Cyclic Matching Pursuits with Multiscale Time-frequency Dictionaries

Sturm, Bob L.; Christensen, Mads Græsbøll

Published in:
Asilomar Conference on Signals, Systems and Computers. Conference Record

DOI (link to publication from Publisher):
10.1109/ACSSC.2010.5757627

Publication date:
2010

Document Version
Early version, also known as pre-print

Cyclic Matching Pursuits with Multiscale Time-frequency Dictionaries

Bob L. Sturm
Dept. Architecture, Design and Media Technology
Aalborg University Copenhagen
Lautrupvang 15, 2750 Ballerup, Denmark
(+45) 99 40 76 33, bst@create.aau.dk

Mads G. Christensen
Dept. of Architecture, Design and Media Technology
Aalborg University
Niels Jernes Vej 14, 3-216 DK-9220 Aalborg, Denmark
(+45) 99 40 97 93, mgc@create.aau.dk

***Abstract*—We generalize cyclic matching pursuit (CMP), propose an orthogonal variant, and examine their performance using multiscale time-frequency dictionaries in the sparse approximation of signals. Overall, we find that the cyclic approach of CMP produces signal models that have a much lower approximation error than existing greedy iterative descent methods such as matching pursuit (MP), and are competitive with models found using orthogonal MP (OMP), and orthogonal least squares (OLS). This implies that CMP is a strong alternative to the more computationally complex approaches of OMP and OLS for modeling high-dimensional signals.**

I. Introduction

Assume modeling a discrete signal $\mathbf{x} \in \mathbb{C}^K$ as $\mathbf{x} = \mathbf{D}\mathbf{s}$ where $\mathbf{D} = [\mathbf{d}_1|\mathbf{d}_2|\cdots|\mathbf{d}_N] \in \mathbb{C}^{K\times N}$ is a dictionary of N atoms that may be parametric, learned, or both. For sparse approximation, we seek a solution $\mathbf{s} \in \mathbb{C}^N$ that has a large concentration of energy in few of its elements, and that gives a small error relative to any noise in the signal. In these respects, we can consider the problem of sparse approximation as either minimizing the number of atoms in the model that has a maximum error $\epsilon \geq 0$

$$\min ||\mathbf{s}||_0 \text{ such that } ||\mathbf{x} - \mathbf{D}\mathbf{s}||_2^2 \leq \epsilon \tag{1}$$

where $||\mathbf{s}||_0$ is the number of non-zero elements in $\mathbf{s}$, or alternatively minimizing the error of the model using a maximum number of atoms $n > 0$

$$\min ||\mathbf{x} - \mathbf{D}\mathbf{s}||_2^2 \text{ such that } ||\mathbf{s}||_0 \leq n. \tag{2}$$

Another possibility, but harder to define precisely, is finding an efficient model that provides a "useful" analysis of the signal, for example, its time-frequency characteristics. To ensure there exists solutions to $\mathbf{x} = \mathbf{D}\mathbf{s}$, one typically uses a dictionary of size $N \gg K$ and with $\text{rank}(\mathbf{D}) = K$, which means that the dictionary is *overcomplete* for the inner product space $\mathbb{C}^K$. From the infinite number of solutions then, we want to find the ones most favorable with respect to the problems in (1) or (2), or in terms of its "descriptiveness" of the signal.

Many methods have been proposed and extensively studied to solve these three problems, either exactly or approximately, e.g., [1]–[11]. One approach is to replace the non-convex function in (1) or (2) with a convex one that favors sparse solutions, e.g., an ℓ_1-norm that is computationally tractable to solve using methods of convex optimization [3], [7]. These approaches, however, are currently too costly for high-dimensional signals, and large dictionaries. Greedy iterative descent methods provide another approach to solving the problems above for high-dimensional signals and large dictionaries. Matching pursuit (MP) [1] is the simplest of these, and can be efficiently implemented for large shift-invariant dictionaries and high-dimensional data, e.g., [12]. Other iterative approaches provide better signal decomposition in terms of reducing the approximation error, e.g., orthogonal MP (OMP) [2], [8], [10], and orthogonal least squares (OLS) [4], [5], but at a higher computational cost than MP. An extension to MP is given by cyclic MP (CMP) [6], where the simple strategy and low complexity of MP are preserved, but post-processing steps are taken to refine and improve the signal model. Originally, CMP was proposed within a framework of estimation for perceptual audio coding over parameteric models using a Fourier dictionary of single-scale discrete atoms with continuously distributed frequencies and complex amplitudes. However, the principle of CMP is applicable to any dictionary, including atoms that are time localized, as in the case with a multiscale time-frequency dictionary.

In this paper, we first review CMP in a single scale Fourier dictionary. Then we generalize it to using any dictionary, and propose an orthogonal extension. We also study the computational complexity of these methods. Finally, we compare the performance of these algorithms for modeling signals in multiscale time-frequency dictionaries.

