Author's Accepted Manuscript

Cooperative Greedy Pursuit Strategies for Sparse Signal Representation by Partitioning

Laura Rebollo-Neira

PII: S0165-1684(16)00058-X
DOI: http://dx.doi.org/10.1016/j.sigpro.2016.02.008
Reference: SIGPRO6058

To appear in: *Signal Processing*

Received date: 3 May 2015
Revised date: 31 January 2016
Accepted date: 5 February 2016

Cite this article as: Laura Rebollo-Neira, Cooperative Greedy Pursuit Strategie for Sparse Signal Representation by Partitioning, *Signal Processing* http://dx.doi.org/10.1016/j.sigpro.2016.02.008

Cooperative Greedy Pursuit Strategies for Sparse Signal Representation by Partitioning

Laura Rebollo-Neira
Mathematics Department
Aston University
B3 7ET, Birmingham, UK

February 10, 2016

Abstract

Cooperative Greedy Pursuit Strategies are considered for approximating a signal partition subjected to a global constraint on sparsity. The approach aims at producing a high quality sparse approximation of the whole signal, using highly coherent redundant dictionaries. The co-operation 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. The advantage of the strategy is illustrated by approximation of music signals using redundant trigonometric dictionaries. In addition to rendering stunning improvements in sparsity with respect to the concomitant trigonometric basis, these dictionaries enable a fast implementation of the approach via the Fast Fourier Transform.

Keywords: Cooperative Greedy Pursuit Strategies; Sparse Representation of Music Signals by Partitioning; Trigonometric Dictionaries.

1 Introduction

Compressible signals, such as audio and vision data, are characterized by samples containing a good deal of redundancy. 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. Most popular transformations for signal processing (e.g. Discrete Cosine Transform and Discrete Wavelet Transform) do not modify the signal size. A sparse representation is achieved, a posteriori, by disregarding the least relevant

points in the transformed domain. Nonetheless, 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 stepwise 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 Pursuit (MP) [4] and Orthogonal Matching Pursuit (OMP) [5] methods. The approach, which in principle consider the stepwise selection of single atoms, has 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]. In that context the reconstruction problem is of a very particular nature, though: 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. 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 and computational complexity 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. The broadly used image compression standard JPEG and the music compression standard MP3, for instance, operate by partitioning as a first step in the compression process.

The central aim of this paper is to tailor pursuit algorithms for approximation by partitioning without significant increment in computational complexity, in comparison with standard applications. The examples presented here clearly show that a global constraint on sparsity, rather than quality constraints on the individual partition's units, may benefit enormously the quality of the signal approximation. This is true even if the actual approximation of each unit is performed individually. Because considerations are restricted to producing high quality approximations, there is no risk that blocking artifacts could appear due to the signal division into non-overlapping pieces.

Notational convention

Throughout the paper $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{N}$ indicate the sets of real, complex, and natural numbers, respectively. Boldface letters are used to indicate Euclidean vectors, whilst standard mathematical fonts indicate components, e.g., $\mathbf{f} \in \mathbb{R}^N$, $N \in \mathbb{N}$ is a vector of components $f(i)$, $i = 1, \ldots, N$. A partition of a signal $\mathbf{f} \in \mathbb{R}^N$ is represented as a set of disjoint pieces $\mathbf{f}\{q\} \in \mathbb{R}^{N_\mathrm{b}}$, $q = 1, \ldots, Q$, which for simplicity are assumed to be all of the same size and such that $QN_\mathrm{b} = N$. The signal is reconstructed from the partition through the operation $\mathbf{f} = \hat{\mathrm{J}}_{q=1}^{Q} \mathbf{f}\{q\}$ where $\hat{\mathrm{J}}$ represents the operator which concatenates the partition. The operation is defined as follows: given $\mathbf{f}\{1\} \in \mathbb{R}^{N_\mathrm{b}}$ and $\mathbf{f}\{2\} \in \mathbb{R}^{N_\mathrm{b}}$, the vector $\mathbf{f} = \mathbf{f}\{1\}\, \hat{\mathrm{J}}\, \mathbf{f}\{2\}$ is a vector in $\mathbb{R}^{2N_\mathrm{b}}$ having components $f(i) = f\{1\}(i)$ for $i = 1, \ldots, N_\mathrm{b}$,

and $f(i) = f\{2\}(i - N_\mathrm{b})$ for $i = N_\mathrm{b} + 1, \ldots, 2N_\mathrm{b}$. Thus $\mathbf{f} = \hat{\mathrm{J}}_{q=1}^{Q} \mathbf{f}\{q\}$ is a vector in $\mathbb{R}^{QN_\mathrm{b}}$ having components $f(i) = f\{q\}(i - (q-1)N_\mathrm{b})$, $i = (q-1)N_\mathrm{b} + 1, \ldots, qN_\mathrm{b}$, $q = 1, \ldots, Q$. Consequently $\langle \mathbf{f}, \mathbf{f} \rangle = \|\mathbf{f}\|^2 = \sum_{q=1}^{Q} \|\mathbf{f}\{q\}\|^2$, were $\langle \cdot, \cdot \rangle$ indicates the Euclidean inner product and $\| \cdot \|$ the induced 2-norm.

Paper contributions

Given a signal partition $\mathbf{f}\{q\} \in \mathbb{R}^{N_\mathrm{b}}$, $q = 1, \ldots, Q$ and a dictionary $\mathcal{D} = \left\{ \mathbf{d}_n \in \mathbb{R}^{N_\mathrm{b}} \, ; \|\mathbf{d}_n\| = 1 \right\}_{n=1}^{M}$ to approximate the elements $\mathbf{f}\{q\}$ in the partition, the following outcome has been recently reported [15]: A very significant gain in the sparsity of a signal approximation may be effectively obtained by a greedy pursuit strategy, if the approximation of each piece $\mathbf{f}\{q\}$ (called 'q-block') is accomplished in a hierarchized manner. Suppose that the $k(q)$-term approximation of $\mathbf{f}\{q\}$ is the atomic decomposition:

$$\mathbf{f}^{\mathrm{a}}\{q\} = \sum_{n=1}^{k(q)} c\{q\}_n \mathbf{d}_{\ell\{q\}_n}, \quad q = 1, \ldots, Q,$$

with the atoms $\mathbf{d}_{\ell\{q\}_n}$, $n = 1, \ldots, k(q)$ selected from the dictionary $\mathcal{D}$, via a stepwise greedy pursuit strategy. Suppose also that the number of atoms to approximate the whole signal $\mathbf{f}$ 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 $\mathbf{f}\{q\}$ to be approximated in each approximation step. Some remarkable results of this strategy, which has been termed Hierarchized Block Wise (HBW) greedy strategy, are illustrated in [15] by approximating images using the greedy algorithms MP and OMP. When these methods are applied in the proposed HBW fashion are called HBW-MP and HBW-OMP, respectively.

While [15] focusses on highlighting the suitability of the HBW-OMP/MP method when the image approximation is carried out in the wavelet domain, this Communication extends the method as well as the range of applicability. The extended techniques are shown to produce hugely sparse high quality approximation of music signals. This is accomplished with trigonometric dictionaries, which are endowed with the additional advantage of enhancing the competitiveness of the approach in terms of computational complexity.

The extension of the idea outlined in [15] comprises:

- A revised version of the approach, which optimizes the ranking of the pieces $\mathbf{f}\{q\}$, $q = 1, \ldots, Q$ for their step wise approximation without significant extra computational cost. Additionally, the selection of atoms is also optimized by including the Optimized Orthogonal Matching Pursuit (OOMP) criterion [16].

- A HBW backward approach for downgrading the approximation when required. This is realized in a stepwise optimized manner, by removing terms in the approximations of the selected blocks.

- An alternative to the HBW strategy which consists of two stages. The first stage involves the approximation of the pieces $\mathbf{f}\{q\}$, $q = 1, \ldots, Q$, up to a tolerance error. The second stage refines the previous approximation by making possible the downgrading of the atomic decomposition of some blocks, called 'donors', and the upgrading of the atomic decompositions of another blocks, called 'receivers'. Since this process is inspired in the Swapping-based-Refinement of greedy strategies introduced in [17], we refer to it as HBW-Swapping-based-Refinement (HBW-SbR).

Further contributions of the paper are

- The illustration of the huge gain in the sparsity of the representation of music signals obtainable by the proposed approach when using trigonometric dictionaries.

- The finding that a mixed trigonometric dictionary having both, discrete cosine and sine components, renders the best sparsity performance for melodic music signals, in comparison with the single discrete cosine (or sine) dictionaries with the same redundancy. This result contrasts with the fact that, the broadly used single discrete cosine basis is the basis yielding best sparsity results for music signals, in comparison to other trigonometric basis of the same nature.

  Note: The discrete cosine/sine components of the mixed dictionary are not just the real and imaginary parts of the complex exponentials in a Fourier series. The involved phase factors make the whole difference in achieving high sparsity.

- The provision of implementation details in the form of algorithms for the proposed strategies and the provision of dedicated algorithms to operate with trigonometric dictionaries via the Fast Fourier Transform (FFT).

A library of MATLAB functions for implementing the proposed methods, and running the numerical examples in this paper, are available for downloading on [18].

