Advances in Applied Mathematics 应用数学进展, 2026, 15(5), 25-34
Published Online May 2026 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2026.155205

Hans汉斯

一种基于差分隐私的张量填充算法

戴森立

温州大学数理学院，浙江 温州

收稿日期：2026年4月9日；录用日期：2026年5月2日；发布日期：2026年5月12日

摘　要

针对高维张量数据填充过程中存在的隐私泄露问题，本文提出一种融合拉普拉斯机制的差分隐私张量填充算法(DP-ADMM-t-SVD)。该算法以ADMM-t-SVD低秩张量填充算法为基础，引入了ε-差分隐私框架，先对观测张量进行逐切片归一化以统一全局敏感度，再对张量在观测位置注入与隐私预算ε相关的拉普拉斯噪声，通过交替方向乘子法迭代求解带隐私约束的张量填充优化问题，实现数据隐私保护与填充精度的双重兼顾。为验证算法可行性，分别在不同尺寸的人工低秩张量数据集和篮球视频现实张量数据集上开展实验，实验结果表明，所提算法在人工和现实数据集上均表现出良好性能：隐私预算ε较小时，算法具备优异的隐私保护效果且恢复精度处于可控范围；ε较大时，恢复精度接近无扰动的基线算法，同时用户可通过调节ε的取值适配不同的隐私保护与恢复精度需求。

关键词

张量填充，差分隐私，拉普拉斯机制，ADMM-t-SVD

A Tensor Completion Algorithm Based on Differential Privacy

Senli Dai

School of Mathematics and Physics, Wenzhou University, Wenzhou Zhejiang

Received: April 9, 2026; accepted: May 2, 2026; published: May 12, 2026

Abstract

To address the privacy leakage issue in the completion of high-dimensional tensor data, this paper proposes a differential privacy-preserving tensor completion algorithm integrated with the Laplacian mechanism, denoted as DP-ADMM-t-SVD. Based on the ADMM-t-SVD low-rank tensor completion algorithm, the proposed algorithm introduces the ε-differential privacy framework. It first performs per-slice normalization on the observed tensor to unify the global sensitivity, and then injects Laplacian

文章引用: 戴森立. 一种基于差分隐私的张量填充算法[J]. 应用数学进展, 2026, 15(5): 25-34.
DOI: 10.12677/aam.2026.155205

noise associated with the privacy budget ε only at the observed positions of the tensor. Furthermore, it iteratively solves the privacy-constrained tensor completion optimization problem via the Alternating Direction Method of Multipliers (ADMM), thus realizing a balanced trade-off between data privacy protection and completion accuracy. To verify the feasibility of the proposed algorithm, experiments are carried out on synthetic low-rank tensor datasets with different dimensions and a real-world tensor dataset from a basketball video. Experimental results show that the proposed algorithm achieves favorable performance on both synthetic and real-world datasets: when the privacy budget ε is small, the algorithm provides an excellent privacy protection effect while keeping the completion accuracy within a controllable range; when ε is large, the completion accuracy is close to that of the non-perturbed baseline algorithm. Meanwhile, users can adjust the value of ε to adapt to different requirements for privacy protection and completion accuracy.

Keywords

Tensor Completion, Differential Privacy, Laplacian Mechanism, Alternating Direction Method of Multipliers-Tensor Singular Value Decomposition

1. 介绍

多维数组，又称张量，是向量和矩阵的一种推广[1]。随着大数据、物联网与人工智能技术的深度融合，多维数据呈现出爆发式增长态势，这类数据普遍具备多模态、多时域、多主体的复杂结构特征，传统向量、矩阵等低维建模方式难以完整保留数据内在的关联特性与结构信息。张量作为矩阵的高阶推广，能够天然刻画多维数据的高阶交互关系，在推荐系统[2]、降维[3]、交通预测[4]、计算机视觉[5]等领域得到广泛应用，成为高维数据建模与分析的核心工具。

然而在实际数据采集、传输与存储过程中，受设备故障、传输损耗、采样缺失、隐私脱敏等因素影响，张量数据普遍存在数据缺失的问题，严重制约后续应用的进行。基于此，张量填充技术应运而生，它通过数据的低秩性、平滑性等先验特征，从部分观测条目中复原完整张量，是一种解决高维数据缺失问题的关键技术。与此同时，张量数据往往包含用户身份、行为轨迹、医疗记录、商业机密等敏感信息，在填充、传输、共享与计算全流程中，极易面临隐私泄露、数据窃取、推断攻击等安全风险，违背数据合规要求与隐私保护准则[6]。

在此背景下，如何设计一种兼顾数据完整性与隐私安全性的张量填充算法，成为数据科学与隐私计算领域的交叉热点。一方面，好的张量填充算法可以充分发挥多维数据的价值；另一方面，严格的隐私保护可以保护用户的隐私不被泄露。二者的结合对推动医疗、金融、政务等敏感领域的数据共享与智能应用具有重要的理论价值与现实意义。

