COOPERATIVE ORTHOGONAL MATCHING PURSUIT STRATEGIES FOR SPARSE APPROXIMATION BY PARTITIONING

LAURA REBOLLO-NEIRA†

Abstract. Cooperative Orthogonal Matching Pursuit strategies are considered for approximating a signal partition, subjected to a global constraint on sparsity. The approach is designed to produce a high quality sparse approximation of the whole signal, using highly coherent redundant dictionaries. The cooperation takes place by ranking the partition units for their sequential stepwise approximation and is realized by means of i) forward steps for the upgrading of an approximation and/or ii) backward steps for the corresponding downgrading.

Key words. Sparse Representations, Greedy Algorithms, Orthogonal Matching Pursuit.

AMS subject classifications. 41A45

1. Introduction. Any kind of information given as a sequence of, say \(N\), samples is called a digital signal of size \(N\). Important classes of signals, such as audio and vision data, are characterized by samples containing a good deal of redundancy, though. Consequently, if properly processed, the signal information content can be accessed from a reduced data set. Transformations for data reduction are said to produce a sparse representation of a signal if they can accurately reproduce the information the signal conveys, with significantly less points that those by which the original signal is given. Amongst the most popular of such transformations are the Discrete Cosine Transform (DCT) and the Discrete Wavelet Transform (DWT). Like most other transformations for signal processing the DCT and DWT do not modify the signal size. A sparse representation is achieved, a posteriori, by disregarding the least relevant points in the transformed domain. On the other hand, much higher sparsity in a signal representation may be achieved by allowing for the expansion of the transformed domain, thereby leaving room for dedicated transformations adapted to the particular signal. Such a framework involves a large redundant set called ‘dictionary’. The aim is to represent a signal as a superposition of, say \(K\), dictionary’s elements, which are called ‘atoms’, with \(K\) much less than the signal size.

Given a redundant dictionary, the problem of finding the sparsest approximation of a signal, up to some predetermined error, is an NP-hard problem [1]. Consequently, the interest in practical applications lies in the finding of those solutions which, being sparse in relation to other possible representations, are also easy and fast to construct. Sparse practical solutions can be found by what are known as Pursuit Strategies. Here the discrimination between two broad categories is in order: i) The Basis Pursuit based approaches, which endeavor to obtain a sparse solution by minimization of the 1-norm [2]. ii) Greedy algorithms which look for a sparse solution by step-wise selection of dictionary’s atoms. Practical greedy algorithms, which originated as regression techniques in statistics [3], have been popularized in signal processing applications as Matching

†Mathematics Department, Aston University, B3 7ET, Birmingham, UK
Pursuit [4] and Orthogonal Matching Pursuit (OMP) [5]. These approaches, which in principle consider the stepwise selection of single atoms, have been extended to multiple atom selection [6]. Dedicated algorithms such as Stagewise Orthogonal Matching Pursuit [7], Compressive Sampling Matching Pursuit [8], and Regularized Orthogonal Matching Pursuit [9] are known to be effective within the context of the emerging theory of sampling, called compressive sensing/sampling. This theory asserts that sparsity of a representation may also lead to more economical data collection [10–14]. However, the reconstruction problem is of a very particular nature: It is assumed that a signal is sparse in an orthogonal basis and the goal is to reconstruct the signal from a reduced number of measures. On the contrary, in this Communication we address the traditional representation matter: the signal is assumed to be completely given by its samples. The aim is to produce a high quality approximation of all those samples, as a K-term superposition of atoms belonging to a highly coherent dictionary [16]. In this case, minimization of the 1-norm is not effective and step wise greedy selection of single atoms benefits sparsity results.

In practice, when trying to approximate real life signals using a redundant dictionary, there is a need to approximate by partitioning. While this requirement normally comes from storage demands, it does not represent a disadvantage. On the contrary, to capture local properties of signals such as images, or music, non-local orthogonal transforms are also applied on partitions to obtain superior results.

In particular, we tailor the OMP greedy algorithm for approximation by partitioning. The examples presented here clearly show that a global constraint on sparsity, rather than quality constraints on the individual partition’s elements, may benefit enormously the quality of the signal approximation. This is true even if the actual approximation of each unit is performed individually.

Notational convention. Throughout the paper \( \mathbb{R} \) and \( \mathbb{N} \) indicate the sets of real and natural numbers, respectively. Boldface letters are used to indicate Euclidean vectors or matrices, whilst standard mathematical fonts indicate components, e.g., \( f \in \mathbb{R}^N \), \( N \in \mathbb{N} \) is a vector of components \( f(i), \ i = 1, \ldots, N \). A partition of a signal \( f \in \mathbb{R}^N \) is represented as a set of disjoint pieces \( f_q \in \mathbb{R}^{N_b} \), \( q = 1, \ldots, Q \), which for simplicity are assumed to be all of the same size and such that \( QN_b = N \), i.e., \( f = \cup_{q=1}^{Q} f_q \). Notice that the definition implies that, for \( q \neq p \) \( \langle f_q, f_p \rangle = 0 \), where the operation \( \langle \cdot, \cdot \rangle \) indicates the Euclidean inner product. Consequently, \( \langle f, f \rangle = \|f\|_2^2 = \sum_{q=1}^{Q} \|f_q\|_2^2 \).

