Fast Algorithm for Sparse Signal Approximation using Multiple Additive Dictionaries

Ray Maleh, Daehyun Yoon, Anna C. Gilbert

Fast Algorithm for Sparse Signal Approximation using Multiple Additive Dictionaries

Ray Maleh
Department of Mathematics
University of Michigan
Ann Arbor, MI, USA
Email: rmaleh@umich.edu

Daehyun Yoon
Department of Electrical Engineering
University of Michigan
Ann Arbor, MI, USA
Email: quann@umich.edu

Anna C. Gilbert
Department of Mathematics
University of Michigan
Ann Arbor, MI, USA
Email: annacg@umich.edu

***Abstract*—There are several models for sparse approximation: one where a signal is a sparse linear combination of vectors over a redundant dictionary and a second model in which a collection of signals is a simultaneous sparse linear combination over a single dictionary. In this work, interpolate between these two models to synthesize a single signal of interest from K highly incoherent dictionaries while enforcing simultaneous sparsity on the K resulting coefficient vectors. We define this as the parallel approximation problem, which arises quite naturally in many applications such as MRI parallel excitation using multiple transmission coils. We present an efficient algorithm to solve the parallel approximation problem called Parallel Orthogonal Matching Pursuit (POMP). We prove its correctness in a general setting and then discuss adaptations needed to make it suitable for use in an MRI parallel excitation setting. We then discuss parallel excitation in more detail and demonstrate how POMP solves the problem as accurately, but much faster, than previously proposed convex optimization methods.**

I. Introduction

There are two predominant paradigms for sparse approximation. In the first setting, $y \in \mathbb{C}^M$ is a signal vector we wish to synthesize as $y = \Phi x$, Φ is a dictionary matrix whose columns correspond to elementary signal vectors, and $x \in \mathbb{C}^N$ is a coefficient vector. The goal is to create an approximation of y using only a few columns of Φ, enforcing sparsity on x. Several algorithms have been designed to attack this problem, including convex optimization [1] and orthogonal matching pursuit (OMP) [6], [7]. Frequently in applications, such as hyperspectral imaging, sparse-gradient image recovery [4], etc., one is required to solve a related sequence of K sparse approximation problems of the form $y_k = \Phi x_k$ with $k = 1, \cdots, K$. If the x_ks have a common support set, then these K problems can be coupled and solved by a specially designed convex program[5] or by an extension of OMP known as Simultaneous Orthogonal Matching Pursuit (SOMP) [3], [5]. In this work, we introduce a new variant of the sparse approximation problem that interpolates between these two paradigms, which we shall refer to as Parallel Approximation. In this problem, we assume that we have a set of sparse coefficient vectors$\{x_1, \cdots, x_K\} \subset \mathbb{C}^N$ with a common support which additively synthesize a single signal using K highly incoherent dictionaries. In other words, we have K dictionary matrices $\{\Phi_1, \cdots, \Phi_K\} \subset \mathbb{C}^{M\times N}$ and a vector to approximate $y \in \mathbb{C}^M$ given by

$$y = \Phi_1 x_1 + \Phi_2 x_2 + \cdots \Phi_K x_K. \quad (1)$$

Our objective is to efficiently compute the x_ks from y and the Φ_ks while enforcing simultaneous sparsity on x_ks. This problem is of profound importance to the medical imaging community. It arises in MRI parallel excitation design with multiple transmission coils. In this application, one tries to select a small but optimal set of 2D frequencies that yield a high quality image scan over a restricted region of interest (ROI); e.g., the shape of a human head. Each dictionary matrix is a 2D discrete Fourier matrix multiplied by a sensitivity matrix induced by a particular transmission coil.

In [9], convex optimization is proposed to solve this parallel approximation problem. Unfortunately, this method is slow and inappropriate for any real-time computation, which drastically limits its usefulness in clinical settings. We propose a new algorithm called Parallel Orthogonal Matching Pursuit (POMP) for solving the parallel approximation problem shown in Equation(1). We prove its correctness in a general setting and then discuss adaptations needed to apply it to parallel MRI excitation, where the dictionary matrices are not incoherent. Then, we discuss the parallel excitation application in more detail and present experimental simulations that demonstrate that POMP performs quite similarly to convex optimization in terms of approximation error but runs significantly faster, suggesting its use in clinical settings. We conclude by suggesting further applications of the parallel approximation problem.

