GENERALIZED DANTZIG SELECTOR FOR LOW-TUBAL-RANK TENSOR RECOVERY

Andong Wang[1,2,3], *Xulin Song*[1,2], *Xiyin Wu*[1,2], *Zhihui Lai*[4] *and Zhong Jin*[1,2*]

[1]School of Computer Science and Engineering, Nanjing University of Science & Technology, China
[2]Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, Nanjing University of Science & Technology, China
[3] Tensor Learning Unit, RIKEN Center for Advanced Intelligence Project (AIP), Japan
[4]School of Computer Science and Software Engineering, Shenzhen University, China

ABSTRACT

Due to the superiority in exploiting the ubiquitous "spatial-shifting" property in modern multi-way data, the recently proposed low-tubal-rank model has been successfully applied for tensor recovery in signal processing and computer vision. In this paper, we define the generalized tensor Dantzig selector to recover a low-tubal-rank tensor from noisy linear measurements. Algorithmically, we develop an efficient algorithm based on the ADMM framework. Statistically, we establish non-asymptotic upper bounds on the estimation error for the problems of tensor completion and compressive sensing. Numerical experiments illustrate that our bounds can predict the scaling behavior of the estimation error. Experiments on real-word datasets show the effectiveness of the proposed model.

Index Terms— Dantzig selector, tensor completion, compressive sensing, tubal nuclear norm, statistical performance

1. INTRODUCTION

As a multi-linear extension of vector and matrix, tensor has intrinsic advantages in modeling multi-way correlations. In many applications, most variations of the data tensor can be dominated by a relatively small number of intrinsic factors, which can be well modeled by tensor low-rankness [1]. Different from the uniquely defined matrix rank, a tensor has multiple definitions of rank function. The most popular tensor ranks are the CP rank [2] and the Tucker rank [3]. Recently, the tensor tubal rank [4], induced by the tensor singular value decomposition (t-SVD) [5], was proposed as a new tensor complexity measure and has been applied in may tensor recovery tasks like image/video inpainting/de-noising [4, 6, 7].

The problem of tensor recovery from a few noisy linear measurements finds many applications in signal processing [8–11]. In general, recovering a tensor from incomplete measurements is an ill-posed problem, especially when the quantity of information carried by the observations does not significantly exceed the degree of freedom (DoF) [12]. To make the problem well-posed, we assume the tubal rank r^* of the underlying tensor $\mathcal{L}^* \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ is low. Then the degree of freedom in $\mathcal{L}^*$ is at most $r^*(d_1 + d_2 - r^*)d_3$ [13], which is significantly smaller than the entry number $d_1 d_2 d_3$. In the noiseless setting, by tensor tubal nuclear norm (TNN) minimization, it is proved that $O(r^*(d_1 + d_2 - r^*)d_3)$ Gaussian measurements are sufficient for exact recovery of $\mathcal{L}^*$, and $O(r^* \max\{d_1, d_2\} d_3 \log^2(d_1 d_3 + d_2 d_3))$ observations sufficient for exact tensor completion provided $\mathcal{L}^*$ satisfies the tensor coherence condition (TIC) [8]. In the noisy setting, one needs $O(r^* \max\{d_1, d_2\} d_3 \log(d_1 d_3 + d_2 d_3))$ observations for approximate tensor completion via an iterative singular tube thresholding algorithm (ISTT) [14]. In [8], an l_2-norm constrained TNN minimization model ("l_2-Con" for short) is also defined for tensor compressive sensing.

In compressive sensing, the Dantzig Selector [15] is an alternative to the l_2-norm constrained model and the regularized approaches (like Lasso) for sparse recovery. Recently, the generalized Dantzig selector has attracted much attention in low-rank matrix recovery [16, 17]. In [17], the generalized matrix Dantzig selector has shown a typically sharper upper bound on the estimation error than the l_2-norm constrained norm minimizer.

In this paper, we first define the generalized tensor Dantzig selector for low-tubal-rank tensor recovery. Then, an algorithm based on the alternating direction method of multipliers (ADMM) is presented. Further, non-asymptotic upper bounds on the estimation error are established for two typical tensor recovery problems, i.e., tensor completion and tensor compressive sensing. Synthetic and real data experiments verify the sharpness of the proposed upper bounds and the effectiveness of the generalized tensor Dantzig selector.

2. NOTATIONS AND PRELIMINARIES

First, the main notations are listed in Table 1. When the field and size of a tensor are not shown explicitly, it is in $\mathbb{R}^{d_1 \times d_2 \times d_3}$.

*✉ *zhongjin@njust.edu.cn*. This work is partially supported by National Natural Science Foundation of China under Grant Nos. 61872188, U1713208, 61602244, 61672287, 61702262, 61773215, 61703209, and by the Natural Science Foundation of Guangdong Province under Grant No. 2017A030313367.

