



**AALBORG UNIVERSITY**  
DENMARK

**Aalborg Universitet**

## **Sparse Approximation and the Pursuit of Meaningful Signal Models with Interference Adaptation**

Sturm, Bob L.; Shynk, John

*Published in:*  
IEEE Transactions on Audio Speech and Language Processing

*DOI (link to publication from Publisher):*  
[10.1109/TASL.2009.2037395](https://doi.org/10.1109/TASL.2009.2037395)

*Publication date:*  
2010

*Document Version*  
Early version, also known as pre-print

[Link to publication from Aalborg University](#)

*Citation for published version (APA):*  
Sturm, B. L., & Shynk, J. (2010). Sparse Approximation and the Pursuit of Meaningful Signal Models with Interference Adaptation. *IEEE Transactions on Audio Speech and Language Processing*, 18(3), 461-472. <https://doi.org/10.1109/TASL.2009.2037395>

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

### **Take down policy**

If you believe that this document breaches copyright please contact us at [vbn@aub.aau.dk](mailto:vbn@aub.aau.dk) providing details, and we will remove access to the work immediately and investigate your claim.

# Sparse Approximation and the Pursuit of Meaningful Signal Models with Interference Adaptation

Bob L. Sturm, *Member, IEEE*, John J. Shynk\*, *Senior Member, IEEE*

## Abstract

In the pursuit of a sparse signal model, mismatches between the signal and the dictionary, as well as atoms poorly selected by the decomposition process, can diminish the efficiency and meaningfulness of the resulting representation. These problems increase the number of atoms needed to model a signal for a given error, and they obscure the relationships between signal content and the elements of the model. To increase the efficiency and meaningfulness of a signal model built by an iterative descent pursuit, such as matching pursuit (MP), we propose integrating into its atom selection criterion a measure of *interference* between an atom and the model. We define interference and illustrate how it describes the contribution of an atom to modeling a signal. We show that for any nontrivial signal, the convergent model created by MP must have as much destructive as constructive interference, i.e., MP cannot avoid correction in the signal model. This is not necessarily a shortcoming of orthogonal variants of MP, such as orthogonal MP (OMP). We derive interference-adaptive iterative descent pursuits and show how these can build signal models that better fit the signal locally, and reduce the corrections made in a signal model. Compared with MP and its orthogonal variants, our experimental results not only show an increase in model efficiency, but also a clearer correspondence between the signal and the atoms of a representation.

B. L. Sturm is with Université Pierre et Marie Curie, UPMC Paris 06, Institut Jean Le Rond d'Alembert (IJLRA), UMR 7910, Équipe Lutheries, Acoustique, Musique (LAM), 11, rue de Lourmel, 75015 Paris, France. E-mail: boblsturm@gmail.com. He was with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA, 93106-9560 USA. J. J. Shynk (\*contact author) is with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA, 93106-9560 USA. E-mail: shynk@ece.ucsb.edu. Phone: 805-893-3977 Fax: 805-893-3262

This work was supported in part by the National Science Foundation under Grant CCF-0729229, and in part by Bourse Chateaubriand N. 634146B.

Manuscript submitted Dec. 19, 2008. Revision submitted Sep. 21, 2009.

## Index Terms

Sparse approximation, orthogonal matching pursuit, signal decomposition, overcomplete dictionary. EDICS:  
AUD-ANSY

## I. INTRODUCTION

Sparse approximation emphasizes the importance of describing a signal using few terms selected from an overcomplete set (*dictionary*) of functions (*atoms*) [1]–[5]. The goal is to construct the  $n$ th-order model of a signal  $\mathbf{x} \in \mathbb{C}^K$  ( $n \ll K$ )

$$\mathbf{x} = \sum_{i=0}^{n-1} a_i \mathbf{h}_i + \mathbf{r}(n) = \mathbf{H}(n) \mathbf{a}(n) + \mathbf{r}(n) \quad (1)$$

where  $\mathbf{H}(n) \triangleq [\mathbf{h}_0 | \mathbf{h}_1 | \cdots | \mathbf{h}_{n-1}]$  is a matrix of  $n$  atoms selected from a dictionary  $\mathcal{D}_N \triangleq \{\mathbf{d}_i \in \mathbb{C}^K : \|\mathbf{d}_i\|_2 = 1\}_{i=1}^N$  of  $N \gg K$  atoms (unless otherwise stated, we assume  $\text{span}\{\mathcal{D}_N\} = \mathbb{C}^K$ ),  $\mathbf{a}(n)$  is a column vector of  $n$  weights, and  $\mathbf{r}(n)$  is the  $n$ th-order residual signal. (The notation  $[\cdot | \cdot]$  denotes a concatenation of column vectors into a matrix.) One criterion with which to build this model is the minimization of a residual (error) subject to a constraint on the number of terms used:

$$\min_{\mathcal{H}_n \subset \mathcal{D}_N} \|\mathbf{x} - \mathbf{H}(n) \mathbf{a}(n)\|_2^2 \text{ subject to } n \ll K \quad (2)$$

where  $\mathcal{H}_n$  is a subset of  $n$  atoms from  $\mathcal{D}_N$ . In other words, find a signal model with small error using a number of terms much smaller than the dimension of the signal. Solving this problem is combinatorially complex in general because it involves examining many  $n$ -tuples of  $\mathcal{D}_N$ . For this reason, other more tractable methods have been proposed to find suboptimal but useful solutions to (2). These include convex optimization principles such as basis pursuit (BP) [3], and greedy iterative descent methods such as matching pursuit (MP) [1], orthogonal MP (OMP) [6] (alluded to in [1]), and optimized OMP (OOMP) [7] (previously presented as orthogonal least squares [8], as well as MP with pre-fitting [9]).

Consider the signals shown as insets in Fig. 1, each of which has been decomposed using MP, OMP, and OOMP, and an overcomplete dictionary of Gabor atoms (modulated Gaussian windows) and Dirac spikes (unit impulses). Each plot shows the residual energy decay as a function of the pursuit iteration. Of the three pursuits, OOMP produces models with the smallest error (the norm in (2)) in nearly all iterations. For this reason we claim that for each of these signals, the models produced by OOMP are more *efficient* than the others because they use fewer atoms to reach the same level of error. What is not evident from the residual energy decay, however, is that a model can contain “artifacts” — roughly,

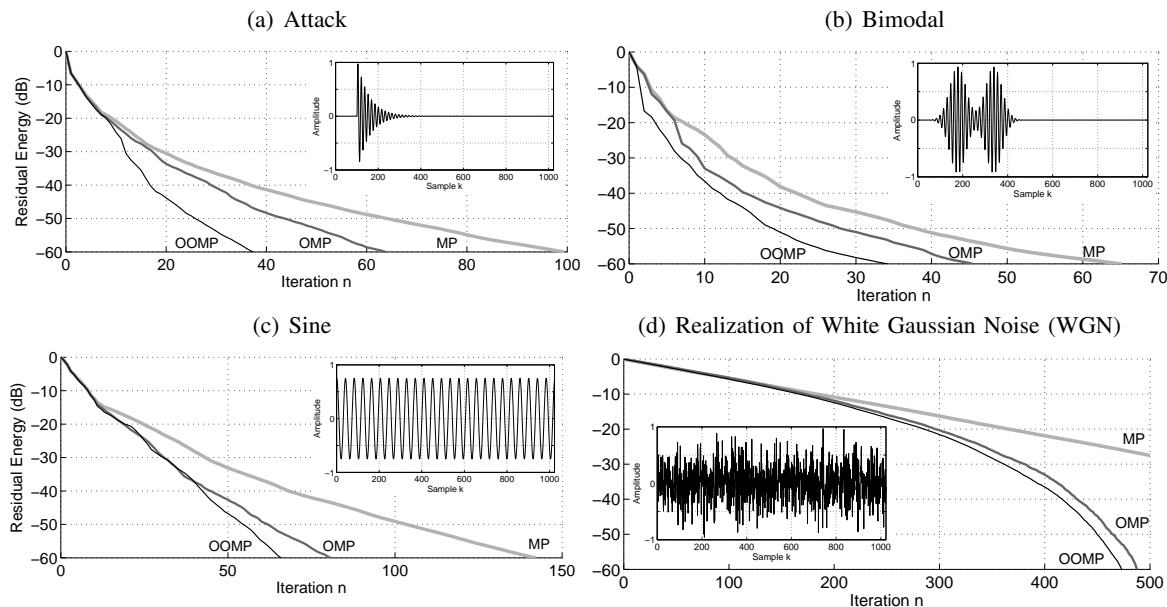


Fig. 1. Normalized residual energy as a function of pursuit iteration for four signals (inset) using MP, OMP, and OOMP (labeled). The dictionary is a union of Gabor atoms and Dirac spikes.

atoms placed where they should not be — due to (i) the process of decomposition in the pursuit, and (ii) mismatches between the model and the signal [10]–[17].

These artifacts become readily apparent when observing the *wivigram* of a representation, which is a superposition of each atom’s Wigner-Ville time-frequency distribution [1], [18]. For example, the model of Attack created by OMP shown in Fig. 2(a) has many atoms over a time where the data is zero. Since these atoms serve only to correct the signal model, we claim that they are of little use for describing the signal, and thus carry little *meaning* about the signal. Figure 2(b) reveals that OOMP models Bimodal with one large-scale atom, and a smaller-scale atom removes the large overshoot introduced to the residual by subtraction of the first atom. Arguably, this model is less efficient and meaningful than one using two Gabor atoms with scales the size of each individual mode. These two examples clearly show that atoms can exist solely to correct the errors introduced by other atoms. This behavior diminishes the efficiency and meaningfulness of a signal model, and consequently limits its use in any application requiring clear and accurate correspondences between the signal and its model.

Others have noted and addressed these problems of iterative descent approaches to sparse approximation, but with solutions that increase the computational complexity of the decomposition or constrain the choice of the dictionary. The high-resolution pursuit algorithm in [10], [11] specifically attempts to address “non-features” introduced by MP by proposing an atom selection criterion based upon minimizing the

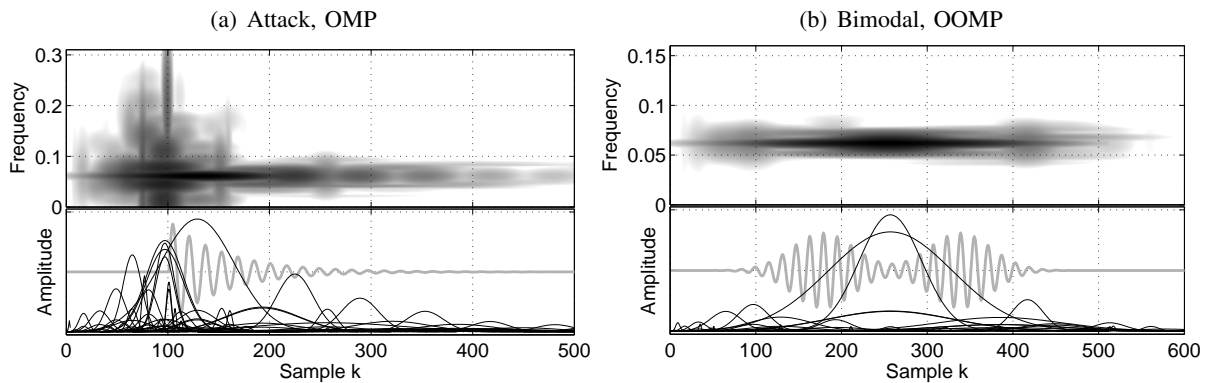


Fig. 2. Top: wivigrams (superposition of the Wigner-Ville time-frequency energy distributions) of atoms in the model created from the dictionary used in Fig. 1. Bottom: time-domain envelopes of atoms (black) overlaid onto original signal (thick gray). (Height of atom envelope is proportional to the square root of its energy in the model.)

worst fit of smaller-scale atoms — which decreases the frequency resolution of the pursuit, and requires specially constructed dictionaries with each atom being a linear combination of smaller atoms (e.g., B-splines). Other approaches construct dictionaries with elements that are more similar to the signal content of interest, for example, modeling asymmetric structures with damped sinusoids [12]. And molecular MP (MMP) [19] considers two different dictionaries for the transient and tonal portions of audio signals. A completely different approach in [13] averages the results of many pursuits of the same signal to facilitate its accurate characterization for analysis; but this does not provide a solution to (2). With all these methods, however, the same problems of correction and model mismatch can occur, though to different extents. In this paper, we address the problem of detecting and avoiding such errors in the iterative descent pursuit of a signal model, without imposing constraints on the dictionary, and while maintaining the simple iterative structure of MP, OMP, and OOMP [17]. Our resulting signal models show an increase in efficiency — fewer atoms required to reach a given error — and meaningfulness — clearer correspondence between features of a signal and its model — especially for models created by OMP and OOMP.

