

Adaptive Algorithm for Sparse Signal Recovery

Fekadu L. Bayisa*, Zhiyong Zhou, Ottmar Cronie, and Jun Yu

Department of Mathematics and Mathematical Statistics, Umeå University, Umeå, Sweden

Abstract

The development of compressive sensing in recent years has given much attention to sparse signal recovery. In sparse signal recovery, spike and slab priors are playing a key role in inducing sparsity. The use of such priors, however, results in non-convex and mixed integer programming problems. Most of the existing algorithms to solve non-convex and mixed integer programming problems involve either simplifying assumptions, relaxations or high computational expenses. In this paper, we propose a new adaptive alternating direction method of multipliers (AADMM) algorithm to directly solve the suggested non-convex and mixed integer programming problem. The algorithm is based on the one-to-one mapping property of the support and non-zero element of the signal. At each step of the algorithm, we update the support by either adding an index to it or removing an index from it and use the alternating direction method of multipliers to recover the signal corresponding to the updated support. Moreover, as opposed to the competing "adaptive sparsity matching pursuit" and "alternating direction method of multipliers" methods our algorithm can solve non-convex problems directly. Experiments on synthetic data and real-world images demonstrated that the proposed AADMM algorithm provides superior performance and is computationally cheaper than the recently developed iterative convex refinement (ICR) and adaptive matching pursuit (AMP) algorithms.

Keywords: Sparsity; adaptive algorithm; sparse signal recovery; spike and slab priors

1. Introduction

Over the past decades, sparse signal recovery from fewer samples has received attention due to increasing impracticality. Sparsity is usually assumed for signal reconstruction and inverse problems (Tropp and Gilbert, 2007). It is often assumed in compressed sensing theory (Donoho, 2006). The presence of sparsity has several applications in signal recovery (Tropp and Gilbert, 2007; Wright et al., 2009), medical image reconstruction (Chaari et al., 2011; Andersen et al., 2014), image classification (Mousavi et al., 2014; Srinivas et al., 2015), and dictionary learning (Sadeghi et al., 2013; Suo et al., 2014).

Lately, there has been a continuous increase in the volume of data generated. For instance, medical imaging systems such as magnetic resonance imaging (MRI) may deliver multidimensional signals. According to Lee et al. (2017), however, magnetic resonance (MR) images may appear blurred due to the long scanning times, which in turn can cause patient inconvenience and unwarranted movements. One of the strategies to accelerate MR scanning is image reconstruction from reduced samples. The key feature in MR image reconstruction is the use of prior information about the signal through the compressibility or sparse representation under an appropriate sparsifying transform such as the wavelet transform and finite-differencing (Lustig et al., 2008). Signal reconstruction by taking the sparsity into account is therefore of great interest.

In general, sparse signal recovery problems are ill-posed problems. However, the sparsity assumption makes sparse signal recovery possible from fewer measurements. Hence, regularizations are often required to promote the

*Corresponding author.

E-mail address: fekadu.bayisa@umu.se

sparsity of the unknown signal. Several algorithms have been proposed to solve these regularized problems. The most recent algorithms are adaptive matching pursuit (Vu et al., 2017) and iterative convex refinement (Mousavi et al., 2015). Greedy algorithms (Tropp and Gilbert, 2007; Mohimani et al., 2009; Mousavi et al., 2013), Bayesian methods (Ji et al., 2008; Dobigeon et al., 2009; Lu et al., 2013a), and general sparse approximation algorithms such as SpaRSA and alternating direction method of multipliers (Wright et al., 2009; Becker et al., 2011; Boyd et al., 2011) have also been exploited for sparse signal recovery.

In Bayesian sparse signal recovery, setting priors for the signal has played a key role in promoting sparsity and improving performance. Examples of priors are spike and slab (Mitchell and Beauchamp, 1988; Lu et al., 2013b; Andersen et al., 2014; Mousavi et al., 2014, 2015; Vu et al., 2017), Bernoulli-Gaussian (Lavielle, 2009), Bernoulli-exponential (Dobigeon et al., 2009), Laplacian (Babacan et al., 2010), and generalized Pareto (Cevher et al., 2010).

Chaari et al. (2013) explored a fully Bayesian sparse signal recovery using Bernoulli-Laplacian priors and obtained sparser solutions in comparison to the Bernoulli-Gaussian prior setting. The increased sparsity is due to the Laplacian term in the model. Motivated by this, we concentrated on the setup of Yen (2011) to promote sparsity by setting spike and slab priors for the signal. We proposed a Bernoulli-Laplace prior for the signal in order to induce sparsity in the signal recovery. Using the proposed prior, we develop a model that is an extension of least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996). The developed model is a more general model and its optimization is known to be a non-convex and mixed integer programming problem where the existing solving methods involve simplifying or relaxation assumptions (Andersen et al., 2014; Yen, 2011; Srinivas et al., 2015; Mousavi et al., 2015). In this work, however, we developed an adaptive algorithm to solve the optimization problem directly in its general form.

To the best of our knowledge, the competitors of our algorithm are adaptive matching pursuit (AMP) (Vu et al., 2017) and iterative convex refinement (ICR) (Mousavi et al., 2015). Our algorithm and AMP are similar in that they involve two stages to solve the optimization problems. The first stage deals with the support of the signal while the second stage is concerned with signal reconstruction for a given support of the signal. We perform these two stages iteratively to obtain the reconstructed signal and an updated support of the signal. These two stages hold for AMP and AADMM. The main difference is the involvement of l_1 -norm in the second stage. In AMP, the second stage does not involve l_1 -norm and one can reconstruct the signal by forward and backward substitution. While AADMM involves l_1 -norm in the second stage with a cost of that the problem can not be solved by forward and backward substitution. However, it is advantageous for a reconstruction problem to involve an l_1 -norm as it enforces sparsity during sparse signal recovery. In contrast to AADMM, ICR uses a relaxation assumption to solve the non-convex and mixed integer programming problem.

The main contributions of our work are the following: 1) We formulate a sparse model using the maximum a posteriori estimation technique. 2) We propose an adaptive alternating direction method of multipliers, hereinafter AADMM, to solve the non-convex and mixed integer programming problem directly. A matching pursuit procedure and an alternating direction method of multipliers are combined to develop a computationally efficient algorithm to solve the optimization problem. Wu and Wang (2012) developed an adaptive sparsity matching pursuit algorithm, which used two nested stages to solve convex problems, for sparse signal reconstruction. Boyd et al. (2011) also devised an alternating direction method of multipliers to solve convex optimization problems. Our algorithm can solve non-convex problems directly, which is a clear added-value that the above-mentioned two algorithms don't have. 3) We compare the performance of the proposed algorithm with the recently developed algorithms. For a given support of the signal, the proposed optimization problem involves an l_1 -norm. Consequently, the proposed optimization problem is not differentiable. As a result, we can not exploit Cholesky decomposition based forward-backward substitution to solve the optimization problem (Boyd et al., 2011; Abur, 1988). In the most recent algorithm AMP, the optimization problem is differentiable for a given support of the signal and its differentiation is available in a simple closed-form to exploit Cholesky decomposition based forward-backward substitution for solving the resulting closed form problem (Vu et al., 2017). Therefore, the most recent algorithm

AMP can not be used to solve our proposed optimization problem. In this regard, the proposed algorithm is advantageous. Besides, the proposed optimization problem involves an l_1 -norm and therefore we expect the signal to be recovered in a sparser form. Here we compare the performance of the proposed algorithm with the recently developed algorithms AMP and ICR. Notice that ICR can be used to solve the proposed optimization while AMP can not be exploited to solve the optimization problem. As signal reconstruction is concerned, however, we can still use AMP to reconstruct the signal generated according to the problem setting in this article in order to compare its performance with AADMM. 4) The developed adaptive algorithm can also be used to reconstruct both unconstrained and constrained (or non-negative) signals. 5) We test our algorithm on both simulated data and real images. The results reveal the merits of the proposed AADMM algorithm.

