Highly Scalable Matching Pursuit Signal Decomposition Algorithm

Daniel Christensen * Santanu Das † Ashok N. Srivastava ‡

Abstract

Matching Pursuit Decomposition (MPD) is a powerful iterative algorithm for signal decomposition and feature extraction. MPD decomposes any signal into linear combinations of its dictionary elements or "atoms". A best fit atom from an arbitrarily defined dictionary is determined through cross-correlation. The selected atom is subtracted from the signal and this procedure is repeated on the residual in the subsequent iterations until a stopping criterion is met. The reconstructed signal reveals the waveform structure of the original signal. However, a sufficiently large dictionary is required for an accurate reconstruction; this in return increases the computational burden of the algorithm, thus limiting its applicability and level of adoption. The purpose of this research is to improve the scalability and performance of the classical MPD algorithm. Correlation thresholds were defined to prune insignificant atoms from the dictionary. The "Coarse-Fine" Grids and Multiple Atom Extraction techniques were proposed to decrease the computational burden of the algorithm. The Coarse-Fine Grids method enabled the approximation and refinement of the parameters for the best fit atom. The ability to extract multiple atoms within a single iteration enhanced the effectiveness and efficiency of each iteration. These improvements were implemented to produce an improved Matching Pursuit Decomposition algorithm entitled MPD++. Disparate signal decomposition applications may require a particular emphasis of accuracy or computational efficiency. The prominence of the key signal features required for the proper signal classification dictates the level of accuracy necessary in the decomposition. The MPD++ algorithm may be easily adapted to accommodate the imposed requirements. Certain feature extraction applications may require rapid signal decomposition. The full potential of MPD++ may be utilized to produce incredible performance gains while extracting only slightly less energy than the standard algorithm. When the utmost accuracy must be achieved, the modified algorithm extracts atoms more conservatively but still exhibits computational gains over classical MPD. The MPD++ algorithm was demonstrated using an over-complete dictionary on real life data. Computational times were reduced by factors of 1.9 and 44 for the emphases of accuracy and performance, respectively. The modified algorithm extracted similar amounts of energy compared to classical MPD. The degree of the improvement in computational time depends on the complexity of the data, the initialization parameters, and the breadth of the dictionary. The results of the research confirm that the three modifications successfully improved the scalability and computational efficiency of the MPD algorithm. Correlation Thresholding decreased the time complexity by reducing the dictionary size. Multiple Atom Extraction also reduced the time complexity by decreasing the number of iterations required for a stopping criterion to be reached. The Course-Fine Grids technique enabled complicated atoms with numerous variable parameters to be effectively represented in the dictionary. Due to the nature of the three proposed modifications, they are capable of being stacked and have cumulative effects on the reduction of the time complexity.

1 Introduction

Structural Health Monitoring (SHM) deals with the detection, localization, classification, and prediction of remaining useful life (RUL) of structures and materials. Numerous Non-Destructive Evaluation (NDE)

*D. Christensen is currently associated with Department of Aerospace Engineering, Iowa State University, IA 50011. (email:aerodc@iastate.edu)

†S. Das is currently associated with University Affiliated Research Center, UC Santa Cruz,NASA Ames Research Park, Moffett Field, CA 94035. (email:Santanu.Das-1@nasa.gov)

‡A. N. Srivastava is the Principal Investigator of the Integrated Vehicle Health Management Project at NASA. He is in the Intelligent Systems Division at NASA Ames Research Center, Moffett Field, CA, 94035 (email: Ashok.N.Srivastava@nasa.gov)

techniques are applicable for SHM; wave-based methods are one of the most commonly used approaches. Actuators are often employed to excite vibrations that propagate through the media; distributed piezoelectric sensors at other locations convert the received wave into a voltage that may be analyzed. As the vibration travels through the medium, the material properties and boundary conditions affect the propagation of the wave. In a damaged state, changes in material properties or structural integrity cause a deviance from the baseline, healthy-state system response. As a result, characteristic signatures regarding the quality, or health, of the structure or material may be detected by the sensors. Unfortunately, the change in response between healthy and damaged conditions is often minor and buried amongst other dominant features.

