NONCONVEX COMPRESSIVE SENSING RECONSTRUCTION FOR TENSOR USING STRUCTURES IN MODES

Xin Ding[⋆] *Wei Chen*[†⋆] *Ian Wassell*[⋆]

[⋆] Computer Laboratory, University of Cambridge, UK
[†] State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, China

ABSTRACT

This paper focuses on the reconstruction of a tensor captured using Compressive Sensing (CS). Instead of processing the signals via vectorization as is done in conventional CS, in tensor CS high dimensional signals are kept in their original formats, which benefits hardware implementation and eases memory requirements. In addition, more structures exist in a tensor along its various dimensions than in its vectorized format. Utilizing these various structures, this paper proposes a general reconstruction approach for tensor CS. Employing the proximity operator of a nonconvex norm function, a special case for a tensor with low rank and sparse structures is elaborated, which is shown to outperform the state-of-art tensor CS reconstruction methods when applied to magnetic resonance imaging and hyper-spectral imaging.

***Index Terms*—** Compressive sensing, tensor reconstruction, sparse and low rank reconstruction.

1. INTRODUCTION

Compressive Sensing (CS) [1, 2, 3], which achieves sensing and compression at the same time, is drawing increasing attention for data acquisition. It relies upon the fact that most practical signals can be represented with only a few non-zero coefficients (i.e., the signal is sparse) in a suitable basis and guarantees the reconstruction of such signals from far fewer samples than that required by the Nyquist sampling theorem.

However, it is challenging to apply CS in modern applications, which often generate signals with multidimensional structures, e.g., Magnetic Resonance Imaging (MRI) and Hyper-Spectral Imaging (HSI). In conventional CS, a multidimensional signal, i.e., a tensor, is mapped in vector format for sensing and reconstruction. Such a vectorized tensor in real-world applications requires an enormous CS sensing matrix, which imposes large demands in terms of processing power and memory size. In addition, sensing the vectorized signal is equivalent to simultaneously multiplexing along all data dimensions, which is difficult to achieve in hardware. Besides, a tensor usually possesses structures along its various dimensions; while the vectorization results in a loss of structure.

In the past few years, extension of CS to 2D signals, i.e., matrices, has been made [4, 5, 6, 7], where a rank deficient model is involved rather than simply assuming a sparse model. However, research concerning CS for tensors is still quite rare. In [8], Kronecker product matrices are proposed to form the CS sensing matrix, thereby the sensing procedure is partitioned along dimensions. The theory presented in this paper paves the way to tensor CS. Using the canonical polyadic decomposition model of a tensor, in [9, 10], a framework for recovering tensors that have sparse factor matrices is proposed. A different approach, based on the Tucker decomposition model that is more intuitively related to practical CS, multidimensional CS is proposed in [11, 12]. The authors assume a sparse Tucker core tensor and propose the Kronecker Orthogonal Matching Pursuit (KOMP) algorithm to reconstruct a tensor from the Kronecker CS samples. It is also extended to exploit the block sparse structure and a Block-based KOMP algorithm is proposed. Other related work lies in the field of tensor completion, where a low rank tensor model is assumed [13, 14]. However, all the aforementioned work utilizes either a sparse or a low rank model for the whole tensor, which does not fully exploit the various structures present in a high dimensional signal.

In this paper, we propose an approach to CS tensor reconstruction, in which diverse structures along each dimension (i.e., mode) of a tensor are exploited. Even though simultaneous structures have been considered for matrices, e.g., the matrix signal is both rank deficient and sparse or group sparse [7, 15, 16], their approaches are not applicable to our problem because they do not separately sense and reconstruct the signals along dimensions using the Kronecker structures, which means that their implementations are limited by the large data size of a tensor. In some other work [17, 18], dictionary learning techniques are extended to tensors to take advantage of structures in modes. However, to the best of our knowledge, there is no prior work focused on exploiting various structures in the tensor modes in the reconstruction process of tensor CS. We solve the proposed multi-structure optimization problem using an approach based upon the Alternating Di-

This work is supported by EPSRC Research Grant EP/K033700/1; the Natural Science Foundation of China (61401018, U1334202); the Fundamental Research Funds for the Central Universities (No. 2014JBM149); the State Key Laboratory of Rail Traffic Control and Safety (RCS2016ZT014) of Beijing Jiaotong University; the Key Grant Project of Chinese Ministry of Education (313006).

rection Method of Multipliers (ADMM) [19], in which we employ nonconvex reconstruction because the convex relaxation commonly used in CS is shown to not perform well in the case of simultaneous structures [20]. The derived algorithm overcomes the difficulty of high memory requirements and computation loads encountered for tensor reconstruction in conventional CS and ADMM, and it provides the flexibility of involving various structures, thereby improving the reconstruction accuracy.

