Low tubal rank tensor recovery using the Bürer-Monteiro factorisation approach. Application to optical coherence tomography

Mohamed Ibrahim Assoweh, Stéphane Chrétien, Brahim Tamadazte

Low tubal rank tensor recovery using the Bürer-Monteiro factorisation approach. Application to optical coherence tomography

Mohamed Ibrahim Assoweh[a,b], Stéphane Chrétien[a,c], Brahim Tamadazte[e]

[a] *Laboratoire de Mathématiques, Universté de Franche Comté, 16 route de Gray, 25030 Besançon, France*
[b] *Université de Djibouti, Campus de Balbala, Croisement RN2-RN5, B.P: 1904, Djibouti*
[c] *National Physical Laboratory, Hampton Road, Teddington, TW11 0LW, UK*
[d] *ERIC Laboratory and UFR ASSP, University of Lyon 2, 5 avenue Mendes-France, 69676 Bron Cedex, France*
[e] *FEMTO-ST Institute, AS2M department, 24 rue Alain Savary, BP 25000, Besançon, France*

Abstract

In this paper, we study the low-tubal-rank tensor completion problem, i.e., the problem of recovering a third-order tensor by observing a subset of its entries, when these entries are selected uniformly at random. We propose a mathematical analysis of an extension of the Bürer-Monteiro factorisation approach to this problem. We then illustrate the use of the Bürer-Monteiro approach on a chalenging OCT reconstruction problem on both synthetic and real world data, using an alternating minimisation algorithm.

Keywords: Tensor completion, t-SVD, Low rank estimation, Bürer-Monteiro approach

1. Introduction

Tensor completion has many applications in various fields of engineering and data science, such as Markov Field analysis, signal and image processing, etc [28], seismic data reconstruction [22], health data analytics [33], compression of hyperspectral images [24], 3D image and video reconstruction from subsampled measurements [25], [17]. From the mathematical viewpoint, tensor completion is currently a very active research trend as well [21], [11], [32], etc. Tensor completion relies on the often encountered property that the sought after tensor is low rank, which is the case in many different applications; see [31], [1], [28], etc. Previous work on *tensor completion* based on partial sampling includes [14], [30], [3], [23], [13], [8], [35], [9], etc.

One of the main ingredient of both theoretical analysis and practical implementations of tensor completion is the singular value decomposition (SVD), which has been recently extended from the matrix to the tensor setting in various directions depending on the properties one intends to preserve [12], [21], [28], [19], etc. One particular approach relies on expressing the original tensor as a sum of rank-1 tensors [21], a construct which looses the orthogonality property of the singular vectors as compared with the matrix SVD. Conversely, the multilinear SVD of [12] enforces this orthogonality property of singular vectors at the price of loosing the standard notion of scalar singular values and replacing it with the notion of core tensors. Another trend is the recent tubal-SVD of [19], which extends the matrix notion of SVD to the higher dimensional setting by generalising matrix products from 2D arrays of scalars to 2D arrays of "tubes" and which uses a specific "tubal" product. As for [12], the singular values are no longer scalars, but become higher dimensional "tubal singular values". As a consequence of the diversity of approaches to the construction of the tensor SVD, the notion of rank is very specific to the definition of the SVD used in the application of interest.

On the computational side, the estimation of low rank tensors has been a topic of increasing interest, due to the high dimensionality of the problem. The approach developed in [2] is an example of efficient algorithm with guaranteed complexity. Convex relaxations based on various extensions of the nuclear norm penalisation approach, originally devised in the matrix case in [29], have also been proposed lately in the tensor case; see e.g. [29], [8], [35], etc. Tighter relaxations have also been proposed and precisely studied in [30], [9]. Factorisation based methods form another group of methods which have been successfully employed in [27], however without clear mathematical underpinnings. In the case of tubal SVD-based approaches, some recent work include [26], [34], [34] where iterative methods are implemented achieving high practical efficiency, while leaving the question of establishing the convergence rigorously to further investigation.

The goal of the present paper is to study the factorisation based approach to the low tubal-rank tensor recovery problem. This approach, also known as the Bürer-Monteiro factorisation approach in the Semi-Definite Programming literature [5], [4], has proved very efficient in practice but also amenable to thorough theoretical analysis [4], [15] in the matrix setting. The main contribution of the present paper is a proof that similar results also hold for the low tubal-rank tensor completion problem, after appropriately generalising some of the main ingredients from [15]. In particular, we prove that all the local minimisers of the least-squares cost function applied to the couple resulting from the factorisation, are in fact global minimisers. The main consequence of this analysis is that one can safely run a gradient-like or alternating optimisation algorithm for the reconstruction of the low tubal-rank tensor of interest.We also illustrate our theoretical results with promising numerical experiments for a challenging OCT reconstruction problem.

We now define more precisely the mathematical problem addressed in this paper.

1.1. The Tensor Completion Problem

All the notations and concepts about tensors are collected in 3.

Tensor Completion is a generalisation of Matrix Completion, an extensively studied problem in data science which went through a sudden surge of interest triggered by the Netflix context [7]. The tensor completion problem is the one of recovering a tensor $\mathcal{M}^\star$ a small number of components of which are observed uniformly at random. More precisely, given $\Omega \subset \{1,\ldots,n_1\} \times \{1,\ldots,n_2\}$ to be a set of indices for the observed entries, we define for any tensor $\mathcal{M}$, the observed tensor $\mathcal{M}_\Omega$ by

$$[\mathcal{M}_\Omega]_{ijk} = \begin{cases} \mathcal{M}_{ijk} & \text{for all } k \in \{1,\ldots,n_3\}, \text{ if } \big(i,j\big) \in \Omega \\ \\ 0 & \text{otherwise.} \end{cases}$$

The probability of selecting one entry will be denoted by $p = 1/(n_1 \times n_2)$.

This recovery problem is hopeless in the general setting, but in many areas of engineering, social, biological etc sciences, the tensor to be recovered can be assumed to be low rank, where by rank, one refers to the appropriate notion among the ones that have been devised in the literature on the mathematics of tensors [21], [28].

Under a low tensor rank assumption, one way of recovering the original $\mathcal{M}^\star$ is to solve the following optimization problem

$$\min_{\mathcal{M}\in\mathbb{R}^{n_1\times n_2\times n_3}} f(\mathcal{M}) = \frac{1}{2p}\|\mathcal{M}-\mathcal{M}^\star\|_\Omega^2 \quad \text{s.t.} \quad \operatorname{rank}(\mathcal{M}) = r \tag{1}$$

for some $r \geq 0$, where $\|\cdot\|_\Omega$ denotes the Frobenius norm restricted to the components indexed by Ω. The main difficulty in addressing this optimisation problem resides in handling the rank constraint, even in the case of 2D tensors, i.e. matrices. In the matrix case, one efficient method which has gained a lot of interest lately is the Bürer-Monteiro factorisation approach [5] which, in the symmetric case $n_1 = n_2 = n$, takes the following form

$$\min_{\mathcal{U} \in \mathbb{R}^{n \times r \times n_3}} f(\mathcal{U}) = \frac{1}{2p} \|\mathcal{U} * \mathcal{U}^\top - \mathcal{M}^\star\|_\Omega^2. \tag{2}$$

Recent work by [15] showed that the Bürer Monteiro approach is able to recover low rank matrices when the observation set up is of the Compressed Sensing type. The work in [15] also showed that for matrix completion, a penalisation term is required to be added to the least squares functional in order to ensure successful recovery via the following optimisation problem:

$$\min_{\mathcal{U} \in \mathbb{R}^{n \times r \times n_3}} \frac{1}{2p} \|\mathcal{U} * \mathcal{U}^\top - \mathcal{M}^\star\|_\Omega^2 + Q(\mathcal{U}) \tag{3}$$

where Q will be specified later.

The goal of the present paper is to study an extension of the penalised Bürer-Monteiro least squares problem (3) in the setting of tensor recovery. As in [15] both the symmetric and non-symmetric settings will be studied.

1.2. Plan of the paper

This chapter is organized as follows. Section 2 introduces some notations and preliminaries of optimality conditions. In Section 3 we define the background of several algebraic structures of third-order tensors. Section 4 presents the main recovery results. In Section 5 we demonstrate the effectiveness of low rank tensor reconstruction for the problem of Optical Coherence Tomography. In Section 7 we prove our main theoretical results.

2. Preliminaries

In this section we provide all the preliminary technical results and notations which will be used in our analysis.

2.1. Notations

In this paper, we focus on real valued third-order tensors in the space $\mathbb{R}^{n_1\times n_2\times n_3}$. We use n_1, n_2, n_3 for tensor dimensions, x for vectors and $X \in \mathbb{R}^{n_1\times n_2}$ for matrices. Tensors are denoted by calligraphic letters, i.e $\mathcal{A} \in \mathbb{R}^{n_1\times n_2\times n_3}$. For a vector x, $\|x\|_2$ denotes its ℓ_2 norm and for a matrix X we use $\|X\|_F$ to denote its Frobenius norm. For a tensor $\mathcal{A}$, we use $\|\mathcal{A}\|_F$ to denote its Frobenius norm which is the square root of the sum of its squared components; see 3 for further details. Throughout the paper, we use $\mathcal{M}^\star$ to denote the original low rank solution to be recovered and we denote by $\sigma_1^\star$ its largest singular value, by $\sigma_r^\star$ its r-th singular value. The ratio $\kappa^\star = \sigma_1^\star/\sigma_r^\star$ be called the condition number.

Given a transformation $\mathcal{H}$, we use the notations $\mathcal{M} : \mathcal{H} : \mathcal{N}$ to denote the quadratic form $\langle \mathcal{M}, \mathcal{H}(\mathcal{N})\rangle$.

2.2. Optimality Conditions

Suppose we are optimizing a function $f(\mathbf{x})$ with no constraints on $\mathbf{x}$. For a point $\mathbf{x}$ to be a local minimum, it must satisfy the first and second order necessary conditions, i.e. $\nabla f(\mathbf{x}) = 0$ ans $\nabla^2 f(\mathbf{x}) \succeq 0$.

If one of the two conditions is not verified, i.e, if we are not a local minimum, it is always possible to follow the gradient and reduce the value of the function. In this case, [16] defines a **strict-saddle** property, which is a quantitative version of the optimality conditions.

Definition 1. *We say f satisfies the $\big(\theta,\gamma,\zeta\big)$-strict saddle property if for any point $\boldsymbol{x}$ at least one of the following is true:*

1. $\|\nabla f(x)\| \geq \theta$.
2. $\lambda_{\min}\Big(\nabla^2 f(x)\Big) \leq -\gamma$.
3. *x is ζ-close to $\mathcal{X}^\star$ where $\mathcal{X}^\star$ is the set of a local minima.*

This definition intuitively says that any point at which the gradient of f is small, is either close to a local minimiser, a local maximiser or a saddle point with at least one significantly negative eigenvalue.

3. Background on tensors and the tubal algebra

3.1. Basic Notations for Third-order Tensor

In this section, we recall the framework introduced by Kilmer and Martin [20] and [19] for a very special class of tensors which is particularly adapted to our setting.

3.1.1. Slices and transposition

A third-order tensor are represented as A and its (i, j, k)th entry is denoted by A_{ijk}.

Definition 2. *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 $\mathbb{R}^{1\times1\times n_3}$ and a tubal vector (t-vector) is an element of $\mathbb{R}^{n_1\times1\times n_3}$

Definition 3. *(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 slide number 2 to the slice number n_3 and then appending the conjugate transposed frontal slice $A^{(1)^\top}$.*

Definition 4. *(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 $B^{(n)}$:*

$$C^{(k)} := (A \cdot B)^{(k)} := A^{(k)}\ B^{(k)}, \quad k = 1, \dots, n_3. \tag{4}$$

We will also need the canonical inner product.

Definition 5. *(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),*

$$\langle A, B\rangle = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3} A_{ijk}B_{ijk}. \tag{5}$$

3.1.2. Convolution and Fourier transform

Definition 6. *(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 $\big(i, j\big)$-th tube is given by*

$$C\big(i, j, :\big) = \sum_{k=1}^{n_2} A\big(i, k, :\big) * B\big(k, j, :\big) \tag{6}$$

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

Definition 7. *(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 $J^i, i = 2, \ldots, n_3$ are zero.*

Definition 8. *(Orthogonal tensor) A tensor $Q \in \mathbb{R}^{n\times n\times n_3}$ is orthogonal if it satisfies*

$$Q^\top * Q = Q * Q^\top = J. \tag{7}$$

The tensor $\hat{A}$ 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

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

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

$$\hat{A}(j_1, j_2, k_3) = \sum_{j_3=1}^{n_3} A(j_1, j_2, j_3)\ \exp\Big(-2\frac{i\pi j_3 k_3}{n_3}\Big),$$

for all j_1, j_2, $1 \le j_1 \le n_1$, $1 \le j_2 \le n_2$. Naturally, one can compute A from $\hat{A}$ via $\text{ifft}\big(\hat{A}, [\,], 3\big)$ using the inverse FFT, which is defined as:

$$A(j_1, j_2, k_3) = \sum_{j_3=1}^{n_3} \hat{A}(j_1, j_2, j_3)\ \exp\Big(2\frac{i\pi j_3 k_3}{n_3}\Big),$$

for all j_1, j_2, $1 \le j_1 \le n_1$, $1 \le j_2 \le n_2$.

Definition 9. *(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*

$$A^{-1} * A = A * A^{-1} = J \tag{8}$$

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

Remark 1. *It is proved in [19] 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*

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

3.2. The t-SVD

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

Definition 10. *(f-diagonal tensor) Tensor* A *is called f-diagonal if each frontal slice* $A^{(i)}$ *is a diagonal matrix.*

Definition 11. *(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*

$$M = U * S * V^\top \tag{9}$$

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 or 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:*

$$\hat{M}^{(i)} = \hat{U}^{(i)} \cdot \hat{S}^{(i)} \cdot \left(\hat{V}^{(i)}\right)^\top. \tag{10}$$

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 2. They will also be called *tubal eigenvalues.*

Definition 12. *The spectrum* $\sigma(A)$ *of the tensor* A *is the tubal vector given by*

$$\sigma(A)_i = S(i, i, :) \tag{11}$$

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

3.3. Some natural Tensor Norms

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

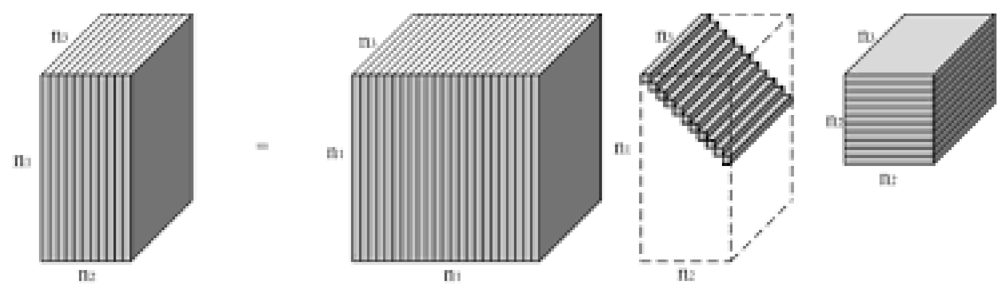

Figure 1: The t-SVD of a tensor

Definition 13. *(Tensor Frobenius norm) The induced Frobenius norm from the inner product defined above is given by,*

$$\|A\|_F = \langle A, A\rangle^{1/2} = \frac{1}{\sqrt{n_3}}\|\hat{A}\|_F = \sqrt{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3} A_{ijk}^2}. \tag{12}$$

Definition 14. *(Tensor spectral norm) The tensor spectral norm $\|A\|_\infty$ is defined as follows:*

$$\|A\|_\infty = \max_i \|\sigma(A)_i\|_2 \tag{13}$$

where $\|.\|_2$ 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 15. *(Tubal nuclear norm) The tensor nuclear norm of a tensor A denoted as $\|A\|_\circledast$ is the sum of singular values of all the frontal slices of A. Moreover,*

$$\begin{aligned}\|A\|_\circledast &= \sum_{i=1}^{\min\{n_1,n_2\}} \sqrt{\sum_{j=1}^{n_3} S(i,i,j)^2} \\ &= \sum_{i=1}^{\min\{n_1,n_2\}} \|\sigma(A)_i\|_2. \end{aligned} \tag{14}$$

Note that by Parseval's equality

$$\sqrt{\sum_{j=1}^{n_3} S(i,i,j)^2} = \frac{1}{\sqrt{n_3}} \sqrt{\sum_{j=1}^{n_3} \hat{S}(i,i,j)^2}. \tag{15}$$

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 $\hat{U}^{(k)}\hat{S}^{(k)}\hat{V}^{(k)^t}$ 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*

$$|\langle A, B\rangle| \le \|A\|_{\circledast} \|B\|_{\infty}.$$

Proof. By Cauchy-Schwartz, we have

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

taking the maximum of $\|\mathcal{F}(B)_{jj}\|_2$ and the sum the slices of $\left(\|\hat{S}_{jj}\|_2\right)^{1/2}$, and apply (15) 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*

$$\|A\|_{\circledast} \le \sqrt{\operatorname{rank}(A)} \|A\|_F.$$

Proof. Again by Cauchy-Schwartz, we have

$$
\begin{aligned}
\|A\|_{\circledast} &= \sum_{j=1}^{\min\{n_1,n_2\}} \|S(j,j,:)\|_2 \\
&= \sum_{j=1}^{\mathrm{rank(A)}} \|S(j,j,:)\|_2 \\
&\le \sqrt{\mathrm{rank}(A)}\Big(\sum_{j=1}^{\min\{n_1,n_2\}} \|S(j,j,;)\|_2^2\Big)^{1/2} \\
&\le \sqrt{\mathrm{rank}(A)}\|A\|_F.
\end{aligned}
$$

□

3.4. 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 [19], 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 $S_{i,i,:}$ as the number of non-zero components of $\hat{S}(i,i,:)$. Then, the easiest way of defining the rank of a tensor is by means of the notion of multirank as follows.

Definition 16. *The multirank of a tensor is the vector* $(\rho_1,\dots,\rho_r)$ *where* r *is the number of nonzero tubal vectors* $S(i,i,:)$, $i = 1,\dots,\min\{n_1,n_2\}$ *where comes from the t-SVD of* M *and* r *is also called the rank of the tensor* M*.*

We now define the range of a tensor.

Definition 17. *Let* j *denote the number of invertible tubal eigenvalues and let* k *denote the number of nonzero non-invertible 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*

$$
\begin{aligned}
\mathcal{R}(M) = \{&\vec{U}^{(1)} * c_1 + \cdots + \vec{U}^{(j+k)} * c_{j+k} \mid c_l \in \mathcal{R}ange(s_l * \cdot), \\
&l \in \{j+1,\cdots,j+k\}\}.
\end{aligned}
$$

Definition 18. *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*

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

4. Main result

In this section, we present our main contribution to the analysis of the Bürer-Monteiro approach to the tensor completion problem. For this purpose, let $\mathcal{M}^\star = \mathcal{U}^\star * \mathcal{V}^{\star\top}$ denote the factorisation of $\mathcal{M}^*$, and for any variable tensor $\mathcal{M}$, we will use the similar factorisation $\mathcal{M} = \mathcal{U} * \mathcal{V}^\top$. We can now define the following objective function of $\mathcal{M}$ expressed as a function of $(\mathcal{U}, \mathcal{V})$:

$$\begin{aligned} f(\mathcal{U}, \mathcal{V}) = {} & 2(\mathcal{M} - \mathcal{M}^\star) : \mathcal{H}_0 : (\mathcal{M} - \mathcal{M}^\star) + \frac{1}{2}\|\mathcal{U}^\top * \mathcal{U} - \mathcal{V}^\top * \mathcal{V}\|_F^2 \\ & + Q_0(\mathcal{U}, \mathcal{V}). \end{aligned} \tag{16}$$

The asymmetric problem can easily be reduced to a symmetric problem as follows. Suppose $\mathcal{M}^\star$ is the optimal solution and its t-SVD is $\mathcal{X}^\star * \mathcal{D}^\star * \mathcal{Y}^{\star\top}$. Let $\mathcal{U}^\star = \mathcal{X}^\star * (\mathcal{D}^\star)^{\frac{1}{2}}$, $\mathcal{V}^\star = \mathcal{Y}^\star * (\mathcal{D}^\star)^{\frac{1}{2}}$ and $\mathcal{M} = \mathcal{U} * \mathcal{V}^\top$ is the current point, we reduce the problem into a symmetric case by using following notations.

$$\mathcal{W} = \begin{pmatrix} \mathcal{U} \\ \mathcal{V} \end{pmatrix}, \mathcal{W}^\star = \begin{pmatrix} \mathcal{U}^\star \\ \mathcal{V}^\star \end{pmatrix}, \mathcal{N} = \mathcal{W} * \mathcal{W}^\top, \mathcal{N}^\star = \mathcal{W}^\star * \mathcal{W}^{\star\top}. \tag{17}$$

In the sequel, we define $\Delta = \mathcal{W} - \mathcal{W}^\star * \mathcal{R}$.

We will also transform the Hessian operator to operate $(n_1 + n_2) \times r \times n_3$ tensors. For this purpose, define the tensors $\mathcal{H}_1$ and $\mathcal{G}$ such that for all $(\mathcal{U}, \mathcal{V})$ we have:

$$\begin{aligned} \mathcal{N} : \mathcal{H}_1 : \mathcal{N} &= \mathcal{M} : \mathcal{H}_0 : \mathcal{M} \\ \mathcal{N} : \mathcal{G} : \mathcal{N} &= \|\mathcal{U}^\top * \mathcal{U} - \mathcal{V}^\top * \mathcal{V}\|_F^2 \end{aligned}$$

where we recall that $\mathcal{N}$ is a function of $(\mathcal{U}, \mathcal{V})$. Now, let $Q(\mathcal{W}) := Q_0(\mathcal{U}, \mathcal{V})$ and we can rewrite the objective function $f(\mathcal{W})$ as

$$\frac{1}{2}\left[(\mathcal{N} - \mathcal{N}^\star) : 4\mathcal{H}_1 : (\mathcal{N} - \mathcal{N}^\star) + \mathcal{N} : \mathcal{G} : \mathcal{N}\right] + Q(\mathcal{W}). \tag{18}$$

The main result of this paper is the following theorem.

Theorem 4.1. *Let* $d = \max\{n_1, n_2\}$*. Assume that*

$$p \geq c_1 \frac{\mu^4 r^6 \left(\kappa^\star\right)^6 \log(d)}{\min\{n_1, n_2\}},$$

for some positive constant c_1*. Choose*

$$\alpha_1^2 = \Theta\big(\frac{\mu r \sigma_1^\star}{n_1}\big), \ \alpha_2^2 = \Theta\big(\frac{\mu r \sigma_1^\star}{n_2}\big)$$

and

$$\lambda_1 = \Theta\big(\frac{n_1}{\mu r \kappa^\star}\big), \ \lambda_2 = \Theta\big(\frac{n_2}{\mu r \kappa^\star}\big).$$

Then with probability at least

$$1 - 2n_1 n_3 \exp\left\{-pn_2\big(\big(1 + \frac{t}{pn_2}\big) \ln\big(1 + \frac{t}{pn_2}\big) - \frac{t}{pn_2}\big)\right\},$$

for tensor completion objective (18) we have

- *All local minima satisfy* $\mathcal{U} * \mathcal{V}^\top = \mathcal{M}^\star$*;*
- *the function is* $\big(\epsilon, c_1\big(\sigma_r^\star\big), C\big(\frac{\epsilon}{\sigma_r^\star}\big)\big)$*-strict saddle for polynomially small* ϵ*.*

for some positives constants C, c_1*.*

Proof. See 7.3. ☐

5. Numerical validation on medical images

In this section, we present some numerical results validating our approach on medical images and volumes. Our experiments were performed on Optical Coherence Tomography (OCT) images, also called "*optical biopsies*" used by clinicians to perform micrometric (at cellular level) characterization of biological tissues in both *in situ* and *ex vivo* settings. Application of OCT in different medical setups such as ophthalmology, dermatology, cardiovascular surgery, etc, is usually considered of high clinical value. However, *in situ* acquisition of high resolution and 3-dimensional optical biopsies is well known to be very challenging in practice. Some well known drawbacks of using

OCT for such medical applications are: long acquisition times (generating artefacts, e.g., under physiological disturbances) for full-resolution volume acquisition. Moreover, preprocessing/processing, transfer and storage of very large datasets (up to 10 Go for a full resolution OCT volume) is one of the main limitations for using OCT-based optical biopsies in some medical applications of interest. The subsampling approach together with the efficient factorisation-based optimisation method proposed in the present paper aim at circumventing these issues.

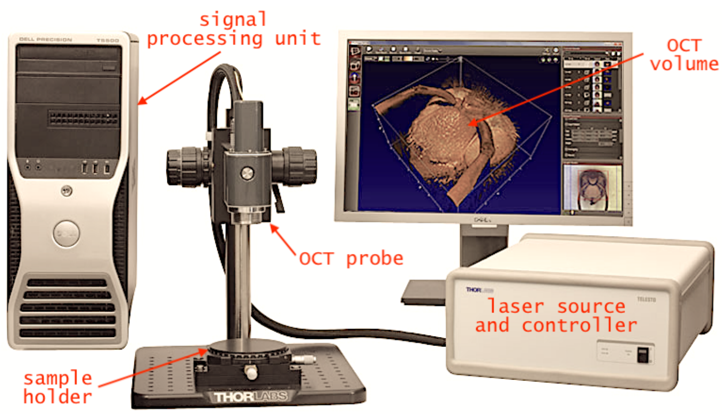

Figure 2: Photography of our OCT imaging system.