Paper contributions. Given a partition \( f_q, q = 1, \ldots, Q \) and a fixed dictionary \( D = \{d_n \in \mathbb{R}^{N_b} : \|d_n\| = 1\}_{n=1}^{M} \) to approximate the partition, the following outcome has been recently reported [15]: A significant gain in the sparsity of the signal approximation may be effectively obtained by a greedy pursuit strategy, if the approximation of each piece \( f_q \) (called ‘q-block’) is accomplished in a hierarchized manner. Suppose that the \( k_q \)-term approximation of \( f_q \) is the atomic decomposition:

\[
f_q^{k_q} = \sum_{n=1}^{k_q} c_n^q d_n^q, \quad q = 1, \ldots, Q,
\]
with the atoms $d_{\ell q}^n$, $n = 1, \ldots, k_q$ selected from the dictionary $D$, via a stepwise greedy pursuit strategy. Suppose also that the number of atoms to approximate the whole signal $f$ is a fixed value $K$, i.e., $K = \sum_{q=1}^Q k_q$. A possibility to handle this constraint is to consider a hierarchized selection of the pieces $f_q$ to be approximated in each approximation step. Some remarkable results of this strategy, which has been termed Hierarquized Block Wise (HBW) greedy strategy, are illustrated in [15] by means of the greedy algorithms Matching Pursuit (MP) [4] and Orthogonal Matching Pursuit (OMP) [5]. When MP and OMP are applied in the proposed HBW fashion are called HBW-MP and HBW-OMP, respectively. If the goal is to achieve high quality approximations with coherent dictionaries, the performance of HBW-OMP, in relation to HBW-MP, is certainly worth the additional computational cost introduced by OMP.

In this paper we extend the method HBW-OMP by considering:

- A revised version of the approach proposed in [15].
- A HBW backward approach for downgrading the approximation when required.
- An alternative to the HBW-OMP strategy which consists of two stages. The first stage involves the approximation of the pieces $f_q$, $q = 1, \ldots, Q$, up to a tolerance error. The second stage refines the previous approximation by allowing the migration of atoms, from the atomic decomposition of some blocks, called ‘donors’, to the atomic decompositions of another blocks, called ‘receivers’. Since this process is inspired in the Swapping-based-Refinement of greedy strategies introduced in [20], we refer to it as HBW-Swapping-Refinement (HBW-SR) of OMP.

### 2. Hierarchized Blockwise OMP, revised.

As already stated, a signal $f \in \mathbb{R}^N$ will be considered to be the composition of $Q$ identical and disjoint blocks $f = \bigcup_{q=1}^Q f_q$, where $f_q \in \mathbb{R}^{N_b}$ assuming that $N_b Q = N$. Given a dictionary $\{d_n \in \mathbb{R}^{N_b}; \|d_n\| = 1\}_{n=1}^M$, with $M > N$, each of the $Q$ blocks $f_q \in \mathbb{R}^{N_b}$ is approximated by an atomic decomposition of the form:

\[
\tag{2.1} f^k_q(i) = \sum_{n=1}^{k_q} c^{k_q,q}(n)d_{\ell q}^n(i) \quad i = 1, \ldots, N_b, q = 1, \ldots, Q.
\]

For each $q$ the HBW-OMP introduced in [15] selects the atoms $d_{\ell q}^n$ as follows: On setting $R^{0,q} = f_q$, $q = 1, \ldots, Q$ at each iteration the algorithm picks the atoms $d_{\ell q+1}^n \in D$ maximizing the absolute value of the inner products $\langle d_n, R^{k,q}\rangle$, $n = 1, \ldots, M$, $q = 1, \ldots, Q$, i.e.,

\[
\tag{2.2} \ell_{q+1}^q = \arg \max_{n=1,\ldots,M} \sum_{i=1}^{N_b} d_n(i)R^{k,q}(i), \quad \text{with } R^{k,q}(i) = f_q(i) - \sum_{n=1}^{k_q} c^{k,q}(n)d_{\ell q}^n(i).
\]

The coefficients $c^{k,q}(n)$, $n = 1, \ldots, k_q$ in (2.2) are such that the 2-norm $\|R^{k,q}\|$ is minimized. This is ensured by requesting that $R^{k,q} = f_q - \tilde{P}_{\ell_q}^{k,q} f_q$, where $\tilde{P}_{\ell_q}^{k,q}$ is the orthogonal projection operator onto $\mathbb{V}_{\ell_q}^{k_q} = \text{span}(d_{\ell q}^n)_{n=1}^{k_q}$. An implementation as in [17] provides us with the representation of
\[ \hat{P}_{\psi_q} f_q \text{ given by,} \]

\[ (2.3) \quad \hat{P}_{\psi_q} f_q = \sum_{n=1}^{k_q} d_{\ell_q}^n \langle b_{n}^{k_q}, f_q \rangle = \sum_{n=1}^{k_q} c_{k_q, q}(n) d_{\ell_q}^n. \]

For a fixed \( q \) the vectors \( b_{n}^{k_q}, n = 1, \ldots, k_q \) are biorthogonal to the selected atoms \( d_{\ell_q}^n, n = 1, \ldots, k_q \) and span the identical subspace, i.e.,

\[ \forall q \quad \mathbb{V}_{k_q} = \text{span}\{b_{n}^{k_q}, n = 1, \ldots, k_q\} = \text{span}\{d_{\ell_q}^n, n = 1, \ldots, k_q\}. \]

