<table>
<thead>
<tr>
<th>Title</th>
<th>Greedy pursuits based gradual weighting strategy for weighted 1-minimization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Narayanan, Sathiya; Sahoo, Sujit Kumar; Makur, Anamitra</td>
</tr>
<tr>
<td>Date</td>
<td>2018</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10220/50384">http://hdl.handle.net/10220/50384</a></td>
</tr>
<tr>
<td>Rights</td>
<td>© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at: <a href="https://doi.org/10.1109/ICASSP.2018.8462645">https://doi.org/10.1109/ICASSP.2018.8462645</a></td>
</tr>
</tbody>
</table>
GREEDY PURSUITS BASED GRADUAL WEIGHTING STRATEGY FOR WEIGHTED $\ell_1$-MINIMIZATION

Sathiya Narayanan$^1$, Sujit Kumar Sahoo$^2$ and Anamitra Makur$^1$

$^1$School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
$^2$Department of Statistics and Applied Probability, National University of Singapore, Singapore

Abstract—In Compressive Sensing (CS) of sparse signals, standard $\ell_1$-minimization can be effectively replaced with Weighted $\ell_1$-minimization (Wℓ1) if some information about the signal or its sparsity pattern is available. If no such information is available, Re-Weighted $\ell_1$-minimization (ReWℓ1) can be deployed. ReWℓ1 solves a series of Wℓ1 problems, and therefore, its computational complexity is high. An alternative to ReWℓ1 is the Greedy Pursuits Assisted Basis Pursuit (GPABP) which employs multiple Greedy Pursuits (GPs) to obtain signal information which in turn is used to run Wℓ1. Although GPABP is an effective fusion technique, it adapts a binary weighting strategy for running Wℓ1, which is very restrictive. In this article, we propose a gradual weighting strategy for Wℓ1, which handles the signal estimates resulting from multiple GPs more effectively compared to the binary weighting strategy of GPABP. The resulting algorithm is termed as Greedy Pursuits assisted Weighted $\ell_1$-minimization (GP-Wℓ1). For GP-Wℓ1, we derive the theoretical upper bound on its reconstruction error. Through simulation results, we show that the proposed GP-Wℓ1 outperforms ReWℓ1 and the state-of-the-art GPABP.

Index Terms—Weighted $\ell_1$-minimization, greedy pursuits assisted basis pursuit

I. INTRODUCTION

Compressive Sensing (CS) ensures the reconstruction of a sparse signal $x \in \mathbb{R}^n$ from its measurement vector $y \in \mathbb{R}^m$ of the form $y = \Phi x + v \in \mathbb{R}^m$, where $\Phi \in \mathbb{R}^{m \times n}$ is a known CS matrix with $m \ll n$ and $v$ is the measurement noise [1]-[2]. Let $K$ denote the signal sparsity level (i.e. there are only $K \ll n$ significant entries in $x$). Typically, $m = O(K \log(n))$ [3]. A sensing matrix $\Phi$ is said to obey the Restricted Isometry Property (RIP) if there exists a constant $\delta_K \in [0,1]$ satisfying

$$(1 - \delta_K)\|x\|^2_2 \leq \|\Phi x\|^2_2 \leq (1 + \delta_K)\|x\|^2_2$$

for all $K$-sparse vectors $x \in \mathbb{R}^n$. Note that $\|\cdot\|_2$ stands for the vector $\ell_2$-norm. The sparse signal can be reconstructed from $y$ using CS reconstruction algorithms. They are broadly classified as convex relaxation methods (such as Basis Pursuit or $\ell_1$-minimization [4]) and Greedy Pursuit (GP) algorithms (such as Orthogonal Matching Pursuit (OMP) [5], Subspace Pursuit (SP) [6], Iterative Hard Thresholding (IHT) [7] and Backtracking-based Adaptive Orthogonal Matching Pursuit (BAOMP) [8]). Among these two classes, GPs are known for their faster convergence. Applications of CS include image/video compression [9]-[10], radar echo recovery [11] and ECG signal reconstruction [12].

