

Stable, Robust and Super Fast Reconstruction of Tensors Using Multi-Way Projections

Cesar F. Caiafa*, *Member, IEEE*, and Andrzej Cichocki, *Fellow, IEEE*

Abstract

In the framework of multidimensional Compressed Sensing (CS), we introduce an analytical reconstruction formula that allows one to recover an N th-order data tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ from a reduced set of multi-way compressive measurements by exploiting its low multilinear-rank structure. Moreover, we show that, an interesting property of multi-way measurements allows us to build the reconstruction based on compressive linear measurements taken only in two selected modes, independently of the tensor order N . In addition, it is proved that, in the matrix case and in a particular case with 3rd-order tensors where the same 2D sensor operator is applied to all mode-3 slices, the proposed reconstruction $\underline{\mathbf{X}}_\tau$ is stable in the sense that the approximation error is comparable to the one provided by the best low-multilinear-rank approximation, where τ is a threshold parameter that controls the approximation error. Through the analysis of the upper bound of the approximation error we show that, in the 2D case, an optimal value for the threshold parameter $\tau = \tau_0 > 0$ exists, which is confirmed by our simulation results. On the other hand, our experiments on 3D datasets show that very good reconstructions are obtained using $\tau = 0$, which means that this parameter does not need to be tuned. Our extensive simulation results demonstrate the stability and robustness of the method when it is applied to real-world 2D and 3D signals. A comparison with state-of-the-arts sparsity based CS methods specialized for multidimensional signals is also included. A very attractive characteristic of the proposed method is that it provides a direct computation, i.e. it is non-iterative in contrast to all existing sparsity based CS algorithms, thus providing super fast computations, even for large datasets.

Index Terms

Compressed Sensing (CS), Kronecker-CS, Low-rank tensor approximation, Multi-way analysis, Tucker model.

I. INTRODUCTION

DURING the last years there has been an increased interest in *Compressed Sensing* (CS), whose aim is the reconstruction of signals based on a set of measurements that is much smaller than the original signal size. Thus, instead to acquire a potentially large signal and compress it, CS suggests that signals can be compressively sampled thus reducing the amount of measurements. More specifically, in standard CS [1], [2], a signal $\mathbf{x} \in \mathbb{R}^n$, which is assumed unavailable, is reconstructed from a reduced set of m linear projections ($m \ll n$) $\mathbf{w} = \Phi \mathbf{x} \in \mathbb{R}^m$, where the sensing matrix $\Phi \in \mathbb{R}^{m \times n}$ is typically random or composed by few selected rows of the Fourier transform matrix [3]. In order to make the CS problem solvable, it is necessary to exploit *a priori* information about the signal of interest \mathbf{x} by imposing some constraints. For example, it is widely assumed that the signal \mathbf{x} is compressible by decomposing it in a known Wavelet basis (*dictionary*). In other words, it is assumed that every signal admits a *sparse* representation on a given dictionary, i.e. combining only few elements, called *atoms*. Under the scope of these *sparse models*, many efficient CS algorithms were developed in order to reconstruct signals from compressive measurements which involve iterative refinements of the solution by means of *Greedy algorithms* or by minimizing the ℓ_1 -norm of the solution (see [4] for an up to date summary of algorithms). These algorithms have found many applications in diverse fields such as in medical imaging, surveillance, machine learning, etc [5].

Cesar F. Caiafa is with Instituto Argentino de Radioastronomía (IAR) - CCT La Plata, CONICET CC5 (1894) V. Elisa, ARGENTINA and Facultad de Ingeniería - UBA, Buenos Aires, ARGENTINA.

E-mail: ccaiafa@fi.uba.ar

Andrzej Cichocki is with Lab. for Advanced Brain Signal Processing (LABSP), Brain Science Institute, RIKEN, 2-1 Hirosawa, Wako, 351-0198, JAPAN and Systems Research Institute, PAS, Warsaw, POLAND.

This paper has supplementary downloadable material available at <http://ieeexplore.ieee.org>., provided by the author. The material includes the Matlab codes needed to reproduce the simulation results. Contact ccaiafa@fi.uba.ar for further questions about this work.

A. Multidimensional CS

Most of the development of CS was focused on problems involving 1D signal or 2D image data encoded in vectors. However, many important applications involve higher dimensional signals or *tensors*. Some data sources are readily generated as tensors such as hyperspectral images, videos, 3D light field displays [6], Magnetic Resonance Imaging (MRI) [7], etc., in other cases, tensors can be synthetically created by a rearrangement of lower dimensional data structures or by mathematical construction [8]. In some applications, like in materials science [9] or in scientific computation [10], the exponential increase in memory and time requirements when the number of dimensions increases, makes impossible to work with full datasets and models with few parameters must be used. Such models are referred as tensor decompositions, which can be obtained by making few inspections of the full-datasets or by taking compressive measurements.

Recently, the Kronecker-CS model [11] has been proposed in order to provide a practical implementation of CS for higher order tensors by exploiting their multidimensional structure. This model explicitly assumes that multidimensional signals have sparse representations using separable dictionaries, usually known as *Kronecker dictionaries*. Kronecker bases are well known and widely used in image processing, for example, given a 3D ($I_1 \times I_2 \times I_3$) image, its associated dictionary $\mathbf{D} \in \mathbb{R}^{I_1 I_2 I_3 \times I_1 I_2 I_3}$ is $\mathbf{D} = \mathbf{D}_3 \otimes \mathbf{D}_2 \otimes \mathbf{D}_1$ where $\mathbf{D}_n \in \mathbb{R}^{I_n \times I_n}$ ($n = 1, 2, 3$) are small dictionaries associated to columns (mode-1), rows (mode-2) and tubes (mode-3), respectively. Moreover, when working with multidimensional signals, the Kronecker structure also arises naturally in the physical implementation of the sensing devices since they can operate on different dimensions or modes of the signal, independently, through separated sensing matrices [11], [12]. This Kronecker structure of the sensing operator/dictionary is equivalent to apply the constrained Tucker model [13] and made possible to implement relatively fast and practical algorithms based on sparsity structures, e.g., on hyperspectral 3D images and video data through a vector ℓ_1 -norm minimization algorithm [11]. More recently, greedy algorithms, especially designed to take advantage of the Kronecker structure and block sparsity of the representations, were proposed in [13] and applied to a variety of multidimensional signal processing problems such as in MRI, hyperspectral imaging and multidimensional inpainting [14]. Also, in [15], the authors developed generalized tensor compressed sensing algorithms by exploiting the Kronecker structure in a similar way as done in [13], [16] but using an ℓ_1 -minimization approach.

B. Exploiting low-rank approximations instead of sparsity

While sparsity is the “working horse” of standard CS, recently a new line of research has been proposed suggesting that, instead of using sparse models as *a priori* information about multidimensional signals, the low-rank approximation property could be exploited, i.e. without any a priori knowledge about the possible bases or factor matrices for each mode (dictionaries). This idea was first explored in [17], where matrices were reconstructed from its under-sampled measurements by solely assuming the existence of a low-rank approximation and by solving a convex optimization problem involving the matrix nuclear norm. These ideas have been extended to tensors, by considering different models for the measurements and using tensor low-rank approximations (based on the CANDECOMP/PARAFAC (CP) model) or low-multilinear-rank approximations (based on the Tucker model). For example, in [18], tensor completion of visual data was analyzed by generalizing the minimization of matrix nuclear norm to the tensor case. In [19], also the problem of estimating missing entries in tensors was considered by assuming that a low-rank CP model is fitted through a weighted least squares problem formulation. In [20], hyperspectral images (3D tensors) are recovered from random linear projections of all channels by using a reconstruction algorithm that combines low-rank and joint-sparse matrix recovery. In [21] the problem of reconstructing tensors having a low multilinear-rank Tucker model, based on a set of linear measurements, was investigated and used for tensor denoising via an iterative multilinear-rank minimization method. In the context of optical-interferometric imaging, in [22], a method for recovering a supersymmetric rank-1 3D tensor from a set of multilinear measurements was proposed by using a convex linear program. In [23], the Kronecker sensing structure was used for tensor compression and a method involving a low-rank model fitting, followed by a per mode ℓ_0/ℓ_1 decompression, was proposed in order to recover a low-rank CP tensor with sparse factors. We refer to Table I for a brief summary of recent approaches to CS involving tensor datasets, where the differences and similarities among the methods are highlighted.

In this work, we extend the ideas and results of our recent conference paper [24], providing a direct (i.e., analytical) reconstruction formula that allows us to recover a tensor from a set of multilinear projections that are obtained by multiplying the data tensor by a different sensing matrix in each mode. This model comes into scene

TABLE I
SUMMARY AND COMPARISON OF AVAILABLE APPROACHES TO CS FOR TENSOR DATASETS

Articles	Tensor order	Data Model	Measurements Model	Algorithm type
Duarte et al [11]	$N \geq 2$	Sparsity, Kronecker dictionaries	Multilinear	ℓ_1 -norm minimization
Caiafa et al [13], [14]	$N \geq 2$	Sparsity, Block-sparsity, Kronecker dictionaries	Multilinear	Greedy
Li et al [15]	$N \geq 2$	Sparsity, Kronecker dictionaries	Multilinear	ℓ_1 -norm minimization
Candes et al [17]	$N = 2$	Low-rank matrix	Linear (missing entries)	Nuclear norm minimization
Liu et al [18]	$N = 3$	Low-rank (CP model)	Linear (missing entries)	Tensor nuclear norm minimization
Acar et al [19]	$N = 3$	Low-rank (CP model)	Linear (missing entries)	Weighted Least Squares
Golbabaee et al [20]	$N = 2$	Joint-sparsity, Low-rank matrices	Linear	Nuclear norm, $\ell_{2,1}$ mixed norm minimization.
Rauhut et al [21]	$N \geq 3$	Low multilinear-rank (Tucker model)	Linear	Iterative Hard Thresholding
Auria et al [22]	$N = 3$	Rank-1 Tensor	Multilinear	Convex programming
Sidiropoulos et al [23]	$N = 3$	Low-rank Sparse CP model	Multilinear	ℓ_0/ℓ_1 -norm decompression
Current article	$N \geq 2$	Low multilinear-rank (Tucker model)	Multilinear	Non-iterative (direct) reconstruction

naturally in many potential applications, for example, in the case of sensing 2D or 3D images by means of a separable operator as developed in [25], [26], [11], [12], i.e., by taking compressive measurements of columns, rows, etc. separately, imposing a Kronecker structure on the sensing operator. The key assumption is that our multidimensional signal is well approximated by a low multilinear-rank Tucker model which is realistic for many structured datasets, specially in the case of multidimensional images. We formulate our multidimensional CS model in a general setting for N -th order tensors and provide theoretical stability analysis and robustness evidence for $N = 2$ and a very important particular case with $N = 3$.

A particularly distinctive and attracting feature of the proposed reconstruction method is that, unlike all other methods listed in Table I, our method is non-iterative. We believe that the present mathematical model could be fully exploited by the next generation of multidimensional compressive sensors for very large datasets. Through extensive simulations on real-world datasets, we also illustrate the relevance of our results in hyperspectral compressive imaging for which the technology is already available [11], [25], [12].

C. Paper organization

This paper is organized as follows: in Section II, tensor notation, definitions and basic results used throughout the paper, are introduced; in Section III, the reconstruction formula is introduced for the ideal case when the tensor of interest admits an exact low multilinear-rank representation; in Section IV, the effect of a more realistic model for signals is analyzed and a modified reconstruction formula is proposed in order to guarantee stable reconstructions; in Section V, several numerical results based on 2D and 3D real-world signals are provided, validating our theoretical results and evaluating the stability and robustness of our proposed reconstruction scheme. The performance is evaluated in terms of computational time, quality of reconstructions and variance of the results over Monte Carlo simulations (robustness). Finally, in section VI, the main conclusions of the present work are outlined.

II. NOTATION, DEFINITIONS AND PRELIMINARY RESULTS

A. Tensor notation and operations

Tensors (multi-way arrays) are denoted by underlined boldface capital letters, e.g. $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is an N -th order tensor of real numbers. Matrices (2D arrays) are denoted by bold uppercase letters and vectors by boldface

lower-case letters, e.g. $\mathbf{X} \in \mathbb{R}^{I_1 \times I_2}$ and $\mathbf{x} \in \mathbb{R}^I$ are a matrix and a vector, respectively. The element (i_1, i_2, \dots, i_N) of a tensor is referred as $x_{i_1 i_2 \dots i_N}$. The Frobenius norm is defined by $\|\underline{\mathbf{X}}\|_F = \sqrt{\sum_{i_1} \dots \sum_{i_N} x_{i_1 i_2 \dots i_N}^2}$. The spectral norm of a matrix \mathbf{A} is denoted by $\|\mathbf{A}\|$ corresponding to its largest singular value.

Given a tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, its mode- n fibers are the vectors obtained by fixing all indices except i_n , which correspond to columns ($n = 1$), rows ($n = 2$), and so on. Mode- n unfolding of a tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ yields a matrix $\mathbf{X}_{(n)} \in \mathbb{R}^{I_n \times \bar{I}_n}$ ($\bar{I}_n = \prod_{m \neq n} I_m$) whose columns are the corresponding mode- n fibers arranged in a particular order, to be more precise, tensor element (i_1, i_2, \dots, i_N) maps to matrix element (i_n, j) , where $j = 1 + \sum_{k \neq n} (i_k - 1) J_k$ with $J_k = \prod_{m \neq n}^{k-1} I_m$ [27].

Given a multidimensional signal (tensor) $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ and a matrix $\Phi \in \mathbb{R}^{J \times I_n}$ the mode- n tensor by matrix product $\underline{\mathbf{Y}} = \underline{\mathbf{X}} \times_n \Phi \in \mathbb{R}^{I_1 \times \dots \times I_{n-1} \times J \times I_{n+1} \times \dots \times I_N}$ is defined by:

$$y_{i_1 \dots i_{n-1} j i_{n+1} \dots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 \dots i_n \dots i_N} \phi_{j i_n}, \quad (1)$$

with $i_k = 1, 2, \dots, I_k$ ($k \neq n$) and $j = 1, 2, \dots, J$. It should be noted that this corresponds to the product of matrix Φ by each one of the mode- n fibers of $\underline{\mathbf{X}}$ since $\mathbf{Y}_{(n)} = \Phi \mathbf{X}_{(n)}$.