II. Parallel Orthogonal Matching Pursuit

We begin with some notation. Let $\{x_1, \cdots, x_K\} \subset \mathbb{C}^N$ be K T-sparse coefficient vectors with common support Λ_{opt}. In other words, for each i not in Λ_{opt}, the vector $[x_1(i), \cdots, x_K(i)]$ is zero. Let $\Phi_1, \Phi_2, \cdots, \Phi_K$ be a sequence of K $M \times N$ dictionary matrices. Let $\Phi = [\Phi_1 | \cdots | \Phi_K]$ be the horizontal concatenation of these matrices. Also, let ϕ_k^i be the ith column (or atom) of matrix Φ_k. We are given a target signal $y \in \mathbb{C}^M$ to approximate in the form shown in Equation 1. We can estimate the x_ks using the following algorithm.

RM and ACG are supported by NSF DMS 0354600 and DARPA/ONR N66001-06-1-2011. DY is supported by NIH R01NS058576. ACG is an Alfred P. Sloan Fellow.

Algorithm: Parallel Orthogonal Matching Pursuit

Inputs: A target signal to approximate : y
Dictionary matrices : Φ_k, $k = 1, \cdots, K$
Number of iterations T.
Outputs: T term approximations $\widetilde{x_k}$'s of the x_k's
Residual r_T.

Initialize residual $r_0 = y$, index set $\Lambda = \emptyset$.
For t from 1 to T {
Let
$$\lambda_t = \operatorname*{argmax}_i \sum_{k=1}^{K} \left|(\phi_k^i)^* r_t\right|. \tag{2}$$
Set $\Lambda_t = \Lambda_{t-1} \cup \{\lambda_t\}$.
Let p_k be the projection of the residual onto the selected vectors $\{\phi_k^i | 1 \leq k \leq K, i \in \Lambda_t\}$.
Set $r_t = r_0 - p_t$. }
Solve for the $\widetilde{x_k}$'s by using the coefficients of the ϕ_k^i's determined when solving for p_T.

Fig. 1. Pseudocode for Parallel Orthogonal Matching Pursuit.

In order to be able to recover the coefficient vectors x_i accurately, it is desirable to have that the columns of the Φ_k's be highly incoherent, or in other words, have their inner products as close to zero as possible. To that end, we define the notions of *cumulative coherence* and *cross cumulative coherence*.

Definition 1. *Let $\Phi_1, \cdots \Phi_K$ denote K $M \times N$ measurement matrices Now fix any one k. Then the* cumulative coherence *of matrix Φ_k, if defined as a function of T, to be*
$$\mu_k(T) = \max_i \max_{\substack{|\Lambda|=T \\ i \notin \Lambda}} \sum_{j \in \Lambda} \left|(\phi_k^i)^* \phi_k^j\right|$$
We define the cross-cumulative coherence of matrix Φ_k to be
$$\nu_k(T) = \max_i \max_{\substack{|\Lambda|=T \\ \ell \neq k}} \sum_{j \in \Lambda} \left|(\phi_k^i)^* \phi_\ell^j\right|.$$

With these definitions in mind, we can prove the following sufficient condition to ensure the correctness of POMP.

Proposition 1. *Suppose we have an ensemble of K dictionaries $\Phi = [\Phi_1, \cdots, \Phi_K]$ with cumulative coherences μ_k and cross cumulative coherences ν_k, $k = 1, \cdots, K$ that satisfy*
$$\sum_{k=1}^{K} \mu_k(T) + \mu_k(T-1) + 2(K-1)\nu_k(T) < 1. \tag{3}$$
Then given any signal y with jointly T sparse coefficient vectors, POMP *will select a correct column index at every iteration and therefore recover y exactly.*