Motivation and Relation to Prior Work: In reconstruction of sparse signals, standard $\ell_1$-minimization can be effectively replaced with Weighted $\ell_1$-minimization (Wℓ1) if some information about the signal is known apriori [13]-[16]. If no such information is available, Re-Weighted $\ell_1$-minimization (ReWℓ1) can be deployed [17]. As ReWℓ1 solves a series of Wℓ1 problems, its computational complexity is high [18]. The Greedy Pursuits Assisted Basis Pursuit (GPABP) algorithm [19], an effective alternative to ReWℓ1 in no prior information scenario, employs multiple GPs to obtain signal information which in turn is used to run Wℓ1. However, GPABP adapts a binary weighting strategy to determine the weight vector for Wℓ1, which is very restrictive [19]-[20]. Given the support estimates resulting from multiple GPs, a gradual weighting could be adapted. This article proposes a gradual weighting strategy for Wℓ1, which handles the signal estimates resulting from multiple GPs more effectively compared to the binary weighting strategy of GPABP. The resulting algorithm is termed as Greedy Pursuits assisted Weighted $\ell_1$-minimization (GP-Wℓ1).

II. PRIOR WORKS

This section briefly describes Wℓ1, ReWℓ1 and GPABP. The actual support of $x$, $T \subset \{1,2,\ldots,n\}$, is defined as the set of indices $i$ where $x(i)$ is non-zero. In [13], Wℓ1 was proposed for reconstructing sparse signals whose prior information is available in the form of partial support. The partial support, say $T_k \subset \{1,2,\ldots,n\}$, is defined as the set of indices $i$ where $x(i)$ is estimated to be non-zero. The Wℓ1 problem is formulated as

$\hat{x} = \arg \min_{\hat{x}} \|\hat{x}\|_{1,w} \text{ s.t. } \|\Phi \hat{x} - y\|_2 \leq \epsilon$ (2)

where $\hat{x}$ is the reconstructed signal, $w \in [0,1]^n$ and $\|\hat{x}\|_{1,w} := \sum_{i=1}^n w(i)|\hat{x}(i)|$ is the weighted $\ell_1$ norm with $w(i) = \omega \in [0,1]$ whenever $i \in T_k$, and $w(i) = 1$ otherwise. Note that, in (2), $\epsilon$ is the error tolerance (due to the presence of noise in $y$). A similar problem to Wℓ1 was reported in [14] but it assumed a probabilistic prior on the support. If no prior information is available, ReWℓ1 can be applied [17]. ReWℓ1 solves a series of Wℓ1 problems where the weights for the next iteration (say $w^{(i+1)}$) are computed from the value of the current solution (say $\hat{x}^{(i)}$) as follows,

$w(i)^{(i+1)} = \frac{1}{|\hat{x}(i)^{(i)}| + \tau}$ for $i = 1,2,\ldots,n$. (3)

The parameter $\tau > 0$ provides stability. In the initial iteration of ReWℓ1, $\hat{x}^{(0)}$ is obtained by solving (2), fixing $\omega = 1$.

The GPABP algorithm [19], tailored specially for the no prior information scenario, employs multiple GPs to form $\tilde{T}_k$ as $\tilde{T}_k = \bigcap_{i=1}^L \tilde{T}_i$, where $L$ is the number of GPs and $\tilde{T}_i$ is the support set estimated by the $i$th GP. Then, $\tilde{T}_k$ is used to run Modified Basis Pursuit (Mod-BP) [15] as follows,

$\hat{x} = \arg \min_{\hat{x}} \|\hat{x}\|_{1,T_k} \text{ s.t. } \|\Phi \hat{x} - y\|_2 \leq \epsilon$ (4)

where $\tilde{T}_k$ is the set compliment of $\tilde{T}_k$, $\hat{x}_{\tilde{T}_k}$ is the subset of $\hat{x}$ formed by extracting the entries of $\hat{x}$ corresponding to the indices in $\tilde{T}_k$ and $\|\cdot\|_1$ stands for the vector $\ell_1$-norm. The GPABP algorithm was shown to outperform fusion-based CS reconstruction algorithms such as fusion-of-algorithms for CS [21] and committee machine approach for CS [22].
III. PROPOSED WEIGHTING STRATEGY FOR Wℓ₁

