ABSTRACT

Compressed sensing shows that a sparse or compressible signal can be reconstructed from a few incoherent measurements. Noting that sparse signals can be well modeled by algebraic-tailed impulsive distributions, in this paper, we formulate the sparse recovery problem in a Bayesian framework using algebraic-tailed priors from the generalized Cauchy distribution (GCD) family for the signal coefficients. We develop an iterative reconstruction algorithm from this Bayesian formulation. Simulation results show that the proposed method requires fewer samples than most existing reconstruction methods to recover sparse signals, thereby validating the use of GCD priors for the sparse reconstruction problem.

Index Terms— Compressed sensing, Bayesian methods, signal reconstruction, nonlinear estimation, impulse noise

1. INTRODUCTION

Compressed sensing (CS) is a novel framework that goes against the traditional data acquisition paradigm. CS demonstrates that a sparse, or compressible, signal can be acquired using a low rate acquisition process that projects the signal onto a small set of vectors incoherent with the sparsity basis [1]. Let \( x \in \mathbb{R}^n \) be an sparse signal, and \( y = \Phi x \) a set of measurements with \( \Phi \) an \( m \times n \) sensing matrix \((m < n)\). The optimal algorithm to recover \( x \) from the measurements is

\[
\min_x \|x\|_0 \quad \text{subject to} \quad \Phi x = y \tag{1}
\]

(optimal in the sense that finds the sparsest vector \( x \) such that is consistent with the measurements). Since noise is always present in real data acquisition systems, the acquisition system can be modeled as

\[
y = \Phi x + r \tag{2}
\]

where \( r \) represents the sampling noise.

The problem in (1) is combinatorial and NP-hard; however, a range of different algorithms have been developed that enable approximate reconstruction of sparse signals from noisy compressive measurements [1, 2, 3]. The most common approach is to use Basis Pursuit Denoising (BPD) [1], which uses the unconstrained convex program

\[
\min_x \|y - \Phi x\|_2^2 + \lambda \|x\|_1, \tag{3}
\]

to estimate a solution of the problem. A family of iterative greedy algorithms ([2] and references therein) are shown to enjoy a similar approximate reconstruction property, generally with less computational complexity. However, these algorithms require more measurements for exact reconstruction than the \( L_1 \) minimization approach.

Recent works show that nonconvex optimization problems can recover a sparse signal with fewer measurements than current geometric methods, while preserving the same reconstruction quality [4, 5]. In [4], the authors replace the \( L_1 \) norm in BPD with the \( L_p \) norms, for \( 0 < p < 1 \), to approximate the \( L_0 \) norm and encourage sparsity in the solution. Candès et al use a re-weighted \( L_1 \) minimization approach to find a sparse solution in [5]. The idea is that giving a large weight to small components encourages sparse solutions.

The CS reconstruction problem can also be formulated in a Bayesian framework (see [6] and references therein), where the coefficients of \( x \) are modeled with Laplacian priors and a solution is iteratively constructed. The basic premise in CS is that a small set of coefficients in the signal have larger value than the rest of the coefficients (ideally zero), yielding a very impulsive characterization. Algebraic-tailed distributions put more mass in very high amplitude values and also in “zero-like” small values, and are therefore more suitable models for sparse coefficients of compressible signals.

In this paper, we formulate the CS recovery problem in a Bayesian framework using algebraic-tailed priors from the generalized Cauchy distribution (GCD) family for the signal coefficients. An iterative reconstruction algorithm is developed from this Bayesian formulation. Simulation results show that GCD priors are a good model for sparse representations. Numerical results also show that the proposed method requires fewer samples than most existing recovery strategies to perform the reconstruction.

2. BAYESIAN MODELING AND GENERALIZED CAUCHY DISTRIBUTION

In Bayesian modeling, all unknowns are treated as stochastic quantities with assigned probability distributions. Consider
the observation model in (2). The unknown signal \( x \) is modeled by a prior distribution \( p(x) \), which represents the a priori knowledge about the signal. The observation \( y \) is modeled by the likelihood function \( p(y|x) \). Modeling the sampling noise as white Gaussian noise and using a Laplacian prior for \( x \), the maximum a posteriori (MAP) estimate of \( x \) is equivalent to finding the solution of (3) [6].

