# Algorithmic guarantees

## 1 General setting

We first introduce the regularity condition on the loss function $\mathcal{L}$ and set $\mathcal{S}$.

**Definition 1.** Let f be a real-valued function. We say $f$ satisfies RCG$(\alpha, \beta, \mathcal{S})$ condition for $\alpha, \beta > 0$ and the set $\mathcal{S}$ if,

$$\langle \nabla f(x) - \nabla f(x'), x - x' \rangle \geq \alpha \|x - x'\|_2^2 + \beta \|\nabla f(x) - \nabla f(x')\|_2^2,$$

for any $x, x' \in \mathcal{S}$.

Define

$$\begin{aligned}
\bar{\lambda} &:= \max \left\{ \sigma_{\max}\left(\mathcal{M}_1(\mathcal{B})\right), \sigma_{\max}\left(\mathcal{M}_2(\mathcal{B})\right), \sigma_{\max}\left(\mathcal{M}_3(\mathcal{B})\right) \right\}, \\
\underline{\lambda} &:= \min \left\{ \sigma_{\min}\left(\mathcal{M}_1(\mathcal{B})\right), \sigma_{\min}\left(\mathcal{M}_2(\mathcal{B})\right), \sigma_{\min}\left(\mathcal{M}_3(\mathcal{B})\right) \right\},
\end{aligned}$$

and $\kappa = \bar{\lambda}/\underline{\lambda}$ can be regarded as a tensor condition number. Here $\mathcal{M}_i$ is the matricization operator with respect to $i$-th mode.

We define some constants related to side information $\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3$ as

$$\begin{aligned}
\gamma_1 &:= \prod_{k=1}^{3} \|\boldsymbol{X}_k\|_F^2, \\
\gamma_2 &:= \prod_{k=1}^{3} \sigma_{\min}(\boldsymbol{X}_k)^2.
\end{aligned}$$

Without loss of generality, we scale the the side information matrices $\boldsymbol{X}_k$ so that $\|\boldsymbol{X}_k\|_\infty \leq 1$ for all $k = 1, 2, 3$.

**Lemma 1.1.** Suppose $f \colon \mathbb{R}^{d_1 \times d_2 \times d_3} \to \mathbb{R}$ is a $\alpha_1$-smooth and $\alpha_2$-strongly function. Define $g \colon \mathbb{R}^{p_1 \times p_2 \times p_3} \to \mathbb{R}$ as $g(\mathcal{B}) = f(\mathcal{B} \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\})$ for all $\mathcal{B} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$. Then, $g$ is $\alpha_1\gamma_1$-smooth and $\alpha_2\gamma_2$-strongly convex function.

*Proof.* First, we prove the strong convexity. By definition, we have

$$f(\mathcal{T}_1) \geq f(\mathcal{T}_2) + \langle \nabla f(\mathcal{T}_2), \mathcal{T}_1 - \mathcal{T}_2 \rangle + \frac{\alpha_2}{2} \|\mathcal{T}_1 - \mathcal{T}_2\|_F^2, \text{ for all } \mathcal{T}_1, \mathcal{T}_2 \in \mathbb{R}^{d_1 \times d_2 \times d_3}.$$

Notice that for any $\mathcal{B} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$, we have $\mathcal{B} \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$. Thus, for any $\mathcal{B}_1, \mathcal{B}_2 \in \mathbb{R}^{p_1 \times p_2 \times p_3}$,

$$\begin{aligned}
& f(\mathcal{B}_1 \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\}) \\
& \geq f(\mathcal{B}_2 \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\}) + \langle \nabla f(\mathcal{B}_2 \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\}), (\mathcal{B}_1 - \mathcal{B}_2) \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\} \rangle + \frac{\alpha_2}{2} \|(\mathcal{B}_1 - \mathcal{B}_2) \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\}\|_F^2 \\
& \geq f(\mathcal{B}_2 \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\}) + \langle \nabla f(\mathcal{B}_2 \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\}) \times \{\boldsymbol{X}_1^T, \boldsymbol{X}_2^T, \boldsymbol{X}_3^T\}, \mathcal{B}_1 - \mathcal{B}_2 \rangle + \frac{\alpha_2\gamma_2}{2} \|\mathcal{B}_1 - \mathcal{B}_2\|_F^2. \quad (1)
\end{aligned}$$

