ORTHOGONAL MATCHING PURSUIT WITH DICTIONARY REFINEMENT FOR MULTITONE SIGNAL RECOVERY

MICHAEL G. MOORE AND MARK A. DAVENPORT*

Abstract. In this paper, we propose a low-cost algorithm for recovering multitone signals from compressive measurements. We introduce a simple and efficient modification to orthogonal matching pursuit. Our approach uses a DFT basis, but refines the frequency estimate obtained at each iteration via a simple gradient descent. We find that by adapting the dictionary in this manner we can realize the benefits of an overcomplete DFT frame without incurring the increased computation. Numerical simulations show that this approach not only outperforms traditional OMP, it even outperforms ℓ_1-minimization unless we incur the computational cost of using a highly overcomplete DFT frame.

Key words. Compressive sensing, multitone signals, overcomplete DFT frames, gradient descent, orthogonal matching pursuit

1. Introduction. Compressive sensing has recently emerged as a framework for acquiring sparse signals [1–3]. If a signal $x \in \mathbb{C}^N$ can be written as a linear combination of $K \ll N$ elements from some dictionary, we say it is K-sparse. Specifically, given an orthonormal basis or dictionary Ψ, we say that a vector x is K-sparse if we can write $x = \Psi\alpha$ where α has at most K nonzeros. In the case where x is exactly or approximately K-sparse, it is often possible to recover the signal from a number of measurements that is much closer to K than N. Specifically, we can acquire M linear measurements of x, represented as $y = \Phi x$, where $\Phi \in \mathbb{C}^{M\times N}$. The vast literature on compressive sensing shows that under appropriate conditions on the product $\Phi\Psi$, it is possible to accurately and efficiently recover the vector α (and hence x) [3].

However, in practice there are many natural signal models which appear "sparse," but for which there is no simple orthonormal basis Ψ that sparsely captures this structure. For example, a common signal model in the context of compressive sensing of analog signals is the "multitone" model, where we suppose that our signal x can be represented as a sum of K complex exponentials:

$$x[n] = \sum_{k=0}^{K-1} a_k e^{j2\pi n f_k} \qquad n \in \{0,\ldots,(N-1)\},\ a_k \in \mathbb{C},\ f_k \in [0,1). \tag{1.1}$$

This signal appears very sparse—it can be described exactly via $2K$ parameters, and appears to be a natural model for "spectral sparsity." Thus we might consider setting Ψ to be the $N \times N$ Discrete Fourier Transform (DFT) matrix, and indeed, this is what is often done in practice (e.g., see [4]). However, in general this will not do a very good job of sparsely representing x due to finite-window effects. To see why this is the case, observe that the Discrete-Time Fourier Transform (DTFT) of a length-N multitone signal is composed of a combination of K modulated Dirichlet functions:

$$X\left(e^{j2\pi f}\right) = \sum_{k=0}^{K-1} a_k e^{j\pi(f_k - f)(N-1)} \frac{\sin(\pi(f_k - f)N)}{N\sin(\pi(f_k - f))}. \tag{1.2}$$

Recall that we can interpret the DFT as simply samples of $X\left(e^{j2\pi f}\right)$ at frequencies $f = \ell/N$ for $\ell = 0,\ldots,N-1$. If the tone frequencies f_k happen to all lie on the

*Georgia Institute of Technology, School of Electrical and Computer Engineering, Atlanta, GA. Email: {mmoore90,mdav}@gatech.edu

so-called "Nyquist grid", i.e., for each f_k there is an ℓ such that $f_k = \ell/N$, then the DFT will indeed sparsely capture the structure in x and have exactly K nonzeros. In general, however, we cannot expect the f_k to lie on this discrete grid. In that case, not only is the DFT of x not sparse, it is not even very compressible due to the slow-decaying behavior of the Dirichlet functions in (1.2).

To address this challenge, a wide range of more sophisticated approaches to recovering such signals from compressive measurements have been proposed [5–8]. These methods can be broadly separated into two categories. One class of methods abandons the DFT in favor of more sophisticated spectral estimation techniques (for example, [5, 6]). Unfortunately, such methods also come with significantly increased computational requirements. The other broad class of methods applies existing sparse recovery algorithms, but replaces the DFT with a redundant or "overcomplete" DFT frame (as in many approaches discussed in [6] as well as the algorithms proposed in [7, 8]). Specifically, these approaches replace the DFT orthobasis with a $D \times N$ matrix whose columns are given by

$$\psi_f = \begin{bmatrix} e^{j2\pi 0 f} \\ e^{j2\pi 1 f} \\ e^{j2\pi 2 f} \\ \vdots \end{bmatrix}, \tag{1.3}$$

where $f \in \{0, 1/D, \ldots, (D-1)/D\}$. Using this frame, we can then apply standard sparse recovery algorithms such as ℓ_1-minimization [1, 2] or greedy/iterative algorithms like orthogonal matching pursuit (OMP) [9, 10], compressive sampling matching pursuit (CoSaMP) [11], or iterative hard thresholding (IHT) [12]. It is important to note that the standard theoretical guarantees for these algorithms do not apply when D is much greater than N since, in this setting, Ψ becomes highly coherent. Despite the lack of theoretical guarantees, in practice these algorithms often perform relatively well.[1]

