SEPARABLE COSPARSE ANALYSIS OPERATOR LEARNING

Matthias Seibert[1], Julian Wörmann[1*], Rémi Gribonval[2], Martin Kleinsteuber[1]*

[1]TU München, Department of Electrical Engineering and Information Technology, Munich, Germany.
[2]PANAMA Project-Team (INRIA & CNRS).
{m.seibert, julian.woermann, kleinsteuber}@tum.de, gribonval@inria.fr

ABSTRACT

The ability of having a sparse representation for a certain class of signals has many applications in data analysis, image processing, and other research fields. Among sparse representations, the cosparse analysis model has recently gained increasing interest. Many signals exhibit a multidimensional structure, e.g. images or three-dimensional MRI scans. Most data analysis and learning algorithms use vectorized signals and thereby do not account for this underlying structure. The drawback of not taking the inherent structure into account is a dramatic increase in computational cost.
We propose an algorithm for learning a cosparse Analysis Operator that adheres to the preexisting structure of the data, and thus allows for a very efficient implementation. This is achieved by enforcing a separable structure on the learned operator. Our learning algorithm is able to deal with multidimensional data of arbitrary order. We evaluate our method on volumetric data at the example of three-dimensional MRI scans.

***Index Terms*—** Cosparse Analysis Model, Analysis Operator Learning, Sparse Coding, Separable Filters

1. INTRODUCTION

Sparse signal models enjoy great popularity for handling various signal processing tasks such as image denoising, super-resolution, compressive sensing, and others. In the last decade, most of the research activity has focused on the well-known synthesis model, where the signal is synthesized via a linear combination of a few signal atoms taken from a dictionary. The synthesis model has a closely related counterpart called the analysis model where signals are characterized by their sparsity in a transformed domain. This transformation is achieved through an operator $\mathbf{\Omega} \in \mathbb{R}^{m\times n}$, with $m \geq n$. When applied to a signal $\mathbf{x} \in \mathbb{R}^n$ it yields the coefficient vector $\boldsymbol{\alpha} \in \mathbb{R}^m$ defined as

$$\boldsymbol{\alpha} = \mathbf{\Omega}\mathbf{x}. \tag{1}$$

In the cosparse analysis model the signal information is encoded in the zero-entries of $\boldsymbol{\alpha}$. The number of zero elements in the coefficient vector, i.e. $p = m - \|\boldsymbol{\alpha}\|_0$, is called the cosparsity of the signal, and we refer to $\mathbf{\Omega}$ as the cosparse analysis operator.

A prominent example for an analytic analysis operator is the finite difference operator in image processing. However, the advantage of the low computational complexity of such an analytically given transformation comes at the cost of a poor adaptation to specific signal classes of interest. It is now well-known that for a particular class, sparser signal representations and thus better reconstruction accuracies can be achieved if the transformation matrix is learned from a representative training set. We follow this approach in our work. Various learning algorithms that aim at learning a cosparse analysis operator have been proposed, cf. [1–5]. In the following, we give a short summary of the algorithms proposed in [3] and [4]. [3] is of importance as, to our knowledge, the only available separable cosparse analysis operator learning scheme is based on the algorithm therein, while [4] is relevant to our cause since we extend its framework to the separable case. The algorithm proposed in [3] is based on the famous K-SVD algorithm for dictionary learning. It consists of two alternating steps. First, a sparse coding step is performed, during which the current cosparse analysis operator is fixed and the best approximation of the noisy training signal with regard to the given cosparsity is computed. Second, each row of the cosparse analysis operator is updated using only information from the training signals orthogonal to the respective row. The general idea in [4] is to use a geometric optimization approach in order to fulfill the requirements demanded of the cosparse analysis operator. A smooth cost function is designed that, in addition to the sparsity promoting function, includes penalty functions that ensure full rank and incoherence of the learned operator. This cost function is minimized by using a geometric conjugate gradient method.

The cosparse analysis operators produced by these learning algorithms are structured only in the sense that they have unit row norm, while no further structure is enforced. These

This work was partially supported by the German Research Foundation (DFG) under grant KL 2189/8-1, the DFG funded Cluster of Excellence CoTeSys, and by the European Research Council, PLEASE project (ERC-StG-2011-277906).

