

Article

Spectrally Sparse Tensor Reconstruction in Optical Coherence Tomography Using Nuclear Norm Penalisation

Mohamed Ibrahim Assoweh ¹, Stéphane Chrétien ^{2,3,4,*}  and Brahim Tamadazte ^{5,6}

¹ Laboratoire de Mathématiques de Besançon, Université de Bourgogne-Franche Comté, 25030 Besançon, France; assowehmohamed@yahoo.fr

² Laboratoire ERIC, UFR ASSP, Université Lyon 2, 69676 Bron, France

³ National Physical Laboratory, Teddington TW11 0LW, UK

⁴ The Alan Turing Institute, London NW1 2DB, UK

⁵ Institute for Intelligent Systems and Robotics, University of Sorbonne, CNRS, UMR 7222, 4 pl. Jussieu, 75005 Paris, France; brahim.tamadazte@femto-st.fr

⁶ FEMTO-ST Institute, University of Bourgogne-Franche Comté, CNRS, 25000 Besançon, France

* Correspondence: stephane.chretien@univ-lyon2.fr or stephane.chretien@npl.co.uk

Received: 24 January 2020; Accepted: 7 April 2020; Published: 18 April 2020



Abstract: Reconstruction of 3D objects in various tomographic measurements is an important problem which can be naturally addressed within the mathematical framework of 3D tensors. In Optical Coherence Tomography, the reconstruction problem can be recast as a tensor completion problem. Following the seminal work of Candès et al., the approach followed in the present work is based on the assumption that the rank of the object to be reconstructed is naturally small, and we leverage this property by using a nuclear norm-type penalisation. In this paper, a detailed study of nuclear norm penalised reconstruction using the tubal Singular Value Decomposition of Kilmer et al. is proposed. In particular, we introduce a new, efficiently computable definition of the nuclear norm in the Kilmer et al. framework. We then present a theoretical analysis, which extends previous results by Koltchinskii Lounici and Tsybakov. Finally, this nuclear norm penalised reconstruction method is applied to real data reconstruction experiments in Optical Coherence Tomography (OCT). In particular, our numerical experiments illustrate the importance of penalisation for OCT reconstruction.

Keywords: tensor completion; tubal SVD; nuclear norm penalisation

1. Introduction

1.1. Motivations and Contributions

3D Image reconstruction from subsamples is a difficult problem at the intersection of the fields of inverse problem, computational statistics and numerical analysis. Following the Compressed Sensing paradigm [1], the problem can be tackled using sparsity priors in the case where sparsity can be proved present in the image to be recovered. Compressed sensing has evolved from the recovery of a sparse vector [2], to the recovery of spectrally sparse matrices [3]. The problem we are considering in the present paper was motivated by Optical Coherence Tomography (OCT), where a 3D volume is to be recovered from a small set of measurements along a ray across the volume, making the problem a kind of tensor completion problem.

The possibility of efficiently addressing the reconstruction problem in Compressed Sensing, Matrix Completion and their avatars greatly depend on formulating it as a problem which can be

relaxed as a convex optimisation problem. For example, the sparsity of a vector is measured by the number of non-zero components in that vector. This quantity is a non-convex function of the vector but a surrogate can be easily found in the form the ℓ_1 -norm. Spectral sparsity of a matrix is measured by the number of nonzero singular values and, similarly, a convex surrogate is the ℓ_1 -norm of the vector of singular values, also called the *nuclear norm*.

The main problem with tensor recovery is that the spectrum is not a well defined quantity and several approaches have been proposed for defining it [4,5]. Different nuclear norms have also been proposed in the literature, based on the various definitions for the spectrum [6–8]. A very interesting approach was proposed in References [9,10]. This approach is interesting in several ways:

- the 3D tensors are considered as matrix of 1D vectors (tubes) and the approach uses a tensor product similar to the classical matrix product, after replacing multiplication of entries by, for example, convolution of the tubes,
- the SVD is fast to compute,
- a specific and natural nuclear norm can be easily defined.

Motivated by the 3D reconstruction problem, our goal in the present paper is to study the natural nuclear norm penalised estimator for tensor completion problem. Our approach extends the method proposed in [11] to the framework developed by [10]. The main contribution of the paper is

- to present the most natural definition of the nuclear norm in the framework of Reference [10],
- to compute the subdifferential of this nuclear norm,
- to present a precise mathematical study of the nuclear norm penalised least-squares reconstruction method,
- to illustrate the efficiency of the approach on the OCT reconstruction problem with real data.

1.2. Background on Tensor Completion

1.2.1. Matrix Completion

After the many successes of Compressed Sensing in inverse problems, Matrix and tensor completion problems have recently taken the stage and become the focus of an extensive research activity. Completion problems have applications in collaborative filtering [12], Machine Learning [13,14], sensor networks [15], subspace clustering [16], signal processing [8], and so forth. The problem is intractable if the matrix to recover does not have any particular structure. An important discovery in References [17,18] is that when the matrix to be recovered has a low rank, then it can be recovered based on a few observations only [18] and using nuclear norm penalised estimation is a reasonably easy problem to solve.

The use of the nuclear norm as a convex surrogate for the rank was first proposed in Reference [19] and further analysed in a series of breakthrough papers [3,11,17,20–22].

1.2.2. Tensor Completion

The matrix completion problem was recently generalised to the problem of tensor completion. A very nice overview of tensors and algorithms is given in Reference [23]. Tensors play a fundamental role in statistics [24] and more recently in Machine Learning [13]. It can be used for multidimensional time series [25], analysis of seismic data [6], Hidden Markov Models [13], Gaussian Mixture based clustering [26], Phylogenetics [27], and much more. Tensor completion is however a more difficult problem from many points of view. First, the rank is NP-hard to compute in full generality [5]. Second, many different Singular Value Decompositions are available. The Tucker decomposition [5] extends the useful sum of rank one decomposition to tensors but it is NP-hard to compute in general. An interesting algorithm was proposed in References [28,29]; see also the very interesting Reference [30]. Another SVD was proposed in Reference [31] with the main advantage that the generalized singular vectors

form an orthogonal family. The usual diagonal matrix is replaced with a so called “core” tensor with nice orthogonality properties and very simple structure in the case of symmetric [32] or orthogonally decomposable tensors [33].

Another very interesting approach, called the t -SVD was proposed recently in [10] for 3D tensors with applications to face recognition [34] and image deblurring, computerized tomography [35] and data completion [36]. In the t -SVD framework, one dimension is assumed to be of a different nature from the others, such as, for example, time. In this setting, the decomposition resembles the SVD closely by using a diagonal tensor instead of a diagonal matrix as for the standard 2D SVD. The t -SVD was also proved to be amenable to online learning [37]. One other interesting advantage of the t -SVD is that the nuclear norm is well defined and its subdifferential is easy to compute.

The t -SVD proposed in Reference [10] is a very attractive representation of tensors in many settings such as image analysis, multivariate time series, and so forth. Obviously, it is not so attractive for the study of symmetric moment or cumulant tensors as studied in Reference [13], but for large image sequences and times series, this approach seems to be extremely relevant.

1.3. Sparsity

One of the main discoveries of the past decade is that sparsity may help in certain contexts where data are difficult or expensive to acquire [1]. When the object A_0 to recover is a vector, the fact that it may be sparse in a certain dictionary will dramatically improve recovery, as demonstrated by the recent breakthroughs in high dimensional statistics on the analysis of estimators obtained via convex optimization, such as the LASSO [38–40], the Adaptive LASSO [41] or the Elastic Net [42]. When A_0 is a matrix, the property of having a low rank may be crucial for the recovery as proved in References [11,43]. Extensions to the tensor case are studied in References [7,44,45]. Tensor completion using the t -SVD framework has been analyzed in Reference [36]. In particular, our results complement and improve on the results in Reference [36]. Reference [46] deserves special mention for using more advanced techniques based on real algebraic geometry and sums of squares decompositions.

The usual estimator in a sparse recovery problem is a nuclear norm penalized least squares such as the one studied here and defined by (27). Several other types of nuclear norms have been used for tensor estimation and completion. In particular, several nuclear norms can be naturally defined such as in for example, References [7,47,48]. It is interesting to notice that, sparsity promoting penalization of the least squares estimator crucially relies on the structure of the subdifferential of the norm involved. See for instance References [11,49] or Reference [50]. In Reference [7] a subset of the subdifferential is studied and then used in order to establish the efficiency of the penalized least squares approach. Another interesting approach is the one in [48] where an outer approximation of the subdifferential is given. In the matrix setting, the work in References [51,52] are famous for providing a neat characterization of the subdifferential of matrix norms or more generally functions of the matrix enjoying enough symmetries. In the 3D or higher dimensional setting, however, the case is much less understood. The relationship between the tensor norms and the norms of the flattening are intricate although some good bounds relating one to the other can be obtained, as, for example, in Reference [53].

The extension of previous results on low rank matrix reconstruction to the tensor setting is nontrivial but is shown to be relatively easy to obtain once the appropriate background is given. In particular, our theoretical analysis will generalise the analysis in Reference [11]. In order to do this, we provide a complete characterisation of the subdifferential of the nuclear norm. Our results will be illustrated by computational experiments for solving a problem in Optical Coherence Tomography (OCT).

1.4. Plan of the Paper

Section 2 presents the necessary mathematical background about tensors and sparse recovery. Section 3 introduces the measurement model and the present our nuclear-norm penalised estimator.

In Section 4 we prove our main theoretical result. In Section 5, our approach is finally illustrated in the context of Optical Coherence Tomography, and some computational results based on real data are presented.

2. Background on Tensors, t -SVD

In this section, we present what is meant by the notion of tensor and the various generalisations of the common objects in linear algebra to the tensor setting. In particular, we will introduce the Singular Value Decomposition proposed in Reference [10] and some associated Schatten norms.

2.1. Basic Notations for Third-Order Tensor

In this section, we recall the framework introduced by Kilmer and Martin [9,10] for a very special class of tensors which is particularly adapted to our setting.

2.1.1. Slices and Transposition

For a third-order tensor A , its (i, j, k) th entry is denoted by A_{ijk} .

Definition 1. The k th-frontal slice of A is defined as

$$A^{(k)} = A(:, :, k).$$

The j th-transversal slice of A is defined as

$$\vec{A}^{(j)} = A(:, j, :).$$

A tubal scalar (t -scalar) is an element of $\mathbb{R}^{1 \times 1 \times n_3}$ and a tubal vector (t -vector) is an element of $\mathbb{R}^{n_1 \times 1 \times n_3}$

Definition 2 (Tensor transpose). The conjugate transpose of a tensor $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ tensor A^t obtained by conjugate transposing each of the frontal slice and then reversing the order of transposed frontal slices starting from the slice number 2 to the slice number n_3 and then appending the conjugate transposed frontal slice $A^{(1)^T}$.

Definition 3 (The “dot” product). The dot product $A \cdot B$ between two tensors $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $B \in \mathbb{R}^{n_2 \times n_4 \times n_3}$ is the tensor $C \in \mathbb{R}^{n_1 \times n_4 \times n_3}$ whose slice $C^{(n)}$ is the matrix product of the slice $A^{(n)}$ with the slice $B^{(n)}$:

$$C^{(k)} := (A \cdot B)^{(k)} := A^{(k)} B^{(k)}, \quad k = 1, \dots, n_3. \quad (1)$$

We will also need the canonical inner product.

Definition 4 (Inner product of tensors). If A and B are third-order tensors of same size $n_1 \times n_2 \times n_3$, then the inner product between A and B is defined as the following (notice the normalization constant of FFT),

$$\langle A, B \rangle = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} A_{ijk} B_{ijk}. \quad (2)$$

2.1.2. Convolution and Fourier Transform

Definition 5 (t -product for circular convolution). The t -product $A * B$ of $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $B \in \mathbb{R}^{n_2 \times n_4 \times n_3}$ is an $n_1 \times n_4 \times n_3$ tensor whose (i, j) -th tube is given by

$$C(i, j, :) = \sum_{k=1}^{n_2} A(i, k, :) * B(k, j, :), \quad (3)$$

where $*$ denotes the circular convolution between two cubes of same size.

Definition 6 (Identity tensor). The identity tensor $J \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ is defined to be a tensor whose first frontal slice J^1 is the $n_1 \times n_1$ identity matrix and all other frontal slices $J^i, i = 2, \dots, n_3$ are zero.

Definition 7 (Orthogonal tensor). A tensor $Q \in \mathbb{R}^{n \times n \times n_3}$ is orthogonal if it satisfies

$$Q^\top * Q = Q * Q^\top = J. \quad (4)$$

\hat{A} is a tensor which is obtained by taking the Fast Fourier Transform (FFT) along the third dimension and we will use the following convention for Fast Transform along the 3rd dimension

$$\hat{A} = \text{fft}(A, [], 3).$$

The one-dimensional FFT along the 3th-dimension is given

$$\hat{A}(j_1, j_2, k_3) = \sum_{j_3=1}^{n_3} A(j_1, j_2, j_3) \exp(-2\frac{i\pi j_3 k_3}{n_3}), \quad \forall j_1, j_2, 1 \leq j_1 \leq n_1, 1 \leq j_2 \leq n_2.$$

Naturally, one can compute A from \hat{A} via $\text{ifft}(\hat{A}, [], 3)$ using the inverse FFT, and is defined:

$$A(j_1, j_2, k_3) = \sum_{j_3=1}^{n_3} \hat{A}(j_1, j_2, j_3) \exp(2\frac{i\pi j_3 k_3}{n_3}), \quad \forall j_1, j_2, 1 \leq j_1 \leq n_1, 1 \leq j_2 \leq n_2.$$

Definition 8 (Inverse of a tensor). The inverse of a tensor $A \in \mathbb{R}^{n \times n \times n_3}$ is written as A^{-1} satisfying

$$A^{-1} * A = A * A^{-1} = J. \quad (5)$$

where J is the identity tensor of size $n \times n \times n_3$.