Let $[d] := \{1, \cdots, d\}, \forall d \in \mathbb{N}_+$. Let $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}, \forall a, b \in \mathbb{R}$. For $i \in [d]$, $\boldsymbol{e}_i \in \mathbb{R}^d$ denotes the standard vector basis whose i_{th} entry is 1 with the others 0. For $(i, j, k) \in [d_1] \times [d_2] \times [d_3]$, the outer product $\boldsymbol{e}_i \circ \boldsymbol{e}_j \circ \boldsymbol{e}_k$ denotes a standard tensor basis in $\mathbb{R}^{d_1 \times d_2 \times d_3}$. The matrix nuclear norm $\|\cdot\|_*$ and spectral norm $\|\cdot\|$ are the sum and maximum of the singular values, respectively. For a 3-way tensor, fft3$(\cdot)$ denotes the *fast Fourier transformation* along *the third mode*. Positive constants are denoted by C, c, c_0, etc. Abbreviation *w.h.p.* is short for "with high probability". The identity operator is denoted by $\mathbb{I}$. Let $\mathbb{S}^{D-1}$ be the unit sphere. For simplicity, let $\tilde{d} = (d_1 + d_2)d_3$ and $D = d_1 d_2 d_3$. We also assume $d_1 \geq d_2$ without loss of generality.

Table 1: Some notations

Notation	Descriptions	Notation	Descriptions
$\mathcal{L}^*$	the true tensor	$\hat{\mathcal{L}}$	proposed estimator
$\underline{\underline{\Delta}}$	$\mathcal{L}^* - \hat{\mathcal{L}}$	$\Vert\mathcal{T}\Vert_{sp}$	tensor spectral norm
$\widetilde{\mathcal{T}}$	fft3$(\mathcal{T})$	$\Vert\mathcal{T}\Vert_\star$	tubal nuclear norm
$\mathcal{T}_{ijk}$	$(i, j, k)_{th}$ entry of $\mathcal{T}$	$\Vert\mathcal{T}\Vert_{\mathrm{F}}$	$\sqrt{\sum_{ijk} \mathcal{T}_{ijk}^2}$
$\mathcal{T}(i, j, :)$	$(i, j)_{th}$ tube of $\mathcal{T}$	$\Vert\mathcal{T}\Vert_\infty$	$\max_{ijk} \lvert\mathcal{T}_{ijk}\rvert$
$\mathcal{T}(:, :, k)$	k_{th} frontal slice of $\mathcal{T}$	$\langle \mathcal{A}, \mathcal{B} \rangle$	$\sum_{ijk} \mathcal{A}_{ijk} \mathcal{B}_{ijk}$
$\boldsymbol{\mathcal{T}}_A(\mathcal{L}^*)$	cone$\{\mathcal{T} \mid \Vert\mathcal{L}^* + \mathcal{T}\Vert_\star \leq \Vert\mathcal{L}^*\Vert_\star\}$	$\boldsymbol{\mathcal{E}}_A(\mathcal{L}^*)$	$\boldsymbol{\mathcal{T}}_A(\mathcal{L}^*) \cap \mathbb{S}^{D-1}$
$\mathbb{C}(r) := \left\{\mathcal{T} \mid \Vert\mathcal{T}\Vert_\infty = 1, \frac{\Vert\mathcal{T}\Vert_\star}{\Vert\mathcal{T}\Vert_{\mathrm{F}}} \leq \sqrt{r}, \frac{\Vert\mathcal{T}\Vert_{\mathrm{F}}^2}{D} \geq 8\sqrt{\log \tilde{d}/(N \log(6/5))}\right\}$			

Then, some concepts related to t-SVD will be defined.

Definition 1 (t-product [11]) *Let $\mathcal{T}_1 \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ and $\mathcal{T}_2 \in \mathbb{R}^{d_2 \times d_4 \times d_3}$. Their t-product $\mathcal{T} := \mathcal{T}_1 * \mathcal{T}_2$ is a tensor in $\mathbb{R}^{d_1 \times d_4 \times d_3}$, whose $(i, j)_{th}$ tube $\mathcal{T}(i, j, :) = \sum_{k=1}^{d_2} \mathcal{T}_1(i, k, :) \bullet \mathcal{T}_2(k, j, :)$, where $\bullet$ denotes the circular convolution [5].*

