STRUCTURED SPARSITY USING BACKWARDS ELIMINATION FOR AUTOMATIC MUSIC TRANSCRIPTION

Nicolas Keriven[*]

CMAP
Ecole polytechnique
Route de Saclay, Palaiseau, France

Ken O'Hanlon, Mark D. Plumbley[†]

Centre for Digital Music
Queen Mary University of London
Mile End Road, London, UK

ABSTRACT

Musical signals can be thought of as being sparse and structured, with few elements active at a given instant and temporal continuity of active elements observed. Greedy algorithms such as Orthogonal Matching Pursuit (OMP), and structured variants, have previously been proposed for Automatic Music Transcription (AMT), however some problems have been noted. Hence, we propose the use of a backwards elimination strategy in order to perform sparse decompositions for AMT, in particular with a proposed alternative sparse cost function. However, the main advantage of this approach is the ease with which structure can be incorporated. The use of group sparsity is shown to give increased AMT performance, while a molecular method incorporating onset information is seen to provide further improvements with little computational effort.

Index Terms— structured sparsity, music transcription, backwards elimination, group sparsity

1. INTRODUCTION

Given a signal $\mathbf{s} \in \mathbb{R}^M$ and a dictionary matrix $\mathbf{D} \in \mathbb{R}^{M\times N}$, sparse approximation seeks a coefficient vector, $\mathbf{x}$, with few active elements such that $\mathbf{s} \approx \mathbf{D}\mathbf{x}$. Ideally, this is performed by finding $\mathbf{x}$ that minimises the sparse cost function

$$\mathcal{C}_{sparse} = \|\mathbf{s} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_0 \tag{1}$$

where $\|\mathbf{x}\|_0 = |\mathbf{x} \neq 0|$. Finding a general solution to the minimisation of (1) is NP-hard [1] and a ℓ_1 norm penalty is used to effect a convex relaxation known as Basis Pursuit Denoising [1] or Lasso [2]. A popular alternative approach is to use greedy methods such as Orthogonal Matching Pursuit (OMP) [3], which build up an approximation by iteratively adding the atom most correlated with the residual energy. Greedy methods can suffer when presented with dictionaries containing atoms that are correlated, and several algorithms have recently been proposed which include backtracking steps [4].

Often in a signal representation, it can be assumed that certain atoms will be active together, and structured sparse methods allow these assumptions to be incorporated. Group sparsity assumes that certain atoms tend to be active in the same coefficient vector, and algorithms such as Group Lasso [5] and Block-OMP [6] are derived from the sparse methodology to solve this problem. Multichannel, or simultaneous, sparsity [7] considers that a similar atom is active in many sensors at once. Molecular sparse representations [8] consider structure from related coefficient vectors such as neighbouring time frames in an audio spectrogram.

Automatic Music Transcription (AMT) seeks machine understanding of a musical signal in terms of pitch-time activity. Spectrogram decompositions are a popular approach for AMT in which the approximation $\mathbf{S} \approx \mathbf{DX}$ is sought where $\mathbf{S} \in \mathbb{R}^{M\times T}$ is the spectrogram, $\mathbf{D} \in \mathbb{R}^{M\times N}$ is a dictionary and $\mathbf{X} \in \mathbb{R}^{N\times T}$ is the activation matrix. While spectrograms may be decomposed in an unsupervised manner using Non-negative Matrix Factorisation (NMF) methods [9] [10], superior AMT results are seen with a supervised NMF approach when a fixed pitch-labelled dictionary is used [10] [11]. Thresholding is often used to ascertain the final binary output for AMT, and a typical strategy is to adapt the threshold to the maximum value of the activation matrix, $\mathbf{X}$, [10]. Greedy sparse algorithms have also been used for AMT. Leveau et al [12] propose a modified molecular Matching Pursuit using a tracking step to decompose a spectrogram with a dictionary of pitch/instrument labeled atoms. Tjoa et al [13] propose the use of OMP with large overcomplete dictionaries consisting of datapoints, and we proposed the use of group and molecular variants of OMP [14].