Remark 1. It is proved in Reference [10] that for any tensor $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $B \in \mathbb{R}^{n_2 \times n_4 \times n_3}$, we have

$$A * B = C \Leftrightarrow \hat{A} \cdot \hat{B} = \hat{C}.$$

2.2. The t -SVD

We finally arrive at the definition of the t -SVD.

Definition 9 (f -diagonal tensor). Tensor A is called f -diagonal if each frontal slice $A^{(i)}$ is a diagonal matrix.

Definition 10 (Tensor Singular Value Decomposition: t -SVD). For $M \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the t -SVD of M is given by

$$M = U * S * V^\top, \quad (6)$$

where U and V are orthogonal tensor of size $n_1 \times n_1 \times n_3$ and $n_2 \times n_2 \times n_3$ respectively. S is a rectangular f -diagonal tensor of size $n_1 \times n_2 \times n_3$, and the entries in S are called the singular values of M . This SVD can be obtained using the Fourier transform as follows:

$$\hat{M}^{(i)} = \hat{U}^{(i)} \cdot \hat{S}^{(i)} \cdot (\hat{V}^{(i)})^\top. \quad (7)$$

This t -SVD is illustrated in Figure 1 below. Notice that the diagonal elements of S , i.e., $S(i, i, :)$ are tubal scalars as introduced in Definition 1. They will also be called *tubal eigenvalues*.

Definition 11. The spectrum $\sigma(A)$ of the tensor A is the tubal vector given by

$$\sigma(A)_i = S(i, i, :) \quad (8)$$

for $i = 1, \dots, \min\{n_1, n_2\}$.

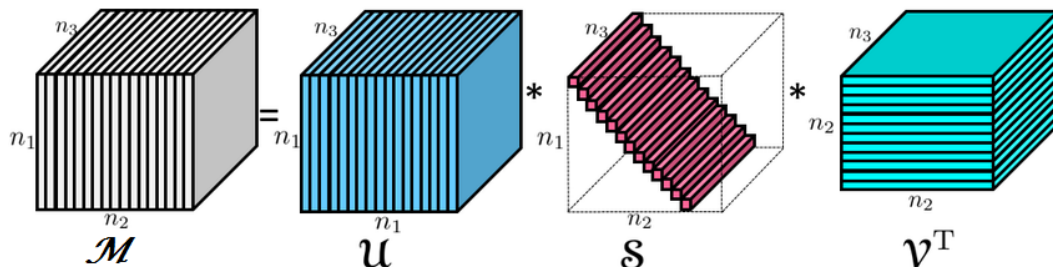


Figure 1. The t -SVD of a tensor.

2.3. Some Natural Tensor Norms

Using the previous definitions, it is easy to define some generalisations of the usual matrix norms.

Definition 12 (Tensor Frobenius norm). The induced Frobenius norm from the inner product defined above is given by,

$$\|A\|_F = \langle A, A \rangle^{1/2} = \frac{1}{\sqrt{n_3}} \|\hat{A}\|_F = \sqrt{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} A_{ijk}^2}. \quad (9)$$

Definition 13 (Tensor spectral norm). The tensor spectral norm $\|A\|_\infty$ is defined as follows:

$$\|A\|_\infty = \max_i \|\sigma(A)_i\|_2 \quad (10)$$

where $\|\cdot\|_2$ is the l_2 -norm.

Proposition 1. Let M be $n_1 \times n_2 \times n_3$ tensor. Therefore

$$\|M\|_\infty = \|\mathcal{F}(M)\|_\infty,$$

where \mathcal{F} corresponds to the Fast Fourier Transform.

Definition 14 (Tubal nuclear norm). The tensor nuclear norm of a tensor A denoted as $\|A\|_\otimes$ is the sum of singular values of all the frontal slices of A . Moreover,

$$\begin{aligned} \|A\|_\otimes &= \sum_{i=1}^{\min\{n_1, n_2\}} \sqrt{\sum_{j=1}^{n_3} S(i, i, j)^2} \\ &= \sum_{i=1}^{\min\{n_1, n_2\}} \|\sigma(A)_i\|_2. \end{aligned} \quad (11)$$

Note that by Parseval's inequality

$$\sqrt{\sum_{j=1}^{n_3} S(i, i, j)^2} = \frac{1}{\sqrt{n_3}} \sqrt{\sum_{j=1}^{n_3} \hat{S}(i, i, j)^2}. \quad (12)$$

Therefore, it is equivalent to define the tubal-nuclear norm via in the Fourier domain. Recall moreover that the $\hat{S}(i, i, j)$ are all non-negative due to the fact that $\hat{U}^{(k)} \hat{S}^{(k)} \hat{V}^{(k)^\top}$ is the SVD of the k th slice of A .

Proposition 2 (Trace duality property). *Let A, B be $n_1 \times n_2 \times n_3$ tensor. Therefore*

$$|\langle A, B \rangle| \leq \|A\|_{\otimes} \|B\|_{\infty}$$

Proof. By Cauchy-Schwartz, we have

$$\begin{aligned} |\langle A, B \rangle| &= |\langle \mathcal{F}(A), \mathcal{F}(B) \rangle| \\ &= |\langle \mathcal{F}(U) \mathcal{F}(S) \mathcal{F}(V^\top), \mathcal{F}(B) \rangle| \\ &= \left| \sum_{i=1}^{n_3} \text{tr} \left(\hat{S}^{(i)} \hat{V}^{(i)\top} \mathcal{F}(B)^{(i)\top} \hat{U}^{(i)} \right) \right| \\ &= \left| \sum_{i=1}^{n_3} \sum_{j=1}^{\min\{n_1, n_2\}} \hat{S}_{jj}^{(i)} \left(\hat{V}^{(i)\top} \mathcal{F}(B)^{(i)\top} \hat{U}^{(i)} \right)_{jj} \right| \\ &\leq \sum_{j=1}^{\min\{n_1, n_2\}} (\|\hat{S}_{jj}\|_2)^{1/2} \left(\|(\hat{V}^\top \mathcal{F}(B)^t \hat{U})_{jj}\|_2 \right)^{1/2} \\ &\leq \sum_{j=1}^{\min\{n_1, n_2\}} (\|\hat{S}_{jj}\|_2)^{1/2} (\|\mathcal{F}(B)_{jj}\|_2)^{1/2}, \end{aligned}$$

taking the maximum of $\|\mathcal{F}(B)_{jj}\|_2$ and the sum the slices of $(\|\hat{S}_{jj}\|_2)^{1/2}$, and apply (12) and inverse of FFT, we obtain the result. \square

Proposition 3. *Given tensor $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. We have*

$$\|A\|_{\otimes} \leq \sqrt{\text{rank}(A)} \|A\|_F.$$

Proof. Again by Cauchy-Schwartz, we have

$$\begin{aligned} \|A\|_{\otimes} &= \sum_{j=1}^{\min\{n_1, n_2\}} \|S(j, j, :)\|_2 \\ &= \sum_{j=1}^{\text{rank}(A)} \|S(j, j, :)\|_2 \\ &\leq \sqrt{\text{rank}(A)} \left(\sum_{j=1}^{\min\{n_1, n_2\}} \|S(j, j, :)\|_2^2 \right)^{1/2} \\ &\leq \sqrt{\text{rank}(A)} \|A\|_F. \end{aligned}$$

\square

Lemma 1. *We have*

$$\|P_{S_1^\perp} \hat{A}^\lambda P_{S_2^\perp}\|_{\otimes} = \max_{\|W\|_{\infty} \leq 1} \langle W, P_{S_1^\perp} \hat{A}^\lambda P_{S_2^\perp} \rangle$$

Proof.

$$\begin{aligned}\langle W, P_{S_1^\perp} \hat{A}^\lambda P_{S_2^\perp} \rangle &= \langle \mathcal{F}(W), \mathcal{F}(P_{S_1^\perp} \hat{A}^\lambda P_{S_2^\perp}) \rangle \quad \text{denote } B = \mathcal{F}(P_{S_1^\perp} \hat{A}^\lambda P_{S_2^\perp}) \text{ and take this t-svd} \\ &= \sum_{i=1}^{n_3} \text{tr} \left(\hat{V}^{(i)\top} \mathcal{F}(W)^{(i)\top} \hat{U}^{(i)} \hat{S}^{(i)} \right) \\ &= \sum_{i=1}^{n_3} \sum_{j=1}^{\min\{n_1, n_2\}} \left(\hat{V}^{(i)\top} \mathcal{F}(W)^{(i)\top} \hat{U}^{(i)} \right)_{jj} \hat{S}_{jj}^{(i)}\end{aligned}$$

If we take $\mathcal{F}(W)$ such as $(\hat{V}^\top \mathcal{F}(W)^\top \hat{U})_{jj}$ is colinear with \hat{S}_{jj} , i.e., $\forall j = 1, \dots, \min\{n_1, n_2\}$

$$\left(\hat{V}^\top \mathcal{F}(W)^\top \hat{U} \right)_{jj} = \alpha_j \hat{S}_{jj} \Leftrightarrow \left(\hat{V} \hat{S}_W \hat{U}^\top \right)_{jj} = \alpha_j \hat{S}_{jj} \text{ with } \alpha_j \leq 1.$$

This means to solve the equation system to determine the α_j . Thus,

$$|\alpha_j| \|\hat{S}_{jj}\|_2 = 1 \implies |\alpha_j| = \frac{1}{\|\hat{S}_{jj}\|_2}.$$

With this result, the remaining of proof follows directly from the proof of Proposition 2. \square

2.4. Rank, Range and Kernel

The rank, the range and the kernel are extremely important notions for matrices. They will play a role in our analysis of the penalised least squares tensor recovery procedure as well.

As noticed in Reference [10], a tubal scalar may have all its entrees different from zero but still be non-invertible. According to the definition, a tubal scalar $a \in \mathbb{R}^{1 \times 1 \times n_3}$ is invertible if there exists a tubal scalar b such that $a * b = b * a = e$. Equivalently, the Fourier transform \hat{a} of a has no coefficient equal to zero. We can define the tubal rank ρ_i of $S_{i,:}$ as the number of non-zero components of $\hat{S}(i, :, :)$. Then, the easiest way to define the rank of a tensor is by means of the notion of multirank as follows.

Definition 15. The multirank of a tensor is the vector (ρ_1, \dots, ρ_r) where r is the number of nonzero tubal vectors $S(i, :, :)$, $i = 1, \dots, \min\{n_1, n_2\}$.

We now define the range of a tensor.

Definition 16. Let j denote the number of invertible tubal eigenvalues and let k denote the number of nonzero noninvertible tubal eigenvalues. The range $\mathcal{R}(M)$ of a tensor $M \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is defined as

$$\mathcal{R}(M) = \left\{ \vec{U}^{(1)} * c_1 + \dots + \vec{U}^{(j+k)} * c_{j+k} \mid c_l \in \text{Range}(s_l * \cdot), l \in \{j+1, \dots, j+k\} \right\}. \quad (13)$$

Definition 17. Let j denote the number of invertible tubal eigenvalues. The kernel $\mathcal{K}(M)$ of a tensor $M \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is defined as

$$\mathcal{K}(M) = \left\{ \vec{V}^{(j+1)} * c_1 + \dots + \vec{V}^{(n_2)} * c_{n_2} \mid s_l * c_l = 0, l \in \{j+1, \dots, j+n_2\} \right\}. \quad (14)$$

3. Measurement Model and the Estimator

3.1. The Observation Model

In the model considered hereafter, the observed data are Y_1, \dots, Y_n given by the following model

$$Y_i = \langle X_i, A_0 \rangle + \xi_i, \quad i = 1, \dots, n,$$

where the notation $\langle \cdot, \cdot \rangle$ stands for the canonical scalar product of tensors. This can be seen as a tensor regression problem X_i , $i = 1, \dots, n$ are some tensors in $\mathbb{R}^{n_1 \times n_2 \times n_3}$ and ξ_i , $i = 1, \dots, n$ are some independent zero mean random variables. Assume that the frontal faces $X^{(i)}$ are i.i.d uniformly distributed on the set

$$\mathcal{X} = \{e_j(n_1)e_k^\top(n_2), 1 \leq j \leq n_1, 1 \leq k \leq n_2\}, \quad (15)$$

where $e_k(n)$ are the canonical basis vectors in \mathbb{R}^n .

Our goal is to recover the tensor A_0 based on the data Y_i , $i = 1, \dots, n$ only for n as small as possible.

Definition 18. For any tensors $A, B \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we define the scalar product

$$\langle A, B \rangle = \sum_{i_1=1, \dots, n_1} \sum_{i_2=1, \dots, n_2} \sum_{i_3=1, \dots, n_3} A_{i_1, i_2, i_3} B_{i_1, i_2, i_3},$$

and the bilinear form

$$\langle A, B \rangle_{L_2(\Pi)} = \frac{1}{n} \sum_{i=1}^n \mathbb{E} [\langle A, X_i \rangle \langle B, X_i \rangle]$$

Here $\Pi = \frac{1}{n} \sum_{i=1}^n \Pi_i$, where Π_i denotes the distribution of X_i . The corresponding semi-norm $\|A\|_{L_2(\Pi)}^2$ is given by

$$\|A\|_{L_2(\Pi)}^2 = \frac{1}{n} \sum_{i=1}^n \mathbb{E} [\langle A, X_i \rangle^2]$$

and will denote by M the tensor given by

$$M = \frac{1}{n} \sum_{i=1}^n (Y_i X_i - \mathbb{E} [Y_i X_i]).$$

3.2. The Estimator

The approach proposed in Reference [11] for low rank matrix estimation which will be extended to tensor estimation in the present paper consists in minimising

$$\hat{A}_\lambda \in \operatorname{argmin}_{A \in \mathbb{A}} L_n(A), \quad (16)$$

where

$$L_n(A) = \|A\|_{L_2(\Pi)}^2 - \left\langle \frac{2}{n} \sum_{i=1}^n Y_i X_i, A \right\rangle + \lambda \|A\|_{\otimes}, \quad (17)$$

where $\|\cdot\|_{\otimes}$ is a tubal tensor nuclear norm that we will introduce in Definition 14.