1.1. Multi-linear Algebra and Notations

A multidimensional array $\underline{\mathbf{X}} \in \mathbb{R}^{N_1 \times \ldots \times N_n}$ is known as a mode-n tensor, where the dimension of its i-th mode is N_i. The mode-i vectors are determined by fixing every index except the one in the mode i. The mode-i unfolding matrix $\mathbf{X}_{(i)} \in \mathbb{R}^{N_i \times N_1 \ldots N_{i-1} N_{i+1} \ldots N_n}$ is obtained by arranging all the mode-i vectors as columns of a matrix. Given a matrix $\mathbf{A} \in \mathbb{R}^{J \times N_k}$, the mode-$k$ tensor by matrix product is defined as $\underline{\mathbf{Z}} = \underline{\mathbf{X}} \times_k \mathbf{A}$, where $\underline{\mathbf{Z}} \in \mathbb{R}^{N_1 \times \ldots \times N_{k-1} \times J \times N_{k+1} \times \ldots \times N_n}$ and it is calculated by: $\underline{\mathbf{Z}} = fold_i(\mathbf{A}\mathbf{X}_{(i)})$, where $fold_i(\cdot)$ is an operator that folds up a matrix along mode i to a tensor. The Kronecker and Hadamard product between two matrices are denoted by $\mathbf{A} \otimes \mathbf{B}$ and $\mathbf{A} \circ \mathbf{B}$, respectively. The Kronecker product of n matrices is denoted by $\overline{\mathbf{A}}$, where $\overline{\mathbf{A}} = \mathbf{A}_n \otimes \ldots \otimes \mathbf{A}_1$. The l_p norm of a vector is defined as: $||\mathbf{x}||_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$ and the l_0 norm is given by the number of nonzero entries. The operator $vec(\cdot)$, $ten(\cdot)$, $diag(\cdot)$ and $svd(\cdot)$ denote vectorization, tensorization, taking a matrix diagonal and singular value decomposition, respectively.

2. TENSOR COMPRESSIVE SENSING

Extending the sensing model in CS, a tensor $\underline{\mathbf{X}} \in \mathbb{R}^{N_1 \times \ldots \times N_n}$ is sampled by:

$$\underline{\mathbf{Y}} = \underline{\mathbf{X}} \times_1 \mathbf{\Phi}_1 \times_2 \mathbf{\Phi}_2 \ldots \times_n \mathbf{\Phi}_n, \tag{1}$$

where $\underline{\mathbf{Y}} \in \mathbb{R}^{M_1 \times \ldots M_n}$ is the measurement, $\mathbf{\Phi}_i \in \mathbb{R}^{M_i \times N_i}$ $(i = 1, \ldots, n)$ are sensing matrices and $M_i < N_i$. A tensor is regarded as K sparse when it can be represented as:

$$\underline{\mathbf{X}} = \underline{\mathbf{S}} \times_1 \mathbf{\Psi}_1 \times_2 \mathbf{\Psi}_2 \ldots \times_n \mathbf{\Psi}_n, \tag{2}$$

where $\mathbf{\Psi}_i \in \mathbb{R}^{N_i \times N_i}$ $(i = 1, \ldots, n)$ are the sparsifying basis, e.g., a Discrete Wavelet Transform (DWT), $\underline{\mathbf{S}} \in \mathbb{R}^{N_1 \times \ldots N_n}$ is the sparse representation which has only K non-zero coefficients. Then the sensing procedure is equivalent to:

$$\underline{\mathbf{Y}} = \underline{\mathbf{S}} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \ldots \times_n \mathbf{A}_n, \tag{3}$$

where $\mathbf{A}_i = \mathbf{\Phi}_i \mathbf{\Psi}_i$ $(i = 1, \ldots, n)$ are the equivalent sensing matrices.

To see the connection of tensor CS to conventional CS, equation (3) has been shown in [12] to be equivalent to:

$$\mathbf{y} = (\mathbf{A}_n \otimes \mathbf{A}_{n-1} \otimes \ldots \otimes \mathbf{A}_1)\mathbf{s}, \tag{4}$$

where $\mathbf{y}$ and $\mathbf{s}$ are the vectorized $\underline{\mathbf{Y}}$ and $\underline{\mathbf{S}}$, respectively. By using the notation $\overline{\mathbf{A}}$, it becomes a conventional CS model.

