Multidim Syst Sign Process (2015) 26:787–802
DOI 10.1007/s11045-014-0282-7

Split Bregman algorithms for sparse group Lasso with application to MRI reconstruction

Jian Zou · Yuli Fu

Received: 7 October 2013 / Revised: 4 January 2014 / Accepted: 28 January 2014 /
Published online: 12 February 2014

Abstract Sparse group lasso, concerning with group-wise and within-group sparsity, is generally considered difficult to solve due to the mixed-norm structure. In this paper, we propose efficient algorithms based on split Bregman iteration to solve sparse group lasso problems, including a synthesis prior form and an analysis prior form. These algorithms have low computational complexity and are suitable for large scale problems. The convergence of the proposed algorithms is also discussed. Moreover, the proposed algorithms are used for magnetic resonance imaging reconstruction. Numerical results show the effectiveness of the proposed algorithms.

Keywords Sparse group lasso · Split Bregman iteration · Convergence · MRI reconstruction

1 Introduction

One of the most important problems in statistical sparse signal processing is to estimate an unknown deterministic sparse vector though a linear transformation (Bruckstein et al. 2009; Candès and Wakin 2008; He et al. 2007, 2009a), i.e.,

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^m$, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{m\times n}$, $m < n$ and $\mathbf{e} \in \mathbb{R}^m$ is the additive noise vector.

A traditional approach to solve (1) is least absolute shrinkage and selection operator (Lasso) (Tibshirani 1996; Zhao and Yu 2006; Zou 2006), which solve the following optimization problem:

J. Zou
School of Information and Mathematics, Yangtze University, Jingzhou 434020, Hubei, China
e-mail: zoujianjz@gmail.com

Y. Fu (✉)
School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641, Guangdong, China
e-mail: fuyuli@scut.edu.cn

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1, \tag{2}$$

where $\|\mathbf{x}\|_1 = \sum_{j=1}^{n} |\mathbf{x}_j|$ and $\mathbf{x}_j$ denotes the jth element of $\mathbf{x}$.

In the above standard lasso model, the nonzero elements can appear anywhere in the vector. However, sparse signals exhibit additional structure in some practical scenarios. One simple but important structured sparsity is represented by group (or block) sparse model, in which the nonzero elements appear in groups rather than arbitrarily spread throughout the vector. In order to estimate these group sparse coefficients, group lasso (Huang and Zhang 2010; Meier et al. 2008; Yuan and Lin 2006) is proposed as follows:

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_{2,1}, \tag{3}$$

where $\|\mathbf{x}\|_{2,1} = \sum_{j=1}^{N} \|\mathbf{x}[j]\|_2$, $\mathbf{x}[j]$ denotes the jth group of $\mathbf{x}$ and N is the number of groups.

Group lasso gives a sparse set of groups, that is, if it includes a group in the model, then all elements in the group will be nonzero. Sometimes, both sparsity of groups and within each group should be considered, for example, in gene expression data, only particularly "important" genes in pathways of interest need identified (Simon et al. 2013), in climate data predictive model, only important climate variables at important locations are selected from the plethora of potential covariates at various spatial locations (Chatterjee et al. 2011). To deal with these problems, sparse group lasso (Friedman et al. 2010; Simon et al. 2013) is investigated as follows:

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda_1\|\mathbf{x}\|_{2,1} + \lambda_2\|\mathbf{x}\|_1. \tag{4}$$

It is possible to recover the signal by (4) when the signal itself is sparse. However, in many applications, e.g. image processing, the signal itself is not sparse, but has a sparse representation in some transform domains, such transforms can be, e.g., Fourier transforms, local cosine transforms, wavelet or framelet transforms. In such cases the synthesis prior model (4) is not applicable. The following analysis prior model should to be employed:

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda_1\|\mathbf{D}\mathbf{x}\|_{2,1} + \lambda_2\|\mathbf{D}\mathbf{x}\|_1 \tag{5}$$

where $\mathbf{D}$ is a sparsifying transform matrix.

Sparse group lasso yields both group-wise and within-group sparsity, and may help to recover the corresponding signal more accurately. Sparse group lasso arise in varied areas of signal processing, e.g., gene Expression (Simon et al. 2013), climate data predictive (Chatterjee et al. 2011) and Video tagging (Zhu et al. 2013). It is generally considered difficult to solve (4) or (5) due to the non-smoothness of the $\ell_{2,1}$ mixed norm and ℓ_1 norm. Existing methods for solving problem (4) base on either generalized gradient descent strategy (Friedman et al. 2010; Simon et al. 2013) or gradient projection strategy (Liu et al. 2009b).

In this paper, new algorithms based on split Bregman iteration will be proposed for solving sparse group lasso problem, including the synthesis prior model (4) and the analysis prior model (5). These algorithms have low computational complexity and are suitable for large scale problems. Then sparse group lasso will be used for MRI reconstruction. Numerical results will demonstrate the effectiveness of the proposed algorithms.

Through the paper, we denote vectors by boldface lowercase letters, e.g., $\mathbf{x}$, matrices by boldface uppercase letters, e.g., $\mathbf{X}$. $\mathbf{I}$ is an identity matrix. For a given matrix $\mathbf{X}$, $\mathbf{X}^T$ denotes

its transpose. $\|\mathbf{x}\|_1$ and $\|\mathbf{x}\|_2$ denote the ℓ_1 and ℓ_2 norm of a vector $\mathbf{x}$,respectively. $\mathbb{R}$ denotes the set of real numbers.

2 Split Bregman iteration for sparse group Lasso

2.1 Split Bregman iteration for sparse group Lasso

Split Bregman iteration, motivated by the Bregman distance, has been shown to be an efficient method for solving various optimization problems arising in many areas such as compressed sensing (Yin et al. 2008), matrix rank minimization (Ma et al. 2011), matrix factorizations (Jiang and Yin 2012), image processing (Cai et al. 2009; Goldstein et al. 2010; Goldstein and Osher 2009; Xu et al. 2014) and many other applications (Smith et al. 2012; Ye and Xie 2011). In the following, we will use split Bregman iteration to deduce algorithms for the proposed model (4) and (5).

Firstly we deal with the analysis prior model (5) since (4) is a special case of (5) when we set $\mathbf{D} = \mathbf{I}$. For (5), the variable is "coupled" between the different norms since $\ell_{2,1}$ mixed-norm and ℓ_1 norm contain the variable $\mathbf{Dx}$ but the ℓ_1-norm contains $\mathbf{Ax}$. Based on variable splitting strategy, we first introduce auxiliary variables and transform (5) into an equivalent problem

$$\begin{aligned} &\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{Ax}\|_2^2 + \lambda_1\|\mathbf{u}\|_{2,1} + \lambda_2\|\mathbf{v}\|_1 \\ &\text{s.t. } \mathbf{u} = \mathbf{Dx}, \mathbf{v} = \mathbf{Dx}. \end{aligned} \tag{6}$$

Note that the variables in the objective function of problem (6) are "de-coupled" since the variables in different norms are independent. (6) can be converted into an unconstrained problem

$$\min_{\mathbf{x},\mathbf{u},\mathbf{v}} \frac{1}{2}\|\mathbf{y}-\mathbf{Ax}\|_2^2 + \lambda_1\|\mathbf{u}\|_{2,1} + \lambda_2\|\mathbf{v}\|_1 + \frac{\mu_1}{2}\|\mathbf{u}-\mathbf{Dx}\|_2^2 + \frac{\mu_2}{2}\|\mathbf{v}-\mathbf{Dx}\|_2^2. \tag{7}$$

We then apply Bregman iteration technique (Yin et al. 2008), reduce problem (7) to a sequence of unconstrained problems as follows:

$$\begin{cases} (\mathbf{x}^{k+1}, \mathbf{u}^{k+1}, \mathbf{v}^{k+1}) \\ = \arg\min_{\mathbf{x},\mathbf{u},\mathbf{v}} \frac{1}{2}\|\mathbf{y}-\mathbf{Ax}\|_2^2 + \lambda_1\|\mathbf{u}\|_{2,1} + \lambda_2\|\mathbf{v}\|_1 \\ + \frac{\mu_1}{2}\|\mathbf{u}-\mathbf{Dx}-\mathbf{b}^k\|_2^2 + \frac{\mu_2}{2}\|\mathbf{v}-\mathbf{Dx}-\mathbf{c}^k\|_2^2, \\ \mathbf{b}^{k+1} = \mathbf{b}^k + (\mathbf{Dx}^{k+1} - \mathbf{u}^{k+1}), \\ \mathbf{c}^{k+1} = \mathbf{c}^k + (\mathbf{Dx}^{k+1} - \mathbf{v}^{k+1}). \end{cases} \tag{8}$$

