AN EXTENDED BLOCK RESTRICTED ISOMETRY PROPERTY FOR SPARSE RECOVERY WITH NON-GAUSSIAN NOISE[*]

Klara Leffler[1)]

Department of Mathematics and Mathematical Statistics, Umeå University, Umeå, Sweden.
Email: klara.leffler@umu.se

Zhiyong Zhou

Department of Statistics, Zhejiang University City College, Hangzhou 310015, China.
Email: zhiyongzhou@zucc.edu.cn

Jun Yu

Department of Mathematics and Mathematical Statistics, Umeå University, Umeå, Sweden.
Email: jun.yu@umu.se

Abstract

We study the recovery conditions of weighted mixed ℓ_2/ℓ_p minimization for block sparse signal reconstruction from compressed measurements when partial block support information is available. We show theoretically that the extended block restricted isometry property can ensure robust recovery when the data fidelity constraint is expressed in terms of an ℓ_q norm of the residual error, thus establishing a setting wherein we are not restricted to Gaussian measurement noise. We illustrate the results with a series of numerical experiments.

Mathematics subject classification: 94A12, 94A20.
Key words: Compressed sensing, Block sparsity, Partial support information, Signal reconstruction, Convex optimization.

1. Introduction

Recovering an unknown signal from significantly fewer measurements is a fundamental aspect in computational sciences today. The key ingredient here is the sparsity of the unknown signal – a realisation that has led to the theory of compressed sensing (CS) [1–3] through which successful recovery of high dimensional (approximately) sparse signals is now possible at a rate significantly lower than the Nyquist sampling rate. This allows an unknown signal $x \in \mathbb{R}^N$ to be successfully recovered via $y = Ax + e \in \mathbb{R}^m, m \ll N$, if x is (approximately) sparse in some transform domain, and the noise e satisfies $\|e\|_2 \leq \epsilon$, for $\epsilon > 0$. In short, recovery is possible if the measurement matrix $A \in \mathbb{R}^{m \times N}$ satisfies the restricted isometry property (RIP): $(1 - \delta_k)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_k)\|x\|_2^2$ for any k-sparse x and some $\delta_k \in [0, 1]$ [1]. Under such conditions, stable and robust recovery is guaranteed via the ℓ_1 minimization

$$\min_z \|z\|_1 \quad \text{s.t.} \quad \|y - Az\|_2 \leq \epsilon. \tag{1.1}$$

The question of how few measurements one might use was answered when it was shown that Gaussian random matrices satisfy the RIP with high probability, provided that $m \geq Ck \log(eN/$

k), for some $C > 0$ [4]. Today, an interesting challenge lies in customizing the recovery process to take into account prior knowledge about e.g. signal structure and properties of noise present. However, so far no unified framework has been proposed for this purpose - something we aim to do in this paper.

In addition to pure sparsity, nonzero signal components may appear in clustered regions, either naturally or as a result of some sparsifying transformation. These 'blocks' occur in many real worlds scenarios such as genetics and image processing [5–7]. Incorporating the block structure into a CS recovery algorithm provides some immediate benefits in terms of reduction of the number of required measurements for stable recovery, and a more robust recovery via better differentiation of recovery artifacts [8].

Let $x[i]$ define the ith block of a vector $x \in \mathbb{R}^N$ over the block index set $\mathcal{I} = \{d_1, \ldots, d_n\}$ such that $N = \sum_{i=1}^n d_i$, and let the blocks be formed sequentially with length d_i of block i

$$x = (\underbrace{x_1 \cdots x_{d_1}}_{x^T[1]} \underbrace{x_{d_1+1} \cdots x_{d_1+d_2}}_{x^T[2]} \cdots \underbrace{x_{N-d_n+1} \cdots x_N}_{x^T[n]})^T. \tag{1.2}$$

We define a signal $x \in \mathbb{R}^N$ as block k-sparse over $\mathcal{I}$ if $x[i]$ is nonzero for at most k indices i, i.e., if $\|x\|_{0,\mathcal{I}} \leq k$, where $\|x\|_{0,\mathcal{I}} = \sum_{i=1}^n \mathbb{I}(\|x[i]\|_2 > 0)$. The block structure of the unknown signal can be incorporated into the recovery process via a mixed minimization scheme using e.g. the ℓ_2/ℓ_1 norm [9], or its nonconvex generalization, the ℓ_2/ℓ_p norm ($0 < p \leq 1$) [5,6], where the mixed ℓ_2/ℓ_p norm is defined as $\|x\|_{2,p} = (\sum_{i=1}^n \|x[i]\|_2^p)^{1/p}$. The sufficient condition for the existence of an exact solution to (1.1) provided by the RIP has been generalized into the block sparse setting, thus guaranteeing exact and robust recovery of block sparse signals via both mixed ℓ_2/ℓ_1 and ℓ_2/ℓ_p minimization [5,9].

It may furthermore be possible to draw an estimate of the support of the largest block components of a signal, e.g., when working with recursive reconstruction of time sequences of sparse spatial signals where support estimates of previous instances can be used to estimate the present ones [10–12]. Given a support estimate $\tilde{T} \subset \{1, \ldots, N\}$, one can incorporate the prior support information via a weighted minimization approach with weights $\omega_i = \omega \in [0, 1]$ when $i \in \tilde{T}$ and $\omega_i = 1$ otherwise [10].

