathbf{x}^{(t)} \right)}.$$

Lemma 2. *The vector $\mathbf{b}$ defined at line 3 of Algorithm 1 obeys*

$$\left\| \overline{\mathbf{x}}|_{\mathcal{T}} - \mathbf{b} \right\|_2 \leq \frac{\left\| \nabla_{\mathcal{T}} f \left(\overline{\mathbf{x}} \right) \right\|_2}{\widetilde{\beta}_{4s} \left(\mathbf{b} \right)} + \frac{\widetilde{\gamma}_{4s} \left(\mathbf{b} \right)}{2 \widetilde{\beta}_{4s} \left(\mathbf{b} \right)} \left\| \overline{\mathbf{x}}|_{\mathcal{T}^c} \right\|_2.$$

Proof of Theorem 1: Since $\mathcal{Z} \subseteq \mathcal{T}$ we have $\mathcal{T}^c \subseteq \mathcal{Z}^c$ and thus

$$\left\| \left(\mathbf{x}^{(t)} - \overline{\mathbf{x}} \right) \Big|_{\mathcal{Z}^c} \right\|_2 \geq \left\| \left(\mathbf{x}^{(t)} - \overline{\mathbf{x}} \right) \Big|_{\mathcal{T}^c} \right\|_2$$

$$= \left\| \overline{\mathbf{x}}|_{\mathcal{T}^c} \right\|_2.$$

Then it follows from Lemma 1 that

$$
\begin{aligned}
\left\| \overline{\mathbf{x}} \big|_{\mathcal{T}^c} \right\|_2 &\leq \frac{\widetilde{\gamma}_{4s}\left(\mathbf{x}^{(t)}\right)}{\widetilde{\beta}_{4s}\left(\mathbf{x}^{(t)}\right)} \left\| \mathbf{x}^{(t)} - \overline{\mathbf{x}} \right\|_2 \\
&\quad + \frac{\left\| \nabla_{\mathcal{R}\backslash\mathcal{Z}} f\left(\overline{\mathbf{x}}\right) \right\|_2 + \left\| \nabla_{\mathcal{Z}\backslash\mathcal{R}} f\left(\overline{\mathbf{x}}\right) \right\|_2}{\beta_{4s}} \\
&\leq \left(\mu_{4s} - 1\right) \left\| \mathbf{x}^{(t)} - \overline{\mathbf{x}} \right\|_2 + 2\epsilon,
\end{aligned}
\tag{5}
$$

where we used the fact that $\alpha_{4s} \geq \alpha_{2s}$ and $\beta_{4s} \leq \beta_{2s}$ to simplify the expressions. Furthermore, we have

$$
\begin{aligned}
\left\| \mathbf{x}^{(t+1)} - \overline{\mathbf{x}} \right\|_2 &= \left\| \mathbf{b}_s - \overline{\mathbf{x}} \right\|_2 \\
&\leq \left\| \mathbf{b}_s - \overline{\mathbf{x}}\big|_{\mathcal{T}} \right\|_2 + \left\| \overline{\mathbf{x}}\big|_{\mathcal{T}^c} \right\|_2 \\
&\leq \left\| \mathbf{b}_s - \mathbf{b} \right\|_2 + \left\| \mathbf{b} - \overline{\mathbf{x}}\big|_{\mathcal{T}} \right\|_2 + \left\| \overline{\mathbf{x}}\big|_{\mathcal{T}^c} \right\|_2 \\
&\leq 2\left\| \mathbf{b} - \overline{\mathbf{x}}\big|_{\mathcal{T}} \right\|_2 + \left\| \overline{\mathbf{x}}\big|_{\mathcal{T}^c} \right\|_2,
\end{aligned}
$$

where the last inequality holds because $\mathbf{b}_s$ is the best s-term approximation of $\mathbf{b}$. Hence, it follows from Lemma 2 that

$$
\begin{aligned}
\left\| \mathbf{x}^{(t+1)} - \overline{\mathbf{x}} \right\|_2 &\leq 2\frac{\left\| \nabla_{\mathcal{T}} f\left(\overline{\mathbf{x}}\right) \right\|_2}{\widetilde{\beta}_{4s}\left(\mathbf{b}\right)} + \frac{\widetilde{\alpha}_{4s}\left(\mathbf{b}\right)}{\widetilde{\beta}_{4s}\left(\mathbf{b}\right)} \left\| \overline{\mathbf{x}}\big|_{\mathcal{T}^c} \right\|_2 \\
&\leq 2\epsilon + \mu_{4s} \left\| \overline{\mathbf{x}}\big|_{\mathcal{T}^c} \right\|_2.
\end{aligned}
$$

