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
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 $\mu_B$, given a block decomposition $\mathcal{D} = \bigcup_{l=1}^{L} B_l$, is

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

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 $\mu_B$ 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 $\text{Card}(I) = m$.

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

$$\mu_1(m) \triangleq \max_{I, \text{Card}(I) = m} \max_{j \in I} \sum_{k \in I} \max_{i,k} | < g_k^i, g_l^j |.$$

Definition 4: A given dictionary $\mathcal{D}$ is said to be block quasi-incoherent, if we can find a block decomposition such that $\mu_1(m)$ grows slowly with $m$.

The block coherence $\mu_B$ considers coherence between two blocks, and the Babel block function $\mu_1(m)$ measures coherence between $m$ blocks. Notice that the Babel block function is bounded by the block coherence: $\mu_1(m) \leq \mu_B$. The definitions of the previous functions is the extension of the coherence $\mu$ 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 represents 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

$$\xi(B_i) \triangleq \min_{I, \text{Card}(I) = \text{rank}(B_i)} \max_k \sum_{l \neq k} | < g_l^i, g_k |,$$

where $B_i$ is a set of independent atoms from $B_i$, i.e., $B_i \subset B_i$ and $\text{span}(B_i) = \text{span}(B_i)$.

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. $B_i$, where the Bore function is minimal, is called $B_i^\perp$. If $\xi(B_i) = 0$, we can find a set $B_i^\perp \subset B_i$ that is an orthogonal basis for $\text{span}(B_i)$. The extension of the Bore function to the dictionary $\mathcal{D}$ is finally defined as $\xi(D) = \max_i \xi(B_i)$.

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 $f \in V_I = \text{span}(B_I)$, under the recovery condition

$$\eta(B_I) \triangleq \max_{g \notin B_i} \| B_I^+ g \|_1 < 1,$$
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 infinite 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 $||B_I^T r_{n-1}||_\infty$ corresponds to the largest inner product in magnitude between $r_{n-1}$ and an atom that does not belong to $B_I$, that means $g \in B_I$. An atom is selected from the correct block $B_i$, $i \in I$, when the following quotient is less than one

$$\rho(r_{n-1}) = \frac{||B_I^T r_{n-1}||_\infty}{||B_I^T r_{n-1}||_\infty} < 1 \quad (5)$$

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, $r_{n-1} = (B_I^T)^T B_I^T r_{n-1}$ and

$$\rho(r_{n-1}) = \frac{||B_I^T (B_I^+)^T B_I^T r_{n-1}||_\infty}{||B_I^T r_{n-1}||_\infty} \leq ||B_I^T (B_I^+)^T||_{\infty, \infty}$$

where the matrix norm $\| \cdot \|_{\infty, \infty}$ is the norm “induced” by the vector norm $\| \cdot \|_\infty$. Using properties of the matrix norm we obtain

$$\rho(r_{n-1}) \leq ||B_I^T (B_I^+)^T||_{\infty, \infty} = ||B_I^+ B_I||_{1,1} = \max_{g \in B_I} ||B_I^+ g||_1,$$

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 dimensional space $\mathcal{V}_I$, and we know that MP in a finite dimension space is exponentially convergent.

\[\square\]

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

$$R \mu_1(p)(m) + \xi(\mathcal{D}) + R \mu_1(p)(m-1) < 1 \quad (6)$$

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 $B_I' = \bigcup_{i \in I} B_i$, the union of the $m$ sets associated to the $m$ blocks $B_i$ in the definition (5) of the $Borel$ function. Now we define $B_I'$ to be a set of linear independent atoms from $B_I'$ such that $|B_I'| = rank(B_I')$. It follows that $\text{span}(B_I') = \text{span}(B_I) = \mathcal{V}_I$, $B_I'$ is a basis for $\mathcal{V}_I$, therefore $r_{n-1} = (B_I')^T B_I^T r_{n-1}$ and

