Greedy adaptive discrimination: component analysis by simultaneous sparse approximation

Jeffrey M. Sieracki[a] and John J. Benedetto[b]

[a]SR2 Group, LLC, PO Box 1011, College Park, MD 20741
[b]Norbert Wiener Center for Harmonic Analysis and Applications,
Department of Mathematics, University of Maryland,
College Park, MD 20742

ABSTRACT

Sparse approximation is typically concerned with generating compact representation of signals and data vectors by constructing a tailored linear combination of atoms drawn from a large dictionary. We have developed an algorithm based on simultaneous matching pursuits that facilitates the concurrent approximation of multiple signals in a common, low-dimensional representation space. The algorithm leads to an effective method of extracting signal components from collections of noisy data, and in particular is robust against jitter as well as additive noise. We illustrate its utility and compare performance in several variations by numerical examples.

Keywords: Sparse Approximation, Matching Pursuits, Wavelets, GAD

1. INTRODUCTION

Much of signal processing is concerned with transforming raw data into representations that facilitate extraction of information. Successful analysis of signals, images, and other ordered datasets is tied closely to the choice of representation. Low-dimensional representations have advantages: by concentrating information into fewer, larger coefficients, dimensionality is reduced, round-off error instabilities are minimized, and decision algorithms become more transparent. Low-dimensional representation can also aid in interpretation of signal structure and facilitate extraction of natural features from the data.

We may describe ordered data as a function f in one or more variables. For example, an audio recording may be described as $f(t)$, an image as $f(x, y)$, and so on. Each such f, generically termed a *signal*, lies in an appropriate Hilbert space H, termed the *signal space*. Decomposing a signal f into a linear combination of components generates a *representation* of f,

$$f = a_0 g_0 + a_1 g_1 + ... + a_i g_i + ... \tag{1}$$

The coefficients a_i depend upon the component *atoms* g_i, and the collection of atoms $\{g_i\} \subset H$ defines a *representation space* for the decomposition.

If $\{g_i\}$ is a fixed, spanning set of H, then for a given $f \in H$ the expansion (1) will generally distribute information over a large set of coefficients. Rather than fixing $\{g_i\}$ a priori, one may instead tailor the representation space to the signal f by adaptively choosing $\{g_i\} \subset H$ in order to reduce the dimensionality of the representation. Typically each g_i belongs to a large dictionary, $\mathcal{D} \subset H$, from which sufficient atoms are selected to approximate f. We may write

$$f = a_0 g_{\gamma_0} + a_1 g_{\gamma_1} + ... + a_n g_{\gamma_n} + R, \tag{2}$$

where R is the residue (if any) after approximation to $n + 1$ terms. When $R = 0$ the representation is exact, otherwise we say (2) is an *approximation* of f. We say that (2) is *sparse* when n is small compared to the dimensionality of the signal space H. In general, each new approximation generates a different representation space since each g_{γ_i} depends on f.

For further information on this manuscript please contact Jeff Sieracki at spiepaper@sr2group.com.

Real-world signal component analysis often relies on comparing one signal to another signal in order to determine similarities and differences. Suppose that we adaptively approximate two signals,

$$f^1 = a_0 g_{\gamma_0^1} + a_1 g_{\gamma_1^1} + ... + a_n g_{\gamma_n^1} + R^1 \tag{3}$$

and

$$f^2 = a_0 g_{\gamma_0^2} + a_1 g_{\gamma_1^2} + ... + a_n g_{\gamma_n^2} + R^2. \tag{4}$$

In general $\left\{g_{\gamma_i^1}\right\} \neq \left\{g_{\gamma_i^2}\right\}$, so it is not possible to compare directly the representation coefficients of one signal with those of the other without further processing. It becomes necessary either to map one generated representation space to another, or to re-represent all the signals of interest in a common space by means of an additional transform. The problem is compounded since in practical applications one often analyzes a large collections of signals $\left\{f^i\right\}_{i=1}^M$ where $M >> 2$. Ideally, we would like to find a representation space such that $g_{\gamma_i^1} = g_{\gamma_i^2}$ for all $i \in (0, n]$ but that simultaneously generates sparse approximations for all signals $\left\{f^i\right\}_{i=1}^M$.

Between 1998 and 2002 we developed a series of algorithms suited to this problem.[1] The new methods have since been used in unpublished work to analyze a variety of synthetic and real-world data types.* The root algorithms are related to generalizations of Mallat and Zhang's method of Matching Pursuits (MP) introduced in 1993.[2,3] Sparse approximation has become a topic of budding interest in harmonic analysis, and recently Tropp, Gilbert, and Strauss[4] have reported a simultaneous approximation algorithm based on a generalization of orthogonal MP. It has also recently come to our attention that Temlyakov and Leviatan independently obtained results for a number of related greedy methods.[5] These authors have taken a different tack from the specific methods we describe here.

