Faster & Greedier: algorithms for sparse reconstruction of large datasets

(Invited Paper)

Michael E. Davies and Thomas Blumensath
Institute for Digital Communications
School of Engineering and Electronics
University of Edinburgh
Scotland, UK
Email: mike.davies@ed.ac.uk, thomas.blumensath@ed.ac.uk

***Abstract*—We consider the problem of performing sparse reconstruction of large-scale data sets, such as the image sequences acquired in dynamic MRI. Here, both conventional L_1 minimization through interior point methods and Orthogonal Matching Pursuit (OMP) are not practical. Instead we present an algorithm that combines fast directional updates based around conjugate gradients with an iterative thresholding step similar to that in StOMP but based upon a weak greedy selection criterion. The algorithm can achieve OMP-like performance and the rapid convergence of StOMP but with MP-like complexity per iteration. We also discuss recovery conditions applicable to this algorithm.**

I. Introduction

A sparse signal expansion is a signal model that uses a linear combination of a small number of elementary waveforms to represent or approximate a signal. Such expansions are of increasing interest in signal processing with applications ranging from source coding [1] to de-noising [3], source separation [4] and signal acquisition [5]. However it is still a serious challenge to calculate sparse reconstructions for very large data sets. Such signals can have millions of samples and several thousand significant coefficients.

Let $\mathbf{x} \in \mathbb{R}^M$ be a vector denoting the signal and $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ a matrix with $M < N$. We will refer to $\mathbf{\Phi}$ as the dictionary and call the column vectors ϕ_i of $\mathbf{\Phi}$ atoms. The challenge is to find a vector $\mathbf{y}$ satisfying the relationship:

$$\mathbf{x} = \mathbf{\Phi}\mathbf{y} + \epsilon. \tag{1}$$

However, because $M < N$, there is typically a continuum of such solutions. It is therefore common to search for a vector $\mathbf{y}$ that optimizes a certain sparsity measure. For example, it is common to look for a vector $\mathbf{y}$ with the smallest number of non-zero elements.

In this paper we are interested in greedy methods, the most popular of which are Matching Pursuit (MP) [6] and Orthogonal Matching Pursuit (OMP) [7]. MP is an algorithm often used for practical large scale applications since there are now very efficient $\mathcal{O}(N \log M)$ (per iteration) implementations [8], [9] whenever fast transforms are available. For general dictionaries MP is $\mathcal{O}(NM)$ (per iteration). On the other hand, OMP has superior performance. Current implementations, however, are much more demanding both in terms of computation time and memory requirement. Recently the authors presented a new set of gradient pursuit algorithms that use directional updates to perform approximate orthogonalization but with a computational cost more akin to MP [2]. Here we present a technique that combines these fast directional updates with a very greedy threshold step that is derived from the weak-Matching Pursuit selection rule.

The resulting algorithm selects multiple atoms at each iteration and has a control parameter that allows one to trade off computational cost (number of iterations) with recovery performance. It is therefore well suited to tackle extremely large signals such as dynamic MRI imaging.

A. Notation

Γ^n will denote a set containing the indices of the elements selected up to and including iteration n. Using this index set as a subscript, the matrix $\mathbf{\Phi}_{\Gamma^n}$ will be a sub-matrix of $\mathbf{\Phi}$ containing only those columns of $\mathbf{\Phi}$ with indices in Γ^n. The same convention is used for vectors. For example, $\mathbf{y}_{\Gamma^n}$ is a sub-vector of $\mathbf{y}$ containing only those elements of $\mathbf{y}$ with indices in Γ^n. In general, the superscript in the subscript of $\mathbf{y}_{\Gamma^n}$ reminds us that we are in iteration n, on occasion, however, we resort to using superscripts (e.g. $\mathbf{y}^n$) to label the iteration. The Gram matrix $\mathbf{G}_{\Gamma^n} = \mathbf{\Phi}_{\Gamma^n}^T \mathbf{\Phi}_{\Gamma^n}$ will also be used frequently. In general, lower case bold face characters represent vectors while upper case bold characters are used for matrices. Individual elements from a vector will be in standard type face with a subscript. For example $\mathbf{g}$ will be used to refer to a gradient vector with g_i denoting the i^{th} element of this vector. Inner products between vectors will often be written using angled brackets, e.g. $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$.