* These authors contributed equally to this work.

matrices allow for a high cosparsity in the analyzed signal. However, the size of the learned operators is ultimately restricted due to limitations in memory and computational power. Separable learning schemes tackle this problem by enforcing additional structure on the learned operators.

For the synthesis model the viability of separable approaches has been examined in [6, 7]. To the best of the authors' knowledge, the only work combining a separable structure with the cosparse analysis model can be found in [8]. The authors suggest a learning scheme for two-dimensional signals, which is based on the Analysis K-SVD algorithm in [3].

We propose a separable learning scheme based on the work in [4]. By introducing separability we are able to deal with data of arbitrarily high order while maintaining its inherent structure. We evaluate our proposed method on volumetric Magnetic Resonance Imaging (MRI) data.

Notation. Matrices are denoted as boldface capital letters such as $\mathbf{X}$. Vectors are boldface small letters, i.e. $\mathbf{x}$. Tensors are denoted as calligraphic letters $\mathcal{X}$, whereas the $(i_1, i_2, \ldots, i_N)$ entry of a tensor with order N is denoted by $x_{i_1 i_2 \ldots i_N}$. Entries of vectors and matrices are indexed accordingly. The ith row of a matrix is denoted by $\boldsymbol{\omega}_{i,:}$. Some additional notation necessary for working with tensors will be introduced in the following section.

2. STRUCTURED COSPARSE ANALYSIS OPERATOR LEARNING

2.1. Tensor Operations

In order to be able to deal with multidimensional data, we have to introduce some techniques from the field of multilinear algebra. First, the so called n-mode product, introduced in [9], provides a way to apply transformations in the form of matrices to the separate modes of the tensor.

Definition 1. *Given the tensor $\mathcal{S} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ and the matrix $\mathbf{U} \in \mathbb{R}^{J_n \times I_n}$, the n-*mode product *is denoted by*

$$\mathcal{S} \times_n \mathbf{U}$$

and results in a $I_1 \times I_2 \times \ldots \times I_{n-1} \times J_n \times I_{n+1} \times \ldots \times I_N$ tensor. The entries of this tensor are defined as

$$(\mathcal{S} \times_n \mathbf{U})_{i_1 i_2 \ldots i_{n-1} j_n i_{n+1} \ldots i_N} = \sum_{i_n=1}^{I_n} s_{i_1 i_2 \ldots i_N} \cdot u_{j_n i_n} \quad (2)$$

for all $j_n = 1, \ldots, J_n$.

To offer a better understanding of this concept, a visualization for a tensor of order 3 is provided in Figure 1. In addition to this product, we will also make use of the so-called *n-mode matrix unfolding* that enables us to one-to-one map a given tensor to a matrix of appropriate size. Since we will exclusively use unfoldings along the Nth mode, we will only give the definition for this specific case.

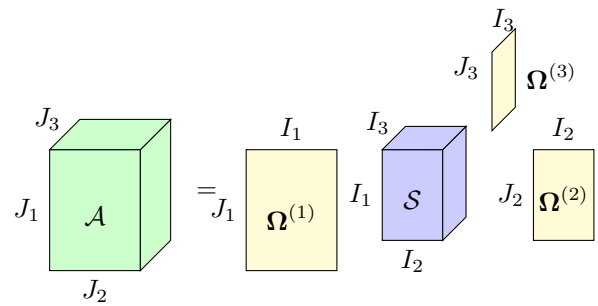


Fig. 1: The n-mode product $\mathcal{A} = \mathcal{S} \times_1 \mathbf{\Omega}^{(1)} \times_2 \mathbf{\Omega}^{(2)} \times_3 \mathbf{\Omega}^{(3)}$ with tensors $\mathcal{A}, \mathcal{S}$ and matrices $\mathbf{\Omega}^{(i)}$, $i = 1, 2, 3$.

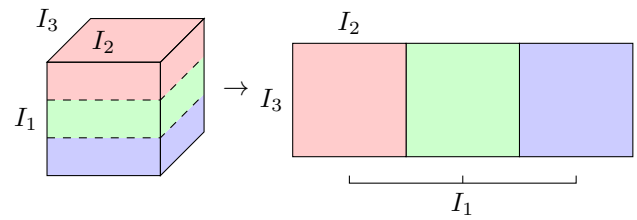


Fig. 2: The 3-mode matrix unfolding of a 3-tensor.