The rest of this paper is organized as follows. In Section II, we present some notation and briefly review the methods of MP, OMP, and OOMP. In Section III, we formally define the concept of *interference*, and demonstrate how it reveals useful information about a signal model, as well as the dictionary and the decomposition process. This provides a means to gauge how an atom contributes to modeling a signal (and its meaningfulness), as well as the performance of a pursuit both locally and globally. Finally, we prove that any convergent signal model created by MP has as much destructive as constructive interference

(unless the signal is a trivial composition of nonoverlapping atoms from the dictionary). In Section IV, we describe how to integrate interference into the atom selection criteria of MP, OMP, and OOMP to create *interference-adaptive iterative descent pursuits*. Several computer simulations are provided in Section V. Finally, Section VI summarizes the conclusions of this paper and ongoing work.

## II. NOTATION AND REVIEW OF PURSUITS

For two vectors in the  $K$ -dimensional complex vector space  $\mathbf{x}, \mathbf{y} \in \mathbb{C}^K$ , we use the following notation to denote an inner product:  $\langle \mathbf{x}, \mathbf{y} \rangle \triangleq \mathbf{y}^H \mathbf{x}$ , where  $^H$  is conjugate transpose. All vectors are column vectors and are denoted using lower-case bold roman letters;  $[\mathbf{w}]_i$  refers to the  $i$ th element of  $\mathbf{w}$ . All matrices are denoted with upper-case bold roman letters; the  $i$ th element in the  $j$ th column of  $\mathbf{A}$  is denoted by  $[\mathbf{A}]_{ij}$ . Finally, the  $\ell_p$ -norm of  $\mathbf{x} \in \mathbb{C}^K$  for  $p \geq 1$  is  $\|\mathbf{x}\|_p^p \triangleq \sum_{i=1}^K |[\mathbf{x}]_i|^p$ , and thus the  $\ell_2$ -norm is given by  $\|\mathbf{x}\|_2^2 = \sum_{i=1}^K |[\mathbf{x}]_i|^2 = \langle \mathbf{x}, \mathbf{x} \rangle = \mathbf{x}^H \mathbf{x}$ .

The  $n$ th-order representation of the signal  $\mathbf{x} \in \mathbb{C}^K$  is denoted by  $\mathcal{X}_n = \{\mathbf{H}(n), \mathbf{a}(n), \mathbf{r}(n)\}$ , which produces the  $n$ th-order model in (1). MP [1] updates  $\mathcal{X}_n$  according to:

$$\mathcal{X}_{n+1} = \left\{ \begin{array}{l} \mathbf{H}(n+1) = [\mathbf{H}(n)|\mathbf{h}_n], \\ \mathbf{a}(n+1) = [\mathbf{a}^T(n)|\langle \mathbf{r}(n), \mathbf{h}_n \rangle]^T, \\ \mathbf{r}(n+1) = \mathbf{x} - \mathbf{H}(n+1)\mathbf{a}(n+1) \end{array} \right\} \quad (3)$$

using the atom selection criterion

$$\mathbf{h}_n = \arg \min_{\mathbf{d} \in \mathcal{D}_N} \|\mathbf{r}(n) - \langle \mathbf{r}(n), \mathbf{d} \rangle \mathbf{d}\|_2^2 \quad (4)$$

$$\begin{aligned} &= \arg \min_{\mathbf{d} \in \mathcal{D}_N} \|\mathbf{r}(n)\|_2^2 + |\langle \mathbf{r}(n), \mathbf{d} \rangle|^2 \|\mathbf{d}\|_2^2 - 2|\langle \mathbf{r}(n), \mathbf{d} \rangle|^2 \\ &= \arg \min_{\mathbf{d} \in \mathcal{D}_N} \|\mathbf{r}(n)\|_2^2 - |\langle \mathbf{r}(n), \mathbf{d} \rangle|^2 \\ &= \arg \max_{\mathbf{d} \in \mathcal{D}_N} |\langle \mathbf{r}(n), \mathbf{d} \rangle| \end{aligned} \quad (5)$$

where  $\|\mathbf{d}\|_2 = 1$  for all elements of the dictionary  $\mathcal{D}_N$ . The last expression shows that minimizing a specific error is equivalent to maximizing a magnitude correlation with the residual. MP guarantees that the  $n$ th atom and the  $n$ th residual are orthogonal, i.e.,  $\langle \mathbf{r}(n), \mathbf{h}_{n-1} \rangle = 0$ . OMP [6] uses the same atom selection criterion in (5), but updates the representation according to:

$$\mathcal{X}_{n+1} = \left\{ \begin{array}{l} \mathbf{H}(n+1) = [\mathbf{H}(n)|\mathbf{h}_n], \\ \mathbf{a}(n+1) = \mathbf{H}^\dagger(n+1)\mathbf{x}, \\ \mathbf{r}(n+1) = \mathbf{x} - \mathbf{H}(n+1)\mathbf{a}(n+1) \end{array} \right\} \quad (6)$$

where the pseudoinverse  $\mathbf{H}^\dagger(n) \triangleq [\mathbf{H}^H(n)\mathbf{H}(n)]^{-1}\mathbf{H}^H(n)$  pre-multiplying  $\mathbf{x}$  gives the optimal weights for the least-squares (LS) projection of  $\mathbf{x}$  onto  $\text{span}\{\mathcal{H}_n\}$ . OOMP [7] updates  $\mathcal{H}_n$  using (6), but its atom selection takes into account the LS projection of the signal onto the updated representation:

$$\begin{aligned} \mathbf{h}_n &= \arg \min_{\mathbf{d} \in \mathcal{D}_N} \left\| \mathbf{x} - [\mathbf{H}(n)|\mathbf{d}] ([\mathbf{H}(n)|\mathbf{d}]^H [\mathbf{H}(n)|\mathbf{d}])^{-1} [\mathbf{H}(n)|\mathbf{d}]^H \mathbf{x} \right\|_2^2 \\ &= \arg \min_{\mathbf{d} \in \mathcal{D}_N} \left\| \mathbf{x} - \left( \mathbf{x}_{\mathcal{H}_n} + \frac{\langle \mathbf{r}(n), \mathbf{d}_{\mathcal{H}_n^\perp} \rangle}{\|\mathbf{d}_{\mathcal{H}_n^\perp}\|_2} \frac{\mathbf{d}_{\mathcal{H}_n^\perp}}{\|\mathbf{d}_{\mathcal{H}_n^\perp}\|_2} \right) \right\|_2^2 \\ &= \arg \min_{\mathbf{d} \in \mathcal{D}_N} \left\| \mathbf{r}(n) - \frac{\langle \mathbf{r}(n), \mathbf{d}_{\mathcal{H}_n^\perp} \rangle}{\|\mathbf{d}_{\mathcal{H}_n^\perp}\|_2} \frac{\mathbf{d}_{\mathcal{H}_n^\perp}}{\|\mathbf{d}_{\mathcal{H}_n^\perp}\|_2} \right\|_2^2 \end{aligned} \quad (7)$$

$$= \arg \max_{\mathbf{d} \in \mathcal{D}_N} \frac{|\langle \mathbf{r}(n), \mathbf{d}_{\mathcal{H}_n^\perp} \rangle|}{\|\mathbf{d}_{\mathcal{H}_n^\perp}\|_2^2} \quad (8)$$

where  $\mathbf{x}_{\mathcal{H}_n} \triangleq \mathbf{H}(n)\mathbf{H}^\dagger(n)\mathbf{x}$ , and  $\mathbf{d}_{\mathcal{H}_n^\perp} \triangleq \mathbf{d} - \mathbf{d}_{\mathcal{H}_n} = \mathbf{d} - \mathbf{H}(n)\mathbf{H}^\dagger(n)\mathbf{d}$ . Essentially, OOMP is MP with the additional step that every element of the dictionary is made orthogonal to the approximation and renormalized before the next atom is selected [20]. OMP and OOMP guarantee, unlike MP, that the  $n$ th residual is orthogonal to the first  $n$  atoms of the model, i.e.,  $\mathbf{H}^H(n)\mathbf{r}(n) = \mathbf{H}^H(n)[\mathbf{x} - \mathbf{x}_{\mathcal{H}_n}] = \mathbf{0}$ . This implies that the new residual is orthogonal to the new approximation as well, i.e.,  $\langle \mathbf{r}(n), \mathbf{x}_{\mathcal{H}_n} \rangle = 0$ . MP, OMP, and OOMP all provide suboptimal but “good” solutions to (2) without combinatorial complexity, which means the solutions may not be the sparsest possible but are beneficial to some desired application.

As long as the dictionary spans the space of the signal, the models created by MP, OMP, and OOMP will be *convergent*, which means that  $\lim_{n \rightarrow \infty} \|\mathbf{x} - \mathbf{H}(n)\mathbf{a}(n)\|_2^2 = 0$ . (In the case of OMP and OOMP, this limit is  $K$  because at each iteration they select an atom that is linearly independent of all others in the model.) One performance measure of the signal model in (1) is its signal-to-residual energy ratio (SRR), defined as

$$\text{SRR}(n) \triangleq \frac{\|\mathbf{x}\|_2^2}{\|\mathbf{x} - \mathbf{H}(n)\mathbf{a}(n)\|_2^2}. \quad (9)$$

Two signal models with the same SRR but different orders differ in their efficiency. For example in Fig. 1, the models generated by OOMP are much more efficient than those produced by MP since the OOMP model orders are smaller than those of MP for the same SRR. (Note that  $-10 \log_{10}[\text{SRR}(n)]$  is plotted in Fig. 1.) Finally, a signal model is meaningful when its atoms directly and clearly correspond to features or structures in the signal. In Fig. 2(a), for example, the atoms to the right of the signal onset correspond to the frequency and mean envelope of the signal, while those to the left pertain to nothing real in the signal. In the next section, we define ways to quantify the efficiency and meaningfulness of a signal model by using a measure of interference.