II. Greedy Algorithms

The algorithms in this paper approximate a vector $\mathbf{x}$ iteratively. In iteration n we calculate an approximation using

$$\hat{\mathbf{x}}^n = \mathbf{\Phi}_{\Gamma^n} \mathbf{y}_{\Gamma^n}, \tag{2}$$

and calculate the approximation error as

$$\mathbf{r}^n = \mathbf{x} - \hat{\mathbf{x}}^n. \tag{3}$$

In each iteration, the approximation error is then used to determine new elements to be selected from $\mathbf{\Phi}$ in order to find a better approximation.

Greedy algorithms such as MP and OMP use a common selection strategy for the new atom at iteration n by choosing the index such that:

$$i^n = \arg_i \max |\mathbf{\Phi}^T \mathbf{r}^{n-1}| \tag{4}$$

(breaking ties deterministically). However they differ significantly on how they update the residual. A broad family of updates that include those of MP and OMP are described next.

A. Directional updates

The Directional Pursuit family of algorithms can be summarized as follows

1) Initialize $\mathbf{r}^0 = \mathbf{x}, \mathbf{y}^0 = \mathbf{0}, \Gamma^0 = \emptyset$
2) for $n = 1; n := n + 1$ till stopping criterion is met
   a) $\mathbf{g}^n = \mathbf{\Phi}^T \mathbf{r}^{n-1}$
   b) $i^n = \arg_i \max |g_i^n|$
   c) $\Gamma^n = \Gamma^{n-1} \cup i^n$
   d) calculate update direction $\mathbf{d}_{\Gamma^n}$
   e) $\mathbf{c}^n = \mathbf{\Phi}_{\Gamma^n} \mathbf{d}_{\Gamma^n}$
   f) $a^n = \langle \mathbf{r}^n, \mathbf{c}^n \rangle / \|\mathbf{c}^n\|_2^2$
   g) $\mathbf{y}^n_{\Gamma^n} := \mathbf{y}^{n-1}_{\Gamma^n} + a^n \mathbf{d}_{\Gamma^n}$
   h) $\mathbf{r}^n = \mathbf{r}^{n-1} - a^n \mathbf{c}^n$
3) Output $\mathbf{r}^n, \mathbf{y}^n$

Different choices of update result in different algorithms with different degrees of complexity. For example, if the direction chosen in step 2d is $\mathbf{d}_{\Gamma^n} = \Phi^\dagger r^{n-1}$ then we get the OMP algorithm. Similarly if we choose $\mathbf{d}_{\Gamma^n} = \epsilon_{i^n}$, where ϵ_k is the Dirac basis in $\mathbb{R}^N$, we get the MP algorithm.

B. Faster: conjugate gradients updates

In [2] the authors proposed that we should choose directions that aim to minimize:

$$f(\mathbf{y}_{\Gamma^n}) = \|\mathbf{x} - \mathbf{\Phi}_{\Gamma^n} \mathbf{y}_{\Gamma^n}\|_2^2, \tag{5}$$

while still being computationally easy to calculate. In particular we could choose:

1) gradient updates (GP):

$$d_{\Gamma^n} = \mathbf{g}_{\Gamma^n} = \mathbf{\Phi}^T_{\Gamma^n} (\mathbf{x} - \mathbf{\Phi}_{\Gamma^n} \mathbf{y}^{n-1}_{\Gamma^n}). \tag{6}$$

2) conjugate gradient updates (ACGP):

$$d_{\Gamma^n} = \mathbf{g}_{\Gamma^n} + \beta d^{n-1}_{\Gamma_n}, \tag{7}$$

where $d^{n-1}_{\Gamma_n}$ denotes the direction $d_{\Gamma^{n-1}}$ extended by padding with a zero into the n-dimensional space. β is chosen to enforce conjugacy between d_{Γ^n} and $d^{n-1}_{\Gamma_n}$:

$$\beta = -\frac{\left\langle (\mathbf{\Phi}_{\Gamma^n} \mathbf{d}^{n-1}_{\Gamma^n}), (\mathbf{\Phi}_{\Gamma^n} \mathbf{g}^n_{\Gamma^n}) \right\rangle}{\|\mathbf{\Phi}_{\Gamma^n} \mathbf{d}^{n-1}_{\Gamma^n}\|_2^2}. \tag{8}$$

Here we will concentrate on the conjugate gradient update direction which proved to be most effective.