Such vectors can be adaptively constructed through the recursion formula

\[ b_{n}^{k_q+1} = b_{n}^{k_q} - b_{n}^{k_q+1} \langle d_{\ell_{k_q+1}}^n, b_{n}^{k_q} \rangle, \quad n = 1, \ldots, k_q, \]

where \( b_{n}^{k_q+1} = w_{n}^{k_q+1}/\|w_{n}^{k_q+1}\|^2 \),

\[ (2.4) \]

with

\[ w_{n}^{k_q+1} = d_{\ell_{k_q+1}}^n - \sum_{n=1}^{k_q} \frac{w_{n}^{k_q}}{\|w_{n}^{k_q}\|^2} \langle w_{n}^{k_q}, d_{\ell_{k_q+1}}^n \rangle \quad \text{and} \quad w_{1}^{k_q} = d_{\ell_1}. \]

For numerical accuracy in the construction of the orthogonal set \( w_{n}^{k_q}, n = 1, \ldots, k_q + 1 \) at least one re-orthogonalization step is usually needed. This implies to recalculate the vectors as

\[ (2.5) \]

The coefficients in (2.3) are obtained from the inner products

\[ c_{k_q, q}(n) = \langle b_{n}^{k_q}, f_q \rangle, \quad n = 1, \ldots, k_q. \]

For a given number \( K \) the algorithm iterates until the condition \( \sum_{q=1}^{Q} k_q = K \), is met. In other words, the algorithm stops when the maximum number of total atoms allowed for the signal approximation is reached. The next proposition shows that, through the above implementation, the hierarchized selection criterion (2.2) can actually be optimized in a stepwise sense, without increasing in any significant way the computational cost.

**Proposition 2.1.** Let \( \ell_{k_q+1}^q \) be the index arising, for each value of \( q \), from the maximization process

\[ (2.6) \]

\[ \ell_{k_q+1}^q = \arg \max_{n=1, \ldots, M} |\langle d_n, R_{k_q}^q \rangle| = \arg \max_{n=1, \ldots, M} \left| \sum_{i=1}^{N_q} d_n(i) R_{n}^{k_q}(i) \right|. \]
with $R^{k+1}_q$ as in (2.2). In order to minimize the square norm of the total residual $\|R^{k+1}\|^2$ at iteration $k+1$ the atom $d_{kq+1}^q$ to be selected should correspond to the block $q^*$ such that

\begin{equation}
q^* = \arg\max_{q=1,...,Q} |\langle \frac{w_{kq+1}^q}{\|w_{kq+1}^q\|^2} R^{k+1}_q \rangle|,
\end{equation}

with $w_{kq+1}^q$ as in (2.4).

**Proof.** Since at iteration $k+1$ the atomic decomposition of only one block is augmented, by one atom, the total residue at iteration $k+1$ is constructed as

$$R^{k+1} = \bigcup_{p=1}^Q R^{k,p} \cup R^{k+1}_q.$$

Then,

$$\|R^{k+1}\|^2 = \sum_{p=1}^Q \|R^{k,p}\|^2 + \|R^{k+1}_q\|^2.$$

Moreover

$$R^{k+1}_q = f^q - \hat{P}_{kq+1} f^q$$

$$= f^q - \sum_{n=1}^{k+1} \frac{w_n^q}{\|w_n^q\|^2} \langle w_n^q, f^q \rangle$$

$$= f^q - \sum_{n=1}^{k+1} \frac{w_n^q}{\|w_n^q\|^2} \langle w_n^q, f^q \rangle - w_{kq+1}^q \frac{\|w_{kq+1}^q\|^2}{\|R^{k+1}_q\|^2} \langle w_{kq+1}^q, f^q \rangle$$

$$= f^q - \hat{P}_{kq} f^q - w_{kq+1}^q \frac{\|w_{kq+1}^q\|^2}{\|R^{k+1}_q\|^2} \langle w_{kq+1}^q, f^q \rangle$$

$$= R^{k+1}_q - w_{kq+1}^q \frac{\|w_{kq+1}^q\|^2}{\|R^{k+1}_q\|^2} \langle w_{kq+1}^q, f^q \rangle.$$

Since $\hat{P}_{kq} w_{kq+1}^q = 0$ it is true that $\langle w_{kq+1}^q, R^{k+1}_q \rangle = \langle w_{kq+1}^q, f^{kq} \rangle$ and therefore

\begin{equation}
\|R^{k+1}\|^2 = \sum_{p=1}^Q \|R^{k,p}\|^2 - \|\frac{w_{kq+1}^q}{\|w_{kq+1}^q\|^2} f^{kq}\|^2.
\end{equation}

From the last equation it follows that $\|R^{k+1}\|^2$ is minimum for the $q^*$-value corresponding to the maximum value of $|\langle \frac{w_{kq+1}^q}{\|w_{kq+1}^q\|^2} f^{kq} \rangle| = |\langle w_{kq+1}^q, R^{kq} \rangle|$. \qed
Notice that, since vectors $w_{k+1}^q$ are needed for the subsequent calculation of coefficients (c.f. (2.3)) the optimized ranking of blocks (2.7) does not require significant extra computational effort. Only $Q-1$ of these vectors will not be otherwise used when the algorithm stops.