This section discusses different setups using progressive subsampling rate ranging from 20% to 80%. In Section 5.1, we present of our spectral domain OCT system (the most popular marketed OCT systems). In Section 5.2, we describe the different experimental scenarios. Our cmputational results are presented in Section 5.3

5.1. OCT Imaging System

As mentioned above, OCT is a well-established medical imaging technique (e.g., for optical biopsy-based diagnosis) that uses a light wave to capture 3-dimensional images of a light-diffusing material (e.g,, biological tissue) with a micrometer (1μm) resolution [18]. OCT is uses low coherence interferometric technique at near-infrared wavelength. Indeed, light absorption of imaged biological tissues is limited in near-infrared light wavelength range, which restricts penetration up to about 1mm. This technique is thus halfway between ultrasonic (resolution of 150μm, penetration of 10cm) and confocal microscopy (resolution of 0.5μm, penetration of 200μm).

The OCT imaging technique allows to retrieve three types of information. Firstly, each position of the light spot on the imaged tissue gives the reflectivity profile (z axis), called A-scan, which can contain information about the structure and spatial dimensions of the sample under study. Secondly, a 2-dimensional slice ($x - z$ axes scan) of the sample (transverse tomography), called B-scan, can be obtained by combining series of A-scan profiles. Finally, combining successive B-scan cross-sections allows acquiring volumetric OCT data ($x - y - z$ axes scan), called C-scan.

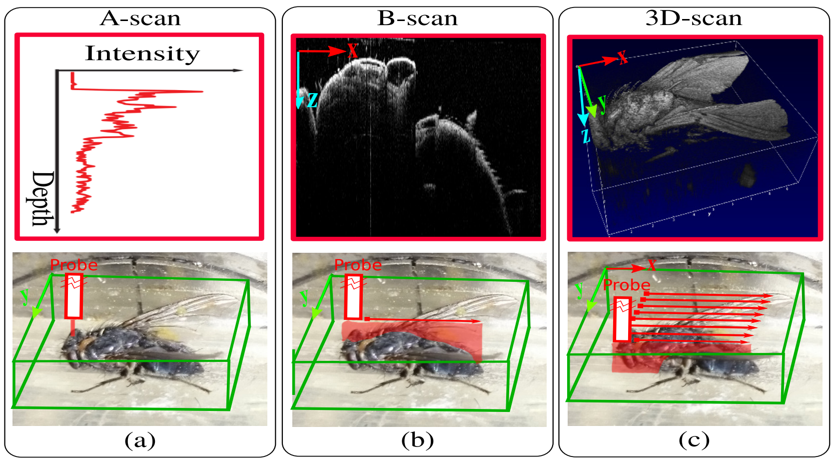

Figure 3: Available acquisition modes in an OCT imaging system: (a) A-scan, (b) B-Scan, and (c) C-scan.

The acquisition of the different types of OCT signals (i.e., A-scan, B-scan, and C-scan) is performed sequentially by moving the light spot on the imaged sample. In other words, it is possible to acquire each single data independently of the others. In particular, the 2 degrees-of-freedom galvanometer integrated in the OCT probe makes it possible to optimise the sampling using any prespecified geometrically constrained protocol. As a result, one of the great advantages of OCT is that it is ideally suited to geometric subsampling in the spirit of compressed sensing.

5.2. Validation Scenarios

The developed materials and methods were implemented in a MatLab framework without taking into account code optimization aspects nor time-computation. The numerical validation of the methods was achieved using two optical biopsies acquired on biological samples: a piece of a grape

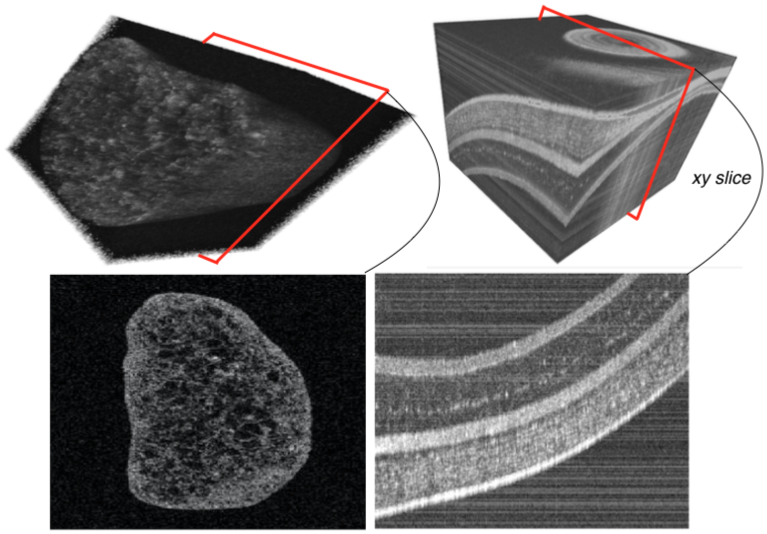

Figure 4: Examples of the OCT volumes of biological samples used to validate the proposed method. (first row): the initial OCT volumes and (second row), B-scan images (100^{th} vertical slice) taken from the original volumes.

(Fig. 4(left)), a sample of fish eye retina (Fig. 4(right)) recorded from a commercial OCT device[1]. Both optical biopsies (considered as low-tubal-rank tensors) have equal size $A_{n_1 \times n_2 \times n_3} = 281 \times 281 \times 281$ voxels. Different scenarios were considered to assess the performance of our algorithm. The sampling rates used for these experiments ranged from 20% to 80% (with a step of 10%) and formed masks that were applied to the original volume (we randomly pick 20% to 80% pixels from the original tensors). Finally, we set the maximum iteration number to be $i_{max} = 10$.

5.3. Obtained Results

Note that instead of illustrating the fully reconstructed OCT volume, we choose to show 2D images (the 100^{th} xz slice of the reconstructed volumes) for a better visualization, with the naked eye, of the quality of the obtained results as can be shown in Fig. 6. Again it can be noticed that the sharpness

[1]The Telesto-II from Thorlabs (https://www.thorlabs.com/thorproduct.cfm?partnumber=TELESTO-II)

of the boundary is will preserved. Furthermore, the recovered data can be improved such as using conventional filters based post-processing methods.

Retina

The retina is chosen because for its translucent characteristics offering very good conditions for acquiring high resolution OCT (optical biopsy) images/volumes and interesting signal-to-noise ratio comparing to most OCT images. The reconstructed data from the subsampled volumes are depicted in Fig. 5

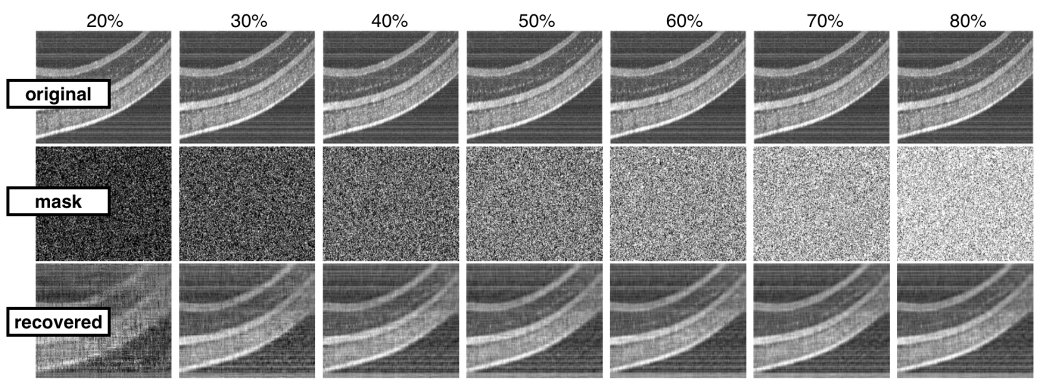

Figure 5: **[sample: retina]** - Reconstructed OCT images (only a 2D slice is shown in this example). First row corresponds to the original slice, second row the subsampled data (ranging from 20% to 80% with a step of 10%) to be reconstructed, and third row the reconstructed slices.

Grape

As mentioned above, the second OCT volume used to assess the performance of the algorithm is recorded by imaging a part of a grape. Even, the grape is also a translucent medium, the signal-to-noise ratio is less important than the one obtained by imaging the retina. The validation scenario is still the same as for the first test, i.e., different subsampling OCT volumes were built using 20% to 80% (with a step of 10%) of the original data. Again it can be noticed that the sharpness of the boundary is preserved. Furthermore, the recovered data can be improved such as using conventional filters based post-processing methods.

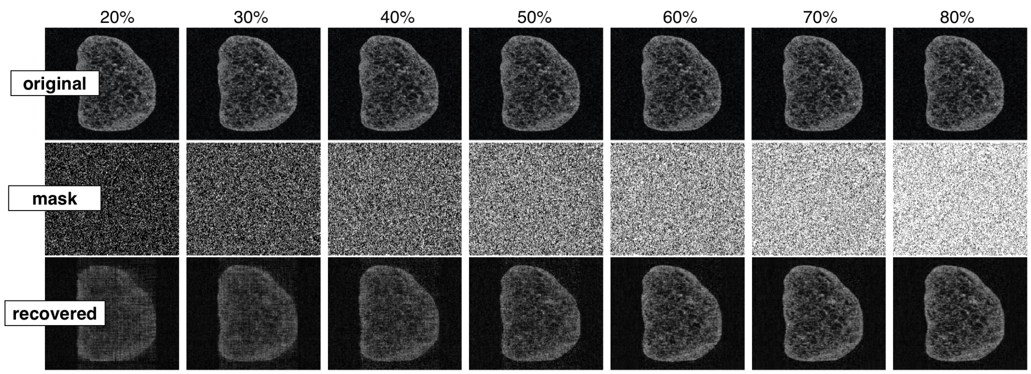

Figure 6: **[sample: grape]** - Reconstructed OCT images (only a 2D slice is shown in this example). First row corresponds to the original slice, second row the subsampled data (ranging from 20% to 80% with a step of 10%) to be reconstructed, and third row the reconstructed slices.

5.4. Evaluation Scores

To quantitatively assess the numerical validation results, we implemented two images similarity scores extensively employed in the image processing community.

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

$$PSNR = 10\log_{10}\Big(\frac{d^2}{MSE}\Big) \tag{19}$$

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

$$MSE = \sum_{i=1}^{n}\sum_{j=1}^{m}\Big(I_o(i,j) - I_r(i,j)\Big)^2 \tag{20}$$

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

- 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

$$SSIM = \Big(s\big(I_r, I_o\big)\Big)\Big(l\big(I_r, I_o\big)\Big(c\big(I_r, I_o\big)\Big) \tag{21}$$

where,

$$l = \frac{2\mu_{I_r}\mu_{I_o} + C_1}{\mu_{I_r}^2 + \mu_{I_o}^2 + C_1}, \tag{22}$$

$$c = \frac{2\sigma_{I_r}\sigma_{I_o} + C_2}{\sigma_{I_r}^2 + \sigma_{I_o}^2 + C_2}, \tag{23}$$

$$s = \frac{2\sigma_{I_r,I_o} + C_3}{\sigma_{I_r}\sigma_{I_o} + C_3}, \tag{24}$$

with μ_{I_r}, μ_{I_o}, σ_{I_r}, σ_{I_o}, and μ_{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.

Tables 1 summarizes the numerical values the PSNR and SSIM computed for each test. The obtained numerical results for both evaluation scores clearly demonstrate the relevance of the proposed approach for this type of images/volumes. As expected, increasing the number of samples significantly improves the quality scores, however, using only 20% sampled data gives unexpectedly good and exploitable recovery. In the range from 30% to 80%, the reconstructed data are faithful to the original ones.

Table 1: Numerical values of the SSIM and PSNR scores.

sample 1: eye							
subsampling rate	20%	30%	40%	50%	60%	70%	80%
PSNR	14.20	17.73	18.44	18.70	18.87	18.99	19.19
SSIM	00.13	00.29	00.34	00.36	00.38	00.38	00.39
sample 1: grape							
subsampling rate	20%	30%	40%	50%	60%	70%	80%
PSNR	19.24	19.69	19.64	20.30	22.07	23.58	24.44
SSIM	00.20	00.25	00.30	00.38	00.37	00.43	00.46

5.5. Impact of the Initialization Parameters on the Quality of the Reconstruction

Note that two initialization parameters have influence on the quality of the reconstruction. They concern the number of iteration "i" and the tubal rank "r". First, in the numerical validation discussed above, both the number of iterations i and tubal rank r were, respectively, fixed to $i_{max} = 20$ and $r = 20$.

5.5.1. *Number of iterations i*

In this section, we varied the values of these parameters and for each pair (i, r), we computed both the PSNR and SSIM values. As can be seen in Fig. 7, the best reconstruction (for $r = 20$) was obtained using only few iterations i.e., $i = 5$.

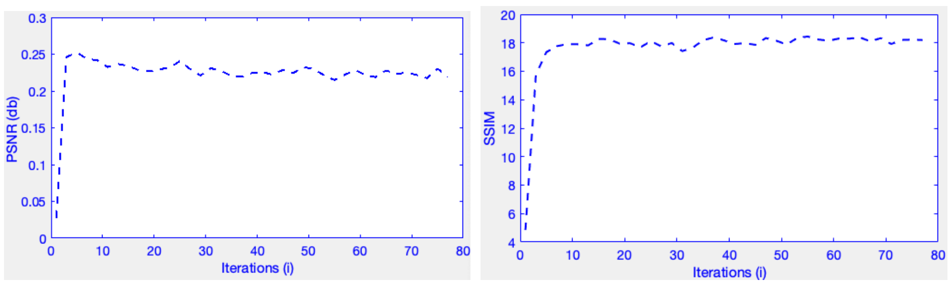

Figure 7: Representation of both the PSNR (right) and the SSIM (left) criteria in function of the number of iterations i.

5.5.2. *Choice of tubal rank r*

According to the above statement, we fixed the iterations number $i = 5$, and we varied the tubal rank r. As can be shown in Fig. 8, the best similarity scores (PSNR and SIMM) are obtained for a $r = 80$.

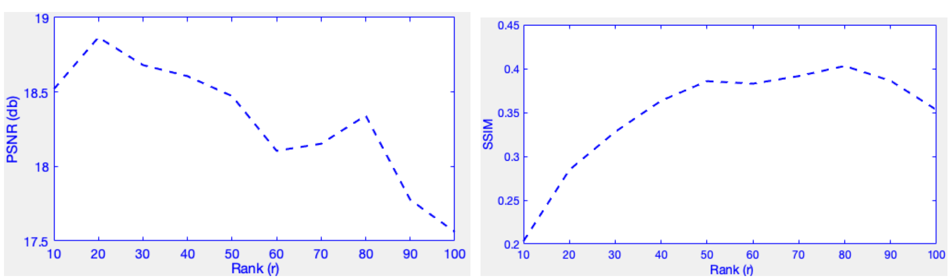

Figure 8: Representation of both the PSNR (right) and the SSIM (left) criteria in function of the tubal rank values r.

The choice of the tubal rank is crucial for efficient image reconstruction, and one needs easy-to-use criteria for swift selection. Many possible methods are available in the literature such as the Bayesian Information Criterion (BIC). One of the main drawbacks of BIC is that only sum of squared errors are taken into account whereas the information about the errors is usually much richer than what is collected in sums of squared errors. In this section,

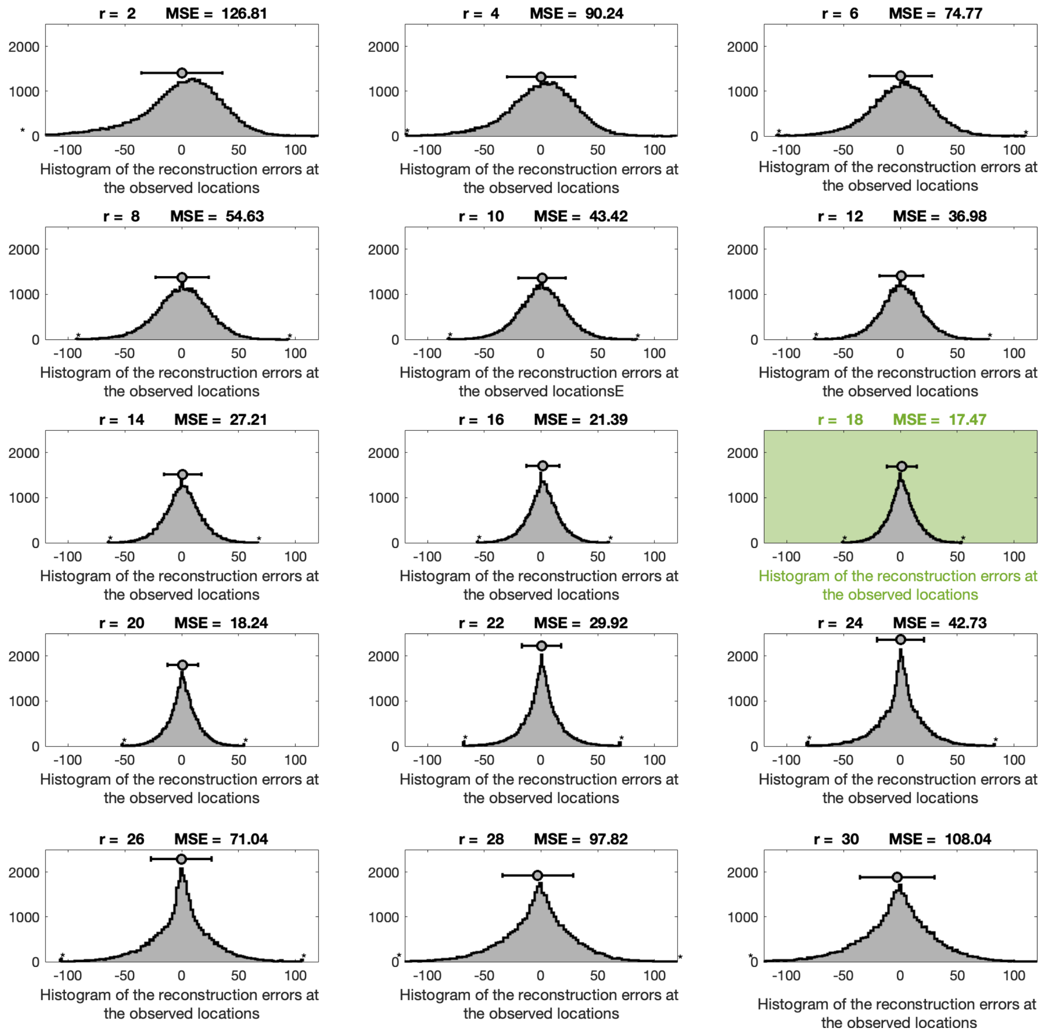

Figure 9: Histograms of the reconstructed error at the observed locations only. The plot in green shows the histogram associated with the smallest MSE over the total tensor image, corroborating the relevance of using histograms for accurate reconstruction.

we report an interesting observation about using histograms of the reconstruction errors at the observed locations as an appropriate proxy for model selection. Our results are given in Figure 10 where the histograms are plotted for various values of the rank. The MSE values are also reported in each figure. We notice that there exists a strong correlation between the shape and support range of the histograms and the quality of the reconstruction as mea-

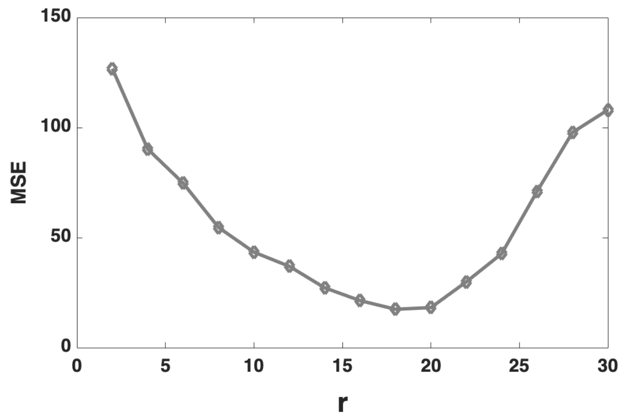

Figure 10: MSE over the whole reconstructed tensor image as a function of the rank

sure by the MSE. Notice in particular that the histogram corresponding to the best MSE is symmetric and has the second smallest support range. The histogram with the smallest support range gives a 4% larger MSE. Also, the shape of the histograms show an interesting structural change as the rank increases, passing from a smooth Gaussian-like behavior to a more spiky Laplacian-like behavior. We thus conclude that the histograms contain all necessary information for accurate results in the problem of low rang tubal tensor reconstruction.

6. Conclusion and Perspectives

In this paper, we studied a low tubal-rank tensor completion problem using non convex optimisation, as initially proposed in [15]. A theoretical extension of the analysis in [16] was provided in order to address the important tubal tensor case. The theoretical results were validated numerically using real data, i.e., OCT volumes acquired in biological samples (a retina and a grape). The obtained results are encouraging and demonstrate the performance of the low tubal-rank tensor completion problem.

Further work will consist in the validation of the method in a physical imaging device. In order to achieve this, it will be important to consider a GPU implementation of the algorithm in order to address the real-time processing inherent challenges. Additional research work can be undertaken in adapting the algorithm to an online setting where the hyperparameters can be learned using e.g. the approach of [10].

[1] A. Anandkumar, R. Ge, D. Hsu, D. M Kakade, and M. Telgarsky, *Tensor decompositions for learning latent variable models*, The Journal of Machine Learning Research, ,**15** 2014, no. 1, 2773–2832.

[2] A. Anandkumar, R. Ge, and M. Janzamin, *Guaranteed non-orthogonal tensor decomposition via alternating rank-*1 *updates*, arXiv, preprint arXiv:1402.5180, 2014.

[3] J. A. Bazerque, G. Mateos, and G. B. Giannakis, *Rank regularization and bayesian inference for tensor completion and extrapolation*, IEEE Trans. on Signal Processing **61**, 2013, no. 22, 5689–5703.

[4] N. Boumal, V. Voroninski, and A. Bandeira, *The non-convex burer-monteiro approach works on smooth semidefinite programs*, Advances in Neural Information Processing Systems, 2016, pp. 2757–2765.

[5] S. Burer and R. DC Monteiro, *A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization*, Mathematical Programming, **95**, 2003, no. 2, 329–357.

[6] Candès, Emmanuel and Recht Benjamin, *Exact matrix ompletion via convex optimization*, Communications of ACM, **55**, 2012, no. 6, 111–119.

[7] Candès, Emmanuel J and Tao, Terence, The power of convex relaxation: Near-optimal matrix completion, IEEE Transactions on Information Theory, 56, no. 5, 2053–2080, 2010.

[8] S. Chrétien and T. Wei, *Convex recovery of tensors using nuclear norm penalization*, Int. Conf. on Latent Variable Analysis and Signal Separation, Springer, 2015, pp. 360–367.

[9] S. Chrétien and T. Wei, *Sensing tensors with gaussian filters*, IEEE Trans. on Information Theory, **63**, 2017, no. 2, 843–852.

[10] Chretien, Stephane and Lohvithee, Manasavee and Sun, Wenjuan and Soleimani, Manuchehr, *Efficient hyper-parameter selection in Total Variation-penalised XCT reconstruction using Freund and Shapire's Hedge approach*, Mathematics, **8**, 2020, no. 4, 493.

[11] Andrzej Cichocki, Danilo Mandic, Lieven De Lathauwer, Guoxu Zhou, Qibin Zhao, Cesar Caiafa, and Huy Anh Phan, *Tensor decompositions for signal processing applications: From two-way to multiway component analysis*, IEEE Signal Processing Mag., **32**, 2015, no. 2, 145–163.

[12] L. De Lathauwer, B. De Moor, and J. Vandewalle, *A multilinear singular value decomposition*, SIAM Journal on Matrix Analysis and Applications, **21**, 2000, no. 4, 1253–1278.

[13] M. Filipović and A. Jukić, *Tucker factorization with missing data with application to low-rank tensor completion*, Multidimensional Systems and Signal Processing, **26**, 2015, no. 3, 677–692.

[14] S. Gandy, B. Recht, and I. Yamada, *Tensor completion and low-n-rank tensor recovery via convex optimization*, Inverse Problems **27**, 2011, no. 2, 025010.

[15] Rong Ge, Chi Jin, and Yi Zheng, *No spurious local minima in nonconvex low rank problems: A unified geometric analysis*, Int. Conf. on Machine Learning, 2017, pp. 1233–1242.

[16] Rong Ge, Furong Huang, Chi Jin, and Yang Yuan, *Escaping From Saddle Points — Online Stochastic Gradient for Tensor Decomposition*, arXiv, preprint arXiv:1503.02101, 2015.

[17] W. Hu, D. Tao, W. Zhang, Y. Xie, and Y. Yang, *The twist tensor nuclear norm for video completion*, IEEE Trans. on Neural Networks and Learning Systems, **28**, 2017, no. 12, 2961–2973.

[18] Fujimoto, James G, Pitris, Costas, Boppart, Stephen A and Brezinski, Mark E, Optical coherence tomography: an emerging technology for biomedical imaging and optical biopsy, Neoplasia, **2**, 1-2, 9–25,2000.

[19] M. E Kilmer, K. Braman, N. Hao, and R. C Hoover, *Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging*, SIAM Journal on Matrix Analysis and Applications, **34**, 2013, no. 1, 148–172.

[20] M. E Kilmer and C. D Martin, *Factorization strategies for third-order tensors*, Linear Algebra and its Applications, **435** 2011, no. 3, 641–658.

[21] T. G Kolda and B. W Bader, *Tensor decompositions and applications*, SIAM review **51**, 2009, no. 3, 455–500.