$$\rho(r_{n-1}) = \frac{||B_I^T r_{n-1}||_\infty}{||B_I^T r_{n-1}||_\infty} \leq ||B_I^T (B_I')^T B_I^T r_{n-1}||_\infty \leq \frac{||B_I^T (B_I')^T B_I^T r_{n-1}||_\infty}{||B_I^T r_{n-1}||_\infty}$$
since \( B_T \subset B_I \) we have that \( \| B_T^T r_{n-1} \|_\infty \geq \| B_T^T r_{n-1} \|_\infty \) and

\[
\rho(r_{n-1}) \leq \frac{\| B_T^T (B_T^*)^T B_T^T r_{n-1} \|_\infty}{\| B_T^T r_{n-1} \|_\infty} \\
\leq \| B_T^T (B_T^*)^T \|_{\infty, \infty} \\
= \max_{g \in B_T} \| B_T^T g \|_1. 
\]

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

\[
\rho(r_{n-1}) \leq \max_{g \in B_T} \| (B_T^T B_T^*)^{-1} B_T^T g \|_1 \\
\leq \| (B_T^T B_T^*)^{-1} \|_{1,1} \max_{g \in B_T} \| B_T^T g \|_1. 
\] (7)

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

\[
\max_{g \in B_T} \| B_T^T g \|_1 = \max_{g \in B_T} \sum_{\psi \in B_T^*} | < \psi, g > | \\
\leq R \mu_{1_n}(m) \quad (8)
\]

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

\[
\| (B_T^T B_T^*)^{-1} \|_{1,1} = \| (I + A)^{-1} \|_{1,1} = \| \sum_{k=0}^{\infty} (-A)^k \|_{1,1} \\
\leq \sum_{k=0}^{\infty} \| A \|_1^k = \frac{1}{1 - \| A \|_{1,1}}. 
\]

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

\[
\| A \|_{1,1} = \max_k \sum_{j \neq k} | < g_j^r, g_k^r > | \\
\leq \xi(D) + R \mu_{1_n}(m - 1) \quad (9)
\]

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

\[
\rho(r_{n-1}) \leq \frac{R \mu_{1_n}(m)}{1 - (\xi(D) + R \mu_{1_n}(m - 1))}. 
\]

So under the condition

\[
R \mu_{1_n}(m) + \xi(D) + R \mu_{1_n}(m - 1) < 1
\]

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.

\( \square \)

Using the bound for the Babel block function \( \mu_{1_n}(m - 1) \leq \mu_{1_n}(m) \leq m \mu_B \) 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

\[
m < \frac{1 - \xi(D)}{2R \mu_B}. 
\]
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 \( R \mu_{1,n}(m) + \xi(\mathcal{D}) + R \mu_{1,n}(m-1) < 1 \), then MP picks up atoms only from the correct blocks at each step and

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

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

\[
\|r_n\|_2^2 = \|r_{n-1}\|_2^2 - \max_{g \in B_I} |\langle r_{n-1}, g \rangle|^2
\]

\[
= \|r_{n-1}\|_2^2 \left(1 - \frac{\max_{g \in B_I} |\langle r_{n-1}, g \rangle|^2}{\|r_{n-1}\|_2^2}\right). \tag{11}
\]

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 \( r \in \mathcal{V}_I = \text{span}(B_I) = \text{span}(B_{I*}) \), where the set \( B_{I*} \) is defined in the proof 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

\[
\|r\|_2^2 = \langle r, r \rangle = \langle \sum_i g_i c_i, r \rangle
\]

\[
\leq \sum_i |\langle g_i, r \rangle| |c_i|
\]

\[
\leq \max_{g \in B_I} |\langle g, r \rangle| \|c\|_1, \tag{13}
\]

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

\[
\|c\|_2^2 = \sum_{i=1}^p c_i^2 = p \sum_{i=1}^p \frac{1}{p} |c_i|^2
\]

\[
\geq p \left( \sum_{i=1}^p \frac{1}{p} |c_i| \right)^2 = \frac{1}{p} \|c\|_1^2,
\]

