

Compressed Sensing with Non-Gaussian Noise and Partial Support Information

Ahmad Abou Saleh, Fady Alajaji, and Wai-Yip Chan

Abstract—We study the problem of recovering sparse and compressible signals using a weighted ℓ_p minimization with $0 < p \leq 1$ from noisy compressed sensing measurements when part of the support is known a priori. To better model different types of non-Gaussian (bounded) noise, the minimization program is subject to a data-fidelity constraint expressed as the ℓ_q ($2 \leq q < \infty$) norm of the residual error. We show theoretically that the reconstruction error of this optimization is bounded (stable) if the sensing matrix satisfies an extended restricted isometry property. Numerical results show that the proposed method, which extends the range of p and q comparing with previous works, outperforms other noise-aware basis pursuit programs. For $p < 1$, since the optimization is not convex, we use a variant of an iterative reweighted ℓ_2 algorithm for computing a local minimum.

Index Terms—Compressed sensing, denoising, nonconvex optimization, sparsity, weighted ℓ_p minimization.

I. INTRODUCTION

THE theory of compressed sensing (CS) is to reliably recover high dimensional sparse or compressible signals from a significantly fewer number of linear measurements than the signal space dimension [1], [2]. Consider a signal \mathbf{x} of dimension N to be sparse in some orthonormal basis $\Psi \in \mathbb{R}^{N \times N}$; in other words, $\mathbf{x} = \Psi \mathbf{u}$, where \mathbf{u} is a K -sparse vector, i.e., there are only K ($K \ll N$) nonzero elements in \mathbf{u} . In CS, the encoder records $M < N$ linear measurements given by $\mathbf{y} = \Phi \mathbf{x}$, where $\Phi \in \mathbb{R}^{M \times N}$ is a measurement matrix that satisfies the restricted isometry property (RIP); Φ satisfies the RIP of order K if there exists $\delta_K \in [0, 1]$ such that, for any K -sparse \mathbf{u} , $(1 - \delta_K) \|\mathbf{u}\|_2^2 \leq \|\Phi \mathbf{u}\|_2^2 \leq (1 + \delta_K) \|\mathbf{u}\|_2^2$, where δ_K is the RIP constant of order K and $\|\cdot\|_2$ denotes the ℓ_2 norm. Roughly speaking, the RIP condition states that any (sufficient) small subset of the columns of the sensing matrix “nearly” forms an orthogonal basis. The condition is satisfied when the elements of the matrix Φ are samples from an independent and identically distributed (i.i.d.) Gaussian process where M is of the order of $K \log_2(N/K)$. In this case, recovery of \mathbf{x}

from \mathbf{y} works by solving the convex basis pursuit (BP) program: $\min_{\hat{\mathbf{x}}} \|\Psi^T \hat{\mathbf{x}}\|_1$ subject to $\Phi \hat{\mathbf{x}} = \mathbf{y}$, where $\|\cdot\|_1$ denotes the ℓ_1 norm ($\|\mathbf{x}\|_1 \triangleq \sum_{i=1}^N |x_i|$) and $(\cdot)^T$ is the transpose operator [3]. In practice, the collected measurements are disturbed by noise \mathbf{n} (i.e., $\hat{\mathbf{y}} = \Phi \mathbf{x} + \mathbf{n}$). Using a noise-aware version of BP, the signal is decoded by solving the convex basis pursuit denoising (BPDN) program: $\min_{\hat{\mathbf{x}}} \|\Psi^T \hat{\mathbf{x}}\|_1$ subject to $\|\Phi \hat{\mathbf{x}} - \hat{\mathbf{y}}\|_2 \leq \epsilon'$, where ϵ' bounds the total amount of noise in the measurements.

In many applications, it is possible to obtain some estimate about the support of the signal. To incorporate such prior information, several modified versions of the BPDN were studied in [4]–[6] based on a weighted ℓ_1 minimization approach with zero weights on the known support. In [7], the author extends the work of [5] to the case of compressible signals and noisy measurements. In [8], an RIP-based condition for recovering compressively sampled signals using partial support information is derived. On the other hand, several works have attempted to close the gap in performance from using ℓ_1 instead of ℓ_0 (combinatorial problem) minimization by solving a nonconvex ℓ_p ($0 < p < 1$) minimization. In [9]–[11], the authors provide sufficient conditions for recoverability of sparse/compressible signals from noisy measurements via such ℓ_p minimization. In [12], the authors study weighted ℓ_p minimization with zero weights on the known support.