From a Bayesian point of view, the ℓ_2 fidelity constraint in (1.1) corresponds to a conditional loglikelihood associated with Gaussian white noise. The measurement noise might however not be Gaussian. This motivates an extension of the existing CS theory to one with a data fidelity constraint expressed in the ℓ_q norm of the residual error. The case with k-sparse signals has been studied in [4, 13, 14] for $q \geq 2$ and in [15] for $0 \leq q < 2$. A sufficient condition for sparse recovery from noisy measurements with non-Gaussian noise is given by an extension of the RIP [13], wherein the measurement matrix $A \in \mathbb{R}^{m \times N}$ is said to satisfy the extended restricted isometry property ($\text{RIP}_{q,2}$) of order k if there exists a $\delta_k \in (0, 1)$ such that $\mu_{q,2}^2(1 - \delta_k)\|x\|_2^2 \leq \|Ax\|_q^2 \leq \mu_{q,2}^2(1 + \delta_k)\|x\|_2^2$, for any k-sparse vector x and some $\mu_{q,2} > 0$. A natural question is whether one can find an optimal ℓ_q constraint for specific noise types. We expand the existing results to investigate possible q-optimality for block-sparse signals with partially known block support.

Consider an arbitrary signal $x \in \mathbb{R}^N$, defined as (1.2), with x^k as its best block k-sparse approximation. Let T_0 be the block support of x^k, where $T_0 \subset \{1, \ldots, n\}$ and $|T_0| \leq k$. Let $\tilde{T} \subset \{1, \ldots, n\}$ be the block support estimate, where $|\tilde{T}| = \rho k$ and $0 \leq \rho \leq a$ for some $a > 1$ and $|\tilde{T} \cap T_0| = \alpha \rho k$ (for interpretation of ρ and α see [10]). We define the weighted mixed ℓ_2/ℓ_p

minimization with an ℓ_q constraint of the fidelity term as

$$\min_z \sum_{i=1}^n \omega_i \|z[i]\|_2^p, \quad \text{s.t. } \|y - Az\|_q \leq \epsilon, \tag{1.3}$$

where $\omega_i = \omega \in [0,1]$ if $i \in \tilde{T}$ and $\omega_i = 1$ otherwise, $0 < p \leq 1$, and $0 \leq q < \infty$. We present a sufficient condition for robust recovery of (approximately) block-sparse signals via a weighted mixed ℓ_2/ℓ_p minimization with an ℓ_q data fidelity constraint in Section 2, thereby suggesting a unified framework for prior data knowledge in terms of structure and noise properties. Our recovery algorithm is based on convex optimization and an iteratively re-weighted least squares (IRLS) approach and is presented in Section 3, together with numerical experiments that verify our results.

2. An Extended Block Restricted Isometry Property

In this section, we introduce the extended block restricted isometry property as a sufficient condition for robust recovery via the weighted mixed ℓ_2/ℓ_p minimization with an ℓ_q error constraint. Extending the RIP has an advantage in the sense of an increased probability to satisfy the property. This means that more sensing matrices can be used in compressed sensing and allows for recovery of signals with more non-zero entries, i.e., less sparse signals, given that the signal sparsity follows a block structure. By taking advantage of explicit block structure, reconstruction performance can be improved [13, 16, 17].

Definition 2.1. *A measurement matrix* $A \in \mathbb{R}^{m \times N}$ *satisfies the* extended block restricted isometry property *(BRIP$_{q,2}$), $q \geq 2$, over $\mathcal{I} = \{d_1, \cdots, d_n\}$ of order k if*

$$\mu_{q,2}^2 (1 - \delta_{k|\mathcal{I}}) \|x\|_2^2 \leq \|Ax\|_q^2 \leq \mu_{q,2}^2 (1 + \delta_{k|\mathcal{I}}) \|x\|_2^2$$

for all $x \in \mathbb{R}^N$ that are block k-sparse over $\mathcal{I}$, some $\mu_{q,2} > 0$ and a BRIP$_{q,2}$ constant $\delta_{k|\mathcal{I}} \in [0,1]$.

Note that for $q = 2$ we have $\mu_{q,2} = 1$, and the definition corresponds to that of the block RIP in [9]. For convenience, we will use the notation δ_k for the BRIP$_{q,2}$ constant $\delta_{k|\mathcal{I}}$ when there is no confusion. The following theorem provides a guarantee of robust recovery as a consequence of the BRIP$_{q,2}$.

Theorem 2.1. *Let the best block k-sparse approximation of $x \in \mathbb{R}^N$, x^k, be supported on the block index set T_0. Define $y = Ax + e \in \mathbb{R}^m$ with $\|e\|_q \leq \epsilon$. Let $\tilde{T} \subset \{1, 2, \ldots, n\}$ be an arbitrary set, and define ρ and α as before such that $|\tilde{T}| = \rho k$ and $|T_0 \cap \tilde{T}| = \alpha \rho k$. Suppose that there exists an $a \in \mathbb{Z}$, with $a \geq (1 - \alpha)\rho$, $a > 1$.*

(i) *For $2 \leq q < \infty$, the measurement matrix A satisfies the BRIP$_{q,2}$ with*

$$(1 - \delta_{(a+1)k})^p - a^{p/2-1} \gamma c_q^p > 0,$$

then the solution $x^\sharp$ of problem (1.3) satisfies

$$\|x^\sharp - x\|_2^p \leq C_1 \frac{\epsilon^p}{\mu_{q,2}^p} + C_2 k^{p/2-1} \tilde{\omega} \tag{2.1}$$

for some positive constants C_1 and C_2.

(ii) *For $0 \leq q < 2$ and exactly block k-sparse x, the measurement matrix A satisfies the* $BRIP_{2,2}$ *with*

$$(1 - \delta_{(a+1)k})^p - \delta_{(2a+1)k}^p a^{p/2-1}\gamma > 0,$$

then the solution $x^\sharp$ of problem (1.3) satisfies

$$\|x^\sharp - x\|_2^p \leq C_3\epsilon^p \tag{2.2}$$

for some positive constant C_3.

