Trigonometric dictionary based codec for music compression with high quality recovery

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

***Abstract*—A codec for compression of music signals is proposed. The method belongs to the class of transform lossy compression. It is conceived to be applied in the high quality recovery range though. The transformation, endowing the codec with its distinctive feature, relies on the ability to construct high quality sparse approximation of music signals. This is achieved by a redundant trigonometric dictionary and a dedicated pursuit strategy. The potential of the approach is illustrated by comparison with the OGG Vorbis format, on a sample consisting of clips of melodic music. The comparison evidences remarkable improvements in compression performance for the identical quality of the decompressed signal.**

Keywords: Music Compression, Hierarchized Block Wise Multichannel Optimized Orthogonal Matching Pursuit. Trigonometric Dictionaries.

I. Introduction

For the most part the techniques for compressing high fidelity music have been developed within the lossless compression framework [1]–[8]. Because lossless music compression algorithms are reversible, which implies that can reproduce the original signal when decompressing the file, the efficiency of those algorithms are compared on the reduction of file size and speed or the process. Conversely, lossy compression introduces irreversible loss and should be compared also taking into account the quality of the decompressed data.

This work focusses on lossy compression of music signals with *high quality recovery*. This means that the recovered signal should be very similar to the original one, with respect to the Euclidean distance of the data points. In other words, the recovered signal should yield a high Signal to Noise Ratio (SNR). The proposed approach is based on the ability to construct a high quality sparse representation of a piece of music. The sparsity is achieved by selecting elements from a redundant trigonometric dictionary, through a dedicated greedy pursuit methodology which approximates simultaneously all the channels of a stereo signal. Pursuit strategies for approximating multiple signals sharing the same sparsity structure are refereed in the literature to as several names: Vector greedy algorithms [9], [10], simultaneous greedy approximations [10], [11], and multiple measurement vectors (MMV). [12], [13]. Following previous work [14], [15], we dedicate greedy methodologies for simultaneous representation to approximate a *partitioned multichannel music signal* subjected to a global constraint on the sparsity.

The use of redundant dictionaries for constructing sparse representations is known to be a successful approach in a variety of signal processing applications [16]–[23]. In particular for compression of facial images [24]–[26]. This paper extends the range of successful applications by presenting a number of examples were the proposed dictionary based codec for compression of music signals yields remarkable results, in relation to file size and quality of the recovered signal.

A. Paper contributions

The central aim of the paper is to produce a proof of concept of the proposed codec. The proposal falls within the usual transform coding scheme. It consists of three main steps:

i) Transformation of the signal.
ii) Quantization of the transformed data.
iii) Bit-stream entropy coding.

However, we move away from the traditional compression techniques at the very beginning. Instead of considering an orthogonal transformation, the first step is realized by approximating the signal using a trigonometric redundant dictionary. In a previous work [15], the dictionary has been proven to yield stunning sparse approximation of melodic music, if processed by the adequate greedy strategy. We demonstrate now that the sparsity renders compression.

The Hierarchized Block Wise Optimized Orthogonal Matching Pursuit (HBW-OOMP) method in [15] is generalized here, to consider the simultaneous approximation of multichannel signals. Within the proposed scheme the advantage of simultaneous approximation is twofold: a)It reduces the processing time at the transformation stage and b)It reduces the number of parameters to be stored, which improves compression performance.

The success of the codec, designed to achieve high quality recovery, is illustrated by comparison with the OGG Vorbis compression format. Accessing the quality of the recovered signal by the classic SNR, a substantial gain in compression, for the same quality of the decompressed signal, is demonstrated on a number of clips of melodic music.

B. Paper Organization

Sec. II introduces the notation and some relevant mathematical background. Sec. III discusses the HBW strategy to approximate simultaneously a multichannel signal. Sec. IV describes a simple compression scheme that benefits from the achieved sparse approximation of the multichannel signal. Sec. V demonstrates the potential of the technique by comparison with the OGG format. The final conclusions are presented in Sec. VI

II. Mathematical background and notational

Throughout the paper $\mathbb{R}$ and $\mathbb{N}$ stand for the sets of real and natural numbers, respectively. Low boldface letters are used to indicate Euclidean vectors and capital boldface letters to indicate matrices. Their corresponding component are represented using standard mathematical fonts, e.g., $\mathbf{f} \in \mathbb{R}^N$, $N \in \mathbb{N}$ is a vector of components $f(i)$, $i = 1, \ldots, N$ and $\mathbf{F} \in \mathbb{R}^{N \times L}$ is a matrix of real entries $F(i, j)$, $i = 1, \ldots, N$, $j = 1, \ldots, L$.

An L-channel signal is represented as a matrix $\mathbf{F} \in \mathbb{R}^{N \times L}$ the columns of which are the channels, indicated as vectors $\mathbf{f}_j \in \mathbb{R}^N$, $j = 1, \ldots, L$. Thus, a single channel reduces to a vector. A partition of a multichannel signal $\mathbf{F} \in \mathbb{R}^{N \times L}$ is realized by a set of disjoint pieces $\mathbf{F}_q \in \mathbb{R}^{N_b \times L}$, $q = 1, \ldots, Q$, which for simplicity are assumed to be all of the same size and such that $QN_b = N$, i.e., for each channel it holds that $\mathbf{f}_j = \hat{\mathrm{J}}_{q=1}^{Q} \mathbf{f}_{q,j}$, where the concatenation operation $\hat{\mathrm{J}}$ is defined as follows: $\mathbf{f}_j$ is a vector in $\mathbb{R}^{QN_b}$ having components $f_j(i) = f_{q,j}(i - (q-1)N_b)$, $i = (q-1)N_b + 1, \ldots, qN_b$, $q = 1, \ldots, Q$. In the adopted notation $f_j(i)$ can also be indicated as $F(i, j)$ and $f_{q,j}(i)$ as $F_q(i, j)$. Hence

$$\|\mathbf{F}\|_F^2 = \sum_{q=1}^{Q} \|\mathbf{F}_q\|_F^2,$$

where each $\mathbf{F}_q$ is a matrix consisting of the channels $\mathbf{f}_{q,j} \in \mathbb{R}^{N_b}$, $j = 1, \ldots, L$, as columns and $\| \cdot \|_F$ indicates the Frobenius norm. Accordingly,

$$\|\mathbf{F}_q\|_F^2 = \sum_{j=1}^{L} \|\mathbf{f}_{q,j}\|^2,$$

where $\|.\|$ indicates the Euclidean norm induced by the Euclidean inner product $\langle \cdot, \cdot \rangle$.

