IJST, Transactions of Electrical Engineering, Vol. 37, No. E1, pp 1-16
Printed in The Islamic Republic of Iran, 2013

BLOCK SUBSPACE PURSUIT FOR BLOCK-SPARSE SIGNAL RECONSTRUCTION*

A. KAMALI[1,**] M. R. AGHABOZORGI SAHAF[2], A. M. DOOST HOOSEINI[3] AND A. A. TADAION[4]

[1,2,4]Dept. of Electrical Engineering, Faculty of Electrical and Computer Engineering,
Yazd University, Safaeieh, Pejouhesh St., P.O.Box 89195-741, Yazd, I. R. of Iran
Email: akamali@stu.yazduni.ac.ir
[3]Dept. of Electrical and Computer Engineering, Isfahan University of Technology, I. R. of Iran

Abstract– Subspace Pursuit (SP) is an efficient algorithm for sparse signal reconstruction. When the interested signal is block sparse, i.e., the nonzero elements occur in clusters, block sparse recovery algorithms are developed. In this paper, a blocked algorithm based on SP, namely Block SP (BSP) is presented. Contrary to the previous algorithms such as Block Orthogonal Matching Pursuit (BOMP) and mixed l_2/l_1-norm, our approach presents better recovery performance and requires less time when non-zero elements appear in fixed blocks in a particular hardware in most of the cases. It is demonstrated that our proposed algorithm can precisely reconstruct the block-sparse signals, provided that the sampling matrix satisfies the block restricted isometry property -which is a generalization of the standard RIP widely used in the context of compressed sensing-with a constant parameter. Furthermore, it is experimentally illustrated that the BSP algorithm outperforms other methods such as SP, mixed l_2/l_1-norm and BOMP. This is more pronounced when the block length is small.

Keywords– Subspace pursuit, block sparsity, compressed sensing, block RIP

1. INTRODUCTION

The most important purpose of the Compressed Sensing (CS) is the reconstruction of an unknown vector from an under-determined system of linear equation [1], [2]. The CS has gained a fast-growing interest in many different communication fields such as channel estimation [3], [4], error-correcting codes [5] and other fields (see e.g.[6]).

Consider the equation $\mathbf{y} = \mathbf{D}\mathbf{x}$, where $\mathbf{x}$ is an unknown K-sparse signal of length N, $\mathbf{D}$ denotes a sampling matrix of size $L \times N$ (where typically $L < N$) and $\mathbf{y}$ denotes the measurement vector of length L. Generally, when $K = N$ one can hope that the solution of the recent equation is unique for a large enough L. It has been shown that if $\mathbf{D}$ is chosen properly and $\mathbf{x}$ is sufficiently sparse, then $\mathbf{x}$ can be recovered from $\mathbf{y} = \mathbf{D}\mathbf{x}$ [7].

In order to obtain the sparsest solution of $\mathbf{y} = \mathbf{D}\mathbf{x}$, the solution for the minimum number of nonzero components should be investigated (that is minimum l_0 norm); however, finding the minimum l_0 norm is an intractable problem as the dimension increases (because of combinatorial search) [8]. In addition, the minimum l_0 norm is too sensitive to noise, which makes the approach less popular.

The two most commonly used signal recovery algorithms are Basis Pursuit (BP), or l_1 - minimization approach [1], [9] and Orthogonal Matching Pursuit (OMP) [10]. In the BP algorithm, the sparse signal is the solution of minimum l_1 norm, i.e. the solution which minimizes $\sum_i |x_i|$. Such a solution can be

*Received by the editors November 5, 2012; Accepted May 20, 2013.
**Corresponding author

easily found by linear programming (LP) methods. The main criteria to find out whether BP can recover the data is the Restricted Isometry Property (RIP) [11]. By utilizing fast LP algorithms, specifically interior-point methods, large-scale problems become tractable, although the process is very slow. The OMP algorithm is one of the conventional greedy algorithms, which is used in CS area because of its simplicity. The RIP has been studied for OMP, similar to BP method. In [12], the authors showed that OMP can exactly recover any K-sparse signal in no more than K steps if the RIP of order $K+1$ with isometry constant $\delta < \frac{1}{3\sqrt{K}}$ is satisfied. OMP is quite fast, but it is a greedy algorithm and does not provide a good estimate of sources. Along with the previous approaches, other algorithms like SL0 [13] have also been suggested that are faster than BP while providing at least the same accuracy.

The conventional sparsity model assumes that the nonzero coefficient elements can be located anywhere in the vector. In some practical scenarios such as when dealing with multi-band signals [14], equalization of sparse communication channels [15] and magnetoencephalography [16], the nonzero coefficients might occur in blocks. In such cases the signals will be referred to as block-sparse signals.

The question which would arise is whether the block-sparse signal exposes better recovery properties than treating the signal as being sparse in the conventional sense. This problem is considered in [17], where it is shown that, a mixed l_2/l_1-norm recovery algorithm, as a suitable extension of the BP method to the block-sparse case [7], guarantees recovery of any block-sparse signal if $\mathbf{D}$ has a small block RIP [17]. However, this algorithm is computationally complex, as we have a complicated optimization problem. For the more general setting of model-based compressive sensing, which includes block-sparsity as a particular case, it is shown in [18] that an extension of the CoSaMP (Compressive Sampling Matching Pursuit) algorithm [19] can yield excellent recovery reconstruction properties for block sparse signals. Other beneficial algorithms are BOMP [7] and BSL0 [20] which generalize the OMP and SL0, respectively. BOMP is a block version of the OMP and does not provide a good estimation of sources. Although the BSL0 has the advantages of SL0, the setting of input parameters causes the BSL0 method to become partly complicated. Like the conventional sparse algorithms, the recovery condition of some block-sparse methods such as BOMP has been considered in [21]. It was shown that if the sampling matrix $\mathbf{D}$ satisfies the Block RIP of order $K+1$ with isometry constant $\delta_d < \frac{1}{1+2\sqrt{K}}$, BOMP can exactly recover block K-sparse signals in no more than K steps.

In this paper, a block version of the Subspace Pursuit (SP) [22] -which is more efficient than CoSaMP [22]- is introduced, termed Block-SP (BSP). This approach is mainly based on the SP and BOMP algorithms. Then, the recovery performance of this method using Block RIP is analyzed. It is illustrated that the proposed algorithm outperforms both BOMP and mixed l_2/l_1-norm algorithms.

The rest of the paper is organized as follows: the next section introduces the basic principles of block sparse signal reconstruction. The proposed method is introduced in Section 3. The main point of formal proof for guaranteed reconstruction performance of BSP method is considered in Section 4. In Section 5, the proposed method is implemented. Also, a comprehensive comparison between this method and the conventional SP, BOMP and mixed l_2/l_1-norm methods is performed and the results are presented. Finally, a brief conclusion is given in Section 6.

2. BLOCK SPARSITY AND BLOCK RIP

In this section, we present some main concepts of block sparse signal recovery, including block sparsity and block RIP. Then, we consider four Lemmas that are a block version of Lemma 1 and Lemma 2 in [22]. The Definitions and Lemmas which are mentioned in this section are used in our algorithm specifically in the next sections.

a) Block sparsity

A block-sparse signal can be stated as follows,

$$\mathbf{x} = [\underbrace{x_1,\dots,x_d}_{\mathbf{x}^T[1]},\underbrace{x_{d+1},\dots,x_{2d}}_{\mathbf{x}^T[2]},\dots,\underbrace{x_{N-d+1},\dots,x_N}_{\mathbf{x}^T[M]}]^T,$$

where $\mathbf{x}^T[i]$, $i=1,\dots,M$ is called the i^{th} block of $\mathbf{x}$ and d is the block size. The signal $\mathbf{x}$ is K-sparse if, at most, K blocks of the signal out of M are nonzero. According to the definition of mixed l_2/l_0-norm [17], block sparsity can also be stated as,

$$\|\mathbf{x}\|_{2,0} = \sum_{l=1}^{M} I(\|\mathbf{x}[l]\|_2 > 0),$$

where the indicator function $I(.)$ is defined as follows,

$$I(\|\mathbf{x}[l]\|_2 \triangleq \begin{cases} 1 & \|\mathbf{x}[l]\|_2 \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

It is clear that when $d=1$, the block sparsity reduces to the conventional sparsity.

b) Block RIP (BRIP)

Before the definition of BRIP, we review the conventional RIP [11]. A matrix $\mathbf{D} \in \mathbb{R}^{L\times N}$ satisfies the RIP of order K if there exists a constant $\delta \in (0,1)$ such that,

$$(1-\delta)\|\mathbf{x}\|_2^2 \le \|\mathbf{D}\,\mathbf{x}\|_2^2 \le (1+\delta)\|\mathbf{x}\|_2^2$$

holds for all $\mathbf{x} \in \mathbb{R}^N$ assuming that$\|\mathbf{x}\|_0 \le K$. In [17], this property is generalized to block-sparse vectors and therefore leads to the following definition.