Proof: Suppose that after t iterations, POMP has selected only correct column indices. Then it follows that the residual r_t in an element of the space $\sum_{k=1}^{K} \operatorname{colspan}(\Phi_{k,\text{opt}})$ where $\Phi_{k,\text{opt}}$ is the $M \times T$ submatrix of Φ_k which consist of the T columns corresponding to the correct column indices, i.e. the non-zero entires of x_k. Now let's write r as
$$r = \sum_{k=1}^{K} \sum_{i=1}^{T} c_{k,t} \phi_{k,\text{opt}}^i$$
where $\phi_{k,\text{opt}}^i$ is the i-th column of $\Phi_{k,\text{opt}}$. Now without loss of generality, we may assume that $\sum_{k=1}^{K} |c_{k,1}| \geq \sum_{k=1}^{K} |c_{k,i}|$ for each i. Otherwise, just reorder the columns of Φ_{opt}. Then we can derive the following useful inequalities based on cumulative coherence and cross cumulative coherence estimates: For any $\phi_{k,\text{opt}}^1$ from the first column of some $\Phi_{k,\text{opt}}$, we have that
$$|(\phi_{k,\text{opt}}^1)^* r| \geq$$
$$|c_{k,1}| - \left(\sum_{k=1}^{K} |c_{k,1}|\right) (\mu_k(T-1) + (K-1)\nu_k(T)).$$
Summing over k gives us that
$$\sum_{k=1}^{K} |(\phi_{k,\text{opt}}^1)^* r| \geq \tag{4}$$
$$\left(\sum_{k=1}^{K} |c_{k,1}|\right) \left(1 - \sum_{k=1}^{K} [\mu_k(T-1) + (K-1)\nu_k(T)]\right).$$
Similarly, for each k, we can define Ψ_k to be the $M \times (N-T)$ submatrix of Φ_k consisting of the incorrect columns. Let ψ_k^i denote the ith column of Ψ_k. Then for a fixed k and i, we can obtain the estimate
$$|(\psi_k^i)^* r| \leq \left(\sum_{k=1}^{K} |c_{k,1}|\right) (\mu_1(T) + (K-1)\nu_1(T)).$$
Again, we keep i fixed and sum over k to obtain:
$$\sum_{k=1}^{K} |(\psi_k^i)^* r| \leq \tag{5}$$
$$\left(\sum_{k=1}^{K} |c_{k,1}|\right) \left(\sum_{k=1}^{K} [\mu_1(T) + (K-1)\nu_1(T)]\right).$$
Now observe that POMP will definitely pick a correct atom if
$$\sum_{k=1}^{K} |(\phi_{k,\text{opt}}^1)^* r| > \sum_{k=1}^{K} |(\psi_k^i)^* r|.$$
Combining inequalities 4 and 5 gives us a sufficient condition for this, which is:
$$\left(\sum_{k=1}^{K} |c_{k,1}|\right) \left(\sum_{k=1}^{K} [\mu_1(T) + (K-1)\nu_1(T)]\right) <$$
$$\left(\sum_{k=1}^{K} |c_{k,1}|\right) \left(1 - \sum_{k=1}^{K} [\mu_k(T-1) + (K-1)\nu_k(T)]\right).$$
Rearranging terms now gives us
$$\sum_{k=1}^{K} \mu_k(T) + \mu_k(T-1) + 2(K-1)\nu_k(T) < 1,$$

which completes the proof. ∎

Observe that if we set $K = 1$, we get the same exact condition that guarantees the correctness of regular orthogonal matching pursuit: $\mu_1(T-1) + \mu_1(T) < 1$ (see [7]). Indeed, if $K = 1$, POMP turns out to be equivalent to OMP. Using the method of proof for 1 and an argument very similar to that in [7], we can also show the following ℓ_2-ℓ_2 performance guarantee for POMP

Proposition 2. *Suppose that y is any signal. Let y_{opt} denote the optimal jointly T-sparse parallel representation of y in $\Phi_1, \cdots, \Phi_K$. Suppose that we have the condition:*

$$\xi(T) := \sum_{k=1}^{K} \mu_k(T) + (K-1)\mu_k(T) < 0.5.$$

Then after T iterations, POMP *will return an estimate $\tilde{y}$ of y satisfying*

$$\|y - \tilde{y}\|_2 \leq \sqrt{1 + \frac{K^2 T\,(1-\xi(T))}{(1-2\xi(T))^2}}\,\|y - y_{opt}\|_2\,.$$

This result states that given a sufficiently incoherent ensemble of dictionaries, then the error generated by POMP's reconstruction of a signal y is $O(K\sqrt{T})$ times worse than the error induced by y's optimal jointly T-sparse representation.

