

Dark Energy in Sparse Atomic Estimations

Bob L. Sturm, *Student Member, IEEE*, John J. Shynk*, *Senior Member, IEEE*,
Laurent Daudet, *Member, IEEE* and Curtis Roads

Abstract—Sparse atomic decomposition algorithms, such as matching pursuit, attempt to find an efficient estimation of a signal using terms (atoms) selected from an overcomplete dictionary. In some cases, atoms can be selected that have energy in regions of the signal that have no energy. Other atoms are then used to destructively interfere with these terms in order to preserve the original waveform. Because some terms may even “disappear” in the reconstruction, we refer to the destructive and constructive interference between the atoms of a sparse atomic estimation as “dark energy.” In this paper, we formally define dark energy for matching pursuit, explore its properties, and present empirical results for decompositions of audio signals. This work demonstrates that dark energy is a useful measure of the interference between the terms of a sparse atomic estimation, and might provide information for the decomposition process.

Index Terms—Sparse overcomplete methods, signal estimation, matching pursuit. EDICS: AUD-ANSY

I. INTRODUCTION

A SPARSE atomic estimation of a signal is a linear combination of functions with finite support called atoms chosen from an overcomplete dictionary. Sparse overcomplete methods include basis pursuit [1], matching pursuit (MP) [2], orthogonal MP [3], high resolution MP [4], [5], and psychoacoustically weighted MP [6]. These methods attempt to estimate a signal in a way that is sparse, efficient, and robust, as well as meaningful. Sparse overcomplete methods have been used in many applications, such as compression and coding of audio [7], speech [8], images [9], and video [10]; analysis of multimodal signals [11]; blind source separation [12]; and modification of audio signals [13].

Sparse decomposition algorithms like MP iteratively build a signal model using atoms selected in a nonlinear and locally greedy manner. Some terms in the estimation, however, arise as a result of this greediness. In the extreme case, some selected atoms may exist in regions where the original signal possesses no energy. These terms remain a part of the estimation but destructively interfere and “disappear” in the reconstruction. Although other atoms may not disappear, they may still exist to “correct” other atoms through constructive

and destructive interference to preserve the waveform of the original signal. Because of the disappearance of atoms in the extreme case, we refer to interference between the terms of a sparse atomic estimation as *dark energy* (DE).

Several researchers have noted the tendency of MP to correct itself using terms that interfere with each other [1], [4], [5], [13], [14]. For example, when using a dictionary of symmetric atoms to decompose an audio signal, MP has a propensity to place atoms preceding transients. A simple example of this is seen in Fig. 1. This behavior affects the efficiency and usefulness of a sparse estimation because some terms may not be real features of a signal [1], [4], [5], and may result in detrimental artifacts [13], [14]. Thus, minimizing the number of these terms or the extent of their corrections should benefit an estimation.

Various methods have been proposed to diminish this behavior, such as using a particular dictionary [14], or instead by selecting atoms in a different manner than MP [1], [4], [5]. There is much evidence that with a dictionary “tuned” to the types of structures expected in a signal, such as a union of short-scale wavelets and large-scale windowed cosines for musical audio signals [15], or even learning the dictionary from a set of test signals [16], that the resulting estimations are much more sparse, efficient, robust, and meaningful. However, these estimations may still not be free of interfering terms since they are built from overcomplete sets of functions. Instead of attempting to mitigate DE, we are here interested in quantifying its existence in a sparse atomic estimation to learn about the signal, the dictionary, the estimation algorithm, and the quality of the results. The ultimate goal of this work is to use DE to improve the sparse atomic estimation of signals.

The rest of this paper is organized as follows. Section II provides an overview of a sparse signal model and briefly reviews MP. In Section III, we present a formal measure of DE and discuss its significance. Section IV presents some empirical results for a set of real audio signals. Finally, conclusions of this work and an outline for future directions are provided in Section V.

II. SPARSE SIGNAL MODEL AND MATCHING PURSUIT

A sparse signal model uses a set of pre-defined functions drawn from an overcomplete dictionary. A signal vector \mathbf{x} in a real K -dimensional subspace $\mathcal{X} \in \mathbb{R}^K$ is modeled by an n th-order linear combination of these functions

$$\mathbf{x} = \tilde{\mathbf{x}}(n) + \mathbf{r}(n) = \mathbf{G}(n)\mathbf{a}(n) + \mathbf{r}(n) \quad (1)$$

where $\mathbf{G}(n) \triangleq [\mathbf{g}_0 | \dots | \mathbf{g}_{n-1}]$ is a $K \times n$ matrix whose columns are functions (atoms) $\{\mathbf{g}_i \in \mathcal{X} : \|\mathbf{g}_i\| = 1\}$ selected from a larger real $K \times N$ matrix \mathbf{D} (dictionary) with rank K

B. L. Sturm and J. J. Shynk (*contact author) are with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA, 93106-9560 USA. E-mail: {boblsturm, shynk}@ece.ucsb.edu. Phone: 805-893-3977. Fax: 805-893-3262.

L. Daudet is with the University Pierre and Marie Curie, Paris 6, IJLRA/LAM, 11 rue de Lourmel 75015, Paris, France. E-mail: daudet@lam.jussieu.fr. Phone: +33 (0)1 53 95 43 35. Fax: +33 (0)1 45 77 16 59.

C. Roads is with the Media Arts and Technology Program, University of California, Santa Barbara, CA 93106-6065 USA. E-mail: clang@mat.ucsb.edu. Phone: 805-893-5244. Fax: 805-893-2930.

This work was supported in part by the National Science Foundation under Grant CCF-0729229.

Manuscript accepted Nov. 7, 2007; revised Dec. 7, 2007.