Definition 1. (**Block RIP**)[17] The matrix $\mathbf{D} \in \mathbb{R}^{L\times N}$ has the BRIP with parameters (K,δ_d) where $\delta_d \in (0,1)$, if for all d-block K-sparse $\mathbf{x} \in \mathbb{R}^N$ we have,

$$(1-\delta_d)\|\mathbf{x}\|_2^2 \le \|\mathbf{D}\,\mathbf{x}\|_2^2 \le (1+\delta_d)\|\mathbf{x}\|_2^2$$

The BRIP can be defined in other forms. However, the block truncation should at first be defined.

Definition 2. (**Block Truncation**) Suppose that $\mathbf{D} \in \mathbb{R}^{L\times N}$, $\mathbf{x} \in \mathbb{R}^N$ and $I \subseteq \{1,\dots,M\}$. Let $|I|$ denote the size of I. The matrix $\mathbf{D}_I \in \mathbb{R}^{L\times|I|d}$ consists of the columns of $\mathbf{D}$ with indices $(i-1)d+1$ to id for $i \in I$ and $\mathbf{x}_I \in \mathbb{R}^{|I|d}$ denotes a vector whose elements start from the $((i-1)d+1)$th element of $\mathbf{x}$ and ends at the id th element for $i \in I$. The space spanned by the columns of $\mathbf{D}_I$ is denoted by span $(\mathbf{D}_I)$.

Now, BRIP can also be expressed in the following form,

A matrix $\mathbf{D} \in \mathbb{R}^{L\times N}$ is stated to satisfy the BRIP with parameters (K,δ_d) for $0 \le \delta_d \le 1$, if for the index set $I \subseteq \{1,\dots,M\}$ ($|I| \le K$) we have,

$$(1-\delta_d)\|\mathbf{x}_I\|_2^2 \le \|\mathbf{D}_I\,\mathbf{x}_I\|_2^2 \le (1+\delta_d)\|\mathbf{x}_I\|_2^2$$

for all $\mathbf{x}_I \in \mathbb{R}^{d|I|}$. We identify δ_{dK}, the block RIP constant, as the infimum of all parameters δ_d for which the BRIP holds, i.e.

$$\delta_{dk} \triangleq \inf\{\delta_d : (1-\delta_d)\|\mathbf{x}_I\|_2^2 \le \|\mathbf{D}_I\,\mathbf{x}_I\|_2^2 \le (1+\delta_d)\|\mathbf{x}_I\|_2^2, \forall |I| \le K, \forall\, \mathbf{x}_I \in \mathbb{R}^{d|I|}\}.$$

Lemma 1. (Uniformity of δ_{dK})

For any two integer numbers K and K', if $K \le K'$, then we have,

$$\delta_{dK} \le \delta_{dK'}. \tag{1}$$

Proof: See Appendix A.

Lemma 2. (Near – orthogonality of blocks)

$|< \mathbf{D}_I \mathbf{a}, \mathbf{D}_J \mathbf{b} >| \le \delta_{d(|I|+|J|)}$ Let $I, J \subseteq \{1,2,\dots,M\}$ be two disjoint sets, $I \cap J = \varnothing$. Assume that $\delta_{d(|I|+|J|)} \le 1$, then all vectors $\mathbf{a} \in \mathbb{R}^{d|I|}$ and $\mathbf{b} \in \mathbb{R}^{d|I|}$,

$$\left|< \mathbf{D}_I \mathbf{a}, \mathbf{D}_J \mathbf{b} >\right| \le \delta_{d(|I|+|J|)} \|\mathbf{a}\|_2 \|\mathbf{b}\|_2 \tag{2}$$

and

$$\left\|\mathbf{D}_I^H \mathbf{D}_J \mathbf{b}\right\|_2 \le \delta_{d(|I|+|J|)} \|\mathbf{b}\|_2 \tag{3}$$

where " H " is the conjugate transpose operator.

Proof: See Appendix B.

Definition 3. (Projection and Residue) Suppose $\mathbf{y} \in \mathbb{R}^L$ and $\mathbf{D}_I \in \mathbb{R}^{L\times|I|d}$. Assume that $\mathbf{D}_I$ is a matrix of full column rank $(|I|d \le L)$, then the Moor-penrose pseudoinverse of $\mathbf{D}_I$ can be denoted by $\mathbf{D}_I^\dagger = (\mathbf{D}_I^H \mathbf{D}_I)^{-1} \mathbf{D}_I^H$. The projection operator of $\mathbf{y}$ onto span($\mathbf{D}_I$) is defined as,

$$\mathbf{y}_p = \operatorname{proj}(\mathbf{y}, \mathbf{D}_I) \stackrel{\Delta}{=} \mathbf{D}_I \mathbf{D}_I^\dagger \mathbf{y},$$

and the residue vector of the projection is defined as,

$$\mathbf{y}_r \stackrel{\Delta}{=} \operatorname{resid}(\mathbf{y}, \mathbf{D}_I) \stackrel{\Delta}{=} \mathbf{y} - \mathbf{y}_p. \tag{4}$$

Lemma 3. (Orthogonality of the residue) For any vector $\mathbf{y} \in \mathbb{R}^L$ and sampling matrix $\mathbf{D}_I \in \mathbb{R}^{L\times|I|d}$ of full column rank $(|I|d \le L)$, if $\mathbf{y}_r = \operatorname{resid}(\mathbf{y}, \mathbf{D}_I)$, then we have,

$$\mathbf{D}_I^H \mathbf{y}_r = 0. \tag{5}$$

Proof: See Appendix C.

Lemma 4. (Approximation of the projection residue) Suppose a matrix $\mathbf{D} \in \mathbb{R}^{L\times N}$, let $I, J \subseteq \{1,2,\dots,M\}$ be two disjoint sets and $\delta_{d(|I|+|J|)} \le 1$. Moreover, let $\mathbf{y} \in \operatorname{span}(\mathbf{D}_I)$, $\mathbf{y}_p = \operatorname{proj}(\mathbf{y}, \mathbf{D}_J)$ and $\mathbf{y}_r = \operatorname{resid}(\mathbf{y}, \mathbf{D}_J)$. Then,

$$\left\|\mathbf{y}_p\right\|_2 \le \frac{\delta_{d(||I||+|J|)}}{1-\delta_{\max(d|I|,d|J|)}} \|\mathbf{y}\|_2 \tag{6}$$

and

$$\left(1 - \frac{\delta_{d(||I||+|J|)}}{1-\delta_{\max(d|I|,d|J|)}}\right) \|\mathbf{y}\|_2 \le \|\mathbf{y}_r\|_2 \le \|\mathbf{y}\|_2 \tag{7}$$

Proof: See Appendix D.

3. THE BSP ALGORITHM

In this section we present our proposed algorithm method. Prior to this, we consider the OMP and SP algorithms. As can be seen in Algorithm 1, the SP method [22] selects K indices related to the largest correlation amplitude in every stage of estimation, whilst only one index is selected in OMP method [10]. In addition to this property of choosing K number of indices which are the estimation of the correct support set T , it would edit and refine its own set of K indices. This would bring the estimated support set close to correct support set (this cannot be done in the family of OMP algorithm). In our proposed BSP algorithm (Algorithm 2), contrary to BOMP, in which in each stage only the block index that is best matched to residual vector is chosen [7], it selects the K block indices which have the largest correlation. In addition, this selected set of K block indices in every iteration stage of estimation would not be constant and would be refined itself. As a result, the estimated support set would become very close to the correct support set in a way similar to the SP algorithm.