Definition 1. *A* dictionary *for $\mathbb{R}^{N_b}$ is an* over-complete *set of normalized to unity elements $\mathcal{D} = \{\mathbf{d}_n \in \mathbb{R}^{N_b}\, ; \|\mathbf{d}_n\| = 1\}_{n=1}^{M}$, which are called* atoms.

Approximation assumption: *Given a dictionary $\mathcal{D}$ and a multichannel signal partitioned into Q blocks $\mathbf{f}_{q,j} \in \mathbb{R}^{N_b}$, $j = 1, \ldots, L, q = 1, \ldots, Q$, as described above, the k_q-term approximation of each block is assumed to be of the form*

$$\mathbf{f}_{q,j}^{k_q} = \sum_{n=1}^{k_q} c_{q,j}^{k_q}(n) \mathbf{d}_{\ell_n^q}, \quad j = 1, \ldots, L,\, q = 1, \ldots, Q, \quad (1)$$

where the atoms $\mathbf{d}_{\ell_n^q}$, $n = 1, \ldots, k_q$ are the same *for all the channels corresponding to a particular block q.*

Before discussing how to select the atoms in (1) it is convenient to review some properties of an orthogonal projector. Let's start by recalling its definition.

Definition 2. *An operator $\hat{P}_{\mathbb{V}_{k_q}^q}$ is an orthogonal projection operation onto $\mathbb{V}_{k_q}^q \subset \mathbb{R}^{N_b}$ if and only if:*

a) *$\hat{P}_{\mathbb{V}_{k_q^q}}$ is idempotent, i.e., $\hat{P}_{\mathbb{V}_{k_q^q}} \hat{P}_{\mathbb{V}_{k_q^q}} = \hat{P}_{\mathbb{V}_{k_q^q}}$.*

b) *$\hat{P}_{\mathbb{V}_{k_q^q}} \mathbf{g} = \mathbf{g}$ if $\mathbf{g} \in \mathbb{V}_{k_q^q}$ and $\hat{P}_{\mathbb{V}_{k_q^q}} \mathbf{g}^\perp = 0$ if $\mathbf{g}^\perp \in \mathbb{V}_{k_q^q}^\perp$, with $\mathbb{V}_{k_q^q}^\perp$ indicating the orthogonal complement of $\mathbb{V}_{k_q^q}$ in $\mathbb{R}^{N_b}$.*

The following properties will be used in the proofs of subsequente theorems.

i) An orthogonal projector operator $\hat{P}_{\mathbb{V}_{k_q^q}}$ is hermitian, i.e. for all $\mathbf{h}$ and $\mathbf{g}$ in $\mathbb{R}^{N_b}$ it is true that $\langle \mathbf{h}, \hat{P}_{\mathbb{V}_{k_q}^q} \mathbf{g} \rangle = \langle \hat{P}_{\mathbb{V}_{k_q}^q} \mathbf{h}, \mathbf{g} \rangle$.

ii) If $\mathbb{V}_{k_q+1}^q$ is constructed as $\mathbb{V}_{k_q+1}^q = \mathbb{V}_{k_q}^q + \mathbf{d}_{\ell_{k_q+1}^q}$, for all $\mathbf{h} \in \mathbb{R}^{N_b}$ it holds that

$$\hat{P}_{\mathbb{V}_{k_q+1}^q} \mathbf{h} = \hat{P}_{\mathbb{V}_{k_q}^q} \mathbf{h} + \mathbf{w}_{k_q+1}^q \frac{\langle \mathbf{w}_{k_q+1}^q, \mathbf{h} \rangle}{\|\mathbf{w}_{k_q+1}^q\|^2},$$

with $\mathbf{w}_{k_q+1}^q = \mathbf{d}_{\ell_{k_q+1}^q} - \hat{P}_{\mathbb{V}_{k_q}^q} \mathbf{d}_{\ell_{k_q+1}^q}$.

Theorem 1. *Let $\mathbf{F}_q$ be a $N_b \times L$ matrix the columns of which are the L signals $\mathbf{f}_{q,j} \in \mathbb{R}_b^N$, $j = 1, \ldots, L$ and let $\mathbf{F}_q^{k_q}$ be the matrix with the corresponding k_q-term approximations $\mathbf{f}_{q,j}^{k_q} \in \mathbb{V}_{k_q^q}$, $j = 1, \ldots, L$. For the error $\|\mathbf{F}_q - \mathbf{F}_q^{k_q}\|_F^2$ to be minimum the k_q-term approximations must satisfy: $\mathbf{f}_{q,j}^{k_q} = \hat{P}_{\mathbb{V}_{k_q^q}} \mathbf{f}_{q,j}$, $j = 1, \ldots, L$.*

Proof. $\|\mathbf{F}_q - \mathbf{F}_q^{k_q}\|_F^2$ can be expressed as

$$\begin{aligned} \|\mathbf{F}_q - \mathbf{F}_q^{k_q}\|_F^2 &= \sum_{j=1}^{L} \langle \mathbf{f}_{q,j} - \mathbf{f}_{q,j}^{k_q}, \mathbf{f}_{q,j} - \mathbf{f}_{q,j}^{k_q} \rangle \\ &= \sum_{j=1}^{L} \|\mathbf{f}_{q,j}\|^2 - 2\langle \mathbf{f}_{q,j}, \mathbf{f}_{q,j}^{k_q} \rangle + \|\mathbf{f}_{q,j}^{k_q}\|^2. (2) \end{aligned}$$