2. 张量填充及相关符号

张量填充(又称为张量补全)，是矩阵填充的高阶拓展，核心目标是在已知部分张量元素的前提下，借助数据内在的结构先验与相关性特征，精准估计缺失元素，复原出完整且贴合真实分布的张量。首先我们介绍一下张量的相关知识。

我们以三阶张量为例进行说明，一个张量的切片定义为固定一个维度的索引，提取剩余两个维度构

成的二维矩阵。张量的定义为，固定其余两个维度的索引，提取剩余的那个维度构成的一维向量。我们分别采用 Matlab 记号 $\mathcal{A}(k,:,:)$，$\mathcal{A}(:,k,:)$，$\mathcal{A}(:,:,k)$ 表示张量的第 k 个水平切片，侧向切片，正面切片。用 $\mathcal{A}(:,i,j)$，$\mathcal{A}(i,:,j)$，$\mathcal{A}(i,j,:)$ 来分别表示模 1 管，模 2 管和模 3 管。有了这个概念后我们引入一种张量之间的运算，叫做张量积，用 $*$ 表示，它的定义如下：对于任意两个张量 $\mathcal{A}\in\mathbb{R}^{I_1\times I_2\times I_3}$，$\mathcal{B}\in\mathbb{R}^{I_2\times I_4\times I_3}$，它们的张量积为：

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

接下来我们介绍一下张量的转置，f -对角张量以及正交张量，为张量的 t-SVD 分解做准备。

张量的转置：一个三阶张量 $\mathcal{A}\in\mathbb{R}^{I_1\times I_2\times I_3}$，它的转置 $\mathcal{A}^{\mathrm{T}}\in\mathbb{R}^{I_2\times I_1\times I_3}$ 为将 $\mathcal{A}$ 的每个正面切片都转置再重新按 I_3 排序为新的张量 $\mathcal{A}^{\mathrm{T}}$。

f -对角张量：如果一个张量的每个正面切片都是对角矩阵，则称这个张量为 f -对角张量。

正交张量：如果一个张量 $\mathcal{Q}\in\mathbb{R}^{I_1\times I_2\times I_3}$ 满足 $\mathcal{Q}*\mathcal{Q}^{\mathrm{T}}=\mathcal{Q}^{\mathrm{T}}*\mathcal{Q}=\mathcal{I}$，那么称张量 $\mathcal{Q}$ 是正交的，其中 $*$ 代表张量积，$\mathcal{I}\in\mathbb{R}^{I_1\times I_1\times I_3}$ 是单位张量，它的第一个正面切片是一个 $I_1\times I_1$ 的单位矩阵，其余正面切片均为零矩阵。

有了上述概念后，我们给出三阶张量的 t-SVD 分解。

对于一个张量 $\mathcal{A}\in\mathbb{R}^{I_1\times I_2\times I_3}$，它的 t-SVD 分解定义为：

$$\mathcal{A}=\mathcal{U}*\mathcal{S}*\mathcal{V}^{\mathrm{T}} \tag{2}$$

其中 $\mathcal{U}$ 和 $\mathcal{V}$ 分别为 $I_1\times I_1\times I_3$ 和 $I_2\times I_2\times I_3$ 的正交张量。$\mathcal{S}$ 是一个大小为 $I_1\times I_2\times I_3$ 的 f -对角张量，$*$ 代表张量积。

对于一个三阶张量 $\mathcal{A}\in\mathbb{R}^{I_1\times I_2\times I_3}$，用 $\tilde{\mathcal{A}}=fft\left[\mathcal{A},[\,],3\right]$ 表示在 Matlab 中沿着第三维度进行离散变化，用 $\mathcal{A}=ifft\left[\tilde{\mathcal{A}},[\,],3\right]$ 表示将 $\tilde{\mathcal{A}}$ 还原成 $\mathcal{A}$。

除此之外，我们再给出另外一个张量相关的定义。$\mathrm{blkdiag}\left(\hat{\mathcal{A}}\right)$ 是一个分块对角矩阵，它的定义如下：

$$\mathrm{blkdiag}\left(\hat{\mathcal{A}}\right)=\begin{bmatrix}\hat{\mathcal{A}}^{(1)} & 0 & \cdots & 0\\ 0 & \hat{\mathcal{A}}^{(2)} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \hat{\mathcal{A}}^{(I_3)}\end{bmatrix} \tag{3}$$

其中 $\hat{\mathcal{A}}^{(i)}$ 是张量 $\hat{\mathcal{A}}$ 的第 i 个正面切片，$i=1,2,\cdots,I_3$。

3. 张量填充优化模型

数学上，给定一个只有部分观测元素的张量 $\mathcal{X}\in\mathbb{R}^{I_1\times\cdots\times I_N}$，低秩张量填充问题可以表述为以下模型：

$$\min_{\mathcal{X}} rank(\mathcal{X})\quad \text{s.t. } P_\Omega(\mathcal{X})=P_\Omega(\mathcal{M}) \tag{4}$$