In most previous works, the CS recovery program uses an ℓ_2 norm fidelity constraint. From a Bayesian standpoint, the ℓ_2 norm used in the fidelity term (of the BPDN) corresponds to the conditional log-likelihood associated with additive white Gaussian noise. However, the noise in the disturbed measurement might not be Gaussian. As an example, quantization distortion is highly non-Gaussian and is essentially bounded by the quantization bin widths. This motivates the need for a new CS decoder that better models different types of noise. The authors in [13] adapt BPDN to integrate the quantization distortion constraint. In [14], the authors study a new class of optimization problems coined basis pursuit dequantizing (BPDQ) that models more faithfully different types of noise than BPDN; this new class of optimization proceeds by minimizing the sparsity of the signal subject to a data-fidelity constraint expressed in the ℓ_q ($2 \leq q < \infty$) norm of the residual error instead of the ℓ_2 norm as in BPDN; the case of $q = 1$ is studied in [15]. The BPDQ of moment q , BPDQ $_q$, is

$$\min_{\hat{\mathbf{x}}} \|\Psi^T \hat{\mathbf{x}}\|_1, \quad \text{subject to } \|\Phi \hat{\mathbf{x}} - \hat{\mathbf{y}}\|_q \leq \tilde{\epsilon} \quad (1)$$

where $\|\mathbf{x}\|_q \triangleq \sum_{i=1}^N |x_i|^q$ and $\tilde{\epsilon}$ bounds the measurement noise. The authors in [14] study the problem in (1) and show that this approach gives some gain over BPDN. However, no prior information on the unknown signal is assumed in [14]. Motivated by

Manuscript received January 22, 2015; accepted April 20, 2015. Date of publication April 27, 2015; date of current version May 01, 2015. This work was supported in part by the NSERC of Canada. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Alex Dimakis.

A. Abou Saleh and W.-Y. Chan are with the Department of Electrical and Computer Engineering, Queen's University, Kingston, ON K7L 3N6 Canada (e-mail: ahmad.abou.saleh@queensu.ca; chan@queensu.ca).

F. Alajaji is with the Department of Mathematics and Statistics and the Department of Electrical and Computer Engineering, Queen's University, Kingston, ON K7L 3N6 Canada (e-mail: fady@mast.queensu.ca).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/LSP.2015.2426654

the fact that the CS algorithm may have prior knowledge about the signal and that disturbance in the measurements might come from quantization distortion, we study weighted ℓ_p ($0 < p \leq 1$) minimization subject to an ℓ_q ($2 \leq q < \infty$) norm fidelity constraint for compressively sampled signals with (not fully accurate) partial support information; we develop RIP-based guarantee conditions for stable recovery. This generalizes the problems in [8], [14] and [16] to a wider range of p and q . We also discuss the implementation aspect for the algorithm for $p < 1$. In this case, the problem is no longer convex; we resort to the use of an iterative reweighted convex problem to solve the original nonconvex optimization.

II. WEIGHTED BASIS PURSUIT DEQUANTIZING

In this section, we present a CS recovery method that incorporates partial knowledge about the sparsity support and models more faithfully different types of measurement noise. Next we assume a canonical basis (Ψ is the identity matrix).

Definition 1: Φ satisfies the extended $RIP_{q,2}$ of order K if $\mu_{q,2}(1 - \delta_K)^{1/2} \|\mathbf{u}\|_2 \leq \|\Phi \mathbf{u}\|_q \leq \mu_{q,2}(1 + \delta_K)^{1/2} \|\mathbf{u}\|_2$, for all K -sparse signals \mathbf{u} , a $\mu_{q,2} > 0$ and an $RIP_{q,2}$ constant $\delta_K \in [0, 1]$ [14]. Note that $\mu_{q,2}$ is equal to 1 for $q = 2$.

Consider a compressible signal $\mathbf{x} \in \mathbb{R}^N$. Let \mathbf{x}_K be K -sparse where the K nonzero coefficients of \mathbf{x}_K are equal to the K largest-in-magnitude coefficients of \mathbf{x} ; we refer to \mathbf{x}_K as the best K term approximation of \mathbf{x} . Let $T_0 = \text{supp}(\mathbf{x}_K)$ be the support of \mathbf{x}_K (the locations of the nonzero coefficients), where $T_0 \subseteq \{1, \dots, N\}$. Let the support estimate \tilde{T} be a subset of $\{1, \dots, N\}$ with cardinality $|\tilde{T}| = \rho K$, where $0 < \rho \leq a$ for some $a \geq 1$, and $\mathbf{x}_{\tilde{T}}$ be a sub-vector of \mathbf{x} with coefficient $\mathbf{x}_{\tilde{T}}(i) = \mathbf{x}(\tilde{T}(i))$, $i = 1, \dots, \rho K$. The generalized recovery process, coined the weighted BPDQ $_{p,q}$ (WBPDQ $_{p,q}$), is given by

$$\min_{\hat{\mathbf{x}}} \|\Psi^T \hat{\mathbf{x}}\|_{p,w}, \quad \text{subject to } \|\Phi \hat{\mathbf{x}} - \hat{\mathbf{y}}\|_q \leq e \quad (2)$$

where $\|\mathbf{x}\|_{p,w} = (\sum_{i=1}^N w_i^p |x_i|^p)^{1/p}$, $w_i = 1$ for $i \in \tilde{T}^c$, $w_i = w$ for $i \in \tilde{T}$, $0 \leq w \leq 1$, $0 < p \leq 1$, $q \geq 2$ and e bounds the total amount of noise in the measurements. Next, we show that WBPDQ $_{p,q}$ is stable if Φ satisfies an $RIP_{q,2}$.

