APPLYING DIFFERENTIAL PRIVACY TO TENSOR COMPLETION

Zheng Wei[1], Zhengpin Li[1], Xiaojun Mao[2] and Jian Wang[1]
[1]School of Data Science, Fudan University, China
[2]School of Mathematical Sciences, Shanghai Jiao Tong University, China

1 Abstract

Tensor completion aims at filling the missing or unobserved entries based on partially observed tensors. However, utilization of the observed tensors often raises serious privacy concerns in many practical scenarios. To address this issue, we propose a solid and unified framework that contains several approaches for applying differential privacy to the two most widely used tensor decomposition methods: i) CANDECOMP/PARAFAC and ii) Tucker decompositions. For each approach, we establish a rigorous privacy guarantee and meanwhile evaluate the privacy-accuracy trade-off. Experiments on synthetic datasets demonstrate that our proposal achieves high accuracy for tensor completion while ensuring strong privacy protections.

2 Problem Formulation

For an tensor $\mathcal{X} \in \mathbb{R}^{I\times J\times K}$, the standard Tucker decomposition is:

$$\mathcal{X} = \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{t=1}^{T} g_{pqt}\mathbf{a}_{:p}\circ\mathbf{b}_{:q}\circ\mathbf{c}_{:t} = [\![\mathcal{G};\mathbf{A},\mathbf{B},\mathbf{C}]\!]$$

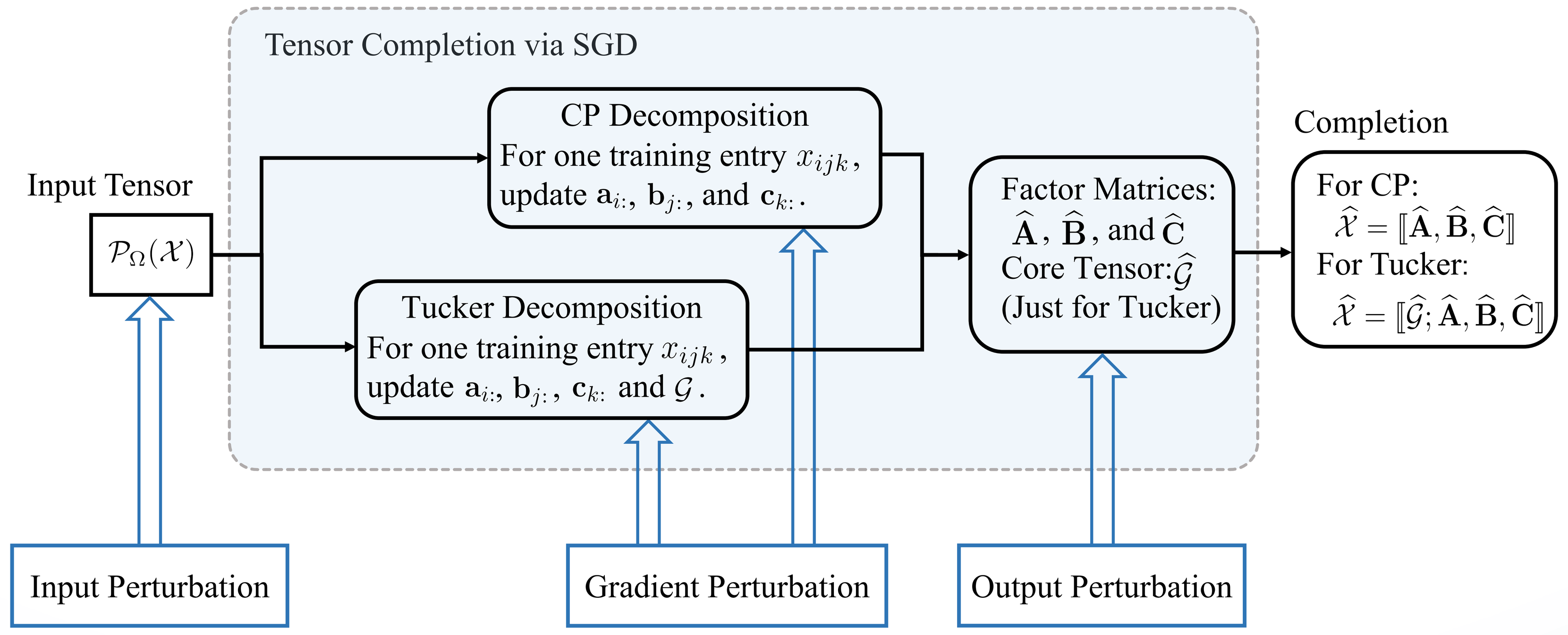
In the special case where $\mathcal{G}$ is superdiagonal (i.e., $g_{pqr}\equiv g_r$) and $P=Q=R$, it would reduce to CP decomposition:

