Recovery Guarantees for Mixed Norm ℓ_{p_1,p_2} Block Sparse Representations

Fardin Afdideh, Ronald Phlypo, Christian Jutten
Univ. Grenoble Alpes, Grenoble INP, CNRS, GIPSA-lab, Grenoble, France
{fardin.afdideh, ronald.phlypo, christian.jutten}@gipsa-lab.grenoble-inp.fr

Abstract—**In this work, we propose theoretical and algorithmic-independent recovery conditions which guarantee the uniqueness of block sparse recovery in general dictionaries through a general mixed norm optimization problem. These conditions are derived using the proposed block uncertainty principles and block null space property, based on some newly defined characterizations of block spark, and (p,p)-block mutual incoherence. We show that there is improvement in the recovery condition when exploiting the block structure of the representation. In addition, the proposed recovery condition extends the similar results for block sparse setting by generalizing the criterion for determining the active blocks, generalizing the block sparse recovery condition, and relaxing some constraints on blocks such as linear independency of the columns.**

Keywords. Block-sparsity, Block-sparse recovery conditions, Block Mutual Incoherence Constant (BMIC), Block Spark, Block Uncertainty Principle (BUP).

I. Introduction

In many research fields in science and technology, e.g., biomedical imaging [1], genomics [2], statistics [3], data conversion [4], sensor networks [5], error correcting codes [6], and superresolution [7], scientists and engineers end up with vastly underdetermined systems of linear equations (USLE), which have an infinite number of solutions, if any. Because of this infinitely many solutions, the problem is said ill-posed. According to prior knowledge about the nature of the data, and consequently the solution, this eligible infinite number of solutions, which results in ambiguity, could be restricted to a smaller class of solutions or, pragmatically, to a unique solution. Commonly, a key prior is the assumption of *sparsity* of the solution. If there exist a solution that is sparse enough, one can derive necessary and sufficient conditions for exact (stable) recovery. These guarantee that the unique solution (an approximate solution) can be found independent of the algorithm used. We consider the following linear model:

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{\beta} \tag{1}$$

where $\boldsymbol{y} \in \mathbb{R}^m$ is the measurement vector, $\boldsymbol{\Phi} \in \mathbb{R}^{m\times n}$ a general dictionary and $\boldsymbol{\beta} \in \mathbb{R}^n$ the representation vector, and $m < n$ since we consider the underdetermined case.

The aforementioned sufficient sparsity condition is determined by a so-called *sparsity level*, which is the upper bound for the number of nonzero entries of the representation vector, and is derived from the dictionary. In other words, when the representation vector is very sparse, the sparsity level is low and vice versa. To define a framework for recovery or identifiability conditions, some properties and characterizations of the dictionary are introduced in the literature, among which the most commonly are: null space property [8], uniform uncertainty principle [9], Mutual Incoherence (MI) [10], spark [8], [11] and related properties [12], [9], [13]. Among all these properties, MI has the great advantage to be simple and tractable.

As mentioned before, the goal is to extract a representation vector which is the sparsest among all solutions, i.e., a representation with the fewest nonzero elements. For a vector $\boldsymbol{\beta}$, the basic sparsity measure is the ℓ_0 pseudo-norm, which is the number of nonzero entries of the vector $\boldsymbol{\beta}$. In general consider the following ℓ_p-norm optimization problem for recovering the sparse representation:

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_p \quad s.t. \quad \boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{\beta} \tag{2}$$

where $\|\boldsymbol{\beta}\|_p = \big(\sum_j |\boldsymbol{\beta}_j|^p\big)^{\frac{1}{p}}$. A usual value for p is $p = 1$, leading to a ℓ_1-norm optimization problem, which can be viewed as a convexification of ℓ_0 pseudo-norm optimization problem. It turns out that under sufficient conditions the solutions to both ℓ_0 and ℓ_1-norm optimization problem are the same and unique [14], [10]. In finding the unique and sparse representations, some researchers focused on a more general case, where $0 < p \leq 1$ [15], [16], [17], [12]. In works where the ℓ_p-norm, $p < 1$, is used as a measure of sparsity [9], the corresponding optimization problem is non-convex and combinatorial. But there is still hope, by approximating the solution of ℓ_p-norm, $p < 1$, optimization problem, and relaxing the optimization problem to ℓ_1-norm ($p = 1$ in (2)). The corresponding optimization problem is convex and can be recast as a Linear Programming (LP) problem. Therefore, it can be efficiently solved by search problems based on either the classical simplex method or recently popular interior point methods.