Hence the use of backwards elimination with structured sparsity for AMT is explored in the rest of this paper. Some background material is outlined in the next section, before the proposed methodology is introduced. A modified sparse cost function, group and molecular sparse approaches, and an alternative thresholding strategy are proposed. Experiments are then described before concluding with pointers to further work.

[*] Nicolas Keriven was visiting at Queen Mary

[†] This research is supported by ESPRC Leadership Fellowship EP/G007144/1

978-1-4799-1180-6/13/$31.00 ©2013 IEEE

2. BACKGROUND

2.1. Non-Negative Least Squares

Non-Negative Least Squares (NNLS) is a well studied constrained least squares problem:

$$\mathbf{x} \leftarrow \min_{x} \|\mathbf{s} - \mathbf{D}\mathbf{x}\|_2^2 \quad s.t. \quad \mathbf{x} \geq 0. \tag{2}$$

The classic NNLS [15] is an active set algorithm that proceeds by adding an atom and performing a least squares backprojection at each iteration. Atoms in the active set with negative backprojection coefficients are ejected. Other active set [16], and gradient-based approaches have been proposed. and thresholded NNLS (T-NNLS) is seen to outperform ℓ_1 minimisation for sparse non-negative representations [17] .

2.2. Backwards Elimination

Backwards elimination is a stepwise approach which starts from a set, Γ, containing many atom indices, and iteratively eliminates an atom with index $\hat{n}$ so that $\Gamma \leftarrow \Gamma \backslash \hat{n}$ where

$$\hat{n} = \arg\min_{n} \Delta\mathbf{r}^n \tag{3}$$

and

$$\Delta\mathbf{r}^n = \|\bar{\mathbf{r}}^n\|_2^2 - \|\mathbf{r}^i\|_2^2 \tag{4}$$

where $\bar{\mathbf{r}}^n$ is the residual given the hypothetical sparse support $\Gamma^n = \Gamma \backslash n$, and $\mathbf{r}^i$ is the residual at the ith iteration. A fast version of the backwards elimination step is proposed as part of the Greedy Sparse Least Squares (GSLS) algorithm [18] :

$$\Delta\mathbf{r}^n = \frac{\mathbf{x}_n^2}{[(\mathbf{D}_\Gamma^T \mathbf{D}_\Gamma)^{-1}]_{n,n}}. \tag{5}$$

where $\mathbf{x}$ is the Least Squares solution vector using the current support, Γ. The elimination criteria (5) is derived using block matrix inverse updates, and can be calculated for all active atoms simultaneously using matrix / vector operations.

2.3. Structured Sparsity

Group sparsity considers that certain atoms tend be active together in the same coefficient vector. Using the set of tuples $\mathcal{L} = \{\mathcal{L}^l\}$ where $\mathcal{L}^l$ contains the indices of the lth group gives the following notation for the lth group of the dictionary, $\mathbf{D}[l]$, and of the coefficient vector, $\mathbf{x}[l]$:

$$\begin{aligned} \mathbf{D}[l] &= [\mathbf{d}_{\mathcal{L}^l(1)}, ..., \mathbf{d}_{\mathcal{L}^l(|\mathcal{L}^l|)}] \\ \mathbf{x}[l] &= [x_{\mathcal{L}^l(1)}, ..., x_{\mathcal{L}^l(|\mathcal{L}^l|)}]^T \end{aligned}$$

where $\mathcal{L}^l(i)$ is the ith member of the lth tuple of the set of tuples $\mathcal{L}$ and $\sum_l |\mathcal{L}^l| = N$.

Algorithms for solving the group sparse problem are derived from the general sparse representations methodology. Greedy methods such as Block-OMP [6] use selection criteria considering all atoms in a group. Group Lasso [5] replaces the ℓ_1-norm penalty term [2] [1] with a mixed $\ell_{p,q}$ vector norm :

$$\|\mathbf{x}\|_{p,q} = \|\mathbf{g}\|_q \tag{6}$$

where $\mathbf{g}_l = \|\mathbf{x}[l]\|_p$. Different combinations of (p, q) are used depending on the desired properties of the decomposition. When few groups are to be active, the $\ell_{2,0}$ is optimal, while the $\ell_{2,1}$-norm relaxation is used [6]. An alternative group sparse penalty, employing the subspace projection norm as the group coeffcient, $\mathbf{g}_l = \|\mathbf{D}[l]\mathbf{x}[l]\|_2$ was proposed in [19]. This approach is referred to here as the $\ell_{\perp,0}$-norm.