II. Cyclic Matching Pursuit

At its core, CMP [6] is identical to MP, except it reassesses the parameters of all atoms of the model in a cyclic manner. It is in principle capable of "correcting" atoms that were selected in a biased way by a greedy pursuit [11]. In this section we review CMP in a continuous Fourier dictionary, generalize it to other dictionaries, and propose an orthogonal variation.

A. Pursuit in an Infinite Discrete Fourier Dictionary

Consider a K-dimensional signal $\mathbf{x} \in \mathbb{C}^K$ that we model as a linear combination of $L > 0$ elements drawn from

$$\mathcal{F}_K \triangleq \Big\{ \mathbf{d}(\omega) = \frac{1}{\sqrt{K}} e^{j\omega\mathbf{k}} : -\pi \leq \omega < \pi, \quad \mathbf{k} \triangleq [0, 1, \ldots, K-1]^T \Big\} \tag{3}$$

where $e^{j\omega\mathbf{k}} \triangleq [e^{j\omega 0}, e^{j\omega 1}, \ldots, e^{j\omega(K-1)}]^T$. We can exactly represent any finite norm $\mathbf{x} \in \mathbb{C}^K$ by at most K elements of $\mathcal{F}_K$ since $\text{span}\{\mathcal{F}_K\} = \mathbb{C}^K$ — for instance, using the discrete Fourier transform of the signal. However, we want to find $L \ll K$ elements of $\mathcal{F}_K$ such that the *total distortion* is minimized

$$\min_{\substack{\Theta\subset\Omega \\ |\Theta|=L}} J(\Theta) \triangleq \min_{\substack{\Theta\subset\Omega \\ |\Theta|=L}} \left|\left|\mathbf{x} - \sum_{i=0}^{L-1} \frac{1}{\sqrt{K}} a_i e^{j\omega_i \mathbf{k}}\right|\right| \tag{4}$$

where $\Omega \triangleq \{\theta = \{\omega, a\} \in [-\pi, \pi) \times \mathbb{C}\}$, and we define the norm here as the Euclidean norm. The notation $|\Theta|$ denotes the cardinality of the set Θ. This cannot be solved directly, and even for a dictionary $\mathcal{D}$ of finite size requires a combinatorial search, i.e., $\binom{|\mathcal{D}|}{L}$.

Instead of minimizing the total distortion in (4) over the entire dictionary, we can use MP and select one atom at a time such that the lth-atom is selected according to

$$\theta_l = \arg\min_{\theta\in\Omega} \left|\left|\mathbf{r}(l) - \frac{1}{\sqrt{K}} a e^{j\omega\mathbf{k}}\right|\right| \tag{5}$$

where the lth-order residual is defined

$$\mathbf{r}(l) \triangleq \mathbf{x} - \sum_{i=0}^{l-1} \frac{1}{\sqrt{K}} a_i e^{j\omega_i \mathbf{k}} \tag{6}$$

with the initialization $\mathbf{r}(0) \triangleq \mathbf{x}$. After L iterations we have the parametric signal model $\Theta_L = \{\theta_l : l = 0, \ldots, L-1\}$. However, if the atoms in $\mathcal{F}_K$ are not orthogonal, this signal model will not be the solution to (4). For instance, we can reduce the norm of (6) by recomputing the best weights given the L atoms by orthogonal projection. Furthermore, greedy atom selection of MP over a non-orthogonal dictionary, along with a lack of fit between signal structures and the dictionary, will produce a model that has artifacts of the MP decomposition process [11]. These artifacts contribute more to fixing the errors of previously selected atoms than modeling the signal.

CMP [6] iteratively decreases the intermediate distortion in (5) by refining the parameters of the model in a cyclic manner. In principle, it operates the same way as any cyclic minimizer [13]: CMP refines each parameter while holding the others fixed. CMP then augments the model and proceeds to refine all the parameters again. Considering that we already have a set of $0 < l \leq L$ parameters $\Theta_l^{(0)} = \{\theta_i = \{\omega_i, a_i\}\}_{i=0}^{l-1}$, which produces the lth-order model in (6), and thus a signal model with the distortion $J(\Theta_l^{(0)}) = ||\mathbf{r}(l)||$, we attempt to refine this model one parameter at a time.