Remark 2.1. We have $\gamma = \omega + (1-\omega)(1+\rho-2\alpha\rho)^{1-p/2}$ and $\tilde{\omega} = \omega\|x_{T_0^c}\|_{2,p}^p + (1-\omega)\|$ $x_{\tilde{T}^c \cap T_0^c}\|_{2,p}^p$ for some given $0 \leq \omega \leq 1$, and $c_q = C_q(A, (a+1)k, ak)$ with

$$C_q(A, k, k') = \min\left\{\left[(\delta_k + \delta_{k+k'})(\delta_{k'} + \delta_{k+k'} + \tilde{q}(1+\delta_k))\right]^{1/2},\right.$$

$$\left.\left[\left(\delta_{k+k'} + \tilde{q}\frac{1+\delta_{k+k'}}{2}\right)\left(\delta_{k+k'} + \tilde{q}\frac{2+\delta_{k'}+\delta_{k+k'}}{2}\right)\right]^{1/2}\right\}. \tag{2.3}$$

where $\tilde{q} = q - 2$.

Remark 2.2. The constants C_1, C_2 and C_3 are explicitly given by the following expressions:

$$C_1 = \frac{2^p(1+a^{p/2-1}\gamma)(1+\delta_{(a+1)k})^{p/2}}{(1-\delta_{(a+1)k})^p - a^{p/2-1}\gamma c_q^p},$$

$$C_2 = 2a^{p/2-1}\left[\frac{c_q^p(1+a^{p/2-1}\gamma)}{(1-\delta_{(a+1)k})^p - a^{p/2-1}\gamma c_q^p} + 1\right],$$

$$C_3 = \frac{2^p(1+a^{p/2-1}\gamma)(1+\delta_{(a+1)k})^{p/2}}{(1-\delta_{(a+1)k})^p - \delta_{(2a+1)k}^p a^{p/2-1}\gamma}.$$

Remark 2.3. For Theorem 2.1 to hold, it is sufficient that

(i) For $2 \leq q < \infty$, the measurements matrix A satisfies

$$\delta_{(a+1)k} < 1 - \left(a^{p/2-1}\left(\omega + (1-\omega)(1+\rho-2\alpha\rho)^{1-p/2}\right)\right)^{1/p}c_q^2.$$

(ii) For $0 \leq q < 2$, the measurements matrix A satisfies

$$\delta_{(2a+1)k} < \frac{1}{1 + \left(a^{p/2-1}\left(\omega + (1-\omega)(1+\rho-2\alpha\rho)^{1-p/2}\right)\right)^{1/p}}.$$

An illustration of how the slightly stronger sufficient condition given in Remark 2.3 varies with choices of ω, α and p for non-Gaussian noise is shown in Fig. 2.1. We observe that the condition on the extended block restricted isometry constant ($BRIC_{q,2}$) weakens as the support estimate
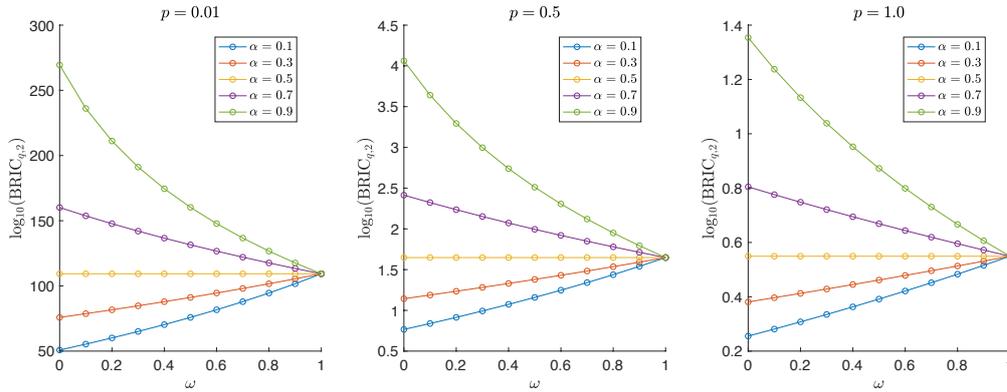
Fig. 2.1. Comparison of the sufficient conditions for recovery for the weighted mixed ℓ_2/ℓ_p minimization with an ℓ_q error constraint for various choices of weights, accuracy levels and p. I all figures, we fix $a = 3$, $\rho = 1$ and $q = 1$.

accuracy α increases, thus allowing for a wider class of measurement matrices A. Furthermore, for each choice of p and $\alpha > 0.5$, the condition weakens with increased weights ω. The opposite occurs for $\alpha < 0.5$, and for $\alpha = 0.5$ the condition remains the same. This is to be expected since the weights are constructed such that zero-expected blocks are penalized more. From another point of view, the sufficient condition on the $\mathrm{BRIC}_{q,2}$ weakens as p decreases, which reflects the benefits of non-convex minimisation.

Note that the result for exactly block-sparse signals follows from letting $\epsilon = 0$, and that letting $q = 2$ reduces the result to Theorem 1 in [7]. A crucial step in the proof of Theorem 2.1 (see Appendix A) is specified in Lemma 2.1, which is based on the geometrical description of the ℓ_q norm.