The complexity of the OMP approach for totally independent approximation of the blocks is dominated by the computation of the inner products in (2.2). The additional stepwise complexity introduced by the HBW selection of the blocks is $O(Q)$, which accounts for the complexity for finding the maximum element of an array of length $Q$. The complexity for the calculation of the vectors (2.4) remains $O(N_b k_q^2)$, for each block in the partition. Obviously, by improving sparsity via the HBW implementation of OMP the calculations for these vectors are reduced. Contrarily, storage becomes more demanding. Indeed, the HBW-OMP approach introduces the need for saving the vectors (2.4) and (2.5), for each block. This implies a requirement for extra storage of $2K$ vectors of size $N_b$. We postpone until Sec. 4 the discussion about how to address this issue when dealing with very large signals. For signals of moderate size storage is not a problem. As the following numerical example illustrates, the improvement in sparsity may result very significant.

**Numerical Example I**

The signal to be approximated is a piece of piano melody shown in Fig 1. It consists of $N = 960512$ samples (20 secs) divided into $Q = 938$ blocks with $N_b = 1024$ samples each. The dictionary $\mathcal{D}$ is the union of a Redundant Discrete Cosine (RDC) Dictionary, $\mathcal{D}_c$, and a Redundant Discrete Sine Dictionary (RDS), $\mathcal{D}_s$, defined below:

- $\mathcal{D}_c = \{ w_c(n) \cos \frac{\pi(2i-1)(n-1)}{2M}, i = 1, \ldots, N \}_{n=1}^M$.
- $\mathcal{D}_s = \{ w_s(n) \sin \frac{\pi(2i-1)(n)}{2M}, i = 1, \ldots, N \}_{n=1}^M$.

![Fig. 1. Piano melody. Credit: Julius O. Smith, Center for Computer Research in Music and Acoustics (CCRMA), Stanford University.](image)
where $w_c(n)$ and $w_s(n)$, $n = 1, \ldots, M$ are normalization factors. For $M = N$ each of the above dictionaries is an orthonormal basis, the Orthogonal Discrete Cosine (ODC) and Orthogonal Discrete Sine (ODS) basis, henceforth to be denoted $B_c$ and $B_s$, respectively. The joint dictionary is an orthonormal basis for $M = N$, the Orthogonal Discrete Cosine-Sine (ODCS) basis, to be denoted $B_{cs}$.

As a measure of approximation quality we use the standard Signal to Noise Ratio (SNR),

$$\text{SNR} = 10 \log_{10} \frac{\|f\|_2^2}{\|f - f_k\|_2^2} = 10 \log_{10} \frac{\sum_{i=1}^{N} |f_q(i)|^2}{\sum_{i=1}^{N} |f_q(i) - f^*_q(i)|^2}.$$ 

This numerical example aims at illustrating the following outcomes in relation to the signal in hand.

1) The approximation power of all the orthogonal basis and dictionaries defined above remarkably improve if, instead of applying OMP independently to approximate each block $f_q$ up to the same SNR, the HBW-OMP approach is applied with an equivalent constraint with regard to the sparsity of the whole signal.

2) For approximating the signal with the OMP greedy strategy, the redundant dictionaries $D_c$ and $D_{cs}$ perform significantly better than any of the orthogonal basis $B_c$, $B_s$ and $B_{cs}$.

In order to demonstrate 1) and 2) each of the blocks $f_q$ is approximated independently with OMP, up to SNR=25 dB. Redundant dictionaries, with redundancy 2 and 4, are simply creating by setting $M = 2N$ and $M = 4N$ in the definitions of $D_c$ and $D_s$, with $N_b = 1024$. They will be denoted as $D_{c2}$, $D_{s2}$ and $D_{c4}$, $D_{s4}$, respectively. Notice that for these dictionaries the inner products in (2.2) can be evaluated via the Fast Fourier Transform (FFT) with zero padding, which reduces the complexity of the calculation to $O(M \log_2 M)$. FFT is also used for calculating the inner products for the dictionaries $D_{cs2} = D_c \cup D_s$ and $D_{cs4} = D_{c2} \cup D_{s2}$. Approximations up to the same quality are performed with each of the orthonormal basis $B_c$, $B_s$ and $B_{cs}$. The results are presented in Table 1.

As can be seen, the SR produced by the redundant dictionaries $D_{c2}$, $D_{c4}$, $D_{cs2}$, and $D_{cs4}$ is substantially larger than that corresponding to any of the orthogonal basis. Moreover, in all the cases the HBW-OMP strategy improves, notoriously, upon the OMP approach applied independently to produce a uniform SNR in the approximation of each block. It is also noticed that the highest sparsity is yielded by dictionaries $D_{cs2}$ and $D_{cs4}$. The next section discusses the HBW backward strategy for reducing coefficients from an approximation.