Feature extraction is a key step for accurate analysis of these aberrations. Feature extraction detects the primary components of the signal while neglecting the inherent noise. However, the necessity of precision in the decomposition complicates the feature extraction algorithm. Failure to avoid mediocre or inaccurate decompositions may result in the false classification of the structure's health. A popular feature extraction technique that addresses these issues is the Matching Pursuit Decomposition algorithm.

2 Motivation

3 Classical MPD

The Matching Pursuit Decomposition (MPD) algorithm was originally developed by Mallat and Zhang [3] for signal decomposition. In MPD, function segments or "atom" are matched to portions of the signal. Once the best matched atom is determined, it is extracted from the signal and the algorithm proceeds onto the next iteration. Any functional representation may be used to define atoms; even convoluted elements, such as chirps or impulse functions, may be represented with MPD. Each atom is normalized to unit energy. As a result, the amplitude of a match is merely the cross-correlation value of the matched atom with the specified portion of the signal. A dictionary is comprised of arbitrarily defined atoms; a wide variety of atoms may inhabit the same composite dictionary.

MPD is a greedy algorithm that extracts the largest amount of energy possible per iteration. Greedy algorithms operate by determining the locally optimum solution for a given iteration in an effort to optimize the global solution [5]. Each iteration is conducted independently with no knowledge of previous iterations and without any concern for future iterations.

Cross-correlation is used to quantify the degree of similarity between atoms and the reference signal. During each iteration, the cross-correlation with the signal is calculated for every atom in the dictionary. The best match from the dictionary is the atom with the largest cross-correlation value and its corresponding time delay. After this optimal match is determined, it is extracted from the signal and stored for later reconstruction or classification.

The residual after k iterations is represented by $R_x^k[n]$, where the signal is notated by $x[n]$. Equation 1 shows the initialized state prior to the first iteration. The best fit atom from dictionary D is determined through cross-correlation with the residual as shown in Eqn. 2. The selected atom is subtracted from the signal (Eqn. 3) and the procedure is repeated on the residual in the subsequent iteration. The i^{th} matched atom is represented by $d_{\gamma i}$ and its corrosponding cross-correlation with $a_{\gamma i}$, where the parameters for a particular atom are notated by γ.

$$R_x^0[n] = x[n] \tag{1}$$

$$a_{\gamma i} = argmax_{d_{\gamma i} \in D} \left| \left\langle R_x^k[n], d_{\gamma i} \right\rangle \right| \tag{2}$$

$$R_x^k[n] = R_x^{k-1}[n] - \left\langle R_x^k[n], d_{\gamma i} \right\rangle a_{\gamma i} \tag{3}$$

The algorithm is repeated until one of three possible criterions are reached: a specified number of matches are removed from the signal, a designated amount of energy is extracted from the signal, or the extracted components remove below a certain threshold amount of energy. These criterions prevent the algorithm from modeling noise or nonexistent features.

A pseudo code for the MPD algorithm is shown in 3. The algorithm begins with the generation of the dictionary. The algorithm continually determines and extracts the best matches until a stopping criterion is met, thus signaling the completion of the decomposition.

Algorithm 1 Matching Pursuit Decomposition

1: Build Dictionary: $D = \{d_{\gamma 1}, d_{\gamma 2} \ldots, d_{\gamma j} \ldots, d_{\gamma n}\}$, where $d_{\gamma n}(t) = \frac{1}{\sqrt{\alpha_n}} d\left(\frac{t-\beta_n}{\alpha_n}\right) e^{j2\pi\kappa_n t}$
2: Initialize K_{stop}, δ_{stop}, $k = 0$, $R_x^0[n] = x[n]$, $E_x^0 = \left\|R_x^0\right\|_2$
MPD routine
3: **while** $k < K_{stop}$ or $E_x^k > \delta_{stop}$ **do**
4: $a_{\gamma j}^k = \left\langle R_x^k[n], d_{\gamma j}^k \right\rangle$
5: Select dictionary element whose time correlation with the $R_x^k[n]$ is maximum
6: $R_x^k[n] = R_x^{k-1}[n] - a_{\gamma j}^k d_{\gamma j}^k[n]$
7: $k = k + 1$, $E_x^k = \left\|R_x^k\right\|_2$
8: **end while**

Once decomposed, the signal may be reconstructed through the synthesis of the matches as shown in Eqn. 4. The number of matches is represented by N.

$$\hat{x} = \sum_{i=1}^{N} a_{\gamma i} d_{\gamma i} \tag{4}$$

As mentioned previously, a sufficiently large dictionary is required for accurate reconstruction. This limitation increases the computational burden of the algorithm.