### III. INTERFERENCE

Ultimately, we want to determine how an atom of  $\mathcal{X}_n$  directly reflects some aspect of  $\mathbf{x}$ , and measure how much it corrects the signal model [14]–[17]. Considering the various effects in the models shown in Fig. 2, one measure of correction is suggested by the energy difference caused by incorporating a new atom into the model. If it removes parts of other atoms, e.g., those preceding the onset seen in Fig. 2(a), then the energy it “imparts” to the model should be less than what is expected if the atom is orthogonal to the other atoms already in the representation. Two vectors  $\mathbf{h}, \mathbf{g} \in \mathbb{C}^K$  *constructively interfere* if  $\|\mathbf{h} + \mathbf{g}\|_2^2 > \|\mathbf{h}\|_2^2 + \|\mathbf{g}\|_2^2$ , and they *destructively interfere* if  $\|\mathbf{h} + \mathbf{g}\|_2^2 < \|\mathbf{h}\|_2^2 + \|\mathbf{g}\|_2^2$ . Thus, the difference  $\|\mathbf{h} + \mathbf{g}\|_2^2 - (\|\mathbf{h}\|_2^2 + \|\mathbf{g}\|_2^2)$  quantifies the amount  $\mathbf{h}$  and  $\mathbf{g}$  interfere with each other, and the sign determines the type of interference.

**Definition 1 (Interference):** *The interference associated with the  $m$ th atom  $\mathbf{h}_m \in \mathcal{H}_n$  of the  $n$ th-order representation  $\mathcal{X}_n = \{\mathbf{H}(n), \mathbf{a}(n), \mathbf{r}(n)\}$  of  $\mathbf{x}$  for  $m = 0, 1, \dots, n - 1$  is*

$$\Delta(m) \triangleq \frac{1}{2} \left[ \|\hat{\mathbf{x}}(n)\|_2^2 - (\|\hat{\mathbf{x}}(n \setminus m)\|_2^2 + \|a_m \mathbf{h}_m\|_2^2) \right] \quad (10)$$

$$= \text{Real} \{ a_m^* \langle \hat{\mathbf{x}}(n \setminus m), \mathbf{h}_m \rangle \} \quad (11)$$

where  $\hat{\mathbf{x}}(n) \triangleq \mathbf{H}(n)\mathbf{a}(n)$ ,  $\hat{\mathbf{x}}(n \setminus m) \triangleq \hat{\mathbf{x}}(n) - a_m \mathbf{h}_m$  (the  $n$ th-order approximation of  $\mathbf{x}$  excluding the  $m$ th atom), and  $a_m$  is the  $(m + 1)$ th element of  $\mathbf{a}(n)$  (also denoted by  $[\mathbf{a}(n)]_{m+1}$ ).  $\square$

Since none of the weights are presumed to be zero (otherwise the atom should be excluded from the representation), we see that a sufficient condition for  $\Delta(m) = 0$  is  $\langle \mathbf{h}_l, \mathbf{h}_m \rangle = 0 \forall \{\mathbf{h}_l, \mathbf{h}_m\}_{l \neq m} \in \mathcal{H}_n$ , i.e.,  $\mathbf{h}_m$  is orthogonal to all other atoms in the set of vectors  $\mathcal{H}_n$ . However, when  $\mathcal{D}_N$  is overcomplete, it can also be the case that while  $\mathbf{h}_m$  is not orthogonal to all other atoms, it is orthogonal to  $\hat{\mathbf{x}}(n) - a_m \mathbf{h}_m$ . Thus, the necessary condition for zero interference associated with the  $m$ th atom in  $\mathcal{X}_n$  is  $\langle \hat{\mathbf{x}}(n \setminus m), \mathbf{h}_m \rangle = 0$ , i.e., it is orthogonal to the linear combination of the  $n - 1$  other atoms of the model.

The significance of interference becomes clear when we use it to divide the atoms of a representation. Figure 3 shows how the atoms of the models seen in Fig. 2 are distributed according to the sign of the interference. For Attack, all atoms with peak amplitudes that precede the signal onset appear in the destructively interfering ( $\Delta(m) < 0$ ) partial representation. These atoms may be well-matched to errors in the model, but are not well-matched to the signal. Surprisingly, almost the entire representation created by OOMP for Bimodal appears in the destructively interfering representation, which means that the contribution of these atoms to the model is reliant on how other atoms correct their contribution.



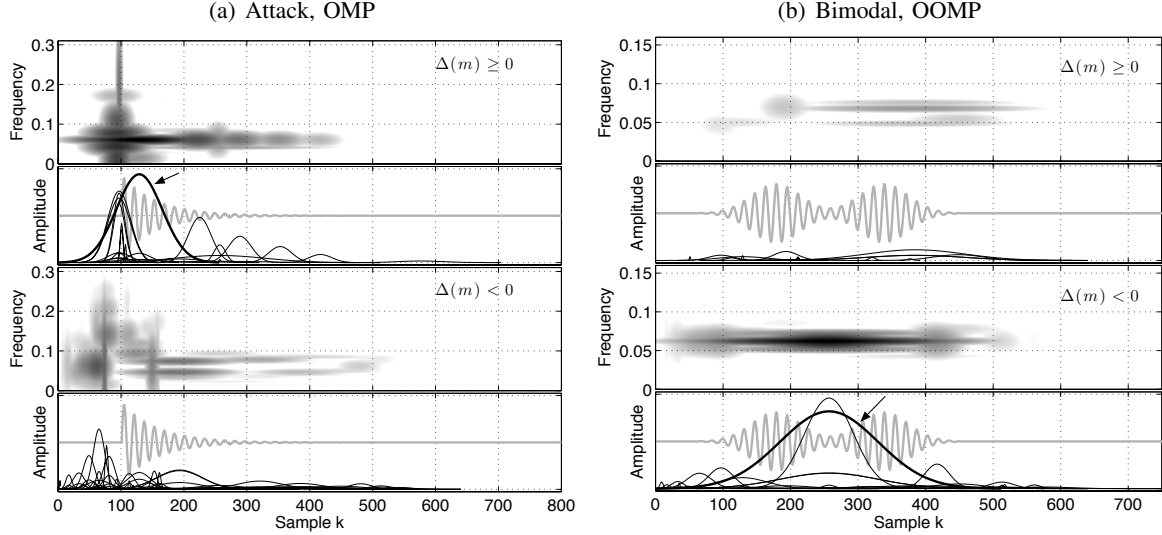


Fig. 3. Models from Fig. 2 partitioned into two submodels based on the sign of the interference  $\Delta(m)$ . Wivigrams are shown above the time-domain atom envelopes for positive and negative interference. Original signal is the thick gray line in the time-domain plots. Arrows denote first atom selected.

This suggests that few atoms in the model contribute in a constructive and clear manner to represent the signal.

To obtain a performance measure for the whole representation, e.g., whether the trend of a representation is more constructive or destructive, we propose a cumulative measure of interference.

**Definition 2 (Cumulative Interference):** *The cumulative interference of the  $m$ th-order model from the  $n$ th-order representation  $\mathcal{X}_n = \{\mathbf{H}(n), \mathbf{a}(n), \mathbf{r}(n)\}$  of  $\mathbf{x}$  for  $m = 1, 2, \dots, n$  is*

$$\Delta_{\Sigma}(m) \triangleq \sum_{l=0}^{m-1} \Delta(l) = \sum_{l=0}^{m-1} \text{Real} \{a_l^* \langle \hat{\mathbf{x}}(n \setminus l), \mathbf{h}_l \rangle\}. \quad (12)$$

For each of the representations of the signals in Fig. 1, the corresponding cumulative interference is shown in black in Fig. 4. The thick gray lines above and below zero show how the positive and negative interference accumulate, respectively. Observe that the model for Sine created by OMP has atoms that on average constructively interfere more than they destructively interfere, while the opposite occurs for the model created by OOMP. For Bimodal, both models from OMP and OOMP have negative cumulative interference, with the model from OOMP having the most extreme values. When the atoms of a representation are ordered according to the iteration at which they are found, cumulative interference provides a historical record of the pursuit. For example, the MP decomposition of Sine is constructive

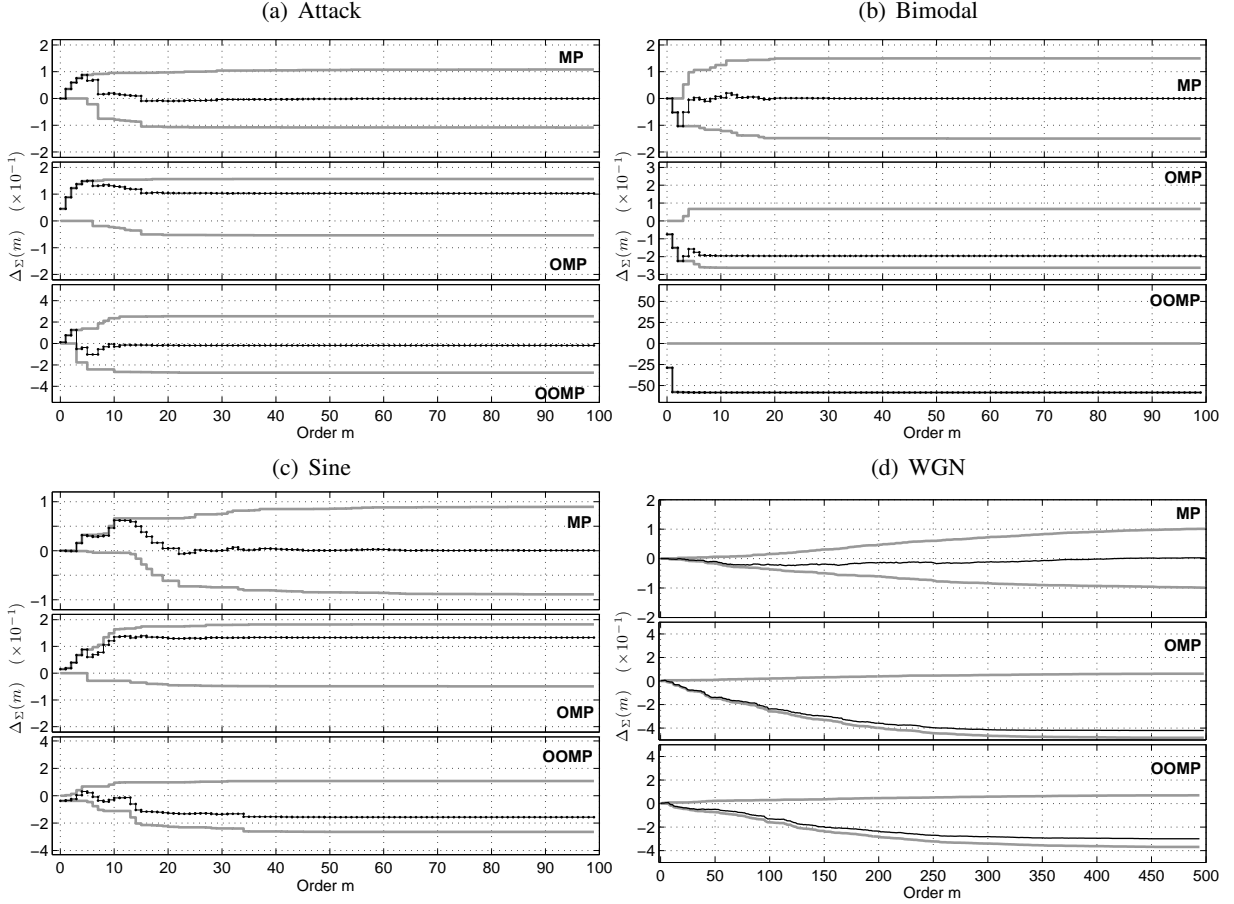


Fig. 4. Cumulative interference  $\Delta_{\Sigma}(m)$  (12) (black) for the representations in Fig. 1 using MP, OMP, and OOMP. Thick gray line above zero is cumulative sum of  $\Delta(m) > 0$ , and that below zero is cumulative sum of  $\Delta(m) < 0$ . (Note changes in the axes ranges.)

until about the 13th iteration. The OOMP model of Bimodal is extremely destructive, which is also seen in Fig. 3(b). This model is built mostly from destructively interfering atoms; the extent of this is given by the *total interference*.