B. Tucker model and multilinear-rank

The *Tucker decomposition model* [28] provides a generalization of the low-rank approximation of matrices to the case of tensors, i.e. for a given tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, we have $\underline{\mathbf{X}} = \underline{\mathbf{X}}_0 + \underline{\mathbf{E}}$, where $\underline{\mathbf{E}}$ is an error tensor and the multilinear-rank- (R_1, R_2, \dots, R_N) tensor approximation $\underline{\mathbf{X}}_0$ is defined as follows (Tucker model):

$$\underline{\mathbf{X}}_0 = \underline{\mathbf{G}} \times_1 \mathbf{A}_1 \times_2 \dots \times_N \mathbf{A}_N, \quad (2)$$

with a *core tensor* $\underline{\mathbf{G}} \in \mathbb{R}^{R_1 \times R_2 \times \dots \times R_N}$ and *factor matrices* $\mathbf{A}_n \in \mathbb{R}^{I_n \times R_n}$ (typically $R_n \ll I_n$). A data tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is said to have *multilinear-rank- (R_1, R_2, \dots, R_N)* if such a decomposition is exact for a set of minimal values (R_1, R_2, \dots, R_N) , i.e. $\underline{\mathbf{X}} = \underline{\mathbf{X}}_0$. We say that a tensor $\underline{\mathbf{G}} \in \mathbb{R}^{R_1 \times R_2 \times \dots \times R_N}$ is *full-rank* if all its unfolded matrices are full-rank, i.e., $\text{rank}(\mathbf{G}_{(n)}) = R_n, \forall n$. A particularly interesting case of the Tucker model is when factor matrices $\mathbf{A}_n = \mathbf{U}_n \in \mathbb{R}^{I_n \times R_n}$ are orthogonal and chosen as the truncated matrices of left singular vectors associated with the unfolding matrices $\mathbf{X}_{(n)} = \mathbf{U}_n \Sigma_n \mathbf{V}_n^T$. In this case, we obtain the so called truncated *Higher Order Singular Value Decomposition* (HOSVD) [28]. It is noted that, in the matrix case, the truncated SVD provides the best low rank approximation having orthogonal factors and a diagonal core matrix.

C. Multi-way Projections

While in classical 1D CS, the set of compressive measurements is obtained by a linear projection, i.e. by multiplying the vector signal by a sensing matrix, in the tensor case, we can exploit its multi-way structure and use devices that provide compressive measurements by multiplying an N th-order tensor by sensing matrices in several modes, similarly or identically to the Kronecker-CS setting [11]. According to the definition of the mode- n product in eqn. (1), multiplying a data tensor by a sensing matrix in the mode- n corresponds to apply a projection to every mode- n fiber. For example, a 2D signal $\mathbf{X} \in \mathbb{R}^{I_1 \times I_2}$ can be compressively sensed by using two sensing matrices, $\Phi_1 \in \mathbb{R}^{R_1 \times I_1}$ and $\Phi_2 \in \mathbb{R}^{R_2 \times I_2}$ for mode-1 and mode-2, respectively, i.e. $\mathbf{W} = \mathbf{X} \times_1 \Phi_1 \times_2 \Phi_2 \in \mathbb{R}^{R_1 \times R_2}$ or, equivalently, $\mathbf{W} = \Phi_1 \mathbf{X} \Phi_2^T$, or $\mathbf{w} = (\Phi_2 \otimes \Phi_1)^T \mathbf{x}$, where $\mathbf{w} \in \mathbb{R}^{R_1 R_2}$ and $\mathbf{x} \in \mathbb{R}^{I_1 I_2}$ are the vectorized versions of matrices \mathbf{W} and \mathbf{X} , respectively. Thus, the objective of Kronecker-CS is to recover the signal \mathbf{X} from the measured data matrix \mathbf{W} .

In this paper, we assume that the following set of compressive multi-way measurements $\underline{\mathbf{Z}}^{(n)} \in \mathbb{R}^{R_1 \times \dots \times R_{n-1} \times I_n \times R_{n+1} \times \dots \times R_N}$ ($n = 1, 2, \dots, N$), are available:

$$\underline{\mathbf{Z}}^{(n)} = \underline{\mathbf{X}} \times_1 \Phi_1 \times_2 \dots \times_{n-1} \Phi_{n-1} \times_{n+1} \Phi_{n+1} \times_{n+2} \dots \times_N \Phi_N, \quad (3)$$

where $\Phi_n \in \mathbb{R}^{R_n \times I_n}$ ($R_n \ll I_n$) are the corresponding mode- n sensing matrices. Note that eqn. (3) indicates that the original tensor is multiplied by the set of sensing matrices in all modes except in mode- n (see Fig. 1 (top)). We also assume available the following core tensor $\underline{\mathbf{W}} \in \mathbb{R}^{R_1 \times R_2 \times \dots \times R_N}$:

$$\underline{\mathbf{W}} = \underline{\mathbf{X}} \times_1 \Phi_1 \times_2 \dots \times_N \Phi_N, \quad (4)$$

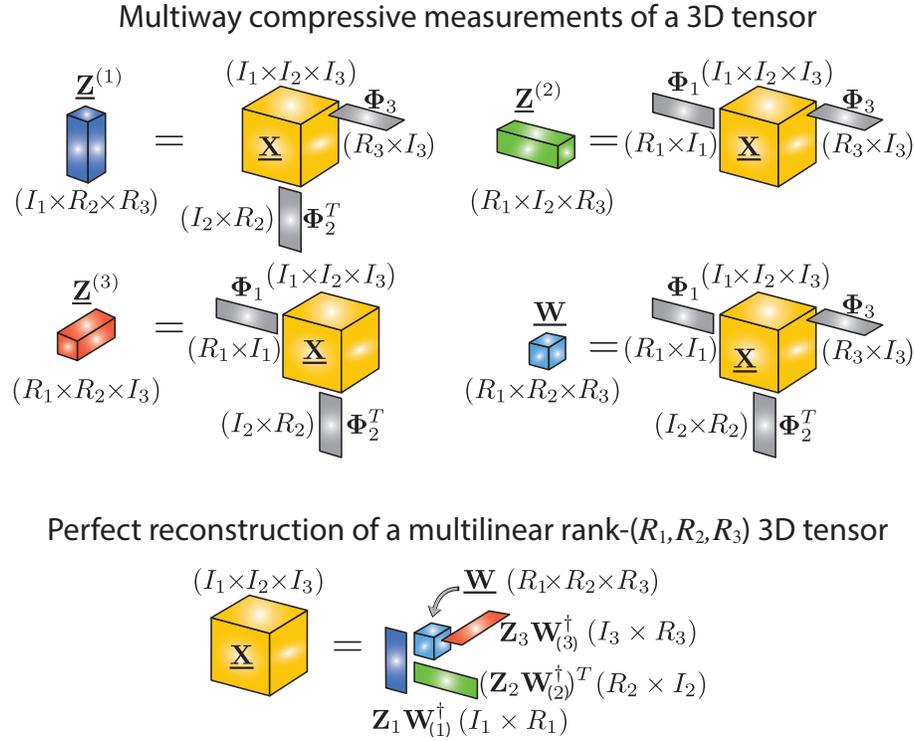


Fig. 1. Multi-way measurements and the reconstruction model for a low multilinear-rank 3D tensor.

which, in fact, is redundant since it can be computed from $\underline{\mathbf{Z}}^{(n)}$ and Φ_n taking into account that $\underline{\mathbf{W}} = \underline{\mathbf{Z}}^{(n)} \times_n \Phi_n$ for any n .

Most of state-of-the-art tensor reconstruction algorithms based on Kronecker-CS [11], [13], [14], [22], [23] assume that the only available measurement tensor is $\underline{\mathbf{W}}$, i.e. the product of the original dataset by the sensing matrices Φ_n ($n=1,2,3$) in all modes simultaneously. On the other hand, our present method requires to have the set of tensor measurements obtained by multiplying the datasets by all the sensing matrices except one, i.e. tensors $\underline{\mathbf{Z}}^{(n)}$ ($n=1,2,3$), not necessarily involving a larger amount of measurements (see experimental comparison in Section V-E). However, it is not difficult to see that already existing hardware implementations of CS imaging systems can be easily adapted in order to provide the kind of measurements required by our method. For example, in the 2D case, our method proposes to collect measurements on columns using a common sensing matrix Φ_1 , and rows using Φ_2 . By using the same ideas of the single-pixel camera developed in [29] and used in [11], our sensing operator can be obtained by using, for example, a linear array of DMDs (Digital Microarray Devices) with random orientations to sense columns and another linear array of DMDs to sense rows. On the other hand, it is also interesting to note that in [25], a different hardware implementation was used to provide real Kronecker-CS, i.e., by applying different sensing matrices to columns and rows which can be used to provide the measurements required by our method. It is also interesting to note that, recently, in [12], a new hardware implementation has been proposed that allows to employ Kronecker-CS with separable sensing operators in space and in spectral domains which could be also used to provide the measurements required by our method.

In some 3D applications, the mode-3 sensing matrix is the identity matrix, i.e. $\Phi_3 = \mathbf{I} \in \mathbb{R}^{I_3 \times I_3}$, because the same sensing operator ($\Phi_2 \otimes \Phi_1$) is applied to each frontal slice of the tensor. This is the case, for instance, in hyperspectral compressive imaging, where each frequency band (a frontal slice) is sensed by applying a different selective filter [11], [30]. This mathematical model is also valid for video sequences, where each frontal slice of the tensor corresponds to a snapshot taken at a given time.

D. The Truncated Moore-Penrose Pseudo-inverse

It is well known that, by using the Moore-Penrose (MP) pseudo-inverse of an ill-conditioned matrix $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2}$, an unstable behavior is produced since the norm $\|\mathbf{A}^\dagger\|$ could be extremely large when the smallest singular value

$\sigma_{\min}(I_1, I_2)$ is close to zero. In order to avoid this problem, the *truncated MP pseudo-inverse* $\mathbf{W}^{*\tau}$ can be used as a regularization technique, which is defined as follows [31]:

$$\mathbf{W}^{*\tau} = \mathbf{V}\mathbf{S}^{*\tau}\mathbf{U}^T, \quad (5)$$

with entries of the diagonal matrix $\mathbf{S}^{*\tau} \in \mathbb{R}^{I_2 \times I_1}$ defined as follows:

$$\sigma_i^* = \begin{cases} 1/\sigma_i, & \text{if } \sigma_i > \tau \\ 0, & \text{if } \sigma_i \leq \tau, \end{cases} \quad (6)$$

where τ is a free threshold parameter (see next sections). It is noted that $\mathbf{W}^{*\tau} \rightarrow \mathbf{W}^\dagger$ as $\tau \rightarrow 0$ and $\|\mathbf{W}^{*\tau}\| \leq 1/\max(\tau, \sigma_R)$. Also, the following properties are easily verified:

$$\mathbf{W}\mathbf{W}^{*\tau}\mathbf{W} = \mathbf{W} + \mathbf{H}, \quad (7)$$

with $\|\mathbf{H}\| \leq \tau$ if $\tau > \sigma_R$, and $\mathbf{H} = \mathbf{0}$ if $\tau \leq \sigma_R$;

$$\mathbf{W}^{*\tau}\mathbf{W}\mathbf{W}^{*\tau} = \mathbf{W}^{*\tau}, \text{ and} \quad (8)$$

$$\|\mathbf{W}\mathbf{W}^{*\tau}\| = \|\mathbf{W}^{*\tau}\mathbf{W}\| = 1. \quad (9)$$

The following lemma, provides a generalization of the property (7) to the 3rd order tensor case.

Lemma II.1. *For a given tensor $\underline{\mathbf{W}} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$ with smallest singular values in each mode σ_{R_1} , σ_{R_2} and σ_{R_3} , respectively, the following property holds:*

$$\underline{\mathbf{W}} \times_1 \mathbf{W}_{(1)}^{*\tau} \times_2 \mathbf{W}_{(2)}^{*\tau} \times_3 \mathbf{W}_{(3)}^{*\tau} = \underline{\mathbf{W}} + \underline{\mathbf{H}}, \quad (10)$$

with

$$\begin{aligned} \underline{\mathbf{H}} &= \underline{\mathbf{0}}, & \text{if } \tau \leq \underline{\sigma}, \\ \|\underline{\mathbf{H}}\|_F &\leq (\sqrt{R_1} + \sqrt{R_2} + \sqrt{R_3})\tau, & \text{if } \tau > \bar{\sigma} \end{aligned}$$

where $\bar{\sigma} = \max(\sigma_{R_n})$ and $\underline{\sigma} = \min(\sigma_{R_n})$ ($n = 1, 2, 3$).

Proof. See proof in the Appendix A. □

III. EXACT LOW MULTILINEAR-RANK TENSOR RECOVERY

The following theorem provides an explicit reconstruction formula, as illustrated in Fig. 1 (bottom), and states the conditions under which the original tensor can be exactly recovered from the set of multi-way measurements $\underline{\mathbf{Z}}^{(n)}$ ($n = 1, 2, \dots, N$) and $\underline{\mathbf{W}}$, defined in equations (3) and (4), respectively.

