Enhancing the Performance of the Bayesian Pursuit Algorithm

B. Deka and P. K. Bora
Department of Electronics and Communication Engineering
Indian Institute of Technology Guwahati
Guwahati, 781039, India
Email: {bhabesh_d, prabin}@iitg.ernet.in

Abstract—Finding sparse solutions to under-determined systems of linear equations has recently got a plethora of applications in the field of signal processing. It is assumed that an ideal noiseless signal has sufficiently sparse representation. But in practice a noisy version of such signal can only be observed. In this paper, we propose a new initialization scheme and a stopping condition for the recently introduced Bayesian Pursuit Algorithm (BPA) for sparse representation in the noisy settings. Experimental results show that the proposed modifications lead to a better quality of sparse solution and faster rate of convergence over the existing BPA especially at low noise levels.

I. INTRODUCTION

Finding relevant sparse solutions of under-determined systems of linear equations in the presence of noise has been used popularly in signal processing community. It has found applications in a wide range of diverse fields. These include image denoising [1], blind source separation (BSS) [2] and compressed sensing [3]. The sparse representation of a signal is modeled by

\[ y = \Phi x + n \] (1)

where \( y \) is an \( m \times 1 \) signal vector, \( x \) is a \( k \times 1 \) sparse coefficient vector, \( \Phi \) is an \( m \times k \) matrix called the dictionary and \( n \) is an \( m \times 1 \) noise vector. It is assumed that \( m < k \) which means that the dictionary is overcomplete. The columns of the dictionary are called atoms.

The representation of a signal with an overcomplete dictionary has the advantage over traditional basis representations because they offer a wider range of generating elements (bases) and hence are more flexible in signal representation. It is most suitable for tasks like signal detection and data compression [4]. A theoretical justification of the use of an overcomplete dictionary has been given in [5]. We assume here that the signal under consideration has a sparse representation by the overcomplete dictionary. Since the dictionary is overcomplete, the under-determined system of linear equations has infinitely many solutions. It is shown that under some sparsity constraints the solution is also unique [6].

Finding the sparsest solution, that is, the solution with the minimum number of nonzero elements, is an NP-hard combinatorial problem [7]. This problem can be solved in a tractable way using two approaches: 1) optimization techniques and 2) greedy algorithms. The first category solves the problem by minimizing a cost function and the second tries to find the nonzero elements directly through correlations. The optimization techniques are broadly divided into two groups, namely, the convex and non-convex optimization methods. The basis pursuit [8] (BP) applies convex optimization that uses the \( \ell^1 \)-norm in place of \( \ell^0 \)-norm as the cost function. The resulting optimization problem is solved by the linear programming approach. The Focal Under-determined System Solver (FOCUSS) [9] is an important technique of non-convex optimization in the noise free case. It uses the \( \ell^p \)-norm with \( p < 1 \) instead of the \( \ell^0 \)-norm. The Bayesian methods such as the Sparse Bayesian Learning (SBL) [10], and Bayesian Compressive Sensing approach (BCS)[11] are also in the non-convex category.

The greedy sparse representation algorithms generally use the correlation between the signal (or residual signal) and the atoms of the dictionary as a measure to find the atoms with nonzero coefficients. There are a number of pursuit algorithms like the Matching Pursuit (MP) [12], Orthogonal Matching Pursuit (OMP) [13], Stage-wise OMP (StOMP) [14], Gradient Pursuit (GP) [15], Stage-wise weak Gradient Pursuit (StGP) [16], Bayesian Pursuit Algorithm (BPA) [7]. The greedy algorithms determine one active atom (eg., in MP) or several active atoms (eg., in StOMP) recursively at a time without solving a hard optimization problem in a multidimensional space. The MP determines it by exhaustively searching the atoms with highest correlations. In StOMP, it is done by comparing the correlations with a threshold [14]. The BPA algorithm is based on a hypothesis testing as an activity measure to decide the nonzero components of the sparse vector in a Bayesian framework.

