

Supplementary Material of “Guaranteed Tensor Recovery Fused Low-rankness and Smoothness”

Hailin Wang, Jiangjun Peng, Wenjin Qin, Jianjun Wang and Deyu Meng, *Member, IEEE*

Abstract

In this supplementary material, we first presents some main preliminaries of the high-order t-SVD framework in Appendix A, then give the detailed analysis and proofs of all theoretical results involved in the main paper, in Appendix B to Appendix G. Besides, we further provide some more experimental results in Appendix H, more extensions to other tasks in Appendix I, and further discussions in Appendix J.

CONTENTS

Appendix A: Preliminaries of High-order t-SVD Framework	2
Appendix B: Proof of Theorem 3	3
B-A Main Preliminaries	3
B-B Proof of Theorem 3	4
Appendix C: Analysis for Remark 1 and Remark 2	5
C-A Main Preliminaries	5
C-B Analysis for Remark 1	6
C-C Analysis for Remark 2	7
Appendix D: Proof of Theorem 4	8
D-A Main Preliminaries	8
D-B Dual Certificates	9
D-C Proof of Theorem 4	11
Appendix E: Proofs of Proposition 1 and Theorem 5	13
E-A Proof of Proposition 1	13
E-B A Key Lemma	14
E-C Proof of Theorem 5	15
Appendix F: Proof of Theorem 6	16
F-A Main Preliminaries	16
F-B Dual Certificates	17
F-C Proof of Theorem 6	18
F-D Proof of Lemma 14	19
F-E Proof of Lemma 15	20
Appendix G: Analysis for Remark 4	22
Appendix H: More Experimental Results	23
H-A Visual Data Inpainting	23
H-B Visual Data Denoising	26
Appendix I: More Experimental Extensions to Other Tasks	26
I-A Hyperspectral Image Anomaly Detection	28
I-B Surveillance Video Foreground/Background Separation	28

H. Wang, J. Peng (co-first author) are with the School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China (email: wanghailin97@163.com, andrew.pengjj@gmail.com). W. Qin and J. Wang are with the School of Mathematics and Statistics, Southwest University, Chongqing, 400715, China (email: qinwenjin2021@163.com, wjj@swu.edu.cn). D. Meng (corresponding author) is with the School of Mathematics and Statistics and Ministry of Education Key Lab of Intelligent Networks and Network Security, Xi'an Jiaotong University, Xian, Shaanxi, China and Macau Institute of Systems Engineering, Macau University of Science and Technology, Macau, China. (email: dymeng@mail.xjtu.edu.cn).

Appendix J: Further Discussions	29
J-A How t-CTV Simultaneously Encodes Two Priors?	30
J-B Evaluating t-CTV along Its Partial Modes	35
J-C Why Better Performance?	36
References	38

APPENDIX A
PRELIMINARIES OF HIGH-ORDER T-SVD FRAMEWORK

TABLE I: Summary of notations and preliminaries of high-order t-SVD framework.

Notations	Descriptions	Notations	Descriptions
$[d] = \{1, 2, \dots, d-1, d\}$	set of the first d natural numbers	$\mathbf{A} \in \mathbb{C}^{n_1 \times n_2}$	matrix
\mathbf{A}^*	conjugate transpose	\mathbf{I}_n	$n \times n$ identity matrix
$\sigma_i(\mathbf{A})$	i -th singular value	$\mathbf{A} \otimes \mathbf{B}$	Kronecker product of \mathbf{A} and \mathbf{B}
$\ \mathbf{A}\ = \max_i \sigma_i(\mathbf{A})$	matrix spectral norm	$\ \mathbf{A}\ _* = \sum_i \sigma_i(\mathbf{A})$	matrix nuclear norm
$\mathcal{A} \in \mathbb{R}^{n_1 \times \dots \times n_d}$	order- d tensor	$\mathcal{A}_{i_1 \dots i_d}$ or $\mathcal{A}(i_1, \dots, i_d)$	(i_1, \dots, i_d) -th entry
$\mathbf{A}_{(k)}$	mode- k unfolding of \mathcal{A}	$\ \mathcal{A}\ _\infty = \max_{i_1 \dots i_d} \mathcal{A}_{i_1 \dots i_d} $	tensor infinity norm
$\ \mathcal{A}\ _1 = \sum_{i_1 \dots i_d} \mathcal{A}_{i_1 \dots i_d} $	tensor ℓ_1 -norm	$\mathcal{A}(:, :, i_3, \dots, i_d)$ or $\mathcal{A}^{(i_3, \dots, i_d)}$	(i_3, \dots, i_d) -th face slice
$\ \mathcal{A}\ _F = (\sum_{i_1 \dots i_d} \mathcal{A}_{i_1 \dots i_d} ^2)^{1/2}$	tensor Frobenius norm	$\langle \mathcal{A}, \mathcal{B} \rangle = \sum \langle \mathcal{A}^{(i_3, \dots, i_d)}, \mathcal{B}^{(i_3, \dots, i_d)} \rangle$	tensor inner product
$\mathfrak{L}(\cdot) : \mathbb{C}^{n_1 \times \dots \times n_d} \rightarrow \mathbb{C}^{n_1 \times \dots \times n_d}$	arbitrary invertible transform	$\mathcal{A} *_{\mathfrak{L}} \mathcal{B}$	transforms \mathfrak{L} based tensor product
$\mathcal{A}^i \in \mathbb{R}^{n_1 \times \dots \times n_{d-1}}$	order- $(d-1)$ tensor constructed by keeping the d -th index of \mathcal{A} fixed at i , i.e., $\mathcal{A}^i := \mathcal{A}(:, \dots, :, i)$.		
\mathbf{A}^k	k -th face slice, $k = i_3 + (i_4 - 1)n_3 + (i_5 - 1)n_3n_4 + \dots + (i_d - 1)\prod_{j=3}^{d-1} n_j$ corresponding to $\mathbf{A}^{(i_3, \dots, i_d)}$.		
$\mathcal{A} \times_n \mathbf{U}$	the mode- n product of tensor \mathcal{A} with matrix \mathbf{U} , $\mathcal{B} = \mathcal{A} \times_n \mathbf{U} \Leftrightarrow \mathbf{B}_{(n)} = \mathbf{U} \cdot \mathbf{A}_{(n)}$.		
$\mathcal{A}_{\mathfrak{L}} \triangleq \mathfrak{L}(\mathcal{A})$	$\mathfrak{L}(\mathcal{A}) = \mathcal{A} \times_3 \mathbf{U}_{n_3} \times_4 \mathbf{U}_{n_4} \dots \times_d \mathbf{U}_{n_d}$, $\mathbf{U}_{n_j} \in \mathbb{C}^{n_j \times n_j}$ denotes an invertible transform matrix.		
$\mathfrak{L}^{-1}(\mathcal{A})$	$\mathfrak{L}^{-1}(\mathcal{A}) = \mathcal{A} \times_d \mathbf{U}_{n_d}^{-1} \times_{d-1} \mathbf{U}_{n_{d-1}}^{-1} \dots \times_3 \mathbf{U}_{n_3}^{-1}$, $\mathfrak{L}^{-1}(\mathfrak{L}(\mathcal{A})) = \mathcal{A}$.		
$\text{circ}(\mathcal{A})$	$\text{circ}(\mathcal{A}) = \begin{bmatrix} \mathcal{A}^1 & \mathcal{A}^{nd} & \mathcal{A}^{nd-1} & \dots & \mathcal{A}^2 \\ \mathcal{A}^2 & \mathcal{A}^1 & \mathcal{A}^{nd} & \dots & \mathcal{A}^3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathcal{A}^{nd} & \mathcal{A}^{nd-1} & \mathcal{A}^{nd-2} & \dots & \mathcal{A}^1 \end{bmatrix} \in \mathbb{R}^{n_1 n_d \times n_2 n_d \times n_3 \times \dots \times n_{d-1}}$ is in block circulant pattern.		
$\text{bcirc}(\mathcal{A})$	a $n_1 \prod_{j=3}^d n_j \times n_2 \prod_{j=3}^d n_j$ block circulant matrix at the base level of the operator $\text{circ}(\mathcal{A})$.		
$\text{unfold}(\mathcal{A})$	$\text{unfold}(\mathcal{A}) = [\mathcal{A}^1, \mathcal{A}^2, \dots, \mathcal{A}^{nd-1}, \mathcal{A}^{nd}]^T \in \mathbb{R}^{n_1 n_d \times n_2 \times \dots \times n_{d-1}}$ by stacking all order- $(d-1)$ tensors \mathcal{A}^i .		
$\text{fold}(\mathcal{A})$	the operation takes $\text{unfold}(\mathcal{A})$ back to order- d tensor form, i.e., $\text{fold}(\text{unfold}(\mathcal{A}), n_d) = \mathcal{A}$.		
$\text{bunfold}(\mathcal{A})$	a $n_1 \prod_{j=3}^d n_j \times n_2$ matrix formed by applying $\text{unfold}(\cdot)$ repeatedly until a block matrix result.		
$\text{bdiag}(\mathcal{A})$	$\text{bdiag}(\mathcal{A}) = \text{diag}(\mathcal{A}^1, \mathcal{A}^2, \dots, \mathcal{A}^{nd}) \in \mathbb{R}^{n_1 n_3 \dots n_d \times n_2 n_3 \dots n_d}$.		
$\mathcal{A} \triangle \mathcal{B}$	face-wise product of two order- d tensor, $\mathcal{C} = \mathcal{A} \triangle \mathcal{B} \Leftrightarrow \text{bdiag}(\mathcal{C}) = \text{bdiag}(\mathcal{A}) \cdot \text{bdiag}(\mathcal{B}) \Leftrightarrow \mathcal{C}^k = \mathbf{A}^k \cdot \mathbf{B}^k$.		

For brevity, we first supplement the basic notations and preliminaries of the high-order t-SVD framework listed in Table I mainly from [1]. More detailed introduction can be found in [1]-[5].

Then we introduce some relevant algebraic properties of t-SVD. For $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times \dots \times n_d}$, $\mathcal{T}_{\mathfrak{L}} := \mathfrak{L} \times_3 \mathbf{U}_{n_3} \times_4 \dots \times_d \mathbf{U}_{n_d}$ is tantamount to $\text{bunfold}(\mathcal{T}_{\mathfrak{L}}) = (\hat{\mathbf{U}} \otimes \mathbf{I}_{n_1}) \cdot \text{bunfold}(\mathcal{T})$, where $\hat{\mathbf{U}} = \mathbf{U}_{n_d} \otimes \mathbf{U}_{n_{d-1}} \otimes \dots \otimes \mathbf{U}_{n_3}$. According to the assumption that the corresponding matrices $\{\mathbf{U}_{n_j}\}_{j=3}^d$ of the linear transform \mathfrak{L} satisfy certain ‘orthogonality’, i.e., $\hat{\mathbf{U}}^* \hat{\mathbf{U}} = \hat{\mathbf{U}} \hat{\mathbf{U}}^* = \ell \cdot \mathbf{I}_{n_3 \dots n_d}$, where $\ell > 0$ is the corresponding constant. Without loss of generality, for $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times \dots \times n_d}$, there have

$$\begin{aligned} \|\mathcal{A}\|_F &= \frac{1}{\sqrt{\ell}} \|\text{bdiag}(\mathcal{A}_{\mathfrak{L}})\|_F, \\ \langle \mathcal{A}, \mathcal{B} \rangle &= \frac{1}{\ell} \langle \text{bdiag}(\mathcal{A}_{\mathfrak{L}}), \text{bdiag}(\mathcal{B}_{\mathfrak{L}}) \rangle. \end{aligned}$$

Besides, from the definition of tensor-tensor product, For $\mathcal{C} = \mathcal{A} *_{\mathfrak{L}} \mathcal{B}$, there has

$$\text{bdiag}(\mathcal{C}_{\mathfrak{L}}) = \text{bdiag}(\mathcal{A}_{\mathfrak{L}}) \cdot \text{bdiag}(\mathcal{B}_{\mathfrak{L}}),$$

which leads to the algorithm of the tensor-tensor product, and also the t-SVD, t-SVT. See Algorithms 1, 2, 3, respectively, in this supplementary material.

Some other notions need to be complemented. The spectral norm of $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$ is defined as

$$\|\mathcal{T}\| := \|\text{bdiag}(\mathcal{T}_{\mathfrak{L}})\|.$$

The t-SVD rank has the following alternative definitions:

$$\text{rank}_{\text{t-SVD}}(\mathcal{T}) = \#\{i : \mathcal{S}(i, i, :, \dots, :) \neq \mathbf{0}\} = \max_{i_3, \dots, i_d} \text{rank}(\mathcal{T}_{\mathfrak{L}}(:, :, i_3, \dots, i_d)).$$

and the TNN has the equal definition:

$$\|\mathcal{T}\|_{*,\mathfrak{L}} = \frac{1}{\ell} \|\text{bdiag}(\mathcal{T}_{\mathfrak{L}})\|_* = \frac{1}{\ell} \sum_{i_3=1}^{n_1} \cdots \sum_{i_d=1}^{n_d} \|\mathcal{T}_{\mathfrak{L}}(:, :, i_3, \dots, i_d)\|_*,$$

where $\|\cdot\|_*$ denotes the matrix nuclear norm. By the way, TNN is proved to be the dual norm of the spectral norm [1].

At last, there have

$$|\langle \mathcal{A}, \mathcal{B} \rangle| = \frac{1}{\ell} |\langle \text{bidag}(\mathcal{A}_{\mathfrak{L}}), \text{bidag}(\mathcal{B}_{\mathfrak{L}}) \rangle| \leq \frac{1}{\ell} \|\text{bidag}(\mathcal{A}_{\mathfrak{L}})\|_{\text{F}} \|\text{bidag}(\mathcal{B}_{\mathfrak{L}})\|_{\text{F}} = \|\mathcal{A}\|_{\text{F}} \|\mathcal{B}\|_{\text{F}},$$

$$\langle \mathcal{A}, \mathcal{B} \rangle = \frac{1}{\ell} \langle \text{bidag}(\mathcal{A}_{\mathfrak{L}}), \text{bidag}(\mathcal{B}_{\mathfrak{L}}) \rangle \leq \frac{1}{\ell} \|\text{bidag}(\mathcal{A}_{\mathfrak{L}})\| \|\text{bidag}(\mathcal{B}_{\mathfrak{L}})\|_* = \|\mathcal{A}\| \|\mathcal{B}\|_{*,\mathfrak{L}},$$

$$\|\mathcal{A} *_{\mathfrak{L}} \mathcal{B}\|_{\text{F}} = \frac{1}{\sqrt{\ell}} \|\text{bdiag}(\mathcal{A}_{\mathfrak{L}}) \cdot \mathcal{B}_{\mathfrak{L}}\|_{\text{F}} \leq \frac{1}{\sqrt{\ell}} \|\text{bdiag}(\mathcal{A}_{\mathfrak{L}})\| \|\text{bdiag}(\mathcal{B}_{\mathfrak{L}})\|_{\text{F}} = \|\mathcal{A}\| \|\mathcal{B}\|_{\text{F}}.$$

We summarize several necessary tensor properties in the following Lemma.

Lemma 1 (Some properties in t-SVD framework). *Under the high-order t-SVD framework with invertible linear transform L , the following properties hold for order- d tensors $\mathcal{T}, \mathcal{A}, \mathcal{B}$ with appropriate shape:*

- 1) $\mathcal{T} = \mathcal{A} *_{\mathfrak{L}} \mathcal{B} \Leftrightarrow \text{bdiag}(\mathcal{T}_{\mathfrak{L}}) = \text{bdiag}(\mathcal{A}_{\mathfrak{L}}) \cdot \text{bdiag}(\mathcal{B}_{\mathfrak{L}})$,
- 2) $\|\mathcal{T}\|_{\text{F}}^2 = \frac{1}{\ell} \|\text{bdiag}(\mathcal{T}_{\mathfrak{L}})\|_{\text{F}}^2$,
- 3) $\langle \mathcal{A}, \mathcal{B} \rangle = \frac{1}{\ell} \langle \text{bdiag}(\mathcal{A}_{\mathfrak{L}}), \text{bdiag}(\mathcal{B}_{\mathfrak{L}}) \rangle$,
- 4) $|\langle \mathcal{A}, \mathcal{B} \rangle| \leq \|\mathcal{A}\|_{\text{F}} \|\mathcal{B}\|_{\text{F}}$,
- 5) $\langle \mathcal{A}, \mathcal{B} \rangle \leq \|\mathcal{A}\| \|\mathcal{B}\|_{*,\mathfrak{L}}$,
- 6) $\|\mathcal{A} *_{\mathfrak{L}} \mathcal{B}\|_{\text{F}} \leq \|\mathcal{A}\| \|\mathcal{B}\|_{\text{F}}$.

Algorithm 1 Order- d product based on any invertible linear transforms [1]

Input: $\mathcal{A} \in \mathbb{R}^{n_1 \times l \times n_3 \times \dots \times n_d}$, $\mathcal{B} \in \mathbb{R}^{l \times n_2 \times n_3 \times \dots \times n_d}$, and the corresponding matrices $\{\mathbf{U}_{n_i}\}_{i=3}^d$ of invertible linear transform L .

Output: $\mathcal{C} = \mathcal{A} *_{\mathfrak{L}} \mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times \dots \times n_d}$.

1. Compute the result of linear transform on \mathcal{A} and \mathcal{B}

$$\mathcal{A}_{\mathfrak{L}} = \mathfrak{L}(\mathcal{A}), \quad \mathcal{B}_{\mathfrak{L}} = \mathfrak{L}(\mathcal{B}).$$

2. Compute each matrix slice of $\mathcal{C}_{\mathfrak{L}}$ by

for $i_3 \in [n_3], \dots, i_d \in [n_d]$ **do**

$$\mathcal{C}_{\mathfrak{L}}(:, :, i_3, \dots, i_d) = \mathcal{A}_{\mathfrak{L}}(:, :, i_3, \dots, i_d) \cdot \mathcal{B}_{\mathfrak{L}}(:, :, i_3, \dots, i_d).$$

end for

3. Compute the result of inverse linear transform on $\mathcal{C}_{\mathfrak{L}}$

$$\mathcal{C} = \mathfrak{L}^{-1}(\mathcal{C}_{\mathfrak{L}}).$$

APPENDIX B PROOF OF THEOREM 3

A. Main Preliminaries

Suppose $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{1 \times 1 \times n_3 \times \dots \times n_d}$ and $\mathcal{C} = \mathcal{A} *_{\mathfrak{L}} \mathcal{B}$, and we have

$$\begin{aligned} \text{bunfold}(\mathcal{C})^T &= \text{bunfold}(\mathcal{A} *_{\mathfrak{L}} \mathcal{B})^T \\ &= \text{bunfold}(\mathcal{B} *_{\mathfrak{L}} \mathcal{A})^T \\ &= ((\hat{\mathbf{U}}^T \otimes I_1) \cdot \text{bdiag}(\mathcal{B}_{\mathfrak{L}}^T) \cdot (\hat{\mathbf{U}} \otimes I_1) \cdot \text{unfold}(\mathcal{A}))^T \\ &= \text{bunfold}(\mathcal{A})^T \cdot \hat{\mathbf{U}}^T \cdot \text{bdiag}(\mathcal{B}_{\mathfrak{L}}^T) \cdot \hat{\mathbf{U}}^{-T} := \text{bunfold}(\mathcal{A})^T \cdot \mathfrak{D}(\mathcal{B}), \end{aligned}$$

where $\mathfrak{D}(\mathcal{B}) := \hat{\mathbf{U}}^T \cdot \text{bdiag}(\mathcal{B}_{\mathfrak{L}}^T) \cdot \hat{\mathbf{U}}^{-T}$ and $\hat{\mathbf{U}}^{-T} := (\hat{\mathbf{U}}^{-1})^T$. Note that $\text{bunfold}(\mathcal{C})^T, \text{bunfold}(\mathcal{A})^T$ are $1 \times n_3 \dots n_4$, and then

$$\text{reshape}(\mathcal{C}, [1, n_3 \dots n_d]) = \text{reshape}(\mathcal{A}, [1, n_3 \dots n_d]) \cdot \mathfrak{D}(\mathcal{B}). \quad (\text{B-1})$$

Algorithm 2 Order- d t-SVD based on any invertible linear transforms [1]

Input: $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times \cdots \times n_d}$, and the corresponding matrices $\{\mathbf{U}_{n_j}\}_{j=3}^d$ of invertible linear transform L .

Output: t-SVD components \mathcal{U} , \mathcal{S} and \mathcal{V} of \mathcal{T} .

1. Compute the result of linear transform on \mathcal{T}

$$\mathcal{T}_\mathcal{L} = \mathcal{L}(\mathcal{T}).$$

2. Compute each matrix slice of $\mathcal{U}_\mathcal{L}$, $\mathcal{S}_\mathcal{L}$ and $\mathcal{V}_\mathcal{L}$ from $\mathcal{T}_\mathcal{L}$ by

for $i_3 \in [n_3], \dots, i_d \in [n_d]$ **do**

$$[\mathbf{U}, \mathbf{S}, \mathbf{V}] = \text{svd}(\mathcal{T}_\mathcal{L}(:, :, i_3, \dots, i_d)),$$

$$\mathcal{U}_\mathcal{L}(:, :, i_3, \dots, i_d) = \mathbf{U}, \mathcal{S}_\mathcal{L}(:, :, i_3, \dots, i_d) = \mathbf{S}, \mathcal{V}_\mathcal{L}(:, :, i_3, \dots, i_d) = \mathbf{V}.$$

end for

3. Compute the result of inverse linear transform on $\mathcal{U}_\mathcal{L}$, $\mathcal{S}_\mathcal{L}$ and $\mathcal{V}_\mathcal{L}$

$$\mathcal{U} = \mathcal{L}^{-1}(\mathcal{U}_\mathcal{L}), \mathcal{S} = \mathcal{L}^{-1}(\mathcal{S}_\mathcal{L}), \mathcal{V} = \mathcal{L}^{-1}(\mathcal{V}_\mathcal{L}).$$

Algorithm 3 Order- d t-SVT based on any invertible linear transforms [1]

Input: $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times \cdots \times n_d}$, $\tau > 0$, and the corresponding matrices $\{\mathbf{U}_{n_i}\}_{i=3}^d$ of invertible linear transform L .

Output: $\mathcal{D} := \text{t-SVT}(\mathcal{T}) = \mathcal{U} *_{\mathcal{L}} \mathcal{S}_\tau *_{\mathcal{L}} \mathcal{V}^\top$.

1. Compute the result of linear transform on \mathcal{T}

$$\mathcal{T}_\mathcal{L} = \mathcal{L}(\mathcal{T}).$$

2. Compute the matrix SVT on each slice of $\mathcal{T}_\mathcal{L}$ by

for $i_3 \in [n_3], \dots, i_d \in [n_d]$ **do**

$$[\mathbf{U}, \mathbf{S}, \mathbf{V}] = \text{svd}(\mathcal{T}_\mathcal{L}(:, :, i_3, \dots, i_d)),$$

$$\mathcal{D}_\mathcal{L}(:, :, i_3, \dots, i_d) = \mathbf{U} \cdot (\mathbf{S} - \tau)_+ \cdot \mathbf{V}^\top.$$

end for

3. Compute the result of inverse linear transform on $\mathcal{D}_\mathcal{L}$

$$\mathcal{D} = \mathcal{L}^{-1}(\mathcal{D}_\mathcal{L}).$$

Let $\hat{\mathcal{A}} \in \mathbb{R}^{n_1 \times 1 \times n_3 \times \cdots \times n_d}$. Consider the product $\hat{\mathcal{C}} = \hat{\mathcal{A}} *_{\mathcal{L}} \mathcal{B}$, since $\hat{\mathcal{C}}(j, 1, :, \dots, :) = \hat{\mathcal{A}}(j, 1, :, \dots, :) *_{\mathcal{L}} \mathcal{B}$, and then based on (B-1), we can have

$$\text{reshape}(\hat{\mathcal{C}}, [n_1, n_3 \cdots n_d]) = \text{reshape}(\hat{\mathcal{A}}, [n_1, n_3 \cdots n_d]) \cdot \mathcal{D}(\mathcal{B}).$$

At last, for $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ with t-SVD rank R , and its t-SVD $\mathcal{T} = \mathcal{U} *_{\mathcal{L}} \mathcal{S} *_{\mathcal{L}} \mathcal{V}^\top$. Denote $\mathcal{C} := \mathcal{S} *_{\mathcal{L}} \mathcal{V}^\top$, and then $\mathcal{T} = \sum_{i=1}^R \mathcal{U}(:, i, :, \dots, :) *_{\mathcal{L}} \mathcal{C}(i, :, :, \dots, :)$. We can then get that

$$\text{reshape}(\mathcal{T}(:, j, :, \dots, :), [n_1, n_3 \cdots n_d]) = \sum_{i=1}^R \text{reshape}(\mathcal{U}(:, j, :, \dots, :), [n_1, n_3 \cdots n_d]) \cdot \mathcal{D}(\mathcal{C}(i, j, :, \dots, :)) \quad (\text{B-2})$$

for any $j = 1, \dots, R$.

B. Proof of Theorem 3

Theorem 3. (optimality principle) Given the t-SVD of $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ by $\mathcal{T} = \mathcal{U} *_{\mathcal{L}} \mathcal{S} *_{\mathcal{L}} \mathcal{V}^\top$, then, $\mathcal{T}_k := \sum_{i=1}^k \mathcal{U}(:, i, :, \dots, :) *_{\mathcal{L}} \mathcal{S}(i, i, :, \dots, :) *_{\mathcal{L}} \mathcal{V}(:, i, :, \dots, :)^\top$ is the best Frobenius norm approximation over all t-SVD rank- k tensor. Moreover, suppose \mathbf{M} is an unfolding matrix of \mathcal{T} , we have,

$$\text{rank}_{\text{t-SVD}}(\mathcal{T}) \leq \text{rank}(\mathbf{M}), \quad (\text{B-3})$$

and

$$\|\mathcal{T} - \mathcal{T}_k\|_\text{F} \leq \|\mathbf{M} - \mathbf{M}_k\|_\text{F}, \quad (\text{B-4})$$

where \mathbf{M}_k is the rank- k approximation of the matrix \mathbf{M} .

Proof: Using t-SVD theorem, i.e., Theorem 1, it is easy to know that any t-SVD rank- k order- d \mathcal{X} can be decomposed as $\mathcal{X} = \mathcal{A} *_{\mathcal{L}} \mathcal{B}$, where $\mathcal{A} \in \mathbb{R}^{n_1 \times k \times n_3 \times \cdots \times n_d}$, $\mathcal{B} \in \mathbb{R}^{k \times n_2 \times n_3 \times \cdots \times n_d}$. Then

$$\|\mathcal{T} - \mathcal{X}\|_\text{F}^2 = \frac{1}{\ell} \|\text{bdiag}(\mathcal{T}_\mathcal{L}) - \text{bdiag}(\mathcal{X}_\mathcal{L})\|_\text{F}^2 = \frac{1}{\ell} \sum_{i_3=1}^{n_3} \cdots \sum_{i_d=1}^{n_d} \|\mathcal{T}_\mathcal{L}(:, :, i_3, \dots, i_d) - \mathcal{X}_\mathcal{L}(:, :, i_3, \dots, i_d)\|_\text{F}^2.$$

By definition, $\mathcal{X}_\mathcal{L}(:, :, i_3, \dots, i_d)$ is a rank- k outer product by $\mathcal{A}_\mathcal{L}(:, :, i_3, \dots, i_d) \cdot \mathcal{B}_\mathcal{L}(:, :, i_3, \dots, i_d)$. Note that the best rank- k approximation to $\mathcal{T}_\mathcal{L}(:, :, i_3, \dots, i_d)$ is $\sum_{i=1}^k \mathcal{U}_\mathcal{L}(:, i, i_3, \dots, i_d) \cdot \mathcal{S}_\mathcal{L}(i, i_3, \dots, i_d) \cdot \mathcal{V}_\mathcal{L}(:, i, i_3, \dots, i_d)^\top$, therefore

$$\|\mathcal{T}_{\mathcal{L}}(:, :, i_3, \dots, i_d) - \sum_{i=1}^k \mathcal{U}_{\mathcal{L}}(:, i, i_3, \dots, i_d) \cdot \mathcal{S}_{\mathcal{L}}(i, i, i_3, \dots, i_d) \cdot \mathcal{V}_{\mathcal{L}}(:, i, i_3, \dots, i_d)^T\|_F^2 \leq \|\mathcal{T}_{\mathcal{L}}(:, :, i_3, \dots, i_d) - \mathcal{X}_{\mathcal{L}}(:, :, i_3, \dots, i_d)\|_F^2,$$

Without loss of generality, let $M = \text{bunfold}(\mathcal{X}) \in \mathbb{R}^{n_1 n_3 \cdots n_d \times n_2}$ be the unfolding matrix along with the mode-2 of \mathcal{T} . For other unfolding matrices, the following analysis is the same. Suppose the t-SVD rank of tensor \mathcal{T} equals R , and the rank of matrix M equals r . Let $M = AB^T$ be its a rank- r factorization, where A is of size $n_1 n_3 \cdots n_d \times r$ and B is $n_2 \times r$. By $\mathcal{T}_{\mathcal{L}} = \mathcal{T} \times_3 U_{n_3} \times_4 \cdots \times_d U_{n_d}$, its (j_3, \dots, j_d) -th slice has the following construction,

$$\begin{aligned} \mathcal{T}_{\mathcal{L}}(:, :, j_3, \dots, j_d) &= \sum_{i_3=1}^{j_3} \cdots \sum_{i_d=1}^{j_d} U_{n_3}(j_3, i_3) \cdots U_{n_d}(j_d, i_d) \cdot \mathcal{T}(:, :, i_3, \dots, i_d) \\ &= \sum_{i_3=1}^{j_3} \cdots \sum_{i_d=1}^{j_d} U_{n_3}(j_3, i_3) \cdots U_{n_d}(j_d, i_d) \cdot (A((j-1)n_1 : jn_1, :) \cdot B^T) \\ &= \left(\sum_{i_3=1}^{j_3} \cdots \sum_{i_d=1}^{j_d} U_{n_3}(j_3, i_3) \cdots U_{n_d}(j_d, i_d) A((j-1)n_1 : jn_1, :) \right) \cdot B^T, \end{aligned}$$

where $j = j_3 + (j_4 - 1)n_3 + \cdots + (j_d - 1)n_3 \cdots n_d$. Such factorization form of $\mathcal{T}_{\mathcal{L}}(:, :, j_3, \dots, j_d)$ means that its rank is bounded by $\min(n_1, r)$ for any j_3, \dots, j_d . Note that

$$\text{rank}_{\text{t-SVD}}(\mathcal{T}) = \max_{i_3, \dots, i_d} \text{rank}(\mathcal{T}_{\mathcal{L}}(:, :, i_3, \dots, i_d)).$$

Thus,

$$R = \text{rank}_{\text{t-SVD}}(\mathcal{T}) = \max_{i_3, \dots, i_d} \text{rank}(\mathcal{T}_{\mathcal{L}}(:, :, i_3, \dots, i_d)) \leq \min(n_1, r) \leq r. \quad (\text{B-5})$$

Let $M = USV^T$ be the matrix SVD of M . Denote $C = SV^T$, and then the rank- k approximation of M_k can be expressed as $M_k = U(:, 1:k) \cdot C(1:k, :)$. Then $M_k(:, j) = \sum_{i=1}^k c_{ij} U(:, i)$ for any $j = 1, \dots, k$, where $c_{ij} = C(i, j)$. Then

$$\text{reshape}(M_k(:, j), [n_1, n_3 \cdots n_d]) = \sum_{i=1}^k c_{ij} \text{reshape}(U(:, i), [n_1, n_3 \cdots n_d]). \quad (\text{B-6})$$

Note that $\frac{1}{\sqrt{\ell}} \hat{U}$ is unitary for the invertible linear transform, and then we have $c_{ij} I_{n_3 \cdots n_d} = \hat{U}^T \text{diag}(c_{ij} \mathbf{e}) \hat{U}^{-T}$, where \mathbf{e} is a vector of all ones with length $n_3 \cdots n_d$. Denote $\mathcal{C} \in \mathbb{R}^{k \times n_2 \times \cdots \times n_d}$ with $\mathcal{C}(i, j, :, \dots, :) := \text{reshape}(c_{ij} \hat{U}^{-1} \mathbf{e}, [1, 1, n_3, \dots, n_d])$. Then

$$c_{ij} I_{n_3 \cdots n_d} = \hat{U}^T \text{diag}(c_{ij} \mathbf{e}) \hat{U}^{-T} = \hat{U}^T \text{bdiag}(\mathcal{L}(\mathcal{C}(i, j, :, \dots, :))^T) \hat{U}^{-T} = \mathcal{D}(\mathcal{C}(i, j, :, \dots, :))$$

Thus (B-6) has the equal form

$$\text{reshape}(M_k(:, j), [n_1, n_3 \cdots n_d]) = \sum_{i=1}^k \text{reshape}(U(:, i), [n_1, n_3 \cdots n_d]) \cdot \mathcal{D}(\mathcal{C}(i, j, :, \dots, :)).$$

Based on (B-2), the above relationship can be combined into a tensor format equivalently,

$$\mathcal{Z}_k := \sum_{i=1}^k \mathcal{Q}(:, i, :, \dots, :) *_{\mathcal{L}} \mathcal{C}(i, :, :, \dots, :) = \mathcal{Q} *_{\mathcal{L}} \mathcal{C},$$

where

$$\mathcal{Z}(:, i, :, \dots, :) = \text{reshape}(M_k(:, i), [n_1, 1, n_3, \dots, n_d]) \quad \text{and} \quad \mathcal{Q}(:, i, :, \dots, :) = \text{reshape}(U(:, i), [n_1, 1, n_3, \dots, n_d]).$$

Since \mathcal{T}_k is the best Frobenius norm approximation over all t-SVD rank k tensors, then

$$\|\mathcal{T} - \mathcal{T}_k\|_F^2 \leq \|\mathcal{T} - \mathcal{Z}_k\|_F^2 = \|M - M_k\|_F^2. \quad (\text{B-7})$$

The proof is completed. ■

APPENDIX C ANALYSIS FOR REMARK 1 AND REMARK 2

A. Main Preliminaries

Given $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, and $M \in \mathbb{R}^{m \times n_k}$, $\mathcal{Y} := \mathcal{T} \times_k M \in \mathbb{R}^{n_1 \times \cdots \times n_{k-1} \times m \times n_{k+1} \times \cdots \times n_d}$ with

$$\mathcal{Y}(i_1, \dots, i_{k-1}, :, i_{k+1}, \dots, i_d) = M \cdot \mathcal{T}(i_1, \dots, i_{k-1}, :, i_{k+1}, \dots, i_d). \quad (\text{C-1})$$

For two matrices $A \in \mathbb{R}^{a \times n_i}$ and $B \in \mathbb{R}^{b \times n_j}$, where $i \neq j$, it holds that

$$(\mathcal{T} \times_i A) \times_j B = (\mathcal{T} \times_j B) \times_i A. \quad (\text{C-2})$$

If $i = j$, i.e., $A \in \mathbb{R}^{a \times n_i}$ and $B \in \mathbb{R}^{b \times n_i}$, then

$$(\mathcal{T} \times_i A) \times_i B = \mathcal{T} \times_i (BA), \quad (\text{C-3})$$

where $a = n_i$ is required for the shape of A .