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 \( \Sigma \) is diagonal with full rank since \( B_{I*} \) has full rank, we can bound the operator \( B_{I*} \) writing

\[
\|r\|_2^2 = \|B_{I*}c\|_2^2 = c^* \Sigma^2 V^* c \quad (y = V^* c)
\]

\[
= y^* \Sigma^2 y = \sum_i \sigma_i^2 y_i^2
\]

\[
\geq \sigma_{\min}^2 \|y\|_2^2 = \sigma_{\min}^2 \|c\|_2^2. \tag{16}
\]
The square singular values of $B_{I^*}$ coincide to the eigenvalues of the Gram matrix $G = B_{I^*}B_{I^*}^T$, indeed $\Sigma^2$ and $G$ are similar matrices. The eigenvalue $\lambda_{\text{min}} = \sigma_{\text{min}}^2$ 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

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

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

$$\max_{g \in B_{I^*}} \frac{|\langle f, g \rangle|^2}{\|g\|_2^2} \geq \frac{1 - \xi(D) - R \mu_{1,n} (m - 1)}{R m},$$

and finally from (11) we end the proof

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

\[ \Box \]

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

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

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

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

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

**Proof of Lemma 1.** Suppose for simplicity $m = 2$. Let $d_i = \text{rank}(B_i)$ and $d_I = \text{rank}(B_I)$, we build an orthogonal basis for $\mathcal{V}_I = \text{span}(B_1 \cup B_2)$ 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 $\mathcal{V}_2$ that are orthogonal to $\mathcal{V}_1$, 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 $[E_1 I_2]$ for $\mathcal{V}_I$. We can generalize this procedure to $m > 2$: $\mathcal{V}_i = \text{span}([E_i | E_i^*])$ and $\mathcal{V}_I = \text{span}([E_1 | \cdots | E_m])$. It is now easy to proof (18), writing $f$ and $f^i$ respect the basis that we defined above

$$f^i = [E_i | E_i^*] \begin{bmatrix} a_i \\ a_i^* \end{bmatrix},$$

$$f = [E_1 | \cdots | E_m] \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix},$$

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

$$\sum_{i=1}^m \|f^i\|_2^2 = \sum_{i=1}^m \|a_i\|_2^2 + \sum_{i=1}^m \|a_i^*\|_2^2 = \|f\|_2^2 + \sum_{i=1}^m \|a_i^*\|_2^2 \geq \|f\|_2^2.$$  

\[ \Box \]
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

$$
\|r_n\|^2 = \|r_{n-1}\|^2 - \max_{g \in B_1} |\langle r_{n-1}, g \rangle|^2.
$$

(19)

In order to characterize the decay of the residual energy, we need a meaningful lower bound for $\max_{\hat{g} \in B_1} |\langle \hat{r}, g \rangle|^2$ where $\hat{r} \in \mathcal{V}_I$. If MP selects an atom from the block $B_j$, it follows that

$$
\max_{\hat{g} \in B_j} |\langle \hat{r}, g \rangle|^2 \geq \max_{\hat{g} \in B_j} |\langle \hat{r}_j, g \rangle|^2 \geq \beta^2 \|\hat{r}_j\|^2_2 \geq \beta^2 \|\hat{r}\|^2_2 \frac{\|r\|^2_2}{m}.
$$

(20)

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

$$
\|\hat{r}\|^2_2 \geq \|r\|^2_2 \frac{m}{m}.
$$

(21)

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

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

Since the dimension of the vector spaces generated by $\text{span}(B_i)$ is supposed to be small, we expect $\beta$ to be close to one. The term $m$ that divides $\beta^2$ could be substituted, taking into account the block incoherent structure of the dictionary. If we have $\beta^2$ 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 $f \in \mathcal{V}_I = \text{span}(B_I)$ with $B_I = B_1 \cup B_2$. The energy of the residual after two iteration is bounded by

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

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

REFERENCES