Since $\mathbf{f}_{q,j}^{k_q}$ is an element of $\mathbb{V}_{k_q}^q$ we can write it as $\mathbf{f}_{q,j}^{k_q} = \hat{P}_{\mathbb{V}_{k_q}^q} \mathbf{f}_{q,j} + \mathbf{g}_{q,j}$ for some $\mathbf{g}_{q,j} \in \mathbb{V}_{k_q}^q$. The corresponding replacements in (2), and the fact that $\mathbf{g}_{q,j} = \hat{P}_{\mathbb{V}_{k_q}^q} \mathbf{g}_{q,j}$ and $\hat{P}_{\mathbb{V}_{k_q}^q}$ is idempotent and hermitian, lead to the expression

$$\|\mathbf{F}_q - \mathbf{F}_q^{k_q}\|_F^2 = \sum_{j=1}^{L} \|\mathbf{f}_{q,j}\|^2 - \langle \mathbf{f}_{q,j}, \hat{P}_{\mathbb{V}_{k_q}} \mathbf{f}_{q,j} \rangle + \|\mathbf{g}_{q,j}\|^2, \quad (3)$$

from where it follows that $\|\mathbf{F}_q - \mathbf{F}_q^{k_q}\|_F^2$ is minimum if $\mathbf{g}_{q,j} = 0$, $j = 1, \ldots, L$ i.e. $\mathbf{f}_{q,j}^{k_q} = \hat{P}_{\mathbb{V}_{k_q}} \mathbf{f}_{q,j}$, $j = 1, \ldots, L$. □

Corollary 1. *The statement of Theorem 1 also minimizes the norm of the total residual $\mathbf{R}^K = \mathbf{F} - \mathbf{F}^K$ in approximating the whole multichannel signal, with $\mathbf{F}^K = \hat{\mathrm{J}}_{q=1}^{Q} \mathbf{F}_q^{k_q}$ and $K = \sum_{q=1}^{Q} k_q$.*

Proof. It readily follows by the a definition of the adopted disjoint partition:

$$\|\mathbf{R}^K\|_F^2 = \sum_{q=1}^{Q} \|\mathbf{F}_q - \mathbf{F}_q^{k_q}\|_F^2$$

is obviously minimum if each $\|\mathbf{F}_q - \mathbf{F}_q^{k_q}\|_F^2$ is minimum. □

Assuming, for the moment, that the sets of indices $\Gamma_q = \{\ell_n^q\}_{n=1}^{k_q}$ labeling the atoms in (1) are known, we recall at this point an effective construction of the required orthogonal projector for optimizing the approximation. Such a projection is given in terms of biorthogonal vectors as follows:

$$\hat{P}_{\mathbb{V}_k^q} \mathbf{f}_{q,j} = \sum_{n=1}^{k_q} \mathbf{d}_{\ell_n^q} \langle \mathbf{b}_n^{k_q,q}, \mathbf{f}_{q,j} \rangle = \sum_{n=1}^{k_q} c_{q,j}^{k_q}(n) \mathbf{d}_{\ell_n^q}. \quad (4)$$

For a fixed q the vectors $\mathbf{b}_n^{k_q,q}$, $n=1,\ldots,k_q$ are biorthogonal to the selected atoms $\mathbf{d}_{\ell_n^q}$, $n=1,\ldots,k_q$ and span the identical subspace, i.e.,

$$\mathbb{V}_{k_q}^q = \text{span}\{\mathbf{b}_n^{k_q,q}\}_{n=1}^{k_q} = \text{span}\{\mathbf{d}_{\ell_n}^q\}_{n=1}^{k_q}.$$

Such vectors can be adaptively constructed, from $\mathbf{b}_1^{1,q} = \mathbf{w}_1^q = \mathbf{d}_{\ell_1}^q$, through the recursion formula [27]:

$$\begin{aligned}\mathbf{b}_n^{k_q+1,q} &= \mathbf{b}_n^{k_q,q} - \mathbf{b}_{k_q+1}^{k_q+1,q}\langle \mathbf{d}_{\ell_{k_q+1}}^q, \mathbf{b}_n^{k_q,q}\rangle, \quad n=1,\ldots,k_q,\\ \mathbf{b}_{k_q+1}^{k_q+1,q} &= \mathbf{w}_{k_q+1}^q/\|\mathbf{w}_{k_q+1}^q\|^2,\end{aligned} \tag{5}$$

with

$$\mathbf{w}_{k_q+1}^q = \mathbf{d}_{\ell_{k_q+1}}^q - \sum_{n=1}^{k_q} \frac{\mathbf{w}_n^q}{\|\mathbf{w}_n^q\|^2}\langle \mathbf{w}_n^q, \mathbf{d}_{\ell_{k_q+1}}^q\rangle. \tag{6}$$

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

$$\mathbf{w}_{k_q+1}^q \leftarrow \mathbf{w}_{k_q+1}^q - \sum_{n=1}^{k_q} \frac{\mathbf{w}_n^q}{\|\mathbf{w}_n^q\|^2}\langle \mathbf{w}_n^q, \mathbf{w}_{k_q+1}^q\rangle. \tag{7}$$

The alternative representation of $\hat{P}_{\mathbb{V}_{k_q}^q}$, in terms of vectors $\mathbf{w}_n^q$, $n=1,\ldots,k_q$, gives the decompositions:

$$\hat{P}_{\mathbb{V}_{k_q}^q}\mathbf{f}_{q,j} = \sum_{n=1}^{k_q} \mathbf{w}_n^q \frac{\langle \mathbf{w}_n^q, \mathbf{f}_{q,j}\rangle}{\|\mathbf{w}_n^q\|^2}, \quad j=1,\ldots,L,\ q=1,\ldots,Q. \tag{8}$$

While these decompositions are not the representations of interest (c.f. (1)) they play a central role in the derivations of the next section.