The paper is organized as follows. In Section 2, we demonstrate the details of the proposed model. The developed adaptive algorithm for sparse signal recovery, evaluation method of the signal recovery, and the results obtained are reported in Section 3, Section 4, and Section 5, respectively. Finally, the conclusions and the future works are presented in Section 6.

2. Problem Formulation

Sparse signal recovery algorithms are used to recover a sparse signal $\mathbf{x} \in \mathbb{R}^{n \times 1}$ from observed measurements $\mathbf{y} \in \mathbb{R}^{m \times 1}$, where $m \ll n$. The basic model for sparse signal recovery is given by

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\epsilon}, \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a measurement matrix, and $\boldsymbol{\epsilon} \in \mathbb{R}^{m \times 1}$ is a Gaussian noise with a variance-covariance structure given by $\sigma^2 \mathbf{I}$. Here \mathbf{I} is an $m \times m$ identity matrix. Since $m \ll n$, the inverse problem in equation (1), which occurs in a number of applications in the field of signal and image processing (Chaari et al., 2011; Pustelnik et al., 2010) is an ill-posed problem.

To allow sparse modeling and regularization, we assume a prior distribution for the unknown signal $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$. Using a Bernoulli random variable ω_i , assume that

$$x_i \mid \omega_i = 1 \sim P_1(x_i), \quad x_i \mid \omega_i = 0 \sim P_0(x_i) \quad \text{and} \quad P_0(0) = 1,$$

where $P_1(\cdot)$ and $P_0(\cdot)$ are probability distributions and $i = 1, 2, \dots, n$. Here ω_i is used to control the structural sparsity of the signal \mathbf{x} . It is also exploited to form a mixture of distributions given by

$$x_i \sim \omega_i P_1(x_i) + (1 - \omega_i) P_0(x_i). \quad (2)$$

The mixture of distributions in equation (2) is an approximation of a spike and slab prior, which was proposed by Mitchell and Beauchamp (1988). The distribution $P_1(x_i)$ can be thought of as a slab while $P_0(x_i)$, which is an approximation of the Dirac delta function centered at the event $x_i = 0$, can be considered as a spike. Since the signal \mathbf{x} is expected to be sparse, one can specify a Laplace distribution peaked at location parameter $\mu = 0$ as a slab. In Bayesian inference, spike and slab priors are the gold standard for inducing sparsity (Titsias and Lázaro-Gredilla, 2011). Our interest here is to explore sparse signal recovery using the following three-level models:

$$\begin{aligned} \mathbf{y} \mid \mathbf{A}, \mathbf{x}, \sigma^2 &\sim \mathcal{N}(\cdot \mid \mathbf{A}\mathbf{x}, \sigma^2 \mathbf{I}), \\ \mathbf{x} \mid \boldsymbol{\omega}, \sigma^2, \lambda &\sim \prod_{i=1}^n \left\{ \omega_i \text{Laplace} \left(\cdot \mid 0, \frac{2\sigma^2}{\lambda} \right) + (1 - \omega_i) P_0(x_i) \right\}, \\ \boldsymbol{\omega} \mid \boldsymbol{\kappa} &\sim \prod_{i=1}^n \text{Bernoulli}(\kappa_i), \end{aligned} \quad (3)$$

where

$$\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_n)', \boldsymbol{\kappa} = (\kappa_1, \kappa_2, \dots, \kappa_n)', \lambda > 0,$$

and the notations $\mathcal{N}(\cdot)$, **Laplace** (\cdot) and **Bernoulli** (\cdot) represent normal, Laplace and Bernoulli distributions, respectively.

A posterior maximization procedure can be used to induce sparsity in the Bayesian frame work (Cevher, 2009; Cevher et al., 2010; Mousavi et al., 2015; Vu et al., 2017).

Proposition 2.1. Assume that σ^2 is known. Using the model set-up in equation (3) and a posterior maximization procedure, one can obtain the regularized optimization problem:

$$\min_{\mathbf{x}, \boldsymbol{\omega}} \left\{ \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \lambda \|\mathbf{x}\|_1 + \sum_{i=1}^n \omega_i \gamma_i \right\}, \quad (4)$$

where

$$\gamma_i = 2\sigma^2 \log \frac{4\sigma^2 (1 - \kappa_i)}{\lambda \kappa_i}. \quad (5)$$

Proof. Let $\Omega = \{\sigma^2, \lambda, \boldsymbol{\kappa}\}$. Using the hierarchical model, the joint posterior density of \mathbf{x} and $\boldsymbol{\omega}$ is given by

$$\begin{aligned} \log f(\mathbf{x}, \boldsymbol{\omega} | \mathbf{y}, \mathbf{A}, \Omega) &\propto \log \left\{ g(\mathbf{y} | \mathbf{A}, \mathbf{x}, \sigma^2) h(\mathbf{x} | \boldsymbol{\omega}, \sigma^2, \lambda) P(\boldsymbol{\omega} | \boldsymbol{\kappa}) \right\}, \\ &= \log \left\{ \mathcal{N}(\cdot | \mathbf{Ax}, \sigma^2 \mathbf{I}) \left\{ \prod_{i=1}^n \text{Laplace} \left(\cdot | 0, \frac{2\sigma^2}{\lambda} \right)^{\omega_i} \times \right. \right. \\ &\quad \left. \left. P_0(x_i)^{(1-\omega_i)} \right\} \prod_{i=1}^n \text{Bernoulli}(\kappa_i) \right\}, \\ &= -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \\ &\quad \sum_{i=1}^n \left\{ -\omega_i \log \frac{4\sigma^2}{\lambda} - \omega_i \log (1 - \kappa_i) + \omega_i \log \kappa_i \right\} + \\ &\quad \sum_{i=1}^n \log (1 - \kappa_i) - \frac{\lambda}{2\sigma^2} \sum_{i=1}^n \omega_i |x_i|. \end{aligned}$$