Careful attention to the equations involved in Approximate Conjugate Gradient Pursuit (ACGP) shows that it can be implemented at a computational cost of $2\mathbf{\Phi} + 2M + 7n$ flops per iteration at the nth step and a storage cost of $N + 2M + 2n$. Here $\mathbf{\Phi}$ denotes one (possibly fast) matrix-vector product using either $\mathbf{\Phi}$ or $\mathbf{\Phi}^T$ and it is these operations that dominate the algorithm's computational and storage costs. Specifically, ACGP only requires such evaluations in calculating $\mathbf{g}_{\Gamma^n}$ and $\mathbf{\Phi}_{\Gamma^n} \mathbf{g}^n_{\Gamma^n}$. All other quantities can be calculated via simple recursions.[1]

It can be seen that the cost per iteration is comparable to the cost of an MP iteration. In contrast, OMP has a computational and storage cost that is quadratic in n.

In [2] it was shown that performance of ACGP on incoherent dictionaries is very similar to that of OMP but at a much reduced cost and can be used on problems whose size precludes the computation and/or storage costs required by OMP.

The full computational time of ACGP is also effected by the number of iterations taken. Empirically we have found that we require only slightly more iterations than the number of atoms selected (in the Shepp-Logan phantom example in [2] we needed about 4100 iterations to find 4000 atoms). However this still leaves us with an overall computational cost dominated by $2n\mathbf{\Phi}$ flops which, when the number of significant coefficients, n is large, may still be too many. We therefore next consider how we might reduce the number of iterations required.

C. Greedier: stagewise selection

Another recent innovation in greedy algorithms was proposed by Donoho et al. [11] where the stagewise OMP algorithm (StOMP) was proposed. This differs from OMP in the selection step. Instead of choosing the atom that best correlates with the current residual, StOMP selects all atoms whose inner product with the residual exceeds a given threshold:

$$\Gamma_n = \Gamma_{n-1} \cup \{k : |\langle \phi_k, \mathbf{r}^{n-1} \rangle| \geq \lambda^n\} \tag{9}$$

where λ^n is the threshold at iteration n. In [11] the authors propose using: $\lambda^n_{stomp} = t||r^{n-1}||_2/\sqrt{M}$ where the constant t is recommended to be between $2 - 3$.

Empirically, our experience has been that the selection step in StOMP is not refined enough for general use. This is particularly the case when the dictionary $\mathbf{\Phi}$ is not draw from a uniform spherical ensemble (for which StOMP is specifically designed). We therefore introduce a different selection step.

1) Weak stagewise selection: A different strategy for a *very* greedy selection step can be motivated by the weak-OMP algorithm (WOMP). WOMP replaces the greedy selection step (4) by a weak greedy selection step - select *any* atom ϕ_i such that:

$$|\langle \phi_i, \mathbf{r}^{n-1} \rangle| \geq \alpha \max |\mathbf{\Phi}^T \mathbf{r}^{n-1}| \tag{10}$$

where $0 \leq \alpha \leq 1$ is a control parameter. WOMP was originally proposed to deal with issues related to infinite dimensional dictionaries and has been shown to correctly select atoms under similar conditions to OMP [13]. Here we

[1] This is one $\mathbf{\Phi}$ less than is reported in [2] since we have subsequently identified a further recursion to remove a matrix vector product

propose to use it on finite dimensional dictionaries to select multiple atoms. To this end we define:

Definition 1 (Weak stagewise selection): At each iteration select *all* atoms ϕ_i that satisfy (10).

Surprisingly it does not seem to have been noticed previously that all the results for WOMP in terms of recovery and stability apply equally well to a stagewise weak-OMP algorithm (SWOMP).

Let us briefly consider the relationship between the StOMP selection and SWOMP selection. Both algorithms select atoms by applying thresholding to $\mathbf{\Phi}^T\mathbf{r}^{n-1}$ as described by (9). For the weak stagewise selection we have:

$$\lambda_{wss}^n = \alpha||\mathbf{\Phi}^T\mathbf{r}^{n-1}||_\infty \tag{11}$$

Using norm inequalities we can see that:

$$\frac{\alpha}{t}\sqrt{M}\sigma_1(\mathbf{\Phi})\lambda_{stomp} \geq \lambda_{wss} \geq \frac{\alpha}{t}\sigma_M(\mathbf{\Phi})\lambda_{stomp} \tag{12}$$

where $\sigma_k(\mathbf{\Phi})$ denotes the kth singular value of $\mathbf{\Phi}$. We thus see that the two thresholds are similar, however the key difference lies in the fact that the SWOMP threshold is a function of the correlation between the atoms and the residual rather than only a function of the residual. This also allows us to extend OMP recovery results to SWOMP. We will present these in the next section. First we describe our *preferred* algorithm.