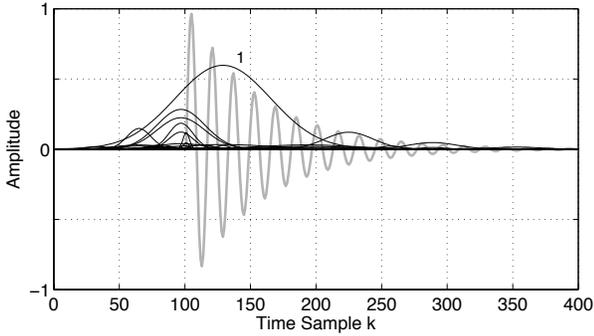


Fig. 1. A transient signal (gray) is decomposed by MP over a dictionary of Gabor atoms (envelopes in black). Several terms of the decomposition exist prior to the signal onset and remove the error introduced by the first selected term (labeled).

(generally $N \gg K$), and $\mathbf{a}(n) \triangleq [a_0, \dots, a_{n-1}]^T$ is a vector of the expansion coefficients. The residual is $\mathbf{r}(n) \triangleq \mathbf{x} - \tilde{\mathbf{x}}(n)$ (with $\mathbf{r}(0) \equiv \mathbf{x}$ and $\tilde{\mathbf{x}}(0) = \mathbf{0}$), the approximation is $\tilde{\mathbf{x}}(n)$, and the estimation is $\{\mathbf{G}(n), \mathbf{a}(n)\}$, all for the n th-order. We emphasize that n is the order of the estimation or the decomposition iteration, and is not a time index of the original signal (for which we use k).

MP is an iterative descent algorithm that offers a tractable method for finding a good solution to (1), though suboptimal in the L_2 sense [2]. It estimates \mathbf{x} by iteratively selecting atoms from a dictionary \mathbf{D} using the following criterion:

$$\mathbf{g}_n = \arg \max_{\mathbf{g} \in \mathbf{D}} |\mathbf{g}^T \mathbf{r}(n)|. \quad (2)$$

At step n once \mathbf{g}_n is found, the estimate is updated by appending the atom and the expansion coefficient to the previous estimate: $\mathbf{G}(n+1) = [\mathbf{G}(n) | \mathbf{g}_n]$ and $\mathbf{a}(n+1) = [\mathbf{a}^T(n), a_n]^T$. When the new expansion coefficient is given by

$$a_n = \mathbf{g}_n^T \mathbf{r}(n), \quad (3)$$

the energy of the new residual $\|\mathbf{r}(n)\|^2 = \|\mathbf{x} - \tilde{\mathbf{x}}(n)\|^2$ is minimized [2]. The decomposition can be stopped after a specified number of steps, or after $\|\mathbf{r}(n)\|^2 \leq \epsilon$ for some $\epsilon > 0$.

Modeling the variety of structures found in real signals by using orthogonal functions is often disadvantageous [2]. Thus, an important advantage of sparse overcomplete methods is the freedom to choose the structure of \mathbf{D} without an orthogonality restriction. Obviously, this has a significant impact on the performance and properties of a decomposition algorithm, as well as the estimations. An overcomplete dictionary that is well-correlated with the expected structures in a class of signals can result in significantly sparser and more efficient and meaningful estimations than can be obtained by using an orthonormal basis [1], [2], [4], [9], [15], [16].

In the computer simulations presented later, we use dictionaries designed from three different types of atoms: Gabor atoms, damped oscillators, and damped sinusoids. Each atom has the same form:

$$g(k; u, s, \omega, \phi) = Aw(k - u; s) \cos[\omega(k - u)T + \phi] \quad (4)$$

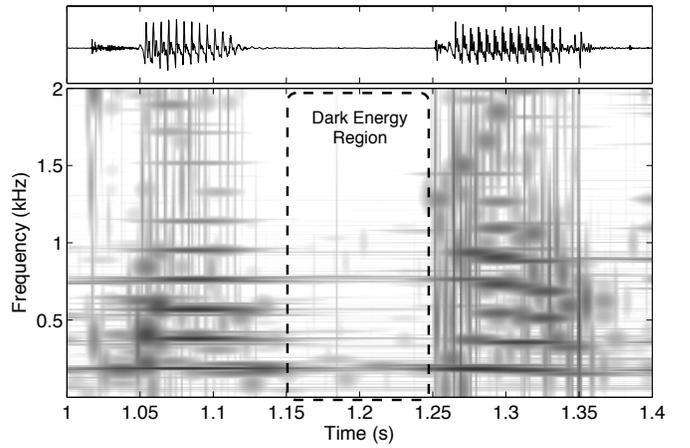


Fig. 2. A portion of speech (top) is decomposed over a dictionary of Gabor atoms. The wivigram (bottom) displays the distribution of energy of the estimation in the time-frequency plane. The dashed region shows instances of dark energy.

where $0 \leq k < K$ is a time index, $0 \leq \omega \leq \pi$ is a normalized frequency, $0 \leq u < K - s/2$ is a translation in samples, $1 \leq s \leq K$ is a scale in samples, T is the sampling period, $0 \leq \phi < 2\pi$ is a phase offset, and the scalar A is set such that $\sum_k |g(\cdot)|^2 = 1$. The $(k+1)$ st element of the atom vector \mathbf{g} is $g(k; u, s, \omega, \phi)$. The function $w(k; s)$ is a window. A Gabor atom uses a truncated and sampled Gaussian window:

$$w(k; s) = \begin{cases} \exp\left(-\frac{(k-s/2)^2}{2(\alpha s)^2}\right), & k = 0, \dots, s-1 \\ 0, & k = s, \dots, K-1 \end{cases} \quad (5)$$

where $\alpha > 0$ controls the variance. A damped oscillator has a truncated exponential window:

$$w(k; s) = \begin{cases} \exp(-\beta k/s), & k = 0, \dots, s-1 \\ 0, & k = s, \dots, K-1 \end{cases} \quad (6)$$

where $\beta > 0$ controls the decay. A damped sinusoid is a damped oscillator with the restriction that $\phi = -\pi/2$ in (4).