Theorem III.1 (Low multilinear-rank case). *If tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ has multilinear-rank- (R_1, R_2, \dots, R_N) and sensing matrices $\Phi_n \in \mathbb{R}^{R_n \times I_n}$ are such that the tensor $\underline{\mathbf{W}} = \underline{\mathbf{X}} \times_1 \Phi_1 \cdots \times_N \Phi_N \in \mathbb{R}^{R_1 \times R_2 \times \dots \times R_N}$ is full-rank, then the following reconstruction formula is exact, i.e. $\hat{\underline{\mathbf{X}}} = \underline{\mathbf{X}}$:*

$$\hat{\underline{\mathbf{X}}} = \underline{\mathbf{W}} \times_1 \mathbf{Z}_1 \mathbf{W}_{(1)}^\dagger \times_2 \cdots \times_N \mathbf{Z}_N \mathbf{W}_{(N)}^\dagger, \quad (11)$$

where “ \dagger ” stands for the MP pseudo-inverse of a matrix and $\mathbf{Z}_n \equiv (\underline{\mathbf{Z}}^{(n)})_{(n)} \in \mathbb{R}^{I_n \times \bar{R}_n}$, with $\bar{R}_n = \prod_{m \neq n} R_m$.

Proof. Let us consider the exact HOSVD decomposition $\underline{\mathbf{X}} = \underline{\mathbf{G}} \times_1 \mathbf{U}_1 \times_2 \cdots \times_N \mathbf{U}_N$, with core tensor $\underline{\mathbf{G}} \in \mathbb{R}^{R_1 \times R_2 \times \dots \times R_N}$ and orthogonal factors $\mathbf{U}_n \in \mathbb{R}^{I_n \times R_n}$, which exists because it is assumed that tensor $\underline{\mathbf{X}}$ has multilinear-rank- (R_1, R_2, \dots, R_N) . We consider, for convenience, a change of bases such that the new factors \mathbf{A}_n satisfy $\Phi_n \mathbf{A}_n = \mathbf{I} \in \mathbb{R}^{R_n \times R_n}$. We can do this by defining the new set of factors as follows:

$$\mathbf{A}_n = \mathbf{U}_n (\Phi_n \mathbf{U}_n)^{-1}, \quad (12)$$

thus, we have $\underline{\mathbf{X}} = \underline{\mathbf{G}} \times_1 \mathbf{A}_1 \times_2 \cdots \times_N \mathbf{A}_N$ where $\underline{\mathbf{G}} = \underline{\mathbf{X}} \times_1 \mathbf{A}_1^\dagger \times_2 \cdots \times_N \mathbf{A}_N^\dagger = \underline{\mathbf{W}}$. Note also that, with these new bases, the multi-way measurements are now simplified to $\underline{\mathbf{Z}}^{(n)} = \underline{\mathbf{W}} \times_n \mathbf{A}_n$ or, equivalently, $\mathbf{Z}_n = \mathbf{A}_n \mathbf{W}_{(n)}$. Taking into account that $\mathbf{W}_{(1)} \mathbf{W}_{(1)}^\dagger \mathbf{W}_{(1)} = \mathbf{W}_{(1)}$, the mode-1 unfolded version of eqn. (11) is:

$$\hat{\mathbf{X}}_{(1)} = \mathbf{A}_1 \mathbf{W}_{(1)} \left(\mathbf{Z}_N \mathbf{W}_{(N)}^\dagger \otimes \cdots \otimes \mathbf{Z}_2 \mathbf{W}_{(2)}^\dagger \right)^T, \quad (13)$$

which, can be written in terms of its associated mode-2 unfolding matrix as:

$$\hat{\mathbf{X}}_{(2)} = \mathbf{Z}_2 \mathbf{W}_{(2)}^\dagger \mathbf{W}_{(2)} \left(\mathbf{Z}_N \mathbf{W}_{(N)}^\dagger \otimes \cdots \otimes \mathbf{Z}_3 \mathbf{W}_{(3)}^\dagger \otimes \mathbf{A}_1 \right)^T.$$

Now, by substituting $\mathbf{Z}_2 = \mathbf{A}_2 \mathbf{W}_{(2)}$ in the previous equation and, by repeating this process for all modes $n = 3, 4, \dots, N$, we finally arrive at:

$$\hat{\mathbf{X}}_{(N)} = \mathbf{A}_N \mathbf{W}_{(N)} (\mathbf{A}_{N-1} \otimes \cdots \otimes \mathbf{A}_1)^T, \quad (14)$$

which proves that $\hat{\mathbf{X}} = \mathbf{X}$, since $\mathbf{W} = \mathbf{G}$. \square

It is noted that, using a simplified notation, the Tucker model of a tensor with size $I \times I \times I$ having multilinear rank (R, R, R) , requires $R^3 + 3IR$ parameters and the suggested reconstruction needs no more than $3IR^2$ measurements corresponding to tensors $\underline{\mathbf{Z}}^{(n)}$, $n = 1, 2, 3$. This is larger than the number of parameters of the associated Tucker model, however, if $R \ll I$ we see that the number of measurements is approximately R times the number of parameters, which is in general much smaller than the total number of entries I^3 .

A. Multi-way Measurements Via Linear Projections Applied to Only Two Selected Modes

It is interesting to note that all multi-way measurements defined in eqn. (3) ($n = 1, 2, \dots, N$) and eqn. (4), can be computed from linear measurements taken only in two selected modes out of $N >$. To show this, suppose that we have at our disposal the linear measurements in modes m_1 and m_2 given by: $\mathbf{Y}_m = \Phi_m \mathbf{X}_{(m)}$, with $m = m_1, m_2$; then, it is easy to see that the mode- m unfolding matrix of each multi-way measurement $\underline{\mathbf{Z}}^{(n)}$, ($n \neq m$), can be written as follows:

$$\begin{aligned} (\underline{\mathbf{Z}}^{(n)})_{(m)} &= \mathbf{Y}_m (\Phi_N^T \otimes \cdots \otimes \Phi_{m+1}^T \otimes \Phi_{m-1}^T \otimes \cdots \\ &\quad \cdots \otimes \Phi_{n+1}^T \otimes \mathbf{I} \otimes \Phi_{n-1}^T \otimes \cdots \otimes \Phi_1^T). \end{aligned}$$

For example, in the important particular case of hyperspectral images, sometimes it is easier to take compressive measurements of columns (mode-1) and rows (mode-2) of a 3rd order tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ by using the corresponding matrices $\Phi_1 \in \mathbb{R}^{R_1 \times I_1}$ and $\Phi_2 \in \mathbb{R}^{R_2 \times I_2}$ as follows:

$$\mathbf{Y}_1 = \Phi_1 \mathbf{X}_{(1)} \text{ and } \mathbf{Y}_2 = \Phi_2 \mathbf{X}_{(2)}. \quad (15)$$

It is easy to see that the multi-way measurements required by our method can be obtained from these two set of linear projections, for example, noting that: $(\underline{\mathbf{Z}}^{(1)})_{(2)} = \mathbf{Y}_2 (\Phi_3^T \otimes \mathbf{I})$, $(\underline{\mathbf{Z}}^{(2)})_{(1)} = \mathbf{Y}_1 (\Phi_3^T \otimes \mathbf{I})$, $(\underline{\mathbf{Z}}^{(3)})_{(1)} = \mathbf{Y}_1 (\mathbf{I} \otimes \Phi_2^T)$, and that tensor $\underline{\mathbf{W}}$ can be obtained using that $\underline{\mathbf{W}} = \underline{\mathbf{Z}}^{(n)} \times_n \Phi_n$ (for any n). In this case, matrix $\Phi_3 \in \mathbb{R}^{R_3 \times I_3}$ is not a real sensing matrix and its rank R_3 must be chosen in order to capture the numerical rank of the dataset in its 3rd mode, i.e. $\mathbf{X}_{(3)}$. In Algorithm 1, the steps to reconstruct a tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ from the linear projections defined in equation (15) are summarized.

Algorithm 1 : Reconstruction of tensor $\underline{\mathbf{X}}$ from linear projections taken in mode-1 and mode-2 (as defined in eqn. (15))

Require: Linear projections $\mathbf{Y}_1 \in \mathbb{R}^{R_1 \times I_2 \times I_3}$, $\mathbf{Y}_2 \in \mathbb{R}^{R_2 \times I_1 \times I_3}$, sensing matrices $\Phi_1, \Phi_2, \Phi_3 \in \mathbb{R}^{R_n \times I_n}$ ($R_n \ll I_n$)

Ensure: Tensor reconstruction $\hat{\underline{\mathbf{X}}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$

- 1: $(\underline{\mathbf{Z}}^{(1)})_{(2)} = \mathbf{Y}_2 (\Phi_3^T \otimes \mathbf{I})$; Mode-2 unfolding of $\underline{\mathbf{Z}}^{(1)}$
 - 2: $\underline{\mathbf{Z}}^{(1)} = \text{tensorize}((\underline{\mathbf{Z}}^{(1)})_{(2)})$; Tensorization
 - 3: $(\underline{\mathbf{Z}}^{(2)})_{(1)} = \mathbf{Y}_1 (\Phi_3^T \otimes \mathbf{I})$; Mode-1 unfolding of $\underline{\mathbf{Z}}^{(2)}$
 - 4: $\underline{\mathbf{Z}}^{(2)} = \text{tensorize}((\underline{\mathbf{Z}}^{(2)})_{(1)})$; Tensorization
 - 5: $(\underline{\mathbf{Z}}^{(3)})_{(1)} = \mathbf{Y}_1 (\mathbf{I} \otimes \Phi_2^T)$; Mode-1 unfolding¹ of $\underline{\mathbf{Z}}^{(3)}$
 - 6: $\underline{\mathbf{Z}}^{(3)} = \text{tensorize}((\underline{\mathbf{Z}}^{(3)})_{(1)})$; Tensorization
 - 7: $\mathbf{Z}_n = (\underline{\mathbf{Z}}^{(n)})_{(n)}$; Mode- n unfolding ($n = 1, 2, 3$)
 - 8: $\underline{\mathbf{W}} = \underline{\mathbf{Z}}^{(n)} \times_n \Phi_n$; for any $n = 1, 2$, or 3
 - 9: Compute pseudo inverses of unfolding matrices $\mathbf{W}_{(n)}$ ($n = 1, 2, 3$)
 - 10: $\hat{\underline{\mathbf{X}}} = \underline{\mathbf{W}} \times_1 \mathbf{Z}_1 \mathbf{W}_{(1)}^\dagger \times_2 \mathbf{Z}_2 \mathbf{W}_{(2)}^\dagger \times_3 \mathbf{Z}_3 \mathbf{W}_{(3)}^\dagger$,
 - 11: **return** $\hat{\underline{\mathbf{X}}}$
-

¹It is noted that the mode-2 unfolding can be used instead as follows: $(\underline{\mathbf{Z}}^{(3)})_{(2)} = \mathbf{Y}_2 (\mathbf{I} \otimes \Phi_1^T) \in \mathbb{R}^{R_2 \times I_3 \times R_1}$.

IV. STABLE RECONSTRUCTIONS OF APPROXIMATELY LOW MULTILINEAR-RANK TENSORS

The reconstruction formula of eqn. (11) is exact for the case where the tensor has multilinear-rank- (R_1, R_2, \dots, R_N) . However, in real world applications, signals usually have not low multilinear-rank, but they admit good low multilinear-rank approximations. Thus, a more realistic model for a signal should be $\underline{\mathbf{X}} = \underline{\mathbf{X}}_0 + \underline{\mathbf{E}}$ where $\underline{\mathbf{X}}_0$ has exact multilinear-rank- (R_1, R_2, \dots, R_N) and $\underline{\mathbf{E}}$ is a tensor error with sufficiently small norm, i.e. $\|\underline{\mathbf{E}}\|_F \leq \epsilon$.

In this case, one may ask about the optimal choice for sensing matrices $\Phi_n \in \mathbb{R}^{R_n \times I_n}$ in order to provide the best low multilinear-rank reconstruction. In the matrix case ($N = 2$), it is not difficult to see that, if sensing matrices are constructed using the first R singular vectors in each mode, then the obtained reconstruction is optimal (best low-rank approximation). To be more specific, given the SVD of the data matrix $\mathbf{X} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T \in \mathbb{R}^{I_1 \times I_2}$ with $\mathbf{U} = (\mathbf{U}_1 \ \mathbf{U}_2)$, $\mathbf{V} = (\mathbf{V}_1 \ \mathbf{V}_2)$ and $\mathbf{\Lambda} = \begin{pmatrix} \Lambda_1 & \mathbf{0} \\ \mathbf{0} & \Lambda_2 \end{pmatrix}$ where $\mathbf{U}_1 \in \mathbb{R}^{I_1 \times R}$ and $\mathbf{V}_1 \in \mathbb{R}^{I_2 \times R}$ are the first R left and right singular values, respectively, and Λ_1 is a diagonal matrix containing the first R singular values in its main diagonal in decreasing order, if we define the sensing matrices as follows: $\Phi_1 = \mathbf{U}_1^T \in \mathbb{R}^{R \times I_1}$, $\Phi_2 = \mathbf{V}_1^T \in \mathbb{R}^{R \times I_2}$, then the reconstruction is given by:

$$\begin{aligned} \hat{\mathbf{X}} &= \mathbf{Z}_1 \mathbf{W}^\dagger \mathbf{Z}_2^T = (\mathbf{X} \Phi_2^T) (\Phi_1 \mathbf{X} \Phi_2^T)^\dagger (\mathbf{X}^T \Phi_1^T) = (\mathbf{X} \mathbf{V}_2) (\mathbf{U}_1^T \mathbf{X} \mathbf{V}_2)^\dagger (\mathbf{X}^T \mathbf{U}_1) = \\ &= (\mathbf{U}_1 \Lambda_1) \Lambda_1^\dagger (\Lambda_1 \mathbf{V}_1^T) = \mathbf{U}_1 \Lambda_1 \mathbf{V}_1^T = \mathbf{X}_0, \end{aligned} \quad (16)$$

where \mathbf{X}_0 is, by definition the truncated SVD which is the best low rank approximation.

However, in practice we do not know the singular vectors because the original dataset is not available so we need to use sensing matrices that are independent from the dataset, for example, by generating them randomly which will give us sub-optimal reconstructions, i.e. $\|\underline{\mathbf{X}} - \hat{\underline{\mathbf{X}}}\|_F \geq \epsilon$.