In the first refinement step for the $(i+1)$th parameter ($0 \leq i \leq l$), we minimize the *intermediate distortion*

$$\theta_i^{(1)} = \arg\min_{\theta} D(\theta; \mathbf{r}_{l\setminus i}) \triangleq \arg\min_{\theta} \left|\left|\mathbf{r}_{l\setminus i} - \frac{a}{\sqrt{K}} e^{j\omega\mathbf{k}}\right|\right| \tag{7}$$

where we define the intermediate residual

$$\begin{aligned}\mathbf{r}_{l\setminus i} &\triangleq \mathbf{x} - \frac{1}{\sqrt{K}} \sum_{m=0}^{i-1} a_m^{(1)} e^{j\omega_m^{(1)}\mathbf{k}} - \frac{1}{\sqrt{K}} \sum_{m=i+1}^{l-1} a_m^{(0)} e^{j\omega_m^{(0)}\mathbf{k}} \\ &= \mathbf{r}_{l\setminus i-1} + \frac{1}{\sqrt{K}} a_i^{(0)} e^{j\omega_i^{(0)}\mathbf{k}}. \end{aligned} \tag{8}$$

with $\mathbf{r}_{l\setminus -1} \triangleq \mathbf{x}$. The refined parameters are given by

$$\omega_i^{(1)}(\mathbf{r}_{l\setminus i}) = \arg\max_{\omega} \left|\left\langle \mathbf{r}_{l\setminus i}, e^{j\omega\mathbf{k}} \right\rangle\right| \tag{9}$$

$$a_i(\omega_i^{(1)}) = \frac{1}{\sqrt{K}} \left\langle \mathbf{r}_{l\setminus i}, e^{j\omega_i^{(1)}\mathbf{k}} \right\rangle. \tag{10}$$

The problem in (9) is equivalent to finding the frequency of the maximum DTFT magnitude (or two if the signal is real). With these new parameters, we replace those of the $(i+1)$th atom, update the intermediate residual, and refine the parameters of the next atom.

Since we are minimizing the distortion (7), the worst we can do in each refinement step is reselect the same atom, and thus we are guaranteed that

$$\cdots \leq D\left(\theta_i^{(2)}; \mathbf{r}_{l\setminus i}\right) \leq D\left(\theta_i^{(1)}; \mathbf{r}_{l\setminus i}\right) \leq D\left(\theta_i^{(0)}; \mathbf{r}_{l\setminus i}\right) \tag{11}$$

in which it is implicit that the residual $\mathbf{r}_{l\setminus i}$ is changing with each refinement. In the first refinement cycle, we refine $\Theta_l^{(0)}$ into $\Theta_l^{(1)}$, and we can perform additional cycles of refinement. As a result of (11) then, we are guaranteed that the total distortion (4) will not increase with each refinement cycle

$$\ldots \leq J\left(\Theta_l^{(2)}\right) \leq J\left(\Theta_l^{(1)}\right) \leq J\left(\Theta_l^{(0)}\right). \tag{12}$$

The second step of CMP is the augmentation of the model by a new atom from $\mathcal{F}_K$. In this case, we simply add to the set of parameters Θ_l those of a new atom as in (5)

$$\begin{aligned}\theta_{l+1}^{(0)} &= \arg\min_{\theta} D\left(\theta; \mathbf{r}(l+1)\right) \\ &\triangleq \arg\min_{\theta} \left|\left|\mathbf{r}(l+1) - \frac{1}{\sqrt{K}} a e^{j\omega\mathbf{k}}\right|\right| \end{aligned} \tag{13}$$

which is solved using (9) and (10) with $\mathbf{r}(l+1)$ defined like in (6). After this, we refine the parameter set $\Theta_{l+1}^{(0)}$, augment again, and repeat until we have arrived at the refined parameters of an Lth-order model. Since the model augmentation guarantees a nonincreasing total distortion (4), we can see

$$J\left(\Theta_L^{(*)}\right) \leq \ldots \leq J\left(\Theta_l^{(*)}\right) \leq \ldots \leq J\left(\Theta_1^{(*)}\right) \tag{14}$$

no matter the number of refinement cycles at each order.

B. Pursuit in Any Dictionary

CMP in [6] is motivated by the need for parametric models of audio signals producing low perceptual distortion, and as such does not directly lend itself to other dictionaries, such as a multiresolution Gabor dictionary. In this subsection, we generalize CMP to sparse approximation using any dictionary. As before, we consider a K-dimensional signal $\mathbf{x} \in \mathbb{C}^K$ that we model as a linear combination of $0 < L \ll K$ elements drawn from $\mathcal{D} \triangleq \{\mathbf{d}_i \in \mathbb{C}^K : ||\mathbf{d}_i|| = 1\}$ such that

$$\min_{\substack{\mathcal{H}\subset\mathcal{D} \\ |\mathcal{H}|=L}} J(\mathcal{H}) = \min_{\substack{\mathcal{H}\subset\mathcal{D} \\ |\mathcal{H}|=L}} \left|\left|\mathbf{x} - \sum_{i=0}^{L-1} a_i \mathbf{h}_i\right|\right| \tag{15}$$

where $\mathbf{h}_i \in \mathcal{H} \subset \mathcal{D}$. As before, solving this problem has combinatorial complexity so we approach it as done in MP, but with the refinement apparatus of CMP.

