噪声低管秩张量恢复中小初始化的力量

Zhiyu Liu[1,2], **Haobo Geng**[3], **Xudong Wang**[1,2],**Yandong Tang**[1],**Zhi Han**[1],**Yao Wang**[4,] *

[1] 机器人与智能系统国家重点实验室，中国科学院沈阳自动化研究所，中国[2] 中国科学院大学，中国[3] 电气与电子工程学院，南洋理工大学，新加坡[4] 智能决策与机器学习中心，管理学院，西安交通大学，中国

arXiv:2603.02729v1 [cs.LG] 3 Mar 2026

摘要

我们研究在 t-乘积框架下从含噪线性测量中恢复低管秩张量 $\boldsymbol{\mathcal{X}}_\star \in \mathbb{R}^{n\times n\times k}$ 的问题。一种广泛采用的策略是将优化变量分解为 $\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{U}}^\top$，其中 $\boldsymbol{\mathcal{U}} \in \mathbb{R}^{n\times R\times k}$，然后应用因子化梯度下降（FGD）求解所得优化问题。由于底层张量 $\boldsymbol{\mathcal{X}}_\star$的管秩 r 通常未知，该方法往往假设 $r < R \leq n$，这一情形称为过参数化。然而，当测量受到某种稠密噪声（例如高斯噪声）污染时，采用常用谱初始化的 FGD 所产生的恢复误差会随高估管秩 R线性增长。为解决这一问题，我们证明使用小初始化可使 FGD 达到近乎极小极大最优恢复误差，即使管秩 R 被显著高估。借助四阶段分析框架，我们分析了这一现象，并建立了迄今已知最尖锐的误差界，该误差界与高估管秩 R无关。此外，我们提供了理论保证，表明一种易于使用的提前停止策略在实践中可以达到已知的最佳结果。上述理论发现均通过一系列仿真和真实数据实验得到验证。

1 引言

近年来，现实世界数据日益增长的复杂性和维度凸显了传统向量和矩阵模型的局限性。作为一种自然的泛化形式，张量提供了一个更具表现力的框架，用于捕捉源自高光谱成像（Han等人，2025）、动态视频序列（Han等人，2024）以及传感器阵列（Rajesh & Chaturvedi，2021；Fu等人，2025）等应用的数据中固有的多维相关性。这些应用的一个共同特征是：当数据以张量形式表示时，其通常具有潜在的低秩结构。利用这一性质，大量逆问题可以有效地重新表述为低秩张量恢复任务。典型实例包括图像修复（Zhang & Aeron，2016；Gilman等人，2022；Yang等人，2022）、压缩成像和视频表示（Wang等人，2017；Baraniuk等人，2017；Wang等人，2018）、基于不完整观测的背景建模（Cao等人，2016；Li等人，2022；Peng等人，2022），以及诸如计算机断层扫描等先进医学成像技术（Liu等人，2024a）。低秩张量恢复的目标是根据少量带噪测量恢复目标张量 $\mathrm{X}_\star$： $y_i = \langle \mathrm{A}_i, \mathrm{X}_\star \rangle + s_i$, $i = 1, 2..., m$，其中 s_i 表示未知噪声。该模型可以简洁地表示为 $\boldsymbol{y} = \mathfrak{M}(\mathrm{X}_\star) + \boldsymbol{s}$，其中$\mathfrak{M}(\mathrm{X}_\star) = [\langle \mathrm{A}_1, \mathrm{X}_\star \rangle, \langle \mathrm{A}_2, \mathrm{X}_\star \rangle, ..., \langle \mathrm{A}_m, \mathrm{X}_\star \rangle]$。由于 $\mathrm{X}_\star$具有低秩结构，因此该问题可以通过秩最小化求解：$\min_{\boldsymbol{\mathcal{X}}} \mathrm{rank}(\mathrm{X})$，s. t. $||\boldsymbol{y} - \mathfrak{M}(\mathrm{X})||_2 \leq \epsilon_s$，其中 $\mathrm{rank}(\cdot)$ 表示张量秩函数， ϵ_s 表示噪声水平。

张量分解有多种方法，例如 CANDECOMP/PARAFAC 分解（CP）（Carroll & Chang，1970年；Harshman，1970年）、Tucker 分解（Tucker，1966）和张量

*通讯作者：*yao.s.wang@gmail.com*

奇异值分解（t-SVD）（Kilmer & Martin，2011）、张量列车（Oseledets，2011）和张量环（Zhao 等人，2016），每种方法对张量秩都有不同的定义。本文采用 t-SVD 及其相关的管秩（Kilmer 等人，2013）。采用 t-SVD 是因为它通过 t-积沿第三维进行循环卷积，从而能够有效捕捉频域结构（Wu 等人，2024）。这一能力使其特别适合处理图像和视频等多维数据（He 等人，2024；Wu & Fan，2024；Wu 等人，2025年；Liu等人，2023年；Han 等人，2023）。此外，t-SVD 可保证最优低秩近似，其方式与矩阵的 Eckart–Young定理直接类似（Eckart & Young，1936年）。在 t-SVD 框架下，由于问题（2）是 NP难，一种常见做法是将管秩约束松弛为张量核范数，从而将原问题改写为管状张量核范数最小化。尽管这种松弛在理论上是合理的，但求解时通常需要反复计算 t-SVD；随着张量维度增大，这些计算会变得越来越昂贵。

为了解决这一问题，近期一种更受欢迎的方法是采用张量 Burer–Monteiro（BM）因子分解，这是矩阵Burer–Monteiro方法（Burer & Monteiro，2003）的高阶扩展。该技术将大型张量表示为两个较小因子张量的 t-积，从而将原问题转化为针对这两个因子的优化问题，通常最小化形式为[1]的目标函数

$$\min_{\boldsymbol{\mathcal{U}}\in\mathbb{R}^{n\times R\times k}} f(\boldsymbol{\mathcal{U}}) = \frac{1}{4m}\left\|\boldsymbol{y} - \mathfrak{M}(\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{U}}^\top)\right\|^2, \ \mathfrak{M}(\cdot): \mathbb{R}^{n\times n\times k} \to \mathbb{R}^m, \tag{1}$$

其中 $*$ 表示张量-张量积。随后可以应用因子化梯度下降及其变体，从而显著降低计算成本（Liu 等，2024b；Karnik 等人，2025）。然而，这类方法通常需要预先知道目标张量的管秩 r，但在实践中往往无法获得。因此，通常会假设一个估计秩 $R > r$，这种情况通常称为过参数化或过秩情形。然而，在含噪低管秩张量恢复中，过参数化可能导致更大的恢复误差。Liu 等（2024b）表明，过参数化情形下的恢复误差会随估计的管秩 R线性增长。当管秩被严重高估时，误差可能会变得很大。此外，当管秩被高估时，FGD 的收敛速度会严重减慢。这引出了一个重要问题：在含噪低管秩张量恢复中，是否可能获得一个仅取决于真实管状秩的误差界 r？

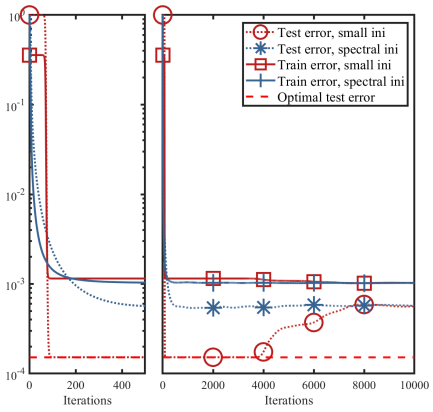


图 1：使用 FGD 并采用谱初始化与小初始化求解问题 (1) 时，训练误差和测试误差的对比。真实张量的管秩为 $r = 2$，过估计秩为 $R = 4$，尺寸为 $n = 20, k = 3$，$m = 5kr$，测量数量为$2n - r$，噪声为 $\sigma = 10^{-3}$。谱初始化遵循 Liu等人（2024b）的方法，而小初始化采用近零起点。训练误差为 $\frac{1}{4m}||\boldsymbol{y} - \mathfrak{M}(\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{U}}^\top)||^2$，测试误差为 $|\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{U}}^\top - \boldsymbol{\mathcal{X}}_\star||_F^2/||\boldsymbol{\mathcal{X}}_\star||_F^2$。“基线”表示在精确秩 $R = r$下的恢复。插图展示了早期（前 500 次迭代）与完整误差曲线。

通过进一步探究这一问题，我们发现采用小初始化时，因子化梯度下降仅依赖于 r，便能以线性速度收敛至近乎极小极大最优的误差，即使管秩被显著高估也是如此。如图 1 所示，在过参数化条件下，采用谱初始化的 FGD 会产生次优的恢复误差，而采用小初始化的 FGD 则能达到与精确管秩设定下相同的误差。然而，随着算法不断迭代，误差会逐渐增大，最终达到与谱初始化相同的水平。我们对这一现象进行了理论分析，并推导出迄今已知的最佳误差界。此外，基于提前停止和验证（Prechelt，1998年；Stone，2018；Ding 等，2025），我们证明了该误差可以达到，并给出了相应的理论保证。

[1]与先前的工作一样，我们假设$\boldsymbol{\mathcal{X}}_\star$ 是一个对称半正定张量。详细说明请参见定义2。

本文的主要贡献总结如下：
最紧的误差上界 我们发现，借助小初始化，FGD 能够在含噪、过参数化的低管秩张量恢复中达到仅依赖于精确管秩的误差。我们为 FGD 建立了全局收敛性，以及仅依赖于真实管状秩的最紧误差界。这显著改进了先前的结果（Liu 等，2024b）。据我们所知，这是首个独立于高估的张量秩的误差界。

极小极大下界与近最优性。 我们推导了含噪管秩张量恢复的信息论极小极大下界，表明任意估计量的均方误差至少为 $\Omega(\frac{nrk\sigma^2}{m})$。将该下界与我们的上界进行比较，表明我们的方法近乎最优；剩余差距仅来自常数因子以及对条件数 κ的依赖。

可实现的恢复误差 将基于验证集的早停方法应用于 FGD，即可在不掌握目标张量任何先验信息的情况下达到该误差界。我们从理论上证明，当验证样本数超过 $\tilde{\mathcal{O}}(r^2\kappa^8)$ 时，验证误差与上界仅相差常数因子。在合成数据集和真实数据集上，我们表明，在过参数化设置下，FGD（小初始化和基于验证集的早停）取得的误差与精确秩设置下相当，并显著优于谱初始化和大型随机初始化。

1.1 相关工作

t-SVD框架下的非凸低管秩张量恢复t-SVD框架下的非凸低管秩张量恢复方法大致可分为两类。第一类旨在通过用非凸替代函数取代管张量核范数来提高恢复精度。第二类则侧重于通过将大型张量分解为

表 1：基于 t-SVD 的几种低管秩张量恢复方法比较。假设噪声向量 $\boldsymbol{s}$ 由均值为零、方差为 σ^2 的高斯随机变量组成。

方法	rate	保证	误差
(Zhang 等，2020)	✗	✓	✗
(Liu 等，2024b)	sub-线性	局部	$\widetilde{\mathcal{O}}\left(\frac{nkR\sigma^2}{m}\right)$
(Karnik 等人，2025)	线性	全局	✗
Ours	线性	全局	$\widetilde{\mathcal{O}}\left(\frac{nkr\sigma^2}{m}\right)$

较小的因子张量来提高计算效率。我们首先讨论第一类方法。这些方法源自管张量核范数，包括t-Schatten-p范数（Kong 等人，2018）、加权t-TNN（Mu 等人，2020）、t-TNN部分和（Jiang 等人，2020）等变体（Qin 等人，2025）。其他方法则采用 Geman 或 Laplace惩罚等非凸函数来替代管张量核范数（Cai 等人，2019；Xu 等人，2019）。值得注意的是，Wang 等人（Wang 等人，2021）提出了一个广义非凸框架，涵盖了多种非凸惩罚函数。然而，这些方法仍依赖于重复计算 t-SVD，计算开销较大，而且往往缺乏理论保证。第二类包括基于分解的方法，即将大型张量分解为两个或三个较小的因子张量，再采用交替最小化（Zhou等，2017年；Liu 等人，2019；He 和 Atia，2023；Wu 等人，2025）、非凸张量范数最小化（Du 等人，2021；Jiang 等人，2023b）、因子化梯度下降（Liu 等，2024b；Karnik 等人，2025）或缩放梯度下降（Feng 等人，2025；Wu，2025；Liu 等人，2025）等优化技术。除这两大类方法外，还有基于随机低秩近似（Qin 等人，2024）和交替投影（Qiu 等人，2022）的方法，用于求解张量恢复问题。

低秩张量恢复中的过参数化基于分解的方法通常需要已知张量秩。然而，在实际应用中，真实秩往往难以确定。因此，通常假设估计秩大于真实秩，这种情形称为过参数化。在矩阵感知中，已有研究表明，即使在过参数化情形下，梯度下降仍能达到最优解（Zhu et al., 2018；St¨oger & Soltanolkotabi, 2021；Soltanolkotabi et al., 2025；Jiang et al., 2023a；Zhuo et al., 2024；Ding 等，2025）。相比之下，关于张量恢复中过参数化情形的研究相对有限。尽管已有许多估计张量秩的方法，但这些方法计算开销较大，且缺乏明确的理论保证（Zhou & Cheung, 2019；Shi等人，2021；Zheng et al., 2023；Zhu et al., 2025）。近期，Liu等人（2024b）研究了高估管秩情形下的低管秩张量恢复

并为 FGD 建立了局部收敛保证和恢复误差界，其中误差取决于高估管秩。Karnik 等人（2025）进一步证明，在无噪声情形下，高估管秩时采用小初始化的 FGD 可以全局收敛。此外，对于 Tucker 分解，Luo 和 Zhang（2024）研究了张量对张量回归中的过参数化情形。然而，在存在噪声时，其恢复误差仍取决于高估的张量秩。我们将所提方法与若干密切相关的工作进行了比较，结果汇总于表 1。

2 预备知识

符号 y、$\boldsymbol{y}$、$\boldsymbol{Y}$、Y 分别表示标量、向量、矩阵和张量。设 $\mathrm{Y} \in \mathbb{R}^{m\times n\times k}$ 为一个三阶张量。我们将其位于位置 (i,j,l) 的元素记为 $\mathrm{Y}(i,j,l)$，并按照 MATLAB 风格索引，将第 l 个正面切片记为 $\boldsymbol{Y}(l) := \mathrm{Y}(:,:,l)$。两个张量 Y 与 Z 的内积由 $\langle \mathrm{Y}, \mathrm{Z}\rangle = \sum_{l=1}^{k} \langle \boldsymbol{Y}(l)\, \boldsymbol{Z}^{(l)}\rangle$ 给出，其中每个 $\boldsymbol{Y}(l)$ 与 $\boldsymbol{Z}^{(l)}$ 为对应的前切片。

对于任意张量 $Y \in \mathbb{R}^{m\times n\times k}$，其沿第三模的离散傅里叶变换得到 $\overline{Y} \in \mathbb{C}^{m\times n\times k}$。在 *MATLAB* 语法中，我们有 $\overline{Y} = \mathit{fft}(Y,[\,],3)$，以及 $Y = \mathit{ifft}(\overline{Y},[\,],3)$。我们将 $\overline{\boldsymbol{Y}} \in \mathbb{C}^{mk\times nk}$ 记为 $\overline{Y}$ 的分块对角矩阵，即 $\overline{\boldsymbol{Y}} = \mathit{bdiag}(\overline{Y}) = \mathit{diag}(\overline{Y}(1);\ \overline{Y}(2);\ \ldots;\ \overline{Y}(k))$。

两个张量的张量-张量积（t-积）$\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{m\times q\times k}$ 和 $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{q\times n\times k}$ 为 $\boldsymbol{\mathcal{Z}} * \boldsymbol{\mathcal{Y}} \in \mathbb{R}^{m\times n\times k}$，其纤维由以下公式给出：$(\boldsymbol{\mathcal{Z}}\boldsymbol{\mathcal{Y}})(i,i',:) = \sum_{p=1}^{q} \boldsymbol{\mathcal{Z}}(i,p,:) * \boldsymbol{\mathcal{Y}}(p,i',:)$，其中 $*$ 表示循环卷积运算，即 $(\boldsymbol{x} * \boldsymbol{y})_i = \sum_{j=1}^{k} x_j y_{i-j(\bmod k)}$.

对于任意张量 $\boldsymbol{\mathcal{Y}} \in \mathbb{C}^{m\times n\times k}$，其共轭转置 $\boldsymbol{\mathcal{Y}}^{\top} \in \mathbb{C}^{n\times m\times k}$ 通过对每个正面切片取共轭转置，并将第 2 至 k 个切片的顺序反转来计算。单位张量以 $\boldsymbol{\mathcal{I}} \in \mathbb{R}^{n\times n\times k}$ 表示，其定义为：第一个正面切片对应于 $n\times n$ 单位矩阵，而其余所有正面切片均由零构成。数学表达式如下：$\boldsymbol{I}^{(1)} = \boldsymbol{I}_{n\times n},\quad \boldsymbol{I}^{(l)} = 0, l = 2,3,\ldots,k.$ 若张量 $\boldsymbol{\mathcal{Q}} \in \mathbb{R}^{n\times n\times k}$ 满足以下条件，则称其为正交张量：$\boldsymbol{\mathcal{Q}}^{\top} * \boldsymbol{\mathcal{Q}} = \boldsymbol{\mathcal{Q}} * \boldsymbol{\mathcal{Q}}^{\top} = \boldsymbol{\mathcal{I}}.$

定理 1 (t-SVD (Kilmer& Martin, 2011))。 设 $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{m\times n\times k}$，则可分解为 $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{Y}}} * \boldsymbol{\mathcal{S}}_{\boldsymbol{\mathcal{Y}}} * \boldsymbol{\mathcal{W}}_{\boldsymbol{\mathcal{Y}}}^{\top}$，其中 $\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{Y}}} \in \mathbb{R}^{m\times m\times k}$、$\boldsymbol{\mathcal{W}}_{\boldsymbol{\mathcal{Y}}} \in \mathbb{R}^{n\times n\times k}$ 为正交张量，且 $\boldsymbol{\mathcal{S}}_{\boldsymbol{\mathcal{Y}}} \in \mathbb{R}^{m\times n\times k}$ 为 *F*-对角张量，即 $\boldsymbol{\mathcal{S}}\mathcal{Y}$ 的所有前切片均为对角矩阵。

对于 $\mathrm{Y} \in \mathbb{R}^{m\times n\times k}$，其管秩 $\mathrm{rank}_t(\mathrm{Y})$ 定义为 $\mathrm{S}\mathcal{Y}$ 的非零对角管，其中 $\mathrm{S}\mathcal{Y}$ 是 Y 的 t-SVD 中的 F-对角张量。即 $\mathrm{rank}_t(\mathrm{Y}) := \#\{i : \mathrm{S}\mathcal{Y}(i,i,:) \neq 0\}$.其平均秩定义为 $\mathrm{rank}_a(\mathrm{Y}) = \frac{1}{k}\sum_{i}^{k} \mathrm{rank}(\boldsymbol{Y}^{(i)})$。张量 $\mathrm{Y} \in \mathbb{R}^{m\times n\times k}$ 的条件数定义为 $\kappa(\mathrm{Y}) = \frac{\sigma_1(\boldsymbol{Y})}{\sigma_{\min}(\boldsymbol{Y})}$，其中 $\boldsymbol{Y}$ 是张量 Y 的分块对角矩阵，且 $\sigma_1(\boldsymbol{Y}) \geq \cdots \geq \sigma_{\min}(\boldsymbol{Y}) > 0$ 表示 $\boldsymbol{Y}$ 的奇异值。对于 $\mathrm{Y} \in \mathbb{R}^{m\times n\times k}$，其谱范数记为 $\|\mathrm{Y}\| := \|\mathrm{bcirc}(\mathrm{Y})\| = \|\boldsymbol{Y}\|$；其弗罗贝尼乌斯范数定义为 $\|\mathrm{Y}\|_F := \sqrt{\sum_{i,j,l} \mathrm{Y}(i,j,l)^2}$；其管张量核范数定义为 $\|\mathrm{Y}\|_* := \|\boldsymbol{Y}\|_*$(Karnik 等人，2025)。

3 主要结果

3.1 因子化梯度下降与 t-RIP

首先，我们给出用于求解问题 (1) 的因子化梯度下降法的详细更新规则：$\boldsymbol{\mathcal{U}}_0 \sim \mathcal{N}(0, \frac{\alpha^2}{R})$，$\boldsymbol{\mathcal{U}}_{t+1} = \boldsymbol{\mathcal{U}}_t - \eta \cdot \frac{1}{m}\mathfrak{M}^{*}\left(\mathfrak{M}(\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^{\top} - \boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^{\top}) - s\right) * \boldsymbol{\mathcal{U}}_t$，其中 $\mathfrak{M}^{*}(\mathbf{e}) = \sum_{i=1}^{m} e_i \boldsymbol{\mathcal{A}}_i$ 且 $\boldsymbol{\mathcal{X}}_{\star} = \boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^{\top}, \boldsymbol{\mathcal{X}} \in \mathbb{R}^{n\times r\times k}$。分析因子化梯度下降收敛性的一个常见假设是 t-RIP，其定义如下：

定义 1 （t-RIP（Zhang 等，2021））。 若线性映射 $\mathfrak{M}: \mathbb{R}^{n\times n\times k} \to \mathbb{R}^{m}$ 满足 (r,δ) 张量限制等距性质（*t-RIP*），则称其对于 $\delta \in [0,1]$ 成立：对于任意张量 $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{n\times n\times k}$，若其管秩为 $\leq r$，则满足以下不等式：$(1-\delta)\|\boldsymbol{\mathcal{Y}}\|_F^2 \leq \|\mathfrak{M}(\boldsymbol{\mathcal{Y}})\|^2/m \leq (1+\delta)\|\boldsymbol{\mathcal{Y}}\|_F^2.$

已有研究表明，当 $m \gtrsim rnk/\delta^2$时，t-RIP 条件以高概率成立（Zhang 等，2021）；前提是算子 $\mathfrak{M}$ 中的每个测量张量 $\boldsymbol{\mathcal{A}}_i$ 的元素均独立地取自均值为 0、方差为 1 的次高斯分布。需要注意的是，该条件已被先前的研究广泛采用（Zhang等人，2020；Liu 等，2024b；Karnik 等人，2025），因此在我们的设定中，这是一个自然且合理的假设。

我们将 FGD 更新分解为

$$\boldsymbol{\mathcal{U}}_{t+1} = \boldsymbol{\mathcal{U}}_t - \eta(\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top - \boldsymbol{\mathcal{X}}_\star) * \boldsymbol{\mathcal{U}}_t + \eta \underbrace{\left(\mathfrak{I} - \frac{\mathfrak{M}^*\mathfrak{M}}{m}\right)(\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top - \boldsymbol{\mathcal{X}}_\star)}_{(a)} * \boldsymbol{\mathcal{U}}_t + \eta \cdot \underbrace{\frac{1}{m}\mathfrak{M}^*(\boldsymbol{s})}_{(b):=\boldsymbol{\mathcal{E}}} * \boldsymbol{\mathcal{U}}_t,$$

其中 $\mathfrak{I}: \mathbb{R}^{n\times n\times k} \to \mathbb{R}^{n\times n\times k}$ 表示恒等映射。随后，分别应用 t-RIP 条件和张量集中界来控制项 (a) 和 (b)。

3.2 理论保证

We 首先，我们通过 FGD 建立求解含噪低管秩张量恢复问题的理论保证 with 小初始化。

定理 2. 假设以下条件成立：*(1)* 线性映射 $\mathfrak{M}$ 满足$(2r+1,\delta)$ *t-RIP*，且 $\delta \le \frac{c}{\sqrt{kr\kappa^4}}$；*(2)* 步长 $\eta \le c\kappa^{-4}||\boldsymbol{\mathcal{X}}||^{-2}$；*(3)* 误差项 $\boldsymbol{\mathcal{E}} := \frac{1}{m}\mathfrak{M}^*(\boldsymbol{s})$ 满足 $||\boldsymbol{\mathcal{E}}|| \le c\kappa^{-2}\sigma_{\min}^2(\boldsymbol{\mathcal{X}})$；*(4)* 初始点 $\boldsymbol{\mathcal{U}}_0$ 的每个元素均为 $i.i.d\,\mathcal{N}(0, \frac{\alpha^2}{R})$；*(5)* $\boldsymbol{\mathcal{X}}_\star \in \mathbb{R}^{n\times n\times k}$ 是一个满管秩 r 张量。在所有这些假设下，以下结论以至少 $1 - ke^{-\tilde{c}R} - \max\{k(\tilde{C}\epsilon)^{R-r+1}, k\epsilon^2\}$ 的概率成立。

1. 当 $R = r$，且初始化尺度满足 $\alpha \lesssim \frac{\sqrt{r}\sigma_{\min}(\boldsymbol{\mathcal{X}})}{k^{23/14}(R\wedge n)\kappa^2}\left(\frac{2\kappa^2\sqrt{rn}}{\tilde{c}_3}\right)^{-10\kappa^2}$ 时，我们有

$$||\boldsymbol{\mathcal{U}}_{\hat{t}} * \boldsymbol{\mathcal{U}}_{\hat{t}}^\top - \boldsymbol{\mathcal{X}}_\star||_F \lesssim \sqrt{r}\kappa^2||\boldsymbol{\mathcal{E}}||, \text{ where } \hat{t} \gtrsim \frac{1}{\eta\sigma_{\min}^2(\boldsymbol{\mathcal{X}})}\ln\left(\frac{\kappa^2 r^{3/2}\sqrt{n}}{\sqrt{k}\alpha\sigma_{\min}(\boldsymbol{\mathcal{X}})}\right).$$

2. 当 $r < R < 3r$ 且初始化尺度 α 满足$\alpha \lesssim$最小 $\left\{\frac{\sigma_{\min}}{R\wedge n}(\boldsymbol{\mathcal{X}})\kappa^2,\ \kappa^{\frac{35}{21}}||\boldsymbol{\mathcal{E}}||^{\frac{16}{21}}((R\wedge n)-r)^{\frac{4}{7}}||\boldsymbol{\mathcal{X}}||^{\frac{11}{21}}\right\} \frac{r}{k^{23/14}}\left(\frac{2\kappa^2\sqrt{rn}}{\tilde{c}_3}\right)^{-10\kappa^2}$ 时，我们有

$$||\boldsymbol{\mathcal{U}}_{\hat{t}} * \boldsymbol{\mathcal{U}}_{\hat{t}}^\top - \boldsymbol{\mathcal{X}}_\star||_F \lesssim \sqrt{r}\kappa^2||\boldsymbol{\mathcal{E}}||, \text{ where } \hat{t} \asymp \frac{1}{\eta\sigma_{\min}^2(\boldsymbol{\mathcal{X}})}\ln\left(\frac{n^{\frac{1}{2}}r^{\frac{5}{2}}\kappa^2||\boldsymbol{\mathcal{X}}||^2}{k^2((R\wedge n)-r)\alpha^2}\right).$$

3. 当 $R \ge 3r$ 且初始化尺度满足$\alpha \lesssim$最小 $\left\{\frac{\sigma_{\min}}{R\wedge n}(\boldsymbol{\mathcal{X}})\kappa^2,\ \kappa^{\frac{35}{21}}||\boldsymbol{\mathcal{E}}||^{\frac{16}{21}}((R\wedge n)-r)^{\frac{4}{7}}||\boldsymbol{\mathcal{X}}||^{\frac{11}{21}}\right\} \frac{1}{k^{23/14}}\left(\frac{2\kappa^2\sqrt{n}}{\tilde{c}_3\sqrt{r}}(R\wedge n)\right)^{-10\kappa^2}$ 时，我们有

$$||\boldsymbol{\mathcal{U}}_{\hat{t}} * \boldsymbol{\mathcal{U}}_{\hat{t}}^\top - \boldsymbol{\mathcal{X}}_\star||_F \lesssim \sqrt{r}\kappa^2||\boldsymbol{\mathcal{E}}||, \text{ where } \hat{t} \asymp \frac{1}{\eta\sigma_{\min}(\boldsymbol{\mathcal{X}})^2}\ln\left(\frac{\sqrt{n}\kappa^2||\boldsymbol{\mathcal{X}}||^2}{k^2((R\wedge n)-r)(R\wedge n)\alpha^2}\right).$$

这里，$c, \tilde{c}, \tilde{c}_3, \epsilon, C$ 是固定的数值常数，并定义 $R\wedge n := \min\{R,n\}$, $\kappa := \kappa(\boldsymbol{\mathcal{X}})$。