2 Hierarchized Blockwise OMP, revised

As already stated, a signal $\mathbf{f} \in \mathbb{R}^N$ will be considered to be the composition of Q identical and disjoint blocks $\mathbf{f} = \hat{\mathrm{J}}_{q=1}^{Q} \mathbf{f}\{q\},$, where $\mathbf{f}\{q\} \in \mathbb{R}^{N_\mathrm{b}}$ assuming that $N_\mathrm{b} Q = N$. Each piece $\mathbf{f}\{q\} \in \mathbb{R}^{N_\mathrm{b}}$ is approximated using a dictionary $\mathcal{D} = \left\{\mathbf{d}_n \in \mathbb{R}^{N_\mathrm{b}} \,;\, \|\mathbf{d}_n\| = 1\right\}_{n=1}^{M}$, with $M > N_\mathrm{b}$, by an atomic decomposition of the form:

$$\mathbf{f}^{\mathrm{a}}\{q\} = \sum_{n=1}^{k(q)} c\{q\}_n \mathbf{d}_{\ell\{q\}_n}, \quad q = 1 \ldots, Q. \tag{1}$$

where the atoms $\mathbf{d}_{\ell\{q\}_n}$, $n = 1, \ldots, k(q)$ are selected from the dictionary to approximate the block q. Thus, each set of selected indices $\Gamma\{q\} = \{\ell\{q\}_n\}_{n=1}^{k(q)}$ is a subset of the set of labels $\{1, \ldots, M\}$ which identify the atoms in $\mathcal{D}$. In general $\Gamma\{q\} \neq \Gamma\{p\}$ and $k(q) \neq k(p)$ if $p \neq q$.

The HBW-OMP approach outlined in [15] selects atoms in (1) as indicated by the OMP approach, i.e.: On setting initially $k(q) = 0$ and $\mathbf{r}^0\{q\} = \mathbf{f}\{q\}$, the algorithm picks the atoms in the atomic decomposition of block q by selecting one by one the indices which satisfy:

$$\ell\{q\}_{k(q)+1} = \underset{n=1,\ldots,M}{\arg\max} \left|\left\langle \mathbf{d}_n, \mathbf{r}^{k(q)}\{q\} \right\rangle\right|, \text{ with } \mathbf{r}^{k(q)}\{q\} = \mathbf{f}\{q\} - \sum_{n=1}^{k(q)} c\{q\}_n \mathbf{d}_{\ell\{q\}_n} \text{ if } k(q) > 0. \tag{2}$$

The coefficients $c\{q\}_n$, $n = 1, \ldots, k(q)$ in (2) are such that the norm $\left\|\mathbf{r}^{k(q)}\{q\}\right\|$ is minimized. This is ensured by requesting that $\mathbf{r}^{k(q)}\{q\} = \mathbf{f}\{q\} - \hat{\mathrm{P}}_{\mathbb{V}_{k(q)}^q} \mathbf{f}\{q\}$, where $\hat{\mathrm{P}}_{\mathbb{V}_{k(q)}^q}$ is the orthogonal projection operator onto $\mathbb{V}_{k(q)}^q = \operatorname{span}\left\{\mathbf{d}_{\ell\{q\}_n}\right\}_{n=1}^{k(q)}$. This condition also guarantees that the set $\left\{\mathbf{d}_{\ell\{q\}_n}\right\}_{n=1}^{k(q)}$ is linearly independent.

An implementation as in [16] provides us with two representations of $\hat{\mathrm{P}}_{\mathbb{V}^q_{k(q)}}$. One of the representations is achieved by orthonormalization of the set $\left\{\mathbf{d}_{\ell\{q\}_n}\right\}_{n=1}^{k(q)}$. The other by biorthogonalization of the same set. We implement the orthogonalization by Gram Schmidt method, including a re-orthogonalization step, as follows: The orthonormal set $\{\widetilde{\mathbf{w}}\{q\}_n\}_{n=1}^{k(q)}$ is inductively constructed from $\widetilde{\mathbf{w}}\{q\}_1 = \mathbf{w}\{q\}_1 = \mathbf{d}_{\ell\{q\}_1}$ through the process:

$$\widetilde{\mathbf{w}}\{q\}_n = \frac{\mathbf{w}\{q\}_n}{\|\mathbf{w}\{q\}_n\|}, \quad \text{with} \quad \mathbf{w}\{q\}_n = \mathbf{d}_{\ell\{q\}_n} - \sum_{i=1}^{n-1} \widetilde{\mathbf{w}}\{q\}_i \left\langle \widetilde{\mathbf{w}}\{q\}_i, \mathbf{d}_{\ell\{q\}_n} \right\rangle. \tag{3}$$

For numerical accuracy at least one re-orthogonalization step is usually needed. This implies to recalculate the above vectors as

$$\mathbf{w}\{q\}_n \leftarrow \mathbf{w}\{q\}_n - \sum_{i=1}^{n-1} \widetilde{\mathbf{w}}\{q\}_i \left\langle \widetilde{\mathbf{w}}\{q\}_i, \mathbf{w}\{q\}_n \right\rangle, \tag{4}$$

with the corresponding normalization giving rise to the orthonormal set $\{\widetilde{\mathbf{w}}\{q\}_n\}_{n=1}^{k(q)}$. Notice that, whilst this set can be used to calculate the orthogonal projection of $\mathbf{f}\{q\}$ onto $\mathbb{V}^q_{k(q)}$ as

$$\hat{\mathrm{P}}_{\mathbb{V}^q_{k(q)}} \mathbf{f}\{q\} = \sum_{n=1}^{k(q)} \widetilde{\mathbf{w}}\{q\}_n \left\langle \widetilde{\mathbf{w}}\{q\}_n, \mathbf{f}\{q\} \right\rangle, \tag{5}$$

this superposition is not the atomic decomposition in terms of the selected atoms. In order to produce such a decomposition we use the other representation for an orthogonal projector given in [16]. Namely,

$$\hat{\mathrm{P}}_{\mathbb{V}^q_{k(q)}} \mathbf{f}\{q\} = \sum_{n=1}^{k(q)} \mathbf{d}_{\ell\{q\}_n} \left\langle \mathbf{b}\{q\}_n, \mathbf{f}\{q\} \right\rangle. \tag{6}$$

For a fixed q the vectors $\mathbf{b}\{q\}_n$, $n = 1, \ldots, k(q)$ in (6) are biorthogonal to the selected atoms, i.e. $\left\langle \mathbf{d}_{\ell\{q\}_n}, \mathbf{b}\{q\}_m \right\rangle = 0$ if $n \neq m$ and 1 if $n = m$, and span the identical subspace i.e., $\mathbb{V}^q_{k(q)} = \mathrm{span}\{\mathbf{b}\{q\}_n\}_{n=1}^{k(q)} = \mathrm{span}\{\mathbf{d}_{\ell\{q\}_n}\}_{n=1}^{k(q)}$. In order to fulfill the last condition *all* the vectors need to be updated when a new atom is introduced in the spanning set. Starting from $\mathbf{b}\{q\}_1 = \mathbf{d}_{\ell\{q\}_1}$ both the calculation and upgrading of the biorthogonal set is attained recursively as follows

$$\begin{aligned} \mathbf{b}\{q\}_{k(q)} &= \frac{\mathbf{w}\{q\}_{k(q)}}{\|\mathbf{w}\{q\}_{k(q)}\|^2}, \quad \text{with} \quad \mathbf{w}\{q\}_{k(q)} \text{ as in (3)} \\ \mathbf{b}\{q\}_n &\leftarrow \mathbf{b}\{q\}_n - \mathbf{b}\{q\}_{k(q)} \left\langle \mathbf{d}_{\ell\{q\}_{k(q)}}, \mathbf{b}\{q\}_n \right\rangle, \quad n = 1, \ldots, (k(q)-1). \end{aligned} \tag{7}$$

These vectors produce the required atomic decomposition for the block q. The coefficients in (1) are the inner products in (6), i.e. $c\{q\}_n = \langle \mathbf{b}\{q\}_n, \mathbf{f}\{q\} \rangle$, $n = 1, \ldots, k(q)$.

The HBW variant of a greedy strategy establishes an order for upgrading the atomic decomposition of the blocks in the partition. Instead of completing the approximation of each block at once, the atomic decomposition to be upgraded at each iteration corresponds to a block, say $q^\star$, selected according to a greedy criterion. In other words: the HBW version of a pursuit method for approximating a signal by partitioning involves two selection instances: a) the selection of dictionary's atoms for approximating each of the blocks in the partition and b)the selection, at each iteration step, of the block where the upgrading of the approximation is to be realized. In [15] the blocks are selected through the process

$$q^\star = \underset{q=1,\ldots,Q}{\arg\max} \left| \left\langle \mathbf{d}_{\ell\{q\}_{k(q)+1}}, \mathbf{r}^{k(q)}\{q\} \right\rangle \right|. \tag{8}$$

We notice, however, that condition (8) for the block's selection can be optimized, in a stepwise sense, without significant computational cost. The next proposition revise the condition (8) considered in [15].

Proposition 1. *Let $\ell\{q\}_{k(q)+1}$ be the index arising, for each value of q, from the maximization process* (2)*. In order to minimize the square norm of the total residual $\|\mathbf{r}^{k+1}\|^2$ at iteration $k+1$ the atomic decomposition to be upgraded should correspond to the block $q^\star$ such that*

$$q^\star = \underset{q=1,\ldots,Q}{\arg\max}\, \chi(q), \quad \textit{where } \chi(q) = \left| \left\langle \widetilde{\mathbf{w}}\{q\}_{k(q)+1}, \mathbf{f}\{q\} \right\rangle \right| = \frac{\left| \left\langle \mathbf{d}_{\ell\{q\}_{k(q)+1}}, \mathbf{r}^{k(q)}\{q\} \right\rangle \right|}{\left\| \mathbf{w}\{q\}_{k(q)+1} \right\|}, \tag{9}$$

with $\mathbf{w}\{q\}_{k(q)+1}$ and $\widetilde{\mathbf{w}}\{q\}_{k(q)+1}$ as given in (3)*.*

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

$$\mathbf{r}^{k+1} = \hat{\mathbb{J}}_{\substack{p=1 \\ p\neq q}}^{Q} \mathbf{r}\{p\}^{k(p)}\, \hat{\mathbb{J}}\, \mathbf{r}\{q\}^{k(q)+1}.$$

Then,

$$\|\mathbf{r}^{k+1}\|^2 = \sum_{\substack{p=1 \\ p\neq q}}^{Q} \|\mathbf{r}\{p\}^{k(p)}\|^2 + \|\mathbf{r}\{q\}^{k(q)+1}\|^2.$$

