Abstract:
A medical image reconstruction method based on sparse Bayesian compressed sensing is presented, and the method employs a hierarchical model of the Laplace prior to model the sparse wavelet coefficients and unknown images. The experiments are designed to compare the Bayesian Compressed Sensing (BCS) method with the Basis Pursuit (BP) algorithm and the Orthogonal Matching Pursuit (OMP) algorithm. The results imply that the presented algorithm exceeds the greedy algorithm and the linear programming such as BP and OMP etc.

Keywords:
Sparse Bayesian; Compressed sensing; Laplace prior; Marginal likelihood; Gaussian distribution; Image reconstruction

1. Introduction

Compressed sensing (CS) is a novel data collection and encoding theory presented by Donoho, Candes and Tao et al. in 2006 [1, 2, 3] which points out that when the original signals or images are sparse or compressible in certain transform domain, one can reconstruct them accurately from a small number of random measurement samples that are much smaller than the number of signal samples in Nyquist-Shannon’s theorem. In the medical application, medical imaging, such as Magnetic Resonance Imaging (MRI) and CT etc, requires faster imaging speed in order to reduce the harm to the human body. It has been proved that most medical images, especially Magnetic Resonance images, are sparse or compressible by certain sparsifying transform basis [4, 5], therefore, employing CS theory to deduce the collected samples and shorten the imaging time is a promising research in the medical imaging.

Although many current CS methods can reconstruct the images well, they have own disadvantages respectively, for instance, greedy algorithm has a fast reconstruction speed but with a bad reconstruction result. Moreover, greedy algorithm requires large number of samples measured which makes the images can not be compressed effectively. Linear programming needs the number of sampling is smaller and with a better reconstruction result than the greedy algorithm, but the reconstruction speed is much slower compared with the greedy algorithm. Especially, the linear programming can not be used to large-scale problem in the practical application.

Bayesian Compressed Sensing (BCS) provides posterior and estimation of model parameters for uncertain information in the reconstruction based on probability statistics, and it has been exploring to solve the sparse reconstruction problems by more and more researchers. BCS was first be presented by Ji et . in 2008 [6] from a Bayesian framework based on the Sparse Bayesian learning and Relevance Vector Machine theory which was developed by Tipping [7] to estimate the underlying signal based on compressed sensing measurements. Taikas et al. [8] proposed an adaptive sparse Bayesian kernel learning method to learn the parameters of the kernels in solving regression and classification. He et al [9] exploited the statistical structure of the wavelet coefficients and constituted a hierarchical Bayesian model via Markov chain Monte Carlo (MCMC) sampling to solve the inverse problem of natural images.

In this paper, we also formulate the sparse reconstruction problem of the images from a Bayesian framework by employing Laplace prior to promote sparsity of the basis coefficients motivated by Li et al. [10] and Babacan et al. [11]. Considering Laplace prior is not conjugate to Gaussian likelihood, the resulting problem is intractable if Laplace prior is utilized on each wavelet coefficient. To solve this problem, a hierarchical approach is employed. More details are described in the

next sections.

2. Bayesian Compressed sensing

Let \( x \) denote the unknown \( N \times 1 \) image signal, which is compressible in certain linear basis \( \Psi \) (here is a wavelet basis), and therefore, it can be expressed as

\[
x = \Psi \omega,
\]

where \( \omega \) is an \( N \times 1 \) sparse coefficients vector with most of the entries being zeros. To acquire the measurements, we take a random undersampling on the sparse wavelet coefficients and obtain the measurement system

\[
y = \Phi \omega + \epsilon,
\]

where \( y \) is the \( M \times 1 \) measurement vector (\( M \ll N \)), \( \Phi \) is \( M \times N \) measurement matrix which subjects to random Gaussian distribution, and \( \epsilon \) represents the acquisition noise. According to CS theory, the unknown image \( x \) can be accurately reconstructed under certain constraint conditions by exploiting the sparsity of the coefficients \( w \) and solving the optimization problem of the \( l_0 \) norm. A common formulation is to solve the \( l_1 \) norm instead of the \( l_0 \) norm as a result of solving \( l_0 \) norm being a NP-hard. So the optimization of \( l_0 \) is transformed into

\[
\omega = \arg \min \| y - \Phi \omega \|^2 + \delta \| \omega \|_1,
\]

where \( \| \| \) denotes the \( l_1 \) norm.