备注 1. *(*恢复误差*)* 我们最终的恢复误差为 $\sqrt{r}\kappa^2||\boldsymbol{\mathcal{E}}||$，它仅取决于噪声项 $\boldsymbol{\mathcal{E}}$ 的谱范数、$\boldsymbol{\mathcal{X}}$ 的条件数 κ 以及真实管状秩 r。我们不对噪声分布作任何具体假设，仅要求 $||\boldsymbol{\mathcal{E}}|| \le c\kappa^{-2}\sigma_{\min}^2(\boldsymbol{\mathcal{X}})$。因此，我们的结果可能适用于广泛的噪声分布。当噪声为高斯噪声时，我们的误差界可化为 *(Liu*等人，*2024b)* 的结果。然而，一个关键区别是，我们的误差界仅取决于真实管状秩 r，而 *Liu* 等人（*2024b*）的误差界则取决于高估管秩 R。

接下来，我们给出一个定理，用以刻画高斯噪声情形下的极小极大误差。定理 3 确立了低管秩张量恢复的基本统计极限。具体而言，对于任意估计过程，在管秩至多为 r 的张量上，均方误差不可能一致地低于 $\Theta(nrk\sigma^2/m)$ 阶。此外，存在一些参数选择，使得误差以常数概率达到这一阶。

定理 3 (**极小极大误差**)。 假设线性映射 $\mathfrak{M}(\mathbb{R}^{n\times n\times k}$ r $s\sim\mathcal{N}\ 0\ \sigma^2\boldsymbol{I}$ $\star\in$) 满足 (, δ) *t-RIP*，$\boldsymbol{\mathcal{X}}$是一个满管秩张量，并且有(,)，那么任意估计量$\boldsymbol{\mathcal{X}}_{est}$均满足

$$\sup_{\boldsymbol{\mathcal{X}}_\star}\mathbb{E}||\boldsymbol{\mathcal{X}}_{est}-\boldsymbol{\mathcal{X}}_\star||_F^2\geq\frac{1}{1+\delta}\frac{nrk\sigma^2}{m},\ \sup_{\boldsymbol{\mathcal{X}}_\star}\mathbb{P}\left(||\boldsymbol{\mathcal{X}}_{est}-\boldsymbol{\mathcal{X}}_\star||_F^2\geq\frac{nrk\sigma^2}{2m(1+\delta)}\right)\geq 1-e^{-\frac{nrk}{16}}.$$

在高斯噪声下的极小极大误差情形中，我们进一步证明，当 $s\sim\mathcal{N}(0,\ \sigma^2)$ 时，采用小初始化的 FGD 可收敛至近乎最优的误差。

推论 1。 （高斯情形下近乎极小极大最优误差）在定理 *2* 的假设下，进一步假设噪声向量 $\boldsymbol{s}$ 的各个元素均服从均值为零、方差为 σ^2 的高斯分布，且测量数量满足 $m\gtrsim nk\kappa^4\frac{\sigma^2}{\sigma_{\min}^4}$（$\boldsymbol{\mathcal{X}}$）。那么，以高概率有 $\|\boldsymbol{\mathcal{U}}_{\hat{t}}*\boldsymbol{\mathcal{U}}^\top_{\hat{t}}-\boldsymbol{\mathcal{X}}_\star\|_F^2\lesssim\frac{nkr\kappa^4\sigma^2}{m}$，其中 $\hat{}\boldsymbol{\mathcal{U}}_t$ 与定理 *2* 中的相同。

注 2。 （采样复杂度）我们对测量数量 m 的假设主要源于 *t-RIP* 条件，该条件要求 $m\gtrsim nkr/\delta^2$。在本文中，我们仅依赖 $(2r+1,\ \delta)$ ***t-RIP***，而不依赖高估管秩 R，这与 *(Karnik* 等人，*2025)* 中的设置一致。相比之下，*(Liu* 等，*2024b)* 要求 $(4R,\ \delta)$ ***t-RIP***，导致随着高估管秩 R 的增加，采样复杂度更高。需要注意的是，这一采样复杂度仅是理论保证所需的；在实践中，如图 *2 (d)* 所示，更小的样本量就已足够。

注 3. *(*与（***Liu*** 等，***2024b***）的比较*)* 本文与（*Liu* 等，*2024b*）均采用因子化梯度下降算法求解低管秩张量恢复问题。它们率先将 *FGD* 应用于这一问题，并对收敛性和恢复误差进行了分析。然而，我们的工作在几个关键方面与其显著不同：（***1***）初始化：他们依赖谱初始化来获得足够好的初始点，以进行理论分析。相比之下，我们的方法只需小型随机初始化即可保证收敛。这两种初始化策略带来了截然不同的分析框架和理论结果。（***2***）收敛率：在（*Liu* 等，*2024b*）中，过参数化条件下的收敛率为次线性，而我们的分析表明，即使在过参数化情形下，收敛率仍保持线性。（***3***）恢复误差：他们的恢复误差取决于过参数化的管秩 R，而我们的误差仅取决于真实管秩 r。（***4***）采样复杂度：他们要求测量算子 $\mathfrak{M}$ 满足（$4R,\delta$）*t-RIP* 条件，而我们只要求$\mathfrak{M}$ 满足（$2r+1,\delta$）*t-RIP* 条件。因此，（*Liu* 等，*2024b*）中的采样复杂度会随着过参数化程度的增加而增长，而我们的要求则较为宽松，且与 R无关。

备注 4. *(*与 ***Karnik*** 等人（***2025***）的比较*)* 另一项相关工作是 *Karnik* 等人（*2025*），该研究探讨了小初始化下的管状张量恢复。我们的工作与其在几个关键方面有所不同。***(1)*** 问题设定：他们关注小初始化的隐式正则化效应，而我们的目标是在小初始化下，为含噪的低管秩张量恢复提供理论保证。***(2)*** 技术工具：他们的分析仅将 *FGD* 轨迹划分为两个阶段——谱阶段和收敛阶段，无法精确刻画噪声演化。相比之下，我们提出了四阶段分解，对轨迹进行了更细致的描述，从而能够追踪噪声在所有阶段的影响。因此，将他们的结果直接扩展到含噪张量情形，并不能得到极小极大最优恢复保证。***(3)*** 理论结果：我们的分析对参数的限制更宽松。例如，我们的定理 *2* 中初始化尺度的上界 α 比 *[Karnik* 等人（*2025*）的定理 *3]1*宽松得多。

注 5。（关于管状秩估计方法的讨论）过去五年中，已有许多采用秩估计策略的低管秩张量恢复方法被提出（*Shi* 等人（*2021*）；*Zheng* 等人（*2023*）；*Zhu* 等人（*2025*））。（***1***）问题设定：我们的目标是，即使指定的管状秩上界超过真实管状秩，也能实现稳定恢复，并确保误差不会随着上界增大而恶化。相比之下，管状秩估计方法旨在识别或近似真实管状秩。（***2***）噪声模型：*Shi* 等人（*2021*）和 *Zhu* 等人（*2025*）研究了存在稀疏噪声时的秩估计，而 *Zheng* 等人（*2023*）则着重于无噪声情形下快速且稳健的秩估计。我们的结果适用于存在次高斯噪声的情形。（***3***）理论保证：据我们所知，上述工作并未在 *t-SVD* 和管状秩设定下提供与秩无关的误差界。我们的主要贡献是建立此类与管状秩无关的保证，并证明其具有近极小极大统计精度。

3.3 证明概要

将$\boldsymbol{\mathcal{X}}$的张量列子空间定义为$\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}} \in \mathbb{R}^{n\times r\times k}$。考虑张量$\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^{\top} * \boldsymbol{\mathcal{U}}_t$及其对应的 t-SVD$\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^{\top} * \boldsymbol{\mathcal{U}}_t = \boldsymbol{\mathcal{V}}_t * \boldsymbol{\mathcal{S}}_t * \boldsymbol{\mathcal{W}}_t^{\top}$，其中$\boldsymbol{\mathcal{W}}_t \in \mathbb{R}^{R\times r\times k}$. 。我们将$\boldsymbol{\mathcal{W}}_{t,\perp} \in \mathbb{R}^{R\times(n-r)\times k}$记作一个张量，其张量列子空间与$\boldsymbol{\mathcal{W}}_t$. 的列子空间正交。于是，我们可以将$\boldsymbol{\mathcal{U}}_t$分解为“信号项”和“过参数化项”：

$$\boldsymbol{\mathcal{U}}_t = \underbrace{\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t * \boldsymbol{\mathcal{W}}_t^{\top}}_{\text{signal term}} + \underbrace{\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp} * \boldsymbol{\mathcal{W}}_{t,\perp}^{\top}}_{\text{over-parameterization term}}. \tag{2}$$

通过这种分解，我们可以分别分析信号项和过参数化项。具体而言，我们考虑以下三个量来研究 FGD 的收敛行为：

- $\sigma_{\min}(\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t)$：信号项的大小；
- $\|\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp}\|$：过参数化项的大小；
- $\|\boldsymbol{\mathcal{V}}^{\top}\boldsymbol{\mathcal{X}}^{\perp} * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t}\|$：信号的列空间与真实值的列空间之间的对齐程度。

接下来，我们将 FGD 的轨迹划分为四个阶段：

I. 对齐阶段： 在此阶段，信号项$\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t$的列空间逐渐与真实值$\boldsymbol{\mathcal{X}}_\star$的列空间对齐，这一点可由$\|\boldsymbol{\mathcal{V}}^{\top}\boldsymbol{\mathcal{X}}^{\perp} * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t}\|$的值不断减小看出。由于采用了小初始化，$\sigma_{\min}(\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t)$和$\|\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp}\|$均保持较小。

II. 信号放大阶段： 此时，$\sigma_{\min}(\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t)$呈指数增长，直至至少达到$\sigma_{\min}(\boldsymbol{\mathcal{X}})\sqrt{10}$，而$\|\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp}\|$则基本保持在初始化的量级。

III. 局部细化阶段： 在此阶段，利用分解式 (3)，误差被分解为 a s

$$\|\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^{\top} - \boldsymbol{\mathcal{X}}_\star\| \le 4\|\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^{\top} * (\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^{\top} - \boldsymbol{\mathcal{X}}_\star)\| + \|\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp}\|^2.$$

过参数化项$\|\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp}\|^2$保持较小，而子空间内误差$\|\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^{\top} * (\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^{\top} - \boldsymbol{\mathcal{X}}_\star)\|$迅速减小，从而使恢复误差达到最低。

IV. 过拟合阶段： 最终，过参数化项$\|\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp}\|$开始增大，导致总体误差$\|\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^{\top} - \boldsymbol{\mathcal{X}}_\star\|_F$上升，并趋近于谱初始化的误差。

小初始化的力量 通过上述四阶段分析，我们可以看到，小初始化发挥着至关重要的作用。具体而言，小初始化确保信号项迅速增大，同时将过参数化项保持在较小的量级，从而减轻过参数化带来的负面影响。特别是在阶段 III 期间，过参数化项$\|\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp}\|^2$保持较小，而$\|\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^{\top} * (\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^{\top} - \boldsymbol{\mathcal{X}}_\star)\|$则快速收敛。此外，由于引入了$\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}$，我们有

$$\|\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^{\top} * (\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^{\top} - \boldsymbol{\mathcal{X}}_\star)\|_F \le \sqrt{r}\|\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^{\top} * (\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^{\top} - \boldsymbol{\mathcal{X}}_\star)\|,$$

这确保了最终恢复误差与超管秩R无关。

注释 6。 我们假设$\boldsymbol{\mathcal{X}}_\star$是对称的，且可以分解为$\boldsymbol{\mathcal{X}}_\star = \boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^{\top}$，这与先前的研究一致（*Liu* 等，*2024b*；*Karnik* 等人，*2025*）。要推广到一般的非对称情形，即$\boldsymbol{\mathcal{X}}_{asym} \in \mathbb{R}^{m\times n\times k}$分解为$\boldsymbol{\mathcal{L}} * \boldsymbol{\mathcal{R}}^{\top}$，则需要进行若干修改。我们在此简要讨论，更多细节见附录 *I*。首先，需要进行对称化步骤，以构造对称张量$\boldsymbol{\mathcal{X}}_{sym} \in \mathbb{R}^{(m+n)\times(m+n)\times k}$及其对应的对称模型。其次，两个因子张量的演化轨迹相互耦合，因此有必要分析额外的不平衡项；在对称情形下则不存在这一问题。

注释 7.（与 *Ding* 等，*2025* 的比较）当$n_3 = 1$时，我们的框架退化为矩阵情形：*t-*积变为矩阵乘法，管秩变为矩阵秩，且$\mathcal{X} = \mathcal{U} * \mathcal{U}^{\top}$退化为$X = UU^{\top}$。在这一特殊情形下，定理 *2* 得出了与矩阵 *FGD* 中报告的相同定性现象：如文献（*Ding* 等，*2025*）所示，小初始化和提前停止可得到不随过度指定的秩而恶化的误差界。然而，将矩阵情形推广到张量情形并非易事，必须应对张量所特有的若干挑战，详见注释 *8*。

注释 8.（张量特有的挑战）首先，在矩阵情形中，值域和核是互补子空间。这一性质对三阶管状张量不再成立。如果真实张量的 *t-SVD* 中含有不可逆纤维，等价地说，如果傅里叶域中的某些频率切片为零，那么值域和核就会共享公共生成元。因此，在这些不可逆纤维上，将梯度更新经典地分解为“信号项”和“过参数化项”的方法会失效。这就需要引入更精细的张量条件数概念，以分别跟踪可识别和不可识别的分量。其次，对于幂法，管状张量的每个频率切片都表现得如同一次独立的矩阵幂迭代，这是（*gle，2013*）中已知的事实。然而，在用于张量恢复的梯度下降中，测量算子及其伴随算子会耦合所有频率切片之间的信息。因此，任一切片的更新都取决于其他所有切片，这使得无法像幂法那样独立分析各个切片。最后，在矩阵情形中，*Candes & Plan*（*2011*）已确立含噪矩阵感知的极小极大误差。据我们所知，然而，张量情形尚无此类极小极大误差分析。

3.4 通过验证进行早停

尽管定理 2 给出了目前已知最紧的误差界，但显然，$\hat{t}$ 的选择取决于对$\boldsymbol{\mathcal{X}}_\star$的先验知识，而这在实践中往往不可得。如图 1 所示，将 $\hat{t}$ 设得过小或过大都会导致误差增大。一种实用的解决方案是利用验证来确定何时停止算法，这是机器学习中的一种常见技术（Prechelt，1998年；Stone，2018；Ding 等，2025）。具体而言，我们将观测数据 $\{\boldsymbol{\mathcal{A}}_i, y_i\}_{i=1}^m$ 随机划分为大小为 m_{train} 的训练集（$\boldsymbol{y}_{\text{train}}, \mathfrak{M}_{\text{train}}$）和大小为 m_{val} 的验证集（$\boldsymbol{y}_{\text{val}}, \mathfrak{M}_{\text{val}}$）。随后，我们使用训练集进行梯度下降。每次迭代后，我们计算验证损失$e_t = \frac{1}{4}||\boldsymbol{y}_{\text{val}} - \mathfrak{M}_{\text{val}}(\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top)||^2$。最终估计值选为 $\check{t} = \arg\min_t e_t$，并将$\boldsymbol{\mathcal{U}}_{\check{t}} * \boldsymbol{\mathcal{U}}_{\check{t}}^\top$作为恢复的张量输出。完整过程见算法 2、附录 D。

接下来，我们给出一个理论保证，证明当 $\check{t} = \arg\min_{1\le t\le T} e_t$ 时，恢复误差 $|\boldsymbol{\mathcal{U}}_{\check{t}} * \boldsymbol{\mathcal{U}}_{\check{t}}^\top - \boldsymbol{\mathcal{X}}_\star||_F$满足定理 2 中给出的界。

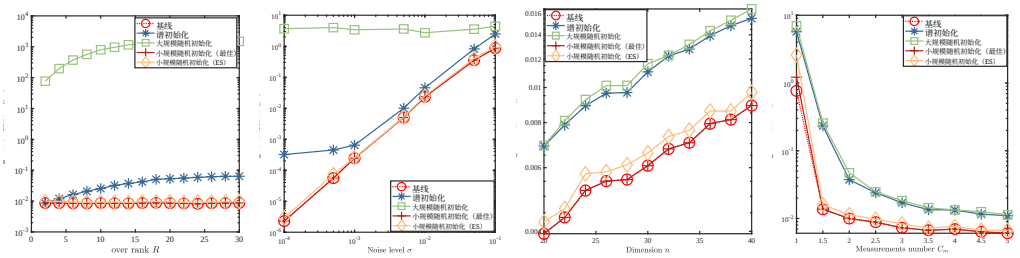
定理 *4.* *假设满足定理 **2** 中相同的条件，但将 **(y, 𝔐)** 替换为 **(y_{train}, $\mathfrak{M}_{train}$)**。此外，假设 $m_{val} \geq C_1{}^{m^2}\boldsymbol{train\ log}_T(rnk\kappa^4)^2$，，并且 T为定理 **2** 中的最大 $\hat{t}$。假设噪声向量 $\boldsymbol{s}$ 的每个分量均独立采样自高斯分布 $\mathcal{N}(0, \sigma^2)$。定义 $\check{t} = \boldsymbol{arg\ min}_{1\le t\le T} e_t$. 那么，以至少 $1 - 2T\exp\left(-\frac{C_2(nkr\kappa^4)}{m_{\text{train}}^2}{}^{2m_{\text{val}}}\right)$的概率，$||\boldsymbol{U}_{\check{t}} * \boldsymbol{U}_{\check{t}}^\top - \boldsymbol{X}_\star||_F^2 \le C\frac{nkr\sigma^2\kappa^4}{m_{train}}$.*

备注 9。 在定理 *2* 中，我们要求 $m_{train} \gtrsim nkr^2\kappa^8$。将其代入条件 $m_{val} \geq C_1{}^{m^2}$ 训练 对数 T $(rnk\kappa^4)^2$，，可得 $m_{val} \gtrsim r^2\kappa^8$ 对数 T。与 m_{train} 相比，这一数值相对较小，因此在实践中是可行的。实验还表明，只需相对较小的 m_{val}，即可获得接近精确管秩设置下的误差。

4 实验

我们进行了一系列实验，结果表明，在过秩设置下，采用小型初始化并结合验证，可以获得与精确参数化设置相当的恢复误差。与采用大型随机初始化或谱初始化的 FGD（Liu 等，2024b）相比，我们的方法实现了最低的恢复误差，凸显了小型初始化的独特有效性。附录 J 中还给出了其他仿真研究和真实数据实验。

实验设置 我们首先生成一个管秩为 r 的真实张量 $\mathcal{X}_\star \in \mathbb{R}^{n\times n\times k}$，具体方法是设置 $\mathcal{X}_\star = \mathcal{X} * \mathcal{X}^\top$，其中 $\mathcal{X} \in \mathbb{R}^{n\times r\times k}$ 的元素独立地取自高斯分布 $\mathcal{N}(0, 1)$。接下来，我们通过设置 $\mathcal{X}_\star \leftarrow \mathcal{X}_\star/||\mathcal{X}_\star||_F$ 对张量进行归一化。我们通过从高斯分布$\mathcal{N}(0, 1)$中独立抽取每个元素来生成测量算子 $\mathfrak{M}$。噪声向量 $\boldsymbol{s}$ 的元素独立地取自 $\mathcal{N}(0, \sigma^2)$。最后，通过测量模型 $\boldsymbol{y} = \mathfrak{M}(\mathcal{X}_\star) + \boldsymbol{s}$ 获得观测值。在所有实验中，我们设定$m = 2C_m nrk$，以及 $n = 30$, $k = 3$, $r = 3$, $m_{\text{val}} = 0.05m$。对于采用小初始化的 FGD，我们将初始化尺度设为 $\alpha = 10^{-10}$。对于采用谱初始化的 FGD，我们遵循原论文中的相同初始化过程。对于采用大初始化的 FGD，我们将初始化尺度设为 $\alpha = 10$，并将其步长设为 $\eta = 0.001$ 以防止发散。我们使用 FGD，并



(a) (b) (c) (d)

图 2：在不同的 r、 σ、 n和 m条件下的性能比较。子图 (a) 展示了所有方法在不同超秩值 R 下的恢复误差，参数设为 $m = 10nrk$、$n = 30$、 $\sigma = 10^{-3}$、 $\eta = 0.1$ 和 $T = 5000$。子图 (b) 展示了在不同噪声水平 σ 下的误差，其中 $m = 10nrk$、 $n = 30$、 $R = 3r$、 $\eta = 0.1$ 和 $T = 5000$。子图 (c) 展示了问题维度 n 变化时的误差，其中 $m = 10nrk$、 $R = 3r$、 $\eta = 0.1$、 $T = 20000$ 和$\sigma = 10^{-3}$。子图 (d) 展示了不同测量数量 C_m 下的性能，其中 $m = 2C_mnrk$、 $n = 30$、 $R = 3r$、 $\eta = 0.01$、 $T = 20000$ 和 $\sigma = 10^{-3}$。

将精确秩作为基线方法，其中“Small random ini (best)”表示采用小规模随机初始化的 FGD 所获得的最小误差，“Small random ini (ES)”表示采用小规模随机初始化并使用验证和提前停止的 FGD 所获得的误差。我们使用相对平方误差（RSE） $\frac{\|\boldsymbol{U}_t\boldsymbol{U}_t^\top - \boldsymbol{X}_\star\|_F^2}{\|\boldsymbol{X}_\star\|_F^2}$来评估不同方法的性能，以及所有实验

P实验重复 P进行 20 次。

不同初始化方法的比较从图 2 中，我们得出以下观察结果：

1. 在所有四种设置中，采用小初始化都能达到与基线方法相同的最小误差，这证明了其有效性。此外，将小初始化与基于验证集的早停相结合，无需任何关于目标张量的先验知识，也能获得非常接近基线方法的误差。这支持了定理中的结论。

2. 对于谱初始化和大型随机初始化，随着高估秩的增大，恢复误差也随之增大，并且始终高于小初始化的误差。大型随机初始化的误差尤其高，这是由于其收敛缓慢。然而，在图 2 (c) 和 (d) 所示的实验中，当迭代次数足够多时，其误差与谱初始化相当。

3. 如图 2 (d) 所示，小初始化还能显著降低采样复杂度。即使在 $m = 3nrk$ 时，它仍能实现较低的误差，明显优于其他初始化方法。

验证与提前停止方法我们验证了验证和提前停止策略的有效性。如图 3 (a) 所示，当验证损失达到最低点时，相对恢复误差最小，这表明使用验证损失作为停止准则是可靠的。图 3 (b) 显示，验证样本过多时，由于训练数据不足，恢复误差会高于可达到的最小误差。相反，如果用于验证的样本过少（不足 5%），基于验证的方法可能变得不可靠，导致恢复误差增大。因此，将总样本的 5%–10% 用于验证是合理的选择。

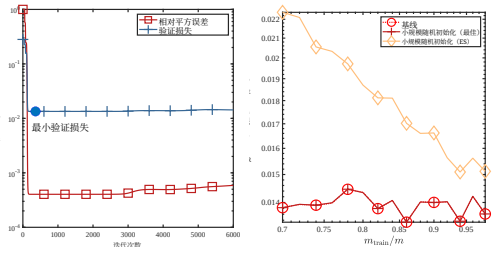


图 3：使用 m =$10nrk$, $R = 3r$, $n = 30$, $\sigma = 10^{-3}$, $\eta = 0.1$对算法进行验证。(a) 验证损失与 RSE 的关系，蓝点标记最小值。(b) 在不同m_{train}下，基于验证的方法的误差与基线和小初始化方法的最小误差进行比较。

张量补全的真实数据实验

我们针对低管秩张量补全问题开展真实数据实验。我们考虑伯努利观测模型下的低管秩张量补全问题。设目标张量为 $\mathcal{X}_\star \in \mathbb{R}^{n_1\times n_2\times n_3}$ ，其管秩 r未知，其中每个元素以概率 p被独立观测。将观测索引集合记为$\boldsymbol{\Omega} \subseteq [n_1] \times [n_2] \times [n_3]$，并且

表 2：在不同采样率和噪声水平下，各种方法的平均峰值信噪比（PSNR）和平均相对误差（RE）对比。“FGD-ES”表示采用提前停止的 FGD，“FGD-best”表示 FGD 在所有迭代次数中达到的最小误差。为简洁起见，我们将 GTNN-HOP$_{0.3}$ 称为 GTNN。

方法	$p=0.2$				$p=0.3$			
	$\sigma=0.07$		$\sigma=0.1$		$\sigma=0.07$		$\sigma=0.1$	
	峰值信噪比 ↑	RE ↓	峰值信噪比 ↑	RE ↓	峰值信噪比 ↑	RE ↓	峰值信噪比 ↑	相对误差 ↓
TCTF	16.5892	0.3175	16.5484	0.3191	20.6744	0.2008	20.6335	0.2024
TNN	21.2692	0.1851	19.7672	0.2188	22.0592	0.1681	20.1682	0.2082
TC-RE	20.9288	0.1921	19.5480	0.2242	21.5387	0.1782	19.8376	0.2161
UTF	16.3227	0.3243	14.8770	0.3802	19.2245	0.2355	17.8283	0.2734
GTNN	22.1092	0.1675	20.3132	0.2051	23.1542	0.1481	21.1111	0.1867
FGD-ES	22.5912	0.1616	21.7977	0.1765	23.6579	0.1426	22.7157	0.1585
FGD-best	**22.7438**	**0.1587**	**21.9268**	**0.1739**	**23.8422**	**0.1395**	**22.8550**	**0.1559**

将观测算子定义为 $\mathfrak{P}_\Omega(\mathrm{A}) = \Omega \odot \mathrm{A}$，其中 $\odot$ 表示Hadamard乘积。目标是从部分含噪观测 $\mathfrak{P}_\Omega(\mathrm{X}_\star + \mathrm{S}_n)$中准确恢复低管秩张量$\mathrm{X}_\star$，其中$\mathrm{S}_n$ 在本文中被假定为高斯噪声，其元素独立同分布地采样自高斯分布 $\mathcal{N}(0,\sigma^2)$。在t-乘积框架下，我们采用Burer–Monteiro分解$\mathrm{L} * \mathrm{R}^\top$，其中$\mathrm{L} \in \mathbb{R}^{n_1\times R\times n_3}$，$\mathrm{R} \in \mathbb{R}^{n_2\times R\times n_3}$。该恢复问题通过最小化如下因子化损失函数来表述：$f(\mathrm{L},\mathrm{R}) = \frac{1}{2p}||\mathfrak{P}_\Omega(\mathrm{L} * \mathrm{R}^\top - \mathrm{X}_\star - \mathrm{S}_n)||_F^2$，该函数可通过关于(L, R)的梯度下降进行优化。

随后，我们在伯克利分割数据集（Martin 等人，2001）上开展彩色图像补全实验。我们随机选取 50 张尺寸为 $481 \times 321 \times 3$ 的彩色图像。我们比较了三类方法：一种凸方法——管状张量核范数最小化（TNN）（Lu等，2018年）；非凸方法——UTF（Du 等人，2021）和 GTNN-HOP（Wang等人，2024）；以及基于秩估计的方法——TCTF（Zhou等，2017年）和 TC-RE（Shi等人，2021）。我们采用峰值信噪比和相对误差作为评估指标；有关更详细的实验设置，请参阅附录J.2。表 2 所示的结果表明，采用小初始化的 FGD 显著优于所有其他方法，而采用提前停止的 FGD 表现稍差，但仍可接受。因此，尽管张量补全问题不要求满足 t-RIP 假设，采用小初始化的 FGD 仍能实现最低的重建误差。在

此外，我们评估了该算法对不同管秩的敏感性。如图4所示，选择不同的 R 对恢复性能的影响很小。因此，当真实秩未知时，为恢复选择一个稍大的秩是一种切实有效的策略。此外，附录J.2中展示了视频补全实验。

5 结论

我们提出一种新颖的方法，即采用小初始化的因子化梯度下降，来解决含噪低管秩张量恢复问题。我们证明，即使管秩被高估，恢复误差仍仅取决于精确管秩 r，而与高估的管秩 R无关。这显著改进了 (Liu等人, 2024b) 中的误差界。据我们所知，这是首个不依赖高估管秩、且近乎极小极大最优的含噪低管秩张量恢复误差界。此外，我们证明，无需任何关于底层张量的先验知识，通过验证和早停程序即可达到该误差界。我们还进行了数值实验，以支持我们的理论发现。

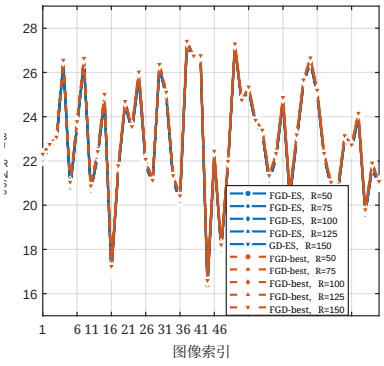


图4：验证 FGD 对不同管秩的敏感性。

致谢

本研究部分获国家自然科学基金委员会项目资助，项目编号为 T2596040、T2596045 和 U23A20343；中国科学院基础研究青年科学家项目资助，项目编号为 YSBR-041；辽宁省“揭榜挂帅”科技计划项目资助，项目编号为 2023JH1/10400045；沈阳自动化研究所基础研究项目资助，项目编号为 2024JC3K01；辽宁省自然科学基金资助，项目编号为 2025-BS-0193；以及 DICP 与 SIA 联合创新基金资助，项目编号为 UN202401。