Simultaneous sparsity considers that a similar atom is active in several channels at once, using algorithms similar to those for group sparsity to decompose several signals simultaneously. Molecular sparsity [8] considers similar atoms existing in different vectors, such as adjacent time frames of a spectrogram that may be correlated, while the relationships may not be explicitly known. Typically molecular algorithms are greedy, and different approaches to molecular clustering can be used. For instance, a tracking approach whereby a molecule is formed by selecting one atom and then selecting adjacent atoms is used in [8] and [12], while an agglomerative clustering approach is used in [20] and [14].

3. METHOD

3.1. Modified sparse cost function

It can be seen (5) that the backwards elimination cost for an atom is scaled to the *square* of the least squares solution coefficients. Further to this it has been observed that NNLS and NMF coefficients scale well, in terms of a thresholding parameter, relative to the use of varying spectrogram transforms. In light of this, a modified ℓ_0 sparse cost function is proposed:

$$\mathcal{C}_{mod} = \|\mathbf{s} - \mathbf{D}\mathbf{x}\|_2 + \lambda\|\mathbf{x}\|_0 \tag{7}$$

differing from the standard sparse cost function (1) through using the residual error norm instead of the least squares error.

It is easily perceived how (3), the backwards elimination criteria performs a local optimisation of $\mathcal{C}_{sparse}$ (1) when $\Delta\mathbf{r}^{\hat{n}} < \lambda$, which can be used as a stopping condition for a backwards elimination strategy where λ represents a threshold. In terms of using the backwards elimination strategy with $\mathcal{C}_{mod}$ (7), the original elimination criteria (3) can be used, affording the use of the fast calculation (5). Only the stopping condition is affected, becoming $\bar{\Delta}\mathbf{r}^{\hat{n}} < \lambda$ where

$$\bar{\Delta}\mathbf{r}^{\hat{n}} = \sqrt{\|\mathbf{r}^i\|_2^2 + \Delta\mathbf{r}^{\hat{n}}} - \|\mathbf{r}^i\|_2. \tag{8}$$

It is proposed to use the NNLS solution vector to initialise the backwards elimination approach, which is then referred to as Backwards From NNLS (BF-NNLS) outlined in Algorithm 1.

Algorithm 1 BF-NNLS ($\mathcal{C}_{mod}$)

```
Input    $\mathbf{D} \in \mathbb{R}^{M \times N}$ , $\mathbf{s} \in \mathbb{R}^{M}$ $\lambda$
Initialise
$\mathbf{x}^0 = \arg\min_x \|\mathbf{s} - \mathbf{D}\mathbf{x}\|_2^2$    s.t.    $\mathbf{x} \geq 0$
$\Gamma = \{j | x_j > 0\}$;    $i = 0$
$\mathbf{r}^0 = \mathbf{s} - \mathbf{D}\mathbf{x}^0$
repeat
  $i = i + 1$;    Select $\hat{n}$ using (3);
  $\Gamma = \Gamma \backslash \hat{n}$ ;    Calculate $\bar{\Delta}\mathbf{r}^{\hat{n}}$ using (8);
  $\|\mathbf{r}^i\|_2 = \|\mathbf{r}^{i-1}\|_2 + \bar{\Delta}\mathbf{r}^{\hat{n}}$
until $\bar{\Delta}\mathbf{r}^{\hat{n}} > \lambda$
Output $\Gamma$
```

3.2. Group backwards elimination

We have previously [14] used OMP-based algorithms with a dictionary comprised of a union of subspaces, each of which represented one note, for AMT. This approach led to greater modelling power in the dictionary and improved AMT performance. A group variant of BF-NNLS that proceeds similarly but uses a group elimination criteria (G-BF-NNLS)

$$\hat{l} = \arg\min_{l} \Delta\mathbf{r}^{[l]} \tag{9}$$

is proposed, where $\Delta\mathbf{r}^{[l]} = \|\bar{\mathbf{r}}^{[l]}\|_2^2 - \|\mathbf{r}^i\|_2^2$, similar to the standard backward elimination cost (4).