**Definition 3 (Total Interference):** *The total interference of the  $n$ th-order representation  $\mathcal{X}_n = \{\mathbf{H}(n), \mathbf{a}(n), \mathbf{r}(n)\}$  of  $\mathbf{x}$  is the sum of all interference terms:*

$$\Delta_{\Sigma}(n) = \sum_{m=0}^{n-1} \Delta(m) = \|\hat{\mathbf{x}}(n)\|_2^2 - \|\mathbf{a}(n)\|_2^2 \quad (13)$$

where  $\hat{\mathbf{x}}(n \setminus m)$  has been substituted into (12). □

The cumulative interference for the models from Fig. 1 is asymptotic to a total interference in each case, as seen in Fig. 4, and for MP it appears to converge to zero. These results are explained in the following theorem.

**Theorem 1:** *The total interference of any convergent additive model of  $\mathbf{x} \in \mathbb{C}^K$  in (1) is the difference between the energy of  $\mathbf{x}$  and the  $\ell_2$ -norm of the representation weights, i.e.,*

$$\lim_{n \rightarrow \infty} \Delta_{\Sigma}(n) = \|\mathbf{x}\|_2^2 - \lim_{n \rightarrow \infty} \|\mathbf{a}(n)\|_2^2. \quad (14)$$

Furthermore, every convergent model produced by MP will have zero total interference:

$$\lim_{n \rightarrow \infty} \Delta_{\Sigma}(n) = 0. \quad (15)$$

□

*Proof:* By definition,  $\lim_{n \rightarrow \infty} \|\mathbf{x} - \mathbf{H}(n)\mathbf{a}(n)\|_p = \lim_{n \rightarrow \infty} \|\mathbf{r}(n)\|_p = 0$  for a convergent model, which implies that  $\lim_{n \rightarrow \infty} \|\hat{\mathbf{x}}(n)\|_p = \|\mathbf{x}\|_p$ . Substituting this into (13) yields (14) when taken to convergence. Now, by the energy conservation property of MP [1],

$$\|\mathbf{x}\|_2^2 = \|\mathbf{r}(n)\|_2^2 + \|\mathbf{a}(n)\|_2^2 \quad (16)$$

for all iterations  $n \geq 0$ . Substituting  $\|\mathbf{x}\|_2^2 - \lim_{n \rightarrow \infty} \|\mathbf{r}(n)\|_2^2$  for  $\lim_{n \rightarrow \infty} \|\mathbf{a}(n)\|_2^2$  into (13) yields

$$\begin{aligned} \lim_{n \rightarrow \infty} \Delta_{\Sigma}(n) &= \|\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2 + \lim_{n \rightarrow \infty} \|\mathbf{r}(n)\|_2^2 \\ &= \lim_{n \rightarrow \infty} \|\mathbf{r}(n)\|_2^2 = 0, \end{aligned} \quad (17)$$

thus completing the proof. Since (16) does not hold for OMP or OOMP in general, (15) does not apply to the models they produce. ■

Theorem 1 demonstrates that for any convergent model of  $\mathbf{x} \in \mathbb{C}^K$  built by MP, regardless of the contents of the dictionary (as long as it spans  $\mathbb{C}^K$ ), its total interference must converge to zero. This implies there will be equal amounts of positive (constructive) interference and negative (destructive) interference in the signal model of  $\mathbf{x}$ , which explains the behavior seen in Fig. 4 for the models of MP. Thus, we cannot expect MP to find a convergent signal model in which atom correction does not play as much a role as constructive modeling of the signal (unless the atoms are mutually orthogonal, which is a trivial scenario). On the other hand, for any other convergent model built by a pursuit for which (15) does not hold, such as OMP and OOMP where it is possible that  $\lim_{n \rightarrow \infty} \|\mathbf{a}(n)\|_2^2 \neq \|\mathbf{x}\|_2^2$ , its total interference can be skewed more toward constructive interference than destructive interference, but only by an amount strictly less than the energy of the signal  $\|\mathbf{x}\|_2^2$  because  $\|\mathbf{a}(n)\|_2^2 > 0$ . In other words,  $\Delta_{\Sigma}(n) < \|\mathbf{x}\|_2^2$  for all  $n$  for OMP and OOMP, but it is not necessarily zero.

Theorem 1 motivates us to consider which kind of total interference — positive (constructive) or negative (destructive) — is “better” in a signal model. Though OOMP is capable of decomposing Bimodal to 60 dB SRR using only 34 atoms, as seen in Fig. 1(b), the resulting model appears to consist mostly of corrections to the few initial atoms selected greedily, as seen in Fig. 3(b). A large amount of negative interference indicates correction due to mismatches between the model and the signal, as well as poor atom selections by the decomposition process. It is thus reasonable to claim that a good signal model lacks terms that require correction through destructive interference — thus maintaining a meaningful correspondence to signal content — and a good pursuit avoids selecting such terms.

We have thus far shown that interference can be useful to study the structure of a sparse approximation. Next we employ interference in the atom selection criterion of a greedy iterative descent pursuit, and demonstrate that the efficiency and meaningfulness of the resulting signal models can be improved.

#### IV. INCORPORATING INTERFERENCE INTO PURSUITS

Our goal is to build a sparse signal model having small error and large positive total interference. Thus, instead of an optimization as in (2), consider the *joint minimization* of the  $n$ th-order model error energy and its total negative interference:

$$\begin{aligned} & \min_{\mathcal{H}_n \subset \mathcal{D}_N} \|\mathbf{x} - \mathbf{H}(n)\mathbf{a}(n)\|_2^2 - \gamma \Delta_{\Sigma}(n) \\ & = \min_{\mathcal{H}_n \subset \mathcal{D}_N} \|\mathbf{x} - \mathbf{H}(n)\mathbf{a}(n)\|_2^2 + \gamma (\|\mathbf{a}(n)\|_2^2 - \|\mathbf{H}(n)\mathbf{a}(n)\|_2^2) \end{aligned} \quad (18)$$

where the weight  $-\infty < \gamma < \infty$  determines the influence of total interference on the solution, we have substituted (13), and we implicitly emphasize that  $n \ll K$  such that  $\hat{\mathbf{x}}(n) \rightarrow \mathbf{x}$  quickly. For an overcomplete dictionary, there will exist many  $\mathcal{H}_n \subset \mathcal{D}_N$  such that  $\|\mathbf{x} - \hat{\mathbf{x}}(n)\|_2 \approx 0$ , and of these we want the one that has the largest difference  $\|\hat{\mathbf{x}}(n)\|_2^2 - \|\mathbf{a}(n)\|_2^2$  when  $\gamma > 0$ . The solution to this formulation maximizes the energy of the approximation of  $\mathbf{x}$  onto the span of  $\mathcal{H}_n$ , and the difference between its energy and that in a “transformed domain,” e.g.,  $\|\mathbf{H}^\dagger(n)\mathbf{x}\|_2^2$  in the case of OMP and OOMP. When  $\gamma > 0$ , we seek a signal model that is on average more constructive than destructive, and vice versa for  $\gamma < 0$  (for  $\gamma = 0$ , (18), of course, reduces to (2)).

Finding a subset of  $n \ll K$  dictionary atoms using (18) leaves us, as before, with a combinatorial algorithm. We want a solution that “points” near  $\mathbf{x}$ , but does so using  $n \ll K$  atoms working together in a constructive manner. Thus, as is done in MP, OMP, and OOMP where an iterative descent approach is used to find a “good” solution to (2) (but not necessarily the best), we will do the same for (18) by

incorporating interference into the atom selection criteria of MP, OMP, and OOMP [17]. Toward this end, consider augmenting the atom selection of MP and OMP (4) as follows:

$$\mathbf{h}_n = \arg \min_{\mathbf{d} \in \mathcal{D}_N} \|\mathbf{r}(n) - \langle \mathbf{r}(n), \mathbf{d} \rangle \mathbf{d}\|_2^2 + \gamma(n)[- \Delta_{\mathbf{d}}] \quad (19)$$

where  $\Delta_{\mathbf{d}}$  is the interference associated with  $\mathbf{d} \in \mathcal{D}_N$  if it augments the  $n$ th-order model in (1), and (potentially time-varying)  $-\infty < \gamma(n) < \infty$  weights the influence and type of interference in the atom selection. With this criterion, the algorithm seeks an atom that minimizes the squared error, and at the same time constructively interferes with the rest of the model when  $\gamma(n) > 0$ , or destructively interferes when  $\gamma(n) < 0$ . We similarly change the atom selection criterion of OOMP in (7) to

$$\mathbf{h}_n = \arg \min_{\mathbf{d} \in \mathcal{D}_N} \|\mathbf{r}(n) - \mathbf{r}_{\mathbf{d}}(n)\|_2^2 + \gamma(n)[- \Delta_{\mathbf{d}}] \quad (20)$$

where  $\mathbf{r}_{\mathbf{d}}(n)$  is the LS projection of the residual onto the orthogonal complement to  $\mathcal{H}_n$  of the dictionary atom  $\mathbf{d}$ , i.e.,

$$\mathbf{r}_{\mathbf{d}}(n) \triangleq \frac{\langle \mathbf{r}(n), \mathbf{d}_{\mathcal{H}_n^\perp} \rangle}{\|\mathbf{d}_{\mathcal{H}_n^\perp}\|_2} \frac{\mathbf{d}_{\mathcal{H}_n^\perp}}{\|\mathbf{d}_{\mathcal{H}_n^\perp}\|_2}. \quad (21)$$

Observe that this selection criterion, and the approach in (18), are similar to the technique of regularization used in optimization theory [21]; here, however, the second term can be negative. The type of interference weighting  $\gamma(n)$  will control the performance of these pursuits, but for simplicity in the computer simulations presented in Section V, we set the interference weighting to be constant  $\gamma(n) = \gamma$ .

#### A. Interference of a New Atom

To employ the atom selection criteria in (19) and (20), we must first find an expression for  $\Delta_{\mathbf{d}}$  — the interference of the new atom with the rest of the model. We see that the interference measure in (10) depends on the signal model order, and with the addition of a new atom, the interference associated with each atom in the model can change. We will now find this change for MP, OMP, and OOMP. Define the column vector  $\underline{\Delta}(n) \triangleq [\Delta(0), \Delta(1), \dots, \Delta(n-1)]^T$ . The following propositions show how the interference of each atom in  $\mathcal{X}_n$  changes with the addition of  $\mathbf{h}_n$  in the representation update of MP (22), and of OMP and OOMP in (26).