III. DARK ENERGY

The signal shown at the top of Fig. 2 is decomposed by MP over a dictionary of Gabor atoms. The lower image shows the superposition of the Wigner-Ville distribution of each atom [2] in the estimation, which we refer to as a *wivigram*. The dashed region contains several atoms where the original signal has little energy and few atoms are expected in the estimation. These atoms, or parts thereof, destructively interfere and disappear in the approximation. The energy in this time-frequency region can be found by a double integration over the wivigram (which is phase-invariant) over all frequencies and the given time limits. Similarly, one may find the energy of the approximation in this region by summing the corresponding squared sample values of $\tilde{\mathbf{x}}(n)$. Figure 3 shows how the energy in the wivigram and of the approximation change as functions of the estimation order n for two different time regions in Fig. 2.

Although MP is nonlinear, energy is conserved at each step

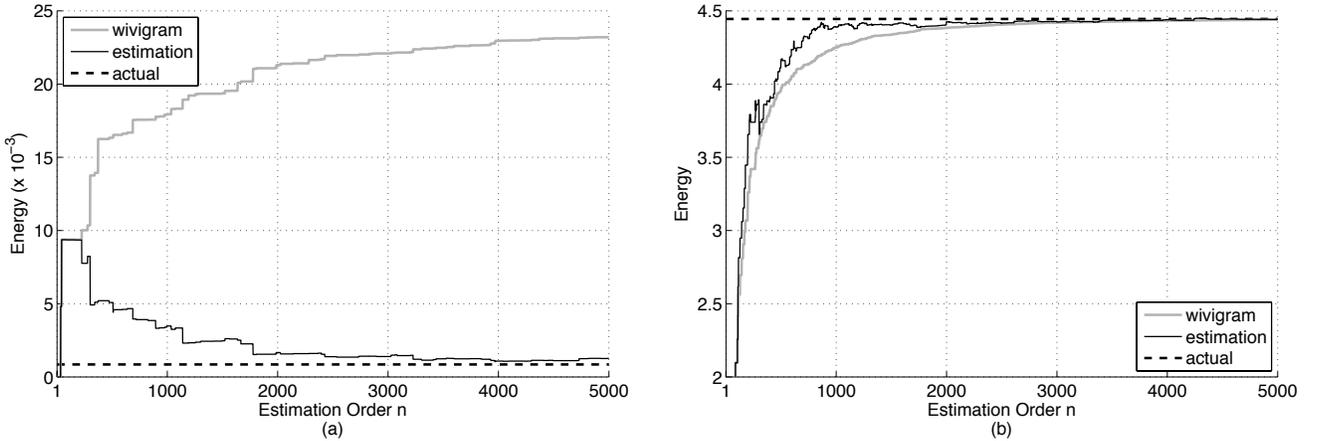


Fig. 3. For the estimation shown in Fig. 2, we plot the energies of the approximation $\tilde{\mathbf{x}}(n)$ (thin line), the actual signal (dashed line) \mathbf{x} , and in the wivigram (thick gray line) over two different time regions. (a) $1.15 \leq t \leq 1.25$ s. (b) $1.0 \leq t \leq 1.4$ s.

as in a linear expansion over an orthogonal basis [2], i.e.,

$$\|\mathbf{x}\|^2 = \|\mathbf{a}(n)\|^2 + \|\mathbf{r}(n)\|^2, \quad n = 0, 1, \dots \quad (7)$$

Observe, however, that in some time regions, as shown in Fig. 3(a), the energy in the wivigram actually grows larger than that of the original signal and the approximation. Over a longer time range of the signal, the energy in the wivigram asymptotically approaches that of the original signal from below. This local discrepancy is due to atoms in the estimation that destructively and constructively interfere with other atoms. It is this behavior of an overcomplete signal decomposition that we wish to characterize. Next, we propose a measure of the interference within a sparse atomic estimation found using MP. This will provide information about how well a dictionary and the decomposition estimate the original signal both globally and locally, as an alternative to the global mean-square error.

A. Definition of Dark Energy in Matching Pursuit

Consider the signal model in (1) where the atoms are selected using (2), and the expansion coefficients are given by (3). At step n we would like to measure the effect of adding the new atom \mathbf{g}_{n-1} to the current approximation $\tilde{\mathbf{x}}(n-1)$ to generate the new approximation $\tilde{\mathbf{x}}(n) = \tilde{\mathbf{x}}(n-1) + a_{n-1}\mathbf{g}_{n-1}$. Although this atom removes the most residual energy, it is not obvious how it changes the energy of the approximation.

Consider the difference between the energy of the new approximation and the energy that would result if the new atom is orthogonal to the current approximation:

$$\Delta(n) \triangleq \|\tilde{\mathbf{x}}(n)\|^2 - (\|\tilde{\mathbf{x}}(n-1)\|^2 + |a_{n-1}|^2) \quad (8)$$

$$= 2a_{n-1}\mathbf{g}_{n-1}^T\tilde{\mathbf{x}}(n-1), \quad n = 1, 2, \dots \quad (9)$$

where (1) and (3) have been substituted. This quantity describes how much energy the new atom “imparts” to the current approximation. If the atoms in the dictionary are orthogonal, then $\Delta(n) = 0$ for all n , and each atom will impart all of its energy. In other words, the difference in the energy between subsequent approximations will always

be equal to the energy of the new scaled atom. This is not the case for an overcomplete (and thus non-orthogonal) dictionary because the terms can interfere with each other. The amount of this interference is measured by the correlation-like relation in (9). An example realization of $\Delta(n)$ is shown in Fig. 4(a). When $\Delta(n) < 0$, the n th selected atom destructively interferes with the current approximation, i.e., the energy of the new approximation is less than that of the current approximation summed with that of the scaled atom. Conversely, $\Delta(n) > 0$ signifies that the new atom constructively interferes with the approximation.