Although \( l_0 \) and \( l_1 \) norms can solve the inverse problem with the deterministic regularization approach, the regularized formulation can not provide the estimation error evaluation. Furthermore, the uncertainty of the input parameters in the prior model and the noise in the measurements can be processed well in a probability data integration manner [6, 7, 11]. The Bayesian method provides a excellent estimate for the uncertainty of input parameters through its probabilistic representations.

2.1. Image and observation models

The observation noise \( \epsilon = y - \Phi \omega \) is an independent Gaussian distribution with zero mean and unknown variance \( \sigma^2 \), and the Gaussian likelihood mode is

\[
p(y | \omega, \beta) = N(y | \Phi \omega, \beta^{-1}),
\]

where \( \beta^{-1} = \sigma^2 \) denoting the inverse of the noise variance (precision) of the Gaussian density function. The unknown noise (observation) variance will be inferred from the measurements during the inversion. Since the conjugate prior for the inverse of noise precision of a Gaussian distribution is Gamma distribution, so we consider Gamma distribution prior on \( \beta_i \) in order to make the analysis easier, that is

\[
p(\beta_i | a_\beta, b_\beta) = \Gamma(\beta_i | a_\beta, b_\beta), i = 1, 2, \ldots, M. \tag{5}
\]

The Gamma distribution is defined as

\[
\Gamma(x | m, n) = \frac{n^m x^{m-1} e^{-nx}}{\Gamma(m)}, x \geq 0. \tag{6}
\]

The above analysis has converted the inverse of CS problem into a linear regression problem with a constraint (prior) of the sparse weights \( \omega \). Since the conjugate prior for the inverse variance of a Gaussian distribution is the Gamma distribution, and, therefore, a Gamma prior is placed on \( \beta \) to estimate it, which makes the analysis simplified greatly.

To obtain the \( l_1 \) regularization formulation in (3), an equivalent conversion form in Bayesian estimation framework is to assign an independent Laplace prior on the coefficients \( \omega \), that is

\[
p(\omega | \lambda) = \frac{1}{2} \exp(-\frac{\lambda}{2} \| \omega \|_1), \tag{7}
\]

where \( \lambda \) is assumed to obey the following Gamma distribution. According to Bayesian theorem, we can estimate \( \omega \) as follows:

\[
\hat{\omega} = \arg \max_{\omega} p(\omega | y) \\
= \arg \max_{\omega} \{p(y | \omega) p(\omega)\} \\
= \arg \min_{\omega} \| y - \Phi \omega \|^2 + \frac{\lambda}{2} \| \omega \|_1. \tag{8}
\]

The solution of Eq. (8) is equivalent to the \( l_1 \) regularization method [12]. This reveals that \( l_1 \) norm constrained optimization problem can employ the Laplace distribution for \( \omega \) to deal with it.

However, the associated Bayesian analysis can not be preformed since the Laplace prior in Eq. (7) is not conjugate to the Gaussian likelihood in Eq. (4). Therefore, a hierarchical prior model is employed [7].

In the first stage of the hierarchical model, independent Gaussian priors are specified on coefficients \( \omega \)

\[
p(\omega | \gamma) = \prod_{i=1}^{N} N(\omega | 0, \gamma_i), \tag{9}
\]

where \( \gamma_i \) is the hyperparameters denoting the variance of the prior distribution and \( \gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_N\} \).