To reconstruct $\underline{\mathbf{S}}$ or $\underline{\mathbf{X}}$ from the measurement $\underline{\mathbf{Y}}$ is an underdetermined problem and it is modeled in CS as a l_0 minimization problem as follows:

$$min\ ||\mathbf{s}||_0\ s.t.\ \underline{\mathbf{Y}} = \underline{\mathbf{S}} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \ldots \times_n \mathbf{A}_n, \tag{5}$$

which is NP-hard and is normally solved by replacing the l_0 norm with a l_1 norm. The reconstruction is guaranteed when the sensing matrix $\overline{\mathbf{A}}$ obeys the Restricted Isometry Property (RIP) [1]. The relationship between the RIP of $\overline{\mathbf{A}}$ and each component $\mathbf{A}_i$ has been analyzed in [8]. Nonconvex optimization, where the l_0 norm in (5) is replaced by a l_p $(0 < p < 1)$ quasi-norm, has been shown [21, 22] to be capable of further reducing the required number of samples for reconstruction, while still being tractable.

Tensor CS enlarges the dimension of the signals that can be processed in a conventional CS system. Specifically, in conventional CS, a n-mode tensor $\underline{\mathbf{X}} \in \mathbb{R}^{N \times \ldots \times N}$ is vectorized as $\mathbf{x} \in \mathbb{R}^{N^n}$, which requires a global multiplexing with a sensing matrix of size $M^n \times N^n$; while in tensor CS, such a sensing matrix is partitioned as n smaller sensing matrices of size $M \times N$, which reduces the memory storage requirements, and also eases hardware implementation. In addition, extending CS reconstruction algorithms to utilize the Kronecker structure reduces the computational complexity. This is primarily because the complexity involved in operations such as $\mathbf{y} - \overline{\mathbf{A}}\mathbf{s}$ is $\mathcal{O}(M^n N^n + M^n)$; while that for $\underline{\mathbf{Y}} - \underline{\mathbf{S}} \times_1 \mathbf{A}_1 \ldots \times_n \mathbf{A}_n$ is $\mathcal{O}[MN^n(1 - M^n/N^n)/(1 - M/N) + M^n]$, which is always smaller than the former.

3. TENSOR CS RECONSTRUCTION USING STRUCTURES IN MODES

High dimensional signals often possess various structures along its dimensions, e.g., sparse, low rank, row sparse, column sparse. Utilizing these structures simultaneously can reduce the degrees of freedom of the reconstruction problem and consequently improve the reconstruction performance [20]. The reconstruction problem that provides the flexibility of exploiting various structures is formulated as:

$$\begin{aligned} \min_{\underline{\mathbf{X}}} & \sum_{i=1}^{n} \alpha_i ||\mathbf{\Omega}_i \mathbf{X}_{(i)}||_{(i-th\ norm)} \\ s.t.\ & \underline{\mathbf{Y}} = \underline{\mathbf{X}} \times_1 \mathbf{\Phi}_1 \ldots \times_n \mathbf{\Phi}_n, \end{aligned} \tag{6}$$

where $\mathbf{\Omega}_i$ denotes an orthogonal transform for the i-th mode, $||\cdot||_{(i-th\ norm)}$ is a general form of norms that can be defined specifically according to the structure of the $i-$th mode and α_i $(i = 1, \ldots, n)$ are the tunning parameters. This problem can be rewritten as:

$$\begin{aligned} \min_{\underline{\mathbf{X}}, \underline{\mathbf{Z}_1}, \ldots, \underline{\mathbf{Z}_n}} & \sum_{i=1}^{n} \alpha_i ||\mathbf{\Omega}_i \mathbf{Z}_{i(i)}||_{(i-th\ norm)} \\ s.t.\ & \underline{\mathbf{Y}} = \underline{\mathbf{X}} \times_1 \mathbf{\Phi}_1 \ldots \times_n \mathbf{\Phi}_n, \\ & \underline{\mathbf{X}} = \underline{\mathbf{Z}_i},\ i = 1, \ldots, n, \end{aligned} \tag{7}$$

where $\underline{\mathbf{Z}_i}$ $(i = 1, \ldots, n)$ are intermediate variables.