4 Modifications to the MPD Algorithm

Modifications addressing two aspects of the algorithm are proposed. One approach involves the implementation of an intelligent, self-adaptive, or reduced-in-size dictionary to assuage the complications caused by abundantly populated dictionaries. The Correlation Thresholding (CT) and Coarse-Fine Grids (CFG) modifications operate along this concept through the reduction of the dictionary's size. A second method, the Multiple Atom Extraction technique, modifies the structure and flow of the algorithm. The improved MPD algorithm (entitled MPD++) and its modifications are discussed in the following sections.

4.1 Correlation Thresholding

Significant time complexity is required while computing the cross-correlation values for each atom. While a sufficiently complete dictionary is required for accurate decomposition, a majority of the atoms may not be used throughout the decomposition process. The removal of these unused atoms from the dictionary reduces the size of the dictionary and potentially offers significant increases in performance with no cost in accuracy.

CT begins by calculating the correlation values for all of the atoms in the dictionary through a standard MPD iteration. After the initial iteration, the correlation ratio for each atom is calculated. The correlation ratio (CR), for a particular atom is defined as that atom's maximum correlation value divided by the largest correlation of all atoms in the dictionary (Eqn. 5). CT applies a threshold to the correlation ratio for each atom in the dictionary. Atoms with correlation ratios above the threshold form the reduced dictionary; atoms that do not meet this threshold are rejected. This reduced dictionary is used in the subsequent iterations instead of the full, base dictionary. With fewer atoms to analyze, the algorithm will require less time per iteration. After a repruning criterion is met, as discussed below, the reduced dictionary will be thrown out and the base dictionary will be repruned to generate a new reduced dictionary.

$$CR_n = \frac{\left\langle R_x^k[n], d_{\gamma i} \right\rangle}{argmax_{d_{\gamma i} \in D} \left| \left\langle R_x^k[n], d_{\gamma i} \right\rangle \right|} \tag{5}$$

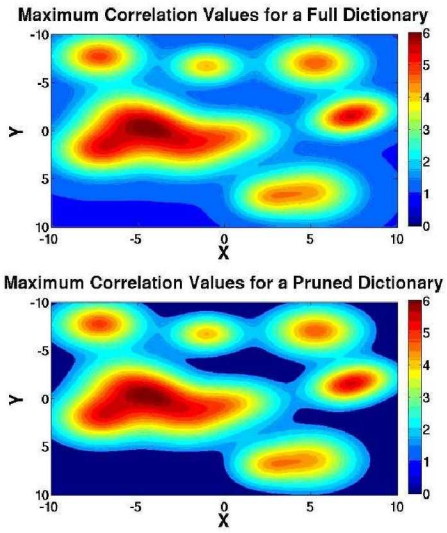

Figure 1: A visualization of Correlation Thresholding

The cross-correlation calculations required for CT are computed recursively during a classical MPD iteration. The only increase in computational complexity due to CT is the determination of offending atoms and their removal from the dictionary. This increase in computational load is trivial compared to the cost of calculating correlation values with the full, unreduced dictionary in the unmodified algorithm. Figure 1 provides a visualization of CT. While this example uses superposed two-dimensional Gaussian distributions, it resembles the results of an actual dictionary. In this example, atoms are generated by varying two variables, X and Y. Each point on the grid represents specific parameters for a particular atom. Local maximums represent atoms that fit well into the signal. If one of the variables is offset marginally from a local maximum, its amplitude decreases slightly. This effect produces the gradual sloping effect observed in the first plot in Fig. 1.While the top plot shows an example base dictionary, the bottom plot shows a reduced dictionary. The dark blue areas mark atoms that were excluded from the reduced dictionary due to a reduction threshold violation.