Theorem 1: Define ρ and α so that $|\tilde{T}| = \rho K$, $|\tilde{T} \cap T_0| = \alpha \rho K$ and $\tilde{T}_\alpha = T_0 \cap \tilde{T}$. Suppose that there exists an $a \geq (1 - \alpha)\rho$, $a \geq 1$ and that Φ satisfies $RIP_{q,2}$ with

$$(1 - \delta_{(a+1)K})^p - c_q^p \gamma > 0 \quad (3)$$

where $\gamma = w^p + (1 - w^p)a^{\frac{p}{2}-1}(1 + \rho - 2\alpha\rho)^{1-\frac{p}{2}}$ and $c_q = C_q(\Phi, (1+a)K, aK)$ with $C_q(\Phi, K, K')$ given by min of

$$\left[(\delta_K + \delta_{K+K'}) (\delta'_K + \delta_{K+K'} + \bar{q}(1 + \delta'_K)) \right]^{\frac{1}{2}}, \quad (4)$$

$$\left[\left(\delta_{K+K'} + \bar{q} \frac{1 + \delta_{K+K'}}{2} \right) \left(\delta_{K+K'} + \bar{q} \frac{2 + \delta'_K + \delta_{K+K'}}{2} \right) \right]^{\frac{1}{2}}$$

and $\bar{q} = q - 2 \geq 0$. Then the solution $\hat{\mathbf{x}}$ to (2) satisfies

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2^p \leq C_0(e^p/\mu_{q,2}^p) + C_1(\tilde{w}K^{\frac{p}{2}-1}) \quad (5)$$

where

$$C_0 = \frac{2^p(1 + \gamma)(1 + \delta_{(a+1)K})^{p/2}}{[(1 - \delta_{(a+1)K})^p - c_q^p \gamma]}$$

$$C_1 = \frac{2c_q^p(1 + \gamma)(a)^{\frac{p}{2}-1}}{[(1 - \delta_{(a+1)K})^p - c_q^p \gamma]} + 2(a)^{\frac{p}{2}-1} \quad (6)$$

and $\tilde{w} = (w^p \|\mathbf{x} - \mathbf{x}_K\|_p^p + (1 - w^p) \|\mathbf{x}_{\tilde{T}^c \cap T_0^c}\|_p^p)$. The first term in the right hand side of (5) depends on the measurement matrix and the measurement noise; the second term, which also depends on the measurement matrix, comes from the compressibility of the signal.

The proof of Theorem 1 is presented in the Appendix; the proof uses a combination of techniques used to prove [2, Theorem 2], [14, Theorem 2] and [8, Theorem 3].

Remark 1: For $w = 1$, $a = 1$ and $p = 1$, Theorem 1 reduces to Theorem 2 in [14]. For $\alpha = 0.5$, $a = 1$ and $p = 1$, the sufficient condition and the bound on the reconstruction error given by Theorem 1 is identical to that of Theorem 2 in [14]; this means that choosing $w = 1$ in (2) gives the best performance for $\alpha = 0.5$.

Next, we discuss some implementation aspects of the optimization problem given by (2). For the case of $p = 1$, the WBPDQ $_{1,q}$ program is convex and can be solved using the toolbox in [17]. For $0 < p < 1$, however, the optimization is no longer convex. Algorithms that have been used to solve such optimization include the projected gradient method [9], [16], the iterative reweighted ℓ_1 method [18] and the iterative reweighted least squares method [19], [20]. We herein use a variant of the iteratively reweighted ℓ_2 algorithm to compute a local minimum. The approach is to replace the weighted ℓ_p objective function in (2) by a weighted ℓ_2 norm as follows

$$\hat{\mathbf{x}}^{(t+1)} = \min_{\hat{\mathbf{x}}} \sum_{i=1}^N m_i^{(t)} \hat{x}_i^2, \quad \text{subject to } \|\Phi \hat{\mathbf{x}} - \hat{\mathbf{y}}\|_q \leq e \quad (7)$$

where $m_i^{(t)} = w_i^p((\hat{x}_i^{(t)})^2 + \epsilon)^{p/2-1}$; the weight $m_i^{(t)}$ is updated at each iteration t and $\epsilon > 0$ is a regularization factor. This process is repeated with decreasing value of ϵ until convergence. Since $p/2 - 1 < 0$, allowing ϵ to be zero, the weights in (7) are undefined whenever the value of the signal is zero ($\hat{x}_i^{(t)} = 0$). To overcome this, we use the approach of regularizing the optimization problem by incorporating a small $\epsilon > 0$. Algorithm 1 is stated as follows

- 1) Set $t = 0$, $\epsilon = \epsilon_0$ and start with an initial solution $\hat{\mathbf{x}}^{(0)}$.
- 2) Set $t = t + 1$; solve for $\hat{\mathbf{x}}^{(t)}$ using (7).
- 3) Decrease ϵ as follow: $\epsilon = \epsilon/10$.
- 4) Evaluate the convergence function $J = \frac{\|\hat{\mathbf{x}}^{(t)} - \hat{\mathbf{x}}^{(t-1)}\|^2}{\|\hat{\mathbf{x}}^{(t-1)}\|^2}$. If J is less than some positive threshold J_{th} or $t > T_{max}$ stop iterating. Else go to step 2.