B. Analysis for Remark 1

Proposition 2. For $\mathcal{T} \in \mathbb{R}^{n_1 \times \dots \times n_d}$, \mathcal{G}_k is its gradient tensor along the k -th mode. Then we have, \mathcal{G}_k is linear with \mathcal{T} by the tensor-tensor product, i.e., there exists the corresponding tensor \mathcal{D}_k such that $\mathcal{G}_k = \mathcal{D}_k *_{\mathfrak{L}} \mathcal{T}$. Further, it yields that

$$R - 1 \leq \text{rank}_{t\text{-SVD}}(\mathcal{G}_k) \leq R, \quad (\text{C-4})$$

where R denotes the t -SVD rank of \mathcal{T} .

Proof: For the first result, based on the definition of gradient tensor, $\mathcal{G}_k = \nabla_k(\mathcal{T}) = \mathcal{T} \times_k D_{n_k}$, $k = 1, \dots, d$, where D_{n_k} sized $n_k \times n_k$ is a row circulant matrix of $(-1/2, 1/2, 0, \dots, 0)$. We consider the t -SVD rank of \mathcal{G}_k in its transform domain, i.e., $\text{rank}_{t\text{-SVD}}(\mathcal{G}_k) = \max_{i_3, \dots, i_d} \text{rank}(\mathcal{L}(\mathcal{G}_k)(:, :, i_3, \dots, i_d))$. In the transform domain,

$$\mathcal{L}(\mathcal{G}_k) = \mathcal{L}(\mathcal{T} \times_k D_{n_k}) = (\mathcal{T} \times_k D_{n_k}) \times_3 U_{n_3} \times_4 \dots \times_d U_{n_d}. \quad (\text{C-5})$$

Case 1: $k = 1, 2$. Using (C-2), (C-5), we have

$$\mathcal{L}(\mathcal{G}_k) = (\mathcal{T} \times_3 U_{n_3} \times_4 \dots \times_d U_{n_d}) \times_k D_{n_k} = \mathcal{L}(\mathcal{T}) \times_k D_{n_k}.$$

Then, for any $i_2 \in [n_2], i_3 \in [n_3], \dots, i_d \in [n_d]$, $\mathcal{L}(\mathcal{G}_1)(:, i_2, i_3, \dots, i_d) = D_{n_1} \cdot \mathcal{L}(\mathcal{T})(:, i_2, i_3, \dots, i_d)$ from (C-1), which leads to the result that each slice of $\mathcal{L}(\mathcal{G}_1)$ satisfies

$$\mathcal{L}(\mathcal{G}_1)(:, :, i_3, \dots, i_d) = D_{n_1} \cdot \mathcal{L}(\mathcal{T})(:, :, i_3, \dots, i_d). \quad (\text{C-6})$$

Thus

$$\begin{aligned} \text{rank}_{t\text{-SVD}}(\mathcal{G}_1) &= \max_{i_3, \dots, i_d} \text{rank}(\mathcal{L}(\mathcal{G}_1)(:, :, i_3, \dots, i_d)) \\ &= \max_{i_3, \dots, i_d} \text{rank}(D_{n_1} \cdot \mathcal{L}(\mathcal{T})(:, :, i_3, \dots, i_d)) \\ &\stackrel{(a)}{\geq} \max_{i_3, \dots, i_d} \{\text{rank}(D_{n_1}) + \text{rank}(\mathcal{L}(\mathcal{T})(:, :, i_3, \dots, i_d)) - n_1\} \\ &\stackrel{(b)}{=} \max_{i_3, \dots, i_d} \{\text{rank}(\mathcal{L}(\mathcal{T})(:, :, i_3, \dots, i_d)) - 1\} \\ &= \text{rank}_{t\text{-SVD}}(\mathcal{T}) - 1 = R - 1, \end{aligned}$$

where (a) is based on the Sylvester rank inequality [6], and (b) follows the fact that $\text{rank}(D_{n_k}) = n_k - 1$. Besides,

$$\begin{aligned} \text{rank}_{t\text{-SVD}}(\mathcal{G}_1) &= \max_{i_3, \dots, i_d} \text{rank}(\mathcal{L}(\mathcal{G}_1)(:, :, i_3, \dots, i_d)) \\ &= \max_{i_3, \dots, i_d} \text{rank}(D_{n_1} \cdot \mathcal{L}(\mathcal{T})(:, :, i_3, \dots, i_d)) \\ &\leq \max_{i_3, \dots, i_d} \min\{\text{rank}(D_{n_1}), \text{rank}(\mathcal{L}(\mathcal{T})(:, :, i_3, \dots, i_d))\} \\ &\leq \max_{i_3, \dots, i_d} \text{rank}(\mathcal{L}(\mathcal{T})(:, :, i_3, \dots, i_d)) \\ &= \text{rank}_{t\text{-SVD}}(\mathcal{T}) = R. \end{aligned}$$

In the same way, $R - 1 \leq \text{rank}_{t\text{-SVD}}(\mathcal{G}_2) \leq R$, where the D_{n_2} is multiplied on the right in (C-6).

Case 2: $k = 3, \dots, d$. In this case, using (C-2), (C-3), and (C-5), we have

$$\begin{aligned} \mathcal{L}(\mathcal{G}_k) &= (\mathcal{T} \times_3 \dots \times_{k-1} U_{n_{k-1}} \times_{k+1} \dots \times_d U_{n_d}) \times_k D_{n_k} \times_k U_{n_k} \\ &= (\mathcal{T} \times_3 \dots \times_{k-1} U_{n_{k-1}} \times_{k+1} \dots \times_d U_{n_d}) \times_k (U_{n_k} D_{n_k}) \\ &= (\mathcal{T} \times_3 \dots \times_d U_{n_d} \times_k U_{n_k}^{-1}) \times_k (U_{n_k} D_{n_k}) \\ &= \mathcal{L}(\mathcal{T}) \times_k (U_{n_k} D_{n_k} U_{n_k}^{-1}). \end{aligned}$$

Denote $\bar{D}_{n_k} := U_{n_k} D_{n_k} U_{n_k}^{-1}$, \bar{D}_{n_k} to be a diagonal matrix since the transform matrix U_{n_k} can make the circulant matrix D_{n_k}

diagonalizable, and $\text{rank}(\bar{\mathbf{D}}_{n_k}) = \text{rank}(\mathbf{D}_{n_k}) = n_k - 1$. Then, based on (C-1),

$$\mathcal{L}(\mathcal{G}_k)(i_1, \dots, i_{k-1}, :, i_{k+1}, \dots, i_d) = \bar{\mathbf{D}}_{n_k} \cdot \mathcal{L}(\mathcal{T})(i_1, \dots, i_{k-1}, :, i_{k+1}, \dots, i_d), \quad (\text{C-7})$$

which can further results in the fact that each slice of $\mathcal{L}(\mathcal{G}_k)$ satisfies

$$\mathcal{L}(\mathcal{G}_k)(:, :, i_3, \dots, i_k, \dots, i_d) = \sum_{j_k=1}^{n_k} \bar{\mathbf{D}}_{n_k}(i_k, j_k) \cdot \mathcal{L}(\mathcal{T})(:, :, i_3, \dots, j_k, \dots, i_d) = \bar{\mathbf{D}}_{n_k}(i_k, i_k) \cdot \mathcal{L}(\mathcal{T})(:, :, i_3, \dots, i_k, \dots, i_d), \quad (\text{C-8})$$

where the second equation uses $\bar{\mathbf{D}}_{n_k}(i_k, j_k) = 0, i_k \neq j_k$. Thus each slice of $\mathcal{L}(\mathcal{G}_k)$ is linear with that of $\mathcal{L}(\mathcal{T})$ (possibly linear with complex multiplication when the transform \mathcal{L} is complex). Thus

$$\text{rank}_{\text{t-SVD}}(\mathcal{G}_k) = \max_{i_3, \dots, i_d} \text{rank}(\mathcal{L}(\mathcal{G}_k)(:, :, i_3, \dots, i_d)) = \max_{i_3, \dots, i_d} \text{rank}(\mathcal{L}(\mathcal{T})(:, :, i_3, \dots, i_d)) = \text{rank}_{\text{t-SVD}}(\mathcal{X}) = R.$$

At last, based on (C-6) and (C-8), it's easy to get that there must exist the corresponding \mathcal{D}_k such that (C-6) and (C-8) can be further written as a form as $\text{bdiag}(\mathcal{L}(\mathcal{G}_k)) = \text{bdiag}(\mathcal{L}(\mathcal{D}_k)) *_{\mathcal{L}} \text{bdiag}(\mathcal{L}(\mathcal{T}))$. It means that the difference operation is equal to a tensor-tensor product $\mathcal{G}_k = \mathcal{D}_k *_{\mathcal{L}} \mathcal{T}$. For examples, consider $d = 3$ and $\mathbf{U}_3 = F_3$ is the 3×3 DFT transform matrix, we can take \mathcal{D}_1 is a tensor of size $n_1 \times n_1 \times 3$, whose first slice $\mathcal{D}_1(:, :, 1) = \mathbf{D}_{n_1}$ and all others are zeros, and then $\mathcal{L}(\mathcal{D}_1)$'s all slices are exactly equal to \mathcal{D}_{n_1} . Then we have

$$\begin{bmatrix} \mathcal{L}(\mathcal{G}_1)(:, :, 1) \\ \mathcal{L}(\mathcal{G}_1)(:, :, 2) \\ \mathcal{L}(\mathcal{G}_1)(:, :, 3) \end{bmatrix} = \begin{bmatrix} \mathbf{D}_{n_1} & & \\ & \mathbf{D}_{n_1} & \\ & & \mathbf{D}_{n_1} \end{bmatrix} \cdot \begin{bmatrix} \mathcal{L}(\mathcal{T})(:, :, 1) \\ \mathcal{L}(\mathcal{T})(:, :, 2) \\ \mathcal{L}(\mathcal{T})(:, :, 3) \end{bmatrix}, \quad (\text{C-9})$$

which is equal to (C-6). This follows that $\text{bdiag}(\mathcal{L}(\mathcal{G}_1)) = \text{bdiag}(\mathcal{L}(\mathcal{D}_1)) *_{\mathcal{L}} \text{bdiag}(\mathcal{L}(\mathcal{T}))$, and thus $\mathcal{G}_1 = \mathcal{T} \times_1 \mathbf{D}_{n_1} = \mathcal{D}_1 *_{\mathcal{L}} \mathcal{T}$. In the same way, we can construct \mathcal{D}_2 such that $\mathcal{G}_2 = \mathcal{D}_2 *_{\mathcal{L}} \mathcal{T}$. In addition, for the difference operation along the third dimension, we can take \mathcal{D}_3 as a tensor of size $n_1 \times n_1 \times 3$ constructed by $\mathcal{D}_3(i, i, :) = F_3^{-1} \cdot \text{diag}(\bar{\mathbf{D}}_3)$ and all other entries are zeros. In the same way, one can verify that $\mathcal{G}_3 = \mathcal{T} \times_3 \mathbf{D}_3 = \mathcal{D}_3 *_{\mathcal{L}} \mathcal{T}$. Note that the constructed difference operation tensors also have circulant structure like in the low dimension cases. It should be indicated that such strict linear transform using the tensor-tensor product is very important, which will be used frequently throughout our theoretical analysis and the algorithm implementation. ■

C. Analysis for Remark 2

Proposition 3. For an order- d tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times \dots \times n_d}$ with t-SVD rank R , it holds that

$$\|\mathcal{T}\|_{\text{TV}} \lesssim \|\mathcal{T}\|_{\text{t-CTV}} \lesssim \sqrt{R} \|\mathcal{T}\|_{\text{TV}}. \quad (\text{C-10})$$

Proof: The proof mainly uses some properties under the t-SVD framework and some inequalities between matrix norms. On the one hand, for any $k \in \Gamma$,

$$\|\mathcal{G}_k\|_{\circledast, \mathcal{L}} = \frac{1}{\ell} \|\text{bdiag}(\mathcal{L}(\mathcal{G}_k))\|_* \geq \frac{1}{\ell} \|\text{bdiag}(\mathcal{L}(\mathcal{G}_k))\|_{\text{F}} = \frac{1}{\sqrt{\ell}} \|\mathcal{G}_k\|_{\text{F}} \geq \frac{1}{\sqrt{\ell}} \frac{1}{\sqrt{n_1 \cdots n_d}} \|\mathcal{G}_k\|_1. \quad (\text{C-11})$$

On the other hand,

$$\begin{aligned} \|\mathcal{G}_k\|_{\circledast, \mathcal{L}} &= \frac{1}{\ell} \|\text{bdiag}(\mathcal{L}(\mathcal{G}_k))\|_* \\ &\leq \frac{1}{\ell} \sqrt{\text{rank}_{\text{t-SVD}}(\mathcal{G}_k)} \sqrt{n_3 \cdots n_d} \|\text{bdiag}(\mathcal{L}(\mathcal{G}_k))\|_{\text{F}} \\ &= \frac{1}{\sqrt{\ell}} \sqrt{\text{rank}_{\text{t-SVD}}(\mathcal{G}_k)} \sqrt{n_3 \cdots n_d} \|\mathcal{G}_k\|_{\text{F}} \\ &\leq \frac{1}{\sqrt{\ell}} \sqrt{R} \sqrt{n_3 \cdots n_d} \|\mathcal{G}_k\|_{\text{F}} \leq \frac{1}{\sqrt{\ell}} \sqrt{R} \sqrt{n_3 \cdots n_d} \|\mathcal{G}_k\|_1. \end{aligned} \quad (\text{C-12})$$

Combine (C-11), (C-12), we have

$$\|\mathcal{G}_k\|_{\text{F}} \lesssim \|\mathcal{G}_k\|_{\circledast, \mathcal{L}} \lesssim \sqrt{R} \|\mathcal{G}_k\|_{\text{F}}, \text{ and } \|\mathcal{G}_k\|_1 \lesssim \|\mathcal{G}_k\|_{\circledast, \mathcal{L}} \lesssim \sqrt{R} \|\mathcal{G}_k\|_1.$$

Recall the definitions of TV and t-CTV norms, and the result is obtained. The proof is completed. ■

APPENDIX D
PROOF OF THEOREM 4

A. Main Preliminaries

First, for $\mathcal{T} \in \mathbb{R}^{n_1 \times \dots \times n_d}$, and $\mathcal{G}_k = \nabla_k(\mathcal{T}) = \mathcal{T} \times_k D_{n_k}$, $k = 1, 2, \dots, n$, the difference operation $\nabla_k(\cdot)$ is a linear operator and its spectral norm $\|\nabla_k\| := \sup_{\|\mathcal{T}\|_F \leq 1} \|\nabla_k(\mathcal{T})\|_F \leq 1^1$. Note that it has been proven that there always exists corresponding difference operation tensor D_k such that $\nabla_k(\mathcal{T}) = D_k *_{\mathfrak{L}} \mathcal{T}$. Denote $\nabla_k^T(\mathcal{T}) := \mathcal{T} \times_k D_{n_k}^T$ as the transpose operator of $\nabla_k(\cdot)$. One can verify that its corresponding difference operation tensor equals D_k^T , i.e., $\nabla_k^T(\mathcal{T}) = D_k^T *_{\mathfrak{L}} \mathcal{T}$. The following property will be used frequently,

$$\langle \mathcal{A}, \nabla_k(\mathcal{B}) \rangle = \langle \nabla_k^T(\mathcal{A}), \mathcal{B} \rangle. \quad (\text{D-1})$$

In addition, $\nabla_k^T(\cdot)$'s spectral norm is also bounded by one, i.e., $\|\nabla_k^T\| \leq 1$.

The $L_{\infty,2}$ norm of the tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times \dots \times n_d}$ is defined as

$$\|\mathcal{T}\|_{\infty,2} = \max \left\{ \max_{i_1} \|\mathcal{T}(i_1, :, :, \dots, :) \|_F, \max_{i_2} \|\mathcal{T}(:, i_2, :, \dots, :) \|_F \right\}. \quad (\text{D-2})$$

Given $\mathcal{T} \in \mathbb{R}^{n_1 \times \dots \times n_d}$ with t-SVD rank R , and its skinny t-SVD $\mathcal{T} = \mathcal{U} *_{\mathfrak{L}} \mathcal{S} *_{\mathfrak{L}} \mathcal{V}^T$, where $\mathcal{U} \in \mathbb{R}^{n_1 \times R \times n_3 \times \dots \times n_d}$ and $\mathcal{V} \in \mathbb{R}^{n_2 \times R \times n_3 \times \dots \times n_d}$ satisfy $\mathcal{U}^T *_{\mathfrak{L}} \mathcal{U} = \mathcal{V}^T *_{\mathfrak{L}} \mathcal{V} = \mathcal{I}_R$, and $\mathcal{S} \in \mathbb{R}^{R \times R \times n_3 \times \dots \times n_d}$ is a f-diagonal tensor. The subdifferential (the set of subgradient) of TNN $\|\cdot\|_{\circledast, \mathfrak{L}}$ at \mathcal{T} is given as [1],

$$\partial \|\mathcal{T}\|_{\circledast, \mathfrak{L}} = \{\mathcal{U} *_{\mathfrak{L}} \mathcal{V}^T + \mathcal{W} | \mathcal{U}^T *_{\mathfrak{L}} \mathcal{W} = \mathcal{O}, \mathcal{W} *_{\mathfrak{L}} \mathcal{V}^T = \mathcal{O}, \|\mathcal{W}\| \leq 1\}, \quad (\text{D-3})$$

where \mathcal{O} is a zero tensor. Denote \mathbb{T} as the linear space of \mathcal{U}, \mathcal{V} , i.e.,

$$\mathbb{T} := \{\mathcal{U} *_{\mathfrak{L}} \mathcal{X}^T + \mathcal{Y} *_{\mathfrak{L}} \mathcal{V}^T | \mathcal{X} \in \mathbb{R}^{n_1 \times R \times n_3 \times \dots \times n_d}, \mathcal{Y} \in \mathbb{R}^{n_2 \times R \times n_3 \times \dots \times n_d}\}, \quad (\text{D-4})$$

and \mathbb{T}^\perp is its orthogonal complement. For any $\mathcal{Z} \in \mathbb{R}^{n_1 \times \dots \times n_d}$, the projections onto \mathbb{T} and \mathbb{T}^\perp are respectively,

$$\mathcal{P}_{\mathbb{T}}(\mathcal{Z}) = \mathcal{U} *_{\mathfrak{L}} \mathcal{U}^T *_{\mathfrak{L}} \mathcal{Z} + \mathcal{Z} *_{\mathfrak{L}} \mathcal{V} *_{\mathfrak{L}} \mathcal{V}^T - \mathcal{U} *_{\mathfrak{L}} \mathcal{U}^T *_{\mathfrak{L}} \mathcal{Z} *_{\mathfrak{L}} \mathcal{V} *_{\mathfrak{L}} \mathcal{V}_k^T, \quad (\text{D-5})$$

$$\mathcal{P}_{\mathbb{T}^\perp}(\mathcal{Z}) = \mathcal{Z} - \mathcal{P}_{\mathbb{T}}(\mathcal{Z}) = (\mathcal{I}_{n_1} - \mathcal{U} *_{\mathfrak{L}} \mathcal{U}^T) *_{\mathfrak{L}} \mathcal{Z} *_{\mathfrak{L}} (\mathcal{I}_{n_2} - \mathcal{V} *_{\mathfrak{L}} \mathcal{V}^T). \quad (\text{D-6})$$

For the above definition, it's easy to verify that $\mathcal{P}_{\mathbb{T}}$ is self-adjoint, i.e.,

$$\langle \mathcal{P}_{\mathbb{T}}(\mathcal{A}), \mathcal{B} \rangle = \langle \mathcal{A}, \mathcal{P}_{\mathbb{T}}(\mathcal{B}) \rangle = \langle \mathcal{P}_{\mathbb{T}}(\mathcal{A}), \mathcal{P}_{\mathbb{T}}(\mathcal{B}) \rangle, \quad \forall \mathcal{A}, \mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times n_3}, \quad (\text{D-7})$$

and $\|\mathcal{P}_{\mathbb{T}}\| = 1$. $\mathcal{P}_{\mathbb{T}^\perp}$ is the same, which will be used frequently.

Let $\mathcal{T}_0 \in \mathbb{R}^{n_1 \times \dots \times n_d}$ be with t-SVD rank R and it satisfies the standard gradient tensor incoherence conditions (17)-(19) (see Definition 10 in the main paper). Denote the corresponding set \mathbb{T}_k of $\mathcal{G}_k = \nabla_k(\mathcal{T})$, $k \in \Gamma$ like (D-4), i.e.,

$$\mathbb{T}_k := \{\mathcal{U}_k *_{\mathfrak{L}} \mathcal{X}^T + \mathcal{Y} *_{\mathfrak{L}} \mathcal{V}_k^T | \mathcal{X} \in \mathbb{R}^{n_1 \times R \times n_3 \times \dots \times n_d}, \mathcal{Y} \in \mathbb{R}^{n_2 \times R \times n_3 \times \dots \times n_d}\}, \quad (\text{D-8})$$

where $\mathcal{U}_k, \mathcal{V}_k$ from the t-SVD form $\mathcal{G}_k = \mathcal{U}_k *_{\mathfrak{L}} \mathcal{S}_k *_{\mathfrak{L}} \mathcal{V}_k^T$. Based on the incoherence conditions, we then have

$$\begin{aligned} \|\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T\|_\infty &= \max_{i_1, i_2} \left\| \sum_{j=1}^R \mathcal{U}_k(i_1, j, :, \dots, :) *_{\mathfrak{L}} \mathcal{V}_k(i_2, j, :, \dots, :)^T \right\|_\infty \\ &\leq \max_{i_1, i_2} \sum_{j=1}^R \|\mathcal{U}_k(i_1, j, :, \dots, :)\|_F \|\mathcal{V}_k(i_2, j, :, \dots, :)\|_F \\ &\leq \max_{i_1, i_2} \frac{1}{2} \sum_{j=1}^R (\|\mathcal{U}_k(i_1, j, :, \dots, :)\|_F^2 + \|\mathcal{V}_k(i_2, j, :, \dots, :)\|_F^2) \\ &= \frac{1}{2} (\max_{i_1} \|\mathcal{U}_k^T *_{\mathfrak{L}} \mathring{\mathbf{e}}_1^{(i_1)}\|_F^2 + \max_{i_2} \|\mathcal{V}_k^T *_{\mathfrak{L}} \mathring{\mathbf{e}}_2^{(i_2)}\|_F^2) \\ &\leq \frac{\mu R}{n_{(2)} \ell}. \end{aligned} \quad (\text{D-9})$$

Note that in the proof of Theorem 4, we do not need the joint incoherence condition (19) given in Definition 10. Directly using the bound in (D-9) is enough.

¹For simplicity, we add a normalization factor $\frac{1}{2}$ in the difference operator throughout the whole proof, i.e., let D_{n_k} be a row circulant matrix of $(-1/2, 1/2, 0, \dots, 0)$. Thus $\|\nabla_k\| \leq 1$. Such setting is convenient for subsequent proof and does not affect the optimization.

B. Dual Certificates

The proof of Theorem 4 needs to use a key variable named dual certification, which is a commonly adopted proof strategy [7]. We present the dual certificate conditions in the section and prove how to construct it via the Golfing scheme [8]. Then the proof of Theorem 4 can be deduced in the next subsection. First, we introduce some lemmas, which are straight results from Lemmas IV.7, IV.8, and IV.9 in [1] since the gradient standard incoherence condition is equivalent to that the gradient tensors \mathcal{G}_k satisfy the standard incoherence condition.

Lemma 2 (Lemma IV.7 in [1]). *Denote the corresponding set \mathbb{T}_k of $\mathcal{G}_k = \nabla_k(\mathcal{T})$, $k \in \Gamma$ in (D-8). Suppose $\Omega \sim \text{Ber}(q)$ and \mathcal{G}_k satisfies the incoherence conditions (18)-(19) in Definition 9. Then, with high probability, it holds that*

$$\|\mathcal{P}_{\mathbb{T}_k} - q^{-1} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega} \mathcal{P}_{\mathbb{T}_k}\| \leq \epsilon, \quad (\text{D-10})$$

provided that $q \geq a_1 \epsilon^{-2} \mu R \log(n_{(1)} \ell) / n_{(2)} \ell$ for some numerical constant $a_1 > 0$.

Lemma 3 (Lemma IV.8 in [1]). *Suppose that $\mathcal{Z} \in \mathbb{R}^{n_1 \times \dots \times n_d}$ is fixed, and $\Omega \sim \text{Ber}(q)$. Then with high probability, it holds that*

$$\|(\mathcal{I} - q^{-1} \mathcal{P}_{\Omega}) \mathcal{Z}\| \leq a_2 \left(\frac{\log(n_{(1)} \ell)}{q} \|\mathcal{Z}\|_{\infty} + \sqrt{\frac{\log(n_{(1)} \ell)}{q}} \|\mathcal{Z}\|_{\infty,2} \right), \quad (\text{D-11})$$

for some numerical constant $a_2 > 0$.

Lemma 4 (Lemma IV.9 in [1]). *Denote the corresponding set \mathbb{T}_k of $\mathcal{G}_k = \nabla_k(\mathcal{T})$, $k \in \Gamma$ in (D-8). Suppose that $\mathcal{Z} \in \mathbb{T}_k$ is fixed and $\Omega \sim \text{Ber}(q)$. Then with high probability, it holds that*

$$\|(\mathcal{I} - q^{-1} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega}) \mathcal{Z}\|_{\infty,2} \leq \frac{1}{2} \|\mathcal{Z}\|_{\infty,2} + \frac{1}{2} \sqrt{\frac{n_{(1)} \ell}{\mu R}} \|\mathcal{Z}\|_{\infty}, \quad (\text{D-12})$$

provided that $q \geq a_3 \mu R \log(n_{(1)} \ell) / n_{(2)} \ell$ for some numerical constant $a_3 > 0$.

Lemma 5 (dual certification). *With the same assumptions in Theorem 4, for any $k \in \Gamma$, there always exists a dual certificate $\mathcal{C}_k \in \mathbb{R}^{n_1 \times \dots \times n_d}$ such that*

- 1) $\mathcal{P}_{\Omega}(\mathcal{C}_k) = \mathcal{C}_k$;
- 2) $\|\nabla_k^T(\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T) - \mathcal{P}_{\mathbb{T}_k}(\mathcal{C}_k)\|_{\text{F}} \leq \frac{1}{4} \sqrt{\frac{p}{\ell}}$;
- 3) $\|\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{C}_k)\| \leq \frac{1}{2}$.

Proof: We use the Golfing scheme [8] to construct the dual certificate which satisfies these conditions. For the choice of p in Theorem 4, we have

$$p \geq c_0 \mu R (\log(n_{(1)} \ell))^2 (n_{(2)} \ell)^{-1} \geq c'_0 \mu R \log(n_{(1)} \ell) (\log_2(n_{(1)} \ell) + 3) (n_{(2)} \ell)^{-1} \geq (n_{(2)} \ell)^{-1} \quad (\text{D-13})$$

for some sufficiently large $c_0 \geq (\frac{3}{\log_2(n_{(1)} \ell)} + 2)c'_0$ with $c'_0 := \max\{\frac{1}{4}a_1, a_2, a_3, \frac{1}{\mu R} \log(n_{(1)} \ell)^{-1} (\log_2(n_{(1)} \ell) + 3)^{-1}\} > 0$. Set $n = \lceil \log_2(n_{(1)} \ell) \rceil + 2$ where $\lceil \cdot \rceil$ denotes round up. Assume that the set Ω of observed entries is generated from $\Omega = \bigcup_{i=1}^n \Omega_i$, where each Ω_i is independent with all others and $\mathbb{P}[(i_1, \dots, i_d) \in \Omega_i] = q := 1 - (1 - p)^{\frac{1}{m}}$. Let $\mathcal{D}_0 := \mathcal{O}$, and for $i = 1, 2, \dots, n$, define

$$\mathcal{D}_i = \mathcal{D}_{i-1} + q^{-1} \mathcal{P}_{\Omega_i} \mathcal{P}_{\mathbb{T}_k} (\nabla_k^T(\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T) - \mathcal{P}_{\mathbb{T}_k} \mathcal{D}_{i-1}). \quad (\text{D-14})$$

Then the dual certificate is given by $\mathcal{C}_k := \mathcal{D}_n$.