Split Bregman technique in (Cai et al. 2009; Goldstein and Osher 2009) show that the first minimization of (8) can solve efficiently by minimizing with respect to $\mathbf{x}$, $\mathbf{u}$, $\mathbf{v}$ separately as follows:

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{Ax}\|_2^2 + \frac{\mu_1}{2}\|\mathbf{Dx}-\mathbf{u}^k+\mathbf{b}^k\|_2^2 + \frac{\mu_2}{2}\|\mathbf{Dx}-\mathbf{v}^k+\mathbf{c}^k\|_2^2, \tag{9}$$

$$\mathbf{u}^{k+1} = \arg\min_{\mathbf{u}} \lambda_1\|\mathbf{u}\|_{2,1} + \frac{\mu_1}{2}\|\mathbf{u}-\mathbf{Dx}^{k+1}-\mathbf{b}^k\|_2^2, \tag{10}$$

$$\mathbf{v}^{k+1} = \arg\min_{\mathbf{v}} \lambda_2\|\mathbf{v}\|_1 + \frac{\mu_2}{2}\|\mathbf{v}-\mathbf{Dx}^{k+1}-\mathbf{c}^k\|_2^2. \tag{11}$$

Equation (9) is convex and differentiable, optimality conditions for $\mathbf{x}$ are easily derived. By differentiating with respect to $\mathbf{x}$ and setting the result equal to zero, we have

$$\left(\mathbf{A}^T\mathbf{A} + \mu_1\mathbf{D}^T\mathbf{D} + \mu_2\mathbf{D}^T\mathbf{D}\right)\mathbf{x} = \mathbf{A}^T\mathbf{y} + \mu_1\mathbf{D}^T(\mathbf{u}^k - \mathbf{b}^k) + \mu_2\mathbf{D}^T(\mathbf{v}^k - \mathbf{c}^k). \quad (12)$$

Equation (10) can be solved by group-wise shrinkage operators, which can be also seen in paper (Deng et al. 2011). Equation (10) is equivalent to

$$\min_{\mathbf{u}} \sum_{j=1}^{N} \left(\lambda_1\|\mathbf{u}[j]\|_2 + \frac{\mu_1}{2}\|\mathbf{u}[j] - (\mathbf{D}\mathbf{x})^{k+1}[j] - \mathbf{b}^k[j]\|_2^2\right), \quad (13)$$

which has a close form solution by shrinkage (or soft thresholding) formula

$$\mathbf{u}[j] = \text{shrink2}\left((\mathbf{D}\mathbf{x})^{k+1}[j] + \mathbf{b}^k[j], \lambda_1/\mu_1\right), \quad j = 1, \ldots, N, \quad (14)$$

where

$$\text{shrink2}(\mathbf{x}, \gamma) = \frac{\mathbf{x}}{\|\mathbf{x}\|_2}\max\{\|\mathbf{x}\|_2 - \gamma, 0\}. \quad (15)$$

Equation (11) is a standard ℓ_1-minimization problem and the solution of (11) is given explicitly by shrinkage (or soft thresholding) formula

$$\mathbf{v}_j = \text{shrink1}\left((\mathbf{D}\mathbf{x})_j^{k+1} + \mathbf{c}_j^k, \lambda_2/\mu_2\right), \quad j = 1, \ldots, n, \quad (16)$$

where

$$\text{shrink1}(x, \gamma) = \text{sgn}(x) \odot \max\{x - \gamma, 0\}. \quad (17)$$

Now, we present a split Bregman algorithm for analysis sparse group lasso problem (SBASGL) as Algorithm 1.

Algorithm 1 SBASGL

Require: $\mathbf{A}, \mathbf{y}, \mathbf{x}^0 = \mathbf{u}^0 = \mathbf{v}^0 = \mathbf{0}$.
Ensure: $\mathbf{x}$.
while "stopping criterion is not met" **do**
$\mathbf{x}^{k+1} = (\mathbf{A}^T\mathbf{A} + \mu_1\mathbf{D}^T\mathbf{D} + \mu_2\mathbf{D}^T\mathbf{D})^{-1}(\mathbf{A}^T\mathbf{y} + \mu_1\mathbf{D}^T(\mathbf{u}^k - \mathbf{b}^k) + \mu_2\mathbf{D}^T(\mathbf{v}^k - \mathbf{c}^k))$;
$\mathbf{u}[j]^{k+1} = \text{shrink2}((\mathbf{D}\mathbf{x})^{k+1}[j] + \mathbf{b}^k[j], \lambda_1/\mu_1), \quad j = 1, \ldots, N$;
$\mathbf{v}_j = \text{shrink1}((\mathbf{D}\mathbf{x})_j^{k+1} + \mathbf{c}_j^k, \lambda_2/\mu_2), \quad j = 1, \ldots, n$;
$\mathbf{b}^{k+1} = \mathbf{b}^k + (\mathbf{D}\mathbf{x}^{k+1} - \mathbf{u}^{k+1})$;
$\mathbf{c}^{k+1} = \mathbf{c}^k + (\mathbf{D}\mathbf{x}^{k+1} - \mathbf{v}^{k+1})$.
end while

For the synthesis prior model (4), similar with the corresponding deduction of SBASGL, we can easily get the split Bregman algorithm for synthesis sparse group lasso problem (SBSSGL) as Algorithm 2.

In Algorithm 1, the update of $\mathbf{x}$ needs to compute $(\mathbf{A}^T\mathbf{A} + \mu_1\mathbf{D}^T\mathbf{D} + \mu_2\mathbf{D}^T\mathbf{D})^{-1}$, which is computationally expensive. Fortunately, noting that the matrix $\mathbf{A}$ and $\mathbf{D}$ are fixed, we can precalculate it. Especially, in some applications, e.g., image processing, transform matrix $\mathbf{D}$ would be curvelet, orthonormal wavelet, and framelet. In these cases $\mathbf{D}^T\mathbf{D} = \mathbf{I}$ and we only need to compute $(\mathbf{A}^T\mathbf{A} + \mu_1\mathbf{I} + \mu_2\mathbf{I})^{-1}$. Then the main computational cost of the algorithms becomes matrix-vector multiplication, which computational complexity is $O(mn)$ when $\mathbf{A}$

Algorithm 2 SBSSGL

Require: $\mathbf{A}, \mathbf{y}, \mathbf{x}^0 = \mathbf{u}^0 = \mathbf{v}^0 = \mathbf{0}$.
Ensure: $\mathbf{x}$.
while "stopping criterion is not met" **do**
$\mathbf{x}^{k+1} = (\mathbf{A}^T\mathbf{A} + \mu_1\mathbf{I} + \mu_2\mathbf{I})^{-1}(\mathbf{A}^T\mathbf{y} + \mu_1(\mathbf{u}^k - \mathbf{b}^k) + \mu_2(\mathbf{v}^k - \mathbf{c}^k))$;
$\mathbf{u}[j]^{k+1} = \text{shrink2}(\mathbf{x}^{k+1}[j] + \mathbf{b}^k[j], \lambda_1/\mu_1), \quad j = 1, \ldots, N$;
$\mathbf{v}_j = \text{shrink1}(\mathbf{x}_j^{k+1} + \mathbf{c}_j^k, \lambda_2/\mu_2), \quad j = 1, \ldots, n$;
$\mathbf{b}^{k+1} = \mathbf{b}^k + (\mathbf{x}^{k+1} - \mathbf{u}^{k+1})$;
$\mathbf{c}^{k+1} = \mathbf{c}^k + (\mathbf{x}^{k+1} - \mathbf{v}^{k+1})$.
end while

is a random matrix (Gaussian, etc.) or $O(n \log n)$ when $\mathbf{A}$ is a partial fast transform matrix (DCT, DFT, etc.).