Assume that σ^2 is known. Then, we have the following optimization problem:

$$\begin{aligned} \max_{\mathbf{x}, \boldsymbol{\omega}} \left\{ 2\sigma^2 \log f(\mathbf{x}, \boldsymbol{\omega} | \mathbf{y}, \mathbf{A}, \Omega) \right\} &= \min_{\mathbf{x}, \boldsymbol{\omega}} \left\{ -2\sigma^2 \log f(\mathbf{x}, \boldsymbol{\omega} | \mathbf{y}, \mathbf{A}, \Omega) \right\}, \\ &= \min_{\mathbf{x}, \boldsymbol{\omega}} \left\{ n\sigma^2 \log 2\pi + n\sigma^2 \log \sigma^2 + \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \right. \\ &\quad \lambda \sum_{i=1}^n \omega_i |x_i| - 2\sigma^2 \sum_{i=1}^n \log (1 - \kappa_i) - 2\sigma^2 \times \\ &\quad \left. \sum_{i=1}^n \left\{ \omega_i \log \frac{4\sigma^2}{\lambda} + \omega_i \log (1 - \kappa_i) - \omega_i \log \kappa_i \right\} \right\}, \\ &= \min_{\mathbf{x}, \boldsymbol{\omega}} \left\{ \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \lambda \|\mathbf{x}\|_1 + \sum_{i=1}^n \omega_i \gamma_i \right\}. \end{aligned}$$

□

Remark 2.1. If all ω_i are identically distributed, then the last term of the optimization problem in equation (4)

becomes a regularized l_0 -norm of the signal. From equation (5), γ_i can be negative for large κ_i and increasing κ_i decreases γ_i . This means that a strong belief in the presence of a non-zero value of the signal decreases the penalty value for the signal.

Remark 2.2. We have proposed a more general optimization problem in comparison to the optimization problem suggested by Yen (2011), Lu et al. (2013b), Lu et al. (2013a), and Srinivas et al. (2015) since they simplified the optimization by assuming the same parameter κ for all the Bernoulli random variables. This assumption allowed them to obtain an optimization problem that involves a regularized l_0 -norm. In addition to the assumption of having the same parameter for all the Bernoulli random variables, Yen (2011) exploited the l_2 -norm of the signal instead of the l_1 -norm in equation (4). The same problem setting as in equation (4) has been utilised by Mousavi et al. (2015) and Vu et al. (2017) except that they exploited the l_2 -norm instead of the l_1 -norm. This makes our approach advantageous as it contains an l_1 -norm, which can enforce sparse signal recovery (Malioutov et al., 2005; Jia et al., 2011). The problem in equation (4) promotes greater generality in capturing sparsity and it is also an extension of the existing LASSO method. Our optimization problem appears in signal recovery (Lu et al., 2013a,b; Chaari et al., 2014), regression (Tibshirani, 1996), image classification (Srinivas et al., 2015), and medical image reconstruction (Lustig et al., 2008). Using the conventional optimization algorithms, we may not be able to solve the proposed optimization problem because it is a non-convex mixed-integer programming problem.

3. Adaptive Algorithm

In this section, we present the details of the methods used to solve the proposed optimization problem. Let \mathbf{a}_i denote the i^{th} column of the matrix \mathbf{A} . Each column of the matrix is assumed to have norm 1. That is, $\|\mathbf{a}_i\|_2^2 = 1$, $i = 1, 2, \dots, n$. If we know the support $S = \{i \mid x_i \neq 0, i = 1, 2, \dots, n\}$ of the signal \mathbf{x} , then the optimization problem in equation (4) is equivalent to:

$$\min_{\mathbf{x}_S} \left\{ \|\mathbf{y} - \mathbf{A}^S \mathbf{x}_S\|_2^2 + \lambda \|\mathbf{x}_S\|_1 + \sum_{i \in S} \gamma_i \right\}, \quad (6)$$

where

$$\mathbf{A}^S = [\mathbf{a}_i : i \in S], \quad \mathbf{x}_S = (x_i : i \in S)^T.$$

Since there is a one-to-one correspondence between S and \mathbf{x}_S , we may solve the optimization problem in equation (4) by finding the support S and solving the optimization problem in equation (6). Based on this idea, an adaptive alternating direction method of multipliers (AADMM) is proposed for the signal recovery. AADMM uses a greedy method to update the support S . For a given support, it uses alternating direction method of multipliers (ADMM) to solve the optimization problem in equation (6). For each iteration of the algorithm, the greedy method updates the support S either by absorbing one of the unselected indices into S or by removing one of the elements from S . We select the option, that is either absorbing or removing an index, that decreases the objective function of the optimization problem in equation (4).

Based on the support S of the signal, consider the following definitions:

$$\delta_S = \sum_{i \in S} \gamma_i \text{ and } g(S) = \min_{\mathbf{x}_S} \left\{ \|\mathbf{y} - \mathbf{A}^S \mathbf{x}_S\|_2^2 + \lambda \|\mathbf{x}_S\|_1 + \delta_S \right\}. \quad (7)$$

For each iteration in the AADMM algorithm, we need to compute the following to update the support:

$$U_S = \min_{i \notin S} \{g(S \cup \{i\}) - g(S)\}, \quad (8)$$

$$\begin{aligned} i^* &= \arg \min_{i \notin S} \{g(S \cup \{i\}) - g(S)\}, \\ V_S &= \min_{j \in S} \{g(S \setminus \{j\}) - g(S)\}, \\ j^* &= \arg \min_{j \in S} \{g(S \setminus \{j\}) - g(S)\}. \end{aligned} \quad (9)$$

Equations (8) and (9) represent the minimization of the change in the cost function by selecting one of unselected indices and by removing one of the already selected indices, respectively. Based on U_S and V_S , we have three possible cases. The first case is that if both U_S and V_S are not less than zero, adding or removing the indices can not decrease the cost function and we can stop the algorithm. If $U_S < V_S$, then we update S by absorbing the index i^* and if $V_S < U_S$, then we update S by removing the index j^* . After each iteration, the procedure guarantees that the cost function decreases and thereby the algorithm converges eventually, being a monotone limit. The optimal support \hat{S} is an element of the set $\{S : U_S \geq 0, \quad V_S \geq 0\}$. However, it is hardly practical to optimize $g(S \cup \{i\})$ and $g(S \setminus \{j\})$ for each i and j . Therefore, we would like to use the upper bounds of U_S and V_S to significantly reduce the computational cost of $g(S \cup \{i\})$ and $g(S \setminus \{j\})$. For each iteration in the AADMM algorithm, the decision to add an index to S or remove an index from S , that is to obtain an updated support, is based on the upper bounds of U_S and V_S . Let S^u denote the updated support for a given iteration in the AADMM algorithm. Using the updated support S^u , we estimate \mathbf{x}_{S^u} and compute the residual $\mathbf{r}_{S^u} = \mathbf{y} - \mathbf{A}^{S^u} \mathbf{x}_{S^u}$. This continues until convergence of the algorithm has been obtained. The following results are the key tools for developing the algorithm. Proposition 3.1 is used to initialise the support while propositions 3.2 and 3.3 are utilised for approximating the computationally expensive U_S and V_S . Once the initial value of the support is obtained, we estimate the signal using ADMM (Boyd et al., 2011).