We present here an extended simultaneous MP algorithm together with an outline of analysis methods by which discriminating feature components may be extracted from collections of data. We refer to the process as Greedy Adaptive Discrimination (GAD). The purpose of this paper is to illustrate the usefulness of the methods with non-trivial synthesized numerical signal data, and to compare several variations on the method to each other and to ordinary MP. Those interested in the associated theory, in both infinite and finite dimensional Hilbert spaces, and in methods of analyzing additional classes of data, we refer to [6].

2. SIMULTANEOUS MATCHING PURSUITS

In the interest of conserving space, we shall not review MP here, except to recall certain definitions. The proposed generalization of MP, termed *Simultaneous Matching Pursuits* (SMP), may be written as follows:

Definition 1 (Simultaneous Matching Pursuits).
Let $\left\{f^i\right\} \subset H$ *be indexed by* $i \in s$, *with* $|s| < \infty$.
Let $\mathcal{D} = \{g_\gamma\}_{\gamma \in \Gamma} \subset H$ *and* $\alpha \in (0, 1]$.

1. *Initialize.*
   Define $R^0 f^i \equiv f^i$.
   Set $n = 0$.

2. *Choose dictionary atom for "almost" best match.*
   Find $g_{\gamma_n} \in \mathcal{D}$ *such that, for some* $i \in s$,

$$\left|\langle R^n f^i, g_{\gamma_n}\rangle\right| \geq \alpha \sup_{\gamma \in \Gamma} \left|\langle R^n f^i, g_\gamma\rangle\right|. \tag{5}$$

3. *Project* $R^n f^i$ *on* g_{γ_n} *to find* $(n+1)^{th}$ *residue for all* $i \in s$ *by*

$$R^n f^i = \langle R^n f^i, g_{\gamma_n}\rangle g_{\gamma_n} + R^{n+1} f^i. \tag{6}$$

*Methods and certain applications disclosed herein are the subject of several US and international pending patent applications. Please contact the first author for more information.

4. Repeat steps 2 and 3 for $(n+1)^{th}$ residue, etc.

SMP in the case of $card(\{f^i\}) = |s| = 1$ is precisely MP as defined by Mallat and Zhang. To show convergence they adapted a proof by Jones from the context of projection pursuit regressions.[7] We note, however, that in step 2, the trivial extension of MP would require that equation (5) hold for all $i \in s$; however, we weaken the requirement to hold only for *some* $i \in s$. The actual choice function used to satisfy step 2 will yield convergence for all $i \in s$. We shall address this shortly. Constraints for the general algorithm are discussed in [6].

The algorithm selects decomposition atoms stepwise from $\mathcal{D}$. By summing the recursion (6) over $n = [0, 1, ..., m-1]$ one may obtain at the m^{th} step,

$$f^i = \underbrace{\sum_{n=0}^{m-1} \langle R^n f^i, g_{\gamma_n} \rangle g_{\gamma_n}}_{\text{approximation}} + \underbrace{R^m f^i}_{\text{error}}. \tag{7}$$

Assuming convergence, then, when $\mathcal{D}$ is complete in H it can be shown that for all $i \in s$,

$$f^i = \sum_{n=0}^{+\infty} \langle R^n f^i, g_{\gamma_n} \rangle g_{\gamma_n} \tag{8}$$

and

$$\left\| f^i \right\|^2 = \sum_{n=0}^{+\infty} \left| \langle R^n f^i, g_{\gamma_n} \rangle \right|^2. \tag{9}$$

Equation (8) comprises a collection of simultaneous representations of the the signals $\{f^i\}$. Equation (7) comprises a simultaneous sparse approximation. Equivalently, we may represent the signals $\{f^i\}$ with a generalized *structure book* by indexing the stepwise inner products with $i \in s$, writing

$$\left\{ \left\{ \langle R^n f^i, g_{\gamma_n} \rangle \right\}_{i \in s}, \gamma_n \right\}_{n \in \mathbb{N}}. \tag{10}$$

Since each signal f^i is expanded over the same atoms g_{γ_n}, we need not store the choices γ_n more than once.