We then propose to solve this problem using the Alternating Direction Method of Multipliers (ADMM) [19]. However, since large dimensions are involved in the problem, the original ADMM becomes impractical as it leads to high storage requirements and a significant computational burden. We thus derive efficient calculations in the ADMM steps to enable implementation. The augmented Lagrangian function of

(7) is defined as:

$$\begin{aligned} &L_\rho(\underline{\mathbf{X}}, \underline{\mathbf{Z}}_1, ..., \underline{\mathbf{Z}}_n, \underline{\mathbf{P}}_1, ..., \underline{\mathbf{P}}_n, \underline{\mathbf{Q}}) \\ =&\sum_{i=1}^{n}(\alpha_i||\mathbf{\Omega}_i\mathbf{Z}_{i(i)}||_{(i-th\ norm)} + < \mathbf{\Omega}_i(\underline{\mathbf{X}} - \underline{\mathbf{Z}}_i),\ \mathbf{\Omega}_i\underline{\mathbf{P}}_i > \\ &+ \frac{\rho}{2}||\mathbf{\Omega}_i(\underline{\mathbf{Z}}_i - \underline{\mathbf{X}})||_F^2) + < \underline{\mathbf{Y}} - \underline{\mathbf{X}} \times_1 \mathbf{\Phi}_1 ... \times_n \mathbf{\Phi}_n,\ \underline{\mathbf{Q}} > \\ &+ \frac{\rho}{2}||\underline{\mathbf{Y}} - \underline{\mathbf{X}} \times_1 \mathbf{\Phi}_1 ... \times_n \mathbf{\Phi}_n||_F^2, \end{aligned} \tag{8}$$

in which $\rho > 0$, and $\mathbf{P}$ and $\mathbf{Q}$ are dual variables. The ADMM then proceeds by iteratively updating $\underline{\mathbf{Z}}_i$s ($i = 1, ..., n$), $\underline{\mathbf{X}}$, $\underline{\mathbf{P}}_i$s ($i = 1, ..., n$) and $\underline{\mathbf{Q}}$ with the others fixed.

Updating each $\underline{\mathbf{Z}}_i$ for $i = 1, ..., n$ yields:

$$\begin{aligned} \underline{\mathbf{Z}}_i{}^{t+1} =& min\ \alpha_i||\mathbf{\Omega}_i\mathbf{Z}_{i(i)}||_{(i-th\ norm)} \\ &+ \frac{\rho}{2}||\mathbf{\Omega}_i\underline{\mathbf{Z}}_i - \mathbf{\Omega}_i(\frac{1}{\rho}\underline{\mathbf{P}}_i{}^t + \underline{\mathbf{X}})||_F^2 \\ =&\mathbf{\Omega}_i^T prox_{\frac{\alpha_i}{\rho},\ (i-th\ norm)}[\mathbf{\Omega}_i(\frac{1}{\rho}\underline{\mathbf{P}}_i{}^t + \underline{\mathbf{X}}^t)], \end{aligned} \tag{9}$$

where $prox(\cdot)$ denotes the proximity operator [23] and is defined for function f as: $prox_{\frac{1}{\rho},\ f}(x) = arg\ min_y f(y) + \frac{\rho}{2}||x - y||_2^2$. The proximity for many widely used functions has been derived in [23, 24, 25], including the l_p norm, the log-penalty and the $l_{p,q}$ norm. The obtained proximity operators for these functions are element-wise thresholding functions, which are simple to compute even for high dimensional data. To take advantage of the various structures in tensor modes, one can define the transform operator $\mathbf{\Omega}_i$ and the i-th norm specifically according to the applications, provided that the proximity for the norm can be calculated.

With $\underline{\mathbf{Z}}_i$ ($i = 1, ..., n$) determined, we then update $\underline{\mathbf{X}}$ by minimizing L_ρ, which yields:

$$\begin{aligned} \underline{\mathbf{X}}^{t+1} = (n\mathbf{I} + \overline{\mathbf{\Phi}}^T\overline{\mathbf{\Phi}})^{-1}\{vec[\sum_{i=1}^{n}(\underline{\mathbf{Z}}_i - \frac{1}{\rho}\underline{\mathbf{P}}_i) \\ + (\frac{1}{\rho}\underline{\mathbf{Q}} + \underline{\mathbf{Y}}) \times_1 \mathbf{\Phi}_1^T ... \times_n \mathbf{\Phi}_n^T]\}. \end{aligned} \tag{10}$$

We observe that in equation (10), the term $(n\mathbf{I} + \overline{\mathbf{\Phi}}^T\overline{\mathbf{\Phi}})$ has the size $N_1N_2...N_n \times N_1N_2...N_n$, which is not feasible to be processed by a conventional computing device and worse, its inverse is even more difficult to calculate having a complexity of $\mathcal{O}[(N_1N_2...N_n)^3]$.