Proof of 1): By the construction of \mathcal{C}_k , it's easy to verify that

$$\mathcal{P}_{\Omega}(\mathcal{C}_k) = \mathcal{P}_{\Omega}(\mathcal{D}_n) = \mathcal{D}_n = \mathcal{C}_k.$$

Proof of 2): Denote $\mathcal{K}_i := \nabla_k^T(\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T) - \mathcal{P}_{\mathbb{T}_k} \mathcal{D}_i$, $i = 0, 1, \dots, m$. By the definition of \mathcal{D}_i , we have $\mathcal{K}_0 = \nabla_k^T(\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T)$, and

$$\begin{aligned} \mathcal{K}_i &= \nabla_k^T(\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T) - \mathcal{P}_{\mathbb{T}_k} \mathcal{D}_i \\ &= \nabla_k^T(\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T) - \mathcal{P}_{\mathbb{T}_k} (\mathcal{D}_{i-1} + q^{-1} \mathcal{P}_{\Omega_i} \mathcal{P}_{\mathbb{T}_k} (\nabla_k^T(\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T) - \mathcal{P}_{\mathbb{T}_k} \mathcal{D}_{i-1})) \\ &= \mathcal{P}_{\mathbb{T}_k} (\nabla_k^T(\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T) - \mathcal{P}_{\mathbb{T}_k} \mathcal{D}_{i-1}) - q^{-1} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega_i} \mathcal{P}_{\mathbb{T}_k} (\nabla_k^T(\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T) - \mathcal{P}_{\mathbb{T}_k} \mathcal{D}_{i-1}) \\ &= (\mathcal{P}_{\mathbb{T}_k} - q^{-1} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega_i} \mathcal{P}_{\mathbb{T}_k}) \mathcal{K}_{i-1}, \end{aligned} \quad (\text{D-15})$$

for $i = 1, \dots, n$. Obviously, $\mathcal{K}_i \in \mathbb{T}_k$ for $i = 0, \dots, n$ since $\mathcal{K}_0 = \nabla_k^T(\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T) \in \mathbb{T}_k$ (there exists a tensor \mathcal{D}_k such that $\nabla_k^T(\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T) = \mathcal{D}_k *_{\mathfrak{L}} \mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T$). Note that Ω_i is independent of \mathcal{K}_{i-1} and by the choice of p , we then have

$$q \geq \frac{p}{n} \geq c'_0 \mu R \log(n_{(1)} \ell) (n_{(2)} \ell)^{-1}. \quad (\text{D-16})$$

Letting $\epsilon = \frac{1}{2}$ in Lemma 2 with $\Omega_i, i = 1, \dots, n$, we can obtain that from (D-15)

$$\|\mathcal{K}_i\|_{\text{F}} \leq \|\mathcal{P}_{\mathbb{T}_k} - q^{-1} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega_i} \mathcal{P}_{\mathbb{T}_k}\| \|\mathcal{K}_{i-1}\|_{\text{F}} \leq \frac{1}{2} \|\mathcal{K}_{i-1}\|_{\text{F}},$$

holds with high probability (w.h.p) for any $i = 1, \dots, n$. Applying the above inequality recursively with $i = n, n-1, \dots, 1$, it gets

$$\|\nabla_k^T (\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T) - \mathcal{P}_{\mathbb{T}_k} \mathcal{C}_k\|_{\text{F}} = \|\mathcal{K}_n\|_{\text{F}} \leq \frac{1}{2} \|\mathcal{K}_{n-1}\|_{\text{F}}^2 \leq \dots \leq (\frac{1}{2})^n \|\mathcal{K}_0\|_{\text{F}}.$$

Note that

$$\|\mathcal{K}_0\|_{\text{F}} = \|\nabla_k^T (\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T)\|_{\text{F}} \leq \|\nabla_k^T\| \|\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T\|_{\text{F}} \leq \|\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T\|_{\text{F}} \leq \sqrt{R}.$$

Thus,

$$\|\nabla_k^T (\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T) - \mathcal{P}_{\mathbb{T}_k} \mathcal{C}_k\|_{\text{F}} \leq (\frac{1}{2})^n \sqrt{R} \leq \frac{1}{4} (\frac{1}{2})^{\log_2(n_{(1)}\ell)} \sqrt{R} = \frac{1}{4} \frac{1}{n_{(1)}\ell} \sqrt{R} \leq \frac{1}{4} \sqrt{\frac{R}{n_{(1)}}} \frac{1}{\sqrt{n_{(2)}\ell}} \frac{1}{\sqrt{\ell}} \leq \frac{1}{4} \sqrt{\frac{p}{\ell}},$$

where the last inequality uses $R \leq n_{(1)}$ and (D-13).

Proof of 3: Note that $\mathcal{C}_k = q^{-1} \sum_{i=1}^n \mathcal{P}_{\Omega_i} \mathcal{P}_{\mathbb{T}_k} \mathcal{K}_{i-1}$ by its construction. We have

$$\begin{aligned} \|\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{C}_k)\| &\leq \sum_{i=1}^n \|\mathcal{P}_{\mathbb{T}_k^\perp}(q^{-1} \mathcal{P}_{\Omega_i} \mathcal{P}_{\mathbb{T}_k}) \mathcal{K}_{i-1}\| \\ &\leq \sum_{i=1}^n \|(\mathcal{I} - q^{-1} \mathcal{P}_{\Omega_i}) \mathcal{P}_{\mathbb{T}_k} \mathcal{K}_{i-1}\| \\ &= \sum_{i=1}^n \|(\mathcal{I} - q^{-1} \mathcal{P}_{\Omega_i}) \mathcal{K}_{i-1}\|, \end{aligned} \tag{D-17}$$

where the last equation uses $\mathcal{K}_i \in \mathbb{T}_k$. Applying (D-11) in Lemma 3 with $\Omega_i, i = 1, \dots, n$ in (D-17), we can get that w.h.p,

$$\begin{aligned} \|\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{C}_k)\| &\leq a_2 \sum_{i=1}^n \left(\frac{\log(n_{(1)}\ell)}{q} \right) \|\mathcal{K}_{i-1}\|_\infty + \sqrt{\frac{\log(n_{(1)}\ell)}{q}} \|\mathcal{K}_{i-1}\|_{\infty,2} \\ &\leq \frac{a_2}{\sqrt{c_0}} \sum_{i=1}^n \left(\frac{n_{(2)}\ell}{\mu R} \right) \|\mathcal{K}_{i-1}\|_\infty + \sqrt{\frac{n_{(2)}\ell}{\mu R}} \|\mathcal{K}_{i-1}\|_{\infty,2}, \end{aligned} \tag{D-18}$$

where the last inequality uses (D-16) and provided that $c_0 > 1$. Now we bound $\|\mathcal{K}_{i-1}\|_\infty$ and $\|\mathcal{K}_{i-1}\|_{\infty,2}, i = 1, \dots, n$. Letting $\epsilon = \frac{1}{2}$ again in Lemma 2 with $\Omega_i, i = 1, \dots, n$, and using (D-15) repeatedly, we can obtain that w.h.p

$$\begin{aligned} \|\mathcal{K}_{i-1}\|_\infty &= \|(\mathcal{P}_{\mathbb{T}_k} - q^{-1} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega_i} \mathcal{P}_{\mathbb{T}_k}) \mathcal{K}_{i-2}\|_\infty \\ &\leq \|(\mathcal{P}_{\mathbb{T}_k} - q^{-1} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega_i} \mathcal{P}_{\mathbb{T}_k})\| \|\mathcal{K}_{i-2}\|_\infty \\ &\quad \vdots \\ &\leq \|(\mathcal{P}_{\mathbb{T}_k} - q^{-1} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega_i} \mathcal{P}_{\mathbb{T}_k})\|^{i-1} \|\mathcal{K}_0\|_\infty \\ &\leq (\frac{1}{2})^{i-1} \|\nabla_k^T (\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T)\|_\infty \\ &\leq (\frac{1}{2})^{i-1} \|\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T\|_\infty \leq (\frac{1}{2})^{i-1} \frac{\mu R}{n_{(2)}\ell}, \end{aligned} \tag{D-19}$$

where the last inequality holds based on (D-9). Using (D-15) again and applying (D-12) in Lemma 4 with Ω_i , we can obtain

that w.h.p

$$\begin{aligned}
\|\mathcal{K}_{i-1}\|_{\infty,2} &= \|(\mathcal{P}_{\mathbb{T}_k} - q^{-1} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega_i} \mathcal{P}_{\mathbb{T}_k}) \mathcal{K}_{i-2}\|_{\infty,2} \\
&= \|(\mathcal{I} - q^{-1} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega_i}) \mathcal{K}_{i-2}\|_{\infty,2} \\
&\leq \frac{1}{2} \|\mathcal{K}_{i-2}\|_{\infty,2} + \frac{1}{2} \sqrt{\frac{n_{(1)}\ell}{\mu R}} \|\mathcal{K}_{i-2}\|_{\infty} \\
&\stackrel{(a)}{=} \frac{1}{2} \|\mathcal{K}_{i-2}\|_{\infty,2} + \frac{1}{2} \sqrt{\frac{n_{(1)}\ell}{\mu R}} \left(\frac{1}{2}\right)^{i-2} \frac{\mu R}{n_{(2)}\ell} \\
&\leq \frac{1}{2} \|\mathcal{K}_{i-2}\|_{\infty,2} + \left(\frac{1}{2}\right)^{i-1} \sqrt{\frac{\mu R}{n_{(2)}\ell}},
\end{aligned} \tag{D-20}$$

for $i = 2, \dots, n$, where (a) uses (D-19). In a recursive way, we have

$$\begin{aligned}
\|\mathcal{K}_{i-1}\|_{\infty,2} &\leq \frac{1}{2} \|\mathcal{K}_{i-2}\|_{\infty,2} + \left(\frac{1}{2}\right)^{i-1} \sqrt{\frac{\mu R}{n_{(2)}\ell}} \\
&\leq \frac{1}{2} \left(\frac{1}{2} \|\mathcal{K}_{i-3}\|_{\infty,2} + \left(\frac{1}{2}\right)^{i-2} \sqrt{\frac{\mu R}{n_{(2)}\ell}} \right) + \left(\frac{1}{2}\right)^{i-1} \sqrt{\frac{\mu R}{n_{(2)}\ell}} \\
&\quad \vdots \\
&\leq \left(\frac{1}{2}\right)^{i-1} \|\mathcal{K}_0\|_{\infty,2} + (i-1) \left(\frac{1}{2}\right)^{i-1} \sqrt{\frac{\mu R}{n_{(2)}\ell}} \\
&= \left(\frac{1}{2}\right)^{i-1} \|\nabla_k^T (\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T)\|_{\infty,2} + (i-1) \left(\frac{1}{2}\right)^{i-1} \sqrt{\frac{\mu R}{n_{(2)}\ell}} \\
&\leq \left(\frac{1}{2}\right)^{i-1} \|\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T\|_{\infty,2} + (i-1) \left(\frac{1}{2}\right)^{i-1} \sqrt{\frac{\mu R}{n_{(2)}\ell}} \\
&\leq i \left(\frac{1}{2}\right)^{i-1} \sqrt{\frac{\mu R}{n_{(2)}\ell}},
\end{aligned} \tag{D-21}$$

where the last inequality uses

$$\begin{aligned}
\|\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T\|_{\infty,2} &= \max \left\{ \max_{i_1} \|\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T(i_1, :, :, \dots, :) \|_{\text{F}}, \max_{i_2} \|\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T(:, i_2, :, \dots, :) \|_{\text{F}} \right\} \\
&= \max \left\{ \max_{i_1} \|(\mathring{\mathbf{e}}_1^{(i_1)})^T *_{\mathfrak{L}} \mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T\|_{\text{F}}, \max_{i_2} \|\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T *_{\mathfrak{L}} \mathring{\mathbf{e}}_2^{(i_2)}\|_{\text{F}} \right\} \\
&\leq \sqrt{\frac{\mu R}{n_{(2)}\ell}}
\end{aligned} \tag{D-22}$$

based on the incoherence conditions (19)-(20). Substituting (D-19) and (D-21) back to (D-18), we get

$$\begin{aligned}
\|\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{C}_k)\| &\leq \frac{a_2}{\sqrt{c_0}} \sum_{i=1}^n \left\{ \frac{n_{(2)}\ell}{\mu R} \cdot \left(\frac{1}{2}\right)^{i-1} \frac{\mu R}{n_{(2)}\ell} + \sqrt{\frac{n_{(2)}\ell}{\mu R}} \cdot i \left(\frac{1}{2}\right)^{i-1} \sqrt{\frac{\mu R}{n_{(2)}\ell}} \right\} \\
&\leq \frac{a_2}{\sqrt{c_0}} \sum_{i=1}^n (i+1) \left(\frac{1}{2}\right)^{i-1} \stackrel{(a)}{=} \frac{a_2}{\sqrt{c_0}} (6-s) \leq \frac{6a_2}{\sqrt{c_0}} \leq \frac{1}{2},
\end{aligned}$$

provided that c_0 is sufficiently large by $c_0 \geq (12a_2)^2$, where $s = \frac{1}{4n_{(1)}\ell} (6 + \log_2(n_{(1)}\ell))$ is a relatively small constant in above equation (a). The proof is completed. \blacksquare

C. Proof of Theorem 4

Theorem 4. Consider t -CTV based TC model (15). Suppose that \mathcal{T}_0 obeys the standard gradient tensor incoherence conditions (17)-(18) and $\Omega \sim \text{Ber}(p)$. Then, there exist universal constants $c_0, c_1, c_2 > 0$ such that \mathcal{T}_0 is the unique solution to model (15) with probability at least $1 - c_1\gamma(n_{(1)}n_3 \cdots n_d)^{-c_2}$, provided that

$$p \geq c_0\mu R(\log(n_{(1)}\ell))^2/n_{(2)}\ell, \tag{D-23}$$

where ℓ is the specific scale factor given in (5), $n_{(1)} := \max\{n_1, n_2\}$ and $n_{(2)} := \min\{n_1, n_2\}$.

Proof: Consider any tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times \dots \times n_d}$ in the feasible region of t-CTV-TC model (15) obeying that $\mathcal{P}_\Omega(\mathcal{T}) = \mathcal{P}_\Omega(\mathcal{T}_0)$ and $\mathcal{T} \neq \mathcal{T}_0$, we're going to prove its objective function, i.e., its the t-CTV norm, is larger than that at the true underlying tensor \mathcal{T}_0 , hence showing \mathcal{T}_0 is the unique minimum solution. To this end that $\|\mathcal{T}\|_{t\text{-CTV}} \geq \|\mathcal{T}_0\|_{t\text{-CTV}}$, we aim to prove that $\|\nabla_k(\mathcal{T})\|_{\circledast, \mathfrak{L}} \geq \|\nabla_k(\mathcal{T}_0)\|_{\circledast, \mathfrak{L}}$ holds for any $k \in \Gamma$ in an equivalent manner.

Let $\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T + \mathcal{W}_k$ be an arbitrary subgradient of the TNN at $\mathcal{G}_k = \nabla_k(\mathcal{T}_0)$ like (D-3), by the convexity of TNN, we have

$$\|\nabla_k(\mathcal{T})\|_{\circledast, \mathfrak{L}} - \|\nabla_k(\mathcal{T}_0)\|_{\circledast, \mathfrak{L}} \geq \langle \mathcal{U}_k * \mathcal{V}_k^T + \mathcal{W}_k, \nabla_k(\mathcal{T}) - \nabla_k(\mathcal{T}_0) \rangle = \langle \nabla_k^T(\mathcal{U}_k * \mathcal{V}_k^T) + \nabla_k^T(\mathcal{W}_k), \mathcal{T} - \mathcal{T}_0 \rangle,$$

where the last equation uses (D-1). Denote $\mathcal{H} = \mathcal{T} - \mathcal{T}_0 \neq \mathcal{O}$, it's easy to see that $\mathcal{P}_\Omega(\mathcal{H}) = \mathcal{O}$. Since the TNN is the dual norm of the operator norm, i.e., $\|\mathcal{A}\|_{\circledast, \mathfrak{L}} = \sup_{\|\mathcal{B}\| \leq 1} \langle \mathcal{B}, \mathcal{A} \rangle$ for any \mathcal{A} , then there always exists a tensor $\nabla_k^T(\tilde{\mathcal{W}}_k)$ obeying $\|\nabla_k^T(\tilde{\mathcal{W}}_k)\| \leq 1$ such that $\langle \nabla_k^T(\tilde{\mathcal{W}}_k), \mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H}) \rangle = \|\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H})\|_{\circledast, \mathfrak{L}}$. Thus we can take $\nabla_k^T(\mathcal{W}_k) = \mathcal{P}_{\mathbb{T}_k^\perp}(\nabla_k^T(\tilde{\mathcal{W}}_k))$, and then

$$\langle \nabla_k^T(\mathcal{W}_k), \mathcal{H} \rangle = \langle \mathcal{P}_{\mathbb{T}_k^\perp}(\nabla_k^T(\tilde{\mathcal{W}}_k)), \mathcal{H} \rangle = \langle \nabla_k^T(\tilde{\mathcal{W}}_k), \mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H}) \rangle = \|\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H})\|_{\circledast, \mathfrak{L}}.$$

Then we have

$$\|\nabla_k(\mathcal{T})\|_{\circledast, \mathfrak{L}} - \|\nabla_k(\mathcal{T}_0)\|_{\circledast, \mathfrak{L}} \geq \|\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H})\|_{\circledast, \mathfrak{L}} + \langle \nabla_k^T(\mathcal{U}_k * \mathcal{V}_k^T), \mathcal{H} \rangle.$$

Introduce the dual certificate variable \mathcal{C}_k in Lemma 5, and it's easy to get that $\langle \mathcal{C}_k, \mathcal{H} \rangle = \mathcal{O}$ since \mathcal{C}_k locates in Ω while \mathcal{H} locates in Ω^\perp . Thus

$$\begin{aligned} \|\nabla_k(\mathcal{T})\|_{\circledast, \mathfrak{L}} - \|\nabla_k(\mathcal{T}_0)\|_{\circledast, \mathfrak{L}} &\geq \|\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H})\|_{\circledast, \mathfrak{L}} + \langle \nabla_k^T(\mathcal{U}_k * \mathcal{V}_k^T), \mathcal{H} \rangle \\ &= \|\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H})\|_{\circledast, \mathfrak{L}} + \langle \nabla_k^T(\mathcal{U}_k * \mathcal{V}_k^T), \mathcal{H} \rangle \\ &= \|\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H})\|_{\circledast, \mathfrak{L}} + \langle \nabla_k^T(\mathcal{U}_k * \mathcal{V}_k^T) - \mathcal{P}_{\mathbb{T}_k}(\mathcal{C}_k), \mathcal{H} \rangle - \langle \mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{C}_k), \mathcal{H} \rangle. \end{aligned} \quad (\text{D-24})$$

Note that $\nabla_k^T(\mathcal{U}_k * \mathcal{V}_k^T) \in \mathbb{T}_k$ since $\nabla_k^T(\cdot)$ is linear, and then

$$\begin{aligned} \|\nabla_k(\mathcal{T})\|_{\circledast, \mathfrak{L}} - \|\nabla_k(\mathcal{T}_0)\|_{\circledast, \mathfrak{L}} &\geq \|\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H})\|_{\circledast, \mathfrak{L}} + \langle \nabla_k^T(\mathcal{U}_k * \mathcal{V}_k^T) - \mathcal{P}_{\mathbb{T}_k}(\mathcal{C}_k), \mathcal{P}_{\mathbb{T}_k}(\mathcal{H}) \rangle - \langle \mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{C}_k), \mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H}) \rangle \\ &\geq \|\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H})\|_{\circledast, \mathfrak{L}} - \|\nabla_k^T(\mathcal{U}_k * \mathcal{V}_k^T) - \mathcal{P}_{\mathbb{T}_k}(\mathcal{C}_k)\|_{\text{F}} \|\mathcal{P}_{\mathbb{T}_k}(\mathcal{H})\|_{\text{F}} - \|\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{C}_k)\| \|\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H})\|_{\circledast, \mathfrak{L}} \\ &\geq \frac{1}{2} \|\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H})\|_{\circledast, \mathfrak{L}} - \frac{1}{4} \sqrt{\frac{p}{\ell}} \|\mathcal{P}_{\mathbb{T}_k}(\mathcal{H})\|_{\text{F}}, \end{aligned} \quad (\text{D-25})$$

where the second inequality holds based on properties in Lemma 1, and the last inequality follows from Lemma 5.

Next, we prove that the right side of (D-25) is positive. First,

$$\|\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H})\|_{\circledast, \mathfrak{L}} = \frac{1}{\ell} \|\text{bdiag}(\mathfrak{L}(\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H})))\|_* \geq \frac{1}{\ell} \|\text{bdiag}(\mathfrak{L}(\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H})))\|_{\text{F}} = \frac{1}{\sqrt{\ell}} \|\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H})\|_{\text{F}} \geq \frac{1}{\sqrt{\ell}} \|\mathcal{P}_\Omega \mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H})\|_{\text{F}}. \quad (\text{D-26})$$

Note that $\mathcal{P}_\Omega(\mathcal{H}) = \mathcal{O}$, and then we can get

$$\|\mathcal{P}_\Omega \mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H})\|_{\text{F}} = \|\mathcal{P}_\Omega(\mathcal{H} - \mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H}))\|_{\text{F}} = \|\mathcal{P}_\Omega \mathcal{P}_{\mathbb{T}_k}(\mathcal{H})\|_{\text{F}}. \quad (\text{D-27})$$

Besides,

$$\begin{aligned} \|\mathcal{P}_\Omega \mathcal{P}_{\mathbb{T}_k}(\mathcal{H})\|_{\text{F}}^2 &= \langle \mathcal{P}_\Omega \mathcal{P}_{\mathbb{T}_k}(\mathcal{H}), \mathcal{P}_\Omega \mathcal{P}_{\mathbb{T}_k}(\mathcal{H}) \rangle \\ &= \langle \mathcal{P}_{\mathbb{T}_k}(\mathcal{H}), \mathcal{P}_\Omega \mathcal{P}_{\mathbb{T}_k}(\mathcal{H}) \rangle \\ &= \langle \mathcal{P}_{\mathbb{T}_k}(\mathcal{H}), \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_\Omega \mathcal{P}_{\mathbb{T}_k}(\mathcal{H}) \rangle. \end{aligned}$$

Then we have

$$\begin{aligned} \|\mathcal{P}_\Omega \mathcal{P}_{\mathbb{T}_k}(\mathcal{H})\|_{\text{F}}^2 &= p \langle \mathcal{P}_{\mathbb{T}_k}(\mathcal{H}), \mathcal{P}_{\mathbb{T}_k}(\mathcal{H}) - (\mathcal{P}_{\mathbb{T}_k}(\mathcal{H}) - p^{-1} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_\Omega \mathcal{P}_{\mathbb{T}_k}(\mathcal{H})) \rangle \\ &= p \langle \mathcal{P}_{\mathbb{T}_k}(\mathcal{H}), \mathcal{P}_{\mathbb{T}_k}(\mathcal{H}) \rangle - p \langle \mathcal{P}_{\mathbb{T}_k}(\mathcal{H}), \mathcal{P}_{\mathbb{T}_k}(\mathcal{H}) - p^{-1} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_\Omega \mathcal{P}_{\mathbb{T}_k}(\mathcal{H}) \rangle \\ &\geq p \langle \mathcal{P}_{\mathbb{T}_k}(\mathcal{H}), \mathcal{P}_{\mathbb{T}_k}(\mathcal{H}) \rangle - p \|\mathcal{P}_{\mathbb{T}_k}(\mathcal{H}) - p^{-1} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_\Omega \mathcal{P}_{\mathbb{T}_k}(\mathcal{H})\| \langle \mathcal{P}_{\mathbb{T}_k}(\mathcal{H}), \mathcal{H} \rangle \\ &= p \langle \mathcal{P}_{\mathbb{T}_k}(\mathcal{H}), \mathcal{P}_{\mathbb{T}_k}(\mathcal{H}) \rangle - p \|\mathcal{P}_{\mathbb{T}_k}(\mathcal{H}) - p^{-1} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_\Omega \mathcal{P}_{\mathbb{T}_k}(\mathcal{H})\| \langle \mathcal{P}_{\mathbb{T}_k}(\mathcal{H}), \mathcal{P}_{\mathbb{T}_k}(\mathcal{H}) \rangle \\ &\geq p(1 - \|\mathcal{P}_{\mathbb{T}_k}(\mathcal{H}) - p^{-1} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_\Omega \mathcal{P}_{\mathbb{T}_k}(\mathcal{H})\|) \|\mathcal{P}_{\mathbb{T}_k}(\mathcal{H})\|_{\text{F}}^2 \\ &\geq \frac{p}{2} \|\mathcal{P}_{\mathbb{T}_k}(\mathcal{H})\|_{\text{F}}^2, \end{aligned} \quad (\text{D-28})$$

where the last inequality comes from (D-10) in Lemma 2 with $\epsilon = \frac{1}{2}$. Combine the (D-26), (D-27) and (D-28), we then have

$$\frac{1}{2} \|\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H})\|_{\circledast, \mathfrak{L}} \geq \frac{1}{2} \frac{1}{\sqrt{\ell}} \|\mathcal{P}_\Omega \mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H})\|_{\text{F}} = \frac{1}{2} \frac{1}{\sqrt{\ell}} \|\mathcal{P}_\Omega \mathcal{P}_{\mathbb{T}_k}(\mathcal{H})\|_{\text{F}} \geq \frac{1}{2} \sqrt{\frac{p}{2\ell}} \|\mathcal{P}_{\mathbb{T}_k}(\mathcal{H})\|_{\text{F}} > \frac{1}{4} \sqrt{\frac{p}{\ell}} \|\mathcal{P}_{\mathbb{T}_k}(\mathcal{H})\|_{\text{F}}.$$

Last, considering the union probability bound, the appropriate probability in Lemma 2, 3, 4 has the same form $1 - c_1(n_1 n_3 \cdots n_d + n_2 n_3 \cdots n_d)^{-c_2}$ with some numerical constants $c_1, c_2 > 0$ since they all rely on the non-commutative Bernstein inequality [9] on the block diagonal matrices. More details can be found in [1], [9]. Thus, for each $k \in \Gamma$, $\|\nabla_k(\mathcal{T})\|_{\circledast, \mathfrak{L}} \geq \|\nabla_k(\mathcal{T}_0)\|_{\circledast, \mathfrak{L}}$ holds with probability at least $1 - c_1(n_{(1)} n_3 \cdots n_d)^{-c_2}$. That follows that with probability at least $(1 - c_1(n_{(1)} n_3 \cdots n_d)^{-c_2})^\gamma \geq 1 - c_1 \gamma (n_{(1)} n_3 \cdots n_d)^{-c_2}$,

$$\|\mathcal{T}\|_{\text{t-CTV}} = \frac{1}{\gamma} \sum_{k \in \Gamma} \|\nabla_k(\mathcal{T})\|_{\circledast, \mathfrak{L}} \geq \frac{1}{\gamma} \sum_{k \in \Gamma} \|\nabla_k(\mathcal{T}_0)\|_{\circledast, \mathfrak{L}} = \|\mathcal{T}_0\|_{\text{t-CTV}},$$

which means that \mathcal{T}_0 is the unique solution to model (15). The proof is then completed. \blacksquare

APPENDIX E PROOFS OF PROPOSITION 1 AND THEOREM 5

A. Proof of Proposition 1

The core of Proposition 1 is a combination of the main results in existing work [10]. Let $\mathcal{T}_0 \in \mathbb{R}^{n_1 \times \cdots \times n_d}$. Suppose to use random Bernoulli sampling $\Omega \sim \text{Ber}(p)$, and m is number of sampling entries. Let $\{\mathbf{e}_i, i = 1, 2, \dots, n_1 \cdots n_d\}$ be the standard basis set in $\mathbb{R}^{n_1 \cdots n_d}$. Given a sampling matrix \mathbf{A} sized $m \times n_1 \cdots n_d$, whose rows are chosen from $\{\mathbf{e}_j, j = i_1 + (i_2 - 1)n_1 + (i_3 - 1)n_1 n_2 + \cdots + (i_d - 1)n_1 \cdots n_d | (i_1, i_2, \dots, i_d) \in \Omega\}$. Then the random Bernoulli sampling $\mathcal{P}_\Omega(\mathcal{T}_0)$ is equal to $\mathcal{A}(\mathcal{T}_0) := \mathbf{A} \text{vec}(\mathcal{T}_0)$, where $\text{vec}(\cdot)$ is the vectorization operator. Then, there is a direct result from the Proposition 3 in [10], as stated in the following.

Lemma 6 (Proposition 3 in [10]). *Denote $\bar{\mathcal{T}}_0 = \frac{\mathcal{T}_0}{\|\mathcal{T}_0\|_{\text{F}}}$. Then with probability at least $1 - \exp(-\frac{m}{4n_1 \cdots n_d \|\mathcal{T}_0\|_\infty^2})$, it holds that*

$$\frac{\|\mathcal{A}(\bar{\mathcal{T}}_0)\|_{\text{F}}}{\sigma_{\min}(\mathcal{A})} \leq \sqrt{\frac{2m}{n_1 \cdots n_d}}. \quad (\text{E-1})$$

Given an order- d tensor \mathcal{T} and a tensor set \mathbb{S} , define the correlation of \mathcal{T} and \mathbb{S} as

$$\rho(\mathcal{T}, \mathbb{S}) := \inf_{\mathcal{X} \in \mathbb{S}, \mathcal{X} \neq \mathcal{O}} \frac{|\langle \mathcal{X}, \mathcal{T} \rangle|}{\|\mathcal{X}\|_{\text{F}} \|\mathcal{T}\|_{\text{F}}}. \quad (\text{E-2})$$

Then, consider the tensor completion model (E-7)'s equal form

$$\min_{\mathcal{T}} f(\mathcal{T}) := \sum_i w_i \|\mathcal{T}\|_{(i)} \quad \text{s.t. } \mathcal{A}(\mathcal{T}) = \mathcal{A}(\mathcal{T}_0). \quad (\text{E-3})$$

There exists a deterministic model failure conclusion by using the geometric properties (E-2) of the regularization norms.

Lemma 7 (Theorem 1 in [10]). *The model (E-7) fails to recover the original tensor \mathcal{T}_0 if the following condition hold,*

$$\rho(\mathcal{T}_0, \partial f(\mathcal{T}_0)) > \frac{\|\mathcal{A}(\bar{\mathcal{T}}_0)\|_{\text{F}}}{\sigma_{\min}(\mathcal{A})}. \quad (\text{E-4})$$

That is to say, \mathcal{T}_0 is not the unique solution of model (E-7).

Further, the following lemma provides lower bounds of the left hand side of (E-4).

Lemma 8 (Proposition 1 in [10]). *Consider the combination of multiobjective norms $f(\mathcal{T}) := \sum_i w_i \|\mathcal{T}\|_{(i)}$ in model (E-7). Let L_i be the Lipschitz constant of $\|\cdot\|_{(i)}$, $\kappa_i = \frac{\|\bar{\mathcal{T}}_0\|_{(i)}}{L_i}$. Then, it holds that*

$$\rho(\mathcal{T}_0, \partial f(\mathcal{T}_0)) \geq \sum_i \bar{w}_i \kappa_i \geq \kappa_{\min}, \quad (\text{E-5})$$

where $\bar{w}_i = \frac{w_i L_i}{\sum_i w_i L_i}$ and $\kappa_{\min} = \min\{\kappa_i\}$.

Combine Lemma 6, Lemma 7 and Lemma 8, the model (E-7) will fail to recover \mathcal{T}_0 when

$$\rho(\mathcal{T}_0, \partial f(\mathcal{T}_0)) \geq \sum_i \bar{w}_i \kappa_i \geq \kappa_{\min} \geq \sqrt{\frac{2m}{n_1 \cdots n_d}} \geq \frac{\|\mathcal{A}(\bar{\mathcal{T}}_0)\|_{\text{F}}}{\sigma_{\min}(\mathcal{A})}. \quad (\text{E-6})$$

This then derives the following Proposition 1.