Definition 2. *The N-mode matrix unfolding of a tensor $\mathcal{S} \in \mathbb{R}^{I_1 \times \ldots \times I_N}$ is denoted by $\mathbf{S}_{(N)} \in \mathbb{R}^{I_N \times (\prod_{j \neq N} I_j)}$. Given the tensor $\mathcal{S} \in \mathbb{R}^{I_1 \times \ldots \times I_N}$, the unfolding matrix $\mathbf{S}_{(N)}$ contains the element $s_{i_1 i_2 \ldots i_N}$ in row i_N and column*

$$(i_1 - 1) I_2 I_3 \ldots I_{N-1} + (i_2 - 1) I_3 I_4 \ldots I_{N-1} + \ldots + i_{N-1}.$$

For a general definition of this mapping we refer the interested reader to [9]. Figure 2 illustrates the unfolding of a third order tensor.

The unfolding makes it possible to rewrite the n-mode product of a tensor equivalently as a matrix-vector product.

$$\begin{aligned} &\mathcal{A} = \mathcal{S} \times_1 \mathbf{\Omega}^{(1)} \ldots \times_N \mathbf{\Omega}^{(N)} \\ \Leftrightarrow\ &\mathbf{A}_{(N)} = \mathbf{\Omega}^{(N)} \cdot \mathbf{S}_{(N)} \cdot \left(\mathbf{\Omega}^{(1)} \otimes \ldots \otimes \mathbf{\Omega}^{(N-1)} \right)^\top \\ \Leftrightarrow\ &\operatorname{vec}(\mathbf{A}_{(N)}) = \left(\mathbf{\Omega}^{(1)} \otimes \ldots \otimes \mathbf{\Omega}^{(N)} \right) \cdot \operatorname{vec}(\mathbf{S}_{(N)}) \quad (3) \end{aligned}$$

Here, vec denotes the operator that stacks the columns of a matrix below one another and $\otimes$ denotes the Kronecker product of matrices. The first reshaping is obtained by unfolding $\mathcal{A}$ and $\mathcal{S}$ along the Nth mode, while the second follows from the definition of the Kronecker product.

We refer to a large, structured cosparse analysis operator $\mathbf{\Omega}$ that can be represented as a concatenation of smaller operators $\{\mathbf{\Omega}^{(1)}, \ldots, \mathbf{\Omega}^{(N)}\}$ as a *separable* operator.

2.2. Derivation of the Cost Function

With the introduced tensor operations we are able to define a way to apply a cosparse analysis operator to a multidimensional signal $\mathcal{S}$ recorded as an N-tensor. Let $\{\mathcal{S}_i\}$,

$i = 1, \ldots, T$ be a set of training signals where each signal is a tensor of order N. The optimal separable cosparse analysis operator for this set is learned by solving the learning problem

$$\arg\min_{\mathbf{\Omega}^{(i)} \in \mathfrak{C}_i} \frac{1}{T} \sum_{j=1}^{T} g\left(\mathcal{S}_j \times_1 \mathbf{\Omega}^{(1)} \ldots \times_N \mathbf{\Omega}^{(N)}\right)^2 \quad (4)$$

with the sparsity promoting function

$$g(\mathcal{A}) := \sum_k \log\left(1 + \nu \alpha_k^2\right). \quad (5)$$

The parameter $\nu > 0$ serves as a weighting factor to control the sparsity, and the sum is taken over all entries α_k of the tensor $\mathcal{A}$. There are two things to point out in the design of Equation (4). First, in the proposed optimization problem we minimize the square of $g(\cdot)$. As shown in [4] this leads to a balanced minimization of expectation and variance of the samples' sparsity. Second, the minimization is performed over a constraint set $\mathfrak{C}_i$ which prevents the trivial solution of the given problem. The role of this constraint set will be discussed in the following.