Recall that nuclear norm penalisation is widely used in sparse estimation when the matrix to be recovered is low rank [1]. Following the success of the application of sparsity to low rank matrix recovery, several extensions of the matrix nuclear norm were proposed in the literature [7,8,47], etc. Another type of nuclear norm was proposed in Reference [36] based on the tubal framework of Kilmer [10]. Our estimator is another nuclear norm penalisation based estimator. As it will be explained in Section 2 below, the nuclear norm used in the present paper has some advantages over

other norms in the context of tubal low rank tensors and the resulting estimator is most relevant in many applications where we want to recover a tensor which is the sum of a small number of rank-one tensors.

4. Main Results

In this section, we provide our main results. First, in Section 4.2, we give a complete characterisation of the subdifferential of the nuclear norm (Definition 14). Then, we propose a statistical study of the nuclear-norm penalised estimator \hat{A}_λ in Section 4.3.

4.1. Preliminary Remarks

4.1.1. Orthogonal Invariance

It is easy to see that the t -nuclear norm is orthogonally invariant. Indeed, consider two orthogonal tensors $O_1 \in \mathbb{R}^{n_1 \times n_1 \times n_3}$, $O_2 \in \mathbb{R}^{n_2 \times n_2 \times n_3}$, $k = 1, \dots, n_3$. Since the product of two orthogonal tensors is itself orthogonal, we have

$$\sigma(A) = \sigma(O_1 * A * O_2^\top). \quad (18)$$

4.1.2. Support of a Tensor

Given that the t -svd of a tensor is

$$A = U * S * V^\top \quad (19)$$

$$= \sum_{k=1}^{\min\{n_1, n_2\}} \vec{U}^{(k)} * \vec{s}_{kk} * \vec{V}^{(k)\top}, \quad (20)$$

with $(\vec{U}^{(1)}, \dots, \vec{U}^{(\min\{n_1, n_2\})})$ is a family of orthonormal matrix in $\mathbb{R}^{n_1 \times n_1}$, and $(\vec{V}^{(1)}, \dots, \vec{V}^{(\min\{n_1, n_2\})})$ a family of orthonormal matrix in $\mathbb{R}^{n_2 \times n_2}$ and $\vec{s}_{kk} = S(k, k, :)$ are the spectrum of A .

The support of A is the couple of linear vector spaces (S_1, S_2) of tubal tensors, where

- S_1 is the linear span of $(\vec{U}^{(1)}, \dots, \vec{U}^{(\min\{n_1, n_2\})})$ and
- S_2 is the linear span of $(\vec{V}^{(1)}, \dots, \vec{V}^{(\min\{n_1, n_2\})})$.

We also let S_j^\perp , $j = 1, 2$ to be the orthogonal complements of S_j and by P_{S_j} , $j = 1, 2$, the projector on linear vector subspace S of tubal tensors.

4.2. The Subdifferential of the t -nuclear Norm

Our first result is a characterisation of the subdifferential of $\|\cdot\|_\otimes$. Recall first the particular case of the matrix nuclear norm $\|\cdot\|_*$. By Corollary 2.5 in [52], we have

$$\partial \|\cdot\|_* = \{UV^\top + W \mid \|W\| \leq 1, U^\top W = 0, VW = 0\}.$$

This result is established in [52] using the Von-Neumann inequality and in particular the equality case of this inequality.

4.2.1. Von-Neumann's Inequality for Tubal Tensors

Theorem 1. Let A, B be $n_1 \times n_2 \times n_3$ tensor. Therefore

$$\langle A, B \rangle \leq \langle S^A, S^B \rangle \quad (21)$$

where S^A is a rectangular f -diagonal tensor, contains all the singular values of A .

Equality holds in (21) if and only if A, B have the same singular tensors.

Proof. Let \mathcal{F} denote the Fast Fourier Transform. We have

$$\begin{aligned}\langle A, B \rangle &= \langle \mathcal{F}(A), \mathcal{F}(B) \rangle \\ &= \left\langle \mathcal{F}\left(U^A * S^A * (V^{A^\top})\right), \mathcal{F}\left(U^B * S^B * (V^{B^\top})\right) \right\rangle \\ &= \left\langle \mathcal{F}(U^A) \cdot \mathcal{F}(S^A) \cdot \mathcal{F}(V^{A^\top}), \mathcal{F}(U^B) \cdot \mathcal{F}(S^B) \cdot \mathcal{F}(V^{B^\top}) \right\rangle \\ &= \sum_{k=1}^{n_3} \left\langle \mathcal{F}(U^A)^{(k)} \cdot \mathcal{F}(S^A)^{(k)} \cdot \left(\mathcal{F}(V^{A^\top})\right)^{(k)}, \right. \\ &\quad \left. \mathcal{F}(U^B)^{(k)} \cdot \mathcal{F}(S^B)^{(k)} \cdot \left(\mathcal{F}(V^{B^\top})\right)^{(k)} \right\rangle.\end{aligned}$$

Thus, using the Von Neumann inequality for matrices, we get

$$\begin{aligned}&\leq \sum_{k=1}^{n_3} \langle \mathcal{F}(S^A)^{(k)}, \mathcal{F}(S^B)^{(k)} \rangle \\ &\leq \langle \mathcal{F}(S^A), \mathcal{F}(S^B) \rangle \\ &= \langle S^A, S^B \rangle,\end{aligned}$$

where the last equality stems from the isometry property of Fast Fourier Transform. \square

Notice that the Von-Neumann inequality was extended in [54] to general tensors and exploited in [55] for the computation of the subdifferential of some tensor norms. In comparison, the case of the t -nuclear norm only need an appropriate use of the matrix Von Neumann inequality.

4.2.2. Lewis's Characterization of the Subdifferential

Theorem 2. Let $f : \mathbb{R}^{\min\{n_1, n_2\}} \mapsto \mathbb{R}$ be a function convex, so :

$$(f \circ \sigma)^* = f^* \circ \sigma.$$

The proof is exactly the same as in [52], Theorem 2.4.

Theorem 3. Let us suppose that the function $f : \mathbb{R}^{\min\{n_1, n_2\}} \rightarrow \mathbb{R}$ is convex. Then, the tensor

$$Y \in \partial(f \circ \sigma)(X) \text{ if and only if } \sigma(Y) \in \partial f(\sigma(X)). \quad (22)$$

The proof is the same as in [52], Corollary 2.5 where the matrix Von Neumann inequality (and more precisely, the exact characterization of the equality case) is replaced with Von Neumann inequality for tubal tensors given by Theorem 4.1.

Theorem 4. Let r denote the number of $\bar{S}_X^{(k)}$ which are non-zero. The subdifferential of the t -nuclear norm is given by

$$\begin{aligned}\partial \|X\|_{\otimes} &= \left\{ U * D(\vec{\mu}) * V^t + W \mid \vec{\mu}_k = \bar{S}_X^{(k)} / \|\bar{S}_X^{(k)}\|_2, k = 1, \dots, r, \right. \\ &\quad \left. \|\vec{\mu}_k\|_2 = 1, r < k \leq \min\{n_1, n_2\} \right\}.\end{aligned} \quad (23)$$

Proof. We only need to rewrite (22) using the well-known formula for the subdifferential of the Euclidean norm. We provide the details for the sake of completeness.

Let \vec{S}_X be a t-vector in $\mathbb{R}^{n_1 \times 1 \times n_3}$.

$$\begin{aligned}\partial f(\sigma(X)) &= \partial \left(\sum_{k=1}^{\min\{n_1, n_2\}} \|\vec{S}_X^{(k)}\|_2 \right) \\ &= \partial \|\vec{S}_X^{(1)}\|_2 \times \dots \times \|\vec{S}_X^{(\min\{n_1, n_2\})}\|_2\end{aligned}\quad (24)$$

where in second equality we use the result of [56], Chapter 16, Section 1, Proposition 16.8.

As is well known, the subgradient of the l_2 -norm is

$$\partial \|w\|_2 = \begin{cases} \left\{ \frac{w}{\|w\|_2} \right\} & \text{if } w \neq 0 \\ \{z, \|z\|_2 \leq 1\} & \text{if } w = 0 \end{cases}$$

and plugg this formula into (24). Therefore

$$\partial(f \circ \sigma)(X) = \left\{ U * D(\vec{\mu}) * V^\top \mid \vec{\mu} \in \partial f(\sigma(X)), U \in \mathbb{R}^{n_1 \times n_1 \times n_3}, V \in \mathbb{R}^{n_2 \times n_2 \times n_3}, X = U * S * V^\top \right\}. \quad (25)$$

Let T be the set of indices j which tubal scalar $\sigma(X)_j \neq 0$. Thus,

$$U * D(\vec{\mu}) * V^\top = U_T * D(\vec{\mu}_T) * V_T^\top + U_{T^c} * D(\vec{\mu}_{T^c}) * V_{T^c}^\top$$

Moreover,

$$\vec{\mu}_T = \begin{bmatrix} \frac{\vec{S}_X^{(1)}}{\|\vec{S}_X^{(1)}\|_2} \\ \vdots \\ \frac{\vec{S}_X^{(\min\{n_1, n_2\})}}{\|\vec{S}_X^{(\min\{n_1, n_2\})}\|_2} \end{bmatrix}. \quad (26)$$

Therefore, $U * D(\vec{\mu}) * V^\top$ is of the form

$$U_T * D(\vec{\mu}_T) * V_T^\top + W$$

with $W = U_{T^c} * D(\vec{\mu}_{T^c}) * V_{T^c}^\top$ and

$$\vec{\mu}_{T^c} \in \{z, \|z\|_2 \leq 1\}^{\times |T^c|}.$$

So we have,

$$\|W\|_\infty \leq 1, U_T^\top * W = 0 \text{ and } W * V_T = 0.$$

□

4.3. Error Bound

Our error bound on tubal-tensor recovery using the approach of Koltchinskii, Lounici and Tsybakov [11] is given in the following theorem.

Theorem 5. Let $\mathbb{A} \subseteq \mathbb{R}^{n_1 \times n_2 \times n_3}$ be a convex set of tensors. Let \hat{A}^λ be defined by

$$\hat{A}^\lambda = \arg \min_{A \in \mathbb{A}} \|A\|_{L_2(\Pi)}^2 - 2 \left\langle \frac{1}{n} \sum_{i=1}^n Y_i X_i, A \right\rangle + \lambda \|A\|_{\otimes}, \quad (27)$$

where we recall that $\|\cdot\|_{\otimes}$ denotes the tensor nuclear norm (see Definition 14). Assume that there exists a constant $\rho > 0$ such that, for all tensor $A \in \mathbb{A} - \mathbb{A} := \{A_1 - A_2 : A_1, A_2 \in \mathbb{A}\}$,

$$\|A\|_{L_2(\Pi)}^2 \geq \rho^{-2} \|A\|_F^2. \quad (28)$$

Let $\lambda \geq 2 \|M\|_\infty$, then

$$\|\hat{A}^\lambda - A_0\|_{L_2(\Pi)}^2 \leq \inf_{A \in \mathbb{A}} \left[\|A - A_0\|_{L_2(\Pi)}^2 + \left(\frac{1 + \sqrt{2}}{2} \right)^2 \rho^2 \lambda^2 \text{rank}(A) \right].$$

Proof. We follow the proof of Reference [11]. We provide all the details for the sake of the completeness.

$$\hat{A}^\lambda \in \arg \min_{A \in \mathbb{A}} L_n(\hat{A}^\lambda) = \|A\|_{L_2(\Pi)}^2 - 2 \left\langle \frac{1}{n} \sum_{i=1}^n Y_i X_i, A \right\rangle + \lambda \|A\|_{\otimes}. \quad (29)$$

Let us compute the directional derivative of L_n at \hat{A}^λ .

$$DL_n(\hat{A}^\lambda; h) = \lim_{t \rightarrow 0} \frac{L_n(\hat{A}^\lambda + th) - L_n(\hat{A}^\lambda)}{t}.$$

The optimality condition of \hat{A}^λ implies $DL_n(\hat{A}^\lambda; A - \hat{A}^\lambda) \geq 0$, $\forall A \in \mathbb{A}$. Thus

$$\begin{aligned} L_n(\hat{A}^\lambda + th) &= \|\hat{A}^\lambda + th\|_{L_2(\Pi)}^2 - 2 \left\langle \frac{1}{n} \sum_{i=1}^n Y_i X_i, \hat{A}^\lambda + th \right\rangle + \lambda \|\hat{A}^\lambda + th\|_{\otimes} \\ &= \|A\|_{L_2(\Pi)}^2 + 2t \langle \hat{A}^\lambda, h \rangle_{L_2(\Pi)} + t^2 \|h\|_{L_2(\Pi)}^2 - 2 \left\langle \frac{1}{n} \sum_{i=1}^n Y_i X_i, \hat{A}^\lambda \right\rangle \\ &\quad - 2t \left\langle \frac{1}{n} \sum_{i=1}^n Y_i X_i, h \right\rangle + \lambda \|\hat{A}^\lambda + th\|_{\otimes} \end{aligned}$$

$$\begin{aligned} \lim_{t \rightarrow 0} \frac{L_n(A + th) - L_n(A)}{t} &= 2 \langle \hat{A}^\lambda, h \rangle_{L_2(\Pi)} - 2 \left\langle \frac{1}{n} \sum_{i=1}^n Y_i X_i, h \right\rangle \\ &\quad + \lambda \left(\|\hat{A}^\lambda + th\|_{\otimes} - \|\hat{A}^\lambda\|_{\otimes} \right) \geq 0 \end{aligned}$$

for all $h \in T_{\mathbb{A}}(\hat{A}^\lambda)$ (the tangent cone to \mathbb{A} at \hat{A}^λ). Taking $h = A - \hat{A}^\lambda$, we have,

$$2 \langle \hat{A}^\lambda, A - \hat{A}^\lambda \rangle_{L_2(\Pi)} - 2 \left\langle \frac{1}{n} \sum_{i=1}^n Y_i X_i, A - \hat{A}^\lambda \right\rangle + \lambda \lim_{t \rightarrow 0} \left(\frac{\|\hat{A}^\lambda + th\|_{\otimes} - \|\hat{A}^\lambda\|_{\otimes}}{t} \right) \geq 0. \quad (30)$$