In particular, we say that a reconstruction method is stable if the obtained error $\|\underline{\mathbf{X}} - \hat{\underline{\mathbf{X}}}\|_F$ is comparable to the input error ϵ , i.e. $\|\underline{\mathbf{X}} - \hat{\underline{\mathbf{X}}}\|_F \sim K\epsilon$ for some constant K . As we will show in this section, the formula given by eqn. (11) may suffer from an unstable behavior, especially in the matrix case ($N = 2$), i.e. generating large output errors even when the input error ϵ is small. We will show that we can solve this unstable behavior by using the truncated pseudo-inverse, as defined in Section II-D, instead of the exact MP pseudo-inverse. In other words, we define the *modified reconstruction* formula as follows:

$$\hat{\underline{\mathbf{X}}}_\tau = \underline{\mathbf{W}} \times_1 \mathbf{Z}_1 \mathbf{W}_{(1)}^{*\tau} \times_2 \cdots \times_N \mathbf{Z}_N \mathbf{W}_{(N)}^{*\tau}, \quad (17)$$

where $\mathbf{W}_{(n)}^{*\tau}$ is the truncated pseudo-inverse of matrix $\mathbf{W}_{(n)}$ (with the threshold parameter τ). It is noted that, when $\tau \leq \sigma_R(\mathbf{W}_n) \ \forall n$, eqn. (17) is equivalent to eqn. (11).

In the next sections we derive theoretical upper bounds for the reconstructions errors for $N = 2$ (matrix case) and for a particular case of a 3D tensor ($N = 3$), where the same 2D sensor is applied to every frontal slice. We also analyze under which conditions the parameter τ can be chosen in order to provide the minimum upper bound (optimal value τ_{opt}).

A. Error Upper Bound for the 2D case ($N=2$)

The following theorem provides an upper bound for the reconstruction of eqn. (17) and shows that the error bound approaches to zero as $\epsilon \rightarrow 0$ if $\tau = 0$ or $\tau \propto \epsilon$ (i.e., τ is proportional to the best approximation error ϵ).

Theorem IV.1. . Let matrix $\mathbf{X} \in \mathbb{R}^{I_1 \times I_2}$ be approximated by a rank- R matrix $\mathbf{X}_0 \in \mathbb{R}^{I_1 \times I_2}$, i.e. $\mathbf{X} = \mathbf{X}_0 + \mathbf{E}$ where $\|\mathbf{E}\| \leq \epsilon$ and, given sensing matrices $\Phi_1 \in \mathbb{R}^{R \times I_1}$, $\Phi_2 \in \mathbb{R}^{R \times I_2}$ such that $\mathbf{W} = \Phi_1 \mathbf{X} \Phi_2^T$ is full-rank, then the following error upper bound holds:

$$\|\mathbf{X} - \hat{\mathbf{X}}_\tau\| \leq \begin{cases} b\epsilon + c\frac{\epsilon^2}{\sigma_R}, & \text{if } \tau \leq \sigma_R, \\ a\tau + b\epsilon + c\frac{\epsilon^2}{\tau}, & \text{if } \tau > \sigma_R, \end{cases} \quad (18)$$

where σ_R is the R -th singular value (smallest) of the matrix \mathbf{W} and constants a , b and c are defined below.

Proof. Let $\mathbf{X}_0 = \mathbf{U}_1 \mathbf{\Gamma} \mathbf{U}_2^T$ be the truncated SVD of \mathbf{X}_0 , and the factors $\mathbf{A}_1, \mathbf{A}_2 \in \mathbb{R}^{I \times R}$ are defined by eqn. (12), then we obtain

$$\mathbf{X}_0 = \mathbf{A}_1 \mathbf{G} \mathbf{A}_2^T \text{ with } \mathbf{G} = \mathbf{W} - \Phi_1 \mathbf{E} \Phi_2^T. \quad (19)$$

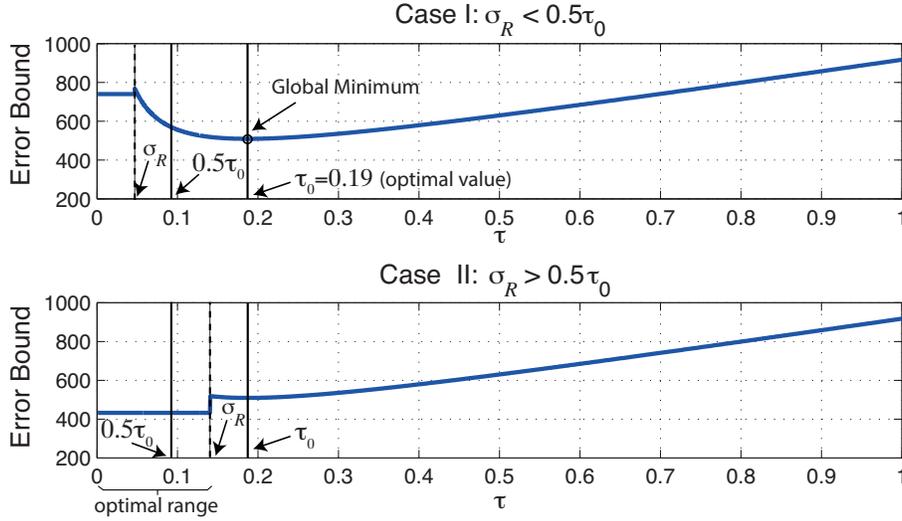


Fig. 2. Matrix Case ($N = 2$): Illustration of the two possible shapes for the error upper bound function depending on the magnitude of the singular value σ_R . Typical values of parameters a, b, c were chosen as in the digital image example of Fig. 3 ($I_1 = I_2 = 512$ and $R = 256$).

The mode-1 measurement matrix is $\mathbf{Z}_1 = \mathbf{X}_{(1)} \Phi_2^T = \mathbf{X}_0 \Phi_2^T + \mathbf{E} \Phi_2^T = \mathbf{A}_1 \mathbf{W}_{(1)} + \mathbf{F}_1$, with $\mathbf{F}_1 = (\mathbf{I} - \mathbf{A}_1 \Phi_1) \mathbf{E} \Phi_2^T$ (where we assumed that $\Phi_n \mathbf{A}_n = \mathbf{I}$), and using a similar analysis for mode-2, we obtain that $\mathbf{Z}_2 = \mathbf{A}_2 \mathbf{W}_{(2)} + \mathbf{F}_2$, with $\mathbf{F}_2 = (\mathbf{I} - \mathbf{A}_2 \Phi_2) \mathbf{E}^T \Phi_1^T$.

Using the property (8) of the truncated pseudo-inverse, the reconstructed matrix becomes:

$$\hat{\mathbf{X}}_\tau = \mathbf{Z}_1 \mathbf{W}^{*\tau} \mathbf{Z}_2^T. \quad (20)$$

Now, by inserting the expressions for \mathbf{Z}_1 and \mathbf{Z}_2 into eqn. (20) we obtain:

$$\hat{\mathbf{X}}_\tau = \mathbf{A}_1 \mathbf{W} \mathbf{W}^{*\tau} \mathbf{W} \mathbf{A}_2^T + \mathbf{A}_1 \mathbf{W} \mathbf{W}^{*\tau} \mathbf{F}_2^T + \mathbf{F}_1 \mathbf{W}^{*\tau} \mathbf{W} \mathbf{A}_2^T + \mathbf{F}_1 \mathbf{W}^{*\tau} \mathbf{F}_2^T. \quad (21)$$

Assuming that $\mathbf{W} \mathbf{W}^{*\tau} \mathbf{W} = \mathbf{W} + \mathbf{H}$ with $\|\mathbf{H}\| \leq \tau$ if $\tau > \sigma_R$, and $\mathbf{H} = \mathbf{0}$, if $\tau \leq \sigma_R$ (see Section II-D), and using $\mathbf{X}_0 = \mathbf{X} - \mathbf{E}$ and equation (19), we obtain:

$$\hat{\mathbf{X}}_\tau - \mathbf{X} = -\mathbf{E} + \mathbf{A}_1 \Phi_1 \mathbf{E} \Phi_2^T \mathbf{A}_2^T + \mathbf{A}_1 \mathbf{H} \mathbf{A}_2^T + \mathbf{A}_1 \mathbf{W} \mathbf{W}^{*\tau} \mathbf{F}_2^T + \mathbf{F}_1 \mathbf{W}^{*\tau} \mathbf{W} \mathbf{A}_2^T + \mathbf{F}_1 \mathbf{W}^{*\tau} \mathbf{F}_2^T. \quad (22)$$

If we apply the spectral norm to the last matrix equation and by using that $\|\mathbf{W}^{*\tau}\| = 1/\max(\tau, \sigma_R)$, we finally obtain

$$\|\mathbf{X} - \hat{\mathbf{X}}_\tau\| \leq \begin{cases} b\epsilon + c\frac{\epsilon^2}{\sigma_R}, & \text{if } \tau \leq \sigma_R \\ a\tau + b\epsilon + c\frac{\epsilon^2}{\tau}, & \text{if } \tau > \sigma_R, \end{cases} \quad (23)$$

where constants are identified as follows:

$$a = \|\mathbf{A}_1\| \|\mathbf{A}_2\|, \quad (24)$$

$$b = 1 + \|\mathbf{A}_1 \Phi_1\| \|\mathbf{A}_2 \Phi_2\| + \|\mathbf{A}_1\| (1 + \|\mathbf{A}_2 \Phi_2\|) \|\Phi_1\| + \|\mathbf{A}_2\| (1 + \|\mathbf{A}_1 \Phi_1\|) \|\Phi_2\|, \quad (25)$$

$$c = (1 + \|\mathbf{A}_1 \Phi_1\|) (1 + \|\mathbf{A}_2 \Phi_2\|) \|\Phi_1\| \|\Phi_2\|. \quad (26)$$

□

It is important to note that the obtained theoretical bound is not tight in general because it is based on matrix inequalities that usually are not tight, such as the triangular inequality, i.e. $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$ and other used matrix inequalities. However, it still is useful since:

- 1) The case $\epsilon = 0$ (exact case) is not realistic since always real world datasets are not low multilinear rank, so a theoretical bound, even if it is not tight is needed to characterize the reconstruction behavior in real applications;
- 2) The bound, as a function of ϵ is asymptotically linear, i.e. of order $\mathcal{O}(\epsilon)$. So as more precise the model is (smaller ϵ), much better reconstructions, and of the same order, are expected, which is intuitive but can

be only proved by a theoretical bound. In particular, it proves that, the bound approaches zero as the best low-rank approximation error $\epsilon \rightarrow 0$, when $\tau = 0$ or $\tau \propto \epsilon$, and perfect reconstruction is obtained when the signal has rank- R , as it was already proved in Theorem III.1;

B. Optimal selection of threshold parameter τ

A quite important question arises here: What is the optimal value of the threshold parameter τ ? Since the objective is to obtain a reconstruction error as small as possible, we can try to minimize the upper bound and hope that the theoretical bound gets its minimum at a point close to the minimum actual error. After a careful look at eqn. (18) as a function of τ , if we define

$$\tau_0 = \epsilon \sqrt{\frac{c}{a}}, \quad (27)$$

we can identify two different cases:

- **CASE I (small σ_R):** When $\sigma_R < \frac{1}{2}\tau_0$, the error bound is a convex function attaining its global minimum at $\tau = \tau_0$. This means that we should use $\tau_{opt} = \tau_0$. Note that the error bound for the original reconstruction formula (11) corresponds to the case of setting $\tau = 0$, which gives us a larger error bound (see Fig. 2 (top)).
- **CASE II (large σ_R):** On the other hand, when $\sigma_R > \frac{1}{2}\tau_0$, the best choice is to set $\tau < \sigma_R$, which corresponds to using the original reconstruction formula of eqn. (11), i.e., with $\tau_{opt} = 0$ (see Fig. 2 (bottom)).

However, it is important to note that, in practice, we are not able to compute the optimal parameter τ_0 because we do not know matrices \mathbf{A}_1 and \mathbf{A}_2 which depends on the SVD decomposition of the original unknown signal. However, we may use a rough overestimated approximation of this parameter by assuming that $\|\mathbf{A}_n \Phi_n\| + 1 \approx \|\mathbf{A}_n \Phi_n\|$ and using the fact that $\|\mathbf{A}_n \Phi_n\| \leq \|\mathbf{A}_n\| \|\Phi_n\|$ which provides us the following rough estimation:

$$\tau_0 \lesssim \epsilon \|\Phi_1\| \|\Phi_2\|. \quad (28)$$

But, the problem still remains challenging since usually we do not know exactly the error of the best low-rank estimation ϵ .

C. Error Upper Bound for a particular 3D case ($N=3$)

In this section we consider the recovery of a 3-rd order tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, for the particular and very important case where the same sensing operator is applied to every frontal slice of a tensor as considered hyperspectral CS imaging or video CS [11]. In other words, we assume that matrix $\Phi_3 = \mathbf{I} \in \mathbb{R}^{I_3 \times I_3}$ (i.e. $R_3 = I_3$) is the identity matrix, thus the following multi-way projections are available: $\mathbf{Z}_1 = \mathbf{X}_{(1)}(\mathbf{I} \otimes \Phi_2)^T \in \mathbb{R}^{I_1 \times R_2 I_3}$, $\mathbf{Z}_2 = \mathbf{X}_{(2)}(\mathbf{I} \otimes \Phi_1)^T \in \mathbb{R}^{I_2 \times R_1 I_3}$ and $\mathbf{Z}_3 = \mathbf{X}_{(3)}(\Phi_2 \otimes \Phi_1)^T \in \mathbb{R}^{I_3 \times R_1 R_2}$. In this case, the core tensor $\underline{\mathbf{W}}$ defined in eqn. (4) becomes $\underline{\mathbf{W}} = \underline{\mathbf{Z}}^{(3)}$. The following theorem is the 3-rd order counterpart of Theorem IV.1 and provides an upper bound for the reconstruction error. In order to make the analysis simpler, we consider that the smallest singular values of tensor $\underline{\mathbf{W}}$ in mode-1 and mode-2 are the same and it is denoted as $\sigma_R = \sigma_{R_1} = \sigma_{R_2}$.