To keep the advantage of low memory requirement in tensor CS, we rewrite the term $(n\mathbf{I} + \overline{\mathbf{\Phi}}^T\overline{\mathbf{\Phi}})^{-1}$ as:

$$(\mathbf{U}_n \otimes ... \otimes \mathbf{U}_1)(n\mathbf{I} + \mathbf{V}_n \otimes ... \otimes \mathbf{V}_1)^{-1}(\mathbf{U}_n^T \otimes ... \otimes \mathbf{U}_1^T), \tag{11}$$

where $\mathbf{U}_i$ and $\mathbf{V}_i$ comes from the eigendecompositions of $\mathbf{\Phi}_i^T\mathbf{\Phi}_i$, i.e., $\mathbf{U}_i\mathbf{V}_i\mathbf{U}_i^T = \mathbf{\Phi}_i^T\mathbf{\Phi}_i$, and the matrices $\mathbf{V}_i$ ($i = 1..., n$) are diagonal. Therefore the term $(n\mathbf{I} + \mathbf{V}_n \otimes ... \otimes \mathbf{V}_1)$ is also diagonal and its inverse can be easily obtained. The computational complexity is now reduced to that of the eigendecompositions, i.e., $\mathcal{O}(\sum_{i=1}^{n} N_i^3)$ and they only need to be calculated once before the iterations start. To avoid the storage of the large matrix in (11), we partition it into small matrices and update $\underline{\mathbf{X}}$ by:

$$\underline{\mathbf{X}}^{t+1} = [ten(\mathbf{v}) \circ (\underline{\mathbf{G}} \times_1 \mathbf{U}_1^T ... \times_n \mathbf{U}_n^T)] \times_1 \mathbf{U}_1 ... \times_n \mathbf{U}_n, \tag{12}$$

where $\mathbf{v} = diag[(n\mathbf{I} + \mathbf{V}_n \otimes ... \otimes \mathbf{V}_1)^{-1}]$, $\underline{\mathbf{G}} = \sum_{i=1}^{n}(\underline{\mathbf{Z}}_i - \frac{1}{\rho}\underline{\mathbf{P}}_i) + (\frac{1}{\rho}\underline{\mathbf{Q}} + \underline{\mathbf{Y}}) \times_1 \mathbf{\Phi}_1^T ... \times_n \mathbf{\Phi}_n^T$. In this way, the complexity of the matrix-vector multiplication in (10) is also reduced, as discussed in Section 2.

In the final steps, the dual variables are updated by:

$$\underline{\mathbf{P}}_i{}^{t+1} = \underline{\mathbf{P}}_i{}^t - \rho(\underline{\mathbf{Z}}_i{}^{t+1} - \underline{\mathbf{X}}^{t+1}), \tag{13}$$

$$\underline{\mathbf{Q}}^{t+1} = \underline{\mathbf{Q}}^t - \rho(\underline{\mathbf{X}}^{t+1} \times_1 \mathbf{\Phi}_1 ... \times_n \mathbf{\Phi}_n - \underline{\mathbf{Y}}). \tag{14}$$

The convergence of convex ADMM has been studied in [19] and a recent work also analyzes the nonconvex and smooth case [26]. The convergence issue for a nonconvex and nonsmooth case is still an open question, but it has been empirically shown to converge reliably for certain cases [6].

4. NONCONVEX CS RECONSTRUCTION FOR TENSORS WITH LOW RANK AND SPARSE MODES

In this section, we consider a special case when a tensor has low rank and sparse modes, i.e., $rank(\mathbf{X}_{(i)}) < r_i$ for some of the modes i and $supp(\mathbf{\Omega}_i\mathbf{X}_{(i)}) < K$ for the others. Such models widely exist in practical applications, e.g., MRI, HSI, video processing and sub-wavelength optical imaging [27].

Here, we employ nonconvex functions to model the sparsity and rank deficiency. Specifically, we adopt the proximal l_p ($p \leq 1$) norm (denoted by $G_{\rho,p}$) as defined in [6], of which the proximity has been derived as a p-shrinkage operator that functions element-wise as follows:

$$shrink_p(t,\ \mu) = max\{0,\ |t| - \mu|t|^{p-1}\}\frac{t}{|t|}. \tag{15}$$

The optimization problem then becomes:

$$\begin{aligned} \min_{\underline{\mathbf{X}}}\ &\sum_{i\in\mathcal{S}}\alpha_i[G_{\rho,p}(\mathbf{\Omega}_i\mathbf{X}_{(i)})] + \sum_{i\in\mathcal{L}}\alpha_i\{G_{\rho,p}[\sigma(\mathbf{X}_{(i)})]\} \\ &s.t.\ \underline{\mathbf{Y}} = \underline{\mathbf{X}} \times_1 \mathbf{\Phi}_1 ... \times_n \mathbf{\Phi}_n, \end{aligned} \tag{16}$$

where $\sigma(\mathbf{X}_{(i)})$ is the vector of singular values of $\mathbf{X}_{(i)}$, the sets $\mathcal{L}$ and $\mathcal{S}$ denote the modes that are low-rank or sparse, respectively. Follow the process in Section 3, the Tensor ADMM algorithm (TADMM) is summarized in Algorithm 1.