2.2 Convergence results

The convergence property of Algorithms 1 and 2 are shown in the following theorems.

Theorem 1 *Suppose that there exists at least one solution* $\mathbf{x}^*$ *of* (5). *Assume* $\mu_1 > 0$ *and* $\mu_2 > 0$. *Then the following property for the split Bregman iteration in Algorithm 1 holds:*

$$\begin{aligned} &\lim_{k\to+\infty} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}^k\|_2^2 + \lambda_1\|\mathbf{D}\mathbf{x}^k\|_{2,1} + \lambda_2\|\mathbf{D}\mathbf{x}^k\|_1 \\ &= \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}^*\|_2^2 + \lambda_1\|\mathbf{D}\mathbf{x}^*\|_{2,1} + \lambda_2\|\mathbf{D}\mathbf{x}^*\|_1. \end{aligned} \tag{18}$$

Furthermore,

$$\lim_{k\to+\infty} \|\mathbf{x}^k - \mathbf{x}^*\|_2 = 0 \tag{19}$$

whenever (5) *has a unique solution.*

Theorem 2 *Suppose that there exists at least one solution* $\mathbf{x}^*$ *of* (4). *Assume* $\mu_1 > 0$ *and* $\mu_2 > 0$. *Then the following property for the split Bregman iteration in Algorithm 2 holds:*

$$\begin{aligned} &\lim_{k\to+\infty} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}^k\|_2^2 + \lambda_1\|\mathbf{x}^k\|_{2,1} + \lambda_2\|\mathbf{x}^k\|_1 \\ &= \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}^*\|_2^2 + \lambda_1\|\mathbf{x}^*\|_{2,1} + \lambda_2\|\mathbf{x}^*\|_1. \end{aligned} \tag{20}$$

Furthermore,

$$\lim_{k\to+\infty} \|\mathbf{x}^k - \mathbf{x}^*\|_2 = 0 \tag{21}$$

whenever (4) *has a unique solution.*

The proof of Theorem 1, which can be found in the "Appendix", is inspired by the work of Cai et al. (2009). However, this proof is much different due to the mixed norm. The proof of Theorem 2 is quite similar to that of Theorem 1 and so is omitted.

3 Numerical experiments

In this section, we present numerical results to illustrate the performance of our split Bregman algorithms for both synthetic data and real-world Data. In this section, we compare the proposed algorithms with the state-of-the-art algorithm sparse learning with efficient projections (SLEPs) (Liu et al. 2009b). SLEP package provides functions for solving a family of sparse learning problems. Particularly, the function "sgLeastR", "glLeastR" and "LeastR" solve sparse group lasso, group lasso and lasso problem, respectively. The algorithm in paper (Simon et al. 2013) is not compared in this section since this algorithm is written using R language. In contrast, both SBSSGL and SLEP are written using MATLAB. All experiments are run in MATLAB 7.8.0 on the PC with Intel Celeron 2.4 GHz and 2 G memory.

3.1 Synthetic data experiment

In this experiment, a sparse group signal is generated as follows: a vector $\mathbf{x} \in \mathbb{R}^n$ is first randomly divided into N groups and then K of the groups are randomly chosen as active ones while the remaining groups are filled with zeros. For the active groups, k of the elements are randomly chosen as active ones and filled with zero mean Gaussian random samples of unit variance while the remaining elements are also filled with zeros. All group lengths are enforced the same in the following two experiments to ensure the number of nonzero elements increase with the active group number. The measurement matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is with Gaussian independent and identically distribution (i.i.d) entries of zero mean and variance $1/(2n)$. The measurement signal is given by $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$, where $\mathbf{e}$ is Gaussian white noise. Then the unknown sparse group signal is reconstructed by sparse group lasso (SBSSGL, sgLeastR), group lasso (glLeastR), and lasso (LeastR). Mean square error (MSE) is used here to quantify the reconstruction performance.

We present two classes of numerical experiments to show how the MSE of each algorithm changes with different parameters.

The first experiment illustrates how the MSE of each algorithm changes along the group-wise sparsity level. Here the group-wise sparsity level is measured by the active group number K. We set $n = 1{,}000$, $m = 500$ and divide $\mathbf{x}$ into $N = 50$ groups with group length $d = 20$, in each active group, 50 % elements are chosen as active ones. Let the active group number K change from 1 to 10. The variance of Gaussian white noise is 0.01. Figure 1a shows how the MSE curves of SBSSGL, sgLeastR, glLeastR and LeastR change respectively, along the group-wise sparsity level. It is seen that the MSE curve of SBSSGL keeps the lowest level among these algorithms.

The second experiment illustrates how the MSE of each algorithm changes along the within-group sparsity level. Here the within-group sparsity level is defined by $r = k/d$, in which k is the number of active elements in each active group and d is the group length. We also set $n = 1{,}000$, $m = 500$, $N = 50$, $d = 20$ and fix $K = 5$. Let the within-group sparsity level r change from 0.1 to 1. The variance of Gaussian white noise is 0.01. Figure 1b shows how the MSE curves of SBSSGL, sgLeastR, glLeastR and LeastR change respectively, along the within-group sparsity level. It is seen that the MSE curve of SBSSGL keeps the lowest level among these algorithms.

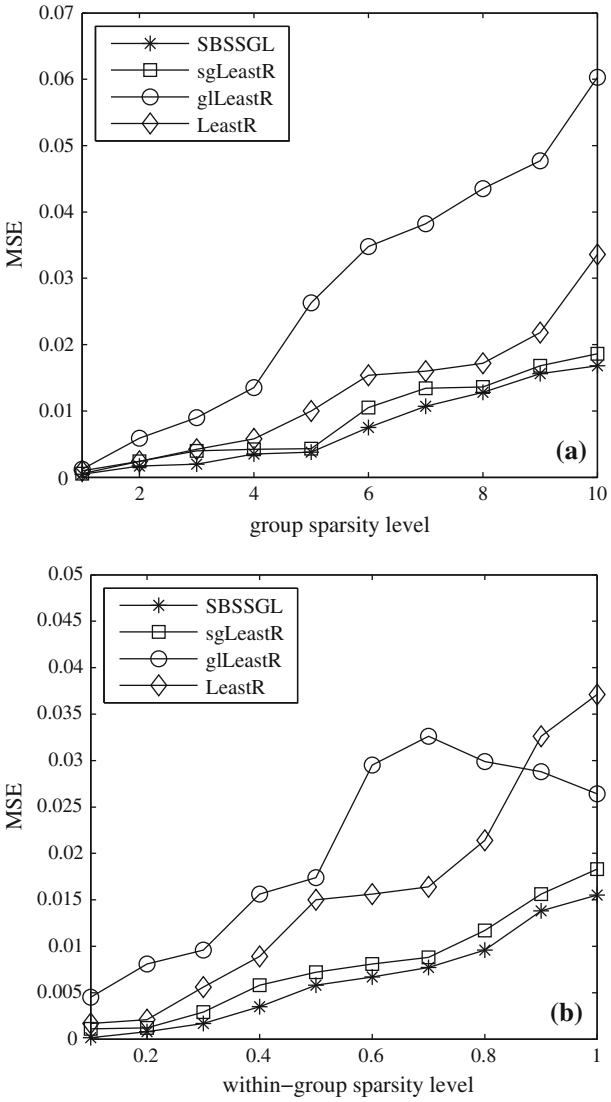

Fig. 1 Comparison of SBSSGL, sgLeastR, glLeastR and LeastR. **a** MSE as a function of the group-wise sparsity level. **b** MSE as a function of the within-group sparsity level. All the results are average of 100 runs

3.2 Real-world data experiment

In this experiment, MRI reconstruction problem (Lustig et al. 2007, 2008; Liu et al. 2009; Lee et al. 2013a, b) is considered as a sparse group lasso problem to illustrate the validity of the proposed algorithm for real-world data.