As shown in (3), the n-mode product which is used in the cost function can be rewritten as a matrix-vector-product. The Kronecker product of the small operators that appears in (3) can be interpreted as a large cosparse analysis operator with additional structure which is applied to the vectorized signal. The algorithm we propose is based on the work by the authors in [4] and extended to a separable case. In [4], the authors enforce certain properties on their unstructured operator $\widehat{\mathbf{\Omega}}$ via regularizers. They demand that

(i) The rows of $\widehat{\mathbf{\Omega}}$ have unit Euclidean norm.

(ii) The operator $\widehat{\mathbf{\Omega}}$ has full rank, i.e. it has the maximal number of linear independent rows.

(iii) The rows of the operator $\widehat{\mathbf{\Omega}}$ are not trivially linearly dependent, i.e. $\boldsymbol{\omega}_{k,:} \neq \pm \boldsymbol{\omega}_{l,:}$ for $k \neq l$.

As stated in [4] condition (i) is realized by the oblique manifold

$$\widehat{\mathrm{OB}}(m,n) = \left\{\mathbf{\Omega} \in \mathbb{R}^{m \times n} \,:\, (\mathbf{\Omega}\mathbf{\Omega}^\top)_{ii} = 1,\, i = 1, \ldots, m\right\}.$$

In the following, we show that for a structured cosparse analysis operator enforcing these constraints on the operator components is equivalent to enforcing them on the large operator. It should be pointed out that the cosparse analysis operators for the separate modes are actually the transposed of an element of the oblique manifold of appropriate size.

It follows from the definition of the Kronecker product that if the rows of $\mathbf{\Omega}^{(1)}$ and $\mathbf{\Omega}^{(2)}$ have unit norm, then the rows of $\mathbf{\Omega}^{(1)} \otimes \mathbf{\Omega}^{(2)}$ have unit norm as well. Furthermore, the rank of the matrix resulting from Kronecker product $\mathbf{\Omega}^{(1)} \otimes \mathbf{\Omega}^{(2)}$ is the product of the rank of $\mathbf{\Omega}^{(1)}$ and $\mathbf{\Omega}^{(2)}$. Finally, if the operators $\mathbf{\Omega}^{(1)}, \mathbf{\Omega}^{(2)}$ do not exhibit trivially linearly dependent rows, then neither does $\mathbf{\Omega}^{(1)} \otimes \mathbf{\Omega}^{(2)}$. By induction the above mentioned properties can be extended to the Kronecker product of finitely many operators. Hence, the properties demanded in [4] are fulfilled for $\mathbf{\Omega}^{(1)} \otimes \ldots \otimes \mathbf{\Omega}^{(N)}$ if the conditions (i), (ii), and (iii) hold for each operator component $\mathbf{\Omega}^{(i)}, i = 1, \ldots, N$. To realize the full rank and the incoherence constraints, we adopt the regularizers

$$h\big(\mathbf{\Omega}^{(i)}\big) = -\tfrac{1}{n_i \log(n_i)} \log\det\left(\tfrac{1}{m_i}(\mathbf{\Omega}^{(i)})^\top \mathbf{\Omega}^{(i)}\right),$$

$$r\big(\mathbf{\Omega}^{(i)}\big) = -\sum_{k<l} \log\left(1 - \left((\boldsymbol{\omega}_{k,:}^{(i)})^\top (\boldsymbol{\omega}_{l,:}^{(i)})\right)^2\right).$$

The sum of these log-barrier functions is itself a logarithmic barrier functions which ensures that $\mathbf{\Omega}^{(i)}$ lies within the intersection of the two proposed constraint sets. With the corresponding weights κ and μ which control the importance of the corresponding constraint, we are able to formulate the optimization problem

$$\begin{aligned} \arg\min_{\mathbf{\Omega}^{(i)}} \frac{1}{T} \sum_{j=1}^{T} & g\left(\mathcal{S}_j \times_1 \mathbf{\Omega}^{(1)} \ldots \times_N \mathbf{\Omega}^{(N)}\right)^2 \\ & + \kappa \sum_{j=1}^{N} h\big(\mathbf{\Omega}^{(j)}\big) + \mu \sum_{j=1}^{N} r\big(\mathbf{\Omega}^{(j)}\big) \\ \text{subject to:} \quad & \mathbf{\Omega}^{(i)} \in \widehat{\mathrm{OB}}(m_i, n_i), \quad i = 1, \ldots, N. \end{aligned} \quad (6)$$

This optimization task is tackled using a geometric conjugate gradient method based on the one introduced in [4]. Our algorithm is an extension in the sense that it is designed to work on the product of oblique manifolds. The line search required in the CG algorithm is performed using the non-monotone line search strategy proposed in [10].