Using block inverse matrices, similar to [18], it is proposed to calculate the backward elimination step by

$$\Delta\mathbf{r}^{[l]} = \frac{\mathbf{x}[l]^T\mathbf{x}[l]}{[(\mathbf{D}_\Gamma^T\mathbf{D}_\Gamma)^{-1}][l,l]} \tag{10}$$

where $\mathbf{Y}[l,l]$ refers to a principal submatrix of the square matrix $\mathbf{Y}$ containing only the elements indexed by the lth block. However, (10) requires a matrix inversion for each group and cannot be calculated for all groups simultaneously as in the single atom case (5). It is worth noting that the size of each group may differ depending on the number of atoms selected during the initial NNLS decomposition.

In a similar fashion to the BF-NNLS algorithm, a modified group sparse cost function is proposed

$$\mathcal{C}_{mod(G)} = \|\mathbf{s} - \mathbf{D}\mathbf{x}\|_2 + \lambda\|\mathbf{x}\|_{\perp,0} \tag{11}$$

and the cost of the group downdate can also be calculated in a similar fashion using a group version of (8).

3.3. Molecular backward elimination

Further structure can be added to the decomposition by considering time-persistence. Here, it is proposed to do this in a straightforward manner, using an onset detector to delineate strips of the spectrogram. While the previous methods performed frame-wise decompositions, here the decompositions consider all time frames of a spectrogram strip simultaneously, leveraging the relationship between time frames in order to cancel spurious eliminations. This can also be considered a simultaneous sparse [7] approach. Similarly, through using several time frames together, with a group structure for each note at each time frame, this method can be seen as similar to that of the Collaborative Hierarchical Lasso (CHi-Lasso) [21].

Considering the set of detected onsets $O = \{o_1, ..., o_Q\} \subset \{1...N\}$, the elimination criteria for a group of pitch-similar atoms across, $\mathbf{S}^q$, the qth strip of the spectrogram is given by

$$\eta = \arg\min_{\eta} \sum_{t=o_q}^{o_{q+1}} \Delta\mathbf{r}_t^{\eta} \tag{12}$$

where η can represent n or $[l]$, in the sparse and group sparse cases respectively. The selection criteria (12) is then used in the Molecular-BF-NNLS, outlined in Algorithm 2, which proceeds similar to BF-NNLS with one notable difference. M-BF-NNLS runs until the support is empty, assigning an elimination value to each pitch-time point in the matrix $\bar{\mathbf{X}}$, which is then thresholded in a similar manner to NNLS.

Algorithm 2 M-(G)BF-NNLS

```
Input
  $\mathbf{D} \in \mathbb{R}^{\mathbf{M} \times \mathbf{N}}$; $K = o_{q+1} - o_q + 1$; $\mathbf{X}^0 \in \mathbb{R}^{N \times K}$
  $\mathbf{S}^q \in \mathbb{R}^{M \times K}$; $\Gamma^q$.
Initialise
  $i = 0$; $\bar{\mathbf{X}} = 0^{N \times K}$; $\mathbf{R}^0 = \mathbf{S}^q - \mathbf{D}\mathbf{X}^0$
repeat
  $i = i + 1$
  Select $\hat{\eta}$ using (12)
  for $t = 1 : K$ do
    $\Gamma_t^q = \Gamma_t^q \backslash \hat{\eta}$
    Calculate $\bar{\Delta}\mathbf{r}_t^{\hat{\eta}}$ using (8)
    $\|\mathbf{r}_t^i\|_2 = \|\mathbf{r}_t^{i-1}\|_2 + \bar{\Delta}\mathbf{r}_t^{\hat{\eta}}$
    $\bar{\mathbf{X}}_{\hat{\eta},t} = \bar{\Delta}\mathbf{r}_t^{\hat{\eta}}$
  end for
until $|\Gamma| = 0$
Output    $\bar{\mathbf{X}}$
```