Algorithm 1. SP Algorithm

Input: sampling matrix $\mathbf{D}$, measurement vector $\mathbf{y}$, sparsity K
Initialize:
- $l = 0$
- $h^l = \mathbf{D}^H \mathbf{y}$
- $T^0 =$ { K first indices in descending sort $|h^l(j)|$}
- $\mathbf{y}_r^l =$ resid$(\mathbf{y}, \mathbf{D}_{T^0})$

Iteration:
1) $l = l + 1$
2) $h^l = \mathbf{D}^H \mathbf{y}_r^{l-1}$
3) $\tilde{T}^l = T^{l-1} \cup$ { K first indices in descending sort $|h^l(j)|$}
4) $\mathbf{x}_p = \mathbf{D}^{\dagger}_{\tilde{T}^l} \mathbf{y}$
5) $T^l =$ { K first indices in descending sort $|\mathbf{x}_p(j)|$}
6) $\mathbf{y}_r^l = \text{resid}(\mathbf{y}, \mathbf{D}_{T^l})$
7) if $\|\mathbf{y}_r^l\|_2 \geq \|\mathbf{y}_r^{l-1}\|_2$, then $T^l = T^{l-1}$ and quit

Output: The estimate signal $\hat{\mathbf{x}}$, as $\hat{\mathbf{x}}_{\{1,\ldots N\}-T^l} = o$ and $\hat{\mathbf{x}}_{T^l} = \mathbf{D}^{\dagger}_{T^l} \mathbf{y}$.

Algorithm 2. BSP Algorithm (proposed algorithm)

Input: sampling matrix $\mathbf{D}$, measurement vector $\mathbf{y}$, sparsity K , block length d
Initialize:
- $l = 0$
- $h^l = \mathbf{D}^H \mathbf{y}$
- $T^0 =$ {d-block K first indices in descending sort $\|h^l[j]\|_2$}
- $\mathbf{y}_r^l =$ resid$(\mathbf{y}, \mathbf{D}_{T^0})$

Iteration:
1) $l = l + 1$
2) $h^l = \mathbf{D}^H \mathbf{y}_r^{l-1}$
3) $\tilde{T}^l = T^{l-1} \cup$ {d-block K first indices in descending sort $\|h^l[j]\|_2$}
4) $\mathbf{x}_p = \mathbf{D}^{\dagger}_{\tilde{T}^l} \mathbf{y}$
5) $T^l =$ {d-block K first indices in descending sort$\|\mathbf{x}_p[j]\|_2$}
6) $\mathbf{y}_r^l = \text{resid}(\mathbf{y}, \mathbf{D}_{T^l})$
7) if $\|\mathbf{y}_r^l\|_2 \geq \|\mathbf{y}_r^{l-1}\|_2$, then $T^l = T^{l-1}$ and quit

Output: The estimate signal $\hat{\mathbf{x}}$, as $\hat{\mathbf{x}}_{\{1,\ldots N\}-T^l} = o$ and $\hat{\mathbf{x}}_{T^l} = \mathbf{D}^{\dagger}_{T^l} \mathbf{y}$.

4. ANALYSIS OF BSP USING BLOCK RIP

In this section, we investigate a sufficient condition for the exact reconstruction of any block-sparse signal using Block RIP. Four theorems are considered that are a block version of corresponding theorems in [22]. However, the main conclusion of our work is included in theorem1.

Theorem 1. Suppose that $\mathbf{x} \in \mathbb{R}^N$ be a d-block K-sparse signal and let its measurement vector be $\mathbf{y} = \mathbf{Dx} \in \mathbb{R}^N$. If the sampling matrix $\mathbf{D}$ satisfies the BRIP with constant

$$\delta_{3dK} < 0.1672, \tag{8}$$

then the BSP algorithm is guaranteed to exactly recover $\mathbf{x}$ from $\mathbf{y}$ via a finite number of iterations.

To prove the main theorem in this paper, the three original theorems should be considered which are as follows.

Theorem 2. The following inequality is valid,

$$\left\|\mathbf{x}_{T-\tilde{T}_l}\right\|_2 \leq \frac{2\delta_{3dk}}{(1-\delta_{3dk})^2}\left\|\mathbf{x}_{T-T^{l-1}}\right\|_2 \tag{9}$$

that $\left\|\mathbf{x}_{T-\tilde{T}_l}\right\|_2$ and $\left\|\mathbf{x}_{T-T^{l-1}}\right\|_2$ stand for residue signal coefficient vector corresponding to the support set estimate $\tilde{T}_l$ and residue signal based upon the estimate of supp($\mathbf{x}$) before the l^{th} iteration of the BSP algorithm, respectively.

Proof: Our proof is divided into two stages. At the first stage, it is shown how the measurement residue $\mathbf{y}_r^{l-1}$ is connected to $\mathbf{x}_{T-T^{l-1}}$,

$$\mathbf{y}_r^{l-1} = \mathbf{D}_{T\cup T^{l-1}}\mathbf{x}_r^{l-1} = \begin{bmatrix}\mathbf{D}_{T-T^{l-1}} & \mathbf{D}_{T^{l-1}}\end{bmatrix}\begin{bmatrix}\mathbf{x}_{T-T^{l-1}} \\ \mathbf{x}_{P,T^{l-1}}\end{bmatrix}, \tag{10}$$

for some $\mathbf{x}_r^{l-1} \in \mathbb{R}^{|T\cup T^{l-1}|}$ and $\mathbf{x}_{P,T^{l-1}} \in \mathbb{R}^{|T^{l-1}|}$. Furthermore,

$$\left\|\mathbf{x}_{P,T^{l-1}}\right\|_2 \leq \frac{\delta_{2dk}}{1-\delta_{2dk}}\left\|\mathbf{x}_{T-T^{l-1}}\right\|_2 \tag{11}$$

To consider (10), it is clear that,

$$\begin{aligned}
\mathbf{y}_r^{l-1} &= \mathrm{resid}(\mathbf{y}, \mathbf{D}_{T^{l-1}}) = \mathrm{resid}(\mathbf{D}_{T-T^{l-1}}\mathbf{x}_{T-T^{l-1}} + \mathbf{D}_{T\cap T^{l-1}}\mathbf{x}_{T\cap T^{l-1}}, \mathbf{D}_{T^{l-1}}) \\
&= \mathrm{resid}(\mathbf{D}_{T-T^{l-1}}\mathbf{x}_{T-T^{l-1}}, \mathbf{D}_{T^{l-1}}) + \mathrm{resid}(\mathbf{D}_{T\cap T^{l-1}}\mathbf{x}_{T\cap T^{l-1}}, \mathbf{D}_{T^{l-1}}) \\
&= \mathrm{resid}(\mathbf{D}_{T-T^{l-1}}\mathbf{x}_{T-T^{l-1}}, \mathbf{D}_{T^{l-1}}) + 0 \\
&\overset{(4)}{=} (\mathbf{D}_{T-T^{l-1}}\mathbf{x}_{T-T^{l-1}}) - \mathbf{D}_{T^{l-1}}((\mathbf{D}_{T^{l-1}}^H\mathbf{D}_{T^{l-1}})^{-1}\mathbf{D}_{T^{l-1}}^H)\mathbf{D}_{T-T^{l-1}}\mathbf{x}_{T-T^{l-1}} \\
&= [\mathbf{D}_{T-T^{l-1}} \quad \mathbf{D}_{T^{l-1}}]\begin{bmatrix}\mathbf{x}_{T-T^{l-1}} \\ \mathbf{x}_{P,T^{l-1}}\end{bmatrix},
\end{aligned}$$

so,

$$\mathbf{x}_{p,T^{l-1}} = -(\mathbf{D}_{T^{l-1}}^H\mathbf{D}_{T^{l-1}})^{-1}\mathbf{D}_{T^{l-1}}^H(\mathbf{D}_{T-T^{l-1}}\mathbf{x}_{T-T^{l-1}}).$$

In addition, we have this BRIP definition, (let $\mathbf{D}_I = \mathbf{D}_{T^{l-1}}^H\mathbf{D}_{T^{l-1}}$)

$$\left\|\mathbf{x}_{P,T^{l-1}}\right\|_2 = \left\|\left(\mathbf{D}_{T^{l-1}}^H\mathbf{D}_{T^{l-1}}\right)^{-1}\mathbf{D}_{T^{l-1}}^H(\mathbf{D}_{T-T^{l-1}}\mathbf{x}_{T-T^{l-1}})\right\|_2$$

$$\leq \frac{1}{1-\delta_{dk}}\left\|\mathbf{D}_{T^{l-1}}^H(\mathbf{D}_{T-T^{l-1}}\mathbf{x}_{T-T^{l-1}})\right\|_2$$

$$\overset{(3)}{\leq} \frac{\delta_{2dk}}{1-\delta_{dk}}\left\|\mathbf{x}_{T-T^{l-1}}\right\|_2$$

$$\overset{(1)}{\leq} \frac{\delta_{2dk}}{1-\delta_{2dk}}\left\|\mathbf{x}_{T-T^{l-1}}\right\|_2$$

and here, proof of the first step is completed. Now, in this step, we show that,

$$\left\|\mathbf{x}_{T-\tilde{T}_l}\right\|_2 \leq \frac{2\delta_{3dk}}{(1-\delta_{3dk})^2}\left\|\mathbf{x}_{T-T^{l-1}}\right\|_2$$