In paper He et al. (2008, 2009b) and Zou et al. (2012), for a 512×512 MR image $\mathbf{I}$ (shown in Fig. 2a), 472 from 512 possible parallel lines in the spatial frequency of $\mathbf{I}$ were extracted, the other 40 lines of 512 were removed, then MR image reconstruction problem was considered as a sparse reconstruction problem (He et al. 2009b) or a multiple measurement

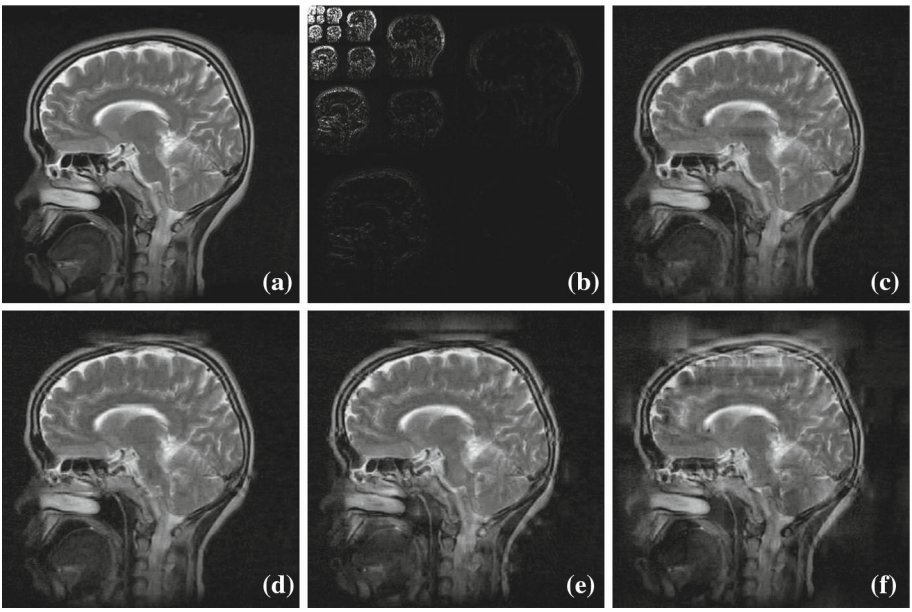

Fig. 2 MRI reconstruction results. **a** Original MR image. **b** Wavelet coefficient. **c** Reconstruction by SBASGL. **d** Reconstruction by sgLeastR. **e** Reconstruction by glLeastR. **f** Reconstruction by LeastR

vector (MMV) problem (He et al. 2008; Zou et al. 2012), which can be seen as a special group (block) sparse problem. However, as far as we can see from Fig. 2b, the elements in the active rows are not all nonzero, in other words, it is also sparse in the active rows. For this reason, we can consider MR image reconstruction problem as a sparse group lasso problem.

Since the MR image itself is not sparse, but has a sparse representation in some transform domains, analysis prior model (5) is used in this experiment, sparsifying transform matrix $\mathbf{D}$ is an orthonormal matrix corresponding to a 2D-DWT for a two-level Daubechies-4 wavelet. Then the image can be rewritten as a sparse group vector for which the group length is just the row length of the image. The measurement matrix was $\mathbf{A} = \mathbf{FD}$ where $\mathbf{F}$ is a $m \times n$ partial discrete Fourier matrix by randomly removing some rows of the 2-D DFT, here, we set $m = 0.5\,n$.

In this experiment, we treat the MR image as sparse group vector and reconstruct it by SBASGL (Fig. 2c) and sgLeastR (Fig. 2d). For comparison, we also treat the MR image as group sparse and standard sparse vectors and reconstruct by glLeastR (Fig. 2e) and LeastR (Fig. 2f), respectively. Peak signal to noise ratio (PSNR) (Wang et al. 2004) is used to quantify the reconstruction performance. PSNR is defined as

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{\mathrm{MAX}^2}{\mathrm{MSE}}\right) \tag{22}$$

where MAX is the maximum possible pixel value of the image and MSE is defined as

$$\mathrm{MSE} = \frac{1}{mn}\sum_{i=1}^{m-1}\sum_{j=1}^{n-1}[\mathbf{I}_{ori}(i,\,j) - \mathbf{I}_{rec}(i,\,j)]^2 \tag{23}$$

Table 1 MRI reconstruction results

Algorithm	PSNR (dB)	CPU time (s)
SBASGL	34.90	33.72
sgLeastR	29.43	41.43
glLeastR	26.14	45.85
LeastR	22.86	50.17

where $\mathbf{I}_{ori}$ and $\mathbf{I}_{rec}$ are the original image and reconstructed image, respectively. m, n are the size of the images. In this paper, $m = n = 512$ and $\mathrm{MAX} = 255$ since the pixels of the images are represented by 8 bits per sample.

Table 1 shows reconstruction results. Considering PSNR and CPU time comprehensively, SBASGL algorithm outperform some others.

In this experiment, the algorithms' performance was evaluated employing PC platform without any hardware accelerated technique. Real-time or fast reconstruction of MR image is essential for the commercial implementations. Parallel computing technique such as GPU (CUDA, etc) has gained a great deal of interest in recent years and has become a standard in commercial computing since it can significantly reduce the computation time (Bilen et al. 2012; Stone et al. 2008). The proposed split Bregman algorithms for MMV can also be accelerated in case of a parallel implementation with GPU, similar to the work in Smith et al. (2012).

4 Conclusion

In this paper, we proposed efficient split Bregman algorithms for solving sparse group lasso problems, including synthesis and analysis prior forms. Split Bregman algorithms "decouple" the variables in different norms and update the variables alternately. They are computationally efficient and suitable for large scale problems. The numerical results demonstrate the validity of the proposed algorithms for both synthetic and real-world data.

Acknowledgments The work is supported partially by Guangdong and National ministry of education ii projection (Grant No. 2012B091100331), International cooperation project of Guangdong Natural Science Foundation (Grant No. 2009B050700020), NOS—Guangdong Union Project (Grant No. U0835003), NOS (Grant Nos. 61104053, 61103122).

Appendix: Proof of Theorem 1

In this appendix, we present the proof of Theorem 1.

Proof Since subproblems (9)–(11) are convex, the first order optimality condition of Algorithm 1 gives as follows:

$$\begin{cases} \mathbf{0} = \mathbf{A}^T(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{y}) + \mu_1\mathbf{D}^T(\mathbf{D}\mathbf{x}^{k+1} - \mathbf{u}^k + \mathbf{b}^k) + \mu_2\mathbf{D}^T(\mathbf{D}\mathbf{x}^{k+1} - \mathbf{v}^k + \mathbf{c}^k), \\ \mathbf{0} = \lambda_1\mathbf{p}^{k+1} + \mu_1(\mathbf{u}^{k+1} - \mathbf{D}\mathbf{x}^k - \mathbf{b}^k), \\ \mathbf{0} = \lambda_2\mathbf{q}^{k+1} + \mu_2(\mathbf{v}^{k+1} - \mathbf{D}\mathbf{x}^k - \mathbf{c}^k), \\ \mathbf{b}^{k+1} = \mathbf{b}^k + (\mathbf{D}\mathbf{x}^{k+1} - \mathbf{u}^{k+1}), \\ \mathbf{c}^{k+1} = \mathbf{c}^k + (\mathbf{D}\mathbf{x}^{k+1} - \mathbf{v}^{k+1}), \end{cases} \tag{24}$$

where $\mathbf{p}^{k+1} \in \partial\|\mathbf{u}^{k+1}\|_{2,1}$ and $\mathbf{q}^{k+1} \in \partial\|\mathbf{v}^{k+1}\|_1$. The definition of subgradients $\partial\|\cdot\|_{2,1}$ and $\partial\|\cdot\|_1$ can be seen in Simon et al. (2013).

Since there exists at least one solution $\mathbf{x}^*$ of (5) by the assumption, by the first order optimality condition, $\mathbf{x}^*$ must satisfy

$$\mathbf{A}^T(\mathbf{A}\mathbf{x}^* - \mathbf{y}) + \lambda_1\mathbf{D}^T\mathbf{p}^* + \lambda_2\mathbf{D}^T\mathbf{q}^* = 0, \tag{25}$$

where $\mathbf{p}^* \in \partial\|\mathbf{u}^*\|_{2,1}$, $\mathbf{q}^* \in \partial\|\mathbf{v}^*\|_1$ and $\mathbf{u}^* = \mathbf{v}^* = \mathbf{D}\mathbf{x}^*$.

