ATOMIC NORMS IN GROUP SLIDING SPARSE RECOVERY

Carlos Buelga Sanchez, David Gregoratti, and Xavier Mestre

Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA)
Parc Mediterrani de la Tecnologia – 08860 – Castelldefels – Barcelona (Spain)
email: carlos.buelga@cttc.es

ABSTRACT

This paper tackles a compressed sensing problem with the unknown signal showing a flexible block sparsity structure, where flexible means that blocks of nonzero elements have no predetermined positions and only their minimum length is known. By capitalizing on the Minkowsky functional, the related support recovery problem is written in terms of a new vector norm that outperforms the classic l_1 norm in describing the considered sparsity structure. Also, the minimum number of measurements that are needed for perfect reconstruction is estimated by the Gaussian-width analysis of the new norm.

I. INTRODUCTION

Big data mining applications and massive deployments of sensing devices, as well as other large dimensional problems, often require dealing with a huge number of unknown variables. In this context, reducing the number of meaningful parameters by introducing a sparsity assumption has proven to be a valid approach [1]. Indeed, besides reducing computational and storing requirements, models with a reduced number of significant parameters are usually much easily understood and manipulated.

As an example, take the classical linear model

$$y = \Phi x + \eta \tag{1}$$

where signal $y \in \mathbb{R}^n$ is a noisy observation of the original (unknown) signal, $x \in \mathbb{R}^M$, through measurement matrix $\Phi \in \mathbb{R}^{n\times M}$. The noise vector is denoted by $\eta \in \mathbb{R}^n$. When the signal is undersampled, that is when $n < M$, reconstruction of x from y is generally impossible. Nevertheless, if we can assume that x is ς-sparse (i.e., it has only ς nonzero entries), then Compressed Sensing (CS) results [2], [3], [4], [5] ensure that a robust estimate $\hat{x}$ of x (that is with a negligible error $\|x - \hat{x}\|$) can be obtained as long as matrix Φ is properly chosen. For instance, this is true for all random matrices with i.i.d. normalized entries such that

$$n > 2\varsigma \log\Big(\mathrm{e}M\varsigma^{-1}\Big). \tag{2}$$

Carlos Buelga was supported by the Spanish goverment under grant TEC2014-59255-C3-1, David Gregoratti and Xavier Mestre by Catalan government under grant 2017-SGR-01479.

Moreover, $\hat{x}$ is a solution to

$$\hat{x} = \arg\min_x \|x\|_{l_1^M} \quad \text{s.t. } \|y - \Phi x\|_{l_2^n} \le \delta$$

where δ is an upper bound on the l_2^n norm of the noise η. Recovery is perfect, that is $\hat{x} = x$, in the noiseless case.

The focus of this paper is on sparse parameter vectors with specific sparsity patterns. It is well known that prior knowledge on the sparsity structure can be exploited to improve the performance of the recovery algorithm, that being in terms of speed, accuracy or observation size [6], [7]. More specifically, we consider a sparsity model where nonzero elements appear in continuous sequences of length s or longer. Block sparsity models frequently appear in machine learning, DNA microarrays, MIMO channel equalization, source localization in sensor networks and Magnetic Resonance Imaging [8], [9]. Note that, as opposed to the classic group lasso [10], [11], strings of active parameters can appear anywhere throughout vector x and not only at specific positions. Also, as already mentioned, only the minimum length of the string is known. The model introduced here resembles the one of [12]. However, their solution capitalizes on heuristics to build a weighted sparsity-inducing norm, where the weights must be tuned on a case-by-case basis. Another option is to view our model as a particular case of the totally unimodular framework in [13]. The authors encode the structure of the group sparsity model in a totally unimodular matrix, (that is, a matrix where every square submatrix has determinant 0 or ± 1) and then use the double Fenchel conjugation. In contrast to [13], we provide a theoretical approach to evaluate the number of measurements needed for robust recovery and an experimental comparisons between the induced norm and the classical l_1 norm. Finally, the (K, C)-Model proposed in [14] (K-sparse vectors in at most C clusters) tackles a similar question, but looks into uniform recovery in terms of a modified restricted isometric constant (RIP) to quantify the number of measurements necessary for this (K, C)-Model. However, [15] only studies uniform recovery, which always requires a higher number of measurements. Here, we are more interested on non-uniform recovery guarantees, which depend on specific characteristics of the sensed signal. On the other hand, checking that a matrix has the RIP property for some constant is NP hard [16]. We follow a different

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

path, and derive our sparsity-inducing norm by applying the *atomic-norm* paradigm introduced in [17]. Also we provide results towards the characterization of the minimum number of measurements that are required for robust recovery.