3. MULTIDIMENSIONAL SIGNAL RECONSTRUCTION

Multidimensional signals can be found in many applications ranging from video, over hyper-spectral imaging to 3D models, medical images, and others. However, even within one signal class each dimension can represent a completely different physical meaning. While in hyper-spectral data the third dimension encodes the spectral component of a scene, volumetric data typically encodes spatial information. In this work we focus on three-dimensional data with a homogeneous physical meaning in all dimensions. An example for data of this kind is volumetric MRI data on which we evaluate our proposed scheme by conducting several signal reconstruction experiments.

For this purpose, we first extract a training set of $T = 20\,000$ non-flat and normalized tensors of size $\mathcal{S}_i \in \mathbb{R}^{5\times5\times5}$ from a synthetic, unbiased MRI brain atlas [11, 12], which is available online[1]. The parameter ν in (5) is set to $\nu = 1\,000$, while the weightings of the constraints read $\kappa = 500$ and $\mu = 0.5$. After learning, we use the operators each of size $\mathbf{\Omega}^{(i)} \in \mathbb{R}^{6\times5}$, $i = 1, 2, 3$ to regularize the reconstruction for-

[1] http://www.bic.mni.mcgill.ca/ServicesAtlases/ICBM152NLin2009

Table 1: MRI volume data reconstruction from measurements corrupted by additive white Gaussian noise with standard deviation σ_{noise}. Achieved PSNR in decibels, t in minutes.

Method	σ_{noise}	*PSNR*	*MSSIM*	t_{learn}	t_{rec}
AKSVD	5	38.41	0.968	980	840
	15	28.83	0.798		
MAOL	5	38.55	0.971	18	12
	15	30.64	0.851		

mulated as an inverse problem. However, since the operators only act on local signal cubes, we first introduce the operator $\mathbf{\Pi}_{(r,c,k)}(\cdot)$ that extracts a small tensor of appropriate size centered at location (r, c, k) from the entire volume. Given the measurements $\mathbf{y} = \mathbf{\Phi}(\mathcal{V}) + \eta \in \mathbb{R}^{J_1 \cdot J_2 \cdot J_3}$ arranged as a vector, where $\mathbf{\Phi}(\mathcal{V})$ models the measurement process and η constitutes some noise, we finally state the reconstruction of the original volumetric signal $\mathcal{V} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ as the inverse problem

$$\mathcal{V}^\star \in \arg\min_{\mathcal{V}} \ \tfrac{1}{2}\|\mathbf{\Phi}(\mathcal{V}) - \mathbf{y}\|_2^2 + \tfrac{\lambda}{M}\sum_{(r,c,k)} g\big(\mathbf{\Pi}_{(r,c,k)}(\mathcal{V}) \times_1 \mathbf{\Omega}^{(1)} \times_2 \mathbf{\Omega}^{(2)} \times_3 \mathbf{\Omega}^{(3)}\big). \quad (7)$$

In all our experiments, we employ the same sparsity promoting function $g(\cdot)$ as given in (5). The parameter $\lambda \in \mathbb{R}^+$ balances between the data fidelity and the regularization and M serves as a normalization factor. Optimization is performed via a conjugate gradient approach. Our algorithm is named MAOL (Multidimensional Analysis Operator Learning).