Now let us consider operation in a finite dimensional Hilbert space H as might be realized in actual computational signal analysis. Let $\mathcal{D}$ be complete in H. We further assume that $\{f^i\}$ is finite and may be indexed by $i \in s \subset \mathbb{N}$. Let $M = |s| = card(\{f^i\})$, so that $s = \{1, 2, ..., M\}$.

One defines the *correlation ratio* of $f \in H$ with respect to a dictionary $\mathcal{D}$ as,

$$\lambda(f) = \frac{\sup_{\gamma \in \Gamma} |\langle f, g_\gamma \rangle|}{\|f\|} \leq 1. \tag{11}$$

Mallat and Zhang showed the following lemma with respect to the correlation ratio.

Lemma 1 (Mallat and Zhang). *Let $\mathcal{D}$ be a complete dictonary in finite dimensional Hilbert space H. Then*

$$I(\lambda) = \inf_{f \in H} \lambda(f) > 0. \tag{12}$$

Mallat and Zhang also showed that we may define a subdictionary $\mathcal{D}_\alpha = \{g_\gamma\}_{\gamma \in \Gamma_\alpha} \subset \mathcal{D}$, for which there exists $\alpha \in (0, 1]$ and $\Gamma_\alpha \subset \Gamma$, a finite subset of indices, such that for any $f \in H$,

$$\sup_{\gamma \in \Gamma_\alpha} |\langle f, g_\gamma \rangle| \geq \alpha \sup_{\gamma \in \Gamma} |\langle f, g_\gamma \rangle|. \tag{13}$$

In our numerical examples, we shall utilize a finite dictionary of Gabor atoms that has this property.

In order to establish an explicit choice function for the algorithm, we define a scalar valued cost function based on a vector p-norm taken over the set of inner products of the stepwise residues with a given dictionary atom.

Definition 2 (P-NORM).
Given $n \in \mathbb{N}$ and $\gamma \in \Gamma$. For each vector $\{\langle R^n f^i, g_\gamma \rangle\}_{i \in s} = (\langle R^n f^1, g_\gamma \rangle, \langle R^n f^2, g_\gamma \rangle, ..., \langle R^n f^{|s|}, g_\gamma \rangle) \in \mathbb{C}^{|s|}$, let the p-norm *for positive integer p be defined by*

$$\left\|\{\langle R^n f^i, g_\gamma \rangle\}\right\|_p = \Big(\sum_{i \in s} \left|\langle R^n f^i, g_\gamma \rangle\right|^p\Big)^{\frac{1}{p}}. \tag{14}$$

We now define our choice function at the n^{th} step by the following operations:

Definition 3 (SMP Choice Function).

1. *Find $g_{\widetilde{\gamma}_n} \in \mathcal{D}_\alpha$ such that,*
$$\left\|\{\langle R^n f^i, g_{\widetilde{\gamma}_n} \rangle\}\right\|_p = \sup_{\gamma \in \Gamma_\alpha} \left\|\{\langle R^n f^i, g_\gamma \rangle\}\right\|_p; \tag{15}$$
2. *Search in a neighborhood $N_{g_{\widetilde{\gamma}_n}}$ of $g_{\widetilde{\gamma}_n}$ for $g_{\gamma_n} \in \mathcal{D}$ where $\left\|\{\langle R^n f^i, g_{\gamma_n} \rangle\}\right\|_p$ reaches a local maximum.*

Γ_α is finite, hence $\widetilde{\gamma}_n$ exists. In our examples we improve this choice by searching a well-defined neighborhood $N_{g_{\widetilde{\gamma}_n}}$ around $g_{\widetilde{\gamma}_n}$. The search may also be accomplished without making the neighborhood explicit by, for example, Newton's Method.

We state without proof the following result, which shows that if we utilize this choice function in the SMP algorithm then convergence is exponential, modulo $M = |s|$ the cardinality of $\{f^i\}$.

Theorem 1 (Exponential Convergence). *Let $\{f^i\}_{i \in s} \subset H$ be a finite subset of a finite dimensional Hilbert space H, with cardinality of $\{f^i\} = |s| = M < \infty$. Then for all $i \in s$ the norms of the residues $\left\|R^n f^i\right\|$ converge to 0 as $n \to \infty$. Furthermore, there exists $i_0 \in s$ such that, for all $k, n > 0$ with $n \geq kM$, the following expression holds uniformly for all $i \in s$:*

$$\left\|R^n f^i\right\| \leq \frac{M^{\frac{1}{p}}}{\alpha I(\lambda)} \left\|f^{i_0}\right\| \left(1 - \frac{\alpha^2 I^2(\lambda)}{M^{\frac{1}{p}}}\right)^{\frac{k}{2}}. \tag{16}$$

For full treatment of this result and its derivation see [6] .