Although $\Delta(n)$ has units of energy, its interpretation as such is difficult because it can be negative. As a result, we define the DE associated with the addition of \mathbf{g}_{n-1} to the current approximation as the magnitude of $\Delta(n)$:

$$\Xi(n) \triangleq |\Delta(n)| = 2|a_{n-1}\mathbf{g}_{n-1}^T\tilde{\mathbf{x}}(n-1)|. \quad (10)$$

DE is thus proportional to the magnitude of the correlation of the selected atom with the current approximation of \mathbf{x} , i.e., it is a measure of the extent to which the selected atom is already represented in $\tilde{\mathbf{x}}(n-1)$. When the DE is nonzero, then \mathbf{g}_{n-1} does not contain entirely new “information” with respect to $\tilde{\mathbf{x}}(n-1)$, even though, via (2), that atom is the most correlated with the residual $\mathbf{r}(n-1)$ of all atoms in \mathbf{D} . Note that unlike (9), DE does not discriminate between constructive and destructive interference; it is a measure of either type of interference for the atom as a whole.

B. Cumulative Interference and Dark Energy

The quantities in (9) and (10) can be summed to provide cumulative measures of the interference between terms in an n th-order estimation. These sums are defined as

$$\Delta_{\Sigma}(n) \triangleq \sum_{m=1}^n \Delta(m) \quad (11)$$

$$\Xi_{\Sigma}(n) \triangleq \sum_{m=1}^n \Xi(m) \quad (12)$$

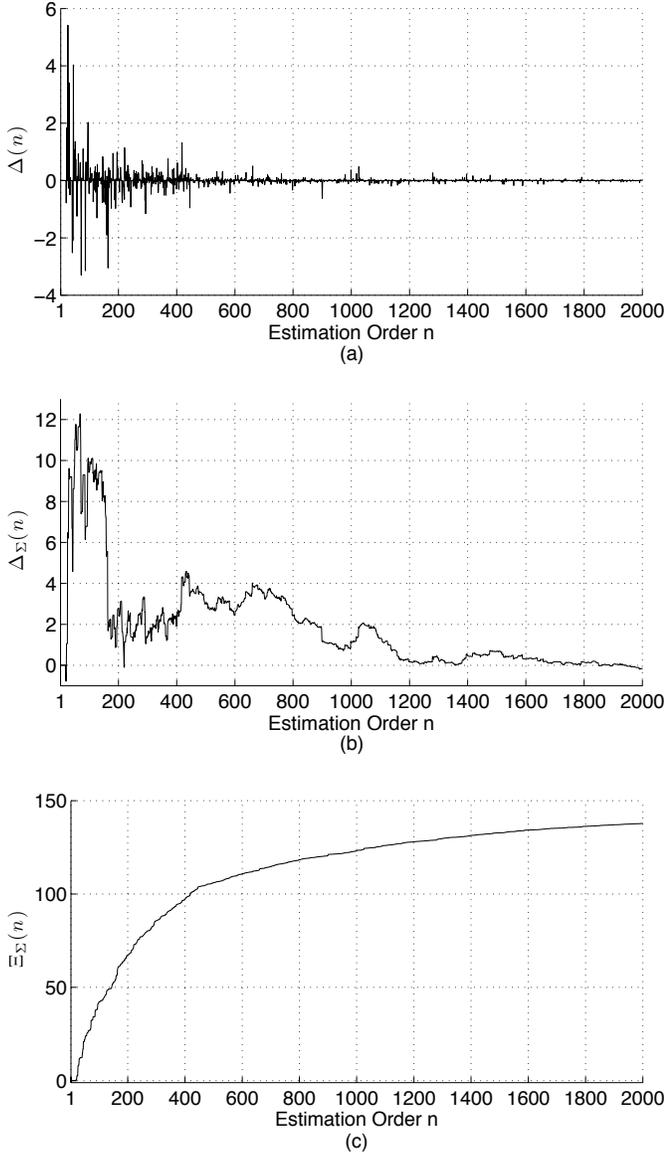


Fig. 4. Interference and DE trajectories for the signal shown in Fig. 2. (a) $\Delta(n)$. (b) $\Delta_{\Sigma}(n)$. (c) $\Xi_{\Sigma}(n)$.

where $\Delta_{\Sigma}(n)$ (which can be positive or negative) is the cumulative interference in the n th-order estimation, and $\Xi_{\Sigma}(n) \geq 0$ is the corresponding cumulative DE (CDE). Examples of these for the signal in Fig. 2 are shown in Fig. 4(b) and Fig. 4(c), respectively. Observe in Fig. 4(b) that significant positive interference occurs in the first hundred iterations. The CDE shown in Fig. 4(c) grows most rapidly over these iterations, but is asymptotically convergent.

C. Convergence of Dark Energy

Assuming that the dictionary \mathbf{D} is complete for some finite-dimensional subspace $\mathcal{X} \in \mathbb{R}^K$, we now show that DE is bounded above by a decaying exponential and converges to zero. Using the Cauchy-Schwartz inequality, we can write

$$\Xi(n+1) = 2|a_n \mathbf{g}_n^T \tilde{\mathbf{x}}(n)| \leq 2|a_n| \|\tilde{\mathbf{x}}(n)\| \quad (13)$$

since each atom has unit norm. Because \mathbf{D} is assumed to be complete, $\lim_{n \rightarrow \infty} \|\tilde{\mathbf{x}}(n)\| = \|\mathbf{x}\|$, and thus

$$\lim_{n \rightarrow \infty} \Xi(n+1) \leq 2\|\mathbf{x}\| \lim_{n \rightarrow \infty} |a_n|, \quad (14)$$

which demonstrates that DE is bounded from above by the expansion coefficients. It is shown in [2] that the magnitude of the expansion coefficients is bounded from above by a decaying exponential, which we reproduce in part below.