D. Faster & Greedier

To get the benefits of both our fast directional updates and the greedy stagewise selection strategy we present the following (preferred) algorithm which we call stagewise conjugate gradient pursuit (StCGP):

Definition 2 (StCGP): For some pre-determined α, perform the following two steps at each iteration:

1) Weak stagewise selection;
2) Conjugate gradient updating of the residual.

The algorithm can be stopped either when a desired number of atoms has been identified or when the residual becomes small enough.

StCGP thus provides a computationally efficient approximation to the SWOMP algorithm. The weakness parameter, α, can be used to provide faster selection but typically at the expense of requiring the dictionary to be more incoherent (see next section). In the extreme case all atoms can be selected in a single step and then SWOMP becomes a simple thresholding algorithm [12]. It should be noted that when the total number of selection steps is extremely small StCGP will benefit from post-processing to fully orthogonalize the residual. However these circumstances tend to occur when the selection step is substantially smaller (in term of iterations and computation) than the orthogonalization procedure.

III. Recovery analysis

We now consider the performance of our fast and greedy algorithms in the context of incoherent and random dictionaries. We will concentrate on the EXACT-SPARSE problem [13] where the signal x is assumed to have some unknown m-sparse representation in the dictionary $\mathbf{\Phi}$ and we are interested in determining an algorithm's ability to correctly find the nonzero coefficients.

A. Worst case analysis

Let us begin with a worst case analysis for the recovery of an exact sparse signal based upon Tropp's exact recovery condition (ERC). In [13] Tropp showed that a sufficient condition for WOMP to select an atom from the correct subset at each iteration is:

$$\max_{k\notin\Gamma^*} ||\mathbf{\Phi}_{\Gamma^*}^\dagger\phi_k||_1 < \alpha \tag{13}$$

where Γ^* defines the correct subset of atoms. Subsequently Gribonval and Vandergheynst [14] noted that the orthogonalization property was never used in Tropp's original arguments and thus 'general' Matching Pursuit and 'general' Weak Matching Pursuit also pick up only correct atoms under the same conditions. This means for example that the worst case analysis of Tropp also applies to GP and ACGP algorithms [2].

We now make a similar observation that the stepwise nature of the pursuit algorithms is also never used in the proofs. We therefore have the immediate Corollary:

Corollary 1: A sufficient condition for either SWOMP or StCGP to recover the sparsest representation of an exact sparse signal is that (13) holds.

These results also translate into recovery conditions in terms of dictionary coherence measures [13].

B. Exact recovery for random dictionaries

A key emerging application domain for sparse reconstruction algorithms is in compressed sensing. Here we are interested in the ability to accurately estimate the significant coefficients and their values to recover some observed signal (or a sparse representation in some transform). Here it has been shown that an important property for compressed sensing to work is for the sensing matrix (dictionary) to have a restricted isometry property:

$$(1-\epsilon) \leq \frac{||\mathbf{\Phi}_T\mathbf{y}_T||}{||\mathbf{y}_T||} \leq (1+\epsilon) \tag{14}$$

for any subsets T, $|T| < 2m$. This essentially limits the ability for the matrix Φ_T to shrink or magnify distances (i.e. it bounds the condition number of the submatrices of Φ). In particular a certain class of random dictionaries have this property and recently Tropp and Gilbert [15] have derived conditions for exact recovery with high probability using OMP. Specifically, they have shown (using our notation):

Theorem 1: (Tropp & Gilbert [15]) Fix $\delta \in (0, 0.36)$ and choose $M \geq Km\log(N/\delta)$ where K is an absolute constant. Suppose that $\mathbf{y}$ is an arbitrary m-sparse signal in $\mathbb{R}^N$ and draw a random $M \times N$ admissible measurement matrix (dictionary) independent from the signal. Given the data $\mathbf{x} = \mathbf{\Phi}\mathbf{y}$, OMP can reconstruct the signal with probability $1-\delta$.

It is interesting to ask how our algorithms might perform in such a situation.