Consider that we have already found l atoms and weights, giving a l-order *representation* $\mathcal{X}_{\text{MP},l} = \{\mathbf{H}(l), \mathbf{a}(l), \mathbf{r}(l)\}$ where $\mathbf{x} = \mathbf{H}(l)\mathbf{a}(l) + \mathbf{r}(l)$, $\mathbf{H}(l) = [\mathbf{h}_0|\mathbf{h}_1|\cdots|\mathbf{h}_{l-1}]$, and $\mathbf{a}(l) \in \mathbb{C}^l$ is a vector of weights. We augment this representation by the rule

$$\mathcal{X}_{\text{MP},l+1} = \left\{ \begin{array}{l} \mathbf{H}(l+1) = [\mathbf{H}(l)|\mathbf{h}_l], \\ \mathbf{a}(l+1) = \left[\mathbf{a}^T(l), \langle \mathbf{r}(l), \mathbf{h}_l \rangle\right]^T, \\ \mathbf{r}(l+1) = \mathbf{x} - \mathbf{H}(l+1)\mathbf{a}(l+1) \end{array} \right\} \quad (16)$$

using the atom selection criterion

$$\mathbf{h}_l = \arg\min_{\mathbf{d}\in\mathcal{D}} ||\mathbf{r}(l) - \langle \mathbf{r}(l), \mathbf{d}\rangle \mathbf{d}|| = \arg\max_{\mathbf{d}\in\mathcal{D}} |\langle \mathbf{r}(l), \mathbf{d}\rangle| \quad (17)$$

which guarantees that the residual norm is non-increasing.

With this augmented model now, we refine the atom choices in the following way. In the first refinement step for the $(i+1)$th atom ($0 \leq i \leq l$) of the representation basis $\mathbf{h}_i \in \mathcal{H}_l^{(0)}$, we minimize the intermediate distortion

$$\mathbf{h}_i^{(1)} = \arg\min_{\mathbf{d}\in\mathcal{D}} D(\mathbf{d}; \mathbf{r}_{l\backslash i}) \stackrel{\Delta}{=} \arg\min_{\mathbf{d}\in\mathcal{D}} \left|\left|\mathbf{r}_{l\backslash i} - \langle \mathbf{r}_{l\backslash i}, \mathbf{d}\rangle \mathbf{d}\right|\right| \quad (18)$$

where the intermediate residual is as in (8)

$$\mathbf{r}_{l\backslash i} \stackrel{\Delta}{=} \mathbf{x} - \sum_{m=0}^{i-1} a_m^{(1)} \mathbf{h}_m^{(1)} - \sum_{m=i+1}^{l} a_m^{(0)} \mathbf{h}_m^{(0)} \quad (19)$$

$$= \mathbf{r}_{l\backslash i-1} + a_i^{(0)} \mathbf{h}_i^{(0)}. \quad (20)$$

With this procedure, we have refined the atoms in $\mathcal{H}_l^{(0)}$ into $\mathcal{H}_l^{(1)}$. As before, the worst we can do in a refinement step is to reselect the same atom, and thus we are guaranteed that the intermediate distortion will not increase

$$\ldots \leq D\left(\mathbf{h}_i^{(2)}; \mathbf{r}_{l\backslash i}\right) \leq D\left(\mathbf{h}_i^{(1)}; \mathbf{r}_{l\backslash i}\right) \leq D\left(\mathbf{h}_i^{(0)}; \mathbf{r}_{l\backslash i}\right) \quad (21)$$

for any number of refinements (it is implicit that the residual $\mathbf{r}_{l\backslash i}$ is changing with each refinement). From this fact, we know the total distortion will not increase as in (14), and so we can thus refine the $(l+1)$th-order model until some stopping criterion is met, e.g., when for some threshold $\eta \geq 0$

$$J\left(\mathcal{H}_l^{(m)}\right) \geq (1-\eta) J\left(\mathcal{H}_l^{(m-1)}\right). \quad (22)$$

C. An Orthogonal Variation in Any Dictionary

It appears to be a simple matter to integrate this refinement process into OMP, where the atom selection is the same as in MP (17), but the representation update rule is

$$\mathcal{X}_{\text{OMP},l+1} = \left\{ \begin{array}{l} \mathbf{H}(l+1) = [\mathbf{H}(l)|\mathbf{h}_l], \\ \mathbf{a}(l+1) = \mathbf{H}^\dagger(l+1)\mathbf{x}, \\ \mathbf{r}(l+1) = \mathbf{x} - \mathbf{H}(l+1)\mathbf{a}(l+1) \end{array} \right\} \quad (23)$$

where $\mathbf{H}^\dagger(l+1) = \left[\mathbf{H}^H(l+1)\mathbf{H}(l+1)\right]^{-1} \mathbf{H}^H(l+1)$, and we know the matrix $\mathbf{H}^H(l+1)\mathbf{H}(l+1)$ is non-singular because $\mathbf{H}(l+1)$ is full rank by construction. Updating the model weights in this way guarantees that $\mathbf{H}^H(l)\mathbf{r}(l) = \mathbf{0}$, i.e., that the signal approximation $\mathbf{H}(l+1)\mathbf{a}(l+1)$ is the orthogonal projection of $\mathbf{x}$ onto the column space of $\mathbf{H}(l+1)$. For a model built by OMP, however, refining each atom by the criterion in (18) does not guarantee (21).