Consider the following ‘‘correlation ratio’’ [2]:

$$\lambda(\mathbf{x}) \triangleq \max_{\mathbf{g} \in \mathbf{D}} \frac{|\mathbf{g}^T \mathbf{x}|}{\|\mathbf{x}\|}. \quad (15)$$

Clearly, $\lambda(\mathbf{x}) \leq 1$; equality is achieved when \mathbf{D} contains $\mathbf{x}/\|\mathbf{x}\|$. From (2) and Lemma 1 in [2],

$$\inf_{\mathbf{x} \in \mathcal{X}} \lambda(\mathbf{x}) \triangleq I(\lambda) > 0. \quad (16)$$

This implies that for $\{\mathbf{x} \in \mathcal{X} : \|\mathbf{x}\| \neq 0\}$, the minimum of the maximum magnitude correlation with any $\mathbf{g} \in \mathbf{D}$ is > 0 . Lemma 2 in [2] relates $I(\lambda)$ to the norm residual as follows:

$$\frac{\|\mathbf{r}(n)\|}{\|\mathbf{x}\|} \leq [1 - I^2(\lambda)]^{n/2} \triangleq \Lambda^n. \quad (17)$$

Thus, the norm of the residual decays exponentially at a rate proportional to $\Lambda \triangleq \sqrt{1 - I^2(\lambda)}$. Note that $0 \leq \Lambda < 1$, and is strictly < 1 because of (16).

Since in MP the expansion coefficients are found using (3), we can write using the Cauchy-Schwartz inequality

$$|a_n| = |\mathbf{g}_n^T \mathbf{r}(n)| \leq \|\mathbf{g}_n\| \|\mathbf{r}(n)\| = \|\mathbf{r}(n)\|. \quad (18)$$

Finally, by substituting (17) into (14), we see that

$$\begin{aligned} \lim_{n \rightarrow \infty} \Xi(n+1) &\leq 2\|\mathbf{x}\| \lim_{n \rightarrow \infty} \|\mathbf{r}(n)\| \\ &\leq 2\|\mathbf{x}\|^2 \lim_{n \rightarrow \infty} \Lambda^n = 0. \end{aligned} \quad (19)$$

This proves that for $\mathbf{x} \in \mathcal{X}$ decomposed over \mathbf{D} , the DE in the estimation converges exponentially to zero. This behavior is seen in Fig. 4(a).

D. Balance of Positive and Negative Interference

Next, we show that for $\mathbf{x} \in \mathcal{X}$ the cumulative interference $\Delta_{\Sigma}(n)$ in (11) converges to zero. This means that in a convergent estimation, there must be equal constructive and destructive interference. We can rewrite (11) as follows:

$$\begin{aligned} \frac{1}{2} \Delta_{\Sigma}(n) &= \sum_{m=0}^{n-1} a_m \mathbf{g}_m^T [\mathbf{x} - \mathbf{r}(m)] \\ &= \sum_{m=0}^{n-1} a_m \mathbf{g}_m^T \mathbf{x} - \sum_{m=0}^{n-1} a_m^2 \\ &= \mathbf{x}^T \sum_{m=0}^{n-1} a_m \mathbf{g}_m - \|\mathbf{a}(n)\|^2 \\ &= \mathbf{x}^T \tilde{\mathbf{x}}(n) - \|\mathbf{a}(n)\|^2 \end{aligned} \quad (20)$$

$$= \mathbf{x}^T \tilde{\mathbf{x}}(n) - \|\mathbf{a}(n)\|^2 \quad (21)$$

where (3) has been substituted in (20). Using (7) yields

$$\frac{1}{2}\Delta_{\Sigma}(n) = \mathbf{x}^T \tilde{\mathbf{x}}(n) - (\|\mathbf{x}\|^2 - \|\mathbf{r}(n)\|^2). \quad (22)$$

In the limit as $\tilde{\mathbf{x}}(n) \rightarrow \mathbf{x}$, the norm of the residual converges to zero and thus

$$\lim_{n \rightarrow \infty} \frac{1}{2}\Delta_{\Sigma}(n) = \|\mathbf{x}\|^2 - (\|\mathbf{x}\|^2 - 0) = 0. \quad (23)$$

This result shows that, as expected, there are equal amounts of constructive and destructive interference in a convergent estimation. The plot $\Delta_{\Sigma}(n)$ in Fig. 4(b) hints at this result.

E. Upper Bound for the Cumulative Dark Energy

Since DE converges exponentially to zero, it necessarily follows that the CDE (12) is also bounded and convergent. We now find an upper bound. From (17) and (18), we can write

$$|a_n| = |\mathbf{g}_n^T \mathbf{r}(n)| \leq \|\mathbf{r}(n)\| \leq \|\mathbf{x}\| \Lambda^n. \quad (24)$$

Substituting this result into (10) yields

$$\Xi(n+1) = 2|a_n| |\mathbf{g}_n^T \tilde{\mathbf{x}}(n)| \quad (25)$$

$$\leq 2\Lambda^n \|\mathbf{x}\| |\mathbf{g}_n^T \mathbf{G}(n) \mathbf{a}(n)|. \quad (26)$$

Using the Cauchy-Schwartz inequality on the last term on the right-hand side, we have that $|\mathbf{g}_n^T \mathbf{G}(n) \mathbf{a}(n)| \leq \|\mathbf{g}_n^T \mathbf{G}(n)\| \|\mathbf{a}(n)\|$. From the energy conservation property in (7), it is easy to see that $\|\mathbf{a}(n)\| \leq \|\mathbf{x}\|$. Also, $\|\mathbf{g}_n^T \mathbf{G}(n)\| < \sqrt{n}$; the upper bound occurs when $\mathbf{G}(n)$ contains n columns of \mathbf{g}_n , but this is prohibited by the selection criterion of MP. Combining these results gives

$$\Xi(n+1) < 2\sqrt{n}\Lambda^n \|\mathbf{x}\|, \quad (27)$$

which after substituting into (12), yields the following upper bound for the CDE:

$$\Xi_{\Sigma}(n) < 2\|\mathbf{x}\| \sum_{m=0}^{n-1} \sqrt{m}\Lambda^m. \quad (28)$$

Since $0 \leq \Lambda < 1$ the CDE necessarily converges for $\mathbf{x} \in \mathcal{X}$. The value at convergence depends upon a complex relationship between the dictionary and the signal, as expressed in (15).