其中 $\mathcal{X}\in\mathbb{R}^{I_1\times\cdots\times I_N}$ 表示观测到的不完整张量。Ω 表示观测到元素的下标集合。$P_\Omega(\mathcal{X})$ 表示投影算子，若 $(i_1,\cdots,i_N)\in\Omega$，则 $P_\Omega(\mathcal{X})_{i_1\cdots i_N}=\mathcal{X}_{i_1\cdots i_N}$，否则 $P_\Omega(\mathcal{X})_{i_1\cdots i_N}=0$。与矩阵填充不同的是，矩阵的秩是唯一定义的，而张量秩的定义并不唯一。例如 CP-秩是基于张量的 CP 分解定义的[7]-[9]，Tucker-秩是基于张量的 Tucker 分解定义的[10]-[12]，Zhang 等人在 2014 年提出了一种利用张量奇异值分解 t-SVD 进行张量填充的方法[13]，该方法是基于张量的管秩提出的，它极大的提高了张量填充的精度 t-SVD 算法主要如下所示：

算法 1：t-SVD 算法

输入：$\mathcal{M}\in\mathbb{R}^{I_1\times\cdots\times I_N}$，$L=n_3n_4\cdots n_N$，$\mathcal{D}=\mathcal{M}$

for $i=3$ to N

$\mathcal{D}\leftarrow fft\left(\mathcal{D},[\,],i\right);$

end for

for $i=1$ to L

$[U,S,V]=\text{svd}\left(\mathcal{D}(:,:,i)\right)$

$\hat{u}(:,:,i)=U;\hat{\mathcal{S}}(:,:,i)=\mathcal{S};\hat{V}(:,:,i)=\mathcal{V};$

end for

for $i=3$ to N

$\mathcal{U}\leftarrow ifft\left(\hat{\mathcal{U}},[\,],i\right);\hat{\mathcal{S}}\leftarrow ifft\left(\hat{\mathcal{S}},[\,],i\right);V\leftarrow ifft\left(\hat{\mathcal{V}},[\,],i\right);$

end for

接下来介绍一下基于 t-SVD 的张量填充算法。根据张量的核范数[14]，(4)式可以被改写为

$$\min\left\|\mathcal{X}\right\|_{TNN}\quad \text{s.t. } P_\Omega\left(\mathcal{X}\right)=P_\Omega\left(\mathcal{M}\right) \tag{5}$$

假设 $\mathcal{Y}=\mathcal{P}_\Omega\cdot\mathcal{M}$ 为采样数据，定义 $\mathcal{G}=\mathcal{F}_3\mathcal{P}_\Omega\mathcal{F}_3^{-1}$，其中 $\mathcal{F}_3$ 和 $\mathcal{F}_3^{-1}$ 分别为张量第三维度上的傅里叶变换算子与逆傅里叶变换算子。由此可得 $\hat{\mathcal{Y}}=\mathcal{G}\left(\hat{\mathcal{M}}\right)$，其中 $\hat{\mathcal{Y}}$ 和 $\hat{\mathcal{M}}$ 分别为 $\mathcal{Y}$ 和 $\mathcal{M}$ 在第三模态上的傅里叶变换结果。因此，式(5)等价于如下优化问题

$$\min\left\|\text{blkdiag}\left(\hat{\mathcal{X}}\right)\right\|_*\quad \text{s.t. } \hat{\mathcal{Y}}=\mathcal{G}\left(\hat{\mathcal{X}}\right) \tag{6}$$

其中 $\hat{\mathcal{X}}$ 为 $\mathcal{X}$ 在第三维度上的傅里叶变换，而张量核范数又满足下式 $\left\|\mathcal{X}\right\|_{TNN}=\left\|\text{blkdiag}\left(\hat{\mathcal{X}}\right)\right\|_*$，因此式(6)可以改写为

$$\min\left\|\text{blkdiag}\left(\hat{\mathcal{Z}}\right)\right\|_*+1_{\hat{\mathcal{Y}}=\mathcal{G}\left(\hat{\mathcal{X}}\right)}\quad \text{s.t. } \hat{\mathcal{X}}-\hat{\mathcal{Z}}=0 \tag{7}$$

其中 1 表示指数函数。这个问题是一个凸优化问题，我们考虑采用交替方向乘子法(ADMM)算法进行求解。首先，定义拉格朗日函数：

$$\mathcal{L}\left(\mathcal{X},\mathcal{Z},\mathcal{Q}\right)=1_{\mathcal{Y}=P_\Omega(\mathcal{X})}+\frac{1}{\rho}\left\|Z\right\|_{TNN}+\left\langle\mathcal{Q}(:),\mathcal{X}(:)-\mathcal{Z}(:)\right\rangle+\frac{1}{2}\left\|\mathcal{X}-\mathcal{Z}\right\|_F^2 \tag{8}$$

固定当前迭代步中的

$$\mathcal{X}^{k+1}=\arg\min_{\mathcal{X}}\left\{1_{Y=P_\Omega(X)}+\mathcal{X}(:)^{\mathrm{T}}\mathcal{Q}^k(:)+\frac{1}{2}\left\|\mathcal{X}-\mathcal{Z}^k\right\|_F^2\right\} \tag{9}$$