To see why, consider that we refine each new atom by using (18), but with the orthogonalized residual

$$\mathbf{r}_{l\backslash i} \stackrel{\Delta}{=} \left[\mathbf{I} - \mathbf{P}_{l\backslash i}(\mathbf{P}_{l\backslash i}^H \mathbf{P}_{l\backslash i})^{-1} \mathbf{P}_{l\backslash i}^H\right] \mathbf{x}$$
$$= \mathbf{r}_{l\backslash i-1} + (\mathbf{P}_{l\backslash i-1}^H \mathbf{P}_{l\backslash i-1})^{-1} \mathbf{P}_{l\backslash i-1}^H \mathbf{h}_i \quad (24)$$

where we define $\mathbf{P}_{l\backslash i} \stackrel{\Delta}{=} [\mathbf{H}'(i)|\mathbf{H}(l\backslash i)]$ constructed by the $(i-1)$ new atoms as the columns of $\mathbf{H}'(i)$, and the $(l-i+1)$ atoms remaining to be refined as the columns of $\mathbf{H}(l\backslash i)$. The matrix $\mathbf{P}_{l\backslash i}$ is full rank since both submatrices have linearly independent columns by construction. Now, the atom refinement criterion (18) will replace the ith atom with one that points the most along $\mathbf{r}_{l\backslash i}$ oblivious to the fact that we will recompute all weights by orthogonally projecting $\mathbf{x}$ onto the span of the refined representation basis. This does not guarantee (21), i.e., that the intermediate residual norm will decrease with each atom replacement.

Instead, we must replace each atom with the one that points the most along the residual projected onto the left null space of $\mathbf{P}_{l\backslash i}$, which is exactly what is done in OLS [4], [5]. Thus, the first replacement of the ith atom is given by

$$\mathbf{h}_i^{(1)} \stackrel{\Delta}{=} \arg\min_{\mathbf{d}\in\mathcal{D}} \left|\left|\mathbf{r}_{l\backslash i} - \frac{\langle \mathbf{r}_{l\backslash i}, \mathbf{d}\rangle}{||\mathbf{d}_{\mathcal{P}_{l\backslash i}^\perp}||^2} \mathbf{d}_{\mathcal{P}_{l\backslash i}^\perp}\right|\right| \quad (25)$$

where

$$\mathbf{d}_{\mathcal{P}_{l\backslash i}^\perp} \stackrel{\Delta}{=} \left[\mathbf{I} - \mathbf{P}_{l\backslash i}(\mathbf{P}_{l\backslash i}^H \mathbf{P}_{l\backslash i})^{-1} \mathbf{P}_{l\backslash i}^H\right] \mathbf{d}. \quad (26)$$

With this atom refinement used to create each $\mathcal{X}_{\text{OLS},l}$ using the same update rule as OMP (23), we are now guaranteed (21), and that $J\left(\mathcal{H}_L^{(*)}\right) \leq \ldots \leq J\left(\mathcal{H}_l^{(*)}\right) \leq \ldots \leq J\left(\mathcal{H}_1^{(*)}\right)$. We call this algorithm cyclic OLS (COLS).

D. Computational Complexity

The computational complexity of these approaches vary. The complexity of MP is $\mathcal{O}(LK \log K)$ [12] to create an L-order model for a length-K real signal over a multiscale complex time-frequency dictionary. This means that for CMP running a maximum of $R \geq 0$ refinement cycles for each iteration, the worst-case complexity is

$$\mathcal{O}\left(K \log K \left[1 + R \sum_{l=2}^{L} l\right]\right)$$
$$= \mathcal{O}\Big(K \log K \big[1 + R\left(L[L+1]/2 - 1\right)\big]\Big) \quad (27)$$

which is quadratic in the model order. By way of comparison, the computational complexity for OMP to create an L-order model of the same signal with the same dictionary, is

$$\mathcal{O}\left(LK \log K + \sum_{l=2}^{L} l^3 + Kl^2 + Kl\right) \quad (28)$$

assuming the cost of inverting a complex matrix of rank $2l$ is $\mathcal{O}(l^3)$ (our matrix includes atoms and their conjugates). This cost is cubic in the model order.