III. Multichannel HBW strategy

The HBW version of pursuit strategies [14], [15] is a dedicated implementation of those techniques, specially designed for approximating a signal partition subjected to a global constraint on sparsity. The approach operates by raking the partition units for their sequential stepwise approximation. In this section we extend the HBW method to consider the case of a multichannel signal, within the approximation assumption specified in Sec. II.

Theorem 2. *Considerer that, for each block q, the k_q-term atomic decompositions* (1) *fulfilling that $\mathbf{f}_{q,j}^{k_q} = \hat{P}_{\mathbb{V}_{k_q}^q}\mathbf{f}_{q,j}$ are known, with $\mathbb{V}_{k_q}^q = \text{span}\{\mathbf{d}_{\ell_n^q}\}_{n=1}^{k_q}$. Let the indices $\ell_{k_q+1}^q \notin \{\ell_n^q\}_{n=1}^{k_q}$ be selected, for each q-value, by the same criterion as the one used to choose the atoms in* (1). *In order to minimize the square norm of the total residual $\mathbf{R}^{K+1}$, with $K=\sum_{q=1}^Q k_q$, the atomic decomposition to be upgraded at iteration $K+1$ should correspond to the block $q^\star$ such that*

$$q^\star = \operatorname*{arg\,max}_{q=1,\ldots,Q} \frac{\sum_{j=1}^L |\langle \mathbf{w}_{k_q+1}^q, \mathbf{f}_{q,j}\rangle|^2}{\|\mathbf{w}_{k_q+1}^q\|^2}, \tag{9}$$

with $\mathbf{w}_1^q = \mathbf{d}_{\ell_1^q}$ and $\mathbf{w}_{k_q+1}^q = \mathbf{d}_{\ell_{k_q+1}^q} - \hat{P}_{\mathbb{V}_{k_q}^q}\mathbf{d}_{\ell_{k_q+1}^q}$.

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

$$\mathbf{R}^{K+1} = \hat{\mathrm{J}}_{\substack{p=1\\p\neq q}}^{Q} \mathbf{R}_p^{k_p}\ \hat{\mathrm{J}}\ \mathbf{R}_q^{k_q+1}.$$

Then,

$$\|\mathbf{R}^{K+1}\|_F^2 = \sum_{\substack{p=1\\p\neq q}}^{Q} \|\mathbf{R}_p^{k_p}\|_F^2 + \|\mathbf{R}_q^{k_q+1}\|_F^2,$$

so that $\|\mathbf{R}^{K+1}\|$ is minimized by the minimum value of $\|\mathbf{R}_q^{k_q+1}\|_F^2$. Moreover, by definition $\|\mathbf{R}_q^{k_q+1}\|_F^2 = \|\mathbf{F}_q - \mathbf{F}_q^{k_q+1}\|_F^2$ and, from (2) and the fact that $\mathbf{f}_{q,j}^{k_q+1} = \hat{P}_{\mathbb{V}_{k_q+1}^q}\mathbf{f}_{q,j}$, we can write:

$$\|\mathbf{R}_q^{k_q+1}\|_F^2 \quad = \quad \sum_{j=1}^L \|\mathbf{f}_{q,j}\|^2 - \|\hat{P}_{\mathbb{V}_{k_q+1}^q}\mathbf{f}_{q,j}\|^2. \tag{10}$$

Then, $\|\mathbf{R}_q^{k_q+1}\|_F^2$ is minimum if $\sum_{j=1}^L \|\hat{P}_{\mathbb{V}_{k_q+1}^q}\mathbf{f}_{q,j}\|^2$ is maximum. Applying the property ii) of an orthogonal projector listed in Sec. II we have:

$$\sum_{j=1}^L \|\hat{P}_{\mathbb{V}_{k_q+1}^q}\mathbf{f}_{q,j}\|^2 = \sum_{j=1}^L \|\hat{P}_{\mathbb{V}_{k_q}^q}\mathbf{f}_{q,j}\|^2 + \sum_{j=1}^L \frac{|\langle \mathbf{w}_{k_q+1}^q, \mathbf{f}_{q,j}\rangle|^2}{\|\mathbf{w}_{k_q+1}^q\|^2},$$

with $\mathbf{w}_{k_q+1}^q = \mathbf{d}_{\ell_{k_q+1}^q} - \hat{P}_{\mathbb{V}_{k_q}^q}\mathbf{d}_{\ell_{k_q+1}^q}$. Because $\hat{P}_{\mathbb{V}_{k_q}^q}\mathbf{f}_{q,j}$ is fixed at iteration $K+1$, we are in a position to conclude that $\|\mathbf{R}^{K+1}\|_F^2$ is minimized by upgrading the atomic decomposition of the block $q^\star$ satisfying (9). □

Theorem (2) gives the HBW prescription optimizing the raking of the blocks in a multichannel signal partition, for their sequential stepwise approximation. This is irrespective of what the criterion for choosing the atoms for the approximation is. The method as a whole depends on that criterion, of course. One can use for example an extension of the Orthogonal Matching Pursuit (OMP) criterion, which when applied to multichannel signals has been termed simultaneous OMP in [11]. According to this criterion the indices of the atoms in the approximation of each block-q are such that

$$\ell_{k_q+1}^q = \operatorname*{arg\,max}_{n=1,\ldots,M} \sum_{j=1}^L |\langle \mathbf{d}_n, \mathbf{r}_{q,j}^{k_q}\rangle|, \tag{11}$$

where $\mathbf{r}_{q,j}^0 = \mathbf{f}_{q,j}$ and $\mathbf{r}_{q,j}^{k_q} = \mathbf{f}_{q,j} - \hat{P}_{\mathbb{V}_{k_q}^q}\mathbf{f}_{q,j}$. Alternatively, the extension of OMP to multichannels which is known as MMV-OMP [12], [13] selects the index fulfilling

$$\ell_{k_q+1}^q = \operatorname*{arg\,max}_{n=1,\ldots,M} \sum_{j=1}^L |\langle \mathbf{d}_n, \mathbf{r}_{q,j}^{k_q}\rangle|^2. \tag{12}$$