Finally, $g$ is $\alpha_2\gamma_2$-strongly convex from (1) because

$$g(\mathcal{B}_1) \geq g(\mathcal{B}_2) + \langle \nabla g(\mathcal{B}_2), \mathcal{B}_1 - \mathcal{B}_2 \rangle + \frac{\alpha_2\gamma_2}{2} \|\mathcal{B}_1 - \mathcal{B}_2\|_F^2, \text{ for all } \mathcal{B}_1, \mathcal{B}_2 \in \mathbb{R}^{p_1 \times p_2 \times p_3}.$$

Smoothness of $g$ is directly followed by

$$\|\nabla g(\mathcal{B}_1) - \nabla g(\mathcal{B}_2)\|_F = \|(\nabla f(\mathcal{B}_1 \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\}) - \nabla f(\mathcal{B}_2 \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\})) \times \{\boldsymbol{X}_1^T, \boldsymbol{X}_2^T, \boldsymbol{X}_3^T\}\|_F$$

$$
\begin{aligned}
&\le \|(\nabla f(\mathcal{B}_1 \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\}) - \nabla f(\mathcal{B}_2 \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\}))\|_F \sqrt{\gamma_1} \\
&\le \alpha_1 \|(\mathcal{B}_1 - \mathcal{B}_2) \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\}\|_F \sqrt{\gamma_2} \\
&= \alpha_1 \gamma_1 \|\mathcal{B}_1 - \mathcal{B}_2\|_F,
\end{aligned}
$$

where the last inequality comes from $\beta$ smoothness of $f$. □

Since negative log-likelihoods of poisson and binomial distribution are not strongly convex and smooth in the unbounded domain. We thus introduce the following assumption on $\mathcal{B}_{\text{true}}$ to ensure that $\mathcal{B}_{\text{true}}$ is in a bounded set.

**Assumption 1.** Suppose $\mathcal{B}_{\text{true}} = \mathcal{C}^* \times \{\boldsymbol{M}_1^*, \boldsymbol{M}_2^*, \boldsymbol{M}_3^*\}$, where $\boldsymbol{M}_k^* \in \mathbb{R}^{p_k \times r_k}$ is a orthogonal matrix for $k = 1, 2, 3$. There exists some constants $\{\mu_k\}_{k=1}^3$, $B$ such that $\|\boldsymbol{M}_k^*\|_{2,\infty}^2 \le \frac{\mu_k r_k}{p_k}$ for $k = 1, 2, 3$ and $\bar{\lambda} \le B\sqrt{\frac{\prod_{k=1}^3 p_k}{\prod_{k=1}^3 \mu_k r_k}}$. Here $\|\boldsymbol{M}_k^*\|_{2,\infty}$ is the largest row-wise $\ell_2$ norm of $\boldsymbol{M}_k^*$.

**Remark 1.** This condition guarantees that $\mathcal{B}_{\text{true}}$ is entry-wise upperbounded by $B$, which guarantees the local strong convexity and smoothness of the negative log-likelihood function.

We define searching spaace $\mathcal{S}$ as follows:

$$
\begin{aligned}
&\mathcal{S} = \mathcal{S}_c \times \mathcal{S}_1 \times \mathcal{S}_2 \times \mathcal{S}_3, \text{ where} \qquad (2) \\
&\mathcal{S}_k = \left\{ (\boldsymbol{M}_k \in \mathbb{R}^{p_k \times r_k} : \|\boldsymbol{M}_k\|_{2,\infty} \le b\sqrt{\frac{\mu_k r_k}{p_k}} \right\} \text{ for } k = 1, 2, 3, \\
&\mathcal{S}_c = \left\{ \mathcal{C} \in \mathbb{R}^{r_1 \times r_2 \times r_3} : \max_k \|\mathcal{M}_k(\mathcal{C})\|_2 \le b^{-3} B \sqrt{\frac{\prod_{k=1}^3 p_k}{\prod_{k=1}^3 \mu_k r_k}} \right\}.
\end{aligned}
$$