Proposition 1. Let an order- d tensor $\mathcal{T}_0 \in \mathbb{R}^{n_1 \times \dots \times n_d}$ with multi structural prior simultaneously. Consider the following general TC model

$$\min_{\mathcal{T}} f(\mathcal{T}) := \sum_i w_i \|\mathcal{T}\|_{(i)} \quad \text{s.t.} \quad \mathcal{P}_{\Omega}(\mathcal{T}) = \mathcal{P}_{\Omega}(\mathcal{T}_0), \quad (\text{E-7})$$

where $\|\cdot\|_{(i)}$ denotes a regularization norm modeling certain corresponding prior with Lipschitz constant L_i , and $w_i > 0$ is the weight parameter. Suppose $\Omega \sim \text{Ber}(p)$ and m is the number of sampling entries. Then, there exist constants $c_0, c_1 > 0$ such that \mathcal{T}_0 is not the unique solution of (E-7) with probability at least $1 - \exp(-\frac{c_1 m}{n_1 \dots n_d \|\mathcal{T}_0\|_{\infty}^2})$, provided that

$$m \leq m_{\text{low}} := c_0 \kappa_{\min}^2 n_1 \dots n_d \leq c_0 (\sum_i \bar{w}_i \kappa_i)^2 n_1 \dots n_d, \quad (\text{E-8})$$

where $\kappa_{\min} = \min\{\kappa_i = \|\bar{\mathcal{T}}_0\|_{(i)}/L_i\}$ and $\bar{\mathcal{T}}_0 = \mathcal{T}_0/\|\mathcal{T}_0\|_{\text{F}}$.

B. A Key Lemma

Obviously, the key to get the corresponding bounds in Theorem 5 lies in the correlation concept κ_i . To that end, we need the following high probability bound.

Lemma 9. For $\mathbf{x} \in \mathbb{R}^n$, denote $\nabla(\mathbf{x})$ as its gradient vector. Suppose $\nabla(\mathbf{x})$ is at most k -sparse, i.e., $\|\nabla(\mathbf{x})\|_0 \leq k$, and its sparse positions are randomly distributed. Then, with probability at least $1 - 2 \exp(-c_0 \epsilon^2 \frac{k}{4n \|\nabla(\mathbf{x})\|_{\infty}^2})$,

$$\|\nabla(\mathbf{x})\|_2 \leq (1 + \epsilon) \sqrt{\frac{k}{n}} \|\mathbf{x}\|_2 \lesssim \sqrt{\frac{k}{n}} \|\mathbf{x}\|_2. \quad (\text{E-9})$$

In other words, $\|\nabla(\mathbf{x})\|_2$ is bounded by $\sqrt{k/n} \|\mathbf{x}\|_2$ with high probability.

Proof: Without loss of generality, assume $\|\mathbf{x}\|_2 = 1$ and $\nabla(\mathbf{x})$ is k -sparse exactly.

1) Introducing randomness: Denote the set $S = \{i_1, i_2, \dots, i_k\}$ as the support set of $\nabla(\mathbf{x})$, where $i_1 < i_2 < \dots < i_k$. Since $\nabla(\mathbf{x})$ is exactly sparse with k non-zero elements, then, it can be verified that \mathbf{x} is in a block form, i.e.,

$$\mathbf{x} = (\dots, \underset{\text{index}: i_1}{x_{i_1}}, \underset{\text{index}: i_2}{x_{i_2}}, \dots, \underset{\text{index}: i_2}{x_{i_2}}, \underset{\text{index}: i_3}{x_{i_3}}, \dots, \underset{\text{index}: i_{k-1}}{x_{i_{k-1}}}, \underset{\text{index}: i_k}{x_{i_k}}, \dots, \underset{\text{index}: i_k}{x_{i_k}}, \dots)^T, \quad (\text{E-10})$$

where the number at the bottom denotes the index number of \mathbf{x} . Thus

$$\|\nabla(\mathbf{x})\|_2^2 = \sum_{i \in S} (\frac{x_{i+1} - x_i}{2})^2 \leq \frac{1}{2} \sum_{i \in S} (x_{i+1}^2 + x_i^2) \doteq \sum_{i \in S} x_i^2. \quad (\text{E-11})$$

Note that the positions of the non-zero entries in $\nabla(\mathbf{x})$ are completely random, which means S is a random uniform subset of $[n] := \{1, 2, \dots, n\}$ with proportion $\frac{k}{n}$. Viewing $\nabla(\mathbf{x})$ as a random variable, it's easy to get,

$$\mathbb{E}(\sum_{i \in S} x_i^2) = \frac{k}{n} \sum_{i \in [n]} x_i^2 = \frac{k}{n}. \quad (\text{E-12})$$

Thus,

$$\mathbb{E}\|\nabla(\mathbf{x})\|_2 \leq \sqrt{\mathbb{E}\|\nabla(\mathbf{x})\|_2^2} \leq \sqrt{\frac{k}{n}}, \quad (\text{E-13})$$

where we use Jensen inequality variant $f(\mathbb{E}X) \leq \mathbb{E}f(X)$ when f is a concave function.

2) Sub-Gaussian concentration: Denote $g := \|\nabla(\mathbf{x})\|_2$ for simplicity, and then g is bounded by $\|\mathbf{x}\|_2 = 1$ using $\|\nabla\| \leq 1$. According to that any bounded random variable and its centering form follow the sub-Gaussian distribution (see Example 2.5.8 and Lemma 2.6.8, [11]), it can be obtained that $g - \mathbb{E}g$ is also sub-Gaussian with sub-Gaussian norm $\|\cdot\|_{\psi 2}$,

$$\|g - \mathbb{E}g\|_{\psi 2} \leq c_1 \|g\|_{\psi 2} \leq c_1 \frac{\|\nabla(\mathbf{x})\|_{\infty}}{\sqrt{\log 2}}, \quad (\text{E-14})$$

where $c_1 > 1$ is a constant. Last, using the tail exponential decay property of sub-Gaussian variable (see Proposition 2.5.2, [11]), for all $t > 0$, there exists an absolute constant $c_2 > 0$, such that

$$\mathbb{P}\{|g - \mathbb{E}g| > t\} \leq 2 \exp(-\frac{c_2 t^2}{\|g - \mathbb{E}g\|_{\psi 2}^2}) \leq 2 \exp(-c_0 \frac{t^2}{\|\nabla(\mathbf{x})\|_{\infty}^2}), \quad (\text{E-15})$$

where the constant $c_0 = c_2 \log 2/c_1^2$. Moreover, note that the sub-Gaussian norm $\|g - \mathbb{E}g\|_{\psi 2}$ is the smallest possible number that makes (E-15) valid (up to an absolute constant). Recall that $\mathbb{E}g \leq \sqrt{\frac{k}{n}}$, by letting $t = \epsilon \sqrt{\frac{k}{n}}$, and then (E-9) holds with probability at least $1 - 2 \exp(-c_0 \epsilon^2 \frac{k}{4n \|\nabla(\mathbf{x})\|_{\infty}^2})$. The proof is then completed. ■

C. Proof of Theorem 5

Theorem 5. For order- d tensor $\mathcal{T}_0 \in \mathbb{R}^{N \times \dots \times N}$ with \mathbf{L} and \mathbf{S} prior structures simultaneously, denote its t -SVD rank as R and gradient tensor \mathcal{G}_k 's sparsity (number of nonzero entries) as S_k , and $S = \min_{k \in \Gamma} \{S_k\}$. Then, the corresponding lower bounds of the following \mathbf{L} and/or \mathbf{S} models satisfy:

Model	$f(\mathcal{T})$	$m_{\text{low}} \lesssim$
\mathbf{L}	$\ \mathcal{T}\ _{\circledast, \mathfrak{L}}$	$N^d \cdot \frac{R}{N}$
\mathbf{S}	$\ \mathcal{T}\ _{\text{TV}}$	$N^d \cdot \frac{S}{N^d}$
$\mathbf{L} + \mathbf{S}$	$\ \mathcal{T}\ _{\circledast, \mathfrak{L}} + \alpha \ \mathcal{T}\ _{\text{TV}}$	$N^d \cdot \min\{\frac{R}{N}, \frac{S}{N^d}\}$
$t\text{-CTV}$	$\ \mathcal{T}\ _{t\text{-CTV}}$	$N^d \cdot \frac{R}{N} \cdot \frac{S}{N^d}$

Proof: 1. For the pure low-rank model $f(\mathcal{T}) = \|\mathcal{T}\|_{\circledast, \mathfrak{L}}$, the Lipschitz constant

$$L_{\|\mathcal{T}\|_{\circledast, \mathfrak{L}}} = \sup \frac{\|\mathcal{T}\|_{\circledast, \mathfrak{L}}}{\|\mathcal{T}\|_{\text{F}}} = \sup \frac{\frac{1}{\ell} \|\text{bdiag}(\mathcal{T}_{\mathfrak{L}})\|_*}{\frac{1}{\sqrt{\ell}} \|\text{bdiag}(\mathcal{T}_{\mathfrak{L}})\|_{\text{F}}} = \frac{1}{\sqrt{\ell}} \sqrt{N^{d-1}}, \quad (\text{E-16})$$

and

$$\|\bar{\mathcal{T}}_0\|_{\circledast, \mathfrak{L}} = \frac{1}{\ell} \|\text{bdiag}(\mathfrak{L}(\bar{\mathcal{T}}_0))\|_* \leq \frac{1}{\ell} \sqrt{RN^{d-2}} \|\text{bdiag}(\mathfrak{L}(\bar{\mathcal{T}}_0))\|_{\text{F}} = \frac{1}{\sqrt{\ell}} \sqrt{RN^{d-2}} \|\bar{\mathcal{T}}_0\|_{\text{F}} = \frac{1}{\sqrt{\ell}} \sqrt{RN^{d-2}}. \quad (\text{E-17})$$

Then, $\kappa = \frac{\|\bar{\mathcal{T}}_0\|_{\circledast, \mathfrak{L}}}{L_{\|\mathcal{T}\|_{\circledast, \mathfrak{L}}}} = \sqrt{\frac{R}{N}}$. From Proposition 1, we have

$$m_{\text{low}} \lesssim N^d \kappa^2 = N^d \frac{R}{N}. \quad (\text{E-18})$$

2. For the pure smooth model $f(\mathcal{T}) = \sum_k \frac{1}{\gamma} \|\nabla_k(\mathcal{T})\|_{\text{F}}$, for any $k \in \Gamma$, the Lipschitz constant

$$L_{\|\nabla_k(\mathcal{T})\|_{\text{F}}} = \sup \frac{\|\nabla_k(\mathcal{T})\|_{\text{F}}}{\|\mathcal{T}\|_{\text{F}}} = \sup \|\nabla_k(\bar{\mathcal{T}})\|_{\text{F}} = \sup \|\nabla_k\| \|\bar{\mathcal{T}}\|_{\text{F}} = \|\bar{\mathcal{T}}\|_{\text{F}} = 1, \quad (\text{E-19})$$

and

$$\|\nabla_k(\bar{\mathcal{T}}_0)\|_{\text{F}} \lesssim \sqrt{\frac{S_k}{N^d}} \|\bar{\mathcal{T}}_0\|_{\text{F}} = \sqrt{\frac{S_k}{N^d}}, \quad (\text{E-20})$$

where we use Lemma 9 since Frobenius norm is entrywise. Then $\kappa_k = \frac{\|\nabla_k(\bar{\mathcal{T}}_0)\|_{\text{F}}}{L_{\|\nabla_k(\mathcal{T})\|_{\text{F}}}} = \sqrt{\frac{S_k}{N^d}}$ and $\kappa_{\min} = \sqrt{\frac{S}{N^d}}$, where $S := \min\{S_k, k \in \Gamma\}$, and thus

$$m_{\text{low}} \lesssim N^d \kappa_{\min}^2 = N^d \frac{S}{N^d}. \quad (\text{E-21})$$

3. For the low-rank plus smooth model $f(\mathcal{T}) = \|\mathcal{T}\|_{\circledast, \mathfrak{L}} + \alpha \|\mathcal{T}\|_{\text{TV}}$, from Proposition 1, combining (E-18) and (E-21), it's easy to get

$$m_{\text{low}} \lesssim N^d \min\left\{\frac{R}{N}, \frac{S}{N^d}\right\}. \quad (\text{E-22})$$

4. Last, for the t-CTV model $f(\mathcal{T}) = \|\mathcal{T}\|_{t\text{-CTV}} = \frac{1}{\gamma} \sum_{k \in \Gamma} \|\nabla_k(\mathcal{T})\|_{\circledast, \mathfrak{L}}$, for any $k \in \Gamma$, the Lipschitz constant

$$L_{\|\nabla_k(\mathcal{T})\|_{\circledast, \mathfrak{L}}} = \sup \frac{\|\nabla_k(\mathcal{T})\|_{\circledast, \mathfrak{L}}}{\|\mathcal{T}\|_{\text{F}}} = \sup \frac{\|\nabla_k(\mathcal{T})\|_{\circledast, \mathfrak{L}}}{\|\nabla_k(\mathcal{T})\|_{\text{F}}} \frac{\|\nabla_k(\mathcal{T})\|_{\text{F}}}{\|\mathcal{T}\|_{\text{F}}} = \frac{1}{\sqrt{\ell}} \sqrt{N^{d-1}} \sup \frac{\|\nabla_k(\mathcal{T})\|_{\text{F}}}{\|\mathcal{T}\|_{\text{F}}} = \frac{1}{\sqrt{\ell}} \sqrt{N^{d-1}}, \quad (\text{E-23})$$

and

$$\|\nabla_k(\bar{\mathcal{T}}_0)\|_{\circledast, \mathfrak{L}} \leq \frac{1}{\sqrt{\ell}} \sqrt{RN^{d-2}} \|\nabla_k(\bar{\mathcal{T}}_0)\|_{\text{F}} \lesssim \frac{1}{\sqrt{\ell}} \sqrt{RN^{d-2}} \sqrt{\frac{S_k}{N^d}} = \frac{1}{\sqrt{\ell}} \sqrt{\frac{RS_k}{N^2}}, \quad (\text{E-24})$$

where uses (E-17) and Lemma 9. Then $\kappa_k = \frac{\|\nabla_k(\bar{\mathcal{T}}_0)\|_{\circledast, \mathfrak{L}}}{L_{\|\nabla_k(\mathcal{T})\|_{\circledast, \mathfrak{L}}}} = \sqrt{\frac{RS_k}{N^{d+1}}}$ and $\kappa_{\min} = \sqrt{\frac{RS}{N^{d+1}}}$, and thus

$$m_{\text{low}} \lesssim N^d \kappa_{\min}^2 = \frac{RS}{N} = N^d \frac{R}{N} \frac{S}{N^d}. \quad (\text{E-25})$$

The proof is then completed. ■

APPENDIX F
PROOF OF THEOREM 6

A. Main Preliminaries

The proof of Theorem 6 is similar to a certain extent to that of Theorem 4, both aiming to show that, for any feasible perturbations of the underlying unique optimal solution, its objective would increase if there exist certain dual certificate conditions. The main difference is that some extra analysis is required on the sparse component in model (16).

First, we need to introduce some necessary preliminaries. Some nations have been introduced in the main preliminaries for the proof of Theorem 4 (see Appendix D). Besides those, for a given scalar x , we use $\text{sgn}(x)$ to denote its sign. By extension, for an order- d tensor \mathcal{T} , let $\text{sgn}(\mathcal{T})$ be a tensor whose entries are the signs of those of \mathcal{T} . Denote Ω as the support set of \mathcal{T} , i.e., the index of these nonzero entries of \mathcal{T} . For the L_1 -norm, its sub-gradient on \mathcal{T} is of the form

$$\partial\|\mathcal{T}\|_1 = \{\text{sgn}(\mathcal{T}) + \mathcal{F}|_{\mathcal{P}_\Omega}(\mathcal{F}) = \mathcal{O}, \|\mathcal{F}\|_\infty \leq 1\}. \quad (\text{F-1})$$

Denote \mathcal{T}' as a trimmed version of \mathcal{T} if the support set of \mathcal{T}' is a subset of that of \mathcal{T} and $\mathcal{T}'(i_1, \dots, i_d) = \mathcal{T}(i_1, \dots, i_d)$ whenever $\mathcal{T}'(i_1, \dots, i_d) \neq 0$, i.e., \mathcal{T}' is a ‘sub’ tensor by setting some entries of \mathcal{T} to be 0. The following lemma asserts that if the t-CTV based TRPCA model (16) recovers the $(\mathcal{T}_0, \mathcal{E}_0)$ from observation $\mathcal{M} = \mathcal{T}_0 + \mathcal{E}_0$, it can also recover $(\mathcal{T}_0, \mathcal{E}'_0)$ from observation $\mathcal{M}_1 = \mathcal{T}_0 + \mathcal{E}'_0$, where \mathcal{E}'_0 is any trimmed version of \mathcal{E}_0 .

Lemma 10. Suppose $(\mathcal{T}_0, \mathcal{E}_0)$ is the unique solution to model (16) with observation $\mathcal{M} = \mathcal{T}_0 + \mathcal{E}_0$. Then, $(\mathcal{T}_0, \mathcal{E}'_0)$ is the unique solution to model (16) with observation $\mathcal{M}_1 = \mathcal{T}_0 + \mathcal{E}'_0$, where \mathcal{E}'_0 is any trimmed version of \mathcal{E}_0 .

Proof: Write $\mathcal{E}'_0 = \mathcal{P}_\Omega(\mathcal{E}_0)$ for some $\Omega_0 \in [n_1] \times \dots \times [n_d]$ and let $(\hat{\mathcal{T}}, \hat{\mathcal{E}})$ be the solution to model (16) with observation $\mathcal{M}_1 = \mathcal{T}_0 + \mathcal{E}'_0$. Then

$$\|\hat{\mathcal{T}}\|_{\text{t-CTV}} + \lambda\|\hat{\mathcal{E}}\|_1 \leq \|\mathcal{T}_0\|_{\text{t-CTV}} + \lambda\|\mathcal{P}_\Omega(\mathcal{E}_0)\|_1, \quad (\text{F-2})$$

and

$$\|\hat{\mathcal{T}}\|_{\text{t-CTV}} + \lambda(\|\hat{\mathcal{E}}\|_1 + \|\mathcal{P}_{\Omega^\perp}(\mathcal{E}_0)\|_1) \leq \|\mathcal{T}_0\|_{\text{t-CTV}} + \lambda\|\mathcal{E}_0\|_1. \quad (\text{F-3})$$

Note that $(\hat{\mathcal{T}}, \hat{\mathcal{E}} + \mathcal{P}_{\Omega^\perp}(\mathcal{E}_0))$ is a feasible solution to model (16) since $\hat{\mathcal{T}} + \hat{\mathcal{E}} + \mathcal{P}_{\Omega^\perp}(\mathcal{E}_0) = \mathcal{T}_0 + \mathcal{E}'_0 + \mathcal{P}_{\Omega^\perp}(\mathcal{E}_0) = \mathcal{M}$. Then we have

$$\|\hat{\mathcal{T}}\|_{\text{t-CTV}} + \lambda\|\hat{\mathcal{E}} + \mathcal{P}_{\Omega^\perp}(\mathcal{E}_0)\|_1 \leq \|\hat{\mathcal{T}}\|_{\text{t-CTV}} + \lambda(\|\hat{\mathcal{E}}\|_1 + \|\mathcal{P}_{\Omega^\perp}(\mathcal{E}_0)\|_1) \leq \|\mathcal{T}_0\|_{\text{t-CTV}} + \lambda\|\mathcal{E}_0\|_1, \quad (\text{F-4})$$

where the first inequality uses triangle inequality and the second one uses (F-3). Based on the optimality of $(\mathcal{T}_0, \mathcal{E}_0)$, there must exist $\hat{\mathcal{T}} = \mathcal{T}_0$ and $\hat{\mathcal{E}} + \mathcal{P}_{\Omega^\perp}(\mathcal{E}_0) = \mathcal{E}_0$, which further leads to $\hat{\mathcal{E}} = \mathcal{P}_\Omega(\mathcal{E}_0) = \mathcal{E}'_0$. The proof is then completed. \blacksquare

In addition, besides the standard incoherence conditions (17)-(18), the proof of Theorem 6 need to use the joint incoherence condition (19), i.e.,

$$\|\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T\|_\infty^2 \leq \frac{\mu R}{n_1 n_2 \ell^2}, \quad (\text{F-5})$$

for $\mathcal{G}_k = \nabla_k(\mathcal{T}_0) = \mathcal{U}_k *_{\mathfrak{L}} \mathcal{S}_k *_{\mathfrak{L}} \mathcal{V}_k^T, k \in \Gamma$. It has been proven that the joint incoherence condition is unavoidable in the typical problems of separating low-rank and sparse components [12]. While in the proof of Theorem 4, such joint incoherence condition can be removed by introducing the bounds in terms of $L_{\infty,2}$ -norm in Lemma 3 and Lemma 4. To establish the theoretical guarantees of the t-CTV-TRPCA model, we use the following bounds as substitution of Lemma 3 and Lemma 4. We omit the proofs since they can be derived easily, and that is just same as the process of Lemma 3 and Lemma 4 in [1], only using the L_∞ -norm not the $L_{\infty,2}$ -norm.

Lemma 11. Suppose that $\mathcal{Z} \in \mathbb{R}^{n_1 \times \dots \times n_d}$ is fixed, and $\Omega \sim \text{Ber}(q)$. Then with high probability, it holds that

$$\|(\mathcal{I} - q^{-1}\mathcal{P}_\Omega)\mathcal{Z}\| \leq \sqrt{\frac{b_2 n_{(1)} \ell \log(n_{(1)} \ell)}{q}} \|\mathcal{Z}\|_\infty \quad (\text{F-6})$$

provided that $q \geq b_2 \log(n_{(1)} \ell) / n_{(2)} \ell$ for some numerical constant $b_2 > 0$.

Lemma 12. Denote the corresponding set \mathbb{T}_k of $\mathcal{G}_k = \nabla_k(\mathcal{T}), k \in \Gamma$ like (D-3). Suppose that $\mathcal{Z} \in \mathbb{T}_k$ is fixed and $\Omega \sim \text{Ber}(q)$. Then with high probability, it holds that

$$\|(\mathcal{I} - q^{-1}\mathcal{P}_{\mathbb{T}_k} \mathcal{P}_\Omega)\mathcal{Z}\|_\infty \leq \epsilon \|\mathcal{Z}\|_\infty, \quad (\text{F-7})$$

provided that $q \geq b_3 \epsilon^{-2} \mu R \log(n_{(1)} \ell) / n_{(2)} \ell$ for some numerical constant $b_3 > 0$.

Moreover, we need to use an assumption which holds with high probability. From Lemma 2, for any $k \in \Gamma$, $\|\mathcal{P}_{\mathbb{T}_k} - (1 -$

$\rho)^{-1} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega^\perp} \mathcal{P}_{\mathbb{T}_k} \| \leq \epsilon$ with high probability provided that $1 - \rho \geq a_0 \epsilon^{-2} \mu R \log(n_{(1)} \ell) / n_{(2)} \ell$ for $\Omega \sim \text{Ber}(1 - \rho)$. Since

$$\mathcal{P}_{\mathbb{T}_k} - (1 - \rho)^{-1} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega^\perp} \mathcal{P}_{\mathbb{T}_k} = (1 - \rho)^{-1} (\mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega} \mathcal{P}_{\mathbb{T}_k} - \rho \mathcal{P}_{\mathbb{T}_k}), \quad (\text{F-8})$$

using $\mathcal{I} = \mathcal{P}_{\Omega} + \mathcal{P}_{\Omega^\perp}$, and based on the triangle inequality, it holds that

$$\|\mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega} \mathcal{P}_{\mathbb{T}_k}\| \leq (1 - \rho) \|\mathcal{P}_{\mathbb{T}_k} - (1 - \rho)^{-1} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega^\perp} \mathcal{P}_{\mathbb{T}_k}\| + \rho \|\mathcal{P}_{\mathbb{T}_k}\| = (1 - \rho)^{-1} \epsilon + \rho. \quad (\text{F-9})$$

Then, with high probability, $\|\mathcal{P}_{\Omega} \mathcal{P}_{\mathbb{T}_k}\|^2 = \|\mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega} \mathcal{P}_{\mathbb{T}_k}\| \leq \rho + \epsilon$, which means, for any $\sigma > 0$,

$$\|\mathcal{P}_{\Omega} \mathcal{P}_{\mathbb{T}_k}\| \leq \sigma \quad (\text{F-10})$$

holds w.h.p. provided that Ω is not too large.

B. Dual Certificates

Lemma 13. *With the same assumptions in Theorem 6, for any $k \in \Gamma$, there always exists a dual certificate pair $\{\mathcal{C}_k, \mathcal{F}\}$ such that*

$$\nabla_k^T (\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}^T) + \nabla_k^T (\mathcal{C}_k) = \lambda (\text{sgn}(\mathcal{E}_0) + \mathcal{F} + \mathcal{P}_{\Omega_0}(\mathcal{B})) \quad (\text{F-11})$$

with $\mathcal{P}_{\mathbb{T}_k}(\mathcal{C}_k) = \mathcal{O}$, $\|\mathcal{C}_k\| \leq \frac{1}{2}$, $k \in \Gamma$, $\mathcal{P}_{\Omega_0}(\mathcal{F}) = \mathcal{O}$, $\|\mathcal{F}\|_\infty \leq \frac{1}{2}$ and $\|\mathcal{P}_{\Omega_0}(\mathcal{B})\|_F \leq \frac{1}{4}$.

Proof: To prove the result, it's sufficient to produce corresponding dual certificates $\mathcal{C}_k, k \in \Gamma$ obeying

$$\begin{cases} \mathcal{C}_k \in \mathbb{T}_k^\perp, \\ \|\mathcal{C}_k\| \leq \frac{1}{2}, \\ \|\mathcal{P}_{\Omega_0}(\nabla_k^T (\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T + \mathcal{C}_k)) - \lambda \text{sgn}(\mathcal{E}_0)\| \leq \frac{\lambda}{4}, \\ \|\mathcal{P}_{\Omega_0^\perp}(\nabla_k^T (\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T + \mathcal{C}_k))\|_\infty < \frac{\lambda}{2}. \end{cases} \quad (\text{F-12})$$

We then construct the certificates satisfying the above conditions. Assume that \mathcal{E}_0 's support set Ω_0 is uniformly distributed among all sets of cardinality m , that means $\Omega_0 \sim \text{Ber}(\rho)$ with $p = \frac{m}{n_1 \dots n_d}$. Then $\Omega_0^\perp \sim \text{Ber}(1 - \rho)$. Let $\Omega_0^\perp = \bigcup_{i=1}^n \Omega_i$, where each Ω_i is independent with all others and $\mathbb{P}[(i_1, \dots, i_d) \in \Omega_i] = q := 1 - (1 - \rho)^{\frac{1}{n}}$.

Now, take a dual certificate \mathcal{C}_k for any $k \in \Gamma$ as an example, and we decompose it as

$$\mathcal{C}_k = \mathcal{C}_k^L + \mathcal{C}_k^S, \quad (\text{F-13})$$

where each component can be constructed in the following strategy.

Construct \mathcal{C}_k^L via the Golfing scheme:

$$\mathcal{C}_k^L = \mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{D}_n), \quad (\text{F-14})$$

where \mathcal{D}_m is obtained by

$$\mathcal{D}_0 = \mathcal{O}, \mathcal{D}_i = \mathcal{D}_{i-1} + q^{-1} \mathcal{P}_{\Omega_i} \mathcal{P}_{\mathbb{T}_k} (\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T - \mathcal{P}_{\mathbb{T}_k} \mathcal{D}_{i-1}), i = 1, \dots, n. \quad (\text{F-15})$$

Construct \mathcal{C}_k^S via the Least Square: Assume $\|\mathcal{P}_{\Omega_0} \mathcal{P}_{\mathbb{T}_k}\| \leq \frac{1}{2}$. Then, $\|\mathcal{P}_{\Omega_0} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega_0}\| < \frac{1}{4}$ and thus, the operator $\mathcal{P}_{\Omega_0} - \mathcal{P}_{\Omega_0} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega_0}$ mapping Ω_0 onto itself is invertible, and its inverse is denoted by $(\mathcal{P}_{\Omega_0} - \mathcal{P}_{\Omega_0} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega_0})^{-1}$. We then set

$$\mathcal{C}_k^S = \frac{\lambda}{4} \mathcal{P}_{\mathbb{T}_k^\perp} (\mathcal{P}_{\Omega_0} - \mathcal{P}_{\Omega_0} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega_0})^{-1} (\text{sgn}(\mathcal{E}_0)). \quad (\text{F-16})$$

This is equivalent to

$$\mathcal{C}_k^S = \frac{\lambda}{4} \mathcal{P}_{\mathbb{T}_k^\perp} \sum_{n \geq 0} (\mathcal{P}_{\Omega_0} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega_0})^n (\text{sgn}(\mathcal{E}_0)). \quad (\text{F-17})$$

Since both \mathcal{C}_k^L and \mathcal{C}_k^S belong to \mathbb{T}_k^\perp and $\mathcal{P}_{\Omega_0}(\mathcal{C}_k^S) = \frac{\lambda}{4} \mathcal{P}_{\Omega_0}(\mathcal{I} - \mathcal{P}_{\mathbb{T}_k})(\mathcal{P}_{\Omega_0} - \mathcal{P}_{\Omega_0} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega_0})^{-1} (\text{sgn}(\mathcal{E}_0)) = \frac{\lambda}{4} \text{sgn}(\mathcal{E}_0)$, we will establish that $\mathcal{C}_k = \mathcal{C}_k^L + \mathcal{C}_k^S$ is a valid dual certificate if it obeys

$$\begin{cases} \|\mathcal{C}_k^L + \mathcal{C}_k^S\| < \frac{1}{2}, \\ \|\mathcal{P}_{\Omega_0}(\nabla_k^T (\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T + \mathcal{C}_k^L))\|_F \leq \frac{\lambda}{4}, \\ \|\mathcal{P}_{\Omega_0^\perp}(\nabla_k^T (\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T + \mathcal{C}_k^L + \mathcal{C}_k^S))\|_\infty \leq \frac{\lambda}{2}. \end{cases} \quad (\text{F-18})$$

This can be done by using the following two lemmas, whose proofs are presented in the end of this section. ■

Lemma 14. *Assume that $\Omega_0 \sim \text{Ber}(\rho)$ with $\rho \leq \rho_s$ for some $0 < \rho_s < 1$. Set $n = \lceil \frac{1}{2} \log_2(n_{(1)}^2 \ell) \rceil + 3$. Then under the orther assumptions of Theorem 6, the \mathcal{C}_k^L constructed by (F-14) obeys*

- 1) $\|\mathcal{C}_k^L\| \leq \frac{1}{4}$;
- 2) $\|\mathcal{P}_{\Omega_0}(\nabla_k^T (\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T + \mathcal{C}_k^L))\|_F \leq \frac{\lambda}{4}$;

$$3) \quad \|\mathcal{P}_{\Omega_0^\perp}(\nabla_k^T(\mathcal{U}_k *_{\mathcal{L}} \mathcal{V}_k^T + \mathcal{C}_k^L))\|_\infty \leq \frac{\lambda}{4}.$$

Lemma 15. Assume that $\Omega_0 \sim \text{Ber}(\rho_s)$, and the signs of \mathcal{E}_0 are independent and identically distributed symmetric (and independent of Ω_0). Then under the orther assumptions of Theorem 6, the \mathcal{C}_k^S constructed by (F-16) obeys

- 1) $\|\mathcal{C}_k^S\| \leq \frac{1}{4}$;
- 2) $\|\mathcal{P}_{\Omega_0^\perp}(\nabla_k^T(\mathcal{C}_k^S))\|_\infty \leq \frac{\lambda}{4}$.

C. Proof of Theorem 6

Theorem 6. Consider t-CTV based TRPCA model (16). Suppose that \mathcal{T}_0 obeys the standard and joint gradient tensor incoherence conditions (17)-(19) and \mathcal{E}_0 's support set, denoted as Ω_0 , is uniformly distributed among all sets of cardinality m . Then, there exist universal constants $c_1, c_2 > 0$ such that $(\mathcal{T}_0, \mathcal{E}_0)$ is the unique solution to (16) with probability at least $1 - c_1\gamma(n_{(1)}n_3 \cdots n_d)^{-c_2}$, provided that

$$\text{rank}_{\text{t-SVD}}(\mathcal{T}_0) \leq \frac{\rho_r n_{(2)} \ell}{\mu \log^2(n_{(1)} \ell)} \quad \text{and} \quad m \leq \rho_s n_1 n_2 \cdots n_d, \quad (\text{F-19})$$

where $\rho_r, \rho_s > 0$ are some numerical constants.