This paper critically examines the BPA and suggests modifications for enhanced performance. The initial solution required for the BPA is given by the \( \ell^2 \)-norm solution which is not sparse. In the proposed modifications, first a low-resolution sparse solution is obtained using the StOMP and then the BPA is initialized with this solution. We also suggest a new criterion for stopping the iterations of the BPA which is more theoretically justified than the one proposed in the BPA.

The rest of the paper is organized as follows. Section II presents a model for the sparse coefficient vector followed by the theoretical foundation of the BPA in section III. Section IV describes the proposed algorithm. The experimental results are presented in section V and finally section VI concludes the
II. MODELING THE SPARSE VECTOR

The model given by equation (1), the noise vector \( \mathbf{n} \) is assumed to be additive white Gaussian noise. The coefficient vector \( \mathbf{x} = [x_1, x_2, \ldots, x_k]^T \) is modeled by the Bernoulli-Gaussian (BG) model [7] where a coefficient is inactive with probability \( p \), and active with probability \( 1 - p \). According to this model, each coefficient \( x_i \) can be written as \( x_i = q_i r_i \) where \( q_i \) represents the activity of each coefficient such that

\[
q_i = \begin{cases} 
0 & \text{if } x_i \text{ is inactive with probability } p \\
1 & \text{if } x_i \text{ is active with probability } (1 - p) 
\end{cases} \tag{2}
\]

and \( r_i \sim N(0, \sigma_e^2) \) is the amplitude of each coefficient. Therefore the coefficient vector \( \mathbf{x} \) can be written as

\[
\mathbf{x} = \mathbf{Qr} \tag{3}
\]

where \( \mathbf{Q} = \text{diag}[q_1, q_2, \ldots, q_k] \) and \( \mathbf{r} = [r_1, r_2, r_3, \ldots, r_k]^T \) is the amplitude vector.

III. BAYESIAN PURSUIT ALGORITHM (BPA)

The BPA determines the activity measure of the atoms by means of a Bayesian hypothesis testing strategy on the correlation of the residual signal with the dictionary atoms. The method is briefly discussed below.

The equation (1) can be rewritten as

\[
\mathbf{y} = \sum_{i=1}^{k} \phi_i x_i + \mathbf{n} \tag{4}
\]

where \( \phi_i \) is the \( i^{\text{th}} \) atom in the dictionary normalized to have unity norm.

The correlation \( z_j \) between \( \mathbf{y} \) and the \( j^{\text{th}} \) atom \( \phi_j \) is given by

\[
z_j \triangleq \langle \mathbf{y}, \phi_j \rangle = x_j + \sum_{i \neq j} x_i b_{ij} + v_j \tag{5}
\]

where \( b_{ij} \triangleq \langle \phi_i, \phi_j \rangle \) and \( v_j \triangleq \langle \mathbf{n}, \phi_j \rangle \). Define \( \mathbf{z} = [z_1, z_2, \ldots, z_k]^T \)

The BPA considers two hypotheses:

- \( H_1 \): the \( j^{\text{th}} \) atom is active
- \( H_2 \): the \( j^{\text{th}} \) atom is inactive

Assuming that the coefficients except the \( j^{\text{th}} \) one is known from the previous estimation, equation (5) can be rewritten as

\[
z_j = \sum_{i \neq j} \hat{x}_i b_{ij} + x_j + \sum_{i \neq j} (x_i - \hat{x}_i) b_{ij} + v_j \tag{6}
\]

where \( \hat{x}_i \) is the estimation of the \( i^{\text{th}} \) coefficient at the current iteration. Define

\[
m_j \triangleq \sum_{i \neq j} \hat{x}_i b_{ij} \\
\gamma_j \triangleq \sum_{i \neq j} (x_i - \hat{x}_i) b_{ij} + v_j \tag{7}
\]