4.1.1 Reduction Threshold

The reduction threshold specifies the minimum correlation ratio that will be admitted in the reduced dictionary. Subsequent iterations use the reduced dictionary to determine the best matched atom. To ensure that no atoms are lost during the algorithm's operation, a base dictionary with all atoms is preserved. When either the reprune or ultimate thresholds are violated, the dictionary is repruned from the base dictionary rather than the current reduced dictionary. This ensures that all atoms receive equal opportunity for representation in the decomposition of the signal. The ultimate and reprune thresholds are determined relative to the reduction threshold as discussed in section 4.1.5.

4.1.2 Ultimate Threshold

The ultimate threshold dictates the absolute minimum correlation ratio value accepted as a match in the decomposition. To ensure an up-to-date reduced dictionary, it must be repruned occasionally from the base dictionary. If the current match's correlation ratio violates the ultimate threshold, the match will be discarded, the dictionary will be repruned, and the iteration will be repeated. Otherwise, the algorithm will extract the match and continue onto the next iteration. A violation of this threshold wastes computational time, but helps to ensure the integrity of the reconstruction. The reduction threshold only determines which atoms form the reduced dictionary. After the construction of the reduced dictionary,

the reduction threshold plays no role in the algorithm until the next repruning. As a result, another threshold must be implemented to dictate when the algorithm is to discard the current reduced dictionary. The ultimate threshold is typically set relative to the reduction threshold to produce the desired emphasis of accuracy or performance; however, these thresholds operate independently in the algorithm.

4.1.3 Reprune Threshold

As mentioned previously, violating the ultimate threshold wastes computational resources. If the dictionary is repruned just before the ultimate threshold is breached, the algorithm may avoid the rejection of a match and repetition of an iteration. The reprune threshold helps to ensure smooth flow of the MPD++ algorithm.

The reprune threshold is the second factor in the repruning of the dictionary. The reprune threshold's purpose is to reprune the base dictionary before a violation of the ultimate threshold occurs. This threshold is typically set slightly above the ultimate threshold. As a result, the dictionary may be repruned before the ultimate threshold is violated. However, setting too high of a reprune threshold will cause the algorithm to excessively reprune the dictionary at the expense of potential performance gains.

If the current match's correlation ratio is below the reprune threshold and above the ultimate threshold, it will be subtracted from the signal and the dictionary will be repruned at the beginning of the following iteration. If the reprune threshold is not offended, the algorithm will simply extract the atom and continue onto the next iteration without changing the dictionary.

4.1.4 Interactions between Reprune and Ultimate Thresholds

The melody of interactions between the reprune and ultimate thresholds directly influence the gains in computational performance and the accuracy of the reconstruction. The optimal thresholds are dependent on the signal being decomposed. However, depending whether the emphasis is on obtaining better accuracy or faster results, the thresholds may be approximated and still produce significant increases in performance. If the emphasis is on accuracy, larger, more conservative values of the reprune and ultimate thresholds are recommended; otherwise if maximum performance gains are required, lower thresholds may be chosen.

4.1.5 Identical Reconstruction versus Maximum Performance

Accuracy is often of primary importance. Unfortunately, standard MPD is often too computationally expensive to employ. Modifications that reduce the computational burden of the algorithm are desirable; however, reconstructions with an improved algorithm are often required be identical to those produced by classical MPD. Correlation Thresholding may be applied without losing accuracy. If the ultimate and reduction thresholds have identical values, the program will discard any matches that have a correlation ratio below those of the atoms that were rejected from the dictionary. This prevents the extraction of any matches when an atom that was removed from the reduced dictionary may offer a better match.

After the extraction of other matches in previous iterations, it is possible that an atom rejected from the reduced dictionary may have a correlation ratio above the reduction threshold. While this was not observed during the testing of the modification, it is plausible. To ensure the highest accuracy obtainable, the ultimate threshold may be set slightly above the reduction threshold. The increase in the confidence and accuracy of the results will sacrifice potential performance gains, but the modified algorithm will still be significantly faster than standard MPD.