The intuition behind this approach is that, by increasing D, the vector α becomes a denser sampling of the DTFT of x. Now, we still cannot expect to be able to recover the full DTFT (α is still not sparse). However, all of the sparse recovery algorithms described above will aim to recover a good sparse approximation to α by attempting to find the largest elements in α. The hope is that the DTFT will have noticeable peaks very close to the frequencies $\{f_k\}_{k=0}^{K-1}$, so that by finding the peaks in α when D is relatively large, we will be synthesizing x with dictionary atoms relatively close to the true tones which generated x.

Unfortunately, in practice we must often set D to be rather large to achieve significant performance gains. This comes at a large increase in the computational complexity of the sparse recovery algorithms (which are all *at least* linear in D). Here, we present a simple modification of OMP using a DFT basis or only mildly overcomplete DFT frame that improves signal recovery for multitone signals while retaining OMP's low computational cost. We hope that this serves as an attractive alternative in implementations that cannot afford the higher complexity of more sophisticated methods.

[1]It is worth noting that there has been some recent theoretical progress regarding our understanding of the performance of algorithms such as ℓ_1-minimization and CoSaMP applied to certain overcomplete dictionaries [7, 8], but our understanding of these scenarios remains incomplete.

Algorithm 1 Orthogonal Matching Pursuit for Multitone Signals

input: y, Φ, $\mathbf{G}(\cdot)$, stopping criterion
initialize: $x^0 = 0$, $\Omega^0 = \emptyset$, $\ell = 0$
while not converged **do**
  $p = \Phi^H(y - \Phi x^\ell)$
  $\Omega^{\ell+1} = \Omega^\ell \cup \mathbf{G}(p)$
  $a^{\ell+1} = \arg\min_z \|y - \Phi\Psi_{\Omega^{\ell+1}} z\|_2$
  $x^{\ell+1} = \Psi_{\Omega^{\ell+1}} a^{\ell+1}$
  $\ell = \ell + 1$
end while
output: $\widehat{x} = x^\ell$, $\widehat{\Omega} = \Omega^\ell$

2. Our Approach. Our approach is based on OMP, described in our context in Algorithm 1. The first step compensates for the fact that our signal is observed through compressive measurements by forming a rough estimate of the signal via $p = \Phi^H y$. Next, we let $\mathbf{G}(p)$ represent a function that takes a vector p and returns some estimate of the dominant frequency in p. For example, the traditional approach would consist of taking a length-D FFT and selecting the frequency corresponding to the entry with the maximum magnitude. We let Ω represent a set of frequencies and define a Fourier frame Ψ_Ω to be the $N \times |\Omega|$ matrix with $|\Omega|$ columns of the form in (1.3) at the frequencies in the set Ω. In words, OMP iteratively chooses a single tone at a time, forms the best representation from the tones chosen so far, and then repeats the process on the residual.

This algorithm is appealing for our approach because it only requires that we estimate a single frequency at each iteration. Traditionally this is performed by computing the length-D FFT and selecting the element with maximum magnitude. Our proposal is to replace this step with an alternative approach. Specifically, if we define $S_p(f) := \left|\sum_{n=0}^{N-1} p[n] e^{-j2\pi n f}\right|^2$, we wish to replace $\mathbf{G}(p)$ with the following optimization program:

$$\widehat{f} = \underset{f \in [0,1)}{\arg\max}\ S_p(f). \tag{2.1}$$

This will return the frequency in $[0, 1)$ of the tone that is most correlated with our signal. Unfortunately this problem is no longer a finite search (as with a DFT frame), nor is it convex, so determining the optimum f is nontrivial. However, thanks to the time-limited nature of our signal model, the DTFT of the signal is smooth (recall (1.2)) and is locally convex in the vicinity of the solution. Thus, with appropriate initialization we can find $\widehat{f}$ using a simple gradient descent. This initialization can be accomplished using an FFT to sample the DTFT. For any frequencies where the corresponding DFT coefficient is sufficiently large, we can then use gradient descent to find a local maximum, and then estimate the global maximum by taking the largest of these local maxima.

