SCHOOL OF ENGINEERING - STI
SIGNAL PROCESSING INSTITUTE
Lorenzo Peotta and Pierre Vandergheynst

CH-1015 LAUSANNE
Telephone: +4121 6932601
Telefax: +4121 6937600
e-mail: `lorenzo.peotta@epfl.ch`

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

MP in Block Quasi-Incoherent Dictionaries

Lorenzo Peotta and Pierre Vandergheynst
Swiss Federal Institute of Technology Lausanne (EPFL)

Signal Processing Institute Technical Report
TR-ITS-2003.009
December 15th, 2003

MP in Block Quasi-Incoherent Dictionaries

Lorenzo Peotta and Pierre Vandergheynst
Signal Processing Institute
Swiss Federal Institute of Technology
Lausanne, Switzerland
{lorenzo.peotta,pierre.vandergheynst}@epfl.ch

I. Block Incoherent Dictionaries

Given a redundant dictionary $\mathcal{D}$, we consider the following L-subset decomposition $\mathcal{D} = \bigcup_{l=1}^{L} B_l$, and we call *blocks* the L subsets of atoms B_l, $l = 1, \cdots, L$. The *block coherence* is defined as the maximum coherence between any two atoms, taken from different blocks.

Definition 1: The block coherence μ_B, given a block decomposition $\mathcal{D} = \bigcup_{l=1}^{L} B_l$, is

$$\mu_B \triangleq \max_{i \neq j} \max_{k,l} |< g_k^i, g_l^j >|, \tag{1}$$

where g_k^i is the k^{th} atom from the block B_i.

Definition 2: A dictionary is then said *block incoherent* if there exists a decomposition such that the *block coherence* μ_B is small.

The block coherence considers similarities between atoms from two different block. In order to refine the analysis of the coherence, we introduce another function, called the $Babel\ block$ function, that represents the coherence between sets of m blocks $B_I = \bigcup_{i \in I} B_i$, with $Card(I) = m$.

Definition 3: Let $\mathcal{D} = \bigcup_{l=1}^{L} B_l$ denote a decomposition, and $B_I = \bigcup_{i \in I} B_i$ represent a set of m blocks. The $Babel\ block$ function is

$$\mu_{1_B}(m) \triangleq \max_{I, s.t. |I| = m} \max_{j \notin I, l} \sum_{i \in I} \max_{k_i} |< g_{k_i}^i, g_l^j >|. \tag{2}$$

Definition 4: A given dictionary $\mathcal{D}$ is said to be *block quasi-incoherent*, if we can find a block decomposition such that $\mu_{1_B}(m)$ grows slowly with m.

The *block coherence* μ_B considers coherence between two blocks, and the *Babel block* function $\mu_{1_B}(m)$ measures coherence between m blocks. Notice that the $Babel\ block$ function is bounded by the $block\ coherence$: $\mu_{1_B}(m) \leq m\mu_B$. The definitions of the previous functions is the extension of the coherence μ and the $Babel$ function $\mu_1(m)$ introduced by Donoho, Huo, Elad [1], [2]. The $Babel$ function was developed and utilized by Troop in the Exact Recovery Theorem [3]. We need now also to consider the coherence within a single block. Generally, a single block B_i has a strong coherence (i.e., the *Babel function* grows quickly). For a more detailed analysis, we are however interested in a function that represent the coherence of a particular subset of functions in B_i, and we call it the $Bore$ function $\xi(B_i)$.

Definition 5: The $Bore$ function related to a block B_i is

$$\xi(B_i) \triangleq \min_{\iota, s.t. |B_\iota| = rank(B_i)} \max_k \sum_{l \neq k} |< g_l^i, g_k^i >|, \tag{3}$$

where B_ι is a set of independent atoms from B_i, i.e., $B_\iota \subset B_i$ and $span(B_\iota) = span(B_i)$.

The $Bore$ function $\xi(B_i)$ indicates how much the atoms in a block "can speak different languages". In other words, it illustrates how close a basis constructed with atoms from block B_i is to an orthogonal basis that spans the range $\mathcal{R}(B_i)$. The set of atoms, i.e. B_ι, where the Bore function is minimal, is called $B_{i^\star}$. If $\xi(B_i) = 0$, we can find a set $B_{i^\star} \subset B_i$ that is an orthogonal basis for $span(B_i)$. The extension of the $Bore$ function to the dictionary $\mathcal{D}$ is finally defined as $\xi(\mathcal{D}) = \max_i \xi(B_i)$.