On the other hand,

$$\begin{aligned} \lim_{t \rightarrow 0} \left(\frac{\|\hat{A}^\lambda + th\|_{\otimes} - \|\hat{A}^\lambda\|_{\otimes}}{t} \right) &= \max_{G \in \partial \|\cdot\|_{\otimes}(\hat{A}^\lambda)} \langle G, A - \hat{A}^\lambda \rangle \\ &= \langle \hat{G}^\lambda, A - \hat{A}^\lambda \rangle \end{aligned} \quad (31)$$

by compacity of $G \in \partial \|\cdot\|_{\otimes}(\hat{A}^\lambda)$; see [57]. Combining (31) with (30), we obtain

$$2\langle \hat{A}^\lambda, \hat{A}^\lambda - A \rangle_{L_2(\Pi)} - 2 \left\langle \frac{1}{n} \sum_{i=1}^n Y_i X_i, \hat{A}^\lambda - A \right\rangle + \lambda \langle \hat{G}, \hat{A}^\lambda - A \rangle \leq 0. \quad (32)$$

Consider an arbitrary tensor $A \in \mathbb{A}$ of tubal rank r with spectral representation

$$A = U * S * V^t = \sum_{j=1}^r \vec{U}^{(j)} * \vec{s}_{jj} * \vec{V}^{(j)t} \quad (33)$$

where $\vec{s}_{jj} = S(j, j, :)$, and with support (S_1, S_2) . Using that

$$\left\langle \frac{1}{n} \sum_{i=1}^n \mathbb{E}[Y_i X_i], \hat{A}^\lambda - A \right\rangle = \langle A_0, \hat{A}^\lambda - A \rangle_{L_2(\Pi)} \quad (34)$$

it thus follows from (32) that

$$2\langle \hat{A}^\lambda - A_0, \hat{A}^\lambda - A \rangle_{L_2(\Pi)} + \lambda \langle \hat{G}^\lambda - G, \hat{A}^\lambda - A \rangle \leq -\lambda \langle G, \hat{A}^\lambda - A \rangle + 2\langle M, \hat{A}^\lambda - A \rangle \quad (35)$$

with

$$M = \frac{1}{n} \sum_{i=1}^n (Y_i X_i - \mathbb{E}[Y_i X_i]).$$

By the monotonicity of subdifferentials of convex functions, we have $\langle \hat{G} - G, \hat{A}^\lambda - A \rangle \geq 0$ (cf. [58], Chapter 4, Section 3, Proposition 9). Therefore

$$2\langle \hat{A}^\lambda - A_0, \hat{A}^\lambda - A \rangle_{L_2(\Pi)} \leq -\lambda \langle G, \hat{A}^\lambda - A \rangle + 2\langle M, \hat{A}^\lambda - A \rangle. \quad (36)$$

Furthermore, by (23), the following representations holds

$$G = \sum_{j=1}^r \vec{U}^{(j)} * \vec{\mu}_j * \vec{V}^{(j)\top} + P_{S_1^\perp} W P_{S_2^\perp}$$

where W is an arbitrary tensor with $\|W\|_\infty \leq 1$ and

$$\begin{aligned} \langle P_{S_1^\perp} W P_{S_2^\perp}, \hat{A}^\lambda - A \rangle &= \langle P_{S_1^\perp} W P_{S_2^\perp}, \hat{A}^\lambda \rangle \\ &= \langle W, P_{S_1^\perp} \hat{A}^\lambda P_{S_2^\perp} \rangle \end{aligned}$$

and using Lemma 3, we can choose W such that

$$\langle P_{S_1^\perp} W P_{S_2^\perp}, \hat{A}^\lambda - A \rangle = \|P_{S_1^\perp} \hat{A}^\lambda P_{S_2^\perp}\|_{\otimes},$$

where in the first equality we used that A has support (S_1, S_2) . For this particular choice of W , (36) implies that

$$\begin{aligned} 2\langle \hat{A}^\lambda - A_0, \hat{A}^\lambda - A \rangle_{L_2(\Pi)} + \lambda \|P_{S_1^\perp} \hat{A}^\lambda P_{S_2^\perp}\|_{\otimes} &\leq -\lambda \left\langle \sum_{j=1}^r \vec{U}^{(j)} * \vec{\mu}_j * \vec{V}^{(j)\top}, \hat{A}^\lambda - A \right\rangle \\ &\quad + 2\langle M, \hat{A}^\lambda - A \rangle \end{aligned} \quad (37)$$

Using the identity

$$2\langle \hat{A}^\lambda - A_0, \hat{A}^\lambda - A \rangle_{L_2(\Pi)} = \|\hat{A}^\lambda - A_0\|_{L_2(\Pi)}^2 + \|\hat{A}^\lambda - A\|_{L_2(\Pi)}^2 - \|A - A_0\|_{L_2(\Pi)}^2$$

and the fact that

$$\left\| \sum_{j=1}^r \vec{U}^{(j)} * \vec{\mu}_j * \vec{V}^{(j)\top} \right\|_\infty = \|D(\vec{\mu})\|_\infty$$

$$\left\langle \sum_{j=1}^r \vec{U}^{(j)} * \vec{\mu}_j * \vec{V}^{(j)\top}, \hat{A}^\lambda - A \right\rangle = \left\langle \sum_{j=1}^r \vec{U}^{(j)} * \vec{\mu}_j * \vec{V}^{(j)\top}, P_{S_1} (\hat{A}^\lambda - A) P_{S_2} \right\rangle$$

we deduce from (37) that

$$\begin{aligned} \|\hat{A}^\lambda - A_0\|_{L_2(\Pi)}^2 + \|\hat{A}^\lambda - A\|_{L_2(\Pi)}^2 + \lambda \|P_{S_1^\perp} \hat{A}^\lambda P_{S_2^\perp}\|_{\otimes} &\leq \|A - A_0\|_{L_2(\Pi)}^2 \\ &+ \lambda \|D(\vec{\mu})\|_\infty \|P_{S_1} (\hat{A}^\lambda - A) P_{S_2}\|_{\otimes} + 2\langle M, \hat{A}^\lambda - A \rangle. \end{aligned} \quad (38)$$

Now, to find an upper bound on $2\langle M, \hat{A}^\lambda - A \rangle$, we use the following decomposition:

$$\begin{aligned} \langle M, \hat{A}^\lambda - A \rangle &= \langle \mathcal{P}_A(M), \hat{A}^\lambda - A \rangle + \langle P_{S_1^\perp} M P_{S_2^\perp}, \hat{A}^\lambda - A \rangle \\ &= \langle \mathcal{P}_A(M), \mathcal{P}_A(\hat{A}^\lambda - A) \rangle + \langle P_{S_1^\perp} M P_{S_2^\perp}, \hat{A}^\lambda \rangle \end{aligned}$$

where $\mathcal{P}_A(M) = M - P_{S_1^\perp} M P_{S_2^\perp}$. This implies, due to the trace duality,

$$\begin{aligned} 2|\langle M, \hat{A}^\lambda - A \rangle| &= 2|\langle \mathcal{P}_A(M), \mathcal{P}_A(\hat{A}^\lambda - A) \rangle + \langle P_{S_1^\perp} M P_{S_2^\perp}, \hat{A}^\lambda \rangle| \\ &\leq 2|\langle \mathcal{P}_A(M), \mathcal{P}_A(\hat{A}^\lambda - A) \rangle| + 2|\langle P_{S_1^\perp} M P_{S_2^\perp}, \hat{A}^\lambda \rangle| \\ &\leq 2\|\mathcal{P}_A(M)\|_F \|\mathcal{P}_A(\hat{A}^\lambda - A)\|_F + 2\|P_{S_1^\perp} M P_{S_2^\perp}\|_\infty \|P_{S_1^\perp} \hat{A}^\lambda P_{S_2^\perp}\|_{\otimes} \\ &\leq \Lambda \|\hat{A}^\lambda - A\|_F + \Gamma \|P_{S_1^\perp} \hat{A}^\lambda P_{S_2^\perp}\|_{\otimes} \end{aligned} \quad (39)$$

where

$$\Lambda = 2\|\mathcal{P}_A(M)\|_F, \text{ and } \Gamma = 2\|P_{S_1^\perp} M P_{S_2^\perp}\|_\infty$$

Using that

$$\mathcal{P}_A(M) = P_{S_1^\perp} M P_{S_2} + P_{S_1} M \text{ and } \text{rank}(P_{S_j}) \leq \text{rank}(A), \quad j = 1, 2$$

we have with $\Delta = \|M\|_\infty$

$$\Lambda \leq 2\sqrt{\text{rank}(\mathcal{P}_A)(M)} \|M\|_\infty \leq 2\sqrt{2\text{rank}(A)} \Delta \leq \sqrt{2\text{rank}(A)} \lambda$$

Thus,

$$2|\langle M, \hat{A}^\lambda - A \rangle| \leq \sqrt{2\text{rank}(A)} \lambda \|\hat{A}^\lambda - A\|_F + 2\Delta \|P_{S_1^\perp} \hat{A}^\lambda P_{S_2^\perp}\|_{\otimes} \quad (40)$$

Due to Proposition 3, we have

$$\|P_{S_1}(\hat{A}^\lambda - A)P_{S_2}\|_{\otimes} \leq \sqrt{\text{rank}(\hat{A}^\lambda - A)}\|P_{S_1}(\hat{A}^\lambda - A)P_{S_2}\|_F \leq \sqrt{\text{rank}(\hat{A}^\lambda - A)}\|\hat{A}^\lambda - A\|_F \quad (41)$$

and using the assumption (28), it follows from (38) and (40) that

$$\begin{aligned} \|\hat{A}^\lambda - A_0\|_{L_2(\Pi)}^2 + \|\hat{A}^\lambda - A\|_{L_2(\Pi)}^2 + (\lambda - 2\Delta)\|P_{S_1^\perp}\hat{A}^\lambda P_{S_2^\perp}\|_{\otimes} &\leq \|A - A_0\|_{L_2(\Pi)}^2 + \\ &+ \left(\|D(\vec{\mu})\|_\infty + \sqrt{2}\right)\rho\lambda\sqrt{\text{rank}(A)}\|\hat{A}^\lambda - A\|_{L_2(\Pi)} \end{aligned} \quad (42)$$

Using

$$\begin{aligned} &\left(\|D(\vec{\mu})\|_\infty + \sqrt{2}\right)\rho\lambda\sqrt{\text{rank}(A)}\|\hat{A}^\lambda - A\|_{L_2(\Pi)} - \|\hat{A}^\lambda - A\|_{L_2(\Pi)}^2 \\ &\leq \frac{1}{4}\left(\|D(\vec{\mu})\|_\infty + \sqrt{2}\right)^2\rho^2\lambda^2\text{rank}(A), \end{aligned}$$

we deduce from (42) that

$$\begin{aligned} \|\hat{A}^\lambda - A_0\|_{L_2(\Pi)}^2 + (\lambda - 2\Delta)\|P_{S_1^\perp}\hat{A}^\lambda P_{S_2^\perp}\|_{\otimes} &\leq \|A - A_0\|_{L_2(\Pi)}^2 \\ &+ \frac{1}{4}\left(\|D(\vec{\mu})\|_\infty + \sqrt{2}\right)^2\rho^2\lambda^2\text{rank}(A) \end{aligned}$$

□

In Appendix A, it is proved that we can set

$$\rho = \sqrt{mn_1n_2n_3^{3/2}} \quad (43)$$

and

$$\lambda = 2n_3\|\mathcal{F}(A_0)\|_{s_\infty} \max \left\{ \sqrt{\frac{t + \log(m)}{nn_1n_2}}; \left(1 + \frac{1}{\sqrt{n_1n_2}}\right) \frac{t + \log(m)}{n} \right\}. \quad (44)$$

5. Numerical Experiments

The proposed methods were numerically validated using 3D OCT images. OCT is widely studied as a medical imaging system in many clinical applications and fundamental research. In numerous clinical purposes, OCT is considered as a very interesting technique for in situ tissue characterization known as “optical biopsy” (in opposition to the conventional physical biopsy). OCT is operating under the principle of low coherence interferometry providing micro-meter spatial resolution at several MHz A-scan (1D optical core in z direction) acquisition rate. In these experiments, we found that depth was a different coordinate from the two other coordinates and the tubal SVD approach appeared particularly relevant. One way of circumventing the problem in the case where the three coordinates have the same properties is to perform the reconstruction three times using the proposed method and take the average of the results.

5.1. Benefits of Subsampling for OCT

The OCT imaging device, as the case of the most medical imaging systems, obeys two key requirements for successful application of compressed sensing methods: (1) medical imaging is naturally compressible by sparse coding in an appropriate transform domain (e.g., wavelet, shearlet transforms, etc.) and (2) OCT scanning system (e.g., Spectral-Domain OCT, the most used) naturally

acquire encoded samples, rather than direct pixel samples (e.g., in spatial-frequency encoding). Therefore, the images resulting from Spectral-Domain OCT are sparse in native representation, hence yielding themselves well to various implementations of the ideas from Compressed Sensing theory.

Certainly, OCT enables high-speed A-scan and B-scan acquisitions (Figure 2B) but presents a serious issue when it comes to acquiring a C-scan volume. Therefore, especially in case of biological sample/tissue examination, using OCT volumes poses the problem of frame-rate acquisition (generating artifacts) as well as the data transfer i.e., several hundred Mo for each volume (Figure 2B).

Indeed, OCT volume data can be considered as a $n_1 \times n_2 \times n_3$ tensor of voxels. Thereby, the mathematical methods and materials related to tensors study are well suited for 3D OCT data.