$$\mathcal{X} \approx \sum_{r=1}^{R} g_r\mathbf{a}_r\circ\mathbf{b}_r\circ\mathbf{c}_r = [\![\mathbf{g};\mathbf{A},\mathbf{B},\mathbf{C}]\!],$$

where $\mathbf{g}\in\mathbb{R}^{R}$. Therefore, in the following parts, we provide theoretical analysis and algorithm procedures of the perturbation methods based on Tucker decomposition. By imposing a F-norm penalty to restrict the complexity of the core tensor, the Tucker decomposition problem can be reformulated as:

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C},\mathcal{G}} f(\mathbf{A},\mathbf{B},\mathbf{C},\mathcal{G}) = \|\mathcal{P}_\Omega(\mathcal{X}-[\![\mathcal{G};\mathbf{A},\mathbf{B},\mathbf{C}]\!])\|_F^2 + \lambda_o(\|\mathbf{A}\|_F^2+\|\mathbf{B}\|_F^2+\|\mathbf{C}\|_F^2)+\lambda_g\|\mathcal{G}\|_F^2,$$

3 Differential Privacy Framework



Based on the stages of tensor completion, we design three perturbation approaches (input, gradient and output) to ensure privacy, respectively.

4 Privacy-preserving Perturbation Approaches

- Private Input Perturbation

Algorithm 1 Private Input Perturbation

Input: $\mathcal{X}$: noisy incomplete tensor, Ω: indexes set of observations, d: rank of tensor, λ_o: regularization parameter for the factor matrices, λ_g: regularization parameter for the core tensor, ϵ: privacy budget

Generate each entry of noise tensor $\mathcal{N}$ by $\mathrm{Lap}(\Delta_{\mathcal{X}}^{(I)}/\epsilon)$
Let $\mathcal{X}' = \{x_{ijk}+n_{ijk}|(i,j,k)\in\Omega\}$
Use $\mathcal{X}'$ as input and obtain estimated $\widehat{\mathbf{A}},\widehat{\mathbf{B}},\widehat{\mathbf{C}}$ and $\widehat{\mathcal{G}}$

Output: Estimated $\widehat{\mathbf{A}}\in\mathbb{R}^{n_1\times d}$, $\widehat{\mathbf{B}}\in\mathbb{R}^{n_2\times d}$, $\widehat{\mathbf{C}}\in\mathbb{R}^{n_3\times d}$ and $\widehat{\mathcal{G}}\in\mathbb{R}^{d\times d\times d}$

- Private Gradient Perturbation

Algorithm 2 Private Gradient Perturbation

Input: $\mathcal{X}$: noisy incomplete tensor, Ω: indexes set of observations, d: rank of tensor, λ_o: regularization parameter for the factor matrices, λ_g: regularization parameter for the core tensor, n: number of iterations, ϵ: privacy budget, η: learning rate, m: clipping constant

```
Initialize random factor matrices A, B, C, G
for n iterations do
    for x_ijk ∈ X do
        a_i: ← a_i: − η∇_{a_i:} f
        b_j: ← b_j: − η∇_{b_j:} f
        ∇_{c_k:} f ← ∇_{c_k:} f / max(1, ||∇_{c_k:} f||_2 / m)
        Sample noise n_i: from p(n_i:) ∝ exp(−ε||n_i:|| / Δ_X^(G))
        c_k: ← c_k: − η(∇_{c_k:} f + n_i:)
        G ← G − η∇_G f
    end for
end for
```