IV. EMPIRICAL RESULTS AND DISCUSSION

In the following computer simulations, ten audio signals (sampled at 44.1 kHz) were decomposed by MP implemented using the Matching Pursuit Toolkit [17]. Each signal was decomposed over one of three dictionaries to a maximum signal-to-residual ratio (SRR) of $20 \log_{10}(\|\mathbf{x}\|/\|\mathbf{r}(n)\|) = 60$ dB. Each dictionary consists of atoms with scales that are powers of two, ranging from 2^2 to $2^{14} = 16,384$ samples, corresponding to the range [0.1, 370] ms. Each atom has a translation in samples that is an integer multiple of one quarter of its scale. For example, atoms with a scale of 32 samples have translations that are integer multiples of 8 samples. Furthermore, each translated atom of scale s has modulation frequency $\omega_i = i2\pi/s$, $i = 0, 1, \dots, s/2$. In the dictionaries of damped oscillators and damped sinusoids, we set $\alpha = 0.1$ and $\beta = 10$ in (5) and (6), respectively.

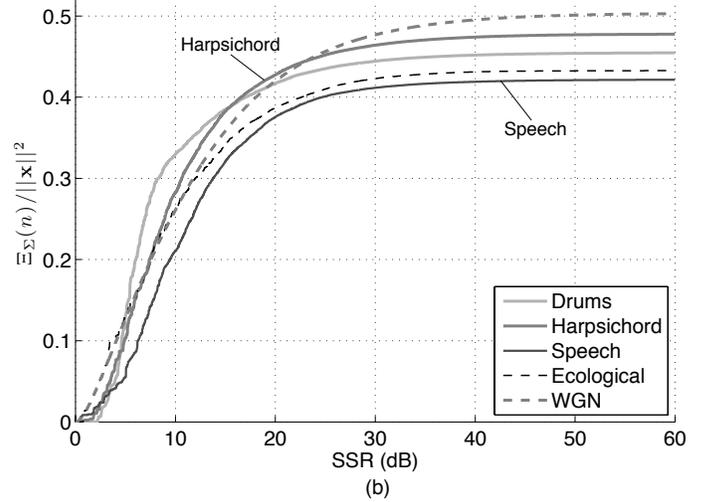
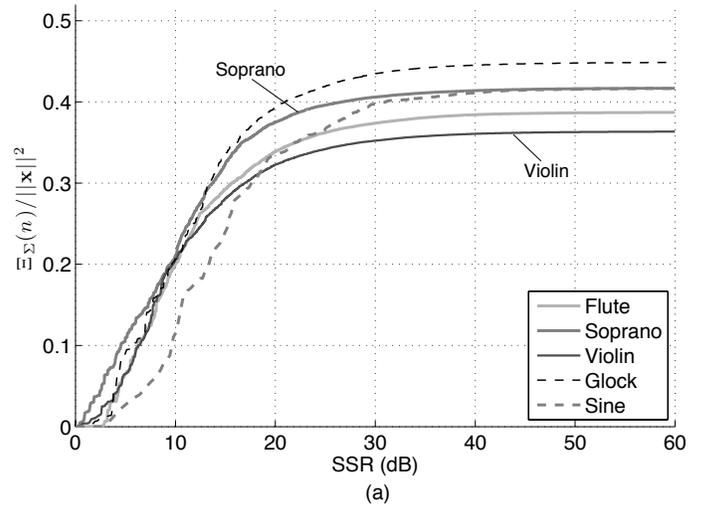


Fig. 5. Convergence of the CDE ratio $\Xi_{\Sigma}(n)/\|\mathbf{x}\|^2$ as a function of SRR for several audio test signals. (a) Flute, Soprano, Violin, Glockenspiel, and Sine. (b) Drums, Harpsichord, Speech, Ecological, and White Gaussian Noise.

Figure 5 shows the ratio of the CDE and the signal energy $\Xi_{\Sigma}(n)/\|\mathbf{x}\|^2$ as a function of the SRR for the decomposition of each signal over a dictionary of Gabor atoms. Note that Ecological is a field recording of birds; Drums includes cymbals, snare, and bass drums; Harpsichord is polyphonic; and Sine is a sine wave. The estimation of Sine has the lowest CDE ratio in the range [0, 17] dB SRR, and that of Drums has the highest in the range [5, 15] dB SRR. This result is intuitive because the residual of Sine is similar to a Gabor atom until most of the energy is represented, and MP begins selecting interfering atoms to preserve the constant envelope of the original waveform. For Drums, since the signal energy is spread widely across the spectrum, many Gabor atoms are needed because each one represents essentially one frequency with a small time support. As a result, there will be more atoms initially interfering with each other than for any of the other signals considered, except perhaps the white Gaussian noise (WGN).

Table I summarizes some statistics of the signals and their

TABLE I
STATISTICS OF THE SIGNALS AND ESTIMATIONS AT ITERATION n FOR $SRR = 60$ dB USING THREE DICTIONARIES. MINIMUM VALUES ACROSS DICTIONARIES ARE DENOTED BY \star .