Tropp, Gilbert and Strauss' recent work has focused on the 1-norm, for which they have obtained interesting results in the context of orthogonal MP.[4] We shall utilize a 2-norm in our examples below and compare it to a 1-norm and ∞-norm. Our results are general permitting the choice of p-norm to be used to tune the algorithm to a given application. Moreover, we do not require the stronger condition of an orthogonal pursuit.

Step one of the SMP Choice Function is sufficient for the algorithm to converge under theorem 1, while step two further optimizes the choice locally and is technically optional. The two-part construction of the choice function parallels that of Mallat and Zhang in MP. We note that many authors using classical MP in practical applications omit the optimization step since it slows execution and sometimes yields only small gains; however, we shall instead take advantage of the optional optimization step to reshape the algorithm to better treat certain classes of degraded data.

We define our improved choice function at the n^{th} step by the following operations:

Definition 4 (GAD Choice Function).

1. *Find $g_{\widetilde{\gamma}_n} \in \mathcal{D}_\alpha$ such that,*
$$\left\|\{\langle R^n f^i, g_{\widetilde{\gamma}_n} \rangle\}\right\|_p = \sup_{\gamma \in \Gamma_\alpha} \left\|\{\langle R^n f^i, g_\gamma \rangle\}\right\|_p; \tag{17}$$
2. *For each $i \in s$, search in a neighborhood $N_{g_{\widetilde{\gamma}_n}}$ of $g_{\widetilde{\gamma}_n}$ for $g_{\gamma_n^i} \in \mathcal{D}$ for which $\left|\langle R^n f^i, g_{\gamma_n^i} \rangle\right|$ reaches a local maximum.*

Correspondingly, we replace the update step relationship (6) with

$$R^n f^i = \langle R^n f^i, g_{\gamma_n^i} \rangle g_{\gamma_n^i} + R^{n+1} f^i. \tag{18}$$

While the SMP Choice Function is sufficient for interesting theoretical results, it suffers from the assumption that strictly identical atoms will be found in each signal f^i. In nearly any real data set, however, we can expect noise, jitter, interference, or other unknown shifts and/or masking in the underlying signal components to be present. The GAD Choice Function allows the algorithm to compensate for these variations by introducing equivalence classes into the simultaneous representation spaces. We shall utilize this property in our numerical examples to effectively de-blur noisy collections of data to deduce properties of the source signal.

In order to store the resulting representation, we further generalized the structure book definition by indexing the atoms as well as the coefficients with $i \in s$, writing

$$\left\{ \left\{ \langle R^n f^i, g_{\gamma_n^i} \rangle \right\}_{i \in s}, \left\{ \gamma_n^i \right\}_{i \in s} \right\} n \in \mathbb{N}. \tag{19}$$

Each subset $\left\{ g_{\gamma_n^i} \right\}_{i \in s}$ comprises an equivalence class of similar atoms.

3. EGAD ANALYSIS

As we shall see numerically, the simultaneous treatment of groups of similar signals using SMP offers a great deal of power in resolving common components. The addition of GAD equivalence classes allows us to better extract commonalities from collections of signals without introducing ghosting or blur through averaging. However, in order to recover estimates of the source components from collections of signals we must reduce each equivalence class to a best estimate of the generating atom.

To accomplish this, we utilize a parametric dictionary.

Definition 5 (Gabor Dictionary). *Let $g(t) = 2^{\frac{1}{4}} e^{-\pi t^2}$ be the Gaussian window function. Let $\gamma = (s, u, \xi)$. The elements of the Gabor Dictionary are those functions $\mathcal{D} = \{g_\gamma(t)\}_{\gamma \in \Gamma} \subset L^2(\mathbb{R})$, defined by*

$$g_\gamma(t) = \frac{1}{\sqrt{s}} g\left(\frac{t-u}{s} \right) e^{i\xi t}. \tag{20}$$

The Gabor dictionary is parameterized by an ordered triplet $\gamma = (s, u, \xi)$, corresponding to scale, position, and modulation frequency. Following Mallat and Zhang, to this dictionary we add the delta functions and the Fourier basis in order to cover the limit cases of small and large scale. We describe the dictionary in continuous form for convenience of explanation. In finite dimensional implementation we uniformly sample and periodize the window function g. We may also recast the dictionary in $\mathbb{R}$ rather than $\mathbb{C}$ by introducing an additional phase parameter ϕ. A finite subdictionary $\mathcal{D}_\alpha$ may be defined by sampling $\mathcal{D}$ under guidance of the classical uncertainty principle. See [2] for details of this dictionary and subdictionary construction.