其解为：

$$\mathcal{X}^{k+1}=\arg\min_{\mathcal{X}:\mathcal{Y}=P_\Omega(\mathcal{X})}\left\|\mathcal{X}-\left(\mathcal{Z}^k-\frac{1}{\rho}\mathcal{Q}^k\right)\right\|_F^2 \tag{10}$$

在迭代得到更新后的 $\mathcal{X}^{k+1}$ 后，接下来更新 $\mathcal{Z}$：

$$\tilde{\mathcal{Z}}^{k+1}=\arg\min_{\tilde{\mathcal{Z}}}\left\{\frac{1}{\rho}\left\|\text{blkdiag}\left(\tilde{\mathcal{Z}}\right)\right\|_{*}+\frac{1}{2}\left\|\tilde{\mathcal{Z}}-\left(\tilde{\mathcal{X}}^{k+1}-\tilde{\mathcal{Q}}^{k}\right)\right\|_{F}^{2}\right\} \tag{11}$$

最后根据得到的 $\mathcal{X}^{k+1}$ 和 $\mathcal{Z}^{k+1}$，得到 $\mathcal{Q}$ 的更新公式如下：

$$\mathcal{Q}^{k+1}=\mathcal{Q}^{k}+\left(\mathcal{X}^{k+1}-\mathcal{Z}^{k+1}\right) \tag{12}$$

其中方程(9)是最小二乘投影到约束上的，方程(10)的解由奇异值阈值给出。$\mathcal{X}(:)$ 和 $\mathcal{Q}^{k}(:)$ 均表示张量的向量化，是一种 Matlab 符号。

4. 加入隐私保护的张量填充算法

在本节中我们将介绍一种隐私保护算法，叫做差分隐私保护策略[15]，并将这种方案应用于张量填充。差分隐私作为数据隐私保护的黄金标准，其核心是通过向数据中注入可控噪声，使得任意一条数据的增减不会显著改变查询/分析结果的分布，从而规避个体信息泄露风险。本文采用纯 ε-差分隐私(ε-DP)框架，结合拉普拉斯噪声机制实现张量观测数据的隐私保护，既保证严格的隐私证明，又兼顾张量填充的精度需求。接下来我们将先介绍一下差分隐私：

定义 1：假设随机算法 $\mathcal{H}$ 的输出落在定义域 $\mathcal{S}$ 中，若对于任意的一个子集 $S\in\mathcal{S}$ 以及两个仅相差一条记录的相邻数据集 D 和 D'，都满足

$$Pr\left(\mathcal{H}\left(D\right)\in S\right)\le \mathrm{e}^{\varepsilon}Pr\left(\mathcal{H}\left(D'\right)\in S\right) \tag{13}$$

则称该算法 $\mathcal{H}$ 满足差分隐私。

拉普拉斯机制[16]是实现 ε-DP 的经典方法，它通过添加符合拉普拉斯分布的噪声，确保查询结果满足 ε-差分隐私，其核心是根据数据的全局敏感度确定噪声尺度。其中 ε 称为隐私预算，ε 越接近 0，$\mathcal{H}$ 在 D、D' 上输出的数据分布越接近，输出结果越不可区分，隐私保护程度越高。反之 ε 越大，隐私保护的程度越小。

定义 2：全局敏感度：给定一个以某个数据集 D 作为输入的查询函数 $f:D\to\mathbb{R}^{d}$，对于一个相邻数据集对 D 和 D'，函数 f 的全局敏感度定义如下：

$$\Delta f=\sup_{D,D'}\left\|f\left(D\right)-f\left(D'\right)\right\| \tag{14}$$

其中 $\mathbb{R}^{d}$ 代表数据集 D 映射到的实数空间，d 为查询函数的维度。

定义 3：拉普拉斯机制：给定一个查询函数 $f:D\to\mathbb{R}^{d}$，拉普拉斯机制 $\mathcal{M}$ 通过 L 向查询结果添加拉普拉斯噪声，它的定义为

$$\mathcal{M}_{L}\left(x\right)=f\left(x\right)+\left(Y_{1},\cdots,Y_{d}\right) \tag{15}$$

这一定义确保了它满足 ε-差分隐私。其中 $Y_{i}\sim Lap\left(\frac{\Delta f}{\varepsilon}\right)\left(i\in\left[d\right]\right)$ 为独立采样的拉普拉斯噪声，它的概率密度函数如下：

$$Pr\left(x\right)=\frac{1}{2\lambda}\mathrm{e}^{-\frac{\left|x\right|}{\lambda}} \tag{16}$$

其中噪声规模 $\lambda=\frac{\Delta f}{\varepsilon}$ 取决于隐私预算 ε 和全局敏感度 Δf，同时全局敏感度 Δf 度量了函数的输出结果对添加、修改或删除数据集中一条记录时可能的最大变化程度。