## 2 Poisson tensor case

Suppose we observe $\mathcal{Y} \in \mathbb{N}^{d_1 \times d_2 \times d_3}$ that satisfies exp(...)

$$
\mathcal{Y}_{ijk} \sim \text{Poisson}\left(\mathcal{B}_{\text{true}} \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\}\right) \text{ independently.}
$$

where $\mathcal{B}_{\text{true}} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ is the low rank tensor parameter whose rank is $(r_1, r_2, r_3)$.

Then we consider the following negative log-likelihood to estimate $\mathcal{B}_{\text{true}}$,

$$
\mathcal{L}(\mathcal{B} | \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3) = \sum_{ijk} \left( -\mathcal{Y}_{ijk} \left[\mathcal{B}_{\text{true}} \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\}\right]_{ijk} + \exp\left( [\mathcal{B}_{\text{true}} \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\}]_{ijk} \right) \right).
$$

**Theorem 2.1.** Suppose Assumption 1 holds and

1. Initialization: $\|\mathcal{B}_{\text{true}} - \mathcal{B}^{(0)}\|_F^2 \le c_1 \frac{\gamma_1 \gamma_2}{(\gamma_1 e^B + \gamma_2 e^{-B})^2} \kappa^{-2} \underline{\lambda}^2$
2. Signal to noise ratio: $\underline{\lambda}^2 \ge c_2 \kappa^2 e^{3B} \sum_{k=1}^3 (d_1 d_2 d_3 r_k / d_k + d_k r_k)$

where $c_1, c_2 > 0$ are universal constants. Then, with probability at least $1 - \exp(c_3 \max_k d_k)$, we have

Write in terms of two terms (statistical term+ algorithmic term) as in T

$$
\|\hat{\mathcal{B}} - \mathcal{B}_{\text{true}}\|_F^2 \le c_4 \left( r_1 r_2 r_3 + \sum_{k=1}^3 d_k r_k \right),
$$

move yellow term as a remark.

for some constants that do not depend on $d_k$ or $r_k$.

*Proof.* Let $\mathcal{L}'(\mathcal{T}) = \sum_{ijk} (-\mathcal{Y}_{ijk} \mathcal{T}_{ijk} + \exp(\mathcal{T}_{ijk}))$ for all $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$. Then we know $\mathcal{L}'$ is $e^B$ smooth and $e^{-B}$-strongly convex [Han et al., 2020]. By Lemma 1.1, $\mathcal{L}(\mathcal{B} | \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3)$ is $\gamma_1 e^B$-smooth and $\gamma_2 e^{-B}$-strongly convex function. Therefore, based on Lemma E.1 in Han et al. [2020], we know $\mathcal{L}(\mathcal{B} | \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3)$

satisfies RCG$(\alpha, \beta, \mathcal{S})$ with $\alpha = \frac{\gamma_1\gamma_2}{\gamma_1 e^B + \gamma_2 e^{-B}}$ and $\beta = \frac{1}{\gamma_1 e^B + \gamma_2 e^{-B}}$, where $\mathcal{S}$ is defined in (2). Therefore, direct application to Theorem 3.1 in Han et al. [2020] with sufficiently large steps $T$, we have

$$\|\mathcal{B}^{(T)} - \mathcal{B}_{\text{true}}\|_F^2 \le C\frac{\kappa^4}{\alpha}\xi^2, \text{ where} \tag{3}$$
$$\xi = \sup_{\substack{\mathcal{T}\in\mathbb{R}^{p_1\times p_2\times p_3}\\ \text{rank}(\mathcal{T})\le(r_1,r_2,r_3)\\ \|\mathcal{T}\|_F^2\le 1}} \langle \nabla\mathcal{L}(\mathcal{B}|\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3), \mathcal{T}\rangle.$$

yellow = y-b'(theta_true) = residual tensor —> subGaussian by our Propositions 2-3 in pg. 7 JCGS supplemen
see equation (3) page 3 in our supplement.