Proposition 3.1. *If $\gamma_i < 0$, then $i \in \hat{S}$.*

Proof. Assume that $i \notin \hat{S}$. Using equation (7), we have

$$g(\hat{S} \cup \{i\}) \leq \|\mathbf{r}_{\hat{S}} - x_i \mathbf{a}_i\|_2^2 + \lambda \|\mathbf{x}_{\hat{S}}\|_1 + \delta_{\hat{S}} + \lambda |x_i| + \gamma_i = g(\hat{S}) + x_i^2 + \lambda |x_i| - 2x_i \mathbf{r}_{\hat{S}}^T \mathbf{a}_i + \gamma_i,$$

which implies that

$$g(\hat{S} \cup \{i\}) - g(\hat{S}) \leq x_i^2 + \lambda |x_i| - 2x_i \mathbf{r}_{\hat{S}}^T \mathbf{a}_i + \gamma_i.$$

Let $h(x) = x^2 + \lambda |x| - 2x \mathbf{r}_{\hat{S}}^T \mathbf{a}_i + \gamma_i$. We see that $h(x)$ is continuous, $\lim_{x \rightarrow \infty} h(x) = \infty$ and $h(0) = \gamma_i < 0$. We observe that there exists a value \bar{x} of x_i such that $\gamma_i < h(\bar{x}) < 0$, whereby we have that $g(\hat{S} \cup \{i\}) - g(\hat{S}) < h(\bar{x}) < 0$. This implies that \hat{S} can not be the optimal solution, which is a contradiction to the assumption. Therefore, i must be in \hat{S} . \square

Proposition 3.2. U_S given in equation (8) satisfies

$$U_S \leq \bar{U}_S = \min_{i \notin S} \begin{cases} \gamma_i - (\mathbf{r}_S^T \mathbf{a}_i)^2 + \lambda \mathbf{r}_S^T \mathbf{a}_i - \frac{\lambda^2}{4}, & \text{if } \mathbf{r}_S^T \mathbf{a}_i > \frac{\lambda}{2}, \\ \gamma_i, & \text{if } |\mathbf{r}_S^T \mathbf{a}_i| \leq \frac{\lambda}{2}, \\ \gamma_i - (\mathbf{r}_S^T \mathbf{a}_i)^2 - \lambda \mathbf{r}_S^T \mathbf{a}_i - \frac{\lambda^2}{4}, & \text{if } \mathbf{r}_S^T \mathbf{a}_i < -\frac{\lambda}{2}, \end{cases} \quad (10)$$

Proof. Since

$$g(S \cup \{i\}) - g(S) \leq x_i^2 + \lambda |x_i| - 2x_i \mathbf{r}_S^T \mathbf{a}_i + \gamma_i,$$

for every x_i , we have that

$$g(S \cup \{i\}) - g(S) \leq \min_{x_i} \{x_i^2 + \lambda|x_i| - 2x_i \mathbf{r}_S^T \mathbf{a}_i + \gamma_i\}.$$

This implies that

$$\min_{i \notin S} \{g(S \cup \{i\}) - g(S)\} \leq \min_{i \notin S} \left\{ \min_{x_i} \{x_i^2 + \lambda|x_i| - 2x_i \mathbf{r}_S^T \mathbf{a}_i + \gamma_i\} \right\}. \quad (11)$$

We obtain the result in equation (10) from equation (11). \square

Proposition 3.3. For a support S , V_S given in equation (9) satisfies

$$V_S \leq \bar{V}_S = \min_{j \in S} \{x_j^2 - \lambda|x_j| + 2x_j \mathbf{r}_S^T \mathbf{a}_j - \gamma_j\}, \quad (12)$$

where $j \in S$ and x_j is the element of the vector \mathbf{x}_S .

Proof. For any $j \in S$,

$$\begin{aligned} g(S) &= \|\mathbf{r}_S\|_2^2 + \lambda\|\mathbf{x}_S\|_1 + \delta_S, \\ &= \|\mathbf{r}_S - x_j \mathbf{a}_j + x_j \mathbf{a}_j\|_2^2 + \lambda\|\mathbf{x}_{S \setminus \{j\}}\|_1 + \delta_{S \setminus \{j\}} + \lambda|x_j| + \gamma_j, \\ &= \|\mathbf{r}_{S \setminus \{j\}}\|_2^2 + \lambda\|\mathbf{x}_{S \setminus \{j\}}\|_1 + \delta_{S \setminus \{j\}} - x_j^2 - 2x_j \mathbf{a}_j^T \mathbf{r}_S + \lambda|x_j| + \gamma_j, \\ &\geq g(S \setminus \{j\}) - x_j^2 - 2x_j \mathbf{a}_j^T \mathbf{r}_S + \lambda|x_j| + \gamma_j. \end{aligned} \quad (13)$$

Equation (13) implies that

$$g(S \setminus \{j\}) - g(S) \leq x_j^2 + 2x_j \mathbf{a}_j^T \mathbf{r}_S - \lambda|x_j| - \gamma_j,$$

which holds for every $j \in S$ and we proved that equation (12) holds. \square

We exploit the computationally cheap \bar{U}_S and \bar{V}_S to update the support and we use the updated support to update the estimation of the signal. This procedure continues until we obtain the optimal support and signal. We have summarised the estimation procedure in Algorithm 1.

The optimization problem in equation (4) with non-negative constraint on the sparse signal can also be solved by AADMM. For the non-negative constraint, we need to modify steps 4 and 6 of Algorithm 1. For step 4, we need to modify the ADMM algorithm provided by [Boyd et al. \(2011\)](#). That is, we replace the soft-thresholding operator $S_\beta(\Lambda)$ in the ADMM algorithm by

$$\max\{0, S_\beta(\Lambda)\},$$

where the soft-thresholding operator is defined by

$$S_\beta(\Lambda) = \begin{cases} \Lambda - \beta, & \Lambda > \beta, \\ 0, & |\Lambda| \leq \beta, \\ \Lambda + \beta, & \Lambda < -\beta. \end{cases}$$

Furthermore, we need to substitute \bar{U}_S of step 6 (that is equation (14)) by

$$\bar{U}_S = \min_{i \notin S} \left\{ \left(\max\left\{0, \mathbf{r}_S^T \mathbf{a}_i - \frac{\lambda}{2}\right\} \right)^2 + \lambda \max\left\{0, \mathbf{r}_S^T \mathbf{a}_i - \frac{\lambda}{2}\right\} - 2 \max\left\{0, \mathbf{r}_S^T \mathbf{a}_i - \frac{\lambda}{2}\right\} \mathbf{r}_S^T \mathbf{a}_i + \gamma_i \right\}.$$