针对低秩张量填充场景，在上节提到的 ADMM-t-SVD 算法的基础上，本考虑对观测阶段的张量数据注入拉普拉斯噪声(输入扰动策略)以提高隐私性，核心步骤如下：

假设原始的观测张量为 $\mathcal{X}\in\mathbb{R}^{n_1\times n_2\times n_3}$，为统一全局敏感度、降低噪声注入的保守性，首先对张量数据进行预处理，对观测张量 $\mathcal{X}$ 进行逐切片归一化：

$$X_n(:,:,i)=\frac{X(:,:,i)}{\max\left(\left|X(:,:,i)\right|\right)}\left(i=1,2,\cdots,n_3\right) \tag{17}$$

其中 X_n 为归一化后的张量，n_3 为张量的切片数量；归一化后每个切片的取值范围被约束在 $[-1,1]$，避免因数据分布差异导致的敏感度高估。

接下来，我们在归一化后的张量 X_n 中添加拉普拉斯噪声，噪声记为 L_noise：

$$\widetilde{X_n}=X_n+L\_noise\odot\Omega \tag{18}$$

其中 $L\_noise=\frac{\Delta f}{\varepsilon}$，$\Delta f$ 为全局敏感度，ε 为隐私预算，$L\_noise\odot\Omega$ 表示仅在观测位置加入噪声，这样可以避免无意义的噪声添加，降低对填充精度的额外影响。Ω 为掩码矩阵，定义如下：

$$\Omega_{i,j,k}=\begin{cases}1, & \text{其他}\\ 0, & \text{如果}\mathcal{X}_{i,j,k}\text{缺失}\end{cases} \tag{19}$$

接下来通过第三节中叙述的张量填充算法进行张量填充，该算法的具体步骤如下：

为适配 ADMM 迭代求解框架，首先对三维张量进行向量化离散化建模。定义向量化算子 $vec(\cdot)$ 将张量展平为列向量：

$$x=vec(\mathcal{X})\in\mathbb{R}^{n_1n_2n_3\times 1} \tag{20}$$

基于掩码张量构建稀疏对角采样矩阵：

$$A=diag\left(vec(\Omega)\right) \tag{21}$$

该矩阵实现对张量观测位置的采样约束。最后构建向量化观测向量：

$$b=vec(\mathcal{X}_{\text{obs}}) \tag{22}$$

其中 $\mathcal{X}_{\text{obs}}$ 为带拉普拉斯噪声的观测张量。

接下来进行算法的迭代步骤，其中 x,z,u 均为迭代过程的中间变量。

在第 k 次迭代时，有

$$\mathcal{X}^{k+1}=P\left(\mathcal{Z}^k-\frac{1}{\rho}\mathcal{Q}^k\right)+\mathcal{Y} \tag{23}$$

其中 $\mathcal{P}$ 为缺失位置补全算子，观测位置固定为 $\mathcal{Y}$。接下来更新 $\mathcal{Z}$：

$$\mathcal{Z}^{k+1}=prox_{\frac{1}{\rho}\|\cdot\|_{TNN}}\left(\mathcal{X}^{k+1}+\frac{1}{\rho}\mathcal{Q}^k\right) \tag{24}$$

最后更新 $\mathcal{Q}$：

$$\mathcal{Q}^{k+1}=\mathcal{Q}^k+\rho\left(\mathcal{X}^{k+1}-\mathcal{Z}^{k+1}\right) \tag{25}$$

得到恢复后的向量为 $\hat{x}$。对向量 $\hat{x}$ 进行重构后得到恢复后的张量 $\hat{\mathcal{X}}$，重构的方法如下：

$$\hat{\mathcal{X}}=reshape\left(\hat{x}.[n_1,n_2,n_3]\right) \tag{26}$$

我们将这种加入了差分隐私的 ADMM-t-SVD 算法称之为 DP-ADMM-t-SVD 算法，他是在一种基于张

量核范数的张量填充方法的基础上加入了差分隐私的方案。通过这样一种差分隐私的方案可以一定程度上避免用户隐私的泄露，并且使用者还可以通过调节隐私预算 ε 的大小来得到自己想要达到的隐私保护效果。

5. 数值实验

在本节中我们通过人工数据集和现实数据集来验证我们所提出的隐私保护算法的可行性。我们将通过对比加入差分隐私前后的误差来验证算法的有效性。我们通过标准误差(RSE)来对比恢复精度，定义如下：

$$\mathrm{RSE}=\frac{\left\|\hat{\mathcal{X}}-\mathcal{X}\right\|_F}{\left\|\mathcal{X}\right\|_F} \tag{27}$$

其中 $\hat{\mathcal{X}}$ 和 $\mathcal{X}$ 分别代表恢复后的矩阵以及真实矩阵。

首先对于人工数据集，我们设定张量维度为 $200\times200\times20$，为了模拟低秩环境，我们设置张量的秩 $r=2$，固定观测率 $p=20\%$，迭代次数 $Iter=40$。迭代次数的选取的依据见图 1。可以看到对于大小分别为 $200\times200\times20$ 和 $500\times500\times50$ 的张量，在迭代次数为 30 次时均达到收敛，因此，我们将迭代次数设置为 40，这保证了算法的收敛性。此外还需要说明的一点是差分隐私的加入不会影响算法的收敛性，我们迭代次数的选择是合理的。