Now we find an upper bound of $\xi$. Notice that

$$\xi = \sup_{\substack{\mathcal{T}\in\mathbb{R}^{p_1\times p_2\times p_3}\\ \text{rank}(\mathcal{T})\le(r_1,r_2,r_3)\\ \|\mathcal{T}\|_F^2\le 1}} \left\langle (\mathcal{Y} - \exp(\mathcal{B}\times\{\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3\}) \times \{\boldsymbol{X}_1^T,\boldsymbol{X}_2^T,\boldsymbol{X}_3^T\}, \mathcal{T}\right\rangle$$

yellow \times {X^T_1, ..., X^T_3} is a (p1,p2,p3) random tensor
with (correlated) sub-Gaussian entries because of linearity properties of subGaussian.

$$= \sup_{\substack{\mathcal{T}\in\mathbb{R}^{p_1\times p_2\times p_3}\\ \text{rank}(\mathcal{T})\le(r_1,r_2,r_3)\\ \|\mathcal{T}\|_F^2\le 1}} \left\langle \left(\mathcal{Y} - \exp(\mathcal{B}\times\{\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3\}\right), \mathcal{T}\times\{\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3\}\right\rangle$$

==> decorrelation ==> apply their results to (p1,p2,p3) tensor

$$\le \sqrt{\gamma_1} \sup_{\substack{\mathcal{T}'\in\mathbb{R}^{d_1\times d_2\times d_3}\\ \text{rank}(\mathcal{T}')\le(r_1,r_2,r_3)\\ \|\mathcal{T}'\|_F^2\le 1}} \left\langle \left(\mathcal{Y} - \exp(\mathcal{B}\times\{\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3\}\right), \mathcal{T}'\right\rangle$$

Therefore, the bound can be improved to (p1,p2,p3) at an extra cost of IIX_kII_{spectral norm}

$$\le C\sqrt{\gamma_1\left(r_1r_2r_3 + \sum_{k=1}^{3} d_k r_k\right)e^B}, \text{ with probability } 1 - C'/(d_1d_2d_3), \tag{4}$$

for some constants $C, C' > 0$. Here the last inequality comes from Lemma 2.1. Plugging (4) into (3) completes the proof. □

**Lemma 2.1** (Lemma E.10, Han et al. [2020])**.** Let $\mathcal{Y}_{ijk} \sim \text{Poisson}(\mathcal{X}_{ijk})$ independently, and each entry of $\mathcal{X}$ is bounded with $|\mathcal{X}_{ijk}| \le B$. Then, with probability at least $1 - c/p_1p_2p_3$

$$\sup_{\substack{\mathcal{T}\in\mathbb{R}^{p_1\times p_2\times p_3}\\ \text{rank}(\mathcal{T})\le(r_1,r_2,r_3)\\ \|\mathcal{T}\|_F^2\le 1}} \langle \mathcal{Y} - \exp(\mathcal{X}), \mathcal{T}\rangle \le C\sqrt{df(\mathcal{X})e^B},$$

where $df(\mathcal{X})$ is the number of free parameters of $\mathcal{X}$.

## 3 Binomial tensor case

Suppose we observe $\mathcal{Y} \in \{0,1\}^{d_1\times d_2\times d_3}$ that satisfies

logistic(…)

$$\mathcal{Y}_{ijk} \sim \text{Bernoulli}\left(\mathcal{B}_{\text{true}}\times\{\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3\}\right) \text{ independently.}$$

where $\mathcal{B}_{\text{true}} \in \mathbb{R}^{d_1\times d_2\times d_3}$ is the low rank tensor parameter whose rank is $(r_1, r_2, r_3)$.