II. ATOMIC-NORM-BASED SIGNAL DESCRIPTION

As mentioned before, the focus of this paper is on vectors that are block sparse of size s, meaning that the nonzero elements appear in groups of at least length s. In this section we derive an atomic norm that promotes this sparsity structure. To start with, let us introduce a couple of definitions that will prove useful in the following.

Definition 1 (Atomic norm [17])**.** Let $\mathcal{A}$ be a compact subset of $\mathbb{R}^M$ and let $\|\cdot\|_{\mathcal{A}}$ be the Minkowsky functional of $\mathcal{A}$, namely

$$\|x\|_{\mathcal{A}} = \inf\{t > 0 : x \in t \cdot \mathrm{conv}(\mathcal{A})\}$$

for all $x \in \mathbb{R}^M$, where $\mathrm{conv}(\mathcal{A})$ denotes the convex hull of $\mathcal{A}$ and $t \cdot \mathcal{A} = \{t \cdot a : a \in \mathcal{A}\}$. It is easy to prove that the above Minkowsky functional is always convex and extended-real valued for all $\mathcal{A}$. Moreover, when central symmetry around the origin is satisfied, that is, $\mathcal{A} = -\mathcal{A}$, the Minkowsky functional is a norm and can be rewritten as

$$\|x\|_{\mathcal{A}} = \inf\left\{\sum_{\alpha\in\mathcal{A}} c_\alpha : x = \sum_{\alpha\in\mathcal{A}} c_\alpha \alpha, \quad c_\alpha \geq 0 \quad \forall \alpha \in \mathcal{A}\right\}.$$

Definition 2. Given a couple of natural numbers (t, u) and $M > 0$ such that $M = tu$, we define $l_p^t\left(l_q^u\right)$ as the vector space $\mathbb{R}^M$ equipped with the norm

$$\|x\|_{l_p^t(l_q^u)} = \left(\sum_{i=1}^{t} \|x^i\|_{l_q^u}^p\right)^{1/p}$$

where $x^i = [x_{u(i-1)+1}, x_{u(i-1)+2}, \ldots, x_{ui}]^{\mathrm{T}}$ $\forall p, q \geq 1$.

Next, let us introduce a (not unique) representation of any vector $x \in \mathbb{R}^{ms}$, for any two integers $m, s > 0$:

$$x = \sum_{i=1}^{ms} a_i = \sum_{i=1}^{ms} \|a_i\|_{l_2^{ms}} \frac{a_i}{\|a_i\|_{l_2^{ms}}} \tag{3}$$

where $a_i \in \mathbb{R}^{ms}$ are vectors of the form

$$a_i = S^{i-1}[a_{1,i}, a_{2,i}, \ldots, a_{s,i}, 0_{m(s-1)}^{\mathrm{T}}]^{\mathrm{T}}$$

for some $a_{1,i}, ..., a_{s,i} \in \mathbb{R}$, and where we introduced the shift operator $S \in \mathbb{R}^{ms\times ms}$, namely a circulant matrix with first row $[0, 0, \ldots, 0, 1]$, as well as the null vector $0_{m(s-1)}$ of length $m(s-1)$. Note that the normalized vector $a_i\|a_i\|_{l_2^{ms}}^{-1}$ belongs to the compact set

$$\mathcal{A}_i = \mathrm{span}\left\{e_i, e_{i\oplus 1}, ..., e_{i\oplus(s-1)}\right\} \cap \mathbb{S}^{ms-1}$$