The generalized Cauchy distribution (GCD) family has algebraic tails which makes it suitable to model many impulsive processes in real life (see [7, 8] and references therein). The PDF of the GCD is given by

\[
f(z) = a \delta^{p} + |z|^p \frac{p}{2}\Gamma\left(\frac{2}{p}\right)/\Gamma\left(\frac{1}{p}\right).
\]

with \( a = p\Gamma(2/p)/2(\Gamma(1/p))^2 \). In this representation, \( \delta \) is the scale parameter and \( p \) is the tail constant. The GCD family contains the meridian distribution [7] and Cauchy distributions as special cases with \( p = 1 \) and \( p = 2 \), respectively. For \( p < 2 \), the tail of the PDF decays slower than in the Cauchy distribution, resulting in a heavier-tailed PDF.

Similar to \( L_p \) norms derived from the generalized Gaussian density (GGD) family, a family of robust metrics are derived from the GCD family [3].

**Definition 1** For \( u \in \mathbb{R}^m \), the \( LL_p \) norm of \( u \) is defined as

\[
\|u\|_{LL_p,\delta} = \sum_{i=1}^{m} \log\left\{ 1 + \delta^{-p}|u_i|^p \right\}, \quad \delta > 0.
\]

The \( LL_p \) norm (quasi-norm) doesn’t over penalize large deviations, and is therefore a robust metric appropriate for impulsive environments [3].

## 3. BAYESIAN COMPRESSED SENSING WITH MERIDIAN PRIORS

Of interest here is the development of a sparse reconstruction strategy using a Bayesian framework. To encourage sparsity in the solution, we propose the use of meridian priors for the signal model. The meridian distribution is a special case of GCD and possesses heavier tails than the Laplacian distribution, thus yielding more impulsive (sparser) signal models and intuitively lowering the number of samples to perform the reconstruction.

We model the sampling noise as independent, zero mean, Gaussian distributed samples with variance \( \sigma^2 \). Using the observation model in (2) the likelihood function becomes

\[
p(y|x; \sigma) = \mathcal{N}(\Phi x, \Sigma), \quad \Sigma = \sigma^2 I.
\]

Assuming the signal \( x \) (or coefficients in a sparse basis) are independent meridian distributed samples yields the following prior

\[
p(x|\delta) = \frac{\delta^n}{2^n} \prod_{i=1}^{n} (\delta + |x_i|)^{-2}
\]

Since \( p(x|y; \sigma, \delta) \propto p(y|x; \sigma)p(x|\delta) \), the MAP estimate, assuming \( \sigma \) and \( \delta \) known, is

\[
\hat{x} = \arg \min_{x} \frac{1}{2}||y - \Phi x||_2^2 + \lambda||x||_{LL_1,\delta}
\]

where \( \lambda = 2\sigma^2 \).

One remark to make is that the \( LL_1 \) norm has been previously used to approximate the \( L_0 \) norm but without making a statistical connection to the signal model. The re-weighted \( L_1 \) approach proposed in [5] is equivalent to finding a solution for the first order approximation of the problem in (8) using a decreasing sequence for \( \delta \).

## 4. FIXED POINT ALGORITHM

In this paper, instead of directly minimizing (8), we develop a fixed point search to find a sparse solution. The fixed point algorithm is based on first order optimality conditions and is inspired from the robust statistics literature [9].