[22] N. Kreimer, A. Stanton, and M. D Sacchi, *Tensor completion based on nuclear norm minimization for 5d seismic data reconstruction*, Geophysics, **78**, 2013, no. 6, V273–V284.

[23] D. Kressner, M. Steinlechner, and B. Vandereycken, *Low-rank tensor completion by riemannian optimization*, BIT Numerical Mathematics, **54**, 2014, no. 2, 447–468.

[24] N. Li and B. Li, *Tensor completion for on-board compression of hyperspectral images*, IEEE Int. Conf. on Image Processing.

[25] Ji Liu, Przemyslaw Musialski, Peter Wonka, and Jieping Ye, *Tensor completion for estimating missing values in visual data*, IEEE Trans. on Pattern Analysis and Machine Intelligence, **35**, 2013, no. 1, 208–220.

[26] X.-Y. Liu, S. Aeron, V. Aggarwal, and X. Wang, *Low-tubal-rank tensor completion using alternating minimization*, Modeling and Simulation for Defense Systems and Applications XI, vol. 9848, Int. Society for Optics and Photonics, 2016, p. 984809.

[27] Yuanyuan Liu and Fanhua Shang, *An efficient matrix factorization method for tensor completion*, IEEE Signal Processing Letters, **20** (2013), no. 4, 307–310.

[28] A. Moitra, *Algorithmic aspects of machine learning*, Cambridge University Press, 2018.

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

[30] B. Romera-Paredes and M. Pontil, *A new convex relaxation for tensor completion*, Advances in Neural Information Processing Systems, 2013, pp. 2967–2975.

[31] H. Sedghi, M. Janzamin, and A. Anandkumar, *Provable tensor methods for learning mixtures of generalized linear models*, Artificial Intelligence and Statistics, 2016, pp. 1223–1231.

[32] Nicholas D Sidiropoulos, Lieven De Lathauwer, Xiao Fu, Kejun Huang, Evangelos E Papalexakis, and Christos Faloutsos, *Tensor decomposition for signal processing and machine learning*, IEEE Trans. on Signal Processing, **65**, 2017, no. 13, 3551–3582.

[33] Y. Wang, R. Chen, J. Ghosh, J. C Denny, A. Kho, Y. Chen, B. A Malin, and J. Sun, *Rubik: Knowledge guided tensor factorization and completion for health data analytics*, Int. Conf. on Knowledge Discovery and Data Mining, ACM, 2015, pp. 1265–1274.

[34] D. Wei, A. Wang, X. Feng, B. Wang, and B. Wang, *Tensor completion based on triple tubal nuclear norm*, Algorithms, **11**, 2018, no. 7, 94.

[35] M. Yuan and C.-H. Zhang, *On tensor completion via nuclear norm minimization*, Foundations of Computational Mathematics, **16**, 2016, no. 4, 1031–1068.

[36] Zemin Zhang and Shuchin Aeron, *Exact tensor completion using t-svd*, IEEE Transactions on Signal Processing, **65**, 2017, no. 6, 1511–1526.

7. Proofs of our results

7.1. Concentration Inequalities for matrix completion

For matrix completion, we need different concentration inequalities for different kinds of matrices. The first type of matrix lies a tangent space and is proved in [6].

Lemma 1. *[6] Let $d = \max\{n_1, n_2\}$. Define the subspace*

$$\mathcal{T} = \{M \in \mathbb{R}^{n_1 \times n_2} \,|M = U^\star X^\top + Y V^{\star\top}, \text{ for some } X \in \mathbb{R}^{n_1 \times r}, Y \in \mathbb{R}^{n_2 \times r}\}.$$

Then, for any $\delta > 0$, as long as sample rate $p \geq \Omega\big(\frac{\mu r}{\delta^2 d} \log(d)\big)$, we will have:

$$\|\frac{1}{p}\mathcal{P}_\mathcal{T}\mathcal{P}_\Omega\mathcal{P}_\mathcal{T} - \mathcal{P}_\mathcal{T}\| \leq \delta.$$

For arbitrary low rank matrix, we will need the following lemma.

Lemma 2. *Suppose that $\Omega \subset [n_1] \times [n_2]$ is the set of edges of a random bipartite graph with $\big(n_1, n_2\big)$ nodes, where any pair nodes on different side is connected with probability p. Let $d = \max\{n_1, n_2\}$, then there exists two universal constants c_1, c_2, for any $\delta > 0$ such that for $p \geq c_1 \frac{\log\left(d\right)}{\min\{n_1, n_2\}}$, then with probability at least $1 - d^{-4}$, we have for any $x, y \in \mathbb{R}^n$:*

$$\frac{1}{p}\sum_{\left(i,j\right)\in\Omega} x_i y_j \leq \|x\|_1\|y\|_1 + c_2\sqrt{\frac{d}{p}}\,\|x\|_2\|y\|_2.$$

This theorem implies following:

Lemma 3. *Let $d = \max\{n_1, n_2\}$. There exists universal constant c_1, c_2, for any $\delta > 0$ so that if $p \geq c_1 \frac{\log\left(d\right)}{\min\{n_1, n_2\}}$ then with probability at least $1 - \frac{1}{2}d^{-4}$, we have for any matrices $X, Y \in \mathbb{R}^{d \times r}$:*

$$\frac{1}{p}\|XY^\top\|_\Omega^2 \leq \|X\|_F^2\|Y\|_F^2 + c_2\sqrt{\frac{d}{p}}\|X\|_F\|Y\|_F \cdot \max_{1\leq i\leq d}\|e_i^\top X\| \cdot \max_{1\leq j\leq d}\|e_j^\top Y\|.$$

On the other hand, for all low rank matrices we also have the following which is tighter for incoherent matrices.

Lemma 4. *[15]Let $d = \max\{n_1, n_2\}$, then with at least probability $1 - e^{\Omega\left(d\right)}$ over random choice of Ω, we have for any rank $2r$ matrices $A \in \mathbb{R}^{n_1 \times n_2}$:*

$$\Big|\frac{1}{p}\|\mathcal{P}_\Omega\big(A\big)\|_\Omega^2 - \|A\|_F^2\Big| \leq C\Big(\frac{d\ r\log\left(d\right)}{p}\|A\|_\infty^2 + \sqrt{\frac{d\ r\log\left(d\right)}{p}}\|A\|_F\|A\|_\infty\Big)$$

for some positive constant C.

Finally, for a matrix with each entry randomly sampled independently with small probability p, next theorem says with high probability, no row can have too many non-zero entries.

Lemma 5. *Let Ω_i denote the support of Ω on the i-th row, let $d = \max\{n_1, n_2\}$. Assume $pn_2 \geq \log\big(2d\big)$, then with probability at least*

$$1 - 2n_1 \exp\left\{-pn_2\Big(\Big(1+\frac{t}{pn_2}\Big)\ln\Big(1+\frac{t}{pn_2}\Big) - \frac{t}{pn_2}\Big)\right\}$$

over random choice of Ω, we have for all $i \in [n_1]$ simultaneously:

$$|\Omega_i| \leq C\ pn_2,$$

for some positive constant C.

7.2. The case of symmetric positive definite problems

We start with the simple definition of tubal symmetry for tensors.

Definition 19. *[19] $\mathcal{A} \in \mathbb{R}^{n\times n\times n_3}$ is a symmetric positive definite if $\hat{\mathcal{A}}^{(i)}$ are Hermitian positive definite for $i = 1, \ldots, n_3$ where $\hat{\mathcal{A}}$ is the Fast Fourier Transform (FFT) of tensor $\mathcal{A}$.*

In the following, we assume that the tensor $\mathcal{M}^\star = \mathcal{U}^\star * \big(\mathcal{U}^\star\big)^\top$ is symmetric and positive semi-definite with $\mathcal{U} \in \mathbb{R}^{n\times r\times n_3}$. The goal is to find the unknown tensor $\mathcal{U}^\star$ solution the following non-convex optimization problem

$$\min_{\mathcal{M}\in\mathbb{R}^{n\times n\times n_3}} \frac{1}{2}\big(\mathcal{M} - \mathcal{M}^\star\big) : \mathcal{H} : \big(\mathcal{M} - \mathcal{M}^\star\big) \text{ s.t. } \operatorname{rank}(\mathcal{M}) = r \qquad (25)$$

where the rank of $\mathcal{M}$ is defined in Section 3.4. Using the factorization idea of Burer and Monteiro [5], the corresponding unconstrained optimization problem with regularization Q can be written as

$$\min_{\mathcal{U}\in\mathbb{R}^{n\times n\times n_3}} \frac{1}{2}\big(\mathcal{U} * \mathcal{U}^\top - \mathcal{M}^\star\big) : \mathcal{H} : \big(\mathcal{U} * \mathcal{U}^\top - \mathcal{M}^\star\big) + Q(\mathcal{U}). \qquad (26)$$

We now present the concept of "direction of improvement" which was introduced in [15].

Definition 20. *(Direction of improvement) Let* $\mathcal{U}, \mathcal{U}^\star \in \mathbb{R}^{n\times r\times n_3}$*, define*

$$\Delta = \mathcal{U} - \mathcal{U}^\star * \mathcal{R}$$

where $\mathcal{R} \in \mathbb{R}^{r\times r\times n_3}$ *is defined as*

$$\mathcal{R} = \underset{\mathcal{Z}^\top * \mathcal{Z} = \mathcal{Z} * \mathcal{Z}^\top = \mathcal{J}}{\mathit{argmin}} \|\mathcal{U} - \mathcal{U}^\star * Z\|_F^2.$$

The direction of improvement is clearly the best direction towards the ground truth solution and the first set to take if one wants to improve the objective value. The direction of improvement is intrumental for proving Lemma 6, which is key to our analysis. This lemma Our version is an adaptation of [15, Lemma 7] to the case of low rank tubal tensor factorisation in the sense proposed by Kilmer. The main technical difficulty of adapting the proof of [15, Lemma 7] is to decouple the slices of the tensor in order to arrive at the same type of computations as in the original version of the result. This is achieved by taking the Fourier transform along the tubes.

Lemma 6. *(Main) Let* Δ *be defined as in (20) and* $\mathcal{M} = \mathcal{U} * \mathcal{U}^\top$*. Then, for any* $\mathcal{U} \in \mathbb{R}^{n\times r\times n_3}$*, we have*

$$\begin{aligned}\Delta : \nabla^2 f(\mathcal{U}) : \Delta = \Delta * \Delta^\top : \mathcal{H} : \Delta * \Delta^\top - 3(\mathcal{M} - \mathcal{M}^\star) : \mathcal{H} : (\mathcal{M} - \mathcal{M}^\star) \\ + 4\langle \nabla f(\mathcal{U}), \Delta\rangle + [\Delta : \nabla^2 Q(\mathcal{U}) : \Delta - 4\langle \nabla Q(\mathcal{U}), \Delta\rangle].\end{aligned} \tag{27}$$

Proof. We have

$$\begin{aligned} f(\mathcal{U}) &= \frac{1}{2}(\mathcal{U} * \mathcal{U}^\top - \mathcal{M}^\star) : \mathcal{H} : (\mathcal{U} * \mathcal{U}^\top - \mathcal{M}^\star) + Q(\mathcal{U}) \\ &= \frac{1}{2}\left\langle \mathcal{U} * \mathcal{U}^\top - \mathcal{M}^\star, \mathcal{H}\Big(\mathcal{U} * \mathcal{U}^\top - \mathcal{U}^\star\Big)\right\rangle + Q(\mathcal{U}) \end{aligned}$$

and we therefore get

$$\begin{aligned} f(\mathcal{U}) &= \frac{1}{2}\langle \mathcal{F}(\mathcal{U} * \mathcal{U}^\top - \mathcal{M}^\star), \mathcal{F}(\mathcal{H}(\mathcal{U} * \mathcal{U}^\top - \mathcal{M}^\star))\rangle + Q(\mathcal{U}) \\ &= \frac{1}{2}\left\langle \mathcal{F}(\mathcal{U}) * \mathcal{F}(\mathcal{U}^\top) - \mathcal{F}(\mathcal{M}^\star), \mathcal{F}(\mathcal{H}) * \mathcal{F}(\mathcal{U}) * \mathcal{F}(\mathcal{U}^\top) - \mathcal{F}(\mathcal{H}) * \mathcal{F}(\mathcal{M}^\star)\right\rangle + Q(\mathcal{U}) \end{aligned}$$

which gives

$$= \frac{1}{2}\sum_{k=1}^{n_3} \langle \mathcal{F}\big(\mathcal{U}\big)^{(k)}\mathcal{F}\big(\mathcal{U}^\top\big)^{(k)} - \mathcal{F}\big(\mathcal{M}^\star\big)^{(k)}, \mathcal{F}\big(\mathcal{H}\big)^{(k)}\mathcal{F}\big(\mathcal{U}\big)^{(k)}\mathcal{F}\big(\mathcal{U}^\top\big)^{(k)}$$

$$- \mathcal{F}\Big(\mathcal{H}\Big)^{(k)}\mathcal{F}\big(\mathcal{M}^\star\big)^{(k)}\rangle + Q(\mathcal{U})$$

and thus

$$\begin{aligned} f(\mathcal{U}) = &\frac{1}{2}\sum_{k=1}^{n_3}\Big\langle \mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\mathcal{U}^\top)^{(k)} - \mathcal{F}(\mathcal{M}^\star)^{(k)}, \mathcal{F}(\mathcal{H})^{(k)}\big(\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\mathcal{U}^\top)^{(k)} \\ &- \mathcal{F}(\mathcal{M}^\star)^{(k)}\big)\Big\rangle + Q(\mathcal{U}) \\ = &\underbrace{\frac{1}{2}\sum_{k=1}^{n_3}\left[\begin{array}{c} \Big(\mathcal{F}\Big(\mathcal{U}\Big)^{(k)}\mathcal{F}\Big(\mathcal{U}^\top\Big)^{(k)} - \mathcal{F}\Big(\mathcal{M}^\star\Big)^{(k)}\Big) \\ : \mathcal{F}\Big(\mathcal{H}\Big)^{(k)} : \\ \Big(\mathcal{F}\Big(\mathcal{U}\Big)^{(k)}\mathcal{F}\Big(\mathcal{U}^\top\Big)^{(k)} - \mathcal{F}\Big(\mathcal{M}^\star\Big)^{(k)}\Big) \end{array}\right]}_{=G(\mathcal{U})} + Q(\mathcal{U}). \end{aligned}$$

Using the fact, with for any $\mathcal{Z} \in \mathbb{R}^{n\times r\times n_3}$, we have:

$$\langle \nabla f(\mathcal{U}), \mathcal{Z}\rangle = \langle \nabla G(\mathcal{U}), \mathcal{Z}\rangle + \langle \nabla Q(\mathcal{U}), \mathcal{Z}\rangle$$

and

$$\mathcal{Z} : \nabla^2 f(\mathcal{U}) : \mathcal{Z} = \mathcal{Z} : \nabla^2 G(\mathcal{U}) : \mathcal{Z} + \mathcal{Z} : \nabla^2 Q(\mathcal{U}) : \mathcal{Z}.$$

So, we need to compute $\langle \nabla G(\mathcal{U}), \mathcal{Z}\rangle$ and $\mathcal{Z} : \nabla^2 G(\mathcal{U}) : \mathcal{Z}$. For this, by expanding the fact, for any $\mathcal{Z} \in \mathbb{R}^{n\times r\times n_3}$, we know that:

$$G\big(\mathcal{U} + \mathcal{Z}\big) = G\big(\mathcal{U}\big) + \langle \nabla G\big(\mathcal{U}\big), \mathcal{Z}\rangle + \frac{1}{2}\mathcal{Z} : \nabla^2 G\big(\mathcal{U}\big) : \mathcal{Z} + O\big(\|\mathcal{Z} * \mathcal{Z}^\top\|^2\big).$$

We obtain,

$$\begin{aligned} \langle \nabla G(\mathcal{U}), \mathcal{Z}\rangle &= \sum_{k=1}^{n_3}\left\{\begin{array}{c} \big[\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\mathcal{U}^\top)^{(k)} - \mathcal{F}(\mathcal{M}^\star)^{(k)}\big] \\ : \mathcal{F}(\mathcal{H})^{(k)} : \\ \big[\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\mathcal{Z}^\top)^{(k)} + \mathcal{F}(\mathcal{Z})^{(k)}\mathcal{F}(\mathcal{U}^\top)^{(k)}\big] \end{array}\right\} \\ &= \big(\mathcal{F}(\mathcal{M}) - \mathcal{F}(\mathcal{M}^\star)\big) : \mathcal{F}(\mathcal{H}) : \big(\mathcal{F}(\mathcal{U}) * \mathcal{F}(\mathcal{Z}^\top) + \mathcal{F}(\mathcal{Z}) * \mathcal{F}(\mathcal{U}^\top)\big) \\ &= \big(\mathcal{M} - \mathcal{M}^\star\big) : \mathcal{H} : \big(\mathcal{U} * \mathcal{Z}^\top + \mathcal{Z} * \mathcal{U}^\top\big) \qquad (28) \end{aligned}$$

and

$$\mathcal{Z} : \nabla^2 G(\mathcal{U}) : \mathcal{Z} = \sum_{k=1}^{n_3} \left\{ \begin{array}{c} \left[\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(Z^\top)^{(k)} + \mathcal{F}(\mathcal{Z})^{(k)}\mathcal{F}(\mathcal{U}^\top)^{(k)}\right] \\ : \mathcal{F}(\mathcal{H})^{(k)} : \\ \left[\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\mathcal{Z}^\top)^{(k)} + \mathcal{F}(\mathcal{Z})^{(k)}\mathcal{F}(\mathcal{U}^\top)^{(k)}\right] \end{array} \right\}$$

$$\begin{aligned} &+ 2\left[\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\mathcal{U}^\top)^{(k)} - \mathcal{F}(\mathcal{M}^\star)^{(k)}\right] : \mathcal{F}(\mathcal{H})^{(k)} : \left[\mathcal{F}(\mathcal{Z})^{(k)}\mathcal{F}(\mathcal{Z}^\top)^{(k)}\right] \\ = &\big(\mathcal{F}(\mathcal{U}) * \mathcal{F}(\mathcal{Z}^\top) + \mathcal{F}(\mathcal{Z}) * \mathcal{F}(\mathcal{U}^\top)\big) : \mathcal{F}(\mathcal{H}) : \big(\mathcal{F}(\mathcal{U}) * \mathcal{F}(\mathcal{Z}^\top) + \mathcal{F}(\mathcal{Z}) * \mathcal{F}(\mathcal{U}^\top)\big) \\ &+ 2\big(\mathcal{F}(\mathcal{M}) - \mathcal{F}(\mathcal{M}^\star)\big) : \mathcal{F}(\mathcal{H}) : \mathcal{F}(\mathcal{Z}) * \mathcal{F}(\mathcal{Z}^\top) \\ = &\big(\mathcal{U} * \mathcal{Z}^\top + \mathcal{Z} * \mathcal{U}^\top\big) : \mathcal{H} : \big(\mathcal{U} * \mathcal{Z}^\top + \mathcal{Z} * \mathcal{U}^\top\big) + 2\big(\mathcal{M} - \mathcal{M}^\star\big) : \mathcal{H} : \mathcal{Z} * \mathcal{Z}^\top. \end{aligned} \tag{29}$$

In the last equality of (28) and (29), we use the linearity of Fourier transform and the inverse of FFT. Let $\mathcal{Z} = \Delta = \mathcal{U} - \mathcal{U}^\star * \mathcal{R}$ and $\mathcal{M} - \mathcal{M}^\star + \Delta * \Delta^\top = \mathcal{U} * \Delta^\top + \Delta * \mathcal{U}^\top$. Indeed,

$$\mathcal{U} * \Delta^\top + \Delta * \mathcal{U}^\top = \mathcal{U} * \mathcal{U}^\top - \mathcal{U} * \mathcal{R}^\top * \mathcal{U}^{\star\top} + \mathcal{U} * \mathcal{U}^\top - \mathcal{U}^\star * \mathcal{R} * \mathcal{U}^\top$$

and using that $\mathcal{R} * \mathcal{R}^\top = \mathcal{J}$, where $\mathcal{J}$ is a identity tensor, we have

$$\begin{aligned} \mathcal{M} - \mathcal{M}^\star + \Delta * \Delta^\top &= \mathcal{U} * \mathcal{U}^\top - \mathcal{U}^\star * \mathcal{U}^{\star\top} + \big(\mathcal{U} - \mathcal{U}^\star * \mathcal{R}\big) * \big(\mathcal{U} - \mathcal{U}^\star * \mathcal{R}\big)^\top \\ &= \mathcal{U} * \mathcal{U}^\top - \mathcal{U} * \mathcal{R}^\top * \mathcal{U}^{\star\top} + \mathcal{U} * \mathcal{U}^\top - \mathcal{U}^\star * \mathcal{R} * \mathcal{U}^\top. \end{aligned}$$

Using

$$\big(\mathcal{M} - \mathcal{M}^\star\big) : \mathcal{H} : \big(\mathcal{M} - \mathcal{M}^\star + \Delta * \Delta^\top\big) = \langle \nabla f(\mathcal{U}), \Delta\rangle - \langle \nabla Q(\mathcal{U}), \Delta\rangle,$$

we have

$$\begin{aligned} \Delta : \nabla^2 f(\mathcal{U}) : \Delta = &\big(\mathcal{U} * \Delta^\top + \Delta * \mathcal{U}^\top\big) : \mathcal{H} : \big(\mathcal{U} * \Delta^\top + \Delta * \mathcal{U}^\top\big) \\ &+2\big(\mathcal{M} - \mathcal{M}^\star\big) : \mathcal{H} : \Delta * \Delta^\top + \Delta : \nabla^2 Q(\mathcal{U}) : \Delta \\ = &\big(\mathcal{M} - \mathcal{M}^\star + \Delta * \Delta^\top\big) : \mathcal{H} : \big(\mathcal{M} - \mathcal{M}^\star + \Delta * \Delta^\top\big) \\ &+2\big(\mathcal{M} - \mathcal{M}^\star\big) : \mathcal{H} : \Delta * \Delta^\top + \Delta : \nabla^2 Q(\mathcal{U}) : \Delta \\ = &\Delta * \Delta^\top : \mathcal{H} : \Delta * \Delta^\top + \big(\mathcal{M} - \mathcal{M}^\star\big) : \mathcal{H} : \big(\mathcal{M} - \mathcal{M}^\star\big) \\ &+4\big(\mathcal{M} - \mathcal{M}^\star\big) : \mathcal{H} : \Delta * \Delta^\top + \Delta : \nabla^2 Q(\mathcal{U}\mathcal{M}) : \Delta \\ = &\Delta * \Delta^\top : \mathcal{H} : \Delta * \Delta^\top - 3\big(\mathcal{M} - \mathcal{M}^\star\big) : \mathcal{H} : \big(\mathcal{M} - \mathcal{M}^\star\big) \\ &+4\big(\mathcal{M} - \mathcal{M}^\star\big) : \mathcal{H} : \big(\mathcal{M} - \mathcal{M}^\star + \Delta * \Delta^\top\big) + \Delta : \nabla^2 Q(\mathcal{U}) : \Delta \\ = &\Delta * \Delta^\top : \mathcal{H} : \Delta * \Delta^\top - 3(\mathcal{M} - \mathcal{M}^\star) : \mathcal{H} : (\mathcal{M} - \mathcal{M}^\star) \\ &+4\langle \nabla f(\mathcal{U}), \Delta\rangle + [\Delta : \nabla^2 Q(\mathcal{U}) : \Delta - 4\langle \nabla Q(\mathcal{U}), \Delta\rangle]. \end{aligned}$$

□

Using the previous lemma, we are now able to prove the following result.