II. Exact Block Selection

Using the definitions defined in Sec. I, we prove in this section that, given a block incoherent dictionary $\mathcal{D}$ and a signal f, the Matching Pursuit (MP) algorithm can recover a block-sparse representation of f. We consider here the restricted problem $(\mathcal{D}, B_I)$-*SPARSE*, which means that f is a linear combination of atoms belonging to a subset of m blocks, $B_I = \bigcup_{i \in I} B_i$.

Firstly, we find a single sufficient condition under which Matching Pursuit recovers atoms from a given set of incoherent blocks B_I. In this case, we say that MP chooses atoms from *correct* blocks B_i $i \in I$. Let represent B_I as an operator or matrix, and let B_I^+ denote its pseudoinverse.

Theorem 1: Let $\mathcal{D}$ a block incoherent dictionary and $B_I = \bigcup_{i \in I} B_i$. If the signal $f \in \mathcal{V}_I = span(B_I)$, under the recovery condition

$$\eta(B_I) \triangleq \max_{g \notin B_I} \|B_I^+ g\|_1 < 1 \tag{4}$$

then we have that MP:

1) picks up atoms only from correct blocks B_i $i \in I$,
2) converges exponentially to f.

Proof of Theorem 1. We follow the proof for "Exact Recovery" theorem [3]. Suppose that $r_{n-1} \in \mathcal{V}_I$. If an atom g_{n-1} from B_I is selected by the Matching Pursuit algorithm, then $r_n = r_{n-1} - < g_{n-1} r_{n-1} > g_{n-1}$ belongs to $\mathcal{V}_I$, with $r_0 = f$. The vector $B_I^T r_{n-1}$ lists the inner products between the residual r_{n-1} and all the atoms from the blocks B_i, $i \in I$; taking the ∞ norm of this vector we have that $\|B_I^T r_{n-1}\|_\infty$ is the largest of these inner products in magnitude, where B_I^T represents the complex conjugate of B_I. The number $\|B_{\bar{I}}^T r_{n-1}\|_\infty$ corresponds to the largest inner product in magnitude between r_{n-1} and an atom that does not belong to B_I, that means $g \in B_{\bar{I}}$. An atom is selected from the correct block B_i, $i \in I$, when the following quotient is less than one

$$\rho(r_{n-1}) \triangleq \frac{\|B_{\bar{I}}^T r_{n-1}\|_\infty}{\|B_I^T r_{n-1}\|_\infty} < 1 \tag{5}$$

By assumption, $r_{n-1} \in \mathcal{V}_I$, and $B_I B_I^+$ is a projector onto the range of B_I. Therefore, using the properties of the pseudoinverse, $r_{n-1} = (B_I^+)^T B_I^T r_{n-1}$ and

$$\begin{aligned} \rho(r_{n-1}) &= \frac{\|B_{\bar{I}}^T (B_I^+)^T B_I^T r_{n-1}\|_\infty}{\|B_I^T r_{n-1}\|_\infty} \\ &\leq \|B_{\bar{I}}^T (B_I^+)^T\|_{\infty,\infty} \end{aligned}$$

where the matrix norm $\|\cdot\|_{p,p}$ is the norm "induced" by the vector norm $\|\cdot\|_p$. Using properties of the matrix norm we obtain

$$\begin{aligned} \rho(r_{n-1}) &\leq \|B_{\bar{I}}^T (B_I^+)^T\|_{\infty,\infty} \\ &= \|B_I^+ B_{\bar{I}}\|_{1,1} \\ &= \max_{g \in B_{\bar{I}}} \|B_I^+ g\|_1, \end{aligned}$$

so $\rho(r_{n-1}) \leq \eta(B_I) < 1$ which means that MP selects an atom from B_I. By induction the first part is proved.

To prove the second part, we just notice that MP is faced with a finite dimentional space $\mathcal{V}_I$, and we know that MP in a finite dimentional space is exponentially convergent.

□

As a corollary, the following theorem gives a condition under which right block selection is in force when f belongs to the *span* of an arbitrary set of m incoherent blocks.

Theorem 2: Let $\mathcal{D}$ a block incoherent dictionary and B_I an arbitrary set of m blocks and $R = \max_i rank(B_i)$. If the signal $f \in \mathcal{V}_I$ and

$$R\,\mu_{1_B}(m) + \xi(\mathcal{D}) + R\,\mu_{1_B}(m-1) < 1 \tag{6}$$

then we have that MP:

1) picks up atoms only from the correct blocks,
2) converges exponentially to f.