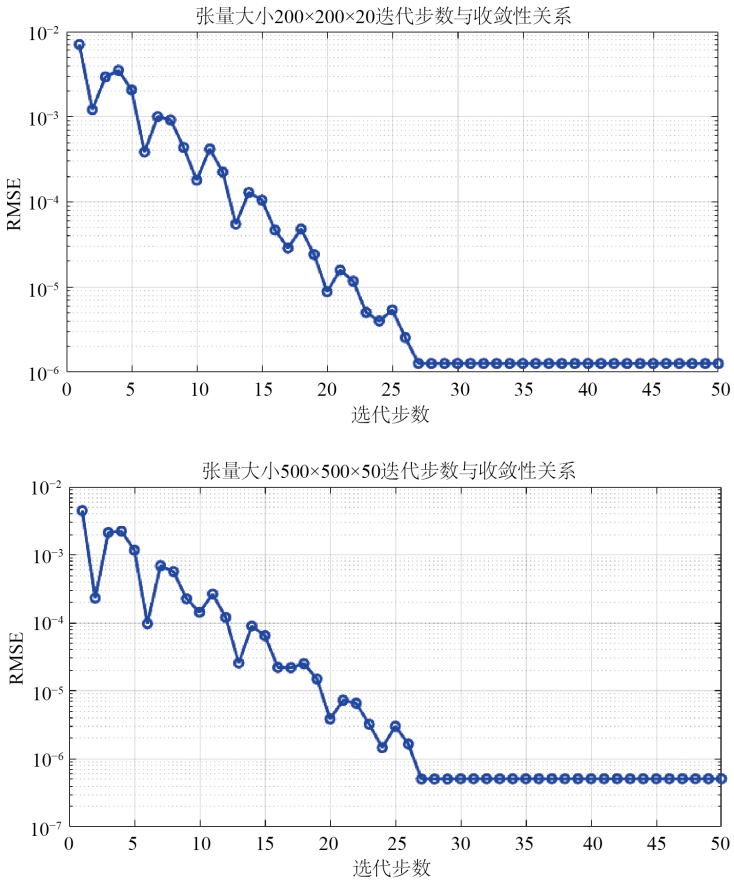

Figure 1. The relationship between number of iterations and convergence performance
图 1. 迭代步数与算法收敛性的关系

具体而言，我们通过调节隐私预算 ε 的大小来进行实验。我们使用无扰动的原版张量填充方法作为基线进行对比。由于引入噪声可能带来不确定性，我们的 RSE 是经过多次试验后取平均得到的。图 2 展示了两种不同大小的张量在加入差分隐私后的恢复精度与基线的对比。

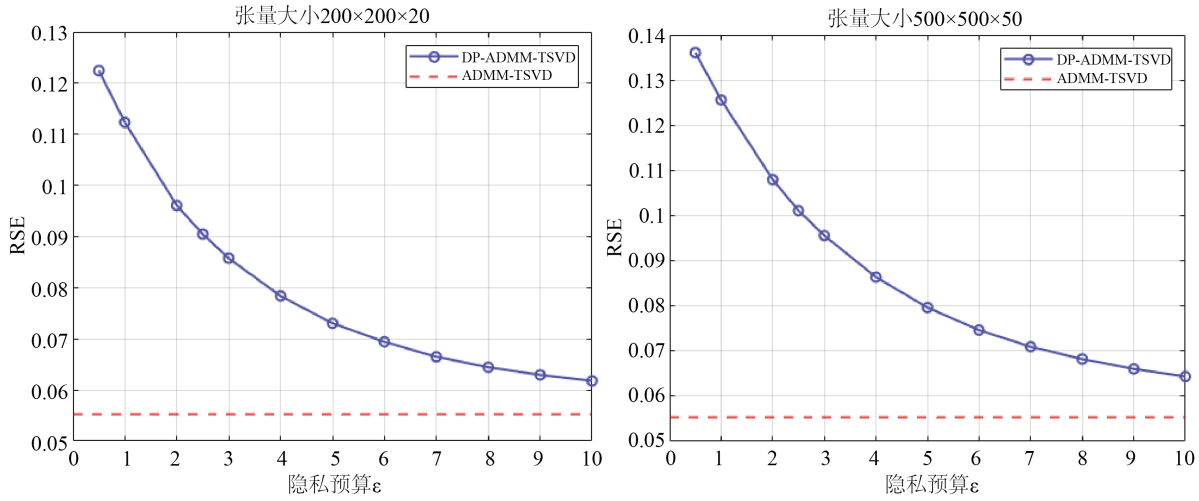

Figure 2. Effect of tensor dimensions on completion effectiveness
图 2. 张量大小对填充效果的影响