	Signal (s)	Order n	MAS	CDE Ratio	Signal (s)	Order n	MAS	CDE Ratio
G	Flute (5.8)	76,742	13,271	0.387 \star	Sine (4.0)	795	199 \star	0.417
	Soprano (11.1)	91,547	8,249 \star	0.417	Speech (8.1)	107,786	13,243	0.422
	Violin (13.9)	163,763	11,805	0.363	Ecological (14.7)	238,092	16,249	0.433
	Glock (8.8)	41,292	4,678 \star	0.449	Harpichord (8.1)	153,834	18,904	0.478
	Drums (5.3)	103,355	19,524	0.455	WGN (4.0)	178,174	44,548	0.503
G \cup O	Flute (5.8)	80,488	13,919	0.400	Sine (4.0)	846	212	0.407 \star
	Soprano (11.1)	96,467	8,692	0.416	Speech (8.1)	116,582	14,323	0.413 \star
	Violin (13.9)	169,723	12,235	0.373	Ecological (14.7)	253,092	17,273	0.441
	Glock (8.8)	43,603	4,939	0.501	Harpichord (8.1)	162,401	19,957	0.484
	Drums (5.3)	110,407	20,856	0.444	WGN (4.0)	191,606	47,902	0.490 \star
G \cup S	Flute (5.8)	76,332	13,200 \star	0.408	Sine (4.0)	817	204	0.415
	Soprano (11.1)	92,007	8,290	0.412 \star	Speech (8.1)	107,582	13,218 \star	0.421
	Violin (13.9)	163,293	11,771 \star	0.359 \star	Ecological (14.7)	235,914	16,101 \star	0.431 \star
	Glock (8.8)	41,604	4,713	0.441 \star	Harpichord (8.1)	152,828	18,780 \star	0.476 \star
	Drums (5.3)	102,957	19,448 \star	0.436 \star	WGN (4.0)	176,292	44,073 \star	0.505

estimations using three different dictionaries, with each signal decomposed to order n such that its $SRR = 60$ dB. These statistics are the signal duration, the estimation order n , the mean number of atoms per second of sound (MAS), and the CDE ratio $\Xi_{\Sigma}(n)/\|\mathbf{x}\|^2$. Each dictionary has atoms with the scales and translations stated above. G consists only of Gabor atoms, $G \cup O$ is a union of G and a set of damped oscillators O, and $G \cup S$ is a union of G and a set of damped sinusoids S. For example, the estimation of Flute (which has a duration of about 5.8 s) using G requires 76,742 atoms, corresponding to a MAS of 13,271 atoms/s, and a CDE ratio of 0.387.

Since WGN, Drums, and Harpsichord are all wideband signals, their estimates should exhibit a large MAS for each dictionary because each atom is modulated by a single frequency. On the other hand, Sine, Glock, and Soprano have relatively few frequency components, and so it is expected that they exhibit low MAS values. All MAS values increase when using $G \cup O$ instead of G. And all but two MAS values decrease when using $G \cup S$ instead of G. The CDE ratio provides a measure of the degree to which an estimation interferes with itself in the signal reconstruction. We expect that signals with soft attacks and symmetric or periodic waveform structures will be more similar to Gabor atoms, and thus will yield less interference. This is verified by the low CDE ratios for the estimations of Flute, Violin, and Soprano using G. Though Sine is periodic and its estimation has the smallest MAS using G, it still has a high CDE ratio because of envelope corrections. The estimation of Glock using G has a low MAS, but we believe that its high CDE ratio is caused by errors created when adapting smooth and symmetric Gabor atoms to the high-energy asymmetric transients. To investigate this, we examined the time positions of the atoms associated with the highest DE. Figure 6 compares these to the times of the transients seen in the wide-band short-time Fourier transform of the signal. This clearly shows that interference in the estimation is concentrated around the transients.

The estimation statistics for these signals change when using different dictionaries. Decomposing over $G \cup O$ instead of G increases the MAS for every signal, while in only five cases does the CDE ratio decrease. And when using $G \cup S$ instead of G the MAS increases for only three signals (Glock, Soprano, Sine), while the CDE ratio increases for only two signals (Flute, WGN) compared with G. This suggests that when using a dictionary that is a union of Gabor and damped sinusoid atoms, sparse estimations of these audio signals exhibit less interference than when using just Gabor atoms. For only Sine, Speech, and WGN do the estimations found using $G \cup O$ result in the smallest CDE ratio; $G \cup S$ performs the best in this respect. One physical reason why $G \cup O$ is not a good dictionary for these signals is that the waveform of a damped oscillator can have a very sharp and unnatural discontinuity due to the free phase term.

V. CONCLUSION AND FUTURE WORK

For overcomplete methods, the fact that nonorthogonal sets of functions are used means that the terms of a signal estimation might constructively and destructively interfere. This is observed in sparse atomic decomposition algorithms, such as matching pursuit (MP), where localized greediness can result in the selection of atoms that “correct” for atoms selected in earlier iterations. In the most extreme case, an atom present in the estimation will completely disappear in the reconstruction – which is the motivation for using the term “dark energy” (DE) to describe the interference between atoms. The existence of these terms negatively affect the efficiency and usefulness of the resulting estimations.