The computational cost of OLS is much higher than that of MP. To find each lth-order projection matrix requires

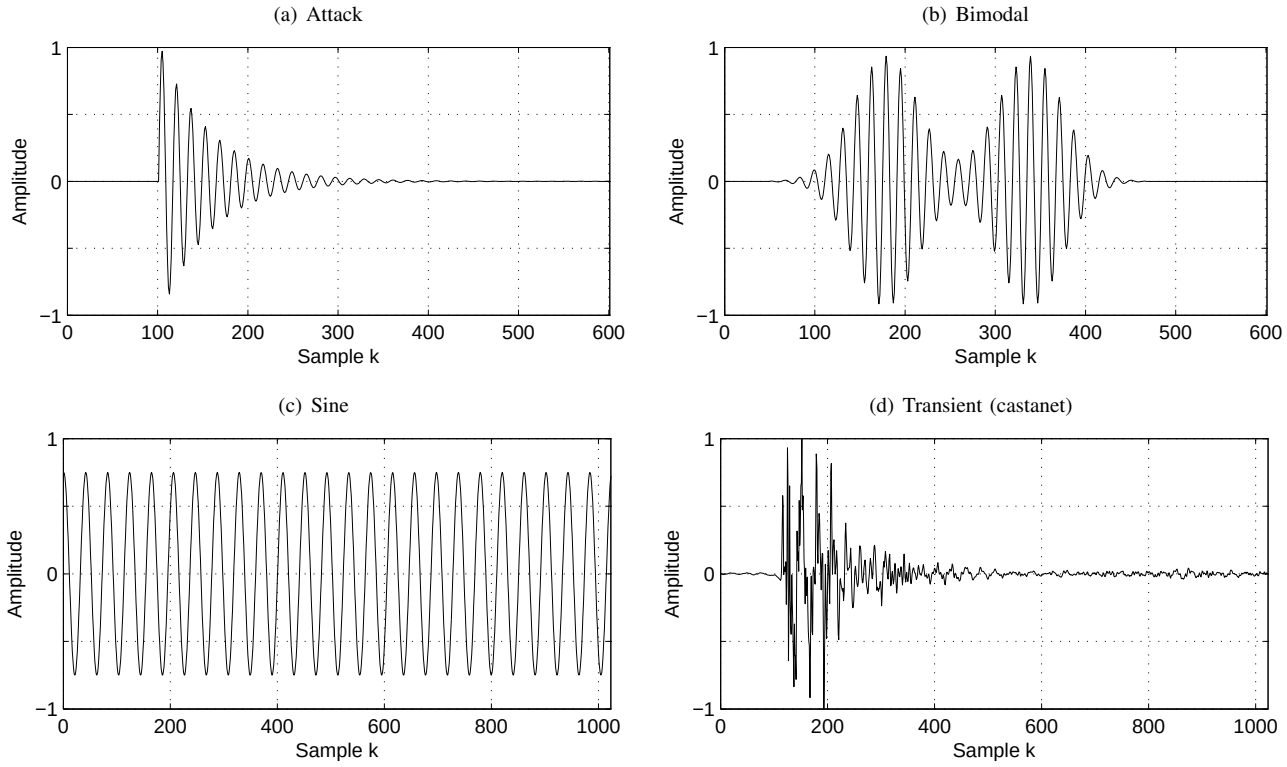

Fig. 1. Four time-domain signals used in our experiments.

$C(l) \triangleq \mathcal{O}(l^3 + l^2K + K^2l)$. Each atom of the dictionary must be projected onto the orthogonal subspace to the lth-order representation basis, which has complexity $\mathcal{O}(NK^2)$. Then, finding the best atom has the same complexity as MP. Each refinement cycle at the lth-order has a complexity $l[C(l-1)+NK^2+K\log K]$. Thus, for COLS running $R \geq 0$ refinement cycles each iteration, the worst-case complexity is

$$\mathcal{O}\left(\sum_{l=2}^{L} C(l) + lRC(l-1) + lR(K\log K + NK^2)\right) \quad (29)$$

which is quartic in the model order. This situation can be ameliorated by updating each projection matrix iteratively using the matrix inversion lemma. Since we are updating each representation basis by concatenating a new column in (23)

$$\begin{aligned}[\mathbf{H}^H(l+1)\mathbf{H}(l+1)]^{-1} &= \left[[\mathbf{H}(l)|\mathbf{h}_l]^H[\mathbf{H}(l)|\mathbf{h}_l]\right]^{-1} \\ &= \begin{bmatrix} \mathbf{H}^H(l)\mathbf{H}(l) & \mathbf{H}^H(l)\mathbf{h}_l \\ \mathbf{h}_l^H\mathbf{H}(l) & 1 \end{bmatrix}^{-1}. \end{aligned} \quad (30)$$