In the second stage of the hierarchy, the Gamma distri-
distribution is specified on $\gamma_i$ with parameters $(1, \frac{\lambda}{2})$ [10],

$$p(\gamma_i | \lambda) = \frac{\lambda}{2} \exp\left(-\frac{\lambda \gamma_i}{2}\right).$$  \hspace{1cm} (10)

In the third stage of the hierarchy, the prior distribution on $\lambda$ is

$$p(\lambda | \nu) = \Gamma(\lambda | \nu, \frac{\nu}{2}).$$  \hspace{1cm} (11)

Based on the above statements, the required Laplace distribution for the coefficients $\omega$ can be obtained by the following

$$p(\omega | \gamma, \lambda) = \int p(\omega | \gamma)p(\gamma | \lambda)d\gamma$$

$$= \prod_{i=1}^{N} p(\omega_i | \gamma_i)p(\gamma_i | \lambda)d\gamma_i$$

$$= \frac{\lambda^{N/2}}{2^N} \exp\left(-\sqrt{\lambda} \sum_{i=1}^{N} |\omega_i|\right).$$  \hspace{1cm} (12)

Combining the above three stages of the hierarchical Bayesian model, the joint distribution can be obtained by

$$p(\omega, \gamma, \lambda, \beta) = p(\gamma | \omega, \beta)p(\beta | \omega)p(\omega | \gamma)p(\gamma | \lambda)p(\lambda),$$  \hspace{1cm} (13)

where $p(\gamma | \omega, \beta)$, $p(\beta | \omega)$, $p(\omega | \gamma)$, $p(\gamma | \lambda)$ and $p(\lambda)$ are defined in (4), (34), (9), (10) and (11) respectively.

2.2. Sparse Bayesian inference

Based on the above described prior distribution for the hyperparameters, Bayesian inference can be performed to compute the posterior distribution over unknown parameters such as $\omega$, $\beta$ and $\lambda$, and the derived hyperparameters. The inference procedure is the following form

$$p(\omega, \gamma, \lambda, \beta | y) = p(\omega | y, \gamma, \beta, \lambda)p(\gamma, \beta, \lambda | y).$$  \hspace{1cm} (14)

According to Bayesian inference, the posterior distribution of weight coefficients $\omega$ obey multivariate Gaussian distribution, that is

$$p(\omega | y, \gamma, \beta, \lambda) = N(\omega | \mu, \Sigma) \propto p(\omega | y, \gamma, \beta, \lambda)p(\gamma, \beta, \lambda | y),$$  \hspace{1cm} (15)

where the parameters

$$\mu = \Sigma \Phi^T y,$$

$$\Sigma = (\Phi \Phi^T + \Lambda)^{-1},$$  \hspace{1cm} (16) \hspace{1cm} (17)

and $\Lambda = \text{diag}(\gamma_i), i = 1, 2, \ldots, N$.

We use type-II maximum marginal likelihood method to estimate the hyperparameters by maximizing the joint distribution $p(y, \gamma, \beta, \lambda)$, which is obtained by the following

$$L(y, \lambda, \beta, \lambda) = \log \int p(y | \omega, \beta)p(\omega | \gamma)p(\gamma | \lambda)p(\beta)p(\lambda)d\omega$$

$$= -\frac{1}{2} \log |C| - \frac{1}{2} \gamma^T C^{-1} y + N \log \lambda - \frac{\lambda}{2} \sum_{i=1}^{N} \gamma_i^2$$

$$- \frac{\nu}{2} \lambda + \frac{\nu}{2} \log \Gamma\left(\frac{\nu}{2}\right) + \nu - 1 \log \lambda$$

$$+ (a_\beta - 1) \log \beta - b_\beta \beta,$$  \hspace{1cm} (18)

where $C = (\beta^{-1}I + \Phi \Lambda^{-1} \Phi^T)$.

To promote the sparsity of $\omega$ and decrease the cost of computation, we employ the method in [11] which updates a single $\gamma_i$ at each iteration instead of updating the whole vector $\gamma$. Accordingly, the matrix $C$ in (18) can be rewritten as

$$C = \beta^{-1}I + \sum_{j \neq i} \gamma_j \phi_j \phi_j^T + \gamma_i \phi_i \phi_i^T$$

$$= C_{-i} + \gamma_i \phi_i \phi_i^T,$$  \hspace{1cm} (19)

where $C_{-i}$ denotes the matrix $C$ with the contribution of basis vector $i$ removed. The terms $|C|$ and $C^{-1}$ can be written as

$$C^{-1} = C_{-i}^{-1} - C_{-i}^{-1} \phi_i \phi_i^T C_{-i}^{-1}$$

$$\frac{1}{1/\gamma_i + \phi_i \phi_i^T C^{-1}_{-i}}.$$

$$|C| = C_{-i} |1 + \gamma_i \phi_i \phi_i^T C_{-i}^{-1} \phi_i|.$$  \hspace{1cm} (20) \hspace{1cm} (21)