Algorithm 1. AADMM Algorithm

- 1: Inputs: $\mathbf{y}, \mathbf{A}, \lambda, \boldsymbol{\gamma}$,
- 2: Initialise the support: $S = \{i \mid \gamma_i < 0\}$,
- 3: **while** true **do**
- 4: Solve \mathbf{x}_S : **ADMM**,
- 5: Update the residual: $\mathbf{r}_S = \mathbf{y} - \mathbf{A}^S \mathbf{x}_S$,
- 6: Compute $[\bar{U}_S, i^*]$, where

$$\bar{U}_S = \min_{i \notin S} \begin{cases} \gamma_i - (\mathbf{r}_S^T \mathbf{a}_i)^2 + \lambda \mathbf{r}_S^T \mathbf{a}_i - \frac{\lambda^2}{4}, & \text{if } \mathbf{r}_S^T \mathbf{a}_i > \frac{\lambda}{2}, \\ \gamma_i, & \text{if } |\mathbf{r}_S^T \mathbf{a}_i| \leq \frac{\lambda}{2}, \\ \gamma_i - (\mathbf{r}_S^T \mathbf{a}_i)^2 - \lambda \mathbf{r}_S^T \mathbf{a}_i - \frac{\lambda^2}{4}, & \text{if } \mathbf{r}_S^T \mathbf{a}_i < -\frac{\lambda}{2}, \end{cases} \quad (14)$$

$$i^* = \arg \min_{i \notin S} \begin{cases} \gamma_i - (\mathbf{r}_S^T \mathbf{a}_i)^2 + \lambda \mathbf{r}_S^T \mathbf{a}_i - \frac{\lambda^2}{4}, & \text{if } \mathbf{r}_S^T \mathbf{a}_i > \frac{\lambda}{2}, \\ \gamma_i, & \text{if } |\mathbf{r}_S^T \mathbf{a}_i| \leq \frac{\lambda}{2}, \\ \gamma_i - (\mathbf{r}_S^T \mathbf{a}_i)^2 - \lambda \mathbf{r}_S^T \mathbf{a}_i - \frac{\lambda^2}{4}, & \text{if } \mathbf{r}_S^T \mathbf{a}_i < -\frac{\lambda}{2}, \end{cases}$$

- 7: Compute $[\bar{V}_S, j^*]$, where

$$\bar{V}_S = \min_{j \in S} \{x_j^2 - \lambda |x_j| + 2x_j \mathbf{r}_S^T \mathbf{a}_j - \gamma_j\},$$

$$j^* = \arg \min_{j \in S} \{x_j^2 - \lambda |x_j| + 2x_j \mathbf{r}_S^T \mathbf{a}_j - \gamma_j\},$$

- 8: Decide:
 - 9: **if** $\min\{\bar{U}_S, \bar{V}_S\} \geq 0$, **then** break the while loop,
 - 10: **else**
 - 11: **if** $\bar{U}_S < \bar{V}_S$, **then**
 - 12: Insert index: $S = S \cup \{i^*\}$,
 - 13: **else**
 - 14: Remove index: $S = S \setminus \{j^*\}$
 - 15: **end if**
 - 16: **end if**
 - 17: **end while**
 - 18: Outputs: $S \rightarrow \hat{\omega}$ and $\mathbf{x}_S \rightarrow \hat{\mathbf{x}}$.
-

The sparse signal recovery performance of the ICR algorithm (Mousavi et al., 2015) have been compared with the algorithms: majorization-minimization algorithm (Yen, 2011), FOCUSS (Gorodnitsky and Rao, 1997), expectation propagation approach for spike and slab recovery (Hernández-Lobato et al., 2015), and Variational Garrote (Kappen and Gómez, 2014). Since the ICR algorithm performs better than these algorithms (Mousavi et al., 2015) and as it can also be used to solve our proposed optimization problem, we choose to compare the sparse signal recovery performance of our algorithm with the recent ICR algorithm. Although AMP can not be used to solve our optimization problem, we can still apply AMP to the signal generated according to our model setup to compare its performance with the proposed AADMM as AMP outperforms ICR and other algorithms, see Vu et al. (2017).

4. Evaluation of the signal recovery

Mean squared error (MSE) is used to evaluate the closeness of the recovered signal to the ground truth. Besides, we utilise the support match level (SML) to measure how much the support of the recovered signal matches that of the support of the ground truth. We also study the effectiveness of the AADMM algorithm in comparison to the most recent ICR and AMP algorithms using MSE, SML, computational time (CT), objective function value

(OFV), and sparsity level (SL). Recall that we only compare our algorithm to the ICR and AMP algorithms since they have the best performance among the competing algorithms, see [Mousavi et al. \(2015\)](#) and [Vu et al. \(2017\)](#).

5. Results

In this section, we present numerical results of our sparse signal recovery algorithm. In comparison to the ICR and AMP algorithms, we also demonstrate the worthiness of the AADMM algorithm.

5.1. Simulation results for unconstrained signals

Our simulation setup for sparse signal recovery was as in [Yen \(2011\)](#); [Beck and Teboulle \(2009\)](#). We have used $\lambda = 2 \times 10^{-4}$, $\sigma^2 = 3.24 \times 10^{-4}$, and randomly generated a Laplacian sparse vector $\mathbf{x} \in \mathbb{R}^{512 \times 1}$ with 30 non-zeros. Using \mathbf{x} , an additive Gaussian noise with variance σ^2 and a randomly generated Gaussian matrix $\mathbf{A} \in \mathbb{R}^{128 \times 512}$, we obtained an observation vector $\mathbf{y} \in \mathbb{R}^{128 \times 1}$ according to equation (1). Using 500 different trials for \mathbf{A} , \mathbf{x} , and ϵ , the averages of the evaluation results are presented in Table 1.

Table 1. Comparison of AADMM, AMP, and ICR algorithms for sparse signal recovery using a true sparsity level = 30. We exploited OFV, MSE, SML, SL, and CT to compare the performance of the algorithms.

Method	OFV	MSE	SML(%)	SL	CT(S)
AADMM	0.058	1.44×10^{-4}	96.256	27.394	0.024
AMP	0.042	1.75×10^{-4}	95.395	32.176	0.122
ICR	0.154	1.06×10^{-3}	92.950	36.066	5.200

It can be seen from Table 1 that AADMM outperforms ICR in several aspects. In terms of OFV, we see that the signal recovery using AADMM achieved a lower OFV. The result in the table also revealed that the AADMM solution is closer to the ground truth in terms of MSE. Besides, AADMM is computationally faster than ICR. We also utilised SML to measure how much the support of the recovered signal matches that of the support of the ground truth. It is clear from the table that AADMM provides sparser solution and a higher SML (96.256%) than ICR.

In comparison to AMP, AADMM achieved a better result in recovering a signal closer to the ground truth. Moreover, AADMM is computationally faster and provided sparser solution. However, AMP achieved a lower OFV as compared to AADMM which may be due to AADMM's sparser signal recovery property.

Using different sparsity levels, we evaluated the algorithms and presented their corresponding evaluation results of the sparse signal recovery problem in Figure 1.

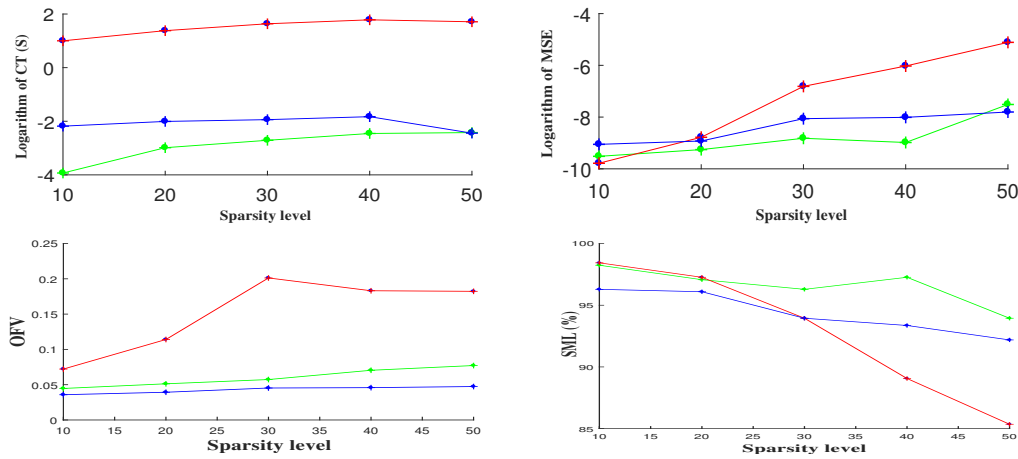


Figure 1. Evaluation of the sparse signal recovery for different sparsity levels using AADMM (green), AMP (blue) and ICR (red). CT, OFV, MSE, and SML are used to evaluate the performance of the algorithms over different sparsity levels. The computational times and MSEs are transformed by logarithm to make their plots visible.