**Proposition 1:** *The interference of the atoms in a representation updated by MP is*

$$\underline{\Delta}(n+1) = \begin{bmatrix} \underline{\Delta}(n) \\ 0 \end{bmatrix} + \text{Real} \begin{bmatrix} a_n \mathbf{A}^H(n) \mathbf{H}^H(n) \mathbf{h}_n \\ a_n^* \langle \hat{\mathbf{x}}(n), \mathbf{h}_n \rangle \end{bmatrix} \quad (22)$$

where  $\mathbf{A}(n) \triangleq \text{diag}[\mathbf{a}(n)]$  and  $a_n \triangleq \langle \mathbf{r}(n), \mathbf{h}_n \rangle$ . □

*Proof:* From the definition of interference for  $\mathcal{X}_{n+1}$  in (11) for  $m = 0, 1, \dots, n$ , the new interference of the  $m$ th atom, denoted by  $\Delta'(m)$ , is given by

$$\Delta'(m) = \text{Real}\{a_m^* \langle \hat{\mathbf{x}}(n+1 \setminus m), \mathbf{h}_m \rangle\} = \begin{cases} \text{Real}\{a_m^* \langle \hat{\mathbf{x}}(n \setminus m) + a_n \mathbf{h}_n, \mathbf{h}_m \rangle\}, & m = 0, \dots, n-1 \\ \text{Real}\{a_n^* \langle \hat{\mathbf{x}}(n+1 \setminus n), \mathbf{h}_n \rangle\}, & m = n \end{cases} \quad (23)$$

where  $\hat{\mathbf{x}}(n+1 \setminus m) = \hat{\mathbf{x}}(n \setminus m)$  for MP for  $m < n$ . Observe that when  $m = n$ ,  $\text{Real}\{a_n^* \langle \hat{\mathbf{x}}(n+1 \setminus n), \mathbf{h}_n \rangle\} = \Delta(n)$ , i.e., the interference of the new atom. For atoms  $m = 0, 1, \dots, n-1$ ,

$$\text{Real}\{a_m^* \langle \hat{\mathbf{x}}(n \setminus m) + a_n \mathbf{h}_n, \mathbf{h}_m \rangle\} = \text{Real}\{a_m^* \langle \hat{\mathbf{x}}(n \setminus m), \mathbf{h}_m \rangle\} + \text{Real}\{a_n a_m^* \langle \mathbf{h}_n, \mathbf{h}_m \rangle\} \quad (24)$$

where the first term is just  $\Delta(m)$ , or the old interference. The argument of the second term can be expressed as

$$\begin{aligned} a_n a_m^* \langle \mathbf{h}_n, \mathbf{h}_m \rangle &= a_n [\mathbf{a}(n)]_{m+1}^* [\mathbf{H}^H(n) \mathbf{h}_n]_{m+1} \\ &= a_n [\mathbf{A}^H(n) \mathbf{H}^H(n) \mathbf{h}_n]_{m+1}. \end{aligned} \quad (25)$$

Combining these into a column vector gives (22) and completes the proof.  $\blacksquare$

Proposition 1 shows that the interference associated with each atom in the updated representation  $\mathcal{X}_{n+1}$  created by MP is a concatenation of the interference associated with the new atom, and an additive factor applied to all other atoms based on how they interfere with the new atom.

**Proposition 2:** *The interference of the atoms in a representation updated by OMP and OOMP is*

$$\underline{\Delta}(n+1) = \begin{bmatrix} \underline{\Delta}(n) \\ \text{Real}\{a_n^* \langle \hat{\mathbf{x}}(n), \mathbf{h}_n \rangle\} \end{bmatrix} + \text{Real} \begin{bmatrix} a_n \mathbf{C}^H(n) \mathbf{b}_n - a_n^* \mathbf{R}(n) \mathbf{b}_n^* \\ -|a_n|^2 \|\mathbf{h}_{n, \mathcal{H}_n}\|_2^2 \end{bmatrix} \quad (26)$$

where  $\mathbf{b}_n \triangleq \mathbf{H}^\dagger(n) \mathbf{h}_n$ ,  $\mathbf{h}_{n, \mathcal{H}_n} \triangleq \mathbf{H}(n) \mathbf{b}_n$ ,  $\mathbf{C}(n) \triangleq \mathbf{A}(n) - a_n \text{diag}[\mathbf{b}_n]$ ,  $a_n \triangleq \langle \mathbf{r}(n), \mathbf{h}_n \rangle / \|\mathbf{h}_{n, \mathcal{H}_n}\|_2^2$  [6], and

$$\mathbf{R}(n) \triangleq \text{diag}\{[\langle \hat{\mathbf{x}}(n \setminus 0), \mathbf{h}_0 \rangle, \langle \hat{\mathbf{x}}(n \setminus 1), \mathbf{h}_1 \rangle, \dots, \langle \hat{\mathbf{x}}(n \setminus n-1), \mathbf{h}_{n-1} \rangle]\} \quad (27)$$

is a diagonal matrix of inner product terms.  $\square$

*Proof:* The new interference of the  $m$ th atom ( $m < n$ ) in the updated representation  $\mathcal{X}_{n+1}$  from (11) is

$$\Delta'(m) = \text{Real}\{[\mathbf{a}(n+1)]_{m+1}^* \langle \hat{\mathbf{x}}(n+1 \setminus m), \mathbf{h}_m \rangle\}. \quad (28)$$

Let us first express  $\hat{\mathbf{x}}(n+1 \setminus m)$  in terms of  $\mathbf{x}(n \setminus m)$  for  $m = 0, 1, \dots, n-1$ . The coefficient update rule of OMP is [6]

$$\mathbf{a}(n+1) \triangleq \begin{bmatrix} \mathbf{a}(n) - a_n \mathbf{b}_n \\ a_n \end{bmatrix}. \quad (29)$$

This yields

$$\begin{aligned}
\hat{\mathbf{x}}(n+1 \setminus m) &= \hat{\mathbf{x}}(n+1) - [\mathbf{a}(n+1)]_{m+1} \mathbf{h}_m \\
&= [\mathbf{H}(n) | \mathbf{h}_n] \begin{bmatrix} \mathbf{a}(n) - a_n \mathbf{b}_n \\ a_n \end{bmatrix} \\
&\quad - [\mathbf{a}(n+1)]_{m+1} \mathbf{h}_m \\
&= \mathbf{H}(n) \mathbf{a}(n) - a_n (\mathbf{H}(n) \mathbf{b}_n - \mathbf{h}_n) \\
&\quad - [\mathbf{a}(n+1)]_{m+1} \mathbf{h}_m \\
&= \hat{\mathbf{x}}(n) - a_n (\mathbf{h}_{n, \mathcal{H}_n} - \mathbf{h}_n) \\
&\quad - [\mathbf{a}(n) - a_n \mathbf{b}_n]_{m+1} \mathbf{h}_m \\
&= \hat{\mathbf{x}}(n \setminus m) + a_n (\mathbf{h}_{n, \mathcal{H}_n^\perp} + \mathbf{h}_m [\mathbf{b}_n]_{m+1})
\end{aligned} \tag{30}$$

where  $\hat{\mathbf{x}}(n \setminus m) = \hat{\mathbf{x}}(n) - [\mathbf{a}(n)]_{m+1} \mathbf{h}_m$  has been used. The inner product of this expression and the  $m$ th atom  $\mathbf{h}_m$  is

$$\begin{aligned}
\langle \hat{\mathbf{x}}(n+1 \setminus m), \mathbf{h}_m \rangle &= \langle \hat{\mathbf{x}}(n \setminus m), \mathbf{h}_m \rangle + a_n \langle \mathbf{h}_{n, \mathcal{H}_n^\perp}, \mathbf{h}_m \rangle \\
&\quad + a_n [\mathbf{b}_n]_{m+1} \|\mathbf{h}_m\|_2^2 \\
&= \langle \hat{\mathbf{x}}(n \setminus m), \mathbf{h}_m \rangle + a_n [\mathbf{b}_n]_{m+1},
\end{aligned} \tag{31}$$

which simplifies because  $\mathbf{h}_{n, \mathcal{H}_n^\perp}$  is orthogonal to all unit-norm atoms in  $\mathcal{H}_n$ . Substituting this expression into (28) yields

$$\begin{aligned}
\Delta'(m) &= \text{Real} \left\{ \left[ \mathbf{a}(n) - a_n \mathbf{b}_n \right]_{m+1}^* \right. \\
&\quad \left. \times \left( \langle \hat{\mathbf{x}}(n \setminus m), \mathbf{h}_m \rangle + a_n [\mathbf{b}_n]_{m+1} \right) \right\} \\
&= \Delta(m) - \text{Real} \left\{ [a_n \mathbf{b}_n]_{m+1}^* \langle \hat{\mathbf{x}}(n \setminus m), \mathbf{h}_m \rangle \right. \\
&\quad \left. + a_n \left[ \mathbf{a}(n) - a_n \mathbf{b}_n \right]_{m+1}^* [\mathbf{b}_n]_{m+1} \right\} \\
&= \Delta(m) - \text{Real} \left\{ a_n^* [\mathbf{R}(n) \mathbf{b}_n^*]_{m+1} \right. \\
&\quad \left. + a_n [\mathbf{C}^H(n) \mathbf{b}_n]_{m+1} \right\}
\end{aligned} \tag{32}$$

where  $\Delta(m) = \text{Real} \{ [\mathbf{a}(n)]_{m+1}^* \langle \hat{\mathbf{x}}(n \setminus m), \mathbf{h}_m \rangle \}$ . This result creates the first row of (26).

The interference associated with the new atom is

$$\begin{aligned}
\Delta(n) &= \text{Real}\{a_n^* \langle \hat{\mathbf{x}}(n+1|n), \mathbf{h}_n \rangle\} \\
&= \text{Real}\{a_n^* \langle \hat{\mathbf{x}}(n+1) - a_n \mathbf{h}_n, \mathbf{h}_n \rangle\} \\
&= \text{Real}\{a_n^* \langle \hat{\mathbf{x}}(n+1), \mathbf{h}_n \rangle\} - |a_n|^2.
\end{aligned} \tag{33}$$

Substituting  $\hat{\mathbf{x}}(n+1) = \mathbf{H}(n+1)\mathbf{a}(n+1)$  yields

$$\begin{aligned}
\Delta(n) &= \text{Real} \left\{ a_n^* \left\langle [\mathbf{H}(n)|\mathbf{h}_n] \begin{bmatrix} \mathbf{a}(n) - a_n \mathbf{b}_n \\ a_n \end{bmatrix}, \mathbf{h}_n \right\rangle \right\} - |a_n|^2 \\
&= \text{Real} \{ a_n^* \langle \hat{\mathbf{x}}(n), \mathbf{h}_n \rangle - |a_n|^2 \langle \mathbf{h}_n, \mathcal{H}_n, \mathbf{h}_n \rangle \} \\
&= \text{Real} \{ a_n^* \langle \hat{\mathbf{x}}(n), \mathbf{h}_n \rangle \} - |a_n|^2 \|\mathbf{h}_n, \mathcal{H}_n\|_2^2.
\end{aligned} \tag{34}$$

This gives the last row of (26) and completes the proof.  $\blacksquare$

Proposition 2 shows that for OMP and OOMP, augmenting the representation by a new atom involves a nontrivial modification to previous interference values, which is caused by the LS projection in (6).