The Gabor dictionary has useful properties for unbiased signal analysis since the generating Gaussian window is well localized in the time-frequency plane. However, the important aspect of this dictionary for our purposes is that the generated atoms behave smoothly under parametric variation. Hence, it becomes meaningful to take averages in these parameters over groups of signals.

We introduce a mean operation in this *parameter space* that generates a new structure book of the form $\{\overline{a}_n, \overline{\gamma}_n\}_{n \in \mathbb{N}}$ from a collection or sub-collection of GAD analyzed signals.

Definition 6 (Parametric Mean). *Let $M = |s|$ be the cardinality of a group of signals. For each n, create a* mean structure book *entry $(\overline{a}_n, \overline{\gamma}_n)$ where $\overline{\gamma}_n = (\overline{s}_n, \overline{u}_n, \overline{\xi}_n, \overline{\phi_n})$ and*

$$\begin{aligned} \overline{a}_n &= \frac{1}{M} \sum_i \langle R^n f^i, g_{\gamma_n^i} \rangle \\ \overline{s}_n &= \frac{1}{M} \sum_i s_n^i \\ \overline{u}_n &= \frac{1}{M} \sum_i u_n^i \end{aligned} \tag{21}$$

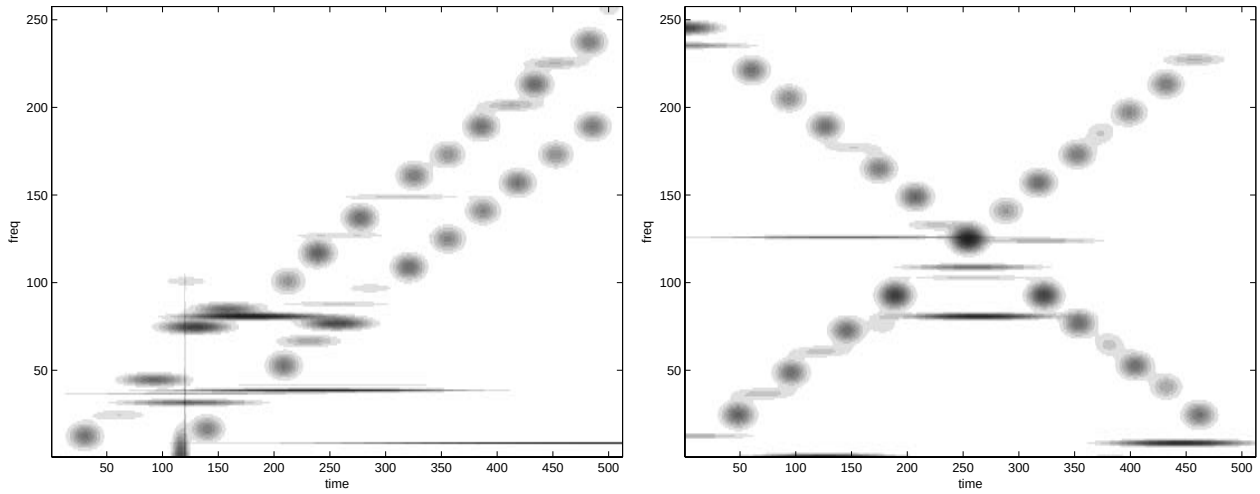

Figure 1. Wigner energy density plots of atoms recovered in GAD analysis of two noisy, jittered data sets using 2-norm choice function and $\Delta_u = \pm 15$.

$$\begin{aligned} \overline{\xi}_n &= \frac{1}{M}\sum_i \xi_n^i \\ \overline{\phi}_n &= \frac{1}{M}\sum_i \phi_n^i. \end{aligned}$$

Other meaningful averages may be defined and are discussed in [6]. We have shown phase ϕ explicitly since our numerical examples will deal with real-valued signals.

The parametric mean offers a method of estimating source values from the scattered parameters of the atoms within an equivalence class. We term the method of estimating parameters of a source signal by analyzing the parameter space of atoms in a GAD equivalence class *estimated parametric GAD*, or eGAD.