Theorem IV.2. . Let tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ be approximated by a multilinear-rank- (R_1, R_2, R_3) tensor $\underline{\mathbf{X}}_0 \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, i.e. $\underline{\mathbf{X}} = \underline{\mathbf{X}}_0 + \underline{\mathbf{E}}$ where $\|\underline{\mathbf{E}}\|_F \leq \epsilon$ and, given sensing matrices $\Phi_1 \in \mathbb{R}^{R_1 \times I_1}$ and $\Phi_2 \in \mathbb{R}^{R_2 \times I_2}$ such that $\underline{\mathbf{W}} = \underline{\mathbf{X}} \times_1 \Phi_1 \times_2 \Phi_2$ is full-rank, then the following error upper bound holds true:

$$\|\underline{\mathbf{X}} - \hat{\underline{\mathbf{X}}}_\tau\|_F \leq \begin{cases} b\epsilon + c\frac{\epsilon^2}{\sigma_R}, & \text{if } \tau \leq \underline{\sigma} \\ a\tau + b\epsilon + c\frac{\epsilon^2}{\tau}, & \text{if } \tau > \bar{\sigma} \end{cases} \quad (29)$$

where $\sigma_R = \sigma_{R_1} = \sigma_{R_2}$, $\bar{\sigma} = \max(\sigma_R, \sigma_{R_3})$ and $\underline{\sigma} = \min(\sigma_R, \sigma_{R_3})$. (i.e., the maximum and minimum of the smallest singular values of the mode- n unfolding matrices $\mathbf{W}_{(n)}$), and constants a , b and c are defined below.

Proof. Following the same line of reasoning as used in the proof of Theorem IV.1, let $\underline{\mathbf{X}}_0 = \underline{\mathbf{I}} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$ be the truncated HOSVD (i.e., having orthogonal factors $\mathbf{U}_n \in \mathbb{R}^{I_n \times R_n}$), which always exists since tensor $\underline{\mathbf{X}}_0$ has multilinear-rank- (R_1, R_2, R_3) , by defining new factors $\mathbf{A}_n \in \mathbb{R}^{I_n \times R_n}$ according to eqn. (12) and noting that $\mathbf{A}_3 = \mathbf{I}$, we obtain:

$$\underline{\mathbf{X}}_0 = \underline{\mathbf{G}} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2, \text{ with } \underline{\mathbf{G}} = \underline{\mathbf{W}} - \underline{\mathbf{E}} \times_1 \Phi_1 \times_2 \Phi_2, \quad (30)$$

where $\underline{\mathbf{W}} = \underline{\mathbf{X}} \times_1 \Phi_1 \times_2 \Phi_2$.

The mode-1 measurement matrix is $\mathbf{Z}_1 = \mathbf{X}_{(1)}(\mathbf{I} \otimes \Phi_2)^T = (\underline{\mathbf{X}}_0)_{(1)}(\mathbf{I} \otimes \Phi_2)^T + \mathbf{E}_{(1)}(\mathbf{I} \otimes \Phi_2)^T = \mathbf{A}_1 \mathbf{W}_{(1)} + \mathbf{F}_1$, with

$$\mathbf{F}_1 = (\mathbf{I} - \mathbf{A}_1 \Phi_1) \mathbf{E}_{(1)} (\mathbf{I} \otimes \Phi_2)^T, \quad (31)$$

and, using a similar analysis as above, we obtain that $\mathbf{Z}_2 = \mathbf{A}_2 \mathbf{W}_{(2)} + \mathbf{F}_2$, with

$$\mathbf{F}_2 = (\mathbf{I} - \mathbf{A}_2 \Phi_2) \mathbf{E}_{(2)} (\mathbf{I} \otimes \Phi_1)^T, \quad (32)$$

for mode-2, and $\mathbf{Z}_3 = \mathbf{W}_{(3)}$ for mode-3. Using these expressions, the reconstructed tensor becomes:

$$\hat{\underline{\mathbf{X}}}_\tau = \underline{\Omega} \times_1 (\mathbf{A}_1 \mathbf{W}_{(1)} + \mathbf{F}_1) \times_2 (\mathbf{A}_2 \mathbf{W}_{(2)} + \mathbf{F}_2) \times_3 \mathbf{W}_{(3)}, \quad (33)$$

where $\underline{\Omega} = \underline{\mathbf{W}} \times_1 \mathbf{W}_{(1)}^{*\tau} \times_2 \mathbf{W}_{(2)}^{*\tau} \times_3 \mathbf{W}_{(3)}^{*\tau}$. The latter equation can be written in the following way

$$\hat{\underline{\mathbf{X}}}_\tau = \underline{\mathbf{B}}_1 + \underline{\mathbf{B}}_2 + \underline{\mathbf{B}}_3 + \underline{\mathbf{B}}_4, \quad (34)$$

with

$$\underline{\mathbf{B}}_1 = \underline{\Omega} \times_1 \mathbf{A}_1 \mathbf{W}_{(1)} \times_2 \mathbf{A}_2 \mathbf{W}_{(2)} \times_3 \mathbf{W}_{(3)}, \quad (35)$$

$$\underline{\mathbf{B}}_2 = \underline{\Omega} \times_1 \mathbf{A}_1 \mathbf{W}_{(1)} \times_2 \mathbf{F}_2 \times_3 \mathbf{W}_{(3)}, \quad (36)$$

$$\underline{\mathbf{B}}_3 = \underline{\Omega} \times_1 \mathbf{F}_1 \times_2 \mathbf{A}_2 \mathbf{W}_{(2)} \times_3 \mathbf{W}_{(3)}, \quad (37)$$

$$\underline{\mathbf{B}}_4 = \underline{\Omega} \times_1 \mathbf{F}_1 \times_2 \mathbf{F}_2 \times_3 \mathbf{W}_{(3)}. \quad (38)$$

By applying Lemma II.1, we have that $\underline{\Omega} \times_1 \mathbf{W}_{(1)} \times_2 \mathbf{W}_{(2)} \times_3 \mathbf{W}_{(3)} = \underline{\mathbf{W}} + \underline{\mathbf{H}}$ and, taking into account eqn. (30) and the fact that $\underline{\mathbf{X}} = \underline{\mathbf{X}}_0 + \underline{\mathbf{E}}$, we obtain that the first term in the right-hand side in eqn. (34) can be written as follows:

$$\underline{\mathbf{B}}_1 = \underline{\mathbf{X}} - \underline{\mathbf{E}} + \underline{\mathbf{E}} \times_1 \mathbf{A}_1 \Phi_1 \times_2 \mathbf{A}_2 \Phi_2 + \underline{\mathbf{H}} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2. \quad (39)$$

Now, by renaming

$$\underline{\mathbf{B}}_5 = -\underline{\mathbf{E}} + \underline{\mathbf{E}} \times_1 \mathbf{A}_1 \Phi_1 \times_2 \mathbf{A}_2 \Phi_2 + \underline{\mathbf{H}} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2, \quad (40)$$

we have that $\underline{\mathbf{B}}_1 = \underline{\mathbf{X}} + \underline{\mathbf{B}}_5$, which implies

$$\hat{\underline{\mathbf{X}}}_\tau - \underline{\mathbf{X}} = \underline{\mathbf{B}}_2 + \underline{\mathbf{B}}_3 + \underline{\mathbf{B}}_4 + \underline{\mathbf{B}}_5, \quad (41)$$

thus, we can find an upper bound of the approximation error by bounding from above each of the terms $\underline{\mathbf{B}}_2, \underline{\mathbf{B}}_3, \underline{\mathbf{B}}_4$ and $\underline{\mathbf{B}}_5$ in the last equation. Using the bounds developed in Appendix B, we finally arrive at eqn. (29), where the constants a, b and c are defined as follows:

$$a = (\sqrt{R_1} + \sqrt{R_2} + \sqrt{I_3}) \|\mathbf{A}_1\| \|\mathbf{A}_2\|,$$

$$b = 1 + \|\mathbf{A}_1 \Phi_1\| \|\mathbf{A}_2 \Phi_2\| + \|\mathbf{A}_1\| (1 + \|\mathbf{A}_2 \Phi_2\|) \|\Phi_1\| + \|\mathbf{A}_2\| (1 + \|\mathbf{A}_1 \Phi_1\|) \|\Phi_2\|,$$

$$c = (1 + \|\mathbf{A}_1 \Phi_1\|) (1 + \|\mathbf{A}_2 \Phi_2\|) \|\Phi_1\| \|\Phi_2\|.$$

□

D. Reconstruction Sensitivity to threshold parameter τ

Theorem IV.2 shows that the Frobenius error bound of the approximation of a 3D tensor has exactly the same flavor as the bound of the spectral norm for the reconstruction of matrices (Theorem IV.1). However, when we apply the method to natural images using random sensing matrices (see section V), the reconstructions of 2D datasets are always more unstable and more sensitive to the choice of threshold parameter τ , compared to the case of 3D datasets. This distinctive behavior is because, in the 2D case, the truncated MP pseudo inverse is applied to a matrix that is usually much more ill-conditioned than the ones considered in the 3D case. To be more specific, in the 2D case, \mathbf{W} is a square random ($R \times R$) matrix, thus it tends to be ill-conditioned compared to the rectangular random matrices used in the 3D case: $\mathbf{W}_{(1)} \in \mathbb{R}^{R_1 \times R_2 I_3}$, $\mathbf{W}_{(2)} \in \mathbb{R}^{R_2 \times I_3 R_1}$ and $\mathbf{W}_{(3)} \in \mathbb{R}^{I_3 \times R_1 R_2}$.

As a consequence and, in agreement with the observed behaviors in the experiments of section V, in the 2D case it is very important to use the proper threshold parameter $\tau_0 \neq 0$ defined in eqn. (27) in order to provide the smallest

reconstruction error and low variability of the results. On the other hand, in the 3D case, the method is always very stable with small reconstruction errors obtained with $\tau = 0$, i.e. computing the standard MP pseudo inverse without truncation. This is a clear advantage of considering a number of dimensions $N > 2$ because, usually, the truncation parameter τ does not need to be tuned and optimal results are obtained by using just $\tau = 0$.

V. NUMERICAL SIMULATIONS OF CS USING REAL-WORLD DATASETS

Here, we analyze various aspects of the proposed CS reconstruction method through numerical simulations and compare the performance with 2D and 3D datasets. All the simulations were performed using Matlab software on an iMac desktop computer, equipped with an Intel Core i5 processor (2.2 GHz) and 8 GB RAM. Matlab codes to reproduce the simulation results are available at <http://web.fi.uba.ar/~ccaiafa/Cesar/Low-Rank-Tensor-CS.html>. In order to evaluate the quality of the reconstructions we use the Peak Signal to Noise Ratio defined as follows: $\text{PSNR (dB)} = 20 \log_{10} (\max(\underline{\mathbf{X}}) / \|\hat{\underline{\mathbf{X}}} - \underline{\mathbf{X}}\|_F)$.

A. Selected illustrative examples for 2D and 3D signals

First, we consider the 2D digital image ‘‘Lena’’ encoded in matrix $\mathbf{X} \in \mathbb{R}^{512 \times 512}$ and assume that compressive Gaussian samples are collected for rows and columns as follows: $\mathbf{Z}_1 = \mathbf{X}\Phi_2^T$ and $\mathbf{Z}_2^T = \Phi_1\mathbf{X}$ with $\Phi_1, \Phi_2 \in \mathbb{R}^{256 \times 512}$ being matrices with independent Gaussian entries. Note that, in this case, we have $I_1 = I_2 = 512$, $R = 256$ and the core matrix $\mathbf{W} = \Phi_1\mathbf{X}\Phi_2^T$ can be computed from the available measurement matrix \mathbf{Z}_1 or \mathbf{Z}_2 multiplying it by the appropriate sensing matrix Φ_1 or Φ_2 , respectively. In Fig. 3, the original digital image is shown in the top-left panel, and right down below, its best rank- R approximation (truncated SVD) and the reconstructions obtained with $\tau = 0$ and $\tau = \tau_0$, as defined in eqn. (27), are also shown. It is clear that the optimal result is obtained with $\tau_{opt} = \tau_0 = 0.19$ with $\text{PSNR}=31.9\text{dB}$, which is considerably much higher than the value obtained with $\tau = 0 = 0.19$ ($\text{PSNR}=17.6\text{dB}$).

We also considered a 3D hyperspectral tensor image $\underline{\mathbf{X}} \in \mathbb{R}^{128 \times 128 \times 32}$, which is actually a patch extracted from ‘‘Ribeira’’, a large hyperspectral image included in a public dataset used in [32]. We also assumed Gaussian sensing matrices $\Phi_1, \Phi_2 \in \mathbb{R}^{64 \times 128}$ which provides us with the following measurement matrices: $\mathbf{Z}_1 = \mathbf{X}_{(1)}(\mathbf{I} \otimes \Phi_2)^T$, $\mathbf{Z}_2 = \mathbf{X}_{(2)}(\mathbf{I} \otimes \Phi_1)^T$ and $\mathbf{Z}_3 = \mathbf{X}_{(3)}(\Phi_2 \otimes \Phi_1)^T$. Note that, in this case, $\mathbf{W}_{(3)} = \mathbf{Z}_{(3)}$. In Fig. 3, the original tensor is shown in the top-right panel as a color RGB display, and right down below, its best multilinear-rank- $(64, 64, 32)$ approximation is shown, which was obtained by applying the Tucker Alternating Least Squares (ALS) algorithm of the Tensor Toolbox [33]. We also show that the obtained reconstructions using $\tau = 0$ ($\text{PSNR}=37.1\text{dB}$) is slightly better than the one obtained with $\tau = \tau_0 = 0.2$ ($\text{PSNR}=33.2\text{dB}$).