Lemma 7. *Given tensors* $\mathcal{U}, \mathcal{U}^\star \in \mathbb{R}^{n\times r\times n_3}$*. Let* $\mathcal{M} = \mathcal{U}*\mathcal{U}^\top$*,* $\mathcal{M}^\star = \mathcal{U}^\star*{\mathcal{U}^\star}^\top$*, and* Δ *be defined as in* (20)*, then we have*

$$\|\Delta * \Delta^\top\|_F^2 \le 2\|\mathcal{M} - \mathcal{M}^\star\|_F^2 \quad \textit{and} \quad \sigma_r^\star\|\Delta\|_F^2 \le \frac{1}{2(\sqrt{2}-1)}\|\mathcal{M} - \mathcal{M}^\star\|_F^2.$$

Proof. We begin to show that

$$\mathcal{U}^\top * \mathcal{U}^\star * \mathcal{R}_\mathcal{U} \text{ is a symmetric } \textbf{PSD} \text{ tensor.} \tag{30}$$

where $\mathcal{R}_\mathcal{U} = \underset{\mathcal{R}*\mathcal{R}^\top=\mathcal{R}^\top*\mathcal{R}=\mathcal{J}}{\arg\min} \|\mathcal{U} - \mathcal{U}^\star * \mathcal{R}\|_F^2$. By developping the Frobenius norm and letting the t-SVD of ${\mathcal{U}^\star}^\top * \mathcal{U}$ be $\mathcal{A} * \mathcal{D} * \mathcal{B}^\top$, we have:

$$\|\mathcal{U} - \mathcal{U}^\star * \mathcal{R}\|_F^2 = \|\mathcal{U} * \mathcal{U}^\top\|_F^2 - 2\langle \mathcal{U}, \mathcal{U}^\star * \mathcal{R}\rangle + \|\mathcal{U}^\star * {\mathcal{U}^\star}^\top\|_F^2.$$

Hence,

$$\begin{aligned}
\underset{\mathcal{R}*\mathcal{R}^\top=\mathcal{R}^\top*\mathcal{R}=\mathcal{J}}{\arg\min} \|\mathcal{U} - \mathcal{U}^\star * \mathcal{R}\|_F^2 &= \underset{\mathcal{R}*\mathcal{R}^\top=\mathcal{R}^\top*\mathcal{R}=\mathcal{J}}{\arg\min} -\langle \mathcal{U}, \mathcal{U}^\star * R\rangle \\
&= \underset{\mathcal{R}*\mathcal{R}^\top=\mathcal{R}^\top*\mathcal{R}=\mathcal{J}}{\arg\min} -\langle \mathcal{F}(\mathcal{U}), \mathcal{F}(\mathcal{U}^\star) * \mathcal{F}(\mathcal{R})\rangle \\
&= \underset{\mathcal{R}*\mathcal{R}^\top=\mathcal{R}^\top*\mathcal{R}=\mathcal{J}}{\arg\min} -\sum_{k=1}^{n_3} \langle \mathcal{F}(\mathcal{U})^{(k)}, \mathcal{F}(\mathcal{U}^\star)^{(k)}\mathcal{F}(\mathcal{R})^{(k)}\rangle \\
&= \underset{\mathcal{R}*\mathcal{R}^\top=\mathcal{R}^\top*\mathcal{R}=J}{\arg\min} -\sum_{k=1}^{n_3} \operatorname{trace}\Big(\mathcal{F}(\mathcal{U})^{(k)\top}\mathcal{F}(\mathcal{U}^\star)^{(k)}\mathcal{F}(\mathcal{R})^{(k)}\Big) \\
&= \underset{\mathcal{R}*\mathcal{R}^\top=\mathcal{R}^\top*\mathcal{R}=\mathcal{J}}{\arg\min} -\sum_{k=1}^{n_3} \operatorname{trace}\Big(\mathcal{F}\big(\mathcal{D}\big)^{(k)}\mathcal{F}\big(\mathcal{A}^\top\big)^{(k)}\mathcal{F}(\mathcal{R})^{(k)}\mathcal{F}\big(\mathcal{B}\big)^{(k)}\Big).
\end{aligned}$$

Since $\mathcal{A}$, $\mathcal{R}$ and $\mathcal{B}$ are orthogonal tensors, then $\mathcal{F}\big(\mathcal{A}^\top\big)^{(k)}$, $\mathcal{F}(\mathcal{R})^{(k)}$ and

$\mathcal{F}\big(\mathcal{B}\big)^{(k)}$ are orthogonal matrices. For any orthogonal tensor $\mathcal{T}$, we have

$$\begin{aligned}
\operatorname{trace}\big(\mathcal{D} * \mathcal{T}\big) &= \langle \mathcal{D}, \mathcal{T} \rangle = \langle \mathcal{F}(\mathcal{D}), \mathcal{F}(\mathcal{T}) \rangle \\
&= \sum_{k=1}^{n_3} \operatorname{trace}\big(\mathcal{F}(\mathcal{D})^{(k)} \mathcal{F}(\mathcal{T})^{(k)}\big) \\
&= \sum_{k=1}^{n_3} \sum_{i=1}^{r} \mathcal{F}(\mathcal{D})_{ii}^{(k)} \mathcal{F}(\mathcal{T})_{ii}^{(k)} \\
&\leq \sum_{k=1}^{n_3} \sum_{i=1}^{r} \mathcal{F}(\mathcal{D})_{ii}^{(k)}
\end{aligned}$$

where the last inequality uses the fact that $\mathcal{F}(\mathcal{D})_{ii}^{(k)}$ is a positive singular values and $\mathcal{T}$ is an orthogonal tensor thus $\mathcal{F}(\mathcal{T})_{ii}^{(k)} \leq 1$. This implies that the maximum of $\mathcal{F}(\mathcal{D})_{ii}^{(k)} \mathcal{F}(\mathcal{T})_{ii}^{(k)}$ is attained at $\mathcal{T} = \mathcal{J}$. In other words, the minimum is attained when

$$-\sum_{k=1}^{n_3} \operatorname{trace}\big(\mathcal{F}\big(\mathcal{D}\big)^{(k)} \mathcal{F}\big(\mathcal{A}^\top\big)^{(k)} \mathcal{F}(\mathcal{R})^{(k)} \mathcal{F}\big(\mathcal{B}\big)^{(k)}\big)$$

is attained when

$$\mathcal{R} = \mathcal{A} * \mathcal{B}^\top.$$

Finally, since

$$\mathcal{U}^\top * \mathcal{U}^\star * \mathcal{R}_{\mathcal{U}} = \mathcal{B} * \mathcal{D} * \mathcal{A}^\top * \mathcal{R}_{\mathcal{U}} = \mathcal{B} * \mathcal{D} * \mathcal{B}^\top$$

we get that $\mathcal{U}^\top * \mathcal{U}^\star * \mathcal{R}_{\mathcal{U}}$ is a symmetric PSD tensor and the proof is completed. □

The following technical result follows the lines of the analysis provided in [15] and shows how one can control the factorisation of differences using the differences of factorisations.

Lemma 8. *Let $\mathcal{U}$ and $\mathcal{Y}$ be two $n \times n \times n_3$ tensors. Let $\mathcal{U}^\top * \mathcal{Y} = \mathcal{Y}^\top * \mathcal{U}$ be a PSD tensor. Then,*

$$\|\big(\mathcal{U} - \mathcal{Y}\big) * \big(\mathcal{U} - \mathcal{Y}\big)^\top\|_F^2 \leq \|\mathcal{U} * \mathcal{U}^\top - \mathcal{Y} * \mathcal{Y}^\top\|_F^2.$$

Proof. Let $\Delta = \mathcal{U} - \mathcal{Y}$, and we have

$$
\begin{aligned}
\|\mathcal{U} * \mathcal{U}^\top - \mathcal{Y} * \mathcal{Y}^\top\|_F^2 =& \|\mathcal{U} * \Delta^\top + \Delta * \mathcal{U}^\top - \Delta * \Delta^\top\|_F^2 \\
=& \text{trace}\Big(\Delta * \mathcal{U}^\top * \mathcal{U} * \Delta^\top + \Delta * \mathcal{U}^\top * \Delta * \mathcal{U}^\top \\
&- \Delta * \mathcal{U}^\top * \Delta * \Delta^\top + \mathcal{U} * \Delta^\top * \mathcal{U} * \Delta^\top\Big) \\
&+ \text{trace}\Big(\mathcal{U} * \Delta^\top * \Delta * \mathcal{U}^\top - \mathcal{U} * \Delta^\top * \Delta * \Delta^\top \\
&- \Delta * \Delta^\top * \mathcal{U} * \Delta^\top - \Delta * \Delta^\top * \Delta * \mathcal{U}^\top\Big) \\
&+ \text{trace}\big(\Delta * \Delta^\top * \Delta * \Delta^\top\big).
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
\text{trace}\big(\Delta * \mathcal{U}^\top * \mathcal{U} * \Delta^\top\big) &= \langle \mathcal{F}(\Delta) \cdot \mathcal{F}(\mathcal{U}^\top), \mathcal{F}(\Delta) \cdot \mathcal{F}(\mathcal{U}^\top) \rangle \\
&= \sum_{k=1}^{n_3} \text{trace}\big(\mathcal{F}(\Delta)^{(k)} \mathcal{F}(\mathcal{U}^\top)^{(k)} \mathcal{F}(\mathcal{U})^{(k)} \mathcal{F}(\Delta^\top)^{(k)}\big) \\
&= \sum_{k=1}^{n_3} \text{trace}\big(\mathcal{F}(\mathcal{U}^\top)^{(k)} \mathcal{F}(\mathcal{U})^{(k)} \mathcal{F}(\Delta^\top)^{(k)} \mathcal{F}(\Delta)^{(k)}\big) \\
&= \text{trace}\big(\mathcal{U}^\top * \mathcal{U} * \Delta^\top * \Delta\big)
\end{aligned}
$$

$$
\begin{aligned}
\text{trace}\big(\Delta * \mathcal{U}^\top * \Delta * \Delta^\top\big) &= \langle \mathcal{F}(\Delta) \cdot \mathcal{F}(\mathcal{U}^\top), \mathcal{F}(\Delta^\top) \cdot \mathcal{F}(\Delta) \rangle \\
&= \sum_{k=1}^{n_3} \text{trace}\big(\mathcal{F}(\Delta)^{(k)} \mathcal{F}(\mathcal{U}^\top)^{(k)} \mathcal{F}(\Delta)^{(k)} \mathcal{F}(\Delta^\top)^{(k)}\big) \\
&= \sum_{k=1}^{n_3} \text{trace}\big(\mathcal{F}(\mathcal{U}^\top)^{(k)} \mathcal{F}(\Delta)^{(k)} \mathcal{F}(\Delta^\top)^{(k)} \mathcal{F}(\Delta)^{(k)}\big) \\
&= \text{trace}\big(\mathcal{U}^\top * \Delta * \Delta^\top * \Delta\big).
\end{aligned}
$$

In a similar manner, we get

- $\text{trace}\big(\Delta * \mathcal{U}^\top * \Delta * \Delta^\top\big) = \text{trace}\big(\mathcal{U} * \Delta^\top * \Delta * \Delta^\top\big) = \text{trace}\big(\Delta * \Delta^\top * \mathcal{U} * \Delta^\top\big) = \text{trace}\big(\Delta * \Delta^\top * \Delta * \mathcal{U}^\top\big)$
- $\text{trace}\big(\Delta * \mathcal{U}^\top * \mathcal{U} * \Delta^\top\big) = \text{trace}\big(\mathcal{U} * \Delta^\top * \Delta * \mathcal{U}^\top\big)$
- $\text{trace}\big(\Delta * \mathcal{U}^\top * \Delta * \mathcal{U}^\top\big) = \text{trace}\big(\mathcal{U} * \Delta^\top * \mathcal{U} * \Delta^\top\big)$.

Therefore using that $\mathcal{U}^\top * \mathcal{Y} = \mathcal{Y}^\top * \mathcal{U}$, we have

$$
\begin{aligned}
&\|\mathcal{U} * \mathcal{U}^\top - \mathcal{Y} * \mathcal{Y}^\top\|_F^2 \\
=&\operatorname{trace}\big(2\mathcal{U}^\top * \mathcal{U} * \Delta^\top * \Delta + \Delta^\top * \Delta * \Delta^\top * \Delta + \mathcal{U}^\top * \Delta * \mathcal{U}^\top * \Delta - 4\mathcal{U}^\top * \Delta * \Delta^\top * \Delta\big) \\
=&\operatorname{trace}\big(2\mathcal{U}^\top * \mathcal{U} * \Delta^\top * \Delta + 2\big(\mathcal{U}^\top * \Delta\big)^2 + \big(\Delta^\top * \Delta\big)^2 - 4\mathcal{U}^\top * \Delta * \Delta^\top * \Delta\big) \\
=&\operatorname{trace}\big(2\mathcal{U}^\top * \big(\mathcal{U} - \Delta\big) * \Delta^\top * \Delta + \big(\frac{1}{\sqrt{2}}\Delta^\top * \Delta - \sqrt{2}\mathcal{U}^\top * \Delta\big)^2 + \frac{1}{2}\big(\Delta^\top * \Delta\big)^2\big) \\
\geq&\operatorname{trace}\big(2\mathcal{U}^\top * \mathcal{Y} * \Delta^\top * \Delta + \frac{1}{2}\big(\Delta^\top * \Delta\big)^2\big) \\
\geq&\frac{1}{2}\|\Delta * \Delta^\top\|_F^2
\end{aligned}
$$

where the last inequality is a consequence of the fact that $\mathcal{U}^\top * \mathcal{Y}$ is a positive semi-definite tensor. $\square$

The next lemma will also be key.

Lemma 9. *Let $\mathcal{U}$ and $\mathcal{Y}$ be two $n \times n \times n_3$ tensors. Let $\mathcal{U}^\top * \mathcal{Y} = \mathcal{Y}^\top * \mathcal{U}$ be a PSD tensor. Then,*

$$
\begin{aligned}
\sigma_{\min}\big(\mathcal{U}^\top * \mathcal{U}\big)\|\mathcal{U} - \mathcal{Y}\|_F^2 &\leq \|\big(\mathcal{U} - \mathcal{Y}\big) * \mathcal{U}^\top\|_F^2 \\
&\leq \frac{1}{2\big(\sqrt{2} - 1\big)}\|\mathcal{U} * \mathcal{U}^\top - \mathcal{Y} * \mathcal{Y}^\top\|_F^2.
\end{aligned}
$$

Proof. Let $\Delta = \mathcal{U} - \mathcal{Y}$, and we have

$$
\begin{aligned}
&\|\mathcal{U} * \mathcal{U} - \mathcal{Y} * \mathcal{Y}^\top\|_F^2 = \|\mathcal{U} * \Delta^\top + \Delta * \mathcal{U}^\top - \Delta * \Delta^\top\|_F^2 \\
&= \operatorname{trace}\big(2\mathcal{U}^\top * \mathcal{U} * \Delta^\top * \Delta + 2\big(\mathcal{U}^\top * \Delta\big)^2 + \big(\Delta^\top * \Delta\big)^2 - 4\mathcal{U}^\top * \Delta * \Delta^\top * \Delta\big) \\
&\geq \operatorname{trace}\big(\big(4 - 2\sqrt{2}\big)\mathcal{U}^\top * \mathcal{Y} * \Delta^\top * \Delta + 2\big(\sqrt{2} - 1\big)\mathcal{U}^\top * \mathcal{U} * \Delta^\top * \Delta\big) \\
&\geq 2\big(\sqrt{2} - 1\big)\|\mathcal{U} * \Delta^\top\|_F^2.
\end{aligned}
$$

where the last inequality uses the positive semidefiniteness of $\mathcal{U}^\top * \mathcal{Y}$. $\square$

Combining 8 and 9, it now within reach to obtain Lemma 7, after replacing $\mathcal{U}$ by $\mathcal{U}^\star * \mathcal{R}_{\mathcal{U}}$ and $\mathcal{Y}$ by $\mathcal{U}$.

Let us now turn to clarifying the interaction between the Hessian and the regulariser. The necessity of using a penalisation (regularisation) comes from the deficiency of the Hessian operator $\mathcal{H}$ in preserving the norm of all low rank tubal tensors. A standard approach to making the Bürer Monteiro successful is to impose some incoherence on the matrix to be reconstructed such as proposed in the following definition.

Definition 21. *[36] Let $\mathcal{M} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and its t-SVD of the form $\mathcal{M} = \mathcal{X} * \mathcal{D} * \mathcal{Y}^\top$. Let $r = rank(\mathcal{M})$. Then, $\mathcal{M}$ is said to satisfy the tensor incoherence property with parameter $\mu > 0$, if*

$$\max_{i=1,\dots,n_1} \|e_i^\top * \mathcal{X}\|_F \leq \sqrt{\frac{\mu r}{n_1}}$$
$$\max_{j=1,\dots,n_2} \|e_j^\top * \mathcal{Y}\|_F \leq \sqrt{\frac{\mu r}{n_2}}$$

where e_i is the $n_1 \times 1 \times n_3$ column basis with $e_{i11} = 1$ and e_j is the $n_2 \times 1 \times n_3$ column basis with $e_{j11} = 1$.

In the following, we will assume that our unknown low rank tensor $\mathcal{M}^\star$ is μ-incoherent.

In the non-convex problem, we try to make sure that the decomposition $\mathcal{U} * \mathcal{U}^\top$ is also incoherent by adding a regularizer of [15], that penalize the function objective when some row of $\mathcal{F}(\mathcal{U})^{(k)}, k = 1 \dots, n_3$ is too large.

$$Q(\mathcal{U}) = \lambda \sum_{k=1}^{n_3} \sum_{i=1}^{n} \left(\|\mathcal{F}(e_i^\top)^{(k)} \mathcal{F}(\mathcal{U})^{(k)}\|_2 - \alpha \right)_+^4 .$$

Here λ, α are parameters that we choose later, $(x)_+ = \max\{x, 0\}$. By adding this regularizer, we can transform the objective function to the unconstrained form

$$\min_{\mathcal{U} \in \mathbb{R}^{n \times r \times n_3}} \quad \frac{1}{2p} \|\mathcal{U} * \mathcal{U}^\top - \mathcal{M}^\star\|_\Omega^2 + Q(\mathcal{U}). \tag{31}$$

Using this fact we begin to show that the regularizer ensures that all rows of $\mathcal{F}(\mathcal{U})^{(k)}, k = 1 \dots, n_3$ are small.

We now study the properties of the gradient and Hessian of the regularizer Q:

Lemma 10. *The gradient and the Hessian of the regularizer*

$$Q(\mathcal{U}) = \lambda \sum_{k=1}^{n_3} \sum_{i=1}^{n} \big(\|\mathcal{F}(e_i^\top)^{(k)} \mathcal{F}(\mathcal{U})^{(k)}\|_2 - \alpha\big)_+^4$$

is:

$$\langle \nabla Q(\mathcal{U}), \mathcal{Z} \rangle = 4\lambda \sum_{k=1}^{n_3} \sum_{i=1}^{n} \big((\|\mathcal{F}(e_i^\top)^{(k)} \mathcal{F}(\mathcal{U})^{(k)}\|_2 - \alpha)_+^3 \\ \times \frac{\mathcal{F}(e_i^\top)^{(k)} \mathcal{F}(\mathcal{U})^{(k)} \mathcal{F}(\mathcal{Z})^{(k)^\top} \mathcal{F}(e_i)^{(k)}}{\|\mathcal{F}(e_i^\top)^{(k)} \mathcal{F}(\mathcal{U})^{(k)}\|_2} \big). \tag{32}$$

$$\mathcal{Z} : \nabla^2 Q(\mathcal{U}) : \mathcal{Z} = \\ 12\lambda \sum_{k=1}^{n_3} \sum_{i=1}^{n} \Big(\Big(\|\mathcal{F}(e_i^\top)^{(k)} \mathcal{F}(\mathcal{U})^{(k)}\|_2 - \alpha\Big)_+^2 \big(\frac{\mathcal{F}(e_i^\top)^{(k)} \mathcal{F}(\mathcal{U})^{(k)} \mathcal{F}(\mathcal{Z})^{(k)^\top} \mathcal{F}(e_i)^{(k)}}{\|\mathcal{F}(e_i^\top)^{(k)} \mathcal{F}(\mathcal{U})^{(k)}\|_2}\big)^2 \Big) \\ + 4\lambda \sum_{k=1}^{n_3} \sum_{i=1}^{n} \big((\|\mathcal{F}(e_i^\top)^{(k)} \mathcal{F}(\mathcal{U})^{(k)}\|_2 - \alpha)_+^3 \tag{33} \\ \times \frac{\|\mathcal{F}(e_i^\top)^{(k)} \mathcal{F}(\mathcal{U})^{(k)}\|_2^2 \, \|\mathcal{F}(e_i^\top)^{(k)} \mathcal{F}(\mathcal{Z})^{(k)}\|_2^2 - \big(\mathcal{F}(e_i^\top)^{(k)} \mathcal{F}(\mathcal{U})^{(k)} \mathcal{F}(Z)^{(k)^\top} \mathcal{F}(e_i)^{(k)}\big)^2}{\|\mathcal{F}(e_i^\top)^{(k)} \mathcal{F}(\mathcal{U})^{(k)}\|_2^3} \big).$$

Proof. Let

$$\varphi(\mathcal{U}) = \sum_{k=1}^{n_3} \sum_{i=1}^{n} h_i\big(\mathcal{F}(\mathcal{U})^{(k)} + t\mathcal{F}(\mathcal{Z})^{(k)}\big) - h_i(\mathcal{U})$$

where

$$h_i(\mathcal{M}) = \big(\|\mathcal{F}(e_i^\top)^{(k)} \mathcal{F}(\mathcal{M})^{(k)}\|_2 - \alpha\big)_+^4.$$

We will have to determine the directional derivative of φ in the direction of $\mathcal{F}(\mathcal{Z})^{(k)}$ for $k = 1, \ldots, n_3$. Suppose that $\|e_i^\top * \mathcal{U}\|_F \geq \alpha$, so for all sufficiently small t and for any $k = 1, \ldots, n_3$, we have

$$\Big(\|\mathcal{F}(e_i^\top)^{(k)} \mathcal{F}(\mathcal{U})^{(k)}\|_2 - \alpha\Big)_+^4 = \Big(\|\mathcal{F}(e_i^\top)^{(k)} \mathcal{F}(\mathcal{U})^{(k)}\| - \alpha\Big)^4.$$

Hence, we have

$$h_i(\mathcal{U}) = g(f_i(\mathcal{U}))\,\text{with}\, g : x \mapsto x^4$$

as well as

$$f_i(\mathcal{U}) = \|\mathcal{F}\big(e_i^\top\big)^{(k)}\mathcal{F}\big(\mathcal{U}\big)^{(k)}\|_2 - \alpha.$$

$$\partial h_i(\mathcal{U}) = \partial g(f_i(\mathcal{U})) \cdot \partial f_i(\mathcal{U})$$

and

$$\partial^2 h_i(\mathcal{U}) = \partial^2(g(f_i(\mathcal{U}))) \cdot \partial f_i(\mathcal{U}) + \partial g(f_i(\mathcal{U})) \cdot \partial^2 f_i(\mathcal{U})$$

with

$$\partial g(f_i(\mathcal{U})) = 4\big(\|\mathcal{F}\big(e_i^\top\big)^{(k)}\mathcal{F}\big(\mathcal{U}\big)^{(k)}\|_2 - \alpha\big)^3,$$

$$\partial f_i\big(\mathcal{U}\big) = \frac{\mathcal{F}\big(e_i^\top\big)^{(k)}\mathcal{F}\big(\mathcal{U}\big)^{(k)}\mathcal{F}\big(\mathcal{Z}\big)^{(k)^\top}\mathcal{F}\big(e_i\big)^{(k)}}{\|\mathcal{F}\big(e_i^\top\big)^{(k)}\mathcal{F}\big(\mathcal{U}\big)^{(k)}\|_2}$$

and

$$\partial^2(g(f_i(\mathcal{U}))) = 12\big(\|\mathcal{F}\big(e_i^\top\big)^{(k)}\mathcal{F}\big(\mathcal{U}\big)^{(k)}\|_2 - \alpha\big)^2 \frac{\mathcal{F}\big(e_i^\top\big)^{(k)}\mathcal{F}\big(\mathcal{U}\big)^{(k)}\mathcal{F}\big(\mathcal{Z}\big)^{(k)^\top}\mathcal{F}\big(e_i\big)^{(k)}}{\|\mathcal{F}\big(e_i^\top\big)^{(k)}\mathcal{F}\big(\mathcal{U}\big)^{(k)}\|_2}$$

and thus,

$$\partial^2 f_i(\mathcal{U}) = \frac{\|\mathcal{F}\big(e_i^\top\big)^{(k)}\mathcal{F}\big(\mathcal{U}\big)^{(k)}\|_2^2\|\mathcal{F}\big(e_i^\top\big)^{(k)}\mathcal{F}\big(\mathcal{Z}\big)^{(k)}\|_2^2 - \big(\mathcal{F}\big(e_i^\top\big)^{(k)}\mathcal{F}\big(\mathcal{U}\big)^{(k)}\mathcal{F}\big(\mathcal{Z}\big)^{(k)^\top}\mathcal{F}\big(e_i\big)^{(k)}\big)^2}{\|\mathcal{F}\big(e_i^\top\big)^{(k)}\mathcal{F}\big(\mathcal{U}\big)^{(k)}\|_2^3}$$

With this result in hand, the remainder of the proof follows in a straightforward manner. □

Lemma 11. *There exists an absolute constant c, such that when the probability p satisfies*

$$p > c_1 \frac{\mu r \ \log(n)}{n}, \ \alpha^2 = \Theta\big(\frac{\mu r \sigma_1^\star}{n}\big) \ \textit{and} \ \lambda = \Theta\big(\frac{n}{\mu r \kappa^\star}\big),$$

we have for any U with $\|\nabla f(\mathcal{U})\|_F \leq \epsilon$ for any polynomial small ϵ, with probability at least

$$1 - 2nn_3 \exp\left\{-pn\big(\big(1+\frac{t}{pn}\big)\ln\big(1+\frac{t}{pn}\big) - \frac{t}{pn}\big)\right\},$$

$$\begin{aligned}\max_{1\leq i\leq n} \|e_i^\top * \mathcal{U}\|_F^2 &= \max_{1\leq i\leq n} \|\mathcal{F}(e_i^\top)\cdot \mathcal{F}(\mathcal{U})\|_F^2 \\ &= \max_{1\leq i\leq n} \sum_{k=1}^{n_3} \|\mathcal{F}\big(e_i^\top\big)^{(k)} \mathcal{F}\big(\mathcal{U}\big)^{(k)}\|_2^2 \leq C\ n_3 \frac{(\mu r)^{1.5}\kappa^\star \sigma_1^\star}{n}\end{aligned}$$

for some constant positive C.