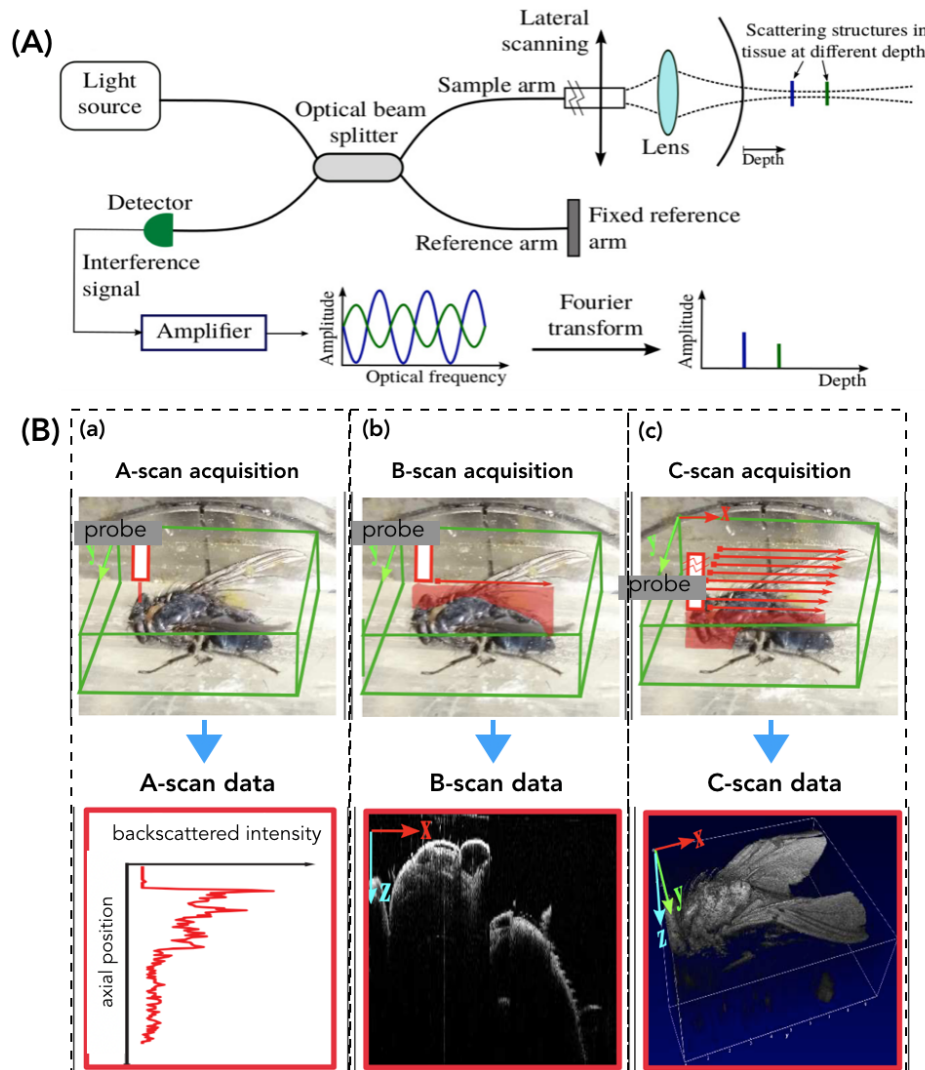


Figure 2. Illustration of the Fourier-Domain Optical Coherence Tomography (OCT) operating. (A) the optical principle and (B) the different available acquisition modes: A-scan (1D optical core), B-Scan (2D image), and C-scan (volume).

5.2. Experimental Set-Up

An OCT imaging system (*Telesto-II* 1325 nm spectral domain OCT) from THORLABS (Figure 3), is used to validate the proposed distortion models. Axial (resp. lateral) resolution is 5.5 μm (resp. 7 μm) and up to 3.5 mm depth. The *Telesto-II* allows a maximum field-of-view of $10 \times 10 \times 3.5 \text{ mm}^3$ with a maximum A-Scan (optical core) acquisition rate of 76 kHz.

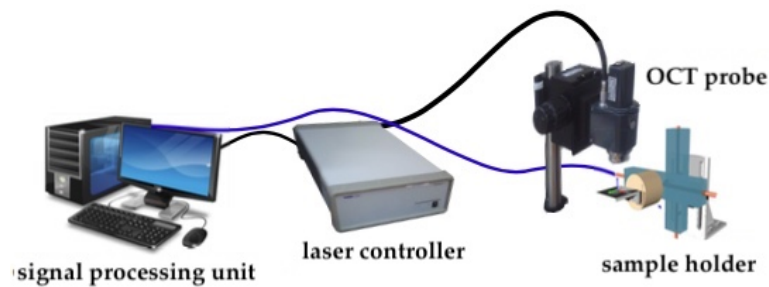


Figure 3. Global view of the OCT acquisition setup.

5.3. Validation Scenario

The method proposed in this paper was implemented in a MATLAB framework without taking into account the code optimization aspects as well as the time-computation. In order to validate experimentally the approach presented here, we acquired different OCT volumes (C-Scan images) of realistic biological samples: for example, a part of a grape (Figure 4 (left)) and a part of a fish eye retina (Figure 4 (right)). The size of the OCT volume used in these numerical validations, for both samples, is $A_{n \times m \times l} = 281 \times 281 \times 281$ voxels.

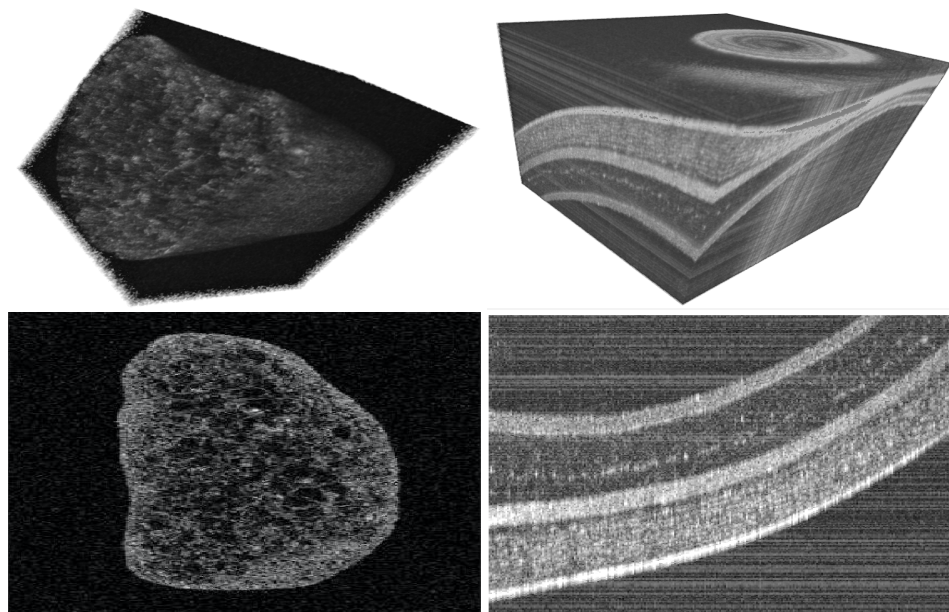


Figure 4. Examples of the OCT volumes of biological samples used to validate the proposed method. (first row) the initial OCT volumes and (second row), B-scan images (100th vertical slice) taken from the initial volumes.

To access the performance of the proposed algorithm, we constructed several undersampled 3D volume using 30%, 50%, 70%, and 90% of the original OCT volume. To do this, we created two types of 3D masks. The first consists of a pseudo-random binary masks M_v using a density random sampling in which the data are selected in a vertical manner (Figure 5 (left)). For the second type of mask, instead of a vertical subsampling of the data, we designed oblique masks M_o as shown in Figure 5 (right)) which are more appropriate in the case of certain imaging systems such as Magnetic Resonance Imaging (MRI), Computerized Tomography scan (CT-scan), and in certain instances, OCT imaging systems as well.

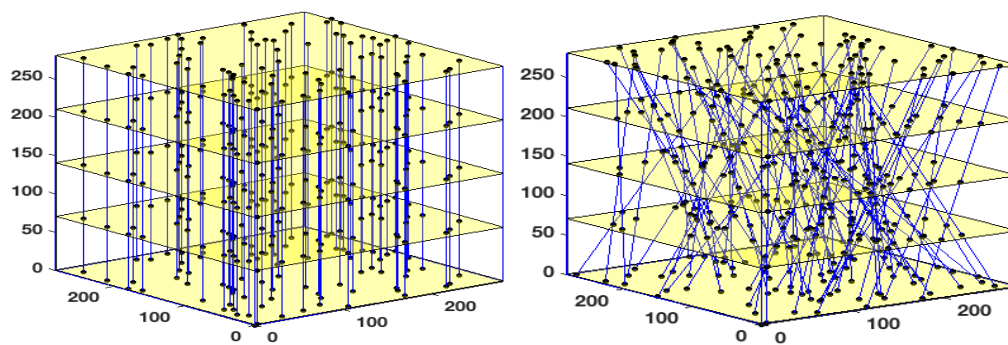


Figure 5. Illustration of the implemented 3D masks used to subsampled the data and creating the 3D masks. (Left) a 3D mask allowing a vertical and random selection of the data, and (right) a 3D mask allowing an oblique selection of the subsampled data.

5.4. Obtained Results

The validation scenarios were carried out as follows—the studied method was applied to each OCT volume (i.e., grape or fish eye) using various subsampling rates (i.e., 30%, 50%, 70%, and 90%). Also, in each case, the vertical M_v or the oblique M_o 3D binary masks were used. The results obtained with our nuclear norm penalised reconstruction approach are discussed below.

5.4.1. Grape Sample

Figures 6 and 7 depict the different reconstructed OCT volumes using the oblique and the vertical binary masks. Instead of illustrating the fully reconstructed OCT volume, we choose to show 2D images (the 100th xy slice of the reconstructed volumes) for a better visualization, with the naked eye, of the quality of the obtained results. It can be highlighted that the sharpness of the boundary is well preserved; however, it loses some features in the zones where the image has low intensity. This is a common effect of most of the compressed sensing and inpainting methods. In order to improve the quality of the recovered data, conventional filters based post-processing could be also be used in order to enhance contrast.

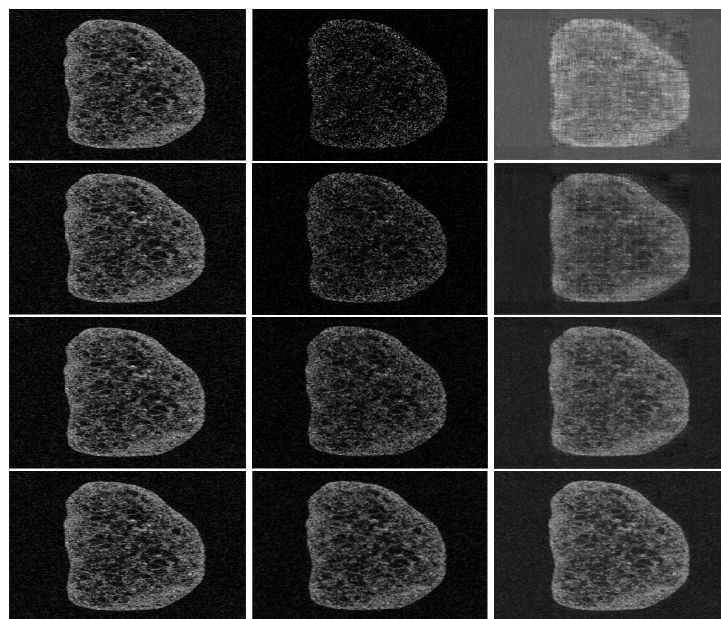


Figure 6. [sample: grape, mask: vertical]—Reconstructed OCT images (only a 2D slice is shown in this example). Each row corresponds to an under-sampling rate: 30% (1st row), 50% (2nd row), 70% (3th row), and 90% (4th row). The first column represents the initial OCT image, the second column the under-sampled data used for the reconstruction, and the third column, the recovered OCT images.

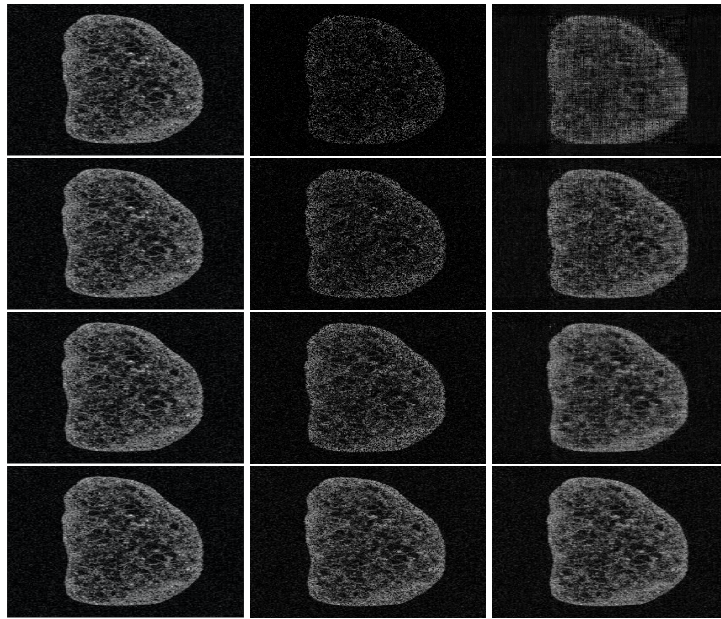


Figure 7. [sample: grape, mask: oblic]—Reconstructed OCT images (only a 2D slice is shown in this example). Each row corresponds to an under-sampling rate: 30% (**1st row**), 50% (**2nd row**), 70% (**3th row**), and 90% (**4th row**). The first column represents the initial OCT image, the second column the under-sampled data used for the reconstruction, and the third column, the recovered OCT images.

5.4.2. Fish Eye Sample

we also performed validation experiments using an optical biopsy of a fish eye (see the images sequence depicted in Figures 8 and 9).