At first, we define,

$$T_\Delta \overset{\Delta}{=} \widetilde{T}^{\,l} - T^{l-1}.$$

Based on the recent definition and considering BSP algorithm , T_Δ contains the K indices corresponding to the largest magnitude entries, the $\mathbf{D}^H\mathbf{y}_r^{l-1}$, therefore,

$$\left\|\mathbf{D}_{T_\Delta}^H\mathbf{y}_r^{l-1}\right\|_2 \geq \left\|\mathbf{D}_T^H\mathbf{y}_r^{l-1}\right\|_2 \geq \left\|\mathbf{D}_{T-T^{l-1}}^H\mathbf{y}_r^{l-1}\right\|_2 \tag{12}$$

By removing the common block columns between $\mathbf{D}_{T_\Delta}$ and $\mathbf{D}_{T-T^{l-1}}$ and noting that $T_\Delta \cap T^{l-1} = \varnothing$, we can write,

$$\left\|\mathbf{D}_{T_\Delta-(T_\Delta\cap(T-T^{l-1}))}^H\mathbf{y}_r^{l-1}\right\|_2 \geq \left\|\mathbf{D}_{T-T^{l-1}-(T_\Delta\cap(T-T^{l-1}))}^H\mathbf{y}_r^{l-1}\right\|_2$$

This implies that,

$$\left\|\mathbf{D}_{T_\Delta-(T_\Delta\cap T)}^H\mathbf{y}_r^{l-1}\right\|_2 \geq \left\|\mathbf{D}_{T-T^{l-1}-(T_\Delta\cap T)}^H\mathbf{y}_r^{l-1}\right\|_2$$

Therefore,

$$\left\|\mathbf{D}_{T_\Delta-T}^H\mathbf{y}_r^{l-1}\right\|_2 \geq \left\|\mathbf{D}_{T-T^{l-1}-T_\Delta}^H\mathbf{y}_r^{l-1}\right\|_2 = \left\|\mathbf{D}_{T-\tilde{T}^l}^H\mathbf{y}_r^{l-1}\right\|_2 \tag{13}$$

The left hand side of (13) can be derived to be equal to,

$$\left\|\mathbf{D}_{T_\Delta-T}^H\mathbf{y}_r^{l-1}\right\|_2 \overset{(10)}{=} \left\|\mathbf{D}_{T_\Delta-T}^H\mathbf{D}_{T\cup T^{l-1}}\mathbf{x}_r^{l-1}\right\|_2$$

$$\overset{(3)}{\leq} \delta_{|(T_\Delta-T)\cup(T\cup T^{l-1})|}\left\|\mathbf{x}_r^{l-1}\right\|_2 = \delta_{|(T_\Delta\cup T\cup T^{l-1})|}\left\|\mathbf{x}_r^{l-1}\right\|_2 \leq \delta_{3dk}\left\|\mathbf{x}_r^{l-1}\right\|_2 \tag{14}$$

And, for the right hand side of (13), using triangular inequality and BRIP definition,

$$\left\|\mathbf{D}_{T-\tilde{T}^l}^H\mathbf{y}_r^{l-1}\right\|_2 = \left\|\mathbf{D}_{T-\tilde{T}^l}^H\mathbf{D}_{T\cup T^{l-1}}\mathbf{x}_r^{l-1}\right\|_2$$

$$= \left\|\mathbf{D}_{T-\tilde{T}^l}^H\mathbf{D}_{(T-\tilde{T}^l)\cup(T\cup T^{l-1}-(T-\tilde{T}^l))}\mathbf{x}_r^{l-1}\right\|_2$$

$$\geq \left\|\mathbf{D}_{T-\tilde{T}^l}^H\mathbf{D}_{T-\tilde{T}^l}(\mathbf{x}_r^{l-1})_{T-\tilde{T}^l}\right\|_2$$

$$-\left\|\mathbf{D}_{T-\tilde{T}^l}^H\mathbf{D}_{(T\cup T^{l-1}-(T-\tilde{T}^l))}(\mathbf{x}_r^{l-1})_{(T\cup T^{l-1}-(T-\tilde{T}^l))}\right\|_2$$

$$\overset{(1),(3)}{\geq} (1-\delta_{dK}) \quad \left\|(\mathbf{x}_r^{l-1})_{T-\tilde{T}^l}\right\|_2 - \delta_{3dk}\left\|\mathbf{x}_r^{l-1}\right\|_2 \tag{15}$$

Combining (13), (14) and (15),

$$\delta_{3dk}\left\|\mathbf{x}_r^{l-1}\right\|_2 \geq (1\text{-}\delta_{dk}) \left\|(\mathbf{x}_r^{l-1})_{T-\tilde{T}^l}\right\|_2 - \delta_{3dk}\left\|\mathbf{x}_r^{l-1}\right\|_2 \Rightarrow \left\|(\mathbf{x}_r^{l-1})_{T-\tilde{T}^l}\right\|_2$$

$$\leq \frac{2\delta_{3dK}}{1-\delta_{dK}} \left\|\mathbf{x}_r^{l-1}\right\|_2 \overset{(1)}{\leq} \frac{2\delta_{3dK}}{1-\delta_{3dK}} \left\|\mathbf{x}_r^{l-1}\right\|_2 \tag{16}$$

Now, the explicit from of $\mathbf{x}_r^{l-1}$ in (10) is manipulated.

Therefore,

$$(\mathbf{x}_r^{l-1})_{T-T^{l-1}} = \mathbf{x}_{T-T^{l-1}} \Rightarrow (\mathbf{x}_r^{l-1})_{T-\tilde{T}} = \mathbf{x}_{T-\tilde{T}} \tag{17}$$

and also,

$$\left\|\mathbf{x}_r^{l-1}\right\|_2 \leq \left\|\mathbf{x}_{T-T^{l-1}}\right\|_2 + \left\|\mathbf{x}_{p,T^{l-1}}\right\|_2$$

$$\overset{(11)}{\leq} (1+\tfrac{\delta_{2dk}}{1-\delta_{2dk}})\left\|\mathbf{x}_{T-T^{l-1}}\right\|_2 \overset{(1)}{\leq} \tfrac{1}{1-\delta_{3dk}}\left\|\mathbf{x}_{T-T^{l-1}}\right\|_2 \tag{18}$$

substitute (17) into (16) and then using (18) shows that,

$$\left\|\mathbf{x}_{T-\tilde{T}^l}\right\|_2 \leq \frac{2\delta_{3dK}}{(1-\delta_{3dK})^2}\left\|\mathbf{x}_{T-T^{l-1}}\right\|_2,$$

which completes the proof of Theorem 2.

Theorem 3. It holds that

$$\left\|\mathbf{x}_{T-T^l}\right\|_2 \leq \frac{1+\delta_{3dK}}{1-\delta_{3dK}}\left\|\mathbf{x}_{T-\tilde{T}^l}\right\|_2,$$

in which $\mathbf{x}_{T-T^l}$ stand for residual signal based upon the estimate of $\operatorname{supp}(\mathbf{x})$ after the l^{th} iteration and $\mathbf{x}_{T-\tilde{T}^l}$ is similar theorem 2.