Definition 2 (t-SVD, tubal rank [11]) *Any $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ has tensor singular value decomposition (t-SVD) $\mathcal{T} := \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$, where $\mathcal{U} \in \mathbb{R}^{d_1 \times d_1 \times d_3}$ and $\mathcal{V} \in \mathbb{R}^{d_2 \times d_2 \times d_3}$ are orthogonal, $\mathcal{S} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ is f-diagonal, $(\cdot)^\top$ denotes the tensor transpose[5]. The tensor tubal rank of $\mathcal{T}$ is defined as the number of non-zero tubes of $\mathcal{S}$ in its t-SVD, i.e., $r_{tubal}(\mathcal{T}) := \sum_i \mathbf{1}(\mathcal{S}(i, i, :) \neq \mathbf{0})$.*

Definition 3 (TNN, tensor spectral norm [7]) *Letting $\widetilde{\mathcal{T}} :=$ fft3$(\mathcal{T})$, the tubal nuclear norm (TNN) and the tensor spectral norm of $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ are respectively defined as*

$$\|\mathcal{T}\|_\star := \sum_{k=1}^{d_3} \frac{\|\widetilde{\mathcal{T}}(:, :, k)\|_*}{d_3}, \ \|\mathcal{T}\|_{\mathrm{sp}} := \max_k \|\widetilde{\mathcal{T}}(:, :, k)\|. \quad (1)$$

TNN and tensor spectral norm are dual norms [8]. TNN has been successfully utilized in tensor recovery problems like tensor completion [11] and tensor robust PCA[7].

3. GENERALIZED TENSOR DANTZIG SELECTOR

The true tensor $\mathcal{L}^*$ is assumed to be low tubal rank, i.e., $r^* = r_{\mathrm{tubal}}(\mathcal{L}^*) \ll d_1 \wedge d_2$. Suppose one observes $N \ll D$ scalars

$$y_i = \langle \mathcal{L}^*, \mathcal{X}_i \rangle + \xi_i, \quad \forall i \in [N], \qquad (2)$$

where $\mathcal{X}_i$'s are known random design tensors, and ξ_i's are *i.i.d.* zero-mean Gaussian noise with known variance σ^2. Let $\boldsymbol{y} = (y_1, \cdots, y_N)^\top$ and $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_N)^\top$. Define the design operator $\mathfrak{X}(\mathcal{T}) := (\langle \mathcal{T}, \mathcal{X}_1 \rangle, \cdots, \langle \mathcal{T}, \mathcal{X}_N \rangle)^\top \in \mathbb{R}^N$ with adjoint operator $\mathfrak{X}^*(\boldsymbol{z}) := \sum_{i=1}^N z_i \mathcal{X}_i, \forall \boldsymbol{z} \in \mathbb{R}^N$. Thus, the observation model (2) can be rewritten as $\boldsymbol{y} = \mathfrak{X}(\mathcal{L}^*) + \boldsymbol{\xi}$. With different design tensors, we consider two classical examples:

Eg.1 *Tensor completion.* In tensor completion, the design tensors $\{\mathcal{X}_i\}$ are *i.i.d.* random tensor bases drawn from uniform distribution on the set $\{\boldsymbol{e}_i \circ \boldsymbol{e}_j \circ \boldsymbol{e}_k, \forall (i, j, k) \in [d_1] \times [d_2] \times [d_3]\}$, which serves as an orthonormal basis in the space of $d_1 \times d_2 \times d_3$ tensors.

Eg.2 *Tensor compressive sensing.* When $\mathfrak{X}$ is a *random Gaussian design*, Model (2) is the tensor compressive sensing model with Gaussian measurements [12]. $\mathfrak{X}$ is named a random Gaussian design when $\mathcal{X}_i$'s are random tensors with *i.i.d.* standard Gaussian entries[18].

To recover the true tensor $\mathcal{L}^*$ from noisy observations $\boldsymbol{y}$, we define the *generalized tensor Dantzig selector* as follows:

$$\hat{\mathcal{L}} \in \operatorname*{argmin}_{\mathcal{L}} \|\mathcal{L}\|_\star \ \text{ s.t. } \ \|\mathfrak{X}^*(\boldsymbol{y} - \mathfrak{X}(\mathcal{L}))\|_{\mathrm{sp}} \leq \lambda, \qquad (3)$$

where λ is a parameter. Model (3) is a generalization of the Dantzig selector [15] to the tensor case.

3.1. An ADMM-based Algorithm

The ADMM framework [19] is applied to solve the proposed model. Let $\mathbb{C}_\lambda := \{\mathcal{T} \mid \|\mathcal{T}\|_{\mathrm{sp}} \leq \lambda\}$. Adding auxiliary variables yields an equivalent formulation to Problem (3):

$$\begin{aligned} &\min_{\mathcal{L}, \mathcal{K}, \mathcal{E}} \quad \|\mathcal{L}\|_\star \\ &\ \text{s.t.} \ \ \mathcal{K} = \mathcal{L}, \ \mathcal{E} + \mathfrak{X}^*\mathfrak{X}(\mathcal{K}) = \mathfrak{X}^*(\boldsymbol{y}), \ \mathcal{E} \in \mathbb{C}_\lambda. \end{aligned} \qquad (4)$$

To solve Problem (4), an ADMM-based algorithm is described in Algorithm 1. Since Algorithm 1 is an instance of ADMM, its convergence behavior exactly follows the convergence analysis in [19].