Lemma 2.1. *Let $u, v \in \mathbb{R}^N$ be block k-sparse and block k'-sparse, respectively, with disjoint supports. If $A \in \mathbb{R}^{m \times N}$ satisfies the $BRIP_{q,2}$ $(2 \leq q < \infty)$ of order $k + k'$ with constant $\delta_{k+k'}$, and of orders k and k' with constants δ_k and $\delta_{k'}$, respectively, then*

$$|\langle J(Au), Av \rangle| \leq \mu_{q,2}^2 C_q(A, k, k') \|u\|_2 \|v\|_2, \tag{2.4}$$

where $J(u)_i = \|u\|_q^{2-q} |u_i|^{q-1} \mathrm{sign}(u_i)$ and $C_q(A, k, k')$ is as defined in (2.3).

The proof of Lemma 2.1 follows from the proof of Lemma 2 in [13] and is hence omitted here.

3. Simulations

In this section, we present a simulation study to illustrate the benefits of a weighted mixed ℓ_2/ℓ_p minimisation with an ℓ_q constraint for recovery of (approximately) block k-sparse signals with partially known block support. For the case of $p = 1$ and $q \geq 1$, the problem is convex and can be solved directly with the CVX toolbox for Matlab [18]. However, for $0 < p < 1$ or $q < 1$ the problem is no longer convex. We therefore adopt an IRLS-inspired approach, a methodology widely used for nonconvex recovery of sparse signals [5, 19, 20]. Our approach

consists in approximating both norms in (1.3) by weighted ℓ_2 norms and hence solves

$$\min_x \|Wx\|_2, \quad \text{s.t.} \begin{cases} \|V(y - Ax)\|_2 \leq \epsilon, & 0 \leq q < 1, \\ \|y - Ax\|_q \leq \epsilon, & 1 \leq q < \infty, \end{cases} \tag{3.1}$$

where V and W are diagonal weight matrices with entries

$$\left(\frac{2}{q} \left(\|y - Ax\|_2^2 + \xi \right)^{q/2-1} \right)^{1/2} \quad \text{and} \quad \left(p\omega_i^{2/p} \left(\|\omega_i^{1/p}x[i]\|_2^2 + \xi \right)^{p/2-1} \right)^{1/2},$$

respectively, updated prior to each iteration. These forms of weight matrices lead to perfect approximations of the desired norms when the solution converges [5, 21–24]. The value of the perturbation parameter ξ is iteratively decreased so that it is negligible when the solution converges. With (3.1) as our objective function, the problem is convex and can be solved using CVX [18] via the following algorithm

---

Algorithm 3.1.

1. *Set $t = 0$, $\xi = 1$, and initialize by solving $x^{(0)} = \arg\min_x \|y - Ax\|_2^2$.*

2. *Set $t = t + 1$ and solve for $x^{(t)}$ using (3.1).*

3. *Decrease ξ as $\xi = 0.9\xi$.*

4. *Evaluate convergence according to $\frac{\|x^{(t)} - x^{(t-1)}\|_2}{\|x^{(t)}\|_2}$ and stop if it is less than 10^{-4}. Otherwise, go to 2.*

In our experiments, the measurement matrix A was generated by creating a $m \times N$ matrix with i.i.d. draws from a standard Gaussian distribution. The measurements y were observed from a noisy model $y = Ax + \sigma u$, where u represents noise drawn from a chosen distribution. The purpose of the experiments was to compare the recovery performance of the mixed ℓ_2/ℓ_p method for block-sparse signals in the presence of different types of noise, with respect to the ℓ_q constraint. In this work, we consider noise from the Gaussian, uniform, beta, Laplace and Poisson distributions, as well as highly impulsive noise.

We simulated block k-sparse signals of length $N = 200$, generated by uniformly choosing k blocks of length $d = 5$ at random and then for these k blocks choosing non-zero values from the standard Gaussian distribution. All signals were corrupted by noise at a level of $\sigma = 0.1$. The performance was measured in terms of the signal to noise ratio

$$SNR = 20 \log_{10} \left(\frac{\|x\|_2}{\|x - x^\sharp\|_2} \right),$$

where x represents the true signal and $x^\sharp$ the reconstruction. In this work, the signal to noise ratio is consistently presented as an average from 20 simulation runs.
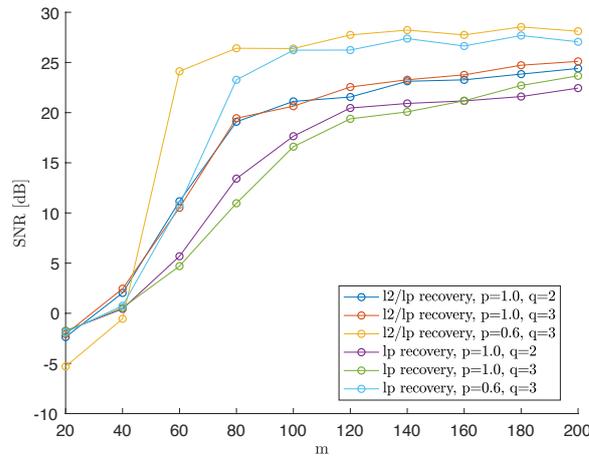
Fig. 3.1. Comparison of reconstruction performance via adaptive and non-adaptive recovery. Signals were corrupted with uniform noise on $[-0.5, 0.5]$ and recovery was performed with $k = 5, \omega = 0.3$ and $\alpha = 0.8$.
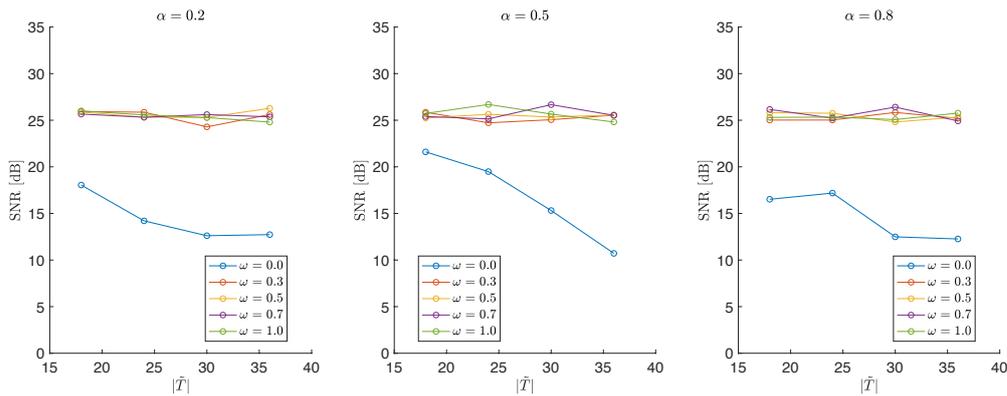