The fastest possible reconstruction while maintaining respectable accuracy may often be the primary focus. If the ultimate threshold is set at a value below the reduction threshold, the improved algorithm will be able to extract more atoms before having to reprune the base dictionary. However, it is possible for the program to select an atom that normally would not have been picked or miss the true best match for that iteration because it was not present in the reduced dictionary. Significant gains in computational performance may be obtained at a decrease in accuracy that is acceptable for many applications. The reprune threshold should still be set slightly above the ultimate threshold and may be above, equal to, or below the reduction threshold. The maximum performance gains are obtainable when both the reprune and ultimate thresholds are below the reduction threshold.

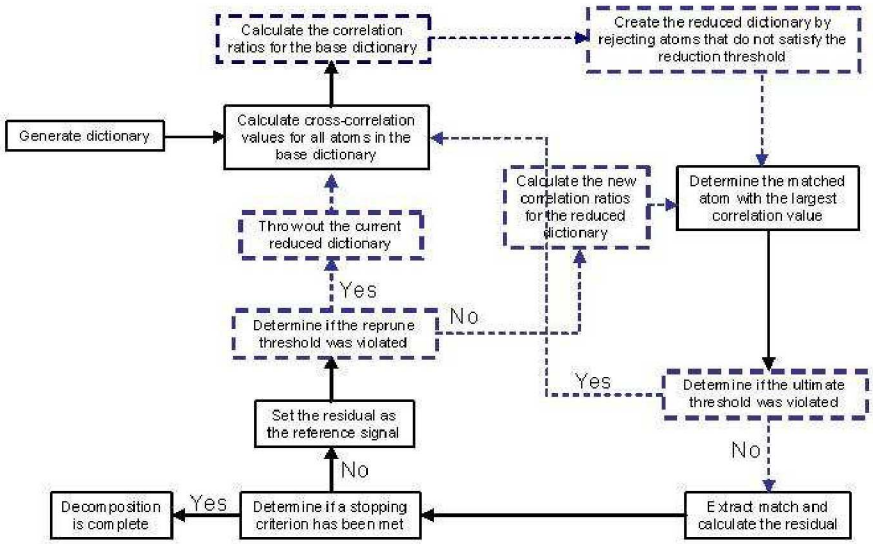

Figure 2: Example of different emphases for Correlation Thresholding

For a balance between accuracy and performance, reprune and ultimate thresholds slightly above and below, respectively, are recommended. Identical reconstruction cannot be guaranteed, but the hybrid emphasis still offers significantly improved performance.

The thresholds and correlation ratio of the matched atom determine the action taken by the algorithm. As long as the correlation ratio is above the ultimate threshold, it will be accepted. Figure 2 shows a flowchart for the CT technique. Blue, dashed arrows and boxes indicate steps and flow paths that differ from classical MPD.

4.1.6 Permanent Deletion Threshold

There are often atoms in the dictionary that are very poor matches with the signal and are excluded from all reduced dictionaries during the decomposition. Removing these atoms offers an additional opportunity to further reduce the dictionary. Any atoms with correlation ratios below the permanent deletion threshold will be removed from the base dictionary and will not be included in future reprunings of the dictionary. Significant performance gains are possible with a permanent deletion threshold; however, the risk of losing accuracy may deter its use in many applications.

4.2 Coarse-Fine Grids

Another method to reduce the size of the dictionary is to sparsify the dictionary using coarse and fine grids. The Coarse-Fine Grids (CFG) technique allows for the approximation and refinement of the parameters for the best fit atom. The best-correlated atom in the coarse grid is determined through a standard MPD iteration, but the atom is not immediately subtracted. Instead, the coarse grid is "zoomed-in" and a fine grid is applied around the selected atom. The best match within the fine

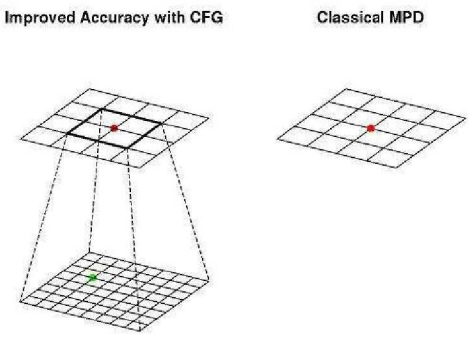

Figure 3: Improved Accuracy with CFG

grid is determined and extracted from the signal. Similar research has been conducted using quadratic interpolation [1], Coarse-Fine Grids and interpolation [2], and Particle Filtering [4].