Moreover, since $\mathbf{r}^{k(q)+1}\{q\} = \mathbf{f}\{q\} - \hat{\mathrm{P}}_{\mathbb{V}^q_{k(q)+1}}\mathbf{f}\{q\}$ using the orthonormal vectors (3) to calculate $\hat{\mathrm{P}}_{\mathbb{V}^q_{k(q)+1}}\mathbf{f}\{q\}$ we have:

$$\|\mathbf{r}^{k(q)+1}\{q\}\|^2 = \|\mathbf{r}^{k(q)}\{q\}\|^2 - \left|\left\langle \widetilde{\mathbf{w}}\{q\}_{k(q)+1}, \mathbf{f}\{q\}\right\rangle\right|^2,$$

which is minimum for the $q^\star$-value corresponding to the maximum value of $\left|\left\langle \widetilde{\mathbf{w}}\{q\}_{k(q)+1}, \mathbf{f}\{q\}\right\rangle\right|$. Since $\mathbf{w}\{q\}_{k(q)+1} = \mathbf{d}_{\ell\{q\}_{k(q)+1}} - \hat{\mathrm{P}}_{\mathbb{V}^q_{k(q)}}\mathbf{d}_{\ell\{q\}_{k(q)+1}}$ and $\hat{\mathrm{P}}_{\mathbb{V}^q_{k(q)}}$ is hermitian (9) follows. □

Notice that, since the vectors $\widetilde{\mathbf{w}}\{q\}_{k(q)+1}$ will be used for subsequent approximations, the optimized ranking of blocks (9) does not require significant extra computational effort. Only $Q-1$ of these vectors will not have been otherwise used when the algorithm stops. Apart from that, the complexity of selecting the blocks remains being that of finding the maximum element of an array of length Q, i.e. $\mathrm{O}(Q)$.

Before presenting the algorithm details we would like to consider also the optimization of the selection of atoms (stage a) of a HBW pursuit strategy). Within our implementation the optimized selection of atoms is readily achievable by the Optimized Orthogonal Matching Pursuit (OOMP) approach [16], which selects the index $\ell\{q\}_{k(q)+1}$ through the maximization process:

$$\ell\{q\}_{k(q)+1} = \underset{n=1,\ldots,M}{\arg\max} \frac{\left|\left\langle \mathbf{d}_n, \mathbf{r}^{k(q)}\{q\}\right\rangle\right|}{(1-s\{q\}_n)^{\frac{1}{2}}}, \quad \text{with} \quad s\{q\}_n = \sum_{i=1}^{k(q)} |\left\langle \mathbf{d}_n, \widetilde{\mathbf{w}}\{q\}_i\right\rangle|^2. \tag{10}$$

The OOMP selection criterion (10) minimizes, in a stepwise sense, the norm of the local error when approximating a single block [16]. The extra computational cost, in comparison to the OMP selection step (c.f. (2)) is the calculation of the sum $s\{q\}_n$ in the denominator of (10), for every atom in the dictionary, $n = 1, \ldots, M$. As becomes clear in Algorithm 3, by saving the sum of previous iterations, at iteration $k(q)+1$ only the inner products $\left\langle \mathbf{d}_n, \widetilde{\mathbf{w}}\{q\}_{k(q)}\right\rangle$, $n = 1, \ldots, M$ need to be calculated.

2.1 Algorithmic Implementation

The implementation details of the HBW-OOMP method is given in Algorithm 1, which is realized through Algorithms 2 and 3. For a fixed number K the algorithm iterates until the condition $\sum_{q=1}^{Q} k(q) = K$ is met.

2.2 Trigonometric dictionaries and sparse representation of music signals

The viability of a sparse representation for a given signal depends in large part on the dictionary choice. In the case of audio signals dictionaries consisting of Gabor atoms [4], combinations of Gabor atoms [19, 20], and Gaussian chirp atoms [21] have been previously considered. In general those dictionaries are very redundant. The greedy algorithm associated with the atom's selection is the MP one, or dedicated versions of MP which exploit further the atom's localization. Less redundancy is involved in dictionaries arising by union of modified discrete cosine transform bases at different scales [22], which have also been explored using the MP algorithm. The particularity of the dictionaries we consider here is that of being appropriate for approximating by partitioning, which motivates the HBW implementation of greedy strategies. A possibility to secure a suitable dictionary for this purpose would involve learning techniques [23–26]. Another possibility, specially for image approximation, is to build the dictionary using atoms of different nature [27]. However, as will be illustrated by numerical examples, when considerations are restricted to music signals the combination of a Redundant Discrete Cosine Dictionary (RDCD) $\mathcal{D}^{\mathrm{c}}$ and a Redundant Discrete Sine Dictionary (RDSD) $\mathcal{D}^{\mathrm{s}}$, defined below, may produce stunning sparsity levels for relatively low redundancy, if the selection is carried out by the proposed techniques.

- $\mathcal{D}^{\mathrm{c}} = \{\frac{1}{w^{c}(n)} \cos \frac{\pi(2i-1)(n-1)}{2M}, i = 1, \ldots, N_{\mathrm{b}}\}_{n=1}^{M}$.
- $\mathcal{D}^{\mathrm{s}} = \{\frac{1}{w^{\mathrm{s}}(n)} \sin \frac{\pi(2i-1)(n)}{2M}, i = 1, \ldots, N_{\mathrm{b}}\}_{n=1}^{M}$,

Algorithm 1 HBW-OOMP Procedure

Input: Signal partition $\mathbf{f}\{q\} \in \mathbb{R}^{N_\mathrm{b}}$, $q = 1, \ldots, Q$. Dictionary $\mathcal{D} = \{\mathbf{d}_n \in \mathbb{R}^{N_\mathrm{b}}; \|\mathbf{d}_n\| = 1\}_{n=1}^{M}$. Number K of total atoms to approximate the whole signal.

Output: Sets $\Gamma\{q\}$, $q = 1, \ldots, Q$ containing the indices of the selected atom for each block. Orthonormal and biorthogonal sets $\{\widetilde{\mathbf{w}}\{q\}_n\}_{n=1}^{k(q)}$ and $\{\mathbf{b}\{q\}_n\}_{n=1}^{k(q)}$, for each block (c.f. (4) and (7) respectively). Coefficients of the atomic decompositions for each block $\{c\{q\}_n\}_{n=1}^{k(q)}$, $q = 1, \ldots, Q$. Approximation of the partition, $\mathbf{f}^\mathrm{a}\{q\}$, $q = 1, \ldots, Q$. Approximation of the whole signal $\mathbf{f}^\mathrm{a}$.

{Initialization}

$j = 0$

for $q = 1 : Q$ **do**

$k(q) = 1$; $\mathbf{r}\{q\} = \mathbf{f}\{q\}$; $s\{q\}_n = 0$, $n = 1, \ldots, M$

{Select the potential first atom for each block, as below}

$\ell\{q\} = \arg\max_{n=1,\ldots,M} |\langle \mathbf{d}_n, \mathbf{f}\{q\}\rangle|$

$\chi(q) = \left|\left\langle \mathbf{d}_{\ell\{q\}}, \mathbf{f}\{q\}\right\rangle\right|$ {Store the maximum for each block}

Set $\mathbf{w}\{q\}_1 = \mathbf{d}_{\ell\{q\}}$; $\widetilde{\mathbf{w}}\{q\}_1 = \mathbf{w}\{q\}_1$; $\mathbf{b}\{q\}_1 = \widetilde{\mathbf{w}}\{q\}_1$; $\Gamma\{q\} = \emptyset$

end for

while $j < K$ **do**

{Select the block to be approximated (c.f.(9)) and store the index of the atom in the approximation of that block, as below}

$q^\star = \arg\max_{q=1,\ldots,Q} \chi(q)$

$\Gamma\{q^\star\} \leftarrow \Gamma\{q^\star\} \cup \ell\{q^\star\}$

{Update the residue corresponding to block $q^\star$, as below}

$\mathbf{r}\{q^\star\} \leftarrow \mathbf{r}\{q^\star\} - \widetilde{\mathbf{w}}\{q^\star\}_{k(q^\star)} \left\langle \widetilde{\mathbf{w}}\{q^\star\}_{k(q^\star)}, \mathbf{f}\{q^\star\}\right\rangle$

if $k(q^\star) > 1$ **then**

{Upgrade the biorthogonal set $\{\mathbf{b}\{q^\star\}_n\}_{n=1}^{k(q^\star)}$ to include the atom $\mathbf{d}_{\ell\{q^\star\}}$ (c.f. (7))}

end if

{Apply Algorithm 2 to select the index $\ell\{q^\star\}$ for a new potential atom for the block $q^\star$}

Increase $k(q^\star) \leftarrow k(q^\star) + 1$

{Compute $\mathbf{w}\{q^\star\}_{k(q^\star)}$ and $\widetilde{\mathbf{w}}\{q^\star\}_{k(q^\star)}$ from $\mathbf{d}_{\ell\{q^\star\}}$, (c.f. (3) and (4)}

{Update the objective functional χ for the block $q^\star$}

$\chi(q^\star) = \left|\left\langle \widetilde{\mathbf{w}}\{q^\star\}_{k(q^\star)}, \mathbf{f}\{q^\star\}\right\rangle\right|$

Increase $j \leftarrow j + 1$

end while

{Calculation of coefficients and approximation}

for $q = 1 : Q$ **do**

$\mathbf{f}^\mathrm{a}\{q\} = \mathbf{f}\{q\} - \mathbf{r}\{q\}$; $c\{q\}_n = \langle \mathbf{b}\{q\}_n, \mathbf{f}\{q\}\rangle$, $n = 1, \ldots, k(q)$

end for

{Concatenation of the partition to produce the whole signal approximation}

$\mathbf{f}^\mathrm{a} = \hat{\mathrm{J}}_{q=1}^{Q} \mathbf{f}^\mathrm{a}\{q\}$

Algorithm 2 Atom Selection Procedure

Input: Residue $\mathbf{r}\{q^\star\}$. Dictionary $\mathcal{D} = \{\mathbf{d}_n \in \mathbb{R}^{N_\mathrm{b}}\,; \|\mathbf{d}_n\| = 1\}_{n=1}^{M}$. Auxiliary sequence $\{s\{q^\star\}_n\}_{n=1}^{M}$ and vector $\widetilde{\mathbf{w}}\{q^\star\}_{k(q^\star)}$ for the upgrading. Set of already selected atoms $\Gamma\{q^\star\}$.