The two hypotheses \( H_1 \) and \( H_2 \) become

\[
H_1 : z_j - m_j = r_j + \gamma_j \\
H_2 : z_j - m_j = \gamma_j
\]

where \( m_j \) is known and \( \gamma_j \) depends on the noise. In [7], the authors consider the posterior densities \( f(H_1|z_j) \) and \( f(H_2|z_j) \). The hypothesis \( H_1 \) is true if

\[
f(H_1|z_j) > f(H_2|z_j) \tag{9}
\]

Using the Bayes’ rule, the above posteriors can be written as

\[
f(H_1|z_j) = \frac{P(H_1) f(z_j|H_1)}{P(z_j)} \text{ and } f(H_2|z_j) = \frac{P(H_2) f(z_j|H_2)}{P(z_j)}
\]

respectively. Substituting these into equation (9) leads to the following decision rule [7] for the hypothesis testing

\[
q_j = \begin{cases} 
1 & \text{if } |z_j - m_j| > \text{Th}_j \\
0 & \text{otherwise} 
\end{cases} \tag{10}
\]

where \( \text{Th}_j \) is given by

\[
\text{Th}_j = \frac{\sigma_{\gamma_j}}{\sigma_r} \sqrt{2(\sigma_r^2 + \sigma_{\gamma_j}^2) \ln \left( \frac{p \sqrt{\sigma_r^2 + \sigma_{\gamma_j}^2}}{(1 - p)\sigma_{\gamma_j}} \right)} \tag{11}
\]

The unknown parameters \( p, \sigma_r \) and \( \sigma_{\gamma_j} \) are estimated from the original signal \( \mathbf{y} \). The estimates of the parameters \( p, \sigma_r \) and \( \sigma_{\gamma_j} \) are [7]

\[
\hat{p} = \frac{\|\mathbf{q}\|_0}{k}, \hat{\sigma}_r = \frac{\|\mathbf{r}\|_2}{\sqrt{k}}, \hat{\sigma}_n = \frac{\|\mathbf{y} - \Phi \mathbf{x}\|_2}{\sqrt{m}} \tag{12}
\]

and \( \sigma_{\gamma_j} \) is estimated as

\[
\sigma_{\gamma_j}^2 = \sigma_n^2 + \sum_{i \neq j} b_{ij}^2 \sigma_{\epsilon,ee}^2 \tag{13}
\]

where \( \sigma_{\epsilon,ee}^2 \) is the variance of the coefficient error \( x_i - \hat{x}_i \). After updating the activity vector \( \mathbf{q} \) by equation (10), the estimation of the amplitude vector \( \mathbf{r} \) is carried out by the Linear Least Square (LLS) estimation as

\[
\hat{\mathbf{r}} = \sigma_r^2 \hat{\mathbf{Q}} \Phi^T (\sigma_r^2 \Phi \hat{\Phi}^T + \sigma_{\epsilon,ee}^2 \mathbf{I})^{-1} \mathbf{y} \tag{14}
\]

where \( \hat{\mathbf{Q}} = \text{diag}[\hat{q}_1, \hat{q}_2, \ldots, \hat{q}_k] \) and \( \hat{\Phi} = [\hat{q}_1, \hat{q}_2, \ldots, \hat{q}_k]^T \) is the updated activity vector. Finally, the \( \hat{x} \) is obtained by equation (3).

IV. PROPOSED MODIFICATION OF THE BPA

A. Motivation

The BPA requires an approximate solution to start the iteration and an estimate for \( p \). The BPA is initialized with the \( \ell^2 \)-norm solution and \( p \) is arbitrarily assigned to 0.8 [17]. We make the following observations:

1) The \( \ell^2 \)-norm solution is not a sparse solution. The initialization of the BPA with the \( \ell^2 \)-norm solution may not be the best approximation. Such initialization is not necessarily the best even in the absence of a priori information. Depending on \( \Phi \), it may be strongly biased towards a particular solution [9].
The BPA is initialized with a constant $p$. It is worthwhile to investigate the role of the initial value of $p$ in the quality of solution at various noise levels.

3) It is important to study the convergence of the BPA under different initial solutions and estimates of $p$.