We begin with the observation that the proof of theorem 1 above [15] can be easily adapted to include a weak stagewise selection step by requiring the greedy selection ratio to always be less than α. The modified result is as follows:

Theorem 2: (SWOMP with random measurements) Assume the same conditions as theorem 1 above but choosing $M \geq K\alpha^{-2}m\log(N/\delta)$. Then SWOMP can reconstruct the signal with probability $1-\delta$.

Extending the result to 'general' MP algorithms does not appear possible since it is necessary in the proof for the algorithms to pick up the correct atoms in at most m steps. Although we always observe that ACGP picks up m atoms in approximately m steps and that StCGP generally takes substantially less steps we have no provable bound for this. However since StCGP is an approximation of SWOMP we typically observe that the performance is similar.

C. Convergence analysis

Although the gradient based directional updates do not fully minimize the residual like OMP it can be shown that under certain circumstances the single optimization step that these updates constitute actually does a pretty good job.

Recall an important condition for correctly retrieving significant atoms is the restricted isometry property, (14). If the dictionary has a small restricted isometry constant, ϵ, then every subdictionary is very nearly orthogonal. Specifically the condition number, κ, of the subdictionary's Gram matrix, $\mathbf{G}_T$ is bounded:

$$\kappa(\mathbf{G}_T) \leq \left(\frac{1+\epsilon}{1-\epsilon}\right)^2 \tag{15}$$

This can be used to explain the good performance of the gradient based updates. A worst case analysis of the gradient line search gives [17]:

$$\begin{aligned}\frac{f(\mathbf{y}^n_{\Gamma^n}) - f(\mathbf{y}^*_{\Gamma^n})}{f(\mathbf{y}^{n-1}_{\Gamma^n}) - f(\mathbf{y}^*_{\Gamma^n})} &\leq \left(\frac{\kappa-1}{\kappa+1}\right)^2 \\ &\leq \left(\frac{2\epsilon}{(1-\epsilon)^2}\right)^2\end{aligned} \tag{16}$$

where $\mathbf{y}^*_{\Gamma^n}$ denotes the least squares solution for (5). Hence for small ϵ the convergence, even of a single gradient iteration, is good. Finally, since the conjugate gradient update minimizes the residual at least as well as the gradient update [18], we can deduce that the CG updates also obey (16).

IV. Numerical experiments

We now examine the performance of StCGP and its ability to exactly recover signals in a simple toy example as well as a more realistic medical image recovery problem. A Matlab implementation of StCGP is available from the second author's web-page.

A. Exact Recovery Performance

To compare the StCGP algorithm to Orthogonal Matching Pursuit, we use a simple toy problem. 10,000 dictionaries of size 128×256 were generated with elements ϕ_i drawn uniformly from the unit sphere. From each dictionary, and at a number of different degrees of sparsity, elements were selected at random and multiplied with unit variance zero mean Gaussian coefficients to generate 10,000 different signals per sparsity level. We first analyze the average performance of various greedy methods in terms of exact recovery of the elements used to generate the signal.

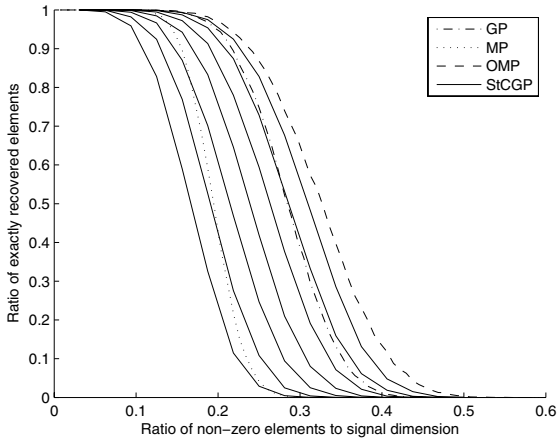

Fig. 1. Comparison between Matching Pursuit (dotted), Orthogonal Matching Pursuit (dashed), Gradient Pursuit (dash-dotted) and Stagewise Conjugate Gradient Pursuit (solid) in terms of exactly recovering the original coefficients. The ordinate shows the fraction of runs in which the algorithms exactly recovered the index set Γ used to generate the data while the abscissa shows the ratio of the size of Γ to the dimension of $\mathbf{x}$. Results averaged over 10 000 runs. The solid lines correspond to (from left to right): $\alpha = 0.7, 0.75, 0.8, 0.85, 0.9, 0.95$ and 1.0.