In our first experiment we evaluate the influence of preserving the multidimensional structure compared to the standard approach of vectorizing each tensor. For this purpose we conducted a simple denoising experiment. The second experiment illustrates the flexibility of our method which is applicable to a wide range of image processing tasks. A prominent linear inverse problem in medical imaging is the reconstruction of MRI images from partial Fourier measurements. Reducing the amount of measurements and thus reducing the scan time for MRI is highly desired for clinical throughput and convenience for patients. Based on the theory of Compressive Sensing (CS), many algorithms have been proposed in the literature so far that exploit the sparsity of this type of data. Typically, sparsity is measured in a fixed analytic transform domain like in the wavelet basis, while additional priors like total-variation further enhance the reconstruction quality. However, recently learning approaches have made their way into this field due to the fact that they adapt the structure of the signal more accurately. The original MRI data has been scaled such that its intensity values lie in the range $[0, 255]$. The performance is measured in terms of PSNR and the mean structural similarity (MSSIM) averaged over all slices in the whole volume. All experiments are implemented in MATLAB and executed on a standard desktop PC.

Table 2: Compressive Sensing MRI volume reconstruction from noiseless radial sampled measurements with an undersampling rate of 20%. Achieved PSNR in decibels.

Method	*PSNR*	*MSSIM*
RecPF	26.06	0.788
DLMRI	27.35	0.818
MAOL	27.90	0.847

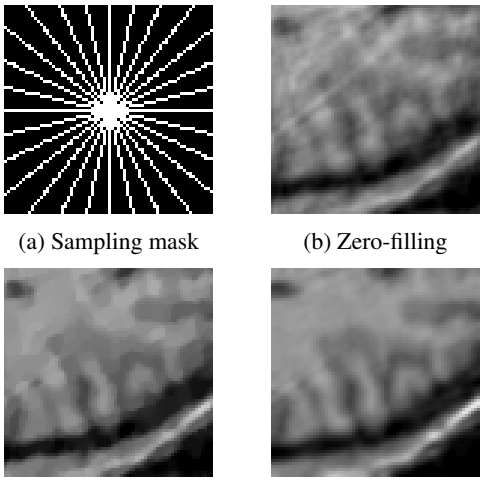
4. NUMERICAL EXPERIMENTS

Denoising aims at reconstructing a clean signal version from noise corrupted data, i.e. we have $\mathbf{\Phi}(\mathcal{V}) = \text{vec}(\mathcal{V})$ and η is additive white Gaussian noise (AWGN) with standard deviation σ_{noise}. The volumetric MRI test data is a representative region of size $70 \times 70 \times 70$ voxels from the T1 weighted BrainWeb phantom available online[2] [13, 14] which has been artificially corrupted by AWGN. We chose the Analysis K-SVD (AKSVD) algorithm proposed in [3] for comparison and learned a single operator $\mathbf{\Omega}_{\text{KSVD}} \in \mathbb{R}^{216 \times 125}$, i.e. the analyzed vector has the same dimension as ours. The operator $\mathbf{\Omega}_{\text{KSVD}}$ is directly learned from 20 000 noise contaminated tensors from the data that has to be reconstructed. Table 1 summarizes the results for two different choices of σ_{noise} as well as the computation time for the operator learning t_{learn} and the reconstruction t_{rec}. We assume σ_{noise} to be known and set $\lambda = 200\sigma_{\text{noise}}$.

While the performance of the two algorithms in the range of moderate AWGN is comparable in terms of PSNR and MSSIM, the separable structure of our operator enables a significant saving of computation time compared to the vectorizing approach.

Regarding our second experiment, the underlying imaging process in MRI consists in measuring Fourier coefficients of a typically transversal slice rather than directly acquiring pixel values. Thus, the measurement process is modeled via $\mathbf{\Phi}(\mathcal{V}) = \mathbf{\Psi}_u(\mathcal{V})$, where $\mathbf{\Psi}_u(\mathcal{V})$ performs a multi-slice 2D undersampled Fourier transform resulting in a vector containing the stacked measurements of each slice. The right choice of the chosen coefficients has a direct impact on the reconstruction quality. Here, we use radial sampling as illustrated in Fig. 3a with an undersampling rate of 20%. The weighting of the regularizer reads $\lambda = 1\,500$. We compare our results to a standard approach (RecPF) [15] that minimizes an error term consisting of a least squares data fitting term combined with total variation and wavelet sparsity regularization. Furthermore, we give the results of the algorithm proposed in [16] (DLMRI) which utilizes an adaptively trained synthesis dictionary on each slice in the reconstruction process. Results on the same noiseless volume as mentioned above are summarized in Table 2. The RecPF algorithm has the fastest execution time, however its accuracy suffers from the fixed an-