The figure shows that our method consistently outperforms ICR and AMP in terms of SML and MSE. It should be noted that AMP achieved a lower OFV in comparison to AADMM, which may be due to the sparser signal recovery property of AADMM and the smoothing term in the optimization problem of AMP. It is worth emphasising that AMP and ICR are also computationally more costly than AADMM.

Figure 2 presents the assessment of the sparse signal recovery problem using different noise levels. We utilised MSE and SML to evaluate the performance of the algorithms for different noise levels.

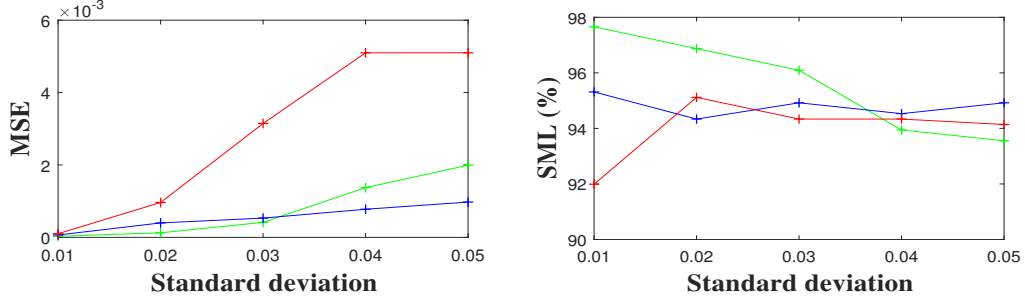


Figure 2. Evaluation of AADMM (green), AMP (blue), and ICR (red) using different noise levels. MSE and SML are used to evaluate the performance of the algorithms over different noise levels.

The figure shows that the recovered sparse signal using the AADMM algorithm is closer to the ground truth for lower noise levels. In terms of the SML, AADMM is more robust than AMP and ICR on most of the noise levels. However, the use of high noise levels has a negative effect on the sparse signal recovery.

Figure 3 demonstrates full histograms of the mean squared errors (MSEs) and standard errors (SEs) for the AADMM, AMP, and ICR algorithms.

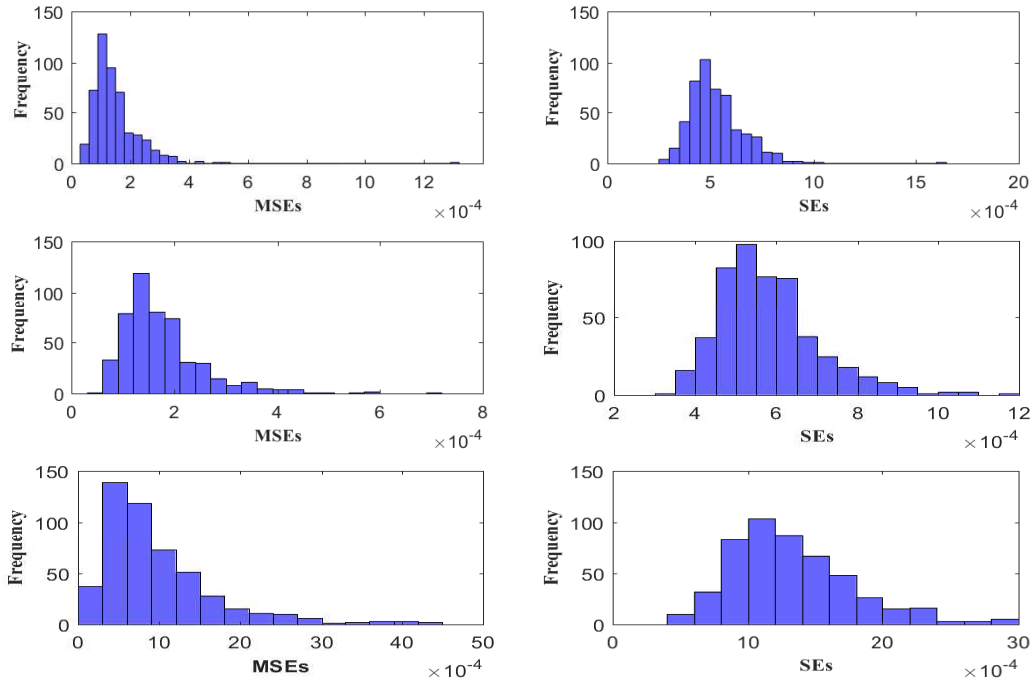


Figure 3. Full histograms of MSEs and SEs for the AADMM (top row), the AMP (middle row), and the ICR (bottom row) algorithms.

The full histograms reveal that AADMM has better sparse signal recovery performance than AMP and ICR; note the differences in scales on the x-axis and the relative locations of the histogram bins.

We examined the convergence of the AADMM algorithm numerically for both unconstrained and constrained (or non-negative) sparse signal recovery. Figure 4 shows the convergence of the AADMM algorithm.

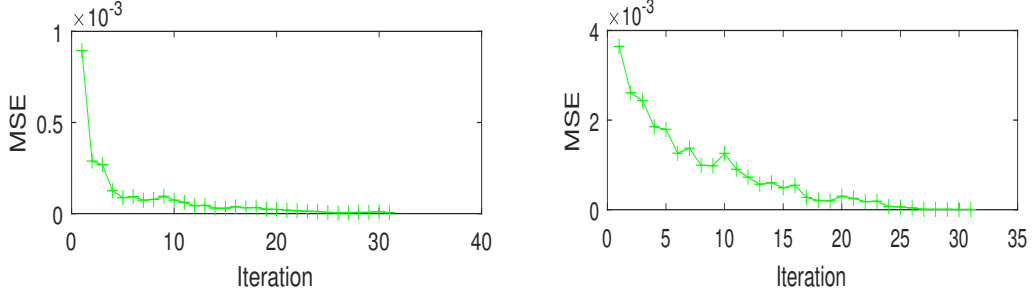


Figure 4. Convergence of AADMM algorithm for unconstrained (left) and constrained (right) signal. MSE against the number of iterations is utilised to assess the convergence of the algorithm.

In addition, we have evaluated the effect of λ on the sparse signal reconstruction performance of our algorithm, see Figure 5.