2.1. Hierarchized Blockwise Backwards Optimized OMP. We extend here the Backward Optimized Orthogonal Matching Pursuit (BOOMP) strategy [19] to select also the blocks
Table 1

| Dict. | OMP  | HBW-OMP |
|-------|------|---------|
|       | SR   | SNR     | SR   | SNR |
| $\mathcal{B}_c$ | 14.38 | 25.0    | 14.38 | 35.19 |
| $\mathcal{D}_{c2}$ | 17.75 | 25.0    | 17.75 | 34.91 |
| $\mathcal{D}_{c4}$ | 19.39 | 25.0    | 19.39 | 34.74 |
| $\mathcal{B}_s$ | 7.65  | 25.0    | 7.65  | 28.31 |
| $\mathcal{D}_{s2}$ | 12.13 | 25.0    | 12.13 | 29.39 |
| $\mathcal{D}_{s4}$ | 13.20 | 25.0    | 13.12 | 29.34 |
| $\mathcal{B}_{cs}$ | 10.77 | 25.0    | 10.77 | 30.09 |
| $\mathcal{D}_{cs2}$ | 20.45 | 25.0    | 20.45 | 35.81 |
| $\mathcal{D}_{cs4}$ | 23.82 | 25.0    | 23.82 | 35.55 |

Comparison of the approximation quality (SNR values) and sparsity (SR values) produced with trigonometric basis $\mathcal{B}_c$, $\mathcal{B}_s$, $\mathcal{B}_{cs}$ and redundant trigonometric dictionaries, $\mathcal{D}_{c2}, \mathcal{D}_{c4}, \mathcal{D}_{s2}, \mathcal{D}_{s4}, \mathcal{D}_{cs2}$, and $\mathcal{D}_{cs4}$. The second column shows the SR resulting when applying the block independent OMP approach to achieve a SNR=25.0dB. The forth column demonstrates the significant gain in SNR rendered by the HBW-OMP strategy for the same sparsity.

from which the atoms are to be removed for downgrading an approximation. The backward strategy is stepwise optimal because it minimizes, at each step, the norm of the error resulting by the elimination of one atom. Before establishing this result let us recall the recursive equations for modifying vectors $b_{k,q}^n$, $n = 1, \ldots, k_q$ to account for the elimination of one atom, say the $j$-th one, from the set $\{d_{k,q}^n\}_{n=1}^{k_q}$. For each $q$, the reduced set of vectors $\{d_{k,q}^n\}_{n=1}^{k_q}$ spanning the reduced subspace $\mathcal{V}_{k_q,j} = \text{span}\{d_{k,q}^n\}_{n=1}^{k_q, n \neq j}$ can be quickly obtained through the adaptive backward equations [18, 19]

$$
(2.9) \quad b_{n/k_q,j,q} = b_{n/k_q,j} - b_{j/k_q,j} \frac{\langle b_{n/k_q,j}, b_{j/k_q,j} \rangle}{\|b_j\|^2}.
$$

Consequently, the coefficients of the atomic decomposition corresponding to the block $q$, from which the atom $d_{k_q,j}^q$ is taken away, need to be modified as

$$
(2.10) \quad c_{k_q,j,q}(n) = c_{k_q,j,q}(n) - c_{k_q,j,q}(j) \frac{\langle b_{j/k_q,j}, d_{k_q,j,q} \rangle}{\|b_j\|^2}, \quad n = 1, \ldots, j - 1, j + 1, \ldots, k_q.
$$

Now the stepwise criterion for shrinking coefficients from the approximation of a signal partitioned into blocks readily follows.
Proposition 2.2. Assume that the approximation of a signal is given as \( f^k = \bigcup_{q=1}^{q^*} f^{k_q,q} \), where \( f^{k_q,q} = \hat{P}_{V_{k_q}} f_q \). The index \( j^* \) and \( q^* \) corresponding to the atoms \( d_{q^*} \) to be removed from the decomposition of \( f^k \), in order to leave an approximation \( f^{k/j} \) such that the error norm \( \|f^k - f^{k/j}\| \) takes its minimum value, satisfies the condition:

\[
(2.11) \quad j^*, q^* = \arg \min_{j=1,\ldots,k_q} \frac{|c_{k_q,q}(j)|}{\|b_{j/q}\|^2}.
\]

Proof. Consider that the atom to be removed corresponds to a particular block, say the \( h \)-th one. Thus, only the atomic decomposition \( f^{k_h,h} \) changes by the atom removal. The approximation of \( f^{k_h,h} \) in \( V_{k_h/j} \) which is optimal, in a minimum distance sense, can be written as \( \hat{P}_{V_{k_h/j}} f^{k_h,h} = \hat{P}_{V_{k_h/j}} \hat{P}_{V_{k_h}} f_h = \hat{P}_{V_{k_h/j}} f_h \). Hence,

\[
(2.12) \quad f^{k_h,h} - f^{k_h,j,h} = \sum_{n=1}^{k_h} d_{q,n} (b_n^{k_h,h}, f^h) - \sum_{n=1}^{k_h} d_{q,n} (b_n^{k_h,j,h}, f^h).
\]