In our simulation, we solve (7) using the toolbox in [17]; we set $J_{th} = 10^{-5}$, $T_{max} = 15$, $\epsilon_0 = 1$ and $\hat{\mathbf{x}}^{(0)}$ to be the solution of the problem for $p = 1$.

Remark 2: We also implemented a variant of the iterative reweighted ℓ_1 method in [18] to solve the nonconvex problem; we replace the objective function in (7) in Algorithm 1 with a weighted ℓ_1 function $\sum_{i=1}^N m_i^{(t)} |\hat{x}_i|$, where $m_i^{(t)} = w_i^p(|\hat{x}_i^{(t)}| + \epsilon)^{p-1}$. Simulation results show that this method gives similar performance as the reweighted ℓ_2 norm method.

III. NUMERICAL RESULTS

We consider \mathbf{x} to be sparse in the canonical basis where the nonzero elements in \mathbf{x} are chosen to be i.i.d. zero-mean

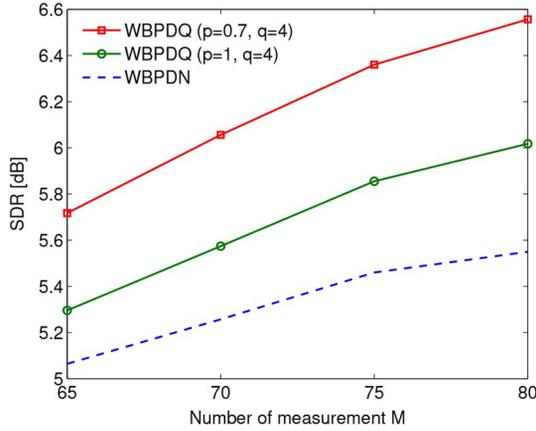


Fig. 1. Performance of $\text{WBPdq}_{p,q}$ with uniform noise for $N = 150$ and $K = 15$; the support knowledge has $|\tilde{T}| = 10$ with 80% accuracy rate. The noise is uniform on $[-0.5, 0.5]$.

Gaussian with unit variance and the signal support is uniformly distributed. The support knowledge \tilde{T} is uniformly chosen with accuracy rate α . Fig. 1 shows the $\text{SDR} \triangleq \frac{\mathbb{E}[\|\mathbf{x}\|^2]}{\mathbb{E}[\|\mathbf{x}-\tilde{\mathbf{x}}\|^2]}$ performance versus the number of measurements M for an accuracy rate of 80% (i.e., $\alpha = 0.8$). We assume that the noise \mathbf{n} is uniform on the interval $[-0.5, 0.5]$ and choose a weight parameter of $w = 0$ in (2). We notice that the $\text{WBPdq}_{p,q}$ ($p = 0.7, q = 4$), solved using Algorithm 1, gives the best performance. Comparing $\text{WBPdq}_{p,q}$ to $\text{WBPdq}_{1,2}$ (i.e., WBPdN), we notice that the SDR gap increases with M . This is from the fact that when M is low, we operate in the region where the constraint on the $\text{RIP}_{q,2}$ constant (3) in Theorem 1 is barely met and hence the benefit of using the ℓ_q norm constraint over the ℓ_2 is not as high as larger M . Although Fig. 1 aims to show the performance of different recovery algorithms over M , it is worth mentioning that the value of q used is not the optimal one.

Fig. 2 shows the SDR performance of $\text{WBPdq}_{p,q}$ ($p = 1, q = 4$) versus w for different accuracy rates α and uniform noise on $[-0.5, 0.5]$. We remark that for low accuracy α , it is beneficial to increase w ; for high accuracy, however, using a lower w gives a better performance. Comparing the gap between $\text{WBPdq}_{1,4}$ and WBPdN over several accuracy rates, we notice that this gap gets wider for higher accuracy. This can be explained from the fact that for higher accuracy, the cardinality of the unknown sparsity pattern is lower and hence less measurements are needed to satisfy the $\text{RIP}_{q,2}$ constraint of Theorem 1; as a result, the benefit of using ℓ_q norm is more visible. Note that for the extreme case of $w = 1$, the SDR gap between $\text{WBPdq}_{1,4}$ and WBPdN is small; this is from the fact that no side information is used at the decoder and since M is relatively small compared to K . Fig. 3 shows the SDR performance as a function of $|\tilde{T}|$ for different values of accuracy α and weight w . We can notice that using a larger support estimate favours better performance for most cases. The only case where increasing $|\tilde{T}|$ decreases the SDR performance is when choosing $w = 0$ (very small) for low accuracy rate α . This can be explained from the fact that for low accuracy α , increasing $|\tilde{T}|$ leads to a relatively more increase in the amount of wrong indices than true ones; as a result, choosing $w = 0$ (i.e., favouring the indices in \tilde{T}) leads to a decrease in the performance. Moreover, the drastic increase