Proof of Theorem 2. The proof is again given by induction. We suppose that $r_{n-1} \in \mathcal{V}_I$. If an atom from B_I is selected, then $r_n \in \mathcal{V}_I$. We indicate with $B_I' = \bigcup_{i \in I} B_{i^\star}$ the union of the m sets associated to the m blocks B_i in the definition (5) of the $Bore$ function. Now we define $B_{I^\star}$ to be a set of linear independent atoms from B_I' such that $|B_{I^\star}| = rank(B_I')$. It follows that $span(B_{I^\star}) = span(B_I) = \mathcal{V}_I$, $B_{I^\star}$ is a basis for $\mathcal{V}_I$, therefore $r_{n-1} = (B_{I^\star}^+)^T B_{I^\star}^T r_{n-1}$ and

$$\begin{aligned} \rho(r_{n-1}) &= \frac{\|B_{\bar{I}}^T r_{n-1}\|_\infty}{\|B_I^T r_{n-1}\|_\infty} \\ &= \frac{\|B_{\bar{I}}^T (B_{I^\star}^+)^T B_{I^\star}^T r_{n-1}\|_\infty}{\|B_I^T r_{n-1}\|_\infty} \end{aligned}$$

since $B_{I^\star} \subset B_I$ we have that $\|B_I^T r_{n-1}\|_\infty \geq \|B_{I^\star}^T r_{n-1}\|_\infty$ and

$$\begin{aligned}\rho(r_{n-1}) &\leq \frac{\|B_{\bar{I}}^T (B_{I^\star}^+)^T B_{I^\star}^T r_{n-1}\|_\infty}{\|B_{I^\star}^T r_{n-1}\|_\infty} \\ &\leq \|B_{\bar{I}}^T (B_{I^\star}^+)^T\|_{\infty,\infty} \\ &= \max_{g \in B_{\bar{I}}} \|B_{I^\star}^+ g\|_1 .\end{aligned}$$

Now we can expand the pseudoinverse and apply the norm bound $\|Ax\|_1 \leq \|A\|_{1,1}\|x\|_1$

$$\begin{aligned}\rho(r_{n-1}) &\leq \max_{g \in B_{\bar{I}}} \|(B_{I^\star}^T B_{I^\star})^{-1} B_{I^\star}^T g\|_1 \\ &\leq \|(B_{I^\star}^T B_{I^\star})^{-1}\|_{1,1} \max_{g \in B_{\bar{I}}} \|B_{I^\star}^T g\|_1 . \qquad (7)\end{aligned}$$

We can easily bound the second term of the right part of (7) using the $Babel\ block$ function

$$\begin{aligned}\max_{g \in B_{\bar{I}}} \|B_{I^\star}^T g\|_1 &= \max_{g \in B_{\bar{I}}} \sum_{\psi \in B_{I^\star}} |<\psi, g>| \\ &\leq R\, \mu_{1_B}(m) \qquad (8)\end{aligned}$$

where $R = \max_i rank(B_i)$. In order to bound the first term of the right part of (7) we use the Von Neumann series to compute the inverse $(B_{I^\star}^T B_{I^\star})^{-1}$. Writing $B_{I^\star}^T B_{I^\star} = \mathcal{I} + A$, where $\mathcal{I}$ is the identity matrix, and under the condition that $\|A\|_{1,1} < 1$, it follows that :

$$\begin{aligned}\|(B_{I^\star}^T B_{I^\star})^{-1}\|_{1,1} &= \|(\mathcal{I} + A)^{-1}\|_{1,1} = \|\sum_{k=0}^{\infty} (-A)^k\|_{1,1} \\ &\leq \sum_{k=0}^{\infty} \|A\|_{1,1}^k = \frac{1}{1 - \|A\|_{1,1}} .\end{aligned}$$