The optimization of the OMP criterion to select the atoms minimizing the norm of the residual error for each block goes with several names, according to the context were it was derived and the actual implementation. In one of the earliest references [28] is called Orthogonal Least Square. In others is called Order Recursive Matching Pursuit (ORMP) and in particular for multichannel signals MMV-ORMP [12]. The implementation we adopt here is termed Optimized Orthogonal Matching Pursuit (OOMP) [27] and will be termed

OOMPMl for multichannel signals. The approach selects the atoms $\ell^q_{k_q+1}$, for the approximation of block q, in other to minimize the norm of the residual error $\|\mathbf{F}_q - \mathbf{F}_q^{k_q+1}\|$ for that block. Those atoms correspond to the indices selected as:

$$\ell^q_{k_q+1} = \operatorname*{arg\,max}_{\substack{n=1,\dots,M\\ n\notin\Gamma_q}} \frac{\sum_{j=1}^L |\langle \mathbf{d}_n, \mathbf{r}^{k_q}_{q,j}\rangle|^2}{1-\sum_{i=1}^{k_q} |\langle \mathbf{d}_n, \tilde{\mathbf{w}}^q_i\rangle|^2}, \quad q=1,\dots,Q, \tag{13}$$

where $\Gamma_q = \{\ell^q_n\}_{n=1}^{k_q}$, $\mathbf{r}^{k_q}_{q,j} = \mathbf{f}_{q,j} - \hat{P}_{\mathbb{V}^q_{k_q}} \mathbf{f}_{q,j}$, with $\mathbf{r}^0_{q,j} = \mathbf{f}_{q,j}$, and $\tilde{\mathbf{w}}^q_i = \frac{\mathbf{w}^q_i}{\|\mathbf{w}^q_i\|}$, with $\mathbf{w}^q_i$, $i=1,\dots,k_q$ as in (6). The proof follows as in Theorem 2, but fixing the value of q and taking the maximization over the index.

As will be discussed in the next section, the used of trigonometric dictionaries reduces the complexity of the calculations in (11), (12), and (13).

The particularity of the OOMPMl implementation being that the coefficients of the atomic decomposition (1) are calculated using vectors (5) which are adaptively upgraded together with the selection of each new atom. For the qth-block the coefficients in the atomic decompositions (1) are computed as:

$$c^{k_q}_{q,j}(n) = \langle \mathbf{b}^{k_q,q}_n, \mathbf{f}_{q,j}\rangle,\ n=1,\dots,k_q,,\ j=1,\dots,L, \tag{14}$$

with $\mathbf{b}^{k_q,q}_n$ as in (5). Thus, when the channels have similar sparsity structure by approximating all of them simultaneously the complexity is reduced.

The HBW-OMPMl/OOMPMl approach is implemented by the following steps:

1) Initialize the algorithm by selecting the 'potential' first atom for the atomic decomposition of every block q, according to criterion (12) or (13). For $q=1,\dots,Q$ set: $k_q=1, \mathbf{w}^q_1 = \mathbf{b}^{1,q}_1 = \mathbf{d}_{\ell^q_1}$.
2) Use criterion (9) for selecting the block $q^\star$ to upgrade the atomic decomposition by incorporating the atom corresponding to the index $\ell^{q^\star}_{k_{q^\star}}$. If $k_{q^\star} > 1$ upgrade vectors (5) for block $q^\star$.
3) Increase $k_{q^\star} \leftarrow k_{q^\star}+1$ and select a new potential atom for the atomic decomposition of block $q^\star$, using the same criterion as in 1). Compute the corresponding $\mathbf{w}^{q^\star}_{k_{q^\star}}$ (c.f. (6)).
4) Check if, for a given K, the condition $\sum_{q=1}^Q k_q = K+1$ has been met. Otherwise repeat steps 2) - 4).
5) For each block, $q=1,\dots,Q$, calculate the coefficients in (1) as in (14).

A. Implementation details with Trigonometric Dictionaries

In [15] we illustrate the clear advantage of approximating music using a mixed dictionary with components $\mathcal{D}^c$ and $\mathcal{D}^s$ as below

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

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

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

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

Fixing $M=2N_b$ a dictionary redundancy four is constructed as $\mathcal{D} = \mathcal{D}^c \cup \mathcal{D}^s$. In addition to yielding highly sparse representation of melodic music, this trigonometric dictionary leaves room for reduction in the computational complexity of the algorithms and also in the storage demands. As discussed below, savings are made possible in an straightforward manner via the Fast Fourier Transform (FFT).

Given a vector $\mathbf{y} \in \mathbb{R}^M$ we define

$$\mathcal{F}(\mathbf{y}, n, M) = \sum_{j=1}^M y(j) e^{\imath\, 2\pi \frac{(n-1)(j-1)}{M}}, \quad n=1,\dots,M. \tag{15}$$

When $M=N_b$ (15) is the Discrete Fourier Transform of vector $\mathbf{y}\in\mathbb{R}^{N_b}$, which can be evaluated using FFT. If $M > N_b$ we can still calculate (15) via FFT by padding with $(M-N_b)$ zeros the vector $\mathbf{y}$. Accordingly, (15) is a useful tool for calculating inner products with the atoms in dictionaries $\mathcal{D}^c$ and $\mathcal{D}^s$. For $n=1,\dots,M$ it holds that

$$\sum_{j=1}^{N_b} \cos\frac{\pi(2j-1)(n-1)}{2M} y(j) = \mathrm{Re}\left(e^{-\imath\,\frac{\pi(n-1)}{2M}} \mathcal{F}(\mathbf{y}, n, 2M)\right) \tag{16}$$

and for $n=2,\dots,M+1$

$$\sum_{j=1}^{N_b} \sin\frac{\pi(2j-1)(n-1)}{2M} y(j) = \mathrm{Im}\left(e^{-\imath\,\frac{\pi(n-1)}{2M}} \mathcal{F}(\mathbf{y}, n, 2M)\right), \tag{17}$$