参考文献

循环矩阵代数中的幂方法与Arnoldi方法。《数值线性代数及其应用》，20(5):809–831，2013。

Richard G Baraniuk、Thomas Goldstein、Aswin C Sankaranarayanan、Christoph Studer、Ashok Veeraraghavan与Michael B Wakin。压缩视频感知：算法、架构与应用。《*IEEE*信号处理杂志》，34(1):52–66，2017。

Samuel Burer与Renato DC Monteiro。一种通过低秩分解求解半定规划的非线性规划算法。《数学规划》，95(2):329–357，2003。

Shuting Cai、Qilun Luo、Ming Yang、Wen Li 和 Mingqing Xiao。通过非凸低秩近似进行张量鲁棒主成分分析。*Applied Sciences*，9(7):1411，2019。

Emmanuel J Candes 和 Yaniv Plan。从最少数量的含噪随机测量中进行低秩矩阵恢复的紧致预言不等式。*IEEE* 信息论汇刊，57 (4):2342–2359，2011。

WenfeiCao、Yao Wang、JianSun、Deyu Meng、Can Yang、Andrzej Cichocki 和 Zongben Xu。用于从压缩测量中进行背景减除的总变分正则化张量鲁棒主成分分析。*IEEE* 图像处理汇刊，25(9):4075–4090，2016。

J Douglas Carroll 和 Jih-JieChang。通过“Eckart-Young”分解的 n 路泛化分析多维尺度分析中的个体差异。*Psychometrika*，35(3):283–319，1970。

Lijun Ding、Zhen Qin、Liwei Jiang、Jinxin Zhou 和 Zhihui Zhu。用于过参数化矩阵与图像恢复的验证方法。载于简约性与学习会议，第 323–350 页。PMLR，2025。

Shiqiang Du、Qingjiang Xiao、Yuqing Shi、Rita Cucchiara 和 Yide Ma。统一低秩张量补全的张量分解与张量核范数方法。*Neurocomputing*，458：204–218，2021。

Carl Eckart 和Gale Young。低秩矩阵近似。*Psychome-trika*，1(3)：211–218，1936年。

Lanlan Feng、Ce Zhu、Yipeng Liu、Saiprasad Ravishankar 和 Longxiu Huang。用于实现鲁棒张量主成分分析的可学习缩放梯度下降。*arXiv* 预印本 *arXiv:2501.04565*，2025。

Mingcheng Fu、Zhi Zheng、Wen-Qin Wang 和 Hing Cheung So。利用累积量张量重构，通过中心对称十字阵列进行三维近场源定位。*IEEE* 航空航天与电子系统汇刊，第 1–14 页，2025。数字对象标识符：10.1109/TAES.2025.3578542。

Kyle Gilman、Davoud Ataee Tarzanagh 和 Laura Balzano。采用 t-SVD 的在线张量补全与跟踪的格拉斯曼流形优化。*IEEE* 信号处理汇刊，70：2152–2167，2022。

Bo Han、Yuheng Jia、Hui Liu 和 Junhui Hou。用于高光谱图像表示的非规则张量低秩表示。*IEEE* 图像处理汇刊，2025。

Zhi Han、Shaojie Zhang、Zhiyu Liu、Yanmei Wang、Junping Yao 和 Yao Wang。结合侧信息的张量鲁棒主成分分析：模型与应用。*IEEE* 视频技术电路与系统汇刊，33(8):3713–3725，2023年。数字对象标识符：10.1109/TCSVT. 2023.3239376。

Zhi Han、Yanmei Wang、Shaojie Zhang、Huijie Fan、Yandong Tang 和 Yao Wang。基于非局部稳健主成分分析的在线视频稀疏噪声去除。*IEEE* 多媒体汇刊，26:7130–7145，2024。

Ra Harshman。PARAFAC过程的基础：多模态因子分析的模型与条件。加州大学洛杉矶分校语音学工作论文，16(1):84，1970。

Chengxun He、Yang Xu、Zebin Wu、Shangdong Zheng 和 Zhihui Wei。多维视觉数据恢复：揭示变换高阶张量奇异值中的全局差异。*IEEE* 图像处理汇刊，33:6409–6424，2024。数字对象标识符：10.1109/TIP.2024. 3475738。

Yicong He 和 George K Atia。基于张量分解和最大相关熵准则的稳健且可并行化的张量补全。载于 *ICASSP 2023-2023 IEEE* 国际声学、语音与信号处理会议（*ICASSP*），第 1–5 页。IEEE，2023。

Liwei Jiang、Yudong Chen 和 Lijun Ding。无模型过参数化非对称矩阵分解中的算法正则化。*SIAM* 数据科学数学杂志，5 (3):723–744，2023年a。数字对象标识符：10.1137/22M1519833。

Tai-Xiang Jiang、Ting-Zhu Huang、Xi-Le Zhao 和 Liang-Jian Deng。通过最小化管状核范数的部分和恢复多维成像数据。计算与应用数学杂志，372:112680，2020。Wei Jiang、Jun Zhang、Changsheng Zhang、Lijun Wang 和 Heng Qi。通过因子张量范数最小化实现鲁棒低管秩张量补全。模式识别，135:109169，2023年b。

Santhosh Karnik、Anna Veselovska、Mark Iwen 和 Felix Krahmer。通过梯度下降实现管状张量分解的隐式正则化。载于第四十二届国际机器学习会议。PMLR，2025。

Misha E Kilmer 和 Carla D Martin。三阶张量的因子分解策略。线性代数及其应用，435(3):641–658，2011。Misha E Kilmer、Karen Braman、Ning Hao 和 Randy C Hoover。作为矩阵算子的三阶张量：一个用于成像应用的理论与计算框架。*SIAM* 矩阵分析与应用杂志，34(1):148–172，2013。Hao Kong、Xingyu Xie 和 Zhouchen Lin。用于低秩张量恢复的 t-schatten-p 范数。*IEEE* 信号处理精选主题杂志，12(6):1405–1419，2018。Zina Li、Yao Wang、Qian Zhao、Shijun Zhang 和 Deyu Meng。用于压缩背景减除的基于张量的在线稳健主成分分析模型。*IEEE* 神经网络与学习系统汇刊，2022。Chunyan Liu、Sui Li、Dianlin Hu、Jianjun Wang、Wenjin Qin、Chen Liu 和 Peng Zhang。用于能谱CT图像重建的融合联合低秩性与平滑性的非局部张量分解。*IEEE* 计算成像汇刊，10:613–627，2024a。Xiao-Yang Liu、Shuchin Aeron、Vaneet Aggarwal 和 Xiaodong Wang。利用交替最小化进行低管秩张量补全。*IEEE* 信息论汇刊，66(3):1714–1737，2019。Xinling Liu、Jingyao Hou、Jiangjun Peng、Hailin Wang、Deyu Meng 和 Jianjun Wang。融合低秩性与局部平滑性的张量压缩感知。人工智能 *AAAI* 会议论文集，37(7):8879–8887，2023年6月。数字对象标识符：10.1609/aaai.v37i7.26067。网址：https://ojs.aaai.org/index.php/AAAI/article/view/26067。

Zhiyu Liu、Zhi Han、Yandong Tang、Xi-Le Zhao 和 Yao Wang。通过因子化梯度下降进行低管秩张量恢复。*IEEE* 信号处理汇刊，2024b。Zhiyu Liu、Zhi Han、Yandong Tang、Jun Fan 和 Yao Wang。通过交替预条件梯度下降高效估计低管秩张量，2025。Canyi Lu、Jiashi Feng、Zhouchen Lin 和 Shuicheng Yan。从高斯测量中精确恢复低管秩张量。载于第*27*届国际人工智能联合会议论文集，第2504–2510页，2018。Yuetian Luo 和 Anru R. Zhang。张量对张量回归：黎曼优化、过参数化、统计计算差距及其相互作用。《统计年鉴》，52(6):2583–2612，2024。数字对象标识符：10.1214/24-AOS2396。网址：https://doi.org/24-AOS2396。D. Martin、C. Fowlkes、D. Tal 和 J. Malik。一个人工分割自然图像数据库及其在评估分割算法和测量生态统计中的应用。载于第*8*届国际计算机视觉会议论文集，第2卷，第416–423页，2001年7月。Yang Mu、Ping Wang、Liangfu Lu、Xuyun Zhang 和 Lianyong Qi。使用张量奇异值分解进行张量补全的加权张量核范数最小化。《模式识别快报》，130:4–11，2020。Ivan V Oseledets。张量列车分解。《*SIAM* 科学计算杂志》，33(5):2295–2317，2011。Jiangjun Peng、Yao Wang、Hongying Zhang、Jianjun Wang 和 Deyu Meng。联合低秩性与局部平滑性及稀疏矩阵的精确分解。*IEEE* 模式分析与机器智能汇刊，45:5766–5781，2022。Lutz Prechelt。提前停止——但何时停止？，第55–69页。Springer Berlin Heidelberg，柏林，海德堡，1998。数字对象标识符：10.1007/3-540-49430-8 3。Wenjin Qin、Hailin Wang、Feng Zhang、Weijun Ma、Jianjun Wang 和 Tingwen Huang。利用随机低秩近似进行非凸鲁棒高阶张量补全。*IEEE* 图像处理汇刊，33:2835–2850，2024。Wenjin Qin、Hailin Wang、Jingyao Hou 和 Jianjun Wang。加速大规模正则化高阶张量恢复。*arXiv* 预印本 *arXiv:2506.09594*，2025。Haiquan Qiu、Yao Wang、Shaojie Tang、Deyu Meng 和 Quanming Yao。快速且可证明的非凸张量稳健主成分分析。载于国际机器学习会议，第18211–18249页。PMLR，2022。G Rajesh 和 Ashvini Chaturvedi。利用鲁棒张量主成分分析，在异构环境无线传感器网络中进行数据重建。*IEEE* 网络信号与信息处理汇刊，7:539–550，2021。Qiquan Shi、Yiu-Ming Cheung 和 Jian Lou。鲁棒张量奇异值分解与恢复及秩估计。*IEEE* 控制论汇刊，52(10):10667–10682，2021。Mahdi Soltanolkotabi、Dominik St¨oger 和 Changzhi Xie。隐式平衡与正则化：过参数化非对称矩阵感知的泛化与收敛保证。*IEEE* 信息论汇刊，71(4):2991–3037，2025。Dominik St¨oger 和 Mahdi Soltanolkotabi。小型随机初始化类似于谱学习：过参数化低秩矩阵重建的优化与泛化保证。神经信息处理系统进展，34:23831–23843，2021。M. Stone。统计预测的交叉验证选择与评估。《皇家统计学会杂志：*B*辑（方法论）》，36(2):111–133，12 2018。国际标准连续出版物号 0035-9246。数字对象标识符：10.1111/j.2517-6161.1974.tb00994.x。

Ledyard R Tucker. 三模因子分析的一些数学注记。*Psychometrika*, 31(3): 279–311, 1966. Hailin Wang, Feng Zhang, Jianjun Wang, Tingwen Huang, Jianwen Huang, and Xinling Liu. 面向低管秩张量恢复的广义非凸方法。*IEEE* 神经网络与学习系统汇刊, 33(8):3305–3319, 2021. Yao Wang, Lin Lin, Qian Zhao, Tianwei Yue, Deyu Meng, and Yee Leung. 通过联合张量 Tucker 分解和加权全变分正则化进行高光谱图像压缩感知。*IEEE* 地球科学与遥感快报, 14(12):2457–2461, 2017. Yao Wang, Deyu Meng, and Ming Yuan. 稀疏恢复：从向量到张量。国家科学评论, 5:756–767, 2018. Zhi-Yong Wang, Hing Cheung So, and Abdelhak M Zoubir. 通过新型诱导稀疏正则项进行低秩张量补全。*IEEE* 信号处理汇刊, 2024. Tong Wu. 通过缩放梯度下降保证非凸低秩张量估计。*arXiv* 预印本 *arXiv:2501.01696*, 2025. Tongle Wu and Jicong Fan. 用于张量补全的平滑张量积。*IEEE* 图像处理汇刊, 33:6483–6496, 2024. 数字对象标识符: 10.1109/TIP.2024.3489272. Tongle Wu, Bin Gao, Jicong Fan, Jize Xue, and W. L. Woo. 基于自适应可学习变换的低秩张量补全。*IEEE* 神经网络与学习系统汇刊, 35(7):8826–8838, 2024. 数字对象标识符: 10.1109/TNNLS.2022.3215974. Tongle Wu, Ying Sun, and Jicong Fan. 从局部测量中进行非凸张量恢复。*AAAI* 人工智能会议论文集, 39(20):21590–21598, 2025年4月。数字对象标识符: 10.1609/aaai.v39i20.35462. 网址 https://ojs.aaai.org/index.php/AAAI/ article/view/35462. Wen-Hao Xu, Xi-Le Zhao, Teng-Yu Ji, Jia-Qing Miao, Tian-Hui Ma, Si Wang, and Ting-Zhu Huang. 基于拉普拉斯函数的低秩张量补全非凸替代函数。信号处理：图像通信, 73:62–69, 2019. Ming Yang, Qilun Luo, Wen Li, and Mingqing Xiao. 通过张量分解和非凸正则化方法补全三维阵列图像数据。*IEEE* 信号处理汇刊, 70:4291–4304, 2022. Feng Zhang, Wendong Wang, Jianwen Huang, Jianjun Wang, and Yao Wang. 基于限制等距性质的低管秩张量恢复性能保证。计算与应用数学杂志, 374:112767, 2020. Feng Zhang, Wendong Wang, Jingyao Hou, Jianjun Wang, and Jianwen Huang. 一大类随机测量系综的张量限制等距性质分析。中国科学：信息科学, 64:1–3, 2021. Zemin Zhang and Shuchin Aeron. 使用 t-SVD 实现精确张量补全。*IEEE* 信号处理汇刊, 65(6):1511–1526, 2016. Qibin Zhao, Guoxu Zhou, Shengli Xie, Liqing Zhang, and Andrzej Cichocki. 张量环分解。*arXiv* 预印本 *arXiv:1606.05535*, 2016. Jingjing Zheng, Wenzhe Wang, Xiaoqin Zhang, and Xianta Jiang. 一种对不准确的秩估计具有鲁棒性的新型张量分解方法。*arXiv* 预印本 *arXiv:2305.11458*, 2023年。Pan Zhou, Canyi Lu, Zhouchen Lin, and Chao Zhang. 用于低秩张量补全的张量分解。*IEEE* 图像处理汇刊, 27(3):1152–1163, 2017. Yang Zhou and Yiu-Ming Cheung. 具有多秩确定功能的贝叶斯低管秩鲁棒张量因子分解。*IEEE* 模式分析与机器智能汇刊, 43(1): 62–76, 2019.

Qile Zhu, Shiqian Wu, Shun Fang, Qi Wu, Shoulie Xie, and Sos Agaian. 利用估计多秩和黎曼优化进行快速张量鲁棒主成分分析。*Applied Intelligence*, 55(1):52, 2025年。

Zhihui Zhu、Qiuwei Li、Gongguo Tang 和 Michael BWakin。低秩矩阵优化中的全局最优性。*IEEE* 信号处理汇刊，66(13):3614–3628，2018。

Jiacheng Zhuo、Jeongyeol Kwon、Nhat Ho 和 Constantine Caramanis。关于过参数化矩阵感知的计算与统计复杂度。机器学习研究期刊，25(169):1–47，2024。

目录

附录结构

附录的组织如下：

- B 节说明了大语言模型的使用情况。
- C 节提供了可复现性声明。
- D 节介绍了支持主要理论结果的补充预备知识。
- E 节给出了定理 2 和推论 1 的详细证明。
- F 节给出了定理 3 的详细证明。
- G 节给出了定理 4 的详细证明。
- H 节介绍了若干技术引理及其证明。
- I 节讨论了向非对称情形的扩展。
- J 节报告了不同噪声分布下的其他模拟结果，以及真实数据实验。

B 大型语言模型的使用

我们仅使用 GPT-5 对本手稿进行语言润色和语法修订。该模型未参与研究构想、算法开发、实验开展或结果分析。作者对本文所呈现的技术内容、理论贡献和实验结果承担全部责任。

C 可复现性声明

本文的所有理论结果均由附录中提供的详细证明充分支持。此外，实验所用代码已包含在补充材料中，以确保能够复现本文报告的所有结果。

D 补充预备知识

对于两个正标量 x， $y, x \lesssim y$ （或 $x \gtrsim y$）表示存在一个普适常数 $z > 0$，使得 $x \leq zy$ （或 $x \geq zy$）成立； $x \asymp y$ 表示存在两个普适常数 z_1， $z_2 > 0$，使得 $z_1 x \leq y \leq z_2 x$成立。

定义2 （对称半正定张量）。 若三阶张量$\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n\times n\times k}$ 满足以下条件，则称其为对称半正定张量：

$$\boldsymbol{\mathcal{A}}^{\top} = \boldsymbol{\mathcal{A}}, \textit{and } \overline{\boldsymbol{\mathcal{A}}}^{(i)} \textit{ is positive semi-definite}.$$

定义3 （分块对角矩阵）。 对于任意张量$\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{m\times n\times k}$，我们将 $\bar{\boldsymbol{Y}} \in \mathbb{C}^{mk\times nk}$ 记为一个分块对角矩阵，其对角线上的第 i 个块为 i 个正面切片 $\bar{\boldsymbol{Y}}_{(i)}$ （属于 $\bar{\boldsymbol{\mathcal{Y}}}$），即：

$$\bar{\boldsymbol{Y}} = \texttt{bdiag}(\bar{\boldsymbol{\mathcal{Y}}}) = \begin{bmatrix} \bar{\boldsymbol{Y}}^{(1)} & & & \\ & \bar{\boldsymbol{Y}}^{(2)} & & \\ & & \ddots & \\ & & & \bar{\boldsymbol{Y}}^{(n_3)} \end{bmatrix}.$$

定义 4 （分块循环矩阵（Kilmer & Martin, 2011））。 对于一个三阶张量$\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1\times n_2\times n_3}$，我们将 $\texttt{bcirc}(\boldsymbol{\mathcal{A}}) \in \mathbb{R}^{n_1 n_3\times n_2 n_3}$ 称为其分块循环矩阵，即，

$$\texttt{bcirc}(\boldsymbol{\mathcal{A}}) = \begin{bmatrix} \boldsymbol{A}^{(1)} & \boldsymbol{A}^{(n_3)} & \cdots & \boldsymbol{A}^{(2)} \\ \boldsymbol{A}^{(2)} & \boldsymbol{A}^{(1)} & \cdots & \boldsymbol{A}^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{A}^{(n_3)} & \boldsymbol{A}^{(n_3-1)} & \cdots & \boldsymbol{A}^{(1)} \end{bmatrix}.$$

定义 5（折叠与展开操作（Kilmer & Martin, 2011））。对于一个三阶张量$\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$，我们有

$$\begin{aligned}&\texttt{unfold}(\mathcal{A}) = [A^{(1)}; A^{(2)}; \cdots; A^{(n_3)}]\\&\texttt{fold}(\texttt{unfold}(\mathcal{A})) = \mathcal{A}.\end{aligned}$$

定义 6（t-积（Kilmer & Martin, 2011））。对于$\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$和$\mathcal{B} \in \mathbb{R}^{n_2 \times q \times n_3}$，$\mathcal{A}$与$\mathcal{B}$的 *t*-积为$\mathcal{C} \in \mathbb{R}^{n_1 \times q \times n_3}$，即：

$$\mathcal{C} = \mathcal{A} * \mathcal{B} = \texttt{fold}(\texttt{bcirc}(\mathcal{A}) \cdot \texttt{unfold}(\mathcal{B})).$$

t-积也可以通过算法 *1* 计算。

定义 7（单位张量（Kilmer & Martin, 2011））。单位张量记作$\mathcal{I} \in \mathbb{R}^{n \times n \times n_3}$，其定义为：第一个正面切片对应于 $n \times n$ 单位矩阵，而其余正面切片均由零构成。数学表达式如下：

$$\mathcal{I}^{(1)} = \boldsymbol{I}_{n \times n}, \quad \mathcal{I}^{(i)} = 0, i = 2, 3, \ldots, n_3.$$

定义 8（正交张量（Kilmer & Martin, 2011））。若张量$\mathcal{Q} \in \mathbb{R}^{n \times n \times n_3}$满足以下条件，则称其为正交张量：

$$\mathcal{Q}^\top * \mathcal{Q} = \mathcal{Q} * \mathcal{Q}^\top = \mathcal{I}.$$

定义 9（F-对角张量（Kilmer & Martin, 2011））。若张量的每个前切片都是对角矩阵，则称该张量为 *F*-对角张量。

定理 5 (t-SVD（Kilmer & Martin, 2011；Lu等，2018年))。设$\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$，则它可以分解为

$$\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top,$$

其中$\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$、$\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$为正交张量，且$\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$为 *F*-对角张量。

定义 10（管秩（Kilmer & Martin, 2011））。对于$\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$，其管秩 $rank_t(\mathcal{A})$ 定义为$\mathcal{S}$中非零对角管的数量，其中$\mathcal{S}$是由$\mathcal{A}$的 *t-SVD* 得到的 *F*-对角张量。也就是说

$$\text{rank}_t(\mathcal{A}) := \#\{i : S(i,i,:) \neq 0\}.$$

算法 1张量-张量积

输入： $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$，$\mathcal{Z} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$。**输出：**

$\mathcal{X} = \mathcal{Y} * \mathcal{Z} \in \mathbb{R}^{n_1 \times n_4 \times n_3}$。

1：计算 $\bar{\mathcal{Y}} = \text{fft}(\mathcal{Y}, [], 3)$ 和 $\bar{\mathcal{Z}} = \text{fft}(\mathcal{Z}, [], 3)$

2：通过以下方式计算$\mathcal{C}$的每个正面切片

$$\bar{\mathbf{X}}^{(i)} = \begin{cases} \bar{\mathbf{Y}}^{(i)}\, \bar{\mathbf{Z}}^{(i)}, & i = 1, \ldots, \left\lceil \dfrac{n_3+1}{2} \right\rceil, \\ \text{conj}\left(\bar{\mathbf{X}}^{(n_3-i+2)}\right), & i = \left\lceil \dfrac{n_3+1}{2} \right\rceil + 1, \ldots, n_3. \end{cases}$$

3：计算 $\mathcal{X} = \text{ifft}(\bar{\mathcal{X}}, [], 3)$。

张量$\mathcal{Y} \in \mathbb{R}^{n \times r \times k}$的 t-SVD 表示为$\mathcal{Y} = \mathcal{V}_\mathcal{Y} * \mathcal{S}_\mathcal{Y} * \mathcal{W}_\mathcal{Y}^\top$。此外，我们将$\mathcal{V}_\mathcal{Y}$定义为$\mathcal{Y}$的张量列子空间，将$\mathcal{V}_{\mathcal{Y}^\perp}$定义为其正交补，即$\mathcal{V}_\mathcal{Y}^\top * \mathcal{V}_{\mathcal{Y}^\perp} = 0$。

基于 t-RIP 条件，我们引入以下两个定义以便于分析。

定义 11。*(S2S-t-RIP)* 若线性映射 $\mathfrak{M} : \mathbb{R}^{n \times n \times k} \to \mathbb{R}^m$ 满足谱到谱(r, δ) 张量限制等距性质（*t-RIP*） *[(r, δ) S2S-t-RIP]*，则对于所有管秩为 $\leq r$ 的张量$\mathcal{Y} \in \mathbb{R}^{n \times n \times k}$，

$$\left\| \left(\mathfrak{I} - \frac{\mathfrak{M}^* \mathfrak{M}}{m} \right) (\mathcal{Y}) \right\| \leq \delta ||\mathcal{Y}||.$$

定义 12。*(S2N-t-RIP)* 若线性映射 $\mathfrak{M}: \mathbb{R}^{n\times n\times k} \to \mathbb{R}^m$ 满足谱到核范数 δ 张量限制等距性质（*t-RIP*） *[δ-S2N-t-RIP]*，则对于所有管秩为 $\leq r$的张量$\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{n\times n\times k}$,

$$\left\| \left(\mathfrak{I} - \frac{\mathfrak{M}^*\mathfrak{M}}{m} \right) (\boldsymbol{\mathcal{Y}}) \right\| \leq \delta ||\boldsymbol{\mathcal{Y}}||_*.$$

接下来，我们给出第 3.4 节所述算法 2 的详细伪代码。

算法 2使用 FGD 和提前停止求解 (1)

输入： 训练数据 ($\boldsymbol{y}_{\text{train}}, \mathfrak{M}_{\text{train}}$)，验证数据 ($\boldsymbol{y}_{\text{val}}, \mathfrak{M}_{\text{val}}$)，初始化尺度 α，步长 η，估计的管秩 R，迭代次数 T

初始化：初始化 $\boldsymbol{\mathcal{U}}_0$,，其中 $\boldsymbol{\mathcal{U}}_0$的每个元素均独立同分布地取自$\mathcal{N}(0, \frac{\alpha^2}{R})$。

对 $t = 0$ 至 $T-1$ **执行**$\boldsymbol{\mathcal{U}}_{t+1} = \boldsymbol{\mathcal{U}}_t - \frac{\eta}{m}\mathfrak{M}^*_{\text{train}}(\mathfrak{M}_{\text{train}}$
$(\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top) - \mathbf{y}_{\text{train}}) * \overline{\boldsymbol{\mathcal{U}}}_t$ 验证损失：
$e_t = \frac{1}{2m}||\boldsymbol{y}_{\text{val}} - \mathfrak{M}_{\text{val}}(\boldsymbol{\mathcal{U}}_t * \overline{\boldsymbol{\mathcal{U}}}_t^\top)||^2$ **结束循环** **输出：**
$\boldsymbol{\mathcal{U}}_{\check{t}}$ 其中 $\check{t} = \arg\min_{1\leq t\leq T} e_t$。

E 定理 2 的证明

在本节中，为便于表述，我们将额外的$\frac{1}{\sqrt{m}}$ 因子吸收到 $\mathfrak{M}$ 中，即$\boldsymbol{\mathcal{A}}_i \leftarrow \boldsymbol{\mathcal{A}}_i/\sqrt{m}$。因此，我们有$(1-\delta)||\boldsymbol{\mathcal{Y}}||_F^2 \leq ||\mathfrak{M}(\boldsymbol{\mathcal{Y}})||^2 \leq (1+\delta)||\boldsymbol{\mathcal{Y}}||_F^2$.