To simplify the model in (18), we specify $\nu = 0$, $a_\beta = 1$ and $b_\beta = 0$. The prior distribution on $\lambda$ in the third stage of the hierarchy is accordingly written as

$$p(\lambda) \propto \frac{1}{\lambda}.$$  \hspace{1cm} (22)

Based on the above modifications and definitions, we substitute (20) and (21) into Eq. (18) and make it as a func-
When $0$, the quantities $q_i, s_i$ and $l(\gamma_i)$ are defined respectively as

$$s_i = \phi_i^T C_i^{-1} \phi_i, \quad q_i = \phi_i^T C_i^{-1} y,$$

and

$$l(\gamma_i) = \frac{1}{2} \left[ \log \frac{1}{1 + \gamma_i s_i} + \frac{q_i^2 \gamma_i}{1 + \gamma_i s_i} - \ln \gamma_i \right].$$

Therefore, the optimal solution of $L(\gamma)$ with respect to the hyperparameter $\gamma$ is transformed into that of $L(\gamma)$ with respect to $\gamma_i$. Due to $C_i^{-1}$ being independent of $\gamma_i$, the quantities $q_i$ and $s_i$ do not depend on $\gamma_i$ known from (24), and thus the single hyperparameter $\gamma_i$ in the terms related of (25) is separated from others, and this provides convenience for the next derivative with respect to $\gamma_i$.

When we hold other hyperparameters fixed and only make the $i$th component changed, the maximum solution of $L(\gamma)$ can be obtained by setting the derivative of $l(\gamma_i)$ to be zero. The analysis of $l(\gamma_i)$ in [11] shows that $L(\gamma)$ has a unique maximum with respect to $\gamma_i$. The derivative of $L(\gamma)$ with respect to $\gamma_i$ is expressed as

$$\frac{dL(\gamma)}{d\gamma_i} = \frac{dL(\gamma_i)}{d\gamma_i} = \frac{1}{2} \left[ - \frac{s_i}{1 + \gamma_i s_i} + \frac{q_i^2 \gamma_i}{1 + \gamma_i s_i} \right].$$

Let $\frac{dL(\gamma_i)}{d\gamma_i} = 0$, the maximum of the solution can be obtained. When $q_i^2 - s_i > \lambda$,

$$\gamma_i = -s_i(1 + 2\lambda) + s_i \sqrt{(s_i + 2\lambda)^2 - 4\lambda (s_i - q_i^2 + \lambda)},$$

when $q_i^2 - s_i \leq \lambda$, $\gamma_i = 0$.

According to [13], it is more efficiency to maintain and update the values of $s_i$ and $q_i$ with

$$S_i = \beta \phi_i^T \phi_i - \beta^2 \phi_i^T \Phi \Sigma \Phi^T \phi_i,$$

$$Q_i = \beta \phi_i^T y - \beta^2 \phi_i^T \Phi \Sigma \Phi^T y,$$

$$s_i = \frac{S_i}{1 - \gamma_i s_i}, q_i = \frac{Q_i}{1 - \gamma_i s_i},$$

where $\Phi$ and $\Sigma$ are only those basis functions that are currently included in the model ($\gamma_i \neq 0$). When $\gamma_i = 0$, the corresponding basis function $\phi_i$ is removed from the current model and $\mu_i$ is set to be zero. When $\gamma_i \neq 0$, the variables $s_i, q_i, \mu_i$ and $\Sigma$ can be updated at each iteration. The estimates of other hyperparameters, such as $\lambda$ and $\beta$, can be obtained by making the derivative of (18) with respect to each hyperparameter to be zero, and thus

$$\lambda = \frac{N - 1}{2 \sum_{i=1}^{N} \gamma_i},$$

$$\beta = \frac{N/2 + 1}{\frac{1}{2} \langle y - \Phi \omega \rangle^2},$$

where the employed values is the current distribution of $\omega$. According to [11], the estimate of the noise precision $\beta$ in (31) is intractable in practical application, thus we employ the method in [11] and also fix it as $\beta = 0.001$, and this makes the convergence is accelerated. The estimate of posterior distribution for coefficients $\omega$ can be implemented by (16) and (17).