Algorithm 1 TADMM

Input: $\underline{\mathbf{Y}}$, $\mathbf{\Phi}_i$, $\mathbf{\Omega}_i$, $\underline{\mathbf{X}}^0$, $\underline{\mathbf{P}}_i{}^0$, $\underline{\mathbf{Q}}^0$, $\underline{\mathbf{Z}}_i{}^0$, $\mathcal{L}$, $\mathcal{S}$, p, ρ, α_i, ($i = 1, ...n$).
Output: $\underline{\mathbf{X}}$.

Repeat
  For $i = 1$ to n **do**
    if $i \in \mathcal{L}$
      $\mathbf{\Gamma}\mathbf{\Lambda}\mathbf{W}^T = svd(\frac{1}{\rho}\mathbf{P}_{i(i)}^t + \mathbf{X}_{(i)}^t)$;
      $\underline{\mathbf{Z}}_i{}^{t+1} = fold_i\{\mathbf{\Gamma} shrink_p[diag(\mathbf{\Lambda}),\ \frac{\alpha_i}{\rho}]\mathbf{W}^T\}$;
    else
      $\underline{\mathbf{Z}}_i{}^{t+1} = shrink_p(\frac{1}{\rho}\underline{\mathbf{P}}_i{}^t + \underline{\mathbf{X}}^t,\ \frac{\alpha_i}{\rho})$;
    end
  end for
  Calculate $\underline{\mathbf{X}}^{t+1}$ using (12);
  Update $\underline{\mathbf{P}}_i{}^{t+1}$ ($i = 1, ..., n$) using (13);
  Update $\underline{\mathbf{Q}}^{t+1}$ using (14).
Until a stopping criteria is met at iteration t.

5. SIMULATION RESULTS

In this section, we evaluate the proposed approach, i.e., TADMM, on 3-mode tensors with low rank and sparse modes in both MRI and HSI applications. In both experiments, we compare TADMM with the state-of-art tensor CS algorithms, which include the sparsity based methods: Kronecker CS

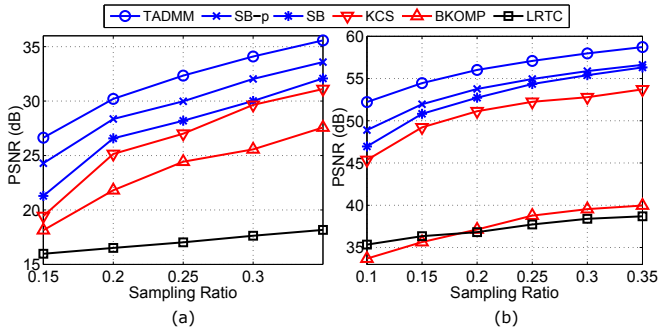

Fig. 1. PSNRs when various sampling ratios are employed for (a) MRI; (b) HSI. ($p = 0.5$). The blue, red and black curves represent the sparse and low rank based methods, sparse based methods and low rank based method.

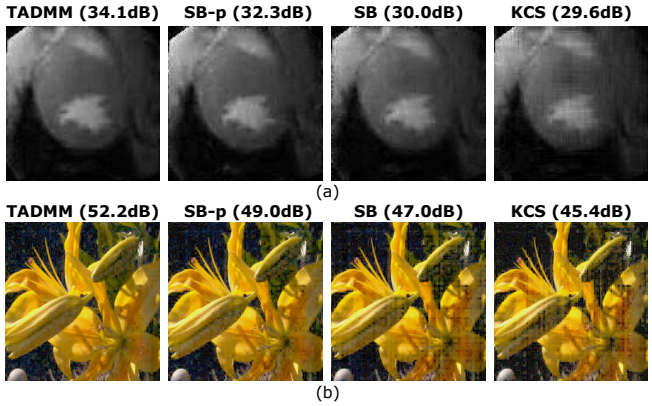

Fig. 2. Reconstruction of (a) MRI (100-th slice; sampling ratio: 0.3, $p = 0.5$); (b) HSI (32 frequency channels are transformed to RGB display; sampling ratio: 0.1, $p = 0.5$).