Output: Index $\ell\{q^\star\}$ corresponding to the selected atom $\mathbf{d}_{\ell\{q^\star\}}$.

{Apply the Auxiliary Procedure, Algorithm 3, to upgrade the sequence $\{s\{q^\star\}_n\}_{n=1}^{M}$ with respect to vector $\widetilde{\mathbf{w}}\{q^\star\}_{k(q^\star)}$}

{Select the index, as below}

$$\ell\{q^\star\} = \underset{\substack{n=1,\ldots,M \\ n\notin\Gamma\{q^\star\}}}{\arg\max} \frac{|\langle \mathbf{d}_n, \mathbf{r}\{q^\star\}\rangle|}{(1 - s\{q^\star\}_n)^{\frac{1}{2}}}$$

Algorithm 3 Auxiliary Procedure

Input: Dictionary $\mathcal{D} = \{\mathbf{d}_n \in \mathbb{R}^{N_\mathrm{b}}\,; \|\mathbf{d}_n\| = 1\}_{n=1}^{M}$. Auxiliary sequence $\{s\{q^\star\}_n\}_{n=1}^{M}$ and vector $\widetilde{\mathbf{w}}\{q^\star\}_{k(q^\star)}$ for the upgrading.

Output: Upgraded sequence $\{s\{q^\star\}_n\}_{n=1}^{M}$, with respect to $\widetilde{\mathbf{w}}\{q^\star\}_{k(q^\star)}$.

for $n = 1 : M$ **do**

$$s\{q^\star\}_n = s\{q^\star\}_n + \left|\left\langle \mathbf{d}_n, \widetilde{\mathbf{w}}\{q^\star\}_{k(q^\star)}\right\rangle\right|^2$$

end for

where $w^{\mathrm{c}}(n)$ and $w^{\mathrm{s}}(n)$, $n = 1, \ldots, M$ are normalization factors as given by

$$w^{\mathrm{c}}(n) = \begin{cases} \sqrt{N_\mathrm{b}} & \text{if} \quad n = 1, \\ \sqrt{\frac{N_\mathrm{b}}{2} + \frac{\sin(\frac{\pi(n-1)}{M})\sin(\frac{2\pi(n-1)N_\mathrm{b}}{M})}{2(1-\cos(\frac{2\pi(n-1)}{M}))}} & \text{if} \quad n \neq 1. \end{cases}$$

$$w^{\mathrm{s}}(n) = \begin{cases} \sqrt{N_\mathrm{b}} & \text{if} \quad n = 1, \\ \sqrt{\frac{N_\mathrm{b}}{2} - \frac{\sin(\frac{\pi n}{M})\sin(\frac{2\pi n N_\mathrm{b}}{M})}{2(1-\cos(\frac{2\pi n}{M}))}} & \text{if} \quad n \neq 1. \end{cases}$$

For $M = N_\mathrm{b}$ each of the above dictionaries is an orthonormal basis, the Orthogonal Discrete Cosine Basis (ODCB) and the Orthogonal Discrete Sine Basis (ODSB), henceforth to be denoted $\mathcal{B}^{\mathrm{c}}$ and $\mathcal{B}^{\mathrm{s}}$ respectively. The joint dictionary is an orthonormal basis for $M = \frac{N_\mathrm{b}}{2}$, the Orthogonal Discrete Cosine-Sine Basis (ODCSB) to be indicated as $\mathcal{B}^{\mathrm{cs}}$. For $M > N_\mathrm{b}$, $\mathcal{D}^{\mathrm{c}}$ is a RDCD and $\mathcal{D}^{\mathrm{s}}$ a RDSD. If $M > \frac{N_\mathrm{b}}{2}$ $\mathcal{D}^{\mathrm{cs}} = \mathcal{D}^{\mathrm{c}} \cup \mathcal{D}^{\mathrm{s}}$ becomes a Redundant Discrete Cosine-Sine Dictionary (RDCSD).

For the sake of discussing a fast calculation of inner products with trigonometric atoms, given a vector $\mathbf{y} \in \mathbb{R}^M$, let's define

$$\mathcal{F}(\mathbf{y}, n, M) = \sum_{j=1}^{M} y(j) e^{-\imath\, 2\pi \frac{(n-1)(j-1)}{M}}, \quad n = 1, \ldots, M. \tag{11}$$

When $M = N_{\rm b}$ (11) is the Discrete Fourier Transform of vector $\mathbf{y} \in \mathbb{R}^{N_{\rm b}}$, which can be evaluated using FFT. If $M > N_{\rm b}$ (11) can still be calculated via FFT, with $\mathcal{F}(\mathbf{y}, n, M)$ indicating that $\mathbf{y}$ is padded with $M - N_{\rm b}$ zeros. Thus, (11) can be used to calculate inner products with the atoms in dictionaries $\mathcal{D}^{\rm c}$ and $\mathcal{D}^{\rm s}$. Indeed,

$$\sum_{j=1}^{N_{\rm b}} \cos \frac{\pi(2j-1)(n-1)}{2M} y(j) = \mathrm{Re}\left(e^{-\imath \frac{\pi(n-1)}{2M}} \mathcal{F}(\mathbf{y}, n, 2M)\right),\ n = 1, \ldots, M. \tag{12}$$

and

$$\sum_{j=1}^{N_{\rm b}} \sin \frac{\pi(2j-1)(n-1)}{2M} y(j) = -\,\mathrm{Im}\left(e^{-\imath \frac{\pi(n-1)}{2M}} \mathcal{F}(\mathbf{y}, n, 2M)\right),\ n = 2, \ldots, M+1, \tag{13}$$

where $\mathrm{Re}(z)$ indicates the real part of z and $\mathrm{Im}(z)$ its imaginary part.

The computation of inner products with trigonometric dictionaries via FFT is outlined in Algorithm 4. For dictionaries $\mathcal{D}^{\rm c}$ and $\mathcal{D}^{\rm s}$ that procedure reduces the complexity for calculating the inner products from $\mathrm{O}(N_{\rm b} M)$, per block, to $\mathrm{O}(2M \log_2 2M)$, per block. For dictionary $\mathcal{D}^{\rm cs}$ the reduction is larger: $\mathrm{O}(M \log_2 M)$, because both the real and imaginary parts of the FFT are used. When Algorithm 1 is applied with trigonometric dictionaries the pieces implemented by Algorithms 2 and 3 should be replaced by Algorithms 6 and 7, given in Appendix A. In addition to speeding the calculations via FFT, Algorithms 6 and 7 avoid storing the dictionary. In the case of the dictionary $\mathcal{D}^{\rm cs}$ it is assumed that both the Cosine and Sine components are of the same size and the first $\frac{M}{2}$ elements of the dictionary are Cosine atoms.

Numerical Example I

The signal approximated in this example is a piece of piano melody shown in Fig 1. It consists of $N = 960512$ samples (20 secs) divided into $Q = 938$ blocks with $N_{\rm b} = 1024$ samples each. As sparsity measure for a signal representation we consider the Sparsity Ratio (SR) defined as $\mathrm{SR} = \frac{N}{K}$, where K is the total number of coefficients in the signal representation. As a measure of approximation quality we use the standard Signal to Noise Ratio (SNR),

$$\mathrm{SNR} = 10 \log_{10} \frac{\|\mathbf{f}\|^2}{\|\mathbf{f} - \mathbf{f}^{\rm a}\|^2} = 10 \log_{10} \frac{\sum_{\substack{i=1 \\ q=1}}^{N_{\rm b}, Q} |f\{q\}(i)|^2}{\sum_{\substack{i=1 \\ q=1}}^{N_{\rm b}, Q} |f\{q\}(i) - f^{\rm a}\{q\}(i)|^2}.$$

Algorithm 4 Computation of inner products with a trigonometric dictionary via FFT. IPTrgFFT procedure: $\mathbf{p}$ =IPTrgFFT($\mathbf{r}, 2M$, Case)

Input: $\mathbf{r} \in \mathbb{R}^{N_\mathrm{b}}$, M, number of elements in the dictionary, and Case ('Cos' or 'Sin').

Output: Vector $\mathbf{p} \in \mathbb{R}^{M}$, with the inner products between $\mathbf{r}$ and 'Cos' or 'Sin' dictionaries.

{Computation of auxiliary vector $\mathbf{t} \in \mathbb{C}^{2M}$ to compute $\mathbf{p}$.}

$\mathbf{t} = \mathrm{FFT}(\mathbf{r}, 2M)$ {FFT with $(2M - N_\mathrm{b})$ zero padding}.