where $(e_i)_{i=1}^{ms}$ and $\mathbb{S}^{ms-1}$ denote the canonical basis and the Euclidean unit sphere of $\mathbb{R}^{ms}$ respectively, and where $i \oplus s = ((i + s - 1) \bmod ms) + 1$. Then, one readily sees that (3) can be used to compute the atomic norm of x with respect to $\mathcal{A} = \{\mathcal{A}_1, ..., \mathcal{A}_{ms}\}$ according to Definition 1, with $M = ms$. Specifically,

$$\|x\|_{\mathcal{A}} = \inf\left\{\sum_{i=1}^{ms} \|a_i\|_{l_2^{ms}} : x = \sum_{i=1}^{ms} a_i, \quad \frac{a_i}{\|a_i\|_{l_2^{ms}}} \in \mathcal{A}_i\right\}. \tag{4}$$

For a vector $x \in \mathbb{R}^{ms}$ that is block sparse of size s, the atomic norm $\|x\|_{\mathcal{A}}$ is expected to be small, since very few atoms in the set $\mathcal{A}$ are necessary to represent x. Moreover, the active blocks of x are allowed to overlap, which is the same as saying that active blocks may be longer than s.

II-A. Closed Expression for the Atomic Norm

The atomic norm for block sparse vectors of size s as defined in (4) is of little help in practical applications. Next, we show that a more useful closed-form expression exists, namely

$$\|x\|_{\mathcal{A}} = \|Hx\|_{l_2^s(l_1^m)} \tag{5}$$

(see Definition 2), with $H \in \mathbb{R}^{ms\times ms}$ a permutation matrix with all zero elements except for those in position (i, j) such that $i = s((j-1) \bmod m) + \lceil j/m \rceil$. In order to show (5), let us define the vector

$$a = [a_{1,1}, a_{2,1}, \ldots, a_{s,1}, a_{1,2}, a_{2,2}, \ldots, a_{s,ms}]^{\mathrm{T}} \in \mathbb{R}^{ms^2}$$

and the matrix $\Lambda = [B_1, B_2, \ldots, B_{ms}] \in \mathbb{R}^{ms\times ms^2}$, where

$$B_i = S^{i-1}\begin{bmatrix} I_{s\times s} \\ 0_{(m-1)s\times s}\end{bmatrix} \in \mathbb{R}^{ms\times s}$$

with $I_{s\times s}$ the identity matrix of size s. Then, the atomic norm of x can be computed as

$$\|x\|_{\mathcal{A}} = \min_a \|a\|_{l_1^{ms}(l_2^s)} \quad \text{s.t. } x = \Lambda a.$$

The Lagrangian of the above minimization problem, with multipliers $\xi \in \mathbb{R}^{ms}$, is given by

$$\begin{aligned}\mathcal{L}(a, \xi) &= \|a\|_{l_1^{ms}(l_2^s)} - \langle \xi, \Lambda a - x\rangle \\ &= \langle \xi, x\rangle - \big(\langle \Lambda^{\mathrm{T}}\xi, a\rangle - \|a\|_{l_1^{ms}(l_2^s)}\big).\end{aligned}$$

Given that $\mathrm{rank}(\Lambda) = ms$, strong duality holds and, thus,

$$\begin{aligned}\|x\|_{\mathcal{A}} &= \max_\xi \inf_a \mathcal{L}(a, \xi) = \max_\xi \big[\langle \xi, x\rangle - \mathcal{F}(\Lambda^{\mathrm{T}}\xi)\big] \\ &= \max_\xi \langle \xi, x\rangle \quad \text{s.t. } \|\Lambda^{\mathrm{T}}\xi\|_{l_\infty^{ms}(l_2^s)} \leq 1\end{aligned} \tag{6}$$

where

$$\mathcal{F}(z) = \sup_a \big(\langle z, a\rangle - \|a\|_{l_1^{ms}(l_2^s)}\big) = \begin{cases} +\infty & \|z\|_{l_\infty^{ms}(l_2^s)} > 1 \\ 0 & \|z\|_{l_\infty^{ms}(l_2^s)} \leq 1\end{cases}$$

is the Fenchel conjugate of $\|\cdot\|_{l_1^{ms}(l_2^s)}$, given by the indicator function of the dual norm (see, e.g., [18, Example 3.26]).