The matrix A has zero diagonal and the values out of diagonal correspond to the inner product between atoms from $B_{I^\star}$, taking into account the structure of $B_{I^\star}$ (it is composed by m incoherent blocks) we can bound the norm using the $Bore$ and $Babel\ block$ function:

$$\begin{aligned}\|A\|_{1,1} &= \max_k \sum_{j \neq k} |<g_j^{I^\star}, g_k^{I^\star}>| \\ &\leq \xi(\mathcal{D}) + R\, \mu_{1_B}(m-1) \qquad (9)\end{aligned}$$

and putting together the bounds (8),(9) into (7) we obtain for $\rho(r_{n-1})$ the bound

$$\rho(r_{n-1}) \leq \frac{R\, \mu_{1_B}(m)}{1 - (\xi(\mathcal{D}) + R\, \mu_{1_B}(m-1))} .$$

So under the condition

$$R\, \mu_{1_B}(m) + \xi(\mathcal{D}) + R\, \mu_{1_B}(m-1) < 1$$

it follows that $\rho(r_{n-1}) < 1$ and MP selects an atom from the correct block B_i, by induction the first part is proved. For the second part, we are in the same condition as in theorem 1.

□

Using the bound for the $Babel\ block$ function $\mu_{1_B}(m-1) \leq \mu_{1_B}(m) \leq m\, \mu_B$ it follows from Theorem 2 that, if the signal f belongs to the $span$ of m blocks, then MP recovers atoms from the correct blocks when

$$m < \frac{1 - \xi(\mathcal{D})}{2R\, \mu_B} .$$

III. Rate of Convergence

Another important factor that determines the quality of a signal expansion is the rate of convergence of the approximation. This rate can also be bounded with the help of the coherence defined previously, in the case of block incoherent dictionaries.

Theorem 3: If the signal $f \in \mathcal{V}_I$ and $R\,\mu_{1_B}(m) + \xi(\mathcal{D}) + R\,\mu_{1_B}(m-1) < 1$, then MP picks up atoms only from the correct blocks at each step and

$$\|r_n\|_2^2 \leq \|f\|_2^2 \left(1 - \frac{1 - \xi(\mathcal{D}) - R\,\mu_{1_B}(m-1)}{R\,m}\right)^n . \tag{10}$$

Proof of Theorem 3. From theorem 2 we know that under condition (6) MP picks up atoms from the right set of blocks B_I. At each step the residual belongs to the space $\mathcal{V}_I$, and the energy of the residual is

$$\begin{aligned} \|r_n\|_2^2 &= \|r_{n-1}\|_2^2 - \max_{g \in B_I} |\langle r_{n-1}, g\rangle|^2 \\ &= \|r_{n-1}\|_2^2 \left(1 - \frac{\max_{g \in B_I} |\langle r_{n-1}, g\rangle|^2}{\|r_{n-1}\|_2^2}\right). \end{aligned} \tag{11}$$

In order to bound the decay of the residual energy, we need a lower bound for

$$\frac{\max_{g \in B_I} |\langle r, g\rangle|^2}{\|r\|_2^2} \tag{12}$$

with $r \in \mathcal{V}_I = span(B_I) = span(B_{I^*})$, where the set B_{I^*} is defined in the prof of theorem 2. Since we can write r like a combination of elements from B_{I^*}, $r = B_{I^*}c = \sum_i g_i c_i$, we have

$$\begin{aligned} \|r\|_2^2 &= \langle r, r\rangle = \left\langle \sum g_i c_i, r \right\rangle \\ &\leq \sum_i |\langle g_i, r\rangle|\,|c_i| \\ &\leq \max_{g \in B_{I^*}} |\langle g, r\rangle|\,\|c\|_1 , \end{aligned} \tag{13}$$

and we obtain the following lower bound for (12)

$$\frac{\max_{g \in B_I} |\langle r, g\rangle|^2}{\|r\|_2^2} \geq \frac{\max_{g \in B_{I^*}} |\langle r, g\rangle|^2}{\|r\|_2^2} \geq \frac{\|r\|_2^2}{\|c\|_1^2}. \tag{14}$$

We wish to change $\|c\|_1$ with $\|c\|_2$ in order to bound (14) with the minimum norm of the operator B_{I^*}. We know that $rank(B_{I^*}) = p \leq Rm$, where $R = \max_i rank(B_i)$, wich means that $\|c\|_0 \leq p \leq Rm$, and using the Gensen inequality we have

$$\begin{aligned} \|c\|_2^2 &= \sum_{i=1}^{p} c_i^2 = p \sum_{i=1}^{p} \frac{1}{p} |c_i|^2 \\ &\geq p \left(\sum_{i=1}^{p} \frac{1}{p} |c_i|\right)^2 = \frac{1}{p} \|c\|_1^2 , \end{aligned}$$

and putting the upper bound $\|c\|_1^2 \leq p\,\|c\|_2^2 \leq Rm\,\|c\|_2^2$ into (14) we obtain

$$\frac{\max_{g \in B_I} |\langle r, g\rangle|^2}{\|r\|_2^2} \geq \frac{\|r\|_2^2}{Rm\,\|c\|_2^2}. \tag{15}$$