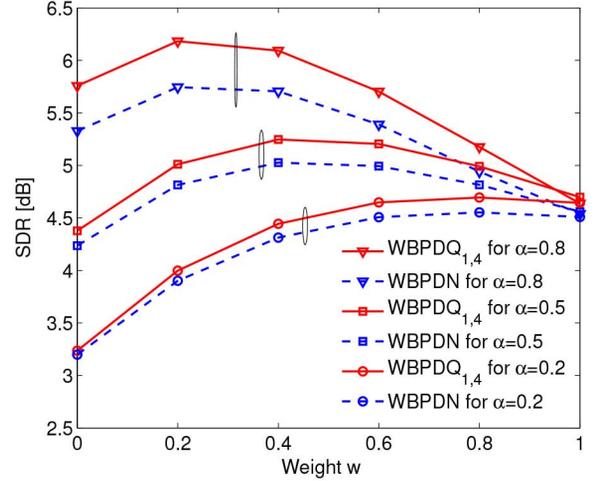


Fig. 2. Performance of the $\text{WBPdq}_{1,q}$ (for $q = 4$) versus w for $N = 150$, $K = 15$ and $M = 75$. The support knowledge has $|\tilde{T}| = 10$ with 80% accuracy rate and the noise is uniform on $[-0.5, 0.5]$.

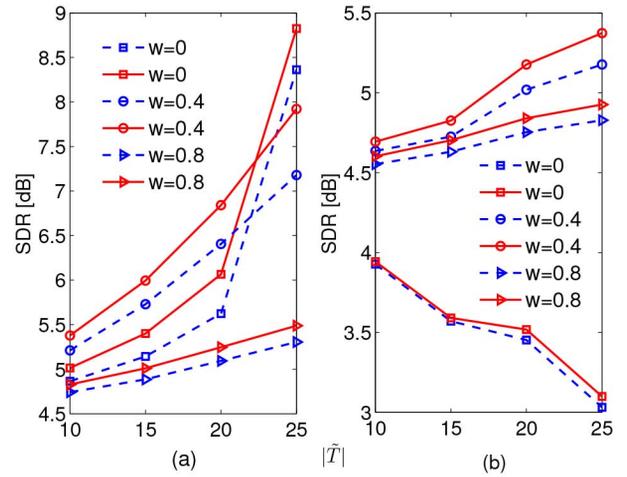


Fig. 3. SDR Performance versus $|\tilde{T}|$ for uniform noise distribution $[-0.5, 0.5]$. The solid curves correspond to $\text{WBPdq}_{1,4}$, while the dashed ones are for WBPdN . The signal has $N = 150$ and $K = 20$. The number of measurements is $M = 80$; Fig. 3 (a) is for $\alpha = 0.8$ while (b) for $\alpha = 0.5$.

in the performance for $|\tilde{T}| = 25$ and $w = 0$ in Fig. 3(a) is from the fact that in this case \tilde{T} contains all indices of the true support T . Fig. 4 shows the SDR performance of $\text{WBPdq}_{p,q}$ ($p = 1$) as a function of q for different measurement noise models; our aim is to show the importance of q with respect to the noise model. As expected, for Gaussian noise, the best performance is shown for $q = 2$. For uniform noise, we notice that the best q is 3. We also study the performance of $\text{WBPdq}_{1,q}$ for the Beta distribution (which subsumes the uniform distribution as a special case), with domain $[0, 1]$ and a probability distribution function (pdf) given by $f(x, \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}$, where B is a normalization constant, and (α, β) denote the distribution parameters. We experiment with different parameters for the Beta distribution, and we notice that choosing higher q gives better performance than choosing $q = 2$ for several cases. More precisely, when the pdf is symmetric and behaves more like a uniform distribution for most of its range ($\alpha = \beta = 0.5$ and $\alpha = \beta = 2$), the best q is greater than 2. For $\alpha = 5$ and $\beta = 1$, the distribution has a similar shape as a limited one sided-Gaussian; this makes $q = 2$ to be the best choice.