[2] http://brainweb.bic.mni.mcgill.ca/brainweb/



(a) Sampling mask (b) Zero-filling

(c) RecPF (d) MAOL

Fig. 3: Sampling mask and reconstruction results for (b) inverse FFT after zero-filling missing Fourier coefficients, (c) the RecPF algorithm, and (d) our proposed separable approach MAOL.

alytical transformations. Both learning approaches DLMRI and MAOL achieve significantly better results, while our approach has a runtime eight times faster than DLMRI. Fig. 3 shows 2D cross sections of the reconstructed volumes.

5. CONCLUSION

We provide a separable cosparse analysis operator learning framework for multidimensional signals called MAOL. Due to the separable nature, the computational complexity of both learning and reconstruction is reduced which makes it possible to train the cosparse analysis operator on larger signal patches compared to methods that generate unstructured operators. MAOL is designed using techniques from multilinear algebra and geometric optimization. The viability of our approach is demonstrated at the hand of experiments performed on volumetric MRI data.

REFERENCES

[1] M. Yaghoobi, S. Nam, R. Gribonval, and M.E. Davies, "Analysis operator learning for overcomplete cosparse representations," in *European Signal Process. Conf.*, 2011.

[2] S. Nam, M.E. Davies, M. Elad, and R. Gribonval, "Cosparse analysis modeling-uniqueness and algorithms," in *IEEE Int. Conf. Acoustics, Speech and Signal Process.*, 2011, pp. 5804–5807.

[3] R. Rubinstein, T. Peleg, and M. Elad, "Analysis K-SVD: A dictionary-learning algorithm for the analysis sparse model," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 661–677, 2013.

[4] S. Hawe, M. Kleinsteuber, and K. Diepold, "Analysis operator learning and its application to image reconstruction," *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2138–2150, 2013.

[5] Y. Chen, R. Ranftl, and T. Pock, "Insights into analysis operator learning: From patch-based sparse models to higher order MRFs," *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1060–1072, 2014.

[6] R. Rigamonti, A. Sironi, V. Lepetit, and P. Fua, "Learning separable filters," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2013, pp. 2754–2761.

[7] S. Hawe, M. Seibert, and M. Kleinsteuber, "Separable dictionary learning," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2013, pp. 438–445.

[8] N. Qi, Y. Shi, X. Sun, J. Wang, and W. Ding, "Two dimensional analysis sparse model," in *IEEE Int. Conf. Image Process.*, 2013.

[9] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. A.*, vol. 21, pp. 1253–1278, 2000.

[10] H. Zhang and W.W. Hager, "A nonmonotone line search technique and its application to unconstrained optimization," *SIAM J. Optimiz.*, vol. 14, no. 4, pp. 1043–1056, 2004.

[11] V.S. Fonov, A.C. Evans, R.C. McKinstry, C.R. Almli, and D.L. Collins, "Unbiased nonlinear average age-appropriate brain templates from birth to adulthood," *NeuroImage*, vol. 47, pp. S102, 2009.

[12] V.S. Fonov, A.C. Evans, K. Botteron, C.R. Almli, R.C. McKinstry, and D.L. Collins, "Unbiased average age-appropriate atlases for pediatric studies," *NeuroImage*, vol. 54, no. 1, pp. 313–327, 2011.

[13] C.A. Cocosco, V. Kollokian, R.K.-S. Kwan, G.B. Pike, and A.C. Evans, "Brainweb: Online interface to a 3D MRI simulated brain database," *NeuroImage*, vol. 5, no. 4, pp. S425, 1997.

[14] R.K.-S. Kwan, A.C. Evans, and G.B. Pike, "MRI simulation-based evaluation of image-processing and classification methods," *IEEE Trans. Med. Imag.*, vol. 18, no. 11, pp. 1085–1097, 1999.

[15] J. Yang, Y. Zhang, and W. Yin, "A fast alternating direction method for TVL1-L2 signal reconstruction from partial fourier data," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 288–297, 2010.

[16] S. Ravishankar and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1028–1041, 2011.