Using the Thin Singular Value Decomposition $B_{I^*} = U\Sigma V^*$, where U and V are orthogonal while Σ is diagonal with full rank since B_{I^*} has full rank, we can bound the operator B_{I^*} writing

$$\begin{aligned} \|r\|_2^2 &= \|B_{I^*}c\|_2^2 = c^* V \Sigma^2 V^* c \qquad (y = V^* c) \\ &= y^* \Sigma^2 y = \sum_i \sigma_i^2 y_i^2 \\ &\geq \sigma_{min}^2 \|y\|_2^2 = \sigma_{min}^2 \|c\|_2^2 . \end{aligned} \tag{16}$$

The square singular values of B_{I^*} coincide to the eigenvalues of the the Gram matrix $G = B_{I^*}^* B_{I^*}$, indeed Σ^2 and G are similar matrixes. The eigenvalue $\lambda_{min} = \sigma_{min}^2$ can be bound using the Geršgorin disc theorem: every eigenvalue of G lies in one of the p discs

$$\text{Disc}_k = \left\{ z : |G_{kk} - z| \leq \sum_{j \neq k} |G_{jk}| \right\}.$$

The matrix G has unitary diagonal since the normalization of the atoms. Taking into account the block incoherent structure of B_{I^*} we can bound the sum above with

$$|1 - \lambda_{min}| \leq \sum_{j \neq k} |G_{jk}| \leq \xi(\mathcal{D}) + R\mu_{1_B}(m-1),$$

and the square minimum singular value $\sigma_{min}^2 \geq 1 - \xi(\mathcal{D}) - R\mu_{1_B}(m-1)$. Putting this bound into (16) and (15) we obtain

$$\frac{\max_{g \in B_I} |\langle r, g \rangle|^2}{\|r\|_2^2} \geq \frac{1 - \xi(\mathcal{D}) - R\mu_{1_B}(m-1)}{Rm},$$

and finally from (11) we end the proof

$$\begin{aligned} \|r_n\|_2^2 &\leq \|r_{n-1}\|_2^2 \left(1 - \frac{1 - \xi(\mathcal{D}) - R\mu_{1_B}(m-1)}{Rm}\right) \\ &\leq \|f\|_2^2 \left(1 - \frac{1 - \xi(\mathcal{D}) - R\mu_{1_B}(m-1)}{Rm}\right)^n. \end{aligned}$$

□

Theorem 4: If the signal $f \in \mathcal{V}_I$ and $R\mu_{1_B}(m) + \xi(\mathcal{D}) + R\mu_{1_B}(m-1) < 1$, then MP picks up atoms only from the correct blocks at each step and

$$\|r_n\|_2^2 \leq \|f\|_2^2 \left(1 - \frac{\beta^2}{m}\right)^n, \tag{17}$$

where $\beta = \min_i \beta_i$, and β_i is related to the redundancy and structure of block B_i [4].

Lemma 1: Let $f \in V_I = span(B_I)$ with $B_I = \bigcup_{i=1}^m B_i$, if we indicate with f^i the projection of f into the space $V_i = span(B_i)$, it follows that

$$\|f\|_2^2 \leq \sum_{i=1}^m \left\|f^i\right\|_2^2. \tag{18}$$

Proof of lemma 1. Suppose for semplicity $m = 2$. Let $d_i = rank(B_i)$ and $d_I = rank(B_I)$, we build an orthogonal basis for $\mathcal{V}_I = span(B_1 \cup B_2)$ taking d_1 orthonormal vectors from $\mathcal{V}_1$, we collect them into the matrix E_1, and $d_I - d_1$ orthonormal vectors from V_2 that are orthogonal to V_1, we collect them into E_2. We build an orthogonal basis for $\mathcal{V}_2$ starting from E_2 and adding $d_1 + d_2 - d_I$ orthonormal vectors from $\mathcal{V}_2$ that are orthogonal to E_2, which we collect into E_2^*. With this notation we have that E_1 is an orthogonal basis for $\mathcal{V}_1$, $[E_2|E_2^*]$ for $\mathcal{V}_2$ and $[E_1|E_2]$ for $\mathcal{V}_I$. We can generalize this procedure to $m > 2$: $\mathcal{V}_i = span([E_i|E_i^*])$ and $\mathcal{V}_I = span([E_1|\cdots|E_m])$. It is now easy to proof (18), writing f^i and f respect the basis that we defined above