In order to generalize the above algorithm, we can replace the correlation criterion in Equation 2 with a general ℓ_p norm as opposed to the ℓ_1 norm shown. We will denote this generalization as POMP$_p$. This is much like the generalization of SOMP seen in [3]. The specific case POMP$_2$ was presented as a solution to the parallel recovery problem in [8]. Another possibility is for each column index i, we can project r_t onto span$\{\phi_1^i, \cdots, \phi_K^i\}$ and then select the i which maximizes the magnitude of the projection. In other words, if we let Φ^i denote the $M \times K$ matrix with columns $\phi_1^i, \cdots, \phi_K^i$, then we select the index i that maximizes $||(\Phi^i)(\Phi^i)^\dagger r_t||_2$ where $(\Phi^i)^\dagger = ((\Phi^i)^*\Phi^i)^{-1}(\Phi^i)^*$. We will denote this modification as POMP$_{\text{proj}}$.

Unfortunately, in many practical applications such as parallel excitation, our measurement dictionaries do not satisfy the sufficient condition $\xi(T) < 0.5$. As a result, we need to modify the basic POMP algorithm to compensate for this sub-optimal incoherence. We do this by changing the atom selection criterion 2 into:

$$A_t = \left\{ i : ||(\Phi^i)^* r_t||_p \geq \tau \max_j ||(\Phi^j)^* r_t||_p \right\}$$

for POMP$_p$ and

$$A_t = \left\{ i : ||(\Phi^i)(\Phi^i)^\dagger r_t||_2 \geq \tau \max_j ||(\Phi^j)(\Phi^j)^\dagger r_t||_2 \right\}$$

for POMP$_{\text{proj}}$ for some threshold $0 < \tau < 1$. Then we set $\Lambda_t = \Lambda_{t-1} \cup A_t$. If the number of chosen is is one, then we set the threshold with respect to the second maximum and update A_t again. In other words, at every iteration, we pick two or more column indices that correlate sufficiently well with the residual. As an easy example of why this is useful, consider the case $K = 1$. Suppose we are given a signal $y \in \mathbb{R}^3$ that is strictly a linear combination of two dictionary vectors ϕ_1 and ϕ_2 that are highly correlated as shown in figure II. We further suppose that ϕ_3 is another dictionary vector that has a tiny component sticking out of the page. In other words, it is not coplanar with ϕ_1 and ϕ_2, but its shortest distance from the plane induced by the latter two vectors is tiny.

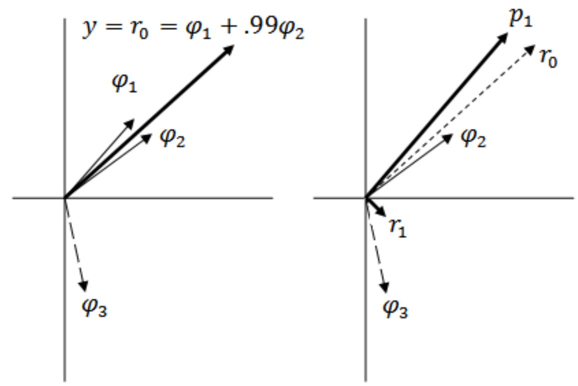

Fig. 2. A demonstration of what goes wrong if POMP is applied onto a signal y that is composed of two highly coherent dictionary vectors ϕ_1 and ϕ_2. ϕ_3 is another dictionary vector that is just barely not coplanar with ϕ_1 and ϕ_2. The first iteration correctly selects ϕ_1. However, because ϕ_2 is highly correlated with ϕ_1, the residual r_1 is projected much closer to ϕ_3 resulting in ϕ_3 being incorrectly chosen in the second iteration.

Because y is 2-sparse, we should only need 2 iterations of regular POMP (or OMP in this case) to recover it. Since ϕ_1 alone is chosen in the first iteration, the residual r_0 will be heavily projected onto ϕ_1, which means that the new residual r_1 will be orthogonal to ϕ_1 and very uncorrelated to ϕ_2. In this particular case, ϕ_3 will be selected instead of ϕ_2 in the second iteration. In order to alleviate this, we choose τ so that ϕ_1 and ϕ_2 will be selected together in one iteration.