We analyze the computational complexity as follows. Updating $\mathcal{L}$ and $\mathcal{E}$ involves the singular tube thresholding operator $\mathfrak{S}_\tau^{\|\cdot\|_\star}(\cdot)$ [14] which costs $O\big(D(d_1 \wedge d_2 + \log d_3)\big)$; By precomputing $(\mathbb{I} + \mathfrak{X}^*\mathfrak{X}\mathfrak{X}^*\mathfrak{X})^{-1}$ and $\mathfrak{X}^*\mathfrak{X}$ which costs $O(D^3 + ND^2)$, the cost of updating $\mathcal{K}$ is $O(D^2)$; Updating $\mathcal{Y}_k$ $(k \leq 2)$ costs $O(D^2)$. Supposing the iteration number is T, the overall computational complexity will be $O\big(D^3 + TD^2 + TD(d_1 \wedge d_2 + \log d_3)\big)$, which is very expensive for large tensors. In some special cases (like tensor completion) where $\langle \mathcal{X}_i, \mathcal{L} \rangle$ operates on an element of $\mathcal{L}$, $(\mathbb{I} + \mathfrak{X}^*\mathfrak{X}\mathfrak{X}^*\mathfrak{X})^{-1}$ and $\mathfrak{X}^*\mathfrak{X}$ can be computed in $O(D)$. Hence, the total complexity of Algorithm 1 will drop to $O\big(TD(\min\{d_1, d_2\} + \log d_3)\big)$.

Algorithm 1: ADMM for Problem (4)

Input: $\{\mathcal{X}_i\}_i, \boldsymbol{y}, \rho, \varepsilon, T_{\max}$.

Output: $\hat{\mathcal{L}} = \mathcal{L}^{t+1}$.

$\mathcal{L}^0 = \mathcal{K}^0 = \mathcal{E}^0 = \mathcal{Y}_1^0 = \mathcal{Y}_2^0 = \mathbf{0}$;

for $t = 0$ to $T_{\max} - 1$ **do**

Update $\mathcal{L}^{t+1} = \mathfrak{S}_{1/\rho}^{\|\cdot\|_\star}(\mathcal{K}^{t+1} + \mathcal{Y}_1^{t+1}/\rho)$, and $\mathcal{E}^{t+1} = \mathcal{E}_0 - \mathfrak{S}_{\lambda}^{\|\cdot\|_\star}(\mathcal{E}_0)$, where $\mathcal{E}_0 = \mathfrak{X}^*(\boldsymbol{y}) - \mathfrak{X}^*\mathfrak{X}(\mathcal{K}^{t+1}) - \mathcal{Y}_2^{t+1}/\rho$.

Update $\mathcal{K}^{t+1} = (\mathbb{I} + \mathfrak{X}^*\mathfrak{X}\mathfrak{X}^*\mathfrak{X})^{-1}\tilde{\mathcal{K}}$, where $\mathcal{K}_0 = \mathfrak{X}^*\mathfrak{X}\big(\mathfrak{X}^*(\boldsymbol{y}) - \mathcal{E}^{t+1} - \mathcal{Y}_2^{t+1}/\rho\big) + \mathcal{L}^{t+1} - \mathcal{Y}_1^{t+1}/\rho$.

Check stopping condition: $\|\mathcal{K}^{t+1} - \mathcal{L}^{t+1}\|_\infty \leq \varepsilon$, $\|\mathcal{E}^{t+1} + \mathfrak{X}^*\mathfrak{X}(\mathcal{K}^{t+1}) - \mathfrak{X}^*(\boldsymbol{y})\|_\infty \leq \varepsilon$, and $\|\mathcal{T}^{t+1} - \mathcal{T}^t\|_\infty \leq \varepsilon$, $\forall \mathcal{T} \in \{\mathcal{L}, \mathcal{K}, \mathcal{E}\}$.