As it is illustrated in these examples, in general, we can say that, working with higher dimensions is an advantage because involved matrices tend to be better conditioned avoiding the difficulty of estimating the optimal τ and allowing us to use just $\tau_{opt} = 0$ (regular MP pseudo inverse).

B. Sensitivity to parameter τ analysis in the 2D and 3D cases

In order to analyze the behavior of the reconstructions as a function of the threshold parameter τ in the 2D and 3D cases, we have generated a matrix signal by using the best rank- R approximation of ‘‘Lena’’ image (512×512) and a tensor signal by using the best multilinear-rank- (R, R, I_3) approximation (Tucker-ALS) of a tensor patch used in the previous section. By adding some Gaussian noise to these base signals, we obtained our models: $\mathbf{X} = \mathbf{X}_0 + \epsilon\mathbf{E}$ (2D case) and $\underline{\mathbf{X}} = \underline{\mathbf{X}}_0 + \epsilon\underline{\mathbf{E}}$ (3D case), where $\|\mathbf{E}\| = \|\underline{\mathbf{E}}\|_F = 1$ and ϵ can be controlled by adjusting the variance of the added noise. We have normalized the matrix \mathbf{X} and tensor $\underline{\mathbf{X}}$ in order to have $\|\mathbf{X}\| = \|\underline{\mathbf{X}}\|_F = 1$.

In Fig. 4 (a), the reconstruction errors $\|\mathbf{X} - \hat{\mathbf{X}}_\tau\|$, obtained for a fixed $\epsilon = 3.1 \times 10^{-4}$, over a total number of 500 simulations (with different sensing matrices in each simulation) are shown for the 2D case and compared against the best low-rank approximation (truncated SVD). We can see that the actual error has a convex shape attaining its minimum approximately at $\tau_{opt} = \tau_0 = 0.19$ (optimal choice). In this case, we obtained an average $\sigma_R = 4.02 \times 10^{-4}$ which means that matrix \mathbf{W} is usually ill-conditioned and $\sigma_R \ll 0.5\tau_0$.

In Fig. 4 (b) the reconstruction errors for the 3D case, $\|\underline{\mathbf{X}} - \hat{\underline{\mathbf{X}}}_\tau\|$, for a fixed $\epsilon = 6.2 \times 10^{-3}$ and the best low multilinear-rank, are shown. In the latter, we observe a totally different behavior compared to the 2D case, instead of a convex function, now the minimum error is obtained by choosing $\tau_{opt} = 0$, which is also slightly lower than

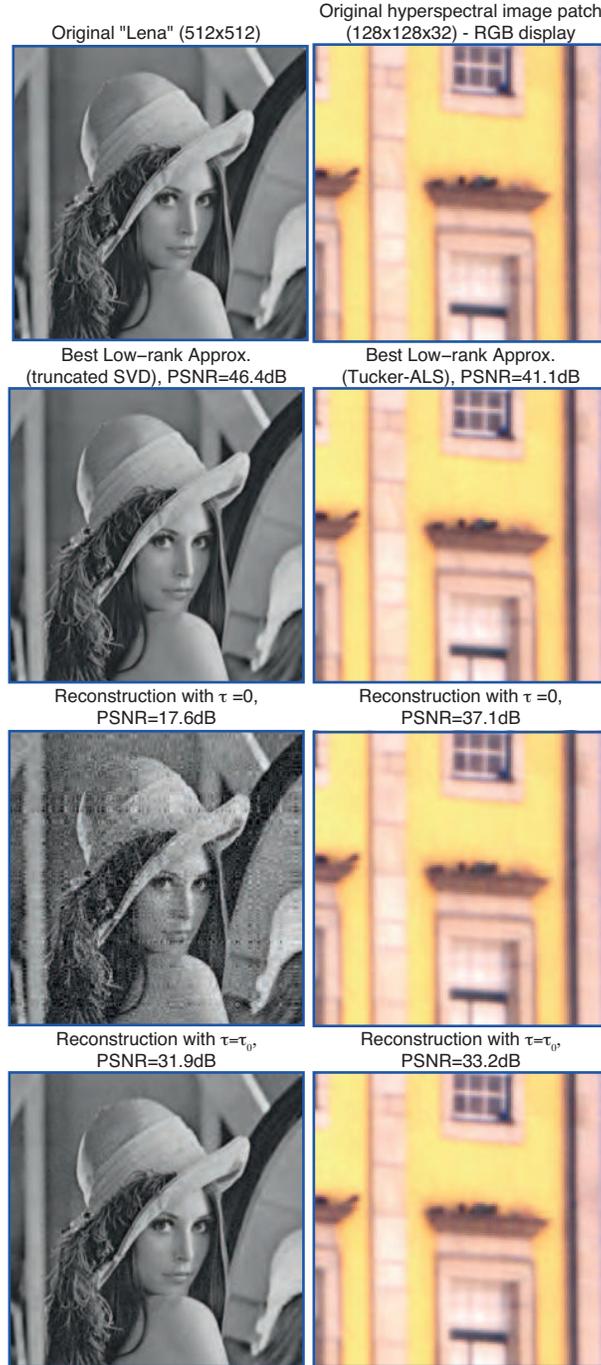


Fig. 3. Original 2D and 3D datasets (top), their best low-rank approximations (2nd row) and the reconstructions obtained by our method using $\tau = 0$ (3rd row) and $\tau = \tau_0$ (bottom). It is highlighted that, unlike in the 2D case, we can use $\tau_{opt} = 0$ as the optimal value (lower error bound) in the 3D case.

using $\tau = \tau_0 = 0.2$. It is noted that, in this case, the average smallest singular value was $\sigma_R = 0.03$, which means that matrices $\mathbf{W}_{(1)}$ and $\mathbf{W}_{(2)}$ are better conditioned than in the 2D case.

Another remarkable property of the proposed reconstruction method is that it provides very low variance of the results (robustness). In the 2D case, the standard deviation s.d. is minimal at $\tau_{opt} = \tau_0$ ($s.d. = 3.0 \times 10^{-4}$) and increases significantly for $\tau < \tau_0$. On the other side, the standard deviation of the errors obtained in the 3D case is very small ($s.d. = 3.0 \times 10^{-4}$) and approximately constant for all the range of the parameter τ values. This means, that our method is robust to the actual random sensing matrices Φ_1 and Φ_2 and, the results in the 2D case is very sensitive to the choice of parameter τ opposed to the 3D case.

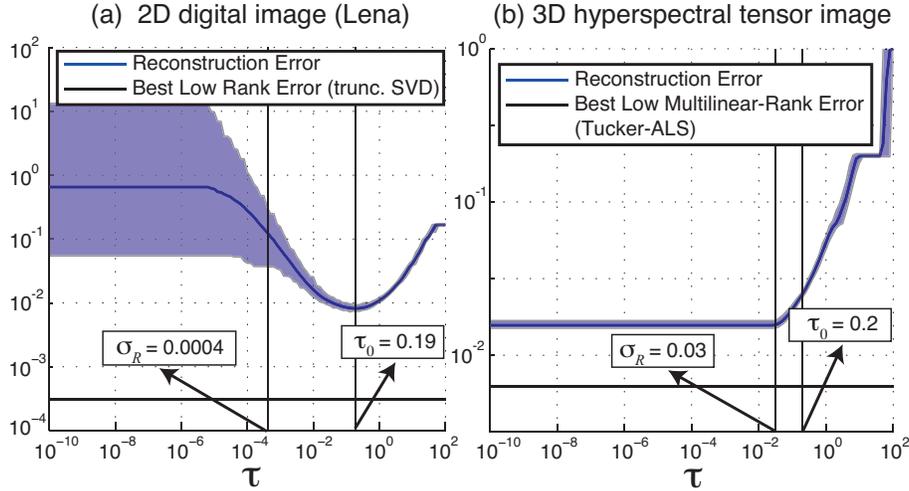


Fig. 4. Approximation errors computed on 500 simulations for “Lena” (512×512) (a) and a $(128 \times 128 \times 32)$ -patch of “Ribeira”. The range of obtained values are shown as shaded areas. It is highlighted that the standard deviation of the error, at the minimum, is $3.0e - 4$ in both cases ((a) and (b)), which means that the method is very robust to the actual choice of random sensing matrices Φ_1 and Φ_2 . Averaged σ_R and τ values are also indicated.

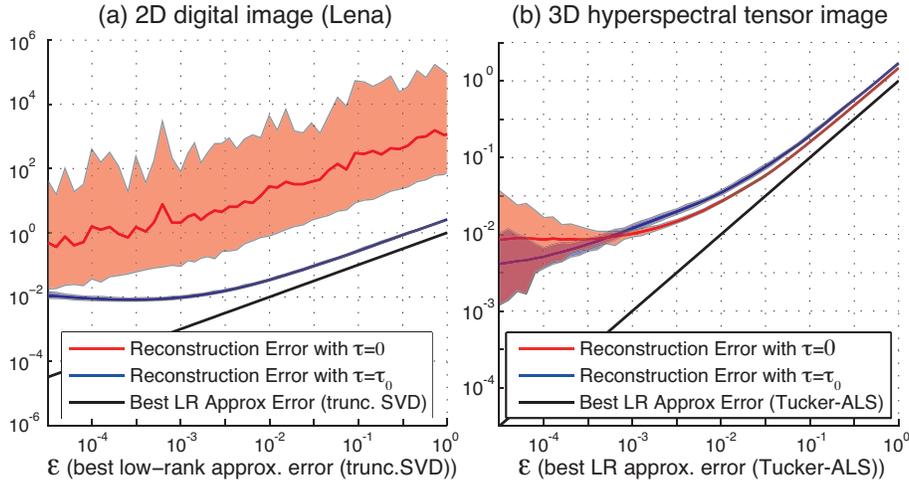


Fig. 5. Approximation errors versus ϵ for 2D case (a) and 3D case (b). It is highlighted that, while $\tau = 0$ gives poorer results (large errors and large deviations) compared to $\tau = \tau_0$ for the 2D case, in the 3D case, similar good results (small errors and deviations) are obtained for both, $\tau = 0$ and $\tau = \tau_0$.

C. Reconstruction error versus model error ϵ

In Fig. 5, using the same model as in the previous section for generating 2D and 3D signals, the performance of the reconstructions using $\tau = \tau_0$ is compared against the case of using $\tau = 0$, as a function of ϵ (best low-rank approximation error). In the 2D case (a), the original reconstruction formula (eqn. (11)) gives always poorer results, i.e., larger errors, and less robust behavior (larger deviations over repeated simulations). On the other side, when using the modified formula (eqn. (17)) with the optimal threshold parameter $\tau_{opt} = \tau_0 = \epsilon \sqrt{c/a}$, the reconstruction errors are robust and much smaller, in fact they are close to the case of the best low-rank approximation given by the truncated SVD for $\epsilon > 0.003$, approximately.

On the other hand, in the 3D case (b), robust (small deviations over repeated simulations) and small errors are obtained with $\tau = 0$, as well as with $\tau = \tau_0$. Thus, our method is less sensitive to the choice of parameter τ in the 3D case compared to the 2D case. Also, for a very small model noise ($\epsilon < 3.1 \times 10^{-4}$), better reconstructions are obtained with $\tau = \tau_0$ while, for a larger model noise ($\epsilon > 3.1 \times 10^{-4}$), using $\tau = 0$ provides slightly better reconstructions.

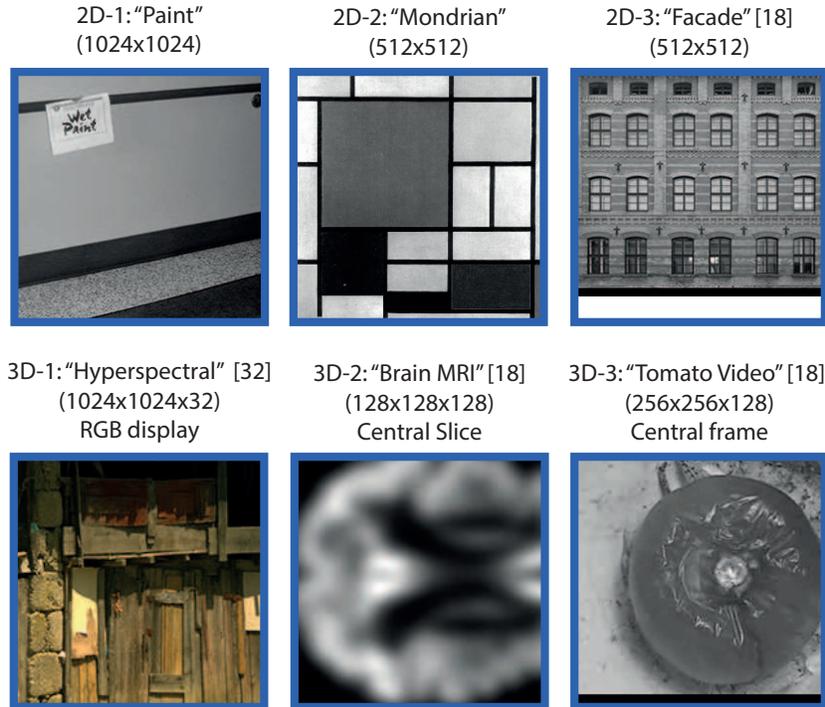


Fig. 6. 2D - images (top) and 3D tensors (bottom) used in our simulations. 3D-1 hyperspectral image corresponds to the scene ‘‘Farme’’ included in a public dataset (<http://personalpages.manchester.ac.uk/staff/david.foster/>)[32]. 3D-2 Brain MRI and 3D-3 Tomato Video datasets were used in the paper [18] and can be obtained from the author’s webpage.