E.1 四个阶段的分析

将$\boldsymbol{\mathcal{X}}$的张量列子空间定义为$\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}} \in \mathbb{R}^{n\times r\times k}$。考虑张量$\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * \boldsymbol{\mathcal{U}}_t$ 及其对应的t-SVD $\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * \boldsymbol{\mathcal{U}}_t = \boldsymbol{\mathcal{V}}_t * \boldsymbol{\mathcal{S}}_t * \boldsymbol{\mathcal{W}}_t^\top$，其中$\boldsymbol{\mathcal{W}}_t \in \mathbb{R}^{r\times R\times k}$. 我们将$\boldsymbol{\mathcal{W}}_{t,\perp}$ 记为一个张量，其张量列子空间与 $\boldsymbol{\mathcal{W}}_t$. 的列子空间正交。然后我们可以将$\boldsymbol{\mathcal{U}}_t$ 分解为“信号项”和“过参数化项”：

$$\boldsymbol{\mathcal{U}}_t = \underbrace{\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t * \boldsymbol{\mathcal{W}}_t^\top}_{\text{signal term}} + \underbrace{\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp} * \boldsymbol{\mathcal{W}}_{t,\perp}^\top}_{\text{over-parameterization term}}. \tag{3}$$

通过这一分解，我们可以分别分析信号项和过参数化项。具体而言，我们考虑以下三个量来研究FGD的收敛行为：

- $\sigma_{\min}(\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t)$：信号项的大小；
- $\|\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp}\|$：过参数化项的大小；
- $\|\boldsymbol{\mathcal{V}}^\top \boldsymbol{\mathcal{X}}^\perp * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t}\|$：信号列空间与真实值列空间之间的对齐程度。

利用这三个指标和恢复误差 $||\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top - \boldsymbol{\mathcal{X}}_\star||_F$，我们将 FGD 轨迹划分为四个阶段，并逐一进行分析。

E.1.1 阶段 I：对齐阶段

在第一阶段，引理 1 指出，若初始化尺度足够小，并且在适当的 t-RIP 条件、步长约束以及噪声谱范数的上界下，信号项几乎与真实张量 $\boldsymbol{\mathcal{X}}_\star$ 的列空间对齐。在此阶段，信号项与过参数化项的幅度都仍然很小，但前者显著大于后者。

引理 1. 固定一个足够小的常数 $c > 0$。设 $\boldsymbol{\mathcal{U}} \in \mathbb{R}^{n\times R\times k}$ 为元素服从独立同分布 $\mathcal{N}(0, \frac{\alpha^2}{R})$ 的随机管状张量，并令 $\epsilon \in (0, 1)$。假设 $\mathfrak{M}: \mathbb{R}^{n\times n\times k} \to \mathbb{R}^m$ 对某个常数 $\delta_1 > 0$ 满足 δ_1-*S2R-t-RIP*。此外，假设

$$\boldsymbol{\mathcal{M}} := \mathfrak{M}^*\mathfrak{M}(\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top) + \boldsymbol{\mathcal{E}} = \boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top + \boldsymbol{\mathcal{E}}_{\boldsymbol{\mathcal{X}}}$$

对于每个 $1 \le j \le k$，其中 $\delta \le c_1\kappa^{-2}$，并且 $\|\boldsymbol{\mathcal{E}}\| \le c_1\kappa^{-2}\sigma^2_{\min}(\boldsymbol{\mathcal{X}})$。令 $\boldsymbol{\mathcal{U}}_0 = \boldsymbol{\mathcal{U}}$，其中

$$\alpha^2 \lesssim \begin{cases} \dfrac{\epsilon(R\wedge n)\|\boldsymbol{\mathcal{X}}\|^2}{k^2n^{3/2}\kappa^2}\left(\dfrac{2\kappa^2kn^{3/2}}{c_3(R\wedge n)^{3/2}\epsilon}\right)^{-15\kappa^2} & \text{if } R \ge 3r \\ \dfrac{\epsilon\|\boldsymbol{\mathcal{X}}\|^2}{k^2n^{3/2}\kappa^2}\left(\dfrac{2\kappa^2kn^{3/2}}{c_3r^{1/2}\epsilon}\right)^{-15\kappa^2} & \text{if } R < 3r \end{cases}$$

假设步长满足 $\eta \le c_2\kappa^{-2}\|\boldsymbol{\mathcal{X}}\|^{-2}$。那么，以至少 $1-p$ 的概率，其中

$$p = \begin{cases} k(\tilde{C}\epsilon)^{R-2r+1} + ke^{-\tilde{c}R} & \text{if } R \ge 2r \\ k\epsilon^2 + ke^{-\tilde{c}R} & \text{if } R < 2r \end{cases}$$

以下陈述成立。经过

$$t_* \lesssim \begin{cases} \dfrac{1}{\eta\min_{1\le j\le k}\sigma_r(\bar{X}^{(j)})^2}\ln\left(\dfrac{2\kappa^2\sqrt{n}}{c_3\epsilon\sqrt{(R\wedge n)}}\right) & \text{if } R \ge 3r \\ \dfrac{1}{\eta\min_{1\le j\le k}\sigma_r(\bar{X}^{(j)})^2}\ln\left(\dfrac{2\kappa^2\sqrt{rn}}{c_3\epsilon}\right) & \text{if } R < 3r \end{cases}$$

次迭代后，有

$$\|\boldsymbol{\mathcal{U}}_{t_*}\| \le 3\|\boldsymbol{\mathcal{X}}\|$$

$$\|\mathcal{V}_{\boldsymbol{\mathcal{X}}^\perp}^\top * \boldsymbol{\mathcal{U}}_{t_*} * \boldsymbol{\mathcal{W}}_*\| \le \epsilon\kappa^{-2}$$

并且对于每个 $1 \le j \le k$，有

$$\sigma_r\left(\overline{\boldsymbol{\mathcal{U}}_{t_*} * \boldsymbol{\mathcal{W}}_{t_*}}^{(j)}\right) \ge \frac{1}{4}\alpha\beta$$

$$\sigma_1\left(\overline{\boldsymbol{\mathcal{U}}_{t_*} * \boldsymbol{\mathcal{W}}_{t_*,\perp}}^{(j)}\right) \le \frac{\kappa^{-2}}{8}\alpha\beta$$

其中

$$\beta \lesssim \begin{cases} \epsilon\sqrt{k}\left(\dfrac{2\kappa^2\sqrt{n}}{c_3\epsilon\sqrt{R\wedge n}}\right)^{10\kappa^2} & \text{if } R \ge 3r \\ \dfrac{\epsilon\sqrt{k}}{r}\left(\dfrac{2\kappa^2\sqrt{rn}}{c_3\epsilon}\right)^{10\kappa^2} & \text{if } R < 3r \end{cases}$$

and

$$\beta \gtrsim \begin{cases} \epsilon\sqrt{k} & \text{if } R \ge 3r \\ \dfrac{\epsilon\sqrt{k}}{r} & \text{if } R < 3r. \end{cases}$$

这里，c_1、c_2、$c_3 > 0$ 是仅取决于 c选择的绝对常数。此外，C、$\tilde{c}$ 是绝对数值常数。

E.1.2 第二阶段：信号放大阶段

在第二阶段中，基于第一阶段的结果，信号项的张量列子空间仍与真实值 $\boldsymbol{\mathcal{X}}_\star$ 的张量列子空间保持良好对齐，即，$\|\boldsymbol{\mathcal{V}}^\top \boldsymbol{\mathcal{X}}^\perp * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{U}}_t} * \boldsymbol{\mathcal{W}}_t\|$仍然很小。与此同时，由 $\sigma_{\min}(\mathcal{V}_{\boldsymbol{\mathcal{X}}}^\top * \boldsymbol{\mathcal{U}}_t)$ 度量的信号项幅值呈指数增长。相比之下，过参数化项 $\|\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp}\|$由于小初始化而保持较小。

引理 2 $c_1\kappa^{-2}\sigma_{\min}^2$ $\mathfrak{M}$ $\mathbb{R}^{n\times n\times k}\to\mathbb{R}^m$ $\eta\le c_1\kappa^{-2}\|\ \|^{-2}$ $2r+1$ δ $c_1>0$ $\|\ \|\le$ $0<\delta\le$。假设步长满足 X，其中某个较小的、$E(X)$，并且：满足 $(\ ,\ )$ t-RIP，常数为 $\frac{c_1}{\kappa^4\sqrt{kr}}$。设 $\gamma\in(0,1\,2)$，并选择迭代次数 t_*，使得 $\sigma_{\min}(U_{t_*}*W_{t*})\ge\gamma$。此外，假设 $\|U_{t_*}*W_{t*},\ _\perp\|\le 2\gamma$、$\|U_{\ t_*}\|\le 3\|X\ \|$、$\gamma\le\frac{c_2\sigma_{\min}(X)}{k^{8/7}\kappa^2(R\wedge n)}$，以及 $\|V^\top X_{\ \perp}*V u_{t_*}*W_{t_*}\|\le c_2\kappa^{-2}$，其中某个较小的 $c_2>0$。设

$$t_1=\min\left\{t\ge t_*:\sigma_{\min}(\mathcal{V}_{\mathcal{X}}^\top*\mathcal{U}_t)\ge\frac{1}{\sqrt{10}}\sigma_{\min}(\overline{\mathcal{X}})\right\},\tag{4}$$

且，则以下结论对所有 $t\in[t_*,t_1]$ 均成立：

$$\sigma_{\min}(\mathcal{V}_{\mathcal{X}}^\top*\mathcal{U}_t)\ge\frac{1}{2}\gamma\left(1+\frac{1}{8}\eta\sigma_{\min}(\mathcal{X})^2\right)^{t-t_*}\tag{5}$$

$$\|\mathcal{U}_t*\mathcal{W}_{t,\perp}\|\le 2\gamma\left(1+80\eta c_2\sigma_{\min}(\mathcal{X})^2\right)^{t-t_*}\tag{6}$$

$$\|\mathcal{U}_t\|\le 3\|\mathcal{X}\|\quad\text{and}\quad\|\mathcal{V}_{\mathcal{X}^\perp}^\top*\mathcal{V}_{\mathcal{U}_t*\mathcal{W}_t}\|\le c_2\kappa^{-2},\tag{7}$$

其中 $t_1-t_*\lesssim\frac{1}{\eta\sigma_{\min}^2}\ln(\frac{\sigma_{\min}}{\gamma})$。

E.1.3 第三阶段：局部细化阶段

当信号项 $\sigma_{\min}(\mathcal{V}_{\mathcal{X}}^\top*\mathcal{U}_t)$ 超过 $\frac{\sigma_{\min}(\mathcal{X})\sqrt{10}}{}$ 时，算法进入第三阶段。在此阶段，恢复误差可分解为

$$\|\mathcal{U}_t*\mathcal{U}_t^\top-\mathcal{X}_\star\|\le 4\|\mathcal{V}_{\mathcal{X}}^\top*(\mathcal{U}_t*\mathcal{U}_t^\top-\mathcal{X}_\star)\|+\|\mathcal{U}_t*\mathcal{W}_{t,\perp}\|^2.$$

由于采用小初始化，过参数化项 $||\mathcal{U}_t*\mathcal{W}_{t,\perp}||^2$ 增长缓慢，而子空间内误差 $||\mathcal{V}_{\mathcal{X}}^\top*(\mathcal{U}_t*\mathcal{U}_t^\top-\mathcal{X}_\star)||$ 迅速减小。此外，由于 $\mathcal{V}_{\mathcal{X}}\in\mathbb{R}^{n\times r\times k}$，我们有

$$\|\mathcal{V}_{\mathcal{X}}^\top*(\mathcal{U}_t*\mathcal{U}_t^\top-\mathcal{X}_\star)\|_F\le\sqrt{r}\|\mathcal{V}_{\mathcal{X}}^\top*(\mathcal{U}_t*\mathcal{U}_t^\top-\mathcal{X}_\star)\|,$$

which解释了为何最终恢复误差仅取决于真实管状秩 r，尽管存在过 - par参数化。

引理 3. 假设引理 *2* 中的假设成立。若 $R>r$，则对于

$$\hat{t}\asymp\frac{1}{\eta\sigma_{\min}(\mathcal{X})^2}\ln\left(\frac{\kappa||\mathcal{X}||}{((R\wedge n)-r)\gamma k}\right)+t_1$$

迭代次数，成立

$$||\mathcal{U}_{\hat{t}}*\mathcal{U}_{\hat{t}}^\top-\mathcal{X}*\mathcal{X}^\top||_F\lesssim\sqrt{r}\kappa^{-3/16}((R\wedge n)-r)^{3/4}k^{3/4}\gamma^{21/16}||\mathcal{X}||^{11/16}+\sqrt{r}\kappa^2||\mathcal{E}||\cdot;\tag{8}$$

若 $R=r$，则对任意 $t\ge t_1$，

$$||\mathcal{U}_t*\mathcal{U}_t^\top-\mathcal{X}*\mathcal{X}^\top||_F\lesssim\sqrt{r}(1-\frac{\eta}{400}\sigma_{\min}^2(\mathcal{X}))^{t-t_1}+\sqrt{r}\kappa^2||\mathcal{E}||.\tag{9}$$

E.1.4 第四阶段：过拟合阶段

The 第四阶段是第三阶段的自然延续。考虑第三阶段中的分解：

$$\|\mathcal{U}_t*\mathcal{U}_t^\top-\mathcal{X}_\star\|\le 4\|\mathcal{V}_{\mathcal{X}}^\top*(\mathcal{U}_t*\mathcal{U}_t^\top-\mathcal{X}_\star)\|+\|\mathcal{U}_t*\mathcal{W}_{t,\perp}\|^2.$$

在第四阶段，过参数化项 $||\mathcal{U}_t*\mathcal{W}_{t,\perp}||^2$ 开始增长，最终主导恢复误差，直至其达到谱初始化的误差水平。

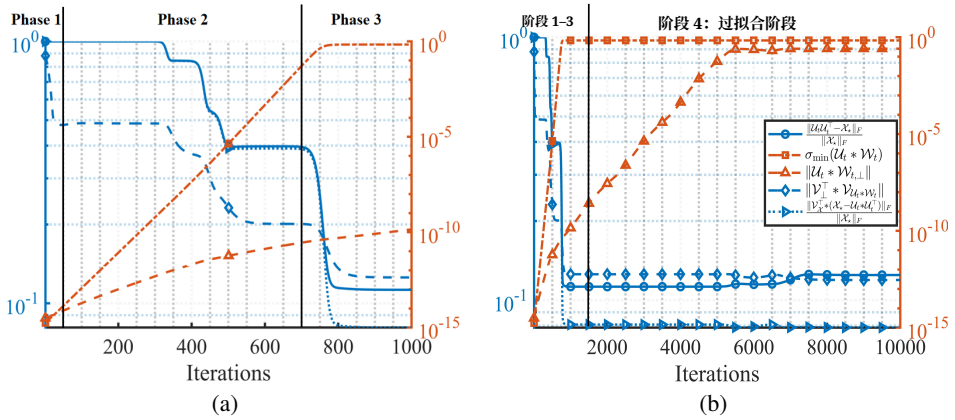


图5：第 3.3 节四阶段收敛分析的验证。左图显示前 1,000 次迭代；右图显示完整的 10,000 次迭代。橙色曲线对应右侧的橙色坐标轴，蓝色曲线对应左侧的蓝色坐标轴。参数设置： $n = 10$, $k = 3, r = 2, R = 10, m = 5knR, \eta = 0.1$，噪声标准差 $\sigma = 0.01$，初始化尺度 $\alpha = 10^{-7}$。

E.2 验证第 3.3 节中的四阶段

我们进行了实验，以验证第 3.3 节所述的四阶段收敛。如图5所示，我们观察到：

- 在阶段 I 中，信号项 $\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t$ 的列空间逐渐与真实值 $\boldsymbol{\mathcal{X}}_\star$的列空间对齐，这可以从 $\|\boldsymbol{\mathcal{V}}^\top \boldsymbol{\mathcal{X}}_\perp * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t}\|$的值不断减小看出。由于采用了小初始化，$\sigma_{\min}(\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t)$ 和 $\|\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp}\|$都保持较小。
- 在第 2 阶段， $\sigma_{\min}(\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t)$ 指数级增长，直至至少达到 $\sigma_{\min}(\boldsymbol{\mathcal{X}})\sqrt{10}$，而$\|\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp}\|$ 则基本保持在初始化时的尺度。
- 在第 3 阶段，过参数化项 $\|\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp}\|^2$ 保持较小，而子空间内误差 $\|\boldsymbol{\mathcal{V}}^\top \boldsymbol{\mathcal{X}} * (\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top - \boldsymbol{\mathcal{X}}_\star)\|$迅速下降，从而使恢复误差达到最低。
- 在第 4 阶段，子空间内误差 $\|\boldsymbol{\mathcal{V}}^\top \boldsymbol{\mathcal{X}} * (\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top - \boldsymbol{\mathcal{X}}_\star)\|$继续下降，但幅度很小；与此同时，过参数化项 $\|\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp}\|$ 迅速增长，并主导总恢复误差，导致整体误差 $\|\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top - \boldsymbol{\mathcal{X}}_\star\|_F$上升。

E.3 定理 2 的证明

由于线性映射 $\mathfrak{M}$ 满足 $(2r+1, \delta)$ t-RIP，根据引理 14， $\mathfrak{M}$ 满足 $(2r, \sqrt{2rk}\delta)$ S2S-t-RIP。因此，

$$
\begin{aligned}
||\boldsymbol{\mathcal{E}}_{\boldsymbol{\mathcal{X}}}|| &= ||(\mathfrak{I} - \mathfrak{M}^*\mathfrak{M})(\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top) + \mathfrak{M}^*(\boldsymbol{s})|| \\
&\le \sqrt{2rk}\delta||\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top|| + ||\mathfrak{M}^*(\boldsymbol{s})|| \\
&\le \sqrt{2rk} \cdot c\kappa^{-4}(rk)^{-1/2} \cdot ||\boldsymbol{\mathcal{X}}||^2 + ||\mathfrak{M}^*(\boldsymbol{s})|| \\
&= \sqrt{2}c\kappa^{-2}\sigma_{\min}(\boldsymbol{\mathcal{X}})^2 + ||\mathfrak{M}^*(\boldsymbol{s})|| \\
&\overset{(a)}{\le} \sqrt{2}c\kappa^{-2}\sigma_{\min}(\boldsymbol{\mathcal{X}})^2 + c_1\kappa^{-2}\sigma_{\min}(\boldsymbol{\mathcal{X}})^2 \\
&\lesssim c\kappa^{-2}\sigma_{\min}(\boldsymbol{\mathcal{X}})^2
\end{aligned}
\tag{10}
$$

其中 (a) 使用假设 $||\boldsymbol{\mathcal{E}}|| \le c_1\kappa^{-2}\sigma_{\min}^2(\boldsymbol{\mathcal{X}})$。

那么，引理 1 以至少 $1 - ke^{-\tilde{c}R} - \max\{k(\tilde{C}\epsilon)^{R-r+1}, k\epsilon^2\}$ 的概率成立。接下来，我们将定理 2 的证明分为三种情况： $R = r$、 $r < R < 3r$ 和 $R \ge 3r$。

E.3.1 情况 1：$R = r$

在这种情况下，根据引理 1 的结果，以下陈述成立：选择

$$\alpha^2 \lesssim \frac{\|\boldsymbol{\mathcal{X}}\|^2}{k^2 n^{3/2}\kappa^2}\left(\frac{2\kappa^2 k n^{3/2}}{\tilde{c}_3 r^{1/2}}\right)^{-15\kappa^2} \text{ and } \tilde{c_3} = c_3\epsilon,$$

然后，经过

$$t_* \lesssim \frac{1}{\eta\sigma_{\min}(\boldsymbol{\mathcal{X}})^2}\ln\left(\frac{2\kappa^2\sqrt{rn}}{\tilde{c}_3}\right)$$

次迭代后，有

$$||\boldsymbol{\mathcal{U}}_{t_*}|| \le 3||\boldsymbol{\mathcal{X}}|| \text{ and } ||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}^\perp} * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{U}}_{t_*} * \boldsymbol{\mathcal{W}}_{t_*}}|| \le c\kappa^{-2} \tag{11}$$

对于每个 $1 \le j \le k$，我们有

$$\sigma_r\left(\overline{\boldsymbol{\mathcal{U}}_{t_*} * \boldsymbol{\mathcal{W}}_{t_*}}^{(j)}\right) \ge \frac{1}{4}\alpha\beta$$

$$\sigma_1\left(\overline{\boldsymbol{\mathcal{U}}_{t_*} * \boldsymbol{\mathcal{W}}_{t_*,\perp}}^{(j)}\right) \le \frac{\kappa^{-2}}{8}\alpha\beta,$$

其中

$$\frac{\sqrt{k}}{r} \lesssim \beta \lesssim \frac{\sqrt{k}}{r}\left(\frac{2\kappa^2\sqrt{nr}}{\tilde{c}_3}\right)^{10\kappa^2}$$

以及 $\tilde{c_3} = \epsilon c_3 = e^{-\tilde{c}/2}c_3$。通过取

$$\alpha \lesssim \frac{\sqrt{r}\sigma_{\min}(\boldsymbol{\mathcal{X}})}{k^{23/14}(R\wedge n)\kappa^2}\left(\frac{2\kappa^2\sqrt{rn}}{\tilde{c}_3}\right)^{-10\kappa^2},$$

我们有 $\gamma = \frac{1}{4}\alpha\beta \lesssim \frac{c_2\sigma_{\min}(\boldsymbol{\mathcal{X}})(R\wedge n)}{k^{8/7}\kappa^2}$。此外，我们有

$$||\boldsymbol{\mathcal{U}}_{t_*} * \boldsymbol{\mathcal{W}}_{t_*,\perp}|| \le \frac{\kappa^{-2}}{8}\alpha\beta \le \frac{\gamma}{2\kappa^2} \le 2\gamma.$$

因此，引理 2 和引理 3 的假设均成立，进而可以利用引理 3 的结果得到：

$$||\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top - \boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top||_F \lesssim \sqrt{r}(1 - \frac{\eta}{400}\sigma_{\min}^2(\boldsymbol{\mathcal{X}}))^{t-t_1} + \sqrt{r}\kappa^2||\boldsymbol{\mathcal{E}}||,$$

对所有 $t \ge t_1$，其中

$$t_1 \lesssim t_* + (t_1 - t_*) \lesssim \frac{1}{\eta\sigma_{\min}(\boldsymbol{\mathcal{X}})^2}\ln\left(\frac{2\kappa^2\sqrt{rn}}{\tilde{c}_3}\right) + \frac{1}{\eta\sigma_{\min}^2(\boldsymbol{\mathcal{X}})}\ln\left(\frac{\sigma_{\min}(\boldsymbol{\mathcal{X}})}{\gamma}\right) \tag{12}$$

$$\overset{(a)}{\lesssim} \frac{1}{\eta\sigma_{\min}(\boldsymbol{\mathcal{X}})^2}\ln\left(\frac{\kappa^2\sqrt{rn}\sigma_{\min}(\boldsymbol{\mathcal{X}})}{\tilde{c}_3\alpha\beta}\right) \tag{13}$$

$$\overset{(b)}{\lesssim} \frac{1}{\eta\sigma_{\min}(\boldsymbol{\mathcal{X}})^2}\ln\left(\frac{\kappa^2 r^{3/2}\sqrt{n}\sigma_{\min}(\boldsymbol{\mathcal{X}})}{\sqrt{k}\alpha}\right), \tag{14}$$

其中，(a) 利用了 $\gamma = \frac{\alpha\beta}{4}$ 这一事实；(b) 利用了 $\beta \gtrsim \frac{\sqrt{k}}{r}$ 这一事实。

然后定义 $\mu := \frac{\eta}{400}\sigma_{\min}^2(\boldsymbol{\mathcal{X}}_\star) \in (0,1)$。利用 $(1-\mu)^s \le e^{-\mu s}$ 这一事实，我们有

$$||\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top - \boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top||_F \lesssim \sqrt{r}e^{-\mu(t-t_1)} + \sqrt{r}\kappa^2||\boldsymbol{\mathcal{E}}||.$$

E.3.2 情况 2：$r < R < 3r$

这种情况的分析方式与前一种情况几乎相同，只是所依据的引理 3 中的结果不同，即当 $R > r$ 时，我们有

$$||\boldsymbol{\mathcal{U}}_{\hat{t}} * \boldsymbol{\mathcal{U}}_{\hat{t}}^\top - \boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top||_F \lesssim \kappa^{-3/16} r^{1/2}((R\wedge n) - r)^{3/4} k^{3/4}\gamma^{21/16}||\boldsymbol{\mathcal{X}}||^{11/16} + \sqrt{r}\kappa^2||\boldsymbol{\mathcal{E}}||,$$

其中

$$\hat{t} \asymp t_1 + \frac{1}{\eta\sigma_{\min}(\boldsymbol{\mathcal{X}})^2}\ln\left(\frac{\kappa||\boldsymbol{\mathcal{X}}||}{((R\wedge n) - r)\gamma k}\right).$$

对 t_1 取情况 1 中的界，我们有

$$\hat{t} \asymp (t_1 - t_*) + t_* + \frac{1}{\eta\sigma_{\min}(\boldsymbol{\mathcal{X}})^2} \ln\left(\frac{\kappa||\boldsymbol{\mathcal{X}}||}{((R\wedge n)-r)\gamma k}\right) \tag{15}$$

$$\asymp \frac{1}{\eta\sigma^2_{\min}(\boldsymbol{\mathcal{X}})} \ln\left(\frac{n^{1/2}r^{5/2}\kappa^2||\boldsymbol{\mathcal{X}}||^2}{k^2[(R\wedge n)-r]\alpha^2}\right). \tag{16}$$

为了得到结果 $||\boldsymbol{\mathcal{U}}_{\hat{t}} * \boldsymbol{\mathcal{U}}_{\hat{t}}^\top - \boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top|| \lesssim \kappa^2||\boldsymbol{\mathcal{E}}||$，我们需要确保 $\kappa^{-3/16} r((R\wedge n) - r)^{3/4} k^{3/4} \gamma^{21/16} ||\boldsymbol{\mathcal{X}}||^{11/16} \leq \kappa^2||\boldsymbol{\mathcal{E}}||$，从而得到

$$\alpha \lesssim \kappa^{35/21}[(R\wedge n)-r]^{-4/7} r k^{-15/14} ||\boldsymbol{\mathcal{X}}||^{-11/21} ||\boldsymbol{\mathcal{E}}||^{16/21} \left(\frac{2\kappa^2\sqrt{rn}}{\tilde{c}_3}\right)^{-10\kappa^2}. \tag{17}$$

利用 $\gamma = \frac{\alpha\beta}{4}$ 和 $\beta \lesssim \frac{\sqrt{k}}{r}\left(\frac{2\kappa^2\sqrt{rn}}{\tilde{c}_3}\right)^{-10\kappa^2}$ 这一事实，为满足假设 $\gamma \lesssim \frac{c_2\sigma_{\min}(\boldsymbol{\mathcal{X}})k^{8/7}\kappa^2(R\wedge n)}{}$

</style>，我们还需要

$$\alpha \lesssim \frac{c_2 r\sigma_{\min}(\boldsymbol{\mathcal{X}})}{\kappa^2(R\wedge n)k^{23/14}} \left(\frac{2\kappa^2\sqrt{rn}}{\tilde{c}_3}\right)^{-10\kappa^2}. \tag{18}$$

结合界（17）和（18），我们得到 α 的界：

$$\alpha \lesssim \min\left\{\frac{r\sigma_{\min}(\boldsymbol{\mathcal{X}})}{k^{23/14}(R\wedge n)\kappa^2}, \frac{r\kappa^{35/21}||\boldsymbol{\mathcal{E}}||^{16/21}}{k^{15/14}[(R\wedge n)-r]^{4/7}||\boldsymbol{\mathcal{X}}||^{11/21}}\right\} \left(\frac{2\kappa^2\sqrt{rn}}{\tilde{c}_3}\right)^{-10\kappa^2} \tag{19}$$

E.3.3 情形 3： $R \geq 3r$

在此情形下，我们同样使用引理 3 的结果。然而，根据引理 1，t_* 和 β 的界有所不同。

具体而言，我们有

$$\begin{aligned} t_* &\lesssim \frac{1}{\eta\sigma_{\min}(\boldsymbol{\mathcal{X}})^2} \ln\left(\frac{2\kappa^2\sqrt{n}}{c_3\epsilon\sqrt{(R\wedge n)}}\right) \\ \epsilon\sqrt{k} \lesssim \beta &\lesssim \epsilon\sqrt{k}\left(\frac{2\kappa^2\sqrt{n}}{c_3\epsilon(R\wedge n)}\right)^{10\kappa^2}, \end{aligned} \tag{20}$$

这意味着

$$\begin{aligned} \hat{t} &\asymp t_* + t_1 - t_* + \hat{t} - t_1 \\ &\asymp \frac{1}{\eta\sigma_{\min}(\boldsymbol{\mathcal{X}})^2} \ln\left(\frac{2\kappa^2\sqrt{n}}{c_3\epsilon\sqrt{(R\wedge n)}}\right) + \frac{1}{\eta\sigma_{\min}(\boldsymbol{\mathcal{X}})^2} \ln\left(\frac{\sigma_{\min}(\boldsymbol{\mathcal{X}})}{\gamma}\right) \\ &\quad + \frac{1}{\eta\sigma_{\min}(\boldsymbol{\mathcal{X}})^2} \ln\left(\frac{\kappa||\boldsymbol{\mathcal{X}}||}{((R\wedge n)-r)\gamma k}\right) \\ &\asymp \frac{1}{\eta\sigma_{\min}(\boldsymbol{\mathcal{X}})^2} \ln\left(\frac{\sqrt{n}\kappa^2||\boldsymbol{\mathcal{X}}||^2}{k^2((R\wedge n)-r)(R\wedge r)\alpha^2}\right). \end{aligned} \tag{21}$$

利用关系 $\gamma = \frac{1}{4}\alpha\beta \lesssim \frac{c_2\sigma_{\min}(\boldsymbol{\mathcal{X}})(R\wedge n),}{\kappa^2}$ 我们得到

$$\alpha \lesssim \frac{\sigma_{\min}(\boldsymbol{\mathcal{X}})}{k^{23/14}(R\wedge n)\kappa^2} \left(\frac{2\kappa^2\sqrt{n}}{\tilde{c}_3\sqrt{(R\wedge n)}}\right)^{-10\kappa^2} \tag{22}$$

此外，根据引理 3 的结果，为了得到 $||\boldsymbol{\mathcal{U}}_{\hat{t}} * \boldsymbol{\mathcal{U}}_{\hat{t}}^\top - \boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top|| \lesssim \kappa^2||\boldsymbol{\mathcal{E}}||$，我们需要对 α 进行如下界定：

$$\begin{aligned} &\alpha \lesssim \kappa^{35/21}[(R\wedge n)-r]^{-4/7} k^{-4/7}\beta^{-1} ||\boldsymbol{\mathcal{X}}||^{-11/21} ||\boldsymbol{\mathcal{E}}||^{16/21} \\ \overset{(a)}{\to}\ &\alpha \lesssim \kappa^{35/21}[(R\wedge n)-r]^{-4/7} k^{-4/7} ||\boldsymbol{\mathcal{X}}||^{-11/21} \frac{1}{\epsilon\sqrt{k}} \left(\frac{2\kappa^2\sqrt{n}}{\tilde{c}_3\sqrt{(R\wedge n)}}\right)^{-10\kappa^2}, \end{aligned} \tag{23}$$