Update $\mathcal{Y}_1^{t+1} = \mathcal{Y}_1^t + \rho(\mathcal{K}^{t+1} - \mathcal{L}^{t+1})$ and $\mathcal{Y}_2^{t+1} = \mathcal{Y}_2^t + \rho\big(\mathcal{E}^{t+1} + \mathfrak{X}^*\mathfrak{X}(\mathcal{K}^{t+1}) - \mathfrak{X}^*(\boldsymbol{y})\big)$.

end for

3.2. Statistical Performance

We analyze statistical performance of the generalized tensor Dantzig selector by establishing non-asymptotic upper bounds on the estimation error for tensor completion and tensor compressive sensing. The proofs are omitted due to page limitation.

We begin with a lemma on the error tensor $\underline{\Delta} := \hat{\mathcal{L}} - \mathcal{L}^*$.

Lemma 1 *Let $\boldsymbol{\mathcal{T}}_A(\mathcal{L}^*) :=$ cone$\{\mathcal{T} \big| \|\mathcal{L}^* + \mathcal{T}\|_\star \leq \|\mathcal{L}^*\|_\star\}$ be the tangent cone of TNN at $\mathcal{L}^*$. For any tensor $\underline{\Delta} \in \boldsymbol{\mathcal{T}}_A(\mathcal{L}^*)$, the restricted norm compatibility inequality holds:*

$$\|\underline{\Delta}\|_\star \leq 2\sqrt{2r^*}\|\underline{\Delta}\|_F. \tag{5}$$

3.2.1. Tensor Completion

In tensor completion, we also assume the true tensor $\mathcal{L}^*$ has l_∞-norm smaller than a constant α to exclude the spiky tensors [14]. Further, we consider a slightly modified estimator

$$\hat{\mathcal{L}} \in \operatorname*{argmin}_{\|\mathcal{L}\|_\infty \leq \alpha} \|\mathcal{L}\|_\star \text{ s.t. } \|\mathfrak{X}^*(\boldsymbol{y} - \mathfrak{X}(\mathcal{L}))\|_{\text{sp}} \leq \lambda. \tag{6}$$

We first give two key lemmas.

Lemma 2 *In tensor completion, if the sample size $N \geq d_1 d_3$, then w.h.p. the quantity $\|\mathfrak{X}^*(\boldsymbol{\xi})\|_{\text{sp}}$ is concentrated around its mean bounded as $\mathbb{E}[\|\mathfrak{X}^*(\boldsymbol{\xi})\|_{\text{sp}}] \leq c_0\sqrt{N \log \tilde{d}/d_2}$.*

Lemma 3 *In tensor completion, it holds w.h.p.*

$$\|\mathfrak{X}(\mathcal{T})\|_F^2 \geq N(2D)^{-1}\|\mathcal{T}\|_F^2 - c_1 r d_1 d_3 \log \tilde{d},$$

for all $\mathcal{T} \in \mathbb{C}(r)$ (see the definition in Table 1).

Then, the estimation error is upper bounded as follows.

Theorem 1 *In tensor completion, if the sample size $N \geq c_2 r d_1 d_3 \log \tilde{d}$ and parameter $\lambda \geq c_3\sqrt{N \log \tilde{d}/d_2}$, then any solution of Problem (6) satisfies w.h.p.*

$$\frac{\|\underline{\Delta}\|_F^2}{D} \leq C_1 \max\left\{(\sigma^2 \vee \alpha^2)\frac{r^* d_1 d_3 \log \tilde{d}}{N}, \alpha^2\sqrt{\frac{\log \tilde{d}}{N}}\right\}.$$

Theorem 1 guarantees that the per-entry estimation error $\|\underline{\Delta}\|_{\text{F}}^2/D \leq O\big(r^* d_1 d_3 \log \tilde{d}/N\big)$, with sample complexity $N = \Omega(r^* d_1 d_3 \log \tilde{d})$. The sample complexity is *near optimal (up to a log factor)* since the DoF of $\mathcal{L}^*$ is at most $r^*(d_1 + d_2 - r^*)d_3$. The error bound can also be proved *near optimal in minimax sense* using a similar argument with Theorem 6 in [14, 20]. The error bound and the sample complexity are in consistence with the results of ISTT [14] which considers the element-wise Bernoulli sampling. When $d_3 = 1$, the error bound for tensor completion degenerates to a matrix completion bound consistent with [21]. As a trade-off of relaxing the strict TIC [8, 11] to the mild l_∞-norm constraint, Theorem 1 cannot guarantee exact tensor completion in the noiseless setting (i.e., $\sigma = 0$) just like [22].