The results are shown in figure 1. We here show the results for MP, GP StCGP, ACGP and OMP. All algorithms were stopped after they had selected exactly the number of elements used to generate the signal. It is clear that weakening the selection criterion reduces (in a controlled manner) the recovery performance. The advantage of this is a reduction in computational cost. This is shown in figure 2. Here the curves correspond to (going from top to bottom): $\alpha = 1.0, 0.95, 0.9, 0.85, 0.8, 0.75$ and 0.7. The top curve indicates that the computational cost for ACGP (StCGP with $\alpha = 1.0$) grows linearly with the number of non-zero coefficients. In contrast for $\alpha < 1.0$ the computational cost grows much more slowly. It should be noted here that these figures do not fully capture the performance of StCGP since the dictionaries used do not have a fast implementation. However they do provide a fair relative comparison between different values of α.

B. Medical Imaging example

One particularly promising application domain of compressed sensing is Magnetic Resonance Imaging (MRI) [16] and we take our next example from this area using the Logan-Shepp phantom. The physical process of acquiring MRI images is equivalent to taking one dimensional slices from the 2-dimension Fourier domain of the image.

For rapid MR imaging it is desirable to take only a subset of

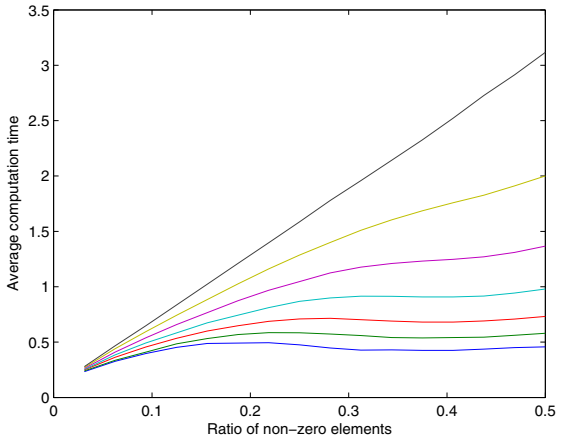

Fig. 2. Comparison of the computation time for StCGP with the different values of alpha as in figure 1. The curves correspond to (going from top to bottom): $\alpha = 1.0, 0.95, 0.9, 0.85, 0.8, 0.75$ and 0.7.

these slices. For example one could take a reduced number of radial lines of the Fourier domain data. In order to reconstruct the original phantom image from the subsampled data we utilize the fact that the image has a sparse representation in the Haar wavelet transform.[2] For this particular image of size 256×256, it was observed that the original image is well approximated (over 300 dB peak signal to noise ratio) using only 4000 of the wavelet coefficients. For this experiment the measurement data consisted of a small number of radial lines sampled from the spatial Fourier domain.

The Logan-Shepp phantom, its sparse representation, and both the fully sampled measurement data and the data subsampled at approximately 15% of Nyquist (42 radial lines in the Fourier domain) are shown in figure 3. This is roughly the limit for OMP to be able to fully reconstruct this image [2].

The performance for GP and ACGP along with OMP and various L_1 methods for this problem were reported in [2] and we refer the reader to this paper for further details. Here we examine the speed and performance of StCGP for α between 0.5 and 1.0. The results are presented in table I. In each case the algorithm was stopped once at least 4000 atoms were selected.[3] Notice that for this data it is possible to obtain an approximate speed up of 80 times using the stagewise algorithm instead of the stepwise version. Even using a relatively conservative value for α, of 0.9, gave an 8 times reduction is computation time.

C. Dynamic MRI example

The above improvements in computation time suggest that StCGP should be a good algorithm for tackling very large-scale problems such as those encountered in dynamic MRI imaging.

[2]It is important to note that we here use a Haar wavelet basis as our sparse representation and not a total variation based constraint as used for example in [5].

[3]These simulations were performed using Matlab running on a 2GHz Pentium PC.