Let $\mathbf{b}^* = \lambda_1\mathbf{p}^*/\mu_1$ and $\mathbf{c}^* = \lambda_2\mathbf{q}^*/\mu_2$. Then (25) can be formulated as

$$\begin{cases} \mathbf{0} = \mathbf{A}^T(\mathbf{A}\mathbf{x}^* - \mathbf{y}) + \mu_1\mathbf{D}^T(\mathbf{D}\mathbf{x}^* - \mathbf{u}^* + \mathbf{b}^*) + \mu_2\mathbf{D}^T(\mathbf{D}\mathbf{x}^* - \mathbf{v}^* + \mathbf{c}^*), \\ \mathbf{0} = \lambda_1\mathbf{p}^* + \mu_1(\mathbf{u}^* - \mathbf{D}\mathbf{x}^* - \mathbf{b}^*), \\ \mathbf{0} = \lambda_2\mathbf{q}^* + \mu_2(\mathbf{v}^* - \mathbf{D}\mathbf{x}^* - \mathbf{c}^*), \\ \mathbf{b}^* = \mathbf{b}^* + (\mathbf{D}\mathbf{x}^* - \mathbf{u}^*), \\ \mathbf{c}^* = \mathbf{c}^* + (\mathbf{D}\mathbf{x}^* - \mathbf{v}^*). \end{cases} \tag{26}$$

Comparing (26) with (24), $(\mathbf{x}^*, \mathbf{u}^*, \mathbf{v}^*, \mathbf{b}^*, \mathbf{c}^*)$ is a fixed point of Algorithm 1. Paper Cai et al. (2009) pointed out that if the unconstrained spilt Bregman iteration converges, it converges to a solution of (4).

Denote the errors by $\mathbf{x}_e^k = \mathbf{x}^k - \mathbf{x}^*$, $\mathbf{u}_e^k = \mathbf{u}^k - \mathbf{u}^*$, $\mathbf{v}_e^k = \mathbf{v}^k - \mathbf{v}^*$, $\mathbf{b}_e^k = \mathbf{b}^k - \mathbf{b}^*$ and $\mathbf{c}_e^k = \mathbf{c}^k - \mathbf{c}^*$.

Subtracting the first equation of (24) by the first equation of (26), we obtain

$$\mathbf{0} = \mathbf{A}^T\mathbf{A}\mathbf{x}_e^{k+1} + \mu_1\mathbf{D}^T(\mathbf{D}\mathbf{x}_e^{k+1} - \mathbf{u}_e^k + \mathbf{b}_e^k) + \mu_2\mathbf{D}^T(\mathbf{D}\mathbf{x}_e^{k+1} - \mathbf{v}_e^k + \mathbf{c}_e^k). \tag{27}$$

Taking the inner product of the left- and right-hand sides with $\mathbf{x}_e^{k+1}$, we have

$$\begin{aligned} \mathbf{0} = {} & \left\|\mathbf{A}\mathbf{x}_e^{k+1}\right\|_2^2 + \mu_1\left\|\mathbf{D}\mathbf{x}_e^{k+1}\right\|_2^2 - \mu_1\left\langle\mathbf{u}_e^k, \mathbf{D}\mathbf{x}_e^{k+1}\right\rangle \\ & + \mu_1\left\langle\mathbf{b}_e^k, \mathbf{D}\mathbf{x}_e^{k+1}\right\rangle + \mu_2\left\|\mathbf{D}\mathbf{x}_e^{k+1}\right\|_2^2 - \mu_2\left\langle\mathbf{v}_e^k, \mathbf{D}\mathbf{x}_e^{k+1}\right\rangle + \mu_2\left\langle\mathbf{c}_e^k, \mathbf{D}\mathbf{x}_e^{k+1}\right\rangle. \end{aligned} \tag{28}$$

Subtracting the second equation of (24) by the second equation of (26) and then taking the inner product of the left- and right-hand sides with $\mathbf{u}_e^{k+1}$, we have

$$\mathbf{0} = \lambda_1\left\langle\mathbf{p}^{k+1} - \mathbf{p}^*, \mathbf{u}^{k+1} - \mathbf{u}^*\right\rangle + \mu_1\left\|\mathbf{u}_e^{k+1}\right\|_2^2 - \mu_1\left\langle\mathbf{u}_e^{k+1}, \mathbf{D}\mathbf{x}_e^{k+1}\right\rangle - \mu_1\left\langle\mathbf{b}_e^k, \mathbf{u}_e^{k+1}\right\rangle \tag{29}$$

where $\mathbf{p}^{k+1} \in \partial\|\mathbf{u}^{k+1}\|_{2,1}$, $\mathbf{p}^* \in \partial\|\mathbf{u}^*\|_{2,1}$.

Similarly, from the third equation of (24) and (26) we have

$$\mathbf{0} = \lambda_2\left\langle\mathbf{q}^{k+1} - \mathbf{q}^*, \mathbf{v}^{k+1} - \mathbf{v}^*\right\rangle + \mu_2\left\|\mathbf{v}_e^{k+1}\right\|_2^2 - \mu_2\left\langle\mathbf{v}_e^{k+1}, \mathbf{D}\mathbf{x}_e^{k+1}\right\rangle - \mu_2\left\langle\mathbf{c}_e^k, \mathbf{v}_e^{k+1}\right\rangle \tag{30}$$

where $\mathbf{q}^{k+1} \in \partial\|\mathbf{v}^{k+1}\|_1$, $\mathbf{q}^* \in \partial\|\mathbf{v}^*\|_1$.

Adding both sides of (28)–(30), we have

$$\begin{aligned}\mathbf{0} = &\left\|\mathbf{A}\mathbf{x}_e^{k+1}\right\|_2^2 + \lambda_1\left\langle \mathbf{p}^{k+1}-\mathbf{p}^*, \mathbf{u}^{k+1}-\mathbf{u}^*\right\rangle + \lambda_2\left\langle \mathbf{q}^{k+1}-\mathbf{q}^*, \mathbf{v}^{k+1}-\mathbf{v}^*\right\rangle \\ &+\mu_1\left(\left\|\mathbf{D}\mathbf{x}_e^{k+1}\right\|_2^2 + \left\|\mathbf{u}_e^{k+1}\right\|_2^2 - \left\langle \mathbf{D}\mathbf{x}_e^{k+1}, \mathbf{u}_e^{k+1}+\mathbf{u}_e^k\right\rangle + \left\langle \mathbf{b}_e^k, \mathbf{D}\mathbf{x}_e^{k+1}-\mathbf{u}_e^{k+1}\right\rangle\right) \\ &+\mu_2\left(\left\|\mathbf{D}\mathbf{x}_e^{k+1}\right\|_2^2 + \left\|\mathbf{v}_e^{k+1}\right\|_2^2 - \left\langle \mathbf{D}\mathbf{x}_e^{k+1}, \mathbf{v}_e^{k+1}+\mathbf{v}_e^k\right\rangle + \left\langle \mathbf{c}_e^k, \mathbf{D}\mathbf{x}_e^{k+1}-\mathbf{v}_e^{k+1}\right\rangle\right).\end{aligned} \tag{31}$$

Furthermore, by subtracting the fourth equation of (24) by the corresponding one in (26), we obtain

$$\mathbf{b}_e^{k+1} = \mathbf{b}_e^k + \mathbf{D}\mathbf{x}_e^{k+1} - \mathbf{u}_e^{k+1}. \tag{32}$$

Taking square norm of both sides of (32) implies

$$\left\|\mathbf{b}_e^{k+1}\right\|_2^2 = \left\|\mathbf{b}_e^k\right\|_2^2 + \left\|\mathbf{D}\mathbf{x}_e^{k+1}-\mathbf{u}_e^{k+1}\right\|_2^2 + 2\left\langle \mathbf{b}_e^k, \mathbf{D}\mathbf{x}_e^{k+1}-\mathbf{u}_e^{k+1}\right\rangle, \tag{33}$$

and further

$$\left\langle \mathbf{b}_e^k, \mathbf{D}\mathbf{x}_e^{k+1}-\mathbf{u}_e^{k+1}\right\rangle = \frac{1}{2}\left(\left\|\mathbf{b}_e^{k+1}\right\|_2^2 - \left\|\mathbf{b}_e^k\right\|_2^2 - \left\|\mathbf{D}\mathbf{x}_e^{k+1}-\mathbf{u}_e^{k+1}\right\|_2^2\right). \tag{34}$$