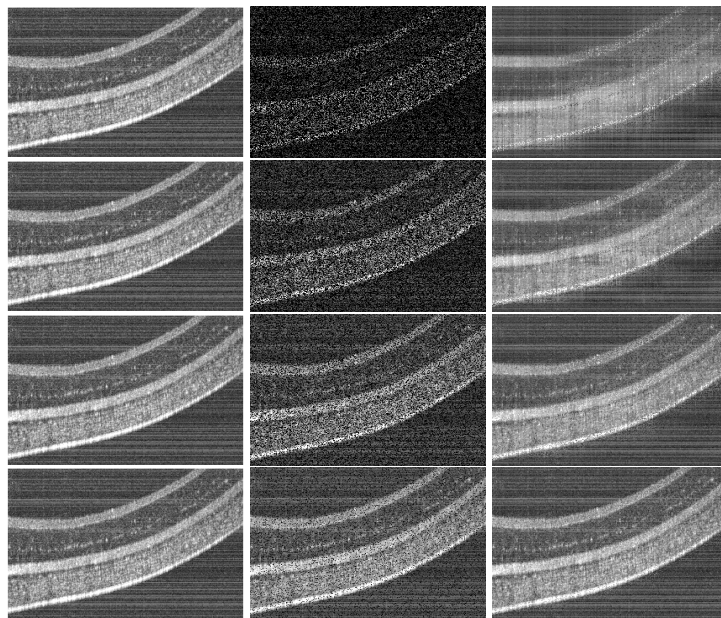


Figure 8. [sample: fish eye retina, mask: vertical]—Reconstructed OCT images (only a 2D slice is shown in this example). Each row corresponds to an under-sampling rate: 30% (**1st row**), 50% (**2nd row**), 70% (**3th row**), and 90% (**4th row**). The first column represents the initial OCT image, the second column the under-sampled data used for the reconstruction, and the third column, the recovered OCT images.

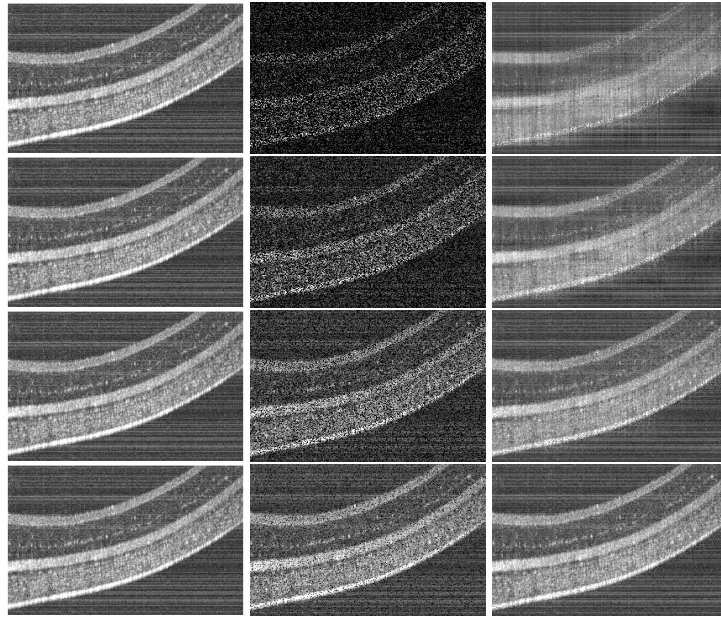


Figure 9. [sample: fish eye retina, mask: oblic]—Reconstructed OCT images (only a 2D slice is shown in this example). Each line corresponds to an under-sampling rate: 30% (**1st line**), 50% (**2nd line**), 70% (**3th line**), and 90% (**4th line**). The first column represents the initial OCT image, the second column the under-sampled data used for the reconstruction, and the third column, the recovered OCT images.

5.5. The Singular Value Thresholding Algorithm

In this section, we describe the algorithm used for the computation of our estimator, namely a tubal tensor version of the fixed point algorithm of Reference [59]. This algorithm is a very simple and scalable iterative scheme which converges to the solution of (17). Each iteration consists of two successive steps:

- a singular value thresholding step where all tubal singular values with norm below the level 2λ are set to zero and the remaining larger singular values are being removed an offset 2λ .
- a relaxed gradient step.

In mathematical terms, the algorithms works as follows:

$$\begin{cases} Z^{(l)} = \text{shrink} \left(A^{(l-1)}, \delta\lambda \right) \\ A^{(l)} = A^{(l-1)} + \delta \mathcal{P}_{\Omega} \left(\sum_{i=1}^n Y_i X_i - Z^{(l)} \right) \end{cases}$$

In the setting of our algorithm, the Shrinkage operator operates as follows:

- compute the circular Fourier transform $\mathcal{F}(A^{(l-1)})$ of the tubal components of $A^{(l-1)}$,
- compute the SVD of all the slices of $\mathcal{F}(A^{(l-1)})$ and forms the tubal singular values,
- sets to zero the tubal singular values whose ℓ_2 -norm lies below the level $2\delta\lambda$ and shrink the other by $2\delta\lambda$,
- recompose the spectrally thresholded matrix and take the inverse Fourier transform of the tubal components.

On the other hand, \mathcal{P}_{Ω} is the operator that assigns to the entries indexed by Ω the observed values and leaves the other values unchanged.

The convergence analysis of [59] directly applies to our setting.

5.6. Analysis of the Numerical Results

In order to quantitatively assess the obtained results using different numerical validation scenarios and OCT images, we implemented two images similarity scores extensively employed in the image processing community. In particular, we use

- the Peak Signal Noise Ratio (PSNR) computed as follows

$$PSNR = 10 \log_{10} \left(\frac{d^2}{MSE} \right) \quad (45)$$

where d is the maximal pixel value in the initial OCT image and the MSE (mean-squared error) is obtained by

$$MSE = \sum_{i=1}^n \sum_{j=1}^m \left(I_o(i, j) - I_r(i, j) \right)^2 \quad (46)$$

with I_o and I_r represent an initial 2D OCT slice (selected from the OCT volume) and the recovered one, respectively.

- The second image similarity score consists of the Structural Similarity Index (SIMM) which allows measuring the degree of similarity between two images. It is based on the computation of three values namely the luminance l , the contrast c and the structural aspect s . It is given by

$$SSIM = \left(s(I_r, I_o) \right) \left(l(I_r, I_o) \right) \left(c(I_r, I_o) \right) \quad (47)$$

where,

$$l = \frac{2\mu_{I_r}\mu_{I_o} + C_1}{\mu_{I_r}^2 + \mu_{I_o}^2 + C_1}, \quad (48)$$

$$c = \frac{2\sigma_{I_r}\sigma_{I_o} + C_2}{\sigma_{I_r}^2 + \sigma_{I_o}^2 + C_2}, \quad (49)$$

$$s = \frac{2\sigma_{I_r I_o} + C_3}{\sigma_{I_r} \sigma_{I_o} + C_3}, \quad (50)$$

with μ_{I_r} , μ_{I_o} , σ_{I_r} , σ_{I_o} , and $\sigma_{I_r I_o}$ are the local means, standard deviations, and cross-covariance for images I_r , I_o . The variables C_1 , C_2 , C_3 are used to stabilize the division with weak denominator.

5.6.1. Illustrating the Rôle of the Nuclear Norm Penalisation

In order to understand the rôle of the nuclear norm penalisation in the reconstruction, we have performed several experiments with different values of the hyper-parameter λ which are reported in Figure 10. This figure shows the performance of the estimator as a function of the ratio between the largest singular value and the smallest selected singular value. This ratio is an implicit function of λ , which is more intuitive than λ for interpreting the results. The smaller the ratio, the smaller the number of singular values incorporated in the estimation.

The corresponding values are reported in Table 1.

Table 1. SME, PSNR, SSIM for various values of the ratio.

Ratio	1	2	3	4	5	6	7	8	9	10
SME	432.65	432.65	432.65	377.12	345.76	345.75	329.06	316.23	323.87	477.43
PSNR	21.76	21.76	21.76	22.36	22.74	22.74	22.95	23.13	23.02	21.34
SSIM	0.57	0.57	0.57	0.58	0.59	0.59	0.60	0.60	0.59	0.53

The results of these experiments show that different weights for the nuclear norm penalisation imply different errors. In these experiments, one sees that the SME is optimised at a ratio equal to 8 and the PSNR is maximised at 8 as well. The SSIM is maximum for values of the ratio equal to 7 and 8. Estimation which does not account for the intrinsic complexity of the object to recover will clearly fail to reconstruct the 3D images properly.

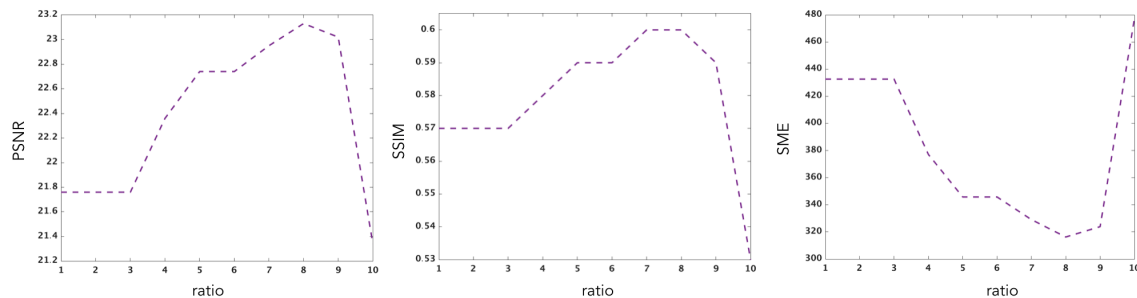


Figure 10. Evolution of the PSNR, SSIM and SME performance criteria as a function of the ratio between the largest singular value and the smallest singular value selected by the penalised estimator.

5.6.2. Performance Results

Tables 2 and 3 summarise the numerical values of MSE, PSNR, SSIM computed for each test, that is, using different undersampling rates and masks, for our two different test samples (grape or fish eye retina). The parameter λ was chosen using the simple and efficient method proposed in [60]. As expected the error decreases as a function of the percentage of observed pixels. The results also show that the estimator is not very sensitive to the orientation of the mask. There seems to be a phase transition after the 70% level, above which the reconstruction accuracy is suddenly improved in terms of PSNR and SSIM, but the method still works satisfactorily at smaller sampling rates.

Table 2. [sample: grape]: Numerical values of the different image similarities (between initial image and reconstructed one): MSE, PSNR and SSIM.

Sample: Part of a Grape				
under-sampling rates	30%	50%	70%	90%
<i>vertical masks</i>				
MSE	536.68	316.23	185.83	62.34
PSNR	20.83	23.13	25.43	30.18
SSIM	0.39	0.59	0.77	0.92
<i>oblic masks</i>				
MSE	725.20	431.91	171.52	54.85
PSNR	19.52	21.77	25.78	30.73
SSIM	0.34	0.55	0.78	0.93

Table 3. [sample: fish eye retina]: Numerical values of the different image similarities (between initial image and reconstructed one): MSE, PSNR and SSIM.

Sample: Fish Eye Retina				
under-sampling rates	30%	50%	70%	90%
<i>vertical masks</i>				
MSE	829.96	570.81	333.13	109.65
PSNR	18.94	20.56	22.50	27.73
SSIM	0.0051	0.65	0.78	0.91
<i>oblic masks</i>				
MSE	828.56	575.90	344.12	99.13
PSNR	18.94	20.52	22.76	28.16
SSIM	0.51	0.65	0.77	0.92

6. Conclusions and Perspectives

In this paper, we studied tensor completion problems based on the framework proposed by Kilmer et al. [10]. We provided some theoretical analysis of the nuclear norm penalised estimator. These theoretical results are validated numerically using realistic OCT data (volumes) of biological samples. These encouraging results with real datasets demonstrate the relevance of the low rank assumption for practical applications. Further research will be undertaken in devising fast algorithms and incorporating other penalties such as, e.g., based on sparsity of the shearlet transform [61].

Author Contributions: Data curation, B.T.; Formal analysis, S.C. and B.T.; Investigation, M.I.A., S.C. and B.T.; Methodology, S.C.; Project administration, B.T.; Supervision, S.C.; Validation, B.T.; Visualization, B.T.; Writing—original draft, M.I.A.; Writing—review and editing, B.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been supported by ANR NEMRO (ANR-14-CE17-0013).

Acknowledgments: The authors would like to thank S. Aeron from Tufts University for interesting discussions and for kindly sharing his code at an initial stage of this research. They would also like to thank the reviewers for their help in improving the paper by their numerous comments and criticisms.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Some Technical Results

Appendix A.1. Calculation of ρ and a value of λ such that $\|M\|_{S_\infty} \leq \lambda$ with high probability

Appendix A.1.1. Computation of ρ

Using the fact $n = m \times n_3$ where m is the number of pixels considered, we have with $i = \{j_1, j_2, j_3\}$

$$\begin{aligned} \|A\|_{L_2(\Pi)}^2 &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\langle A, X_i \rangle^2 \right] \\ &= \frac{1}{mn_3} \sum_{j_3=1}^{n_3} \sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} \langle A, X_{j_1, j_2, j_3} \rangle^2 \frac{1}{n_1 n_2} \\ &= \frac{1}{mn_3} \sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} \frac{1}{n_1 n_2} \sum_{j_3=1}^{n_3} \hat{A}_{j_1, j_2, j_3}^2 \\ &= \frac{1}{mn_3} \sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} \frac{1}{n_1 n_2 n_3^{1/2}} \sum_{j_3=1}^{n_3} A_{j_1, j_2, j_3}^2 \\ &= \frac{1}{mn_1 n_2 n_3^{3/2}} \|A\|_F^2. \end{aligned}$$

Thus we can set

$$\rho = \sqrt{mn_1 n_2 n_3^{3/2}}. \quad (\text{A1})$$

Appendix A.1.2. Control of the Stochastic Error $\|M\|_{S_\infty}$

In this part, we will show that a possible value for the coefficient λ can be taken as

$$\lambda = 2n_3 \|\mathcal{F}(A_0)\|_{S_\infty} \max \left\{ \sqrt{\frac{t + \log(m)}{nn_1 n_2}}; \left(1 + \frac{1}{\sqrt{n_1 n_2}} \right) \frac{t + \log(m)}{n} \right\}. \quad (\text{A2})$$

In order to prove this bound, we will need a large deviation inequality given by the following result

Proposition A1 (Theorem 1.6 in [62]). Let Z_1, \dots, Z_n be independent random variables with dimension $n_1 \times n_2$ that satisfy $\mathbb{E}[Z_i] = 0$ and $\|Z_i\|_{S_\infty} \leq U$ almost surely for some constant U and all $i = 1, \dots, n$. We define

$$\sigma_Z = \max \left\{ \left\| \frac{1}{n} \sum_{i=1}^n \mathbb{E}[Z_{ij} Z_{ij}^\top] \right\|_{S_\infty}^{1/2}, \left\| \frac{1}{n} \sum_{i=1}^n \mathbb{E}[Z_{ij}^\top Z_{ij}] \right\|_{S_\infty}^{1/2} \right\}$$

Then, for all $t > 0$, with probability at least $1 - e^{-t}$, we have

$$\left\| \frac{1}{n} \sum_{i=1}^n Z_i \right\|_{S_\infty} \leq 2 \max \left\{ \sigma_Z \sqrt{\frac{t + \log m}{n}}, U \frac{t + \log m}{n} \right\}$$

where $m = n_1 + n_2$.