In the literature thus far, most efforts have attempted to mitigate the necessity of such correction terms. In our work, we instead seek to quantify the extent to which a sparse atomic estimation exhibits constructive and destructive interference in order to gauge the efficiency of an estimation, the effectiveness of the algorithm, and the fit between the signal and that of

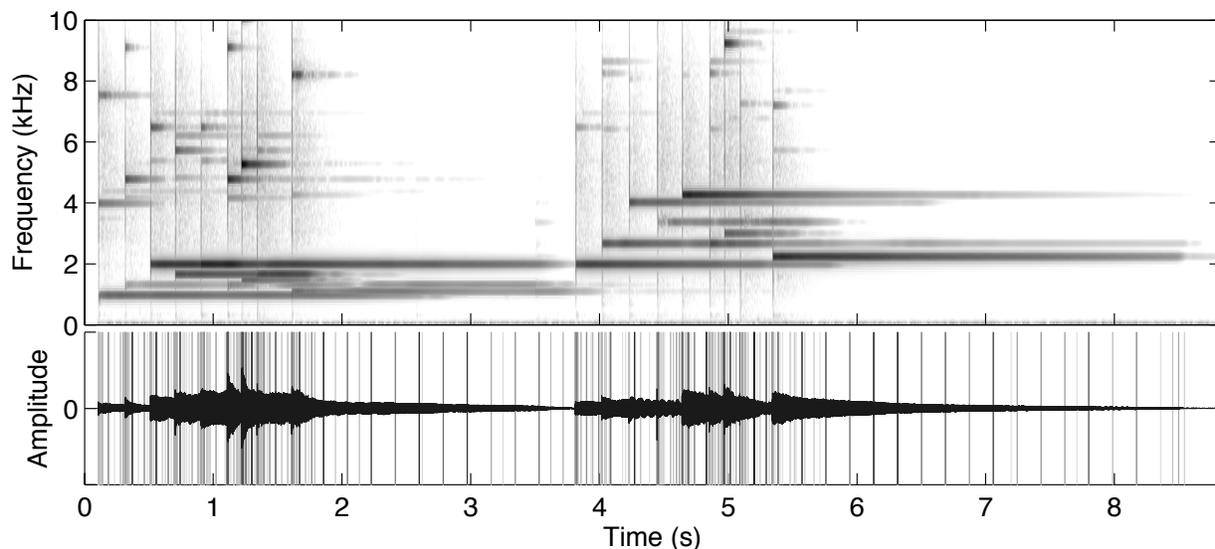


Fig. 6. The wide-band short-time Fourier transform (top) of a music signal (Glock) reveals the locations of transients. The signal waveform (bottom) is superimposed on lines showing the locations of the Gabor atoms in the estimation that interfere the most. The darkness of each line is scaled by the DE of the atom.

the dictionary. In this paper, we have defined the concept of DE for sparse atomic estimations of signals found specifically using MP, proposed a measure for it, and have analytically and empirically investigated its properties.

Future work will explore the utility of DE to improve the efficiency and usefulness of sparse overcomplete methods. For example, DE could provide an intuitive way of controlling the greediness of MP and making better atom selections. The existence of DE in a signal estimation also points to locations of structures in the signal that may not be well-representable by the dictionary. This information could be used to segment the signal into regions that do and do not correlate well with the dictionary. Once the decomposition algorithm has represented those signal aspects that “make sense” with respect to the dictionary, the residual can then be decomposed using a different dictionary, or be modeled differently altogether.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Rémi Gribonval for assisting in the implementation of the damped sinusoid atoms in the MPTK framework, and the reviewers for their helpful comments.

REFERENCES

- [1] 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.
- [2] S. Mallat and Z. Zhang, “Matching pursuits with time-frequency dictionaries,” *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.
- [3] 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, Nov. 1993, pp. 40–44.
- [4] 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.
- [5] S. Jaggi, W. Carl, S. Mallat, and A. S. Willsky, “High resolution pursuit for feature extraction,” *Applied and Computational Harmonic Analysis*, vol. 5, pp. 428–449, 1998.
- [6] R. Heusdens, R. Vafin, and W. B. Kleijn, “Sinusoidal modeling using psychoacoustic-adaptive matching pursuits,” *IEEE Signal Process. Lett.*, vol. 9, no. 8, pp. 262–265, 2002.
- [7] P. Vera-Candeas, N. Ruiz-Reyes, J. C. Cuevas-Martínez, M. Rosa-Zurera, and F. López-Ferreras, “Sinusoidal modeling using perceptual matching pursuits in the bark scale for parametric audio coding,” *IEE Proc.-Visual Image Signal Process.*, vol. 153, no. 4, pp. 431–435, Aug. 2006.
- [8] Ç. Ö. Etemoğlu and V. Cuperman, “Matching pursuits sinusoidal speech coding,” *IEEE Trans. Speech Audio Process.*, vol. 11, no. 5, pp. 413–424, Sept. 2003.
- [9] R. M. Figueras i Ventura, P. Vandergheynst, and P. Frossard, “Low-rate and flexible image coding with redundant representations,” *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 726–739, Mar. 2006.
- [10] O. Al-Shaykh, E. Miloslavsky, T. Nomura, R. Neff, and A. Zakhor, “Video compression using matching pursuits,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 123–143, Feb. 1999.
- [11] G. Monaci, O. D. Escoda, and P. Vandergheynst, “Analysis of multi-modal signals using redundant representations,” in *IEEE Int. Conf. Image Process.*, vol. 3, Genova, Italy, Sept. 2005, pp. 46–49.
- [12] P. Sugden and N. Canagarajah, “Underdetermined noisy blind separation using dual matching pursuits,” in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, Montreal, Quebec, Canada, May 2004, pp. 557–560.
- [13] B. L. Sturm, L. Daudet, and C. Roads, “Pitch-shifting audio signals using sparse atomic approximations,” in *Proc. ACM Workshop Audio Music Comput. Multimedia*, Santa Barbara, CA, Oct. 2006, pp. 45–52.
- [14] 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.
- [15] L. Daudet and B. Torrèsani, “Hybrid representations for audiophonic signal encoding,” *Signal Processing*, vol. 82, no. 11, pp. 1595–1617, Nov. 2002.
- [16] 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.
- [17] S. Krstulovic and R. Gribonval, “MPTK: Matching pursuit made tractable,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, Toulouse, France, Apr. 2006, pp. 496–499.