Using (2.9) and the fact that \( d_{q,j} ^h \in V_{k_h} \), we have

\[
\begin{align*}
(2.13) \quad f^{k_h,h} - f^{k_h,j,h} &= \sum_{n=1}^{k_h} d_{q,n} (b_n^{k_h,h} - b_n^{k_h,j,h}, f^h) + \sum_{n=1}^{k_h} d_{q,n} (b_n^{k_h,j,h}, f^h) d_{q,n} (b_n^{k_h,h}, f^h) \\
&= \sum_{n=1}^{k_h} d_{q,n} (b_n^{k_h,h}, f^h) - \sum_{n=1}^{k_h} d_{q,n} (b_n^{k_h,j,h}, f^h) d_{q,n} (b_n^{k_h,h}, f^h) - d_{q,n} (b_n^{k_h,j,h}, f^h)
\end{align*}
\]

from where it follows that \( \|f^{k_h,h} - f^{k_h,j,h}\| \) is minimized by removing the atom corresponding to the indices \( j^* \) and \( q^* \) such that:

\[
(2.14) \quad j^*, q^* = \arg \min_{j=1,\ldots,k_h} \frac{|\langle b_j^{k_h,h} \rangle |}{\|b_j^{k_h,h}\|^2} = \arg \min_{j=1,\ldots,k_h} \frac{|c_{j,k_h}^{k_h,h}(j)|}{\|b_j^{k_h,h}\|^2}.
\]

Since \( f^k - f^{k/j} = f^{k_h,h} - f^{k_h,j,h} \) the proof is concluded. \( \square \)
Numerical Example II. In order to illustrate the backward approach, we change the information in the last two columns of Table 1 to have the sparsity of all the approximations corresponding to SNR=25dB. For this we downgrade the HBW-OMP approximation to degrade the previous quality. As seen in Table 2, the result is that the previous sparsity increases considerably. For the $\mathcal{D}_{cs4}$ dictionary the SR improves 113%, in comparison with OMP for the same dictionary. Notice also that the best SR result, for dictionary $\mathcal{D}_{cs4}$, is 96% better that the best SR result for an orthogonal basis.

3. Refinement by Swaps. The suitability of the HBW strategy for dealing with the class of signals we are considering is clear. An alternative also worth exploring is to maintain the block independent approximation but allowing possible eventual HBW refinements.

Considering that the approximation (2.1) of each block in a signal partition is known, the goal is to improve the signal approximation, while keeping the total number of atoms $K = \sum_{q=1}^{Q} k_q$ fixed, without much increment in the computational burden. The proposed refinement consists of interchanging atoms between blocks as follows:

i) Use criterion (2.14) to remove one atom from the block $q^\ast$. Let us call this block a ‘donor’ block and indicate it with the index $q_d$.

ii) Use criterion (2.7) to incorporate one atom in the approximation of the block $q^\ast$. Let us call this block a ‘receiver’ and indicate it with the index $q_r$.

iii) Denote by $\delta_d$ the square norm of the error introduced at step i) by downgrading the approximation of the donor block $q_d$. Denote by $\delta_r$ the square norm of the gain in improving the approximation of the receiver $q_r$. Proceed according to the following rule:

| Dict. | OMP | HBW-BOOMP |
|-------|-----|------------|
|       | SR  | SNR        | SR  | SNR        |
| $\mathcal{B}_c$ | 14.38 | 25.0 | 25.56 | 25.0 |
| $\mathcal{D}_{c2}$ | 17.60 | 25.0 | 34.05 | 25.0 |
| $\mathcal{D}_{c4}$ | 19.25 | 25.0 | 37.01 | 25.0 |
| $\mathcal{B}_r$ | 7.65 | 25.0 | 13.67 | 25.0 |
| $\mathcal{D}_{s2}$ | 12.03 | 25.0 | 25.74 | 25.0 |
| $\mathcal{D}_{s4}$ | 13.11 | 25.0 | 28.18 | 25.0 |
| $\mathcal{B}_{cs}$ | 10.77 | 25.0 | 19.94 | 25.0 |
| $\mathcal{D}_{cs2}$ | 20.21 | 25.0 | 42.43 | 25.0 |
| $\mathcal{D}_{cs4}$ | 23.53 | 25.0 | 50.20 | 25.0 |

Comparison of sparsity (SR values) for the same SNR. The HBW-BOOMP results are obtained by degrading the HBW-OMP approximation in Table 1 to SNR=25dB.
If $\delta_r > \delta_d$ accept the change and repeat steps i) and ii). Otherwise stop. Notice that for successive implementation of the steps i) and ii) above, we need: to downgrade the vectors $b_{n,q}^{k_q}$, $n = 1, \ldots k_q$ at step i) and to upgrade these vectors at step ii). The downgrading is realized as in (2.9), this procedure does not change. However, when an upgrading is to be realized after a downgrading, the calculation of vectors $w_{k_q/j+1}^{q}$ entails to compute:

\begin{equation}
   w_{k_q/j+1}^{q} = d_{k_q/j+1}^{q} - \hat{P}_{k_q/j} d_{k_q/j+1}^{q},
\end{equation}

where $\hat{P}_{k_q/j}$ is the orthogonal projection onto the subspace $V_{k_q/j}$ spanned by the atoms $\{d_n^{k_q}\}_{n=1, n\neq j}$. We can take different routes for computing this projector. One possibility is to make use of the downgraded vectors (2.9) to calculate $\hat{P}_{k_q/j} d_{k_q/j+1}^{q}$ in (3.1) as below

\begin{equation}
   w_{k_q/j+1}^{q} = d_{k_q/j+1}^{q} - \hat{P}_{k_q/j} d_{k_q/j+1}^{q} = d_{k_q/j+1}^{q} - \sum_{n=1}^{k_q/j} d_{n}^{q} (b_{n,q}^{k_q/j,q}, d_{k_q/j+1}^{q}).
\end{equation}