Before proving this Theorem, when considering the BSP algorithm, it can be concluded that, $\mathbf{x}_p = \mathbf{D}_{\tilde{T}^l}^{\dagger}\mathbf{y}$ is interpreted as projection coefficient vector. Let us define smear vector as,

$$\boldsymbol{\varepsilon} = \mathbf{x}_p - \mathbf{x}_{\tilde{T}^l},$$

which is non-zero when $T \nsubseteq \tilde{T}^l$ [22]. We must show that $\|\boldsymbol{\varepsilon}\|_2$ is small and then consider the main problem in Theorem 3. To get the desired result, the proof is divided into three sections. In the first section, it can be shown that,

$$\|\boldsymbol{\varepsilon}\|_2 \leq \tfrac{\delta_{3dk}}{1-\delta_{3dk}}\left\|\mathbf{x}_{T-\tilde{T}^l}\right\|_2, \tag{19}$$

In the second section, it is shown that,

$$\left\|\mathbf{x}_{T\cap(\tilde{T}^l-T^l)}\right\|_2 \leq 2\|\boldsymbol{\varepsilon}\|_2, \tag{20}$$

and finally,

$$\left\|\mathbf{x}_{T-T^l}\right\|_2 \leq \frac{1+\delta_{3dK}}{1-\delta_{3dK}}\left\|\mathbf{x}_{T-\tilde{T}^l}\right\|_2. \tag{21}$$

Proof: To prove (19), note that $\mathbf{x}$ is supported on T, i.e, $\mathbf{x}_{T^c} = 0$,

$$\mathbf{x}_p = \mathbf{D}^{\dagger}_{\tilde{T}^l}\mathbf{y} = \mathbf{D}^{\dagger}_{\tilde{T}^l}\mathbf{D}_T\mathbf{x}_T = \mathbf{D}^{\dagger}_{\tilde{T}^l}\mathbf{D}_{T\cap\tilde{T}^l}\mathbf{x}_{T\cap\tilde{T}^l}$$
$$+\mathbf{D}^{\dagger}_{\tilde{T}^l}\mathbf{D}_{T-\tilde{T}^l}\mathbf{x}_{T-\tilde{T}^l} = \mathbf{D}^{\dagger}_{\tilde{T}^l}[\mathbf{D}_{T\cap\tilde{T}^l}\quad \mathbf{D}_{\tilde{T}^l-T}]\begin{bmatrix}\mathbf{x}_{T\cap\tilde{T}^l}\\ 0\end{bmatrix}$$
$$+\mathbf{D}^{\dagger}_{\tilde{T}^l}\mathbf{D}_{T-\tilde{T}^l}\mathbf{x}_{T-\tilde{T}^l} = \mathbf{D}^{\dagger}_{\tilde{T}^l}\mathbf{D}_{\tilde{T}^l}\mathbf{x}_{\tilde{T}^l} + \mathbf{D}^{\dagger}_{\tilde{T}^l}\mathbf{D}_{T-\tilde{T}^l}\mathbf{x}_{T-\tilde{T}^l}$$
$$= \mathbf{x}_{\tilde{T}^l} + \mathbf{D}^{\dagger}_{\tilde{T}^l}\mathbf{D}_{T-\tilde{T}^l}\mathbf{x}_{T-\tilde{T}^l}. \tag{22}$$

In addition to,

$$\|\boldsymbol{\varepsilon}\|_2 = \left\|\mathbf{x}_p - \mathbf{x}_{\tilde{T}^l}\right\|_2 \overset{(22)}{=} \left\|\left(\mathbf{D}^H_{\tilde{T}^l}\mathbf{D}_{\tilde{T}^l}\right)^{-1}(\mathbf{D}^H_{\tilde{T}^l}(\mathbf{D}_{T-\tilde{T}^l}\mathbf{x}_{T-\tilde{T}^l}))\right\|_2$$
$$\overset{\text{BRIP,(3)}}{\leq} \frac{\delta_{3dk}}{1-\delta_{dk}}\left\|\mathbf{x}_{T-\tilde{T}^l}\right\|_2 \leq \frac{\delta_{3dk}}{1-\delta_{3dk}}\left\|\mathbf{x}_{T-\tilde{T}^l}\right\|_2.$$

To prove (20), assume an arbitrary index set $T' \subseteq \tilde{T}^l$ that $T \cap T' = \varnothing$ and $|T'| = K$. It is clear that T' exists because $|\tilde{T}^l - T| \geq K$.
Since,

$$(\mathbf{x}_p)_{T'} = (\boldsymbol{\varepsilon} + \mathbf{x}_{\tilde{T}^l})_{T'} = \boldsymbol{\varepsilon}_{T'} + (\mathbf{x}_{\tilde{T}^l})_{T'} = \boldsymbol{\varepsilon}_{T'} + \mathbf{x}_{T'} = \boldsymbol{\varepsilon}_{T'}.$$

so, we have

$$\left\|(\mathbf{x}_p)_{T'}\right\|_2 \leq \|\boldsymbol{\varepsilon}\|_2$$

According to the BSP algorithm, $\tilde{T}^l - T^l$ is chosen to contain the K smallest projection coefficients. Therefore,

$$\left\|(\mathbf{x}_p)_{\tilde{T}^l - T^l}\right\|_2 \leq \left\|(\mathbf{x}_p)_{T'}\right\|_2 \leq \|\boldsymbol{\varepsilon}\|_2. \tag{23}$$

By decomposition of the left hand side of (23),

$$\left\|(\mathbf{x}_p)_{\tilde{T}^l - T^l}\right\|_2 = \left\|(\boldsymbol{\varepsilon} + \mathbf{x}_{\tilde{T}^l})_{\tilde{T}^l - T^l}\right\|_2 = \left\|\boldsymbol{\varepsilon}_{\tilde{T}^l - T^l} + \mathbf{x}_{\tilde{T}^l - T^l}\right\|_2$$
$$\geq \left\|\mathbf{x}_{\tilde{T}^l - T^l}\right\|_2 - \left\|\boldsymbol{\varepsilon}_{\tilde{T}^l - T^l}\right\|_2 \Rightarrow \left\|\mathbf{x}_{\tilde{T}^l - T^l}\right\|_2 \leq \left\|(\mathbf{x}_p)_{\tilde{T}^l - T^l}\right\|_2 + \|\boldsymbol{\varepsilon}\|_2 \tag{24}$$

By combination of (23) and (24) and $\mathbf{x}_{\tilde{T}^l - T^l} = \mathbf{x}_{T\cap(\tilde{T}^l - T^l)}$, we have,

$$\left\|\mathbf{x}_{T\cap(\tilde{T}^l - T^l)}\right\|_2 \leq 2\|\boldsymbol{\varepsilon}\|_2.$$

This derives the proof (20). Note that this result shows that the energy focused on the error signal components is small.
Finally to prove (21), since

$$\mathbf{x}_{T-T^l} = [\mathbf{x}^H_{T\cap(\tilde{T}^l - T^l)}, \mathbf{x}^H_{T-\tilde{T}^l}]^H,$$

we have,

$$\left\|\mathbf{x}_{T-T^l}\right\|_2 \leq \left\|\mathbf{x}_{T\cap(\tilde{T}^l-T^l)}\right\|_2 + \left\|\mathbf{x}_{T-\tilde{T}^l}\right\|_2 \overset{(20)}{\leq} 2\|\boldsymbol{\varepsilon}\|_2 + \left\|\mathbf{x}_{T-\tilde{T}^l}\right\|_2$$
$$\overset{(19)}{\leq} \left(\frac{2\delta_{3dk}}{1-\delta_{3dk}}+1\right)\left\|\mathbf{x}_{T-\tilde{T}^l}\right\|_2 = \frac{1+\delta_{3dk}}{1-\delta_{3dk}}\left\|\mathbf{x}_{T-\tilde{T}^l}\right\|_2$$

which completes the proof.

Theorem 4 is now considered, which helps to prove the sufficient condition presented in Theorem 1. Before it, we note that in the BSP algorithm the residue that was shown as "$\mathbf{y}_r$" should be decreased in each iteration, i.e.,

$$\left\|\mathbf{y}_r^l\right\|_2 < \left\|\mathbf{y}_r^{l-1}\right\|_2$$

Theorem 4. For each iteration of the BSP algorithm we have,

$$\left\|\mathbf{y}_r^l\right\|_2 \leq \frac{\sqrt{1-\delta_{dK}^2}}{1-\delta_{dK}-\delta_{2dK}} C_K \left\|\mathbf{y}_r^{l-1}\right\|_2,$$

where $C_K = \dfrac{2\delta_{3dK}(1+\delta_{3dK})}{(1-\delta_{3dK})^3}$.