Then applying (5) and simplifying the resulting inequality yield

$$
\begin{aligned}
\left\| \mathbf{x}^{(t+1)} - \overline{\mathbf{x}} \right\|_2 &\leq 2\epsilon + \mu_{4s} \left(\left(\mu_{4s} - 1\right) \left\| \mathbf{x}^{(t)} - \overline{\mathbf{x}} \right\|_2 + 2\epsilon \right) \\
&\leq \left(\mu_{4s}^2 - \mu_{4s}\right) \left\| \mathbf{x}^{(t)} - \overline{\mathbf{x}} \right\|_2 + 2\left(\mu_{4s} + 1\right)\epsilon,
\end{aligned}
$$

which is the desired result. ∎

Lemma 3 (Bounded Sparse Projection). *For any $\mathbf{x} \in \mathbb{R}^n$ the vector $\max\left\{1, \frac{r}{\|\mathbf{x}_s\|_2}\right\} \mathbf{x}_s$ is a solution to the minimization*

$$
\arg\min_{\mathbf{w}} \ \left\| \mathbf{x} - \mathbf{w} \right\|_2 \quad \text{s.t.} \ \left\| \mathbf{w} \right\|_2 \leq r \ \text{and} \ \left\| \mathbf{w} \right\|_0 \leq s.
\tag{6}
$$

Proof: Given an index set $\mathcal{S} \subseteq [n]$ we can write $\left\| \mathbf{x} - \mathbf{w} \right\|_2^2 = \left\| \mathbf{x} - \mathbf{w}\big|_{\mathcal{S}} \right\|_2^2 + \left\| \mathbf{x}\big|_{\mathcal{S}^c} \right\|_2^2$ for vectors $\mathbf{w}$ with $\mathrm{supp}\left(\mathbf{w}\right) \subseteq \mathcal{S}$. Therefore, the solution to

$$
\arg\min_{\mathbf{w}} \ \left\| \mathbf{x} - \mathbf{w} \right\|_2 \quad \text{s.t.} \ \left\| \mathbf{w} \right\|_2 \leq r \ \text{and} \ \mathrm{supp}\left(\mathbf{w}\right) \subseteq \mathcal{S}
$$

is simply obtained by projection of $\mathbf{x}|_{\mathcal{S}}$ onto the sphere of radius r, i.e.,

$$\mathrm{P}_{\mathcal{S}}\left(\mathbf{x}\right) = \max\left\{1, \frac{r}{\|\mathbf{x}|_{\mathcal{S}}\|_2}\right\} \mathbf{x}|_{\mathcal{S}}.$$

Therefore, to find a solution to (6) it suffices to find the index set $\mathcal{S}$ with $|\mathcal{S}| = s$ and thus the corresponding $\mathrm{P}_{\mathcal{S}}\left(\mathbf{x}\right)$ that minimize $\|\mathbf{x} - \mathrm{P}_{\mathcal{S}}\left(\mathbf{x}\right)\|_2$. Note that we have

$$\|\mathbf{x} - \mathrm{P}_{\mathcal{S}}\left(\mathbf{x}\right)\|_2^2 = \|\mathbf{x}|_{\mathcal{S}} - \mathrm{P}_{\mathcal{S}}\left(\mathbf{x}\right)\|_2^2 + \|\mathbf{x}|_{\mathcal{S}^c}\|_2^2$$

$$= \left(\|\mathbf{x}|_{\mathcal{S}}\|_2 - r\right)_+^2 + \|\mathbf{x}|_{\mathcal{S}^c}\|_2^2$$

$$= \begin{cases} \|\mathbf{x}\|_2^2 - \|\mathbf{x}|_{\mathcal{S}}\|_2^2 & , \|\mathbf{x}|_{\mathcal{S}}\|_2 < r \\ \|\mathbf{x}\|_2^2 + r^2 - 2r\|\mathbf{x}|_{\mathcal{S}}\|_2 & , \|\mathbf{x}|_{\mathcal{S}}\|_2 \geq r \end{cases}.$$