其中，(a) 使用了 β 的上界。结合这两个上界 (22) (23)，我们得到 α 的上界：

$$\alpha \lesssim \min\left\{\frac{\sigma_{\min}(\boldsymbol{\mathcal{X}})}{k^{23/14}(R\wedge n)\kappa^2}, \frac{\kappa^{35/21}||\boldsymbol{\mathcal{E}}||^{16/21}}{k^{15/14}||\boldsymbol{\mathcal{X}}||^{11/21}[(R\wedge n)-r]^{4/7}}\right\}\left(\frac{2\kappa^2\sqrt{n}}{\tilde{c}_3\sqrt{(R\wedge n)}}\right)^{-10\kappa^2}. \quad (24)$$

因此，我们完成了定理 2 的证明。

E.4 推论 1 的证明

推论 1 的证明可直接由定理 2 与 $||\boldsymbol{\mathcal{E}}||$的谱范数界推出。注意

$$||\mathfrak{M}^*(s)|| \overset{(a)}{\lesssim} \sqrt{\frac{nk}{m}}\sigma \overset{(b)}{\leq} c\kappa^{-2}\sigma^2_{\min}(\boldsymbol{\mathcal{X}}), \quad (25)$$

其中 (a) 使用 (Liu 等，2024b) 中的结果；(b) 使用假设 $m \gtrsim nk\kappa^4\sigma^2/\sigma^2_{\min}(\boldsymbol{\mathcal{X}})$。因此，定理 2 中的假设 (3) 得以满足。然后我们可以直接使用定理 2 中的结果得到

$$||\boldsymbol{\mathcal{U}}_{\hat{t}} * \boldsymbol{\mathcal{U}}_{\hat{t}}^\top - \boldsymbol{\mathcal{X}}_\star||_F^2 \lesssim r\kappa^4||\boldsymbol{\mathcal{E}}||^2 \lesssim \frac{nkr\sigma^2\kappa^4}{m}. \quad (26)$$

E.5 引理 1 的证明

引理 1 基于 [(Karnik 等人，2025)，引理 D.8 和引理 D.9], 证明，其中将 $\mathcal{M} := \mathfrak{M}^*\mathfrak{M}(\boldsymbol{\mathcal{X}})$ 替换为 $\mathfrak{M}^*\mathfrak{M}(\boldsymbol{\mathcal{X}}) + \boldsymbol{\mathcal{E}}$，其中 $\boldsymbol{\mathcal{E}} = \mathfrak{M}^*(\boldsymbol{s})$。

Le引理 4. 假设线性映射 $\mathfrak{M}: \mathbb{R}^{n\times n\times k} \to \mathbb{R}^m$ 满足 $(2,\delta_1)$ *t-RIP*，并定义 t^* *as*

$$t^* = \min\left\{j \in \mathbb{N} : ||\tilde{\boldsymbol{\mathcal{U}}}_{j-1} - \boldsymbol{\mathcal{U}}_{j-1}|| \geq ||\tilde{\boldsymbol{\mathcal{U}}}_{j-1}||\right\}.$$

那么，对于所有 $1 \leq t \leq t^*$，我们有

$$||\boldsymbol{\mathcal{E}}_t^{\boldsymbol{\mathcal{U}}}|| = ||\boldsymbol{\mathcal{U}}_t - \tilde{\boldsymbol{\mathcal{U}}}_t|| \leq 8(1+\delta_1\sqrt{k})\sqrt{(R\wedge n)}\frac{\alpha^3}{||\mathcal{M}||}||\boldsymbol{\mathcal{U}}||^3(1+\eta||\mathcal{M}||)^{3t}.$$

证明。 该引理的证明基于 [(Karnik 等人，2025)，引理 D.1]。结合引理 14、引理 15 和引理 16 的结果，我们可以推导出该定理。与 [(Karnik 等人，2025)，引理 D.1], 相比，该引理利用不等式 $|\boldsymbol{\mathcal{U}}_{j-1}||_F \leq \sqrt{(R\wedge n)}||\boldsymbol{\mathcal{U}}_{j-1}||$，降低了对第三维 k 的依赖，从而得到关于 $\|\boldsymbol{\mathcal{E}}_t^{\boldsymbol{\mathcal{U}}}\|$的更紧上界。 □

引理 5。 假设线性映射 $\mathfrak{M}: \mathbb{R}^{n\times n\times k} \to \mathbb{R}^m$ 满足 $(2,\delta_1)$ *t-RIP*。考虑张量$\boldsymbol{\mathcal{M}} := \mathfrak{M}^*\mathfrak{M}(\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top) + \boldsymbol{\mathcal{E}} \in \mathbb{R}^{n\times n\times k}$以及 $\tilde{\boldsymbol{\mathcal{U}}}_t := (\boldsymbol{\mathcal{I}} + \eta\boldsymbol{\mathcal{M}})^t * \boldsymbol{\mathcal{U}}_0$. 设$\boldsymbol{\mathcal{M}} \in \mathbb{C}^{nk\times nk}$是张量 $\mathcal{M}$对应的分块对角矩阵，其主特征向量为 $v_1 \in \mathbb{C}^{nk}$，则有

$$t^* \geq \left\lfloor \frac{\ln\left(\frac{||\boldsymbol{\mathcal{M}}||\cdot||\overline{\boldsymbol{\mathcal{U}}_0}^H v_1||_{l_2}}{8(1+\delta_1\sqrt{k})\sqrt{(R\wedge n)}\alpha^3||\boldsymbol{\mathcal{U}}||^3}\right)}{2\ln(1+\eta||\boldsymbol{\mathcal{M}}||)} \right\rfloor.$$

证明。 此引理的证明可通过将引理 4 的结果纳入 [(Karnik 等人，2025) 的证明，即引理 D.2]，得到。 □

引理 6. 假设 $\mathfrak{M}: \mathbb{R}^{n\times n\times k} \to \mathbb{R}^m$ 满足某个 $\delta_1 > 0$ 下的 $\delta_1\sqrt{k}$*-S2N-t-RIP*。此外，假设

$$\mathcal{M} := \mathfrak{M}^*\mathfrak{M}(\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top) + \boldsymbol{\mathcal{E}} = \boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top + \underbrace{\mathfrak{M}^*\mathfrak{M}(\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top) + \boldsymbol{\mathcal{E}} - \boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top}_{\boldsymbol{\mathcal{E}}_{\boldsymbol{\mathcal{X}}}}$$

对于每个 $1 \le j \le k$ 和 $\delta \le c_1\kappa^2$.，有 $||E_X^{(j)}|| \le \delta\lambda_r(X^{(j)}X^{(j)^H})$。将 $\mathfrak{M}$ 的 *t-SVD* 记为 $\boldsymbol{\mathcal{V}}_{\mathcal{M}} * \boldsymbol{\mathcal{S}}_{\mathcal{M}} * \boldsymbol{\mathcal{W}}_{\mathcal{M}}^{\top}$，进而定义 $\boldsymbol{\mathcal{L}} := \boldsymbol{\mathcal{V}}_{\mathcal{M}}(:, 1:\ r, :) \in \mathbb{R}^{n\times r\times k}$，并定义初始化 $\boldsymbol{\mathcal{U}}_0 = \alpha\boldsymbol{\mathcal{U}}$，其尺度参数满足：

$$\alpha^2 \le \frac{c||\boldsymbol{\mathcal{X}}||^2}{12\sqrt{(R\wedge n)k}\kappa^2||\boldsymbol{\mathcal{U}}||^3}\left(\frac{2\kappa^2||\boldsymbol{\mathcal{U}}||^3}{c_3\sigma_{\min}(\overline{\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{L}}}^{\top} * \boldsymbol{\mathcal{U}}})}\right)^{-48\kappa^2} \min\left\{\sigma_{\min}(\overline{\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{L}}}^{\top} * \boldsymbol{\mathcal{U}}}), ||\overline{\boldsymbol{\mathcal{U}}_0}^{H} v_1||_{l_2}\right\},$$

其中，$v_1 \in \mathbb{C}^{nk}$ 是矩阵 $\mathcal{M} \in \mathbb{R}^{nk\times nk}$.的主特征向量

假设学习率 η 满足 $\eta \le c_3\kappa^{-2}||\boldsymbol{\mathcal{X}}||^{-2}$，那么经过 t_* 次迭代后，

$$t_* \asymp \frac{1}{\eta \max_{1\le j\le k}\sigma_r(\overline{X}^{(j)})^2}\ln\left(\frac{2\kappa^2||\boldsymbol{\mathcal{U}}||}{c_3\sigma_{\min}(\overline{\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{L}}}^{\top} * \boldsymbol{\mathcal{U}}})}\right)$$

以下结论成立：

$$||\boldsymbol{\mathcal{U}}_{t_*}|| \le 3||\boldsymbol{\mathcal{X}}||, \quad ||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}^{\perp}} * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{U}}_{t_*}} * \boldsymbol{\mathcal{W}}_{t_*}|| \le c\kappa^{-2}$$

并且对于每个 $1 \le j \le k$，有

$$\begin{aligned}\sigma_r\left(\overline{\boldsymbol{\mathcal{U}}_{t_*} * \boldsymbol{\mathcal{W}}_{t_*}}^{(j)}\right) &\ge \frac{1}{4}\alpha\beta \\ \sigma_r\left(\overline{\boldsymbol{\mathcal{U}}_{t_*} * \boldsymbol{\mathcal{W}}_{t_*,\perp}}^{(j)}\right) &\le \frac{\kappa^{-2}}{8}\alpha\beta\end{aligned} \tag{27}$$

其中 β 满足 $\sigma_{\min}(\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{L}}}^{\top} * \boldsymbol{\mathcal{U}}) \le \beta \le \sigma_{\min}(\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{L}}}^{\top} * \boldsymbol{\mathcal{U}})\left(\frac{2\kappa^2||\boldsymbol{\mathcal{U}}||^3}{c_3\sigma_{\min}(\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{L}}}^{\top} * (\boldsymbol{\mathcal{V}}\boldsymbol{\mathcal{U}})}\right)^{10\kappa^2}$.

证明。该引理的证明依赖于 [(Karnik 等人，2025) 的结果，即引理 D.7]。[(Karnik 等人，2025) 中引理 D.7] 的第一个条件为：

$$\gamma := \frac{\alpha \max_{1\le j\le k}\sigma_{r+1}(\overline{Z}_t^{(j)})||\boldsymbol{\mathcal{U}}|| + ||\boldsymbol{\mathcal{E}}_t^{\boldsymbol{\mathcal{U}}}||}{\min_{1\le j\le k}\sigma_r(\overline{Z}_t^{(j)})} \cdot \frac{1}{\sigma_{\min}\left(\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{L}}}^{\top} * \boldsymbol{\mathcal{U}}\right)} \le c_2\kappa^2.$$

根据 γ的定义，只需证明

$$\max_{1\le j\le k}\sigma_{r+1}(\overline{Z}_t^{(j)})||\boldsymbol{\mathcal{U}}|| \le \frac{c_3}{2\kappa^2}\min_{1\le j\le k}\sigma_r(\overline{Z}_t^{(j)})\sigma_{\min}\left(\overline{\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{L}}}^{\top} * \boldsymbol{\mathcal{U}}}\right) \tag{28}$$

and

$$||\boldsymbol{\mathcal{E}}_t^{\boldsymbol{\mathcal{U}}}|| \le \frac{c_3}{2\kappa^2}\alpha\min_{1\le j\le k}\sigma_r(\overline{Z}_t^{(j)})\sigma_{\min}\left(\overline{\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{L}}}^{\top} * \boldsymbol{\mathcal{U}}}\right). \tag{29}$$

由于对于 $\boldsymbol{\mathcal{Z}}_t = (\boldsymbol{\mathcal{I}} + \eta\mathcal{M})^t$，傅里叶域中的变换会产生以下分块

$$\overline{Z}_t^{(j)} = (Id + \eta\overline{M}^{(j)})^t,$$

结合不等式 (28) 的结果可得

$$\frac{2\kappa^2||\boldsymbol{\mathcal{U}}||}{c_3\sigma_{\min}(\overline{\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{L}}}^{\top} * \boldsymbol{\mathcal{U}}})} \le \frac{\min_{1\le j\le k}\sigma_r(\overline{Z}_t^{(j)})}{\max_{1\le j\le k}\sigma_{r+1}(\overline{Z}_t^{(j)})} = \left(\frac{1+\eta\min_{1\le j\le k}\sigma_r(\overline{M}^{(j)})}{1+\eta\max_{1\le j\le k}\sigma_{r+1}(\overline{M}^{(j)})}\right)^t. \tag{30}$$

对不等式两边取对数可得

$$\ln\left(\frac{2\kappa^2||\boldsymbol{\mathcal{U}}||}{c_3\sigma_{\min}(\overline{\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{L}}}^{\top} * \boldsymbol{\mathcal{U}}})}\right) \le t\ln\left(\frac{1+\eta\min_{1\le j\le k}\sigma_r(\overline{M}^{(j)})}{1+\eta\max_{1\le j\le k}\sigma_{r+1}(\overline{M}^{(j)})}\right). \tag{31}$$

因此，如果我们令 t_* 为

$$t_* := \left\lceil \ln\left(\frac{2\kappa^2||\boldsymbol{\mathcal{U}}||}{c_3\sigma_{\min}(\overline{\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{L}}}^\top * \boldsymbol{\mathcal{U}}})}\right) \Big/ \ln\left(\frac{1+\eta \min\limits_{1\le j\le k}\sigma_r(\overline{M}^{(j)})}{1+\eta \max\limits_{1\le j\le k}\sigma_{r+1}(\overline{M}^{(j)})}\right) \right\rceil, \tag{32}$$

则条件 (28) 将在傅里叶域的每个块中得到满足。为简化记号，我们定义

$$\phi := \ln\left(\frac{2\kappa^2||\boldsymbol{\mathcal{U}}||}{c_3\sigma_{\min}(\overline{\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{L}}}^\top * \boldsymbol{\mathcal{U}}})}\right). \tag{33}$$

然后我们使用引理 4 来证明第二个条件，即不等式 (29) 成立。要使用引理 4，我们需要保证 $t_* \le t^*$. 如引理 5 所证，我们有

$$t^* \ge \left\lfloor \frac{\ln\left(\frac{||\boldsymbol{\mathcal{M}}||\cdot||\overline{\boldsymbol{\mathcal{U}}_0}^H v_1||_{l_2}}{8(1+\delta_1\sqrt{k})\sqrt{(R\wedge n)}\alpha^3||\boldsymbol{\mathcal{U}}||^3}\right)}{2\ln(1+\eta||\boldsymbol{\mathcal{M}}||)} \right\rfloor. \tag{34}$$

为了保证 $t_* \le t^*$，我们需要证明

$$\frac{\phi}{\ln\left(\frac{1+\eta\min_{1\le j\le k}\sigma_r(\overline{M}^{(j)})}{1+\eta\max 1\le j\le k\sigma_{r+1}(\overline{M}^{(j)})}\right)} \le \frac{1}{2}\cdot\frac{\ln\left(\frac{||\boldsymbol{\mathcal{M}}||\cdot||\overline{\boldsymbol{\mathcal{U}}_0}^H v_1||_{l_2}}{8(1+\delta_1\sqrt{k})\sqrt{(R\wedge n)}\alpha^3||\boldsymbol{\mathcal{U}}||^3}\right)}{2\ln(1+\eta||\boldsymbol{\mathcal{M}}||)}. \tag{35}$$

为证明该不等式，我们首先界定 $\ln(1+\eta||\boldsymbol{\mathcal{M}}||)/\ln\left(\frac{1+\eta_{\min 1\le j\le k}\sigma_r(M^{(j)})\underline{\max 1\le j\le k}\sigma_{r+1}(M^{(j)})}{1+\eta\qquad\qquad\underline{\qquad\qquad}}\right)$。利用事实 $\frac{x}{1+x} \le \ln(1+x) \le x$，我们有

$$\frac{\ln(1+\eta||\boldsymbol{\mathcal{M}}||)}{\ln\left(\frac{1+\eta\min_{1\le j\le k}\sigma_r(\overline{M}^{(j)})}{1+\eta\max 1\le j\le k\sigma_{r+1}(\overline{M}^{(j)})}\right)} \le \frac{||\mathcal{M}||(1+\eta\min_{1\le j\le k}\sigma_r(\overline{M}^{(j)})}{\min_{1\le j\le k}\sigma_r(\overline{M}^{(j)}) - \max_{1\le j\le k}\sigma_{r+1}(\overline{M}^{(j)})}. \tag{36}$$

利用假设 $\delta \le \frac{1}{3}$ 以及 $\eta \le c_3\kappa^{-2}||\boldsymbol{\mathcal{X}}||^{-2}$ 和 [（Karnik 等人，2025）的结果，引理 D.6], 我们有

$$\begin{aligned}\frac{||\mathcal{M}||(1+\eta\min_{1\le j\le k}\sigma_r(\overline{M}^{(j)})}{\min_{1\le j\le k}\sigma_r(\overline{M}^{(j)}) - \max_{1\le j\le k}\sigma_{r+1}(\overline{M}^{(j)})} &\le \frac{(1+\delta)||\boldsymbol{\mathcal{T}}||}{(1-\delta)\lambda_r(\overline{T}^{(j)})}\left(1 + c_3(1+\delta)\left(\frac{\lambda_1(\overline{X}^{(j)})}{\kappa||\boldsymbol{\mathcal{X}}||}\right)^2\right) \\ &\le \kappa^2\frac{1+\delta}{1-2\delta}(1+c_3(1+\delta)\frac{1}{\kappa^2}) \overset{(a)}{\le} 5\kappa^2,\end{aligned} \tag{37}$$

其中 (a) 利用了 $\delta \le 1/3$ 与 c_3 充分小这一事实。因此，我们有

$$\frac{\ln(1+\eta||\boldsymbol{\mathcal{M}}||)}{\ln\left(\frac{1+\eta\min_{1\le j\le k}\sigma_r(\overline{M}^{(j)})}{1+\eta\max 1\le j\le k\sigma_{r+1}(\overline{M}^{(j)})}\right)} \le 5\kappa^2. \tag{38}$$

借助这一上界，我们回顾不等式 (35)

$$20\kappa^2\cdot\ln\left(\frac{2\kappa^2||\boldsymbol{\mathcal{U}}||}{c_3\sigma_{\min}(\overline{\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{L}}}^\top * \boldsymbol{\mathcal{U}}})}\right) \le \ln\left(\frac{||\boldsymbol{\mathcal{M}}||\cdot||\overline{\boldsymbol{\mathcal{U}}_0}^H v_1||_{l_2}}{8(1+\delta_1\sqrt{k})\sqrt{(R\wedge n)}\alpha^3||\boldsymbol{\mathcal{U}}||^3}\right), \tag{39}$$

这等于

$$\left(\frac{2\kappa^2||\boldsymbol{\mathcal{U}}||}{c_3\sigma_{\min}(\overline{\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{L}}}^\top * \boldsymbol{\mathcal{U}}})}\right)^{20\kappa^2} \le \frac{||\boldsymbol{\mathcal{M}}||\cdot||\overline{\boldsymbol{\mathcal{U}}_0}^H v_1||_{l_2}}{8(1+\delta_1\sqrt{k})\sqrt{(R\wedge n)}\alpha^3||\boldsymbol{\mathcal{U}}||^3} \overset{(a)}{=} \frac{||\boldsymbol{\mathcal{M}}||\cdot||\overline{\boldsymbol{\mathcal{U}}}^H v_1||_{l_2}}{8(1+\delta_1\sqrt{k})\sqrt{(R\wedge n)}\alpha^2||\boldsymbol{\mathcal{U}}||^3}, \tag{40}$$

其中，(a) 利用了以下事实: $||\boldsymbol{\mathcal{U}}_0{}^H v_1||_{l_2}/\alpha = ||\boldsymbol{\mathcal{U}}^H v_1||_{l_2}$. 为证明不等式 (40)，我们选择 α 为

$$\alpha^2 \leq \left(\frac{2\kappa^2||\boldsymbol{\mathcal{U}}||}{c_3\sigma_{\min}(\overline{\boldsymbol{\mathcal{V}}_{\mathcal{L}}^\top * \boldsymbol{\mathcal{U}}})}\right)^{-20\kappa^2} \cdot \frac{||\boldsymbol{\mathcal{M}}|| \cdot ||\overline{\boldsymbol{\mathcal{U}}}^H v_1||_{l_2}}{8(1+\delta_1\sqrt{k})\sqrt{(R\wedge n)}||\boldsymbol{\mathcal{U}}||^3} \tag{41}$$

鉴于 $\delta \leq \frac{1}{3}$ 且 $||\mathcal{M}|| \geq \frac{2}{3}||\boldsymbol{\mathcal{X}}||^2$，我们将 α 设得更小，即

$$\alpha^2 \leq \left(\frac{2\kappa^2||\boldsymbol{\mathcal{U}}||}{c_3\sigma_{\min}(\overline{\boldsymbol{\mathcal{V}}_{\mathcal{L}}^\top * \boldsymbol{\mathcal{U}}})}\right)^{-20\kappa^2} \cdot \frac{||\boldsymbol{\mathcal{X}}||^2 \cdot ||\overline{\boldsymbol{\mathcal{U}}}^H v_1||_{l_2}}{16\sqrt{(R\wedge n)k}||\boldsymbol{\mathcal{U}}||^3}. \tag{42}$$

因此 $t_* \leq t^*$ 成立，则 [（Karnik 等人，2025）引理D.7] 中的条件成立。因此，利用 [（Karnik 等人，2025）引理 D.7],的结果，我们有

$$\begin{aligned} ||\boldsymbol{\mathcal{E}}_{t_*}^{\boldsymbol{\mathcal{U}}}|| &\leq 8(1+\delta_1\sqrt{k})\sqrt{(R\wedge n)}\frac{\alpha^3}{||\mathcal{M}||}||\boldsymbol{\mathcal{U}}||^3(1+\eta||\mathcal{M}||)^{3t_*} \\ &\overset{(a)}{\leq} 12\sqrt{(R\wedge n)k}\frac{\alpha^3}{||\mathcal{M}||}||\boldsymbol{\mathcal{U}}||^3(1+\eta||\mathcal{M}||)^{3t_*}, \end{aligned} \tag{43}$$

其中 (a) 利用了 $\delta \leq \frac{1}{3}$ 且 $||\mathcal{M}|| \geq \frac{2}{3}||\boldsymbol{\mathcal{X}}||^2$ 这一事实，该事实来自 [（Karnik 等人，2025）引理 D.6]。因此，利用 $Z_t^{(j)} = (Id + \eta M^{(j)})^t$，不等式 (29) 在以下条件下成立

$$12\sqrt{(R\wedge n)}\frac{\alpha^3}{||\mathcal{M}||}||\boldsymbol{\mathcal{U}}||^3(1+\eta||\mathcal{M}||)^{3t_*} \leq \frac{c_3}{2\kappa^2}\min_{1\leq j\leq k}\sigma_r\left((Id+\eta\overline{M}^{(j)})^{t_*}\right)\sigma_{\min}\left(\overline{\boldsymbol{\mathcal{V}}_{\mathcal{L}}^\top * \boldsymbol{\mathcal{U}}}\right), \tag{44}$$

这等价于

$$\alpha^2 \leq c_3\frac{\sigma_{\min}\left(\overline{\boldsymbol{\mathcal{V}}_{\mathcal{L}}^\top * \boldsymbol{\mathcal{U}}}\right)||\boldsymbol{\mathcal{X}}||^2}{12\sqrt{(R\wedge n)k}\kappa^2||\boldsymbol{\mathcal{U}}||^3} \cdot \frac{(Id+\eta\sigma_r(\overline{M}^{(j)}))^{t_*}}{(1+\eta||\mathcal{M}||)^{3t_*}}. \tag{45}$$

注意

$$\frac{Id+\eta\sigma_r(\overline{M}^{(j)})}{(1+\eta||\mathcal{M}||)^{3t_*}} = \exp\left(t_*\ln\left(\frac{(Id+\eta\sigma_r(\overline{M}^{(j)}))}{(1+\eta||\mathcal{M}||)^3}\right)\right) \geq \exp\left(-3t_*\ln(1+\eta||\mathcal{M}||^3)\right). \tag{46}$$

根据 t_* 的定义，即 $t_* = \left\lceil \phi \Big/ \ln\left(\frac{1+\eta}{1+\eta}^{\text{最小值}_{1\leq j\leq k}\sigma_r(M^{(j)})\underline{\text{最大值}}_{1\leq j\leq k}\sigma_{r+1}(M^{(j)})}\right)\right\rceil$ 以及不等式 (38)，我们得到

$$\exp\left(-3t_*\ln(1+\eta||\mathcal{M}||^3)\right) \geq \exp\left(-15\phi\kappa^2\right) = \left(\frac{2\kappa^2||\boldsymbol{\mathcal{U}}||}{c_3\sigma_{\min}(\overline{\boldsymbol{\mathcal{V}}_{\mathcal{L}}^\top * \boldsymbol{\mathcal{U}}})}\right)^{-15\kappa^2}. \tag{47}$$

结合不等式 (45) 和 (47)，我们选择

$$\alpha^2 \leq c_3\frac{\sigma_{\min}\left(\overline{\boldsymbol{\mathcal{V}}_{\mathcal{L}}^\top * \boldsymbol{\mathcal{U}}}\right)||\boldsymbol{\mathcal{X}}||^2}{12\sqrt{(R\wedge n)k}\kappa^2||\boldsymbol{\mathcal{U}}||^3} \cdot \left(\frac{2\kappa^2||\boldsymbol{\mathcal{U}}||}{c_3\sigma_{\min}(\overline{\boldsymbol{\mathcal{V}}_{\mathcal{L}}^\top * \boldsymbol{\mathcal{U}}})}\right)^{-15\kappa^2}. \tag{48}$$

由此 α，不等式 (29) 成立，并且满足 [(Karnik 等人，2025)，引理 D.7]的条件，从而得到

$$||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{V}}^\perp}^\top * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t}|| \overset{(a)}{\leq} 14(\delta+\gamma) \leq c\kappa^{-2}, \tag{49}$$

其中，(a) 使用了以下假设： $\delta \leq c_1\kappa^{-2}$ 和 $\eta \leq c_3\kappa^{-2}||\boldsymbol{\mathcal{X}}||^{-2}$，并将常数$c_1$ 和 c_3 设得足够小。此外，对于每个 $1 \leq j \leq k$，利用 [（Karnik 等人，2025）中的结果，引理 D.7], ，我们有

$$\begin{aligned} \sigma_{\min}(\overline{\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t}^{(j)}) &\geq \frac{1}{4}\alpha\beta \\ \sigma_1(\overline{\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp}}^{(j)}) &\geq \frac{\kappa^{-2}}{8}\alpha\beta, \end{aligned} \tag{50}$$

其中 $\beta := \min_{1\le j\le k} \sigma_r(Z_t^{(j)})\sigma_{\min}(\overline{\mathcal{V}_{\mathcal{L}}^{\perp} * \mathcal{U}})$。

接下来，我们证明 β 和 $||\mathcal{U}_{t_*}||$的界。

考虑 $\beta := \min_{1\le j\le k} \sigma_r(Z_t^{(j)})\sigma_{\min}(\overline{\mathcal{V}_{\mathcal{L}}^{\perp} * \mathcal{U}})$。根据 $Z_t^{(j)}$ 的定义和不等式 (38)，我们有

$$\begin{aligned}(1+\eta\sigma_r(\overline{M}^{(j)}))^{t_*} &\le \exp\left(t_* \ln(1+\eta\sigma_r(\overline{M}^{(j)}))\right) \le \exp\left(t_* \ln(1+\eta||\mathcal{M}||)\right) \\ &\le \exp\left(2\phi \max_{1\le j\le k} \frac{\ln(1+\eta||\mathcal{M}||)}{\frac{1+\eta\sigma_r(\overline{M}^{(j)})}{1+\eta\sigma_{r+1}(\overline{M}^{(j)})}}\right) \le \exp\left(10\phi\kappa^2\right) \\ &= \left(\frac{2\kappa^2||\mathcal{U}||}{c_3\sigma_{\min}(\overline{\mathcal{V}_{\mathcal{L}}^{\top} * \mathcal{U}})}\right)^{10\kappa^2}\end{aligned} \tag{51}$$

对所有 $1 \le j \le k$.均成立

于是我们有

$$\beta \le \sigma_{\min}(\overline{\mathcal{V}_{\mathcal{L}}^{\top} * \mathcal{U}}) \left(\frac{2\kappa^2||\mathcal{U}||}{c_3\sigma_{\min}(\overline{\mathcal{V}_{\mathcal{L}}^{\top} * \mathcal{U}})}\right)^{10\kappa^2}. \tag{52}$$