4. NUMERICAL EXAMPLES

In order to illustrate the usefulness of the GAD analysis, we synthesized two complex data sets. A pair of model signals was constructed consisting of chirps, truncated sine waves, and transients. The first model signal comprised a linear rising chirp spanning the 512 point time window; an additional chirp offset by 100 time points from the first; an 8 cycle sinusoid; a 37.5 cycle truncated sinusoid spanning points 50-384; an 80 cycle truncated sinusoid spanning points 92-256; and a small transient created by sampling a Daubechies' 6 wavelet in a 12 point window, beginning at position 117. The second model signal comprised the same linear rising chirp; a descending chirp with the same properties reversed in time; a truncated 8 cycle sinusoid spanning points 384-512; a truncated 80 cycle sinusoid spanning points 192-333; a truncated 125 cycle sinusoid spanning points 150-384; and an unmodulated 128-scale gaussian centered on point 128.

From each model signal an ensemble of 10 independent data vectors was generated by adding gaussian noise ($\sigma = .5$) and jittering randomly (with uniform distribution) in time by up to ± 12 points. The components of the model signals were of unit amplitude with the exception of the transient which was amplified by 10. Our intention was to simulate real-world data in which signals are often obscured in noise and only weakly temporally correlated.

An eGAD type SMP analysis was performed on each ensemble of data using a 2-norm GAD choice function that allowed for a variation of up to $\Delta_u = \pm 15$ points in the position of each atom across data vectors. Figure 1 shows the Wigner energy distribution[8,2] of the atoms selected by the analysis. For display within printable