3.2.2. Tensor Compressive Sensing

For tensor compressive sensing from noisy Gaussian measurements, we first come up with two lemmas:

Lemma 4 *The quantity $\|\mathfrak{X}^*(\boldsymbol{\xi})\|_{\text{sp}}$ is concentrated around its mean bounded as $\mathbb{E}[\|\mathfrak{X}^*(\boldsymbol{\xi})\|_{\text{sp}}] \leq \sigma\sqrt{d_3 N}(\sqrt{d_1} + \sqrt{d_2})$, w.h.p.*

Lemma 5 *If $\mathfrak{X}$ is a random Gaussian design, then for any $\underline{\Delta} \in \boldsymbol{\mathcal{E}}_A(\mathcal{L}^*) := \boldsymbol{\mathcal{T}}_A(\mathcal{L}^*) \cap \mathbb{S}^{D-1}$, it holds w.h.p.*

$$\|\mathfrak{X}(\underline{\Delta})\|_2^2 \geq N - c_0'\sqrt{N} r^*(d_1 + d_2 - r^*)d_3.$$

Then, a non-asymptotic error bound is established.

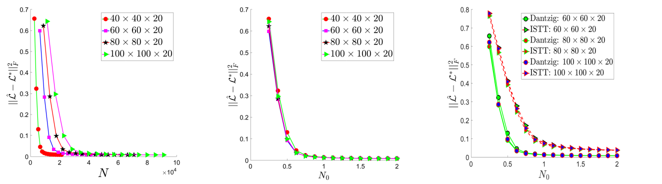
Theorem 2 *In random Gaussian design, if the sample size $N \geq c_1' r^*(d_1 + d_2 - r^*)$ and parameter $\lambda \geq c_2'\sigma\sqrt{d_3 N}(\sqrt{d_1} + \sqrt{d_2})$, then it holds for any solution $\hat{\mathcal{L}}$ of Problem (3) w.h.p.*

$$\|\underline{\Delta}\|_F^2 \leq C_2\sigma^2\frac{r^*(d_1 + d_2)d_3}{N}.$$

Theorem 2 demonstrates that whenever the sample size $N = \Omega(r^*(d_1 + d_2 - r^*)d_3)$, we have $\|\underline{\Delta}\|_{\text{F}}^2 \leq O\big(r^*(d_1 + d_2)d_3 N^{-1}\big)$. In the noiseless setting (i.e., $\sigma = 0$), Theorem 2 guarantees exact recovery with $O(r^*(d_1 + d_2 - r^*))$ samples, which is *order optimal* to the DoF of $\mathcal{L}^*$. When $d_3 = 1$, the problem degenerates to matrix sensing and the error bound $O(r^*(d_1 + d_2)N^{-1})$ is consistent with the bounds in [23]. The sample complexity is consistent with the result of the l_2-norm constrained TNN minimization (l_2-Con) defined by Eq. (7) in [8]. Comparison with l_2-Con on synthetic datasets (shown in Fig. 2.c) demonstrates that the proposed model can obtain solutions of higher precision.

4. EXPERIMENTS

We first verify correctness of Theorems 1 and 2. The true tensors $\mathcal{L}^* \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ with tubal rank r^* are generated by $\mathcal{L}^* = \mathcal{P} * \mathcal{Q}$, where $\mathcal{P} \in \mathbb{R}^{d_1 \times r^* \times d_3}$ and $\mathcal{Q} \in \mathbb{R}^{r^* \times d_2 \times d_3}$ are *i.i.d.* sampled from $\mathcal{N}(0, 1)$. We consider the tensors with



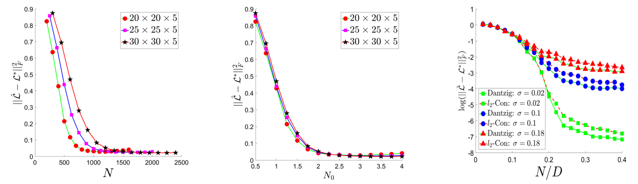
(a) (b) (c)

Fig. 1: Tensor completion. (a): error vs sample size N with $r^* = \lceil \log^{1/2} d \rceil$; (b): error versus rescaled sample size $N_0 = N/(r^* d_1 d_3 \log \tilde{d})$; (c): comparison with ISTT [13].

square frontal slices (i.e. $d_1 = d_2 = d$) for simplicity. For each setting, we run $T = 20$ trials and report the estimation error.