最后，我们证明 $||\mathcal{U}_{t_*}|| \le 3||\mathcal{U}||$.。根据$\mathcal{U}_{t_*} = \mathcal{Z}_{t_*} * \mathcal{U}_0 + \mathcal{E}_{t_*}^{\mathcal{U}}$的定义，我们有

$$||\mathcal{U}_{t_*}|| = \alpha||\mathcal{Z}_{t_*}|| \cdot ||\mathcal{U}|| + ||\mathcal{E}_{t_*}^{\mathcal{U}}||. \tag{53}$$

由不等式 (29)，我们得到

$$||\mathcal{E}_t^{\mathcal{U}}|| \le \frac{c_3}{2\kappa^2}\alpha||\mathcal{Z}_t||\sigma_{\min}\left(\overline{\mathcal{V}_{\mathcal{L}}^{\top} * \mathcal{U}}\right) \le \frac{c_3}{2\kappa^2}\alpha||\mathcal{Z}_t||\sigma_{\min}(\overline{\mathcal{V}_{\mathcal{L}}^{H}})\sigma_{\max}(\overline{\mathcal{U}}) \le \alpha||\mathcal{Z}_t||||\mathcal{U}||, \tag{54}$$

由此可得

$$\begin{aligned}||\mathcal{U}_{t_*}|| &\le 2\alpha||\mathcal{Z}_t||||\mathcal{U}|| \le 2\alpha(1+\eta||\mathcal{M}||)^{t_*}||\mathcal{U}|| \\ &= 2\alpha \ln(t_*(1+\eta||\mathcal{M}||))||\mathcal{U}|| \overset{(a)}{\le} 2\alpha||\mathcal{U}|| \left(\frac{2\kappa^2||\mathcal{U}||}{c_3\sigma_{\min}(\overline{\mathcal{V}_{\mathcal{L}}^{\top} * \mathcal{U}})}\right)^{10\kappa^2} \\ &\overset{(b)}{\le} 2||\mathcal{U}||c_3\sqrt{\frac{\sigma_{\min}\left(\overline{\mathcal{V}_{\mathcal{L}}^{\top} * \mathcal{U}}\right)||\mathcal{X}||^2}{12\sqrt{(R\wedge n)k}\kappa^2||\mathcal{U}||^3}} \cdot \left(\frac{2\kappa^2||\mathcal{U}||}{c_3\sigma_{\min}(\overline{\mathcal{V}_{\mathcal{L}}^{\top} * \mathcal{U}})}\right)^{-15\kappa^2/2} \\ &= 2c_3||\mathcal{X}||\sqrt{\frac{\sigma_{\min}\left(\overline{\mathcal{V}_{\mathcal{L}}^{\top} * \mathcal{U}}\right)}{12\sqrt{(R\wedge n)k}\kappa^2||\mathcal{U}||}} \cdot \left(\frac{2\kappa^2||\mathcal{U}||}{c_3\sigma_{\min}(\overline{\mathcal{V}_{\mathcal{L}}^{\top} * \mathcal{U}})}\right)^{-15\kappa^2/2} \le 3||\mathcal{U}||,\end{aligned} \tag{55}$$

其中，(a) 使用不等式 (51)；(b) 使用不等式 (48)。引理 1 可直接由引理 6 以及 [(Karnik 等人, 2025) 中引理 D.9] 所采用的证明策略推出。 □

E.6 引理2的证明

注意，对于 $t = t_*$，利用这些假设可直接得到这四个不等式成立。在证明$t+1$情形之前，我们先对 $\left\|(\mathfrak{M}^*\mathfrak{M} - \mathfrak{I})\left(\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top\right) + \boldsymbol{\mathcal{E}}\right\|$进行如下界定：

$$
\begin{aligned}
&||\left(\mathfrak{M}^*\mathfrak{M} - \mathfrak{I}\right)\left(\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top\right) + \boldsymbol{\mathcal{E}}|| \\
\leq &||\left(\mathfrak{M}^*\mathfrak{M} - \mathfrak{I}\right)\left(\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t * \boldsymbol{\mathcal{W}}_t^\top * \boldsymbol{\mathcal{U}}_t^\top\right)|| \\
&\quad + ||\left(\mathfrak{M}^*\mathfrak{M} - \mathfrak{I}\right)\left(\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp} * \boldsymbol{\mathcal{W}}_{t,\perp}^\top * \boldsymbol{\mathcal{U}}_t^\top\right)|| + ||\boldsymbol{\mathcal{E}}|| \\
(a)\ \leq &\delta\sqrt{kr}||\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t * \boldsymbol{\mathcal{W}}_t^\top * \boldsymbol{\mathcal{U}}_t^\top|| + \delta\sqrt{k}||\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp} * \boldsymbol{\mathcal{W}}_{t,\perp}^\top * \boldsymbol{\mathcal{U}}_t^\top||_* + ||\boldsymbol{\mathcal{E}}|| \\
&\leq \delta\sqrt{kr}\left(||\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top|| + ||\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t * \boldsymbol{\mathcal{W}}_t^\top * \boldsymbol{\mathcal{U}}_t^\top||\right) + \delta\sqrt{k}||\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp} * \boldsymbol{\mathcal{W}}_{t,\perp}^\top * \boldsymbol{\mathcal{U}}_t^\top||_* + ||\boldsymbol{\mathcal{E}}|| \\
&= \delta\sqrt{kr}\left(||\boldsymbol{\mathcal{X}}||^2 + ||\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t||^2\right) + \delta\sqrt{k}||\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp} * \boldsymbol{\mathcal{W}}_{t,\perp}^\top * \boldsymbol{\mathcal{U}}_t^\top||_* + ||\boldsymbol{\mathcal{E}}|| \\
&\leq \delta\sqrt{kr}\left(||\boldsymbol{\mathcal{X}}||^2 + ||\boldsymbol{\mathcal{U}}_t||^2\right) + \delta\sqrt{k}||\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp} * \boldsymbol{\mathcal{W}}_{t,\perp}^\top * \boldsymbol{\mathcal{U}}_t^\top||_* + ||\boldsymbol{\mathcal{E}}|| \\
(b)\ \leq &\delta\sqrt{kr}\left(||\mathrm{X}||^2 + 9||\mathrm{X}||^2\right) + \delta\sqrt{k^3}((R\wedge n) - r)||\mathrm{U}_t * \mathrm{W}_{t,\perp} * \mathrm{W}_{t,\perp}^\top * \mathrm{U}_t^\top|| + ||\mathrm{E}|| \\
\leq &10\delta\sqrt{kr}\kappa^2\underline{\sigma}_{\min}^2(\mathrm{X}) + \delta\sqrt{k^3}((R\wedge n) - r)||\mathrm{U}_t * \mathrm{W}_{t,\perp}||^2 + ||\mathrm{E}||\ (c)\ \leq 10c_1\kappa^{-2}\sigma_{\min}^2(\mathrm{X}) + 4\delta\sqrt{k^3} \\
&((R\wedge n) - r)\gamma^2(1 + 80\eta c_2\sqrt{k}\underline{\sigma}_{\min}^2(\mathrm{X}))^{2(t-t_*)} + c_1\kappa^{-2}\sigma_{\min}^2(\mathrm{X})\ (d)\ \leq 10c_1\kappa^{-2}\sigma_{\min}^2(\mathrm{X}) + 8\delta\sqrt{k^3} \\
&((R\wedge n) - r)\gamma^{7/4}\sigma_{\min}(\mathrm{X})^{1/4} + \overline{c_1}\kappa^{-2}\sigma_{\min}^2(\mathrm{X})\ (e)\ \leq 40c_1\kappa^{-2}\overline{\sigma}_{\min}^2(\mathrm{X}),
\end{aligned}
\tag{56}
$$

其中，(a) 使用了如下假设： $\mathfrak{M}$ 满足 $(r, \delta\sqrt{kr})$ S2S-t-RIP 和 $\sqrt{k}\delta$-S2N-t-RIP；(b) 由假设 $||\mathrm{U}_t|| \leq 3|\mathrm{X}||$以及 $||\mathrm{U}_t * \mathrm{W}_{t,\perp} * \mathrm{W}_{t,\perp}^\top * \mathrm{U}_t^\top||_* \leq k((R\wedge n) - r)||\mathrm{U}_t * \mathrm{W}_{t,\perp} * \mathrm{W}_{t,\perp}^\top * \mathrm{U}_t^\top||$得出；(c) 使用了假设 $\delta \leq \frac{c_1}{\kappa^4\sqrt{kr}}$ 以及归纳假设；(d) 使用了 t_1 和 t_* 的定义；(e) 使用了假设 $\gamma \leq \frac{c_2\sigma_{\min}(\mathrm{X})}{k^{8/7}\kappa^2(R\wedge n)}$，并选取足够小的$c_2$。借助该不等式，可以将 $||(\mathrm{A}^*\mathrm{A} - \mathrm{I})\left(\mathrm{X} * \mathrm{X}^\top - \mathrm{U}_t * \mathrm{U}_t^\top\right)||$在 [(Karnik 等人，2025) 的引理 E.1–引理 E.7]中替换为 $||(\mathfrak{M}^*\mathfrak{M} - \mathfrak{I})\left(\mathrm{X} * \mathrm{X}^\top - \mathrm{U}_t * \mathrm{U}_t^\top\right) + \mathrm{E}||$，因为它们具有相同的上界。

通过选择足够小的 c_2，结合引理 2 中的其他假设，我们满足了 [(Karnik 等人，2025)，引理 E.6] 中的假设，因此可以直接使用 [(Karnik 等人，2025)，引理 E.6] 中的结果来证明 $|||\boldsymbol{\mathcal{U}}_{t+1}|| \leq 3||\boldsymbol{\mathcal{X}}||$。

此外，满足 [(Karnik 等人，2025)，引理 E.1] 中的假设后，我们使用 [(Karnik 等人，2025)，引理 E.1] 中的结果来证明归纳假设 (5)：

$$
\begin{aligned}
\sigma_{\min}(\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * \boldsymbol{\mathcal{U}}_{t+1}) &\geq \sigma_{\min}(\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * \boldsymbol{\mathcal{U}}_{t+1} * \boldsymbol{\mathcal{W}}_{t+1}) \\
&\geq \sigma_{\min}(\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * \boldsymbol{\mathcal{U}}_{t+1})\left(1 + \frac{1}{4}\eta\sigma_{\min}(\boldsymbol{\mathcal{X}})^2 - \eta\sigma_{\min}(\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * \boldsymbol{\mathcal{U}}_t)^2\right) \\
&\geq \sigma_{\min}(\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * \boldsymbol{\mathcal{U}}_{t+1})\left(1 + \frac{1}{4}\eta\sigma_{\min}(\boldsymbol{\mathcal{X}})^2 - 0.1\eta\sigma_{\min}(\boldsymbol{\mathcal{X}})^2\right) \\
&\geq \sigma_{\min}(\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * \boldsymbol{\mathcal{U}}_{t+1})\left(1 + \frac{1}{8}\eta\sigma_{\min}(\boldsymbol{\mathcal{X}})^2\right) \\
&\geq \left(1 + \frac{1}{8}\eta\sigma_{\min}(\boldsymbol{\mathcal{X}})^2\right) \cdot \frac{1}{2}\gamma\left(1 + \frac{1}{8}\eta\sigma_{\min}(\boldsymbol{\mathcal{X}})^2\right)^{t-t_*} \\
&= \frac{1}{2}\gamma\left(1 + \frac{1}{8}\eta\sigma_{\min}(\boldsymbol{\mathcal{X}})^2\right)^{(t+1)-t_*}.
\end{aligned}
\tag{57}
$$

该不等式意味着 $\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^{\top} * \boldsymbol{\mathcal{U}}_{t+1}$ 的所有奇异值均为正；结合引理 2 和方程 (56) 中的假设， [(Karnik 等人，2025)，引理 E.3] 中的假设也得以满足。因此，我们可以使用 [(Karnik 等人，2025)，引理 E.3] 中的结果来证明归纳假设 (6)：

$$
\begin{aligned}
&||\overline{\boldsymbol{\mathcal{U}}_{t+1} * \boldsymbol{\mathcal{W}}_{t+1,\perp}}^{(j)}|| \\
&\leq \left(1 - \frac{\eta}{2}||\overline{\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp}}^{(j)}||^2 + 9\eta||\overline{\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}^\perp}^{\top} * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t}}^{(j)}|| \cdot ||\boldsymbol{\mathcal{X}}||^2\right) ||\overline{\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp}}^{(j)}|| \\
&\quad + 2\eta||(\mathfrak{M}^*\mathfrak{M} - \mathfrak{I})(\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^{\top} - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^{\top}) + \boldsymbol{\mathcal{E}}|| \cdot ||\overline{\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp}}^{(j)}|| \\
&\leq \left(1 - \frac{\eta}{2} \cdot 4\gamma^2(1 + 80\eta c_2 \sigma_{\min}^2(\boldsymbol{\mathcal{X}})) + 9\eta c_2 \kappa^{-2}||\boldsymbol{\mathcal{X}}||^2\right) ||\overline{\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp}}^{(j)}|| \\
&\quad + 2\eta \cdot 40 c_1 \kappa^{-2}\sigma_{\min}(\boldsymbol{\mathcal{X}})^2 ||\overline{\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp}}^{(j)}|| \\
&\leq \left(1 - 2\eta\gamma^2(1 + 80\eta c_2 \sigma_{\min}^2(\boldsymbol{\mathcal{X}})) + 9\eta c_2 \sigma_{\min}(\boldsymbol{\mathcal{X}})^2\right) ||\overline{\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp}}^{(j)}|| \\
&\quad + 80\eta \cdot c_1 \kappa^{-2}\sigma_{\min}(\boldsymbol{\mathcal{X}})^2 ||\overline{\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp}}^{(j)}|| \\
&\leq (1 + 80 c_2 \eta \sigma_{\min}(\boldsymbol{\mathcal{X}})^2) ||\overline{\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp}}^{(j)}|| \\
&\leq 2\gamma(1 + 80 c_2 \eta \sigma_{\min}(\boldsymbol{\mathcal{X}})^2)^{t+1-t_*}.
\end{aligned}
\tag{58}
$$

注意，对于任意分块对角矩阵 $A = \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_n \end{bmatrix}$，我们有 $||A|| \leq \max_i ||A_i||$。于是我们有 $||\overline{\boldsymbol{\mathcal{U}}_{t+1} * \boldsymbol{\mathcal{W}}_{t+1,\perp}}^{\top}|| \leq \max_j ||\overline{\boldsymbol{\mathcal{U}}_{t+1} * \boldsymbol{\mathcal{W}}_{t+1,\perp}}^{(j)}||$ 因为 $\boldsymbol{\mathcal{U}}_{t+1} * \boldsymbol{\mathcal{W}}_{t+1,\perp}$ 是一个分块对角矩阵。因此，我们完成了归纳假设 (6) 的证明。

接下来，我们通过 [(Karnik 等人，2025)，引理 E.5] 来证明 $||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^{\top} * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{U}}_{t+1} * \boldsymbol{\mathcal{W}}_{t+1}}|| \leq c_2 \kappa^{-2}$。注意，利用引理2的假设以及归纳假设 (5)-(7)，可以满足 [(Karnik 等人，2025)，引理 E.5] 中的假设。

$$
\begin{aligned}
&||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^{\top} * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{U}}_{t+1} * \boldsymbol{\mathcal{W}}_{t+1}}|| \\
&\leq (1 - \frac{\eta}{4}\sigma_{\min}(\boldsymbol{\mathcal{X}})^2)||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^{\top} * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t}|| + 150\eta||\left(\mathfrak{M}^*\mathfrak{M} - \mathfrak{I}\right)\left(\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^{\top} - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^{\top}\right) + \boldsymbol{\mathcal{E}}|| \\
&\quad + 500\eta^2||\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^{\top} - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^{\top}||^2 \\
&\leq (1 - \frac{\eta}{4}\sigma_{\min}(\boldsymbol{\mathcal{X}})^2)c_2\kappa^{-2} + 150\eta \cdot 40 c_1 \kappa^{-2}\sigma_{\min}^2(\boldsymbol{\mathcal{X}}) + 500\eta^2(||\boldsymbol{\mathcal{X}}||^2 + ||\boldsymbol{\mathcal{U}}_t||^2)^2 \\
&\leq (1 - \frac{\eta}{4}\sigma_{\min}(\boldsymbol{\mathcal{X}})^2)c_2\kappa^{-2} + 6000 c_1 \eta\kappa^{-2}\sigma_{\min}^2(\boldsymbol{\mathcal{X}}) + 500\eta^2(||\boldsymbol{\mathcal{X}}||^2 + 9||\boldsymbol{\mathcal{X}}||^2)^2 \\
&= (1 - \frac{\eta}{4}\sigma_{\min}(\boldsymbol{\mathcal{X}})^2)c_2\kappa^{-2} + 6000 c_1 \eta\kappa^{-2}\sigma_{\min}^2(\boldsymbol{\mathcal{X}}) + 50000\eta^2||\boldsymbol{\mathcal{X}}||^4 \\
&\leq (1 - \frac{\eta}{4}\sigma_{\min}(\boldsymbol{\mathcal{X}})^2)c_2\kappa^{-2} + 6000 c_1 \eta\kappa^{-2}\sigma_{\min}^2(\boldsymbol{\mathcal{X}}) + 50000\eta \cdot c_1\kappa^{-4}||\boldsymbol{\mathcal{X}}||^{-2} \cdot ||\boldsymbol{\mathcal{X}}||^4 \\
&\leq (1 - \frac{\eta}{4}\sigma_{\min}(\boldsymbol{\mathcal{X}})^2)c_2\kappa^{-2} + 6000 c_1 \eta\kappa^{-2}\sigma_{\min}^2(\boldsymbol{\mathcal{X}}) + 50000\eta \cdot c_1\kappa^{-4}||\boldsymbol{\mathcal{X}}||^{-2} \cdot ||\boldsymbol{\mathcal{X}}||^4 \\
&\leq (1 - \frac{\eta}{4}\sigma_{\min}(\boldsymbol{\mathcal{X}})^2)c_2\kappa^{-2} + 6000 c_1 \eta\kappa^{-2}\sigma_{\min}^2(\boldsymbol{\mathcal{X}}) + 50000\eta c_1\kappa^{-2}\sigma_{\min}(\boldsymbol{\mathcal{X}})^2 \\
&\leq (1 - \frac{\eta}{4}\sigma_{\min}(\boldsymbol{\mathcal{X}})^2)c_2\kappa^{-2} + 56000\eta c_1 \kappa^{-2}\sigma_{\min}(\boldsymbol{\mathcal{X}})^2.
\end{aligned}
\tag{59}
$$

取足够小的 c_1（即 $c_1 \lesssim \frac{4}{56000}$），我们有 $||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^{\top} * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{U}}_{t+1} * \boldsymbol{\mathcal{W}}_{t+1}}|| \leq c_2\kappa^{-2}$。因此，我们完成了归纳证明。

E.7 引理 3 的证明

利用 t_1 的定义（方程 (4)）以及

$$\sigma_{\min}(\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * \boldsymbol{\mathcal{U}}_{t_1}) \geq \frac{1}{2}\gamma\left(1+\frac{1}{8}\eta\sigma_{\min}(\boldsymbol{\mathcal{X}})^2\right)^{t_1-t_*}$$

由引理 2，我们有

$$\frac{1}{\sqrt{10}}\sigma_{\min}(\boldsymbol{\mathcal{X}}) \geq \sigma_{\min}(\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * \boldsymbol{\mathcal{U}}_{t_1}) \geq \frac{1}{2}\gamma\left(1+\frac{1}{8}\eta\sigma_{\min}(\boldsymbol{\mathcal{X}})^2\right)^{t_1-t_*},$$

由此可得

$$t_1 - t_* \leq \frac{\log\left(\frac{2}{\gamma\sqrt{10}}\sigma_{\min}(\boldsymbol{\mathcal{X}})\right)}{\log\left(1+\frac{1}{8}\eta\sigma_{\min}(\boldsymbol{\mathcal{X}})^2\right)} \overset{(a)}{\leq} \frac{16}{\eta\sigma_{\min}(\boldsymbol{\mathcal{X}})^2}\log\left(\frac{2}{\gamma\sqrt{10}}\sigma_{\min}(\boldsymbol{\mathcal{X}})\right), \tag{60}$$

其中在 (a) 中，我们利用了如下事实：对 $0 \leq x < 1$，有 $\frac{1}{\log(1+x)} \leq \frac{2}{x}$。

因此，我们将 $||\boldsymbol{\mathcal{U}}_{t_1} * \boldsymbol{\mathcal{W}}_{t_1,\perp}||$ 界定为

$$\begin{aligned}
||\boldsymbol{\mathcal{U}}_{t_1} * \boldsymbol{\mathcal{W}}_{t_1,\perp}|| &\leq 2\gamma\left(1+80\eta c_2\sigma_{\min}(\boldsymbol{\mathcal{X}})^2\right)^{t_1-t_*} \\
&\overset{(a)}{\leq} 2\gamma\left(\frac{2}{\sqrt{10}}\cdot\frac{\sigma_{\min}(\boldsymbol{\mathcal{X}})}{\gamma}\right)^{1280c_2} \\
&\overset{(b)}{\leq} 2\gamma\left(\frac{2}{\sqrt{10}}\cdot\frac{\sigma_{\min}(\boldsymbol{\mathcal{X}})}{\gamma}\right)^{1/64} \\
&\overset{(c)}{\leq} 3\gamma^{63/64}\sigma_{\min}(\boldsymbol{\mathcal{X}})^{1/64} \leq 3\gamma^{7/8}\sigma_{\min}(\boldsymbol{\mathcal{X}})^{1/8},
\end{aligned} \tag{61}$$

其中 (a) 由方程 (60) 得出；(b) 使用了 c_2 被选取得足够小这一假设；(c) 使用了 $\sigma_{\min}(\boldsymbol{\mathcal{X}}) \geq \gamma$ 这一事实。

然后我们将引理 3 的证明分为两种情形：精确秩情形和过参数化（过秩）情形。

过秩情形： 设 $\hat{t} := t_1 + \left\lceil \frac{300}{\eta\sigma_{\min}^2(\boldsymbol{\mathcal{X}})}\ln\left(16k\frac{\kappa^{1/4}}{R\wedge n}(()-r)\cdot\frac{||\boldsymbol{\mathcal{X}}||^{7/4}}{\gamma^{7/4}}\right)\right\rceil$。我们首先陈述关于 $t_1 \leq t \leq \hat{t}$ 的归纳假设：

$$\sigma_{\min}(\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t) \geq \sigma_{\min}(\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * \boldsymbol{\mathcal{U}}_t) \geq \frac{\sigma_{\min}(\boldsymbol{\mathcal{X}})}{\sqrt{10}}, \tag{62}$$

$$||\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp}|| \leq (1+80\eta c_2\sigma_{\min}^2(\boldsymbol{\mathcal{X}}))^{t-t_1}||\boldsymbol{\mathcal{U}}_{t_1} * \boldsymbol{\mathcal{W}}_{t_1,\perp}||, \tag{63}$$

$$||\boldsymbol{U}_t|| \leq 3||\boldsymbol{X}\,||, \tag{64}$$

$$||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}^\perp}^\top * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t}|| \leq c_2\kappa^{-2}, \tag{65}$$

$$||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * (\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top)|| \leq 10(1-\frac{\eta}{400}\sigma_{\min}^2(\boldsymbol{\mathcal{X}}))^{t-t_1}||\boldsymbol{\mathcal{X}}||^2 \tag{66}$$

$$+ 18\eta||\boldsymbol{\mathcal{X}}||^2||\boldsymbol{\mathcal{E}}||\sum_{\tau=t_1+1}^{t}(1-\frac{\eta}{200}\sigma_{\min}^2(\boldsymbol{\mathcal{X}}))^{\tau-t_1-1}. \tag{67}$$

当 $t = t_1$ 时，不等式 (62)、(64) 和 (65) 可由引理 2 得出。至于不等式 (63)，显然，当 $t = t_1$ 时它成立。当 $t = t_1$ 时，我们有

$$\begin{aligned}
||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * (\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top - \boldsymbol{\mathcal{U}}_{t_1} * \boldsymbol{\mathcal{U}}_{t_1}^\top)|| &= ||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * (\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top - \boldsymbol{\mathcal{U}}_{t_1} * \boldsymbol{\mathcal{W}}_{t_1} * \boldsymbol{\mathcal{W}}_{t_1}^\top * \boldsymbol{\mathcal{U}}_{t_1}^\top)|| \\
&\leq ||\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top|| + ||\boldsymbol{\mathcal{U}}_{t_1} * \boldsymbol{\mathcal{W}}_{t_1} * \boldsymbol{\mathcal{W}}_{t_1}^\top * \boldsymbol{\mathcal{U}}_{t_1}^\top|| \\
&\leq ||\boldsymbol{\mathcal{X}}||^2 + ||\boldsymbol{\mathcal{U}}_{t_1}||^2||\boldsymbol{\mathcal{W}}_{t_1}||^2 \overset{(a)}{\leq} 10||\boldsymbol{\mathcal{X}}||^2,
\end{aligned}$$

其中 (a) 由不等式 (64) 得出。接下来，我们要证明这些不等式在步骤 $t+1$ 时也成立。为此，我们需要对项 $\left\|(\mathfrak{M}^*\mathfrak{M}-\mathfrak{I})\left(\boldsymbol{\mathcal{X}}*\boldsymbol{\mathcal{X}}^\top-\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{U}}_t^\top\right)+\boldsymbol{\mathcal{E}}\right\|$ 进行界定

$$
\begin{aligned}
&||\left(\mathfrak{M}^*\mathfrak{M}-\mathfrak{I}\right)\left(\boldsymbol{\mathcal{X}}*\boldsymbol{\mathcal{X}}^\top-\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{U}}_t^\top\right)+\boldsymbol{\mathcal{E}}||\\
(a)\quad &\le 10\delta\sqrt{rk}\kappa^2\sigma_{\min}^2(\boldsymbol{\mathcal{X}})+\delta\sqrt{k^3}((R\wedge n)-r)||\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{W}}_{t,\perp}||^2+||\boldsymbol{\mathcal{E}}||\\
(b)\quad &\le 10c_1\kappa^{-2}\sigma_{\min}^2(\boldsymbol{\mathcal{X}})+\delta\sqrt{k^3}((R\wedge n)-r)(1+80\eta c_2\sigma_{\min}^2(\boldsymbol{\mathcal{X}}))^{2(t-t_1)}||\boldsymbol{\mathcal{U}}_{t_1}*\boldsymbol{\mathcal{W}}_{t_1,\perp}||^2\\
&\quad +c_1\kappa^{-2}\sigma_{\min}^2(\boldsymbol{\mathcal{X}})\\
(c)\quad &\le 10c_1\kappa^{-2}\sigma_{\min}^2(\boldsymbol{\mathcal{X}})+9\delta\sqrt{k^3}((R\wedge n)-r)(1+80\eta c_2\sigma_{\min}^2(\boldsymbol{\mathcal{X}}))^{2(\hat{t}-t_1)}\gamma^{7/4}\sigma_{\min}(\boldsymbol{\mathcal{X}})^{1/4}\\
&\quad +c_1\kappa^{-2}\sigma_{\min}^2(\boldsymbol{\mathcal{X}})\\
(d)\quad &\le 10c_1\kappa^{-2}\sigma_{\min}^2(\boldsymbol{\mathcal{X}})+9\delta\sqrt{k^3}((R\wedge n)-r)\left(\frac{\kappa^{1/4}}{16k((R\wedge n)-r)}\cdot\frac{||\boldsymbol{\mathcal{X}}||^{7/4}}{\gamma^{7/4}}\right)^{\mathcal{O}(c_2)}\\
&\quad +c_1\kappa^{-2}\sigma_{\min}^2(\boldsymbol{\mathcal{X}})\\
(e)\quad &\le 40c_1\kappa^{-2}\sigma_{\min}^2(\boldsymbol{\mathcal{X}}),
\end{aligned}
\tag{68}
$$

其中，(a) 使用方程 (56) 的结果；(b) 使用归纳假设 (63) 和 $||\boldsymbol{\mathcal{E}}||$ 的假设；(c) 使用 (61) 的结果和归纳假设 (63)；(d) 使用 $\hat{t}$ 的定义；(e) 使用 c_2 足够小这一假设。

因此，定理 E.1（Karnik 等人，2025）中界 (62)、(64) 和 (65) 所需的条件均已满足。由此可调用相应结果，得出不等式 (62)、(64) 和 (65) 在迭代 $t+1$ 时也成立。

注意，根据归纳假设 (62)，$\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top*\boldsymbol{\mathcal{U}}_{t+1}*\boldsymbol{\mathcal{W}}_t$ 的所有奇异值均为正，因此 [（Karnik 等人，2025）引理 E.3] 的假设均已满足。因此，我们使用 [（Karnik 等人，2025）引理 E.3] 的结果来证明归纳假设 (63)，这与证明不等式 (6) 的方法完全相同。我们直接给出结果，不再提供详细证明：

$$
||\boldsymbol{\mathcal{U}}_{t+1}*\boldsymbol{\mathcal{W}}_{t+1}||\le(1+80c_2\eta\sigma_{\min}^2(\boldsymbol{\mathcal{X}}))^{t+1-t_1}||\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{W}}_t||. \tag{69}
$$