Proof: Consider any tensor pair $(\mathcal{T}, \mathcal{E})$ in the feasible region of t-CTV-TC model (16) obeying that $\mathcal{T} + \mathcal{E} = \mathcal{M}$ and $\mathcal{T} \neq \mathcal{T}_0$, $\mathcal{E} \neq \mathcal{E}_0$. We want to prove its objective function is larger than that at the true underlying tensor pair $(\mathcal{T}_0, \mathcal{E}_0)$, hence showing that $(\mathcal{T}_0, \mathcal{E}_0)$ is the unique minimum solution. To this end that $\|\mathcal{T}\|_{\text{t-CTV}} + \lambda\|\mathcal{E}\|_1 \geq \|\mathcal{T}_0\|_{\text{t-CTV}} + \mathcal{E}_0$, i.e., $\frac{1}{\gamma} \sum_k \|\nabla_k(\mathcal{T})\|_{*,\mathcal{L}} + \lambda\|\mathcal{E}\|_1 \geq \frac{1}{\gamma} \sum_k \|\nabla_k(\mathcal{T}_0)\|_{*,\mathcal{L}} + \lambda\|\mathcal{E}_0\|_1 \geq$, we aim to prove that $\|\nabla_k(\mathcal{T})\|_{*,\mathcal{L}} + \lambda\|\mathcal{E}\|_1 \geq \|\nabla_k(\mathcal{T}_0)\|_{*,\mathcal{L}} + \lambda\|\mathcal{E}_0\|_1$ holds for any $k \in \Gamma$ equally.

Let $\mathcal{U}_k *_{\mathcal{L}} \mathcal{V}_k^T + \mathcal{W}_k$ be an arbitrary sub-gradient of the TNN at $\mathcal{G}_k = \nabla_k(\mathcal{T}_0)$ like (D-3), $k \in \Gamma$, and $\text{sgn}(\mathcal{E}_0) + \mathcal{F}$ be an arbitrary sub-gradient of the L_1 -norm at \mathcal{E}_0 . By the convexity of TNN and L_1 -norm, we have

$$\|\nabla_k(\mathcal{T})\|_{*,\mathcal{L}} + \lambda\|\mathcal{E}\|_1 - (\|\nabla_k(\mathcal{T}_0)\|_{*,\mathcal{L}} + \lambda\|\mathcal{E}_0\|_1) \geq \langle \mathcal{U}_k *_{\mathcal{L}} \mathcal{V}_k^T + \mathcal{W}_k, \nabla_k(\mathcal{T}) - \nabla_k(\mathcal{T}_0) \rangle + \lambda \langle \text{sgn}(\mathcal{E}_0) + \mathcal{F}, \mathcal{E} - \mathcal{E}_0 \rangle. \quad (\text{F-20})$$

Denote $\mathcal{H} = \mathcal{T} - \mathcal{T}_0$, and then $\mathcal{E} - \mathcal{E}_0 = (\mathcal{M} - \mathcal{T}) - (\mathcal{M} - \mathcal{T}_0) = -\mathcal{H}$. Since the TNN is the dual norm of the operator norm, i.e., $\|\mathcal{A}\|_{*,\mathcal{L}} = \sup_{\|\mathcal{B}\| \leq 1} \langle \mathcal{B}, \mathcal{A} \rangle$ for any \mathcal{A} , then there always exists a tensor $\tilde{\mathcal{W}}_k$ obeying $\|\tilde{\mathcal{W}}_k\| \leq 1$ such that $\langle \tilde{\mathcal{W}}_k, \mathcal{P}_{\mathbb{T}_k^\perp}(\nabla_k(\mathcal{H})) \rangle = \|\mathcal{P}_{\mathbb{T}_k^\perp}(\nabla_k(\mathcal{H}))\|_{*,\mathcal{L}}$. Thus we can take $\mathcal{W}_k = \mathcal{P}_{\mathbb{T}_k^\perp}(\tilde{\mathcal{W}}_k)$, and then

$$\langle \mathcal{W}_k, \nabla_k(\mathcal{H}) \rangle = \langle \mathcal{P}_{\mathbb{T}_k^\perp}(\tilde{\mathcal{W}}_k), \nabla_k(\mathcal{H}) \rangle = \langle \tilde{\mathcal{W}}_k, \mathcal{P}_{\mathbb{T}_k^\perp}(\nabla_k(\mathcal{H})) \rangle = \|\mathcal{P}_{\mathbb{T}_k^\perp}(\nabla_k(\mathcal{H}))\|_{*,\mathcal{L}}. \quad (\text{F-21})$$

Besides, we can also take $\mathcal{F} = -\text{sgn}(\mathcal{P}_{\Omega_0^\perp}(\mathcal{H}))$ obeying $\|\mathcal{F}\|_\infty \leq 1$ and thus

$$\langle \mathcal{F}, \mathcal{E} - \mathcal{E}_0 \rangle = \langle -\text{sgn}(\mathcal{P}_{\Omega_0^\perp}(\mathcal{H})), -\mathcal{H} \rangle = \|\mathcal{P}_{\Omega_0^\perp}(\mathcal{H})\|_1. \quad (\text{F-22})$$

Substitute (F-21), (F-22) to (F-20), and then

$$\begin{aligned} & \|\nabla_k(\mathcal{T})\|_{*,\mathcal{L}} + \lambda\|\mathcal{E}\|_1 - (\|\nabla_k(\mathcal{T}_0)\|_{*,\mathcal{L}} + \lambda\|\mathcal{E}_0\|_1) \\ & \geq \|\mathcal{P}_{\mathbb{T}_k^\perp}(\nabla_k(\mathcal{H}))\|_{*,\mathcal{L}} + \|\mathcal{P}_{\Omega_0^\perp}(\mathcal{H})\|_1 + \langle \mathcal{U}_k *_{\mathcal{L}} \mathcal{V}_k^T, \nabla_k(\mathcal{H}) \rangle + \langle \lambda \text{sgn}(\mathcal{E}_0), -\mathcal{H} \rangle \\ & = \|\mathcal{P}_{\mathbb{T}_k^\perp}(\nabla_k(\mathcal{H}))\|_{*,\mathcal{L}} + \|\mathcal{P}_{\Omega_0^\perp}(\mathcal{H})\|_1 + \langle \nabla_k^T(\mathcal{U}_k *_{\mathcal{L}} \mathcal{V}_k^T) - \lambda \text{sgn}(\mathcal{E}_0), \mathcal{H} \rangle, \end{aligned} \quad (\text{F-23})$$

where the last equation uses (D-1). Considering $\langle \nabla_k^T(\mathcal{U}_k *_{\mathcal{L}} \mathcal{V}_k^T) - \lambda \text{sgn}(\mathcal{E}_0), \mathcal{H} \rangle$ and introducing the dual certificate condition in Lemma 13, we have

$$\begin{aligned} |\langle \nabla_k^T(\mathcal{U}_k *_{\mathcal{L}} \mathcal{V}_k^T) - \lambda \text{sgn}(\mathcal{E}_0), \mathcal{H} \rangle| &= |\langle \nabla_k^T(\mathcal{C}_k), \mathcal{H} \rangle + \lambda \langle \mathcal{F}, \mathcal{H} \rangle + \lambda \langle \mathcal{P}_{\Omega_0}(\mathcal{B}), \mathcal{H} \rangle| \\ &= |\langle \mathcal{C}_k, \nabla_k(\mathcal{H}) \rangle + \lambda \langle \mathcal{P}_{\Omega_0^\perp}(\mathcal{F}), \mathcal{H} \rangle + \lambda \langle \mathcal{P}_{\Omega_0}(\mathcal{B}), \mathcal{H} \rangle| \\ &= |\langle \mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{C}_k), \mathcal{P}_{\mathbb{T}_k^\perp}(\nabla_k(\mathcal{H})) \rangle + \lambda \langle \mathcal{P}_{\Omega_0^\perp}(\mathcal{F}), \mathcal{H} \rangle + \lambda \langle \mathcal{P}_{\Omega_0}(\mathcal{B}), \mathcal{H} \rangle| \\ &\leq \|\mathcal{C}_k\| \|\mathcal{P}_{\mathbb{T}_k^\perp}(\nabla_k(\mathcal{H}))\|_{*,\mathcal{L}} + \lambda \|\mathcal{F}\|_\infty \|\mathcal{P}_{\Omega_0^\perp}(\mathcal{H})\|_1 + \lambda \|\mathcal{P}_{\Omega_0}(\mathcal{B})\|_F \|\mathcal{P}_{\Omega_0}(\mathcal{H})\|_F \\ &\leq \frac{1}{2} (\|\mathcal{P}_{\mathbb{T}_k^\perp}(\nabla_k(\mathcal{H}))\|_{*,\mathcal{L}} + \lambda \|\mathcal{P}_{\Omega_0^\perp}(\mathcal{H})\|_1) + \frac{\lambda}{4} \|\mathcal{P}_{\Omega_0}(\mathcal{H})\|_F. \end{aligned} \quad (\text{F-24})$$

Thus,

$$\|\nabla_k(\mathcal{T})\|_{*,\mathcal{L}} + \lambda\|\mathcal{E}\|_1 - (\|\nabla_k(\mathcal{T}_0)\|_{*,\mathcal{L}} + \lambda\|\mathcal{E}_0\|_1) \geq \frac{1}{2} (\|\mathcal{P}_{\mathbb{T}_k^\perp}(\nabla_k(\mathcal{H}))\|_{*,\mathcal{L}} + \lambda \|\mathcal{P}_{\Omega_0^\perp}(\mathcal{H})\|_1) - \frac{\lambda}{4} \|\mathcal{P}_{\Omega_0}(\mathcal{H})\|_F. \quad (\text{F-25})$$

Since

$$\|\mathcal{P}_{\Omega_0}(\mathcal{H})\|_F \leq \|\mathcal{P}_{\Omega_0} \mathcal{P}_{\mathbb{T}_k}(\mathcal{H})\|_F + \|\mathcal{P}_{\Omega_0} \mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H})\|_F \leq \frac{1}{2} \|\mathcal{H}\|_F + \|\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H})\|_F \leq \frac{1}{2} (\|\mathcal{P}_{\Omega_0}(\mathcal{H})\|_F + \|\mathcal{P}_{\Omega_0^\perp}(\mathcal{H})\|_F + \|\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H})\|_F), \quad (\text{F-26})$$

where the third inequality uses (F-10) by letting $\sigma = \frac{1}{2}$. It further leads to the result that

$$\|\mathcal{P}_{\Omega_0}(\mathcal{H})\|_{\text{F}} \leq \|\mathcal{P}_{\Omega_0^\perp}(\mathcal{H})\|_{\text{F}} + 2\|\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H})\|_{\text{F}} \leq \|\mathcal{P}_{\Omega_0^\perp}(\mathcal{H})\|_1 + 2\sqrt{\ell}\|\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{H})\|_{\otimes, \mathfrak{L}}. \quad (\text{F-27})$$

Combine (F-24) and (F-27), we get

$$\|\nabla_k(\mathcal{T})\|_{\otimes, \mathfrak{L}} + \lambda\|\mathcal{E}\|_1 - (\|\nabla_k(\mathcal{T}_0)\|_{\otimes, \mathfrak{L}} + \lambda\|\mathcal{E}_0\|_1) \geq \left(\frac{1}{2} - \frac{\lambda\sqrt{\ell}}{2}\right)\|\mathcal{P}_{\mathbb{T}_k^\perp}(\nabla_k(\mathcal{H}))\|_{\otimes, \mathfrak{L}} + \frac{\lambda}{4}\|\mathcal{P}_{\Omega_0^\perp}(\mathcal{H})\|_1 \geq 0, \quad (\text{F-28})$$

where we use that λ has been set as $\frac{1}{\sqrt{n_{(1)}\ell}}$. At last, unifying the probability for each $k \in \Gamma$, the conclusion that $(\mathcal{T}_0, \mathcal{E}_0)$ is the unique solution to the t-CTV-TRPCA model (16) exactly with probability at least $(1 - c_1(n_{(1)}n_3 \cdots n_d)^{-c_2})^\gamma \geq 1 - c_1\gamma(n_{(1)}n_3 \cdots n_d)^{-c_2}$ for some numerical constant $c_1, c_2 > 0$. The proof is then completed. ■

D. Proof of Lemma 14

Proof: The proof of Lemma 14 is similar with that of Lemma 5. First, denote $\mathcal{K}_i := \mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T - \mathcal{P}_{\mathbb{T}_k}(\mathcal{D}_i)$, $i = 0, 1, \dots, n$, where \mathcal{D}_i is constructed as (F-15). Then we can verify that

$$\mathcal{K}_i = (\mathcal{P}_{\mathbb{T}_k} - q^{-1}\mathcal{P}_{\mathbb{T}_k}\mathcal{P}_{\Omega_i}\mathcal{P}_{\mathbb{T}_k})\mathcal{K}_{i-1}, \quad (\text{F-29})$$

for $i = 1, \dots, n$. When

$$q \geq b_3\epsilon^{-2}\mu R \log(n_{(1)}\ell)/n_{(2)}\ell \quad (\text{F-30})$$

for some numerical constant $b_3 > 0$, considering (F-29), from Lemma 2, we have that $\|\mathcal{K}_i\|_{\text{F}} \leq \epsilon\|\mathcal{K}_{i-1}\|_{\text{F}}$ holds for $i = 1, \dots, n$. It further follows that $\|\mathcal{K}_i\|_{\text{F}} \leq \epsilon^i\|\mathcal{K}_0\|_{\text{F}} = \epsilon^i\sqrt{R}$. Note that $\mathcal{K}_i \in \mathbb{T}_k$ for any $0, 1, \dots, n$ since $\mathcal{K}_0 \in \mathbb{T}_k$. Considering (F-29) as $\mathcal{K}_i = (\mathcal{I} - q^{-1}\mathcal{P}_{\mathbb{T}_k}\mathcal{P}_{\Omega_i})\mathcal{K}_{i-1}$ equally, when q obeys the same estimate, from Lemma 12, we have that $\|\mathcal{K}_i\|_{\infty} \leq \epsilon\|\mathcal{K}_{i-1}\|_{\infty}$ holds for $i = 1, \dots, n$. It further leads to that $\|\mathcal{K}_i\|_{\infty} \leq \epsilon^i\|\mathcal{K}_0\|_{\infty} = \epsilon^i\|\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T\|_{\infty}$.

Proof of 1): Note that $\mathcal{D}_n = q^{-1}\sum_{i=1}^n \mathcal{P}_{\Omega_i}\mathcal{P}_{\mathbb{T}_k}\mathcal{K}_{i-1} = q^{-1}\sum_{i=1}^n \mathcal{P}_{\Omega_i}\mathcal{K}_{i-1}$ based their construction manners. We have

$$\begin{aligned} \|\mathcal{C}_k^L\| &= \|\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{D}_n)\| \leq \sum_{i=1}^n \|q^{-1}\mathcal{P}_{\mathbb{T}_k^\perp}\mathcal{P}_{\Omega_i}\mathcal{K}_{i-1}\| = \sum_{i=1}^n \|\mathcal{P}_{\mathbb{T}_k^\perp}(q^{-1}\mathcal{P}_{\Omega_i}\mathcal{K}_{i-1} - \mathcal{K}_{i-1})\| \\ &\leq \sum_{i=1}^n \|q^{-1}\mathcal{P}_{\Omega_i}\mathcal{K}_{i-1} - \mathcal{K}_{i-1}\| \\ &\leq \sqrt{\frac{b_2 n_{(1)} \ell \log(n_{(1)}\ell)}{q}} \sum_{i=1}^n \|\mathcal{K}_{i-1}\|_{\infty} \\ &\leq \sqrt{\frac{b_2 n_{(1)} \ell \log(n_{(1)}\ell)}{q}} \sum_{i=1}^n \epsilon^{i-1} \|\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T\|_{\infty} \\ &\leq 2\sqrt{\frac{b_2 n_{(1)} \ell \log(n_{(1)}\ell)}{q}} \|\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T\|_{\infty}, \end{aligned} \quad (\text{F-31})$$

where the third inequality is from Lemma 11 and the last one uses $\sum_{i=1}^n (\frac{1}{2})^{i-1} \leq 2$ by letting $\epsilon = \frac{1}{2}$ in Lemma 12. Combining the joint incoherence condition (F-5), we have that

$$\|\mathcal{C}_k^L\| \leq C\sqrt{\frac{\mu R \log(n_{(1)}\ell)}{n_{(2)}\ell}}\epsilon \leq \frac{1}{4} \quad (\text{F-32})$$

holds for some constant $C > 0$ since the logarithmic factor is far more less than the linear factor.

Proof of 2): By the construction of \mathcal{C}_k , it's easy to verify that $\mathcal{P}_{\Omega_0^\perp}(\mathcal{D}_n) = \mathcal{D}_n$. Thus

$$\mathcal{P}_{\Omega_0}(\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T + \mathcal{C}_k^L) = \mathcal{P}_{\Omega_0}(\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T + \mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{D}_n)) = \mathcal{P}_{\Omega_0}(\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T - \mathcal{P}_{\mathbb{T}_k}(\mathcal{D}_n)) = \mathcal{P}_{\Omega_0}(\mathcal{K}_n). \quad (\text{F-33})$$

Then we can get

$$\|\mathcal{P}_{\Omega_0}(\mathcal{K}_n)\|_{\text{F}} \leq \|\mathcal{K}_n\|_{\text{F}} \leq \epsilon^n \sqrt{R}. \quad (\text{F-34})$$

Letting $\epsilon = \frac{1}{2}$ in Lemma 12, note that $n = \lceil \frac{1}{2} \log_2(n_{(1)}^2 \ell) \rceil + 3$, then

$$\epsilon^n \sqrt{R} \leq \frac{1}{8} \sqrt{\frac{R}{n_{(1)}^2 \ell}} = \frac{1}{8} \frac{1}{\sqrt{n_{(1)} \ell}} \sqrt{\frac{R}{n_{(1)}}} \leq \frac{\lambda}{8} < \frac{\lambda}{4}. \quad (\text{F-35})$$

Thus

$$\|\mathcal{P}_{\Omega_0}(\nabla_k^T(\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T + \mathcal{C}_k^L))\|_F \leq \|\nabla_k^T(\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T + \mathcal{C}_k^L)\|_F \leq \|\nabla_k^T\| \|\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T + \mathcal{C}_k^L\|_F \leq \|\mathcal{K}_n\| < \frac{\lambda}{4}, \quad (\text{F-36})$$

where we use $\|\nabla_k^T\| \leq 1$, and this proves the claim.

Proof of 3): Using $\mathcal{D}_n = q^{-1} \sum_{i=1}^n \mathcal{P}_{\Omega_i} \mathcal{K}_{i-1}$ again, it holds that

$$\|\mathcal{D}_n\|_\infty \leq q^{-1} \sum_{i=1}^n \|\mathcal{P}_{\Omega_i} \mathcal{K}_{i-1}\|_\infty \leq q^{-1} \sum_{i=1}^n \|\mathcal{K}_{i-1}\|_\infty \leq q^{-1} \sum_{i=1}^n \epsilon^{i-1} \|\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T\|_\infty. \quad (\text{F-37})$$

Note that $\|\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T\|_\infty \leq \sqrt{\frac{\mu R}{n_1 n_2 \ell^2}}$ by the joint incoherence condition (F-5), and then we get

$$\|\mathcal{D}_n\|_\infty \leq C' \frac{\epsilon^2}{\sqrt{\mu R (\log(n_{(1)} \ell))^2}}, \quad (\text{F-38})$$

for some numerical constant $C' > 0$ whenever q obeys (F-30). Using that $\lambda = \frac{1}{\sqrt{n_{(1)} \ell}}$, thus when ϵ is sufficiently small (note that the setting is consistent with the aforementioned results, since the results (F-32), (F-35) also hold with smaller ϵ)

$$\epsilon \leq \sqrt{\frac{1}{8C'}} (\mu R (\log(n_{(1)} \ell))^2 (n_{(1)} \ell)^{-1})^{\frac{1}{4}} < 1, \quad (\text{F-39})$$

we can get that $\|\mathcal{D}_n\| \leq \frac{1}{8} \frac{1}{\sqrt{n_{(1)} \ell}} = \frac{\lambda}{8}$. At last, recall that $\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T + \mathcal{C}_k^L = \mathcal{K}_n + \mathcal{D}_n$, thus

$$\|\mathcal{P}_{\Omega_0^\perp}(\nabla_k^T(\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T + \mathcal{C}_k^L))\|_\infty \leq \|\nabla_k^T(\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T + \mathcal{C}_k^L)\|_\infty \leq \|\mathcal{U}_k *_{\mathfrak{L}} \mathcal{V}_k^T + \mathcal{C}_k^L\|_\infty \leq \|\mathcal{K}_n\|_\infty + \|\mathcal{D}_n\|_\infty \leq \frac{\lambda}{8} + \frac{\lambda}{8} = \frac{\lambda}{4}. \quad (\text{F-40})$$

The proof is then completed. \blacksquare

E. Proof of Lemma 15

Proof: **Proof of 1):** Denote $\mathcal{M} = \text{sgn}(\mathcal{E}_0)$. Then it is distributed as

$$\mathcal{M}(i_1, \dots, i_d) = \begin{cases} 1, & \text{w.p. } \rho_s/2, \\ 0, & \text{w.p. } 1 - \rho_s, \\ -1, & \text{w.p. } \rho_s/2. \end{cases} \quad (\text{F-41})$$

By construction of \mathcal{C}_k^S in (F-17), we have

$$\|\mathcal{C}_k^S\| = \left\| \frac{\lambda}{4} \mathcal{P}_{\mathbb{T}_k^\perp} \mathcal{M} + \frac{\lambda}{4} \mathcal{P}_{\mathbb{T}_k^\perp} \sum_{n \geq 1} (\mathcal{P}_{\Omega_0} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega_0})^n \mathcal{M} \right\| \leq \frac{\lambda}{4} \|\mathcal{M}\| + \frac{\lambda}{4} \left\| \sum_{n \geq 1} (\mathcal{P}_{\Omega_0} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega_0})^n \mathcal{M} \right\| := \frac{\lambda}{4} \|\mathcal{M}\| + \frac{\lambda}{4} \|\mathfrak{R}(\mathcal{M})\|, \quad (\text{F-42})$$

where $\mathfrak{R}(\mathcal{M}) := \sum_{n \geq 1} (\mathcal{P}_{\Omega_0} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega_0})^n \mathcal{M}$ for simplicity.

For the former $\frac{\lambda}{4} \|\mathcal{M}\|$, from Lemma 4.1 in [13], there exists a function $\psi(\rho_s)$ satisfying $\lim_{\rho_s \rightarrow 0^+} \psi(\rho_s) = 0$ such that $\|\mathcal{M}\| \leq \psi(\rho_s) \sqrt{n_1 n_3}$ where the induced spectral norm is under the DFT based t-SVD (i.e., the corresponding transform constant is n_3). Note that this claim can be easily extended to general transform with constant factor ℓ in the same process. Therefore, it holds that $\frac{\lambda}{4} \|\mathcal{M}\| \leq \frac{\lambda}{4} \psi(\rho_s) \sqrt{n_1 \ell}$ is sufficiently small given $\lambda = 1/\sqrt{n_{(1)} \ell}$ and ρ_s is relatively small.

For the latter $\frac{\lambda}{4} \|\mathfrak{R}(\mathcal{M})\|$, note that

$$\|\mathfrak{R}(\mathcal{M})\| = \|\overline{\mathfrak{L}(\mathfrak{R}(\mathcal{M}))}\| = \sup_{x \in \mathbb{S}^{n_2 n_3 \cdots n_d-1}} \|\overline{\mathfrak{L}(\mathfrak{R}(\mathcal{M}))} \cdot x\|_2. \quad (\text{F-43})$$

The optimal x to (F-43) is an eigenvector of $\overline{\mathfrak{L}(\mathfrak{R}(\mathcal{M}))}^* \cdot \overline{\mathfrak{L}(\mathfrak{R}(\mathcal{M}))}$. Since $\overline{\mathfrak{L}(\mathfrak{R}(\mathcal{M}))}$ is a block diagonal matrix, then the optimal x has a block sparse structure, i.e., x belongs to set $\mathbb{K} = \{x \in \mathbb{R}^{n_2 n_3 \cdots n_d} | x = [x_1^T, \dots, x_j^T, \dots, x_J^T]^T, J = n_3 \cdots n_d, x_j \in \mathbb{R}^{n_2}, \text{ and there exists } k \text{ such that } x_k \neq \mathbf{0}, \text{ and } x_j = \mathbf{0}, j \neq k\}$, i.e., $\|x\|_2 = \|x_k\|_2 = 1$. Note that the covering number (i.e., the number of ϵ -net) of the sphere \mathbb{S}^{n-1} satisfies $N(\mathbb{S}^{n-1}, \epsilon) \leq (1 + \frac{2}{\epsilon})^n$ (see Proposition 4.2.12 and Corollary 4.2.13 in [11]). Let \mathbb{N}_0 be the $\frac{1}{2}$ -net for \mathbb{S}^{n_2-1} at most 5^{n_2} , and then the $\frac{1}{2}$ -net for \mathbb{K} , denoted as \mathbb{N} , has the size at most $n_3 \cdots n_d 5^{n_2}$. We have

$$\|\mathfrak{R}(\mathcal{M})\| = \|\text{bdiag}(\mathfrak{L}(\mathfrak{R}(\mathcal{M})))\| = \sup_{x, y \in \mathbb{K}} \langle x, \text{bdiag}(\mathfrak{L}(\mathfrak{R}(\mathcal{M})))y \rangle = \sup_{x, y \in \mathbb{K}} \langle \text{bdiag}^*(xy^T), \mathfrak{L}(\mathfrak{R}(\mathcal{M})) \rangle, \quad (\text{F-44})$$

where bdiag^* denotes the joint operator of bdiag that maps the block diagonal matrix xy^T back to a tensor of size $n_2 \times n_2 \times$

$n_3 \times \cdots n_d$. Let $\mathcal{Z}' = \text{bdiag}^*(\mathbf{x}\mathbf{y}^\top)$ and $\mathcal{Z} = \mathfrak{L}^{-1}(\mathcal{Z}')$. Then

$$\left\| \frac{1}{4} \mathfrak{R}(\mathcal{M}) \right\| = \frac{1}{4} \sup_{\mathbf{x}, \mathbf{y} \in \mathbb{K}} \langle \mathfrak{L}(\mathcal{Z}), \mathfrak{L}(\mathfrak{R}(\mathcal{M})) \rangle = \frac{1}{4} \sup_{\mathbf{x}, \mathbf{y} \in \mathbb{K}} \ell(\mathcal{Z}, \mathfrak{R}(\mathcal{M})) = \frac{1}{4} \sup_{\mathbf{x}, \mathbf{y} \in \mathbb{K}} \ell(\mathfrak{R}(\mathcal{Z}), \mathcal{M}) \leq \sup_{\mathbf{x}, \mathbf{y} \in \mathbb{N}} \ell(\mathfrak{R}(\mathcal{Z}), \mathcal{M}). \quad (\text{F-45})$$

For a fixed pair (\mathbf{x}, \mathbf{y}) of unit vectors, define the random variable $X(\mathbf{x}, \mathbf{y}) = \langle \ell(\mathfrak{R}(\mathcal{Z}), \mathcal{M}) \rangle$. Note that Ω_0 is also the support set of \mathcal{M} , the signs of \mathcal{M} are independent and identically distributed symmetric and the Hoeffding's inequality (see Theorem 2.2.2 in [11]) gives

$$\mathbb{P}(|X(\mathbf{x}, \mathbf{y})| > t | \Omega_0) \leq 2 \exp\left(\frac{-2t^2}{\|\ell(\mathfrak{R}(\mathcal{Z}))\|_{\text{F}}^2}\right). \quad (\text{F-46})$$

Since $\|\ell(\mathfrak{R}(\mathcal{Z}))\|_{\text{F}} \leq \ell\|\mathfrak{R}\|\|\mathcal{Z}\|_{\text{F}} = \sqrt{\ell}\|\mathfrak{R}\|\|\mathcal{Z}'\|_{\text{F}} = \sqrt{\ell}\|\mathfrak{R}\|$, we have

$$\mathbb{P}\left(\sup_{\mathbf{x}, \mathbf{y} \in \mathbb{N}} |X(\mathbf{x}, \mathbf{y})| > t | \Omega_0\right) \leq 2|\mathbb{N}|^2 \exp\left(\frac{-2t^2}{\ell\|\mathfrak{R}\|_{\text{F}}^2}\right). \quad (\text{F-47})$$

Hence,

$$\mathbb{P}(\|\mathfrak{R}(\mathcal{M})\| > t | \Omega_0) \leq 2|\mathbb{N}|^2 \exp\left(\frac{-2t^2}{\ell\|\mathfrak{R}\|_{\text{F}}^2}\right). \quad (\text{F-48})$$

Recall that $\|\mathscr{P}_{\mathbb{T}_k} \mathscr{P}_{\Omega_0}\| \leq \sigma$ holds with high probability in (F-10). Then $\|\mathfrak{R}\| \leq \sum_{n \geq 1} \|\mathscr{P}_{\Omega_0} \mathscr{P}_{\mathbb{T}_k} \mathscr{P}_{\Omega_0}\|^n = \sum_{n \geq 1} \|\mathscr{P}_{\mathbb{T}_k} \mathscr{P}_{\Omega_0}\|^{2n} \leq \sum_{n \geq 1} (\sigma)^{2n} < \frac{\sigma^2}{1-\sigma^2}$. Therefore, unconditionally, it holds that

$$\mathbb{P}\left(\left\|\frac{1}{4} \mathfrak{R}(\mathcal{M})\right\| > t\right) \leq 2|\mathbb{N}|^2 \exp\left(-\frac{2\gamma^2 t^2}{\ell}\right) + \mathbb{P}(\|\mathscr{P}_{\mathbb{T}_k} \mathscr{P}_{\Omega_0}\| \geq \sigma) \leq 2(n_3 \cdots n_d)^2 5^{2n_2} \exp\left(-\frac{2\gamma^2 t^2}{\ell}\right) + \mathbb{P}(\|\mathscr{P}_{\mathbb{T}_k} \mathscr{P}_{\Omega_0}\| \geq \sigma), \quad (\text{F-49})$$

where $\gamma := \frac{1-\sigma^2}{\sigma^2}$ is sufficiently large. Set $t = C\sqrt{n_1 \ell}$ for some small positive constant, and the above probability inequality implies that $\frac{\lambda}{4}\|\mathfrak{R}(\mathcal{M})\|$ is sufficiently small. Note that the bounds for $\frac{\lambda}{4}\|\mathcal{M}\|$ and $\frac{\lambda}{4}\|\mathfrak{R}(\mathcal{M})\|$ are consistently small provided that $|\Omega_0|$ is not too large, i.e., $m \leq \rho_s n_1 \cdots n_d$ and ρ_s is relatively small, and then $\|\mathcal{C}_k^S\| \leq \frac{1}{4}$ is obtained.

Proof of 2): For $(i_1, i_2, \dots, i_d) \in \Omega_0^\perp$, the (i_1, i_2, \dots, i_d) -th element of $\mathscr{P}_{\Omega_0^\perp}(\nabla_k^\top(\mathcal{C}_k^S))$ can be expressed as

$$\langle \mathscr{P}_{\Omega_0^\perp}(\nabla_k^\top(\mathcal{C}_k^S)), \mathbf{e}_{i_1, i_2, \dots, i_d} \rangle = \langle \mathcal{C}_k^S, \nabla_k(\mathbf{e}_{i_1, i_2, \dots, i_d}) \rangle = \frac{1}{2} (\langle \mathcal{C}_k^S, \mathbf{e}_{i_1, \dots, i'_k, \dots, i_d} \rangle - \langle \mathcal{C}_k^S, \mathbf{e}_{i_1, \dots, i_k, \dots, i_d} \rangle), \quad (\text{F-50})$$

where $\mathbf{e}_{i_1, i_2, \dots, i_d} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is the tensor bias whose (i_1, i_2, \dots, i_d) -th entry equal 1 and the rest ones equals 0, and $i'_k = i_k - 1$.