Since $\hat{P}_{k_q/j}$ is hermitian, according to [21] the re-orthogonalization step for $w_{k_q/j+1}^{q}$ is recommended to be realized by computing $\hat{P}_{k_q/j+1} w_{k_q/j+1}^{q}$ as in

\begin{equation}
   w_{k_q/j+1}^{q} = w_{k_q/j+1}^{q} - \hat{P}_{k_q/j} w_{k_q/j+1}^{q} = w_{k_q/j+1}^{q} - \sum_{n=1}^{k_q/j} b_{n,q}^{k_q/j,q} (d_{n}^{q}, w_{k_q/j+1}^{q}).
\end{equation}

An alternative to numerically implement equation (3.1) could be, for instance, to use the plane rotation approach for downgrading orthogonal projection matrices [22, 23]. Notice that the HBW-SR-OMP refinement implemented as proposed above possesses the desirable feature of acting only if the approximation can be improved by the proposed swaps.

**Numerical Example III**

We illustrate the HBW-SR-OMP refinement by applying it to the signal in Fig. 2, which is a flute exercise consisting of $N = 96256$ samples divided into $Q = 94$ blocks with $N_b = 1024$ samples each. Table 3 shows the improvement in quality obtained by applying the HBW-SR-OMP approach on outputs of the OMP approximation. Notice that in all the cases the SNR results are practically equivalent to those obtained by the HBW-OMP approach.

**4. Processing of large signals.** Following up the discussion at the end of Sec. 2, we outline now a processing scheme which makes it possible the application of HBW techniques on very large signals. As already mentioned, the requirement of having to store matrices (2.4) and (2.5), for each block, restricts the amount of blocks to be processed, with a standard laptop for instance.
Fig. 2. Flute exercise: $N = 96256$ samples.

| Dict. | SR  | OMP | Swaps | HBW-SR-OMP | HBW-OMP |
|-------|-----|-----|-------|------------|---------|
| $B_c$ | 10.14 | 25.00 | 1427 | 26.79 | 26.79 |
| $D_{c2}$ | 14.27 | 25.00 | 894 | 26.73 | 26.67 |
| $D_{c4}$ | 15.52 | 25.00 | 842 | 26.69 | 26.61 |
| $B_s$ | 7.26 | 25.00 | 2620 | 26.73 | 26.73 |
| $D_{s2}$ | 12.10 | 25.00 | 1396 | 27.31 | 27.27 |
| $D_{s4}$ | 13.17 | 25.00 | 1364 | 27.33 | 27.25 |
| $B_{cs}$ | 8.08 | 25.00 | 1822 | 26.65 | 26.65 |
| $D_{cs2}$ | 29.00 | 25.00 | 565 | 27.38 | 27.37 |
| $D_{cs4}$ | 35.16 | 25.00 | 457 | 27.36 | 27.35 |

Table 3. Comparison of quality (SRN values) for a fixed sparsity: that corresponding to $SRN=25$ with the OMP approach. The forth column shows the number of swaps and the fifth columns the SRN achieved by those swaps through the HBW-SR-OMP refinement to the outputs yielded in second column. For further comparison the last column shows the results corresponding to the HBW-OMP approach.

Hence, the whole signal needs to be sub-partitioned into segments of convenient size, say $N_s$, such that $f = \cup_{s=1}^{S} f_s$, with $SN_s = N$. In order to enable the HBW approximation every segment $f_s$ is partitioned into smaller blocks of size $N_b << N_s$. In other words, the processing of large signals is realized by chopping the signal into segments and applying HBW-OMP on the partition of each segment. Nevertheless, the question as to how to set the sparsity constraint in this situation needs further consideration. One could, of course, require the same sparsity in every segment, unless information advising otherwise were available. While uniform sparsity guarantees the same
sparsity on the whole signal, in particular cases, where the nature of the signal changes over its range of definition, it would not render the best approximation. In order to avoid encountering this type of situation we introduce a preliminary step: the randomization of all the small blocks in the signal partition. The implementation is carried out as follows:

i) Given a signal \( f \) split it into \( Q \) blocks of size \( N_b \). Apply an invertible random permutation \( \Pi \) to scramble the block’s location (to bottom graph in Figure 3 provides a visual illustration of how the signal in the top graph looks after this step). Let’s denote the re-arranged signal by \( \tilde{f} = \cup_{q=1}^{Q} \tilde{f}_q \).

ii) Group the blocks \( \tilde{f}_q, q = 1, \ldots, Q \) into, say \( S \) segments, \( \tilde{g}_s \), of size \( N_s = \frac{N_b Q}{S} \), so that \( \tilde{f} = \cup_{s=1}^{S} \tilde{g}_s \), where \( \tilde{g}_s = \cup_{q=1}^{N_s} \tilde{f}_q \).

iii) Approximate each segment \( \tilde{g}_s \) independently by the relevant HBW strategy to obtain \( \tilde{g}_s^K = \cup_{q=1}^{N_s} \tilde{f}_q^k \) \( s = 1, \ldots, N_s \) and \( \tilde{f}^K = \cup_{s=1}^{S} \tilde{g}_s^K \).

iv) Reverse the permutation of the block location to obtain the approximated signal \( \hat{f}^K \).