Fig. 3.2. Reconstruction performance versus support estimate size for various choices of weights, accuracy levels and sparsity levels. Signals were corrupted with uniform noise on $[-0.5, 0.5]$ and recovery was performed with $p = 0.6$, $q = 3$, $k = 6$ and $m = 80$.

The results indicate an overall higher reconstruction performance for values of $p < 1$, active block treatment and adaptive choices of q, especially for lower proportions of available measurements, see Fig. 3.1. We observe that the pairings of α and ω seen in Fig. 2.1 do not entirely match with experimental results, which indicate an overall lower performance for very low values of ω, see Figs. 3.2-3.3. This behaviour can partly be explained by situations where we cannot recover the full support of the unknown signal. Lower weight choices will then be more harshly penalised. In these situations, we expect intermediate weight choices to result in the best performance [10]. The results furthermore indicate that less sparse signals are more sensitive to choices of weights, in particular as the estimated support size increases. Moreover, we observe a small increase in performance from lower to higher accuracy levels for lower proportions of available measurements. As an example, for signals with block-sparsity $k = 6$, the
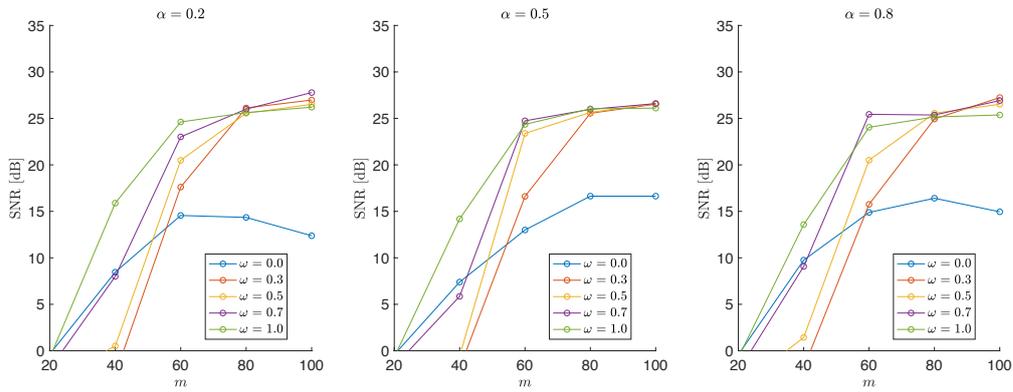
Fig. 3.3. Reconstruction performance versus number of measurements for various choices of weights, accuracy levels and sparsity levels. Signals were corrupted with uniform noise on $[-0.5, 0.5]$ and recovery was performed with $p = 0.6$, $q = 3$, $k = 6$ and $\rho = 1$.

relative improvement in SNR from $\alpha = 0.2$ to $\alpha = 0.8$ when $\omega = 0.7$ and $m = 40$ is 13.4%, see Fig. 3.3.

Fig. 3.4 illustrates the effect of the ℓ_q constraint for different noise models. As expected, optimal reconstruction performance was achieved at $q = 2$ for standard Gaussian noise and at $q = 1$ for Laplacian noise (see, e.g., [4]). In addition, optimal q-values for several other types of noise distributions were found, thereby clearly illustrating the benefits of the adaptive ℓ_q norm.
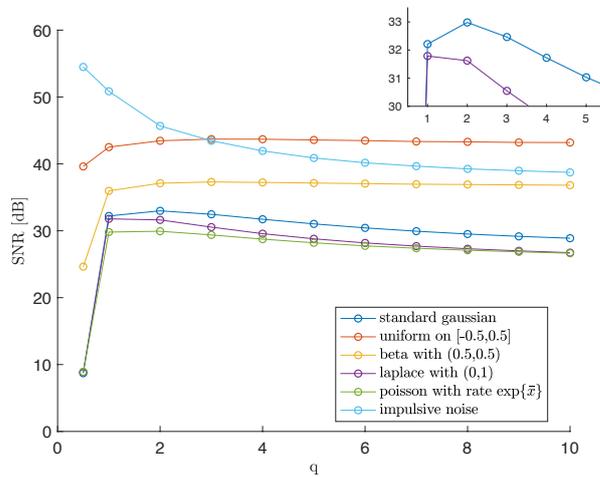

Fig. 3.4. Reconstruction performance for different noise types: standard Gaussian, uniform on $[-0.5, 0.5]$, beta with $(0.5, 0.5)$, Laplace with $(0, 1)$, Poisson with rate $\exp\{\bar{x}\}$, and highly impulsive noise. Recovery was performed with $p = 0.6$, $m = 80$, $k = 5$, $d = 5$, $\omega = 0.3$, and 80% support knowledge accuracy. Top right: Zoom of performance for standard Gaussian and Laplace noise.