Observe that

$$\mathscr{P}_{\Omega_0^\perp} \mathcal{C}_k^S = \frac{\lambda}{4} \mathscr{P}_{\Omega_0^\perp} \mathscr{P}_{\mathbb{T}_k^\perp} \sum_{n \geq 0} (\mathscr{P}_{\Omega_0} \mathscr{P}_{\mathbb{T}_k} \mathscr{P}_{\Omega_0})^n \mathcal{M} = -\frac{\lambda}{4} \mathscr{P}_{\Omega_0^\perp} \mathscr{P}_{\mathbb{T}_k} \sum_{n \geq 0} (\mathscr{P}_{\Omega_0} \mathscr{P}_{\mathbb{T}_k} \mathscr{P}_{\Omega_0})^n \mathcal{M}. \quad (\text{F-51})$$

Then $\langle \mathcal{C}_k^S, \mathbf{e}_{i_1, \dots, i_k, \dots, i_d} \rangle = \frac{\lambda}{4} \langle \mathcal{M}, \mathcal{Q}(i_1, \dots, i_k, \dots, i_d) \rangle$, where

$$\mathcal{Q}(i_1, \dots, i_k, \dots, i_d) := - \sum_{n \geq 0} (\mathscr{P}_{\Omega_0} \mathscr{P}_{\mathbb{T}_k} \mathscr{P}_{\Omega_0})^n \mathscr{P}_{\Omega_0} \mathscr{P}_{\mathbb{T}_k} (\mathbf{e}_{i_1, \dots, i_k, \dots, i_d}). \quad (\text{F-52})$$

Conditional on Ω_0 , the signs of \mathcal{M} are independent and identically distributed symmetric, and the Hoeffding's inequality gives

$$\mathbb{P}(|\langle \mathcal{C}_k^S, \mathbf{e}_{i_1, \dots, i_k, \dots, i_d} \rangle| > \frac{t\lambda}{4} |\Omega_0|) \leq 2 \exp\left(-\frac{2t^2}{\|\mathcal{Q}(i_1, \dots, i_k, \dots, i_d)\|_{\text{F}}^2}\right), \quad (\text{F-53})$$

and

$$\mathbb{P}\left(\sup_{i_1, \dots, i_d} |\langle \mathcal{C}_k^S, \mathbf{e}_{i_1, \dots, i_k, \dots, i_d} \rangle| > \frac{t\lambda}{4} |\Omega_0|\right) \leq 2n_1 \cdots n_d \exp\left(-\frac{2t^2}{\sup_{i_1, \dots, i_d} \|\mathcal{Q}(i_1, \dots, i_k, \dots, i_d)\|_{\text{F}}^2}\right), \quad (\text{F-54})$$

On the event $\{\|\mathscr{P}_{\Omega_0} \mathscr{P}_{\mathbb{T}_k}\| \leq \sigma\}$ that holds with high probability, we have

$$\|\mathscr{P}_{\Omega_0} \mathscr{P}_{\mathbb{T}_k} (\mathbf{e}_{i_1, \dots, i_k, \dots, i_d})\|_{\text{F}} \leq \|\mathscr{P}_{\Omega_0} \mathscr{P}_{\mathbb{T}_k}\| \|\mathscr{P}_{\mathbb{T}_k} (\mathbf{e}_{i_1, \dots, i_k, \dots, i_d})\|_{\text{F}} \leq \sigma \sqrt{\frac{2\mu R}{n_{(2)} \ell}}, \quad (\text{F-55})$$

where the second inequality uses $\|\mathscr{P}_{\mathbb{T}_k} (\mathbf{e}_{i_1, \dots, i_k, \dots, i_d})\|_{\text{F}} \leq \sqrt{\frac{2\mu R}{n_{(2)} \ell}}$ from Lemma VI.1 in [1] under the standard incoherence conditions. On the same event, we have $\|\sum_{n \geq 0} (\mathscr{P}_{\Omega_0} \mathscr{P}_{\mathbb{T}_k} \mathscr{P}_{\Omega_0})^n\| < \frac{1}{1-\sigma^2}$ and thus $\|\mathcal{Q}(i_1, \dots, i_k, \dots, i_d)\|_{\text{F}}^2 \leq \frac{2\sigma^2}{(1-\sigma^2)^2} \frac{\mu R}{n_{(2)} \ell}$.

Thus, unconditionally,

$$\mathbb{P}(\sup_{i_1, \dots, i_d} |\langle \mathcal{C}_k^S, \mathbf{e}_{i_1, \dots, i_k, \dots, i_d} \rangle| > \frac{t\lambda}{4}) \leq 2n_1 \cdots n_d \exp(-\frac{n_{(2)}\ell\delta^2 t^2}{\mu R}) + \mathbb{P}(\|\mathcal{P}_{\Omega_0} \mathcal{P}_{\mathbb{T}_k}\| \geq \sigma), \quad (\text{F-56})$$

where $\delta := \frac{1-\sigma^2}{\sigma}$ is sufficiently large. This proves that $\sup_{i_1, \dots, i_d} |\langle \mathcal{C}_k^S, \mathbf{e}_{i_1, \dots, i_k, \dots, i_d} \rangle| \leq \frac{c\lambda}{4}$ w.h.p for some sufficiently small positive constant $c < 1$ provided that $\mu R \leq \rho_r n_{(2)} \ell (\log(n_{(1)} \ell))^{-2}$ and ρ_r is sufficiently small, and $|\Omega_0|$ is not too large, i.e., $m \leq \rho_s n_1 \cdots n_d$ and ρ_s is relatively small.

If $(i_1, \dots, i'_k, \dots, i_d) \in \Omega_0^\perp$, in the same way, w.h.p., $\sup_{i_1, \dots, i_d} |\langle \mathcal{C}_k^S, \mathbf{e}_{i_1, \dots, i'_k, \dots, i_d} \rangle| \leq \frac{c\lambda}{4} \leq \frac{\lambda}{4}$, and this proves the claim by using the triangle inequality on (F-50), i.e.,

$$\begin{aligned} \|\mathcal{P}_{\Omega_0^\perp}(\nabla_k^T(\mathcal{C}_k^S))\|_\infty &= \sup_{i_1, \dots, i_d} |\langle \mathcal{P}_{\Omega_0^\perp}(\nabla_k^T(\mathcal{C}_k^S)), \mathbf{e}_{i_1, i_2, \dots, i_d} \rangle| \\ &\leq \frac{1}{2} \sup_{i_1, \dots, i'_k, \dots, i_d} |\langle \mathcal{C}_k^S, \mathbf{e}_{i_1, \dots, i'_k, \dots, i_d} \rangle| + \frac{1}{2} \sup_{i_1, \dots, i_k, \dots, i_d} |\langle \mathcal{C}_k^S, \mathbf{e}_{i_1, \dots, i_k, \dots, i_d} \rangle| \\ &\leq \frac{1}{2} \left(\frac{\lambda}{4} + \frac{\lambda}{4} \right) = \frac{\lambda}{4}. \end{aligned} \quad (\text{F-57})$$

If $(i_1, \dots, i'_k, \dots, i_d) \in \Omega_0$, observe that $\mathcal{P}_{\Omega_0}(\mathcal{C}_k^S) = \frac{\lambda}{4} \mathcal{M}$. Then

$$|\langle \mathcal{C}_k^S, \mathbf{e}_{i_1, \dots, i'_k, \dots, i_d} \rangle| = |\langle \mathcal{P}_{\Omega_0}(\mathcal{C}_k^S), \mathbf{e}_{i_1, \dots, i'_k, \dots, i_d} \rangle| = |\langle \frac{\lambda}{4} \mathcal{M}, \mathbf{e}_{i_1, \dots, i'_k, \dots, i_d} \rangle| = \frac{\lambda}{4}. \quad (\text{F-58})$$

This proves the claim similarly. The proof is then completed. \blacksquare

APPENDIX G ANALYSIS FOR REMARK 4

We first give the following recognized general convergence result for two-block ADMM iterates which has been proven in [14], where the assumptions are relatively gentle and normally satisfied in many problems.

Proposition 2. *Considering the following optimization problem with two variables*

$$\min_{x, z} f(x) + g(z) \quad \text{s.t.} \quad Ax + Bz = c. \quad (\text{G1})$$

Its optimal value is denoted by

$$p^* = \inf\{f(x) + g(z) | Ax + Bz = c\}, \quad (\text{G2})$$

and the augmented Lagrangian function is

$$L_\rho(x, z, y) = f(x) + g(z) + y^T(Ax + Bz - c) + \frac{\rho}{2} \|Ax + Bz - c\|_2^2, \quad (\text{G3})$$

where y is the dual variable and $\rho > 0$ is a positive scalar. The ADMM of (G1) consists of the iterations

$$\begin{cases} x^{k+1} := \arg \min_x L_\rho(x, z^k, y^k), \\ z^{k+1} := \arg \min_z L_\rho(x^{k+1}, z, y^k), \\ y^{k+1} := y^k + \rho(Ax^{k+1} + Bz^{k+1} - c). \end{cases} \quad (\text{G4})$$

Then, with assumptions that the objective functions f, g is convex, closed and proper, and the unaugmented Lagrangian L_0 has a saddle point, the ADMM (G4) satisfy

- 1) Residual convergence. $r^k := Ax^k + Bz^k - c \rightarrow 0$ as $k \rightarrow \infty$, i.e., the iterates approach feasibility.
- 2) Objective convergence. $f(x^k) + g(z^k) \rightarrow p^*$ as $k \rightarrow \infty$, i.e., the objective function of the iterates approaches the optimal value.
- 3) Dual variable convergence. $y^k \rightarrow y^*$ as $k \rightarrow \infty$, where y^* is a dual optimal point.

Remark 4. Denote matrices Y, A, X_1, B respectively as

$$\begin{bmatrix} 0 \\ \vdots \\ 0 \\ \overline{\mathcal{L}(\mathcal{P}_\Omega(\mathcal{T}_0))} \end{bmatrix}, \begin{bmatrix} \overline{\mathcal{I}_{\mathcal{L}}} & & & \\ & \ddots & & \\ & & \overline{\mathcal{I}_{\mathcal{L}}} & \overline{\mathcal{I}_{\mathcal{L}}} \end{bmatrix}, \begin{bmatrix} \overline{\mathcal{L}(\mathcal{G}_{k_1})} \\ \vdots \\ \overline{\mathcal{L}(\mathcal{G}_{k_\gamma})} \\ \overline{\mathcal{K}_{\mathcal{L}}} \end{bmatrix}, \begin{bmatrix} \overline{\mathcal{L}(\mathcal{D}_{k_1})} \\ \vdots \\ \overline{\mathcal{L}(\mathcal{D}_{k_\gamma})} \\ -\overline{\mathcal{I}_{\mathcal{L}}} \end{bmatrix},$$

and $X_2 := \overline{\mathcal{I}_{\mathcal{L}}}$. Then, t-CTV-TC model's reformulation (25) can be converted to a standard two-block ADMM with constraint condition $AX_1 - BX_2 = Y$, and the updating of (X_1, X_2) is equal to the iteration system in Algorithm 1 since there exists uniqueness between the original and transform domains. Note that it has proved that general two-block ADMM with closed,

proper and convex objective is convergent in residual, objective and variables, see [14]. This directly yields the convergence of the ADMM iterates in Algorithm 1.

Proof: Using the transition forms, $AX_1 - BX_2 = Y$ can be equivalently expressed as:

$$\overline{\mathcal{L}(\mathcal{G}_k)} = \overline{\mathcal{L}(\mathcal{D}_k)} \cdot \overline{\mathcal{T}_{\mathcal{L}}}, \quad \forall k \in \Gamma, \quad (\text{G5})$$

and

$$\overline{\mathcal{K}_{\mathcal{L}}} + \overline{\mathcal{T}_{\mathcal{L}}} = \overline{\mathcal{L}(\mathcal{P}_{\Omega}(\mathcal{T}_0))}. \quad (\text{G6})$$

Based on the property that $\mathcal{T} = \mathcal{A} *_{\mathcal{L}} \mathcal{B}$ is equivalent to $\text{bdiag}(\mathcal{T}_{\mathcal{L}}) = \text{bdiag}(\mathcal{A}_{\mathcal{L}}) \cdot \text{bdiag}(\mathcal{B}_{\mathcal{L}})$ given in Lemma 1, then (G5) is equivalent to

$$\mathcal{G}_k = \mathcal{D}_k *_{\mathcal{L}} \mathcal{T} = \nabla_k(\mathcal{T}), \quad \forall k \in \Gamma. \quad (\text{G7})$$

For (G6), its tensor format is $\mathcal{K}_{\mathcal{L}} + \mathcal{T}_{\mathcal{L}} = \overline{\mathcal{L}(\mathcal{P}_{\Omega}(\mathcal{T}_0))}$. Using the inverse transform on both sides, then

$$\mathcal{K} + \mathcal{T} = \mathcal{P}_{\Omega}(\mathcal{T}_0). \quad (\text{G8})$$

Thus the t-CTV-TC model's reformulation (25) can be converted to a standard two-block ADMM with constraint condition $AX_1 - BX_2 = Y$.

Consider the update of the two variables X_1 and X_2 . Since the equation $AX_1 - BX_2 = Y$ is of the block form, it's easy to get that the updating of $\overline{\mathcal{L}(\mathcal{G}_k)}$, $k \in \Gamma$, and $\overline{\mathcal{K}_{\mathcal{L}}}$ are separated and rely on X_2 . More importantly, since there exists uniqueness between the original and transform domain (only need using the transform operation), the updating of X_1 is equivalent to the updating of \mathcal{G}_k , $k \in \Gamma$ and \mathcal{K} as given in Algorithm 1 (see (30) and (31) in the main paper). Similarly, the updating of X_2 is equivalent to the updating of \mathcal{T} in Algorithm 1 (see (29)). Thus such reformulation is totally equal to the original form.

Last, directly using the existing acknowledged convergence results of two-block ADMM [14], i.e., Proposition 2, it can be easily deduced that the ADMM iterates of Algorithm 1 is convergent in residual, objective and variables since the assumptions can be easily verified to be satisfied. The same process is appropriate for Algorithm 2. The proof is then completed. ■

APPENDIX H MORE EXPERIMENTAL RESULTS

A. Visual Data Inpainting

This subsection presents more experimental results on visual data inpainting. First, to further verify the superiority of our t-CTV method when the sampling rate (SR) is extremely small, Fig. 18 shows two visual examples on image "lena" and "butterfly" when SR = 5%, 2%, 1% and 0.5%. Table II further lists the PSNR values of the recovered results on 10 random images among the 50 test images. These results show that our t-CTV method indeed has an evidently better performance in cases that the sampling rate is very small. This can be rationally explained by the fact that it has a smaller lower bound in sampling complexity as Theorem 7 states. In addition, we supplement more detailed results of other visual data's inpainting experiments. Fig. 2 shows some recovered videos via different methods when SR = 5%, 10% and 20%. Table III lists the recovery quantitative metric ERGAS for HSIs inpainting results by all competing methods when SR = 10% and Fig. 3 depicts the corresponding recovered HSIs. Fig 4 plots more recovered examples of the MRI/CT medical image completion when SR = 5%, 10% and 20%. These results further validate the advantage of the proposed t-CTV method in different inpainting tasks.

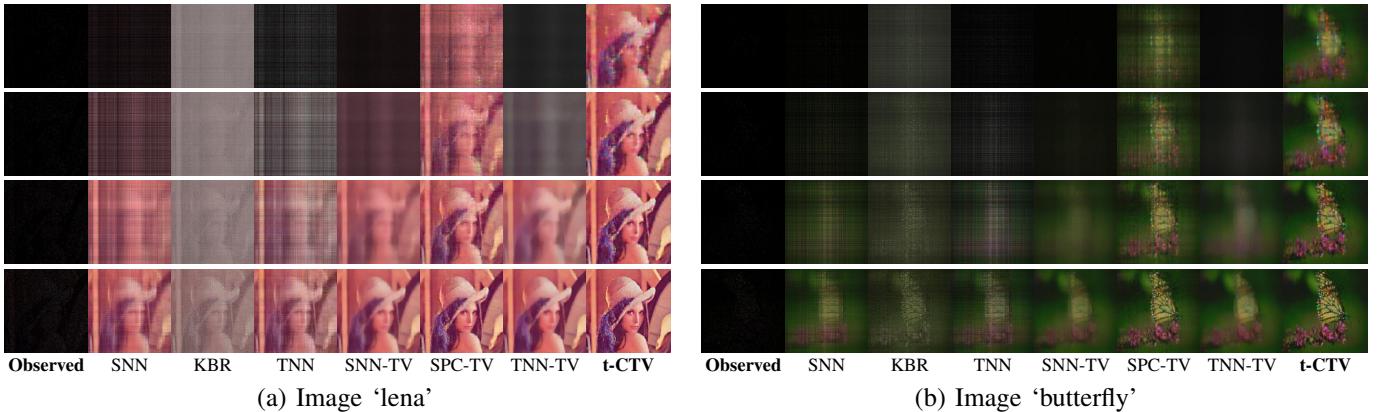


Fig. 1: Color image inpainting results obtained by all competing methods on test image "lena" and "butterfly". From top to bottom: SR = 0.05%, 1%, 2%, 5%.

TABLE II: Eight color images' inpainting quantitative results of all competing methods when SR = 1%. The best and second best results on each line are highlighted in **bold** and underline, respectively, for easy comparison.

Image Method	Image 1		Image 2		Image 3		Image 4		Image 5		Image 6		Image 7		Image 8		Average	
	PSNR	SSIM																
SNN	4.36	0.054	7.20	0.057	7.93	0.091	7.99	0.073	6.36	0.092	7.09	0.088	6.61	0.084	6.48	0.077	6.75	0.077
BCPF	13.94	0.138	13.03	0.081	6.23	0.000	16.94	0.311	13.93	0.123	14.58	0.134	5.21	0.001	4.88	0.000	11.09	0.099
KBR	15.14	<u>0.256</u>	<u>13.92</u>	0.149	14.47	0.177	17.62	<u>0.327</u>	14.44	0.211	14.12	0.194	<u>11.73</u>	0.132	12.48	0.135	14.24	0.198
IRTNN	4.79	0.029	8.58	0.042	9.21	0.049	9.96	0.094	7.48	0.046	7.98	0.048	7.53	0.046	7.71	0.052	7.91	0.051
TNN	7.43	0.055	9.31	0.065	9.79	0.085	10.94	0.133	8.89	0.080	8.83	0.079	7.86	0.075	8.50	0.077	8.94	0.081
MF-TV	2.84	0.022	5.85	0.018	6.97	0.041	6.76	0.032	5.27	0.037	5.93	0.048	5.93	0.042	5.51	0.030	5.63	0.033
SNN-TV	4.42	0.173	7.31	0.093	8.07	0.166	8.10	0.090	6.50	0.235	7.19	0.192	6.70	0.172	6.50	0.142	6.85	0.158
SPC-TV	9.97	0.126	10.88	0.126	9.22	0.091	14.00	0.257	10.74	0.199	11.36	0.177	7.52	0.076	9.00	0.115	10.34	0.146
TNN-TV	10.25	0.401	11.32	<u>0.199</u>	11.74	<u>0.309</u>	13.30	0.180	11.27	<u>0.489</u>	10.97	<u>0.378</u>	9.58	<u>0.336</u>	10.44	<u>0.288</u>	11.11	<u>0.322</u>
t-CTV	18.67	0.589	18.42	0.295	18.87	0.490	19.24	0.391	19.76	0.632	20.00	0.568	17.60	0.552	17.08	0.469	18.70	0.498

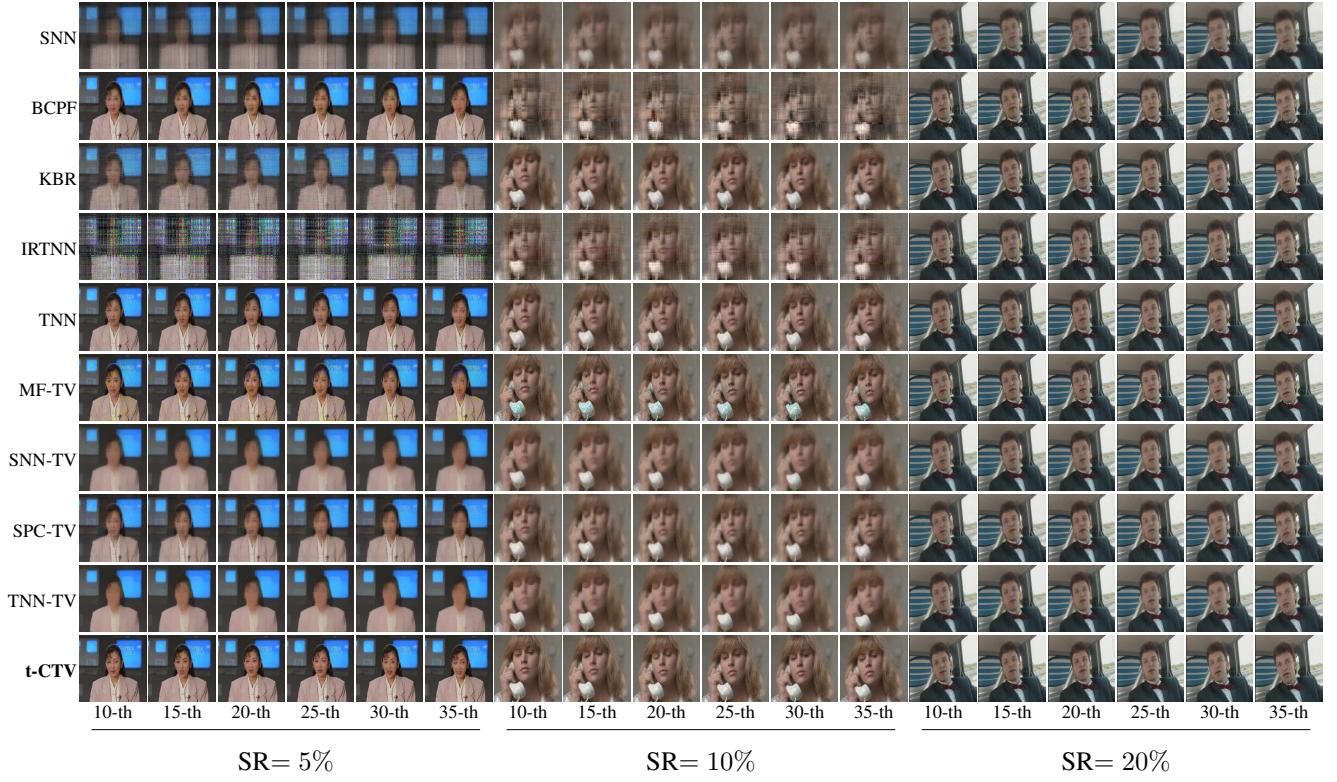


Fig. 2: Color videos inpainting results obtained by all competing methods under SR = 5%, 10% and 20% on the “akiyo”, “suzie” and “carphone” videos. For each video, its 10-th to 35-th frames with interval 5 are shown in the figure.

TABLE III: ERGAS comparison of all competing methods on HSI inpainting under SR = 10%

Method	HSIs Data					Average
	Cuprite	DCMall	KSC	Pavia	PaviaU	
SNN	123.4	364.3	201.1	327.2	290.7	261.3
BCPF	48.9	142.7	92.4	101.5	94.5	96.0
KBR	12.7	<u>61.3</u>	32.8	33.4	31.5	34.4
IRTNN	848.4	734.6	322.9	72.4	72.7	410.2
TNN	33.8	172.9	118.8	87.6	85.9	99.8
MF-TV	32.1	382.0	181.9	213.0	120.8	186.0
SNN-TV	109.9	367.6	179.5	279.0	246.5	236.5
SPC-TV	63.5	144.3	89.4	112.8	91.0	100.2
TNN-Tv	93.1	313.0	161.2	223.5	193.6	196.9
t-CTV	12.7	60.2	<u>34.4</u>	<u>37.3</u>	<u>32.4</u>	<u>35.4</u>

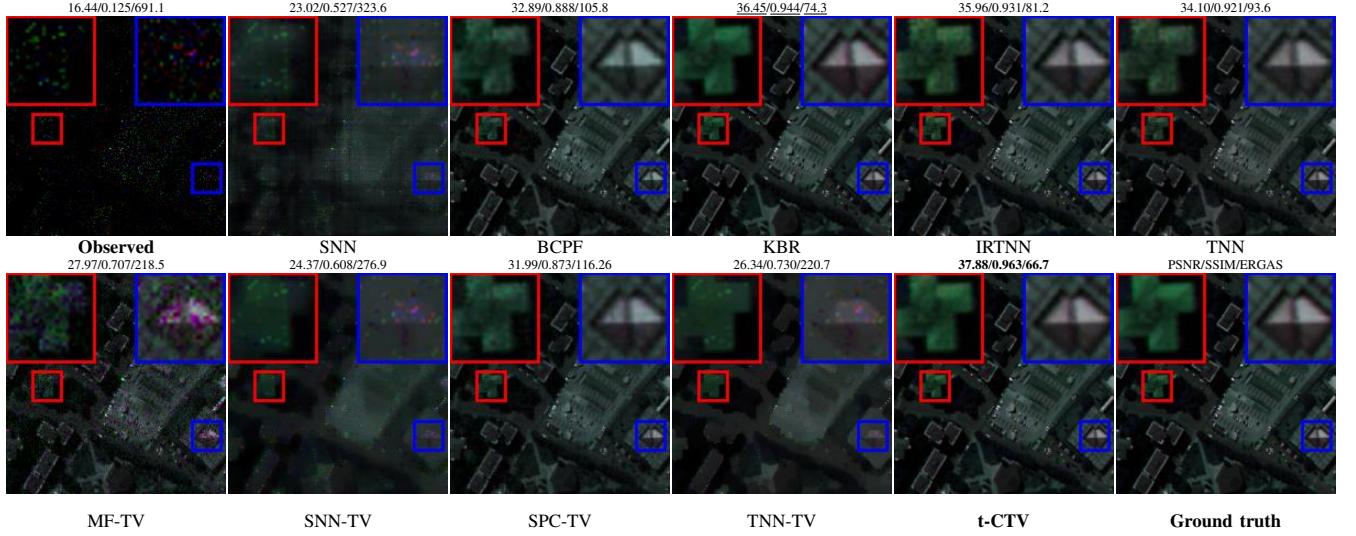


Fig. 3: HSI inpainting results of all competing methods under $SR = 10\%$. The displayed pseudo-color image uses bands 3-44-25 as R-G-B.

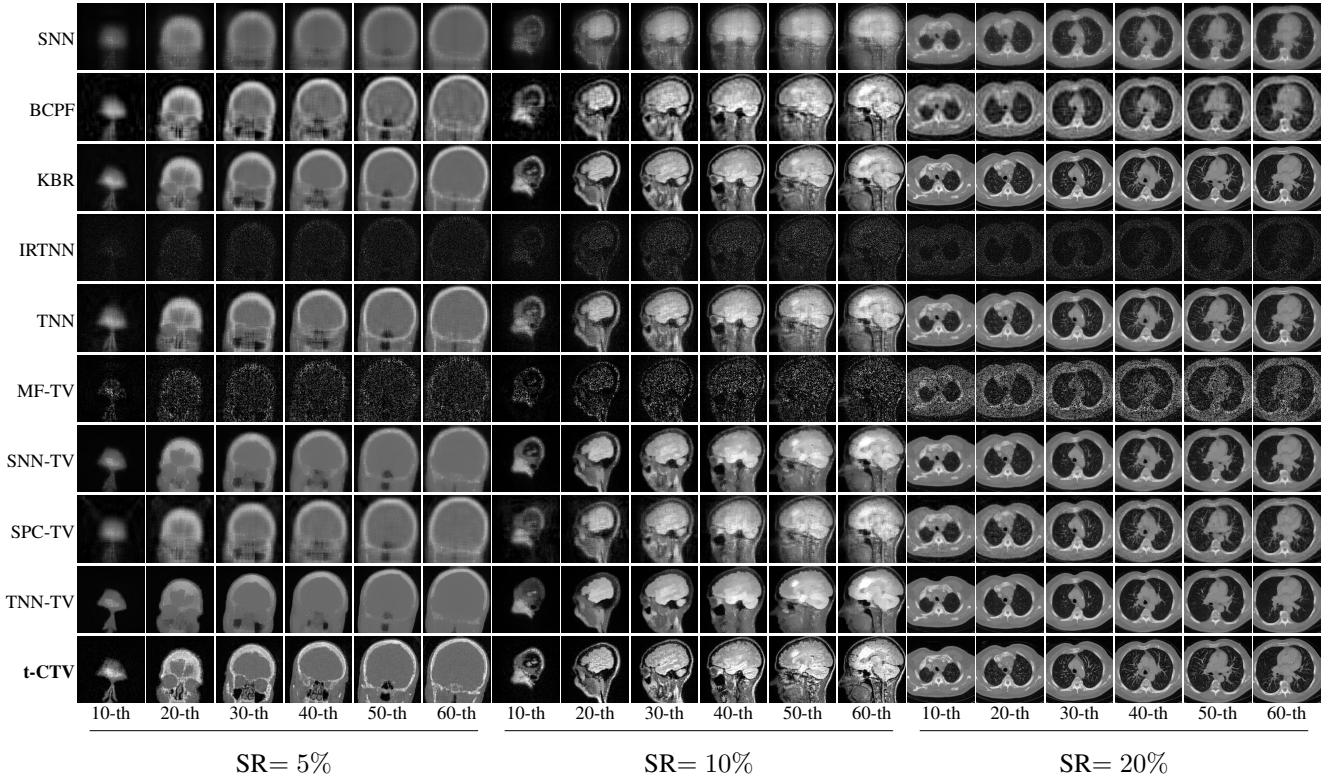


Fig. 4: MRI and CT image inpainting results obtained by all competing methods under $SR = 5\%$, 10% and 20% . for each instance, the 10-th to 60-th bands with interval 5 of the recovered MRI/CT images are shown.

TABLE IV: Eight color images' denoising quantitative results of all competing methods for sparse noise with percentage 0.4. The best and second best results on each line are highlighted in **bold** and underline, respectively.

Image Method	Image 1		Image 2		Image 3		Image 4		Image 5		Image 6		Image 7		Image 8		Average	
	PSNR	SSIM																
SNN	22.40	0.729	20.71	0.473	23.06	<u>0.688</u>	27.00	0.857	22.91	<u>0.728</u>	22.94	<u>0.708</u>	21.54	<u>0.692</u>	20.56	<u>0.654</u>	22.64	<u>0.691</u>
TNN	22.57	0.471	18.37	0.373	21.54	0.461	26.43	0.812	23.91	0.485	22.33	0.439	20.39	0.362	20.23	0.425	21.97	0.479
KBR	21.66	0.567	18.82	0.407	20.41	0.513	27.04	0.866	22.72	0.608	20.73	0.490	18.62	0.417	20.25	0.514	21.28	0.548
LRTV	18.35	0.288	17.95	0.383	19.51	0.379	19.29	0.472	18.91	0.266	19.20	0.315	18.59	0.306	18.38	0.357	18.77	0.346
LRTDTV	21.35	0.412	18.15	0.402	20.28	0.434	20.96	0.573	20.39	0.313	18.57	0.312	17.65	0.303	19.86	0.433	19.65	0.398
TLR-HTV	<u>24.71</u>	0.722	<u>21.62</u>	<u>0.548</u>	<u>24.86</u>	0.669	<u>27.72</u>	0.864	<u>25.93</u>	0.672	<u>24.75</u>	0.608	<u>23.02</u>	0.543	<u>22.71</u>	0.618	<u>24.42</u>	0.656
t-CTV	28.01	0.875	23.22	0.660	28.78	0.851	30.60	0.919	30.10	0.855	29.13	0.867	28.10	0.851	25.85	0.813	27.97	0.836

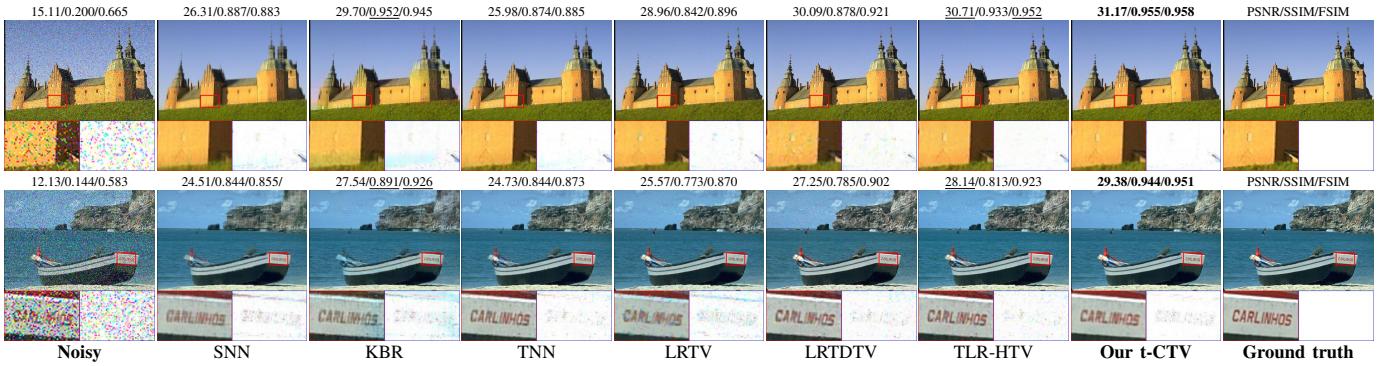


Fig. 5: Color image denoising results of all competing methods. From top to bottom: the noise level is sparse noise with percentage 0.1, 0.2.