Then we consider the following negative log-likelihood to estimate $\mathcal{B}_{\text{true}}$,

$$\mathcal{L}(\mathcal{B}|\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3) = -\sum_{ijk}\left(\mathcal{Y}_{ijk}\left[\mathcal{B}_{\text{true}}\times\{\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3\}\right]_{ijk} + \log\left(1+\exp\left(\left[\mathcal{B}_{\text{true}}\times\{\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3\}\right]_{ijk}\right)\right)\right).$$

**Theorem 3.1.** Suppose Assumption 1 holds and

1. Initialization: $\|\mathcal{B}_{\text{true}} - \mathcal{B}^{(0)}\|_F^2 \le c_1\frac{\min(\gamma_1/\gamma_2, \gamma_2/\gamma_1)}{e^B+3}\kappa^{-2}\underline{\lambda}^2$

2. Signal to noise ratio: $\underline{\lambda}^2 \geq c_2\kappa^2 e^{3B}\sum_{k=1}^{3}(d_1d_2d_3r_k/d_k + d_kr_k)$

where $c_1, c_2 > 0$ are universal constants. Then, with probability at least $1 - c_3/(d_1d_2d_3)$, we have

$$\|\hat{\mathcal{B}} - \mathcal{B}_{\text{true}}\|_F^2 \leq c_4\left(r_1r_2r_3 + \sum_{k=1}^{3} d_kr_k\right),$$

for some constants that do not depend on $d_k$ or $r_k$.

*Proof.* Let $\mathcal{L}'(\mathcal{T}) = -\sum_{ijk}\left(\mathcal{Y}_{ijk}\mathcal{T}_{ijk} + \log(1+\exp(\mathcal{T}_{ijk}))\right)$ for all $\mathcal{T} \in \mathbb{R}^{d_1\times d_2\times d_3}$. Then we know $\mathcal{L}'$ is $\frac{1}{e^B+3}$-smooth and $\frac{1}{4}$-strongly convex [Han et al., 2020]. By Lemma 1.1, $\mathcal{L}(\mathcal{B}|\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3)$ is $\frac{\gamma_1}{e^B+3}$-smooth and $\frac{\gamma_2}{4}$-strongly convex function. By Lemma E.1 in Han et al. [2020], we set

$$\alpha = \frac{\min(\gamma_1,\gamma_2)}{2(e^B+3)} \leq \frac{\frac{\gamma_1\gamma_2}{4(e^B+3)}}{\frac{\gamma_1}{e^B+3}+\frac{\gamma_2}{4}} \text{ and } \beta = \frac{1}{2\max(\gamma_1,\gamma_2)} \leq \frac{1}{\frac{\gamma_1}{e^B+3}+\frac{\gamma_2}{4}}$$

and $\mathcal{L}(\mathcal{B}|\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3)$ satisfies RCG$(\alpha,\beta,\mathcal{S})$ with $\mathcal{S}$ is defined in (2). Therefore, direct application to Theorem 3.1 in Han et al. [2020] with sufficiently large steps $T$, we have

$$\|\mathcal{B}^{(T)} - \mathcal{B}_{\text{true}}\|_F^2 \leq C\frac{\kappa^4}{\alpha}\xi^2, \text{ where}$$
$$\xi = \sup_{\substack{\mathcal{T}\in\mathbb{R}^{p_1\times p_2\times p_3}\\ \text{rank}(\mathcal{T})\leq(r_1,r_2,r_3)\\ \|\mathcal{T}\|_F^2\leq 1}} \langle \nabla\mathcal{L}(\mathcal{B}|\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3),\mathcal{T}\rangle.$$