3. Experiments and discussions

To demonstrate the performance of the proposed method, a 2D image experiment is presented in this section. Without loss of generality, we use a 32×32 small phantom image to be reconstructed in our experiment. The experiments are performed in a Intel (R) core (TM) i7 CPU, 4.0 GHz, 32 bit operating system Windows 7 PC. Due to the medical image is compressible, we transform the image $X$ to the wavelet basis $\Psi$ and obtain the sparse coefficients $\omega$. Then we measure the sparse wavelet coefficients with a random Gaussian measurement matrix $\Phi_{MxN}$, where $M = 600$, $N = 1024$. The reconstructed
To demonstrate the advantage of Bayesian CS over other algorithms such as BP and OMP, we take the experiments to compare the relative reconstruction error and the number of measurements between the presented algorithm, BP and OMP. The compared BP source code was obtained from http://www.acm.caltech.edu/l1magic/ and compared OMP algorithm source code was obtained from http://www.eee.hku.hk/wsha/freecode/freecode.htm. The experiment is implemented with an adaptive kernel learning, and the measurements are tested from $M = 500$ to $M = 680$. The reconstruction error is calculated as $\|Y - X_0\|_2 / \|X_0\|_2$, where $X$ is the reconstructed image and $X_0$ is the original image. The experimental results are shown in Figure 3.

Figure 1 reveals that the presented method outperforms the other two algorithms. The reconstruction error is the lowest among the three methods, the reconstruction error for OMP method is the highest and BP is in-between. Therefore, a conclusion is drawn that under the same reconstruction error condition, BCS method needs the littlest measurement number, OMP method needs the most measurement number, and BP method is in-between.

Under the same number of measurements, here $M = 600$ in our experiments, the reconstruction time for BCS is 0.0738s, for BP 0.6789s and for OMP 0.1724s. It suggests that the reconstruction time of BCS method is the same as BP, but shorter than that of OMP. A detailed comparison and summary is reported in Table 1. Due to the random of the measurement matrix, the provided data is based on the average statistics of 20 times measurements.

Table 1. Further suggests that the proposed method performs well with more little reconstruction error, more little reconstruction time and little measurement number compared with BP and OMP under the same test conditions.
In summary, experimental results of simulation with phantom image reconstruction suggests that the presented method provides improved performance in reconstruction quality compared with the traditional greedy algorithms and linear programming algorithms such as OMP and BP.

<table>
<thead>
<tr>
<th>Phantom</th>
<th>Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td># Nonzeros</td>
<td>BCS</td>
</tr>
<tr>
<td>Time (s)</td>
<td>0.0738</td>
</tr>
<tr>
<td>Error</td>
<td>0.057</td>
</tr>
</tbody>
</table>

4. Conclusions

A medical image reconstruction model is presented based on compressed sensing theory from Bayesian perspective, and the reconstruction results and reconstruction error are demonstrated in this paper. A hierarchical form of Laplace prior is used to estimate the unknown sparse coefficients of the image rejected to wavelet field, and the unknown image can be reconstructed well with very little samples. To test the implementation of this proposed method and the advantage compared with other CS methods, the experiments are designed to compare the reconstruction error, the number of measurements and the running time among the proposed method, BP and OMP, and corresponding conclusions are drawn as follows:

- Under the same reconstruction error conditions, BCS needs the littlest measurement number, OMP needs the most measurement number and BP is in-between.
- Under the same measurement number conditions, the running time for BCS and OMP is the same, and the running time for BP is the longest;
- Under the same measurement number conditions, the reconstruction error is the lowest, BP is the in-between, and OMP is the highest.

References