In applications such as EEG/MEG source reconstruction, multi-band signals [18], gene expression levels [2], in addition to sparsity, structural/geometrical constraints may be available.

For example, in EEG/MEG, we know that each dipole is a 3D moment vector. So, among infinitely many solutions, only the *active blocks* of dimension three, would be of interest. Such sparse representations are referred to as *block k-sparse*

This work have been partly supported by the ERC project CHESS: 2012-ERC-AdG-320684.

representation [19]. In addition to the aforementioned practical interest, from a mathematical point of view, assuming the block-wise structure for the representation leads to weakened recovery conditions. In fact, for the same number of nonzero elements in the representation, assuming the block structure guarantees the uniqueness of the representation with a higher sparsity level [19].

In this paper, a general theoretical and algorithmic-independent framework for the recovery of a block-sparse signal representation is proposed. The mentioned generality is in terms of structure of the dictionary, the norm in the corresponding optimization problem, the dimension of blocks, and even the parameters of the proposed block sparsity measures. The columns of the dictionary can be linearly dependent and the blocks do not need to be orthonormal bases. There is no constraint on the relationship between the number of columns and the number of rows of the dictionary or between the number of rows of the dictionary and the size of each of the blocks of the dictionary. The recovery conditions are proposed for the exact solution of the optimization problem, therefore the uniqueness of the representation is the center of our focus. Through the aforementioned generalizations we can relax some constraints and extend the results of Eldar [19]. We are interested in exact recovery, i.e., the values within the support. In other words, determining the indices of active blocks, which is attained after exact recovery, is in the scope of this study. In addition, we are not constrained to only Euclidean norm criterion for determining the active blocks.

II. Block Sparsity

As mentioned, the linear system of equations is underdetermined, $m < n$ in (1). To define the block-sparsity, the following block-wise structure of the representation vector is assumed. The blocks are assumed to share the same length d (≥ 1), without loss of generality. Define the k^{th} block as:

$$\boldsymbol{\beta}[k] = [\boldsymbol{\beta}_1[k], \cdots, \boldsymbol{\beta}_d[k]]^T \tag{3}$$

And the whole representation vector is viewed as a concatenation of K individual blocks:

$$\boldsymbol{\beta} = \left[\boldsymbol{\beta}^T[1], \cdots, \boldsymbol{\beta}^T[k], \cdots, \boldsymbol{\beta}^T[K]\right]^T \tag{4}$$

where $Kd = n$. Similarly, the following block-wise structure is assumed for the dictionary $\boldsymbol{\Phi}$ with the k^{th} block defined as the d columns of matrix $\boldsymbol{\Phi}$:

$$\boldsymbol{\Phi}[k] = [\boldsymbol{\Phi}_1[k], \cdots, \boldsymbol{\Phi}_d[k]] \tag{5}$$

with, $\boldsymbol{\Phi}_j[k] \in \mathbb{R}^m$. The whole dictionary is viewed as a concatenation of all of the individual blocks:

$$\boldsymbol{\Phi} = [\boldsymbol{\Phi}[1], \cdots, \boldsymbol{\Phi}[k], \cdots, \boldsymbol{\Phi}[K]] \tag{6}$$

where $\boldsymbol{\Phi}[k] \in \mathbb{R}^{m\times d}$. It is assumed that $\boldsymbol{\Phi}_j[k], \forall j, k$ has unit ℓ_2 norm.

In the framework of block sparsity, the fewest active blocks are of interest. Any active block has at least one nonzero element in $\boldsymbol{\beta}$ which results in its nonzero ℓ_p-norm ($0 \leq p \leq +\infty$). To our knowledge, p is always assigned to two, here we use the ℓ_p-norm in a general case of $0 \leq p$. A representation, $\boldsymbol{\beta}$, is called *block k-sparse*, if it has at most k active blocks: $\|\boldsymbol{\beta}\|_{p,0} \leq k$. The mixed norm $\ell_{p,0}$ (see Table I for its definition) measures the activity of each block in ℓ_p-norm sense and the sparsity of the active blocks in ℓ_0 pseudo-norm sense. From a mathematical point of view, not exploiting the block structure of the representation is equivalent to conventional *kd-sparse* representation $\|\boldsymbol{\beta}\|_0 \leq kd$.