Output: Estimated $\widehat{\mathbf{A}}\in\mathbb{R}^{n_1\times d}$, $\widehat{\mathbf{B}}\in\mathbb{R}^{n_2\times d}$, $\widehat{\mathbf{C}}\in\mathbb{R}^{n_3\times d}$ and $\widehat{\mathcal{G}}\in\mathbb{R}^{d\times d\times d}$

- Private Output Perturbation

Algorithm 3 Private Output Perturbation

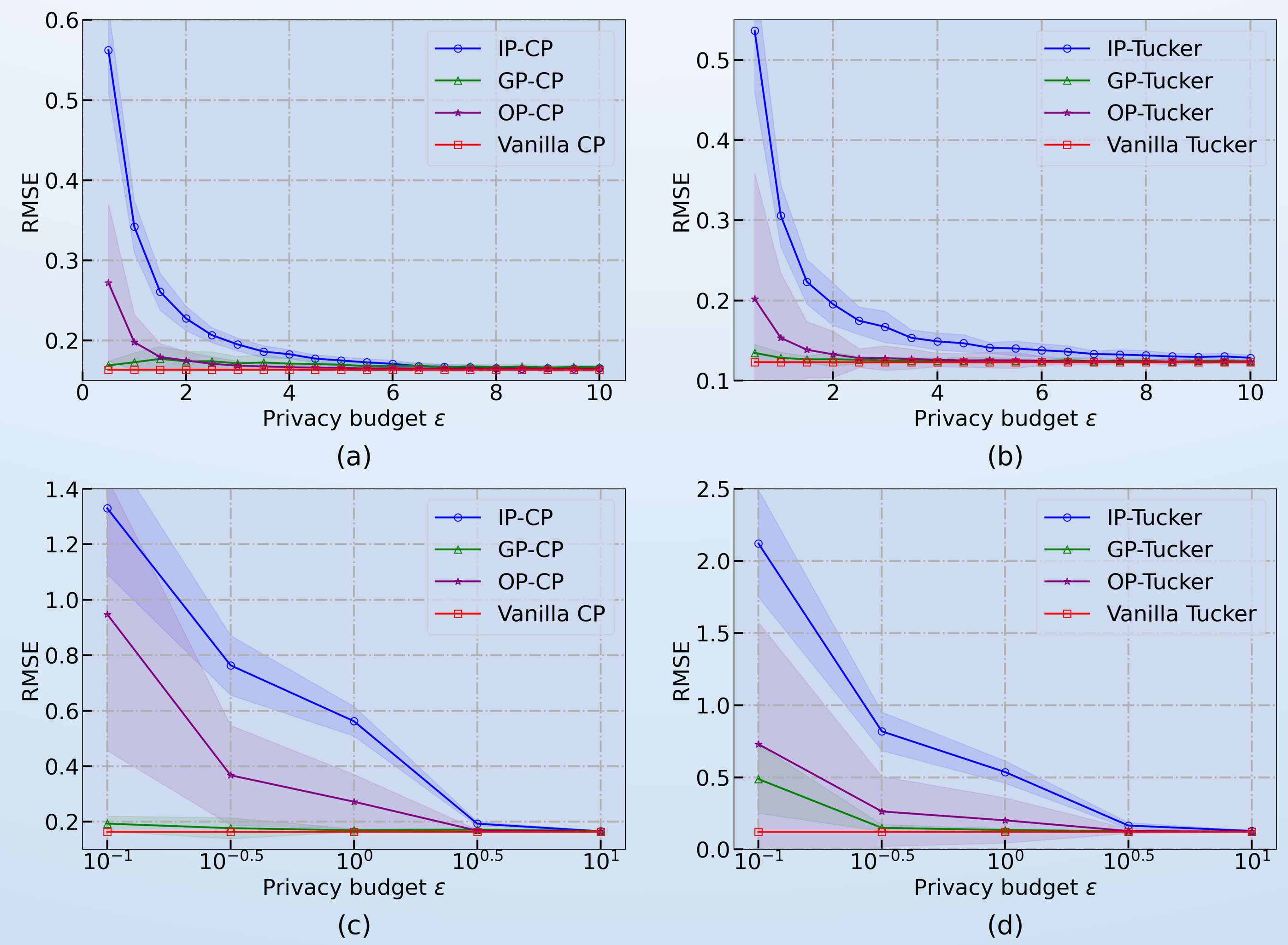
Input: $\mathcal{X}$: noisy incomplete tensor, Ω: indexes set of observations, d: rank of tensor, ϵ: privacy budget