### B. Interference Adaptation in MP, OMP, and OOMP

The propositions above provide expressions for  $\Delta_{\mathbf{d}}$ , the interference associated with any atom of the dictionary if it is selected to be part of the model, and allows us to incorporate *interference adaptation* into MP, OMP, and OOMP. Considering a real signal and real dictionary for simplicity, the atom selection criterion in (19) becomes

$$\begin{aligned}
\mathbf{h}_n &= \arg \min_{\mathbf{d} \in \mathcal{D}_N} \|\mathbf{r}(n)\|_2^2 - |\langle \mathbf{r}(n), \mathbf{d} \rangle|^2 + \gamma(n) [-\langle \mathbf{r}(n), \mathbf{d} \rangle \langle \hat{\mathbf{x}}(n), \mathbf{d} \rangle] \\
&= \arg \max_{\mathbf{d} \in \mathcal{D}_N} |\langle \mathbf{r}(n), \mathbf{d} \rangle|^2 + \gamma(n) \langle \mathbf{r}(n), \mathbf{d} \rangle \langle \hat{\mathbf{x}}(n), \mathbf{d} \rangle
\end{aligned} \tag{35}$$

where we have substituted  $\mathbf{d}$  for  $\mathbf{h}_n$  in the last row of (22) to produce  $\Delta_{\mathbf{d}} = \langle \mathbf{r}(n), \mathbf{d} \rangle \langle \hat{\mathbf{x}}(n), \mathbf{d} \rangle$ , and converted (19) to an equivalent maximization problem. Similarly the criterion in (20) becomes

$$\begin{aligned}
\mathbf{h}_n &= \arg \min_{\mathbf{d} \in \mathcal{D}_N} \|\mathbf{r}(n)\|_2^2 - \frac{|\langle \mathbf{r}(n), \mathbf{d}_{\mathcal{H}_n^\perp} \rangle|^2}{\|\mathbf{d}_{\mathcal{H}_n^\perp}\|_2^4} \\
&\quad + \gamma(n) \left( -\frac{\langle \mathbf{r}(n), \mathbf{d}_{\mathcal{H}_n^\perp} \rangle \langle \hat{\mathbf{x}}(n), \mathbf{d} \rangle}{\|\mathbf{d}_{\mathcal{H}_n^\perp}\|_2^2} + \frac{|\langle \mathbf{r}(n), \mathbf{d}_{\mathcal{H}_n^\perp} \rangle|^2}{\|\mathbf{d}_{\mathcal{H}_n^\perp}\|_2^4} \|\mathbf{d}_{\mathcal{H}_n}\|_2^2 \right) \\
&= \arg \max_{\mathbf{d} \in \mathcal{D}_N} \frac{|\langle \mathbf{r}(n), \mathbf{d}_{\mathcal{H}_n^\perp} \rangle|^2}{\|\mathbf{d}_{\mathcal{H}_n^\perp}\|_2^4} + \gamma(n) \left( \frac{\langle \mathbf{r}(n), \mathbf{d}_{\mathcal{H}_n^\perp} \rangle \langle \hat{\mathbf{x}}(n), \mathbf{d} \rangle}{\|\mathbf{d}_{\mathcal{H}_n^\perp}\|_2^2} - \frac{|\langle \mathbf{r}(n), \mathbf{d}_{\mathcal{H}_n^\perp} \rangle|^2}{\|\mathbf{d}_{\mathcal{H}_n^\perp}\|_2^4} \frac{\|\mathbf{d}_{\mathcal{H}_n}\|_2^2}{\|\mathbf{d}_{\mathcal{H}_n^\perp}\|_2^2} \right)
\end{aligned} \tag{36}$$



where we have substituted  $\mathbf{d}$  for  $\mathbf{h}_n$  in the last row of (26), defined  $a_n \triangleq \langle \mathbf{r}(n), \mathbf{d}_{\mathcal{H}_n^\perp} \rangle / \|\mathbf{d}_{\mathcal{H}_n^\perp}\|_2^2$  as in (26), and again converted (20) to an equivalent maximization problem. Finally, we ignore the last term of (36) by assuming that  $\|\mathbf{d}_{\mathcal{H}_n}\|_2^2 \approx 0$ , and consequently  $\|\mathbf{d}_{\mathcal{H}_n^\perp}\|_2^2 \approx 1$ , until  $\|\mathbf{r}(n)\|_2^2$  becomes small, so that the criterion becomes

$$\mathbf{h}_n = \arg \max_{\mathbf{d} \in \mathcal{D}_N} \frac{|\langle \mathbf{r}(n), \mathbf{d}_{\mathcal{H}_n^\perp} \rangle|^2 + \gamma(n) \langle \mathbf{r}(n), \mathbf{d}_{\mathcal{H}_n^\perp} \rangle \langle \hat{\mathbf{x}}(n), \mathbf{d} \rangle}{\|\mathbf{d}_{\mathcal{H}_n^\perp}\|_2^2}. \quad (37)$$

This is done for the following reasons. First, we are interested more in the sign and the approximate value of the interference than in its exact value because we want to stress finding an atom that is positively correlated with the residual and the approximation. Second, we want an expression that is similar in form to the selection criterion for MP and OMP in (35), with the only difference being the emphasis on the space orthogonal to  $\text{span}\{\mathcal{H}_n\}$ .

Though atoms are now selected differently, their weights are still computed using (3) for MP, and (6) for OMP and OOMP, simply because we still want to reduce the error as much as possible given the selected atom (in the case of MP), or the selected set of atoms (in the case of OMP and OOMP). Thus, the only change to the pursuits is the atom selection criterion. Compared to their original formulations, the proposed MP/OMP atom selection criterion in (35) and that for OOMP in (37) require only the additional step of finding  $\{\langle \hat{\mathbf{x}}(n), \mathbf{d} \rangle : \mathbf{d} \in \mathcal{D}_N\}$  — i.e., the inner products of the current approximation and each element in the dictionary. Depending on the construction of the dictionary, several ways exist for reducing the complexity of this procedure, such as using the frequency-domain techniques in [17], [22]. We can also apply the optimization procedure proposed in [1] using the pre-computed Gramian of the dictionary because  $\langle \hat{\mathbf{x}}(n), \mathbf{d} \rangle = [\mathbf{d}^T \mathbf{H}(n)] \mathbf{a}(n)$ , which is just a linear combination of values from that matrix. The additional computational complexity, however, becomes less of an issue when the interference-adaptive pursuit finds a good model in fewer iterations than the unmodified approaches.

When  $\gamma(n) = 0$ , the selection criteria in (35) and (37) reduce, of course, to (5) and (8), respectively. When  $\gamma(n) > 0$ , atoms that constructively interfere with those in  $\mathcal{X}_n$  have an advantage over other atoms; and when  $\gamma(n) < 0$ , atoms that destructively interfere with those in  $\mathcal{X}_n$  have the advantage. Figure 5 shows a simple diagram of how an interference-adaptive pursuit behaves. After the first iteration, standard MP, OMP, and OOMP select the atom that has the largest projection onto the residual  $\mathbf{r}(1)$  (and, in the case of OOMP, the atom that points least in the direction of the first atom). As seen in the figure, this atom might be nearly orthogonal to the original signal  $\mathbf{x}$ , in which case one could argue that it serves as a correction to the contribution of the first atom. With  $\gamma(1) \neq 0$ , atom selection is influenced by the approximation as well as the residual. Any atom with a large positive projection onto the residual will

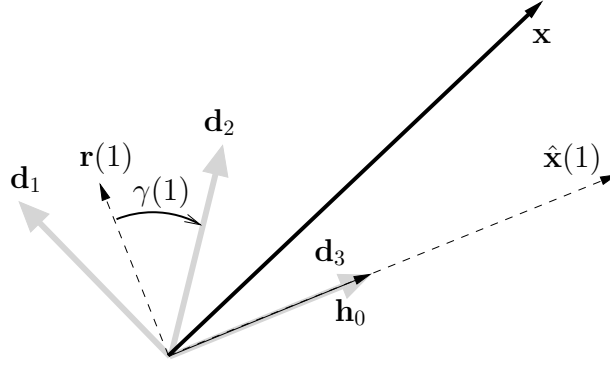


Fig. 5. Geometric interpretation of an interference-adaptive pursuit.  $\mathbf{x} \in \mathbb{R}^2$  is modeled by three dictionary elements  $\{\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3\}$  (gray arrows). First atom selected is  $\mathbf{d}_3$ , becoming  $\mathbf{h}_0$  in  $\mathcal{H}_1$ , which creates the first-order approximation  $\hat{\mathbf{x}}(1) = \mathbf{H}(1)\mathbf{a}(1)$  and the residual  $\mathbf{r}(1) = \mathbf{x} - \hat{\mathbf{x}}(1)$ . Selection of the second atom is influenced by the interference weight: for  $\gamma(1) > 0$ , the pursuit begins to favor  $\mathbf{d}_2$  over  $\mathbf{d}_1$ , which is nearly orthogonal to  $\mathbf{x}$ .

have a small projection onto the approximation, and vice versa, the importance of which depends on the magnitude of  $\gamma(1)$ . In effect, a positive weighting for the interference in (35) and (37) makes the residual appear closer to the original signal  $\mathbf{x}$  than does the original atom selection criterion in (5) and (8). Thus, instead of selecting  $\mathbf{d}_1$  in this example, the pursuit will select  $\mathbf{d}_2$ ; and if using the update rule for OMP and OOMP in (6), the model will converge.

When the interference weight is large and positive, e.g.,  $\gamma(n) \gg \|\mathbf{x}\|_2$ , the interference term will dominate the atom selection. For MP and OMP, the atom selection in (35) becomes

$$\begin{aligned} \lim_{\gamma(n) \rightarrow \infty} \left[ \arg \max_{\mathbf{d} \in \mathcal{D}_N} |\langle \mathbf{r}(n), \mathbf{d} \rangle|^2 + \gamma(n) \langle \mathbf{r}(n), \mathbf{d} \rangle \langle \hat{\mathbf{x}}(n), \mathbf{d} \rangle \right] \\ = \arg \max_{\mathbf{d} \in \mathcal{D}_N} \langle \mathbf{x} - \hat{\mathbf{x}}(n), \mathbf{d} \rangle \langle \hat{\mathbf{x}}(n), \mathbf{d} \rangle \\ = \arg \max_{\mathbf{d} \in \mathcal{D}_N} \langle \mathbf{x}, \mathbf{d} \rangle \langle \hat{\mathbf{x}}(n), \mathbf{d} \rangle - |\langle \hat{\mathbf{x}}(n), \mathbf{d} \rangle|^2 \end{aligned} \quad (38)$$

and for OOMP, the atom selection in (37) becomes

$$\begin{aligned} \lim_{\gamma(n) \rightarrow \infty} \left[ \arg \max_{\mathbf{d} \in \mathcal{D}_N} \frac{|\langle \mathbf{r}(n), \mathbf{d}_{\mathcal{H}_n^\perp} \rangle|^2 + \gamma(n) \langle \mathbf{r}(n), \mathbf{d}_{\mathcal{H}_n^\perp} \rangle \langle \hat{\mathbf{x}}(n), \mathbf{d} \rangle}{\|\mathbf{d}_{\mathcal{H}_n^\perp}\|_2^2} \right] \\ = \arg \max_{\mathbf{d} \in \mathcal{D}_N} \frac{\langle \mathbf{x}, \mathbf{d} \rangle \langle \hat{\mathbf{x}}(n), \mathbf{d} \rangle - |\langle \hat{\mathbf{x}}(n), \mathbf{d} \rangle|^2}{\|\mathbf{d}_{\mathcal{H}_n^\perp}\|_2^2}. \end{aligned} \quad (39)$$

In both of these results, we see that the algorithm attempts to select an atom that lies between the original signal  $\mathbf{x}$  and the current approximation  $\hat{\mathbf{x}}(n)$ , but not so much that it is colinear with  $\hat{\mathbf{x}}(n)$ . When the

interference weight is small and negative, e.g.,  $\gamma(n) \ll -\|\mathbf{x}\|_2$ , then these two scenarios are reversed, and (38) and (39) become

$$\arg \max_{\mathbf{d} \in \mathcal{D}_N} |\langle \hat{\mathbf{x}}(n), \mathbf{d} \rangle|^2 - \langle \mathbf{x}, \mathbf{d} \rangle \langle \hat{\mathbf{x}}(n), \mathbf{d} \rangle \quad (40)$$

and

$$\arg \max_{\mathbf{d} \in \mathcal{D}_N} \frac{|\langle \hat{\mathbf{x}}(n), \mathbf{d} \rangle|^2 - \langle \mathbf{x}, \mathbf{d} \rangle \langle \hat{\mathbf{x}}(n), \mathbf{d} \rangle}{\|\mathbf{d}_{\mathcal{H}_n^+}\|_2^2}, \quad (41)$$

respectively. From these expressions, it is clear that the algorithm will find an atom pointing in a direction opposite to either  $\mathbf{x}$  or  $\hat{\mathbf{x}}(n)$ , i.e., an atom that contributes to the model by destructively interfering with it. These behaviors at the two extremes of  $\gamma(n)$  clearly embody the notions of constructive and destructive interference.