In this work, we first introduce the exact recovery sufficient condition for the desired block-sparse solution to the following $\ell_{p,0}$ mixed norm optimization problem:

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_{p,0} \quad s.t. \quad \boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{\beta} \tag{7}$$

This problem is used in two cases $0 \leq p$ and $1 \leq p$. Then, we generalize the results to the following ℓ_{p_1,p_2} mixed norm optimization problem:

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_{p_1,p_2} \quad s.t. \quad \boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{\beta} \tag{8}$$

where the activity of blocks is measured by ℓ_{p_1}-norm ($1 \leq p_1$) and the sparsity of the blocks by ℓ_{p_2}-norm ($0 \leq p_2 \leq 1$), see Table I for ℓ_{p_1,p_2} mixed norm definition.

It is obvious that, if the size of the blocks, d, is chosen to be 1, then the block sparse representation problem reduces to the conventional sparse representation. To define the block sparsity conditions, it is necessary to introduce some notations.

TABLE I

$\|\boldsymbol{\beta}\|_{p_1,p_2}$ FOR DIFFERENT VALUES OF p_1 AND p_2. $I(.)$ IS THE INDICATOR FUNCTION

	$p_2 = 0$	$0 < p_2 < +\infty$	$p_2 = +\infty$
$p_1 = 0$	$\sum_{k=1}^{K} I\left(\sum_{j=1}^{d} I\left(\boldsymbol{\beta}_j[k]\right)\right)$	$\left(\sum_{k=1}^{K}\left\|\sum_{j=1}^{d} I\left(\boldsymbol{\beta}_j[k]\right)\right\|^{p_2}\right)^{\frac{1}{p_2}}$	$\max_{1\leq k\leq K}\left\{\sum_{j=1}^{d} I\left(\boldsymbol{\beta}_j[k]\right)\right\}$
$0 < p_1 < +\infty$	$\sum_{k=1}^{K} I\left(\left(\sum_{j=1}^{d}\left\|\boldsymbol{\beta}_j[k]\right\|^{p_1}\right)^{\frac{1}{p_1}}\right)$	$\left(\sum_{k=1}^{K}\left\|\sum_{j=1}^{d}\left\|\boldsymbol{\beta}_j[k]\right\|^{p_1}\right\|^{\frac{p_2}{p_1}}\right)^{\frac{1}{p_2}}$	$\max_{1\leq k\leq K}\left\{\left(\sum_{j=1}^{d}\left\|\boldsymbol{\beta}_j[k]\right\|^{p_1}\right)^{\frac{1}{p_1}}\right\}$
$p_1 = +\infty$	$\sum_{k=1}^{K} I\left(\max_{1\leq j\leq d}\left\{\left\|\boldsymbol{\beta}_j[k]\right\|\right\}\right)$	$\left(\sum_{k=1}^{K}\left\|\max_{1\leq j\leq d}\left\{\left\|\boldsymbol{\beta}_j[k]\right\|\right\}\right\|^{p_2}\right)^{\frac{1}{p_2}}$	$\max_{1\leq k\leq K}\left\{\max_{1\leq j\leq d}\left\{\left\|\boldsymbol{\beta}_j[k]\right\|\right\}\right\}$

Definition 1. *The* **Block Support** *and* **Block Cardinality** *of a representation vector,* $\boldsymbol{\beta} \in \mathbb{R}^n$*, are defined as:*

$$\forall p \geq 0, \quad \textit{block-supp}(\boldsymbol{\beta}) = T' = \left\{k : \|\boldsymbol{\beta}[k]\|_p \neq 0, 1 \leq k \leq K\right\}$$
$$\forall p \geq 0, \quad \textit{block-card}(T') = |T'| = \|\boldsymbol{\beta}\|_{p,0}$$

Definition 2. *The* **Block Kernel** *of a dictionary,* $\boldsymbol{\Phi} \in \mathbb{R}^{m\times n}$*, is its usual kernel:*

$$\textit{block-ker}(\boldsymbol{\Phi}) = \left\{\boldsymbol{x} \in \mathbb{R}^n, \sum_{k=1}^{K} \boldsymbol{\Phi}[k]\boldsymbol{x}[k] = \boldsymbol{\Phi}\boldsymbol{x} = 0\right\} = ker(\boldsymbol{\Phi})$$

Definition 3. *The* **Block Spark** *of a dictionary* $\forall p \geq 0$ *is:*

$$\textit{block-spark}(\boldsymbol{\Phi}) = \min_{\substack{\boldsymbol{x}\in \textit{block-ker}(\boldsymbol{\Phi})\\ \boldsymbol{x}\neq \mathbf{0}}} \|\boldsymbol{x}\|_{p,0} \tag{9}$$

If $d = 1$, as expected, block spark is equal to spark.