Similarly, from the fifth equation of (24) and (26) we have

$$\left\langle \mathbf{c}_e^k, \mathbf{D}\mathbf{x}_e^{k+1}-\mathbf{v}_e^{k+1}\right\rangle = \frac{1}{2}\left(\left\|\mathbf{c}_e^{k+1}\right\|_2^2 - \left\|\mathbf{c}_e^k\right\|_2^2 - \left\|\mathbf{D}\mathbf{x}_e^{k+1}-\mathbf{v}_e^{k+1}\right\|_2^2\right). \tag{35}$$

Substituting (34) and (35) into (31), we have

$$\begin{aligned}&\frac{\mu_1}{2}\left(\left\|\mathbf{b}_e^k\right\|_2^2 - \left\|\mathbf{b}_e^{k+1}\right\|_2^2\right) + \frac{\mu_2}{2}\left(\left\|\mathbf{c}_e^k\right\|_2^2 - \left\|\mathbf{c}_e^{k+1}\right\|_2^2\right) \\ &= \left\|\mathbf{A}\mathbf{x}_e^{k+1}\right\|_2^2 + \lambda_1\left\langle \mathbf{p}^{k+1}-\mathbf{p}^*, \mathbf{u}^{k+1}-\mathbf{u}^*\right\rangle + \lambda_2\left\langle \mathbf{q}^{k+1}-\mathbf{q}^*, \mathbf{v}^{k+1}-\mathbf{v}^*\right\rangle \\ &\quad + \frac{\mu_1}{2}\left(\left\|\mathbf{D}\mathbf{x}_e^{k+1}-\mathbf{u}_e^k\right\|_2^2 + \left\|\mathbf{u}_e^{k+1}\right\|_2^2 - \left\|\mathbf{u}_e^k\right\|_2^2\right) \\ &\quad + \frac{\mu_2}{2}\left(\left\|\mathbf{D}\mathbf{x}_e^{k+1}-\mathbf{v}_e^k\right\|_2^2 + \left\|\mathbf{v}_e^{k+1}\right\|_2^2 - \left\|\mathbf{v}_e^k\right\|_2^2\right)\end{aligned} \tag{36}$$

Summing (36) from $k = 0$ to $k = K$ yields

$$\begin{aligned}
&\frac{\mu_1}{2}\left(\left\|\mathbf{b}_e^0\right\|_2^2 - \left\|\mathbf{b}_e^{K+1}\right\|_2^2\right) + \frac{\mu_2}{2}\left(\left\|\mathbf{c}_e^0\right\|_2^2 - \left\|\mathbf{c}_e^{K+1}\right\|_2^2\right) \\
&= \sum_{k=0}^{K}\left\|\mathbf{A}\mathbf{x}_e^{k+1}\right\|_2^2 \\
&\quad + \lambda_1\sum_{k=0}^{K}\left\langle \mathbf{p}^{k+1} - \mathbf{p}^*, \mathbf{u}^{k+1} - \mathbf{u}^*\right\rangle + \lambda_2\sum_{k=0}^{K}\left\langle \mathbf{q}^{k+1} - \mathbf{q}^*, \mathbf{v}^{k+1} - \mathbf{v}^*\right\rangle \\
&\quad + \frac{\mu_1}{2}\left(\sum_{k=0}^{K}\left\|\mathbf{D}\mathbf{x}_e^{k+1} - \mathbf{u}_e^k\right\|_2^2 + \left\|\mathbf{u}_e^{K+1}\right\|_2^2 - \left\|\mathbf{u}_e^0\right\|_2^2\right) \\
&\quad + \frac{\mu_2}{2}\left(\sum_{k=0}^{K}\left\|\mathbf{D}\mathbf{x}_e^{k+1} - \mathbf{v}_e^k\right\|_2^2 + \left\|\mathbf{v}_e^{K+1}\right\|_2^2 - \left\|\mathbf{v}_e^0\right\|_2^2\right).
\end{aligned} \tag{37}$$

Since $\mathbf{p}^{k+1} \in \partial\|\mathbf{u}^{k+1}\|_{2,1}$, $\mathbf{p}^* \in \partial\|\mathbf{u}^*\|_{2,1}$ and $\|\cdot\|_{2,1}$ is convex, for $\forall k$,

$$\begin{aligned}
\left\langle \mathbf{p}^{k+1} - \mathbf{p}^*, \mathbf{u}^{k+1} - \mathbf{u}^*\right\rangle &= \|\mathbf{u}^{k+1}\|_2^2 - \|\mathbf{u}^*\|_2^2 - \left\langle \mathbf{p}^*, \mathbf{u}^{k+1} - \mathbf{u}^*\right\rangle \\
&\quad + \|\mathbf{u}^*\|_2^2 - \|\mathbf{u}^{k+1}\|_2^2 - \left\langle \mathbf{p}^{k+1}, \mathbf{u}^* - \mathbf{u}^{k+1}\right\rangle \geq 0,
\end{aligned} \tag{38}$$

where the last inequation is by the definition of subgradient. Similarly, $\left\langle \mathbf{q}^{k+1} - \mathbf{q}^*, \mathbf{v}^{k+1} - \mathbf{v}^*\right\rangle \geq 0$.

Therefore, all terms involved in (37) are nonnegative. We have

$$\begin{aligned}
&\frac{\mu_1}{2}\left(\left\|\mathbf{b}_e^0\right\|_2^2 + \left\|\mathbf{u}_e^0\right\|_2^2\right) + \frac{\mu_2}{2}\left(\left\|\mathbf{c}_e^0\right\|_2^2 + \left\|\mathbf{v}_e^0\right\|_2^2\right) \\
&\geq \sum_{k=0}^{K}\left\|\mathbf{A}\mathbf{x}_e^{k+1}\right\|_2^2 \\
&\quad + \lambda_1\sum_{k=0}^{K}\left\langle \mathbf{p}^{k+1} - \mathbf{p}^*, \mathbf{u}^{k+1} - \mathbf{u}^*\right\rangle + \lambda_2\sum_{k=0}^{K}\left\langle \mathbf{q}^{k+1} - \mathbf{q}^*, \mathbf{v}^{k+1} - \mathbf{v}^*\right\rangle \\
&\quad + \frac{\mu_1}{2}\sum_{k=0}^{K}\left\|\mathbf{D}\mathbf{x}_e^{k+1} - \mathbf{u}_e^k\right\|_2^2 + \frac{\mu_2}{2}\sum_{k=0}^{K}\left\|\mathbf{D}\mathbf{x}_e^{k+1} - \mathbf{v}_e^k\right\|_2^2.
\end{aligned} \tag{39}$$

From (39) and $\mu_1 > 0$, $\mu_2 > 0$, $\lambda_1 > 0$, $\lambda_2 > 0$, we can get

$$\sum_{k=0}^{+\infty}\left\|\mathbf{A}\mathbf{x}_e^{k+1}\right\|_2^2 < +\infty. \tag{40}$$

Equation (40) imply that $\lim_{k\to+\infty}\|\mathbf{A}\mathbf{x}_e^k\|_2^2 = 0$, and

$$\lim_{k\to+\infty}\left\|\mathbf{A}\mathbf{x}_e^k\right\|_2^2 = \lim_{k\to+\infty}\|\mathbf{A}\mathbf{x}^k - \mathbf{A}\mathbf{x}^*\|_2^2 = \lim_{k\to+\infty}\left\langle \mathbf{A}^T(\mathbf{A}\mathbf{x}^k - \mathbf{A}\mathbf{x}^*), \mathbf{x}^k - \mathbf{x}^*\right\rangle = 0. \tag{41}$$

Recall the definition of the Bregman distance (Goldstein and Osher 2009; Cai et al. 2009), for any convex function J, we have

$$D_J^{\mathbf{p}}(\mathbf{x}, \mathbf{z}) + D_J^{\mathbf{q}}(\mathbf{z}, \mathbf{x}) = \langle \mathbf{q} - \mathbf{p}, \mathbf{x} - \mathbf{z}\rangle \tag{42}$$

where $\mathbf{p} \in \partial J(\mathbf{z})$ and $\mathbf{q} \in \partial J(\mathbf{x})$.