where $\mathrm{Re}(z)$ indicates the real part of z and $\mathrm{Im}(z)$ its imaginary part. The assistance of the FFT for performing the inner products (16) and (17) reduces the complexity in calculating the maximizing function in (12), (which is also the numerator in (13)) from $2MN_b$ to $2M(\log_2 2M+1)$. Storing the sums $S^{k_q-1}_n = \sum_{i=1}^{k_q-1} |\langle \mathbf{d}_n, \tilde{\mathbf{w}}^q_i\rangle|^2$, $n=1\dots,M$ the denominator in the right hand side of (13) involves the calculation of $|\langle \mathbf{d}_n, \tilde{\mathbf{w}}^q_{k_q}\rangle|^2$, $n=1\dots,M$, which can also be computed via FFT. Hence the complexity for the calculation of the denominator in (13) is of the same order as that for the calculation of the numerator. Criterion (13) in general yields higher sparsity, hence it reduces the cost in calculating vectors (5).

The MATLAB function implementing the HBW-OOMPMl approach, named HBW-OOMPMlTrgFFT when dedicated to the above trigonometric dictionary, has been made available on [29].

IV. A SIMPLE CODING STRATEGY

Previously to entropy encoding the coefficients resulting from approximating a signal by partitioning, the real numbers need to be converted into integers. This operation is known as quantization. For the numerical example of Sec. V we adopt a simple uniform quantization technique: The absolute value coefficients $|c_{q,j}(n)|$, $n=1\dots,k_q$, $q=1,\dots,Q$, $j=1,\dots,L$ are converted to integers as follows:

$$c^\Delta_{q,j}(n) = \lfloor \frac{|c_{q,j}(n)|}{\Delta} + \frac{1}{2} \rfloor, \tag{18}$$

where $\lfloor x \rfloor$ indicates the largest integer number smaller or equal to x and Δ is the quantization parameter. The signs of the coefficients, represented as $\mathbf{s}_{q,j}$, $q = 1, \ldots, Q$, $j = 1, \ldots, L$, are encoded separately using a binary alphabet. As for the indices of the atoms, which are common to the atomic decompositions of all the channels, they are firstly sorted in ascending order $\ell_i^q \to \tilde{\ell}_i^q$, $i = 1, \ldots, k_q$, which guarantees that, for each q value, $\tilde{\ell}_i^q < \tilde{\ell}_{i+1}^q$, $i = 1, \ldots, k_q - 1$. This order of the indices induces an order in the coefficients, $\mathbf{c}_{q,j}^\Delta \to \tilde{\mathbf{c}}_{q,j}^\Delta$ and in the corresponding signs $\mathbf{s}_{q,j} \to \tilde{\mathbf{s}}_{q,j}$. The advantage introduced by the ascending order of the indices is that they can be stored as smaller positive numbers by taking differences between two consecutive values. Indeed, by defining $\delta_i^q = \tilde{\ell}_i^q - \tilde{\ell}_{i-1}^q$, $i = 2, \ldots, k_q$ the follow string stores the indices for block q with unique recovery $\tilde{\ell}_1^q, \delta_2^q, \ldots, \delta_{k_q}^q$. The number '0' is then used to separate the string corresponding to different blocks and entropy code a long string, st_{ind}, which is built as

$$st_{\text{ind}} = [\tilde{\ell}_1^1, \ldots, \delta_{k_1}^1, 0, \cdots, 0, \cdots, \tilde{\ell}_1^{k_Q}, \ldots, \delta_{k_Q}^Q]. \quad (19)$$

The corresponding quantized magnitude of the coefficients of each channel are concatenated in the L strings st_{cf}^j, $j = 1, \ldots, L$ as follows:

$$st_{\text{cf}}^j = [\tilde{c}_{1,j}^\Delta(1), \ldots, \tilde{c}_{1,j}^\Delta(k_1), \cdots, \tilde{c}_{k_Q,j}^\Delta(1), \ldots, \tilde{c}_{k_Q,j}^\Delta(k_Q)]. \quad (20)$$

Using '0' to store a positive sign and '1' to store negative one, the signs are placed in the L strings, $st_{\text{sg}}^j, j = 1, \ldots, L$ as

$$st_{\text{sg}}^j = [\tilde{s}_{1,j}(1), \ldots, \tilde{s}_{1,j}(k_1), \cdots, \tilde{s}_{k_Q,j}(1), \ldots, \tilde{s}_{k_Q,j}(k_Q)]. \quad (21)$$

The next encoding/decoding scheme summarizes the above described procedure.

Encoding

- Given a partition $\mathbf{F}_q \in \mathbb{R}^{N_b \times L}, q = 1, \ldots, Q$ of a multichannel signal, where for each q the channels $\mathbf{f}_{q,j} \in \mathbb{R}^{N_b}$, $j = 1, \ldots, L$ are placed as columns of $\mathbf{F}_q$, approximate simultaneously all the channels through the HBW-OOMPMlTrgFFT approach using $K = \sum_{q=1}^{Q} k_q$ atoms to obtain:

$$\mathbf{f}_{q,j}^{k_q} = \sum_{n=1}^{k_q} c_{q,j}(n) \mathbf{d}_{\ell_n}, \quad j = 1, \ldots, L, \, q = 1 \ldots, Q. \quad (22)$$

- Quantize, as in (18), the absolute vale coefficients in the above equation to obtain $c_{q,j}^\Delta(n)$, $n = 1, \ldots, k_q$, $j = 1, \ldots, L$, $q = 1, \ldots, Q$.
- For each q, sort the indices $\ell_1^q, \ldots, \ell_{k_q}$ in ascending oder to have a new order $\tilde{\ell}_1^q, \ldots, \tilde{\ell}_{k_q}$ and the re-ordered sets $\tilde{s}_{q,j}(1), \ldots, \tilde{s}_{q,j}(k_q)$, and $\tilde{c}_{q,j}(1), \ldots, \tilde{c}_{q,j}(k_q)$, to create the strings: st_{ind}, as in (19), and st_{cf}^j, and st_{sg}^j, $j = 1, \ldots, L$ as in (20) and (21), respectively. All these strings are encoded, separately, using arithmetic coding.