Definition 4. *The* **(p,p)-Block Mutual Incoherence Constant (BMIC$_{p,p}$)** *of a dictionary, is defined* $\forall p \geq 1$ *as:*

$$\begin{aligned} M_{p,p}(\boldsymbol{\Phi}) &= \max_{\substack{k,k'\neq k\\ \boldsymbol{x}\neq\mathbf{0}}} \frac{1}{d} \frac{\left\|\boldsymbol{\Phi}^{\dagger}[k]\boldsymbol{\Phi}[k']\boldsymbol{x}[k']\right\|_p}{\|\boldsymbol{x}[k']\|_p} \\ &= \max_{k,k'\neq k} \frac{1}{d} \left\|\boldsymbol{\Phi}^{\dagger}[k]\boldsymbol{\Phi}[k']\right\|_{p\to p} \end{aligned} \tag{10}$$

where $\boldsymbol{\Phi}^{\dagger}[k]$ is Moore-Penrose pseudo-inverse of $\boldsymbol{\Phi}[k]$, and $\|\boldsymbol{\Phi}\|_{p\to p} = \max_{\boldsymbol{x}\neq\mathbf{0}} \|\boldsymbol{\Phi}\boldsymbol{x}\|_p / \|\boldsymbol{x}\|_p$ is the operator norm. If $d = 1$, as expected, $M_{p,p}$ is equal to conventional MI, which is the maximum pairwise correlation between the atoms of the dictionary.

A. Block Uncertainty Principle and Exact Block-Sparse Recovery Condition using Block Spark

In general, the columns of the dictionary do not need to be linearly independent. Therefore, spark was defined [11] according to the smallest number of the columns which are linearly dependent. In the literature, to approach the problem of determining the sufficient conditions for unique sparse recovery, a different problem inspired by the concept of uncertainty principle is considered [20], [10], [21]. Consider the problem (2) for $p = 1$, and suppose $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$ are two distinct representations of the nonzero signal $\boldsymbol{y}$, in the dictionary $\boldsymbol{\Phi}$. The uncertainty principle of redundant solutions states that a nonzero signal can not have multiple sparse representations. In other words, there is a limit on the sparsity level of the representations $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$, namely $\|\boldsymbol{\beta_0}\|_0 + \|\boldsymbol{\beta_1}\|_0 \geq spark(\boldsymbol{\Phi})$. The mentioned uncertainty principle has been stated and demonstrated for different cases of the dictionary. At first, the dictionary was considered as a concatenation of two orthonormal bases [10], [21]. Then, this uncertainty principle was generalized to dictionaries which arise from the union of more than two orthonormal bases [8]. Finally, it was generalized to dictionaries which can be the concatenation of less structured blocks, or frames [11].

Using the aforementioned uncertainty principle, in different cases of dictionary, and the simple criterion of $spark(\boldsymbol{\Phi})$, the uniqueness of the sparse solution can be demonstrated if $\|\boldsymbol{\beta_0}\|_0 < spark(\boldsymbol{\Phi})/2$. For introducing the block sparse recovery conditions, generalizing ideas of [10], [21], we propose the following Lemma based on Block Spark (BS), called *Block Uncertainty Principle (BUP-BS)*:

Lemma 1. (BUP-BS) *For any general dictionary,* $\boldsymbol{\Phi}$*, and for any arbitrary nonzero signal,* $\boldsymbol{y}$*, with two distinct representations,* $\boldsymbol{\beta_0}$ *and* $\boldsymbol{\beta_1}$*, i.e.* $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{\beta_0} = \boldsymbol{\Phi}\boldsymbol{\beta_1}$*, we have:*

$$\|\boldsymbol{\beta_0}\|_{p,0} + \|\boldsymbol{\beta_1}\|_{p,0} \geq \textit{block-spark}(\boldsymbol{\Phi}), \quad \forall p \geq 0 \tag{11}$$

Proof. Here, we used one of the properties of the ℓ_0 pseudo-norm operator for vector space and generalized it to the block structure and derived the corresponding triangle inequality, $\|\boldsymbol{\beta_0}\|_{p,0} + \|\boldsymbol{\beta_1}\|_{p,0} \geq \|\boldsymbol{\beta_0} - \boldsymbol{\beta_1}\|_{p,0}$. Since $(\boldsymbol{\beta_0} - \boldsymbol{\beta_1}) \in \textit{block-ker}(\boldsymbol{\Phi})$, we have $\|\boldsymbol{\beta_0} - \boldsymbol{\beta_1}\|_{p,0} \geq \textit{block-spark}(\boldsymbol{\Phi})$. ☐