For all valid $\mathcal{S}$ with $\|\mathbf{x}|_{\mathcal{S}}\|_2 < r$ we have $\|\mathbf{x}\|_2^2 - \|\mathbf{x}|_{\mathcal{S}}\|_2^2 > \|\mathbf{x}\|_2^2 - r^2$. Similarly, for all valid $\mathcal{S}$ with $\|\mathbf{x}|_{\mathcal{S}}\|_2 < r$ we have $\|\mathbf{x}\|_2^2 + r^2 - 2r\|\mathbf{x}|_{\mathcal{S}}\|_2 \leq \|\mathbf{x}\|_2^2 - r^2$. Furthermore, both $\|\mathbf{x}\|_2^2 - \|\mathbf{x}|_{\mathcal{S}}\|_2^2$ and $\|\mathbf{x}\|_2^2 + r^2 - 2r\|\mathbf{x}|_{\mathcal{S}}\|_2$ are decreasing functions of $\|\mathbf{x}|_{\mathcal{S}}\|_2$. Therefore, $\|\mathbf{x} - \mathrm{P}_{\mathcal{S}}\left(\mathbf{x}\right)\|_2^2$ is a decreasing function of $\|\mathbf{x}|_{\mathcal{S}}\|_2$. Hence, $\|\mathbf{x} - \mathrm{P}_{\mathcal{S}}\left(\mathbf{x}\right)\|_2$ attains its minimum at $\mathcal{S} = \mathrm{supp}\left(\mathbf{x}_s\right)$. ■

APPENDIX B

ON NON-CONVEX FORMULATION OF [8]

[8] derived accuracy guarantees for

$$\arg\max_{\mathbf{x}} \ \langle \mathbf{y}, \mathbf{A}\mathbf{x} \rangle \quad \text{s.t. } \mathbf{x} \in \mathcal{K}$$

as a solver for the 1-bit CS problem, where $\mathcal{K}$ is a subset of the unit Euclidean ball. While their result [8, Theorem 1.1] applies to both convex and non-convex sets $\mathcal{K}$, the focus of their work has been on the set $\mathcal{K}$ that is the intersection of a centered ℓ_1-ball and the unit Euclidean ball. Our goal, however, is to examine the other interesting choice of $\mathcal{K}$, namely the intersection of canonical sparse subspaces and the unit Euclidean ball. The estimator in this case can be written as

$$\arg\max_{\mathbf{x}} \ \langle \mathbf{y}, \mathbf{A}\mathbf{x} \rangle \quad \text{s.t. } \|\mathbf{x}\|_0 \leq s \text{ and } \|\mathbf{x}\|_2 \leq 1. \tag{7}$$

We show that a solution to the optimization above can be obtained explicitly.

Lemma 4. *A solution to (7) is* $\widehat{\mathbf{x}} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{y}\right)_s / \left\|\left(\mathbf{A}^{\mathrm{T}}\mathbf{y}\right)_s\right\|_2$.

Proof: For $\mathcal{I} \subseteq [n]$ define

$$\widehat{\mathbf{x}}\left(\mathcal{I}\right) := \arg\max_{\mathbf{x}} \ \langle \mathbf{y}, \mathbf{A}\mathbf{x} \rangle \quad \text{s.t. } \mathbf{x}|_{\mathcal{I}^c} = 0 \text{ and } \|\mathbf{x}\|_2 \leq 1.$$

Furthermore, choose

$$\widehat{\mathcal{I}} \in \arg\max_{\mathcal{I}} \ \langle \mathbf{y}, \mathbf{A}\widehat{\mathbf{x}}(\mathcal{I}) \rangle \quad \text{s.t. } \mathcal{I} \subseteq [n] \ \text{and} \ |\mathcal{I}| \le s.$$