From the definition of Λ, one sees that (6) requires that $\xi_i^2 + \xi_{i\oplus 1}^2 + \cdots + \xi_{i\oplus(s-1)}^2 = 1$, for all $i = 1, 2, \ldots, ms$, and, in turn, $\xi_i^2 = \xi_j^2$ for all $i \equiv j \pmod s$. Then, claim (5) stems from

$$\|x\|_{\mathcal{A}} = \max_{\widetilde{\xi}\in\mathbb{S}^{s-1}} \left(\sum_{i=1}^{s} \widetilde{\xi}_i \sum_{j\equiv i(mod\, s)} |x_j|\right) = \|Hx\|_{l_2^s(l_1^m)}.$$

III. NUMBER OF MEASUREMENTS FOR PERFECT/ROBUST RECOVERY

The objective of this section is to study the minimum number of measurements that are necessary in order to ensure that the signal is recovered with overwhelming probability when the measurement matrix is a random isotropic map. Our approach will be based on the characterization of the Gaussian width [17] of the atomic norm formulated above.

In order to introduce the concept of Gaussian width, let us recall the definition of the tangent cone of our atomic norm with respect to the point x, namely

$$\mathcal{T}_{\mathcal{A}}(x) = \operatorname{cone}\left\{z - x : \|z\|_{\mathcal{A}} \leq \|x\|_{\mathcal{A}}\right\}.$$

Note that the tangent cone is generated by the set of descending directions of our norm $\|\cdot\|_{\mathcal{A}}$ with respect to the reference point x. On the other hand, we can define the normal cone as the collection of directions that have a negative scalar product with (form an obtuse angle with respect to) the elements of $\mathcal{T}_{\mathcal{A}}(x)$, that is

$$\mathcal{N}_{\mathcal{A}}(x) = \left\{w \in \mathbb{R}^{ms} : \langle w, z - x\rangle \leq 0 \quad \|z\|_{\mathcal{A}} \leq \|x\|_{\mathcal{A}}\right\}.$$

Observe that, by using the permutation matrix H defined as above and using conventional duality theory [19], the normal cone for our specific choice of norm can be reformulated as:

$$\begin{aligned} \mathcal{N}_{\mathcal{A}}(x) &= H^{-1}\left(\mathcal{N}_{l_2^s(l_1^m)}(Hx)\right) \\ &= \bigcup_{t\geq 0}\left\{H^{-1}w \in \mathbb{R}^{ms}\right\} \end{aligned} \tag{7}$$

$$\text{where } \begin{cases} w_i = t\cdot \operatorname{sgn}((Hx)_i)\left\|(Hx)^i\right\|_{l_1^m} & \text{if } (Hx)_i \neq 0 \\ |w_i| \leq t\left\|(Hx)^i\right\|_{l_1^m} & \text{if } (Hx)_i = 0 \end{cases}$$

and where $(Hx)^i$ denotes the m-dimensional block to which the element $(Hx)_i$ belongs.

The interest of the tangent cone is revealed in the following two propositions (respectively formulated for the noiseless and the noisy signal model), which are a straightforward generalization of the equivalent result for the l_1 norm.

Proposition 1. *[17, Proposition 2.1] Let $\Phi : \mathbb{R}^{ms} \longrightarrow \mathbb{R}^n$ be a linear operator, $x \in \mathbb{R}^{ms}$ and define $y = \Phi x$. Then, the following statements are equivalent:*

1) *x is the unique solution to the convex optimization problem*
$$x = \arg\min_{y=\Phi z} \|z\|_{\mathcal{A}} \tag{8}$$
2) *$\ker(\Phi) \cap \mathcal{T}_{\mathcal{A}}(x) = \{0\}$.*