Theorem 1. *For any general dictionary, and* $\forall p \geq 0$ *if*

$$\|\boldsymbol{\beta_0}\|_{p,0} < \frac{\textit{block-spark}(\boldsymbol{\Phi})}{2} \tag{12}$$

then $\boldsymbol{\beta_0}$ *is the unique solution to the optimization problem (7).*

Proof. Suppose $y = \boldsymbol{\Phi}\boldsymbol{\beta_0} = \boldsymbol{\Phi}\boldsymbol{\beta_1}$. Since it is assumed that the number of active blocks of the candidate solution is less than $\textit{block-spark}(\boldsymbol{\Phi})/2$, from Lemma 1 it is concluded that any alternative solution necessarily has more than $\textit{block-spark}(\boldsymbol{\Phi})/2$ active blocks. ☐

As mentioned before, from a mathematical point of view, exploiting the block structure information of the representation leads to improved recovery conditions, i.e. conditions with higher sparsity level. For investigating this phenomenon, consider a block sparse representation, $\boldsymbol{\beta_0}$, which satisfies $\|\boldsymbol{\beta_0}\|_{p,0} < \textit{block-spark}(\boldsymbol{\Phi})/2$. On the other hand, we have, $\|\boldsymbol{\beta_0}\|_0 \leq d \times \|\boldsymbol{\beta_0}\|_{p,0}$, therefore $\|\boldsymbol{\beta_0}\|_0 \leq d \times \textit{block-spark}(\boldsymbol{\Phi})/2$. If we furthermore show that $d \times \textit{block-spark}(\boldsymbol{\Phi}) \geq spark(\boldsymbol{\Phi})$, the benefit of using block structure information will be proved.

Corollary 1. *Let* $\boldsymbol{\Phi}$ *being a general dictionary and* d *being the size of each of the blocks, we have:*

$$\forall d \geq 1, \quad d \times \textit{block-spark}(\boldsymbol{\Phi}) \geq spark(\boldsymbol{\Phi}) \tag{13}$$

Proof. Take $\boldsymbol{x}^\star \overset{\text{def}}{=} \operatorname{argmin}_{\substack{\boldsymbol{x}\in ker(\boldsymbol{\Phi})\\ \boldsymbol{x}\neq\mathbf{0}}} \|\boldsymbol{x}\|_0$ and $\boldsymbol{x}_b^\star \overset{\text{def}}{=} \operatorname{argmin}_{\substack{\boldsymbol{x}\in \textit{block-ker}(\boldsymbol{\Phi})\\ \boldsymbol{x}\neq\mathbf{0}}} \|\boldsymbol{x}\|_{p,0} \quad p \geq 0$. By definition, $\|\boldsymbol{x}^\star\|_0 \leq \|\boldsymbol{x}_b^\star\|_0$, with equality occurring when all of the elements in each of the blocks are nonzero.

It is clear that in a block structured vector, the number of nonzero elements is less than or equal to d times the number of active blocks, with equality occurring when all of the elements in each of the blocks are nonzero:

$$\|\boldsymbol{x}_b^\star\|_0 \leq d \times \|\boldsymbol{x}_b^\star\|_{p,0} \tag{14}$$

Hence, we have, $\|\boldsymbol{x}^\star\|_0 \leq d \times \|\boldsymbol{x}_b^\star\|_{p,0}$. But, $spark(\boldsymbol{\Phi}) = \|\boldsymbol{x}^\star\|_0$, and $\textit{block-spark}(\boldsymbol{\Phi}) = \|\boldsymbol{x}_b^\star\|_{p,0}$. ☐

Therefore, Theorem 1 improves the conventional spark-based recovery conditions of Donoho [11], and Gribonval [8] by weakening the condition.

B. Exact Block-Sparse Recovery Condition using Block Null Space Property

The following Theorem in a special case of $d = 1$, ℓ_1-norm of kernel and for dictionaries being a concatenation of two orthonormal bases, was used in [10] and [21]. Then, in [8], the results of [10] and [21] were generalized to ℓ_p-norm of kernel ($0 \leq p \leq 1$) and union of orthonormal bases. By refining and generalizing the ideas from Donoho [10], Elad [21] and Gribonval [8], we propose the following Theorem, called *Block Null Space Property (BNSP)*, which includes more general case of $d \geq 1$, ℓ_{p_1,p_2} mixed norm ($p_1 \geq 1$, $0 \leq p_2 \leq 1$) of block kernel and a general dictionary.