Tensor Completion. For $\mathcal{L}^*$ with unit F-norm, we add *i.i.d.* Gaussian $\mathcal{N}(0, \sigma^2)$ noise with $\sigma = 0.1/\sqrt{D}$ to keep a *constant signal-to-noise ratio (SNR)*. By choosing $d_1 = d_2 = d \in \{40, 60, 80, 100\}$, $d_3 \in \{20, 30, 40\}$ and $r^* = \lceil \log^{1/2} d \rceil$, we consider 12 different problem sizes. In this constant SNR setting, the upper bound proposed in Theorem 1 is shown to scale like $O(r^* d_1 d_3 \log \tilde{d}/N)$ w.h.p. (following [13]). Thus, if the bound is sharp, it is expected that the estimation error would have the same scaling behavior. Equivalently, if the estimation error is plotted versus the *rescaled sample size* defined as $N_0 := N/r^* d_1 d_3 \log \tilde{d}$, then it is expected that the curves of different tensor sizes should be well aligned. The results are shown in Fig. 1. In particular, plots of the error against the sample size N are shown in Fig. 1.a with $d \in \{40, 60, 80, 100\}$ and $d_3 = 20$, reflecting the intuition that the error decreases as N increases. Plots of the error versus the rescaled sample size N_0 are shown in Fig. 1.b. Since for other problem sizes, similar scaling behaviors are also be observed, the results are omitted. As expected, the error curves align well in Fig. 1, which validates the sharpness of the proposed upper bound in Theorem 1. The proposed estimator (Dantzig) is also compared with ISTT [13] in Fig. 1.c. We can see the proposed estimator achieves higher precision than ISTT.

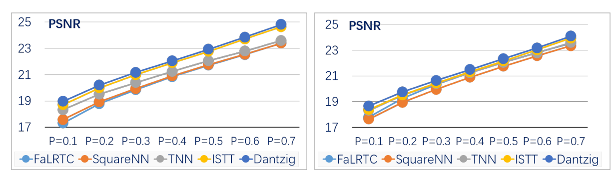
Tensor Compressive Sensing. We follow the procedure for matrix sensing in [23]. Before sensing $\mathcal{L}^*$ with random Gaussian design $\mathfrak{X}$, we first normalize it such that $\|\mathcal{L}^*\|_{\mathrm{F}} = 1$. Then, the noise variables are generated from *i.i.d.* Gaussian $\mathcal{N}(0, \sigma^2)$ with $\sigma = 0.1$. By choosing $d_1 = d_2 = d \in \{20, 25, 30\}$, $d_3 \in \{5, 10\}$ and the tubal rank $r^* = \lceil \log^{1/2} d \rceil$, we consider 6 different problem sizes. Fig. 2 summarizes the results. Particularly, Fig. 2.a plots the estimation error versus the raw sample size N, which shows the consistency that estimation error decreases when sample size increases. Fig. 2.b plots the error against the rescaled sample size $N_0 = N/(r^*(d_1 + d_2)d_3)$. It is clearly shown in Fig. 2.b that the errors are well aligned, validating the sharpness of the proposed error bounds. We also compare the proposed estimator (Dantzig) with l_2-Con [8]. For tensors $\mathcal{L}^*$ of size $20 \times 20 \times 5$ with tubal rank $r^* = 1$, we vary the noise level $\sigma \in \{0.02, 0.1, 0.18\}$ along with the observation ratio $N/D \in 0.02 : 0.02 : 0.4$. The averaged estimation error are shown in log scale in Fig. 2.c, showing that the proposed estimator is more accurate than l_2-Con.



(a) (b) (c)

Fig. 2: Tensor compressive sensing. (a): error versus sample size N with $r^* = \lceil \log^{1/2} d \rceil$. (b): error versus rescaled sample size $N_0 = N/(r^*(d_1 + d_2)d_3)$. (c): comparison with l_2-Con [8] by varying $\sigma \in \{0.02, 0.1, 0.18\}$ and $N/D \in 0.02 : 0.02 : 0.4$ for $\mathcal{L}^* \in \mathbb{R}^{20\times 20\times 5}$ of tubal rank $r^* = 1$.

Point Cloud Data Inpainting. We conduct inpainting experiments on a sequence of point cloud data acquired from a Velodyne HDL-64E LiDAR [1] [24]. The distance and intensity data are formated into two tensors in $\mathbb{R}^{64\times 436\times 80}$. The proposed estimator (Dantzig) is compared with four nuclear norm based tensor completion models including FaLRTC[2] [25], SquareNN[3] [12], TNN[4] [11] and ISTT [13]. The quality of inpainting is measured by the peak signal-to-noise ratio (PSNR). Given $\mathcal{T} \in \mathbb{R}^{d_1\times d_2\times d_3}$, we let the sampling ratio p vary from 0.1 to 0.7 and the noise follows *i.i.d.* Gaussian $\mathcal{N}(0, \sigma^2)$ where $\sigma = 0.2\|\mathcal{T}\|_{\mathrm{F}}/\sqrt{D}$. For quantitative comparison, the PSNR values are shown in Fig. 3. We can see that the proposed estimator outperform other nuclear norm based competitors.



(a) distance data (b) intensity data

Fig. 3: Results of point cloud data inpainting.