Now we find an upper bound of $\xi$. Notice that

y-b'(Theta_true) ==> sub-Gaussian

$$\begin{aligned}
\xi &= \sup_{\substack{\mathcal{T}\in\mathbb{R}^{p_1\times p_2\times p_3}\\ \text{rank}(\mathcal{T})\leq(r_1,r_2,r_3)\\ \|\mathcal{T}\|_F^2\leq 1}} \left\langle \left(-\mathcal{Y} + \frac{1}{1+\exp(-\mathcal{B}\times\{\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3\})}\right)\times\{\boldsymbol{X}_1^T,\boldsymbol{X}_2^T,\boldsymbol{X}_3^T\},\mathcal{T}\right\rangle \\
&= \sup_{\substack{\mathcal{T}\in\mathbb{R}^{p_1\times p_2\times p_3}\\ \text{rank}(\mathcal{T})\leq(r_1,r_2,r_3)\\ \|\mathcal{T}\|_F^2\leq 1}} \left\langle -\mathcal{Y} + \frac{1}{1+\exp(-\mathcal{B}\times\{\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3\})},\mathcal{T}\times\{\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3\}\right\rangle \\
&\leq \sqrt{\gamma_1} \sup_{\substack{\mathcal{T}'\in\mathbb{R}^{d_1\times d_2\times d_3}\\ \text{rank}(\mathcal{T}')\leq(r_1,r_2,r_3)\\ \|\mathcal{T}'\|_F^2\leq 1}} \left\langle -\mathcal{Y} + \frac{1}{1+\exp(-\mathcal{B}\times\{\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3\})},\mathcal{T}'\right\rangle \\
&\leq C\sqrt{\gamma_1\left(r_1r_2r_3 + \sum_{k=1}^{3} d_kr_k\right)}, \text{ with probability } 1 - C'/(d_1d_2d_3),
\end{aligned}$$

for some constants $C, C' > 0$. Here the last inequality comes from (D.27) in the proof of Theorem 4.5 in Han et al. [2020]. ∎

# 4 Sub-Gaussian case with initial points assumption

Suppose we observe $\mathcal{Y} \in \mathbb{N}^{d_1\times d_2\times d_3}$ that satisfies

$$\mathcal{Y}_{ijk} \sim \text{Sub-Gaussian}\left(\mathcal{B}_{\text{true}}\times\{\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3\},\sigma\right) \text{ independently.}$$

where $\mathcal{B}_{\text{true}} \in \mathbb{R}^{d_1\times d_2\times d_3}$ is the low rank tensor parameter whose rank is $(r_1, r_2, r_3)$.

Then we consider the following negative log-likelihood to estimate $\mathcal{B}_{\text{true}}$,

$$\mathcal{L}(\mathcal{B}|\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3) = \frac{1}{2}\|\mathcal{Y} - \mathcal{B}_{\text{true}} \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\}\|_F^2$$

Following the same proof technique in Section 2,3, we have the following theorem.

**Theorem 4.1.** Suppose Assumption 1 holds and

1. Initialization: $\|\mathcal{B}_{\text{true}} - \mathcal{B}^{(0)}\|_F^2 \leq c_1 \frac{\gamma_2}{\gamma_1} \kappa^{-2} \underline{\lambda}^2$
2. Signal to noise ratio: $\underline{\lambda}/\sigma \geq C_1 d_{\max}^{3/4} r_{\max}^{1/4}$

where $c_1, c_2 > 0$ are universal constants. Then, with probability at least $1 - c_3/(d_1 d_2 d_3)$, we have

$$\|\hat{\mathcal{B}} - \mathcal{B}_{\text{true}}\|_F^2 \leq c_4 \sigma^2 \left( r_1 r_2 r_3 + \sum_{k=1}^{3} d_k r_k \right),$$

for some constants that do not depend on $d_k$ or $r_k$.

**Remark 2.** Notice that our error bound terms and probability have changed from $p_1, p_2, p_3$ to $d_1, d_2, d_3$. The main reason is that we did not consider structure of $\mathcal{T} \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\}$ whose degree of freedom is $r_1 r_2 r_3 + \sum_{i=1}^{3} p_k r_k$ when we calculate $\xi$ we did not consider $\mathcal{T}'$ structure. Instead, we regard $\mathcal{T} \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\}$ as $\mathcal{T}'$ whose degree of freedom is $r_1 r_2 r_3 + \sum_{i=1}^{3} d_k r_k$ to apply lemmas in the reference directly.

Improvable. See earlier comments.

# References


Rungang Han, R. Willett, and Anru Zhang. An optimal statistical and computational framework for generalized tensor estimation. *ArXiv*, abs/2002.11255, 2020.