Let \( x^* \) be a stationary point of (8), then the first order optimality condition is

\[
\Phi^T \Phi x^* - \Phi^T y + \lambda \nabla ||x^*||_{LL_1,\delta} = 0.
\]

Noting that the gradient \( \nabla ||x^*||_{LL_1,\delta} \), can be expressed as

\[
\nabla ||x^*||_{LL_1,\delta} = W(x^*) x^*,
\]

where \( W(x) \) is a diagonal matrix with diagonal elements given by

\[
W_{ii}(x) = (\delta + |x_i|)|x_i|^{-1},
\]

the first order optimality condition, (9), is equivalent to

\[
\Phi^T \Phi x^* - \Phi^T y + \lambda W(x^*) x^* = 0.
\]

Solving for \( x^* \) we find the fixed point function

\[
x^* = [\Phi^T \Phi + \lambda W(x^*)]^{-1} \Phi^T y
\]

\[
= W^{-1}(x^*) \Phi^T [\Phi W^{-1}(x^*) \Phi^T + \lambda I]^{-1} y.
\]

The fixed point search uses the solution at previous iteration as input to update the solution. The estimate at iteration time \( t + 1 \) is given by

\[
\hat{x}_{t+1} = W^{-1}(\hat{x}_t) \Phi^T [\Phi W^{-1}(\hat{x}_t) \Phi^T + \lambda I]^{-1} y.
\]

The fixed point algorithm turns out to be a reweighted least squares recursion, which iteratively finds a solution and updates the weight matrix using (11).
A fast way to estimate $\delta$ from $x$ is using order statistics (although more elaborate estimates can be used as in [8]). Let $X$ be a meridian distributed random variable with zero location and scale parameter $\delta$ and denote the $r$-th quartile of $X$ as $Q_r(\delta)$. The interquartile distance is $Q_{0.75}(\delta) - Q_{0.25}(\delta) = 2\delta$, thus, a fast estimate of $\delta$ is half the interquartile distance of the samples $x$. Let $Q^t_r(\delta)$ denote the $r$-th quartile of the estimate $\hat{x}_t$ at time $t$, then the estimate of $\delta$ at iteration time $t$ is given by

$$\hat{\delta}_t = 0.5(Q^t_{0.75}(\delta) - Q^t_{0.25}(\delta)).$$  \hspace{1cm} (15)

To summarize, the final algorithm is depicted in Algorithm 1, where $J$ is the maximum number of iterations and $\gamma$ is a tolerance parameter for the error between subsequent solutions. To prevent numerical instabilities we pre-define a minimum value for $\delta$ denoted as $\delta_{\text{min}}$. We start the recursion with the LS solution ($W = 1$) and we also assume a known noise variance, $\sigma^2$ (recall $\gamma = 2\sigma^2$). The resulting algorithm is coined meridian Bayesian compressed sensing (MBCS).

**Algorithm 1 MBCS**

Require: $\lambda$, $\delta_{\text{min}}$, $\gamma$ and $J$.
1: Initialize $t = 0$ and $\hat{x}_0 = \Phi^T(\Phi\Phi^T + \lambda I)^{-1}y$.
2: while $\|\hat{x}_t - \hat{x}_{t-1}\|_2 > \gamma$ or $t < J$ do
3: Update $\delta_t$ and $W$.
4: Compute $\hat{x}_{t+1}$ as in equation (14).
5: $t \leftarrow t + 1$
6: end while
7: return $\hat{x}$

As mentioned in the last section the reweighted $L_1$ approach of [5] and MBCS minimize the same objective. Moreover, the reweighted $L_1$ may require fewer iterations to converge, but the computational cost of one iteration of MBCS is substantially lower than the computational cost of an iteration of reweighted $L_1$, thereby resulting in a faster algorithm.

5. EXPERIMENTAL RESULTS

In this section we present numerical experiments that illustrate the effectiveness of MBCS for sparse signal reconstruction. For all experiments we use random Gaussian measurements matrices with normalized columns and $\delta_{\text{min}} = 10^{-8}$ in the algorithm. The reconstruction SNR (R-SNR) is used as the performance metric in most experiments.