Case 'Cos'

$p(n) = \frac{1}{w^c(n)} \mathrm{Re}(e^{\imath \frac{\pi(n-1)}{2M}} t(n))$, $n = 1, \ldots, M$ {(c.f. (12))}

Case 'Sin'

$p(n-1) = -\frac{1}{w^s(n)} \mathrm{Im}(e^{\imath \frac{\pi(n-1)}{2M}} t(n))$, $n = 2, \ldots, M+1$ {(c.f.(13))}

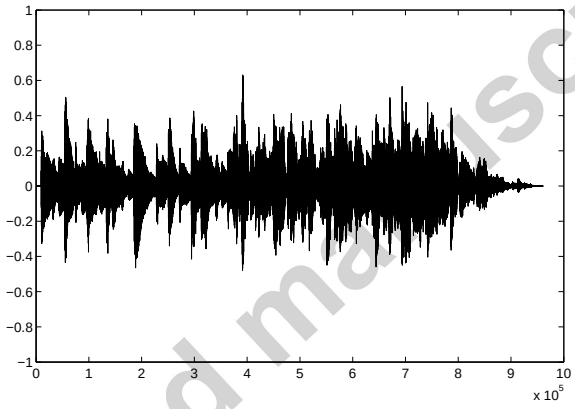

Figure 1: Piano melody. Credit: Julius O. Smith, Center for Computer Research in Music and Acoustics (CCRMA), Stanford University.

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/OOMP independently to approximate each block $\mathbf{f}\{q\}$ up to the same SNR, the HBW-OMP/OOMP approach is applied to match the sparsity of the whole signal.

2) For approximating the signal with the OMP/OOMP greedy strategies, the redundant dictionaries $\mathcal{D}^{\mathrm{c}}$ and $\mathcal{D}^{\mathrm{cs}}$ perform significantly better than any of the orthogonal basis $\mathcal{B}^{\mathrm{c}}, \mathcal{B}^{\mathrm{s}}$ and $\mathcal{B}^{\mathrm{cs}}$.

Dict.	OMP		HBW-OMP		OOMP		HBW-OOMP	
	SR	SNR	SR	SNR	SR	SNR	SR	SNR
$\mathcal{B}^{\rm c}$	14.38	25.0	14.38	35.19	14.38	25.0	14.38	35.19
$\mathcal{D}^{\rm c2}$	17.75	25.0	17.75	34.91	18.47	25.0	18.47	35.08
$\mathcal{D}^{\rm c4}$	19.39	25.0	19.39	34.64	19.80	25.0	19.80	34.95
$\mathcal{B}^{\rm s}$	7.65	25.0	7.65	28.31	7.65	25.0	7.65	28.31
$\mathcal{D}^{\rm s2}$	12.13	25.0	12.13	29.39	12.63	25.0	12.63	29.39
$\mathcal{D}^{\rm s4}$	13.17	25.0	13.17	29.34	13.42	25.0	13.42	29.33
$\mathcal{B}^{\rm cs}$	10.77	25.0	10.77	30.09	10.77	25.0	10.77	30.09
$\mathcal{D}^{\rm cs2}$	20.45	25.0	20.45	35.81	22.08	25.0	22.08	35.56
$\mathcal{D}^{\rm cs4}$	23.83	25.0	23.83	35.56	26.18	25.0	26.18	36.37

Table 1: Comparison of the approximation quality (SNR values) and sparsity (SR values) produced with trigonometric basis $\mathcal{B}^{\rm c}, \mathcal{B}^{\rm s}, \mathcal{B}^{\rm cs}$ and redundant trigonometric dictionaries, $\mathcal{D}^{\rm c2}, \mathcal{D}^{\rm c4}, \mathcal{D}^{\rm s2}, \mathcal{D}^{\rm s4}, \mathcal{D}^{\rm cs2}$, and $\mathcal{D}^{\rm cs4}$. The second column shows the SR resulting when applying the block independent OMP approach to achieve a SNR=25.0dB. The fifth column demonstrates the significant gain in SNR rendered by the HBW-OMP strategy for the same sparsity. Further improvements of the same nature are shown in the last four columns corresponding to the approaches OOMP and HBW-OOMP.

In oder to demonstrate 1) and 2) each of the blocks $\mathbf{f}\{q\}$ is approximated independently with OMP/OOMP, up to SNR=25 dB. Redundant dictionaries, with redundancy 2 and 4, are simply creating by setting $M = 2N_{\rm b}$ and $M = 4N_{\rm b}$ in the definitions of $\mathcal{D}^{\rm c}$ and $\mathcal{D}^{\rm s}$, with $N_{\rm b} = 1024$. They will be denoted as $\mathcal{D}^{\rm c2}$, $\mathcal{D}^{\rm s2}$ and $\mathcal{D}^{\rm c4}$, $\mathcal{D}^{\rm s4}$, respectively. Approximations of the same quality are performed with each of the orthonormal basis $\mathcal{B}^{\rm c}, \mathcal{B}^{\rm s}$ and $\mathcal{B}^{\rm cs}$, and the mixed dictionaries $\mathcal{D}^{\rm cs2} = \mathcal{B}^{\rm c} \cup \mathcal{B}^{\rm s}$ and $\mathcal{D}^{\rm cs4} = \mathcal{D}^{\rm c2} \cup \mathcal{D}^{\rm s2}$. The results are presented in Table 1. As can be seen, the SR produced by the redundant dictionaries $\mathcal{D}^{\rm c2}, \mathcal{D}^{\rm c4}, \mathcal{D}^{\rm cs2}$, and $\mathcal{D}^{\rm cs4}$ is substantially larger than that corresponding to any of the orthogonal basis. Moreover, in all the cases the HBW-OMP/OOMP strategies improve notoriously (up to 11 dB) upon the OMP/OOMP approaches applied independently to produce a uniform SNR in the approximation of each block.

It is noticed that the highest sparsity is attained by the dictionary $\mathcal{D}^{\rm cs4}$ which, as already discussed, involves the most effective implementation via FFT. It is also seen that HBW-OOMP over performs HBW-OMP in terms of sparsity (the running times are similar). This will appear even more clearly in Table 2, where the approximation of both methods are downgraded to produce a SNR of 25dB. The next section discusses the HBW backward strategy which allows for removing atoms from the

signal approximation, in a stepwise optimized fashion.

2.3 Hierarchized Blockwise Backwards OOMP

We extend here the Backward Optimized Orthogonal Matching Pursuit (BOOMP) strategy [28] to select also the blocks from which the atoms are to be removed for downgrading the approximation of a whole signal. The BOOMP strategy is stepwise optimal because it minimizes, at each step, the norm of the error resulting by downgrading, by one atom, an atomic decomposition. This result readily follows from the recursive equations for modifying vectors $\{\mathbf{b}\{q\}_n\}_{n=1}^{k(q)}$ to account for the elimination of, say the j-th atom, from the set $\{\mathbf{d}_{\ell\{q\}_n}\}_{n=1}^{k(q)}$. For each q, the reduced set of vectors $\{\mathbf{b}\{q\}_n\}_{\substack{n=1\\n\neq j}}^{k(q)}$ spanning the reduced subspace $\mathbb{V}_{k(q)/j}^q = \mathrm{span}\{\mathbf{d}_{\ell\{q\}_n}\}_{\substack{n=1\\n\neq j}}^{k(q)}$ can be quickly obtained through the adaptive backward equations [28, 29]

$$\mathbf{b}\{q\}_n \leftarrow \mathbf{b}\{q\}_n - \mathbf{b}\{q\}_j \frac{\langle \mathbf{b}\{q\}_n, \mathbf{b}\{q\}_j \rangle}{\|\mathbf{b}\{q\}_j\|^2}, \quad n = 1, \ldots, j-1, j+1, \ldots, k(q). \tag{14}$$

Consequently, the coefficients of the atomic decomposition corresponding to the block q, from which the atom $\mathbf{d}_{\ell\{q\}_j}$ is taken away, are modified as

$$c\{q\}_n \leftarrow c\{q\}_n - c\{q\}_j \frac{\langle \mathbf{b}\{q\}_j, \mathbf{b}\{q\}_n \rangle}{\|\mathbf{b}\{q\}_j\|^2}, \quad n = 1, \ldots, j-1, j+1, \ldots, k(q). \tag{15}$$

For a fixed value of q, the BOOMP criterion removes the coefficient $c\{q\}_{j^\diamond}$ such that [28]

$$j^\diamond\{q\} = \underset{j=1,\ldots,k(q)}{\arg\min} \frac{|c\{q\}_j|}{\|\mathbf{b}\{q\}_j\|}. \tag{16}$$

The equivalent criterion applies to the selection of the $q^\diamond$-block for the downgrading of the whole approximation. The proof of the next proposition parallels the proof of (16) given in [28].

Proposition 2. *Assume that the approximation of a signal is given as* $\mathbf{f}^{\mathrm{a}} = \hat{\mathrm{J}}_{q=1}^{Q} \mathbf{f}^{\mathrm{a}}\{q\}$, *where* $\mathbf{f}^{\mathrm{a}}\{q\} = \hat{\mathrm{P}}_{\mathbb{V}_{k(q)}^q} \mathbf{f}\{q\}$. *Let* $j^\diamond\{q\}$, $q = 1, \ldots, Q$ *be the indices which satisfies* (16) *for each* q. *The block* $q^\diamond$, *from where the atom* $\mathbf{d}_{\ell_{\{q^\diamond\}_{j^\diamond\{q^\diamond\}}}}$ *is to be removed, in order to leave an approximation* $\mathbf{f}_\downarrow^{\mathrm{a}}$ *such that* $\|\mathbf{f}^{\mathrm{a}} - \mathbf{f}_\downarrow^{\mathrm{a}}\|$ *takes its minimum value, satisfies the condition:*

$$q^\diamond = \underset{q=1,\ldots,Q}{\arg\min} \frac{|c\{q\}_{j^\diamond\{q\}}|}{\|\mathbf{b}\{q\}_{j^\diamond\{q\}}\|}. \tag{17}$$

Proof. Since at each step only one atom is removed from a particular block, say the q-th one, it holds that $\|\mathbf{f}^{\mathrm{a}} - \mathbf{f}^{\mathrm{a}}_{\downarrow}\| = \|\mathbf{f}^{\mathrm{a}}\{q\} - \mathbf{f}^{\mathrm{a}}_{\downarrow}\{q\}\|$. The identical steps as in [28] lead to the expression

$$\mathbf{f}^{\mathrm{a}}\{q\} - \mathbf{f}^{\mathrm{a}}_{\downarrow}\{q\} = \mathbf{b}\{q\}_{j^{\diamond}\{q\}} \frac{\left\langle \mathbf{b}\{q\}_{j^{\diamond}\{q\}}, \mathbf{f}\{q\} \right\rangle}{\|\mathbf{b}\{q\}_{j^{\diamond}\{q\}}\|^2}. \tag{18}$$

Thus $\|\mathbf{f}^{\mathrm{a}}\{q\} - \mathbf{f}^{\mathrm{a}}_{\downarrow}\{q\}\|^2 = \dfrac{\left|\left\langle \mathbf{b}\{q\}_{j^{\diamond}\{q\}}, \mathbf{f}\{q\} \right\rangle\right|^2}{\|\mathbf{b}\{q\}_{j^{\diamond}\{q\}}\|^2}$. Since $\left\langle \mathbf{b}\{q\}_{j^{\diamond}\{q\}}, \mathbf{f}\{q\} \right\rangle = c\{q\}_{j^{\diamond}\{q\}}$ it follows that $\|\mathbf{f}^{\mathrm{a}} - \mathbf{f}^{\mathrm{a}}_{\downarrow}\|$ is minimized by removing the atom $j^{\diamond}\{q^{\diamond}\}$ corresponding to the block $q^{\diamond}$ satisfying (17). □

The HBW-BOOMP algorithm for downgrading a given atomic decomposition of a signal is presented in Algorithm 5.