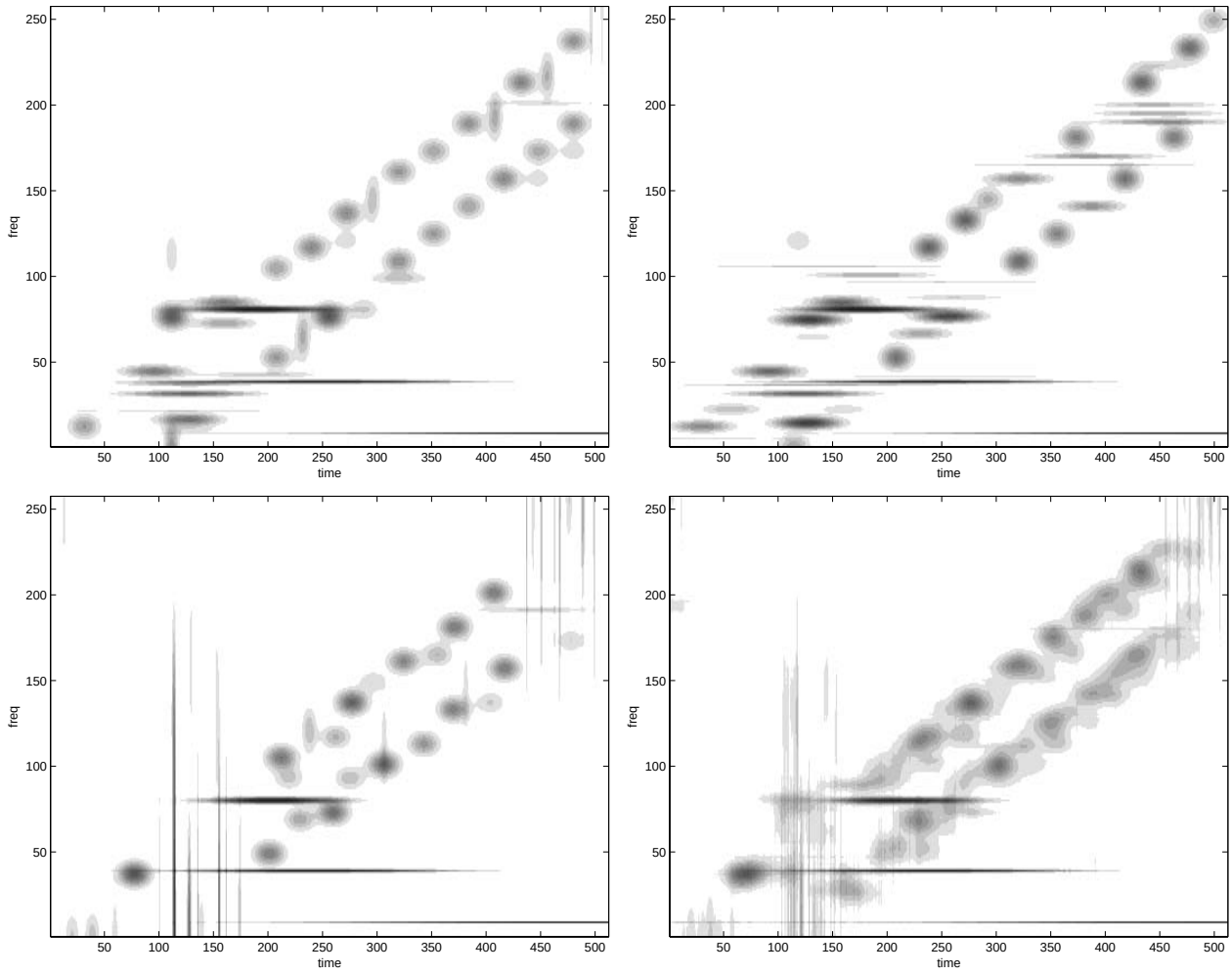

Figure 2. Wigner energy density of atoms selected in analysis of first dataset using 2-norm with $\Delta_u = 0$ (top-left), 1-norm with $\Delta_u = \pm 15$ (top-right), ∞-norm with $\Delta_u = \pm 15$ (bottom-left), ordinary MP averaged in T-F plane (bottom-right).

greyscale levels, the amplitude range of all plots was compressed by clipping the largest amplitudes and was decluttered by thresholding those below the noise floor.

In analysis of the first dataset both chirps are modelled fairly well with Gabor atoms in a stepwise fashion. The 8 cycle sinusoid is somewhat obscured early in time where it is crossed by the chirps. The 37.5 cycle sinusoid is captured reasonably with a Guassian-windowed approximation, as is the 80 cycle sinusoid whose end points are obscured by overlap with the chirps. The main peak of the transient is visible as well, with some indication of the extent of its brief, rapid oscillation.

In analysis of the second dataset both chirps are again modelled fairly well within the capacity of the dictionary. The truncated sinusoids are well modelled, though the 125 cycle feature is broken up by interaction with the crossing chirps. The unmodulated Gaussian bump is captured as well.

To examine the performance of other variations on the SMP analysis, we reanalyzed the first data set using first a 2-norm SMP choice function with no improvement search, then a 1-norm GAD choice function with $\Delta_u = \pm 15$, and last an ∞-norm GAD choice function with $\Delta_u = \pm 15$. Finally we performed individual MP on each vector in the ensemble and averaged the Wigner distributions in the T-F plane to produce a comparison plot. The results are shown in figure 2.

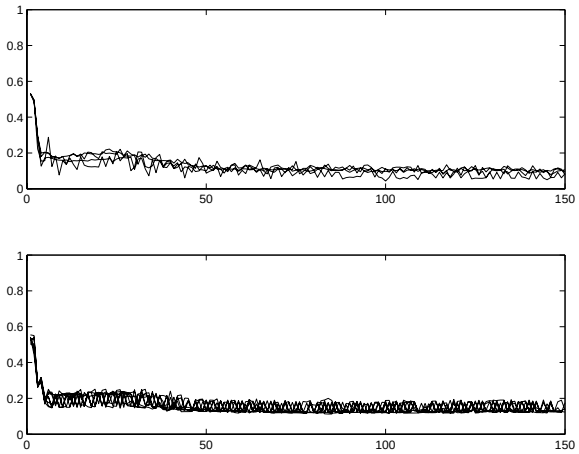

Figure 3. Mean ensemble correlation ratio of stepwise residues for GAD analysis by each method (top) and correlation ratio of stepwise residues for MP analysis of individual signals in ensemble (bottom) for first 150 steps.

The 2-norm choice function without the GAD improvement step (top-left) captures the structure but the organization is clearly not as well-resolved as in figure 1. The jitter is apparent in the chirp approximations, and the transient is resolved only weakly by its mid-frequency components. Contrasting with figure 1, it is clear that the GAD improvement search step not only de-blurs the approximation overall, but also enables the resolution of the time-localized transient.

The 1-norm choice function (top-right) appears to do a poor job with jitter even with the GAD improvement step implemented. Mid- and high-frequency components of the two chirps are merged in several places, the chirp steps are more irregular, and the transient is obscured. Nonetheless, the GAD improvement operation has resulted in reasonably linear organization of the chirp elements. We note that in other simulations using ensembles with less extreme jitter, the 1-norm performed very similarly to the 2-norm, with both showing reasonably good resolution and organization.

The ∞-norm choice function (bottom-left) performs poorly. T-F areas of high complexity are not modelled well. The transient-chirp interaction region is not accurate, and the high frequency chirp regions seems to be dominated by interaction with noise components. In addition, the chirps are not clearly linear. The ∞-norm is the most greedy norm. It tends to favor the best fit to any available signal and does not appear to be a good choice in this application. Where we wish to balance the fit to the ensemble rather than closely match individual signals, the 2-norm does a much better job.