(KCS) [8] and Block sparse KOMP (BKOMP) [11]; the low rank based method: Low Rank Tensor Completion (LRTC) [13]; and the sparse and low rank based method: Split Bregman algorithm (SB) [15]. For the SB algorithm, as it was originally designed for vectorized signals, we modify it to utilize the tensor Kronecker product so that it can be applied to large dimensional signals. We also extend it to utilize the proximal p-norm so that it becomes a nonconvex method, denoted by SB-p. We employ the SPGL1 solver [28] for the KCS method.

In both MRI and HSI experiments, for the TADMM algorithm, mode 1 and 2 of the tensor are sparse in a 4 level DWT basis and mode 3 is low rank in the original domain. A 3D 4 level DWT is utilized for the sparsity based methods. The sensing matrices $\mathbf{\Phi}_1$, $\mathbf{\Phi}_2$ are set as i.i.d Gaussian sensing matrices with various sample ratios and $\mathbf{\Phi}_3$ is an identity matrix. To initialize the ADMM algorithm, we set $\underline{\mathbf{X}}^0 = \underline{\mathbf{P}_i}^0 = \underline{\mathbf{Z}_i}^0 = \mathbf{0}$ $(i = 1, ..., n)$ and $\underline{\mathbf{Q}}^0 = \underline{\mathbf{Y}}$. The parameters are chosen empirically as $\alpha_1 = \alpha_2 = \alpha_3 = 1$, $\rho^0 = 1e-4$, $\rho^{t+1} = 1.15\rho^t$ [13] (increasing ρ gradually improves the rate of convergence). The stopping criteria for all the methods are as follows: 1) the normalized consecutive change on $\underline{\mathbf{X}}$ is small; 2) the normalized residue $\underline{\mathbf{R}} = \underline{\mathbf{Y}} - \underline{\mathbf{X}} \times_1 \mathbf{\Phi}_1 ... \times_n \mathbf{\Phi}_n$ is small; 3) the normalized consecutive change on $\underline{\mathbf{R}}$ is small.

Table 1. Accuracy and Running Time Comparison when sampling ratio is 0.2.

Algorithm	PSNR (dB)	Time (h)
TADMM ($p = 0.2$)	56.22	1.73
TADMM ($p = 0.7$)	55.64	1.69
TADMM ($p = 1$)	53.95	1.39
SB ($p = 0.2$)	53.51	1.74
SB ($p = 0.7$)	53.84	1.70
SB ($p = 1$)	52.73	1.35
KCS	51.13	25.1
BKOMP	37.13	0.77
LRTC	36.82	0.54

The Peak Signal to Noise Ratio (PSNR) is employed for the evaluations. The experiments are conducted on a Macbook pro with a 2.6GHz Intel Core i5 CPU and 8GB RAM.

We employ a cardiac sequence [29] with size $128 \times 128 \times 128$ for the MRI experiments and Fig. 1(a) shows the PSNR results when various sampling ratios (i.e., $M_1M_2/(N_1N_2)$) are used. The superiority of the proposed approach is clear. The BKOMP and LRTC methods perform poorly because of model mismatch. The reconstructed 100-th slice of the best four approaches when the sampling ratio is 0.3 is shown in Fig. 2(a) for visual comparison.

For the HSI experiments, Scene 4 in the natural scenes HSI database [30] with size $1024 \times 1024 \times 32$ (i.e., 32 frequency channels) is employed. The PSNR results using various sampling ratios are shown in Fig. 1(b) and the reconstruction results when the sampling ratio is 0.1 are displayed in RGB in Fig. 2(b). Our approach is shown to outperform the others in the HSI application. We then vary the value of p and compare the accuracy and running time in Table 1, where it can be observed that smaller p leads to more accurate reconstruction and slower execution time. Note that $p = 1$ corresponds to the case of convex relaxation. The TADMM algorithm has the highest accuracy during all the tests and a similar execution time to the SB method. In addition, we observe that the sparse based basis pursuit algorithm KCS runs significantly slower than the others; while the BKOMP and LRTC methods converge faster but lead to inaccurate solutions as their assumed models do not closely match the application.

6. CONCLUSION

In this paper, an ADMM based tensor CS reconstruction approach utilizing various structures in tensor modes is proposed. The general method can be applied to any structures as long as the proximity operators for the associated norms are derivable. In addition, efficient calculations that ease memory requirements and reduce computational complexity for large data are derived. In particular, the TADMM algorithm for a tensor with sparse and low rank modes is presented in detail, in which the nonconvex shrinkage operators are used. The experiments for MRI and HSI applications demonstrate the superiority of the proposed method compared to the state-of-art tensor CS approaches.