An algorithm that accomplishes this task is given in Algorithm 2. $\dot{S}_p(f)$ denotes the derivative of $S_p(f)$ with respect to f. This particular implementation uses the secant method for the gradient descent stage. Note that one should choose $D \geq N$ to avoid undersampling the DTFT, which can lead to a poor frequency estimate as peaks may be missed entirely.

Even with sufficient sampling, straddle losses in the DFT can lead to slight mis-

Algorithm 2 Gradient Descent for Maximizing $S_p(f)$

input: p, N, D, $\Lambda = \{0, 1/D, \ldots, (D-1)/D\}$
initialize: $\widehat{f} = \arg\max_{f \in \Lambda} S_p(f)$
for all $f^0 \in \Lambda$ such that $\frac{S_p(f^0)}{S_p(\widehat{f})} > \left(\frac{\sin(\pi N/2D)}{N \sin(\pi/2D)}\right)^2$ **do**
  initialize: $f^1 = f^0 + \mathrm{sgn}(\dot{S}_p(f^0))/2D$; $\ell = 1$
  while $\dot{S}_p(f^\ell) \neq 0$ **do**
    update: $f^{\ell+1} = f^\ell + \dot{S}_p(f^\ell)\frac{f^\ell - f^{\ell-1}}{\dot{S}_p(f^\ell) - \dot{S}_p(f^{\ell-1})}$; $\ell = \ell + 1$
  end while
  if $S_p(f^\ell) > S_p(\widehat{f})$ **then**
    update: $\widehat{f} = f^\ell$
  end if
end for
output: $\widehat{f}$

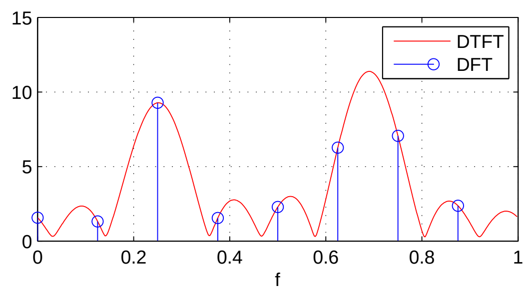

Fig. 2.1. *Multitone signal where straddle loss complicates maximization*

representations of peaks. Straddle loss arises from the fact that we are sampling a continuous function (the DTFT) and the maximum of this sampling is unlikely to be the maximum of the function. Thus it is necessary to check all peaks of magnitude close to the largest to find the true global maximum, hence the "for-all" loop in the algorithm. If this fact is neglected, one may instead find a local maximum that is slightly lower than, though within the straddle loss of, the true maximum. An illustration of this issue is given in Figure 2.1, where the largest peak in the DFT is not in the vicinity of the largest peak in the DTFT. Due to the structure of OMP, this is likely of minimal consequence in the final recovery and could probably be omitted without catastrophic failure (a peak that is nearly missed in one iteration becomes an even better candidate in the next). Moreover, increasing D beyond N reduces the worst-case straddle loss, meaning that initializations near nonglobal maxima are less likely to be considered.

3. Simulation Results. To evaluate the performance of this method, we compare it to standard OMP and ℓ_1-minimization using a fixed DFT frame. This was tested using 1000 trials with a frequency-sparse signal fitting the model presented in (1.1) with $|a_k| = 1$ and $f_k \sim U[0, 1]$. Because OMP is an iterative algorithm, this choice for a_k is something of a worst-case scenario. When tones are of varied amplitude it is easier pick out the stronger, more obvious tones before removing weaker tones that are distorted by the stronger ones. Each signal in the simulation has $K = 10$ tones, is of length $N = 512$, and is observed through a random sampling matrix Φ

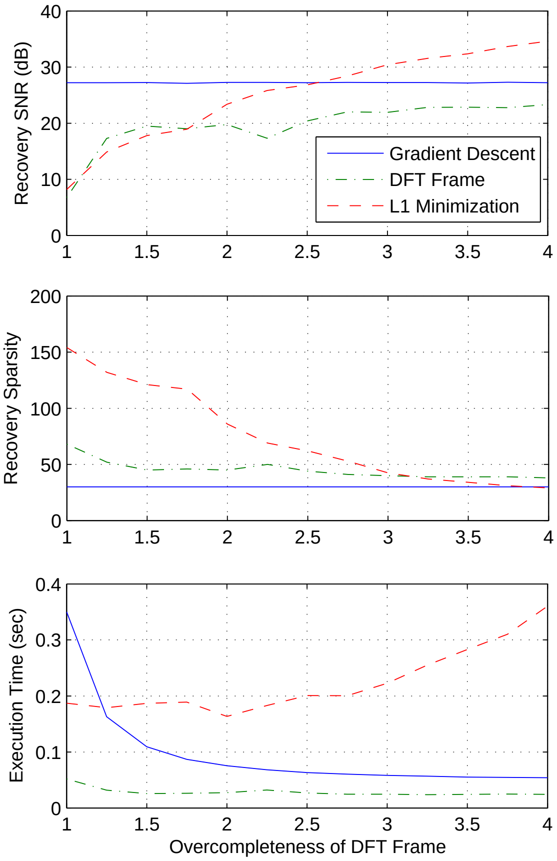

FIG. 3.1. *Median recovery signal-to-noise ratio, median recovery sparsity, and average execution time by overcompleteness.*