3.4. Signal adaptive thresholding

Typically, thresholding of the activation matrix, $\mathbf{X}$, is performed to determine the AMT output. A common approach [10] is to adapt the threshold to the signal using $\lambda = \delta \times \max \mathbf{X}$, where δ is a parameter used in common across many pieces. The maximum activation value may be spurious, and it may be more robust to use a value that is more indicative of the signal in general. To this end the value $M_{m\%}(\mathbf{X})$ relating the mean of the highest $m\%$ positive values of $\mathbf{X}$ is used:

$$\lambda = \delta \times M_{m\%}(\mathbf{X}). \tag{13}$$

4. EXPERIMENTS

Transcription experiments were performed on a subset of MAPS [22], a database of MIDI-aligned piano pieces. The subset used, *EnStDkCl*, was recorded live on a Disklavier, and similar to [14] [23], the first $30sec$ of each piece was used. The selected pieces were downsampled to $22.05kHz$, and spectrograms were formed in two transforms; an STFT with a window size of $92ms$ with a 75% overlap and an ERBT [10] [23] with dimension 512 interpolated onto a 23*ms* grid.

MAPS also contains samples of isolated notes, which are used to form the dictionaries used for the experiments. We follow the experimental setup in [14], in which fixed dictionaries were concatenated from pitch-labeled subdictionaries $\mathbf{D}^{\eta} \in \mathbb{R}^{M \times P}$, each of which is learnt from a single note using Euclidean NMF [9]. Two different dictionaries were learnt for each transform, *single atom dictionaries* for the standard sparse case $(P = 1)$, and *subspace dictionaries* for the group sparse case with $P = 5$, as this groupsize was previously observed to perform well in group sparse AMT decompositions [14].

Frame based measures were used to compare the performance of the various algorithms, whereby the MIDI ground truth was compared with the AMT sparse support, or piano roll, at each time frame. A true positive tp was registered when a point in the pitch-time domain is supported by both the the ground truth and the AMT output, and false positives fp and false negatives fn are registered when the pitch-time point is supported only in the AMT output, and in the ground truth, respectively. Using these classifications, the following metrics are used to measure the performance of the algorithms: Precision, $\mathcal{P} = \frac{\# \, tp}{\# \, tp + \# \, fp}$, Recall, $\mathcal{R} = \frac{\# \, tp}{\# \, tp + \# \, fn}$ and $\mathcal{F}$-measure, $\mathcal{F} = 2 \times \frac{\mathcal{P} \times \mathcal{R}}{\mathcal{P} + \mathcal{R}}$.

For all decompositions an initial NNLS was performed using the Fast-NNLS [16] algorithm, an optimised version of the classic NNLS algorithm [15]. A threshold was used with all algorithms. In the case of T-NNLS and the Molecular-BF-NNLS algorithms thresholding was performed on the activation matrix, while in the case of the (G)BF-NNLS, the threshold was applied as a stopping condition. The value of the threshold was calculated using the coefficients of the initial NNLS activation matrix using the adaptive thresholding (13), with $m = 15\%$, for all algorithms. A range of values of $\delta \in \{0, ..., 50\}$*dB* was used, and the results presented consider the optimal $\mathcal{F}$-measure found at δ_{opt} across all pieces.

4.1. Modified Sparse Cost Function

The first set of experiments compare T-NNLS with BF-NNLS using both $\mathcal{C}_{sparse}$ (1) and $\mathcal{C}_{mod}$ (7) sparse cost functions. Following initial NNLS decompositions in both transforms with the *single atom dictionaries*, the relevant thresholding and eliminations were performed for each approach.