If no prior signal/support information is available, either ReWℓ₁ or GRAPAB can be considered. High computational complexity of ReWℓ₁ is not affordable in many practical scenarios. On the other hand, GRAPAB has a limitation: the weighting is only binary (i.e. \(w = 0, 1\)). In this article, in order to have more number of weights, we propose a novel weighting strategy which assigns the weight for each location based on the number of GPs picking that location. This gradual weighting strategy is expected to give more precise weights compared to that of the binary weighting strategy. The resulting algorithm is termed as GP-Wℓ₁. Algorithm 1 shows the step-by-step procedure in GP-Wℓ₁. The GP-Wℓ₁ algorithm employs \(L\) GPs to obtain \(T_i\), \(i = 1, 2, \ldots, L\). The weight corresponding to each location \(i\) is computed as follows,

\[
w'(i) = 1 - \frac{G_i}L, \quad i = 1, 2, \ldots, n,
\]

where \(G_i\) is the number of GPs picking the location \(i\). Then, the weight vector \(w' = [w'(1) \ w'(2) \ldots \ w'(n)]\) is used to obtain \(\hat{x}\) as follows,

\[
\hat{x} = \arg \min_{\tilde{x}} \|\tilde{x}\|_{1,w'} \text{ s.t. } \|\Phi \tilde{x} - y\|_2 \leq \epsilon.
\]

Using forward and reverse triangle inequalities, the above expression becomes,

\[
\sum_{l=0}^{L} w'(l) \|x_{N_l} + h_{N_l}\|_1 \leq \sum_{l=0}^{L} w'(l) \|x_{N_l} + h_{N_l}\|_1 + 2 \sum_{l=0}^{L} w'(l) \|x_{N_l}\|_1.
\]

Noting that the weights range from 0 to 1,

\[
\|h_{T^-}\|_1 \leq \|h_{N_0} + N_1 + \ldots + N_{L-1}\|_1 + 2 \sum_{l=0}^{L-1} w'(l) \|x_{N_l}\|_1.
\]

Let \(\tilde{T} = (T - N_L) \cup (T^- - N_0)\). Then,

\[
\|h_{T^-}\|_1 \leq \|h_{\tilde{T}}\|_1 + 2 \sum_{l=0}^{L-1} w'(l) \|x_{N_l}\|_1.
\]

Using (5), (6), and (7), we have

\[
\|h_{\tilde{T}}\|_1 \leq \sqrt{N} \|x_1\|_2 + 2 \sum_{l=0}^{L-1} w'(l) \|x_{N_l}\|_1.
\]

The following theorem gives the upper bound on GP-Wℓ₁'s reconstruction error.

**Theorem 1:** Let \(x \in \mathbb{R}^n\) and let \(x_K\) be its best \(K\)-term approximation, supported on \(T\). Assume \(L\) GPs are used for weights estimation and let \([1, 2, \ldots, n]\) be partitioned into \(L + 1\) disjoint subsets \(N_l\), \(l = 0, 1, \ldots, L\) where \(N_l\) denotes the set of locations picked by exactly \(l\) (out of \(L\)) algorithms. Suppose that there exists an \(a \in \mathbb{R}_+\) that obeys \(|N_1 \cup N_2 \cup \ldots \cup N_L| = aK\), and the matrix \(\Phi\) obeys RIP with

\[
\delta_{aK} + \frac{a}{\sqrt{1 - \beta'} + a(1 - \alpha')} \delta_{(a+1)K} \leq \frac{a}{\sqrt{1 - \beta'} + a(1 - \alpha')} - 1,
\]

for \(\alpha' = \frac{|N_1 \cup N_2 \cup \ldots \cup N_L|\cap T}{aK}\) and \(\beta' = \frac{|N_0\cap T|}{K}\). Then the solution \(\hat{x}\) to (6) obeys

\[
\|\hat{x} - x\|_2 \leq D_1 \epsilon + D_2 \sum_{l=0}^{L-1} w(l) \|x_{N_l}\|_1,
\]

where \(D_1\) and \(D_2\) are constants (given in the proof below) that depend on \(\Phi\), \(a\), \(\alpha'\) and \(\beta'\).