The next lemma gives a bound of the stochastic error for tensor completion.

Lemma A1. Let $X^{(i)}$ be i.i.d uniformly distributed on \mathcal{X} . Then for any $t > 0$ with probability at least $1 - \exp(-t)$, we have

$$\|M\|_{S_\infty} \leq 2n_3 \|\mathcal{F}(A_0)\|_{S_\infty} \max \left\{ \sqrt{\frac{t + \log(m)}{nn_1n_2}}, \left(1 + \frac{1}{\sqrt{n_1n_2}}\right) \frac{t + \log(m)}{n} \right\} \quad (\text{A3})$$

In order to prove this lemma, we will need the following lemma

Lemma A2. We draw $X_i^{(j)}$ with a uniform random position on $\{1, \dots, n_1\} \times \{1, \dots, n_2\}$ with null entries everywhere except one input equals 1 at the position (k, l) . Observe that

$$\begin{aligned} \|\mathcal{F}(X_i^{(j)})\|_{S_\infty} &= 1, \quad \left\| \mathbb{E}[\mathcal{F}(X_i^{(j)})] \right\|_{S_\infty} = \sqrt{\frac{1}{n_1n_2}} \text{ and} \\ \sigma_Z &= \max \left\{ \left\| \mathbb{E}[ZZ^\top] \right\|_{S_\infty}^{1/2}, \left\| \mathbb{E}[Z^\top Z] \right\|_{S_\infty}^{1/2} \right\} \leq \frac{\|\mathcal{F}(A_0)\|_{S_\infty}}{\sqrt{n_1n_2}} \end{aligned} \quad (\text{A4})$$

Proof. Let $X_i^{(j)}$ of the form

$$X_i^{(j)} = \begin{bmatrix} 0 & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \dots & 0 \end{bmatrix}. \quad (\text{A5})$$

Determine its Fourier transform and spectral norm of expectation of $\mathcal{F}(X_i^{(j)})$. Thus,

$$\mathcal{F}(X_i^{(j)}) = \begin{bmatrix} 0 & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & e^{-2i\pi j \frac{k_3}{n_3}} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \dots & 0 \end{bmatrix}. \quad (\text{A6})$$

Moreover,

$$\begin{aligned}\mathbb{E} [\mathcal{F}(X_i^{(j)})] &= \frac{1}{n_1 n_2} \begin{bmatrix} e^{-2i\pi j \frac{1}{n_3}} & e^{-2i\pi j \frac{2}{n_3}} & \dots & e^{-2i\pi j \frac{n_2}{n_3}} \\ e^{-2i\pi j \frac{n_2+1}{n_3}} & \dots & \dots & e^{-2i\pi j \frac{n_2+n_2}{n_3}} \\ e^{-2i\pi j \frac{n_2+n_2+1}{n_3}} & \dots & \dots & e^{-2i\pi j \frac{n_2+n_2+n_2}{n_3}} \\ \vdots & \vdots & \vdots & \vdots \\ e^{-2i\pi j \frac{(n_1-2) \times n_2+1}{n_3}} & \dots & \dots & e^{-2i\pi j \frac{(n_1-1) \times n_2}{n_3}} \\ e^{-2i\pi j \frac{(n_1-1) \times n_2+1}{n_3}} & \dots & \dots & e^{-2i\pi j \frac{n_1 \times n_2}{n_3}} \end{bmatrix} \\ &= DLN,\end{aligned}$$

with

$$D = \begin{bmatrix} 1 & & & \\ & \left(e^{-\frac{2i\pi j}{n_3}}\right)^{n_2} & & \\ & & \ddots & \\ & 0 & & 0 \\ & & & \left(e^{-\frac{2i\pi j}{n_3}}\right)^{(n_1-1)n_2} \end{bmatrix}, \quad L = \begin{bmatrix} 1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ 1 \end{bmatrix} \text{ and}$$

$$N = \begin{bmatrix} e^{-\frac{2i\pi j}{n_3}} & \left(e^{-\frac{2i\pi j}{n_3}}\right)^2 & \dots & \left(e^{-\frac{2i\pi j}{n_3}}\right)^{n_2} \end{bmatrix}$$

So, we have $\|\mathcal{F}(X_i^{(j)})\|_{S_\infty} = 1$, and

$$\begin{aligned}\|\mathbb{E} [\mathcal{F}(X_i^{(j)})]\|_{S_\infty} &= \frac{1}{n_1 n_2} \|DLN\| \\ &\leq \frac{1}{n_1 n_2} \|D\|_{S_\infty} \|L\|_2 \|N\|_2 \\ &\leq \frac{1}{\sqrt{n_1 n_2}}\end{aligned}$$

□

Proof of Lemma A1. Consider the tensor completion under uniform sampling at random with Gaussian error. Recall that in this case we assume that the pairs (X_i, Y_i) are i.i.d. Using the fact $Y_i = \langle X_i, A_0 \rangle + \varepsilon_i$ and $\mathbb{E} [\varepsilon_i X_i | X_i] = 0$, we have

$$\begin{aligned}\|M\|_{S_\infty} &= \left\| \frac{1}{n} \sum_{i=1}^n (Y_i X_i - \mathbb{E} [Y_i X_i]) \right\|_{S_\infty} \\ &= \left\| \frac{1}{n} \sum_{i=1}^n (\langle X_i, A_0 \rangle + \varepsilon_i) X_i - \mathbb{E} [\langle X_i, A_0 \rangle X_i] \right\|_{S_\infty} \\ &\leq \left\| \frac{1}{n} \sum_{i=1}^n \langle X_i, A_0 \rangle X_i - \mathbb{E} [\langle X_i, A_0 \rangle X_i] \right\|_{S_\infty} + \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i X_i \right\|_{S_\infty} \\ &= \Lambda_1 + \Lambda_2\end{aligned}$$

In the following, we treat the terms Λ_1 and Λ_2 separately in the lemma and proposition below. Before proceeding, notice that Λ_1 is the Schatten norm of a quadratic function of X_1, \dots, X_n . Furthermore, $\langle X_i, A_0 \rangle$ is the Fourier Transform of the tube (i_1, i_2) for the frequency k_3 . Using the Property 1, we have

$$\begin{aligned}\Lambda_1 &= \left\| \frac{1}{n} \sum_{i=1}^n \mathcal{F}(X_i) \langle \mathcal{F}(X_i), \mathcal{F}(A_0) \rangle - \mathbb{E} [\langle \mathcal{F}(X_i), \mathcal{F}(A_0) \rangle \mathcal{F}(X_i)] \right\|_{S_\infty} \\ &= \left\| \sum_{k_3=1}^{n_3} \left(\frac{1}{n} \sum_{i=1}^n \mathcal{F}(X_i)^{(k_3)} \langle \mathcal{F}(X_i), \mathcal{F}(A_0) \rangle - \mathbb{E} [\langle \mathcal{F}(X_i), \mathcal{F}(A_0) \rangle \mathcal{F}(X_i)^{(k_3)}] \right) \right\|_{S_\infty} \\ &\leq \sum_{k_3=1}^{n_3} \left\| \frac{1}{n} \sum_{i=1}^n \mathcal{F}(X_i)^{(k_3)} \langle \mathcal{F}(X_i), \mathcal{F}(A_0) \rangle - \mathbb{E} [\langle \mathcal{F}(X_i), \mathcal{F}(A_0) \rangle \mathcal{F}(X_i)^{(k_3)}] \right\|_{S_\infty}\end{aligned}$$

□

Define the operator $\Gamma^{(j)}$ which takes the slice j of a tensor (i.e tensor $\mathcal{F}(X_i)$) and puts it in the same place in a zero tensor. The following proposition

Proposition A2. Let T be a null tensor except at slice j . Therefore

$$\|T\|_{S_\infty} = \|T^{(j)}\|_{S_\infty} \quad (\text{A7})$$

Using (A7), we have

$$\Lambda_1 \leq n_3 \left\| \frac{1}{n} \sum_{i=1}^n \mathcal{F}(X_i)^{(j)} \langle \mathcal{F}(A_0), \mathcal{F}(X_i) \rangle - \mathbb{E} [\langle \mathcal{F}(A_0), \mathcal{F}(X_i) \rangle \mathcal{F}(X_i)^{(j)}] \right\|_{S_\infty}$$

Lemma A3. Let $X_i^{(j)}$ be i.i.d a uniform random position on $\{1, \dots, n_1\} \times \{1, \dots, n_2\}$. Then, for all $t > 0$, with probability at least $1 - e^{-t}$, we have

$$\Lambda_1 \leq 2 \max \left\{ \left\| \mathcal{F}(A_0) \right\|_{S_\infty} \sqrt{\frac{t + \log(m)}{n_1 n_2}} n, \left\| \mathcal{F}(A_0) \right\|_{S_\infty} \left(1 + \sqrt{\frac{1}{n_1 n_2}} \right) \frac{t + \log(m)}{n} \right\} \quad (\text{A8})$$

Proof. To prove this lemma, we apply the Proposition A1 for the random variables $Z_{ij} = \mathcal{F}(X_i)^{(j)} \langle \mathcal{F}(A_0), \mathcal{F}(X_i) \rangle - \mathbb{E} [\langle \mathcal{F}(A_0), \mathcal{F}(X_i) \rangle \mathcal{F}(X_i)^{(j)}]$.

Moreover, using the facts

- $\mathbb{E} [Z_{ij}] = \mathbb{E} [\mathcal{F}(X_i)^{(j)} \langle \mathcal{F}(A_0), \mathcal{F}(X_i) \rangle - \mathbb{E} [\langle \mathcal{F}(A_0), \mathcal{F}(X_i) \rangle \mathcal{F}(X_i)^{(j)}]] = 0.$
- $\|Z_{ij}\|_{S_\infty} = \left\| \mathcal{F}(X_i)^{(j)} \langle \mathcal{F}(A_0), \mathcal{F}(X_i) \rangle - \mathbb{E} [\langle \mathcal{F}(A_0), \mathcal{F}(X_i) \rangle \mathcal{F}(X_i)^{(j)}] \right\|_{S_\infty}$

$$\begin{aligned}&\leq \left\| \mathcal{F}(X_i)^{(j)} \right\|_{S_\infty} |\langle \mathcal{F}(A_0), \mathcal{F}(X_i) \rangle| \\&\quad + |\langle \mathcal{F}(A_0), \mathcal{F}(X_i) \rangle| \left\| \mathbb{E} [\mathcal{F}(X_i)^{(j)}] \right\|_{S_\infty} \\&\leq \left\| \mathcal{F}(X_i)^{(j)} \right\|_{S_\infty} \left\| \mathcal{F}(A_0) \right\|_{S_\infty} \left\| \mathcal{F}(X_i)^{(j)} \right\|_{S_\infty} \\&\quad + \left\| \mathcal{F}(A_0) \right\|_{S_\infty} \left\| \mathcal{F}(X_i)^{(j)} \right\|_{S_\infty} \left\| \mathbb{E} [\mathcal{F}(X_i)^{(j)}] \right\|_{S_\infty} \\&\leq \left\| \mathcal{F}(A_0) \right\|_{S_\infty} \left(1 + \sqrt{\frac{1}{n_1 n_2}} \right)\end{aligned}$$

•

$$\begin{aligned}\mathbb{E} \left[Z Z^\top \right] &= \mathbb{E} \left[\langle \mathcal{F}(A_0), \mathcal{F}(X_i) \rangle^2 \mathcal{F}(X_i)^{(i)} \mathcal{F}(X_i)^{(i)\top} \right] \\ &\quad - \mathbb{E} \left[\langle \mathcal{F}(A_0), \mathcal{F}(X_i) \rangle \mathcal{F}(X_i)^{(i)} \right] \mathbb{E} \left[\langle \mathcal{F}(A_0), \mathcal{F}(X_i) \rangle \mathcal{F}(X_i)^{(i)} \right]^\top\end{aligned}$$

and

$$\begin{aligned}\mathbb{E} \left[Z^\top Z \right] &= \mathbb{E} \left[\langle \mathcal{F}(A_0), \mathcal{F}(X_i) \rangle^2 \mathcal{F}(X_i)^{(i)\top} \mathcal{F}(X_i)^{(i)} \right] \\ &\quad - \mathbb{E} \left[\langle \mathcal{F}(A_0), \mathcal{F}(X_i) \rangle \mathcal{F}(X_i)^{(i)} \right]^\top \mathbb{E} \left[\langle \mathcal{F}(A_0), \mathcal{F}(X_i) \rangle \mathcal{F}(X_i)^{(i)} \right]\end{aligned}$$

Using the duality trace and Jensen's inequality, we have

$$\begin{aligned}\sigma_Z^2 &\leq \max \left\{ \left\| \mathbb{E} \left[\langle \mathcal{F}(A_0), \mathcal{F}(X_i) \rangle^2 \mathcal{F}(X_i)^{(i)} \mathcal{F}(X_i)^{(i)\top} \right] \right\|_{S_\infty}, \right. \\ &\quad \left. \left\| \mathbb{E} \left[\langle \mathcal{F}(A_0), \mathcal{F}(X_i) \rangle^2 \mathcal{F}(X_i)^{(i)\top} \mathcal{F}(X_i)^{(i)} \right] \right\|_{S_\infty} \right\} \\ &\leq \left\| \mathcal{F}(A_0) \right\|_{S_\infty}^2 \left\| \mathcal{F}(X_i) \right\|_{S_\infty}^2 \mathbb{E} \left[\left\| \mathcal{F}(X_i) \right\|_{S_\infty}^2 \right] \\ &\leq \frac{\left\| \mathcal{F}(A_0) \right\|_{S_\infty}^2}{n_1 n_2}\end{aligned}$$

Thus, (A8) follows from (A1). \square

Now, we study the bound of Λ_2 . For this purpose, we use the proposition below is an immediate consequence of the matrix Gaussian's inequality of Theorem 1.6 of [62].