The results are shown in Table 1, where it is seen that BF-NNLS with $\mathcal{C}_{mod}$ (7) outperforms the other methods. It is also seen that δ_{opt} is consistent across transforms for T-NNLS and BF-NNLS ($\mathcal{C}_{mod}$). Using $\mathcal{C}_{sparse}$ (1) with BF-NNLS produces worse results than T-NNLS and a large discrepancy in the values of δ_{opt} between the two transforms. A considerable difference is seen between the results for STFT and ERBT, as previously observed [23].

	STFT		ERBT	
	δ_{opt}	$\mathcal{F}$	δ_{opt}	$\mathcal{F}$
T-NNLS	15	64.3	14	68.5
BF-NNLS ($\mathcal{C}_{sparse}$)	5	64	31	67.8
BF-NNLS ($\mathcal{C}_{mod}$)	27	65.7	27	69.8

Table 1. Comparison of T-NNLS with BF-NNLS

4.2. Group Sparsity

Group sparse decompositions were run using the *subspace dictionaries*, with Group T-NNLS (GT-NNLS) and G-BF-NNLS. GT-NNLS refers to a NNLS decomposition for which the group coefficients are calculated by

$$\mathbf{g}_{l,n} = \|\mathbf{D}[l]\mathbf{x}_n[l]\|_2 \tag{14}$$

and for which thresholding is performed on $\mathbf{G}$ in a similar manner to T-NNLS.

Results for these experiments are presented in Table 2 alongside the results of the algorithms using *single atom dictionaries* for comparison. The GT-NNLS provides a small improvement over the T-NNLS, however this is less of an improvement than seen with BF-NNLS. The ability of the *subspace dictionary* to afford better modeling of the signal is only exploited when the backwards elimination strategy and its explicit group sparse penalty is introduced, resulting in improvements of 6 to 7%, a large enhancement. Results not displayed show the improvement produced by using the modified sparse cost function in the group case (11) is larger than that for the the standard case (1) shown in the last section.

4.3. Molecular Approach

Experiments were run using the M-BF-NNLS, in the sparse and group sparse frameworks. A phase-based onset detector [24] was used to delineate the strips of the spectrogram, in which molecular decompositions take place. The experimental setup is the same as previous sections, with decompositions performed on both transforms, with both groupsizes.

The results are shown in Table 2 where they can be compared with the results from the standard and group sparse approaches. In both frameworks the molecular approach is seen to improve upon previous results. In the case of the STFT the results are more enhanced being of the order of 2.5%, and bringing these results close to those of the ERBT.

	STFT			ERBT		
	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$
T-NNLS	66.4	62.2	64.3	72.7	64.6	68.5
GT-NNLS	67.1	63.6	65.3	69.1	70.0	69.6
BF-NNLS	69.6	62.3	65.7	75.1	65.3	69.8
GBF-NNLS	76.7	68.1	72.2	78.1	73.2	75.6
M-BF-NNLS	71.7	64.7	68.0	75.0	68.2	71.4
M-GBF-NNLS	78.0	71.9	74.8	79.0	74.5	76.7
β-NMF	73.0	69.8	71.0	75.5	74.2	74.9

Table 2. AMT with different algorithms.

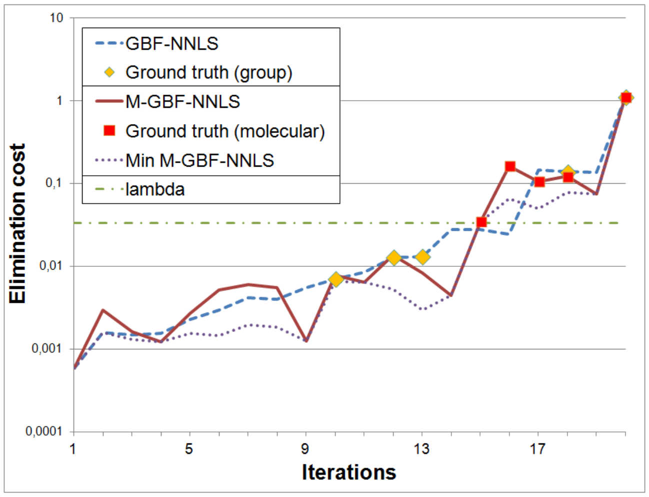

Fig. 1. Elimination cost at each iterations in a selected single time-frame, for GBF-NNLS and M-GBF-NNLS. Min M-GBF-NNLS indicates the smallest elimination cost during M-GBF-NNLS. Iterations where ground truth atoms are selected are outlined.