The last plot (bottom-right) was produced by performing MP on individual signals and then averaging the Wigner energy distributions of the resulting atoms in the T-F plane. This method has become a standard tool for those doing signal analysis with MP. We see that the plot captures the sinusoids reasonably well, as well as the mid sections of the linear chirps. Both the transient region and upper portion of the chirps are obscured by noise features. The lower portion of the first chirp seems to be poorly captured as well. In the MP plots favor, the averaging naturally suppresses some of the spurious atoms that must be thresholded in the GAD plots. The blurring effect also helps blend the discrete atoms of the chirps so that their smooth linearity can be inferred, but, conversely, it would be hard to characterize their exact spread from the plot.

Figure 3 shows the convergence behavior of the two classes of analysis methods over the first 150 steps. The top plot illustrates the mean correlation ratio of the stepwise residues for the four SMP methods; closer analysis shows that all four variations converged to near the noise floor within less than 50 steps. The bottom plot illustrates the same measure for each of the individual MP analysis overlaid. MP converged much slower. The simultaneous algorithm had ten examples of each signal to work with at each iteration while the individual MP analysis only had one. This multiple sampling of the target data enables the GAD algorithm to rapidly distinguish signal from noise.

5. CONCLUSIONS

Redundancy of information speeds convergence for SMP methods when compared to classical MP. In addition, the use of eGAD equivalence classes both de-blurs the recovered approximation and enables resolution of a time-localized transient that is otherwise obscured by jitter. Empirically, the 2-norm choice function outperformed both the 1-norm and ∞-norm in approximating the underlying model signal. In this instance, the 2-norm represents a successful compromise between greedy approximation of individual data vectors and oversimplifying the signal structure by choosing too few signal atoms to represent the group.

In the context of noise and jitter in the simulated data, the 2-norm GAD choice function provides the best performance. However, the blurring effect of the averaged MP plot has the advantage of filling in the T-F gaps so that the characteristics of the dictionary are less dominant. Based on the Wigner plots alone, the preferred analysis will depend on application.

The most significant difference between the GAD results and the final MP average, however, is the dimensionality of the representation. All of the GAD parametric mean signals displayed here are modelled in less than 50 atoms. The MP T-F average requires thousands of points to capture the same or less information.

This low dimensional representation of the source signal provides other advantages. Data is stored compactly and parameters of the source components can be analyzed directly simply by examining the numerical values. For example, rather than estimating the frequency of a sinusoid by examining the Wigner plot as one would with ordinary MP averaging, we can instead simply examine the parametric mean structure book and read off the value of the parameter.

In the numerical simulation, the 2-norm, $\Delta_u = \pm 15$ GAD analysis of the first dataset estimated the frequencies of the 8, 37.5, and 80 cycle sinusoids precisely. While it located the transient peak in time at 116.5 with a width of 16. These are the first four atoms that were selected by the analysis.

In our examples, we deliberately choose certain model components because they could not be simply modelled by the selected dictionary in order to challenge the algorithm. Clearly other dictionaries may be used to examine other parameters or to further optimize the representation of a given signal class.

GAD methods have additional application beyond the scope of this paper. The theoretical justification for the claims in these particular examples is the basis for our joint work [6], where the general theory yields analogous behavior for a large class of similar phenomena as well as establishing the reconstruction theory and algorithms.

REFERENCES

1. J. M. Sieracki, *Greedy Adaptive Discrimination: Signal component analysis by simultaneous matching pursuits with application to ECoG signature detection.*, PhD dissertation, University of Maryland, College Park, 2002.
2. S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing* **41**, pp. 3397–3415, Dec. 1993.
3. G. M. Davis, S. Mallat, and M. Avelanedo, "Greedy adaptive approximations," *J. of Constr. Approx.* **13**, pp. 57–98, 1997.
4. J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Simultaneous sparse approximation via greedy pursuit," *Proc. ICASP, Philidelphia* , 2005.
5. D. Leviatan and V.N. Temlyakov, "Simultanious approximations by greedy algorithms," *IMI Research Report* **2003:02**, Univ. of South Carolina at Columbia, 2003.
6. J. M. Sieracki and J. J. Benedetto, "Greedy adaptive discrimination by simultaneous matching pursuits," *In preparation* .
7. L. K. Jones, "On a conjecture of Huber concerning the convergence of projection pursuit regression," *Ann. Statist.* **15**(2), pp. 880–882, 1987.
8. L. Cohen, "Time-frequency distributions – A review," *Proc. IEEE* **77**, pp. 941–981, 1989.