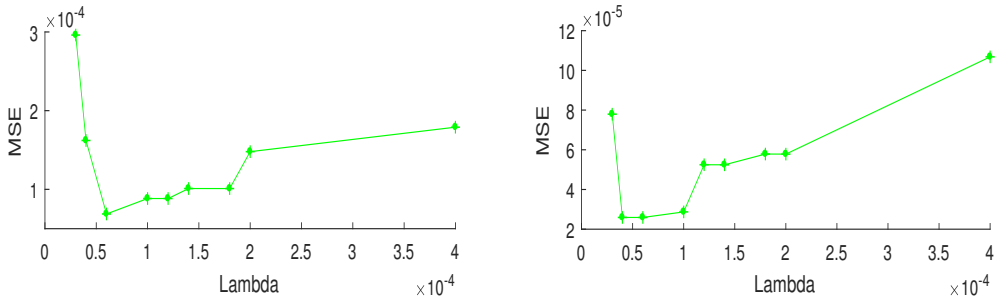


Figure 5. Effect of λ on the sparse signal reconstruction performance of AADMM for unconstrained (left) and constrained (right) signal.

As expected, the sparse signal reconstruction performance of AADMM depends on the regularizing parameter λ and the reconstruction error can be minimized with a properly chosen λ . Note that we have compared the performance of ICR, AMP, and AADMM under parameter settings similar to those in (Yen, 2011; Beck and Teboulle, 2009; Mousavi et al., 2015; Vu et al., 2017).

5.2. Real image Recovery

We applied our algorithm to a signal with non-negative constraints. The non-negative constraint assumption on the signal allows us to explicitly enforce a non-negative constraint during the reconstruction of the signal. We utilised the algorithms on the well-known handwritten digit images MNIST (LeCun et al., 2013). The digit images are real data, naturally sparse and fit into the spike and slab model. Each of the digit images (0 to 9) has a size of 28×28 . The maximum and the minimum sparsity levels of the digit images are 200 and 96, respectively, which shows that the images are sparse. We are interested to recover a sparse signal $\mathbf{x} \in \mathbb{R}^{784 \times 1}$ from undersampled random measurement \mathbf{y} . For a Gaussian random matrix \mathbf{A} and the maximum sparsity level 200, we can approximately determine the length of the random measurement \mathbf{y} required for the successful recovery of a sparse signal \mathbf{x} (Foucart and Rauhut, 2013). Based on this, we randomly generated a Gaussian random matrix $\mathbf{A} \in \mathbb{R}^{550 \times 784}$ and an additive Gaussian noise with variance $\sigma^2 = 3.24 \times 10^{-4}$ to obtain a random measurement \mathbf{y} according to equation (1). Using $\lambda = 2 \times 10^{-4}$, the results obtained for the signal recovery problem (the images of the digits) and the evaluation of the algorithms can be found in Figure 6 and Table 2, respectively.

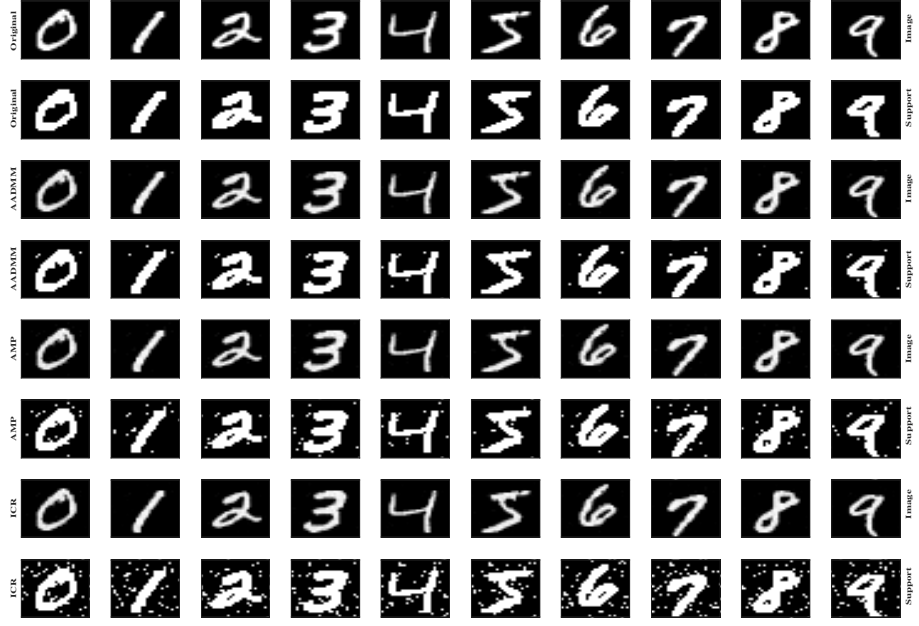


Figure 6. Sparse signal recovery of the real image using AADMM, AMP, and ICR. The first and the second rows show images of the original data and their supports. The third to the eighth rows present the recovered images and their supports by AADMM, AMP, and ICR, respectively.

Table 2. Evaluation of the algorithms for real image recovery. The averages of MSEs, SMLs, and CTs are employed to assess the performance of the algorithms for real image recovery.

Method	MSE	SML(%)	CT (S)
AADMM	1.32×10^{-4}	98.495	1.020
AMP	1.74×10^{-4}	97.360	1.228
ICR	1.46×10^{-4}	94.133	8.734

Figure 6 shows the images of the original data and their supports (the first and the second rows), images of the recovered signal and their supports by AADMM (the third and the fourth rows), images of the reconstructed signal and their supports by AMP (the fifth and the sixth rows), and images of the reconstructed signal and their supports by ICR (the last two rows). The images of the supports give better information about the performance of the sparse signal recovery algorithms. Based on the images of the supports, it is clear that AADMM has better performance than AMP and ICR. Table 2 presents the average of MSEs, SMLs, and CTs of the reconstruction problem. It can be seen that AADMM outperforms AMP and ICR and it is also important to note that AADMM is faster than AMP and, in particular, much faster than ICR.

6. Conclusions

In this paper, we have developed an algorithm (AADMM) to optimize a hard non-convex optimization problem and applied it in sparse signal recovery. Unlike the recent algorithm (ICR), AADMM does not simplify the optimization by considering a history of solutions at previous iterations. That is, AADMM solves the problem in its general form. The most recent AMP algorithm can not be used to directly solve this problem, which means that AADMM is a more general problem solving algorithm. Our evaluation of the algorithm on simulated data and real-world image data shows that AADMM has superior practical merit on ICR. Further, AADMM has better performance than AMP, in particular with respect to obtaining sparser signal recovery.

Regarding future work, one part is to adopt the proposed AADMM algorithm into the compressive sensing MR image reconstruction framework, where the MR images are not sparse in themselves but sparse under a

specific transformation. The key issue in MR image reconstruction is to obtain sparser recovery of MR images, which is the main reason for designing our algorithm. A further reason is to generalize the spike and slab prior to incorporate structural sparsity for sparse signal recovery and to develop its corresponding sparse signal recovery algorithm.

Acknowledgments

This work is supported by the Swedish Research Council grant (Reg. No. 340-2013-5342). We would like to thank three anonymous referees and the Editor for their detailed and insightful comments and suggestions that help to improve the quality of this paper.

Disclosure of Conflicts of Interest

The authors have no relevant conflicts of interest to disclose.