The results are also compared with decompositions using supervised β-NMF with the *single atom dictionaries* with $\beta = 0.5$. This set up has been shown to give state of the art supervised NMF results for AMT in [10], where it is noted that using more than one atom to represent a note decreases the AMT performance. It is seen, in both transforms, that the β-NMF outperforms all other algorithms using the *single atom dictionaries*. However all group BF-NNLS algorithms outperform β-NMF, an effect that is further enhanced with molecular group decompositions.

Further inspection of the results shows that the molecular method tends to more efficiently capture ends of long notes giving better $\mathcal{R}$, while maintaining $\mathcal{P}$. It is worth pointing out that, fundamentally, the molecular and unstructured approaches are the same, differing only in the fact that some temporal information is presented to the molecular selection criteria. However, this extra information does cause changes, tending to rule out spurious elements. An example of this interaction is shown in Figure 1 describing how $\bar{\Delta}\mathbf{r}$ behaves along iterations in a single time-frame for GBF-NNLS and M-GBF-NNLS. For both algorithms, there is no mathematical guarantee that $\bar{\Delta}\mathbf{r}$ will increase at each iteration, as the support changes over time and the atoms are not orthogonal. However, for GBF-NNLS, the elimination cost tends to increase monotonically, with exceptions typically found among the non-relevant atoms far below the stopping threshold, with very limited variation. On the contrary, M-GBF-NNLS doesn't necessarily select the atom with the smallest elimination cost at each iteration, and the corresponding curve displays a greater variability. In Figure 1, the iterations at which ground truth atoms are selected are indicated for each algorithm. Here it is seen that M-GBF-NNLS selects these correct atoms at a later stage than in the GBF-NNLS, and they therefore have a greater elimination cost.

	Detected Onsets			Ground truth		
	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$
M-BF-NNLS	75.0	68.2	71.4	75.3	68.4	71.7
M-GBF-NNLS	79.0	74.5	76.7	79.5	74.8	77.1

Table 3. Effect of onset detection on molecular algorithm using ERBT.

The onsets detection performed prior to the molecular method can sometimes perform poorly, particularly when the onset rate is high. In order to ascertain the level at which this may effect the overall performance, a comparison was made with the molecular algorithm using the ground truth onsets. Results given in Table 3 show the improvement in performance using the ground truth onsets to be relatively small.

4.4. Adaptive threshold

A small improvement in the $\mathcal{F}$-measure when using the proposed adaptive thresholding method is observed, of between 0.1 and 0.8%. Enhancements are more pronounced in the group case. While the improvement can seem insignificant, it is seen to be robust, holding for all algorithms, and computationally inexpensive.

5. CONCLUSIONS AND FURTHER WORK

We have proposed a backwards elimination approach to perform sparse decompositions in the context of AMT, using a modified sparse cost function. This approach was then extended to the group sparse framework, bringing the performance in line with other state-of-the-art decomposition methods. The modified sparse cost functions were seen to be apt, affording improved performance and consistent relative thresholding across transforms, and their importance may extend beyond the context of AMT. The proposed backwards elimination approach improves on OMP-based approaches, allowing easy determination of a stopping condition with improved time-continuity observed in decompositions, while suffering relatively in terms of computational expense. A variation on decomposition-adaptive thresholding was also proposed, effecting a mild but consistent improvement. Further work will focus on further adapting the threshold by incorporating some local measures of the decomposition.

A molecular variant of the BF-NNLS approach, using onsets to delineate areas of the spectrogram, was also proposed, showing further improvements in AMT performance particularly in the case of the STFT. This approach demonstrates the ease with which structure can be incorporated in the backward elimination framework, and further work will investigate if this may be applicable for the purpose of multi-instrument AMT. Multi-instrument signals are seen to display co-activity of instruments at many active points of NMF-based decompositions [10], and greedy sparse approaches have been seen to be relatively successful in this scenario [12]. We believe that the proposed backwards elimination method can outperform such OMP-based approaches.