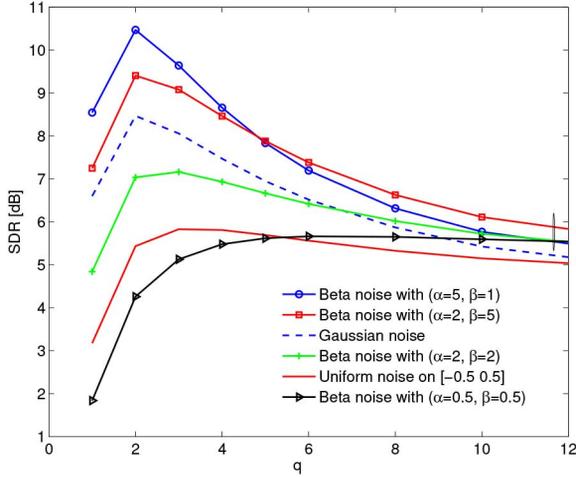


Fig. 4. Performance of the $WBPDP_{1,q}$ versus q for zero-mean Beta and uniform noise distribution. The signal has $N = 150$ and $K = 15$; the support knowledge has $|\tilde{T}| = 10$ with 80% accuracy rate and $M = 75$. Note that every curve corresponds to a different noise level.

Moreover, choosing $p = 0.7$ leads to similar conclusions; reducing p , in principle, gives a better performance. It is worth mentioning that in Fig. 4, the curves do not have the same noise levels; changing the noise level does not affect the curve trend (the curve simply shifts vertically). For future work, we will investigate different penalty functions to increase the recovery tolerance against large (unbounded) noise [21], [22].

APPENDIX

Proof of Theorem 1: Let $\hat{\mathbf{x}} = \mathbf{x} + \mathbf{h}$ be a minimizer of the weighted ℓ_p minimization problem subject to an ℓ_q norm as given in (2). As a result, we have $\|\hat{\mathbf{x}}\|_{p,w} \leq \|\mathbf{x}\|_{p,w}$. By the choice of weights in (2) and using the forward and the reverse triangle inequalities on $\|\mathbf{x} + \mathbf{h}\|_{p,w} \leq \|\mathbf{x}\|_{p,w}$, we obtain after some manipulations (similar to [8])

$$\|\mathbf{h}_{T_0^c}\|_p^p \leq w^p \|\mathbf{h}_{T_0}\|_p^p + (1-w^p) \|\mathbf{h}_{T_0 \cup \tilde{T} \setminus \tilde{T}_\alpha}\|_p^p + 2 \left(w^p \|\mathbf{x}_{T_0^c}\|_p^p + (1-w^p) \|\mathbf{x}_{\tilde{T}^c \cap T_0^c}\|_p^p \right). \quad (8)$$

Similar to [2], we next sort the coefficients of $\mathbf{h}_{T_0^c}$ and partition T_0^c into finite disjoint subsets T_j , $j \in \{1, 2, \dots, \lceil (N-K)/(aK) \rceil\}$ each of size aK , where $a \geq 1$. This means that T_1 indexes the aK largest in magnitude coefficients of $\mathbf{h}_{T_0^c}$, T_2 indexes the second aK largest in magnitude coefficients of $\mathbf{h}_{T_0^c}$ and so on. As a result $\|\mathbf{h}_{T_j}\|_2^2 \leq (aK)^{\frac{p}{2}-1} \|\mathbf{h}_{T_{j-1}}\|_p^p$. Moreover, by denoting $T_{01} = T_0 \cup T_1$ we can write

$$\|\mathbf{h}_{T_{01}}\|_2^p \leq \sum_{j \geq 2} \|\mathbf{h}_{T_j}\|_2^p \leq (aK)^{\frac{p}{2}-1} \sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_p^p \leq (aK)^{\frac{p}{2}-1} \|\mathbf{h}_{T_0^c}\|_p^p. \quad (9)$$

Combining (8) and (9), we obtain the following

$$\|\mathbf{h}_{T_{01}}\|_2^p \leq (aK)^{\frac{p}{2}-1} \left(w^p \|\mathbf{h}_{T_0}\|_p^p + (1-w^p) \|\mathbf{h}_{T_0 \cup \tilde{T} \setminus \tilde{T}_\alpha}\|_p^p + 2 \left(w^p \|\mathbf{x}_{T_0^c}\|_p^p + (1-w^p) \|\mathbf{x}_{\tilde{T}^c \cap T_0^c}\|_p^p \right) \right). \quad (10)$$

Lemma 1: [14] Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^N$ be K -sparse and K' -sparse, respectively, with disjoint supports. If Φ is $RIP_{q,2}$ of order $K + K'$ with constant $\delta_{K+K'}$ and of orders K and K' with constants δ_K and $\delta'_{K'}$, respectively, then

$$|\langle \mathbf{J}(\Phi \mathbf{u}), \Phi \mathbf{v} \rangle| \leq \mu_{q,2}^2 C_q(\Phi, K, K') \|\mathbf{u}\|_2 \|\mathbf{v}\|_2 \quad (11)$$

where $(\mathbf{J}(\mathbf{u}))_i = \|\mathbf{u}\|_q^{2-q} |u_i|^{q-1} \text{sign}(u_i)$ and $C_q(\Phi, K, K')$ is as defined in (4).