4.2.1 Improving Accuracy versus Time Complexity

In scenarios that require the highest accuracy possible, the implementation of CFG produces satisfying results. In accuracy mode, classical MPD and MPD++ begin with the same dictionary, represented by a grid as shown in Fig. 3. After the best fit atom is determined with MPD++ (the red dot), the algorithm generates a fine grid around the selected atom. From this fine grid, the best match (the green dot) is determined and extracted from the signal. Although this technique increases the accuracy of the extraction, the generation and testing of a fine grid increases computational time. Multiple fine grids may be implemented to additionally increase the accuracy of the reconstruction and extracted features. Each fine grid is essentially a "zoomed-in" version of the previous, coarser grid. With an emphasis on accuracy, the CFG modification provides MPD++ an advantage over standard MPD by allowing the modified algorithm to select atoms that the latter cannot.

If the reduction of the time complexity is of utmost concern, a second approach offers significant reduction in computational demands of MPD at a minor decrease in accuracy and extracted energy. In this approach, MPD++ and classical MPD begin with the same grid. In the modified algorithm, the coarse grid is generated by 'zooming-out' or sparsifying the base grid as shown on the left in Fig. 4. After the best fit atom (the red dot) is determined in the coarse grid, a fine grid is generated (middle image in Fig. 4). The final best match (the green dot) in this fine grid is extracted from the signal and the algorithm proceeds onto the next iteration. The fine grid for MPD++ maintains the same resolution as the classical MPD's grid. This approach allows both the modified and standard MPD iterations to select identical atoms. Due to the sparsity of the coarse grid, MPD++ has a significant advantage computationally. A minor loss of accuracy occurs when the fine grid around the selected coarse grid point does not contain the true best atom. For most applications, the loss in accuracy is negligible; however, the increases in computational efficiency are substantial.

4.2.2 Multi-Parameter Atoms

Due to the nature of the Coarse-Fine Grids technique, the greatly sparsified dictionary allows atoms with numerous parameters to be readily adopted into the MPD algorithm. Without CFG, atoms with numerous parameters causes the dictionary size to increase exponentially with the number of parameters and severely limits the effectiveness of standard MPD. The Coarse-Fine Grids technique allows complex atoms to be represented in the dictionary while maintaining a reasonable dictionary size. With the CFG technique, more convoluted element types may be implemented before the computational burden becomes unbearable.

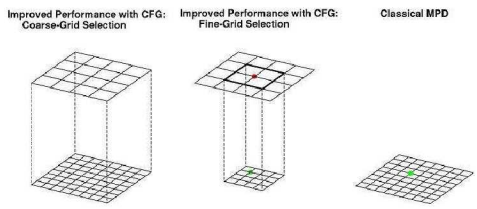

Figure 4: Improved Performance with CFG

4.2.3 Choice of Grids

The selection of grid ranges and increments is only marginally dependent on knowledge of the signal. The sampling frequency and length of the signal often dictate the ranges and increment sizes for the atom parameters. In fact, frequency is dependent solely on the sampling frequency. The frequencies of the atoms typically range between a single coarse increment above zero and below the Nyquist Criterion. This ensures that a broad range of frequency values inhabit the dictionary; a sufficiently complete dictionary is required for decomposition. Other parameters, such as time scale, should be chosen so that the widest atom stretches across a significant portion of the signal while the smallest atom is capable of modeling localized behavior.

Some parameters are more sensitive to poor matching with an excessively coarse grid. Frequency is a notable example of this effect; an overly coarse frequency grid may cause the algorithm to jump over the true best-correlated atom and instead pick a suboptimal match. This occurs because the atom on the coarse grid closest to the best match may be out of phase with the matched signal for a substantial portion of the matched signal segment. Other parameters, such phase shift or time scale may be less sensitive to this effect. A phase shift grid may be very coarse because the atom may be matched to a slightly offset position in the signal with a nearly identical correlation value. Whether a particular parameter requires a smaller increment is dependent on the nature of the parameter.