**Proof:** Let \(\hat{x} = x + h\) be the minimizer of (6). Then

\[
\|x + h\|_{1,w} \leq \|x\|_{1,w'}.
\]

Moreover, due to the choice of weights (given by (5)),

\[
\sum_{l=0}^{L} w'(l) \|x_{N_l}\|_1 \leq \sum_{l=0}^{L} w'(l) \|x_{N_l}\|_1 + 2 \sum_{l=0}^{L} w'(l) \|x_{N_l}\|_1.
\]

By adding \(\sum_{l=0}^{L} (1 - w'(l)) \|x_{N_l}\|_1\) on both sides,

\[
\|h_{T^-}\|_1 \leq \sum_{l=0}^{L} w'(l) \|x_{N_l}\|_1 + 2 \sum_{l=0}^{L} w'(l) \|x_{N_l}\|_1.
\]

Noting that the weights range from 0 to 1,

\[
\|h_{T^-}\|_1 \leq \|h_{N_0} + N_1 + \ldots + N_{L-1}\|_1 + 2 \sum_{l=0}^{L-1} w'(l) \|x_{N_l}\|_1.
\]
As in [13], the coefficients of \( h_T \) are sorted in the order of decreasing magnitude partitioning \( T^c \) into disjoint sets \( T_j \), \( j \in \{1, 2, \ldots\} \) each of size \( aK \), where \( a > 1 \). That is, \( T_1 \) indexes the \( aK \) largest magnitude coefficients of \( h_T \), \( T_2 \) indexes the second \( aK \) largest magnitude coefficients of \( h_T \), and so on. This gives
\[
\|h_T\|_2 \leq (aK)^{-1/2} \|h_{T_1}\|_1.
\]
Let \( T_{01} = T \cup T_1 \), then using the triangle inequality in the above gives
\[
\|h_{T_{01}}\|_2 \leq (aK)^{-1/2} \|h_{T^c}\|_1.
\]
Let \( T_0 = T \cup T_1 \), then using the triangle inequality in the above gives
\[
\|h_{T_0}\|_2 \leq (aK)^{-1/2} \|h_{T^c}\|_1.
\]
Since \( x \) and \( \hat{x} \) are feasible, \( \|\Phi h\|_2 \leq 2\epsilon \) and \( \|\Phi h_{T_{01}}\|_2 \leq 2\epsilon + \sum_{j>1} \|\Phi h_{T_j}\|_2 \leq 2\epsilon + \sqrt{1 + \delta_{aK}} \sum_{j>1} \|h_{T_j}\|_2 \)
where the last inequality follows RIP. Since \( \sum_{j>1} \|h_{T_j}\|_2 = \|h_{T^c}\|_2 \),
using (10) in the above equation gives
\[
\|\Phi h_{T_{01}}\|_2 \leq 2\epsilon + \sqrt{1 + \delta_{aK}} \|h_{T^c}\|_1.
\]
Using RIP on the R.H.S. of the above we get
\[
\sqrt{1 - \delta_{aK}} \|h_{T_{01}}\|_2 \leq 2\epsilon + \sqrt{1 + \delta_{aK}} \|h_{T^c}\|_1.
\]
Since \( T_1 \) contains the largest \( aK \) coefficients of \( h_{T^c} \), \( \|h_{T^c}\|_2 \leq \|h_{T_{01}}\|_2 \), and therefore (8) becomes
\[
\|h_{T^c}\|_1 \leq \sqrt{(1 - \beta + a(1 - \alpha'))K} \|h_{T_{01}}\|_2 + 2 \sum_{l=0}^{L-1} w'(l) \|x_{N_l \cap T^c}\|_1.
\]
Combining (12) and (13) gives
\[
\|h_{T_{01}}\|_2 \leq 2\epsilon + 2 \sqrt{1 + \delta_{aK}} \sum_{l=0}^{L-1} w'(l) \|x_{N_l \cap T^c}\|_1\]
\[
\sqrt{1 - \delta_{aK}} \leq \sqrt{1 - \beta + a(1 - \alpha')}K \sqrt{1 + \delta_{aK}}
\]
Finally, using the relation \( \|h\|_2 \leq \|h_{T_{01}}\|_2 \leq \|h_{T^c}\|_2 \) and substituting (14), (10) and (13) in it gives
\[
\|h\|_2 \leq D_1 \epsilon + D_2 \sum_{l=0}^{L-1} w'(l) \|x_{N_l \cap T^c}\|_1
\]
where the constants
\[
D_1 = \frac{2(1 + \sqrt{1 - \beta + a(1 - \alpha')})}{\sqrt{1 - \delta_{aK}}} \frac{\sqrt{1 + \delta_{aK}}}{\sqrt{1 - \delta_{aK}}}
\]
and
\[
D_2 = \frac{2 \sqrt{1 - \delta_{aK}}}{\sqrt{1 - \delta_{aK}}} \frac{\sqrt{1 - \beta + a(1 - \alpha')}}{\sqrt{1 + \delta_{aK}}}
\]
with the condition that the denominator is positive, equivalently
\[
\delta_{aK} + \frac{a}{\sqrt{1 - \beta + a(1 - \alpha')}} \delta_{aK} < 1.
\]