Proof. We first show that the regulariser forces the tensor $\mathcal{U}$ to have small rows, i.e, prove the Lemma 11. By Lemma 10, we know that:

$$\begin{aligned}\nabla Q(\mathcal{U}) = 4\lambda \sum_{k=1}^{n_3}\sum_{i=1}^{n} \big(\big(\|\mathcal{F}\big(e_i^\top\big)^{(k)} \mathcal{F}\big(\mathcal{U}\big)^{(k)}\|_2 - \alpha\big)_+^3 \\ \times \frac{\mathcal{F}\big(e_i\big)^{(k)} \mathcal{F}\big(e_i^\top\big)^{(k)} \mathcal{F}\big(\mathcal{U}\big)^{(k)}}{\|\mathcal{F}\big(e_i^\top\big)^{(k)} \mathcal{F}\big(\mathcal{U}\big)^{(k)}\|_2}\big).\end{aligned} \tag{34}$$

Using this formula, we have

$$\begin{aligned}\nabla f\big(\mathcal{U}\big) &= \frac{2}{p}\big(M - M^\star\big)_\Omega * \mathcal{U} + \nabla Q(\mathcal{U}) \\ &= \frac{2}{p}\sum_{k=1}^{n_3} \big(\mathcal{F}\big(\mathcal{M}\big)^{(k)} - \mathcal{F}\big(\mathcal{M}^\star\big)^{(k)}\big)_\Omega \mathcal{F}\big(\mathcal{U}\big)^{(k)} + \nabla Q(\mathcal{U}).\end{aligned} \tag{35}$$

Let us study the potential consequence of having $\|e_{i^\star}^\top * \mathcal{U}\|_F \geq 2\alpha$. Consider the gradient along $e_{i^\star} * e_{i^\star}^\top * \mathcal{U}$ direction. Since $\|\nabla f(\mathcal{U})\|_F \leq \epsilon$, we have

$$\langle \nabla f(\mathcal{U}), e_{i^\star} * e_{i^\star}^\top * \mathcal{U}\rangle = \langle e_{i^\star}^\top * \nabla f(\mathcal{U}), e_{i^\star}^\top * \mathcal{U}\rangle \leq \epsilon \|e_{i^\star}^\top * \mathcal{U}\|_F.$$

Therefore, with probability larger than

$$1 - 2nn_3 \exp\left\{-pn\big(\big(1+\frac{t}{pn}\big)\ln\big(1+\frac{t}{pn}\big) - \frac{t}{pn}\big)\right\},$$

using equalities (34) and (35) the followings holds:

$$\epsilon\|e_{i^\star}^\top * \mathcal{U}\|_F = \epsilon\|\mathcal{F}\big(e_{i^\star}^\top\big)\cdot\mathcal{F}\big(\mathcal{U}\big)\|_F = \epsilon\sum_{k=1}^{n_3}\|\mathcal{F}\big(e_{i^\star}^\top\big)^{(k)}\mathcal{F}\big(\mathcal{U}\big)^{(k)}\|_2.$$

Now, for any $k = 1, \ldots, n_3$, we have

$$\begin{aligned}
&\epsilon\|\mathcal{F}\big(e_{i^\star}^\top\big)^{(k)}\mathcal{F}\big(\mathcal{U}\big)^{(k)}\|_2\\
&\overset{\#}{\geq} 4\lambda\big(\|\mathcal{F}(e_{i^\star}^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2-\alpha\big)_+^3\|\mathcal{F}(e_{i^\star}^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2\\
&-\frac{2}{p}\langle\mathcal{F}\big(e_{i^\star}^\top\big)^{(k)}\big(\mathcal{F}(\mathcal{M}^\star)^{(k)}\big)_\Omega, \mathcal{F}\big(e_{i^\star}^\top\big)^{(k)}\big(\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\mathcal{U}^\top)^{(k)}\big)_\Omega\rangle\\
&\overset{\#\#}{\geq}\frac{\lambda}{2}\|\mathcal{F}(e_{i^\star}^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2^4\\
&-2\cdot\frac{1}{\sqrt{p}}\|\mathcal{F}(e_{i^\star}^\top)^{(k)}\big(\mathcal{F}(\mathcal{M}^\star)^{(k)}\big)_\Omega\|_2\cdot\frac{1}{\sqrt{p}}\|\mathcal{F}(e_{i^\star}^\top)^{(k)}\big(\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\mathcal{U}^\top)^{(k)}\big)_\Omega\|_2\\
&\overset{\#\#\#}{=}\frac{\lambda}{2}\|\mathcal{F}(e_{i^\star}^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2^4\\
&-\frac{2}{\sqrt{p}}\sqrt{\sum_{k=1}^{n_3}\Big\|\mathcal{F}\big(e_{i^\star}^\top\big)^{(k)}(\mathcal{F}\big(\mathcal{M}^\star\big)^{(k)})_\Omega\Big\|_2^2}\times\frac{1}{\sqrt{p}}\sqrt{\sum_{k=1}^{n_3}\|\mathcal{F}\big(e_{i^\star}^\top\big)^{(k)}\big(\mathcal{F}\big(\mathcal{U}\big)^{(k)}\mathcal{F}\big(\mathcal{U}^\top\big)^{(k)}\big)_\Omega\|_2^2}\\
&\overset{\#\#\#\#}{\geq}\frac{\lambda}{2}\|\mathcal{F}(e_{i^\star}^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2^4\\
&-2\sum_{k=1}^{n_3}\sqrt{1+0.01}\|\mathcal{F}\big(e_{i^\star}^\top\big)^{(k)}\mathcal{F}\big(\mathcal{M}^\star\big)^{(k)}\|_2\cdot C\sqrt{n}\sum_{k=1}^{n_3}\|\mathcal{F}\big(\mathcal{U}\big)^{(k)}\mathcal{F}\big(\mathcal{U}^\top\big)^{(k)}\|_\infty\\
&\overset{\#\#\#\#\#}{\geq}\frac{\lambda}{2}\|\mathcal{F}(e_{i^\star}^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2^4 - C\sqrt{\mu r}\sigma_1^\star\sum_{k=1}^{n_3}\|\mathcal{F}\big(e_{i^\star}^\top\big)^{(k)}\mathcal{F}\big(\mathcal{U}\big)^{(k)}\|_2^2
\end{aligned}$$

where we used the relation

$$\begin{aligned}
&\langle\mathcal{F}\big(e_{i^\star}^\top\big)^{(k)}\big(\mathcal{F}\big(\mathcal{U}\big)^{(k)}\mathcal{F}\big(\mathcal{U}^\top\big)^{(k)}\big)_\Omega\mathcal{F}\big(\mathcal{U}\big)^{(k)}, \mathcal{F}\big(e_{i^\star}^\top\big)^{(k)}\mathcal{F}\big(\mathcal{U}\big)^{(k)}\rangle\\
&\qquad = \|\mathcal{F}\big(e_{i^\star}^\top\big)^{(k)}\big(\mathcal{F}\big(\mathcal{U}\big)^{(k)}\mathcal{F}\big(\mathcal{U}^\top\big)^{(k)}\big)_\Omega\|_2^2 \geq 0
\end{aligned}$$

in (#); the Cauchy-Schwartz inequality in (##); the isometry of the FFT in (###); (1) and (5) in (####) and the μ-incoherence of $\mathcal{M}^\star$ in (#####). Therefore, we obtain:

$$\|e_{i^\star}^\top * \mathcal{U}\|_F^3 \leq Cn_3\frac{\sqrt{\mu r}\ \sigma_1^\star}{\lambda}\|e_{i^\star}^\top * \mathcal{U}\|_F + \frac{2\epsilon}{\lambda}.$$

By choosing ϵ sufficiently small we can impose $\left(\epsilon/\lambda\right)^{\frac{2}{3}} \leq \dfrac{\sqrt{\mu r}\ \sigma_1^\star}{\lambda}$ and obtain

$$\max_{1\leq i\leq n} \|e_i^\top * \mathcal{U}\|_F^2 \leq c \max\left\{\alpha^2, n_3 \frac{\sqrt{\mu r}\ \sigma_1^\star}{\lambda}\right\}.$$

Finally, substituting our choice of α^2 and λ, the proof is completed. □

We now show that the Hessian operator satisfies that when $\mathcal{U}$ and $\mathcal{U}^\star$ are not close to each other, the terms involving the Hessian operator $\mathcal{H}$ in Equation (27) are significantly negative.

Lemma 12. *When the probability* $p \geq c_1\big(\frac{\mu^3 r^4 \left(\kappa^\star\right)^4 \log n}{n}\big)$, *by choosing* $\alpha^2 = \Theta\big(\frac{\mu r \sigma_1^\star}{n}\big)$ *and* $\lambda = \Theta\big(\frac{n}{\mu r \kappa^\star}\big)$ *with probability at least*

$$1 - 2nn_3 \exp\left\{-pn\big(\big(1+\frac{t}{pn}\big)\ln\big(1+\frac{t}{pn}\big) - \frac{t}{pn}\big)\right\},$$

for all $\mathcal{U}$ *with* $\|\nabla f(\mathcal{U})\|_F \leq \epsilon$ *for polynomially small* ϵ, *we have*

$$\begin{aligned}\Delta * \Delta^\top &: \mathcal{H} : \Delta * \Delta^\top - 3(\mathcal{M}-\mathcal{M}^\star) : \mathcal{H} : (\mathcal{M}-\mathcal{M}^\star)\\ &\leq -0.3 \sum_{k=1}^{n_3} \sigma_r^\star \|\mathcal{F}\big(\Delta\big)^{(k)}\|_F^2.\end{aligned}$$

Proof. Introduce

$$\Delta = \mathcal{U} - \mathcal{U}^\star.$$

Note that when Δ is not incoherent, the Hessian will still preserve norm for frontal faces like $\mathcal{F}(\Delta)^{(k)}\mathcal{F}(\mathcal{U}^\top)^{(k)}$, but but it will not necessarily preserve the norm of frontal faces such as $\mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^\top)^{(k)}$. Hence, we use different concentration lemmas in different cases.

First with the choice of α, λ and using Lemma 11 we know that with probability larger than

$$1 - 2n \exp\left\{-pn\big(\big(1+\frac{t}{pn}\big)\ln\big(1+\frac{t}{pn}\big) - \frac{t}{pn}\big)\right\},$$

the maximum the Euclidean norm of any row of $\mathcal{F}\big(\mathcal{U}\big)^{(k)}$ for $k = 1, \ldots, n_3$ is small as well:

$$\max_{1\le i\le n} \|\mathcal{F}\big(e_i^\top\big)^{(k)}\mathcal{F}\big(\mathcal{U}\big)^{(k)}\|_2^2 \le C\ \frac{(\mu r)^{1.5}\kappa^\star\sigma_1^\star}{n}.$$

Let us now split the analysis into two cases.

<u>Case 1:</u> $\|\mathcal{F}(\Delta)^{(k)}\|_F^2 \le \sigma_r^\star/4$, for any $k = 1, \ldots, n_3$.

In this case, Δ is small and $\Delta * \Delta^\top$ is small too but $\mathcal{H}$ not preserve norm very well for frontal slides $\mathcal{F}\big(\Delta\big)^{(k)}\mathcal{F}\big(\Delta^\top\big)^{(k)}$. Using the choice of p and by Lemma 1, we have

$$\frac{1}{p}\|\mathcal{U}^\star * \Delta^\top\|_\Omega^2 \ge \big(1-\delta\big)\|\mathcal{U}^\star * \Delta^\top\|_F^2 \ge \big(1-\delta\big)\sigma_r^\star\|\Delta\|_F^2.$$

On the other hand,

$$\frac{1}{p}\|\Delta * \Delta^\top\|_\Omega^2 = \sum_{k=1}^{n_3}\frac{1}{p}\|\mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\|_\Omega^2.$$

Using Lemma 3, for any $k = 1, \ldots, n_3$, we have for some positive constant C:

$$\begin{aligned}\frac{1}{p}\|\mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\|_\Omega^2 &\le \|\mathcal{F}(\Delta)^{(k)}\|_F^4 + C\sqrt{\frac{n}{p}}\cdot\frac{(\mu r)^{1.5}\kappa^\star\sigma_1^\star}{n}\|\mathcal{F}(\Delta)^{(k)}\|_F^2\\ &\le \|\mathcal{F}(\Delta)^{(k)}\|_F^4 + \frac{\sigma_r^\star}{4}\|\mathcal{F}(\Delta)^{(k)}\|_F^2\\ &\le \frac{\sigma_r^\star}{2}\|\mathcal{F}(\Delta)^{(k)}\|_F^2.\end{aligned}$$

Using these facts, we obtain

$$
\begin{aligned}
&\Delta * \Delta : \mathcal{H} : \Delta * \Delta^\top - 3\big(\mathcal{M} - \mathcal{M}^\star\big) : \mathcal{H} : \big(\mathcal{M} - \mathcal{M}^\star\big) \\
=&\sum_{k=1}^{n_3} \mathcal{F}(\Delta)^{(k)} \mathcal{F}(\Delta^\top)^{(k)} : \mathcal{F}(\mathcal{H})^{(k)} : \mathcal{F}(\Delta)^{(k)} \mathcal{F}(\Delta^\top)^{(k)} \\
&-3\big(\mathcal{F}(\mathcal{M})^{(k)} - \mathcal{F}(\mathcal{M}^\star)^{(k)}\big) : \mathcal{F}(\mathcal{H})^{(k)} : \big(\mathcal{F}(\mathcal{M})^{(k)} - \mathcal{F}(\mathcal{M}^\star)^{(k)}\big) \\
=&\sum_{k=1}^{n_3} \mathcal{F}(\Delta)^{(k)} \mathcal{F}(\Delta^\top)^{(k)} : \mathcal{F}(\mathcal{H})^{(k)} : \mathcal{F}(\Delta)^{(k)} \mathcal{F}(\Delta^\top)^{(k)} \\
&-3\big(\mathcal{F}(\mathcal{U}^\star)^{(k)} \mathcal{F}(\Delta^\top)^{(k)} + \mathcal{F}(\Delta)^{(k)} \mathcal{F}({\mathcal{U}^\star}^\top)^{(k)} + \mathcal{F}(\Delta)^{(k)} \mathcal{F}(\Delta^\top)^{(k)}\big) : \mathcal{F}(\mathcal{H})^{(k)} : \\
&\big(\mathcal{F}(\mathcal{U}^\star)^{(k)} \mathcal{F}(\Delta^\top)^{(k)} + \mathcal{F}(\Delta)^{(k)} \mathcal{F}({\mathcal{U}^\star}^\top)^{(k)} + \mathcal{F}(\Delta)^{(k)} \mathcal{F}(\Delta^\top)^{(k)}\big) \\
\leq&\sum_{k=1}^{n_3} -12\ \mathcal{F}(\mathcal{U}^\star)^{(k)} \mathcal{F}(\Delta^\top)^{(k)} : \mathcal{F}(\mathcal{H})^{(k)} : \mathcal{F}(\Delta)^{(k)} \mathcal{F}(\Delta^\top)^{(k)} \\
&-12\ \mathcal{F}(\mathcal{U}^\star)^{(k)} \mathcal{F}(\Delta^\top)^{(k)} : \mathcal{F}(\mathcal{H})^{(k)} : \mathcal{F}(\mathcal{U}^\star)^{(k)} \mathcal{F}(\Delta^\top)^{(k)} \\
\leq&\sum_{k=1}^{n_3} -\frac{12}{p}\big(\|\mathcal{F}(\mathcal{U}^\star)^{(k)} \mathcal{F}(\Delta^\top)^{(k)}\|_\Omega^2 - \|\mathcal{F}(\mathcal{U}^\star)^{(k)} \mathcal{F}(\Delta^\top)^{(k)}\|_\Omega \|\mathcal{F}(\Delta)^{(k)} \mathcal{F}(\Delta^\top)^{(k)}\|_\Omega\big) \\
\leq&\sum_{k=1}^{n_3} -12\sqrt{1-\delta}\big(\sqrt{1-\delta} - \sqrt{2/3}\big)\sigma_r^\star \|\mathcal{F}(\Delta)^{(k)}\|_F^2
\end{aligned}
$$

where we use the fact

$$
\frac{1}{p}\ \|\mathcal{F}(\Delta)^{(k)} \mathcal{F}(\Delta^\top)^{(k)}\|_\Omega^2 \leq \frac{\sigma_r^\star}{2}\ \|\mathcal{F}(\Delta)^{(k)}\|_F^2
$$

and

$$
-\frac{1}{p}\ \|\mathcal{F}(U^\star)^{(k)} {\mathcal{F}(\Delta)^{(k)}}^\top\|_\Omega^2 \leq -\big(1-\delta\big)\sigma_r^\star\ \|\mathcal{F}(\Delta)^{(k)}\|_F^2.
$$

Thus, taking $p \geq c_1 \frac{\mu^3 r^4 \left(\kappa^\star\right)^4 \log n}{n}$, we get

$$
\Delta * \Delta : \mathcal{H} : \Delta * \Delta^\top - 3\big(\mathcal{M} - \mathcal{M}^\star\big) : \mathcal{H} : \big(\mathcal{M} - \mathcal{M}^\star\big) \leq \sum_{k=1}^{n_3} -1.2\sigma_r^\star \|\mathcal{F}(\Delta)^{(k)}\|_F^2.
$$

<u>Case 2:</u> $\|\mathcal{F}(\Delta)^{(k)}\|_F^2 \geq \frac{\sigma_r^\star}{4}$, for any $k = 1, \ldots, n_3$.

Using Lemma (4) with high probability and with the choice of p that we have just made, we have

$$
\begin{aligned}
\frac{1}{p}\|\Delta * \Delta^\top\|_\Omega^2 &= \sum_{k=1}^{n_3} \frac{1}{p}\|\mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\|_\Omega^2 \\
&\le \sum_{k=1}^{n_3} \|\mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\|_F^2 \\
&+ C\Big(\frac{nr\log(n)}{p}\|\mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\|_\infty^2 \\
&+ \sqrt{\frac{nr\log(n)}{p}}\|\mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\|_F\|\mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\|_\infty\Big) \\
&\le \sum_{k=1}^{n_3} \|\mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\|_F^2 \\
&+ C\Big(\frac{nr\log(n)}{p}\cdot\frac{\big(\mu r\big)^3\big(\kappa^\star\sigma_1^\star\big)^2}{n^2} + \sqrt{\frac{nr\log(n)}{p}\cdot\frac{\big(\mu r\big)^3\big(\kappa^\star\sigma_1^\star\big)^2}{n^2}}\|\mathcal{F}(\Delta)^{(k)}\|_F^2\Big) \\
&\le \sum_{k=1}^{n_3} \|\mathcal{F}\big(\Delta\big)^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\|_F^2 + \frac{\big(\sigma_r^\star\big)^2}{80} + \frac{\sigma_r^\star}{20}\|\mathcal{F}(\Delta)^{(k)}\|_F^2 \\
&\le \sum_{k=1}^{n_3} \|\mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\|_F^2 + 0.1\sigma_r^\star\|\mathcal{F}(\Delta)^{(k)}\|_F^2.
\end{aligned}
$$

In the second inequality, we used

$$\|\mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\|_\infty^2 \le C\frac{\big(\mu r\big)^3\big(\kappa^\star\sigma_1^\star\big)^2}{n^2}$$

for some positive constant C. In the third inequality, we use for some positive constant c_1

$$p \ge c_1\frac{\mu^3 r^4\big(\kappa^\star\big)^4\log n}{n} \text{ and } \kappa = \sigma_1^\star/\sigma_r^\star.$$

Again, using Lemma 4, we have that, with high probability,

$$
\begin{aligned}
\frac{1}{p}\|\mathcal{M}-\mathcal{M}^\star\|_\Omega^2 &= \sum_{k=1}^{n_3}\frac{1}{p}\|\mathcal{F}(\mathcal{M})^{(k)}-\mathcal{F}(\mathcal{M}^\star)^{(k)}\|_\Omega^2 \\
&\geq \sum_{k=1}^{n_3}\|\mathcal{F}(\mathcal{M})^{(k)}-\mathcal{F}(\mathcal{M}^\star)^{(k)}\|_F^2 \\
&+ C\big(\frac{nr\log(n)}{p}\|\mathcal{F}(\mathcal{M})^{(k)}-\mathcal{F}(\mathcal{M}^\star)^{(k)}\|_\infty^2 \\
&+ \sqrt{\frac{nr\log(n)}{p}}\|\mathcal{F}(\mathcal{M})^{(k)}-\mathcal{F}(\mathcal{M}^\star)^{(k)}\|_F\|\mathcal{F}(\mathcal{M})^{(k)}-\mathcal{F}(\mathcal{M}^\star)^{(k)}\|_\infty\big) \\
&\geq \sum_{k=1}^{n_3}\|\mathcal{F}(M)^{(k)}-\mathcal{F}(\mathcal{M}^\star)^{(k)}\|_F^2 \\
&+ C\big(\frac{nr\log(n)}{p}\cdot\frac{\big(\mu r\big)^3\big(\kappa^\star\sigma_1^\star\big)^2}{n^2} \\
&+ \sqrt{\frac{nr\log(n)}{p}\cdot\frac{\big(\mu r\big)^3\big(\kappa^\star\sigma_1^\star\big)^2}{n^2}}\|\mathcal{F}(\mathcal{M})^{(k)}-\mathcal{F}(\mathcal{M}^\star)^{(k)}\|_F\big) \\
&\geq \sum_{k=1}^{n_3}\|\mathcal{F}(\mathcal{M})^{(k)}-\mathcal{F}(\mathcal{M}^\star)^{(k)}\|_F^2 - \frac{\big(\sigma_r^\star\big)^2}{80} - \frac{\sigma_r^\star}{20}\|\mathcal{F}(\mathcal{M})^{(k)}-\mathcal{F}(\mathcal{M}^\star)^{(k)}\|_F \\
&\geq \sum_{k=1}^{n_3} 0.95\ \|\mathcal{F}(\mathcal{M})^{(k)}-\mathcal{F}(\mathcal{M}^\star)^{(k)}\|_F^2 - 0.1\ \sigma_r^\star\|\mathcal{F}(\Delta)^{(k)}\|_F^2.
\end{aligned}
$$

The third inequality, again we use $p \geq c_1\big(\frac{\mu^3 r^4\big(\kappa^\star\big)^4\log n}{n}\big)$ and $\kappa = \sigma_1^\star/\sigma_r^\star$. This facts implies

$$
\begin{aligned}
&\Delta * \Delta^\top : \mathcal{H} : \Delta * \Delta^\top - 3(\mathcal{M}-\mathcal{M}^\star) : \mathcal{H} : (\mathcal{M}-\mathcal{M}^\star) \\
&= \sum_{k=1}^{n_3} \frac{1}{p}\big(\|\mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\|_\Omega^2 - 3\ \|\mathcal{F}(\mathcal{M})^{(k)} - \mathcal{F}(\mathcal{M}^\star)^{(k)}\|_\Omega^2\big) \\
&\leq \sum_{k=1}^{n_3} \|\mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\|_F^2 + 0.1\sigma_r^\star\|\mathcal{F}(\Delta)^{(k)}\|_F^2 \\
&- 3\left(0.95\|\mathcal{F}(\mathcal{M})^{(k)} - \mathcal{F}(\mathcal{M}^\star)^{(k)}\|_F^2 - 0.1\,\sigma_r^\star\|\mathcal{F}(\Delta)^{(k)}\|_F^2\right) \\
&\leq \sum_{k=1}^{n_3} -0.85\|\mathcal{F}(\mathcal{M})^{(k)} - \mathcal{F}(\mathcal{M}^\star)^{(k)}\|_F^2 + 0.4\,\sigma_r^\star\|\mathcal{F}(\Delta)^{(k)}\|_F^2 \\
&\leq \sum_{k=1}^{n_3} -0.3\,\sigma_r^\star\|\mathcal{F}(\Delta)^{(k)}\|_F^2.
\end{aligned}
$$

The two last inequalities, we use the two bounds of 7. ☐

Now, we need to bound the terms with the regularizer in (27).