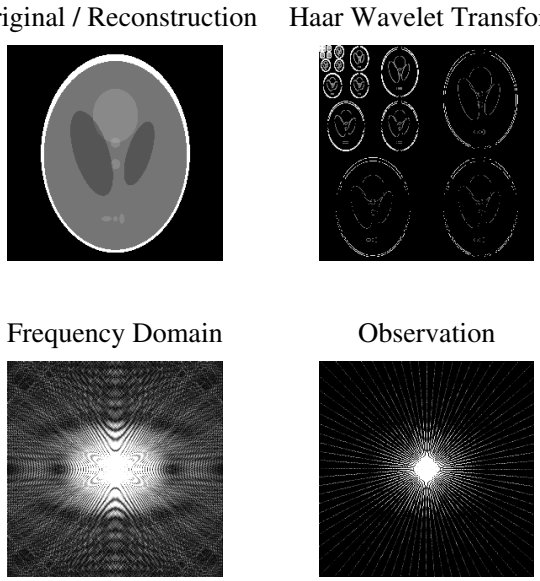

Fig. 3. Magnetic Resonance Imaging (MRI) example. Original phantom image (top left), Fourier domain representation (bottom left), observation of 15% of the frequency coefficients sampled along 42 radial lines (bottom right) and sparse representation in Haar wavelet domain (top right).

TABLE I
THE NUMBER OF ITERATIONS, THE APPROXIMATE COMPUTATION TIME AND PSNR PERFORMANCE (DB) FOR DIFFERENT VALUES OF α IN STCGP.

α	0.5	0.6	0.7	0.8	0.9	1.0
No. of iterations	51	81	214	293	474	4087
computation time (sec.)	19.4	33.3	82	114	182	1562
PSNR (dB)	59	79	311	311	309	301

For this example we used n subsampled version of a fully sampled MRI image sequence of a beating mouse heart to simulate rapid imaging. The sequence consisted of 8 consecutive 256×256 images of the heart and we used a 3-dimensional Haar basis as the sparse representation. As in the previous example, measurements were taken using equally spaced radial lines in the spatial Fourier domain for each image. To add a degree of randomness the orientation of the lines was selected uniformly at random for each image.

Figure 4 shows a plot of the original image sequence as well as the reconstructed image sequence using StCGP with $\alpha = 0.7$ and stopping after selecting $20,000$ atoms. The overall PSNR of the reconstruction was 31.3dB. Furthermore the reconstruction took 50 minutes (44 iterations), which is a speed up of approximately 450 times (based on iteration count) compared with the stepwise algorithm!

V. CONCLUSIONS

We have presented new algorithms that exploit faster updates by approximately orthogonalizing the residual and greedier selection steps that pick up multiple atoms at once. The latter step is controlled by a weakness parameter, α that provides a useful mechanism by which the greediness of the algorithm can be controlled, moving from an iterative stepwise atomic