Remark 1: If \( x \) is exactly sparse (i.e. \( \|x_{N_l \cap T^c}\|_1 = 0 \)) and \( y \) is noiseless (i.e. \( \epsilon = 0 \)), then \( \|h\|_2 = 0 \) which implies the reconstruction is exact.

Remark 2: For theorem 1 to hold, it is sufficient if \( \Phi \) satisfies
\[
\delta_{aK} < \frac{a - (1 - \beta + a(1 - \alpha'))}{a + (1 - \beta + a(1 - \alpha'))}.
\]

Compared to the standard \( \ell_1 \)-minimization [2] which requires \( \Phi \) to satisfy \( \delta_{aK} < \frac{\sqrt{a}}{a+1} \), GP-W\( \ell_1 \) has a weaker requirement provided
\[
\beta > a(1 - \alpha').
\]
As per the definitions of \( a, \alpha', \) and \( \beta', \) the above condition holds when the ingredient GPs in GP-W\( \ell_1 \) (i.e., GP\( _1 \) in step 1 of algorithm 1) estimate the signal support such that
\[
|(N_1 \cup N_2 \cup \ldots \cup N_L \cap T^c) | + |N_L \cap T^c | > |N_1 \cup N_2 \cup \ldots \cup N_L|.
\]

There are two favourable scenarios for the above condition to hold. First, every GP in GP-W\( \ell_1 \) should pick most of the correct non-zero locations. This will lead to the LHS of (20) being close to 2\( K \). Second, all GPs should estimate nearly the same signal support. More importantly, the union of wrong locations picked by GPs (i.e. \( (N_1 \cup N_2 \cup \ldots \cup N_L) \cap T^c \)) should be a small-sized set. This will lead to the RHS of (20) being close to \( K \). Consider a typical reconstruction example where \( n = 50, L = 3 \) and \( T = \{2, 4, 6, 8, 10, 12\} \). Let the signal estimates resulting from \( L \) GPs be \( T_1 = \{2, 4, 6, 8, 10, 11\}, T_2 = \{2, 4, 6, 8, 12, 15\} \) and \( T_3 = \{2, 4, 6, 8, 11, 12\} \). This will result in \( N_1 = \{10, 15\}, N_2 = \{11, 12\} \) and \( N_3 = \{2, 4, 6, 8\} \). Therefore, \( \|N_L \cap T | = 4 > |N_1 \cup N_2 \cup \ldots \cup N_L| \cap T^c | = 6 \). This condition in (20) holds, for this reconstruction scenario, GP-W\( \ell_1 \) has a more favorable RIP requirement compared to that of the standard \( \ell_1 \)-minimization.