The first experiment shows the validity of the joint estimation approach of MBCS. Meridian distributed signals (in the canonical basis) with length $n = 1000$ and $\delta \in \{10^{-3}, 10^{-2}, 10^{-1}\}$ are used. The signals are sampled taking $m = 200$ measurements and zero mean Gaussian distributed sampling noise with variance $\sigma^2 = 10^{-2}$ is added. Table 1 shows the average R-SNR for 200 repetitions. The performance loss is of 6 dB approximately in the worst case, but fully automated MBCS still yields a good reconstruction.

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>Known $\delta$</th>
<th>Estimated $\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^{-3}$</td>
<td>9.91</td>
<td>8.16</td>
</tr>
<tr>
<td>$10^{-2}$</td>
<td>21.5</td>
<td>17.58</td>
</tr>
<tr>
<td>$10^{-1}$</td>
<td>30.69</td>
<td>24.98</td>
</tr>
</tbody>
</table>

The next set of experiments compare MBCS with current reconstruction strategies for noiseless and noisy samples. The algorithms used for comparison are $L_1$ minimization [1], re-weighted $L_1$ minimization [5], re-weighted least squares (RWLS) to approach $L_p$ [4], and compressive sampling matching pursuit (CoSaMP) [2]. We use $k$-sparse signals ($k$ nonzero coefficients) in the canonical basis of length $n = 1000$, in which the amplitudes of the nonzero coefficients are Gaussian distributed with zero mean and standard deviation $\sigma_x = 10$. Each experiment is averaged over 200 repetitions.

The first experiment compares MBCS in a noiseless setting for different sparsity levels, fixing $m = 200$. We use the probability of exact reconstruction as a measure of performance, where a reconstruction is considered exact if $\|\hat{x} - x\|_\infty \leq 10^{-4}$. The results are shown in Fig. 1 (Middle). Results show that MBCS outperforms CoSaMP and $L_1$ minimization (giving larger probability of success for larger values of $k$) and yielding a slightly better performance than $L_p$ minimization. It is of notice that MBCS has similar performance to reweighted $L_1$, since they are minimizing the same objective, but with a different approach.

The second experiment shows the robustness of the proposed method against sampling noise varying the number of samples ($m$) and fixing $k = 10$. The sampling noise is Gaussian distributed with variance $\sigma^2 = 10^{-2}$. Results are presented in Fig. 1 (Right). In this case MBCS outperforms all other reconstruction strategies, yielding a larger R-SNR for fewer samples with a good approximation for 60 samples and above. Moreover, the R-SNR of MBCS is better than reweighted $L_1$ minimization. An explanation for this that $L_1$ minimization methods suffer from bias problems needing a de-biasing step after the solution is found (see [3] and references therein).

As an experiment with real signals, we present an example utilizing a $256 \times 256$ image. We use a Daubechies db8 wavelets as sparse basis and the number of measurements, $m$, is set to $256 \times 256 / 4$ (25% of the number of pixels of the original image). Fig. 1 (Right) shows a zoom of the normalized histogram of the coefficients along with a plot of meridian and Laplacian distributions. It can be noticed that the meridian is a better fit for the tails of the coefficient distribution. Fig. 2 (Left) shows the original image, Fig. 2 (Middle) the reconstructed image by $L_1$ minimization, and Fig. 2 (Right) the reconstructed image by MBCS. The R-SNR is 15.2 dB and 19.3 dB for $L_1$ minimization and MBCS, respectively. This example shows the effectiveness of MBCS to model and recover sparse representations of real signals.
6. CONCLUSIONS

In this paper, we formulate the CS recovery problem in a Bayesian framework using algebraic-tailed priors from the GCD family for the signal coefficients. An iterative reconstruction algorithm, referred to as MBCS, is developed from this Bayesian formulation. Simulation results show that the proposed method requires fewer samples than most existing reconstruction algorithms for compressed sensing, thereby validating the use of GCD priors for sparse reconstruction problems. Methods to estimate the sampling noise variance are still an open problem. A future research direction is to explore the use of GCD priors with $p$ different from 1 to give more flexibility in the sparsity model.

7. REFERENCES