Proposition 2. *[17, Proposition 2.2] Let $\Phi : \mathbb{R}^{ms} \longrightarrow \mathbb{R}^n$ be a linear operator, $x \in \mathbb{R}^{ms}$ and define $y = \Phi x + \eta$ where $\|\eta\| \leq \delta$, $\delta > 0$. If $\|\Phi z\|_{l_2^n} \geq \epsilon \|z\|_{l_2^{ms}}$ for some $\epsilon > 0$, for all $z \in \mathcal{T}_{\mathcal{A}}(x)$ then, any solution $\widetilde{x}$ of the convex optimization problem:*

$$\widetilde{x} = \arg\min_{y=\Phi z} \|z\|_{\mathcal{A}} \tag{9}$$

satisfies $\|x - \widetilde{x}\|_{l_2^{ms}} \leq \frac{2\delta}{\epsilon}$.

The above two propositions establish a direct relationship between recovery conditions for a certain deterministic measure map Φ. However, in many applications it is very difficult to design such matrix in a deterministic way so as to guarantee perfect recovery according to the above deterministic criteria. For this reason, one typically relies on randomized approaches, according to which the measurement matrix Φ is randomly built according to some statistical constraints. When this is the case, one can only guarantee recovery up to a certain probability, but this is enough in practice to guarantee a typical good performance of the compressed sensing method. In order to characterize the chance of recovery when the measurement matrix is random, it is helpful to consider the concept of Gaussian width of a given set $T \subset \mathbb{R}^{ms}$, which is defined as

$$\omega(T) = \mathbb{E}\sup_{x\in T}\langle g, x\rangle$$

where g is a standardized Gaussian random variable. The main interest of the random width is summarized in the following proposition, which is based on concentration inequalities for Lipschitz functions and the Gordon lemma.

Theorem 3. *[17, Theorem 3.2] Let $\Phi : \mathbb{R}^{ms} \longrightarrow \mathbb{R}^n$ be a random matrix with i.i.d. zero mean and variance-$1/n$ entries and consider $x \in \mathbb{R}^{ms}$. Let T denote the spherical part of the tangent cone in the atomic norm, that is $T = \mathcal{T}_{\mathcal{A}}(x) \cap \mathbb{S}^{ms-1}$ and define $\lambda_n = \frac{\sqrt{2}\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}$. Then, the two following statements hold true.*

1) *If $y = \Phi x$, then x is the unique solution to the convex optimization problem in (8) with probability*
$$1 - \exp\left(-\frac{1}{2}\left(\lambda_n - \omega(T)\right)^2\right)$$
provided that $n \geq \omega^2(T) + 1$.
2) *If $y = \Phi x + \eta$ with $\|\eta\|_{l_2^n} \leq \delta$ for some $\delta > 0$, then given $0 < \epsilon < 1$, any solution $\widetilde{x}$ of the convex optimization problem in (9) satisfies $\|x - \widetilde{x}\|_{l_2^{ms}} \leq \frac{2\delta}{\epsilon}$ with probability*
$$1 - \exp\left(-\frac{1}{2}\left(\lambda_n - \omega(T) - \sqrt{n}\epsilon\right)^2\right)$$
provided that $n \geq \frac{\omega^2(T)+3/2}{(1-\epsilon)^2}$.

Observe that in both the noiseless and the noisy models one can ensure recovery with overwhelming probability as long as the number of measurements is larger than a quantity that is proportional to $\omega^2(T)$, i.e. the square of the Gaussian width of the spherical part of the tangent cone. It is therefore

very important to characterize this quantity for the selected atomic norm. As shown in [19], the square of the Gaussian width of $T = \mathcal{T}_{\mathcal{A}}(x) \cap \mathbb{S}^{ms-1}$ can be upper bounded by

$$\omega^2(T) \leq \mathbb{E} \min_{z \in \mathcal{N}_{\mathcal{A}}(x)} \|g - z\|^2.$$