Proof:

$$\left\|\mathbf{y}_r^l\right\|_2 = \left\|\mathrm{resid}(\mathbf{y},\mathbf{D}_{T^l})\right\|_2 = \left\|\mathrm{resid}\left(\mathbf{D}_{T-T^l}\mathbf{x}_{T-T^l},\mathbf{D}_{T^l}\right) + \mathrm{resid}\left(\mathbf{D}_{T^l}\mathbf{x}_{T^l},\mathbf{D}_{T^l}\right)\right\|_2$$
$$= \left\|\mathrm{resid}\left(\mathbf{D}_{T-T^l}\mathbf{x}_{T-T^l},\mathbf{D}_{T^l}\right)+0\right\|_2 \overset{(7)}{\leq} \left\|\mathbf{D}_{T-T^l}\mathbf{x}_{T-T^l}\right\|_2 \leq \sqrt{1+\delta_{dK}}\left\|\mathbf{x}_{T-T^l}\right\|_2$$
$$\overset{(21)}{\leq} \sqrt{1+\delta_{dk}}\,\frac{1+\delta_{3dk}}{1-\delta_{3dk}}\left\|x_{T-\tilde{T}^l}\right\|_2 \overset{(9)}{\leq} \sqrt{1+\delta_{dk}}\,\frac{2\delta_{3dk}(1+\delta_{3dk})}{(1-\delta_{3dk})^3}\left\|\mathbf{x}_{T-T^{l-1}}\right\|_2 \tag{25}$$

In addition

$$\left\|\mathbf{y}_r^{l-1}\right\|_2 = \left\|\mathrm{resid}(\mathbf{y},\mathbf{D}_{T^{l-1}})\right\|_2 = \left\|\mathrm{resid}\left(\mathbf{D}_{T-T^{l-1}}\mathbf{x}_{T-T^{l-1}},\mathbf{D}_{T^{l-1}}\right)\right\|_2$$
$$\overset{(7)}{\geq} \frac{1-\delta_{dk}-\delta_{2dk}}{1-\delta_{dk}}\left\|\mathbf{D}_{T-T^{l-1}}\mathbf{x}_{T-T^{l-1}}\right\|_2 \geq \frac{1-\delta_{dk}-\delta_{2dk}}{1-\delta_{dk}}\sqrt{1-\delta_{dk}}\left\|\mathbf{x}_{T-T^{l-1}}\right\|_2$$
$$= \frac{1-\delta_{dk}-\delta_{2dk}}{\sqrt{1-\delta_{dk}}}\left\|\mathbf{x}_{T-T^{l-1}}\right\|_2 \Rightarrow \left\|\mathbf{x}_{T-T^{l-1}}\right\|_2 \leq \frac{\sqrt{1-\delta_{dk}}}{1-\delta_{dk}-\delta_{2dk}}\left\|\mathbf{y}_r^{l-1}\right\|_2 \tag{26}$$

Substituting (26) in (25),

$$\left\|\mathbf{y}_r^l\right\|_2 \leq \frac{\sqrt{1-\delta_{dk}{}^2}}{1-\delta_{dk}-\delta_{2dk}}.\frac{2\delta_{3dk}(1+\delta_{3dk})}{(1-\delta_{3dk})^3}\left\|\mathbf{y}_r^{l-1}\right\|_2 \tag{27}$$

Now, theorem 1 can be proven. From iteration stopping criterion, we have,

$$\frac{\sqrt{1-\delta_{dk}{}^2}}{1-\delta_{dk}-\delta_{2dk}}.\frac{2\delta_{3dk}(1+\delta_{3dk})}{(1-\delta_{3dk})^3} < 1 \tag{28}$$

To solve the above Inequality, there are several methods which can be used.

Method 1: Substitute δ_{3dK} for δ_{dK} and δ_{2dK} and change it into an inequality with only one variable and considering the permitted limits for variable δ_{3dK}, it can be calculated (as it was calculated in [22]). However, there are different methods with higher accuracy such as those shown below.

Method 2: Using Maclaurin series expansion we have,

$$(1-x)^{1/2} = 1-1/2x-(\alpha),$$

$$\text{for some } \alpha > 0 \quad (\text{if } x > 0) \quad \Rightarrow \qquad 1 - 1/2x > (1-x)^{1/2}. \tag{29}$$

Therefore,

$$\frac{\sqrt{1-\delta_{dK}^2}}{1-\delta_{dK}-\delta_{2dK}} \overset{(29)}{<} \frac{1-1/2\delta_{dK}^2}{1-\delta_{dK}-\delta_{2dK}}. \tag{30}$$

Hence, sufficient condition for the Inequality (28) is

$$\frac{1-1/2\delta_{dK}^2}{1-\delta_{dK}-\delta_{2dK}} \cdot \frac{2\delta_{3dK}(1+\delta_{3dK})}{(1-\delta_{3dK})^3} < 1 \tag{31}$$

It is now possible to write δ_{3dK} versus δ_{dK} and δ_{2dK}, as $0 < \delta_{dK} \le \delta_{2dK} \le \delta_{3dK} < 1$ and $\delta_{dK} + \delta_{2dK} \ne 1/2$, the above problem would be changed into a non-linear optimization problem where different methods can be used to solve them.

Method 3: There is an easier method to solve this problem and its accuracy is dependent on the steps that are used. Since the limits of our results are known $(0 < \delta_{3dK} < 1)$, all of the answers can be checked using simple computer programming precisely (for example, 0.0001) and the final answer can be achieved. The relationship $\delta_{dK} \le \delta_{2dK} \le \delta_{3dK}$ is also always considered.

Method 3 has been selected to solve the Inequality (28). Then the result is obtained as $\delta_{3dK} < 0.1672$, so the correctness of Theorem 1 is confirmed. As can be observed, the dependency of the upper limit to the value of K was eliminated and we reached the upper limit of δ_{3dK} which was also independent of K.

5. SIMULATION RESULTS

With the purpose of evaluating the complexity and accuracy of our proposed algorithm, some simulation experiments have been carried out which are considered in the following. In our simulation, N and L values are assumed to be 1000 and 400, respectively. Based on uniqueness theorem [23], the unique condition of the sparsest solution is: $\|\mathbf{x}\|_0 = (K \times d) < \frac{L}{2} = 200$.

In order to change the block sparsity for a certain block length (d), the value of K is varied from 1 to $\frac{200}{d}$ (for distinctive values of d = 2,5,8). For every value of K in a certain value of d, source vector of $\mathbf{x}$ should be artificially generated. With the purpose of generating this block K-sparse signal, at first, the nonzero K block location is randomly selected. Then, a value of normal distribution $N(0,1)$ is chosen for every d element of selected block. The rest of K selected blocks are considered to be equal to zero. For the sampling matrix $\mathbf{D}$, elements are randomly chosen on the normal distribution and later it's columns elements are normalized to unity. Considering vector $\mathbf{x}$ and matrix $\mathbf{D}$, the measurement vector $\mathbf{y}$ is calculated. Based on the value of matrix $\mathbf{D}$, vector $\mathbf{y}$ and our proposed algorithm, the vector $\mathbf{x}(\hat{\mathbf{x}})$ can be estimate. These experiments for each distinct value of d are repeated 200 times. For each iteration, the dimension of vectors and matrices stay constant, whilst the values of source vector $\mathbf{x}$ and sampling matrix $\mathbf{D}$ are randomly selected. The results are as follows:

A) In figures 1,2, and 3, the frequency of exact reconstruction which is the criteria for the accuracy [5] against different levels are drawn for three different conditions d =2, d =5 and d =8. Our interest in these figures is the block sparsity level at which the frequency of exact reconstruction drops below 1. This level is named "critical block sparsity". As it can be concluded from the results of simulation, critical block sparsity of BSP algorithm by far exceeds the BOMP and mixed l_2/l_1-norm algorithms. We also see experimentally that BSP has better performance than SP when the source signal is block-sparse. The advantages of the BSP algorithm compared to the other considered methods, even though the block length is small, are more distinguishable.

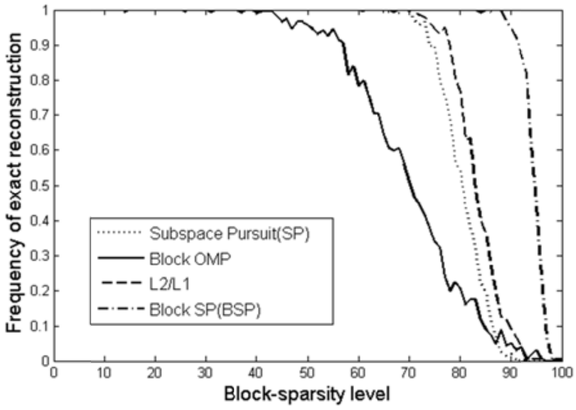

Fig. 1. performance of SP, BOMP, L2/L1, and BSP for d=2.