TABLE V: Five HSIs' denoising quantitative results of all competing methods for sparse noise with percentage 0.4 (ERG. denotes ERGAS). The best and second best results on each line are highlighted in **bold** and underline, respectively.

HSI Method	Cuprite			DCMall			KSC			Pavia			PaviaU			Average		
	PSNR	ERG.	SSIM	PSNR	ERG.	SSIM	PSNR	ERG.	SSIM									
SNN	44.28	15.9	0.993	<u>32.62</u>	90.3	0.966	<u>36.87</u>	69.2	<u>0.974</u>	35.95	77.2	0.984	<u>32.76</u>	<u>137.2</u>	<u>0.978</u>	<u>36.50</u>	78.0	<u>0.979</u>
TNN	41.46	22.6	0.980	31.03	107.4	0.939	36.31	77.8	0.932	<u>36.21</u>	<u>73.6</u>	0.968	<u>36.50</u>	70.9	0.960	36.30	70.5	0.955
KBR	38.88	29.0	0.981	26.32	154.1	0.899	33.30	101.6	0.923	31.04	129.2	0.870	32.10	114.9	0.946	32.33	105.8	0.924
LRTV	36.56	40.0	0.928	26.48	159.6	0.862	32.94	109.6	0.846	31.55	124.0	0.901	33.25	98.5	0.913	32.15	106.3	0.890
LRTDTV	40.23	24.4	0.971	<u>29.76</u>	104.7	0.933	<u>36.88</u>	<u>65.5</u>	0.944	34.74	83.7	0.958	<u>36.69</u>	<u>63.8</u>	0.961	35.66	<u>68.4</u>	0.953
TLR-SSTV	37.00	34.8	0.946	28.06	129.8	0.909	33.87	92.3	0.909	31.49	121.6	0.920	33.33	95.0	0.932	32.75	94.7	0.923
t-CTV	47.56	12.6	0.994	37.50	65.3	0.980	42.38	44.9	0.980	38.92	55.3	0.986	40.51	46.3	0.986	41.37	44.9	0.985

B. Visual Data Denoising

This subsection demonstrates more experimental results on visual data denoising. Table IV and Table V list the detailed denoising quantitative results of all competing methods for sparse noise with percentage 0.4 on 8 color images and 5 HSIs, respectively. Fig 5 and Fig. 6 further show the visual denoising results for easy observation. Fig. 7 plots the specific PSNR values of the 10 test videos' denoising results via different methods under different levels of sparse noise. Fig. 8 shows two examples under the percentage of sparse noise is 0.1 and 0.2. From these results, it's evident to verify that our t-CTV method competes over other competing methods, further substantiating its effectiveness on these tasks.

APPENDIX I MORE EXPERIMENTAL EXTENSIONS TO OTHER TASKS

In this section, we show some extensions that the proposed t-CTV regularizer is applied in other two tasks, hyperspectral image anomaly detection and surveillance video foreground/background separation. The former aims to detect interesting anomaly targets and separate them from the surrounding natural background from observed hyperspectral image, and the latter aims to separate the nearly fixed background and dynamic foreground from a video sequence. The two tasks are very similar since in many cases since the anomalies or foreground subjects are often very small and distributed sparsely in space and the background part can normally assumed as a joint low-rank and smooth tensor. They both can be seen as typical variations of the TRPCA model where we can use the t-CTV to model the underlying joint low-rank and smooth background part and use certain sparsity promoting regularizer to model the rest part. Here, we adopt the $L_{2,1}$ -norm to encode the group sparsity of

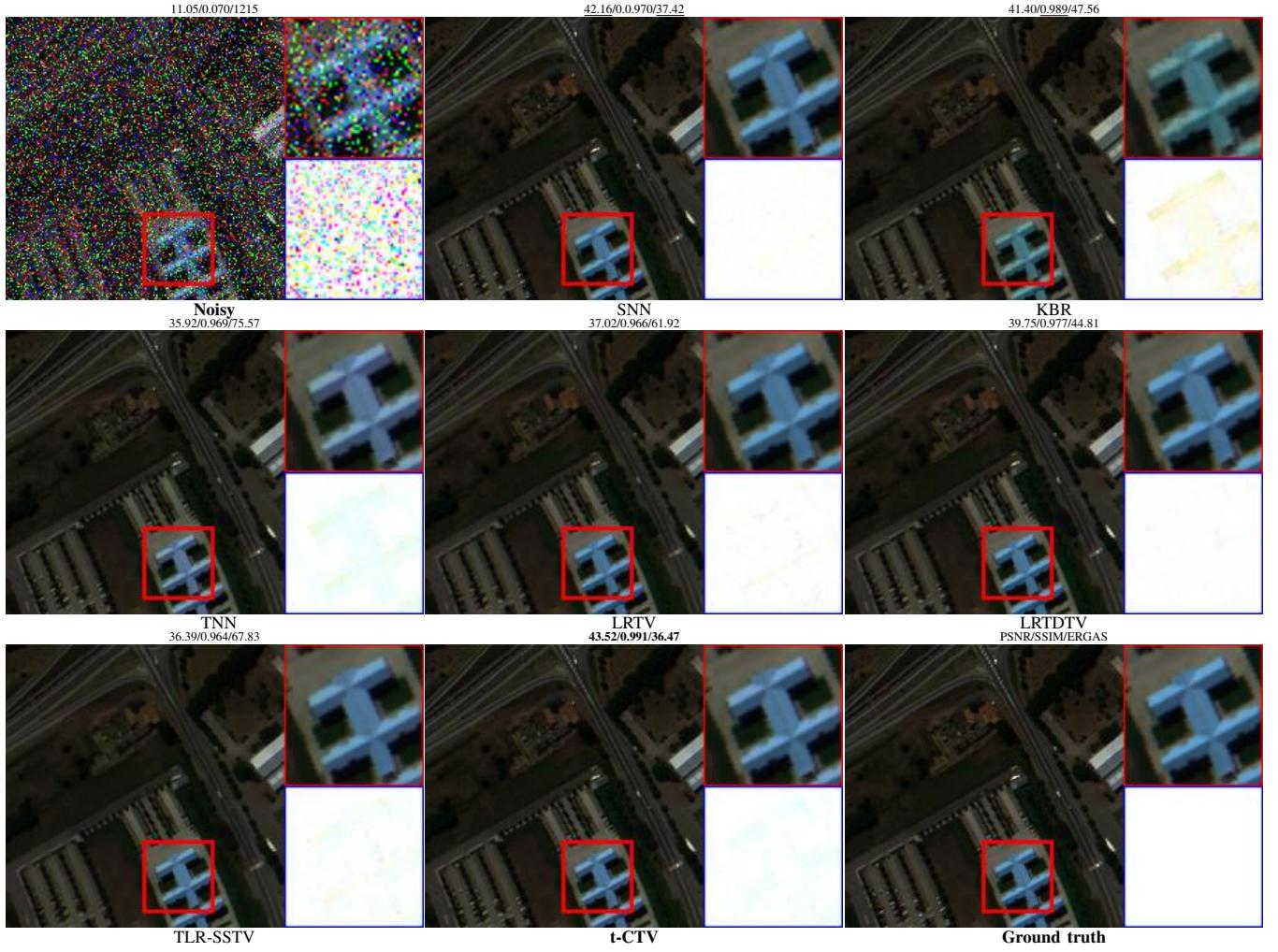


Fig. 6: Denoising results of all competing methods on HSI "PaviaU" with sparse noise when the percentage is 0.2.

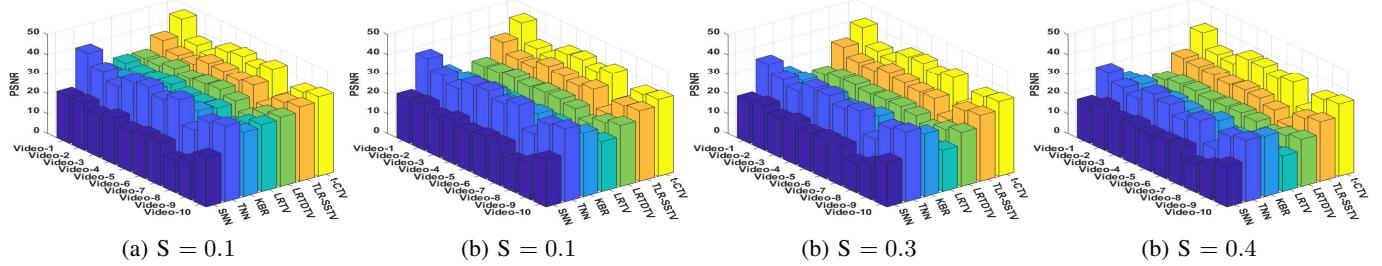


Fig. 7: Performance comparison on PSNR values of recovered color videos via all competing methods in video denoising.

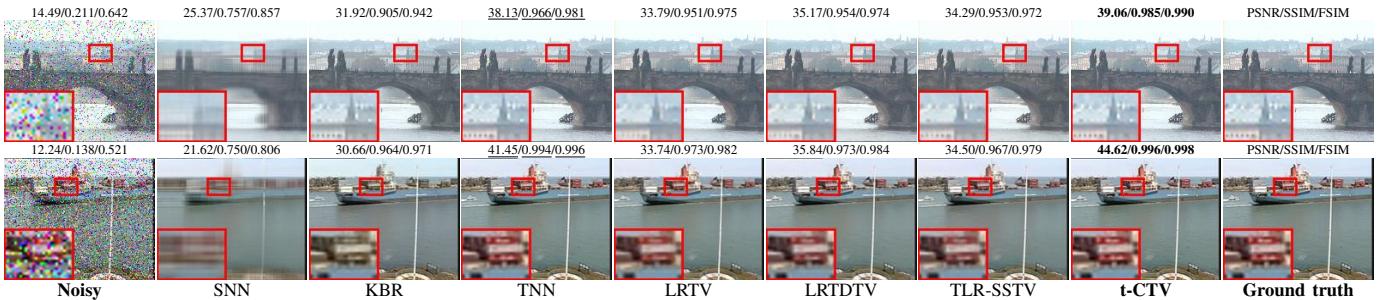


Fig. 8: Color video denoising results obtained by all competing methods with sparse noise when the percentage is 0.1, 0.2 from top to bottom.

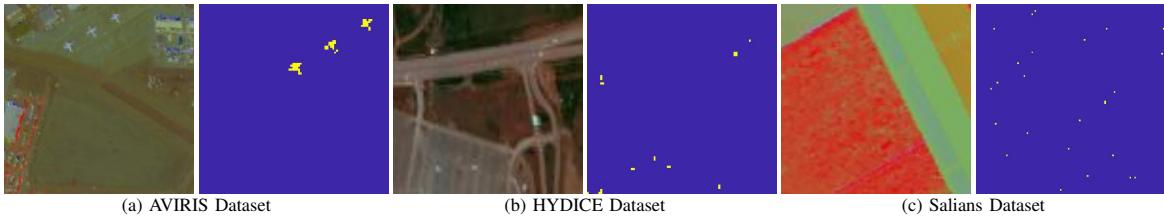


Fig. 9: Typical hyperspectral images in AD tasks. Left: false color image; Right: ground truth of the anomaly map.

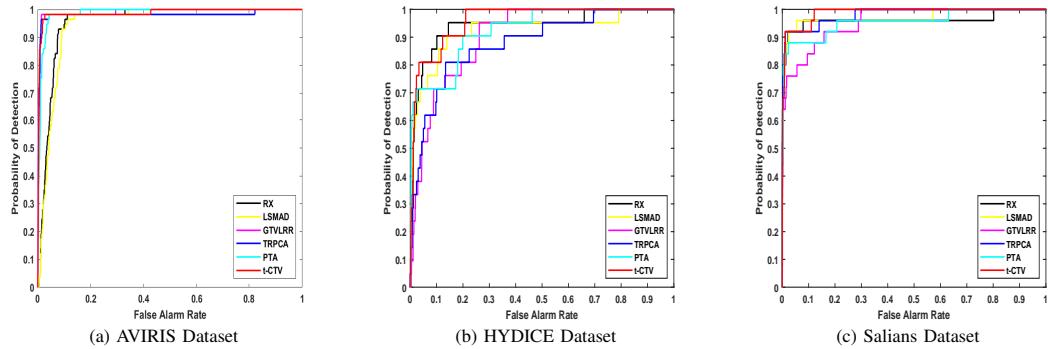


Fig. 10: The ROC curves of different competing methods on the three datasets.

the anomaly/foreground tensor along the spectral/time dimension. Then the t-CTV based model can be formulated, as

$$\min_{\mathcal{T}, \mathcal{E}} \|\mathcal{T}\|_{t\text{-CTV}} + \lambda \|\mathcal{E}\|_{2,1}, \quad \text{s.t., } \mathcal{M} = \mathcal{T} + \mathcal{E}, \quad (\text{H1})$$

where $\mathcal{M} \in \mathbb{R}^{M \times N \times B}$ is the fully observed hyperspectral image or video, \mathcal{T} and \mathcal{E} represent the background tensor and anomaly/foreground tensor, respectively, and $\|\mathcal{E}\|_{2,1} := \sum_{b=1}^B \|\mathcal{E}(:, :, b)\|_F$. We omit the optimization of model (H1) since it is similar to the algorithm 2 in the main text. The only difference is that the $L_{2,1}$ proximal term can also guarantee the corresponding sub-problem possessing the close-formed solution.

A. Hyperspectral Image Anomaly Detection

We select three simulated hyperspectral images with anomaly groundtruth to test. The first image is a dataset that was collected by AVIRIS in San Diego. The water absorption regions, low-SNR, and bad bands are removed and last retain 189 bands. A sub-image is used that is located in the top-left corner with a spatial size of 100×100 . The second one is from the HYDICE, and the original image is sized $307 \times 307 \times 210$. We preserve 175 bands after removing the low-SNR and water vapor absorption bands. An 80×100 sub-image is cropped from the top-right corner, and the cars and roofs in the image scene are considered as anomalies. The last one is the simulated Saliens dataset, which is in size of $120 \times 120 \times 204$. The false color images and the corresponding ground truth maps of the three used datasets are showed in Fig. 9.

Six related hyperspectral AD methods are selected to make comparisons, including the classic RX [15], two geometric modeling based methods LSMAD [16] and GTVLRR [17], two tensor modeling based methods TRPCA [18] and PTA [19]. For these competing methods, we set the associated parameters by corresponding released codes. For our model (H1), the trade-off parameter λ can be set as the recommended one in the main theory of t-CTV based TRPCA and make some slight adjustment for better performance. In the experiments, the λ is set around 0.01 to 0.1. The detection results are evaluated by the *receiver operating characteristic* (ROC) curve between the false alarm rate and the probability of detection. The more closely the ROC curves are located to the upper left, the better anomaly detection performance is, and the larger the corresponding *area under curve* (AUC) is. Fig. 10 gives the ROC curves of different competing methods on the three datasets, and the corresponding AUC values are listed in Table VI, from which we can find that our t-CTV based hyperspectral AD method achieves the best performance. This can be rationally explained by that our t-CTV can well encode the underlying background tensor's low-rank and smooth structures. Furthermore, we show the 2-D plots of the detection results in Fig. 11. It can be seen that our detect results mark the anomalies more significantly and meanwhile reduce other background components, revealing the superiority of our method.

B. Surveillance Video Foreground/Background Separation

We select the Li dataset for experiments which involves 9 video sequences, named airport, boot, curtain, escalator, fountain, shopping mall, switch light, tree and water surface, respectively. Each one of them is collected using a fixed survivance camera

TABLE VI: The AUC values obtained by different competing methods in hyperspectral image anomaly detection task on the three datasets. The best one is in **bold** and the second is in underline.

Dataset	RX	LSMAD	GTVLRR	TRPCA	PTA	t-CTV
AVIRIS	0.9549	0.9471	0.9913	0.9801	0.9866	0.9871
HYDICE	0.9405	0.9224	<u>0.9436</u>	0.8801	0.9247	0.9604
Saliens	0.9609	0.9721	0.9571	<u>0.9818</u>	0.9571	0.9884

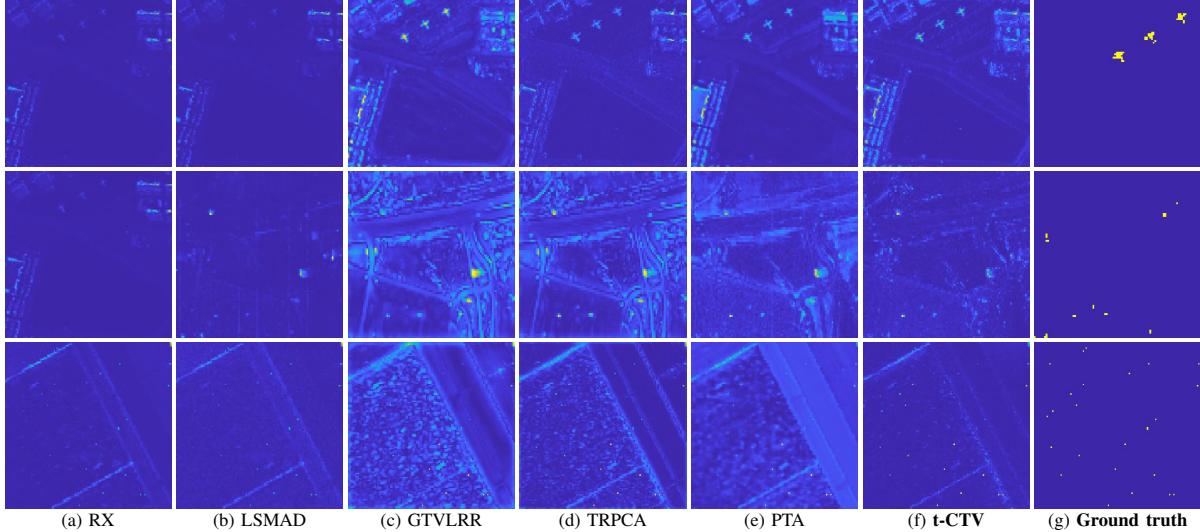


Fig. 11: 2-D plots of the detection results obtained by (a) RX, (b) LSMAD, (c) GTVLRR, (d) TRPCA, (e) PTA and (f) the proposed t-CTV method in hyperspectral image anomaly detection task on the three datasets, (g) the ground truth of anomaly map. From upper to bottom, the dataset is AVIRIS, HYDICE and Saliens, respectively.

TABLE VII: The AUC values obtained by different competing methods in surveillance video foreground/background separation task on the nine datasets. The best one is in **bold** and the second is in underline.

Method	Dataset									Average
	airp.	boot.	curt.	esca.	foun.	shop.	swit.	tree.	wate.	
RPCA	0.8721	0.9168	0.8722	0.9052	0.9418	0.9459	0.9130	0.8917	0.8345	0.8992
VBRPCA	0.8884	0.9172	0.8699	0.9047	<u>0.9401</u>	0.9474	0.8959	0.8869	0.8378	0.8986
RegL1	0.8977	0.9249	0.8899	0.4159	0.9194	0.9443	0.8819	<u>0.8873</u>	<u>0.8920</u>	0.8503
WNNM	0.8994	0.9230	<u>0.8901</u>	0.9079	0.9172	0.9434	0.8836	0.8871	0.8915	<u>0.9047</u>
PRMF	0.8895	0.9217	0.8842	0.9062	0.9154	0.9432	0.8839	0.8867	0.8804	0.9012
MoG	0.8975	<u>0.9231</u>	0.8831	0.9012	0.9098	0.9431	0.8841	0.8829	0.8740	0.8998
LRMR	0.9103	0.9123	0.8531	<u>0.9065</u>	0.8797	0.9202	0.9099	0.8618	0.8583	0.8902
t-CTV	0.9056	0.9054	0.9155	0.9113	0.9085	0.9085	0.9379	0.8682	0.9378	0.9109

in a stable scene. Each video contains 20 frames with pre-given truths of foreground. Eight related methods for surveillance video foreground/background separation are selected to make comparisons, including RPCA [20], VBRPCA [21], EegL1 [22], WNNM [23], PRMF [24], MoG [25] and LRMR [26]. For parameter settings, we follow the instrument according to the original paper suggested. For our method, we choose $\lambda = 0.1$.

In Table VII, we list the AUC values between the separated foreground and corresponding truth of various methods. It shows that our t-CTV method reaches the best performances in average. Besides, Fig.12 gives some separated background and foreground results obtained by various methods. It can be seen that the separated backgrounds of our t-CTV method have less moving subjects (see the second and third examples), and meanwhile the foregrounds are closer to the ground truth. These experimental results further show the superiority of our t-CTV model.

APPENDIX J FURTHER DISCUSSIONS

In this section, we make some further detailed discussions to explain how the proposed single t-CTV simultaneously encode two priors of low-rankness and smoothness, and why it achieves good performance in visual tensor data recovery tasks.



Fig. 12: The separated results of background (upper) and foreground (bottom) obtained by different competing methods in surveillance video foreground/background separation task on boot, curtain and switch light datasets. For the ground truth, the upper is one typical frame of input video, the bottom is its corresponding truth foreground map.

A. How t-CTV Simultaneously Encodes Two Priors?

The characteristics of the t-CTV, and its connections with the related low-rank regularizer and smooth regularizer have been briefly analyzed in Section 4 in the main paper. Here, we make some further detailed supplements to discuss the t-CTV regularization term through both analysis and experiments, showing its connections and differences with related low-rankness prior regularizer *tensor nuclear norm* (TNN) and smoothness prior regularizer *total variation* (TV) norms. Note that all the analysis can be degraded to the matrix sense, which makes the analysis more clearly and easily understood.

Let us first look back to the definitions. The t-CTV norm is very simple that in a form of nuclear norm on the gradient tensors/matrices (called as gradient maps below) of the tensor/matrix data, thus making ones easily associate with the (tensor) nuclear norm on the original (tensor) data that is a classic and direct encoding term for low-rankness, and the other type of smoothness encoding term TV norms including the typical anisotropy one and isotropy one which are L_1 and L_2 (Frobenius norm in formal) on the gradient maps, respectively. Therefore, we need to take efforts to analyze/discuss the potential connections and differences between the t-CTV and TNN as well as TV norms.

Before moving forward, we make some supplements about the low-rankness regularizer, nuclear norm, and the smoothness regularizer, TV norm:

- For the former, the nuclear norm is clear, as the convex surrogate of the rank (even the convex envelope of the corresponding rank function under SVD in matrix case or tensor SVD [2], [27] in tensor case). The consistency between nuclear norm and rank function is well analyzed in [7][3], similar as that between L_1 norm and L_0 norm in classic compressed sensing field [28]. In one word, nuclear norm encodes the low-rankness directly and well.
- For the latter, things are relatively complicated. From the original view, smoothness refers to a certain gradual trend of change in most local areas. “*It assumes that physical properties in a neighborhood of space or in an interval of time present some coherence and generally do not change abruptly. For example, the surface of a table is flat, a meadow presents a texture of grass, and a temporal event does not change abruptly over a short period of time*”, quoted from the Chapter 1.3.3, The Smoothness Prior, in [29]. Essentially, such slow changing phenomenon corresponds to that the values of derivative in most neighborhoods or intervals should be very small if we think of the underlying “surface” as continuous. The smaller, the smoother or flatter. Therefore, the smoothness regularization in general framework can be viewed as an energy term of the derivative, which tends to lower energy for achieving the desired effect. In [29], the author called it as a prior energy string.

The above understanding of promoting low energy of the trend of change, described by derivative mathematically, corresponds to the TV norm minimization for image data, where the gradient maps are the discrete form of the first-order derivatives along the spatial directions. More specifically, the isotropy TV norm minimization constraints the whole gradient maps into a low energy that is defined by the square distance on the difference values of the neighbor elements.

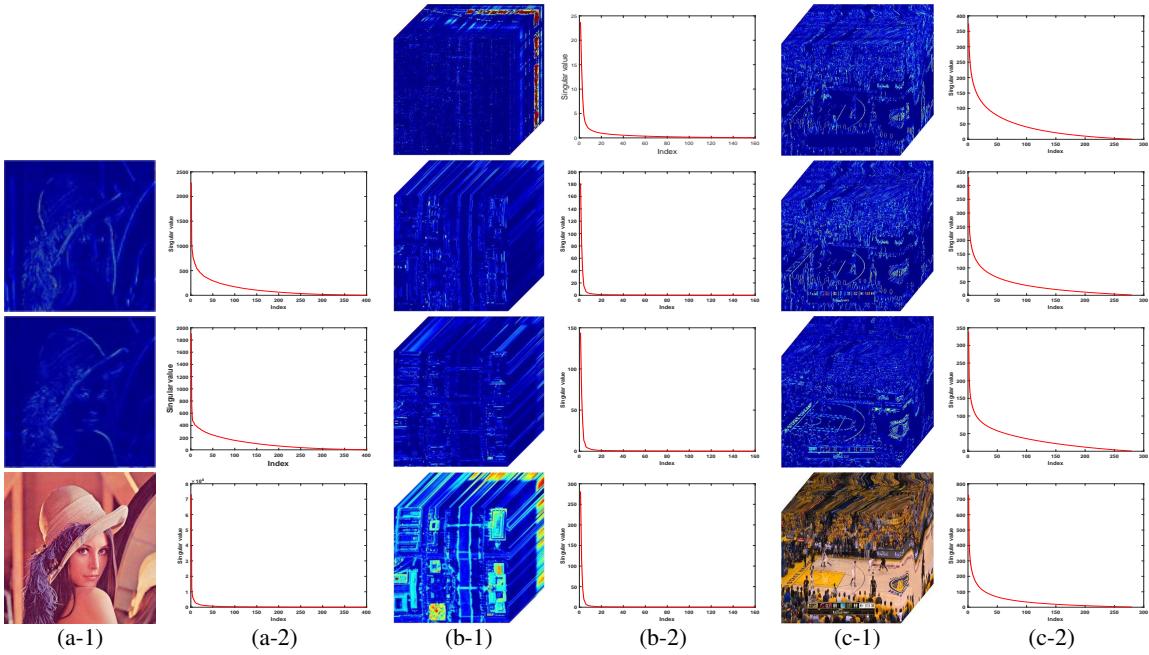


Fig. 13: The curve of singular values of various visual data and corresponding gradient maps. From bottom to upper, plots corresponds to original data, and the gradient maps along the 1st, 2nd and the spectral/time direction. One can see the consistency in low-rankness between the original data and its gradient maps.

A little bit differently, the anisotropy TV norm, which is defined by the absolute distance on the local difference values, can further promote the local elements' difference values as zero exactly. Compared with the isotropy one, the anisotropy TV norm makes the local adjacent pixels are more possible to be equal. That can tend to be more locally flat.

In fact, except defining the smooth regularization using the L_1 or L_2 distance that are conducive to calculation, there have other TV forms can be also used as a type of smooth regularization, like the nonconvex TV [30]–[32] and fractional TV [33]–[35]. The nonconvex TVs are defined by the sum of the element L_q distance of the gradient maps, where q is usually set from 0 to 2, see [30], [31], or other nonconvex penalty function, like logarithm based TV [32]. These TV terms can also be viewed as certain prior energy string like the statement in [29], with the same purpose of controlling the energy of the discrete gradient maps. So the same for the fractional TVs. Differently, they are defined on the fractional derivative for modeling more complicated smooth structures. In actual, choosing which TV variant is data dependent.

To sum up, we think that the core of encoding smoothness prior is to constraint the energy of the measures of the trend of local changes. Normally, the norms are suitable for this purpose and also easy in optimization, thus inducing these TV norms.

Now, we discuss the connections between the t-CTV and TNN, TV norms as follows:

- From the perspective of low-rankness prior encoding, things are very clear. Intuitively, the t-CTV norm constrains the TNN metric of correlated gradient tensors \mathcal{G}_k (see Definition 8 in the main paper), which directly promotes \mathcal{G}_k 's low-rankness property, and then encodes the similar prior structure of the original tensor since there exists strict consistency in low-rankness between the original data and its gradient maps. This can be analyzed by that the gradient maps have the nearly same rank compared with original tensor data since there is only a difference linear operation, i.e., for $\mathcal{T} \in \mathbb{R}^{n_1 \times \dots \times n_d}$ with t-SVD rank R , it can be verified that

$$R - 1 \leq \text{rank}_{\text{t-SVD}}(\mathcal{G}_k) \leq R, \quad (\text{I-1})$$

where \mathcal{G}_k is the gradient tensor along its k -th mode. Besides, it can be also observed that many actual data's gradient maps have the singular values of decay, e.g., RGB images, hyperspectral images, and videos. This is similar to the case of original data. See Fig. 13 for examples. One can see such consistency in low-rankness evidently. Thus, the t-CTV can indirectly induce the expected low-rankness prior structure of the original tensor using the low-rankness metric on the correlated gradient tensors.

- From the perspective of smoothness prior encoding, note that the t-CTV is a well-defined norm in the gradient domain, and thus it measures as an energy control term that tends the discrete first-order derivatives of the tensor data along certain smoothness prior modes being small numerical values to some extent. This is similar to the existing TV norms and the understandings behind. From the general thought of the smoothness regularization, i.e., the view of energy minimization of the measures of the trend of local changes, we can find the t-CTV norm (i.e., nuclear norm on the gradient maps) and

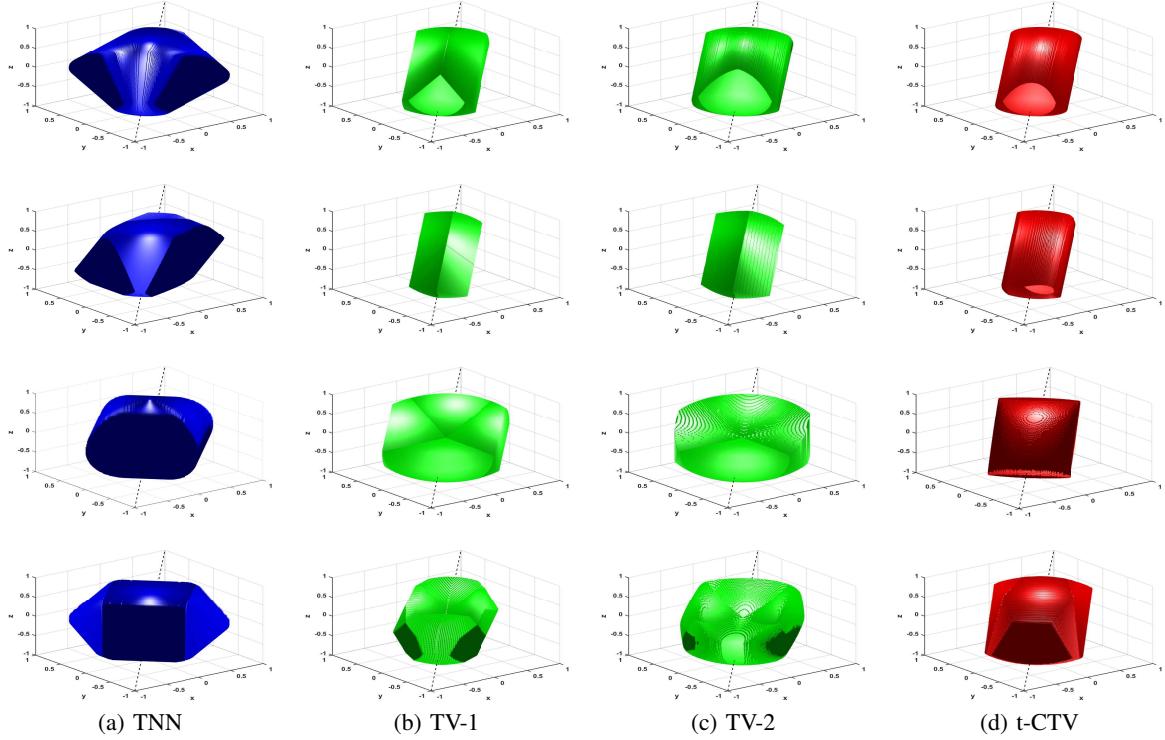


Fig. 14: Manifolds of TNN, TV-1, TV-2 and t-CTV norms. One can see that the manifolds of t-CTV are similar to that of the TV norms in the whole shape.

the TV norms (i.e., L_1 , L_2 norms or other norms on the gradient maps) are consistent in controlling the trend of changes in local areas since they are all certain measure constraint on the gradient maps. They all tend to be smaller when the original data becomes smoother, and take the minimum value of zero only if data is absolutely flat. This indicates, to some extent, that the t-CTV has potential to constraint the smoothness prior structure of the target tensor like the two typical isotropy and anisotropy TV norms, and other TV variants like the nonconvex ones and fractional ones. Take a simple analysis in matrix case. For matrix data X with rank r , G is a gradient map of X whose rank is also r , then the t-CTV norm is $\|G\|_* = \sum_{i=1}^r \sigma_i$, where $\sigma_i > 0$ is the i -th singular value of G . Then consider the isotropy TV norm of X , which is the Frobenius norm of G , i.e., $\|G\|_F = \sqrt{\sum_{i=1}^r \sigma_i^2}$. From their forms of t-CTV norm and isotropy TV norm, it is expected that the t-CTV norm minimization could promote or control the TV norm minimization partly. At least, they all constructed on the gradient maps, and these gradient maps themselves involve the local information.