Another way to interpret the atom selection procedures in (35) and (37) is in terms of maximum *a posteriori* estimation, where an atom is selected based on the following principle:

$$\max_{\mathbf{d} \in \mathcal{D}_N} \log P\{\mathbf{r}(n) | \mathbf{d} \in \mathcal{X}_{n+1}\} + \log P\{\mathbf{d} \in \mathcal{X}_{n+1}\} \quad (42)$$

where  $P\{\mathbf{r}(n) | \mathbf{d} \in \mathcal{X}_{n+1}\}$  is a conditional probability of the  $n$ th-order residual, and  $P\{\mathbf{d} \in \mathcal{X}_{n+1}\}$  is the *a priori* probability of  $\mathbf{d}$  belonging to  $\mathcal{X}_{n+1}$ . The first term can be likened to the first term in (35) and (37), where the atom  $\mathbf{d}$  most correlated with the residual is emphasized; and the second term functions like the second term in (35) and (37). Each pursuit begins with no information about the model, and thus  $P\{\mathbf{d} \in \mathcal{X}_1\}$  is uniform in the dictionary. With successive updates of the model, the prior  $P\{\mathbf{d} \in \mathcal{X}_n\}$  should show a larger probability mass for atoms that are positively correlated with the signal model, and a smaller probability mass for atoms that negatively interfere with the signal model (and zero probability for atoms already selected). Of course, we cannot directly interpret the second component of (35) and (37) as a probability because it can be negative, but this alternative viewpoint helps illuminate how the pursuit of a signal model can be informed by interference.

## V. COMPUTER SIMULATIONS

We tested the interference-adaptive algorithms for the signals in Fig. 1 (except WGN) using the same dictionary of Gabor atoms and Dirac spikes. Figure 6 shows the number of atoms found by OMP and OOMP as a function of the interference weight  $\gamma$  for specific SRRs. We observed significant degradation in the models produced by MP using constant interference adaptation ( $\gamma$  constant), so we do not include those results. Models produced using interference-adaptive OMP and OOMP, however, can become much more efficient. For Attack in Fig. 6(a), the OMP model order at SRR = 60 dB decreases from 65 with

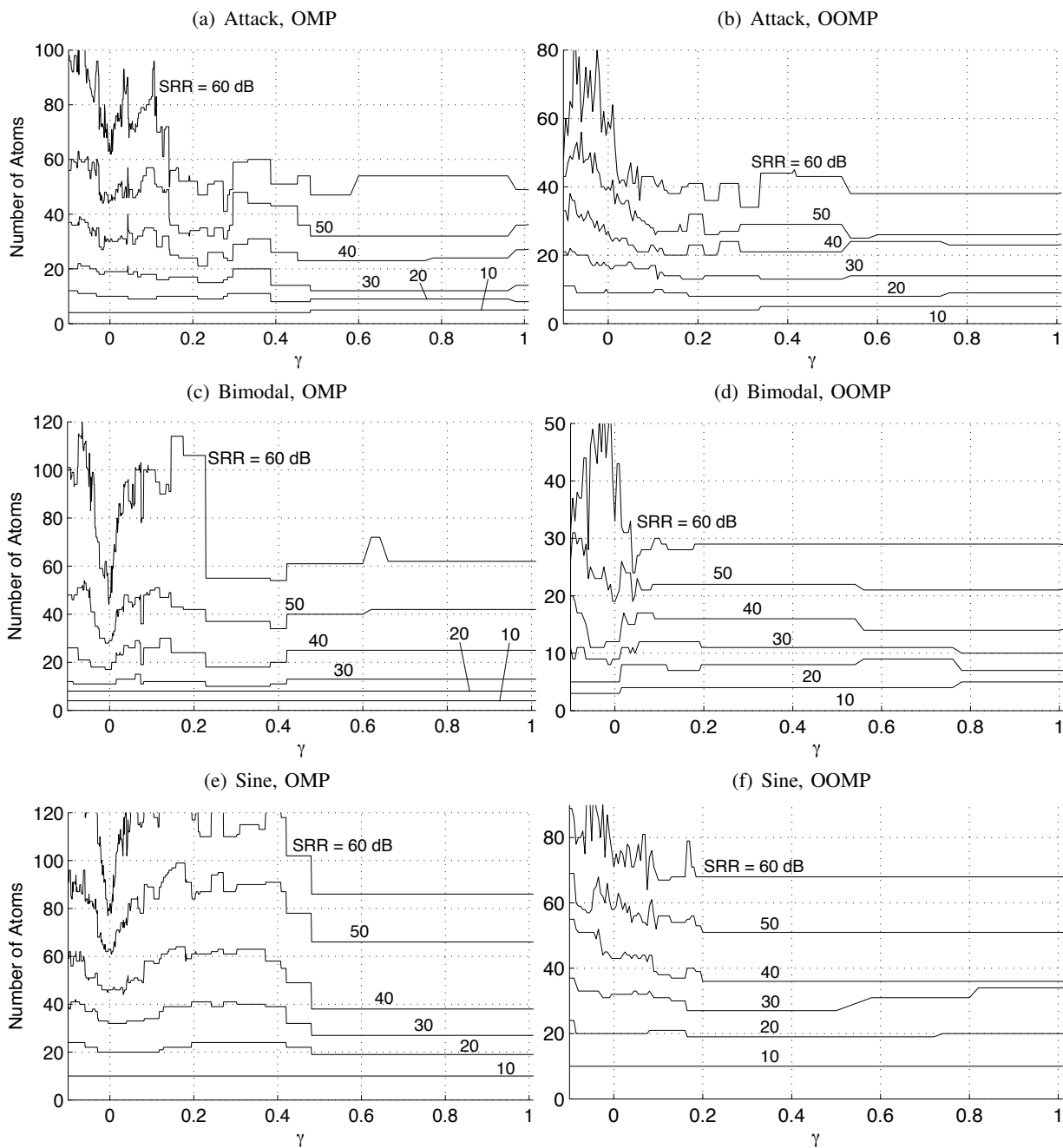


Fig. 6. Number of atoms to reach a specific SRR as function of  $\gamma$  for three signals in Fig. 1 modeled by OMP and OOMP using the same dictionary but with interference adaptation. Interference adaptation is not used when  $\gamma = 0$ . (Note changes in axes ranges.)

no interference adaptation ( $\gamma = 0$ ) to 42 atoms when  $\gamma = 0.28$ ; and the OOMP model order for Bimodal in Figure 6(d) decreases from 33 with no interference adaptation to 24 when  $\gamma = 0.04$ . On the other

hand, we see a lack of improvement in efficiency in models produced by OMP for Bimodal and Sine in Figs. 6(c) and 6(e), where the number of atoms required to reach  $\text{SRR} = 60$  dB is not reduced by interference adaptation. However, in the case of Sine in Fig. 6(e), we do see improvement at lower SRRs. For example, at  $\text{SRR} = 40$  dB the model order decreases from 46 to 38 when  $\gamma > 0.48$ . This leads us to posit that in building a signal model interference adaptation can become a hindrance, at which time  $\gamma(n)$  can be set to zero. (Indeed, later experiments revealed that if interference adaptation is considered only in the initial steps of an MP decomposition and then turned off, some of the resulting models show an increase in efficiency.) It is also clear from Fig. 6 that the nonlinear decompositions of these unit-norm signals ( $\|\mathbf{x}\|_2^2 = 1$ ) are much more sensitive to interference adaptation when  $\gamma$  is small.

Though a model produced by OMP or OOMP with interference adaptation can show a lack of improvement in efficiency, it might still better reflect the structures of the signal. Figure 7 shows three models produced by OMP with interference adaptation at small model orders and three values of  $\gamma$ . For Attack in Fig. 7(a), all the atoms appearing before the signal onset disappear, and we see a short-scale atom at the time of the onset, and atoms of longer scales in the decay. For Bimodal in Fig. 7(b), the model changes from one of much destructive interference to a more constructive solution using several smaller scale atoms. And for Sine in Fig. 7(c), although in Fig. 6(e) there is only a slight decrease in the number of atoms when  $\gamma > 0.48$  at  $\text{SRR} < 40$  dB, we argue that the new model is improved because it consists of nearly constant-amplitude half-period sinusoids that are overlapped and windowed in the middle region — which is what one expects for the most efficient and meaningful model of a stationary sinusoid using modulated symmetric windows. We also observe that the signal envelope is more uniform when using interference adaptation.

In each of these representations, the parameters of the first atom selected must be the same except for its weight. The difference with each pursuit, however, is how that first atom contributes to the modeling of the signal. Observe that for Attack and Bimodal, the amplitude of this atom becomes attenuated with order. The interference adaptive pursuit comes to “view” it as less constructive than other atoms; and with the augmentation of the signal model, the LS projection reduces the contribution of this first atom. In other words, this atom gradually becomes less relevant to the model. A post-processing step to remove such atoms is currently being considered [23].