The ranges and increments may be determined for most parameters when only the duration and sampling frequency of the signal are known. However, a significant reduction in the time complexity is possible when additional knowledge of the signal is applied.

4.3 Multiple Atom Extraction

With even the most effective, abridged dictionary, MPD is fundamentally limited by the nature of its own existence. The iterative basis guarantees that numerous calculations must be performed to obtain the final decomposition. The extraction of multiple atoms per iteration is proposed to reduce the computational burden of the algorithm. The Multiple Atom Extraction (MAE) modification alters the flow and organization of the MPD algorithm instead of merely addressing the encumbering effects caused by excessively large dictionaries.

Classical MPD determines and extracts the single best-correlated atom. However, MPD++ determines a specified number of atoms with the largest correlation values, referred to as the "top atoms". Next, the algorithm subtracts the absolute best-correlated atom from the dictionary; this atom is called the "primary atoms" and all other top atoms are called "secondary atoms". Then, the algorithm determines whether the next best correlated secondary atom overlaps the footprint of a previously removed match. Any overlap will result in a rejection of the atom; however, if no overlap exists, the atom will also be subtracted from the dictionary. The process is repeated until all of the stored top atoms are either extracted or rejected. When this condition is reached, the algorithm will proceed onto the next

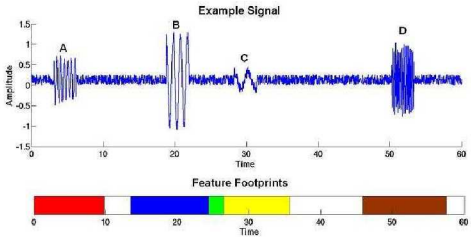

Figure 5: Example of Multiple Atom Extraction

iteration.

An example of Multiple Atom Extraction is shown below in Fig. 5. This signal shows four sinusoidal segments of different frequencies with the addition of white noise. Due to the greedy nature of MPD and the normalization of the atoms to unit energy, the features with the largest amplitude are extracted first. In this example, the algorithm will extract feature B first, proceeded by the extraction of features D and then A. An overlap, represented by the green region, exists between the footprints of features B and C as shown in the lower plot in Fig. 5. The extraction of feature C will be rejected. With Top Atom Tracking mode, discussed below, the algorithm will reattempt the extraction of feature C. Since the extraction of feature B did not influence the correlation value of feature C, this final feature will be extracted. It may be seen that the footprints of features A, B, and D do not overlap.

4.3.1 Buffer Range

The buffer range is a specified portion of the atom's time window that is padded to each side of the extraction site. The buffer range's purpose is to prevent secondary atoms from being extracted too close to a previously extracted atom. Without a buffer range, the extraction of a match may influence the correlation values of other atoms in the same proximity; this may result in detrimental effects on the quality of future extractions. For the fastest deconstruction, no buffer range is necessary and respectable results are obtainable. When accuracy is emphasized, the application of a buffer range is recommended to produce more conservative results. A reduction in time complexity relative to classical MPD is still observed. The footprint of a match is defined as the matched atom's time window plus the buffer zones on both sides of the extraction site.

4.3.2 Attempted Number of Atoms to Extract per Iteration

As the algorithm attempts to extract more atoms per iteration, the time complexity and amount of extracted energy tends to decrease. For conservative, accurate reconstructions, it is recommended to attempt extraction on a small number of atoms per iteration. For maximum performance, numerous atoms may be extracted in each iteration. In fact, if an excessively large number of top atoms are stored, the modification offers incredible performance increases with only a minor loss in extracted energy.

4.3.3 MAE Cutoff Level

A popular stopping criterion is to specify the number of matches to extract. As the algorithm progresses and nears this limit, it is possible that secondary atoms may remove less energy than the primary atom in the subsequent iteration. This undesirable consequence may adversely affect the signal's decomposition and the accuracy of the extracted features. As a result, a cutoff level may be specified to maintain the quality of the decomposition. When there are fewer unextracted atoms remaining than the cutoff

level, the algorithm will deactivate MAE mode and will extract the remaining atoms individually. This increases the accuracy and amount of energy extracted at the cost of potential performance gains.