Numerical Example II

In order to illustrate the backward approach, we change the information in the last four columns of Table 1 to have the sparsity of all the approximations corresponding to SNR=25dB. For this we downgrade the HBW-OMP/OOMP approximations to degrade the previous quality. As seen in Table

Dict.	OMP		HBW-BOOMP		OOMP		HBW-BOOMP	
	SR	SNR	SR	SNR	SR	SNR	SR	SNR
$\mathcal{B}^{\mathrm{c}}$	14.38	25.0	25.56	25.0	14.38	25.0	25.56	25.0
$\mathcal{D}^{\mathrm{c2}}$	17.60	25.0	34.05	25.0	18.47	25.0	36.34	25.0
$\mathcal{D}^{\mathrm{c4}}$	19.25	25.0	37.01	25.0	19.80	25.0	39.06	25.0
$\mathcal{B}^{\mathrm{s}}$	7.65	25.0	13.67	25.0	7.65	25.0	13.67	25.0
$\mathcal{D}^{\mathrm{s2}}$	12.03	25.0	25.74	25.0	12.63	25.0	27.56	25.0
$\mathcal{D}^{\mathrm{s4}}$	13.11	25.0	28.18	25.0	13.42	25.0	29.59	25.0
$\mathcal{B}^{\mathrm{cs}}$	10.77	25.0	19.94	25.0	10.77	25.0	19.94	25.0
$\mathcal{D}^{\mathrm{cs2}}$	20.21	25.0	42.43	25.0	22.08	25.0	48.48	25.0
$\mathcal{D}^{\mathrm{cs4}}$	23.53	25.0	50.19	25.0	26.18	25.0	59.68	25.0

Table 2: Comparison of sparsity (SR values) for the same SNR. The HBW-BOOMP results are obtained by degrading the HBW-OMP/OOMP approximation in Table 1 to SNR=25dB.

2, the result is that the sparsity increases drastically. The SR improves up to 128% (for the $\mathcal{D}^{\mathrm{cs4}}$ dictionary) in comparison with the standard application of the same approaches without ranking the blocks. Notice also that the best SR result (for dictionary $\mathcal{D}^{\mathrm{cs4}}$), is 133% higher than the best SR result for an orthogonal basis (the $\mathcal{B}^{\mathrm{c}}$ basis).

Algorithm 5 HBW-BOOMP

Input: Biorthogonal sets $\{\mathbf{b}\{q\}_n\}_{n=1}^{k(q)}$, $q = 1, \ldots, Q$. Coefficients in the approximation of each block $\{c\{q\}_n\}_{n=1}^{k(q)}$, $q = 1, \ldots, Q$. Sets of indices in the decomposition of each block $\Gamma\{q\}$, $q = 1, \ldots, Q$. Number of total atoms for the downgraded approximation, K_d.

Output: Downgraded biorthogonal sets. Coefficients of the downgraded atomic decompositions for each block. Indices of the remaining atoms.

{Initialization}

$K = 0$;

for $q = 1 : Q$ **do**

$k(q) = |\Gamma\{q\}|$ {Number of elements in $\Gamma\{q\}$}

$K = K + k(q)$

{Select the index of the potential first atom to be removed in each block, as below}

$$j^{\diamond}\{q\} = \underset{n=1,\ldots,k(q)}{\arg\min} \frac{|c\{q\}_n|}{\|\mathbf{b}\{q\}_n\|}$$

$$\chi_{\downarrow}(q) = \frac{|c\{q\}_{j^{\diamond}\{q\}}|}{\|\mathbf{b}\{q\}_{j^{\diamond}\{q\}}\|}$$ {Store the minimum}

end for

$i = K$

while $i > K_{\mathrm{d}}$ **do**

{Select the block to be downgraded, as below}

$$q^{\diamond} = \underset{q=1,\ldots,Q}{\arg\min}\, \chi_{\downarrow}(q)$$

{Apply backward biorthogonalization (c.f. (14)) to downgrade the biorthogonal set corresponding to block $q^{\diamond}$, with respect to $\mathbf{b}\{q^{\diamond}\}_{j^{\diamond}\{q^{\diamond}\}}$ and shift indices}

{Update the coefficients corresponding to block $q^{\diamond}$ (c.f. 15) and shift indices}

Downgrade the set $\Gamma\{q^{\diamond}\} \leftarrow \Gamma\{q^{\diamond}\}/\ell\{q^{\diamond}\}_{j^{\diamond}\{q^{\diamond}\}}$ and shift indices

Decrease $k(q) \leftarrow k(q) - 1$

{Select the new potential atom to be removed from block $q^{\diamond}$, as below}

$$j^{\diamond}\{q^{\diamond}\} = \underset{n=1,\ldots,k(q)}{\arg\min} \frac{|c\{q^{\diamond}\}_n|}{\|\mathbf{b}\{q^{\diamond}\}_n\|}$$

{Update the objective functional $\chi_{\downarrow}$ for the block $q^{\diamond}$, as below}

$$\chi_{\downarrow}(q^{\diamond}) = \frac{|c\{q^{\diamond}\}_{j^{\diamond}\{q^{\diamond}\}}|}{\|\mathbf{b}\{q^{\diamond}\}_{j^{\diamond}\{q^{\diamond}\}}\|}$$

Decrease $i \leftarrow i - 1$

end while

3 Refinement by Swaps

The suitability of the HBW strategy for approximating the class of signals we are considering is a notable fact. However, the possibility of approximating every element of the partition completely independent of the others has the convenient feature of leaving room for straightforward parallelization. This is particularly important for implementations on Graphics Processing Units (GPU), for instance. Accordingly, an alternative worth exploring is to maintain the block independent approximation but allowing for possible eventual HBW refinements.

Assuming that the approximation (1) of each block in a signal partition is known, the goal now is to improve the signal approximation maintaining unaltered the total number of atoms K. The proposed refinement consists of movement of atoms controlled by the following instructions:

i) Use criterion (17) to remove one atom from the block $q^{\diamond}$. Let us call this block a 'donor' block and indicate it with the index q_{d}.

ii) Use criterion (9) to incorporate one atom in the approximation of the block $q^{\star}$. Let us call this block a 'receiver' and indicate it with the index q_{r}.

iii) Denote by δ_{d} the square norm of the error introduced at step i)by downgrading the approximation of the donor block q_{d}. Denote by δ_{r} the square norm of the gain in improving the approximation of the receiver q_{r}. Proceed according to the following rule:

If $\delta_{\mathrm{r}} > \delta_{\mathrm{d}}$ accept the change and repeat steps i) and ii). Otherwise stop.

The above procedure is termed HBW-SbR-OMP/OOMP according to whether the selection of the atom at stage ii) is realized through the OMP (c.f. (2)) or OOMP (c.f. (10)) criterion.

Notice that the HBW-SbR-OMP/OOMP methods, implemented as proposed above, possess the desirable feature of acting *only* if the approximation can be improved by the proposed swap. The possible situation for which $q_{\mathrm{r}} = q_{\mathrm{d}}$ corresponds to a swap in the original SbR-OMP/OOMP approach [17], which may take place here as a particular case.

Implementation Remark: For successive implementation of the steps i) and ii) one needs: to downgrade the vectors $\mathbf{b}\{q^{\diamond}\}_n$, $n = 1, \ldots k(q)$ at step i) and to upgrade these vectors at step ii). In order to allow for step ii) the orthogonal set $\{\mathbf{w}\{q^{\diamond}\}_n\}_{n=1}^{k(q)}$ has to be downgraded as well. This can be realized effectively by the plane rotation approach [30, 31].

Numerical Example III

We illustrate the HBW-SbR-OMP/OOMP approaches by applying them to the signal in Fig. 2, which is a flute exercise consisting of $N = 96256$ samples divided into $Q = 94$ blocks with $N_{\mathrm{b}} = 1024$ samples each.

Table 3 and 4 show the improvement in quality obtained by applying the HBW-SbR-OMP/OOMP approaches on outputs of the OMP/OOMP approximations. Notice that in most cases the SNR results are practically equivalent to those obtained by the HBW-OMP/OOMP approaches.