Our target application, parallel excitation, suffers from this problem because excitation profiles (discussed in the next section) consist of connected regions that concentrate all their energy in adjacent low frequencies. Thus, to compete with existing convex optimization strategies, we incorporate this thresholding into our POMP algorithm. From experiments, we have found that values of τ on the order of .95 to .99 have been optimal. Now with this framework in mind, we describe in the next section the MRI parallel excitation problem and how POMP can be used to efficiently solve it.

III. Parallel Excitation in MRI

In MRI, we reconstruct an image of the proton (H^+) density of the target object from its 2D-Fourier transform samples. Before acquiring these Fourier samples, we must select the region of imaging volume, which is typically in the form of a thin slice. This process is called slice-selective excitation. For example, a 2D MRI image, $I(x,y)$, is $I(x,y) = \int_{-\infty}^{\infty} p(x,y,z)w(x,y,z)dz$ where x, y, z are spatial coordinates, and $p(x,y,z)$ is the proton density and $w(x,y,z)$

is the characteristic function implemented by the excitation process to specify the imaging volume. In the particular case of parallel excitation with slice-selective subpulses[10], the function $w(x, y, z)$ is a separable function that can be rewritten as $w(x, y, z) = d(x, y)s(z)$, where $s(z)$ is the Fourier transform of the slice-selective subpulse and $d(x, y)$ is the inplane excitation profile. In order to develop image contrast from the proton density only, we need to make $d(x, y)$ constant in the imaging region of interest. Otherwise, irregularities in the inplane excitation profile will appear in the reconstructed MRI image and the image quality will be poor. In [10], the inplane excitation profile, d, is determined by

$$d = \Phi_1 b_1 + \Phi_2 b_1 + \cdots \Phi_K b_k,$$

where $\Phi_i = S_i F$, S_i is a diagonal matrix of the i-th coil's sensitivity pattern, and F is a 2D discrete Fourier matrix restricted to the support of d. Each element of b_i represents a complex amplitude of a subpulse transmitted by the i-th coil. The net pulse transmitted from the coil is the concatenation of all subpulses with non-zero weights; therefore, finding a sparse solution for the b_ks is crucial for generating a short pulse, which is important in fast imaging. Here, the elements of d are identically equal to one because we want to correct for the unwanted image contrast otherwise developed by coil sensitivities. A visual depiction of this problem is shown in Fig. 4.

We ran our POMP algorithm on this problem and compared its performance with the convex optimization approach in [9]. In our experiments, the region of interest d is a uniform circular pattern with radius 10.125cm in a viewing area of size 24cm by 24cm over a discrete uniform grid of size 64x64. We used 8 transmission coils for the pulse design. We ran both algorithms on a computer with Intel Core2 Quad CPU 2.4GHz, 4GB RAM and Matlab 7. We used the SeDuMi convex optimization package (http://sedumi.mcmaster.ca) for the convex optimization as in [9]. In Figure 3, we plot the normalized mean squared error(NRMSE) for these two algorithms as a function of sparsity T. As seen in Fig. 3, the NRMSE curves show that our method has compatible accuracy to convex optimization. Also, Table 1 shows that our method runs much faster than convex optimization on the same problem.

Algorithm	Time
Convex Optimization	$\approx$ 48 *minutes*
POMP	$<$ 25 *seconds*

TABLE I
RUNTIME OF CONVEX OPTIMIZATION AND POMP METHODS.