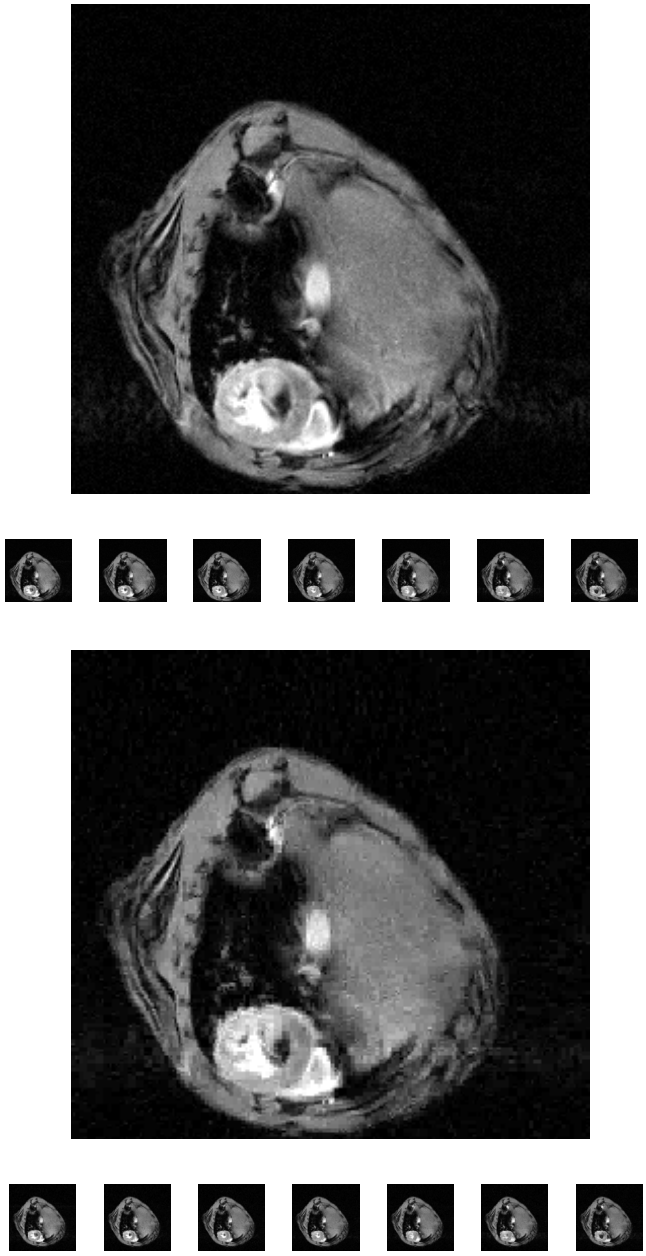

Fig. 4. Dynamic MRI example. The original image sequence (top) and the image sequence reconstructed from 20% of the measurement data (bottom) using StCGP with 44 iterations and $\alpha = 0.7$.

selection to a single thresholding step. Thus one can directly trade computational complexity for recovery performance. The combined technique provides an extremely fast algorithm for tackling large datasets such as those encountered in dynamic MRI.

A further question to examine is whether or not the directional updates can also be used to replace the orthogonalization step in another recently proposed stagewise greedy algorithm: ROMP [19]. ROMP has theoretical recovery performance that is better than OMP and more akin to L_1 minimization. A potential problem of using approximate orthogonalization is that ROMP is allowed to select a small but contained number of incorrect atoms. Since we do not have a bound on the number of iterations needed by the gradient pursuit techniques it may not be possible to ensure that the algorithm does not select too many incorrect atoms.

ACKNOWLEDGMENT

This research was supported by EPSRC grant D000246/1. MED acknowledges support of his position from the Scottish Funding Council and their support of the Joint Research Institute with the Heriot-Watt University as a component part of the Edinburgh Research Partnership.

The authors would like to thank Ian Marshall and Yuehui Tao for providing access to the MRI image data.

REFERENCES

[1] V. K. Goyal, M. Vetterli, and N. T. Thao, "Quantized overcomplete expansions in R^N: Analysis, synthesis and algorithms," *IEEE Transactions on Information Theory*, vol. 44, pp. 16–31, Jan. 1998.

[2] T. Blumensath and M. E. Davies, "Gradient Pursuits," *to appear in IEEE Transactions on Signal Processing*, 2007.

[3] C. Fevotte and S. Godsill, "Sparse linear regression in unions of bases via Bayesian variable selection," *IEEE Signal Processing Letters*, vol. 13, no. 7, pp. 441–444, 2006.

[4] M. Davies and N. Mitianoudis, "A simple mixture model for sparse overcomplete ICA," *IEE Proc.-Vision, Image and Signal Processing*, vol. 151, pp. 35–43, August 2004.

[5] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information.," *IEEE Transactions on Information Theory*, vol. 52, pp. 489–509, Feb 2006.

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

[7] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in 27^{th} *Asilomar Conf. on Signals, Systems and Comput.*, Nov. 1993.

[8] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 1999.

[9] S. Krustulovic and R. Gribonval, "MPTK: Matching pursuit made tractable," in *Proc. Int. Conf. on Acoustic Speech and Signal Processing*, (Toulouse, France), May 2006.

[10] S. Mallat, G. Davis, and Z. Zhang, "Adaptive time-frequency decompositions," *SPIE Journal of Optical Engineering*, vol. 33, pp. 2183–2191, July 1994.

[11] D. Donoho, Y. Tsaig, I. Drori, and J. Starck, "Sparse solutions of underdetermined linear equations by stagewise orthogonal matching pursuit.," 2006.

[12] K. Schnass and P. Vandergheynst, "Average performance analysis for thresholding," *IEEE Signal Processing Letters*, vol 14, no 11, 2007.

[13] J. A. Tropp, "Greed is good: Algorithmic Results for Sparse Approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[14] R. Gribonval and P. Vandergheynst, "On the exponential convergence of matching pursuits in quasi-incoherent dictionaries," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 255–261, 2006.

[15] J. A. Tropp and A. C. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," *submitted,*, 2006.

[16] M. Lustig, D. L. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *submitted*, 2007.

[17] J. R. Shewchuk, "An introduction to the conjugate gradient method without the agonizing pain," *technical report, School of Computer Science, Carnegie Mellon University*, 1994.

[18] H. Crowder and P. Wolfe, "Linear Convergence of the conjugate gradient method," *IBM Journal of Research and Development*, vol 16, pp.431–433, 1972.

[19] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *submitted*, 2007.