Finally, Fig. 8 shows the wivigrams of the signal models in Fig. 7 but to the order where  $\text{SRR} = 60$  dB. Observe in Fig. 8(a) for OMP that the time-frequency distribution of energy shows a model that more directly reflects features of the signal: a sharp onset and a slow decay at a well-defined frequency. In the model of Bimodal shown in Fig. 8(b) for OMP, the two modes are resolved into two more time-

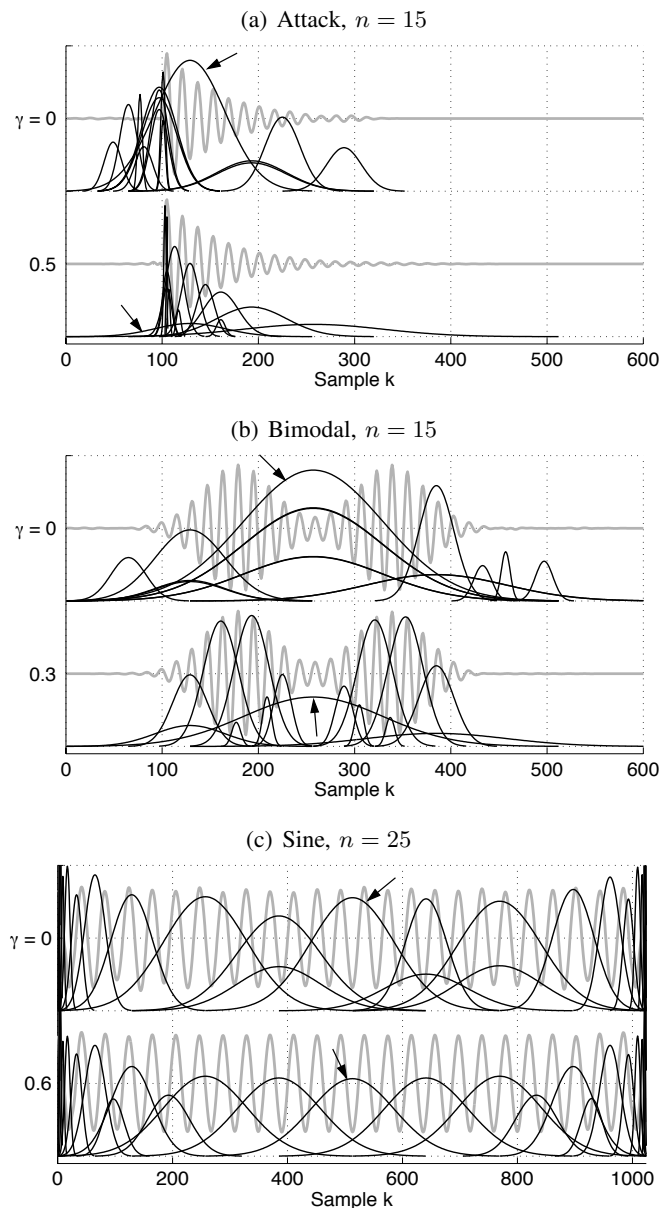


Fig. 7. Atom envelopes (black) of  $n$ th-order models created using OMP for three signals in Fig. 1, but with various values of  $\gamma$  in (35). Interference adaptation has no effect when  $\gamma = 0$ . Reconstructed signals are thick gray in background. Envelope heights are proportional to the square-root of the atom energy. Arrows point to first atom selected.

localized regions of energy. The wivigram of Sine using OMP in Fig. 8(c) shows a clearer image of the signal without the multitude of spurious atoms that appear in the representation created by OMP without interference adaptation, and without the effects around the signal edges.

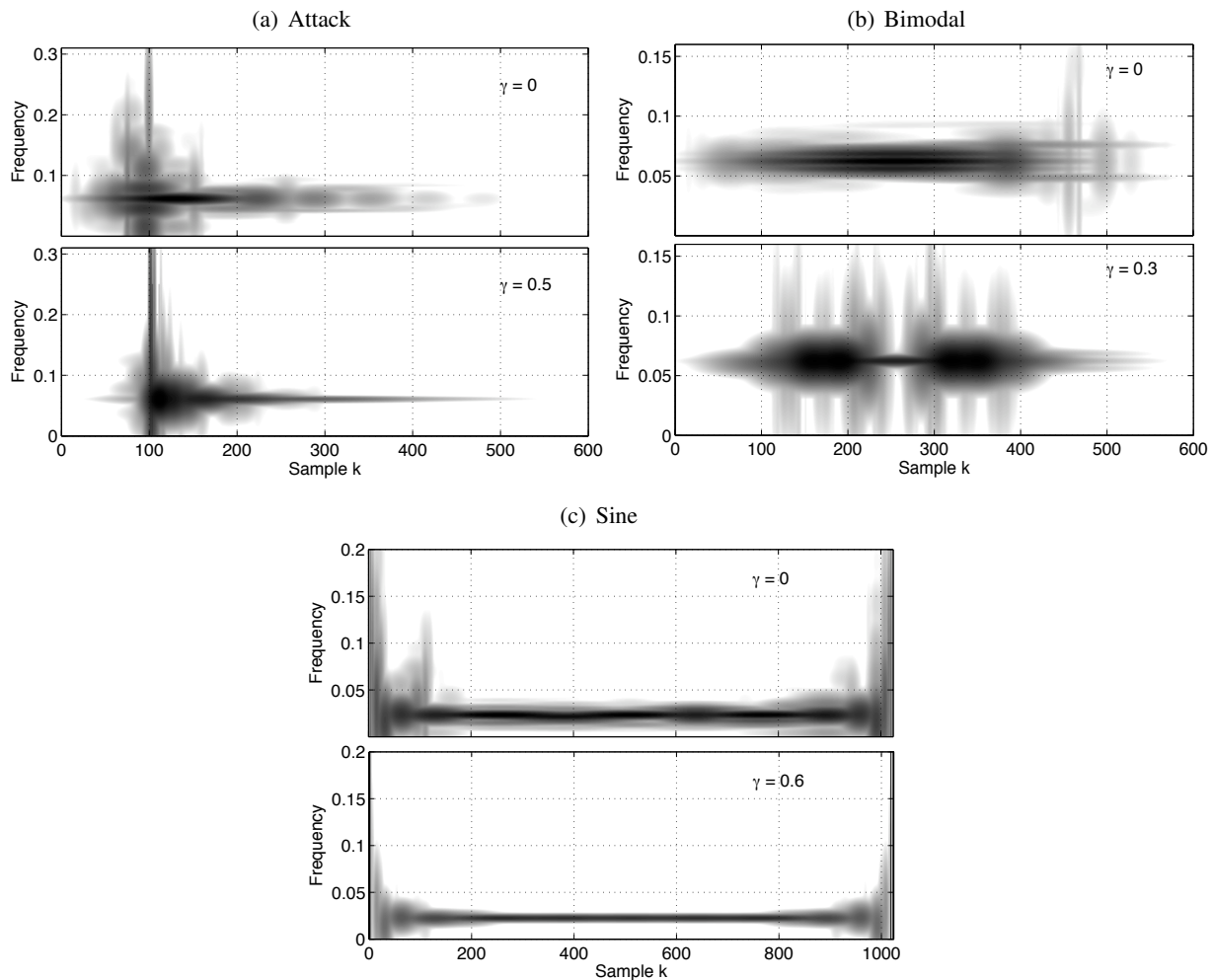


Fig. 8. Wivigrams of signal models seen in Fig. 7 created using OMP and interference adaptation for various values of  $\gamma$ . Interference adaptation has no effect when  $\gamma = 0$ .

## VI. CONCLUSION

We have presented the concept of interference in the sparse approximation of signals, and have demonstrated that its incorporation in the atom selection criterion of an iterative descent pursuit can increase the efficiency and meaningfulness of the resulting signal models. The interference of an atom in a representation describes how and to what extent it contributes to the modeling of the signal: atoms with positive interference can be viewed as representative of the signal and working in concert with other elements of the model, while atoms with negative interference can be viewed as being more “corrective” of the model than representative of the signal. Furthermore, we have proven that any convergent model created by MP must have equal amounts of destructive and constructive interference. Even though other

iterative descent approaches to sparse approximation do not suffer from this result (e.g., OMP and OOMP), there still exists a limit on the amount of constructive interference possible in an additive model of the form in (1). It makes sense that to increase the efficiency of a signal model, we should minimize the amount of correction that occurs in the decomposition. We proposed interference-adaptive iterative descent pursuits, which our computer simulations clearly show increase not only the efficiency, but also the meaningfulness of the resulting signal models, especially for those found using OMP and OOMP.

Our current work aims for a clearer understanding of the role of the interference weighting  $\gamma(n)$ , and we are developing algorithms to adapt it to the signal, or to particular content of interest within the signal. We also expect that pruning a representation of atoms that do not contribute to the signal model will increase its efficiency; various pruning strategies are currently under investigation [23]. We are also investigating the applicability of interference outside the realm of iterative descent pursuit methods, such as to the convex optimization principle of BP [3]. Finally, we are investigating the possibility of using interference to aid in the learning of dictionaries, for example, employing OMP with interference adaptation in K-SVD [24].

#### ACKNOWLEDGMENTS

The authors would like to thank Mads Christensen, Curtis Roads, and Laurent Daudet for helpful discussions. The comments from the reviewers were especially helpful for refining the figures and the language accompanying the results. This work was supported in part by the National Science Foundation under Grant CCF-0729229, and in part by Bourse Chateaubriand N. 634146B.

#### REFERENCES

- [1] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.
- [2] B. Rao, "Signal processing with the sparseness constraint," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, vol. 3, May 1998, pp. 1861–1864.
- [3] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, Aug. 1998.
- [4] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inform. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.
- [5] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Info. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
- [6] Y. Pati, R. Rezaifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, vol. 1, Pacific Grove, CA, Nov. 1993, pp. 40–44.



- [7] L. Rebollo-Neira and D. Lowe, "Optimized orthogonal matching pursuit approach," *IEEE Signal Process. Lett.*, vol. 9, no. 4, pp. 137–140, Apr. 2002.
- [8] S. Chen and J. Wigger, "Fast orthogonal least-squares algorithm for efficient subset model selection," *IEEE Trans. Signal Process.*, vol. 43, no. 7, pp. 1713–1715, July 1995.
- [9] P. Vincent and Y. Bengio, "Kernel matching pursuit," *Machine Learning*, vol. 48, no. 1, pp. 165–187, July 2002.
- [10] R. Gribonval, E. Bacry, S. Mallat, P. Depalle, and X. Rodet, "Analysis of sound signals with high resolution matching pursuit," in *Proc. IEEE-SP Int. Symp. Time-Freq. Time-Scale Anal.*, Paris, France, June 1996, pp. 125–128.
- [11] S. Jaggi, W. C. Karl, S. Mallat, and A. S. Willsky, "High resolution pursuit for feature extraction," *Applied and Computational Harmonic Analysis*, vol. 5, no. 4, pp. 428–449, Oct. 1998.
- [12] M. M. Goodwin and M. Vetterli, "Matching pursuit and atomic signal models based on recursive filter banks," *IEEE Trans. Signal Process.*, vol. 47, no. 7, pp. 1890–1902, July 1999.
- [13] P. J. Durka, D. Ircha, and K. J. Blinowska, "Stochastic time-frequency dictionaries for matching pursuit," *IEEE Trans. Signal Process.*, vol. 49, no. 3, pp. 507–510, Mar. 2001.
- [14] B. L. Sturm, J. J. Shynk, L. Daudet, and C. Roads, "Dark energy in sparse atomic estimations," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 3, pp. 671–676, Mar. 2008.
- [15] B. L. Sturm, J. J. Shynk, and L. Daudet, "A short-term measure of dark energy in sparse atomic estimations," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 2007, pp. 1126–1129.
- [16] —, "Measuring interference in sparse atomic estimations," in *Proc. Conf. Info. Sciences Syst.*, Princeton, NJ, Mar. 2008, pp. 961–966.
- [17] B. L. Sturm, "Sparse approximation and atomic decomposition: Considering atom interactions in evaluating and building signal representations," Ph.D. dissertation, University of California, Santa Barbara, CA, Mar. 2009.
- [18] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, 3rd ed. Amsterdam, The Netherlands: Academic Press, Elsevier, 2009.
- [19] L. Daudet, "Sparse and structured decompositions of signals with the molecular matching pursuit," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 5, pp. 1808–1816, Sept. 2006.
- [20] M. M. Goodwin, "Adaptive signal models: Theory, algorithms, and audio applications," Ph.D. dissertation, University of California, Berkeley, Berkeley, CA, Fall 1997.
- [21] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.
- [22] R. Gribonval, "Approximations non-linéaires pour l'analyse des signaux sonores," Ph.D. dissertation, Université de Paris IX Daupine, Paris, France, Sep. 1999.
- [23] B. L. Sturm, J. J. Shynk, and D. H. Kim, "Pruning a sparse approximation using interference," in *Proc. Conf. Info. Sciences Syst.*, Baltimore, MD, Mar. 2009.
- [24] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov 2006.