4. Conclusion

The objective of this work was to introduce a noise-aware model-based approach to compressed sensing for block-sparse signals. We have shown theoretically that the reconstruction error of the presented optimization is bounded by the noise level if the measurement matrix satisfies an extended block restricted isometry property, and experimentally that our method exhibits a substantial increase in reconstruction quality for highly under-sampled signals compared to existing methods that do not incorporate signal structure. Furthermore, our method provides guidance concerning optimal norm constraints for different noise models.

A. Proof of Theorem 2.1

Proof. [The case when $2 \leq q < \infty$] Let $x^{\sharp} = x + h$ be a solution of the optimization problem (1.3), with x as the unknown true signal. Let x_T denote a vector equal to x on the block index set T and zero elsewhere. Then

$$\omega\|x_{\tilde{T}} + h_{\tilde{T}}\|_{2,p}^p + \|x_{\tilde{T}^c} + h_{\tilde{T}^c}\|_{2,p}^p \leq \omega\|x_{\tilde{T}}\|_{2,p}^p + \|x_{\tilde{T}^c}\|_{2,p}^p.$$

Let $\tilde{T}_\alpha = T_0 \cap \tilde{T}$. Applying the forward and reverse triangle inequalities together with some manipulations yields

$$\begin{aligned}\|h_{T_0^c}\|_{2,p}^p \leq &\omega\|h_{T_0}\|_{2,p}^p + (1-\omega)\|h_{(T_0 \cup \tilde{T})\setminus\tilde{T}_\alpha}\|_{2,p}^p \\ &+ 2\Big(\omega\|x_{T_0^c}\|_{2,p}^p + (1-\omega)\|x_{\tilde{T}^c \cap T_0^c}\|_{2,p}^p\Big).\end{aligned} \tag{A.1}$$

Next, decompose $h_{T_0^c}$ by partitioning T_0^c into disjoint block index sets T_j ($j \geq 1$) such that each T_j consists of ak blocks ($a > 1$), i.e. T_1 indexes the ak blocks with the largest ℓ_2 norm of $h_{T_0^c}$, T_2 indexes the ak blocks with the second largest ℓ_2 norm of $h_{T_0^c}$, and so on. As a result $h_{T_0^c} = \sum_{j\geq 1} h_{T_j}$, with $\|h_{T_j}\|_2^p \leq (ak)^{p/2-1}\|h_{T_{j-1}}\|_{2,p}^p$ so that

$$\sum_{j\geq 2}\|h_{T_j}\|_2^p \leq (ak)^{p/2-1}\|h_{T_0^c}\|_{2,p}^p,$$

and since $\|h_{(T_0 \cup T_1)^c}\|_2 \leq (ak)^{1/2-1/p}\sum_{j\geq 1}\|h_{T_j}\|_{2,p}$ we get

$$\|h_{(T_0 \cup T_1)^c}\|_2^p \leq (ak)^{p/2-1}\|h_{T_0^c}\|_{2,p}^p. \tag{A.2}$$

Let $T_{01} = T_0 \cup T_1$, and combine expressions (A.1) and (A.2) into

$$\begin{aligned}\|h_{T_{01}^c}\|_2^p \leq &(ak)^{p/2-1}\Big[\omega\|h_{T_0}\|_{2,p}^p + (1-\omega)\|h_{(T_0 \cup \tilde{T})\setminus\tilde{T}_\alpha}\|_{2,p}^p \\ &+ 2\Big(\omega\|x_{T_0^c}\|_{2,p}^p + (1-\omega)\|x_{\tilde{T}^c \cap T_0^c}\|_{2,p}^p\Big)\Big].\end{aligned} \tag{A.3}$$

From the definition of J we have

$$\|Ah_{T_{01}}\|_q^2 \leq \langle J(Ah_{T_{01}}), Ah\rangle + \sum_{j\geq 2}|\langle J(Ah_{T_{01}}), Ah_{T_j}\rangle|.$$

Taking the pth power and using the fact that $(b+c)^p \leq b^p + c^p$, for any $b, c > 0$ and $0 < p \leq 1$, we get

$$\|Ah_{T_{01}}\|_q^{2p} \leq \Big(\langle J(Ah_{T_{01}}), Ah\rangle\Big)^p + \sum_{j\geq 2}\Big(|\langle J(Ah_{T_{01}}), Ah_{T_j}\rangle|\Big)^p. \tag{A.4}$$

As x and $x^\sharp$ are both feasible solutions of (1.3), we have

$$\|Ah\|_q \le \|Ax - y\|_q + \|Ax^\sharp - y\|_q \le 2\epsilon,$$

and since A satisfies $\mathrm{BRIP}_{q,2}$, applying Holder's inequality with parameters $r = \frac{q}{1-q}$ and $s = q$ yields

$$\langle J(Ah_{T_{01}}), Ah \rangle \le 2\epsilon\mu_{q,2}\big(1 + \delta_{(a+1)k}\big)^{1/2}\|h_{T_{01}}\|_2. \tag{A.5}$$

Lemma 2.1 furthermore gives

$$|\langle J(Ah_{T_{01}}), Ah_{T_j} \rangle| \le \mu_{q,2}^2 c_q \|h_{T_{01}}\|_2 \|h_{T_j}\|_2 \tag{A.6}$$

with $c_q = C_q(A, (a+1)k, ak)$. Combining (A.4)–(A.6), and applying the $\mathrm{BRIP}_{q,2}$, we get

$$\begin{aligned}
&\mu_{q,2}^p(1 - \delta_{(a+1)k})^p\|h_{T_{01}}\|_2^p \\
&\le (2\epsilon)^p(1 + \delta_{(a+1)k})^{p/2} + \mu_{q,2}^p c_q^p \|h_{T_{01}^c}\|_2^p.
\end{aligned} \tag{A.7}$$