We can investigate the behavior of the right-hand side of the above equation by using the characterization of the normal cone in (7), namely $\omega^2(T) \leq \min_{t \geq 0} (\eta_1(t) + \eta_2(t))$ where

$$\eta_1(t) = \mathbb{E}\left(\sum_{i \in \mathrm{supp}(x)} \left(g_i - \mathrm{sgn}(x_i) t \left\| (Hx)^i \right\|_{l_1^m} \right)^2 \right)$$

$$\eta_2(t) = \mathbb{E}\left(\sum_{i \notin \mathrm{supp}(x)} \min_{|z_i| \leq t \|Hx^i\|_{l_1^n}} (g_i - z_i)^2 \right).$$

The above follows from the fact $\omega^2(T) \leq \eta_1(t) + \eta_2(t)$ for every $t \geq 0$ and in particular for the minimum. Let us first focus on $\eta_1(t)$, which can be upper bounded as

$$\begin{aligned} \eta_1(t) &= \sum_{i \in \mathrm{supp}(x)} \left(1 + t^2 \left\| (Hx)^i \right\|_{l_1^m} \right) \\ &\leq ps + t^2 p \|Hx\|^2_{l_2^s(l_1^n)} = ps\left(1 + \frac{t^2}{s} \|x\|^2_{\mathcal{A}} \right) \end{aligned}$$

where p denotes the number of active blocks in x. Regarding $\eta_2(t)$, we can upper bound it as:

$$\mathbb{E} \sum_{i \notin \mathrm{supp}(x)} S^2_{t \left\| (Hx)^i \right\|_{l_1^m}}(g_i) \leq m\mathbb{E} \sum_{i=1}^{s} S^2_{t \left\| (Hx)^i \right\|_{l_1^m}}(g_i)$$

where $S_t^2(g) = \min_{|z| \leq t}(g - z)^2$ is the square of the soft-thresholding operator. Considering the change of variables $t \frac{\|x\|_{\mathcal{A}}}{\sqrt{s}} \mapsto t$ we readily obtain the upper bound for the Gaussian width

$$\min_{t \geq 0} \left[ps\left(1 + t^2\right) + m\mathbb{E}\left(\sum_{i=1}^{s} S^2_{t\sqrt{s}\frac{\|(Hx)^i\|_{l_1^m}}{\|x\|_{\mathcal{A}}}}(g_i) \right) \right].$$

There are simple cases in which we can obtain a closed form for the above estimate of $\omega^2(T)$. For instance, assume that the support of x has p active and disjoint blocks with spikes of the same amplitude, then $\sqrt{s} \left\| (Hx)^i \right\|_{l_1^m} = \|x\|_{\mathcal{A}}$ $\forall i$ and the above expression simplifies to

$$\omega^2(T) \leq \min_{t \geq 0} \left[ps\left(1 + t^2\right) + m\mathbb{E}\left(\sum_{i=1}^{s} S_t^2(g_i) \right) \right]$$

Using the same argument as in [19] we obtain the following useful estimate $\omega^2(T) \leq 2ps \log\left(\mathrm{e}\frac{s}{p}\right)$. More generally, let us assume that there exists a constant $0 < \vartheta < 1$ such that $\vartheta \leq \left\| (Hx)^i \right\|_{l_1^m} \|Hx\|^{-1}_{l_\infty^s(l_1^m)}$ by an application of Hölder's inequality, we get:

$$\vartheta \leq \sqrt{s} \left\| (Hx)^i \right\|_{l_1^m} \|x\|^{-1}_{\mathcal{A}}.$$

Thanks to monoticity of the soft-thresholding operator,

$$\begin{aligned} \omega(T)^2 &\leq \min_{t \geq 0} \left[ps\left(1 + t^2\right) + m\mathbb{E}\left(\sum_{i=1}^{s} S^2_{t\vartheta}(g_i) \right) \right] \\ &\leq \min_{t \geq 0} \left[ps\left(1 + t^2\right) + ms \exp\left(\frac{-t^2\vartheta^2}{2} \right) \right]. \end{aligned}$$

Therefore, by taking $t = \frac{1}{\vartheta}\sqrt{2\log\left(\frac{m}{p}\right)}$, we obtain the following bound for the Gaussian width

$$\omega^2(T) \leq \left(1 + \frac{2}{\vartheta^2} \log\left(\frac{m}{p}\right) \right) ps + ps \leq \frac{2}{\vartheta^2} ps \log\left(\mathrm{e}\frac{m}{p}\right).$$

Note that, when the reordered blocks show an almost flat l_1^m norm and, in turn, $\vartheta \approx 1$, the bound above is essentially the same as the one for the classic CS case in (2), for $M = ms$ and $\varsigma = ps$. Conversely, as the blocks are more irregular and $\vartheta \to 0$, we see that the Gaussian width tends to increase and, with it, the number of measurements required for perfect reconstruction.