Decoding

- Reverse the arithmetic coding to recover strings $st_{\text{ind}}, st_{\text{cf}}^j, st_{\text{sg}}^j, j = 1, \ldots, L$.
- Invert the quantization step as $|\tilde{c}_{q,j}^{\text{r}}(n)| = \Delta \tilde{c}_{q,j}^\Delta(n)$.
- Recover the partition of each channel through the liner combination

$$\mathbf{f}_{q,j}^{\text{r},k_q} = \sum_{n=1}^{k_q} \tilde{s}_{q,j}(n) |\tilde{c}_{q,j}^{\text{r}}(n)| \mathbf{d}_{\tilde{\ell}_n^q}.$$

- Assemble the recovered signal for each channel as

$$\mathbf{f}_j^{\text{r}} = \hat{\text{J}}_{q=1}^{Q} \mathbf{f}_{q,j}^{\text{r},k_q}, \; j = 1, \ldots, L$$

As already mentioned, the quality of the recovered signal is assessed by the SNR measure, which is calculated as

$$\text{SNR} = \log_{10} \frac{\|\mathbf{F}\|_F^2}{\|\mathbf{F} - \mathbf{F}^{\text{r}}\|_F^2} = 10 \log_{10} \frac{\sum_{j=1}^{L} \|\mathbf{f}_j\|^2}{\sum_{j=1}^{L} \|\mathbf{f}_j - \mathbf{f}_j^{\text{r}}\|^2}.$$

V. Numerical Example

This section is dedicated to illustrate the potential of the proposed codec for compressing melodic music with high quality recovery. The comparison with the state of the art is realized with respect to the OGG format [30]. The reasons being: a) OGG is free licence. b) It is known to recover a signal of audible quality comparable to MP3 (superior for some opinions) from a file of the same size. c)For a high quality setting (e.g. more than 90%) OGG produces a high SNR, which implies that the recovered signal is close to the original signal in the sense of the usual Euclidean distance.

The test clips, originally in WAV format and all sampled at 44100 Hz, are listed in Table I. All the clips are stereo, with two channels. The SNR, in all the cases, is fixed by setting the OGG quality 90%. For comparison purposes the Trigonometric Dictionary Codec (TDC) is tuned to reproduce the same SNR in each case. The partition unit is fixed as $N_b = 1024$ sample points. The approximation routine is set to produce a SNR a few dBs higher than the required one and the quantization parameter Δ is tuned to match the OGG's SNR. The file sizes are shown in Table I. The sizes corresponding to the TDC are obtained using the Arith06 MATLAB function at the entropy coding step. The function is available on [31]. Figure 1 shows the comparison bars in kbps (kilobit per sec) of the compression rate for the clips of Table I and the corresponding SNR values.

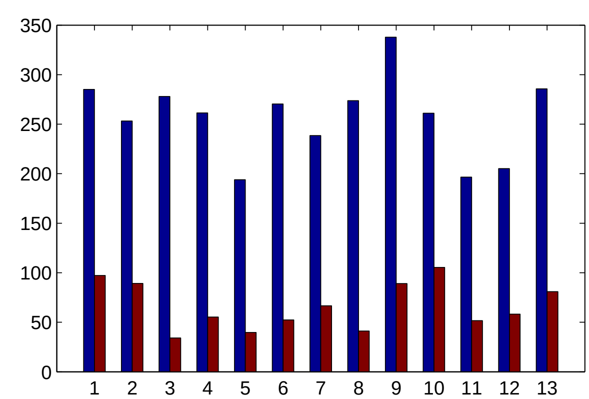

Fig. 1. Comparison bars, OGG vs TDC for the clips of Table 1. The vertical axis corresponds to the compression rate in kbps.

Clip	SNR	OGG	TDC
C1 Electric Guitar	32.10dB	212KB	72KB
C2 Harmonics Guitar	35.30dB	359KB	126KB
C3 Classic Guitar	35.84dB	723KB	89KB
C4 Pop Piano Chord	32.53dB	261KB	55KB
C5 Cathedral Organ	34.38dB	537KB	110KB
C6 Orchestra Horns	39.07dB	629KB	122KB
C7 Ascending Jazz	33.38dB	102KB	28KB
C8 Orchestrated	35.56dB	205KB	31KB
C9 Classic Orchestra	34.63dB	125KB	33KB
C10 Trumpet Sax	34.25dB	78KB	32KB
C11 Orchestra Entrance	34.21dB	74KB	19KB
C12 Piazzola (Orches.)	31.90dB	205KB	58KB
C13 Chopin (Piano)	35.44dB	414KB	58KB

TABLE I
COMPARISON OF FILE SIZES (IN KILOBYTES) COMPRESSING THE CLIPS TO PRODUCE THE SAME SNR. THE SNR VALUES ARISE BY SETTING 90% QUALITY FOR THE OGG COMPRESSION. MOST OF THE CLIPS ARE FROM `free-loops.com`. C3 AND C9 AND C13 ARE FROM SAMPLE WAV FILES ON `onclassical.com`

.

VI. CONCLUSIONS

The proof of concept of the proposed TDC has been presented. Comparisons with the OGG Vorbis standard at 90% quality demonstrate the potential of the proposed codec: For all the short clips of melodic music that have been tested (in addition to those in Table I) the TDC achieves remarkable reduction in the file size for the identical quality.

Because this work focusses on assessing compression vs quality, the pursuit technique which has been applied at the approximation stage aims at producing high sparsity. At the entropy coding step compression performance was prioritized over speed. In addition to Arith06, the arithmetic encoder Arith07 and Huffman encoder Huff06, implemented by the MATLAB functions available on [31], have been tested. All these entropy coding techniques produce similar outputs.