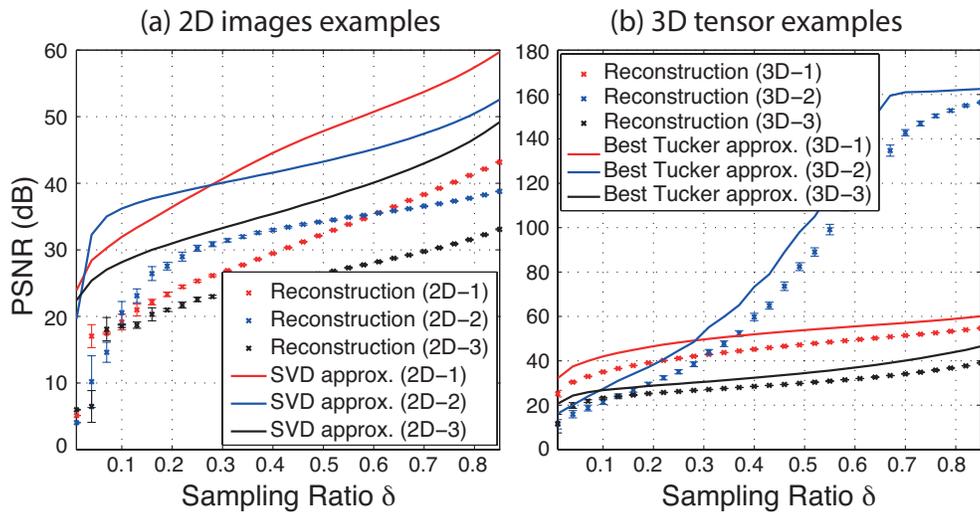


Fig. 7. Performance of reconstructions for 2D (a) and 3D (b) datasets. Mean values plus/minus the standard deviations over 100 Monte Carlo simulations are shown. PSNRs associated with the best low rank (truncated SVD) and the best multilinear-rank (obtained through the Tucker-ALS algorithm) approximations are also shown for reference.

D. Reconstruction error versus sampling ratio δ

In Fig. 7, we analyze, through 100 Monte Carlo simulations, the approximation errors obtained by our method and we compare it to the best possible low-rank approximation as a function of the sampling ratio δ , i.e. the size of the non-redundant measurement data divided by the size of the original dataset, for several 2D ($I \times I$) images (see Fig. 6-top) and different types of 3D ($I \times I \times I_3$) data tensors: a hyperspectral image [32], a Magnetic Resonance Image (MRI) of a brain [18], and a video sequence [18] (see Fig. 6-bottom).

We note that, to compute the reconstruction, matrices \mathbf{Z}_1 , \mathbf{Z}_2 and \mathbf{W} are redundant, in fact matrix \mathbf{W} can be

computed directly from measurements \mathbf{Z}_1 as follows: $\mathbf{W} = \Phi_1 \mathbf{Z}_1$. Also, by assuming the following block notation: $\mathbf{Z}_1^T = (\mathbf{Z}_{1,1}, \mathbf{Z}_{1,2})$, $\mathbf{Z}_2^T = (\mathbf{Z}_{2,1}, \mathbf{Z}_{2,2})$, $\Phi_1 = (\Phi_{1,1}, \Phi_{1,2})$ and $\Phi_2 = (\Phi_{2,1}, \Phi_{2,2})$, with $\mathbf{Z}_{1,1}, \mathbf{Z}_{2,1}, \Phi_{1,1}, \Phi_{2,1} \in \mathbb{R}^{R \times (I-R)}$ and $\mathbf{Z}_{1,2}, \mathbf{Z}_{2,2}, \Phi_{1,2}, \Phi_{2,2} \in \mathbb{R}^{R \times R}$, it is not difficult to prove that, in the 2D case, $\mathbf{Z}_{1,2}$ can be actually computed from \mathbf{Z}_2 and $\mathbf{Z}_{1,1}$ as follows:

$$\mathbf{Z}_{1,2} = (\Phi_2 \mathbf{Z}_2 - \mathbf{Z}_{1,1} \Phi_{1,1}^T) (\Phi_{1,2}^{-1})^T, \quad (42)$$

where matrix $\Phi_{1,2}$ is assumed to be invertible. Thus, the size of the minimal non-redundant measurement data required is given by $2RI - R^2$ (i.e. the sum of the sizes of matrices \mathbf{Z}_2 and $\mathbf{Z}_{1,1}$), which means that the sampling ratio can be defined as follows:

$$\delta = 2\frac{R}{I} - \left(\frac{R}{I}\right)^2. \quad (43)$$

It is easy to see that, the same sampling ratio formula is also valid for the 3D case when the sensing matrix in the mode-3 is equal to the identity matrix.

In Fig. 7 (a), the obtained PSNRs using $\tau = \tau_0$ for all the 2D datasets are shown while, in Fig. 7 (b), the obtained PSNRs using $\tau = 0$ for all the 3D datasets are shown. These results demonstrate the robustness of the method regarding the actual selection of sensing matrices Φ_1, Φ_2 , since the observed variance of PSNR is very small. It is noted that, the gap between the actual PSNR and the best (ideal) Low-Rank (LR) approximation is much smaller in the 3D case compared to the 2D case. In particular, dataset 3D-2 (Brain MRI), showed a very strong low rank structure allowing us to obtain very good reconstructions compared to the other two datasets. For instance, for $\delta = 0.4$, the obtained PSNR is around 60dB for dataset 3D-2, 45dB for dataset 3D-1 (hyperspectral image) and 28dB for dataset 3D-3 (video sequence).

E. Comparison against state-of-the-arts sparsity based algorithms

Here we compare our proposed reconstruction method against state-of-the-arts sparsity based CS algorithms for multidimensional datasets using random sensing matrices of types: a) Gaussian, and b) Bernoulli, i.e. with entries being +1 or -1 with equal probability.

In Table II (rightmost columns), we show the obtained PSNRs by applying our method to all the 2D datasets shown in Fig. 6 (top) using $R = 0.2I$, and we compare them against the results obtained by applying the Kronecker-CS (leftmost columns), consisting in using the SPGL1 algorithm [11], [26] and taking into account the Kronecker structure using Daubechies Wavelet bases. At first glance, it would seem that our proposed method uses more measurements than Kronecker-CS algorithms because we use two tensor measurements, i.e. $\underline{\mathbf{Z}}^{(1)}, \underline{\mathbf{Z}}^{(2)}$, instead of only one tensor $\underline{\mathbf{W}} \in \mathbb{R}^{R \times R}$. In order to make a more complete and fair comparison, in the central columns of Table II, we show the results of applying the Kronecker-CS method using a larger R in order to attain the same sampling ratio $\delta = 0.36$ as in our method. It is interesting to note that, even in the case of having the same sampling ratio, our method gives the best results in all cases except with the 2D-1 ‘‘Paint’’ dataset, indicating that, its sparse model is richer than its low-rank model. On the other hand, datasets 2D-2 and 2D-3 have stronger low rank structures, which allows our proposed method to obtain better reconstructions.

In Table III, we show the obtained PSNRs by applying our method to the 3D datasets shown in Fig. 6 (bottom), and we compare them against the results obtained by applying the N-Way Block Sparse Orthogonal Matching Pursuit (NBOMP) algorithm developed in [13] for the same $R = 0.125$ (leftmost columns) and same sampling ratio $\delta = 0.23$ (central columns). Note that, for example, the hyperspectral image in dataset 3D-1 is so large that it is almost impossible to apply other CS algorithms, even Kronecker-CS methods using a standard computer.

In Table IV and Table V, we compare the computation times required in each case showing that our direct method provides an extremely faster computation. For a fixed sampling ratio δ , our method is 5 orders of magnitude faster in the 2D case, and 2 orders of magnitude faster in the 3D case. For example, dataset 2D-1 ‘‘Paint’’ required almost one hour to be reconstructed by using the Kronecker-CS algorithm and our method takes only 41 milliseconds to build the reconstruction.

VI. CONCLUSIONS AND DISCUSSION

We have provided a new CS reconstruction formula for multidimensional signals assuming that a set of multi-way projections are available and that a good low multilinear-rank approximation exists. Compared to existing sparsity based CS methods, our model does not require to assume sparsity neither a dictionary based representation.

TABLE II

RECONSTRUCTION QUALITY (PSNR) OF THE PROPOSED METHOD AND THE KRONECKER-CS ALGORITHM [11] FOR 2D SIGNALS USING GAUSSIAN AND BERNOULLI SENSING MATRICES.

	Kronecker-CS $R = 0.2I$ $\delta = 0.04$		Kronecker-CS $R = 0.6I$ $\delta = 0.36$		New Method $R = 0.2I$ $\delta = 0.36$	
Data	Gaussian	Bernoulli	Gaussian	Bernoulli	Gaussian	Bernoulli
2D-1	22.8dB	22.3dB	31.7dB	30.6dB	28.3dB	28.4dB
2D-2	9.8dB	12.6dB	26.8dB	26.6dB	32.5dB	32.3dB
2D-3	13.9dB	12.7dB	21.9dB	21.7dB	24.0dB	24.2dB

TABLE III

RECONSTRUCTION QUALITY (PSNR) OF THE PROPOSED METHOD AND N-WAY BLOCK-SPARSE OMP (NBOMP) ALGORITHM [11] FOR 3D SIGNALS USING GAUSSIAN AND BERNOULLI SENSING MATRICES.

	NBOMP-CS $R = 0.125I$ $\delta = 0.016$		NBOMP-CS $R = 0.48I$ $\delta = 0.23$		New Method $R = 0.125I$ $\delta = 0.23$	
Data	Gaussian	Bernoulli	Gaussian	Bernoulli	Gaussian	Bernoulli
3D-1	21.9dB	22.4dB	39.8dB	39.9dB	40.6dB	40.5dB
3D-2	6.3dB	6.1dB	27.7dB	29.2dB	33.0dB	34.4dB
3D-3	3.6dB	3.6dB	23.1dB	22.2dB	25.8dB	25.7dB

TABLE IV

COMPUTATIONAL TIME COMPARISON BETWEEN THE PROPOSED METHOD AND KRONECKER-CS ALGORITHM [11] FOR 2D SIGNALS.

	Kronecker-CS $R = 0.2I$ $\delta = 0.04$	Kronecker-CS $R = 0.6I$ $\delta = 0.36$	New Method $R = 0.2I$ $\delta = 0.36$
Data	Time (sec.)	Time (sec.)	Time (sec.)
2D-1	5.4×10^2	3.5×10^3	4.1×10^{-2}
2D-2	1.2×10^2	6.7×10^2	8.8×10^{-3}
2D-3	1.2×10^2	7.7×10^2	9.3×10^{-3}

TABLE V

COMPUTATIONAL TIME COMPARISON BETWEEN THE PROPOSED METHOD AND N-WAY BLOCK-SPARSE OMP (NBOMP) ALGORITHM [13] FOR 3D SIGNALS.

	NBOMP-CS $R = 0.125I$ $\delta = 0.016$	NBOMP-CS $R = 0.48I$ $\delta = 0.23$	New Method $R = 0.125I$ $\delta = 0.23$
Data	Time (sec.)	Time (sec.)	Time (sec.)
3D-1	6.8	4.0×10^3	1.1×10^1
3D-2	8.2	1.8×10^1	4.4×10^{-1}
3D-3	4.1×10^1	2.4×10^2	1.8

In sparsity based CS, [26], [11], [13], available theoretical guarantees are based on sparsity levels and properties of the sensing/dictionary matrix such as RIP (restricted isometry property) and coherence. It is known that, if a signal has an exact sparse representation, then classical CS algorithms such as Matching Pursuit (MP) or Basis Pursuit (BP) are able to provide an exact reconstruction and, if the signal has an approximate sparse representation, then MP and BP provide reconstructions that are stable, i.e., they are close to the original signal with high probability. In our approach, we prove that, if the signal has an exact low multilinear-rank representation, then the proposed reconstruction is exact and, in the realistic case that the signals have only approximate low multilinear-rank representations, the reconstruction error has an upper bound that is of order $\mathcal{O}(\epsilon)$ where ϵ is the model error.

Our simulation results showed that, the present method has the following significant advantages:

- 1) It is *super fast* because it does not involve iterations making it potentially suitable for large-scale problems;
- 2) It is *stable*, in the sense that tensors which are well approximated by a Tucker model, are also well reconstructed by the proposed method,
- 3) It is *robust* because the reconstruction performance is not sensitive to the used Gaussian/Bernoulli sensing matrices.

Moreover, we have shown that, working with higher dimensions ($N > 2$), in particular with $N = 3$, gives us additional advantages because the involved matrices are better conditioned compared to the 2D case, providing more stable results and allowing us to use the standard MP pseudo inverse, i.e. without truncation.

We have shown the applicability of our method to the case of hyperspectral compressive imaging for which the technology is already available [11], [25] and applied it also to other kinds of 3D datasets such as MRI images and video sequences.

Our model is presented in a general setting which makes it hopefully applicable to next generation of multi-dimensional sensors and new methods for big-data processing under the assumption of the existence of a good multilinear rank representation and the availability of multilinear compressive measurements.