---

**Fig. 3.** The top graph shows a piece (1433600 samples) of Piazzola music interpreted by Gidon Kremer and orchestra. The bottom graph shows the resulting signal after the randomization of the blocks.

**Numerical Example IV**

The signal to be approximated is a 32.5 secs (1433600 samples) piece of Piazzola music, shown in the top graph of Figure 3. It is partitioned into \( Q = 1400 \) blocks of 1024 points each. After randomization (bottom graph of the same figure) the blocks are grouped into 50 segments, each of which is independently approximated by the HBW-OMP approach to produce a SR=11.39. The resulting SNR of the approximated signal is 29.1 dB. Equivalent approximation quality is obtained with other realizations of the random process. For appreciation of this result the original signal was also approximated by HBW-OMP, but without segmentation. In that case that quality corresponding to SR=11.39 is only 1.8% higher i.e., SNR=29.6 dB. The conclusion is that, even
if the largest the segments the less the distortion for the same sparsity, applying HBW-OMP on segments of a large signal is still significantly more beneficial than the independent approximation of the blocks. The latter yields a SNR of only 25 dB.

REFERENCES

[1] B. K. Natarajan, “Sparse Approximate Solutions to Linear Systems”, SIAM Journal on Computing, 24, 227–234 (1995).
[2] S. S. Chen, D. L. Donoho, M. A Saunders, “Atomic Decomposition by Basis Pursuit”, SIAM Journal on Scientific Computing, 20, 33–61 (1998).
[3] J. H. Friedman and W. Stuetzle, “Projection Pursuit Regression”, Journal of the American Statistical Association, 76, 817–823 (1981).
[4] S. G. Mallat and Z. Zhang, “Matching Pursuits with Time-Frequency Dictionaries”, IEEE Trans. Signal Process., 41, 3397–3415 (1993).
[5] Y.C. Pati, R. Rezaifar, and P.S. Krishnaprasad, “Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition,” Proc. of the 27th ACSSC, 1, 40–44 (1993).
[6] Y. Eldar, P. Kuppinger and H. Bölcskei, “Block-Sparse Signals: Uncertainty Relations and Efficient Recovery”, IEEE Trans. Signal Process., 58, 3042–3054 (2010).
[7] D. L. Donoho , Y. Tsaig, I. Drori, J. Starck, “Stagewise Orthogonal Matching Pursuit”, IEEE Transactions on Information Theory, 58, 1094–1121 (2006).
[8] D. Needell and J.A. Tropp, “CoSaMP: Iterative signal recovery from incomplete and inaccurate samples”, Applied and Computational Harmonic Analysis, 26, 301–321 (2009).
[9] D. Needell and R. Vershynin, “Signal Recovery From Incomplete and Inaccurate Measurements via Regularized Orthogonal Matching Pursuit”, IEEE Journal of Selected Topics in Signal Processing, 4, 310–316 (2010).
[10] D. L. Donoho, “Compressed sensing”, IEEE Trans. Inf. Theory, 52, 1289–1306 (2006).
[11] J. Cands, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” IEEE Trans. Inf. Theory, 52, 489 –509 (2006).
[12] R. Baraniuk, “Compressive sensing”, IEEE Signal Processing Magazine, 24, 118–121, (2007).
[13] E. Candès and M. Wakin, “An introduction to compressive sampling”, IEEE Signal Processing Magazine, 25, 21 – 30 (2008).
[14] R. Baraniuk, “More Is less: Signal processing and the data deluge”, Science, 331, 717 – 719 (2011).
[15] L. Rebollo-Neira, R. Matioli, S. Bibi, “Hierarchized block wise image approximation by greedy pursuit strategies,” IEEE Signal Process. Letters, 20, 1175–1178 (2013).
[16] J.A. Tropp, “Greed is good: algorithmic results for sparse approximation”, IEEE Trans. Inf. Theory, 50, 2231–2242 (2004).
[17] L. Rebollo-Neira and D. Lowe, “Optimized orthogonal matching pursuit approach”, IEEE Signal Process. Letters, 9, 137–140 (2002).
[18] L. Rebollo-Neira, “Backward adaptive biorthogonalization”, Int. J. Math. Math. Sci. 1843–1853 (2004).
[19] M. Andrle, L. Rebollo-Neira, E. Sagianos, “Backward-optimized orthogonal matching pursuit approach”, IEEE Signal Process. Lett., 11, 705–708 (2004).
[20] M. Andrle and L. Rebollo-Neira, “A swapping-based refinement of orthogonal matching pursuit strategies”, Signal Processing, 86, 480–495 (2006).
[21] M. Andrle and L. Rebollo-Neira, “Experiments on orthogonalization by biorthogonal representations of orthogonal projectors”, Journal of Computational and Applied Mathematics, 205, 545–551 (2007).
[22] Å. Björck, “Numerical Methods for Least Squares Problems”, SIAM, Philadelphia, 1996.
[23] G.W. Stewart, “On the stability of sequential updates and downdates”, IEEE Trans. Signal Process., 43, 2642 – 2648 (1995).