From the definition of \mathbf{J} , similar to [14], we can write

$$\begin{aligned} \|\Phi \mathbf{h}_{T_{01}}\|_q^2 &= \langle \mathbf{J}(\Phi \mathbf{h}_{T_{01}}), \Phi \mathbf{h}_{T_{01}} \rangle \\ &= \langle \mathbf{J}(\Phi \mathbf{h}_{T_{01}}), \Phi \mathbf{h} \rangle - \sum_{j \geq 2} \langle \mathbf{J}(\Phi \mathbf{h}_{T_{01}}), \Phi \mathbf{h}_{T_j} \rangle \\ &\leq \langle \mathbf{J}(\Phi \mathbf{h}_{T_{01}}), \Phi \mathbf{h} \rangle + \sum_{j \geq 2} |\langle \mathbf{J}(\Phi \mathbf{h}_{T_{01}}), \Phi \mathbf{h}_{T_j} \rangle|. \end{aligned} \quad (12)$$

Taking the p th power and using the fact that for any $b, c \geq 0$ and $0 < p < 1$, $(b+c)^p < b^p + c^p$, we obtain

$$\|\Phi \mathbf{h}_{T_{01}}\|_q^{2p} \leq (\langle \mathbf{J}(\Phi \mathbf{h}_{T_{01}}), \Phi \mathbf{h} \rangle)^p + \sum_{j \geq 2} (|\langle \mathbf{J}(\Phi \mathbf{h}_{T_{01}}), \Phi \mathbf{h}_{T_j} \rangle|)^p. \quad (13)$$

Using the fact that \mathbf{x} and $\hat{\mathbf{x}}$ are feasible solutions, we have $\|\Phi \mathbf{h}\|_q \leq \|\Phi \mathbf{x} - \hat{\mathbf{y}}\|_q + \|\Phi \hat{\mathbf{x}} - \hat{\mathbf{y}}\|_q \leq 2e$; applying Holder's inequality with parameters $r = \frac{q}{q-1}$ and $s = q$, we get

$$\begin{aligned} \langle \mathbf{J}(\Phi \mathbf{h}_{T_{01}}), \Phi \mathbf{h} \rangle &\leq \|\mathbf{J}(\Phi \mathbf{h}_{T_{01}})\|_r \|\Phi \mathbf{h}\|_s \\ &= \|\Phi \mathbf{h}_{T_{01}}\|_q \|\Phi \mathbf{h}\|_q \leq 2e \|\Phi \mathbf{h}_{T_{01}}\|_q \\ &\leq 2e \mu_{q,2} (1 + \delta_{(a+1)K})^{\frac{1}{2}} \|\mathbf{h}_{T_{01}}\|_2 \end{aligned} \quad (14)$$

where the last inequality follows from the definition of $RIP_{q,2}$ and by noting that $\mathbf{h}_{T_{01}}$ is an $(a+1)K$ sparse vector. Using Lemma 1 and the fact that $\mathbf{h}_{T_{01}}$ is $(1+a)K$ sparse and \mathbf{h}_{T_j} is aK sparse for $j \geq 2$, we can write the following

$$|\langle \mathbf{J}(\Phi \mathbf{h}_{T_{01}}), \Phi \mathbf{h}_{T_j} \rangle| \leq \mu_{q,2}^2 c_q \|\mathbf{h}_{T_{01}}\|_2 \|\mathbf{h}_{T_j}\|_2 \quad (15)$$

where $c_q = C_q(\Phi, (1+a)K, aK)$. Combining (13), (14) and (15), $\|\Phi \mathbf{h}_{T_{01}}\|_q^{2p}$ can be upper bounded. By lower bounding $\|\Phi \mathbf{h}_{T_{01}}\|_q^{2p}$ using $RIP_{q,2}$ and applying it to the upper bound found on $\|\Phi \mathbf{h}_{T_{01}}\|_q^{2p}$, we obtain after some manipulations

$$\|\mathbf{h}_{T_{01}}\|_2^p \leq \frac{(2e)^p (1 + \delta_{(a+1)K})^{p/2}}{\mu_{q,2}^p (1 - \delta_{(a+1)K})} + \frac{c_q^p}{(1 - \delta_{(a+1)K})^p} \|\mathbf{h}_{T_{01}}\|_2^p. \quad (16)$$