Lemma 13. *By choosing* $\alpha^2 = \Theta\big(\frac{\mu r \sigma_1^\star}{n}\big)$ *and* $\lambda\alpha^2 \leq C\ \sigma_r^\star$ *for some positive constant* C*, we have:*

$$
\frac{1}{4}\left[\Delta : \nabla^2 Q\big(\mathcal{U}\big) : \Delta - 4\langle \nabla Q\big(\mathcal{U}\big), \Delta\rangle\right] \leq 0.1\sigma_r^\star \sum_{k=1}^{n_3} \|\mathcal{F}\big(\Delta\big)^{(k)}\|_F^2.
$$

Proof. We know that:

$$
\langle \nabla Q(\mathcal{U}), \mathcal{Z}\rangle = 4\lambda \sum_{k=1}^{n_3}\sum_{i=1}^{n} \big(\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 - \alpha\big)_+^3 \frac{\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\mathcal{Z}^\top)^{(k)}\mathcal{F}(e_i)^{(k)}}{\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2}
$$

$$
\begin{aligned}
\mathcal{Z} : \nabla^2 Q(\mathcal{U}) : \mathcal{Z} &= 12\lambda \sum_{k=1}^{n_3}\sum_{i=1}^{n} \big(\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 - \alpha\big)_+^2 \Big(\frac{\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\mathcal{Z}^\top)^{(k)}}{\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2}\Big)^2 \\
&+ 4\lambda \sum_{k=1}^{n_3}\sum_{i=1}^{n} \big(\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 - \alpha\big)_+^3 \\
&\times \frac{\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2^2\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{Z})^{(k)}\|_2^2 - \big(\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\mathcal{Z}^\top)^{(k)}\mathcal{F}(e_i)^{(k)}\big)^2}{\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2^3}.
\end{aligned}
$$

Using this facts with $\mathcal{Z} = \Delta = \mathcal{U} - \mathcal{U}^\star * \mathcal{R}$, we have:

$$
\begin{aligned}
&\frac{1}{4}\left[\Delta : \nabla^2 Q(\mathcal{U}) : \Delta - 4\langle \nabla Q(\mathcal{U}), \Delta\rangle\right] \\
&= \underbrace{\lambda \sum_{k=1}^{n_3}\sum_{i=1}^{n} \begin{matrix} \left(\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 - \alpha\right)_+^3 \\ \times \\ \frac{\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2^2\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\Delta)^{(k)}\|_2^2 - \left(\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\mathcal{F}(e_i)^{(k)}\right)^2}{\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2^3} \end{matrix}}_{=A_1} \\
&+ \underbrace{3\lambda \sum_{k=1}^{n_3}\sum_{i=1}^{n} \left(\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 - \alpha\right)_+^2 \left(\frac{\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\mathcal{F}(e_i)^{(k)}}{\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2}\right)^2}_{=A_2} \\
&- \underbrace{4\lambda \sum_{k=1}^{n_3}\sum_{i=1}^{n} \left(\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 - \alpha\right)_+^3 \frac{\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\mathcal{F}(e_i)^{(k)}}{\|\mathcal{F}(e_i)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2}}_{=A_3}.
\end{aligned}
$$

Furthermore, using the incoherence property of $\mathcal{M}^\star$, we have for any $k = 1, \ldots, n_3$

$$
\begin{aligned}
\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)} - \mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\Delta)^{(k)}\|_2 =& \|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U}^\star)^{(k)}\mathcal{F}(\mathcal{R})^{(k)}\|_2 \\
=& \|\mathcal{F}(e_i^\top)\mathcal{F}(\mathcal{U}^\star)^{(k)}\|_2 \\
\leq& \sqrt{\frac{\mu r \sigma_1^\star}{n}}.
\end{aligned}
$$

By choosing $\alpha > C_1\sqrt{\frac{\mu r \sigma_1^\star}{n}}$ for some large constant C_1 and when $\|\mathcal{F}(e_i^\top)\mathcal{F}(\mathcal{U})^{(k)}\|_2 - \alpha > 0$, we have for any $k = 1, \ldots, n_3$

$$
\begin{aligned}
&\mathcal{F}\left(e_i^\top\right)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\mathcal{F}(e_i)^{(k)} \\
=& \mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\mathcal{U}^\top)^{(k)}\mathcal{F}(e_i)^{(k)} - \mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\mathcal{R}^\top)^{(k)}\mathcal{F}({\mathcal{U}^\star}^\top)^{(k)}\mathcal{F}(e_i)^{(k)} \\
\geq& \|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2^2 - \|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U}^\star)^{(k)}\|_2 \\
\geq& \left(1 - \frac{1}{C_1}\right)\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2^2.
\end{aligned}
$$

The last inequality, we use the fact

$$
\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U}^\star)^{(k)}\|_2 < \frac{\alpha}{C_1}
$$

and

$$\alpha < \|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 \ \Rightarrow \|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U}^\star)^{(k)}\|_2 < \frac{1}{C_1}\ \|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2.$$

Further, we have

$$\begin{aligned}
&\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\Delta)^{(k)}\|_2\\
\le&\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2\big(\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2+\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U}^\star)^{(k)}\|_2\big)\\
\le&\big(1+\frac{1}{C_1}\big)\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2^2.
\end{aligned}$$

Now, we need to bound the summation $A_1 + A_2 + A_3$ as follows to get a bound $A_1 + 0.1A_3$ and $A_2 + 0.9A_3$. Thus,

$$\begin{aligned}
A_1+0.1A_3 =& \lambda\sum_{k=1}^{n_3}\sum_{i=1}^{n}\Bigg(\frac{\begin{array}{c}\big(\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2-\alpha\big)^3\\ \times\\ \|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2^2\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\Delta)^{(k)}\|_2^2\\ -\big(\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\mathcal{F}(e_i)^{(k)}\big)^2\end{array}}{\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2^3}\Bigg)\\
&-0.4\lambda\sum_{k=1}^{n_3}\sum_{i=1}^{n}\big(\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2-\alpha\big)^3\frac{\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\mathcal{F}(e_i)^{(k)}}{\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|}\\
&\le \lambda\sum_{k=1}^{n_3}\sum_{i=1}^{n}\big(\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2-\alpha\big)^3\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2\\
&\times\left[\big(1+\frac{1}{C_1}\big)^2-\big(1-\frac{1}{C_1}\big)^2-0.4\big(1-\frac{1}{C_1}\big)\right]\\
&< 0.
\end{aligned}$$

Moreover,

$$\begin{aligned}
&A_2+0.9A_3\\
&=3\lambda\sum_{k=1}^{n_3}\sum_{i=1}^{n}\big(\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2-\alpha\big)^2\big(\frac{\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\mathcal{F}(e_i)^{(k)}}{\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2}\big)^2\\
&-3.6\lambda\sum_{k=1}^{n_3}\sum_{i=1}^{n}\big(\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2-\alpha\big)^3\frac{\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\mathcal{F}(e_i)^{(k)}}{\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2}
\end{aligned}$$

Denote i-the summand of the frontal faces of $A_2 + 0.9A_3$ as $A_2 + 0.9A_3 = \sum_{k=1}^{n_3} \sum_{i=1}^{n} B_i^k$, with $B_i = A_2^{(i)} + 0.9\, A_3^{(i)}$.

<u>Case 1:</u> for i such that $\|\mathcal{F}\big(e_i^\top\big)^{(k)}\mathcal{F}\big(U\big)^{(k)}\|_2 \geq 9\alpha$ and $C_1 \geq 100$, we have:

$$
\begin{aligned}
B_i^k &= 3\lambda\big(\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 - \alpha\big)^2 \frac{\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\mathcal{F}(e_i)^{(k)}}{\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2} \\
&\quad \times \left[\frac{\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\mathcal{F}(e_i)^{(k)}}{\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2} - 1.2\big(\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 - \alpha\big)\right] \\
&\quad \leq 3\lambda\big(\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 - \alpha\big)^2 \frac{\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\mathcal{F}(e_i)^{(k)}}{\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2} \\
&\quad \times \left[\big(1 + \frac{1}{C_1}\big)\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 - 1.2\big(\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 - \alpha\big)\right] \\
&\quad \leq 0.
\end{aligned}
$$

Because:

- $\big(\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 - \alpha\big)^2 > 0$
- $\dfrac{\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\mathcal{F}(e_i)^{(k)}}{\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2} \leq \big(1 + \dfrac{1}{C_1}\big)\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 \geq 0$
- $\left[\big(1 + \dfrac{1}{C_1}\big)\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 - 1.2\big(\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 - \alpha\big)\right] \leq 0$

<u>Case 2:</u> for i such that $\alpha < \|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 < 9\alpha$, we call this set

$I = \{i|\, \alpha < \|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}\big(\mathcal{U}\big)^{(k)}\|_2 < 9\alpha\}$ and we have for each frontal face:

$$\begin{aligned}
\sum_{i\in I} B_i^{(k)} &\leq 3\lambda \sum_{i\in I} \big(\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 - \alpha\big)_+^2 \frac{\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\mathcal{F}(e_i)^{(k)}}{\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2} \\
&\times \left[\big(1+\frac{1}{C_1}\big)\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 - 1.2\big(\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 - \alpha\big)_+\right] \\
&\leq 3\lambda \sum_{i\in I} \big(\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 - \alpha\big)_+^2 \frac{\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\mathcal{F}(e_i)^{(k)}}{\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2} \\
&\times \big(1+\frac{1}{C_1}\big)\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 \\
&\leq 3\lambda \sum_{i\in I} \big(\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 - \alpha\big)_+^2 \big(1+\frac{1}{C_1}\big)^2 \|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(U)^{(k)}\|_2^2 \\
&\leq 3\lambda\ |I|\ 64\ \alpha^2 . \big(1+\frac{1}{C_1}\big)^2\ 81\ \alpha^2 \\
&\leq 3\ 10^4\ |I|\ \lambda\ \alpha^4.
\end{aligned}$$

In sum, we obtain:

$$\frac{1}{4}\big[\Delta : \nabla^2 Q\big(\mathcal{U}\big) : \Delta - 4\langle \nabla Q\big(\mathcal{U}\big), \Delta\rangle\big] \leq c_2\, \lambda\, |I|\alpha^4$$

for some large constant c_2. Finally, remains to determine with the property of the set I on each front face:

$$\begin{aligned}
\sigma_r^\star \|\mathcal{F}(\Delta)^{(k)}\|_F^2 =& \sigma_r^\star \|\mathcal{F}\big(\mathcal{U}\big)^{(k)} - \mathcal{F}\big(\mathcal{U}^\star\big)^{(k)}\mathcal{F}\big(\mathcal{R}\big)^{(k)}\|_F^2 \\
&= \sigma_r^\star\big(\|\mathcal{F}\big(\mathcal{U}\big)^{(k)}\|_F^2 + \|\mathcal{F}\big(\mathcal{U}^\star\big)^{(k)}\mathcal{F}\big(\mathcal{R}\big)^{(k)}\|_F^2 \\
&\quad - 2\langle \mathcal{F}\big(\mathcal{U}\big)^{(k)}, \mathcal{F}\big(\mathcal{U}^\star\big)^{(k)}\mathcal{F}\big(\mathcal{R}\big)^{(k)}\rangle\big) \\
&\geq \sigma_r^\star \|\mathcal{F}\big(\mathcal{U}\big)^{(k)}\|_F^2 \\
&= \sigma_r^\star \sum_{i\in I} \|\mathcal{F}\big(e_i^\top\big)^{(k)}\mathcal{F}\big(\mathcal{U}\big)^{(k)}\|_2^2 \\
&\geq \sigma_r^\star \alpha^2 |I|.
\end{aligned}$$

Therefore, as long as $\lambda\alpha^2 \leq \sigma_r^\star / C_2$ for some large constant C_2 (which is satisfied by our choice of λ) we obtain:

$$\frac{1}{4}\left[\Delta : \nabla^2 Q\big(\mathcal{U}\big) : \Delta - 4\langle \nabla Q\big(U\big), \Delta\rangle\right] \leq 0.1 \sum_{k=1}^{n_3} \sigma_r^\star \|\mathcal{F}(\Delta)^{(k)}\|_F^2.$$

$\square$

Combing these lemmas, we are now ready to prove the main theorem for symmetric tensor completion.

Theorem 7.1. *Take the sample rate p such that*

$$p \geq c_1 \frac{\mu^3 r^4 \big(\kappa^\star\big)^4 \log(n)}{n},$$

for some positive constant and choose

$$\alpha^2 = \Theta\big(\frac{\mu r \sigma_1^\star}{n}\big) \text{ and } \lambda = \Theta\big(\frac{n}{\mu r \kappa^\star}\big).$$

Then with probability at least

$$1 - 2nn_3 \exp\left\{-pn\big(\big(1+\frac{t}{pn}\big)\ln\big(1+\frac{t}{pn}\big) - \frac{t}{pn}\big)\right\},$$

we have

- *All local minima of* (31) *satisfy $\mathcal{U} * \mathcal{U}^\top = \mathcal{M}^\star$;*
- *the function is $\big(\epsilon, \Omega\big(\sigma_r^\star\big), O\big(\frac{\epsilon}{\sigma_r^\star}\big)\big)$-strict saddle for polynomially small ϵ.*

Proof. We know by 12:

$$\Delta * \Delta^\top : \mathcal{H} : \Delta * \Delta^\top - 3(\mathcal{M} - \mathcal{M}^\star) : \mathcal{H} : (\mathcal{M} - \mathcal{M}^\star) \leq -0.3 \sum_{k=1}^{n_3} \sigma_r^\star \|\mathcal{F}(\Delta)^{(k)}\|_F^2.$$

Further, by 13, we have:

$$\frac{1}{4}\left[\Delta : \nabla^2 Q\big(\mathcal{U}\big) : \Delta - 4\langle \nabla Q\big(\mathcal{U}\big), \Delta\rangle\right] \leq 0.1 \sigma_r^\star \sum_{k=1}^{n_3} \|\mathcal{F}(\Delta)^{(k)}\|_F^2.$$

Using this facts, for any $\mathcal{U}$ with small gradient satisfying $\|\nabla f(\mathcal{U})\|_F \le \epsilon$, we have

$$\Delta : \nabla^2 f(\mathcal{U}) : \Delta \le -0.2 \sum_{k=1}^{n_3} \sigma_r^\star \|\mathcal{F}(\Delta)^{(k)}\|_F^2 + 4\,\epsilon\|\Delta\|_F.$$

That is, if $\mathcal{U}$ is not close to $\mathcal{U}^\star$, that is, $\|\Delta\|_F \ge \dfrac{40\epsilon}{\sigma_r^\star}$, we have

$$\begin{aligned}
\Delta : \nabla^2 f(\mathcal{U}) : \Delta &\le -0.2 \sum_{k=1}^{n_3} \sigma_r^\star \|\mathcal{F}(\Delta)^{(k)}\|_F^2 + 0.1\sigma_r^\star\|\Delta\|_F \\
&\le -0.2 \sum_{k=1}^{n_3} \sigma_r^\star \|\mathcal{F}(\Delta)^{(k)}\|_F^2 + 0.1\sigma_r^\star \sqrt{\sum_{k=1}^{n_3} \|\mathcal{F}(\Delta)^{(k)}\|_F^2} \\
&\le -0.1 \sum_{k=1}^{n_3} \sigma_r^\star \|\mathcal{F}(\Delta)^{(k)}\|_F^2.
\end{aligned}$$

This proves $\left(\epsilon, 0.1\sigma_r^\star, \frac{40\epsilon}{\sigma_r^\star}\right)$-strict saddle property. By taking $\epsilon = 0$, then all stationary points with $\|\Delta\|_F \neq 0$ are saddle points. This means all local minima are global minima (satisfying $\mathcal{U} * \mathcal{U}^\top = \mathcal{M}^\star$), which finishes the proof. □

7.3. Proof of Theorem 4.1

The proof is split into two steps.

7.3.1. Study of the Hessian

Furthermore, we have $Q(\mathcal{W}) = Q_1(\mathcal{U}) + Q_2(\mathcal{V})$.

Lemma 14. *Let $\Delta, \mathcal{N}, \mathcal{N}^\star$ be defined as in Definition **??**. Then, for any $\mathcal{W} \in \mathbb{R}^{(n_1+n_2)\times r\times n_3}$, the Hessian of the objective (18) satisfies:*

$$\begin{aligned}
\Delta : \nabla^2 f(\mathcal{W}) : \Delta \le{}& \Delta * \Delta^\top : \mathcal{H} : \Delta * \Delta^\top - 3(\mathcal{N} - \mathcal{N}^\star) : \mathcal{H} : (\mathcal{N} - \mathcal{N}^\star) \\
&+ 4\langle \nabla f(\mathcal{W}), \Delta\rangle + [\Delta : \nabla^2 Q(\mathcal{W}) : \Delta - 4\langle \nabla Q(\mathcal{W}), \Delta\rangle]
\end{aligned} \tag{36}$$

where

$$\mathcal{H} = 4\mathcal{H}_1 + \mathcal{G}.$$

Further, if $\mathcal{H}_0$ satisfies

$$\mathcal{M} : \mathcal{H}_0 : \mathcal{M} \in \left(1 \pm \delta\right) \|\mathcal{M}\|_F^2$$

for some tensor $\mathcal{M} = \mathcal{U} * \mathcal{V}^\top$, let $\mathcal{W}$ and $\mathcal{N}$ be defined as in (17), then

$$\mathcal{N} : \mathcal{H} : \mathcal{N} \in \left(1 \pm 2\delta\right) \|\mathcal{N}\|_F^2.$$

Proof. We know that the objective function with $\mathcal{N} = \mathcal{W} * \mathcal{W}^\top$ is:

$$\begin{aligned} f(\mathcal{W}) &= \frac{1}{2}\left[\left(\mathcal{N} - \mathcal{N}^\star\right) : 4\mathcal{H}_1 : \left(\mathcal{N} - \mathcal{N}^\star\right) + \mathcal{N} : \mathcal{G} : \mathcal{N}\right] + Q(\mathcal{W}) \\ &= \frac{1}{2} \sum_{k=1}^{n_3} \left\{ \begin{array}{c} \left(\mathcal{F}(\mathcal{W})^{(k)} \mathcal{F}(\mathcal{W}^\top)^{(k)} - \mathcal{F}(\mathcal{N}^\star)^{(k)}\right) \\ : 4\mathcal{F}(\mathcal{H}_1)^{(k)} : \\ \left(\mathcal{F}(\mathcal{W})^{(k)} \mathcal{F}(\mathcal{W}^\top)^{(k)} - \mathcal{F}(\mathcal{N}^\star)^{(k)}\right) \end{array} \right\} \\ &+ \mathcal{F}(\mathcal{W})^{(k)} \mathcal{F}(\mathcal{W}^\top)^{(k)} : \mathcal{F}(\mathcal{G})^{(k)} : \mathcal{F}(\mathcal{W})^{(k)} \mathcal{F}(\mathcal{W}^\top)^{(k)} + Q(\mathcal{W}). \end{aligned}$$

Determine the gradient and the Hessian of f.

$$f(\mathcal{W}) = G(\mathcal{W}) + Q(\mathcal{W})$$

with

$$\begin{aligned} G(\mathcal{W}) = \frac{1}{2} \sum_{k=1}^{n_3} &\left\{ \begin{array}{c} \left(\mathcal{F}(\mathcal{W})^{(k)} \mathcal{F}(\mathcal{W}^\top)^{(k)} - \mathcal{F}(\mathcal{N}^\star)^{(k)}\right) \\ : 4\mathcal{F}(\mathcal{H}_1)^{(k)} : \\ \left(\mathcal{F}(\mathcal{W})^{(k)} \mathcal{F}(\mathcal{W}^\top)^{(k)} - \mathcal{F}(\mathcal{N}^\star)^{(k)}\right) \end{array} \right\} \\ &+ \mathcal{F}(\mathcal{W})^{(k)} \mathcal{F}(\mathcal{W}^\top)^{(k)} : \mathcal{F}(\mathcal{G})^{(k)} : \mathcal{F}(W)^{(k)} \mathcal{F}(\mathcal{W}^\top)^{(k)}. \end{aligned}$$

Using that, for any $\mathcal{Z} \in \mathbb{R}^{(n_1+n_2) \times r \times n_3}$

$$\begin{aligned} \langle \nabla f(\mathcal{W}), \mathcal{Z} \rangle &= \langle \nabla G(\mathcal{W}), \mathcal{Z} \rangle + \langle \nabla Q(\mathcal{W}), \mathcal{Z} \rangle \\ \mathcal{Z} : \nabla^2 f(\mathcal{W}) : \mathcal{Z} &= \mathcal{Z} : \nabla^2 G(\mathcal{W}) : \mathcal{Z} + \mathcal{Z} : \nabla^2 Q(W) : \mathcal{Z}. \end{aligned}$$

We now need to compute $\langle \nabla G(\mathcal{W}), \mathcal{Z} \rangle$ and $\mathcal{Z} : \nabla^2 G(\mathcal{W}) : \mathcal{Z}$. For this purpose, using the fact, for any $\mathcal{Z} \in \mathbb{R}^{\left(n_1+n_2\right) \times r \times n_3}$, we obtain that:

$$G(\mathcal{W} + \mathcal{Z}) = G(\mathcal{W}) + \langle \nabla G(\mathcal{W}), \mathcal{Z} \rangle + \frac{1}{2} \mathcal{Z} : \nabla^2 G(\mathcal{W}) : \mathcal{Z} + O(\|\mathcal{Z} * \mathcal{Z}^\top\|^2).$$

we obtain:

$$\langle \nabla G(\mathcal{W}), \mathcal{Z} \rangle = \sum_{k=1}^{n_3} \left\{ \begin{array}{c} \big(\mathcal{F}(\mathcal{W})^{(k)}\mathcal{F}(\mathcal{W}^\top)^{(k)} - \mathcal{F}(\mathcal{N}^\star)^{(k)}\big) \\ : 4\,\mathcal{F}(\mathcal{H}_1)^{(k)} : \\ \big(\mathcal{F}(\mathcal{W})^{(k)}\mathcal{F}(\mathcal{Z}^\top)^{(k)} + \mathcal{F}(\mathcal{Z})^{(k)}\mathcal{F}(\mathcal{W}^\top)^{(k)}\big) \end{array} \right\}$$

$$+ \sum_{k=1}^{n_3} \mathcal{F}\big(\mathcal{W}\big)^{(k)}\mathcal{F}\big(\mathcal{W}^\top\big)^{(k)} : \mathcal{F}\big(\mathcal{G}\big)^{(k)} : \big(\mathcal{F}(\mathcal{W})^{(k)}\mathcal{F}(\mathcal{Z}^\top)^{(k)} + \mathcal{F}(\mathcal{Z})^{(k)}\mathcal{F}(\mathcal{W}^\top)^{(k)}\big)$$

$$= \big(\mathcal{N} - \mathcal{N}^\star\big) : 4\,\mathcal{H}_1 : \big(\mathcal{W} * \mathcal{Z}^\top + \mathcal{Z} * \mathcal{W}^\top\big) + \mathcal{N} : \mathcal{G} : \big(\mathcal{W} * \mathcal{Z}^\top + \mathcal{Z} * \mathcal{W}^\top\big)$$

and

$$\mathcal{Z} : \nabla^2 G(\mathcal{W}) : \mathcal{Z} = \sum_{k=1}^{n_3} \left\{ \begin{array}{c} \big(\mathcal{F}(\mathcal{W})^{(k)}\mathcal{F}(\mathcal{Z}^\top)^{(k)} + \mathcal{F}(\mathcal{Z})^{(k)}\mathcal{F}(\mathcal{W}^\top)^{(k)}\big) \\ : \big(4\,\mathcal{F}(\mathcal{H}_1)^{(k)} + \mathcal{F}\big(\mathcal{G}\big)^{(k)}\big) : \\ \big(\mathcal{F}(\mathcal{W})^{(k)}\mathcal{F}(\mathcal{Z}^\top)^{(k)} + \mathcal{F}(\mathcal{Z})^{(k)}\mathcal{F}(\mathcal{W}^\top)^{(k)}\big) \end{array} \right\}$$

$$+ \sum_{k=1}^{n_3} \left\{ \begin{array}{c} 2\big(\mathcal{F}(\mathcal{N})^{(k)} - \mathcal{F}(\mathcal{N}^\star)^{(k)}\big) \\ : 4\,\mathcal{F}(\mathcal{H}_1)^{(k)} : \\ \mathcal{F}(\mathcal{Z})^{(k)}\mathcal{F}(\mathcal{Z}^\top)^{(k)} \end{array} \right\}$$

$$+ \sum_{k=1}^{n_3} 2\ \mathcal{F}(\mathcal{N})^{(k)} : \mathcal{F}(\mathcal{G})^{(k)} : \mathcal{F}(\mathcal{Z})^{(k)}\mathcal{F}(\mathcal{Z}^\top)^{(k)}$$

$$= \big(\mathcal{W} * \mathcal{Z}^\top + \mathcal{Z} * \mathcal{W}^\top\big) : \big(4\,\mathcal{H}_1 + \mathcal{G}\big) : \big(\mathcal{W} * \mathcal{Z}^\top + \mathcal{Z} * \mathcal{W}^\top\big)$$

$$+ 2\ \big(\mathcal{N} - \mathcal{N}^\star\big) : 4\mathcal{H}_1 : \mathcal{Z} * \mathcal{Z}^\top + 2\mathcal{N} : \mathcal{G} : \mathcal{Z} * \mathcal{Z}^\top.$$

Let $\mathcal{Z} = \Delta = \mathcal{W} - \mathcal{W}^\star * \mathcal{R}$ and $\mathcal{H} = 4\mathcal{H}_1 + \mathcal{G}$. By noting that: $\mathcal{N} - \mathcal{N}^\star + \Delta * \Delta^\top = \mathcal{W} * \Delta^\top + \Delta * \mathcal{W}^\top$, we have

$$\begin{aligned}
\langle \nabla f(\mathcal{W}), \Delta \rangle &= \big(\mathcal{N} - \mathcal{N}^\star\big) : 4\mathcal{H}_1 : \big(\mathcal{W} * \Delta^\top + \Delta * \mathcal{W}^\top\big) + \mathcal{N} : \mathcal{G} : \big(\mathcal{W} * \Delta^\top + \Delta * \mathcal{W}^\top\big) \\
&\quad + \langle \nabla Q(\mathcal{W}), \Delta \rangle \\
&= \big(\mathcal{N} - \mathcal{N}^\star\big) : \mathcal{H} : \big(\mathcal{W} * \Delta^\top + \Delta * \mathcal{W}^\top\big) - \big(\mathcal{N} - \mathcal{N}^\star\big) : \mathcal{G} : \big(\mathcal{W} * \Delta^\top + \Delta * \mathcal{W}^\top\big) \\
&\quad + \mathcal{N} : \mathcal{G} : \big(\mathcal{W} * \Delta^\top + \Delta * \mathcal{W}^\top\big) + \langle \nabla Q(\mathcal{W}), \Delta \rangle \\
&= \big(\mathcal{N} - \mathcal{N}^\star\big) : \mathcal{H} : \big(\mathcal{W} * \Delta^\top + \Delta * \mathcal{W}^\top\big) + \mathcal{N}^\star : \mathcal{G} : \big(\mathcal{N} - \mathcal{N}^\star + \Delta * \Delta^\top\big) \\
&\quad + \langle \nabla Q(\mathcal{W}), \Delta \rangle \\
&= \big(\mathcal{N} - \mathcal{N}^\star\big) : \mathcal{H} : \big(\mathcal{W} * \Delta^\top + \Delta * \mathcal{W}^\top\big) + 2\,\mathcal{N}^\star : \mathcal{G} : \mathcal{N} + \langle \nabla Q(\mathcal{W}), \Delta \rangle.
\end{aligned}$$