实验结果表明无论是对于较小的张量 200 × 200 × 20 还是较大的张量 500 × 500 × 50，我们所提出的差分隐私张量填充均有较好的效果，具体来讲，在隐私预算 ε 较小时起到了比较好的隐私保护效果，但同时恢复精度又在可控范围内；而当隐私预算 ε 较大时，恢复精度与无扰动的张量填充恢复精度相当，这与我们的预期相符。

对于现实数据集，我们使用了一段黑白灰度篮球视频，这段视频展示了 1.6 秒的比赛，包含 40 张 AVI 格式的数字图像。该视频的每张图像都是包含 144×256 个像素的黑白图像，因此该视频可以看作是一个张量 $\mathcal{X} \in \mathbb{R}^{144\times256\times40}$。为了验证我们的算法，我们选取观测率 $p = 50\%$，并在观测张量上添加噪声，并使用 ADMM-t-SVD 方法进行张量填充。图 3 展示了隐私预算 ε 取 3 和 10 时的恢复精度。根据图 3 可以看出 ε 取 3 时图像噪声更大，隐私保护较强，但恢复精度也较差；ε 取 10 时图像噪声更小，隐私保护较弱，但恢复精度相对较好。表 1 展示了不同隐私预算下，对于视频恢复的误差。从表中也可以看出，对于视频恢复任务来说取 $\varepsilon \geq 10$ 时差分隐私提供弱隐私保护，恢复精度较高；$\varepsilon \leq 10$ 时，差分隐私提供较强的隐私保护，恢复精度较低。实验结果表明，对于现实数据集，我们的差分隐私方案仍然是可行的，使用者可以根据不同的隐私需求以及恢复精度的要求选择合适的隐私预算。

Table 1. Numerical results on the influence of privacy budget on tensor completion
表 1. 隐私预算对张量填充影响的数值结果

张量大小	观测值 p	隐私预算 ε	迭代次数	标准误差
144 × 256 × 40	0.5	不启用	40	16.443142
144 × 256 × 40	0.5	0.5	40	98.233999
144 × 256 × 40	0.5	1	40	87.989532
144 × 256 × 40	0.5	3	40	61.244712

续表

144 × 256 × 40	0.5	5	40	47.814743
144 × 256 × 40	0.5	10	40	33.427845
144 × 256 × 40	0.5	100	40	17.344673

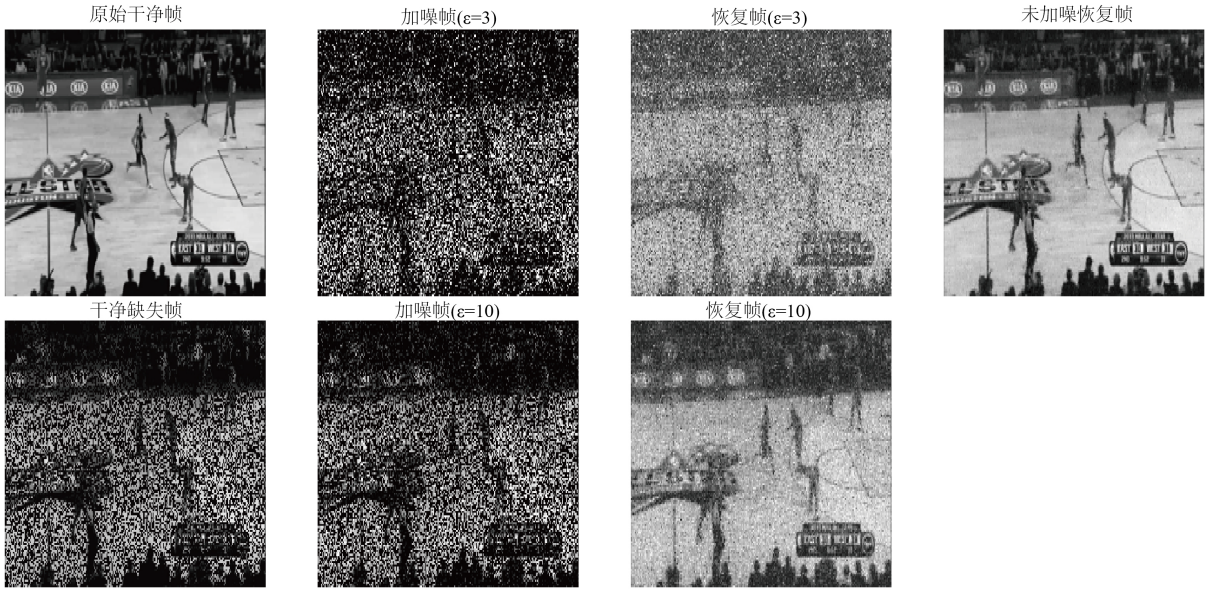

Figure 3. Influence of different privacy budgets on video recovery
图 3. 不同隐私预算对视频恢复的影响