Using (10) in (16), the fact that $\|\mathbf{h}_{T_0}\|_p^p \leq (aK)^{1-\frac{p}{2}} \|\mathbf{h}_{T_{01}}\|_2^p$, $\|\mathbf{h}_{T_0 \cup \tilde{T} \setminus \tilde{T}_\alpha}\|_p^p \leq [(1 + \rho - 2\alpha\rho)K]^{1-\frac{p}{2}} \|\mathbf{h}_{T_{01}}\|_2^p$, $\|\mathbf{h}_{T_0 \cup \tilde{T} \setminus \tilde{T}_\alpha}\|_2 \leq \|\mathbf{h}_{T_{01}}\|_2$ and $\|\mathbf{h}_{T_0}\|_2 \leq \|\mathbf{h}_{T_{01}}\|_2$, we derive the following after some manipulations

$$\|\mathbf{h}_{T_{01}}\|_2^p \leq \frac{(2e)^p (1 + \delta_{(a+1)K})^{p/2} + 2c_q^p (aK)^{\frac{p}{2}-1} \tilde{w} \mu_{q,2}^p}{\mu_{q,2}^p [(1 - \delta_{(a+1)K})^p - c_q^p \gamma]} \quad (17)$$

Using (17) and (10), we bound $\|\mathbf{x} - \hat{\mathbf{x}}\|_2^p = \|\mathbf{h}\|_2^p \leq \|\mathbf{h}_{T_{01}}\|_2^p + \|\mathbf{h}_{T_{01}^c}\|_2^p$; this completes the proof provided that the denominator in (17) is positive which reduces to (3). \blacksquare

REFERENCES

- [1] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [2] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure and Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [3] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 31–61, 1998.
- [4] R. V. Borries, C. Miosso, and C. Potes, "Compressed sensing using prior information," in *IEEE Int. Workshop on Computational Advances in Multi-Sensor Adaptive Process*, St. Thomas, Virgin Islands, Dec. 2007, pp. 121–124.
- [5] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4595–4607, Sep. 2010.
- [6] M. A. Khajehnejad, W. Xu, A. S. Avestimehr, and B. Hassibi, "Weighted ℓ_1 minimization for sparse recovery with prior information," in *IEEE Int. Symp. Inform. Theory*, Seoul, Korea, Jun. 2009.
- [7] L. Jacques, "A short note on compressed sensing with partially known signal support," *Signal Process.*, vol. 90, no. 12, pp. 3308–3312, Dec. 2010.
- [8] M. P. Friedlander, H. Mansour, R. Saab, and O. Yilmaz, "Recovering compressively sampled signals using partial support information," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1122–1134, Feb. 2012.
- [9] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, no. 10, pp. 707–710, Oct. 2007.
- [10] R. Chartrand and V. Staneva, "Restricted isometry properties and non-convex compressive sensing," *Inv. Probl.*, vol. 24, no. 035020, pp. 1–14, May 2008.
- [11] R. Saab, R. Chartrand, and O. Yilmaz, "Stable sparse approximations via nonconvex optimization," in *IEEE Int. Conf. on Acoustics, Speech and Signal Process*, Las Vegas, NV, Apr. 2008.
- [12] T. Ince, A. Nacaroglu, and N. Watsuji, "Nonconvex compressed sensing with partially known signal support," *Signal Process.*, vol. 93, no. 1, pp. 338–344, Jan. 2013.
- [13] W. Dai and O. Milenkovic, "Information theoretical and algorithmic approaches to quantized compressive sensing," *IEEE Trans. Commun.*, vol. 59, no. 7, pp. 1857–1866, Jul. 2011.
- [14] L. Jacques, D. K. Hammond, and J. M. Fadili, "Dequantizing compressed sensing: When oversampling and non-gaussian constraints combine," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 559–571, Jan. 2011.
- [15] J. Fuchs, "Fast implementation of a $\ell_1 - \ell_1$ regularized sparse representations algorithm," in *IEEE Int. Conf. on Acoustics, Speech and Signal Process*, Taipei, Taiwan, Apr. 2009.
- [16] N. Ghadermarzy, H. Mansour, and O. Yilmaz, "Non-convex compressed sensing using partial support information," *Sampling Theory Signal Image Process.*, vol. 13, no. 3, pp. 251–272, 2014.
- [17] M. Grant and S. Boyd, *CVX: Matlab software for disciplined convex programming, version 2.1*, Mar. 2014 [Online]. Available: <http://cvxr.com/cvx>
- [18] X. Chen and W. Zhou, "Convergence of reweighted ℓ_1 minimization algorithms and unique solution of truncated ℓ_p minimization," Dept. Appl. Math., Hong Kong Polytechnic Univ., 2010.
- [19] B. D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Trans. Signal Process.*, vol. 47, no. 1, pp. 187–200, Jan. 1999.
- [20] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *IEEE Int. Conf. on Acoustics, Speech and Signal Process*, Las Vegas, Nevada, Apr. 2008.
- [21] G. R. Arce, D. Otero, A. B. Ramirez, and J.-L. Paredes, "Reconstruction of sparse signal from ℓ_1 dimensionality-reduced Cauchy random projections," in *IEEE Int. Conf. on Acoustics, Speech and Signal Process*, Dallas, Texas, Mar. 2010.
- [22] R. E. Carrillo, K. E. Barner, and T. C. Aysal, "Robust sampling and reconstruction methods for sparse signals in the presence of impulsive noise," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 392–408, Apr. 2010.