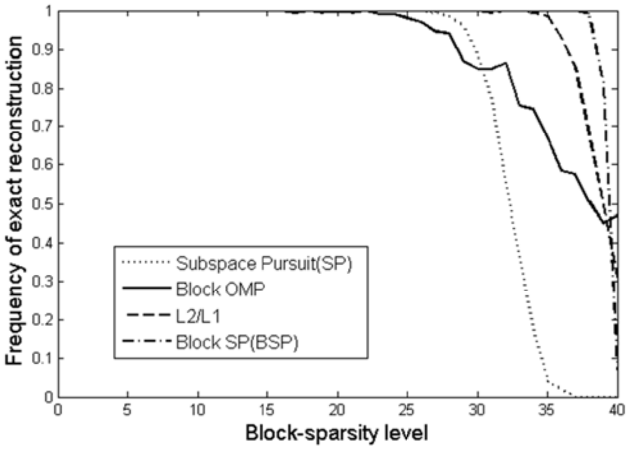

Fig. 2. performance of SP, BOMP, L2/L1, and BSP for d=5

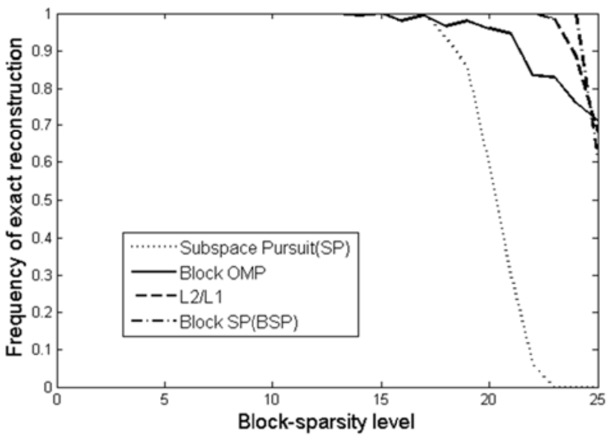

Fig. 3. performance of SP, BOMP, L2/L1, and BSP for d=8

B) In this part, we consider the complexity of the proposed method and compare it with other methods. For evaluation of complexity, the CPU time is used. Although it is not an exact criterion, it gives us an approximate of the complexity. Our simulation is presented in MATLAB 7.10 environment using an Intel core 2 Duo 2.8GHz CPU with 4GB of memory, and under MicrosoftWindows 7 operating system. In Table 1, the average CPU time for different block length and block sparsity levels for all algorithms are tabulated.

It is concluded from numerical results that BSP is faster than BOMP. Consequently, selecting K

indices of blocks whilst considering the error occurance (BSP) is a faster method than selecting K indices in the K numbers of separate stages (BOMP). L2/L1 is the slowest because of using convex optimization methods. In addition, it can be seen that BSP has a better performance than SP in the recovery of block sparse signals. This is because BSP works directly with the block version of SP.

Table 1. The average CPU time for all algorithms

Block Length (d)	Block Sparsity (K)	SP (second)	BOMP (second)	L2/L1 (second)	BSP (second)
2	1	0.015	0.006	9.98	0.012
2	20	0.07	0.03	11.1	0.018
2	100	0.075	0.47	12.1	0.064
5	1	0.0178	0.004	5.3	0.007
5	8	0.068	0.014	5.7	0.008
5	40	0.088	0.19	7.6	0.057
8	1	0.03	0.005	4.2	0.006
8	5	0.06	0.01	4.4	0.007
8	25	0.08	0.12	5.9	0.06

6. CONCLUSION

In this paper, for block-sparse signals recovery, an efficient new method is introduced, namely BSP. The motivation for studying block-sparse signals is that in many applications the nonzero elements of the interested signal appear in clusters. It is experimentally shown that BSP produces better recovery performance than SP, BOMP, and L2/L1 as well, because it is usually faster than them. In addition, the recovery performance of BSP using Block RIP was analyzed. It was shown that if sampling matrix $\mathbf{D}$ satisfies Block RIP with the constant parameter $\delta_{3dK} < 0.1672$, then BSP can exactly reconstruct any block K-sparse signal.

Acknowledgment: This paper is the results of a research project that was supported by Telecommunication Research Center with grant number 500/6475. Hereby, we appreciate the financial support of the Telecommunication Research Center.

REFERENCES

1. Candès, E. J., Romberg, J. K. & Tao, T. (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, Vol. 52, No. 2, pp. 489-509.
2. Donoho, D. L. (2006). Compressed sensing. *IEEE Trans. Inf. Theory*, Vol. 52, No. 4, pp. 1289-1306.
3. Meng, J., Li, Y., Nguyen, N., Yin, W. & Han, Z. (2012). Compressive sensing based high-resolution channel estimation for OFDM system. *IEEE Trans. Signal Process.*, Vol. 6, No. 1, pp. 15-25.
4. Nikoofar, H. R. & Sharafat, A. R. (2010). Modulation classification for burst-mode QAM signals in multipath fading channels. *IJST, Transactions of Electrical Engineering*, Vol. 34, No. B3, pp. 257-274.
5. Candès, E. J., Rudelson, M., Tao, T. & Vershynin, R. (2005). Error correction via linear programming. *Proc. 46th Annu. IEEE Symp. Foundations Computer Science (FOCS)*, pp. 668-681.
6. http://www.dsp.ece.rice.edu/cs/.
7. Eldar, Y. C., Kuppinger, P. & B¨olcskei, H. (2010). Block-sparse signals: Uncertainty relations and efficient recovery. *IEEE Trans. Signal Process.*, Vol. 58, No. 6, pp. 3042-3054.
8. Candès, E. J. & Tao, T. (2005). Decoding by linear programming. *IEEE Trans. Inf. Theory*, Vol. 51, No. 12, pp. 4203-4215.
9. Chen, S. S., Donoho, D. L. & Saunders, M. A. (1999). Atomic decomposition by Basis Pursuit. *SIAM J. Sci. Comput.*, Vol. 20, No. 1, pp. 33-61.

10. Mallat, S. G. & Zhang, Z. (1993). Matching pursuits and time-frequency dictionaries. *IEEE Trans. Signal Process.*, Vol. 41, No. 12, pp. 3397-3415.
11. Candès, E. J., Romberg, J. & Tao, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, Vol. 59, pp. 1207-1223.
12. Davenport, M. A. & Wakin, M. B. (2010). Analysis of orthogonal matching pursuit using the restricted isometry property. *IEEE Trans. Inf. Theory*, Vol. 56, No. 9, pp. 4395-4401.
13. Mohimani, H., Babaie-Zadeh, M. & Jutten, C. (2009). A fast approach for overcomplete sparse decomposition based on smoothed l_0 norm. *IEEE Trans. Signal Process.*, Vol. 57, No. 1, pp. 289-301.
14. Mishali, M. & Eldar, Y. C. (2009). Blind multi-band signal reconstruction: Compressed sensing for analog signals. *IEEE Trans. Signal Process.*, Vol. 57, No. 3, pp. 993-1009.
15. Cotter, S. & Rao, B. (2002). Sparse channel estimation via matching pursuit with application to equalization. *IEEE Trans. on Comm.*, Vol. 50, No. 3, pp. 374-377.
16. Stojnic, M., Parvaresh, F. & Hassibi, B. (2010). On the reconstruction of block-sparse signals with an optimal number of measurements. *IEEE Trans. Signal Process.*, Vol. 57, No. 8, pp. 3075–3085.
17. Eldar, Y. C. & Mishali, M. (2009). Robust recovery of signals from a structured union of subspaces. *IEEE Trans. Inf. Theory,* Vol. 55, No. 11, pp. 5302-5316.
18. Baraniuk, R. G., Cevher, V., Duarte, M. F. & Hegde, C. (2010). Model-based compressive sensing. *IEEE Trans. Inf. Theory,* Vol. 56, No. 4, pp. 1982-2001.
19. Needell, D. & Tropp, J. A. (2009). CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, Vol. 26, No. 3, pp. 301-321.
20. Hamidi Ghalehjegh, S., Babaie-Zadeh, M. & Jutten, C. (2010). Fast block-sparse decomposition based on SL0. *9th Int. Conf. Latent Variable Analysis and Signal Separation (LVA/ICA),* Saint Malo, France.
21. Wang, J. J., Li, G., Zhang, H. & Wang, X. (2011). Analysis of block OMP using block RIP. *CoRR,* vol. abs/1104.1071.
22. Dai, W. & Milenkovic, O. (2009). Subspace pursuit for compressive sensing signal reconstruction. *IEEE Trans. Inf. Theory,* Vol. 55, No. 5, pp. 2230-2249.
23. Donoho, D. L. & Elad, M. (2003). Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization. *Proc. Natl. Acad. Sci.*, Vol. 100, No. 5, pp. 2197-2202.