The last equality, we expand $\mathcal{N} - \mathcal{N}^\star + \Delta * \Delta^\top$ and we use that

$$\mathcal{N}^\star : \mathcal{G} : \mathcal{N}^\star = \mathcal{N}^\star : \mathcal{G} : \mathcal{W}^\star * \mathcal{W}^\top = 0$$

due to fact

$$\mathcal{U}^{\star^\top} * \mathcal{U}^\star = \mathcal{V}^{\star^\top} * \mathcal{V}^\star.$$

Now, the Hessian along the direction Δ is:

$$\begin{aligned}\Delta : \nabla^2 f(\mathcal{W}) : \Delta = &\big(\mathcal{W} * \Delta^\top + \Delta * \mathcal{W}^\top\big) : \mathcal{H} : \big(\mathcal{W} * \Delta^\top + \Delta * \mathcal{W}^\top\big)\\ &+ 2\big(\mathcal{N} - \mathcal{N}^\star\big) : 4\mathcal{H}_1 : \Delta * \Delta^\top\\ &+ 2\mathcal{N} : \mathcal{G} : \Delta * \Delta^\top + \Delta : \nabla^2 Q(\mathcal{W}) : \Delta.\end{aligned} \tag{37}$$

We are interested in the first term of (37) with

$$\mathcal{N} - \mathcal{N}^\star + \Delta * \Delta^\top = \mathcal{W} * \Delta^\top + \Delta * \mathcal{W}^\top,$$

we have:

$$\begin{aligned}&\big(\mathcal{W} * \Delta^\top + \Delta * \mathcal{W}^\top\big) : \mathcal{H} : \big(\mathcal{W} * \Delta^\top + \Delta * \mathcal{W}^\top\big)\\ &= \sum_{k=1}^{n_3}\left\{\begin{array}{c}\big(\mathcal{F}(\mathcal{W})^{(k)}\mathcal{F}(\Delta^\top)^{(k)} + \mathcal{F}(\Delta)^{(k)}\mathcal{F}(\mathcal{W}^\top)^{(k)}\big)\\ : \mathcal{F}(\mathcal{H})^{(k)} :\\ \big(\mathcal{F}(\mathcal{W})^{(k)}\mathcal{F}(\Delta^\top)^{(k)} + \mathcal{F}(\Delta)^{(k)}\mathcal{F}(\mathcal{W}^\top)^{(k)}\big)\end{array}\right\}\\ &= \sum_{k=1}^{n_3}\mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^\top)^{(k)} : \mathcal{F}(\mathcal{H})^{(k)} : \mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\\ &\quad + 2\big(\mathcal{F}(\mathcal{N})^{(k)} - \mathcal{F}(\mathcal{N}^\star)^{(k)}\big) : \mathcal{F}(\mathcal{H})^{(k)} : \big(\mathcal{F}(\mathcal{W})^{(k)}\mathcal{F}(\Delta^\top)^{(k)} + \mathcal{F}(\Delta)^{(k)}\mathcal{F}(\mathcal{W}^\top)^{(k)}\big)\\ &\quad - \big(\mathcal{F}(\mathcal{N})^{(k)} - \mathcal{F}(\mathcal{N}^\star)^{(k)}\big) : \mathcal{F}(\mathcal{H})^{(k)} : \big(\mathcal{F}(\mathcal{N})^{(k)} - \mathcal{F}(\mathcal{N}^\star)^{(k)}\big)\\ &= \sum_{k=1}^{n_3}\mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^\top)^{(k)} : \mathcal{F}(\mathcal{H})^{(k)} : \mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\\ &- \big(\mathcal{F}(\mathcal{N})^{(k)} - \mathcal{F}(\mathcal{N}^\star)^{(k)}\big) : \mathcal{F}(\mathcal{H})^{(k)} : \big(\mathcal{F}(\mathcal{N})^{(k)} - \mathcal{F}(\mathcal{N}^\star)^{(k)}\big)\\ &\quad - 4\mathcal{F}(\mathcal{N}^\star)^{(k)} : \mathcal{F}(\mathcal{G})^{(k)} : \mathcal{F}(\mathcal{N})^{(k)} + 2\langle\nabla f(\mathcal{W}), \Delta\rangle - 2\langle\nabla Q(\mathcal{W}), \Delta\rangle.\end{aligned}$$

For the sum of second and third terms of (37), we have:

$$
\begin{aligned}
&2\big(\mathcal{N}-\mathcal{N}^{\star}\big) : 4\mathcal{H}_1 : \Delta * \Delta^{\top} + 2\mathcal{N} : \mathcal{G} : \Delta * \Delta^{\top} \\
&= \sum_{k=1}^{n_3} 2\big(\mathcal{F}(\mathcal{N})^{(k)} - \mathcal{F}(\mathcal{N}^{\star})^{(k)}\big) : 4\,\mathcal{F}(\mathcal{H}_1)^{(k)} : \mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^{\top})^{(k)} \\
&+ 2\,\mathcal{F}(\mathcal{N})^{(k)} : \mathcal{F}(\mathcal{G})^{(k)} : \mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^{\top})^{(k)} \\
&= \sum_{k=1}^{n_3} 2\big(\mathcal{F}(\mathcal{N})^{(k)} - \mathcal{F}(\mathcal{N}^{\star})^{(k)}\big) : \mathcal{F}(\mathcal{H})^{(k)} : \mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^{\top})^{(k)} \\
&+ 2\,\mathcal{F}(\mathcal{N}^{\star})^{(k)} : \mathcal{F}(\mathcal{G})^{(k)} : \mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^{\top})^{(k)} \\
&= \sum_{k=1}^{n_3} -2\big(\mathcal{F}(\mathcal{N})^{(k)} - \mathcal{F}(\mathcal{N}^{\star})^{(k)}\big) : \mathcal{F}(\mathcal{H})^{(k)} : \big(\mathcal{F}(\mathcal{N})^{(k)} - \mathcal{F}(\mathcal{N}^{\star})^{(k)}\big) \\
&\quad + 2\big(\mathcal{F}(\mathcal{N})^{(k)} - \mathcal{F}(\mathcal{N}^{\star})^{(k)}\big) : \mathcal{F}(\mathcal{H})^{(k)} : \big(\mathcal{F}(\mathcal{W})^{(k)}\mathcal{F}(\Delta^{\top})^{(k)} + \mathcal{F}(\Delta)^{(k)}\mathcal{F}(\mathcal{W}^{\top})^{(k)}\big) \\
&\quad + 2\,\mathcal{F}(\mathcal{N}^{\star})^{(k)} : \mathcal{F}(\mathcal{G})^{(k)} : \mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^{\top})^{(k)} \\
&= \sum_{k=1}^{n_3} -2\big(\mathcal{F}(N)^{(k)} - \mathcal{F}(N^{\star})^{(k)}\big) : \mathcal{F}(\mathcal{H})^{(k)} : \big(\mathcal{F}(N)^{(k)} - \mathcal{F}(N^{\star})^{(k)}\big) \\
&- 2\mathcal{F}(N^{\star})^{(k)} : \mathcal{F}(\mathcal{G})^{(k)} : \mathcal{F}(N)^{(k)} + 2\langle \nabla f(\mathcal{W}), \Delta\rangle - 2\langle \nabla Q(\mathcal{W}), \Delta\rangle.
\end{aligned}
$$

To sum up, we have

$$
\begin{aligned}
&\Delta : \nabla^2 f(\mathcal{W}) : \Delta \\
&= \sum_{k=1}^{n_3} \mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^{\top})^{(k)} : \mathcal{F}(\mathcal{H})^{(k)} : \mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^{\top})^{(k)} \\
&- 3\big(\mathcal{F}(\mathcal{N})^{(k)} - \mathcal{F}(\mathcal{N}^{\star})^{(k)}\big) : \mathcal{F}(\mathcal{H})^{(k)} : \big(\mathcal{F}(\mathcal{N})^{(k)} - \mathcal{F}(\mathcal{N}^{\star})^{(k)}\big) \\
&\quad - 6\mathcal{F}(\mathcal{N}^{\star})^{(k)} : \mathcal{F}(\mathcal{G})^{(k)} : \mathcal{F}(\mathcal{N})^{(k)} \\
&+ 4\langle \nabla f(\mathcal{W}), \Delta\rangle + \big[\Delta : \nabla^2 Q(\mathcal{W}) : \Delta - 4\langle \nabla Q(\mathcal{W}), \Delta\rangle\big] \\
&\le \sum_{k=1}^{n_3} \mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^{\top})^{(k)} : \mathcal{F}(\mathcal{H})^{(k)} : \mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta)^{(k)^{\top}} \\
&- 3\big(\mathcal{F}(\mathcal{N})^{(k)} - \mathcal{F}(\mathcal{N}^{\star})^{(k)}\big) : \mathcal{F}(\mathcal{H})^{(k)} : \big(\mathcal{F}(\mathcal{N})^{(k)} - \mathcal{F}(\mathcal{N}^{\star})^{(k)}\big) \\
&+ 4\langle \nabla f(\mathcal{W}), \Delta\rangle + \big[\Delta : \nabla^2 Q(\mathcal{W}) : \Delta - 4\langle \nabla Q(\mathcal{W}), \Delta\rangle\big].
\end{aligned}
$$

The rest of the proof follows from the following lemma [2]. $\square$

Lemma 15. *Let for any* $k = 1, \dots, n_3$

$$\mathcal{F}(\mathcal{N})^{(k)} := \begin{bmatrix} \mathcal{F}(\mathcal{U})^{(k)} \cdot \mathcal{F}(\mathcal{U}^\top)^{(k)} & \mathcal{F}(\mathcal{U})^{(k)} \cdot \mathcal{F}(\mathcal{V}^\top)^{(k)} \\ \mathcal{F}(\mathcal{V})^{(k)} \cdot \mathcal{F}(\mathcal{U}^\top)^{(k)} & \mathcal{F}(\mathcal{V})^{(k)} \cdot \mathcal{F}(\mathcal{V}^\top)^{(k)} \end{bmatrix} \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)}. \tag{38}$$

If $\mathcal{F}(\mathcal{H}_0)^{(k)}$ *satisfies:*

$$\big(1-\delta\big)\|\mathcal{F}(\mathcal{U})^{(k)} \cdot \mathcal{F}(\mathcal{V}^\top)^{(k)}\|_F^2 \leq \left\{ \begin{array}{c} \mathcal{F}(\mathcal{U})^{(k)} \cdot \mathcal{F}(\mathcal{V}^\top)^{(k)} \\ : \mathcal{F}(\mathcal{H}_0)^{(k)} : \\ \mathcal{F}(\mathcal{U})^{(k)} \cdot \mathcal{F}(\mathcal{V}^\top)^{(k)} \end{array} \right\} \leq \big(1+\delta\big)\|\mathcal{F}(\mathcal{U})^{(k)} \cdot \mathcal{F}(\mathcal{V}^\top)^{(k)}\|_F^2. \tag{39}$$

Then, we have

$$\big(1-2\delta\big)\|\mathcal{F}(\mathcal{N})^{(k)}\|_F^2 \leq \mathcal{F}(\mathcal{N})^{(k)} : \mathcal{F}(\mathcal{H})^{(k)} : \mathcal{F}(\mathcal{N})^{(k)} \leq \big(1+2\delta\big)\|\mathcal{F}(\mathcal{N})^{(k)}\|_F^2.$$

Proof. Knowing that $\mathcal{H}_0$ preserves the norm $\mathcal{M}$, which is the off-diagonal of $\mathcal{N}$ and $\mathcal{G}$ the norm of the diagonal components of $\mathcal{N}$, we have, for any $k = 1, \dots, n_3$

$$\begin{aligned}
&\mathcal{F}(\mathcal{N})^{(k)} : \mathcal{F}(\mathcal{H})^{(k)} : \mathcal{F}(\mathcal{N})^{(k)} \\
&= \mathcal{F}(\mathcal{N})^{(k)} : 4\mathcal{F}(\mathcal{H}_1)^{(k)} : \mathcal{F}(\mathcal{N})^{(k)} + \mathcal{F}(\mathcal{N})^{(k)} : \mathcal{F}(\mathcal{G})^{(k)} : \mathcal{F}(\mathcal{N})^{(k)} \\
&= 4\,\mathcal{F}(\mathcal{U})^{(k)} \cdot \mathcal{F}(\mathcal{V}^\top)^{(k)} : \mathcal{F}(\mathcal{H}_0)^{(k)} : \mathcal{F}(\mathcal{U})^{(k)} \cdot \mathcal{F}(\mathcal{V}^\top)^{(k)} \\
&\qquad + \big(\|\mathcal{F}(\mathcal{U})^{(k)} \cdot \mathcal{F}(\mathcal{U}^\top)^{(k)}\|_F^2 + \|\mathcal{F}(\mathcal{V})^{(k)} \cdot \mathcal{F}(\mathcal{V}^\top)^{(k)}\|_F^2 \\
&\qquad - 2\langle \mathcal{F}(\mathcal{U})^{(k)} \cdot \mathcal{F}(\mathcal{U}^\top)^{(k)}, \mathcal{F}(\mathcal{V})^{(k)} \cdot \mathcal{F}(\mathcal{V}^\top)^{(k)} \rangle\big) \\
&= 4\,\mathcal{F}(\mathcal{U})^{(k)} \cdot \mathcal{F}(\mathcal{V}^\top)^{(k)} : \mathcal{F}(\mathcal{H}_0)^{(k)} : \mathcal{F}(\mathcal{U})^{(k)} \cdot \mathcal{F}(\mathcal{V}^\top)^{(k)} \\
&\qquad + \big(\|\mathcal{F}(\mathcal{U})^{(k)} \cdot \mathcal{F}(\mathcal{U}^\top)^{(k)}\|_F^2 + \|\mathcal{F}(\mathcal{V})^{(k)} \cdot \mathcal{F}(\mathcal{V}^\top)^{(k)}\|_F^2 \\
&\qquad - 2\|\mathcal{F}(\mathcal{U})^{(k)} \cdot \mathcal{F}(\mathcal{V}^\top)^{(k)}\|_F^2\big).
\end{aligned}$$

Using (39), we obtain by calculating:

$$\big(1-2\delta\big)\|\mathcal{F}(\mathcal{N})^{(k)}\|_F^2 \leq \mathcal{F}(\mathcal{N})^{(k)} : \mathcal{F}(\mathcal{H})^{(k)} : \mathcal{F}(\mathcal{N})^{(k)} \leq \big(1+2\delta\big)\|\mathcal{F}(\mathcal{N})^{(k)}\|_F^2.$$

$\square$

[2] Lemma 19 in [15] is false as stated

7.3.2. End of the proof

We first prove that the regularisation enforces that the rows of $\mathcal{F}\big(\mathcal{U}\big)^{(k)}$, $\mathcal{F}\big(\mathcal{V}\big)^{(k)}$, for $k = 1 \dots n_3$, cannot be too large.

Lemma 16. *Let $d = \max\{n_1, n_2\}$, there exists an absolute constant c_1, when sample rate*

$$p > c_1 \frac{\mu r}{\min\{n_1, n_2\}} \log(d),$$

$\alpha_1^2 = \Theta\big(\frac{\mu r \sigma_1^\star}{n_1}\big), \alpha_2^2 = \Theta\big(\frac{\mu r \sigma_1^\star}{n_2}\big)$ and $\lambda_1 = \Theta\big(\frac{n_1}{\mu r \kappa^\star}\big), \lambda_2 = \Theta\big(\frac{n_2}{\mu r \kappa^\star}\big)$, we have for any $\mathcal{W}$ with $\|\nabla f(\mathcal{W})\|_F \leq \epsilon$ for any polynomially small ϵ, with probability at least

$$1 - 2n_1 n_3 \exp\left\{ -pn_2 \big(\big(1 + \frac{t}{pn_2}\big) \ln\big(1 + \frac{t}{pn_2}\big) - \frac{t}{pn_2} \big) \right\}$$

$$\max_{1 \leq i \leq n_1} \|e_i^\top * \mathcal{U}\|_F^2 \leq C \ n_3 \frac{\mu^2 r^{2.5} (\kappa^\star)^2 \sigma_1^\star}{n_1},$$

$$\max_{1 \leq j \leq n_2} \|e_j^\top * \mathcal{V}\|_F^2 \leq C \ n_3 \frac{\mu^2 r^{2.5} (\kappa^\star)^2 \sigma_1^\star}{n_2}$$

for some constant positive C.

Proof. In this proof, by symmetry, without loss of generality, we can assume that for any $k = 1, \dots, n_3$

$$\sqrt{n_1} \max_{1 \leq i \leq n_1} \|\mathcal{F}\big(e_i^\top\big)^{(k)} \mathcal{F}\big(\mathcal{U}\big)^{(k)}\|_2 \geq \sqrt{n_2} \max_{1 \leq j \leq n_2} \|\mathcal{F}\big(e_j^\top\big)^{(k)} \mathcal{F}\big(\mathcal{V}\big)^{(k)}\|_2.$$

By calculating the gradient, we can write the gradient as:

$$\nabla f(\mathcal{W}) = \frac{4}{p} \begin{pmatrix} \big(\mathcal{M} - \mathcal{M}^\star\big)_\Omega * \mathcal{V} \\ \big(\mathcal{M} - \mathcal{M}^\star\big)_\Omega * \mathcal{U} \end{pmatrix} + \begin{pmatrix} \mathcal{U} * \big(\mathcal{U}^\top * \mathcal{U} - \mathcal{V}^\top * \mathcal{V}\big) \\ \mathcal{V} * \big(\mathcal{V}^\top * \mathcal{V} - \mathcal{U}^\top * \mathcal{U}\big) \end{pmatrix} + \nabla Q(\mathcal{W})$$

where

$$\nabla Q(\mathcal{W}) = 4\lambda_1 \sum_{k=1}^{n_3} \sum_{i=1}^{n_1} \left(\|\mathcal{F}\big(e_i^\top\big)^{(k)} \mathcal{F}\big(\mathcal{W}\big)^{(k)}\|_2 - \alpha_1\right)_+^3$$
$$\times \frac{\mathcal{F}\big(e_i\big)^{(k)} \mathcal{F}\big(e_i^\top\big)^{(k)} \mathcal{F}\big(\mathcal{W}\big)^{(k)}}{\|\mathcal{F}\big(e_i\big)^{(k)} \mathcal{F}\big(\mathcal{W}\big)^{(k)}\|_2^2}$$
$$+ 4\lambda_2 \sum_{k=1}^{n_3} \sum_{i=n_1+1}^{n_2} \left(\|\mathcal{F}\big(e_i^\top\big)^{(k)} \mathcal{F}\big(\mathcal{W}\big)^{(k)}\|_2 - \alpha_2\right)_+^3$$
$$\times \frac{\mathcal{F}\big(e_i\big)^{(k)} \mathcal{F}\big(e_i^\top\big)^{(k)} \mathcal{F}\big(\mathcal{W}\big)^{(k)}}{\|\mathcal{F}\big(e_i\big)^{(k)} \mathcal{F}\big(\mathcal{W}\big)^{(k)}\|_2^2}.$$

Using the fact $\langle \nabla Q(\mathcal{W}), \mathcal{W}\rangle \geq 0$, thus, for any point $\mathcal{W}$ with gradient $\|\nabla f(\mathcal{W})\|_F \leq \epsilon$, we have:

$$\begin{aligned}
\epsilon\|\mathcal{W}\|_F &\geq \langle \nabla f(\mathcal{W}), \mathcal{W}\rangle \\
&= \|\mathcal{U}^\top * \mathcal{U} - \mathcal{V}^\top * \mathcal{V}\|_F^2 + \frac{4}{p}\Big\langle \big(\mathcal{M} - \mathcal{M}^\star\big)_\Omega, \mathcal{M}\Big\rangle + \Big\langle \nabla Q(\mathcal{W}), \mathcal{W}\Big\rangle \\
&\geq \|\mathcal{U}^\top * \mathcal{U} - \mathcal{V}^\top * \mathcal{V}\|_F^2 - \frac{4}{p}\Big\langle \big(\mathcal{M}^\star\big)_\Omega, \big(\mathcal{M}\big)_\Omega\Big\rangle \\
&\geq \|\mathcal{U}^\top * \mathcal{U} - \mathcal{V}^\top * \mathcal{V}\|_F^2 - 4 \cdot \frac{1}{\sqrt{p}}\|\mathcal{M}^\star\|_\Omega \cdot \frac{1}{\sqrt{p}}\|\mathcal{M}\|_\Omega \\
&= \|\mathcal{U}^\top * \mathcal{U} - \mathcal{V}^\top * \mathcal{V}\|_F^2 - 4 \cdot \frac{1}{\sqrt{p}}\sqrt{\sum_{k=1}^{n_3} \|\mathcal{F}\big(\mathcal{M}^\star\big)^{(k)}\|_\Omega^2} \cdot \frac{1}{\sqrt{p}}\sqrt{\sum_{k=1}^{n_3} \|\mathcal{F}\big(\mathcal{M}\big)^{(k)}\|_\Omega^2} \\
&\geq \|\mathcal{U}^\top * \mathcal{U} - \mathcal{V}^\top * \mathcal{V}\|_F^2 - C\ \sqrt{n_1 n_2}\sum_{k=1}^{n_3} \|\mathcal{F}\big(\mathcal{M}^\star\big)^{(k)}\|_F \cdot \sum_{k=1}^{n_3} \|\mathcal{F}\big(M\big)^{(k)}\|_\infty
\end{aligned}$$

where in the last inequality, we use Lemma 1 and Lemma 5. Let

$$i^\star = \arg\max_{1\leq i\leq n_1} \|e_i^\top * \mathcal{U}\|_F \tag{40}$$

and

$$j^\star = \arg\max_{1\leq j\leq n_2} \|e_j^\top * \mathcal{V}\|_F. \tag{41}$$

542 Using these facts and recalling that, by assumption,

- $\sqrt{n_1}\|\mathcal{F}(e_{i^\star}^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 \geq \sqrt{n_2}\|\mathcal{F}(e_{j^\star}^\top)^{(k)}\mathcal{F}(\mathcal{V})^{(k)}\|_2$
- $\|\mathcal{M}^\star\|_F = \|\mathcal{F}(\mathcal{M}^\star)\|_F = \sqrt{\sum_{k=1}^{n_3}\|\mathcal{F}(\mathcal{M}^\star)^{(k)}\|_F^2} \leq n_3\sigma_1^\star\sqrt{r}$
- $\|\mathcal{M}\|_\infty = \|\mathcal{F}(\mathcal{M})\|_\infty \leq \sum_{k=1}^{n_3}\|\mathcal{F}(e_{i^\star}^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2\|\mathcal{F}(e_{j^\star}^\top)^{(k)}\mathcal{F}(\mathcal{V})^{(k)}\|_2$

we have, for some positive constant C:

$$\begin{aligned}\|\mathcal{U}^\top * \mathcal{U} - \mathcal{V}^\top * \mathcal{V}\|_F^2 \leq C\ n_1 n_3 \sigma_1^\star\sqrt{r}\sum_{k=1}^{n_3}\|\mathcal{F}(e_{i^\star}^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2^2 \\ + C\ \epsilon d\|e_{i^\star}^\top * \mathcal{U}\|_F,\end{aligned} \tag{42}$$

where in the second term of (42), we use:

$$\begin{aligned}\|\mathcal{W}\|_F &\leq \|\mathcal{U}\|_F + \|\mathcal{V}\|_F \\ &\leq \sum_{k=1}^{n_3}\|\mathcal{F}(\mathcal{U})^k\|_F + \|\mathcal{F}(\mathcal{V})^k\|_F \\ &\leq \sum_{k=1}^{n_3}\sqrt{n_1}\max_i\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 + \sqrt{n_2}\max_j\|\mathcal{F}(e_j^\top)^{(k)}\mathcal{F}(\mathcal{V})^{(k)}\|_2 \\ &\leq 2\sum_{k=1}^{n_3}\sqrt{n_1}\,\|\mathcal{F}(e_{i^\star}^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 \\ &\leq d\sum_{k=1}^{n_3}\ \|\mathcal{F}(e_{i^\star}^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 \text{ where } d = \max\{n_1, n_2\} \\ &= d\|e_{i^\star}^\top * \mathcal{U}\|_F.\end{aligned}$$

In the case $\|e_{i^\star}^\top * \mathcal{U}\|_2 \geq 2\alpha_1$, consider $\langle e_{i^\star}^\top * \nabla f(\mathcal{U}), e_{i^\star}^\top * \mathcal{U}\rangle$ as:

$$
\begin{aligned}
&\epsilon\|e_{i^\star}^\top * \mathcal{U}\|_F \\
&\geq \langle e_{i^\star}^\top * \nabla f(\mathcal{U}), e_{i^\star}^\top * \mathcal{U}\rangle \\
&= \left\langle e_{i^\star}^\top * (\frac{4}{p}\big(\mathcal{M}-\mathcal{M}^\star\big)_\Omega * \mathcal{V} + \mathcal{U} * \big(\mathcal{U}^\top * \mathcal{U} - \mathcal{V}^\top * \mathcal{V}\big) + \nabla Q_1(\mathcal{U})\big), e_{i^\star}^\top * \mathcal{U}\right\rangle \\
&\geq \frac{\lambda_1}{2}\Big(\|e_{i^\star}^\top * \mathcal{U}\|_F - \alpha_1\Big)_+^3 \|e_{i^\star}^\top * \mathcal{U}\|_F - \frac{4}{p}\langle e_{i^\star}^\top * (\mathcal{M}^\star)_\Omega, e_{i^\star}^\top * \big(\mathcal{M}\big)_\Omega\rangle \\
&- \|\mathcal{U}^\top * \mathcal{U} - \mathcal{V}^\top * \mathcal{V}\|_F \|e_{i^\star}^\top * \mathcal{U}\|_F^2 \\
&\geq \frac{\lambda_1}{2}\|e_{i^\star}^\top * \mathcal{U}\|_F^4 - 4\frac{1}{\sqrt{p}}\|e_{i^\star}^\top * (\mathcal{M}^\star)_\Omega\| \cdot \frac{1}{\sqrt{p}}\|e_{i^\star}^\top * \big(\mathcal{M}\big)_\Omega\| - \|\mathcal{U}^\top * \mathcal{U} - \mathcal{V}^\top * \mathcal{V}\|_F \cdot \|e_{i^\star}^\top * \mathcal{U}\|_F^2 \\
&\geq \frac{\lambda_1}{2}\|e_{i^\star}^\top * \mathcal{U}\|_F^4 - \|\mathcal{U}^\top * \mathcal{U} - \mathcal{V}^\top * \mathcal{V}\|_F \cdot \|e_{i^\star}^\top * \mathcal{U}\|_F^2 \\
&- 4\sqrt{1+0.01}\sum_{k=1}^{n_3}\|\mathcal{F}(e_{i^\star}^\top)^{(k)}\mathcal{F}(\mathcal{M}^\star)^{(k)}\|_2 \cdot C\ \sqrt{n_2}\sum_{k=1}^{n_3}\|\mathcal{F}(\mathcal{M})^{(k)}\|_\infty \\
&\geq \frac{\lambda_1}{2}\|e_{i^\star}^\top * \mathcal{U}\|_F^4 - \|\mathcal{U}^\top * \mathcal{U} - \mathcal{V}^\top * \mathcal{V}\|_F \cdot \|e_{i^\star}^\top * \mathcal{U}\|_F^2 \\
&\quad - C\ \sqrt{\mu r}\sigma_1^\star \sum_{k=1}^{n_3}\|\mathcal{F}(e_{i^\star}^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2^2
\end{aligned}
$$

where in the last inequality, we use (1) and (5). Further, using (42), we have:

$$
\begin{aligned}
\lambda_1\sum_{k=1}^{n_3}\|\mathcal{F}(e_{i^\star}^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2^3 \leq 2\epsilon &+ C\ \sqrt{\mu r}\sigma_1^\star\sum_{k=1}^{n_3}\|\mathcal{F}(e_{i^\star}^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2 \\
&+ C\ \sqrt{n_1 n_3 \sigma_1^\star} r^{\frac{1}{4}}\sum_{k=1}^{n_3}\|\mathcal{F}(e_{i^\star}^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2^2 \\
&+ C\sqrt{\epsilon d}\sum_{k=1}^{n_3}\|\mathcal{F}(e_{i^\star}^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\|_2^{\frac{3}{2}}.
\end{aligned}
$$

By choosing ϵ to be polynomially small, we have:

$$
\begin{aligned}
\sqrt{\frac{n_1}{n_2}}\max_{1\leq j\leq n_2}\|e_{j^\star}^\top * \mathcal{V}\|_F &\leq \max_{1\leq i\leq n_1}\|e_{i^\star}^\top * \mathcal{U}\|_F^2 \\
&\leq c\max\left\{\alpha_1^2, \frac{\sqrt{\mu r}.\sigma_1^\star}{\lambda_1}, \frac{n_1 n_3 \sigma_1^\star\sqrt{r}}{\lambda_1^2}\right\}
\end{aligned}
$$

for some positive constant c. Finally, substituting the choice of α^2 and λ_1, we finished the proof. $\square$

Now, we prove, that the Hessian $\mathcal{H}$ terms in (36) is negative when $\mathcal{W} \neq \mathcal{W}^\star$.