4) In the BPA, the stopping criterion heuristically assumes that $\sigma_{r,c}$ decreases linearly with a coefficient $\alpha$ whose value is taken to be close to unity after every iteration. Finally, it becomes zero and $\sigma_{r,c} \approx \sigma_n$. Therefore, there is scope to design for a better stopping criterion for the convergence of the algorithm.

B. Performance of the BPA for different initial solutions and estimates of $p$

To address the issues in 1-3, the BPA is initialized with different initial solutions and estimates of $p$. The following cases are considered:

1) Take the true solution as the initial solution and the true $p$ as the initial estimate of $p$.
2) Take the $\ell^2$-norm solution $x = \Phi^\dagger y$ and the true value of $p$ where $\Phi^\dagger$ is the pseudo inverse of $\Phi$ given by $\Phi^\dagger = (\Phi \Phi^T)^{-1}$.
3) Take an approximate sparse solution obtained by running different pursuit algorithms namely, the StOMP and the OMP for a few iterations as the initial solution and a true value of $p$.
4) Take an approximate sparse solution obtained by the StOMP for a few iterations as the initial solution. Estimate $p$ from the approximate sparse solution according to equation (12).

The following observations are made on a case-by-case basis:

1) The root mean square error (RMSE) of the sparse representation (detailed in the results section) is lower compared to the BPA at different noise levels.
2) The RMSE is higher compared to the BPA at different noise levels. This shows that the accuracy of the final BPA solution depends upon the initial solution.
3) A better quality of solution is obtained with the StOMP initialization compared to the OMP initialization in terms of the RMSE at different noise levels.

C. Stopping Criterion

In practice, to speed up the algorithm, same variance and threshold are used for all the co-efficients which reduces the number of computations by $\frac{1}{m}$ [18]. Thus the activity measure in equation (10) leads to a index set $I_s$ indicating the indices of the active atoms in the current iteration. Now we project the vector $y$ on the columns of $\Phi$ belonging to the support $I_s$. Let $\Phi_I$ denote the matrix with columns chosen using the index set $I_s$. We then obtain the sparse representation $x_s$ given by

$$x_s |_{I_s} = (\Phi_I^T \Phi_I)^{-1} \Phi_I^T y$$

(15)

Then we compute the residual using $r_s = y - \Phi x_s$. The algorithm stops if $||r_s||^2_2 \leq (C.\sigma_n)^2$ where $C$ is a constant multiplier.

D. The Modified Algorithm

On the basis of the above discussion, we propose a two-stage algorithm for finding a sparse representation using the StOMP and the BPA. The proposed algorithm initializes the BPA with an approximate sparse solution obtained by the StOMP with only a few iterations. The modified BPA with the proposed initialization scheme and stopping criterion is shown by a block diagram in Fig. 1. The algorithm can be summarized as follows

Algorithm 1 Modified BPA

1) Take an initial sparse solution $x^{(0)}$ from the StOMP with 3 iterations.
2) Estimate $\hat{p}^{(0)}$, $\hat{\sigma}_r^{(0)}$ and $\hat{\sigma}_n^{(0)}$ using equation (12).
3) Find the set of active atoms and the sparse solution using the BPA.
4) Update $p$, $\sigma_r$, and $\sigma_n$.
5) Compute the residual as discussed in Section IV-C.
6) Check the stopping condition, if it is true then stop the BPA iteration and take the current solution as the final solution otherwise go to step 3 and repeat the steps 3-6.

V. EXPERIMENTAL RESULTS

A number of experiments are carried out to study the performance of the proposed algorithm. We use a random dictionary matrix with normalized columns. The dictionary
is drawn according to a uniform distribution in [-1, 1]. The number of atoms is taken as \( k = 512 \) and the signal length as \( m = 256 \). For initializing the sparse coefficient vector we use the BG model (as mentioned in section II) with the probability \( p = 0.9 \) for the inactive coefficients and all active coefficients are set to be Gaussian distributed with unit variance. Thus 51 atoms are made active for sparse representation of the signal. This is similar to the procedure adopted in [7] to make a fair comparison to the performance of the proposed algorithm. To generate the observed signal we consider the AWGN noise of varying levels in the range \( \sigma_n = 0.02, 0.025, 0.03, \ldots, 0.2 \).