Defining $\mathbf{G}_l \triangleq \mathbf{H}^H(l)\mathbf{H}(l)$, $\mathbf{g}_l \triangleq \mathbf{H}^H(l)\mathbf{h}_l$, by the matrix inversion lemma we know

$$[\mathbf{H}^H(l+1)\mathbf{H}(l+1)]^{-1} = \begin{bmatrix} [\mathbf{I} + b\mathbf{G}_l^{-1}\mathbf{g}_l\mathbf{g}_l^H]\mathbf{G}_l^{-1} & -\mathbf{G}_l^{-1}\mathbf{g}_l/b \\ -b\mathbf{g}_l^H\mathbf{G}_l^{-1} & 1/b \end{bmatrix} \quad (31)$$

where $b \triangleq 1 - \mathbf{g}_l^H\mathbf{G}_l^{-1}\mathbf{g}_l$. This reduces the computational complexity of the each matrix inversion to be quadratic in l. We can apply a similar argument to computing the projection matrix of each matrix with a column removed, which is quadratic in $(l-1)$. Thus, the complexity of COLS reduces to be cubic in the model order

$$\mathcal{O}\left(\sum_{l=2}^{L} l^2 + lR(l-1)^2 + lR(K\log K + NK^2)\right). \quad (32)$$

We can also take into consideration that each $\mathbf{g}_l$ might be found using tabulated inner products of each pair of atoms. Or we can use a closed form solution if it exists. Also, much of the outer product $\mathbf{g}_l\mathbf{g}_l^H$ will be zeros when the dictionary consists of time-localized atoms.

III. Experimental Results

The residual decays of the four signals seen in Fig. 1 using five different greedy sparse approximation algorithms is shown in Fig. 2: MP and OMP [9], LoCOMP [10], OLS, CMP, and COLS, using a dictionary that is a union of Dirac functions (scale and translations of 1 sample), and complex Gabor atoms. The scale/hop in samples of each Gabor atom in the dictionary is $\{4/2, 8/2, 16/4, 32/8, 64/16, 128/32, 256/64, 512/128\}$. The modulation frequencies possible for each Gabor atom of scale s samples are $\omega \in \mathbb{Z}_{[0,s/2]}\pi/s$. We find the optimal complex amplitude of each real atom by maximizing the projection of the residual onto the best complex atom and its conjugate [14]. We do not add any noise to any of these signals. We run the refinement cycles of CMP up to either $R = 10$ times, or until the difference in total distortion between a cycle is less than 0.1%, i.e., $\eta = 0.001$ in (22). It is clear from Fig. 2 that for these four signals CMP results in a residual energy decay that is competitive with the computationally more complex OMP, and OLS algorithms. CMP produces models with residual energy decays superior to OMP and sometimes OLS. COLS performs the best in all these tests, but has the highest computational cost.