Note: The MATLAB functions for implementing the TDC and running the numerical examples of Sec. V, can be downloaded on [29].

Acknowledgements

The author is grateful to Karl Skretting for making available the Arith06, Arith07, and Huff06 functions for entropy coding [31].

REFERENCES

[1] M. Hans and R.W Schafer, "Lossless compression of digital audio" *IEEE Signal Processing Magazine,* **18**, 21–32 (2001).
[2] T. Liebchen and Y.A. Reznik, "MPEG-4 ALS: an Emerging Standard for Lossless Audio Coding," in Proceedings of the Data Compression Conference, 439–448, Snowbird, Utah, March 2004.
[3] Yu Rongshan, S. Rahardja, Xiao Lin, and Ko Chi Chung, "A fine granular scalable to lossless audio coder" *IEEE Transactions on Audio, Speech, and Language Processing,* **14**, 1352–1363 (2006).
[4] H. Huang, P. Franti, D. Huang, and S. Rahardja, "Cascaded RLS-LMS prediction in MPEG-4 lossless audio coding", *IEEE Transactions on Audio, Speech, and Language Processing*, **16**, 554–562 (2008).
[5] F. Ghido and I. Tabus, "Benchmarking of Compression and Speed Performance for Lossless Audio Compression Algorithms",in International Conference on Acoustics, Speech and Signal Processing, ICASSP (2008).
[6] N. Nowak and W. Zabierowski, "Methods of sound data compression Comparison of different standards", CAD Systems in Microelectronics (CADSM), 11th International Conference The Experience of Designing and Application of, 431–434 (2011).
[7] A. Avramovic, G. Banjac, and J. Galic, "Lossless audio compression using modular arithmetic and performance-based adaptation", Telecommunications Forum (TELFOR), 20th, 1256–1259 (2012).
[8] F. Ghido and I. Tabus, "Sparse Modeling for Lossless Audio Compression", IEEE Transactions on Audio, Speech, and Language Processing, **21**, 14–28 (2013).
[9] A. Lutoborski, and V. N. Temlyakov, "Vector greedy algorithms", *Journal of Complexity*, **19**, 458–473 (2003).
[10] D. Leviatan and V. N. Temlyakov, "Simultaneous approximation by greedy algorithms", *Advances in Computational Mathematics,* **25**, 73–90 (2006).
[11] A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit", *Signal Processing*, **86** 572–588 (2006).
[12] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse Solutions to Linear Inverse Problems with Multiple Measurement Vectors," *IEEE Transactions on Signal Processing*, **53** 2477–2488 (2005).
[13] Jie Chen, Xiaoming Huo, "Theoretical Results on Sparse Representations of Multiple-Measurement Vectors", *IEEE Transactions on Signal Processing,* **54**, 4634–4643 (2006).
[14] L. Rebollo-Neira, R. Matiol, S. Bibi, "Hierarchized block wise image approximation by greedy pursuit strategies," *IEEE Signal Process. Letters*, **20**, 1175–1178 (2013).
[15] L. Rebollo-Neira, "Cooperative Greedy Pursuit Strategies for Sparse Signal Representation by Partitioning", arXiv:1501.05971, 2015.
[16] M. D. Plumbley, S. A. Abdallah, T. Blumensath, M. E. Davies, "Sparse Representations of Polyphonic Music",*Signal Processing*, **86**, 417–431, (2006).
[17] R. M. Figueras i Ventura, P. Vandergheynst and P. Frossard, "Low rate and scalable image coding with redundant representations", *IEEE Transactions on Image Processing*, **15**, 726–739 (2006).
[18] J. Mairal, M. Eldar and G. Sapiro, "Sparse Representation for Color Image Restoration", *IEEE Trans. Image Proces.,* **17**, 53 – 69 (2008).
[19] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Yi Ma, "Robust Face Recognition via Sparse Representation" *IEEE Transactions on Pattern Analysis and Machine Intelligence*",**31**, 210 – 227 (2009).
[20] M. Elad, Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing, Springer (2010).
[21] J. Wright, Yi Ma, J. Mairal, G. Sapiro, T.S. Huang, and S. Yan, "Sparse Representation for Computer Vision and Pattern Recognition", *Proc. of the IEEE* **98**, 1031 – 1044 (2010).
[22] Jianchao Yang, I. L., Urbana, J. Wright, T.S. Huang, and Yi Ma, "Image Super-Resolution Via Sparse Representation" *IEEE Transactions on Image Processing*, **19**, 2861–287 (2010).
[23] M. D. Plumbley, T. Blumensath, L. Daudet, R. Gribonval, and M. E. Davies, "Sparse representations in audio and music: From coding to source separation," *Proc. IEEE,* **98,** 995–1005 (2010).
[24] O. Bryt and M. Elad, "Compression of facial images using the K-SVD algorithm" *Journal of Visual Communication and Image Representation*, **19** 270–282 (2008).
[25] J. Zepeda, C. Guillemot, E. Kijak, "Image Compression Using Sparse Representations and the Iteration-Tuned and Aligned Dictionary", *IEEE Journal of Selected Topics in Signal Processing,* **5**,1061–1073 (2011).
[26] I. Ram, I. Cohen, and M. Elad, Facial Image Compression using Patch-Ordering-Based Adaptive Wavelet Transform, *IEEE Signal Processing Letters,* **21**,
[27] L. Rebollo-Neira and D. Lowe, "Optimized orthogonal matching pursuit approach", *IEEE Signal Process. Letters*, **9**, 137–140 (2002).
[28] S. Chen, S. A. Billings, W. Luo, "Orthogonal least squares methods and their application to non-linear system identification". *International Journal of Control* (1989).
[29] `http://www.nonlinear-approx.info/examples/node03.html`
[30] `https://xiph.org/ogg`
[31] `http://www.ux.uis.no/~karlsk/proj99`