Now let $J(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2$, $\nabla J(\mathbf{x}) = \mathbf{A}^T(\mathbf{A}\mathbf{x} - \mathbf{y})$, from (42) we have

$$D_J^{\nabla J(\mathbf{x}^*)}(\mathbf{x}^k, \mathbf{x}^*) + D_J^{\nabla J(\mathbf{x}^k)}(\mathbf{x}^*, \mathbf{x}^k) = \left\langle \mathbf{A}^T(\mathbf{A}\mathbf{x}^k - \mathbf{A}\mathbf{x}^*), \mathbf{x}^k - \mathbf{x}^* \right\rangle = 0. \tag{43}$$

Equation (43) together with the nonnegativity of the Bregman distance, implies that $\lim_{k\to+\infty} D_J^{\nabla J(\mathbf{x}^*)}(\mathbf{x}^k, \mathbf{x}^*) = 0$, i.e.,

$$\lim_{k\to+\infty} \frac{1}{2}\|\mathbf{A}\mathbf{x}^k - \mathbf{y}\|_2^2 - \frac{1}{2}\|\mathbf{A}\mathbf{x}^* - \mathbf{y}\|_2^2 - \left\langle \mathbf{A}^T(\mathbf{A}\mathbf{x}^* - \mathbf{y}), \mathbf{x}^k - \mathbf{x}^* \right\rangle = 0. \tag{44}$$

Equation (39) also leads to

$$\sum_{k=0}^{+\infty} \left\langle \mathbf{p}^{k+1} - \mathbf{p}^*, \mathbf{u}_e^{k+1} - \mathbf{u}_e^* \right\rangle < +\infty, \tag{45}$$

and hence

$$\lim_{k\to+\infty} \left\langle \mathbf{p}^{k+1} - \mathbf{p}^*, \mathbf{u}_e^{k+1} - \mathbf{u}_e^* \right\rangle = 0. \tag{46}$$

Let $J(\mathbf{u}) = \|\mathbf{u}\|_{2,1}$ and $\mathbf{p} \in \partial\|\mathbf{u}\|_{2,1}$, we can get $\lim_{k\to+\infty} D_{\|\cdot\|_{2,1}}^{\mathbf{p}^*}(\mathbf{u}^k, \mathbf{u}^*) = 0$, i.e.,

$$\lim_{k\to+\infty} \|\mathbf{u}^k\|_{2,1} - \|\mathbf{u}^*\|_{2,1} - \left\langle \mathbf{u}_e^k - \mathbf{u}_e^*, \mathbf{p}^* \right\rangle = 0 \tag{47}$$

Similarly, let $J(\mathbf{v}) = \|\mathbf{v}\|_1$ and $\mathbf{q} \in \partial\|\mathbf{v}\|_1$ we can get

$$\lim_{k\to+\infty} \|\mathbf{v}^k\|_1 - \|\mathbf{v}^*\|_1 - \left\langle \mathbf{v}_e^k - \mathbf{v}_e^*, \mathbf{q}^* \right\rangle = 0 \tag{48}$$

Moreover, (39) also leads to $\sum_{k=0}^{+\infty} \|\mathbf{D}\mathbf{x}_e^{k+1} - \mathbf{u}_e^k\|_2^2 < +\infty$, which implies that $\lim_{k\to+\infty} \|\mathbf{D}\mathbf{x}_e^{k+1} - \mathbf{u}_e^k\|_2^2 = 0$. By $\mathbf{D}\mathbf{x}^* = \mathbf{u}^*$, we conclude that

$$\begin{aligned} \lim_{k\to+\infty} \|\mathbf{D}\mathbf{x}_e^{k+1} - \mathbf{x}_e^k\|_2^2 &= \lim_{k\to+\infty} \|\mathbf{D}\mathbf{x}^{k+1} - \mathbf{D}\mathbf{x}^* - \mathbf{u}^k + \mathbf{u}^*\|_2^2 \\ &= \lim_{k\to+\infty} \|\mathbf{D}\mathbf{x}^{k+1} - \mathbf{u}^k\|_2^2 = 0. \end{aligned} \tag{49}$$

Since $\|\cdot\|_{2,1}$ is continuous, by (47) and (49), for $\mathbf{x}$, we have

$$\lim_{k\to+\infty} \|\mathbf{D}\mathbf{x}^k\|_{2,1} - \|\mathbf{D}\mathbf{x}^*\|_{2,1} - \left\langle \mathbf{D}\mathbf{x}^k - \mathbf{D}\mathbf{x}^*, \mathbf{p}^* \right\rangle = 0. \tag{50}$$

Similarly with (49), we have $\lim_{k\to+\infty} \|\mathbf{D}\mathbf{x}^{k+1} - \mathbf{v}^k\|_2^2 = 0$ and since $\|\cdot\|_1$ is continuous, for $\mathbf{x}$, by (48), we also have

$$\lim_{k\to+\infty} \|\mathbf{D}\mathbf{x}^k\|_1 - \|\mathbf{D}\mathbf{x}^*\|_1 - \left\langle \mathbf{D}\mathbf{x}^k - \mathbf{D}\mathbf{x}^*, \mathbf{q}^* \right\rangle = 0. \tag{51}$$

Summing (44), (50) and (51)yields

$$\begin{aligned} 0 &= \lim_{k\to+\infty} \frac{1}{2}\left\|\mathbf{A}\mathbf{x}^k - \mathbf{y}\right\|_2^2 + \lambda_1 \left\|\mathbf{D}\mathbf{x}^k\right\|_{2,1} + \lambda_2 \left\|\mathbf{D}\mathbf{x}^k\right\|_1 \\ &\quad - \frac{1}{2}\left\|\mathbf{A}\mathbf{x}^* - \mathbf{y}\right\|_2^2 - \lambda_1 \left\|\mathbf{D}\mathbf{x}^*\right\|_{2,1} - \lambda_2 \left\|\mathbf{D}\mathbf{x}^*\right\|_1 \\ &\quad - \left\langle \mathbf{D}\mathbf{x}^k - \mathbf{D}\mathbf{x}^*, \lambda_1\mathbf{p}^* + \lambda_2\mathbf{q}^* + \mathbf{A}^T(\mathbf{A}\mathbf{x}^* - \mathbf{y})\right\rangle \\ &= \lim_{k\to+\infty} \frac{1}{2}\left\|\mathbf{A}\mathbf{x}^k - \mathbf{y}\right\|_2^2 + \lambda_1 \left\|\mathbf{D}\mathbf{x}^k\right\|_{2,1} + \lambda_2 \left\|\mathbf{D}\mathbf{x}^k\right\|_1 \\ &\quad - \frac{1}{2}\left\|\mathbf{A}\mathbf{x}^* - \mathbf{y}\right\|_2^2 - \lambda_1 \left\|\mathbf{D}\mathbf{x}^*\right\|_{2,1} - \lambda_2 \left\|\mathbf{D}\mathbf{x}^*\right\|_1. \end{aligned} \tag{52}$$

where the last equality comes from (25). So (18) is proved.

Next, we prove (19) whenever (3) has a unique solution. It is proved by contradiction.

Define $\Phi(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda_1\|\mathbf{D}\mathbf{x}\|_{2,1} + \lambda_2\|\mathbf{D}\mathbf{x}\|_1$. Then $\Phi(\mathbf{x})$ is convex and lower semi-continuous. Since $\mathbf{x}^*$ is the unique minimizer, we have $\Phi(\mathbf{x}) > \Phi(\mathbf{x}^*)$ for $\mathbf{x} \neq \mathbf{x}^*$.