**IV. EXPERIMENTAL RESULTS**

Signal reconstruction is performed using three methods: ReW\( \ell_1 \) [18], GPABP [19], and the proposed GP-W\( \ell_1 \). Both GPABP and GP-W\( \ell_1 \) involve four GPs (OMP [5], SP [6], IHT [7] and BAOMP [8]). The cvx solver [23] is used for the implementation of W\( \ell_1 \) in all
three methods. For ReW\(_{\ell_1}\), \(\tau\) is fixed as 0.1 (a value slightly smaller than the expected non-zero magnitudes of \(x\)).

**Synthetic Sparse Signals:** For our experiments on synthetic signals, Gaussian sparse signals of length \(n = 250\) are generated. First experiment presents the reconstruction error as a function of Measurement Ratio (MR = \(m/n\)). An \(m \times n\) Gaussian random measurement matrix is generated to acquire the sparse signal. In this experiment, \(y\) is corrupted by a noise such that its Signal to Measurement Noise Ratio (SMNR) is 20 dB. The SMNR is defined as

\[
\text{SMNR (in dB)} = 10 \log_{10} \frac{\|x\|_2^2}{\|v\|_2^2}.
\]

The constraint of W\(_{\ell_1}\) is fixed as \(\|\Phi \hat{x} - y\|_2 \leq 0.001\). Sparsity level \(K\) is fixed as 30 and the MR is varied between 0.2 and 0.48. For each value of MR, 250 independent trials are performed to obtain the average results. Fig. 2 shows the average Mean Square Error (MSE = \(\frac{1}{n}\|x - \hat{x}\|_2^2\)) as a function of MR. For MRs greater than 0.3, GP-W\(_{\ell_1}\)'s convergence is much faster than that of the ReW\(_{\ell_1}\) and comparable to that of the GPABP. Next experiment reports the effect of \(L\). The experimental set-up is the same as that of the previous experiment and the GP-W\(_{\ell_1}\)'s performance is recorded for \(L = 2, 3\) and 4. In each trial, support set estimates are obtained for all four GPs mentioned above. In the case of \(L < 4\), only \(L\) randomly chosen support set estimates are used for weights estimation. It can be seen from fig. 4 that the accuracy of GP-W\(_{\ell_1}\) increases with an increase in \(L\).

**Compressible ECG signals:** For conducting an experiment on real world signal, the leads (i.e. ECG signals) are extracted from records 100, 101, 102 and 103 from the MIT-BIH Arrhythmia database [24]. These signals are divided into chunks of \(n = 250\) samples (with their amplitudes ranging from 0 to 255). The sparse representations of these signals are obtained using the discrete cosine transform. We fixed \(K = \lfloor \frac{m}{\log m} \rfloor\). The measurement vector \(y\) is obtained using an \(m \times n\) Gaussian random matrix \(\Phi\). Then, \(y\) is corrupted by a noise such that its SMNR is 20 dB. Fig. 5 shows the average MSE as a function of MR, for three methods (ReW\(_{\ell_1}\), GPABP and GP-W\(_{\ell_1}\)). For MRs greater than 0.35, GP-W\(_{\ell_1}\) gives the least MSE.

**V. CONCLUSION**

We proposed a gradual weighting strategy for W\(_{\ell_1}\), which handles the signal estimates resulting from multiple GPs more effectively compared to the binary weighting strategy of GPABP. For the proposed GP-W\(_{\ell_1}\), we derived the theoretical upper bound on its reconstruction error. Experimental results showed that, in reconstruction of sparse signals with no prior information, GP-W\(_{\ell_1}\) outperforms GPABP and ReW\(_{\ell_1}\).
REFERENCES