APPENDIX A: PROOF OF LEMMA II.1

Let us consider the mode-1 unfolding of the left-hand side of eqn. (10), and apply the property (7) to matrix $\mathbf{W}_{(1)}$, i.e. $\mathbf{W}_{(1)}\mathbf{W}_{(1)}^{*\tau}\mathbf{W}_{(1)} = \mathbf{W}_{(1)} + \mathbf{P}_{(1)}$, then we obtain:

$$\begin{aligned} & \mathbf{W}_{(1)}\mathbf{W}_{(1)}^{*\tau}\mathbf{W}_{(1)}(\mathbf{W}_{(3)}\mathbf{W}_{(3)}^{*\tau} \otimes \mathbf{W}_{(2)}\mathbf{W}_{(2)}^{*\tau})^T = \\ & \mathbf{W}_{(1)}(\mathbf{W}_{(3)}\mathbf{W}_{(3)}^{*\tau} \otimes \mathbf{W}_{(2)}\mathbf{W}_{(2)}^{*\tau})^T + \mathbf{P}_{(1)}(\mathbf{W}_{(3)}\mathbf{W}_{(3)}^{*\tau} \otimes \mathbf{W}_{(2)}\mathbf{W}_{(2)}^{*\tau})^T. \end{aligned}$$

The mode-2 representation of the last equation is as follows:

$$\mathbf{W}_{(2)}\mathbf{W}_{(2)}^{*\tau}\mathbf{W}_{(2)}(\mathbf{W}_{(3)}\mathbf{W}_{(3)}^{*\tau} \otimes \mathbf{I})^T + \mathbf{W}_{(2)}\mathbf{W}_{(2)}^{*\tau}\mathbf{P}_{(1)}(\mathbf{W}_{(3)}\mathbf{W}_{(3)}^{*\tau} \otimes \mathbf{I})^T.$$

Now, by applying the property (7) to matrix $\mathbf{W}_{(2)}$, i.e. $\mathbf{W}_{(2)}\mathbf{W}_{(2)}^{*\tau}\mathbf{W}_{(2)} = \mathbf{W}_{(2)} + \mathbf{Q}_{(2)}$, and using the mode-3 representation of the last equation, we arrive at:

$$\mathbf{W}_{(3)}\mathbf{W}_{(3)}^{*\tau}\mathbf{W}_{(3)} + \mathbf{W}_{(3)}\mathbf{W}_{(3)}^{*\tau}\mathbf{Q}_{(2)} + \mathbf{W}_{(3)}\mathbf{W}_{(3)}^{*\tau}\mathbf{P}_{(1)}(\mathbf{W}_{(2)}\mathbf{W}_{(2)}^{*\tau} \otimes \mathbf{I})^T.$$

Using again the property (7), i.e. $\mathbf{W}_{(3)}\mathbf{W}_{(3)}\mathbf{W}_{(3)}^{*\tau} = \mathbf{W}_{(3)} + \mathbf{R}_{(3)}$, and writing the last equation in tensor format, we finally obtain: $\underline{\mathbf{W}} + \underline{\mathbf{H}}$, with

$$\underline{\mathbf{H}} = \underline{\mathbf{R}} + \underline{\mathbf{Q}} \times_3 \mathbf{W}_{(3)}\mathbf{W}_{(3)}^{*\tau} + \underline{\mathbf{P}} \times_2 \mathbf{W}_{(2)}\mathbf{W}_{(2)}^{*\tau} \times_3 \mathbf{W}_{(3)}\mathbf{W}_{(3)}^{*\tau},$$

which, by using elementary properties of the Frobenius and spectral norms and the fact that $\|\mathbf{P}_{(1)}\|$, $\|\mathbf{Q}_{(2)}\|$, $\|\mathbf{R}_{(3)}\| \leq \tau$ ($\tau > \bar{\sigma}$) and $\|\mathbf{P}_{(1)}\| = \|\mathbf{Q}_{(2)}\| = \|\mathbf{R}_{(3)}\| = 0$ ($\tau \leq \underline{\sigma}$), implies that $\|\underline{\mathbf{H}}\|_F \leq \sqrt{R_3}\tau + \sqrt{R_2}\tau + \sqrt{R_1}\tau$ if $\tau > \bar{\sigma}$, and $\underline{\mathbf{H}} = \underline{\mathbf{0}}$ if $\tau \leq \underline{\sigma}$.

APPENDIX B: BOUNDS FOR TENSORS $\underline{\mathbf{B}}_2$, $\underline{\mathbf{B}}_3$, $\underline{\mathbf{B}}_4$ AND $\underline{\mathbf{B}}_5$

By considering the mode-2 unfolding of eqn. (36) we have that

$$\|\underline{\mathbf{B}}_2\|_F = \|\mathbf{F}_2 \mathbf{W}_{(2)}^{*\tau} \mathbf{W}_{(2)} (\mathbf{W}_{(3)} \mathbf{W}_{(3)}^{*\tau} \otimes \mathbf{A}_1 \mathbf{W}_{(1)} \mathbf{W}_{(1)}^{*\tau})^T\|_F,$$

from which, by applying the property (9) of section II-D, and using the fact that $\|\mathbf{A}\mathbf{B}\| \leq \|\mathbf{A}\| \|\mathbf{B}\|_F$, that spectral norm is a sub-multiplicative norm, and that the spectral norm of the Kronecker product of \mathbf{A} and \mathbf{B} is equal to the product of the norm of \mathbf{A} and that of \mathbf{B} , namely, $\|\mathbf{A} \otimes \mathbf{B}\| = \|\mathbf{A}\| \|\mathbf{B}\|$, we obtain that $\|\underline{\mathbf{B}}_2\|_F \leq \|\mathbf{A}_1\| \|\mathbf{F}_2\|_F$. Now, using the definition of matrix \mathbf{F}_2 (see eqn. (32)), we have that $\|\mathbf{F}_2\|_F \leq \|\mathbf{I} - \mathbf{A}_2 \Phi_2\| \|\mathbf{I} \otimes \Phi_1\| \|\underline{\mathbf{E}}\|_F$ which implies the following bound for $\underline{\mathbf{B}}_2$

$$\|\underline{\mathbf{B}}_2\|_F \leq \|\mathbf{A}_1\| (1 + \|\mathbf{A}_2 \Phi_2\|) \|\Phi_1\| \epsilon. \quad (44)$$

Analogously, we obtain the following bound for $\underline{\mathbf{B}}_3$:

$$\|\underline{\mathbf{B}}_3\|_F \leq \|\mathbf{A}_2\| (1 + \|\mathbf{A}_1 \Phi_1\|) \|\Phi_2\| \epsilon. \quad (45)$$

By considering the mode-1 representation of eqn. (38) we obtain the following inequality:

$$\|\underline{\mathbf{B}}_4\|_F \leq \|\mathbf{F}_1 \mathbf{W}_{(1)}^{*\tau} \mathbf{W}_{(1)}\|_F \|\mathbf{W}_{(3)} \mathbf{W}_{(3)}^{*\tau} \otimes \mathbf{F}_2 \mathbf{W}_{(2)}^{*\tau}\|_F,$$

and using properties of the MP pseudo-inverse and the already obtained bounds for $\|\mathbf{F}_1\|_F$ and $\|\mathbf{F}_2\|_F$ we finally arrive at:

$$\|\underline{\mathbf{B}}_4\|_F \leq (1 + \|\mathbf{A}_1 \Phi_1\|) (1 + \|\mathbf{A}_2 \Phi_2\|) \frac{\|\Phi_1\| \|\Phi_2\| \epsilon^2}{\max(\tau, \sigma_R)}.$$

From the definition of $\underline{\mathbf{B}}_5$ in eqn. (40) and by using Lemma II.1, we derive the following bounds: $\|\underline{\mathbf{B}}_5\|_F \leq \epsilon(1 + \|\mathbf{A}_1 \Phi_1\| \|\mathbf{A}_2 \Phi_2\|) + \tau(\sqrt{R_1} + \sqrt{R_2} + \sqrt{T_3}) \|\mathbf{A}_1\| \|\mathbf{A}_2\|$, if $\tau > \bar{\sigma}$, and $\|\underline{\mathbf{B}}_5\|_F \leq \epsilon(1 + \|\mathbf{A}_1 \Phi_1\| \|\mathbf{A}_2 \Phi_2\|)$, if $\tau \leq \underline{\sigma}$.

ACKNOWLEDGMENT

The authors thank anonymous Reviewers for their insightful critical comments and useful suggestions. This work was partially supported by grants from ANPCyT (Agencia Nacional de Promoción Científica y Tecnológica), PICT 2012 #1519; and CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas), PIP #114-201101-00021.

REFERENCES

- [1] E. Candès and T. Tao, "Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?" *Information Theory, IEEE Transactions on*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
- [2] D. Donoho, "Compressed Sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [3] E. Candès, J. Romberg, and T. Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.
- [4] F. Marvasti, A. Amini, F. Haddadi, M. Soltanolkotabi, B. Khalaj, A. Aldroubi, S. Sanei, and J. Chambers, "A Unified Approach to Sparse Signal Processing," *EURASIP Journal on Advances in Signal Processing*, vol. 2012, no. 1, p. 44, 2012.
- [5] Y. Eldar and G. Kutyniok, "Compressed Sensing: Theory and Applications," *New York: Cambridge Univ. Press*, vol. 20, p. 12, 2012.
- [6] G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar, "Tensor Displays: Compressive Light Field Synthesis Using Multilayer Displays with Directional Backlighting," *ACM Trans. Graph.*, vol. 31, no. 4, p. 80, 2012.
- [7] Y. Yu, J. Jin, F. Liu, and S. Crozier, "Multidimensional Compressed Sensing MRI Using Tensor Decomposition-Based Sparsifying Transform," *PLoS ONE*, vol. 9, no. 6, p. e98441, Jun. 2014.
- [8] A. Cichocki, D. Mandic, C. F. Caiafa, A. H. Phan, G. Zhou, Q. Zhao, and L. De Lathauwer, "Tensor Decompositions for Signal Processing Applications," *IEEE Signal Processing Magazine (in press)*, pp. 1–43, Dec. 2013.
- [9] V. Nico, D. Otto, S. Laurent, and L. De Lathauwer, "Breaking the Curse of Dimensionality using Decompositions of Incomplete Tensors," *Internal Report ESAT STADIUS KU Leuven*, vol. 13, pp. 1–4, Sep. 2013.
- [10] I. V. Oseledets and E. E. Tyrtshnikov, "Breaking the Curse of Dimensionality, Or How to Use SVD in Many Dimensions," *SIAM J. Scientific Computing*, vol. 31, no. 5, pp. 3744–3759, 2009.
- [11] M. F. Duarte and R. Baraniuk, "Kronecker Compressive Sensing," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 494–504, Feb. 2012.
- [12] Y. August, C. Vachman, Y. Rivenson, and A. Stern, "Compressive Hyperspectral Imaging by Random Separable Projections in both the Spatial and the Spectral Domains," *Applied optics*, vol. 52, no. 10, pp. D46–D54, Apr. 2013.
- [13] C. F. Caiafa and A. Cichocki, "Computing Sparse Representations of Multidimensional Signals Using Kronecker Bases," *Neural Computation*, pp. 186–220, Dec. 2012.

- [14] —, “Multidimensional Compressed Sensing and their Applications,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 3, no. 6, pp. 355–380, Oct. 2013.
- [15] Q. Li, D. Schonfeld, and S. Friedland, “Generalized Tensor Compressive Sensing,” *Multimedia and Expo (ICME)*, 2013.
- [16] C. F. Caiafa and A. Cichocki, “Block Sparse Representations of Tensors Using Kronecker Bases,” *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pp. 2709–2712, 2012.
- [17] E. Candes and Y. Plan, “Matrix Completion with Noise,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 925–936, 2010.
- [18] J. Liu, P. Musialski, P. Wonka, and J. Ye, “Tensor Completion for Estimating Missing Values in Visual Data,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 208–220, Jan. 2012.
- [19] E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Mørup, “Scalable Tensor Factorizations for Incomplete Data,” *Chemometrics and Intelligent Laboratory Systems*, vol. 106, no. 1, pp. 41–56, Mar. 2011.
- [20] M. Golbabaee and P. Vanderghelynst, “Hyperspectral Image Compressed Sensing via Low-rank and Joint-sparse Matrix Recovery,” in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 2741–2744.
- [21] H. Rauhut and R. Schneider, “Low-rank Tensor Recovery via Iterative Hard Thresholding,” *Proc. SAMPTA 2013*, Jun. 2013.
- [22] A. Auria, R. Carrillo, J.-P. Thiran, and Y. Wiaux, “Tensor Optimization for Optical-interferometric Imaging,” *Monthly Notices of the Royal Astronomical Society*, vol. 437, no. 3, pp. 2083–2091, Jan. 2014.
- [23] N. D. Sidiropoulos and A. Kyrillidis, “Multi-Way Compressed Sensing for Sparse Low-Rank Tensors,” *IEEE Signal Processing Letters*, vol. 19, no. 11, pp. 757–760, Nov. 2012.
- [24] C. F. Caiafa and A. Cichocki, “Fast and Stable Recovery of Approximately Low Multilinear Rank Tensors From Multi-Way Compressive Measurements,” in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*.
- [25] R. Robucci, J. Gray, L. K. C. L. K. Chiu, J. Romberg, and P. Hasler, “Compressive Sensing on a CMOS Separable-Transform Image Sensor,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1089–1101, May 2010.
- [26] Y. Rivenon and A. Stern, “Compressed Imaging with a Separable Sensing Operator,” *Signal Processing Letters, IEEE*, vol. 16, no. 6, pp. 449–452, 2009.
- [27] T. Kolda and B. Bader, “Tensor Decompositions and Applications,” *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.
- [28] L. De Lathauwer, B. De Moor, and J. Vandewalle, “A Multilinear Singular Value Decomposition,” *SIAM Journal on Matrix Analysis and Applications*, vol. 21, pp. 1253–1278, 2000.
- [29] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. Baraniuk, “Single-Pixel Imaging via Compressive Sampling,” *IEEE signal processing magazine*, vol. 25, no. 2, pp. 83–91, Mar. 2008.
- [30] R. M. Willett, M. F. Duarte, M. A. Davenport, and R. G. Baraniuk, “Sparsity and Structure in Hyperspectral Imaging : Sensing, Reconstruction, and Target Detection,” *IEEE signal processing magazine*, vol. 31, pp. 116–126, Jan. 2014.
- [31] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*, ser. Numerical Aspects of Linear Inversion. SIAM, 1998.
- [32] D. H. Foster, S. M. Nascimento, and K. Amano, “Information Limits on Neural Identification of Colored Surfaces in Natural Scenes,” *Visual Neuroscience*, vol. 21, May 2004.
- [33] B. W. Bader, T. G. Kolda, and others, “MATLAB Tensor Toolbox Version 2.5,” Sandia National Laboratories, Tech. Rep., Feb. 2012.