Solve the problem via SGD and obtain estimated $\widehat{\mathbf{A}},\widehat{\mathbf{B}},\widehat{\mathbf{C}}$ and $\widehat{\mathcal{G}}$
Sample noise matrix $\mathbf{N}$, whose rows are sampled from $\exp(-\epsilon\|\mathbf{n}_{i:}\|/\Delta_{\mathcal{X}}^{(O)})$
$\widehat{\mathbf{C}} \leftarrow \widehat{\mathbf{C}} + \mathbf{N}$

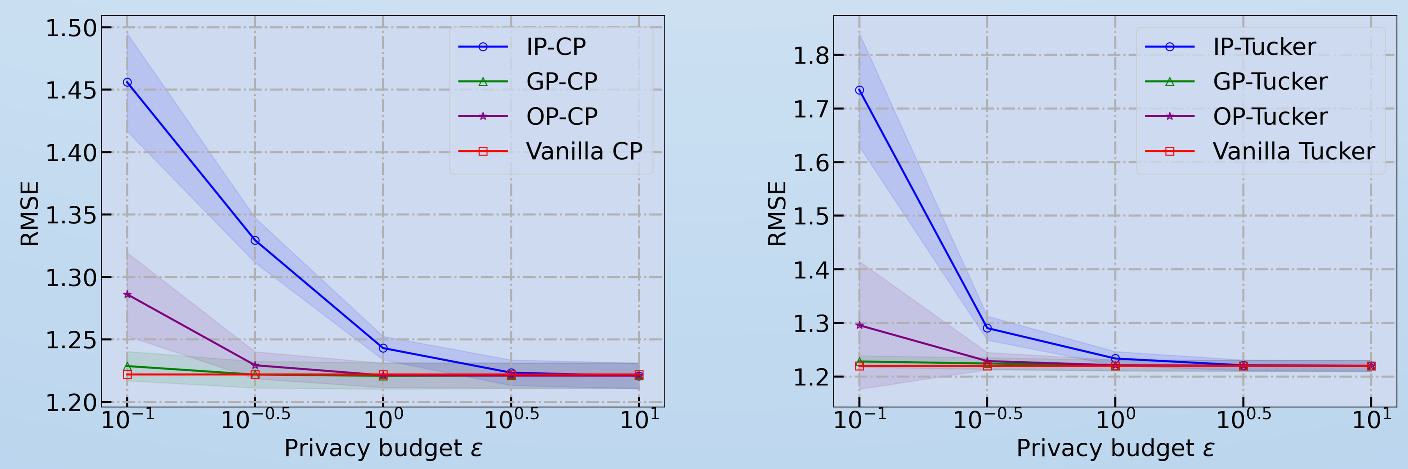
Output: Estimated $\widehat{\mathbf{A}}\in\mathbb{R}^{n_1\times d}$, $\widehat{\mathbf{B}}\in\mathbb{R}^{n_2\times d}$, $\widehat{\mathbf{C}}\in\mathbb{R}^{n_3\times d}$ and $\widehat{\mathcal{G}}\in\mathbb{R}^{d\times d\times d}$

5 Experiments

- Evaluate our proposal on synthetic datasets. The reported results are the average over 50 runs.



- Experimental validation on ML-100K through 10 runs.



(a) CP Decomposition (b) Tucker Decomposition

6 Conclusions

We propose a unified framework for applying differential privacy to tensor completion. In addition, for each privacy-preserving approach in our framework, we provide complete algorithm procedures and theoretical analysis. Experimental results on synthetic and real datasets demonstrate that the proposed approaches can yield high accuracy, while ensuring strong privacy protections.

7 Comments

Zheng Wei and Zhengpin Li contributed equally to this work (Emails: zwei19@fudan.edu.cn; lizp21@m.fudan.edu.cn). Xiaojun Mao and Jian Wang are corresponding authors (Emails: maoxj@sjtu.edu.cn; jian_wang@fudan.edu.cn).