APPENDIX A. PROOF OF LEMMA 1

Based on Definition 1, for all d -block K' -sparse $\mathbf{x}'$ we have,

$$(1-\delta_d)\ \ \|\mathbf{x}'\|_2^2 \le \|\mathbf{D}\mathbf{x}'\|_2^2 \le (1+\delta_d)\,\|\mathbf{x}'\|_2^2$$

and if we let $\delta_{dK'}$ be the infimum of all δ_d,

$$(1-\delta_{dK'})\ \ \|\mathbf{x}'\|_2^2 \le \|\mathbf{D}\mathbf{x}'\|_2^2 \le (1+\delta_{dK'})\,\|\mathbf{x}'\|_2^2$$

while $K \le K'$ implies that all d -block K -sparse $\mathbf{x} \in \mathrm{P}^N$ can satisfy the above statement, therefore,

$$(1-\delta_{dK'})\ \ \|\mathbf{x}\|_2^2 \le \|\mathbf{D}\mathbf{x}\|_2^2 \le (1+\delta_{dK'})\,\|\mathbf{x}\|_2^2$$

since δ_{dK} is defined as the infimum of all parameters δ_d that satisfy the recent inequality, we have,

$$\delta_{dK} \le \delta_{dK'}.$$

APPENDIX B. PROOF OF LEMMA 2

In the first part of Lemma, it is clear that if $\|\mathbf{a}\|_2 = 0$ or $\|\mathbf{b}\|_2 = 0$, then it is satisfied. Now, suppose they are both non-zero values and define

$$\mathbf{a}' = \mathbf{a}/\ \|\mathbf{a}\|_2\ ,\ \mathbf{b}' = \mathbf{b}/\ \|\mathbf{b}\|_2\ ,\ \mathbf{x}' = \mathbf{D}_I\mathbf{a}',\quad \mathbf{y}' = \mathbf{D}_J\mathbf{b}'.$$

Considering the definition of BRIP and the following value,

$$\left\| \begin{bmatrix} \mathbf{a}' \\ \mathbf{b}' \end{bmatrix} \right\|_2^2 = 2,$$

we have,

$$2(1-\delta_{d(|I|+|J|)}) \le \| \mathbf{x}'+\mathbf{y}' \|_2^2 = \left\| [\mathbf{D}_I \quad \mathbf{D}_J] \begin{bmatrix} \mathbf{a}' \\ \mathbf{b}' \end{bmatrix} \right\|_2^2 \le 2(1+\delta_{d(|I|+|J|)}), \qquad \text{(B.1)}$$

and similarly,

$$2(1-\delta_{d(|I|+|J|)}) \le \| \mathbf{x}'-\mathbf{y}' \|_2^2 = \left\| [\mathbf{D}_I \quad \mathbf{D}_J] \begin{bmatrix} \mathbf{a}' \\ -\mathbf{b}' \end{bmatrix} \right\|_2^2 \le 2(1+\delta_{d(|I|+|J|)}). \qquad \text{(B.2)}$$

From (B.1) and (B.2) we have,

$$<\mathbf{x}',\mathbf{y}'> = \frac{\| \mathbf{x}'+\mathbf{y}' \|_2^2 - \| \mathbf{x}'-\mathbf{y}' \|_2^2}{4} \le \delta_{d(|I|+|J|)},$$

and

$$= $$

$$-<\mathbf{x}',\mathbf{y}'> \quad = \frac{\| \mathbf{x}'-\mathbf{y}' \|_2^2 - \| \mathbf{x}'+\mathbf{y}' \|_2^2}{4} \le \delta_{d(|I|+|J|)},$$

Therefor

$$\frac{|<\mathbf{D}_I\mathbf{a},\mathbf{D}_J\mathbf{b}>}{\|\mathbf{a}\|_2\|\mathbf{b}\|_2} = \ |<\mathbf{x}',\mathbf{y}'>| \le \delta_{d(|I|+|J|)},$$

and then,

$$|<\mathbf{D}_I\mathbf{a},\mathbf{D}_J\mathbf{b}>| \le \ \ |<\mathbf{D}_I\mathbf{a},\mathbf{D}_J\mathbf{b}>| \le \|\mathbf{a}\|_2\|\mathbf{b}\|_2\,\delta_{d(|I|+|J|)}.$$

To prove the second part of Lemma, it should be assumed that for any $\mathbf{q} \in \mathbb{R}^{d|I|}$: $\|\mathbf{q}\|_2 = 1$

$$\| \mathbf{D}_I^H \mathbf{D}_J \mathbf{b} \|_2 = \max_{\mathbf{q}} |(\mathbf{D}_I\mathbf{q})^H \mathbf{D}_J\mathbf{b}| = \max_{\mathbf{q}} |<\mathbf{D}_I\mathbf{q},\mathbf{D}_J\mathbf{b}>|$$

$$\le \max_{\mathbf{q}} \delta_{d(|I|+|J|)} \|\mathbf{q}\|_2\|\mathbf{b}\|_2 = \delta_{d(|I|+|J|)}\|\mathbf{b}\|_2$$

APPENDIX C. PROOF OF LEMMA 3

We have,

$$\mathbf{D}_I^H\mathbf{y}_r = \mathbf{D}_I^H(\mathbf{y}-\mathbf{D}_I(\mathbf{D}_I^H\mathbf{D}_I)^{-1}\mathbf{D}_I^H\mathbf{y}) = \mathbf{D}_I^H\mathbf{y} - (\mathbf{D}_I^H\mathbf{D}_I)(\mathbf{D}_I^H\mathbf{D}_I)^{-1}\mathbf{D}_I^H\mathbf{y} = 0.$$

APPENDIX D. PROOF OF LEMMA 4

Suppose that $\mathbf{y}_p = \mathbf{D}_J\mathbf{D}_J^H\mathbf{y} = \mathbf{D}_J\mathbf{x}_p$ and $\mathbf{y} = \mathbf{D}_I\mathbf{x}$. From (2) and BRIP definition we have,

$$|<\mathbf{y}_p,\mathbf{y}>|=|<\mathbf{D}_J\mathbf{x}_p,\mathbf{D}_I\mathbf{x}>| \le \delta_{d(|I|+|J|)}\|\mathbf{x}_p\|_2\|\mathbf{x}\|_2$$

$$\le \delta_{d(|I|+|J|)}\frac{\|\mathbf{y}_p\|_2}{\sqrt{1-\delta_{d|J|}}}\frac{\|\mathbf{y}\|_2}{\sqrt{1-\delta_{d|I|}}} \le \frac{\delta_{d(|I|+|J|)}}{1-\delta_{\max(d|I|,d|J|)}}\|\mathbf{y}_p\|_2\|\mathbf{y}\|_2 \tag{D.1}$$

In addition , for the left side of the above inequality we can write,

$$<\mathbf{y}_p,\mathbf{y}>=<\mathbf{y}_p,\mathbf{y}_p+\mathbf{y}_r> = \|\mathbf{y}_p\|_2^2 \tag{D.2}$$

Therefore, from (D.1) and (D.2) we have,

$$\|\mathbf{y}_p\|_2 \le \frac{\delta_{d(|I|+|J|)}}{1-\delta_{max(d|I|,d|J|)}}\|\mathbf{y}\|_2$$

In order to prove the Eq. (7) , triangular inequality is used as follows,

$$\|\mathbf{y}_r\|_2 = \|\mathbf{y}-\mathbf{y}_p\|_2 \ge \|\mathbf{y}\|_2 - \|\mathbf{y}_p\|_2$$

$$\overset{(6)}{\ge}\left(1-\frac{\delta_{d(|I|+|J|)}}{1-\delta_{max(d|I|,d|J|)}}\right)\|\mathbf{y}\|_2 \tag{D.3}$$

furthermore,

$$\|\mathbf{y}_r\|_2^2 + \|\mathbf{y}_p\|_2^2 = \|\mathbf{y}\|_2^2 \tag{D.4}$$

where $\|\mathbf{y}_p\|_2^2 \ge 0$

From (D.3) and (D.4) we have,

$$\left(1-\frac{\delta_{d(|I|+|J|)}}{1-\delta_{max(d|I|,d|J|)}}\right)\|\mathbf{y}\|_2 \le \|\mathbf{y}_r\|_2 \le \|\mathbf{y}\|_2$$