Now, suppose that (19) does not hold. So, there exists a subsequence $\mathbf{x}^{k_i}$ such that $\|\mathbf{x}^{k_i} - \mathbf{x}^*\|_2 > \epsilon$ for some $\epsilon > 0$ and all i. Then, $\Phi(\mathbf{x}^{k_i}) > \min\{\Phi(\mathbf{x}) : \|\mathbf{x}^{k_i} - \mathbf{x}^*\|_2 = \epsilon\}$. Let $\mathbf{z}$ be the intersection of $\{\mathbf{z} : \|\mathbf{z} - \mathbf{z}^*\|_2 = \epsilon\}$ and the line from $\mathbf{x}^*$ to $\mathbf{x}^{k_i}$. Indeed, there exists a positive number $t \in (0, 1)$ such that $\mathbf{z} = t\mathbf{x}^* + (1-t)\mathbf{x}^{k_i}$. By the convexity of Φ and the definition of $\mathbf{x}^*$, we have

$$\begin{aligned} \Phi(\mathbf{x}^{k_i}) &\geq t\Phi(\mathbf{x}^*) + (1-t)\Phi(\mathbf{x}^{k_i}) \geq \Phi(t\mathbf{x}^* + (1-t)\mathbf{x}^{k_i}) = \Phi(\mathbf{z}) \\ &\geq \min\{\Phi(\mathbf{x}) : \|\mathbf{x}^{k_i} - \mathbf{x}^*\|_2 = \epsilon\}. \end{aligned} \tag{53}$$

Denote $\tilde{\mathbf{x}} = \min\{\Phi(\mathbf{x}) : \|\mathbf{x}^{k_i} - \mathbf{x}^*\|_2 = \epsilon\}$. By (20), we have

$$\Phi(\mathbf{x}^*) = \lim_{k\to+\infty} \Phi(\mathbf{x}^{k_i}) \geq \Phi(\tilde{\mathbf{x}}) > \Phi(\mathbf{x}^*), \tag{54}$$

which is a contradiction. So (19) holds whenever (5) has a unique solution. □

References

Bilen, C., Wang, Y., & Selesnick, I. W. (2012). High-speed compressed sensing reconstruction in dynamic parallel MRI using augmented lagrangian and parallel processing. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, *2*(3), 370–379.

Bruckstein, A. M., Donoho, D. L., & Elad, M. (2009). From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review*, *51*(1), 34–81.

Cai, J. F., Osher, S., & Shen, Z. (2009). Split Bregman methods and frame based image restoration. *Multiscale Modeling & Simulation*, *8*(2), 337–369.

Candès, E. J., & Wakin, M. B. (2008). An introduction to compressive sampling. *IEEE Signal Processing Magazine*, *25*(2), 21–30.

Chatterjee, S., Banerjee, A., & Ganguly, A. (2011). Sparse group lasso for regression on land climate variables. In: 2011 IEEE 11th international conference on data mining workshops (ICDMW) (pp. 1–8). IEEE.

Deng, W., Yin, W., & Zhang, Y. (2011). *Group sparse optimization by alternating direction method*. TR11-06, Department of Computational and Applied Mathematics, Rice University.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). A note on the group lasso and a sparse group lasso. arXiv, preprint arXiv:10010736.

Goldstein, T., & Osher, S. (2009). The split Bregman method for ℓ_1-regularized problems. *SIAM Journal on Imaging Sciences*, *2*(2), 323–343.

Goldstein, T., Bresson, X., & Osher, S. (2010). Geometric applications of the split Bregman method: segmentation and surface reconstruction. *Journal of Scientific Computing*, *45*(1–3), 272–293.

He, Z., Xie, S., Ding, S., & Cichocki, A. (2007). Convolutive blind source separation in the frequency domain based on sparse representation. *IEEE Transactions on Audio, Speech, and Language Processing*, *15*(5), 1551–1563.

He Z, Cichocki, A., Zdunek, R., & Cao, J. (2008). CG-M-FOCUSS and its application to distributed compressed sensing. In: *Advances in Neural Networks-ISNN 2008* (pp. 237–245). IEEE.

He, Z., Cichocki, A., Li, Y., Xie, S., & Sanei, S. (2009a). K-hyperline clustering learning for sparse component analysis. *Signal Processing*, *89*(6), 1011–1022.

He, Z., Cichocki, A., Zdunek, R., & Xie, S. (2009b). Improved FOCUSS method with conjugate gradient iterations. *IEEE Transactions on Signal Processing*, *57*(1), 399–404.

Huang, J., & Zhang, T. (2010). The benefit of group sparsity. *The Annals of Statistics*, *38*(4), 1978–2004.

Jiang, L., & Yin, H. (2012). Bregman iteration algorithm for sparse nonnegative matrix factorizations via alternating ℓ_1-norm minimization. *Multidimensional Systems and Signal Processing*, *23*(3), 315–328.

Lee, D. H., Hong, C. P., & Lee, M. W. (2013a). Sparse magnetic resonance imaging reconstruction using the Bregman iteration. *Journal of the Korean Physical Society*, *62*(2), 328–332.

Lee, D. H., Hong, C. P., Lee, M. W., & Han, B. S. (2013b). Rapid 2D phase-contrast magnetic resonance angiography reconstruction algorithm via compressed sensing. *Journal of the Korean Physical Society*, *63*(5), 1072–1076.

Liu, B., King, K., Steckner, M., Xie, J., Sheng, J., & Ying, L. (2009a). Regularized sensitivity encoding (SENSE) reconstruction using bregman iterations. *Magnetic Resonance in Medicine*, *61*(1), 145–152.

Liu, J., Ji, S., & Ye, J. (2009b). *SLEP: Sparse learning with efficient projections*. Arizona State University.

Lustig, M., Donoho, D., & Pauly, J. M. (2007). Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, *58*(6), 1182–1195.

Lustig, M., Donoho, D. L., Santos, J. M., & Pauly, J. M. (2008). Compressed sensing MRI. *IEEE Signal Processing Magazine*, *25*(2), 72–82.

Ma, S., Goldfarb, D., & Chen, L. (2011). Fixed point and Bregman iterative methods for matrix rank minimization. *Mathematical Programming*, *128*(1–2), 321–353.

Meier, L., Van De Geer, S., & Bühlmann, P. (2008). The group lasso for logistic regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*(1), 53–71.

Simon, N., Friedman, J., Hastie, T., & Tibshirani, R. (2013). A sparse-group lasso. *Journal of Computational and Graphical Statistics*, *22*(2), 231–245.

Smith, D. S., Gore, J. C., Yankeelov, T. E., & Welch, E. B. (2012). Real-time compressive sensing MRI reconstruction using GPU computing and split bregman methods. *International Journal of Biomedical Imaging*, *2012*, 1–6.

Stone, S. S., Haldar, J. P., Tsao, S. C., Wm, Hwu, Sutton, B. P., Liang, Z. P., et al. (2008). Accelerating advanced MRI reconstructions on GPUs. *Journal of Parallel and Distributed Computing*, *68*(10), 1307–1318.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B (Methodological)*, pp. 267–288.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612.

Xu, J., Feng, X., & Hao, Y. (2014). A coupled variational model for image denoising using a duality trategy and split Bregman. *Multidimensional Systems and Signal Processing*, *25*(1), 83–94.

Ye, G. B., & Xie, X. (2011). Split Bregman method for large scale fused lasso. *Computational Statistics & Data Analysis*, *55*(4), 1552–1569.

Yin, W., Osher, S., Goldfarb, D., & Darbon, J. (2008). Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing. *SIAM Journal on Imaging Sciences*, *1*(1), 143–168.

Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *68*(1), 49–67.

Zhao, P., & Yu, B. (2006). On model selection consistency of lasso. *The Journal of Machine Learning Research*, *7*, 2541–2563.

Zhu, X., Huang, Z., Cui, J., & Shen, H. (2013). Video-to-shot tag propagation by graph sparse group lasso. *IEEE Transactions on Multimedia*, *15*(3), 633–646.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, *101*(476), 1418–1429.

Zou, J., Fu, Y., & Xie, S. (2012). A block fixed point continuation algorithm for block-sparse reconstruction. *IEEE Signal Processing Letters*, *19*(6), 364–367.

Jian Zou received his Ph.D. degree in signal and information processing from the South China University of Technology in 2013. Now, he is an Assistant Professor in Yangtze University. His research activities are focused on adaptive signal processing and numerical optimization.

Yuli Fu received his Ph.D. degree in control theory and engineering from the Huazhong University of Science and Technology, China, in 2000. From 2000 to 2002, he was a postdoctoral researcher in the School of Electronic and Information Engineering at the South China University of Technology. Now, he is a Professor in the same University. His research activities are focused on adaptive signal processing and artificial intelligent.