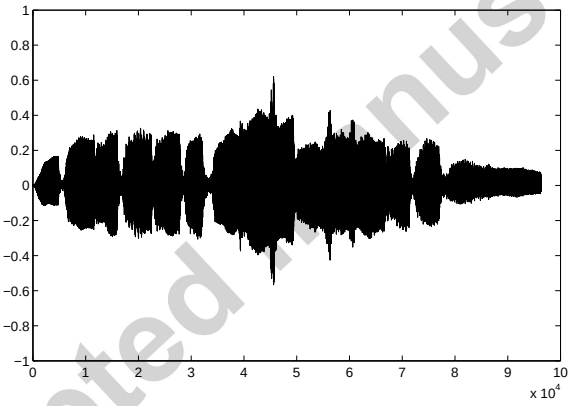

Figure 2: Flute exercise: $N = 96256$ samples.

4 Processing of large signals

As already discussed, the additional computational complexity introduced by the HBW raking of Q blocks is only $\mathrm{O}(Q)$ (over the complexity of a standard pursuit algorithm implemented without ranking the blocks). Storage does become more demanding though. The need for saving vectors (7) and (3), for instance, implies to have to store $2K$ vectors of size N_{b}. This requirement restricts the amount of blocks to be processed with a small system such as a standard laptop. Hence, we outline

Dict.	SR	OMP	Swaps	HBW-SbR-OMP	HBW-OMP
$\mathcal{B}^{\mathrm{c}}$	10.13	25.0	1426	26.80	26.79
$\mathcal{D}^{\mathrm{c2}}$	14.26	25.0	1013	26.69	26.67
$\mathcal{D}^{\mathrm{c4}}$	15.58	25.0	945	26.63	26.61
$\mathcal{B}^{\mathrm{s}}$	7.27	25.0	2612	26.72	26.73
$\mathcal{D}^{\mathrm{s2}}$	12.10	25.0	1397	27.24	27.27
$\mathcal{D}^{\mathrm{s4}}$	13.16	25.0	1435	27.27	27.25
$\mathcal{B}^{\mathrm{cs}}$	8.08	25.0	1822	26.65	26.65
$\mathcal{D}^{\mathrm{cs2}}$	29.00	25.0	636	27.38	27.37
$\mathcal{D}^{\mathrm{cs4}}$	35.17	25.0	518	27.36	27.36

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-SbR-OMP refinement to the outputs yielding the second column. For further comparison the last column shows the results corresponding to the HBW-OMP approach.

Dict.	SR	OOMP	Swaps	HBW-SbR-OOMP	HBW-OOMP
$\mathcal{B}^{\mathrm{c}}$	10.13	25.0	1426	26.80	26.79
$\mathcal{D}^{\mathrm{c2}}$	14.83	25.0	865	26.66	26.68
$\mathcal{D}^{\mathrm{c4}}$	15.80	25.0	929	26.60	26.61
$\mathcal{B}^{\mathrm{s}}$	7.27	25.0	2612	26.72	26.73
$\mathcal{D}^{\mathrm{s2}}$	12.58	25.0	1531	27.14	27.27
$\mathcal{D}^{\mathrm{s4}}$	13.34	25.0	1445	27.27	27.26
$\mathcal{B}^{\mathrm{cs}}$	8.08	25.0	1822	26.65	26.65
$\mathcal{D}^{\mathrm{cs2}}$	33.47	25.0	556	27.64	27.63
$\mathcal{D}^{\mathrm{cs4}}$	42.05	25.0	228	27.10	27.69

Table 4: Same description as in Table 3 but for the OOMP, HBW-SbR-OOMP, and HBW-OOMP approaches.

here a processing scheme which makes it possible the application of HBW techniques on very large signals. For this the whole signal needs to be divided into large segments of convenient size, say N_{s}. More specifically, a number of, say P blocks, of size N_{b} are grouped together to form S larger segments $\mathbf{g}\{s\} = \hat{\mathrm{J}}_{q=(s-1)P+1}^{sP}\mathbf{f}\{q\}$, $s = 1, \ldots, S$, where it is assumed that $N_{\mathrm{b}}P = N_{\mathrm{s}}$ and $SN_{\mathrm{s}} = N_{\mathrm{b}}Q$. Each segment $\mathbf{g}\{s\} \in \mathbb{R}^{N_{\mathrm{s}}}$ is approximated with a HBW strategy as an entire individual signal. The approximation of the original signal is assembled at the very end of the process by the operation $\mathbf{f}^{\mathrm{a}} = \hat{\mathrm{J}}_{s=1}^{S}\mathbf{g}^{\mathrm{a}}\{s\}$. In other words, the processing of large signals is realized by chopping the signal into segments and processing each segment independently of the others. Nevertheless, the question as to how to set the sparsity constraint for each segment 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 $\mathbf{f}$ split it into Q blocks of size N_{b}. Apply an invertible random permutation Π to scramble the block's location (the 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 $\tilde{\mathbf{f}} = \hat{\mathrm{J}}_{q=1}^{Q}\tilde{\mathbf{f}}\{q\}$.

ii) Group every P of the blocks $\tilde{\mathbf{f}}\{q\}$, $q = 1, \ldots, Q$ to have S segments $\tilde{\mathbf{g}}\{s\} = \hat{\mathrm{J}}_{q=(s-1)P+1}^{sP}\tilde{\mathbf{f}}\{q\}$, $s = 1, \ldots, S = \frac{Q}{P}$.

iii) Approximate each of the segments $\tilde{\mathbf{g}}\{s\}$ independently of the others, as it each segment were an entire signal, i.e., obtain $\tilde{\mathbf{g}}^{\mathrm{a}}\{s\} = \hat{\mathrm{J}}_{q=(s-1)P+1}^{sP}\tilde{\mathbf{f}}^{\mathrm{a}}\{q\}$, $s = 1, \ldots, S$.

iv) Assemble the segments to have the approximation of the whole signal as $\tilde{\mathbf{f}}^{\mathrm{a}} = \hat{\mathrm{J}}_{s=1}^{S}\tilde{\mathbf{g}}^{\mathrm{a}}\{s\}$.

v) Reverse the permutation of the block's location to obtain, from $\tilde{\mathbf{f}}^{\mathrm{a}}$, the approximation of the original signal $\mathbf{f}^{\mathrm{a}}$.

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/OOMP approach to produce a SR=11.53. The resulting SNR of the approximated signal is 29.12 dB with HBW-OMP and 30.15 dB

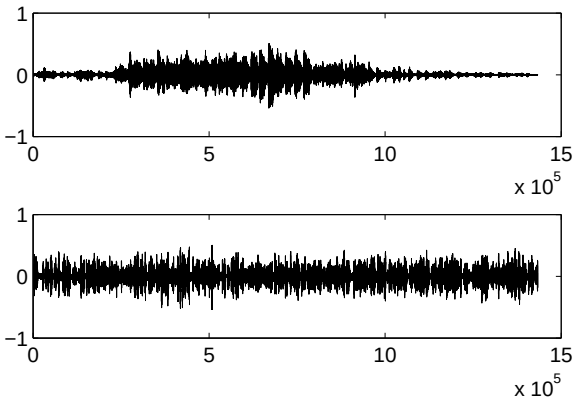

Figure 3: The top graph shows a piece (1433600 samples) of Piazzola music interpreted by Gidon Kremmer and orchestra. The bottom graph shows the resulting signal after the randomization of the blocks.

with HBW-OOMP. 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/OOMP, but without segmentation. In that case the quality corresponding to SR=11.53 is only 1.9% higher (SNR=29.66 dB with HBW-OMP and SNR=30.73 dB with HBW-OOMP). The conclusion is that, even if the largest the segments the less the distortion for the same sparsity, applying HBW-OMP/OOMP on segments of a large signal may be still significantly more advantageous than the independent approximation of the blocks. Indeed, in this example the latter yields a SNR of 25 dB with OMP and 26.1 dB with OOMP.

5 Conclusions

Cooperative greedy pursuit strategies for approximating a signal partition subjected to a global constraint on sparsity have been considered. The cooperation between partition units was realized in several ways: i)By ranking the partition units for their sequential stepwise approximation (HBW-OMP/OOMP) ii)By ranking the partition units for stepwise downgrading of the whole approximation (HBW-BOOMP) iii)By allowing for downgrading of some partition units (donors) and upgrading of another partitions units (receivers), to improve the approximation quality (HBW-SbR-OMP/OOMP). Comparisons of the OMP and OOMP criteria indicate that for the type of dictionaries

and signals considered here, the OOMP criterion renders improvements of sparsity up to 20% in relation to the equivalent strategy involving OMP.

The HBW algorithms maintain the complexity comparable with the totally independent approximation of each partition unit. The examples presented here, using trigonometric dictionaries for achieving high quality approximation of musics signals, provide a clear illustration of the benefits arising by the proposed strategies. Indeed, for a piano melody the increment in sparsity is 128% higher than the sparsity produced by approximating each partition unit up to same quality. Comparison of trigonometric dictionaries results, against results obtained with trigonometric basis, give an improvement of equivalent order: 133%. The mixed Cosine-Sine dictionary has yielded the best sparsity results, not only for the test signals included in this paper but for a significant number of other signals, all in the category of melodic music.

Because the proposed HBW implementation of pursuit strategies requires storage of order $N_{\mathrm{b}}K$ (N_{b} being the length of each block in the partition and K the number of total atoms to approximate the whole signals) a simple procedure for processing large signal was proposed: Previous to dividing the signal into independent segments, each of which to be approximated with the identical constraint of sparsity, the randomization of the partition units is advised.

Acknowledgements

The author is indebted to an anonymous Referee. Her/his comments, criticisms, and suggestions have been of much help to present the paper in the present form.

Appendix A

Algorithm 6 Atom Selection Procedure via FFT

Input: Residue $\mathbf{r}\{q^\star\}$. Dictionary Case ('Cos', 'Sin' or mixed 'Cos-Sin') and number of elements, M. Sequence $\{s\{q^\star\}_n\}_{n=1}^{M}$ if Case is 'Cos' or 'Sin', and $\{s^{\mathrm{c}}\{q^\star\}_n\}_{n=1}^{\frac{M}{2}}, \{s^{\mathrm{s}}\{q^\star\}_n\}_{n=1}^{\frac{M}{2}}$ if Case is 'Cos-Sin'. Vector $\widetilde{\mathbf{w}}\{q^\star\}_{k(q^\star)}$ to upgrade the sequences. Set of already selected atoms $\Gamma\{q^\star\}$ (sets $\Gamma^{\mathrm{c}}\{q^\star\}$ and $\Gamma^{\mathrm{s}}\{q^\star\}$ with the atoms in each component if Case is 'Cos-Sin').