$$\begin{aligned} f^i &= [E_i|E_i^*] \left[\begin{array}{c} a_i \\ \hline a_i^* \end{array}\right] \\ f &= [E_1|\cdots|E_m] \left[\begin{array}{c} a_1 \\ \hline \vdots \\ \hline a_m \end{array}\right], \end{aligned}$$

the energy of f projected into $\mathcal{V}_i$ is $\left\|f^i\right\|_2^2 = \|a_i\|_2^2 + \|a_i^*\|_2^2$, and we can conclude

$$\begin{aligned} \sum_{i=1}^m \left\|f^i\right\|_2^2 &= \sum_{i=1}^m \|a_i\|_2^2 + \sum_{i=1}^m \|a_i^*\|_2^2 \\ &= \|f\|_2^2 + \sum_{i=1}^m \|a_i^*\|_2^2 \\ &\geq \|f\|_2^2. \end{aligned}$$

□

Proof of Theorem 4. By induction we know that the sequence of residuals $r_n \in \mathcal{V}_I$. The normalization of the atoms implies

$$\|r_n\|_2^2 = \|r_{n-1}\|_2^2 - \max_{g\in B_I} |\langle r_{n-1}, g\rangle|^2 . \quad (19)$$

In order to caracterize the decay of the residual energy, we need a meaningfull lower bound for $\max_{g\in B_I} |\langle r, g\rangle|^2$ where $r \in \mathcal{V}_I$. If MP selects an atom from the block B_j, it foolows that

$$\begin{aligned} \max_{g\in B_I} |\langle r, g\rangle|^2 &\geq \max_{g\in B_j} \left|\langle r^j, g\rangle\right|^2 \\ &\geq \beta^2 \left\|r^j\right\|_2^2 \\ &\geq \beta^2 \frac{\|r\|_2^2}{m} \end{aligned} \quad (20)$$

where $\beta = \min(\beta_i)$ and β_i is the structural redundancy factor of the block B_i [4]. Inequality (20) can be derived analizing the case of residual r with energy uniformly spread in all spaces $\mathcal{V}_i$, that means $\left\|r^i\right\|_2^2 = K$ for all $i \in I$. Using *lemma 1* it follows that

$$\left\|r^j\right\|_2^2 \geq \frac{\|r\|_2^2}{m}. \quad (21)$$

When the energy is not uniformly spread, it means there is at least one component r^k, $k \in I$, with energy bigger than (21) and MP will selects an atom from B_k. Putting (20) in (19) we end the proof

$$\begin{aligned} \|r_n\|_2^2 &\leq \|r_{n-1}\|_2^2 - \beta^2 \frac{\|r_{n-1}\|_2^2}{m} \\ &\leq \|f\|_2^2 \left(1 - \frac{\beta^2}{m}\right)^n . \end{aligned}$$

□

Since the dimension of the vector spaces generated by $span(B_i)$ is supposed to be small, we expect β to be close to one. The term m that divides β^2 could be substituted, taking into account the block incoherent structure of the dictionary. If we have β^2 close to one, and m replaced by $h(m) \ll m$, we thus prove the good approximation behavior of Matching Pursuit for structured signals, that we observe on experimental results.

We claim that taking track of the blocks selected, due to the block quasi-incoherent structure of the dictionary, the energy bound (21) can be refined. Here there is some arguments for the case of $f \in \mathcal{V}_I = span(B_I)$ with $B_I = B_1 \cup B_2$. The energy of the residual after two iteration is bounded by

$$\|r_2\|_2^2 \leq \|f\|_2^2 \left(1 - \beta^2 \left(1 - \mu_B\right)\right) ,$$

with $\beta = \min(\beta_1, \beta_2)$.

References

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

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

[3] Tropp, J., "Greed is good : Algorithmic results for sparse approximation," Texas Institute for Computational Engineering and Sciences, Tech. Rep., 2003.

[4] Frossard P. and Vandergheynst P., "Redundancy in Non-Orthogonal Transform," in *Proceedings of IEEE International Symposium on Information Theory*, June 2001.