5. CONCLUSION

The generalized tensor Dantzig selector is defined for low-tubal-rank tensor recovery. Then, an ADMM-based algorithm is developed. Further, upper bounds on the estimation error are established for tensor completion and compressive sensing. Synthetic experiments verify the theory. The effectiveness of estimator is demonstrated on real datasets.

[1]Frame Nos. 65-144 of the *Scenario B* and *Scenario B-additional* datasets: http://www.mrt.kit.edu/z/publ/download/velodynetracking/dataset.html

[2]http://www.cs.rochester.edu/u/jliu/publications.html

[3]https://sites.google.com/site/mucun1988/publi

[4]https://github.com/jamiezeminzhang/

References

[1] G. Donald and Z. Qin, "Robust low-rank tensor recovery: Models and algorithms," *SIAM Journal on Matrix Analysis and Applications*, vol. 35, no. 1, pp. 225–253, 2014.

[2] Richard A Harshman, "Foundations of the parafac procedure: Models and conditions for an "explanatory" multimodal factor analysis," 1970.

[3] Ledyard R Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.

[4] Z. Zhang, G. Ely, S. Aeron, N. Hao, and M. Kilmer, "Novel methods for multilinear data completion and denoising based on tensor-svd," in *CVPR*, 2014, pp. 3842–3849.

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

[6] X-Y. Liu, S. Aeron, V. Aggarwal, et al., "Low-tubal-rank tensor completion using alternating minimization," *arXiv preprint arXiv:1610.01690*, 2016.

[7] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization," in *CVPR*, 2016, pp. 5249–5257.

[8] C. Lu, J. Feng, Z. Lin, et al., "Exact low tubal rank tensor recovery from gaussian measurements," in *IJCAI*, 2018, pp. 1948–1954.

[9] L. Yuan, Q. Zhao, and J. Cao, "High-order tensor completion for data recovery via sparse tensor-train optimization," in *ICASSP*. IEEE, 2018, pp. 1258–1262.

[10] Q. Zhao, L. Zhang, and A. Cichocki, "Bayesian cp factorization of incomplete tensors with automatic rank determination," *IEEE TPAMI*, vol. 37, no. 9, pp. 1751–1763, 2015.

[11] Z. Zhang and S. Aeron, "Exact tensor completion using t-svd," *IEEE TSP*, vol. 65, no. 6, pp. 1511–1526, 2017.

[12] C. Mu, B. Huang, J. Wright, and D. Goldfarb, "Square deal: Lower bounds and improved relaxations for tensor recovery," in *ICML*, 2014, pp. 73–81.

[13] A. Wang and Z. Jin, "Near-optimal noisy low-tubal-rank tensor completion via singular tube thresholding," in *ICDM workshop*, 2017, pp. 553–560.

[14] A. Wang, D. Wei, B. Wang, and Z. Jin, "Noisy low-tubal-rank tensor completion through iterative singular tube thresholding," *IEEE Access*, vol. 6, pp. 35112–35128, 2018.

[15] P. J. Bickel, Ya'acov Ritov, and A. B. Tsybakov, "Simultaneous analysis of lasso and dantzig selector," *The Annals of Statistics*, vol. 37, no. 4, pp. 1705–1732, 2009.

[16] S. Chen and A. Banerjee, "Structured matrix recovery via the generalized dantzig selector," in *NIPS*, 2016, pp. 3252–3260.

[17] S. Gunasekar, A. Banerjee, and J. Ghosh, "Unified view of matrix completion under general structural constraints," in *NIPS*, 2015, pp. 1180–1188.

[18] R. Tomioka, T. Suzuki, K. Hayashi, and H. Kashima, "Statistical performance of convex tensor decomposition," in *NIPS*, 2011, pp. 972–980.

[19] S. Boyd, N. Parikh, et al., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[20] A. Wang, Z. Lai, and Z. Jin, "Noisy low-tubal-rank tensor completion," *Neurocomputing*, vol. 330, pp. 267–279, 2019.

[21] O. Klopp, "Noisy low-rank matrix completion with general sampling distribution," *Bernoulli*, vol. 20, no. 1, pp. 282–303, 2014.

[22] S. Negahban and M. J. Wainwright, "Restricted strong convexity and weighted matrix completion: Optimal bounds with noise," *JMLR*, vol. 13, no. May, pp. 1665–1697, 2012.

[23] S. Negahban and M. J. Wainwright, "Estimation of (near) low-rank matrices with noise and high-dimensional scaling," *The Annals of Statistics*, pp. 1069–1097, 2011.

[24] F. Moosmann and C. Stiller, "Joint self-localization and tracking of generic objects in 3d range data," in *ICRA*, Karlsruhe, Germany, May 2013, pp. 1138–1144.

[25] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE TPAMI*, vol. 35, no. 1, pp. 208–220, 2013.