4.3.4 Reuse Atoms

Features may sometimes appear recurrently throughout a signal; measurements with a periodic basis are a primary example. The Multiple Atom Extraction modfication only stores an atom's single best correlation value. As a result, the algorithm only detects the first appearance of a recurring feature. The "resuse atoms" option forces the program to search for multiple locations in the signal where an atom may generate large correlation values.

For a given atom, the search for well-correlated matches beyond the first typically requires an increase in computational cost. For complicated or non-periodic signals that lack repetition, such as chirps, a single atom is seldomly matched multiple times. As a result, searching for recurrent features costs computational resources while offering no increase in extracted energy. However, signals with redundant features may observe an increase in performance rather than a cost.

For an unknown signal, it is recommended to force the program to search for recurrent features to determine the best decomposition. Searching for repetition on a non-periodic signal does not affect the amount of energy extracted.

4.3.5 Top Atom Tracking

Sometimes the algorithm will uncover atoms that overlap but model different parts of the signal, such as different frequency components. The "Top Atom Tracking" (TAT) mode offers previously rejected secondary atoms an additional chance to be extracted. If TAT is active, the algorithm will store an extra top atom to use as a reference; extraction will never be attempted on this extra stored atom. After the first pass through the list of the top atoms and the extraction of as many atoms as possible, the algorithm will recompute the correlation values for the top atoms that were rejected in the previous pass due to a footprint overlap. If these atoms have a correlation value larger than the extra top atom stored, the atoms will be extracted; otherwise these atoms will be permanently rejected from the current iteration. This process of rechecking secondary atoms repeats until all secondary atoms are either extracted or rejected. Therefore, numerous atoms overlapping each other have a chance to be extracted.

4.3.6 Cumulative Effects of Multiple Modifications

All three modifications are cumulative and may result in phenomenal decreases in the computational burden of MPD. The modifications may be customized to achieve the accuracy or time complexity required for a particular application.

5 Conclussion

All three modifications resulted in significant improvements in the time complexity. Due to the nature of the three proposed modifications, they are capable of being stacked and have cumulative effects on the reduction of the time complexity. The goal of improving standard MPD has been achieved. Reductions in computational time of 1.9 and 44 were observed for the emphases of accuracy and performance, respectively. Substantial performance increases, such as a 300 times improvement in time complexity, may be achieved at a slight cost in accuracy. However, a significant performance increase is still possible if more accuracy is desired. The MPD++ algorithm will be tested in future work to determine the quality of the extracted features. The modified algorithm is significantly faster; however, it must be proven that valuable information for SHM analysis is not being lost. The algorithm will also be tested on other signals. While performance and accuracy improvements may be observed for many signals; the magnitude of improvement observed in this tested signal is not necessarily typical or representative of true gains.

Acknowledgment

This research was conducted as part of an internship through NASA Undergraduate Student Research Program (USRP). The improved algorithm was tested on a signal courtesy of the Air Force Research Laboratory. This work was supported through funding from the NASA Aeronautics Research Mission Directorate, Aviation Safety Program, Integrated Vehicle Health Management project. The authors thank Drs. Kanishka Bhaduri and Nikunj Oza for valuable discussions and suggestions.

References

[1] P. Leveau and L. Daudet, *Model-based matching pursuit - estimation of chirp factors and scale of gabor atoms with iterative extension*, SPARS (2005).

[2] Michaels Lu, Yinghui and Jennifer, *Numerical implementation of matching pursuit for the analysis of complex ultrasonic signals*, 2008.

[3] Stphane Mallat and Zhifeng Zhang, *Matching pursuits with time-frequency dictionaries*, 1993.

[4] A. Chattopadhyay S. Das, I. Kyriakides and A. Papandreou-Suppappola, *Monte carlo matching pursuit decomposition method for damage quantification in composite structures*, Journal of Intelligent Material Systems and Structures (2008 (accepted)).

[5] R. L. Rivest C. E. Leiserson T. H. Cormen, C. Stein, *ntroduction to algorithms*, McGraw-Hill Higher Education, 2001.