Output: Selected index $\ell\{q^\star\}$ for a potential new atom for block $\{q^\star\}$.

Cases single 'Cos' or 'Sin'

{Call IPTrgFFT procedure, Algorithm 4, to calculate inner products, and apply Algorithm 7 to upgrade the sequence $\{s\{q^\star\}_n\}_{n=1}^{M}$ with respect to vector $\widetilde{\mathbf{w}}\{q^\star\}_{k(q^\star)}$}

$\mathbf{p}$=IPTrgFFT($\mathbf{r}\{q^\star\}, 2M$, Case),

{Select the index for a new potential atom for the block $q^\star$, as below}

$$\ell\{q^\star\} = \underset{\substack{n=1,\ldots,M \\ n\notin\Gamma\{q^\star\}}}{\arg\max} \frac{|p\{q^\star\}(n)|}{(1-s\{q^\star\}_n)^{\frac{1}{2}}}$$

Case 'Cos-Sin'

{Same precedure as for the previous case but for both 'Cos' and 'Sin' cases},

$\mathbf{p}^{\mathrm{c}}\{q^\star\}$=IPTrgFFT($\mathbf{r}\{q^\star\}, M$ 'Cos'); $\mathbf{p}^{\mathrm{s}}\{q^\star\}$=IPTrgFFT($\mathbf{r}\{q^\star\}, M$, 'Sin')

{Apply Algorithm 7 to upgrade the sequences $\{s^{\mathrm{c}}\{q^\star\}_n\}_{n=1}^{\frac{M}{2}}$ and $\{s^{\mathrm{s}}\{q^\star\}_n\}_{n=1}^{\frac{M}{2}}$ and selec the index}

$$\ell^{\mathrm{c}}\{q^\star\} = \underset{\substack{n=1,\ldots,\frac{M}{2} \\ n\notin\Gamma^{\mathrm{c}}\{q^\star\}}}{\arg\max} \frac{|p^{\mathrm{c}}\{q^\star\}(n)|}{(1-s^{\mathrm{c}}\{q^\star\}_n)^{\frac{1}{2}}}; \quad \ell^{\mathrm{s}}\{q^\star\} = \underset{\substack{n=1,\ldots,\frac{M}{2} \\ n\notin\Gamma^{\mathrm{s}}\{q^\star\}}}{\arg\max} \frac{|p^{\mathrm{s}}\{q^\star\}(n)|}{(1-s^{\mathrm{s}}\{q^\star\}_n)^{\frac{1}{2}}}$$

{Evaluate the maximum value, as below}

$$\mu^{\mathrm{c}}\{q^\star\} = \frac{|p^{\mathrm{c}}\{q^\star\}(\ell^{\mathrm{c}}\{q^\star\})|}{(1-s^{\mathrm{c}}\{q^\star\}_{\ell^{\mathrm{c}}\{q^\star\}})^{\frac{1}{2}}}; \quad \mu^{\mathrm{s}}\{q^\star\} = \frac{|p^{\mathrm{s}}\{q^\star\}(\ell^{\mathrm{s}}\{q^\star\})|}{(1-s^{\mathrm{s}}\{q^\star\}_{\ell^{\mathrm{s}}\{q^\star\}})^{\frac{1}{2}}}$$

$$\mu\{q^\star\} = \max(\mu^{\mathrm{c}}\{q^\star\}, \mu^{\mathrm{s}}\{q^\star\})$$

if $\mu = \mu^{\mathrm{c}}$ **then**

$\ell\{q^\star\} = \ell^{\mathrm{c}}\{q^\star\}$

else

$\ell\{q^\star\} = \ell^{\mathrm{s}}\{q^\star\} + \frac{M}{2}$

end if

Algorithm 7 Auxiliary Procedure via FFT

Input: Vector $\widetilde{\mathbf{w}}\{q^\star\}_{k(q^\star)}$. Dictionary Case: 'Cos', 'Sin' or 'Cos-Sin', and number of elements M. Sequence $\{s\{q^\star\}_n\}_{n=1}^{M}$, for Cases 'Cos' and 'Sin', or $\{s^{\mathrm{c}}\{q^\star\}_n\}_{n=1}^{\frac{M}{2}}$, and $\{s^{\mathrm{s}}\{q^\star\}_n\}_{n=1}^{\frac{M}{2}}$, for Case 'Cos-Sin'.

Output: Upgraded sequence, $\{s\{q^\star\}_n\}_{n=1}^{M}$, for Cases 'Cos' and 'Sin', or $\{s^{\mathrm{c}}\{q^\star\}_n\}_{n=1}^{\frac{M}{2}}$ and $\{s^{\mathrm{s}}\{q^\star\}_n\}_{n=1}^{\frac{M}{2}}$ for Case 'Cos-Sin'.

Case 'Cos' or 'Sin'

{Call IPTrgFFT procedure, Algorithm 4, to calculate inner products with vector $\widetilde{\mathbf{w}}\{q^\star\}_{k(q^\star)}$}

$\mathbf{p}\{q^\star\}$=IPTrgFFT($\widetilde{\mathbf{w}}\{q^\star\}_{k(q^\star)}, 2M$, Case)

{Upgrade the sequence, as below}

for $n = 1 : M$ **do**

$s\{q^\star\}_n = s\{q^\star\}_n + |p\{q^\star\}(n)|^2$

end for

Case 'Cos-Sin'

{Call IPTrgFFT procedure, Algorithm 4, to calculate the inner products between vector $\widetilde{\mathbf{w}}\{q^\star\}_{k(q^\star)}$ and each of the dictionary components: Cases 'Cos' and 'Sin'}

$\mathbf{p}^c$=IPTrgFFT($\widetilde{\mathbf{w}}\{q^\star\}_{k(q^\star)}, M$, 'Cos')

$\mathbf{p}^s$=IPTrgFFT($\widetilde{\mathbf{w}}\{q^\star\}_{k(q^\star)}, M$, 'Sin')

{Upgrade the sequences, as below}

for $n = 1 : \frac{M}{2}$ **do**

$s^{\mathrm{c}}\{q^\star\}_n = s^{\mathrm{c}}\{q^\star\}_n + |p^{\mathrm{c}}\{q^\star\}(n)|^2$

$s^{\mathrm{s}}\{q^\star\}_n = s^{\mathrm{s}}\{q^\star\}_n + |p^{\mathrm{s}}\{q^\star\}(n)|^2$

end for

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, and 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. Rezaiifar, 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. Biölcskei, "Block-Sparse Signals: Uncertainty Relations and Efficient Recovery", *IEEE Trans. Signal Process.*, **58**, 3042–3054 (2010).

[7] D. L. Donoho , Y. Tsaig , I. Drori , and 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] E. Candès, 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 Magazin*, **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. Matiol, and S. Bibi, "Hierarchized block wise image approximation by greedy pursuit strategies," *IEEE Signal Process. Letters*, **20**, 1175–1178 (2013).

[16] L. Rebollo-Neira and D. Lowe, "Optimized orthogonal matching pursuit approach", *IEEE Signal Process. Letters*, **9**, 137–140 (2002).

[17] M. Andrle and L. Rebollo-Neira, "A swapping-based refinement of orthogonal matching pursuit strategies", *Signal Processing*, **86**, 480–495 (2006).

[18] `http://www.nonlinear-approx.info/examples/node01.html`

[19] R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit", *IEEE Trans. Signal Process.,* **51**, 101–111 (2003).

[20] P. Leveau, E. Vincent, G. Richard, and L. Daudet, "Instrument-Specific Harmonic Atoms for Mid-Level Music Representation", *IEEE Trans Audio, Speech and Language Process.*, **16**, 116–128 (2008).

[21] R. Gribonval, "Fast Matching Pursuit with a multiscale dictionary of Gaussian Chirps", *IEEE Trans. Signal Process.,* **49**, 994–1001 (2001).

[22] E. Ravelli, G. Richard, and L. Daudet, "Union of MDCT Bases for Audio Coding", *IEEE Trans Audio, Speech and Language Process.*, **16**, 1361–1372 (2008).

[23] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T. Lee, and T. J. Sejnowski, "Dictionary learning algorithms for sparse representation", *Neural Comp.*, **15**, 349 – 396 (2003).

[24] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation", *IEEE Trans. Signal Process.* **54**, 4311–4322 (2006).

[25] R. Rubinstein, M. Zibulevsky, and M. Elad, "Double Sparsity: Learning Sparse Dictionaries for Sparse Signal Approximation", *IEEE Trans. Signal Process.* **58**, 1553–1564 (2010).

[26] I. Tošić and P. Frossard, "Dictionary Learning: What is the right representation for my signal?", *IEEE Signal Process. Magazine*, **28**, 27–38 (2011).

[27] L. Rebollo-Neira and J.Bowley, Sparse representation of astronomical images, *Journal of The Optical Society of America A,* **30**, 758–768 (2013).

[28] M. Andrle, L. Rebollo-Neira, and E. Sagianos, "Backward-optimized orthogonal matching pursuit approach", *IEEE Signal Proc. Let.*,**11**,705–708 (2004).

[29] L. Rebollo-Neira, "Backward adaptive biorthogonalization", *Int. J. Math. Math. Sci.* 1843–1853 (2004).

[30] Å. Björck, "Numerical Methods for Least Squares Problems", SIAM, Philadelphia, 1996.

[31] G.W. Stewart, "On the stability of sequential updates and downdates", *IEEE Trans. Signal Process.*, **43**, 2642–2648 (1995).