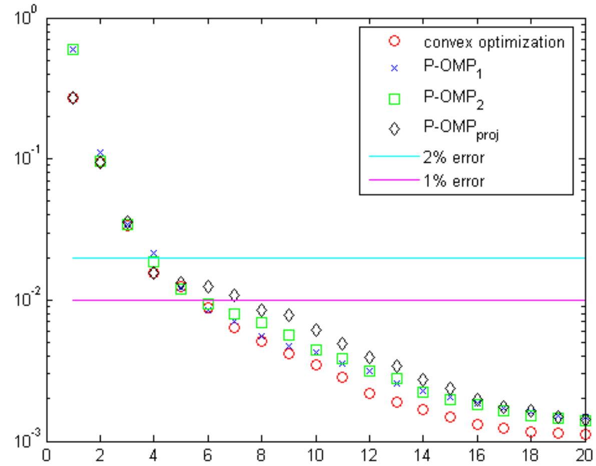

Fig. 3. NRMSE of convex optimziation and POMP methods

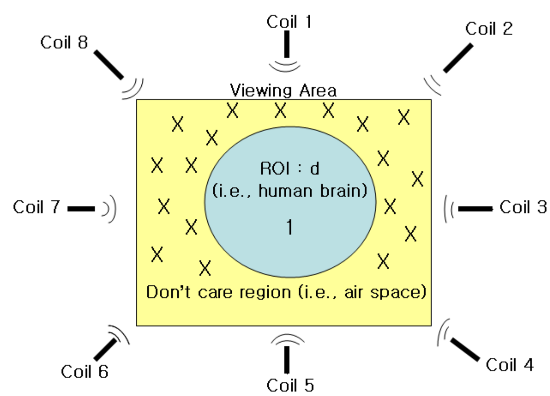

Fig. 4. 8-coil parallel excitation example

IV. Conclusion

As we have demonstrated, Parallel Orthogonal Matching Pursuit provides us with a highly accurate and efficient tool for solving the parallel excitation problem in MRI. However, the applications of POMP are not limited to this. The algorithm is useful in many other imaging applications such as RADAR and SONAR where multiple detectors can simultaneously be used to paint a big picture of objects in the air or sea. In addition, POMP may be useful in coding theory when performing error correction on codewords originating from multiple sources. This particular application is currently an area of active research. Furthermore, thinking of POMP as a compressive sensing tool allows for the designing of measurement matrices that will exploit the sparsity (or low information rate) of highly related signals $x_1, \cdots, x_K$ in order to compress them into one short measurement vector y whose size is possibly logarithmic with respect to the original signals' dimension (see [2]). This is a feat that should be considered highly impressive by the signal processing and data compression communities.

References

[1] E. J. Candes, J. Romberg, T. Tao, *Stable Signal Recovery from Incomplete and Innacurate Measurements*, Comm. Pure Appl. Math, 59 1207-1223.

[2] E. J. Candes, M. Wakin, *An Introduction to Compressive Sensing*, IEEE Signal Processing Magazine, March 2008, 21-30.

[3] R. Gribonval, H. Rauhut, K. Schnass, P. Vandergheynst, *Atoms of All Channels, Unite! Average Case Analysis of Multi-channel Sparse*

Recovery Using Greedy Algorithms, Journal of Fourier analysis and Applications, 2008.

[4] R. Maleh, A. C. Gilbert, *Multichannel Image Estimation Via Simultaneous Orthogonal Matching Pursuit*, SSP Proceedings, 2007.

[5] J. Tropp, *Algorithms for Simultaneous Sparse Approximation (Parts I and II)*, Signal Processing, special issue "Sparse approximations in signal and image processing," Vol. 86, pp. 589-602, Apr. 2006.

[6] J. Tropp, A. C. Gilbert, *Signal Recovery from Partial Information via Orthogonal Matching Pursuit*, Submitted for publication, April 2005.

[7] J. Tropp, *Greed is Good: Algorithmic Results for Sparse Approximation*, IEEE Trans. Info. Theory, Vol. 50, Num. 10, 2231-2242, Oct. 2004.

[8] A. C. Zelinski, V. K. Goyal, E. Adalsteinsson, *Simultaneously Sparse Solutions to Linear Inverse Problems with Multiple System Matrices and a Single Observation Vector*, SIAM Journal of Scientific Computing 2008, under review.

[9] A. C. Zelinski, L. L. Wald, K. Setsompop, V. K. Goyal, E. Adalsteinsson, *Sparsity-Enforced Slice-Selective MRI RF Excitation Pulse Design*, IEEE Transactions on Medical Imaging volume 27, Issue 9, pp. 1213-1229, Sept. 2008

[10] Z. Zhang, Magentic Resonance in Medicine, *Reduction of transmitter B1 inhomogeneity with transmit SENSE slice-select pulses* , Vol. 57, Issue 5, pp. 842-847, Apr. 2007