7. REFERENCES

[1] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Information Theory*, Feb. 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Information Theory*, April 2006.

[3] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Magazine*, vol. 25, no. 2, pp. 21 –30, March 2008.

[4] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.

[5] E. Candès and B. Recht, "Exact matrix completion via convex optimization," *Commun. ACM*, vol. 55, no. 6, pp. 111–119, June 2012.

[6] R. Chartrand, "Nonconvex splitting for regularized low-rank+ sparse decomposition," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5810–5819, 2012.

[7] M. Golbabaee and P. Vandergheynst, "Compressed sensing of simultaneous low-rank and joint-sparse matrices," *arXiv*, 2012.

[8] M. F. Duarte and R. G. Baraniuk, "Kronecker compressive sensing," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 494–504, Feb 2012.

[9] N. D. Sidiropoulos and A. Kyrillidis, "Multi-way compressed sensing for sparse low-rank tensors," *IEEE Signal Process. Letters*, vol. 19, no. 11, pp. 757–760, 2012.

[10] S. Friedland, Q. Li, and D. Schonfeld, "Compressive sensing of sparse tensors," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4438–4447, Oct 2014.

[11] C. F. Caiafa and A. Cichocki, "Computing sparse representations of multidimensional signals using kronecker bases," *Neural Comput.*, vol. 25, no. 1, pp. 186–220, Jan. 2013.

[12] C. F. Caiafa and A. Cichocki, "Multidimensional compressed sensing and their applications," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 3, no. 6, pp. 355–380, 2013.

[13] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. PAMI*, vol. 35, no. 1, pp. 208–220, 2013.

[14] H. Rauhut, R. Schneider, and Z Stojanac, "Low rank tensor recovery via iterative hard thresholding," in *Proc. ICSTA*, 2013.

[15] A. Gogna, A. Shukla, H. K. Agarwal, and A. Majumdar, "Split Bregman algorithms for sparse/joint-sparse and low-rank signal recovery: Application in compressive hyperspectral imaging," in *Proc. IEEE ICIP*. IEEE, 2014, pp. 1302–1306.

[16] Y. Liu, M. D. Vos, and S. V. Huffel, "Compressed sensing of multichannel eeg signals: The simultaneous cosparsity and low-rank optimization," *IEEE Trans. Biomedical Engineer.*, vol. 62, no. 8, pp. 2055–2061, Aug 2015.

[17] M. Seibert, J. Wormann, R. Gribonval, and M. Kleinsteuber, "Separable cosparse analysis operator learning," in *Proc. EUSIPCO*. IEEE, 2014, pp. 770–774.

[18] F. Roemer, G. D. Galdo, and M. Haardt, "Tensor-based algorithms for learning multidimensional separable dictionaries," in *Proc. IEEE ICASSP*. IEEE, 2014, pp. 3963–3967.

[19] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[20] S. Oymak, A. Jalali, M. Fazel, Y. C Eldar, and B. Hassibi, "Simultaneously structured models with application to sparse and low-rank matrices," *IEEE Trans. Information Theory*, vol. 61, no. 5, pp. 2886–2908, 2015.

[21] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Letters*, vol. 14, no. 10, pp. 707–710, 2007.

[22] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inverse Problems*, vol. 24, no. 3, pp. 035020, 2008.

[23] P. L. Combettes and J. C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point algorithms for inverse problems in science and engineering*, pp. 185–212. Springer, 2011.

[24] M. Kowalski, "Sparse regression using mixed norms," *Applied and Comput. Harmonic Analysis*, vol. 27, no. 3, pp. 303–324, 2009.

[25] I. Selesnick, "Penalty and shrinkage functions for sparse signal processing," *Connexions*, 2012.

[26] M. Hong, Z. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *arXiv*, 2014.

[27] A. Szameit, Y. Shechtman, E. Osherovich, E. Bullkich, P. Sidorenko, H. Dana, S. Steiner, E. B. Kley, S. Gazit, T. Cohen-Hyams, et al., "Sparsity-based single-shot subwavelength coherent diffractive imaging," *Nature materials*, vol. 11, no. 5, pp. 455–459, 2012.

[28] E. V. D. Berg and M. P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.

[29] C. Qiu, W. Lu, and N. Vaswani, "Real-time dynamic MR image reconstruction using Kalman filtered compressed sensing," in *Proc. IEEE ICASSP*, 2009, pp. 393–396.

[30] D. Foster, K. Amano, S. Nascimento, and M. J. Foster, "Frequency of metamerism in natural scenes," *JOSA A*, vol. 23, no. 10, pp. 2359–2372, 2006.