Theorem 2. (BNSP) *Let $\boldsymbol{\Phi}$ being a general dictionary and $T' \subset \{1, \cdots, K\}$ a set of block indices. For $p_1 \geq 1$, $0 \leq p_2 \leq 1$ and $T'_0 = $ block-supp$(\boldsymbol{\beta_0})$ define:*

$$P_{p_1,p_2}(T', \boldsymbol{\Phi}) = \max_{\substack{\boldsymbol{x} \in \textit{block-ker}(\boldsymbol{\Phi}) \\ \boldsymbol{x} \neq \boldsymbol{0}}} \frac{\sum_{k \in T'} \left| \sum_{j=1}^{d} |\boldsymbol{x}_j[k]|^{p_1} \right|^{\frac{p_2}{p_1}}}{\sum_{k=1}^{K} \left| \sum_{j=1}^{d} |\boldsymbol{x}_j[k]|^{p_1} \right|^{\frac{p_2}{p_1}}} \tag{15}$$

with the convention of $x^0 = \begin{cases} 1 & x \neq 0 \\ 0 & x = 0 \end{cases}$.

If $P_{p_1,p_2}(T', \boldsymbol{\Phi}) < \frac{1}{2}$, then for all $\boldsymbol{\beta_0}$ such that $T'_0 \subset T'$, $\boldsymbol{\beta_0}$ is the unique solution to the minimization problem (8).

Proof (sketch). Under the assumption of $P_{p_1,p_2}(T', \boldsymbol{\Phi}) < \frac{1}{2}$ and $T'_0 \subset T'$, to show that $\boldsymbol{\beta_0}$ is the unique solution to the minimization problem (8), we need to prove that:

$$\forall \boldsymbol{x} \in \textit{block-ker}(\boldsymbol{\Phi}), \quad \|\boldsymbol{\beta_0}\|_{p_1,p_2}^{p_2} < \|\boldsymbol{\beta_0} + \boldsymbol{x}\|_{p_1,p_2}^{p_2} \tag{16}$$

We may prove the quasi-triangle inequality $\|\boldsymbol{a} + \boldsymbol{b}\|_{p_1}^{p_2} - \|\boldsymbol{a}\|_{p_1}^{p_2} \geq -\|\boldsymbol{b}\|_{p_1}^{p_2}$ for $1 \leq p_1$ and $0 \leq p_2 \leq 1$. Therefore, for proving the necessary condition (16), after dividing the whole blocks to on-block-support ($\in T'$), and off-block-support ($\notin T'$) and using the derived quasi-triangle inequality, it is sufficient to prove:

$$0 < \sum_{k \notin T'} \left| \sum_{j=1}^{d} |\boldsymbol{x}_j[k]|^{p_1} \right|^{\frac{p_2}{p_1}} - \sum_{k \in T'} \left| \sum_{j=1}^{d} |\boldsymbol{x}_j[k]|^{p_1} \right|^{\frac{p_2}{p_1}}$$

But the above inequality is exactly the initial assumption of the proof, $P_{p_1,p_2}(T', \boldsymbol{\Phi}) < \frac{1}{2}$. □

This Theorem determines sufficient conditions on T' by determining a 50% upper threshold on the concentration of the ℓ_{p_1}-norm of blocks of block kernel $\boldsymbol{x}$ in block support T', such that guarantee the uniqueness of the solution to the minimization problem (8).

C. Block Uncertainty Principle and Exact Block-Sparse Recovery Condition using Block Mutual Incoherence Constant

In general, $spark(\boldsymbol{\Phi})$ and *block-spark*$(\boldsymbol{\Phi})$ are computationally intractable, in other words, it is impossible in polynomial time to check the identifiability of the model through the recovery conditions. Here, we use the block-wise extension of the conventional element-wise MI, which is called Block MIC (BMIC$_{\boldsymbol{p,p}}$), $M_{p,p}(\boldsymbol{\Phi})$ (10). Now, thanks to this proposed characterization of the dictionary, we can overcome to intractability of the *block-spark*$(\boldsymbol{\Phi})$, of course, with the expense of making the recovery conditions more restrictive. First, by proposing the following Lemma, we investigate the relation between *block-spark*$(\boldsymbol{\Phi})$ and $M_{p,p}(\boldsymbol{\Phi})$.