Lemma 17. *Let $d = \max\{n_1, n_2\}$, when sample rate $p \geq \Omega\big(\frac{\mu^4 r^6 (\kappa^\star)^6 \log d}{\min\{n_1, n_2\}}\big)$, by choosing $\alpha_1^2 = \Theta\big(\frac{\mu r \sigma_1^\star}{n_1}\big)$, $\alpha_2^2 = \Theta\big(\frac{\mu r \sigma_1^\star}{n_2}\big)$ and $\lambda_1 = \Theta\big(\frac{n_1}{\mu r \kappa^\star}\big)$, $\lambda_2 = \Theta\big(\frac{n_2}{\mu r \kappa^\star}\big)$ with probability at least*

$$1 - 2n_1 n_3 \exp\left\{-pn_2\big(\big(1 + \frac{t}{pn_2}\big)\ln\big(1 + \frac{t}{pn_2}\big) - \frac{t}{pn_2}\big)\right\}$$

for all $\mathcal{W}$ with $\|\nabla f(\mathcal{W})\|_F \leq \epsilon$ and for polynomially small ϵ, we have

$$\Delta * \Delta^\top : \mathcal{H} : \Delta * \Delta^\top - 3(\mathcal{M} - \mathcal{M}^\star) : \mathcal{H} : (\mathcal{M} - \mathcal{M}^\star) \leq -0.3 \sum_{k=1}^{n_3} \sigma_r^\star \|\mathcal{F}\big(\Delta\big)^{(k)}\|_F^2.$$

Proof. As the symmetric case, we are interested in studying the two cases on the norm of Δ and we use the different inequalities of concentrations. Using the choices of α and λ and (16), we know when ϵ is polynomially small with high probability:

$$\max_{1 \leq i \leq n_1} \|e_i^\top * \mathcal{U}\|_F^2 \leq C\ n_3 \frac{\mu^2 r^{2.5} (\kappa^\star)^2 \sigma_1^\star}{n_1},$$

and

$$\max_{1 \leq j \leq n_2} \|e_j^\top * \mathcal{V}\|_F^2 \leq C\ n_3 \frac{\mu^2 r^{2.5} (\kappa^\star)^2 \sigma_1^\star}{n_2}.$$

In the following, we denote for any $k = 1, \ldots, n_3$,

$$\mathcal{F}(\Delta)^{(k)} = \big(\mathcal{F}(\Delta_{\mathcal{U}}^\top)^{(k)}, \mathcal{F}(\Delta_{\mathcal{V}}^\top)^{(k)}\big)^\top$$

and we have

$$\|\mathcal{F}(\Delta_{\mathcal{U}})^{(k)}\|_F \leq \|\mathcal{F}(\Delta)^{(k)}\|_F$$

and

$$\|\mathcal{F}(\Delta_{\mathcal{V}})^{(k)}\|_F \leq \|\mathcal{F}(\Delta)^{(k)}\|_F.$$

We now split the analysis into two cases.

Case 1: $\|\mathcal{F}(\Delta)^{(k)}\|_F^2 \leq \frac{\sigma_r^\star}{40}$, for any $k = 1, \ldots, n_3$. By (1) and (15), we have:

$$\frac{1}{p}\|\mathcal{W}^\star * \Delta^\top\|_\Omega^2 \geq \big(1 - 2\delta\big)\|\mathcal{W}^\star * \Delta^\top\|_F^2 \geq \big(1 - 2\delta\big)\sigma_r^\star\|\Delta\|_F^2.$$

Furthermore, we know:

$$\frac{1}{p}\|\Delta_\mathcal{U} * \Delta_\mathcal{V}^\top\|_\Omega^2 = \frac{1}{p}\sum_{k=1}^{n_3}\|\mathcal{F}(\Delta_\mathcal{U})^{(k)}\mathcal{F}(\Delta_\mathcal{V}^\top)^{(k)}\|_\Omega^2.$$

By (3) and with the choice of p, for any $k = 1, \ldots, n_3$, we have:

$$\begin{aligned}
\frac{1}{p}\|\mathcal{F}(\Delta_\mathcal{U})^{(k)}\mathcal{F}(\Delta_\mathcal{V}^\top)^{(k)}\|_\Omega^2 &\leq \big(1 + \delta\big)\|\mathcal{F}(\Delta_\mathcal{U})^{(k)}\|_F^2\|\mathcal{F}(\Delta_\mathcal{V}^\top)^{(k)}\|_F^2 \\
&\quad + C\sqrt{\frac{d}{p}} \cdot \frac{\mu^2 r^{2.5}(\kappa^\star)^2\sigma_1^\star}{\sqrt{n_1 n_2}}\|\mathcal{F}(\Delta_\mathcal{U})^{(k)}\|_F\|\mathcal{F}(\Delta_\mathcal{V})^{(k)}\|_F \\
&\leq \big(1 + \delta\big)\|\mathcal{F}(\Delta)^{(k)}\|_F^4 + \frac{\sigma_r^\star}{4}\|\mathcal{F}(\Delta)^{(k)}\|_F^2 \\
&\leq \frac{\sigma_r^\star}{20}\|\mathcal{F}(\Delta)^{(k)}\|_F^2.
\end{aligned}$$

So, using that for $k = 1, \ldots, n_3$,

$$\|\mathcal{F}(\Delta_\mathcal{U})^{(k)}\|_F^2 \leq \|\mathcal{F}(\Delta)^{(k)}\|_F^2 \leq \frac{\sigma_r^\star}{40}$$

and

$$\|\mathcal{F}(\Delta_\mathcal{V})^{(k)}\|_F^2 \leq \|\mathcal{F}(\Delta)^{(k)}\|_F^2 \leq \frac{\sigma_r^\star}{40}$$

we obtain:

$$\begin{aligned}
&\Delta * \Delta^\top : \mathcal{H} : \Delta * \Delta^\top \\
&= \frac{4}{p}\|\Delta_\mathcal{U} * \Delta_\mathcal{V}^\top\|_\Omega^2 + \big(\|\Delta_\mathcal{U} * \Delta_\mathcal{U}^\top\|_F^2 + \|\Delta_\mathcal{V} * \Delta_\mathcal{V}^\top\|_F^2 - 2\|\Delta_\mathcal{U} * \Delta_\mathcal{V}^\top\|_F^2\big) \\
&\leq \sum_{k=1}^{n_3}\frac{1}{4}\sigma_r^\star\|\mathcal{F}\big(\Delta\big)^{(k)}\|_F^2.
\end{aligned}$$

Using this facts, we obtain

$$
\begin{aligned}
&\Delta * \Delta^\top : \mathcal{H} : \Delta * \Delta^\top - 3\big(\mathcal{N} - \mathcal{N}^\star\big) : \mathcal{H} : \big(\mathcal{N} - \mathcal{N}^\star\big) \\
&= \sum_{k=1}^{n_3} \mathcal{F}(\Delta)^{(k)} \mathcal{F}(\Delta^\top)^{(k)} : \mathcal{F}(\mathcal{H})^{(k)} : \mathcal{F}(\Delta)^{(k)} \mathcal{F}(\Delta^\top)^{(k)} \\
&\ - 3\big(\mathcal{F}(\mathcal{N})^{(k)} - \mathcal{F}(\mathcal{N}^\star)^{(k)}\big) : \mathcal{F}(\mathcal{H})^{(k)} : \big(\mathcal{F}(N)^{(k)} - \mathcal{F}(\mathcal{N}^\star)^{(k)}\big) \\
&= \sum_{k=1}^{n_3} \mathcal{F}(\Delta)^{(k)} \mathcal{F}(\Delta^\top)^{(k)} : \mathcal{F}(\mathcal{H})^{(k)} : \mathcal{F}(\Delta)^{(k)} \mathcal{F}(\Delta^\top)^{(k)} \\
&- \left\{ \begin{array}{c} 3\big(\mathcal{F}(\mathcal{W}^\star)^{(k)} \mathcal{F}(\Delta^\top)^{(k)} + \mathcal{F}(\Delta)^{(k)} \mathcal{F}({\mathcal{W}^\star}^\top)^{(k)} + \mathcal{F}(\Delta)^{(k)} \mathcal{F}(\Delta^\top)^{(k)}\big) \\ : \mathcal{F}(\mathcal{H})^{(k)} : \\ \big(\mathcal{F}(\mathcal{W}^\star)^{(k)} \mathcal{F}(\Delta^\top)^{(k)} + \mathcal{F}(\Delta)^{(k)} \mathcal{F}({\mathcal{W}^\star}^\top)^{(k)} + \mathcal{F}(\Delta)^{(k)} \mathcal{F}(\Delta^\top)^{(k)}\big) \end{array} \right\} \\
&\le \sum_{k=1}^{n_3} -12\Big(\mathcal{F}(\mathcal{W}^\star)^{(k)} \mathcal{F}(\Delta^\top)^{(k)} : \mathcal{F}(\mathcal{H})^{(k)} : \mathcal{F}(\Delta)^{(k)} \mathcal{F}(\Delta^\top)^{(k)} \\
&+ \mathcal{F}(\mathcal{W}^\star)^{(k)} \mathcal{F}(\Delta^\top)^{(k)} : \mathcal{F}(\mathcal{H})^{(k)} : \mathcal{F}(\mathcal{W}^\star)^{(k)} \mathcal{F}(\Delta^\top)^{(k)}\Big) \\
&\le \sum_{k=1}^{n_3} -\frac{12}{p}\Big(\|\mathcal{F}(\mathcal{W}^\star)^{(k)} \mathcal{F}(\Delta^\top)^{(k)}\|_\Omega^2 - \|\mathcal{F}(\mathcal{W}^\star)^{(k)} \mathcal{F}(\Delta^\top)^{(k)}\|_\Omega \|\mathcal{F}(\Delta)^{(k)} \mathcal{F}(\Delta^\top)^{(k)}\|_\Omega\Big) \\
&= \sum_{k=1}^{n_3} -\frac{12}{p}\|\mathcal{F}(\mathcal{W}^\star)^{(k)} \mathcal{F}(\Delta^\top)^{(k)}\|_\Omega \big(\|\mathcal{F}(\mathcal{W}^\star)^{(k)} \mathcal{F}(\Delta^\top)^{(k)}\|_\Omega - \|\mathcal{F}(\Delta)^{(k)} \mathcal{F}(\Delta^\top)^{(k)}\|_\Omega\big) \\
&\le \sum_{k=1}^{n_3} -12\sqrt{1-2\delta}\big(\sqrt{1-2\delta} - \sqrt{1/4}\big)\sigma_r^\star \|\mathcal{F}(\Delta)^{(k)}\|_F^2 \\
&\le -1.2 \sum_{k=1}^{n_3} \sigma_r^\star \|\mathcal{F}(\Delta)^{(k)}\|_F^2.
\end{aligned}
$$

Case 2: $\|\mathcal{F}(\Delta)^{(k)}\|_F^2 \ge \frac{\sigma_r^\star}{40}$, for any $k = 1, \ldots, n_3$. By Lemma 4 with high probability with the choice of p, we have:

$$
\begin{aligned}
\frac{1}{p}\|\Delta_{\mathcal{U}} * \Delta_{\mathcal{V}}^{\top}\|_{\Omega}^{2} &= \sum_{k=1}^{n_3} \frac{1}{p}\|\mathcal{F}(\Delta_{\mathcal{U}})^{(k)}\mathcal{F}(\Delta_{\mathcal{V}}^{\top})^{(k)}\|_{\Omega}^{2} \\
&\leq \sum_{k=1}^{n_3} \|\mathcal{F}(\Delta_{\mathcal{U}})^{(k)}\mathcal{F}(\Delta_{\mathcal{V}}^{\top})^{(k)}\|_{F}^{2} \\
&+ C\big(\frac{dr\log(d)}{p} \ \|\mathcal{F}(\Delta_{\mathcal{U}})^{(k)}\mathcal{F}(\Delta_{\mathcal{V}}^{\top})^{(k)}\|_{\infty}^{2} \\
&+ \sqrt{\frac{dr\log(d)}{p}}\|\mathcal{F}(\Delta_{\mathcal{U}})^{(k)}\mathcal{F}(\Delta_{\mathcal{V}}^{\top})^{(k)} \ \|_{F} \ \|\mathcal{F}(\Delta_{\mathcal{U}})^{(k)}\mathcal{F}(\Delta_{\mathcal{V}}^{\top})^{(k)}\|_{\infty}\big) \\
&\leq \sum_{k=1}^{n_3} \|\mathcal{F}(\Delta_{\mathcal{U}})^{(k)}\mathcal{F}(\Delta_{\mathcal{V}}^{\top})^{(k)}\|_{F}^{2} \\
&+ C\big(\frac{dr\log(d)}{p} \cdot \frac{\mu^4 r^5 (\kappa^{\star})^4 (\sigma_1^{\star})^2}{n_1 n_2} + \sqrt{\frac{dr\log(d)}{p} \cdot \frac{\mu^4 r^5 (\kappa^{\star})^4 (\sigma_1^{\star})^2}{n_1 n_2}} \ \|\mathcal{F}(\Delta)^{(k)}\|_{F}^{2}\big) \\
&\leq \sum_{k=1}^{n_3} \|\mathcal{F}(\Delta_{\mathcal{U}})^{(k)}\mathcal{F}(\Delta_{\mathcal{V}}^{\top})^{(k)}\|_{F}^{2} + \frac{\left(\sigma_r^{\star}\right)^2}{1000} + \frac{\sigma_r^{\star}}{1000} \ \|\mathcal{F}(\Delta)^{(k)}\|_{F}^{2} \\
&\leq \sum_{k=1}^{n_3} \|\mathcal{F}(\Delta_{\mathcal{U}})^{(k)}\mathcal{F}(\Delta_{\mathcal{V}}^{\top})^{(k)}\|_{F}^{2} + 0.01\sigma_r^{\star} \ \|\mathcal{F}(\Delta)^{(k)}\|_{F}^{2}.
\end{aligned}
$$

Again using Lemma 4 with high probability, we have

$$
\begin{aligned}
\frac{1}{p}\|\mathcal{M}-\mathcal{M}^{\star}\|_{\Omega}^{2} &= \sum_{k=1}^{n_3}\frac{1}{p}\|\mathcal{F}(\mathcal{M})^{(k)}-\mathcal{F}(\mathcal{M}^{\star})^{(k)}\|_{\Omega}^{2} \\
&\geq \sum_{k=1}^{n_3}\|\mathcal{F}(\mathcal{M})^{(k)}-\mathcal{F}(\mathcal{M}^{\star})^{(k)}\|_{F}^{2} \\
&+ C\big(\frac{d\,r\log(d)}{p}\|\mathcal{F}(\mathcal{M})^{(k)}-\mathcal{F}(\mathcal{M}^{\star})^{(k)}\|_{\infty}^{2} \\
&+ \sqrt{\frac{d\,r\log(d)}{p}}\|\mathcal{F}(\mathcal{M})^{(k)}-\mathcal{F}(\mathcal{M}^{\star})^{(k)}\|_{F}\times\|\mathcal{F}(\mathcal{M})^{(k)}-\mathcal{F}(\mathcal{M}^{\star})^{(k)}\|_{\infty}\big) \\
&\geq \sum_{k=1}^{n_3}\|\mathcal{F}(\mathcal{M})^{(k)}-\mathcal{F}(\mathcal{M}^{\star})^{(k)}\|_{F}^{2} \\
&+ C\big(\frac{d\,r\log(d)}{p}\cdot\frac{\mu^4 r^5(\kappa^{\star})^4(\sigma_1^{\star})^2}{n_1 n_2} \\
&+ \sqrt{\frac{d\,r\log(d)}{p}\cdot\frac{\mu^4 r^5(\kappa^{\star})^4(\sigma_1^{\star})^2}{n_1 n_2}}\|\mathcal{F}(\mathcal{M})^{(k)}-\mathcal{F}(\mathcal{M}^{\star})^{(k)}\|_{F}\big) \\
&\geq \sum_{k=1}^{n_3}\|\mathcal{F}(\mathcal{M})^{(k)}-\mathcal{F}(\mathcal{M}^{\star})^{(k)}\|_{F}^{2}-\frac{\left(\sigma_r^{\star}\right)^2}{1000} \\
&- \frac{\sigma_r^{\star}}{1000}\|\mathcal{F}(\mathcal{M})^{(k)}-\mathcal{F}(\mathcal{M}^{\star})^{(k)}\|_{F} \\
&\geq \sum_{k=1}^{n_3}0.95\|\mathcal{F}(\mathcal{M})^{(k)}-\mathcal{F}(\mathcal{M}^{\star})^{(k)}\|_{F}^{2}-0.01\sigma_r^{\star}\|\mathcal{F}(\Delta)^{(k)}\|_{F}^{2}.
\end{aligned}
$$

Thus, we get:

$$
\begin{aligned}
&\Delta * \Delta^\top : \mathcal{H} : \Delta * \Delta^\top - 3(\mathcal{N}-\mathcal{N}^\star) : \mathcal{H} : (\mathcal{N}-\mathcal{N}^\star) \\
&= \sum_{k=1}^{n_3} \mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^\top)^{(k)} : \mathcal{F}(\mathcal{H}) : \mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^\top)^{(k)} \\
&- 3\big(\mathcal{F}(\mathcal{N})^{(k)} - \mathcal{F}(\mathcal{N}^\star)^{(k)}\big) : \mathcal{F}(\mathcal{H}) : \big(\mathcal{F}(\mathcal{N})^{(k)} - \mathcal{F}(\mathcal{N}^\star)^{(k)}\big) \\
&\le \sum_{k=1}^{n_3} \|\mathcal{F}(\Delta)^{(k)}\mathcal{F}(\Delta^\top)^{(k)}\|_F^2 + 0.04\sigma_r^\star\|\mathcal{F}(\Delta)^{(k)}\|_F^2 \\
&- \big(0.98\|\mathcal{F}(\mathcal{N})^{(k)} - \mathcal{F}(\mathcal{N}^\star)^{(k)}\|_F^2 - 0.04\,\sigma_r^\star\|\mathcal{F}(\Delta)^{(k)}\|_F^2\big) \\
&\le \sum_{k=1}^{n_3} 0.94\|\mathcal{F}(\mathcal{N})^{(k)} - \mathcal{F}(\mathcal{N}^\star)^{(k)}\|_F^2 + 0.12\,\sigma_r^\star\|\mathcal{F}(\Delta)^{(k)}\|_F^2 \\
&\le \sum_{k=1}^{n_3} -0.3\,\sigma_r^\star\|\mathcal{F}(\Delta)^{(k)}\|_F^2
\end{aligned}
$$

and the proof is completed. □

Now, it remains to bound the regularizer the terms of the (36).

Lemma 18. *By choosing* $\alpha_1^2 = \Theta\big(\frac{\mu r\sigma_1^\star}{n_1}\big), \alpha_2^2 = \Theta\big(\frac{\mu r\sigma_1^\star}{n_2}\big)$ *and*

$$\lambda_1\ \alpha_1^2 \le C_2\ \sigma_r^\star,\ \ \lambda_2\ \alpha_2^2 \le C_2\ \sigma_r^\star,$$

for some positive constant C_2*, we have:*

$$\frac{1}{4}\big[\Delta : \nabla^2 Q(\mathcal{W}) : \Delta - 4\langle \nabla Q(\mathcal{W}), \Delta\rangle\big] \le 0.1\,\sigma_r^\star \sum_{k=1}^{n_3} \|\mathcal{F}(\Delta)^{(k)}\|_F^2.$$

Proof. Define

$$Q_1(\mathcal{U}) = \lambda_1 \sum_{k=1}^{n_3}\sum_{i=1}^{n_1} \big(\|\mathcal{F}(e_i^\top)^{(k)}\mathcal{F}(\mathcal{U})^{(k)}\| - \alpha_1\big)_+^4$$

and

$$Q_2(\mathcal{V}) = \lambda_2 \sum_{k=1}^{n_3}\sum_{j=1}^{n_2} \big(\|\mathcal{F}(e_j^\top)^{(k)}\mathcal{F}(\mathcal{V})^{(k)}\| - \alpha_2\big)_+^4$$

and

$$Q(\mathcal{W}) = Q_1(\mathcal{U}) + Q_2(\mathcal{V}).$$

Proceeding as in the proof of Lemma (13) of the symmetric case, we obtain:

$$\frac{1}{4}\left[\Delta_{\mathcal{U}} : \nabla^2 Q_1(\mathcal{U}) : \Delta_{\mathcal{U}} - 4\langle \nabla Q_1(\mathcal{U}), \Delta_{\mathcal{U}}\rangle\right] \leq 0.1 \sum_{k=1}^{n_3} \sigma_r^\star \|\mathcal{F}(\Delta_{\mathcal{U}})^{(k)}\|_F^2$$

$$\frac{1}{4}\left[\Delta_{\mathcal{V}} : \nabla^2 Q_2(\mathcal{V}) : \Delta_{\mathcal{V}} - 4\langle \nabla Q_2(\mathcal{V}), \Delta_{\mathcal{V}}\rangle\right] \leq 0.1 \sum_{k=1}^{n_3} \sigma_r^\star \|\mathcal{F}(\Delta_{\mathcal{V}})^{(k)}\|_F^2$$

Then, using

$$\|\mathcal{F}(\Delta_{\mathcal{U}})^{(k)}\|_F^2 \leq \|\mathcal{F}(\Delta)^{(k)}\|_F^2$$

and

$$\|\mathcal{F}(\Delta_{\mathcal{V}})^{(k)}\|_F^2 \leq \|\mathcal{F}(\Delta)^{(k)}\|_F^2,$$

we have:

$$\frac{1}{4}\left[\Delta : \nabla^2 Q(\mathcal{W}) : \Delta - 4\langle \nabla Q(\mathcal{W}), \Delta\rangle\right] \leq 0.1\, \sigma_r^\star \sum_{k=1}^{n_3} \|\mathcal{F}(\Delta)^{(k)}\|_F^2.$$

$\square$