The ak ($a > 1$) largest coefficients of $h_{T_0^c}$ are contained in the set T_1, and we have $|(T_0 \cup \tilde{T})\backslash \tilde{T}_\alpha| = (1 + \rho - 2\alpha\rho)k$ as well as $|\tilde{T} \backslash \tilde{T}_\alpha| = (1 - \alpha)\rho k \le ak$, so

$$\begin{aligned}
\|h_{(T_0 \cup \tilde{T})\backslash \tilde{T}_\alpha}\|_{2,p}^p &\le \big[(1 + \rho - 2\alpha\rho)k\big]^{1-p/2}\|h_{T_{01}}\|_2^p, \\
\|h_{T_0}\|_{2,p}^p &\le k^{1-p/2}\|h_{T_0}\|_2^p \le k^{1-p/2}\|h_{T_{01}}\|_2^p.
\end{aligned}$$

Together with (A.3) and (A.7) we thus get

$$\begin{aligned}
&\mu_{q,2}^p(1 - \delta_{(a+1)k})^p\|h_{T_{01}}\|_2^p \\
&\le (2\epsilon)^p(1 + \delta_{(a+1)k})^{p/2} + \mu^2 c_q^p a^{p/2-1}\Big(\gamma\|h_{T_{01}}\|_2^p + 2k^{p/2-1}\tilde{\omega}\Big),
\end{aligned}$$

where $\tilde{\omega} = \omega\|x_{T_0^c}\|_{2,p}^p + (1-\omega)\|x_{\tilde{T}^c \cap T_0^c}\|_{2,p}^p$ and $\gamma = \omega + (1-\omega)\big[(1+\rho-2\alpha\rho)\big]^{1-p/2}$. Rearranging the terms yields

$$\begin{aligned}
&\mu_{q,2}^p\Big[(1 - \delta_{(a+1)k})^p - a^{p/2-1}\gamma c_q^p\Big]\|h_{T_{01}}\|_2^p \\
&\le (2\epsilon)^p(1 + \delta_{(a+1)k})^{p/2} + 2\mu_{q,2}^p c_q^p(ak)^{p/2-1}\tilde{\omega}.
\end{aligned}$$

Therefore, if $(1 - \delta_{(a+1)k})^p - a^{p/2-1}\gamma c_q^p > 0$, we get

$$\|h_{T_{01}}\|_2^p \le \frac{(2\epsilon)^p(1 + \delta_{(a+1)k})^{p/2} + 2\mu_{q,2}^p c_q^p(ak)^{p/2-1}\tilde{\omega}}{\mu_{q,2}^p\Big[(1 - \delta_{(a+1)k})^p - a^{p/2-1}\gamma c_q^p\Big]}. \tag{A.8}$$

So in the end

$$\begin{aligned}
\|x - x^\sharp\|_2^p \le{}& \frac{2^p(1 + a^{p/2-1}\gamma)(1 + \delta_{(a+1)k})^{p/2}}{(1 - \delta_{(a+1)k})^p - a^{p/2-1}\gamma c_q^p}\frac{\epsilon^p}{\mu_{q,2}^p} \\
&+ 2(ak)^{p/2-1}\tilde{\omega}\left[\frac{c_q^p(1 + a^{p/2-1}\gamma)}{(1 - \delta_{(a+1)k})^p - a^{p/2-1}\gamma c_q^p} + 1\right],
\end{aligned}$$

which completes the proof for $q \ge 2$.

[The case when $0 \leq q < 2$] According to the arguments in the first part of the proof, with the changed assumption that the measurement matrix A satisfies the $\mathrm{BRIP}_{2,2}$, we have

$$\|Ah_{T_{01}}\|_2^2 \leq \langle Ah_{T_{01}}, Ah \rangle + \sum_{j \geq 2} |\langle Ah_{T_{01}}, Ah_{T_j} \rangle|.$$

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

$$\|Ah_{T_{01}}\|_2^{2p} \leq \left(\langle Ah_{T_{01}}, Ah \rangle \right)^p + \sum_{j \geq 2} \left(|\langle Ah_{T_{01}}, Ah_{T_j} \rangle| \right)^p$$

where

$$\langle Ah_{T_{01}}, Ah \rangle \leq (1 + \delta_{(a+1)k})^{1/2} \|h_{T_{01}}\|_2 \|Ah\|_2,$$
$$|\langle Ah_{T_{01}}, Ah_{T_j} \rangle| \leq \delta_{(2a+1)k} \|h_{T_{01}}\|_2 \|h_{T_j}\|_2.$$

Combining these results, we get

$$(1 - \delta_{(a+1)k})^p \|h_{T_{01}}\|_2^p \leq (1 + \delta_{(a+1)k})^{p/2} \|Ah\|_2^p + \delta_{(2a+1)k}^p \|h_{T_{01}^c}\|_2^p. \tag{A.9}$$