Lemma 2. *For any general dictionary,*

$$\textit{block-spark}(\boldsymbol{\Phi}) \geq 1 + (d \times M_{p,p}(\boldsymbol{\Phi}))^{-1}, \quad \forall p \geq 1 \tag{17}$$

Proof (sketch). Because $\boldsymbol{x} \in$ *block-ker*$(\boldsymbol{\Phi})$, we have for all k, $\boldsymbol{x}[k] = -\sum_{k' \neq k} \boldsymbol{\Phi}^{\dagger}[k]\boldsymbol{\Phi}[k']\boldsymbol{x}[k']$. Taking $\|.\|_p$ from both sides and using the triangular inequality for $p \geq 1$, we have $\|\boldsymbol{x}[k]\|_p \leq \sum_{k' \neq k} \left\|\boldsymbol{\Phi}^{\dagger}[k]\boldsymbol{\Phi}[k']\boldsymbol{x}[k']\right\|_p$. It follows from (10), that $\|\boldsymbol{x}[k]\|_p \leq d \times M_{p,p}(\boldsymbol{\Phi}) \sum_{k' \neq k} \|\boldsymbol{x}[k']\|_p$. Adding $d \times M_{p,p}(\boldsymbol{\Phi}) \|\boldsymbol{x}[k]\|_p$ to both sides and summing over nonzero blocks of $\boldsymbol{x}$, we obtain:

$$d \times M_{p,p}(\boldsymbol{\Phi}) \|\boldsymbol{x}\|_{p,0} \|\boldsymbol{x}\|_{p,1} \geq (1 + d \times M_{p,p}(\boldsymbol{\Phi})) \|\boldsymbol{x}\|_{p,1}$$

Then,

$$\|\boldsymbol{x}\|_{p,0} \geq 1 + (d \times M_{p,p}(\boldsymbol{\Phi}))^{-1}$$

which proves the Lemma. □

Lemma 3. (BUP-BMIC$_{\boldsymbol{p,p}}$) *For any general dictionary $\boldsymbol{\Phi}$, with BMIC$_{\boldsymbol{p,p}}$, $M_{p,p}(\boldsymbol{\Phi})$, and for any arbitrary nonzero signal $\boldsymbol{y}$, with two distinct representations $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$, the following inequality holds true $\forall p \geq 1$:*

$$\|\boldsymbol{\beta_0}\|_{p,0} + \|\boldsymbol{\beta_1}\|_{p,0} \geq 1 + (d \times M_{p,p}(\boldsymbol{\Phi}))^{-1} \tag{18}$$

Proof. This follows from Lemma 1 and Lemma 2. □

Theorem 3. *For any general dictionary, $\forall p \geq 1$ if*

$$\|\boldsymbol{\beta_0}\|_{p,0} < \frac{1 + (d \times M_{p,p}(\boldsymbol{\Phi}))^{-1}}{2} \tag{19}$$

then $\boldsymbol{\beta_0}$ is the unique solution to the optimization problem (7).

Proof. Similar to proof of Theorem 1. □

Corollary 2. *If the columns in each of the blocks are orthogonal to each other, i.e., $\boldsymbol{\Phi}^H[k]\boldsymbol{\Phi}[k] = \boldsymbol{I}_d$ for $1 \leq k \leq K$, then $\forall p \geq 1$:*

$$M_{p,p}(\boldsymbol{\Phi}) = \max_{k,k' \neq k} \frac{1}{d} \left\|\boldsymbol{\Phi}^H[k]\boldsymbol{\Phi}[k']\right\|_{p \to p} \tag{20}$$

Proof. Using the pseudo-inverse property of $\boldsymbol{A}^{\dagger} = (\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H$, we have:

$$\begin{aligned} M_{p,p}(\boldsymbol{\Phi}) &= \max_{k,k'\neq k} \frac{1}{d} \left\| \boldsymbol{\Phi}^{\dagger}[k]\boldsymbol{\Phi}[k'] \right\|_{p\to p} \\ &= \max_{k,k'\neq k} \frac{1}{d} \left\| (\boldsymbol{\Phi}^{H}[k]\boldsymbol{\Phi}[k])^{-1}\boldsymbol{\Phi}^{H}[k]\boldsymbol{\Phi}[k'] \right\|_{p\to p} \\ &= \max_{k,k'\neq k} \frac{1}{d} \left\| \boldsymbol{\Phi}^{H}[k]\boldsymbol{\Phi}[k'] \right\|_{p\to p} \end{aligned}$$

The last relation follows from $\boldsymbol{\Phi}^H[k]\boldsymbol{\Phi}[k] = \boldsymbol{I}_d$. □

If in (20), we choose $p = 2$, the right hand side of the inequality would be the block-coherence proposed by Eldar [19], $M_{Eldar}(\boldsymbol{\Phi})$. The block sparse recovery condition for ensuring the uniqueness of the solution of their proposed mixed ℓ_2/ℓ_1-optimization program (L-OPT) [22], Block Matching Pursuit (BMP), and Block Orthogonal Matching Pursuit (BOMP) [19], in the special case of Corollary 2, where there exists intra-block orthogonality, is $\|\boldsymbol{\beta_0}\|_{2,0} < (1 + (dM_{Eldar}(\boldsymbol{\Phi}))^{-1})/2$ which is very similar to Theorem 3. On the other hand, from Corollary 2 in a special setting of $p = 2$ we have $M_{2,2}(\boldsymbol{\Phi}) = M_{Eldar}(\boldsymbol{\Phi})$. Therefore Theorem 3 in a special setting of $p = 2$ is equal to Eldar's recovery condition. In other words, theoretically and independent of the recovery algorithm, unique recovery of representations are guaranteed.