References

References

- Abur, A. (1988). A parallel scheme for the forward/backward substitutions in solving sparse linear equations. *IEEE transactions on power systems*, 3(4):1471–1478.
- Andersen, M. R., Winther, O., and Hansen, L. K. (2014). Bayesian inference for structured spike and slab priors. *Adv. Neural Inf. Process. Syst.*, pages 1745–1753.
- Babacan, S., Molina, R., and Katsaggelos, A. (2010). Bayesian compressive sensing using laplace priors. *IEEE Trans. Image Process.*, 19(1):53–63.
- Beck, A. and Teboulle, M. (2009). A fast iterative shrinkagethresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202.
- Becker, S., J. Bobin, J., and Candé, E. J. (2011). NESTA: A fast and accurate first-order method for sparse recovery. *SIAM J. Imag. Sci.*, 4(1):1–39.
- Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122.
- Cevher, V. (2009). Learning with compressible priors. in *Advances in Neural Information Processing Systems*, pages 261–269.
- Cevher, V., Indyk, P., Carin, L., and Baraniuk, R. G. (2010). Sparse signal recovery and acquisition with graphical models. *Signal Processing Magazine, IEEE*, 27(6):92–103.
- Chaari, L., Batatia, H., Dobigeon, N., and Tournieret, J.-Y. (2014). A hierarchical sparsity-smoothness bayesian model for $l_0 + l_1 + l_2$ regularization. in *Proc. IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, 24(6):1901–1905.
- Chaari, L., Pesquet, J.-C., Benazza-Benyahia, A., and Ciuciu, P. (2011). A wavelet-based regularized reconstruction algorithm for SENSE parallel MRI with applications to neuroimaging. *Med. Image Anal.*, 15(2):185–201.
- Chaari, L., Tournieret, J.-Y., and Batatia, H. (2013). Sparse bayesian regularization using bernoulli-laplacian priors. in *Proc. EUSIPCO*, pages 1–5.
- Dobigeon, N., Hero, A. O., and Tournieret, J.-Y. (2009). Hierarchical bayesian sparse image reconstruction with application to MRFM. *IEEE Trans. Image Process.*, 18(9):2059–2070.
- Donoho, D. L. (2006). Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289–1306.
- Foucart, S. and Rauhut, H. (2013). *A mathematical introduction to compressive sensing*. Birkhauser.
- Gorodnitsky, I. F. and Rao, B. D. (1997). Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm. *IEEE Transactions on signal processing*, 45(3):600–616.
- Hernández-Lobato, J. M., Hernández-Lobato, D., and Suárez, A. (2015). Expectation propagation in linear regression models with spike-and-slab priors. *Machine Learning*, 99(3):437–487.
- Ji, S., Xue, Y., and Carin, L. (2008). Bayesian compressive sensing. *IEEE Trans. Signal Process.*, 56(6):2346–2356.
- Jia, X., Men, C., Lou, Y., and Jiang, S. B. (2011). Beam orientation optimization for intensity modulated radiation therapy using adaptive $l_{2,1}$ -minimization. *Phys. Med. Biol.*, 56:6205–6222.
- Kappen, H. J. and Gómez, V. (2014). The variational garrote. *Machine Learning*, 96(3):269–294.
- Lavielle, M. (2009). Bayesian deconvolution of bernoulli-gaussian processes. *Signal Process.*, 33(1):67–79.
- LeCun, Y., Cortes, C., and Burges, C. J. C. (2013). MNIST dataset. <http://yann.lecun.com/exdb/mnist/>.
- Lee, S. H., Lee, Y. H., Song, H.-T., and Suh, J.-S. (2017). Rapid acquisition of magnetic resonance imaging of the shoulder using three-dimensional fast spin echo sequence with compressed sensing. *Magnetic Resonance Imaging*, 42:152–157.

- Lu, X., Wang, Y., and Yuan, Y. (2013a). Sparse coding from a bayesian perspective. *IEEE Trans. Neural Netw. Learn. Syst.*, 24(6):929–939.
- Lu, X., Yuan, Y., and Yan, P. (2013b). Sparse coding for image denoising using spike and slab prior. *Neurocomputing*, 106:12–20.
- Lustig, M., Donoho, D. L., Santos, J. M., and Pauly, J. M. (2008). Compressed sensing MRI. *IEEE Signal Process. Mag.*, 25(2):72–82.
- Malioutov, D., Cetin, M., and Willsky, A. S. (2005). A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE transactions on signal processing*, 53(8):3010–3022.
- Mitchell, T. J. and Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Amer. Statist. Assoc.*, 83(404):1023–1032.
- Mohimani, H., Babaie-Zadeh, M., and Jutten, C. (2009). A fast approach for overcomplete sparse decomposition based on smoothed l^0 norm. *IEEE Transactions on Signal Processing*, 57(1):289–301.
- Mousavi, A., Maleki, A., and Baraniuk, R. G. (2013). Asymptotic analysis of lassos solution path with implications for approximate message passing. in *arXiv preprint arXiv:1309.5979*.
- Mousavi, H. S., Monga, V., and Tran, T. D. (2015). Iterative convex refinement for sparse recovery. *IEEE Signal Processing Letters*, 22(11):1903–1907.
- Mousavi, H. S., Srinivas, U., Monga, V., Suo, Y., Dao, M., and Tran, T. D. (2014). Multi-task image classification via collaborative, hierarchical spike and slab priors. in *Proc. IEEE Conf. on Image Processing*, pages 4236–4240.
- Pustelnik, N., Chaux, C., Pesquet, J.-C., and Comtat, C. (2010). Parallel algorithm and hybrid regularization for dynamic PET reconstruction. *IEEE*, pages 2423–2427.
- Sadeghi, M., Babaie-Zadeh, M., and Jutten, C. (2013). Dictionary learning for sparse representation: A novel approach. *IEEE Signal Process. Lett.*, 20(12):1195–1198.
- Srinivas, U., Suo, Y., Dao, M., Monga, V., and Tran, T. D. (2015). Structured sparse priors for image classification. *IEEE Trans. Image Processing*, 24(6):1763–1776.
- Suo, Y., Dao, M., Tran, T., Mousavi, H., Srinivas, U., and Monga, V. (2014). Group structured dirty dictionary learning for classification. in *Proc. IEEE Conf. on Image Processing*, pages 150–154.
- Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *J. Royal Statistical Soc. B*, 58(1):267–288.
- Titsias, M. K. and Lázaro-Gredilla, M. (2011). Spike and slab variational inference for multi-task and multiple kernel learning. In *NIPS’2011*.
- Tropp, J. A. and Gilbert, A. C. (2007). Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inf. Theory*, 53(12):4655–4666.
- Vu, T. H., Mousavi, H. S., and Monga, V. (2017). Adaptive matching pursuit for sparse signal recovery. in *Proc. Int. Conf. Acoust., Speech Signal Process.*, pages 4331–4335.
- Wright, S. J., Nowak, R. D., and Figueiredo, M. A. (2009). Sparse reconstruction by separable approximation. *IEEE Trans. Signal Process.*, 57(7):2479–2493.
- Wu, H. and Wang, S. (2012). Adaptive sparsity matching pursuit algorithm for sparse reconstruction. *IEEE Signal Processing Letters*, 19(8):471–474.
- Yen, T.-J. (2011). A majorization-minimization approach to variable selection using spike and slab priors. *Ann. Statist.*, 39(3):1748–1775.