If we consider x as exactly block k-sparse and supported on T_0, then by adopting (A.3) we have

$$\|h_{T_{01}^c}\|_2^p \leq a^{p/2-1} \gamma \|h_{T_{01}}\|_2^p \tag{A.10}$$

with $\gamma = \omega + (1 - \omega)(1 + \rho - 2\alpha\rho)^{1-p/2}$. Together, (A.9) and (A.10) yield

$$\|h_{T_{01}}\|_2^p \leq \frac{(1 + \delta_{(a+1)k})^{p/2}}{(1 - \delta_{(a+1)k})^p - \delta_{(2a+1)k} a^{p/2-1} \gamma} \|Ah\|_2^p,$$

where $\|Ah\|_2 \leq \|Ax^\sharp - y\|_2 + \|Ax - y\|_2 \leq \|Ax^\sharp - y\|_q + \|Ax - y\|_q \leq 2\epsilon$ for $0 \leq q < 2$. So, finally, we get

$$\|x - x^\sharp\|_2^p \leq (1 + a^{p/2-1} \gamma) \frac{(1 + \delta_{(a+1)k})^{p/2}}{(1 - \delta_{(a+1)k})^p - \delta_{(2a+1)k}^p a^{p/2-1} \gamma} (2\epsilon)^p$$

which completes the proof. $\square$

References

[1] E.J. Candés and T. Tao, Decoding by linear programming, *IEEE Transactions on Information Theory*, **51**:12 (2005), 4203–4215.

[2] E.J. Candés, J.K. Romberg and T. Tao, Stable signal recovery from incomplete and inaccurate measurements, *Communications on Pure and Applied Mathematics*, **LIX** (2006), 1207–1223.

[3] D.L. Donoho, Compressed sensing, *IEEE Transactions on Information Theory*, **52**:4 (2006), 1289–1306.

[4] S. Dirksen, G. Lecué and H. Rauhut, On the gap between restricted isometry properties and sparse recovery conditions, *IEEE Transactions on Information Theory*, **64**:8 (2016), 5478-5487.

[5] Y. Wang, J. Wang and Z. Xu, On recovery of block-sparse signals via mixed l_2/l_q ($0 < q \leq 1$) norm minimization, *EURASIP Journal of Advances in Signal Processing*, **2013**:1 (2013), 76–92.

[6] Y. Wang, J. Wang and Z. Xu, Restricted p-isometry properties of nonconvex block-sparse compressed sensing, *Signal Processing*, **104** (2014), 188–196.

[7] Z. Zhou and J. Yu, Recovery analysis for weighted mixed ℓ_2/ℓ_p minimization with $0 < p \leq 1$, *Journal of Computational and Applied Mathematics*, **352** (2019), 210–222.

[8] R.G. Baranuik, V. Cevher, M.F. Duarte and C. Hegde, Model-based compressive sensing, *IEEE Transactions on Information Theory*, **56**:4 (2010), 1982–2001.

[9] Y.C. Eldar and M. Mishali, Robust recovery of signals from a structured unoin of subspaces, *IEEE Transactions on Information Theory*, **55**:11 (2009), 5302–5316.

[10] M.P. Friedlander, H. Mansour, R. Saab and Ö. Yilmaz, Recovering compressively sampled signals using partial support information, *IEEE Transactions on Information Theory*, **58**:2 (2012), 1122–1134.

[11] L. Jacques, A short note on compressed sensing with partially known signal support, *Signal Processing*, **90** (2010), 3308–3312.

[12] N. Vaswani and W. Lu, Modified-cs: Modifying copressive sensing for problems with partially known support, *IEEE Transactions on Signal Processing*, **58**:9 (2010), 4595–4607.

[13] L. Jacques, D.K. Hammond and J.M. Fadili, Dequantizing compressed sensing: When oversampling and non-gaussian constraints combine, *IEEE Transactions on Information Theory*, **57**:1 (2011), 559–571.

[14] A.A. Saleh, F. Alajaji and W.Y. Chan, Compressed sensing with non-gaussian noise and partial support information, *IEEE Signal Processing Letters*, **22**:10 (2015), 1703–1707.

[15] F. Wen, P. Liu, Y. Liu, R.C. Qiu and W. Yu, Robust sparse recovery in impulsive noise via $\ell_p - \ell_1$ optimization, *IEEE Transactions on Signal Processing*, **65**:1 (2017), 105–118.

[16] Y. Gao and M. Ma, A new bound on the block restricted isometry constant in compressed sensing, *Journal of Inequalities and Applications*, **174** (2017).

[17] Y. Li and W. Chen, The high order block rip condition for signal recovery, *Journal of Computational Mathematics*, **37**:1 (2019), 61–75.

[18] M.C. Grant and S. Boyd, The cvx users' guide, Technical Report Release 2.1, CVX Research, Inc., 2017.

[19] R. Chartrand and W. Yin, Iteratively reweighted algorithms for compressive sensing, *33rd International Conference in Acoustics, Speech and Signal Processing (ICASSP)*, (2008).

[20] Y. Shen and S. Li, Restricted p-isometry property and its application for nonconvex compressive sensing, *Advances in Computational Mathematics*, **37**:3 (2012), 441–452.

[21] K.P. Bube and R.T. Langan, Hybrid ℓ^1/ℓ^2 minimization with applications to tomography, *Geophysics*, **62**:4 (1997), 1183–1195.

[22] I. Daubechies, R. Devore, M. Fornasier and C.S. Güntürk, Iteratively reweighted least squares minimization for sparse recovery, *Communications on Pure and Applied Mathematics*, **LXIII** (2010), 0001–0038.

[23] A. Majumdar and R.K. Ward, Non-convex compressed sensing from noisy measurements, *The Open Signal Processing Journal*, **2** (2009), 40–44.

[24] B. Wohlberg and P. Rodríguez, An iteratively reweighted norm algorithm for minimization of total variation functionals, *IEEE Signal Processing Letters*, **14**:12 (2007), 948–951.