III. Conclusion

In this work, the sufficient conditions for unique recovery of block sparse recovery of an arbitrary signal $\boldsymbol{y}$, in a general arbitrary dictionary $\boldsymbol{\Phi}$, using a general mixed norm optimization problem (7) and (8), are proposed. In this study, the dictionary is general and not restricted to be a union of two or more orthonormal bases. The corresponding optimization problem is a general ℓ_{p_1,p_2} ($1 \leq p_1$ and $0 \leq p_2 \leq 1$) of which $\ell_{p,0}$ ($1 \leq p$) is a special case. In addition, the proposed characterizations of $P_{p_1,p_2}(T', \boldsymbol{\Phi})$, $M_{p,p}(\boldsymbol{\Phi})$, and *block-spark*$(\boldsymbol{\Phi})$ are introduced in their general case where $P_{p_1,p_2}(T', \boldsymbol{\Phi})$ is defined for $1 \leq p_1$ and $0 \leq p_2 \leq 1$, $M_{p,p}(\boldsymbol{\Phi})$ for $1 \leq p$ and *block-spark*$(\boldsymbol{\Phi})$ for $0 \leq p$. The properties of Block Null Space Property (BNSP), and two Block Uncertainty Principles (BUP-BS and BUP-BMIC$_{p,p}$) are defined in their general case and introduced to deduce recovery conditions for block-sparse representations. We demonstrated that the proposed block-sparse recovery conditions generalizes previous results by Donoho [10], [11], Elad [21], Gribonval [8], and Eldar [19].

We will pursue the study of our proposed recovery condition since preliminary results have shown improvement with respect to Eldar's under non-orthogonal setting. Future research will also focus on: (1) stable or robust recovery conditions, i.e., when $\|\boldsymbol{y} - \boldsymbol{\Phi\beta}\|_2 < e$, and (2) considering the case of multiple dictionaries sharing the same representation $\boldsymbol{\beta}$.

References

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

[2] F. Parvaresh, H. Vikalo, S. Misra, and B. Hassibi, "Recovering sparse signals using sparse measurement matrices in compressed dna microarrays," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 3, pp. 275–285, June 2008.

[3] S. A. V. D. Geer and H. C. V. Houwelingen, "High-dimensional data: $p \gg n$ in mathematical statistics and bio-medical applications," *Bernoulli*, vol. 10, no. 6, pp. 939–942, 2004.

[4] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[5] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4036–4048, Sept 2006.

[6] J. Feldman, "Decoding error-correcting codes via linear programming," Ph.D. dissertation, 2003.

[7] D. L. Donoho, "Superresolution via sparsity constraints," *SIAM J. Math. Anal.*, vol. 23, no. 5, pp. 1309–1331, 1992.

[8] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

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

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

[11] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *in Proc. Natl. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.

[12] M. Davies and R. Gribonval, "The restricted isometry property and ℓ^p sparse recovery failure," *SPARS09*, Apr. 2009.

[13] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[14] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 43, no. 1, pp. 129–159, 2001.

[15] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Letters*, vol. 14, no. 10, pp. 707–710, Oct. 2007.

[16] M. Davies and R. Gribonval, "Restricted isometry constants where p sparse recovery can fail for $0 < p \leq 1$," Tech. Rep., Jul. 2008.

[17] S. Foucart and M.-J. Lai, "Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$," *App. Comput. Harmonic Anal.*, vol. 26, no. 3, pp. 395 – 407, 2009.

[18] M. Mishali and Y. Eldar, "From theory to practice: Sub-nyquist sampling of sparse wideband analog signals," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 375–391, Apr. 2010.

[19] Y. Eldar, P. Kuppinger, and H. Blcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.

[20] D. L. Donoho and P. B. Stark, "Uncertainty principles and signal recovery," *SIAM J. Appl. Math.*, vol. 49, no. 3, pp. 906–931, 1989.

[21] M. Elad and A. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2558–2567, Sep. 2002.

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