Then $\widehat{\mathbf{x}}\left(\widehat{\mathcal{I}}\right)$ would be a solution to (7). Using the fact that $\langle \mathbf{y}, \mathbf{A}\mathbf{x} \rangle = \langle \mathbf{A}^{\mathrm{T}}\mathbf{y}, \mathbf{x} \rangle$, straightforward application of Cauchy-Schwarz inequality shows that $\widehat{\mathbf{x}}(\mathcal{I}) = \left.\left(\mathbf{A}^{\mathrm{T}}\mathbf{y}\right)\right|_{\mathcal{I}} / \left\|\left.\left(\mathbf{A}^{\mathrm{T}}\mathbf{y}\right)\right|_{\mathcal{I}}\right\|_2$ for which we have

$$\langle \mathbf{y}, \mathbf{A}\widehat{\mathbf{x}}(\mathcal{I}) \rangle = \left\|\left.\left(\mathbf{A}^{\mathrm{T}}\mathbf{y}\right)\right|_{\mathcal{I}}\right\|_2.$$

Thus, we obtain $\widehat{\mathcal{I}} = \operatorname{supp}\left(\left(\mathbf{A}^{\mathrm{T}}\mathbf{y}\right)_s\right)$ and thereby $\widehat{\mathbf{x}}\left(\widehat{\mathcal{I}}\right) = \widehat{\mathbf{x}}$, which proves the claim. ∎

REFERENCES

[1] J. N. Laska, P. T. Boufounos, and R. G. Baraniuk, "Finite range scalar quantization for compressive sensing," in *Proceedings of International Conferenece on Sampling Theory and Applications (SampTA)*, Toulouse, France, May 18-22 2009.

[2] W. Dai, H. V. Pham, and O. Milenkovic, "Distortion-rate functions for quantized compressive sensing," in *IEEE Information Theory Workshop on Networking and Information Theory, 2009. ITW 2009*, Jun. 2009, pp. 171 –175.

[3] J. Sun and V. Goyal, "Optimal quantization of random measurements in compressed sensing," in *IEEE International Symposium on Information Theory, 2009. ISIT 2009*, Jul. 2009, pp. 6 –10.

[4] A. Zymnis, S. Boyd, and E. Candès, "Compressed sensing with quantized measurements," *IEEE Signal Processing Letters*, vol. 17, no. 2, pp. 149 –152, Feb. 2010.

[5] L. Jacques, D. Hammond, and J. Fadili, "Dequantizing compressed sensing: when oversampling and non-gaussian constraints combine," *IEEE Transactions on Information Theory*, vol. 57, no. 1, pp. 559 –571, Jan. 2011.

[6] J. N. Laska, P. T. Boufounos, M. A. Davenport, and R. G. Baraniuk, "Democracy in action: Quantization, saturation, and compressive sensing," *Applied and Computational Harmonic Analysis*, vol. 31, no. 3, pp. 429–443, Nov. 2011.

[7] P. Boufounos and R. Baraniuk, "1-bit compressive sensing," in *Information Sciences and Systems, 2008. CISS 2008. 42nd Annual Conference on*, Mar. 2008, pp. 16 –21.

[8] Y. Plan and R. Vershynin, "Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach," *IEEE Transactions on Information Theory*, vol. 59, no. 1, pp. 482–494, Jan. 2013.

[9] P. Boufounos, "Greedy sparse signal reconstruction from sign measurements," in *Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers, 2009*, Nov. 2009, pp. 1305 –1309.

[10] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[11] J. Laska, Z. Wen, W. Yin, and R. Baraniuk, "Trust, but verify: Fast and accurate signal recovery from 1-bit compressive measurements," *IEEE Transactions on Signal Processing*, vol. 59, no. 11, pp. 5289 –5301, Nov. 2011.

[12] L. Jacques, J. Laska, P. Boufounos, and R. Baraniuk, "Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors," *IEEE Transactions on Information Theory*, vol. PP, 2013, `arXiv:1104.3160 [cs.IT]`.

[13] M. Yan, Y. Yang, and S. Osher, "Robust 1-bit compressive sensing using adaptive outlier pursuit," *Signal Processing, IEEE Transactions on*, vol. 60, no. 7, pp. 3868 –3875, Jul. 2012.

[14] Y. Plan and R. Vershynin, "One-bit compressed sensing by linear programming," *Comm. Pure Appl. Math.*

[15] A. Ai, A. Lapanowski, P. Yaniv, and R. Vershynin, "One-bit compressed sensing with non-gaussian measurements," *Linear Algebra and its Applications*, 2013, to appear.

[16] S. Bahmani, B. Raj, and P. Boufounos, "Greedy sparsity-constrained optimization," *Journal of Machine Learning Research*, 2012, to appear. arxiv:1203.5483v1 [stat.ML].