where $M = 100$ of the N samples are randomly selected for observation:

$$y_m = x_{n_m} \qquad n_m \in \{0, \ldots, N-1\} \qquad m \in \{0, \ldots, M-1\}. \tag{3.1}$$

We quantify our performance via the recovery signal-to-noise ratio (SNR) given by $20 \log_{10} \left(\frac{\|x\|_2}{\|x - \widehat{x}\|_2} \right)$.

In Figure 3.1 we illustrate the performance of our gradient descent approach compared to standard OMP and ℓ_1-minimization as a function of the overcompleteness ratio D/N. Empirically, and as expected, in terms of SNR the performance of our approach shows no dependence on D when $D \geq N$. Thus, we can choose D to be a small, convenient value and avoid the computational burden involved in dealing with large D. The experimental results show that for small choices of D our approach outperforms ℓ_1-minimization both in terms of the computation required as well as the accuracy of the recovery $\widehat{x}$. For D sufficiently large, as suggested by [8], ℓ_1-minimization does achieve superior performance. However, we see that this comes at the cost of a notable increase in computational effort (because our approach performs identically for any D, we can choose whatever value grants the fastest execution).

In Figure 3.2 we show the sorted magnitudes of the coefficients recovered by each method for a single iteration of the above simulation. As one might expect, OMP

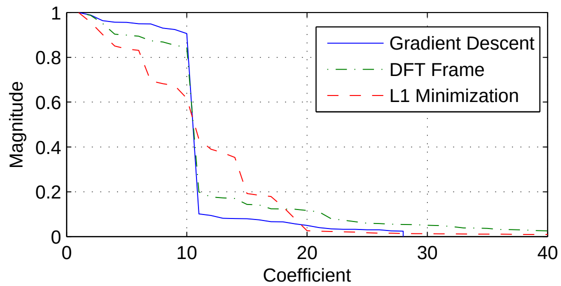

FIG. 3.2. *Coefficient falloff of recovery algorithms for instance of multitone signal with $D = 2.5N$, where our method and ℓ_1 minimization achieve similar recovery quality.*

spends the first $K = 10$ coefficients hitting very close to each peak in the DTFT. After this, it uses the remainder of its coefficients suppressing the residual parts left over from the imperfect "hits" on the true frequencies. Hence the sharp transition in the coefficient magnitudes. ℓ_1-minimization exhibits a much more gradual decay profile. Our method is generally able to fit more energy into the earlier (and hence less into the later) coefficients than OMP with a DFT frame because we can place the earlier coefficients more precisely. Thus, beyond an improved SNR when D is not significantly larger than N, an additional advantage of our method is that the recovered coefficients have an added layer of "interpretability" and a closer degree of correspondence to the underlying parameters of our signal model.

REFERENCES

[1] E. Candès, "Compressive sampling," in *Proc. Int. Congress of Math.*, Madrid, Spain, Aug. 2006.

[2] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] M. Davenport, M. Duarte, Y. Eldar, and G. Kutyniok, "Introduction to compressed sensing," in *Compressed Sensing: Theory and Applications*, Y. Eldar and G. Kutyniok, Eds. Cambridge, UK: Cambridge University Press, 2012.

[4] J. Tropp, J. Laska, M. Duarte, J. Romberg, and R. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse, bandlimited signals," *IEEE Trans. Inform. Theory*, vol. 56, no. 1, pp. 520–544, 2010.

[5] M. Davenport and M. Wakin, "Compressive sensing of analog signals using discrete prolate spheroidal sequences," *Appl. Comput. Harmon. Anal.*, vol. 33, no. 3, pp. 438–472, 2012.

[6] M. Duarte and R. Baraniuk, "Spectral compressive sensing," *to appear in Appl. Comput. Harmon. Anal.*, 2012.

[7] M. Davenport, D. Needell, and M. Wakin, "Signal space CoSaMP for sparse recovery with redundant dictionaries," *Arxiv preprint arxiv:1208.0353*, 2012.

[8] G. Tang, B. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *Arxiv preprint arxiv:1207.6053*, 2012.

[9] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[10] Y. Pati, R. Rezaifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 1993.

[11] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.

[12] T. Blumensath and M. Davies, "Iterative hard thresholding for compressive sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, 2009.