接下来我们证明不等式 (67)。注意，由于不等式 (68)，引理 7 中的条件 (76) 得以满足。此外，利用归纳假设 (62)、(64) 和 (65)，引理 7 的其他条件也得以满足。因此，我们有

$$
\begin{aligned}
&||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top*(\boldsymbol{\mathcal{X}}*\boldsymbol{\mathcal{X}}^\top-\boldsymbol{\mathcal{U}}_{t+1}*\boldsymbol{\mathcal{U}}_{t+1}^\top)||\\
&\overset{(a)}{\le}\left(1-\frac{\eta}{200}\sigma_{\min}(\boldsymbol{\mathcal{X}})^2\right)||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top*(\boldsymbol{\mathcal{X}}*\boldsymbol{\mathcal{X}}^\top-\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{U}}_t^\top)||\\
&+\frac{\eta}{100}\sigma_{\min}(\boldsymbol{\mathcal{X}})^2||\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{W}}_{t,\perp}*\boldsymbol{\mathcal{W}}_{t,\perp}^\top*\boldsymbol{\mathcal{U}}_t^\top||+18\eta||\boldsymbol{\mathcal{X}}||^2||\boldsymbol{\mathcal{E}}||\\
&\overset{(b)}{\le}10\left(1-\frac{\eta}{200}\sigma_{\min}(\boldsymbol{\mathcal{X}})^2\right)(1-\frac{\eta}{400}\sigma_{\min}^2(\boldsymbol{\mathcal{X}}))^{t-t_1}||\boldsymbol{\mathcal{X}}||^2\\
&\quad+\frac{\eta}{100}\sigma_{\min}(\boldsymbol{\mathcal{X}})^2||\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{W}}_{t,\perp}*\boldsymbol{\mathcal{W}}_{t,\perp}^\top*\boldsymbol{\mathcal{U}}_t^\top||\\
&\quad+18\eta||\boldsymbol{\mathcal{X}}||^2||\boldsymbol{\mathcal{E}}||\sum_{\tau=t_1+1}^{t+1}(1-\frac{\eta}{200}\sigma_{\min}^2(\boldsymbol{\mathcal{X}}))^{\tau-t_1-1},
\end{aligned}
\tag{70}
$$

其中，步骤 (a) 根据引理 7 的结果得出；步骤 (b) 使用归纳假设 (67)。注意，不等式 (67) 在 $t+1$ 时成立

$$
||\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{W}}_{t,\perp}*\boldsymbol{\mathcal{W}}_{t,\perp}^\top*\boldsymbol{\mathcal{U}}_t||_*\le\frac{1}{4}(1-\frac{\eta}{400}\sigma_{\min}^2(\boldsymbol{\mathcal{X}}))^{t-t_1}||\boldsymbol{\mathcal{X}}||^2. \tag{71}
$$

利用算子范数与管张量核范数之间的关系，我们得到

$$\begin{aligned}||\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp} * \boldsymbol{\mathcal{W}}_{t,\perp}^\top * \boldsymbol{\mathcal{U}}_t||_* &\leq ((R\wedge n)-r)||\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp}||^2 \\ &\overset{(a)}{\leq} k((R\wedge n)-r)(1+80\eta c_2\sigma_{\min}^2(\boldsymbol{\mathcal{X}}))^{2(t-t_1)}||\boldsymbol{\mathcal{U}}_{t_1} * \boldsymbol{\mathcal{W}}_{t_1,\perp}||^2 \\ &\overset{(b)}{\leq} 9k((R\wedge n)-r)(1+80\eta c_2\sigma_{\min}^2(\boldsymbol{\mathcal{X}}))^{2(t-t_1)}\sigma_{\min}(\boldsymbol{\mathcal{X}})^{1/4}\gamma^{7/4}\end{aligned} \tag{72}$$

其中，(a) 使用归纳假设 (63)；(b) 使用不等式 (61)。

接下来，我们需要对项 $||\boldsymbol{\mathcal{U}}_{t_1} * \boldsymbol{\mathcal{W}}_{t_1,\perp}||$ 进行界定。

结合方程 (72) 和 (61)，我们注意到，当 c_2 足够小时，不等式 (71) 成立，并且

$$9k((R\wedge n)-r)\gamma^{7/4}\sigma_{\min}(\boldsymbol{\mathcal{X}})^{1/4} \leq \left(1+\frac{\eta}{350}\sigma_{\min}(\boldsymbol{\mathcal{X}})^2\right)^{t-t_1}||\boldsymbol{\mathcal{X}}||^2$$

利用 $\ln(1+x)\geq \frac{x}{1-x}$ 这一事实，只要 $t \leq \hat{t} \equiv t_1 + \left\lceil \frac{\kappa^{1/4}}{\eta\sigma_{\min}^2(\boldsymbol{\mathcal{X}})\ln\left(\frac{\kappa^{1/4}}{9k}\frac{||\boldsymbol{\mathcal{X}}||^{7/4}}{R\wedge n((()-r)||\boldsymbol{\mathcal{X}}_{\gamma^{7/4}}||^{7/4}}\right)}\right\rceil$，该不等式就成立。因此，我们完成了过秩情形的归纳证明。

接下来，我们证明 $||\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top||_F$ 的上界：

$$\begin{aligned}||\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top||_F &\overset{(a)}{\leq} 4||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * (\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top - \boldsymbol{\mathcal{U}}_{\hat{t}} * \boldsymbol{\mathcal{U}}_{\hat{t}}^\top)||_F + ||\boldsymbol{\mathcal{U}}_{\hat{t}} * \boldsymbol{\mathcal{W}}_{\hat{t},\perp} * \boldsymbol{\mathcal{W}}_{\hat{t},\perp}^\top * \boldsymbol{\mathcal{U}}_{\hat{t}}^\top||_* \\ &\overset{(b)}{\lesssim} \sqrt{r}(1-\frac{\eta}{400}\sigma_{\min}^2(\boldsymbol{\mathcal{X}}))^{\hat{t}-t_1}||\boldsymbol{\mathcal{X}}||^2 \\ &\quad + \sqrt{r}\eta||\boldsymbol{\mathcal{X}}||^2||\boldsymbol{\mathcal{E}}||\sum_{\tau=t_1+1}^{\hat{t}}(1-\frac{\eta}{200}\sigma_{\min}^2(\boldsymbol{\mathcal{X}}))^{\tau-t_1-1} \\ &\overset{(c)}{\lesssim} \sqrt{r}\left(\frac{\kappa^{1/4}}{9k((R\wedge n)-r)}\frac{||\boldsymbol{\mathcal{X}}||^{7/4}}{\gamma^{7/4}}\right)^{-3/4}||\boldsymbol{\mathcal{X}}||^2 + \sqrt{r}\kappa^2||\boldsymbol{\mathcal{E}}|| \\ &\lesssim \kappa^{-3/16}r^{1/2}k^{3/4}((R\wedge n)-r)^{3/4}\gamma^{21/16}||\boldsymbol{\mathcal{X}}||^{11/16} + +\sqrt{r}\kappa^2||\boldsymbol{\mathcal{E}}||,\end{aligned} \tag{73}$$

其中，(a) 用到了引理 8 的结果；(b) 由不等式 (67) 和 (71) 得出；(c) 用到了 $\hat{t}$ 的定义。

精确秩情形： 由于 $R=r$，我们有 $\boldsymbol{\mathcal{U}}_t = \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t * \boldsymbol{\mathcal{W}}_t^\top$ 以及 $\boldsymbol{\mathcal{W}}_{t,\perp}=0$。采用与过参数化情形类似的方法，可以证明归纳假设 (62)-(65) 对所有 $t\geq t_1$ 均成立。对于归纳假设 (67)，注意到

$$\boldsymbol{\mathcal{U}}_t\boldsymbol{\mathcal{W}}_{t,\perp}\boldsymbol{\mathcal{W}}_{t,\perp}^\top\boldsymbol{\mathcal{W}}_t^\top = 0,$$

这意味着 (67) 对所有 $t\geq t_1$ 也成立。因此，我们得出：

$$\begin{aligned}||\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top - \boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top||_F &\lesssim \sqrt{r}(1-\frac{\eta}{400}\sigma_{\min}^2(\boldsymbol{\mathcal{X}}))^{t-t_1} \\ &\quad + \sqrt{r}\eta||\boldsymbol{\mathcal{X}}||^2||\boldsymbol{\mathcal{E}}||\sum_{\tau=t_1+1}^{t+1}(1-\frac{\eta}{200}\sigma_{\min}^2(\boldsymbol{\mathcal{X}}))^{\tau-t_1-1} \\ &\lesssim \sqrt{r}(1-\frac{\eta}{400}\sigma_{\min}^2(\boldsymbol{\mathcal{X}}))^{t-t_1} + \sqrt{r}\kappa^2||\boldsymbol{\mathcal{E}}||.\end{aligned} \tag{74}$$

E.8 引理 4 的证明

引理 7. 假设满足以下假设：

$$\begin{aligned}||\boldsymbol{\mathcal{U}}_t|| &\leq 3||\boldsymbol{\mathcal{X}}|| \\ \eta &\leq c\kappa^{-2}||\boldsymbol{\mathcal{X}}||^{-2} \\ \sigma_{\min}(\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t) &\geq \frac{1}{\sqrt{10}}\sigma_{\min}(\boldsymbol{\mathcal{X}}) \\ ||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}^\perp}^\top * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{W}}_t}|| &\leq c\kappa^{-2}\end{aligned} \tag{75}$$

and

$$||(\mathfrak{I}-\mathfrak{M}*\mathfrak{M})(\boldsymbol{\mathcal{X}}*\boldsymbol{\mathcal{X}}^\top-\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{U}}_t)|| \\ \leq c\kappa^{-2}\left(||\boldsymbol{\mathcal{X}}*\boldsymbol{\mathcal{X}}-\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{W}}_t*\boldsymbol{\mathcal{W}}_t^\top*\boldsymbol{\mathcal{U}}_t^\top||+||\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{W}}_{t,\perp}*\boldsymbol{\mathcal{W}}_{t,\perp}^\top*\boldsymbol{\mathcal{U}}_t||_*\right), \tag{76}$$

其中常数 $c>0$ 取得足够小。则有

$$||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top*(\boldsymbol{\mathcal{X}}*\boldsymbol{\mathcal{X}}^\top-\boldsymbol{\mathcal{U}}_{t+1}*\boldsymbol{\mathcal{U}}_{t+1}^\top)|| \leq \left(1-\frac{\eta}{200}\sigma_{\min}(\boldsymbol{\mathcal{X}})^2\right)||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top*(\boldsymbol{\mathcal{X}}*\boldsymbol{\mathcal{X}}^\top-\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{U}}_t^\top)|| \\ +\frac{\eta}{100}\sigma_{\min}(\boldsymbol{\mathcal{X}})^2||\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{W}}_{t,\perp}*\boldsymbol{\mathcal{W}}_{t,\perp}^\top*\boldsymbol{\mathcal{U}}_t^\top||+18\eta||\boldsymbol{\mathcal{X}}||^2||\boldsymbol{\mathcal{E}}||. \tag{77}$$

引理 7 的证明。 为了证明引理 4，我们首先引入一个关键的辅助引理，并给出其证明。

引理 8。 在引理 *4* 的假设下，以下不等式成立：

$$\left|\left|\left|\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}^\perp}^\top*\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{U}}_t^\top\right|\right|\right| \leq 3\left|\left|\left|\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}^\perp}^\top*(\boldsymbol{\mathcal{X}}*\boldsymbol{\mathcal{X}}^\top-\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{U}}_t^\top)\right|\right|\right|+\left|\left|\left|\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{W}}_{t,\perp}*\boldsymbol{\mathcal{W}}_{t,\perp}^\top*\boldsymbol{\mathcal{U}}_t^\top\right|\right|\right| \\ \left|\left|\left|\boldsymbol{\mathcal{X}}*\boldsymbol{\mathcal{X}}^\top-\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{U}}_t^\top\right|\right|\right| \leq 4\left|\left|\left|\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}^\perp}^\top*(\boldsymbol{\mathcal{X}}*\boldsymbol{\mathcal{X}}^\top-\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{U}}_t^\top)\right|\right|\right|+\left|\left|\left|\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{W}}_{t,\perp}*\boldsymbol{\mathcal{W}}_{t,\perp}^\top*\boldsymbol{\mathcal{U}}_t^\top\right|\right|\right| \\ \left|\left|\left|\boldsymbol{\mathcal{X}}*\boldsymbol{\mathcal{X}}^\top-\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{W}}_t*\boldsymbol{\mathcal{W}}_t^\top*\boldsymbol{\mathcal{U}}_t^\top\right|\right|\right| \leq 4\left|\left|\left|\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}^\perp}^\top*(\boldsymbol{\mathcal{X}}*\boldsymbol{\mathcal{X}}^\top-\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{U}}_t^\top)\right|\right|\right|, \tag{78}$$

其中 $|||\cdot|||$表示张量范数，例如谱范数。

证明。 前两个不等式是根据 (Karnik 等人，2025) 中的引理 E.7 推导得出的。利用矩阵范数之间的等价性，将引理 E.7 中的弗罗贝尼乌斯范数替换为谱范数，即可得到所需结果。

接下来，我们给出第三个不等式的证明。我们将 $\boldsymbol{\mathcal{X}}*\boldsymbol{\mathcal{X}}^\top-\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{W}}_t*\boldsymbol{\mathcal{W}}_t^\top*\boldsymbol{\mathcal{U}}_t^\top$分解为

$$\underbrace{\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}*\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top*(\boldsymbol{\mathcal{X}}*\boldsymbol{\mathcal{X}}^\top-\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{W}}_t*\boldsymbol{\mathcal{W}}_t^\top*\boldsymbol{\mathcal{U}}_t^\top)}_{\boldsymbol{\mathcal{Z}}_1}+\underbrace{\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}^\perp}*\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}^\perp}^\top*(\boldsymbol{\mathcal{X}}*\boldsymbol{\mathcal{X}}^\top-\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{W}}_t*\boldsymbol{\mathcal{W}}_t^\top*\boldsymbol{\mathcal{U}}_t^\top)}_{\boldsymbol{\mathcal{Z}}_2}.$$

对于$\boldsymbol{\mathcal{Z}}_1$,，我们有

$$\begin{aligned}\boldsymbol{\mathcal{Z}}_1&=\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}*\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top*\boldsymbol{\mathcal{X}}*\boldsymbol{\mathcal{X}}^\top-\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}*\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top*\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{W}}_t*\boldsymbol{\mathcal{W}}_t^\top*\boldsymbol{\mathcal{U}}_t^\top \\ &\overset{(1)}{=}\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}*\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top*\boldsymbol{\mathcal{X}}*\boldsymbol{\mathcal{X}}^\top-\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}*\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top*\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{U}}_t^\top,\end{aligned} \tag{79}$$

其中，式（1）用到了这一事实：

$$\begin{aligned}&\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}*\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top*\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{U}}_t^\top \\ &\quad=\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}*\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top*[\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{W}}_t*\boldsymbol{\mathcal{W}}_t^\top+\boldsymbol{\mathcal{W}}_{t,\perp}*\boldsymbol{\mathcal{W}}_{t,\perp}^\top]*[\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{W}}_t*\boldsymbol{\mathcal{W}}_t^\top+\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{W}}_{t,\perp}*\boldsymbol{\mathcal{W}}_{t,\perp}^\top]^\top \\ &\quad=\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}*\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top*\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{W}}_t*\boldsymbol{\mathcal{W}}_t^\top*\boldsymbol{\mathcal{U}}_t^\top+\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}*\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top*\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{W}}_{t,\perp}*\boldsymbol{\mathcal{W}}_{t,\perp}^\top*\boldsymbol{\mathcal{U}}_t^\top \\ &\quad=\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}*\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top*\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{W}}_t*\boldsymbol{\mathcal{W}}_t^\top*\boldsymbol{\mathcal{U}}_t^\top+\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}*\boldsymbol{\mathcal{V}}_t*\boldsymbol{\mathcal{S}}_t*\boldsymbol{\mathcal{W}}_t^\top*\boldsymbol{\mathcal{W}}_{t,\perp}*\boldsymbol{\mathcal{W}}_{t,\perp}^\top*\boldsymbol{\mathcal{U}}_t^\top \\ &\quad=\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}*\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top*\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{W}}_t*\boldsymbol{\mathcal{W}}_t^\top*\boldsymbol{\mathcal{U}}_t^\top.\end{aligned} \tag{80}$$

因此，我们有

$$|||\boldsymbol{\mathcal{Z}}_1|||=\left|\left|\left|\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}*\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top*\boldsymbol{\mathcal{X}}*\boldsymbol{\mathcal{X}}^\top-\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}*\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top*\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{U}}_t^\top\right|\right|\right| \leq \left|\left|\left|\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top*(\boldsymbol{\mathcal{X}}*\boldsymbol{\mathcal{X}}^\top-\boldsymbol{\mathcal{U}}_t*\boldsymbol{\mathcal{U}}_t^\top)\right|\right|\right|.$$

接下来，我们对该项$\mathcal{Z}_2$,进行界定

$$\begin{aligned}\|\mathcal{Z}_2\| &= \left\|\mathcal{V}_{\mathcal{X}^\perp} * \mathcal{V}_{\mathcal{X}^\perp}^\top * (\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{W}_t * \mathcal{W}_t^\top * \mathcal{U}_t^\top) * (\mathcal{V}_{\mathcal{X}} * \mathcal{V}_{\mathcal{X}}^\top + \mathcal{V}_{\mathcal{X}^\perp} * \mathcal{V}_{\mathcal{X}^\perp}^\top)\right\| \\ &\le \left\|\mathcal{V}_{\mathcal{X}^\perp} * \mathcal{V}_{\mathcal{X}^\perp}^\top * (\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{W}_t * \mathcal{W}_t^\top * \mathcal{U}_t^\top) * \mathcal{V}_{\mathcal{X}}\right\| \\ &\quad + \left\|\mathcal{V}_{\mathcal{X}^\perp} * \mathcal{V}_{\mathcal{X}^\perp}^\top * (\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{W}_t * \mathcal{W}_t^\top * \mathcal{U}_t^\top) * \mathcal{V}_{\mathcal{X}^\perp} * \mathcal{V}_{\mathcal{X}^\perp}^\top\right\| \\ &\overset{(a)}{\le} \left\|(\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top) * \mathcal{V}_{\mathcal{X}}\right\| + \underbrace{\left\|\mathcal{V}_{\mathcal{X}^\perp}^\top * \mathcal{U}_t * \mathcal{W}_t * \mathcal{W}_t^\top * \mathcal{U}_t^\top * \mathcal{V}_{\mathcal{X}^\perp}\right\|}_{\mathcal{Z}_3},\end{aligned} \tag{81}$$

其中，(a) 利用了以下事实：$\mathcal{V}_{\mathcal{X}}^\top * \mathcal{U}_t * \mathcal{W}_{t,\perp} = 0$以及$\mathcal{V}_{\mathcal{X}}^\top * \mathcal{U}_t * \mathcal{U}_t^\top = \mathcal{V}_{\mathcal{X}}^\top * \mathcal{U}_t * \mathcal{W}_t * \mathcal{W}_t^\top * \mathcal{U}_t^\top$。对于项$\mathcal{Z}_3$,，我们有

$$\begin{aligned}&\left\|\mathcal{V}_{\mathcal{X}^\perp}^\top * \mathcal{U}_t * \mathcal{W}_t * \mathcal{W}_t^\top * \mathcal{U}_t^\top * \mathcal{V}_{\mathcal{X}^\perp}\right\| \\ &= \left\|\mathcal{V}_{\mathcal{X}^\perp}^\top * \mathcal{V}_{\mathcal{U}_t * \mathcal{W}_t} * \mathcal{V}_{\mathcal{U}_t * \mathcal{W}_t}^\top * \mathcal{U}_t * \mathcal{W}_t * \mathcal{W}_t^\top * \mathcal{U}_t^\top * \mathcal{V}_{\mathcal{X}^\perp}\right\| \\ &= \left\|\mathcal{V}_{\mathcal{X}^\perp}^\top * \mathcal{V}_{\mathcal{U}_t * \mathcal{W}_t} * \left(\mathcal{V}_{\mathcal{X}}^\top * \mathcal{V}_{\mathcal{U}_t * \mathcal{W}_t}^\top\right)^{-1} * \mathcal{V}_{\mathcal{X}}^\top * \mathcal{V}_{\mathcal{U}_t * \mathcal{W}_t} * \mathcal{V}_{\mathcal{U}_t * \mathcal{W}_t}^\top * \mathcal{U}_t * \mathcal{W}_t * \mathcal{W}_t^\top * \mathcal{U}_t^\top * \mathcal{V}_{\mathcal{X}^\perp}\right\| \\ &\le ||\mathcal{V}_{\mathcal{X}^\perp}^\top * \mathcal{V}_{\mathcal{U}_t * \mathcal{W}_t}|| \cdot ||(\mathcal{V}_{\mathcal{X}}^\top * \mathcal{V}_{\mathcal{U}_t * \mathcal{W}_t})^{-1}|| \left\|\mathcal{V}_{\mathcal{X}}^\top * \mathcal{V}_{\mathcal{U}_t * \mathcal{W}_t} * \mathcal{V}_{\mathcal{U}_t * \mathcal{W}_t}^\top * \mathcal{U}_t * \mathcal{W}_t * \mathcal{W}_t^\top * \mathcal{U}_t^\top * \mathcal{V}_{\mathcal{X}^\perp}\right\| \\ &= \frac{||\mathcal{V}_{\mathcal{X}^\perp}^\top * \mathcal{V}_{\mathcal{U}_t * \mathcal{W}_t}||}{\sigma_{\min}(\mathcal{V}_{\mathcal{X}}^\top * \mathcal{V}_{\mathcal{U}_t * \mathcal{W}_t})} \left\|\mathcal{V}_{\mathcal{X}}^\top * \mathcal{U}_t * \mathcal{U}_t^\top * \mathcal{V}_{\mathcal{X}^\perp}\right\| \\ &= \frac{||\mathcal{V}_{\mathcal{X}^\perp}^\top * \mathcal{V}_{\mathcal{U}_t * \mathcal{W}_t}||}{\sigma_{\min}(\mathcal{V}_{\mathcal{X}}^\top * \mathcal{V}_{\mathcal{U}_t * \mathcal{W}_t})} \left\|\mathcal{V}_{\mathcal{X}}^\top * \left(\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top\right) * \mathcal{V}_{\mathcal{X}^\perp}\right\| \\ &\le 2\left\|\mathcal{V}_{\mathcal{X}}^\top * \left(\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top\right)\right\|.\end{aligned} \tag{82}$$

因此，第三个不等式成立。 □

基于引理 8 的结果，我们接着证明引理 7。我们利用更新公式将$\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_{t+1} * \mathcal{U}_{t+1}^\top$分解为五项

$$\mathcal{U}_{t+1} = \mathcal{U}_t + \eta[(\mathfrak{M}^*\mathfrak{M})(\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top) + \mathcal{E}] * \mathcal{U}_t :$$

$$\begin{aligned}&\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_{t+1} * \mathcal{U}_{t+1}^\top \\ &= \underbrace{(\mathcal{I} - \eta\mathcal{U}_t * \mathcal{U}_t^\top) * (\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top) * (\mathcal{I} - \eta\mathcal{U}_t * \mathcal{U}_t^\top)}_{\mathcal{K}_1} \\ &+ \eta\underbrace{[(\mathfrak{I} - \mathfrak{M}^*\mathfrak{M})(\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top) + \mathcal{E}] * \mathcal{U}_t * \mathcal{U}_t^\top}_{\mathcal{K}_2} \\ &+ \eta\underbrace{\mathcal{U}_t * \mathcal{U}_t * [(\mathfrak{I} - \mathfrak{M}^*\mathfrak{M})(\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top) + \mathcal{E}]}_{\mathcal{K}_3} \\ &- \eta^2\underbrace{\mathcal{U}_t * \mathcal{U}_t^\top * (\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top) * \mathcal{U}_t * \mathcal{U}_t^\top}_{\mathcal{K}_4} \\ &- \eta^2\underbrace{[(\mathfrak{M}^*\mathfrak{M})(\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top) + \mathcal{E}] * \mathcal{U}_t * \mathcal{U}_t^\top * [(\mathfrak{M}^*\mathfrak{M})(\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top) + \mathcal{E}]}_{\mathcal{K}_5}.\end{aligned} \tag{83}$$

我们现在分别界定这些项。**界定$\mathcal{K}_1$：**我们注意
到

$$\begin{aligned}
&\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * (\boldsymbol{\mathcal{I}} - \eta \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top) * (\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top) * (\boldsymbol{\mathcal{I}} - \eta \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top) \\
&= \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * (\boldsymbol{\mathcal{I}} - \eta \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top) * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}} * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * (\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top) * (\boldsymbol{\mathcal{I}} - \eta \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top) \\
&\quad + \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * (\boldsymbol{\mathcal{I}} - \eta \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top) * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}^\perp} * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}^\perp}^\top * (\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top) * (\boldsymbol{\mathcal{I}} - \eta \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top) \\
&= \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * (\boldsymbol{\mathcal{I}} - \eta \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top) * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}} * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * (\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top) * (\boldsymbol{\mathcal{I}} - \eta \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top) \\
&\quad + \eta * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}^\perp} * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}^\perp}^\top * \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top * (\boldsymbol{\mathcal{I}} - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top) \\
&= (\boldsymbol{\mathcal{I}} - \eta \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}) * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * (\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top) * (\boldsymbol{\mathcal{I}} - \eta \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top) \\
&\quad + \eta * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}^\perp} * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}^\perp}^\top * \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top * (\boldsymbol{\mathcal{I}} - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top).
\end{aligned} \tag{84}$$

细节
因此，我们得到

$$\begin{aligned}
||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * \boldsymbol{\mathcal{K}}_1|| &= ||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * (\boldsymbol{\mathcal{I}} - \eta \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top) * (\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top) * (\boldsymbol{\mathcal{I}} - \eta \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top)|| \\
&\leq \left(1 - \frac{\eta}{40}\sigma_{\min}^2(\boldsymbol{\mathcal{X}})\right) ||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * (\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top)|| \\
&\quad + \eta \frac{\sigma_{\min}^2(\boldsymbol{\mathcal{X}})}{400} ||\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp} * \boldsymbol{\mathcal{W}}_{t,\perp}^\top * \boldsymbol{\mathcal{U}}_t||.
\end{aligned} \tag{85}$$

界定$\mathcal{K}_2$：注意

$$\begin{aligned}
||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * \boldsymbol{\mathcal{K}}_2|| &= ||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * [(\boldsymbol{\mathcal{I}} - \boldsymbol{\mathfrak{M}}^* \boldsymbol{\mathfrak{M}})(\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top) + \boldsymbol{\mathcal{E}}] * \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top|| \\
&\leq \left(||(\boldsymbol{\mathcal{I}} - \boldsymbol{\mathfrak{M}}^* \boldsymbol{\mathfrak{M}})(\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top)|| + ||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * \boldsymbol{\mathcal{E}}||\right) ||\boldsymbol{\mathcal{U}}_t||^2 \\
&\overset{(1)}{\leq} 9\left(||(\boldsymbol{\mathcal{I}} - \boldsymbol{\mathfrak{M}}^* \boldsymbol{\mathfrak{M}})(\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top)|| + ||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * \boldsymbol{\mathcal{E}}||\right) ||\boldsymbol{\mathcal{X}}||^2 \\
&\overset{(2)}{\leq} 9c\sigma_{\min}^2(\boldsymbol{\mathcal{U}}) \left(||\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_t * \boldsymbol{\mathcal{W}}_t^\top * \boldsymbol{\mathcal{U}}_t^\top|| + ||\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp} * \boldsymbol{\mathcal{W}}_{t,\perp}^\top * \boldsymbol{\mathcal{U}}_t||_*\right) \\
&\quad + 9||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * \boldsymbol{\mathcal{E}}||||\boldsymbol{\mathcal{X}}||^2 \\
&\overset{(3)}{\leq} 9c\sigma_{\min}^2(\boldsymbol{\mathcal{U}}) \left(||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * (\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top)|| + ||\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp} * \boldsymbol{\mathcal{W}}_{t,\perp}^\top * \boldsymbol{\mathcal{U}}_t||_*\right) \\
&\quad + 9||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * \boldsymbol{\mathcal{E}}||||\boldsymbol{\mathcal{X}}||^2
\end{aligned} \tag{86}$$

其中，(1) 使用假设 $|\boldsymbol{\mathcal{U}}_t|| \leq 3||\boldsymbol{\mathcal{X}}||$；(2) 使用假设 (76)；(3) 使用引理 8 的结果。取一个较小的常数 $c > 0$，我们得到

$$\begin{aligned}
&||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * [(\boldsymbol{\mathcal{I}} - \boldsymbol{\mathfrak{M}}^* \boldsymbol{\mathfrak{M}})(\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top) + \boldsymbol{\mathcal{E}}] * \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top|| \\
&\leq \frac{1}{1000}\sigma_{\min}^2(\boldsymbol{\mathcal{X}}) \left(||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * (\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top)|| + ||\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp} * \boldsymbol{\mathcal{W}}_{t,\perp}^\top * \boldsymbol{\mathcal{U}}_t||_*\right) \\
&\quad + 9||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * \boldsymbol{\mathcal{E}}||||\boldsymbol{\mathcal{X}}||^2.
\end{aligned} \tag{87}$$