- To more intuitively observe the connections between the defined t-CTV norm and the pure low-rank as well as the pure smooth regularizer, i.e., TNN and TV norms, here, we plot their manifold diagrams. The manifold polytope constrains the solution space based corresponding regularization norm and plotted on some simple two-mode and three-mode tensors. The corresponding manifolds are depicted in Fig. 14 for easy visual comparison. Specifically, from top to bottom of the figure, the manifolds of constant penalty are respectively for

1) 3×3 circulant matrix:

$$\begin{bmatrix} x & z & y \\ y & x & z \\ z & y & x \end{bmatrix}, \quad (\text{I-2})$$

2) 3×3 symmetric matrix:

$$\begin{bmatrix} x & y & z \\ y & z & y \\ z & y & x \end{bmatrix}, \quad (\text{I-3})$$

3) 2×2 random matrix:

$$\begin{bmatrix} 0 & x \\ y & z \end{bmatrix}, \quad (\text{I-4})$$

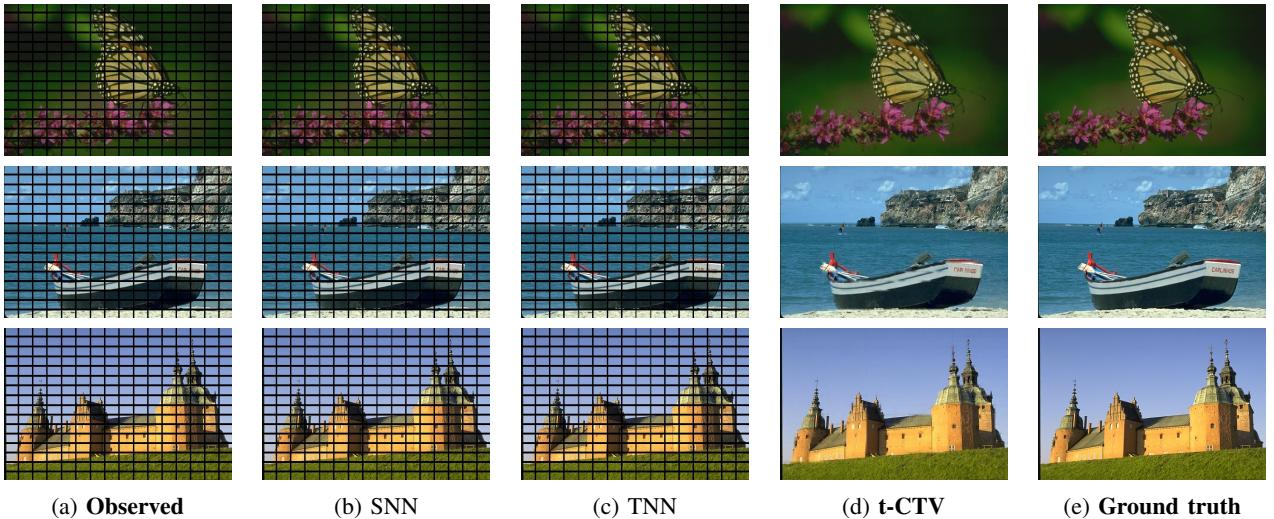


Fig. 15: Color image inpainting results by classical pure low-rankness regularization methods SNN [12], TNN [13] and the proposed t-CTV method in entire rows and columns missing case. One can see that the t-CTV model can handle this while the pure low-rank based ones are completely failed since they only model the structure correlations in whole, indicating the t-CTV model encodes beyond the low-rankness although it consists of the low-rank metric terms.

4) $2 \times 2 \times 2$ random tensor \mathcal{T} with slices

$$\mathcal{T}(:,:,1) = \begin{bmatrix} 0 & x \\ y & z \end{bmatrix}, \quad \mathcal{T}(:,:,2) = \begin{bmatrix} y & z \\ x & 0 \end{bmatrix}. \quad (\text{I-5})$$

One can see that the manifold of the t-CTV norm has evident similarity to that of the TV norm in shape. Note that the manifold polytope can be used to observe the characteristics in constraining the solution space based corresponding regularization norm. From this view, one can find the close connections in encoding smoothness prior between the t-CTV and TV norms. About this point, we construct an experiment in the subsequent for further verification. In addition, there also exist some similar characteristics between the manifold of the TNN and t-CTV norms, like the frontal surfaces in the last column. These observations intuitively depict the encoding capability of the t-CTV norm for both low-rankness and smoothness priors simultaneously.

In summary, it is expected that the t-CTV can finely encode both low-rankness and smoothness priors. It should be noted that although the t-CTV is not the unique encoding term for low-rankness nor smoothness prior, surprisingly, it achieves evidently better performance for joint low-rank and smooth tensor recovery than existing pure low-rankness, smoothness and their combined models. See the experiments part in the main paper. Below, we discuss some differences or advantages of the t-CTV, compared with the pure low-rankness regularizer, the pure smoothness regularizer, and their combined models.

- Compared with pure low-rankness regularization model, typically the SNN [36], TNN [1], we find that our t-CTV minimization model can handle structural missing tasks while the classical methods with a low-rankness regularization term only failed completely. An example for RGB image inpainting is given in Fig. 15. One can see that the t-CTV model can handle the entire columns and rows missing case while the pure low-rank based ones completely failed since they only model the structure correlations in whole. There exist some related analysis in [7], [37] pointing out that the nuclear norm can not work for such entire columns and rows missing case since it only models the whole correlation that cannot affect the local structures. However, although our proposed t-CTV consists of the low-rank metric terms (i.e., the TNN) intuitively, it still performs well in this case. This indicates that the t-CTV model encodes beyond the low-rankness in the whole structure.
- Compared with pure smoothness regularization model, the advantages of the t-CTV are very clear since it considers the correlation in whole structure. The reason is that the smoothness regularization usually need to embed into a low-rank tensor/matrix framework and only itself based models' performance are normally not well enough since the correlation of whole structure, i.e., low-rankness, is basic and important. That is also why these always emerging the type of low-rankness plus smoothness models in the field.
- Compared with the combined low-rankness and smoothness model, in a form of an additional trade-off parameter, the advantage of the proposed t-CTV model is parameter free. It is better than the existing low-rank plus smooth models no matter what value the trade-off parameter in these low-rank plus smooth methods set. More details can be found in the next subsection where the Fig. 19 gives the experimental results in image inpainting tasks.

At last, we show an interesting experiment showing each iterative solution of the pure low-rank model (i.e., TNN), pure

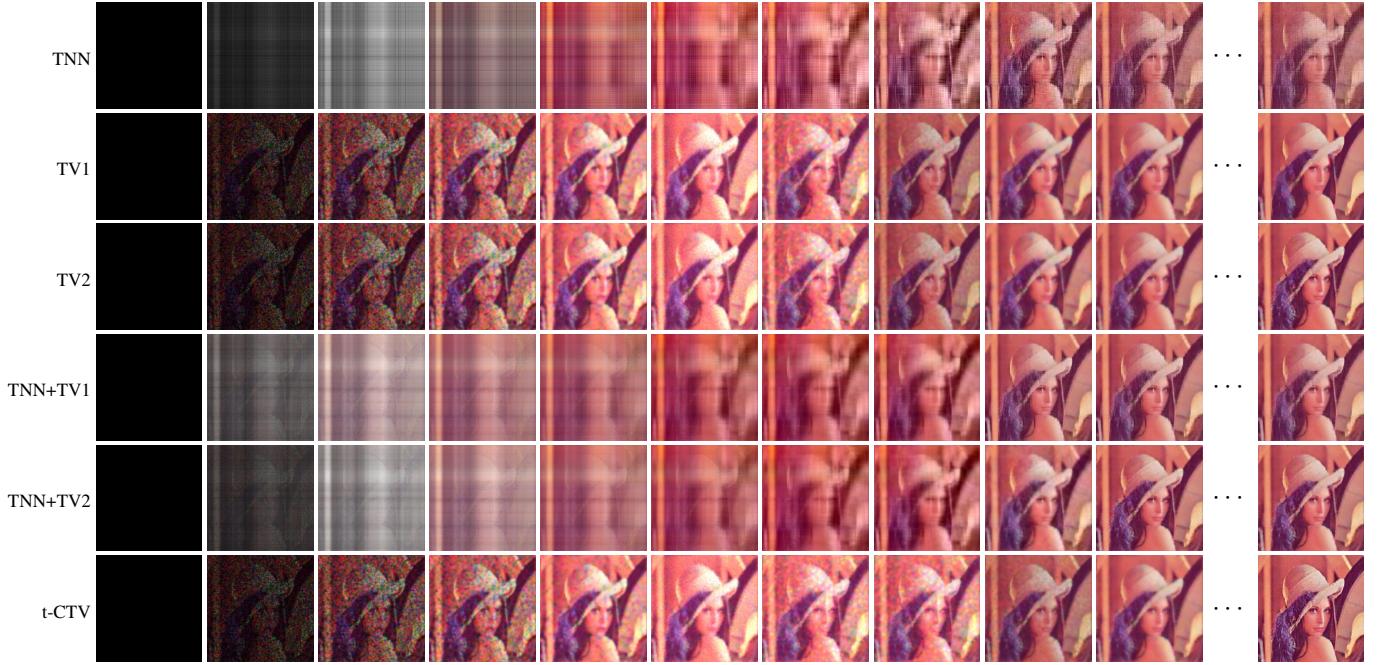


Fig. 16: Comparison the iterative solution of TNN, TV, TNN+TV, and t-CTV models for image inpainting. One can see that the iterative solutions of the t-CTV model are very similar to that of the TV models (TV1 and TV2). There exist many local spots in their iterative solutions, meaning that the updating process promote the local smoothness structure. This indicates that the t-CTV model can encode the smoothness prior to some extent.

smooth model (i.e., TV), low-rank plus smooth model (TNN+TV) and the t-CTV model, observing that the t-CTV model indeed has the ability in constraining local structures that is very similar to the isotropy TV model, and achieving adaptive trade-off between the whole and local information that is very different from the pure low-rank model.

- We use the same optimization framework ADMM solver with zero initialization for the TNN minimization, two TV norm (isotropy TV, dubbed as TV1, anisotropy TV, dubbed TV2) minimization, and their combined models (TNN+TV1, TNN+TV2), and the t-CTV model in RGB image inpainting experiment. The selected iterative solutions of these models are shown in Fig. 16. We can clearly see that the iterative solutions of the model involving the low-rankness term TNN (i.e., TNN, TNN+TV1, TNN+TV2) have the very similar characteristics in updating, where many vertical and horizontal strips can be observed. Instead, the iterative solutions of the t-CTV model are very similar to that of the TV models. There are many local spots in their iterative solutions, meaning that the updating process promotes the local smoothness structure. This indicates that the t-CTV model can indeed encode the smoothness prior to some extent.
- To further observe the property of the t-CTV in encoding smoothness prior, we compute a quantity named ‘Degree of Smoothness’ of each iterative solution, defined as the ratio of the number of values less than a given threshold value ζ in the gradient maps. The changing curves of ‘Degree of Smoothness’ with threshold value $\zeta = 10^{-2}, 10^{-3}$ are plotted in Fig. 17, where the black line is the baseline value of the ground truth. The larger value of ‘Degree of Smoothness’ means a smoother image. One can see that when the threshold value is set as 10^{-2} that is a relatively normal and suitable level, the iterative updating solutions of the t-CTV model first have a large ‘Degree of Smoothness’ like the iterations of the TV term involved models (TV1, TV2, TNN+TV1, TNN+TV2), and finally converges to the same level of the ground truth. Compared with the TV involved models (TV1, TV2, TNN+TV1, TNN+TV2), their solutions have the ‘Degree of Smoothness’ larger than that of the ground truth, meaning that there exists over smoothing drawbacks for these TV involved models. Instead, the pure low-rank model, TNN, models only the global low-rankness while ignores the local smoothness, making the final ‘Degree of Smoothness’ is very small.
- Besides, we also plot the cases of ‘Degree of Smoothness’ with smaller threshold values 10^{-3} . One can observe that the anisotropy TV (i.e., TV1) model and the combined low-rank and smooth one (i.e., TNN+TV1) obtain the ‘Degree of Smoothness’ larger than the baseline, meaning the over smoothing results. In other words, they make many adjacent elements in local smooth area exactly equal, which we have analyzed before since the relied L_1 norm has the ability to promote the absolute sparsity. By contrast, the other type of TV norm, isotropy TV or TV2, can also encode the local smoothness but not the absolute flat in local area, whose involved models (TV2, TNN+TV2) have the smaller value of ‘Degree of Smoothness’. Still, the t-CTV model can tend the solution with smoothness whose value of ‘Degree of Smoothness’ is larger than that of the TNN which entirely ignores the local smoothness.

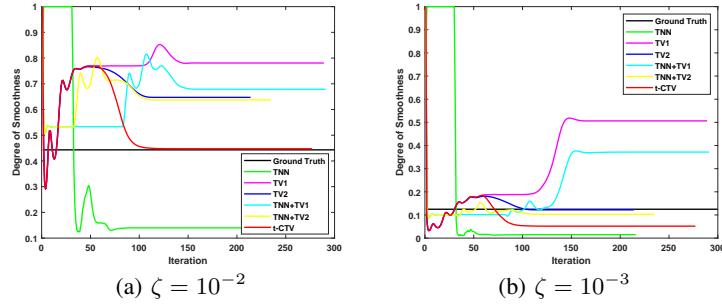


Fig. 17: Comparison of the 'Degree of Smoothness' in each iteration of the TNN, TV, TNN+TV and t-CTV models with threshold value $\zeta = 10^{-2}, 10^{-3}$. One can see that the 'Degree of Smoothness' of the t-CTV converges to that of the ground truth while the TV involved models, especially the TV1, obtain the over smoothing results, and the pure low-rank model TNN has no ability to promote smoothness prior.

B. Evaluating t-CTV along Its Partial Modes

It should be noted that the defined t-CTV is a combination of multiple tensor nuclear norms imposed on the gradient tensor along different modes. Thus, from Prop. 1, it can be deduced that there exists the optimal direction or mode such that the lower bound of sampling complexity only relies on this term. Back to (22) in Prop. 1, the κ_{\min} becomes $\min\{\|\nabla_k(\tilde{T}_0)\|_{\otimes, \mathcal{L}} / L_k, k \in \Gamma\}$, where L_k is the Lipschitz constant of $\|\nabla_k(\cdot)\|_{\otimes, \mathcal{L}}$ that is the same for all k . This means the lower bound is determined by the smallest "partial" t-CTV that represents the most smooth and also low-rank one along such mode. In the subsequent proposed Theorem 5, the result for "whole" t-CTV model is also consistent with this, where we assume the gradient tensor's sparsity (number of nonzero entries) along the k -th mode as S_k , and $S = \min_{k \in \Gamma}\{S_k\}$. Then the corresponding result is only related to S . Similar findings have also been analyzed in reference [1][2].

Considering the recovery performance in experiments, i.e., the evaluation index of accuracy like PSNR, it should be emphasized that this theoretical result could only reflect the general tendency yet not fully correspond to actual performances. Note that the results in Prop. 1 mainly discuss the lower bound of sampling complexity for totally failed recovery, which is different from the recovery error evaluation in actual experiments. In fact, stacking multiple effective regularizers could always further improve recovery performance in terms of certain evaluation metrics. Yet the derived conclusion in Prop. 1 does be helpful for us to select partial of tensor modes in the t-CTV norm, i.e., we should choose a few ones along which the tensor is more significantly low-rank and smooth.

To more intuitively demonstrate this point, Fig. 18 gives an example for RGB image inpainting where we consider 7 cases: using the two spatial modes and one RGB mode, respectively, and their combined cases. At the same time, the results obtained by the baseline models TNN and TNN+TV is also given. It shows that the mode 3, i.e., the RGB mode, has little effect for performance improvement. This can be rationally explained by that the RGB mode is significantly different from the other two spatial modes that have well smoothness. Besides, the used "partial" t-CTV in color image inpainting, including only the modes 1 and 2, is much better than the baselines.

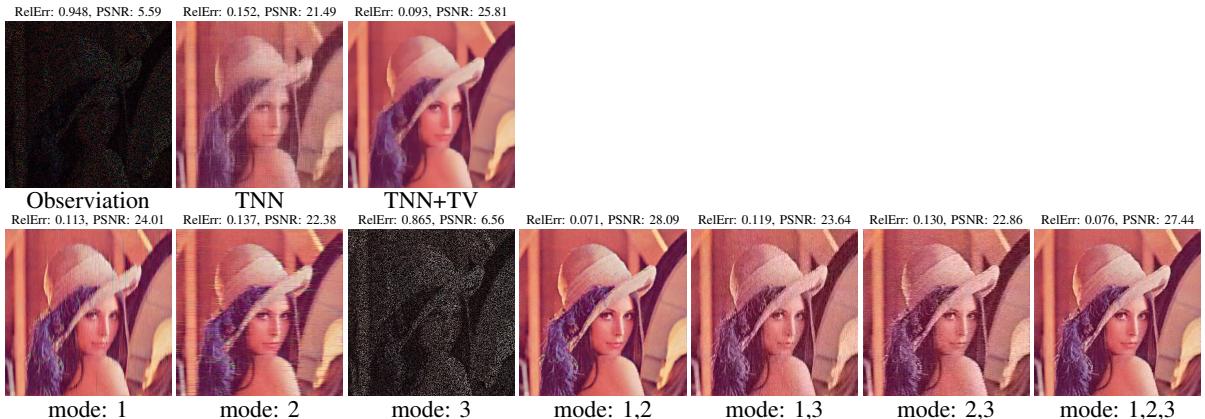


Fig. 18: Comparison of the color image inpainting results with proposed t-CTV model using different mode settings, see the second row. The first row depicts the observation and the results of two baseline models TNN and TNN+TV. The missing rate in this experiment is 90%.

C. Why Better Performance?

Firstly, the most important reason should be the usage of t-CTV. The t-CTV is a single regularizer that can characterize the low-rank and smooth prior simultaneously, rationally avoiding the trade-off parameter. Proposition 1 and Theorem 7 provide a theoretical explanation, and the simulated phase transition results Fig. 4 and Fig. 5 in the main paper have already confirmed this point. Here, we show some experimental results on real data to further support this conclusion. Consider the TC task via the TNN plus TV method, i.e.,

$$\|\mathcal{T}\|_{\otimes, \mathfrak{L}} + \alpha \|\mathcal{T}\|_{\text{TV}} \quad \text{s.t.} \quad \mathcal{P}_{\Omega}(\mathcal{T}) = \mathcal{P}_{\Omega}(\mathcal{T}_0). \quad (\text{I-6})$$

Under different settings of the trade-off parameter α from 10^{-6} to 10^6 , we compare the recovered results of the TNN plus TV method and our t-CTV method in the task of color image inpainting. Fig. 19 shows the mean PSNR values of TNN plus TV under SR = 5%, 10%, 30% and 50%, and also the PSNR values of our t-CTV method and related pure TNN and TV methods for comparison. It can be evidently observed that although the TNN plus TV method degrades to the pure TNN when α is very small and to the pure TV when α is very large, and achieves a better performance when α locates with certain moderate values. Yet in all SR cases, our t-CTV is always significantly better than all the TNN, TV, and the TNN plus TV methods with evidently higher PSNR values.

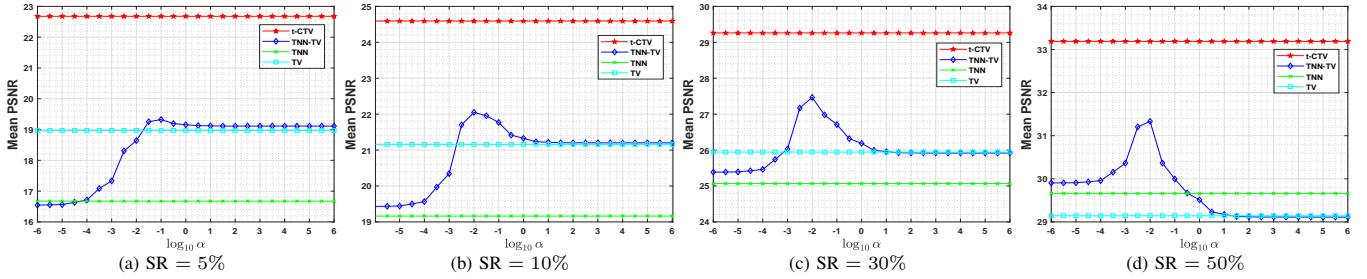


Fig. 19: Performance comparison in terms of mean PSNR among t-CTV, TNN, TV, and TNN plus TV model with trade-off parameter α varying from 10^{-6} to 10^6 .

TABLE VIII: Comparison of t-CTV and CTV in inpainting tasks on color videos. Each value averaged over all 10 videos.

SR	5%		10%		30%		50%	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
CTV	26.21	0.803	28.12	0.858	32.67	0.941	36.53	0.973
t-CTV	31.04	0.891	34.15	0.933	40.72	0.980	45.79	0.993

TABLE IX: Comparison of t-CTV and CTV in denoising tasks on color images. Each value is averaged over all 50 images.

Noise	S = 0.1		S = 0.2		S = 0.3		S = 0.4	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
CTV	28.90	0.880	28.12	0.862	27.37	0.841	26.57	0.814
t-CTV	31.66	0.932	30.56	0.913	29.36	0.884	27.97	0.836

TABLE X: Comparison of DFT based t-CTV (t-CTV(F)) and DCT based t-CTV (t-CTV(C)) in inpainting tasks

SR	5%		10%		20%		40%	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Color Images								
t-CTV(F)	22.67	0.682	24.59	0.763	27.18	0.849	31.21	0.931
t-CTV(C)	20.80	0.581	24.19	0.738	27.38	0.850	31.81	0.938
Color Videos								
t-CTV(F)	31.04	0.891	34.15	0.933	37.92	0.966	43.25	0.988
t-CTV(C)	31.63	0.905	35.13	0.945	39.35	0.974	45.05	0.991

Secondly, a fundamental and necessary point is that our method directly processes these visual data in tensor format just as their original form, making the structural information is not compromised to a large extent. Here, we conduct two experiments to show the benefit of tensor modeling. Table VIII and Table IX list the quantitative comparison of the t-CTV based TC and TRPCA models with corresponding degrade models in matrix space, respectively. Specifically, the tensor data is degraded as

TABLE XI: Comparison of DFT based t-CTV (t-CTV(F)) and DCT based t-CTV (t-CTV(C)) in denoising tasks

Method	S = 0.1		S = 0.2		S = 0.3		S = 0.4	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Color Images								
t-CTV(F)	31.66	0.932	30.56	0.913	29.36	0.884	27.97	0.836
t-CTV(C)	33.08	0.949	31.63	0.929	30.12	0.898	28.41	0.844
Hyperspectral Images								
t-CTV(F)	46.32	0.993	44.92	0.991	43.27	0.989	41.37	0.985
t-CTV(C)	46.90	0.996	45.66	0.994	43.82	0.992	41.99	0.988

its unfolding matrix in models (17) and (18) and the t-CTV is degraded as CTV, i.e.,

$$\|\mathcal{T}\|_{\text{CTV}} = \frac{1}{\gamma} \sum_{k \in \Gamma} \|G_k\|_*, \quad (\text{I-7})$$

where G_k is the unfolding matrix of the corresponding gradient tensor \mathcal{G}_k along the k -th dimension and $\|\cdot\|_*$ denotes the matrix nuclear norm. From these results, one can easily see the superiority of our modeling manner based on tensors, where the PSNR outperforms nearly $5 \sim 8$ dB in color video inpainting task and $1 \sim 3$ dB in color image denoising task.

Last, it should be indicated that the usage of high-order t-SVD framework is also very important. Note that the t-SVD framework is orientation dependent and relies on the specific transforms like DFT, DCT since the induced TNN has an equivalent form

$$\|\mathcal{T}\|_{\circledast, \mathfrak{L}} = \frac{1}{\ell} \|\text{bdiag}(\mathfrak{L}(\mathcal{T}))\|_*. \quad (\text{I-8})$$

Such a characteristic makes the t-SVD able to well extract the low-dimensional structure of visual tensor data since many visual tensors are first constructed by two spatial dimensions as the main information agent and then equip another spectral, temporal or directional dimensions to rich the information. At the same time, the transform operation along the spectral or temporal dimension acts as an aggregator of the inherent multi-dimensional structural characteristics. As a supplement, we present some experimental results when using the DCT transform. Table X and XI demonstrate the quantitative comparison of our proposed t-CTV method in visual data inpainting and denoising under the DFT based t-SVD and the DCT based t-SVD. It shows that the recovery performance can be further improved under the DCT based t-SVD framework. This phenomenon is consistent with the results in [1]. In fact, it is still an open topic that which transform should be the best at present [3], which is an important issue worthy to be explored in our future research.

REFERENCES

- [1] W. Qin, H. Wang, F. Zhang, J. Wang, X. Luo, and T. Huang, "Low-rank high-order tensor completion with applications in visual data," *IEEE Trans. Image Process.*, vol. 31, pp. 2433–2448, 2022.
- [2] M. E. Kilmer and C. D. Martin, "Factorization strategies for third-order tensors," *Linear Alg. Appl.*, vol. 435, no. 3, pp. 641–658, 2011.
- [3] M. E. Kilmer, L. Horesh, H. Avron, and E. Newman, "Tensor-tensor algebra for optimal representation and compression of multiway data," *Proc Natl Acad Sci U S A*, vol. 118, no. 28, 2021.
- [4] C. D. Martin, R. Shafer, and B. LaRue, "An order-p tensor factorization with applications in imaging," *SIAM J. Sci. Comput.*, vol. 35, no. 1, pp. A474–A490, 2013.
- [5] E. Kershner, M. Kilmer, and S. Aeron, "Tensor–tensor products with invertible linear transforms," *Linear Alg. Appl.*, vol. 485, pp. 545–570, 2015.
- [6] G. Matsaglia and G. PH Styan, "Equalities and inequalities for ranks of matrices," *Linear Multilinear Algebra*, vol. 2, no. 3, pp. 269–292, 1974.
- [7] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.
- [8] D. Gross, "Recovering low-rank matrices from few coefficients in any basis," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1548–1566, 2011.
- [9] J. A. Tropp, "User-friendly tail bounds for sums of random matrices," *Found. Comput. Math.*, vol. 12, no. 4, pp. 389–434, 2012.
- [10] S. Oymak, A. Jalali, M. Fazel, Y. C. Eldar, and B. Hassibi, "Simultaneously structured models with application to sparse and low-rank matrices," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2886–2908, 2015.
- [11] R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018. [Online]. Available: <https://books.google.com.hk/books?id=TahxDwAAQBAJ>
- [12] Y. Chen, "Incoherence-optimal matrix completion," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2909–2923, 2015.
- [13] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis with a new tensor nuclear norm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 925–938, 2019.
- [14] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.
- [15] I. S. Reed and X. Yu, "Adaptive multiple-band cfar detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1760–1770, 1990.
- [16] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, "Anomaly detection in hyperspectral images based on low-rank and sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 1990–2000, 2015.
- [17] T. Cheng and B. Wang, "Graph and total variation regularized low-rank representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 391–406, 2019.
- [18] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Joint reconstruction and anomaly detection from compressive hyperspectral images using mahalanobis distance-regularized tensor RPCA," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2919–2930, 2018.
- [19] L. Li, W. Li, Y. Qu, C. Zhao, R. Tao, and Q. Du, "Prior-based tensor approximation for anomaly detection in hyperspectral imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, 2020.
- [20] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 1–37, 2011.
- [21] S. D. Babacan, M. Luessi, R. Molina, and A. K. Katsaggelos, "Sparse bayesian methods for low-rank matrix estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 3964–3977, 2012.
- [22] Y. Zheng, G. Liu, S. Sugimoto, S. Yan, and M. Okutomi, "Practical low-rank matrix approximation under robust 1 1-norm," in *CVPR*, 2012, pp. 1410–1417.
- [23] S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng, and L. Zhang, "Weighted nuclear norm minimization and its applications to low level vision," *Int. J. Comput. Vision*, vol. 121, no. 2, pp. 183–208, 2017.
- [24] N. Wang, T. Yao, J. Wang, and D.-Y. Yeung, "A probabilistic approach to robust matrix factorization," in *ECCV*, 2012, pp. 126–139.
- [25] D. Meng and F. De La Torre, "Robust matrix factorization with unknown noise," in *ICCV*, 2013, pp. 1337–1344.
- [26] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4729–4743, 2013.
- [27] C. Lu, X. Peng, and Y. Wei, "Low-rank tensor completion with a new tensor nuclear norm induced by invertible linear transforms," in *CVPR*, 2019, pp. 5996–6004.
- [28] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [29] S. Z. Li, *Markov random field modeling in computer vision*. Springer Science & Business Media, 2012.
- [30] M. Hintermüller and T. Wu, "Nonconvex tv γ -q-models in image restoration: Analysis and a trust-region regularization-based superlinearly convergent solver," *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1385–1415, 2013.
- [31] X. Zhang, M. Bai, and M. K. Ng, "Nonconvex-tv based image restoration with impulse noise removal," *SIAM J. Imag. Sci.*, vol. 10, no. 3, pp. 1627–1667, 2017.
- [32] B. Zhang, G. Zhu, and Z. Zhu, "A tv-log nonconvex approach for image deblurring with impulsive noise," *Signal Processing*, vol. 174, p. 107631, 2020.
- [33] D. Chen, S. Sun, C. Zhang, Y. Chen, and D. Xue, "Fractional-order tv-l2 model for image denoising," *Central European Journal of Physics*, vol. 11, no. 10, pp. 1414–1422, 2013.
- [34] Z. Ren, C. He, and Q. Zhang, "Fractional order total variation regularization for image super-resolution," *Signal Processing*, vol. 93, no. 9, pp. 2408–2421, 2013.
- [35] W. Lian and X. Liu, "Non-convex fractional-order tv model for impulse noise removal," *J. Comp. Appl. Math.*, vol. 417, p. 114615, 2023.
- [36] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, 2012.
- [37] C. Li, W. He, L. Yuan, Z. Sun, and Q. Zhao, "Guaranteed matrix completion under multiple linear transformations," in *CVPR*, 2019, pp. 11136–11145.