6. 结论

本文提出了一种基于拉普拉斯机制的张量填充方法。该方法通过在观测数据上加入噪声来达到隐私保护的效果，并融入 ADMM-t-SVD 张量填充方案来达成隐私保护的张量填充。通过人工数据集和现实数据集，我们验证了该方法的可行性，并且它还存在一定的优势，即使用者可以根据自身需要来调节隐私预算 ε，具有能动性。此外，除了本文所提到的 ADMM-t-SVD 方案外，我们这一差分隐私方案它还适用于别的张量填充方案，例如 HoMP [17]，TLNM-TQR [18]等矩阵填充算法。未来的研究方向可以考虑其他不同的隐私保护方案，例如联邦学习来增强数据的隐私性，亦可以考虑通过高性能计算来加速张量填充的速度，实现数据隐私和运算时间的双增强。

参考文献

[1] Kolda, T.G. and Bader, B.W. (2009) Tensor Decompositions and Applications. *SIAM Review*, **51**, 455-500. https://doi.org/10.1137/07070111x

[2] Yang, J., Fu, C., Liu, X. and Walid, A. (2022) Recommendations in Smart Devices Using Federated Tensor Learning. *IEEE Internet of Things Journal*, **9**, 8425-8437. https://doi.org/10.1109/jiot.2021.3116505

[3] Zhang, J. and Jiang, J. (2016) Decomposition-Based Tensor Learning Regression for Improved Classification of Multimedia. *Journal of Visual Communication and Image Representation*, **41**, 260-271. https://doi.org/10.1016/j.jvcir.2016.10.006

[4] Zhu, M., Liu, X., Tang, F., Qiu, M., Shen, R., Shu, W., *et al*. (2016) Public Vehicles for Future Urban Transportation. *IEEE Transactions on Intelligent Transportation Systems*, **17**, 3344-3353. https://doi.org/10.1109/tits.2016.2543263

[5] Liu, J., Musialski, P., Wonka, P. and Ye, J. (2013) Tensor Completion for Estimating Missing Values in Visual Data.

IEEE Transactions on Pattern Analysis and Machine Intelligence, **35**, 208-220. https://doi.org/10.1109/tpami.2012.39

[6] Wei, Z., Li, Z., Mao, X. and Wang, J. (2022) Applying Differential Privacy to Tensor Completion. *ICASSP* 2022-2022 *IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), Singapore, 23-27 May 2022, 3923-3927. https://doi.org/10.1109/icassp43922.2022.9747066

[7] Carroll, J.D. and Chang, J. (1970) Analysis of Individual Differences in Multidimensional Scaling via an N-Way Generalization of "Eckart-Young" Decomposition. *Psychometrika*, **35**, 283-319. https://doi.org/10.1007/bf02310791

[8] Harshman, R.A. (1970) Foundations of the PARAFAC Procedure: Models and Conditions for an "Explanatory" Multi-Modal Factor Analysis. *UCLA Working Papers in Phonetics*, **16**, 1-84.

[9] Hitchcock, F.L. (1927) The Expression of a Tensor or a Polyadic as a Sum of Products. *Journal of Mathematics and Physics*, **6**, 164-189. https://doi.org/10.1002/sapm192761164

[10] De Lathauwer, L., De Moor, B. and Vandewalle, J. (2000) A Multilinear Singular Value Decomposition. *SIAM Journal on Matrix Analysis and Applications*, **21**, 1253-1278. https://doi.org/10.1137/s0895479896305696

[11] Kroonenberg, P.M. and de Leeuw, J. (1980) Principal Component Analysis of Three-Mode Data by Means of Alternating Least Squares Algorithms. *Psychometrika*, **45**, 69-97. https://doi.org/10.1007/bf02293599

[12] Tucker, L.R. (1966) Some Mathematical Notes on Three-Mode Factor Analysis. *Psychometrika*, **31**, 279-311. https://doi.org/10.1007/bf02289464

[13] Zhang, Z., Ely, G., Aeron, S., Hao, N. and Kilmer, M. (2014) Novel Methods for Multilinear Data Completion and De-Noising Based on Tensor-SVD. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 3842-3849. https://doi.org/10.1109/cvpr.2014.485

[14] Friedland, S. and Lim, L.H. (2018) Nuclear Norm of Higher-Order Tensors. *Mathematics of Computation*, **87**, 1255-1281. https://doi.org/10.1090/mcom/3239

[15] Dwork, C., McSherry, F., Nissim, K. and Smith, A. (2006) Calibrating Noise to Sensitivity in Private Data Analysis. In: Halevi, S. and Rabin, T., Eds., *Lecture Notes in Computer Science*, Springer, 265-284. https://doi.org/10.1007/11681878_14

[16] Dwork, C. and Roth, A. (2014) The Algorithmic Foundations of Differential Privacy. *Foundations and Trends® in Theoretical Computer Science*, **9**, 211-487. https://doi.org/10.1561/0400000042

[17] Yang, Y., Mehrkanoon, S. and Suykens, J.A.K. (2015) Higher Order Matching Pursuit for Low Rank Tensor Learning. arXiv:1503.02216.

[18] Zheng, Y. and Xu, A. (2021) Tensor Completion via Tensor QR Decomposition and L2,1-Norm Minimization. *Signal Processing*, **189**, Article 108240. https://doi.org/10.1016/j.sigpro.2021.108240