Proposition A3. Let Z_1, \dots, Z_n be independent random variables with dimensions $n_1 \times n_2$, and $\{\varepsilon_i\}$ be a finite sequence of independent standard normal. Define

$$\sigma := \max \left\{ \left\| Z_i Z_i^\top \right\|_{S_\infty}, \left\| Z_i^\top Z_i \right\|_{S_\infty} \right\}$$

Then, for all $t \geq 0$,

$$\left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i Z_i \right\|_{S_\infty} \leq 2\sigma \sqrt{\frac{t + \log(m)}{n}}$$

where $m = n_1 + n_2$

Using this fact and (A7), we have

$$\begin{aligned}\Lambda_2 &= \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i X_i \right\|_{S_\infty} \\ &= \left\| \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^{n_3} \mathcal{F}(\varepsilon_i)^{(k)} \mathcal{F}(X_i)^{(k)} \right\|_{S_\infty} \\ &\leq n_3 \left\| \frac{1}{n} \sum_{i=1}^n \mathcal{F}(\varepsilon_i)^{(j)} \mathcal{F}(X_i)^{(j)} \right\|_{S_\infty} \\ &\leq 2 n_3 \sigma \sqrt{\frac{t + \log(m)}{n}}\end{aligned}$$

References

1. Candès, E.J. Mathematics of sparsity (and a few other things). In Proceedings of the International Congress of Mathematicians, Seoul, Korea, 13–21 August 2014.
2. Candès, E.J.; Romberg, J.; Tao, T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory* **2006**, *52*, 489–509. [\[CrossRef\]](#)
3. Recht, B.; Fazel, M.; Parrilo, P.A. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.* **2010**, *52*, 471–501. [\[CrossRef\]](#)
4. Hackbusch, W. *Tensor Spaces and Numerical Tensor Calculus*; Springer Science & Business Media: Berlin, Germany, 2012; Volume 42.
5. Landsberg, J.M. *Tensors: Geometry and Applications*; American Mathematical Society: Providence, RI, USA, 2012; Volume 128.
6. Sacchi, M.D.; Gao, J.; Stanton, A.; Cheng, J. Tensor Factorization and its Application to Multidimensional Seismic Data Recovery. In *SEG Technical Program Expanded Abstracts*; Society of Exploration Geophysicists: Houston, TX, USA, 2015.
7. Yuan, M.; Zhang, C.H. On tensor completion via nuclear norm minimization. *Found. Comput. Math.* **2016**, *16*, 1031–1068. [\[CrossRef\]](#)
8. Signoretto, M.; Van de Plas, R.; De Moor, B.; Suykens, J.A. Tensor versus matrix completion: A comparison with application to spectral data. *Signal Process. Lett. IEEE* **2011**, *18*, 403–406. [\[CrossRef\]](#)
9. Kilmer, M.E.; Martin, C.D. Factorization strategies for third-order tensors. *Linear Algebra Its Appl.* **2011**, *435*, 641–658. [\[CrossRef\]](#)
10. Kilmer, M.E.; Braman, K.; Hao, N.; Hoover, R.C. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. *SIAM J. Matrix Anal. Appl.* **2013**, *34*, 148–172. [\[CrossRef\]](#)
11. Koltchinskii, V.; Lounici, K.; Tsybakov, A.B. Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *Ann. Stat.* **2011**, *39*, 2302–2329. [\[CrossRef\]](#)
12. Candès, E.J.; Plan, Y. Matrix completion with noise. *Proc. IEEE* **2010**, *98*, 925–936. [\[CrossRef\]](#)
13. Anandkumar, A.; Ge, R.; Hsu, D.; Kakade, S.M.; Telgarsky, M. Tensor decompositions for learning latent variable models. *J. Mach. Learn. Res.* **2014**, *15*, 2773–2832.
14. Signoretto, M.; Dinh, Q.T.; De Lathauwer, L.; Suykens, J.A. Learning with tensors: A framework based on convex optimization and spectral regularization. *Mach. Learn.* **2014**, *94*, 303–351. [\[CrossRef\]](#)
15. So, A.M.C.; Ye, Y. Theory of semidefinite programming for sensor network localization. *Math. Program.* **2007**, *109*, 367–384. [\[CrossRef\]](#)
16. Eriksson, B.; Balzano, L.; Nowak, R. High-rank matrix completion and subspace clustering with missing data. *arXiv* **2011**, arXiv:1112.5629.
17. Candès, E.J.; Recht, B. Exact matrix completion via convex optimization. *Found. Comput. Math.* **2009**, *9*, 717–772. [\[CrossRef\]](#)
18. Singer, A.; Cucuringu, M. Uniqueness of low-rank matrix completion by rigidity theory. *SIAM J. Matrix Anal. Appl.* **2010**, *31*, 1621–1641. [\[CrossRef\]](#)
19. Fazel, M. Matrix Rank Minimization with Applications. Ph.D. Thesis, Stanford University, Stanford, CA, USA, 2002.
20. Gross, D. Recovering low-rank matrices from few coefficients in any basis. *Inf. Theory IEEE Trans.* **2011**, *57*, 1548–1566. [\[CrossRef\]](#)
21. Recht, B. A simpler approach to matrix completion. *J. Mach. Learn. Res.* **2011**, *12*, 3413–3430.
22. Candès, E.J.; Tao, T. The power of convex relaxation: Near-optimal matrix completion. *Inf. Theory IEEE Trans.* **2010**, *56*, 2053–2080. [\[CrossRef\]](#)
23. Kolda, T.G.; Bader, B.W. Tensor decompositions and applications. *SIAM Rev.* **2009**, *51*, 455–500. [\[CrossRef\]](#)
24. McCullagh, P. *Tensor Methods in Statistics*; Chapman and Hall: London, UK, 1987; Volume 161.
25. Rogers, M.; Li, L.; Russell, S.J. Multilinear Dynamical Systems for Tensor Time Series. *Adv. Neural Inf. Process. Syst.* **2013**, *26*, 2634–2642.
26. Ge, R.; Huang, Q.; Kakade, S.M. Learning mixtures of gaussians in high dimensions. *arXiv* **2015**, arXiv:1503.00424.

27. Mossel, E.; Roch, S. Learning nonsingular phylogenies and hidden Markov models. In Proceedings of the Thirty-Seventh Annual ACM Symposium on Theory of Computing, Baltimore, MD, USA, 22–24 May 2005; pp. 366–375.
28. Jennrich, R. A generalization of the multidimensional scaling model of Carroll and Chang. *UCLA Work. Pap. Phon.* **1972**, *22*, 45–47.
29. Kroonenberg, P.M.; De Leeuw, J. Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika* **1980**, *45*, 69–97. [[CrossRef](#)]
30. Bhaskara, A.; Charikar, M.; Moitra, A.; Vijayaraghavan, A. Smoothed analysis of tensor decompositions. In Proceedings of the 46th Annual ACM Symposium on Theory of Computing, New York, NY, USA, 1–3 June 2014; pp. 594–603.
31. De Lathauwer, L.; De Moor, B.; Vandewalle, J. A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.* **2000**, *21*, 1253–1278. [[CrossRef](#)]
32. Kolda, T.G. Symmetric Orthogonal Tensor Decomposition is Trivial. *arXiv* **2015**, arXiv:1503.01375.
33. Robeva, E.; Seigal, A. Singular Vectors of Orthogonally Decomposable Tensors. *arXiv* **2016**, arXiv:1603.09004.
34. Hao, N.; Kilmer, M.E.; Braman, K.; Hoover, R.C. Facial recognition using tensor-tensor decompositions. *SIAM J. Imaging Sci.* **2013**, *6*, 437–463. [[CrossRef](#)]
35. Semerci, O.; Hao, N.; Kilmer, M.E.; Miller, E.L. Tensor-based formulation and nuclear norm regularization for multienergy computed tomography. *Image Process. IEEE Trans.* **2014**, *23*, 1678–1693. [[CrossRef](#)]
36. Zhang, Z.; Aeron, S. Exact tensor completion using t-svd. *IEEE Trans. Signal Process.* **2017**, *65*, 1511–1526. [[CrossRef](#)]
37. Pothier, J.; Girson, J.; Aeron, S. An algorithm for online tensor prediction. *arXiv* **2015**, arXiv:1507.07974.
38. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. Methodol.* **1996**, *58*, 267–288. [[CrossRef](#)]
39. Bickel, P.J.; Ritov, Y.; Tsybakov, A.B. Simultaneous analysis of Lasso and Dantzig selector. *Ann. Stat.* **2009**, *37*, 1705–1732. [[CrossRef](#)]
40. Candès, E.J.; Plan, Y.; others. Near-ideal model selection by l1 minimization. *Ann. Stat.* **2009**, *37*, 2145–2177. [[CrossRef](#)]
41. Zou, H. The adaptive lasso and its oracle properties. *J. Am. Stat. Assoc.* **2006**, *101*, 1418–1429. [[CrossRef](#)]
42. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. Stat. Methodol.* **2005**, *67*, 301–320. [[CrossRef](#)]
43. Gaïffas, S.; Lécué, G. Weighted algorithms for compressed sensing and matrix completion. *arXiv* **2011**, arXiv:1107.1638.
44. Xu, Y. On Higher-order Singular Value Decomposition from Incomplete Data. *arXiv* **2014**, arXiv:1411.4324.
45. Raskutti, G.; Yuan, M. Convex Regularization for High-Dimensional Tensor Regression. *arXiv* **2015**, arXiv:1512.01215.
46. Barak, B.; Moitra, A. Noisy Tensor Completion via the Sum-of-Squares Hierarchy. *arXiv* **2015**, arXiv:1501.06521.
47. Gandy, S.; Recht, B.; Yamada, I. Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Probl.* **2011**, *27*, 025010. [[CrossRef](#)]
48. Mu, C.; Huang, B.; Wright, J.; Goldfarb, D. Square Deal: Lower Bounds and Improved Convex Relaxations for Tensor Recovery. In Proceedings of the International Conference on Machine Learning, Beijing, China, 21–26 June 2014.
49. Amelunxen, D.; Lotz, M.; McCoy, M.B.; Tropp, J.A. Living on the edge: Phase transitions in convex programs with random data. *Inf. Inference* **2014**, *3*, 224–294. [[CrossRef](#)]
50. Lécué, G.; Mendelson, S. Regularization and the small-ball method I: Sparse recovery. *arXiv* **2016**, arXiv:1601.05584.
51. Watson, G.A. Characterization of the subdifferential of some matrix norms. *Linear Algebra Its Appl.* **1992**, *170*, 33–45. [[CrossRef](#)]
52. Lewis, A.S. The convex analysis of unitarily invariant matrix functions. *J. Convex Anal.* **1995**, *2*, 173–183.
53. Hu, S. Relations of the nuclear norm of a tensor and its matrix flattenings. *Linear Algebra Its Appl.* **2015**, *478*, 188–199. [[CrossRef](#)]
54. Chrétien, S.; Wei, T. Von Neumann’s inequality for tensors. *arXiv* **2015**, arXiv:1502.01616

55. Chrétien, S.; Wei, T. On the subdifferential of symmetric convex functions of the spectrum for symmetric and orthogonally decomposable tensors. *Linear Algebra Its Appl.* **2018**, *542*, 84–100. [[CrossRef](#)]
56. Heinze H.; Bauschke, P.L. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*; Canada Mathematical Society: Ottawa, ON, Canada, 2011.
57. Borwein, J.; Lewis, A.S. *Convex Analysis and Nonlinear Optimization: Theory and Examples*; Springer Science & Business Media: Berlin, Germany, 2010.
58. Jean-Pierre Aubin, I.E. *Applied Nonlinear Analysis*; Dover Publications: Mineola, NY, USA, 2006; Volume 518.
59. Ma, S.; Goldfarb, D.; Chen, L. Fixed point and Bregman iterative methods for matrix rank minimization. *Math. Program.* **2011**, *128*, 321–353. [[CrossRef](#)]
60. Chrétien, S.; Lohvithee, M.; Sun, W.; Soleimani, M. Efficient hyper-parameter selection in Total Variation-penalised XCT reconstruction using Freund and Shapire's Hedge approach. *Mathematics* **2020**, *8*, 493. [[CrossRef](#)]
61. Guo, W.; Qin, J.; Yin, W. A new detail-preserving regularization scheme. *SIAM J. Imaging Sci.* **2014**, *7*, 1309–1334. [[CrossRef](#)]
62. Tropp, J.A. User-Friendly tail bounds for sums of random matrices. *arXiv* **2010**, arXiv:1004.4389v7.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).