6. REFERENCES

[1] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, Dec. 1998.

[2] R. Tibrishani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, pp. 267–288, 1994.

[3] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proceedings of 27th Annual Asilomar Conference Signals, Systems, and Computers*, 1993, vol. 1, pp. 40–44.

[4] B. Varadarajan, S. Khudanpur, and T. D. Tran, "Stepwise optimal subspace pursuit for improving sparse recovery," *IEEE Signal Processing Letters*, vol. 18, no. 1, pp. 27–30, January 2011.

[5] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society*, vol. 68, no. 1, pp. 49–67, Feb. 2006.

[6] Y. C. Eldar, P. Kuppinger, and H. Bolsckei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3042–3054, June 2010.

[7] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst, "Atoms of All Channels, Unite! Average Case Analysis ofMulti-Channel Sparse Recovery Using Greedy Algorithms," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 655–687, 2008.

[8] L. Daudet, "Sparse and structured decompositions of signals with the molecular matching pursuit," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 5, pp. 1808–1816, Sept. 2006.

[9] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems (NIPS)*, 2000, pp. 556–562.

[10] E. Vincent, N. Bertin, and R. Badeau, "Adaptive harmonic spectral decomposition for multiple pitch estimation," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 18, no. 3, pp. 528–537, March 2010.

[11] A. Dessein, A. Cont, and G. Lemaitre, "Real-time polyphonic music transcription with non-negative matrix factorization and beta-divergence," in *11th International Society for Music Information Retrieval Conference (ISMIR)*, 2010, pp. 489–494.

[12] P. Leveau, E. Vincent, G. Richard, and L. Daudet, "Instrument-specific harmonic atoms for mid-level music representation," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 1, pp. 116–128, Jan. 2008.

[13] S. K. Tjoa and K. J. Ray Liu, "Factorization of overlapping harmonic sounds using approximate matching pursuit," in *Proceedings of International Society for Music Information Retrieval Conference (ISMIR)*, 2011, pp. 257–262.

[14] K. O'Hanlon, H. Nagano, and M. D. Plumbley, "Structured sparsity for automatic music transcription," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2012*, 2012, pp. 441–444.

[15] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, Prentice Hall, 1974.

[16] R. Bro and S. De Jong, "A fast non-negativity-constrained least squares algorithm," *Journal of Chemometrics*, vol. 11, no. 5, pp. 393–401, September 1997.

[17] M. Slawski and M. Hein, "Sparse recovery by thresholded non-negative least squares," in *Advances in Neural Information Processing Systems (NIPS 24)*, 2011, pp. 1926–1934.

[18] B. Moghaddam, A. Gruber, Y. Weiss, and S. Avidan, "Sparse regression as a sparse eigenvalue problem," in *Information Theory and Applications Workshop, 2008*, 2008, pp. 219 –225.

[19] E. Elhamifar and R. Vidal, "Block sparse recovery via convex optimisation," *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 4094–4107, August 2012.

[20] B. L. Sturm, J. J. Shynk, and S. Gauglitz, "Agglomerative clustering in sparse atomic decompositions of audio signals," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2008, pp. 97–100.

[21] P.Sprechmann, I. Ramirez, P. Cancela, and G. Sapiro, "Collaborative sources identification in mixed signals via hierarchical sparse coding," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2011, pp. 5816–5819.

[22] V. Emiya, R. Badeau, and B. David, "Multipitch estimation of piano sounds using a new probabilistic spectral smoothness principle," *IEEE Transactions on Audio Speech and Language Processing*, vol. 18, no. 6, pp. 1643–1654, Aug. 2010.

[23] K. O'Hanlon and M. D. Plumbley, "Row-weighted decompositions for automatic music transcription," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013.

[24] J.P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and Mark B. Sandler, "A tutorial on onset detection in music signals," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 1035–1047, 2005.