To evaluate the performance of the proposed algorithm, we have considered mainly the following measures:

- Representation error or RMSE: \( \| \hat{x}_{0,e} - \hat{x}_0 \|_2 \) versus the noise level \( \sigma_n \).
- Reconstruction error: \( \| \Phi \hat{x}_{0,e} - y_0 \|_2 \) versus noise level \( \sigma_n \).
- Average number of iterations for convergence versus noise levels.

All the results presented are obtained by averaging the results with 100 different realizations of the dictionary, sparse vector and noise vector. The initial values of the parameters \( p, \sigma_r, \) and \( \sigma_n \) are computed according to the procedure discussed in section IV. For the BPA we have chosen the initialization exactly same as given in [7]. For OMP and StOMP algorithms, (implemented by SparseLab\(^1\)) we use 50 and 20 iterations respectively. The sensitivity parameter in StOMP for threshold selection is chosen as 0.5. The BPA is run for 20 iterations and for convergence the parameter \( \alpha \) is selected as 0.95.

The first set of experiments are carried out to study the performance of the BPA for different initial solutions and estimates of \( p \). We compare the representation error or RMSE for the BPA as well as other pursuit algorithms. First, we initialize the BPA with the true solution and a true value of \( p \). The results are shown in Fig. 2(a). Next, the BPA is initialized with an \( \ell^2 \)-norm solution and a true value of \( p \) and the results are shown in Fig. 2(b). Finally, the result for the BPA initialized with a StOMP and the true value of \( p \) is shown in Fig. 3(a). From these results we conclude that the final BPA solution is affected by the initial solution. The BPA initialized with a low-resolution sparse solution leads to better quality of solution and approaches the true sparse solution.

The second set of experiments are carried out by initializing the BPA with the proposed initialization. We consider a low-resolution sparse initialization of the BPA by StOMP with 3 iterations and an estimate of \( p \) from such initial solution. Fig. 3(b) shows the performance of OMP, StOMP, BPA and the proposed algorithm in terms of representation errors. It is clearly observed that the proposed algorithm outperforms the BPA in terms of the representation error when initialized

\(^1\)The codes SolveOMP, SolveStOMP are publicly available at http://sparselab.standford.edu
by the StOMP at different noise levels. We have also varied the number of iterations of StOMP in the initialization of the proposed algorithm in our experiments. It is observed that the results for the proposed algorithm are very similar even if the number of iterations of StOMP in the initialization are increased.

Fig. 4 shows the reconstruction error versus noise level for various sparse algorithms. The proposed algorithm outperforms the BPA at different noise levels. Although the proposed algorithm outperforms the StOMP in terms of both the representation and reconstruction errors at low noise levels, the StOMP is a better choice at higher noise levels than either of the BPA or the proposed algorithm.

Finally, to study the convergence rate of the proposed algorithm vs. that of the BPA, we have computed the average number of iterations required for convergence of the proposed algorithm at different noise levels. Fig. 5 clearly indicates that the proposed algorithm converges with a very less no. of iterations than the BPA at different noise levels. On an average, the proposed algorithm converges with less than six iterations for various noise levels.

VI. CONCLUSION

In this paper, we have proposed a modified version of the BPA. The main idea behind this algorithm is that we can improve the performance of the sparse representation as well as the rate of convergence of the original BPA by a sparse initialization and redefining the stopping condition of the original BPA. Simulations show the advantage of the proposed modification in terms of representation error, reconstruction error, and the rate of convergence than the state-of-the-art sparse representation algorithms especially at lower levels of noise. Work is also currently in progress to apply the proposed algorithm for the removal of additive white Gaussian noise from grayscale images and the preliminary results in this direction are also very encouraging.

REFERENCES