IV. NUMERICAL EXAMPLES

We give some numerical examples that illustrate the performance of the atomic norm as compared to the traditional l_1 norm. We consider the linear model proposed in (1), with $M = 200$ and where the unknown signal x is block sparse of size $s = 10$ (thus, $m = 20$). We assume that the number of active blocks is $p = 2$ and that the nonzero elements take the value $+1$. The entries of the measurement matrix $\Phi \in \mathbb{R}^{n \times M}$ are generated as i.i.d. zero mean Gaussian random variables with variance $1/n$. Similarly, the additive noise is also Gaussian and white, distributed as $\eta \sim \mathcal{N}(0, 0.04 \cdot I_{n \times n})$. Under this setup, we study the signal support recovery properties of the regularized minimization problem

$$\min_x \|y - \Phi x\|^2_{l_2^n} + \lambda \|x\|_{\mathsf{X}} \tag{10}$$

with X either l_1^M, the classic l_1-norm, or $\mathcal{A}$, the atomic norm presented in this paper and given by (5). More specifically, for all $j = 1, 2, \ldots, M$, we compute $\gamma_j = \|x^j\|_{l_2^s}$, with $x^j = [x_j, x_{j\oplus 1}, \ldots, x_{j\oplus(s-1)}]$, and check whether the two indices j with the highest value of γ_j correspond to the start of the two active blocks in the true signal x. We use the interior point method for solving (10) in both cases. The plots in Fig. 1 show the (empirical, out of 150 experiments) probability of correctly recovering both blocks as a function of the regularizer parameter λ, with $\lambda \in [0, 5]$ (note that, in our setup, $5 \approx 2\|\Phi^{\mathrm{T}} y\|_\infty$: above this value the solution to (10) is identically null in the l_1-norm case [20]). In the top graph, the measurement size is set to $n = 133$, which guarantees perfect reconstruction for both norms, according to (2) and Theorem 3, respectively. We see that the atomic-norm algorithm is much more robust with respect to the choice of the regularizer parameter and almost perfect reconstruction is achieved for a wide range of values of

λ. Conversely, the l_1-norm-based algorithm guarantees a recovery probability above 90% only for few values of λ. The bottom graph, on the other hand, reports the case where $n = 100$, which is below the minimum number of measurements for perfect reconstruction. Indeed, we see that the classic algorithm shows a perfect-recovery probability of, at most, 80%, while the proposed norm still achieves around 90% recovery probability for almost the entire range of λ.

V. CONCLUSION

In this paper, we have derived a norm that is more suitable than the classical l_1 norm in CS problems where sparsity obeys a specific block structure, that we assumed known. We have seen that this norm consists in a rearrangement of the l_2 (l_1) norm and that it preserves the same Gaussian-width bounds and the same number of measurements for exact and robust recovery as the l_1 norm. However, as supported by numerical simulations, the new norm has proved to be much more reliable in the recovery of the unknown signal support by means of the regularized least squares problem. Moreover, as evinced by the reported results, the atomic norm is much more robust with respect to reduced measurement size as well as with respect to the choice of the regularization parameter λ, which is usually critical in the classical l_1-norm case.

VI. REFERENCES

[1] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. C. University Press, 2012.

[2] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory.*, vol. 52, pp. 1289–1306, Apr. 2006.

[3] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements." *Commun. Pure Appl. Math.*, vol. 59, pp. 1207–1223.

[4] ——, "Robust uncertainly principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory.*, vol. 52, pp. 489–509, Feb. 2006.