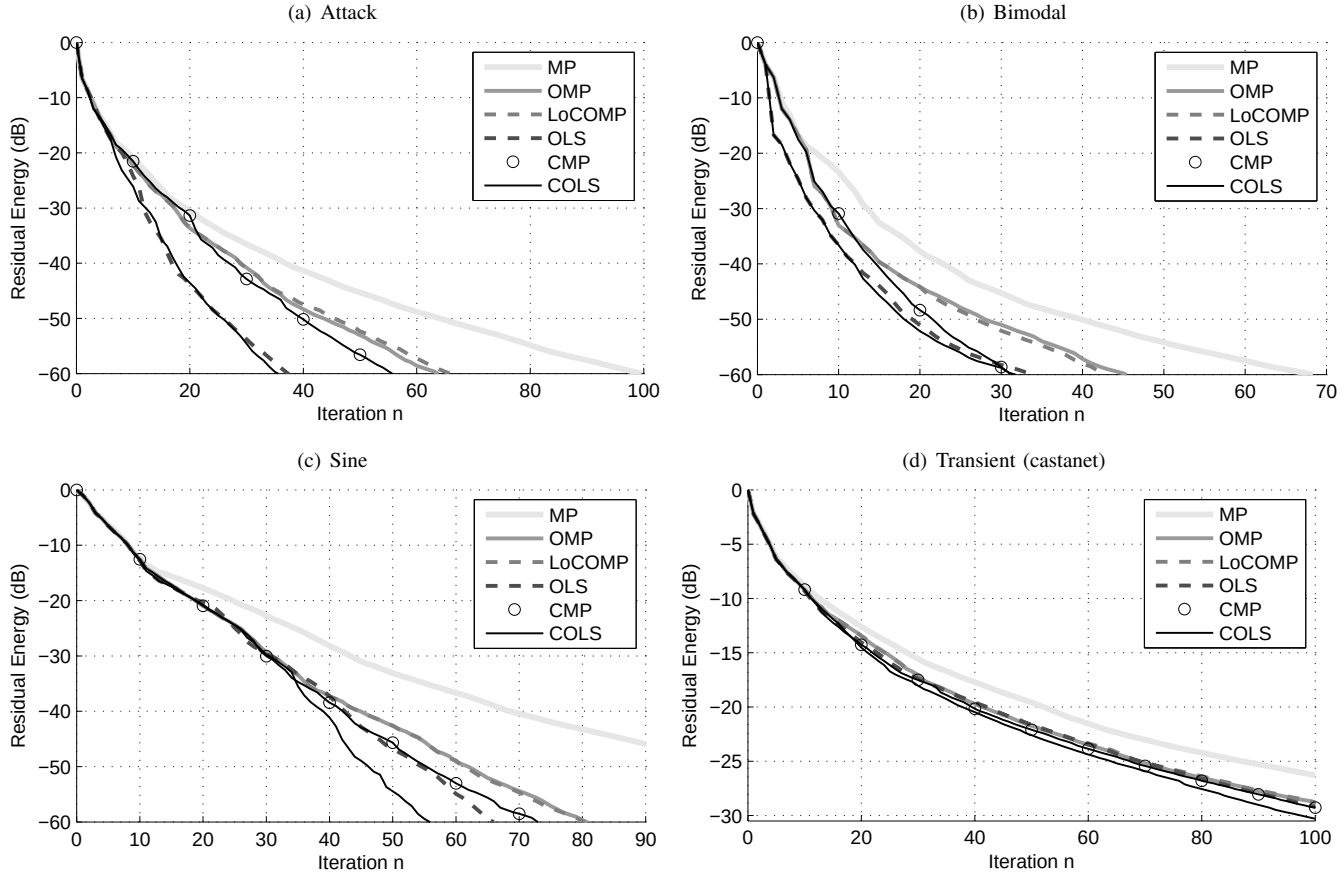

Fig. 2. Residual energy decay of each signal in Fig. 1 for six different greedy decompositions algorithms. MP: Matching Pursuit; OMP: Orthogonal MP; LoCOMP: Low-complexity OMP; OLS: Orthogonal Least-squares; CMP: Cyclic MP; COLS: Cyclic OLS. The dictionary is a union of Gabor atoms and Dirac spikes. Note the differences in the axes.

IV. Conclusion

We have explored CMP using any dictionary, and have proposed the orthogonal variant COLS. We found in our tests that CMP produces models with residual energy decays superior to OMP and sometimes even OLS. Thus, we can take a simple greedy method like MP and improve its model error by cyclicly replacing the atoms. Though CMP increases the computational cost of MP, it does so only as a very simple augmentation of the basic algorithm, thus making it amenable to fast implementations of MP [12]. Our tests also showed that COLS performs better than all other greedy iterative descent approaches with respect to the residual energy. Its computational complexity is high, however. Our future work involves implementing CMP within the MPTK framework [12], and studying its performance in creating models of real high-dimensional audio signals using large time-frequency dictionaries. We are also studying the effects of the cyclic refinement process on the convergence properties of MP.

References

[1] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[2] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1993, vol. 1, pp. 40–44.

[3] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, Aug. 1998.

[4] L. Rebollo-Neira and D. Lowe, "Optimized orthogonal matching pursuit approach," *IEEE Signal Process. Lett.*, vol. 9, no. 4, pp. 137–140, Apr. 2002.

[5] T. Blumensath and M. E. Davies, "On the difference between orthogonal matching pursuit and orthogonal least squares," Tech. Rep., University of Edinburgh, Edinburgh, Scotland, U.K., Mar. 2007.

[6] M. G. Christensen and S. H. Jensen, "The cyclic matching pursuit and its application to audio modeling and coding," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 2007.

[7] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Analysis Applications*, vol. 14, no. 5-6, pp. 877–905, 2008.

[8] T. Blumensath and M. E. Davies, "Gradient pursuits," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2370–2382, June 2008.

[9] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Academic Press, Elsevier, Amsterdam, 3rd edition, 2009.

[10] B. Mailhé, R. Gribonval, F. Bimbot, and P. Vandergheynst, "A low complexity orthogonal matching pursuit for sparse signal approximation with shift-invariant dictionaries," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, Taipei, Taiwan, Apr. 2009, pp. 3445–3448.

[11] B. L. Sturm and J. J. Shynk, "Sparse approximation and the pursuit of meaningful signal models with interference adaptation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 3, pp. 461–472, Mar. 2010.

[12] S. Krstulovic and R. Gribonval, "MPTK: Matching pursuit made tractable," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Toulouse, France, Apr. 2006, vol. 3, pp. 496–499.

[13] P. Stoica and Y. Selén, "Cyclic minimizers, majorization techniques, and the expectation-maximization algorithm: A refresher," *IEEE Sig. Process. Mag.*, vol. 21, no. 1, pp. 112–114, Jan. 2004.

[14] R. Gribonval, *Approximations Non-linéaires pour l'Analyse des Signaux Sonores*, Ph.D. thesis, Université de Paris IX Daupine, Paris, France, Sep. 1999.