[5] D. Chafai, O. Gudon, G. Lecu, and A. Pajor, *Interactions Between Compressed sensing, Random Matrices and High-dimensional Geometry*. Panoramas et Synthses, 2012.

[6] J. Huang, T. Zhang, and D. Metaxas, "Learning with structured sparsity," *Journal of Machine Learning Research.*, vol. 12, pp. 3371–3412, Apr. 2011.

[7] G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity," *CoRR*, vol. abs/1006.3056, 2010.

[8] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans on Information Theory*, vol. 56, no. 4, pp. 1982–2001, April 2010.

[9] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans on Informarion Theory*, vol. 55, no. 11, pp. 5302–5316, Nov 2009.

[10] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "A sparse-group lasso," *Journal of Computational and Graphical Statistics*, vol. 22, no. 2, pp. 231–245, 2013.

[11] S. v. d. G. Lukas Meier and P. Bhlmann, "The group lasso for logistic regression," *J. R. Statist. Soc*, vol. 2, no. 70, pp. 53–71, 2008.

[12] R. Jenatton, J.-Y. Audibert, and F. Bach, "Structured variable selection with sparsity-inducing norms," *J. Mach. Learn. Res.*, vol. 12, pp. 2777–2824, Jul. 2011.

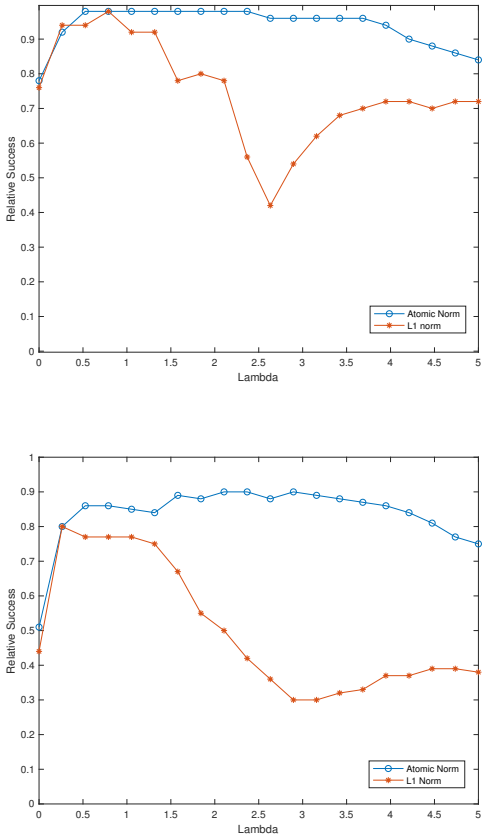

Fig. 1. Relative success of the support recovery with noise distributed as $\eta \sim \mathcal{N}(0, 0.04 \cdot I_{n\times n})$. Top: $n = 133$ measurements. Bottom: $n = 100$ measurements.

[13] M. E. Halabi and V. Cevher, "A totally unimodular view of structured sparsity," *Proceedings of the* 18^{th} *International Conference on Artificial Intelligence and Statistics*, vol. 38, pp. 223–231, 2015.

[14] V. Cevher, P. Indyk, C. Hegde, and R. G Baraniuk, "Recovery of clustered sparse signals from compressive measurements," *Sampling Theory and Applications*, 2009.

[15] T. Blumensath and M. E. Davies, "Sampling theorems for signals from the union of finite-dimensional linear subspaces," *IEEE Transactions on Information Theory*, vol. 55, no. 4, pp. 1872–1882, April 2009.

[16] A. S. Bandeira, E. Dobriban, D. G. Mixon, and W. F. Sawin, "Certifying the restricted isometry property is hard," *https://arxiv.org/pdf/1204.1580.pdf*, 2012.

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

[18] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[19] S. Foucart and H. Rauhut, *A mathematical Introduction to compressive Sensing*. Birkhauser, 2013.

[20] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale l_1-regularized least squares," *J. Selected Topics in Signal Processing.*, vol. 1, pp. 606–617, Dec. 2007.