界定$\mathcal{K}_3$： 与$\mathcal{K}_2$, 类似，我们有

$$\begin{aligned}
&||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t * [(\boldsymbol{\mathcal{I}} - \boldsymbol{\mathfrak{M}}^* \boldsymbol{\mathfrak{M}})(\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^\top - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top) + \boldsymbol{\mathcal{E}}]|| \\
&\leq \frac{1}{1000}\sigma_{\min}^2(\boldsymbol{\mathcal{X}}) \left(||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * (\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^\top)|| + ||\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp} * \boldsymbol{\mathcal{W}}_{t,\perp}^\top * \boldsymbol{\mathcal{U}}_t||_*\right) \\
&\quad + 9||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^\top * \boldsymbol{\mathcal{E}}||||\boldsymbol{\mathcal{X}}||^2.
\end{aligned} \tag{88}$$

界定$\mathcal{K}_4$： 注意

$$\begin{aligned}
&||\mathcal{V}_{\mathcal{X}}^\top * \mathcal{U}_t * \mathcal{U}_t^\top * (\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top) * \mathcal{U}_t * \mathcal{U}_t^\top|| \\
&\qquad \leq ||\mathcal{U}_t||^4 ||\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top|| \\
&\qquad \overset{(1)}{\lesssim} ||\mathcal{X}||^4 ||\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top|| \\
&\qquad \overset{(2)}{\lesssim} ||\mathcal{X}||^4 \left(||\mathcal{V}_{\mathcal{X}}^\top * (\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top)|| + ||\mathcal{U}_t * \mathcal{W}_{t,\perp} * \mathcal{W}_{t,\perp}^\top * \mathcal{U}_t^\top|| \right),
\end{aligned} \tag{89}$$

其中，(1) 使用假设 $||\mathcal{U}_t|| \leq 3||\mathcal{X}||$；(2) 使用引理 8 的结果。然后结合假设 $\eta \leq c\kappa^{-2}||\mathcal{X}||^{-2}$，可得

$$\begin{aligned}
&\eta^2 ||\mathcal{V}_{\mathcal{X}}^\top * \mathcal{U}_t * \mathcal{U}_t^\top * (\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top) * \mathcal{U}_t * \mathcal{U}_t^\top|| \\
&\leq \frac{\eta}{200}\sigma_{\min}^2(\mathcal{X}) ||\mathcal{V}_{\mathcal{X}}^\top * (\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top)|| + \eta \frac{\sigma_{\min}^2(\mathcal{X})}{1000} ||\mathcal{U}_t * \mathcal{W}_{t,\perp} * \mathcal{W}_{t,\perp}^\top * \mathcal{U}_t^\top||.
\end{aligned} \tag{90}$$

界定$\mathcal{K}_5$： 注意

$$\begin{aligned}
&||(\mathfrak{M}^* \mathfrak{M})(\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top)|| \\
&\qquad \leq ||\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * (\mathcal{W}_t * \mathcal{W}_t^\top + \mathcal{W}_{t,\perp} * \mathcal{W}_{t,\perp}^\top) * \mathcal{U}_t^\top|| \\
&\qquad\quad + ||(\mathfrak{M}^* \mathfrak{M} - \mathfrak{I})(\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top)|| \\
&\qquad \overset{(a)}{\leq} \left(||\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{W}_t * \mathcal{W}_t^\top * \mathcal{U}_t^\top|| + ||\mathcal{U}_t * \mathcal{W}_{t,\perp} * \mathcal{W}_{t,\perp}^\top \mathcal{U}_t^\top|| \right) \\
&\qquad\quad + c\kappa^{-2} \left(||\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{W}_t * \mathcal{W}_t^\top * \mathcal{U}_t^\top|| + ||\mathcal{U}_t * \mathcal{W}_{t,\perp} * \mathcal{W}_{t,\perp}^\top \mathcal{U}_t^\top||_* \right) \\
&\qquad \leq 2 \left(||\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{W}_t * \mathcal{W}_t^\top * \mathcal{U}_t^\top|| + ||\mathcal{U}_t * \mathcal{W}_{t,\perp} * \mathcal{W}_{t,\perp}^\top \mathcal{U}_t^\top||_* \right) \\
&\qquad \leq 2 \left(||\mathcal{X}||^2 + ||\mathcal{U}_t * \mathcal{W}_t||^2 + ||\mathcal{U}_t * \mathcal{W}_{t,\perp} * \mathcal{W}_{t,\perp}^\top \mathcal{U}_t^\top||_* \right) \\
&\qquad \overset{(b)}{\leq} 2 \left(||\mathcal{X}||^2 + 2||\mathcal{U}_t||^2 \right),
\end{aligned} \tag{91}$$

其中，(a) 使用假设 (76)；(b) 使用假设 $||\mathcal{U}_t * \mathcal{W}_{t,\perp} * \mathcal{W}_{t,\perp}^\top \mathcal{U}_t^\top|| \leq ||\mathcal{U}_t||^2$。于是有

$$\begin{aligned}
&||\mathcal{V}_{\mathcal{X}}^\top * \mathcal{K}_5|| = ||\mathcal{V}_{\mathcal{X}}^\top * [(\mathfrak{M}^* \mathfrak{M})(\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top) + \mathcal{E}] * \mathcal{U}_t * \mathcal{U}_t^\top * [(\mathfrak{M}^* \mathfrak{M})(\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top) + \mathcal{E}]|| \\
&\leq \left(||(\mathfrak{M}^* \mathfrak{M})(\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top)|| + ||\mathcal{E}|| \right) \cdot ||\mathcal{U}_t||^2 \cdot \left(||(\mathfrak{M}^* \mathfrak{M})(\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top)|| + ||\mathcal{E}|| \right) \\
&\overset{(a)}{\leq} 4 \left(||\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{W}_t * \mathcal{W}_t^\top * \mathcal{U}_t^\top|| + ||\mathcal{U}_t * \mathcal{W}_{t,\perp} * \mathcal{W}_{t,\perp}^\top * \mathcal{U}_t^\top||_* + ||\mathcal{E}|| \right) ||\mathcal{U}_t||^2 \left(||\mathcal{X}||^2 + 2||\mathcal{U}_t||^2 + ||\mathcal{E}|| \right) \\
&\overset{(b)}{\leq} 432 \left(||\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{W}_t * \mathcal{W}_t^\top * \mathcal{U}_t^\top|| + ||\mathcal{U}_t * \mathcal{W}_{t,\perp} * \mathcal{W}_{t,\perp}^\top * \mathcal{U}_t^\top||_* + ||\mathcal{E}|| \right) ||\mathcal{U}_t||^4 \\
&\overset{(c)}{\leq} 1728 \left(||\mathcal{V}_{\mathcal{X}}^\top * (\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top)|| + ||\mathcal{U}_t * \mathcal{W}_{t,\perp} * \mathcal{W}_{t,\perp}^\top * \mathcal{U}_t^\top||_* + ||\mathcal{E}|| \right) ||\mathcal{U}_t||^4,
\end{aligned} \tag{92}$$

其中，(a) 使用了方程（91 $||\mathcal{U}_t|| \leq 3||\mathcal{X}||$ $\eta \leq c\kappa^{-2}||\mathcal{X}||^{-2}$ 8 $||\mathcal{E}||$ $||\mathcal{U}_t||^2$ $\mathcal{U}\mathcal{X}$ $\mathcal{X}$的结果；(b) 使用了假设和；(c) 使用了引理的结果。基于这些结果和假设，我们得到

$$\begin{aligned}
&\eta^2 ||\mathcal{V}_{\mathcal{X}}^\top * [(\mathfrak{M}^* \mathfrak{M})(\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top) + \mathcal{E}] * \mathcal{U}_t * \mathcal{U}_t^\top * [(\mathfrak{M}^* \mathfrak{M})(\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top) + \mathcal{E}]|| \\
&\leq \frac{\eta}{1000}\sigma_{\min}^2(\mathcal{X}) ||\mathcal{V}_{\mathcal{X}}^\top * (\mathcal{X} * \mathcal{X}^\top - \mathcal{U}_t * \mathcal{U}_t^\top)|| + \frac{\eta}{400}\sigma_{\min}^2(\mathcal{X}) ||\mathcal{U}_t * \mathcal{W}_{t,\perp} * \mathcal{W}_{t,\perp}^\top * \mathcal{U}_t^\top||_* \\
&\quad + \frac{\eta}{400}\sigma_{\min}^2(\mathcal{X}) ||\mathcal{E}||.
\end{aligned} \tag{93}$$

结合这五项的界，我们得到

$$||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^{\top} * (\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^{\top} - \boldsymbol{\mathcal{U}}_{t+1} * \boldsymbol{\mathcal{U}}_{t+1}^{\top})|| \leq \left(1 - \frac{\eta}{200}\sigma_{\min}(\boldsymbol{\mathcal{X}})^2\right) ||\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}^{\top} * (\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{X}}^{\top} - \boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{U}}_t^{\top})|| + \frac{\eta}{200}\sigma_{\min}(\boldsymbol{\mathcal{X}})^2 ||\boldsymbol{\mathcal{U}}_t * \boldsymbol{\mathcal{W}}_{t,\perp} * \boldsymbol{\mathcal{W}}_{t,\perp}^{\top} * \boldsymbol{\mathcal{U}}_t^{\top}|| + 18\eta ||\boldsymbol{\mathcal{X}}||^2 ||\boldsymbol{\mathcal{E}}||. \tag{94}$$

□

F 定理 3 的证明

低管秩张量恢复的极小极大误差界的证明沿用矩阵情形的证明（Candes & Plan,2011）。我们从一个标准引理开始，它刻画了在线性模型中估计向量 $x \in \mathbb{R}^n$ 的极小极大风险

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{s} \tag{95}$$

其中 $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ 和 $\boldsymbol{s}$ 的各项独立同分布，服从高斯分布 $\mathcal{N}(0,\ \sigma^2)$。对于此类模型，我们有以下引理给出其极小极大误差界。

引理 9。 设 $\lambda_i(\boldsymbol{A}^{\top}\boldsymbol{A})$ 为矩阵 $\boldsymbol{A}^{\top}\boldsymbol{A}$的特征值。那么

$$\inf_{\hat{\boldsymbol{x}}} \sup_{\boldsymbol{x} \in \mathbb{R}^n} \mathbb{E}||\hat{\boldsymbol{x}} - \boldsymbol{x}||_{l_2}^2 = \sigma^2 \operatorname{trace}\left((\boldsymbol{A}^{\top}\boldsymbol{A})^{-1}\right) = \sum_i \frac{\sigma^2}{\lambda_i(\boldsymbol{A}^{\top}\boldsymbol{A})}. \tag{96}$$

特别地，如果其中一个特征值为零，则极小极大风险无界。

证明。 我们将论证分为三部分：(A) 通过贝叶斯风险得到下界；(B) 在非奇异情形下，由普通最小二乘估计量达到上界。

(A) 下界（贝叶斯论证）
固定 $\tau > 0$，并考虑高斯先验分布 $\boldsymbol{x} \sim \mathcal{N}(0, \tau^2 I_n)$。在该先验分布下，给定 $\boldsymbol{y}$ 时 $\boldsymbol{x}$ 的后验协方差矩阵为

$$\boldsymbol{\Sigma}_{\text{post}}(\tau) = \left(\frac{1}{\sigma^2}\boldsymbol{A}^{\top}\boldsymbol{A} + \frac{1}{\tau^2}I_n\right)^{-1}.$$

（对于后验均值估计量而言的）贝叶斯风险等于后验协方差矩阵的迹 e:

$$\begin{aligned} R_{\text{Bayes}}(\tau) &:= \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}} \left\|\boldsymbol{x} - \mathbb{E}[\boldsymbol{x} \mid \boldsymbol{y}]\right\|_2^2 = \operatorname{tr}\left(\Sigma_{\text{post}}(\tau)\right) \\ &= \sum_{i=1}^{n} \frac{1}{\lambda_i/\sigma^2 + 1/\tau^2} = \sigma^2 \sum_{i=1}^{n} \frac{1}{\lambda_i + \sigma^2/\tau^2}, \end{aligned} \tag{97}$$

为方便起见，我们将 $\lambda_i(\boldsymbol{A}^{\top}\boldsymbol{A})$ 记作 λ_i。
对于任意估计量 $\hat{\boldsymbol{x}}$ 和任意先验分布 π，都有标准的极小极大/贝叶斯不等式

$$\inf_{\hat{\boldsymbol{x}}} \sup_{\boldsymbol{x}} \mathbb{E}\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|^2 \geq \inf_{\hat{\boldsymbol{x}}} \mathbb{E}_{\boldsymbol{x} \sim \pi} \mathbb{E}\left[\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|^2\right] = R_{\text{Bayes}}(\tau),$$

因为对 $\boldsymbol{x}$ 取上确界至少不小于任意先验分布 π 下的平均值。因此，对于每个 $\tau > 0$,

$$\inf_{\hat{\boldsymbol{x}}} \sup_{\boldsymbol{x}} \mathbb{E}\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|^2 \geq \sigma^2 \sum_{i=1}^{n} \frac{1}{\lambda_i + \sigma^2/\tau^2}.$$

如果某些 $\lambda_i = 0$，那么右侧等于 $+\infty$，因为 $\tau \to \infty$ （事实上，对应的求和项为 $\sigma^2/$（ $0 + \sigma^2/\tau^2$）$= \tau^2 \to \infty$），因此此时极小极大风险为无穷大。否则，如果所有 $\lambda_i > 0$,

发送 $\tau \to \infty$。对于每个固定的 i，函数 $\tau \mapsto \sigma^2/(\lambda_i + \sigma^2/\tau^2)$ 关于 τ 单调递增，并在 $\tau \to \infty$ 时收敛到 σ^2/λ_i。由单调收敛定理（或有限和的连续性）可得

$$\inf_{\hat{\boldsymbol{x}}} \sup_{\boldsymbol{x}} \mathbb{E}\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|^2 \geq \lim_{\tau \to \infty} R_{\text{Bayes}}(\tau) = \sigma^2 \sum_{i=1}^{n} \frac{1}{\lambda_i}.$$

(B) 上界（最小二乘法达到该上界）。
假设 $\lambda_i > 0$ 对所有 i都成立，即 $\text{rank}(A) = n$。考虑普通最小二乘估计量

$$\hat{\boldsymbol{x}}_{\text{LS}} = (\boldsymbol{A}^\top \boldsymbol{A})^{-1} \boldsymbol{A}^\top \boldsymbol{y}.$$

代入 $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{s}$ 得

$$\hat{\boldsymbol{x}}_{\text{LS}} - \boldsymbol{x} = (\boldsymbol{A}^\top \boldsymbol{A})^{-1} \boldsymbol{A}^\top \boldsymbol{s}.$$

由于 $\boldsymbol{s} \sim \mathcal{N}(0, \sigma^2 I_m)$，误差 $\hat{\boldsymbol{x}}_{\text{LS}} - \boldsymbol{x}$ 是均值为零的高斯变量，其协方差为

$$\mathbb{E}\big[(\hat{\boldsymbol{x}}_{\text{LS}} - \boldsymbol{x})(\hat{\boldsymbol{x}}_{\text{LS}} - \boldsymbol{x})^\top\big] = (\boldsymbol{A}^\top \boldsymbol{A})^{-1} \boldsymbol{A}^\top (\sigma^2 \boldsymbol{I_m}) \boldsymbol{A} (\boldsymbol{A}^\top \boldsymbol{A})^{-1} = \sigma^2 (\boldsymbol{A}^\top \boldsymbol{A})^{-1}.$$

因此，$\hat{\boldsymbol{x}}_{\text{LS}}$ 的均方风险（对任意固定的 $\boldsymbol{x}$）等于

$$\mathbb{E}\|\hat{\boldsymbol{x}}_{\text{LS}} - \boldsymbol{x}\|^2 = \text{tr}\left(\sigma^2 (\boldsymbol{A}^\top \boldsymbol{A})^{-1}\right) = \sigma^2 \sum_{i=1}^{n} \frac{1}{\lambda_i}.$$

这表明

$$\inf_{\hat{\boldsymbol{x}}} \sup_{\boldsymbol{x}} \mathbb{E}\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|^2 \leq \sup_{\boldsymbol{x}} \mathbb{E}\|\hat{\boldsymbol{x}}_{\text{LS}} - \boldsymbol{x}\|^2 = \sigma^2 \sum_{i=1}^{n} \frac{1}{\lambda_i}.$$

结合 (A) 和 (B) 即得所述恒等式。 □

接下来证明极小极大误差界。定义秩为$_t$ r 的张量集合如下:

$$\boldsymbol{\mathcal{D}}_r = \{\boldsymbol{\mathcal{X}} : \text{rank}_t(\boldsymbol{\mathcal{X}}) = r,\ \boldsymbol{\mathcal{X}} \in \mathbb{R}^{n \times n \times k}\},$$

以及形如 $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{Y}} * \boldsymbol{\mathcal{R}}$ 的张量集合如下:

$$\boldsymbol{\mathcal{D}}_{\boldsymbol{\mathcal{Y}}} = \{\boldsymbol{\mathcal{X}} : \boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{Y}} * \boldsymbol{\mathcal{R}},\ \boldsymbol{\mathcal{X}} \in \mathbb{R}^{n \times r \times k},\ \boldsymbol{\mathcal{Y}} \in \mathbb{R}^{n \times r \times k},\ \boldsymbol{\mathcal{Y}}^\top * \boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{I}}\}.$$

注意，集合 $\boldsymbol{\mathcal{D}}_r$ 远大于集合 $\boldsymbol{\mathcal{D}}_{\boldsymbol{\mathcal{Y}}}$。因此，

$$\inf_{\boldsymbol{\mathcal{X}}_{est}} \sup_{\boldsymbol{\mathcal{X}}:\text{rank}_t(\boldsymbol{\mathcal{X}})=r} \mathbb{E}||\boldsymbol{\mathcal{X}}_{est} - \boldsymbol{\mathcal{X}}||_F^2 \geq \inf_{\boldsymbol{\mathcal{X}}_{est}} \sup_{\boldsymbol{\mathcal{X}}:\boldsymbol{\mathcal{X}}=\boldsymbol{\mathcal{Y}}*\boldsymbol{\mathcal{R}}} \mathbb{E}||\boldsymbol{\mathcal{X}}_{est} - \boldsymbol{\mathcal{X}}||_F^2. \tag{98}$$

对于固定的正交张量 $\boldsymbol{\mathcal{Y}}$，定义正交投影张量 $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{Y}}} = \boldsymbol{\mathcal{Y}} * \boldsymbol{\mathcal{Y}}^\top$，其满足 $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{Y}}}^2 = \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{Y}}}$，$\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{Y}}}^\top = \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{Y}}}$。然后固定估计量 $\boldsymbol{\mathcal{X}}_{est}$，对于任意 $\boldsymbol{\mathcal{X}} \in \boldsymbol{\mathcal{D}}_{\boldsymbol{\mathcal{Y}}}$，有:

$$\begin{aligned} ||\boldsymbol{\mathcal{X}}_{est} - \boldsymbol{\mathcal{X}}||_F^2 &= ||\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{Y}}} * \boldsymbol{\mathcal{X}}_{est} - \boldsymbol{\mathcal{X}} + (\boldsymbol{\mathcal{I}} - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{Y}}}) * \boldsymbol{\mathcal{X}}_{est}||_F^2 \\ &= ||\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{Y}}} * \boldsymbol{\mathcal{X}}_{est} - \boldsymbol{\mathcal{X}}||_F + ||(\boldsymbol{\mathcal{I}} - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{Y}}}) * \boldsymbol{\mathcal{X}}_{est}||_F^2 \\ &\quad + 2\langle \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{Y}}} * \boldsymbol{\mathcal{X}}_{est} - \boldsymbol{\mathcal{X}}, (\boldsymbol{\mathcal{I}} - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{Y}}}) * \boldsymbol{\mathcal{X}}_{est} \rangle \\ &\overset{(a)}{=} ||\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{Y}}} * \boldsymbol{\mathcal{X}}_{est} - \boldsymbol{\mathcal{X}}||_F + ||(\boldsymbol{\mathcal{I}} - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{Y}}}) * \boldsymbol{\mathcal{X}}_{est}||_F^2, \end{aligned} \tag{99}$$

其中（a）利用张量列子空间$\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{Y}}} * \boldsymbol{\mathcal{X}}_{est} - \boldsymbol{\mathcal{X}}$和（$\boldsymbol{\mathcal{I}} - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{Y}}}$）$* \boldsymbol{\mathcal{X}}_{est}$正交这一事实，这意味着它们的内积为零。因此，我们可以直接得到

$$\begin{aligned} ||\boldsymbol{\mathcal{X}}_{est} - \boldsymbol{\mathcal{X}}||_F^2 &\geq ||\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{Y}}} * \boldsymbol{\mathcal{X}}_{est} - \boldsymbol{\mathcal{X}}||_F^2 = ||\boldsymbol{\mathcal{Y}} * \boldsymbol{\mathcal{Y}}^\top * \boldsymbol{\mathcal{X}}_{est} - \boldsymbol{\mathcal{Y}} * \boldsymbol{\mathcal{R}}||_F^2 \\ &= ||\boldsymbol{\mathcal{Y}}^\top * \boldsymbol{\mathcal{X}}_{est} - \boldsymbol{\mathcal{R}}||_F^2. \end{aligned} \tag{100}$$

令$\boldsymbol{\mathcal{R}}_{est} = \boldsymbol{\mathcal{Y}}^\top * \boldsymbol{\mathcal{X}}_{est}$，则有

$$\inf_{\boldsymbol{\mathcal{X}}_{est}} \sup_{\boldsymbol{\mathcal{X}}:\boldsymbol{\mathcal{X}}=\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{R}}} \mathbb{E}||\boldsymbol{\mathcal{X}}_{est} - \boldsymbol{\mathcal{X}}||_F^2 \geq \inf_{\boldsymbol{\mathcal{R}}_{est}} \sup_{\boldsymbol{\mathcal{R}}} \mathbb{E}||\boldsymbol{\mathcal{R}}_{est} - \boldsymbol{\mathcal{R}}||_F^2. \tag{101}$$

因此，极小极大风险的下界不低于根据数据估计$\boldsymbol{\mathcal{R}}$ 的风险

$$\boldsymbol{y} = \mathfrak{M}_{\boldsymbol{\mathcal{Y}}}(\mathrm{vec}(\boldsymbol{\mathcal{R}})) + \boldsymbol{s}, \tag{102}$$

其中 $\mathfrak{M}_{\boldsymbol{\mathcal{Y}}}$: $\mathbb{R}^{rnk} \to \mathbb{R}^m$，vec 表示向量化算子。然后，我们可以应用引理 9 的结果，证明极小极大风险的下界为

$$\inf_{\boldsymbol{\mathcal{X}}_{est}} \sup_{\boldsymbol{\mathcal{X}}:\mathrm{rank}_t(\boldsymbol{\mathcal{X}})=r} \mathbb{E}||\boldsymbol{\mathcal{X}}_{est} - \boldsymbol{\mathcal{X}}||_F^2 \geq \sum_i^{rnk} \frac{\sigma^2}{\lambda_i(\mathfrak{M}_{\boldsymbol{\mathcal{Y}}}^* \mathfrak{M}_{\boldsymbol{\mathcal{Y}}})}. \tag{103}$$

然后，我们可以利用以下引理和 t-RIP 假设来界定项 (103)。

引理 10. 设$\boldsymbol{\mathcal{Y}}$ 是一个 $n \times r \times k$ 正交张量，假设线性映射 $\mathfrak{M}(\cdot)$ 满足(r, δ) *t-RIP*，则 $\mathfrak{M}_{\boldsymbol{\mathcal{Y}}}^* \mathfrak{M}_{\boldsymbol{\mathcal{Y}}}$ 的所有特征值均属于区间 $[m(1-\delta)\ m(1+\delta)]$。

证明。 根据定义，我们有

$$\begin{aligned} \lambda_{\min}(\mathfrak{M}_{\boldsymbol{\mathcal{Y}}}^* \mathfrak{M}_{\boldsymbol{\mathcal{Y}}}) &= \inf_{||\,\mathrm{vec}(\boldsymbol{\mathcal{R}})||_F=1} \left\langle \mathrm{vec}(\boldsymbol{\mathcal{R}}), \mathfrak{M}_{\boldsymbol{\mathcal{Y}}}^* \mathfrak{M}_{\boldsymbol{\mathcal{Y}}}(\mathrm{vec}(\boldsymbol{\mathcal{R}})) \right\rangle \\ \lambda_{\max}(\mathfrak{M}_{\boldsymbol{\mathcal{Y}}}^* \mathfrak{M}_{\boldsymbol{\mathcal{Y}}}) &= \sup_{||\,\mathrm{vec}(\boldsymbol{\mathcal{R}})||_F=1} \left\langle \mathrm{vec}(\boldsymbol{\mathcal{R}}), \mathfrak{M}_{\boldsymbol{\mathcal{Y}}}^* \mathfrak{M}_{\boldsymbol{\mathcal{Y}}}(\mathrm{vec}(\boldsymbol{\mathcal{R}})) \right\rangle. \end{aligned} \tag{104}$$

注意到

$$\langle \boldsymbol{\mathcal{R}}, \mathfrak{M}_{\boldsymbol{\mathcal{Y}}}^* \mathfrak{M}_{\boldsymbol{\mathcal{Y}}}(\mathrm{vec}(\boldsymbol{\mathcal{R}})) \rangle = ||\mathfrak{M}_{\boldsymbol{\mathcal{Y}}}(\mathrm{vec}(\boldsymbol{\mathcal{R}}))||^2 = ||\mathfrak{M}(\boldsymbol{\mathcal{Y}} * \boldsymbol{\mathcal{R}})||^2,$$

于是，我们可以利用 (r, δ) t-RIP 对 $||\mathfrak{M}(\boldsymbol{\mathcal{Y}} * \boldsymbol{\mathcal{R}})||^2$ 进行界定

$$m(1-\delta)||\boldsymbol{\mathcal{Y}} * \boldsymbol{\mathcal{R}}||_F^2 \leq ||\mathfrak{M}(\boldsymbol{\mathcal{Y}} * \boldsymbol{\mathcal{R}})||^2 \leq m(1+\delta)||\boldsymbol{\mathcal{Y}} * \boldsymbol{\mathcal{R}}||_F^2.$$

由于 $||\ *\ ||_F = ||\ ||_F = 1$ $\mathfrak{M}_{\boldsymbol{\mathcal{Y}}}^* \mathfrak{M}_{\boldsymbol{\mathcal{Y}}}$ $[m\ 1-\delta\ \ m\ 1+\boldsymbol{\mathcal{Y}}\boldsymbol{\mathcal{R}}\boldsymbol{\mathcal{R}}$，因此其特征值由 δ $(\ )\ (\ )]$界定。 □

结合引理 10 和方程 (103) 的结果，我们有

$$\inf_{\boldsymbol{\mathcal{X}}_{est}} \sup_{\boldsymbol{\mathcal{X}}:\mathrm{rank}_t(\boldsymbol{\mathcal{X}})=r} \mathbb{E}||\boldsymbol{\mathcal{X}}_{est} - \boldsymbol{\mathcal{X}}||_F^2 \geq \sum_i^{rnk} \frac{\sigma^2}{\lambda_i(\mathfrak{M}_{\boldsymbol{\mathcal{Y}}}^* \mathfrak{M}_{\boldsymbol{\mathcal{Y}}})} \geq \frac{1}{1+\delta} \frac{nrk\sigma^2}{m}, \tag{105}$$

由此完成了定理 3 中第一个不等式的证明。

接下来我们证明定理 3 中的第二个不等式。首先引入一个技术性引理。

引理 11 (引理 3.14，见 (Candes & Plan, 2011))。 假设 $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{A}, \boldsymbol{s}$ 遵循线性模型 *(95)*，其中 $\boldsymbol{s} \sim \mathcal{N}(0,\ \sigma^2 \boldsymbol{I})$，则

$$\inf_{\hat{\boldsymbol{x}}} \sup_{\boldsymbol{x} \in \mathbb{R}^n} \mathbb{P}\left(||\hat{\boldsymbol{x}} - \boldsymbol{x}||^2 \geq \frac{1}{2||\boldsymbol{A}||^2} n\sigma^2 \right) \geq 1 - e^{-n/16}. \tag{106}$$

根据引理 11、10 和线性模型的结果

$$\boldsymbol{y} = \mathfrak{M}_{\boldsymbol{\mathcal{Y}}}(\mathrm{vec}(\boldsymbol{\mathcal{R}})) + \boldsymbol{s},\ \boldsymbol{\mathcal{R}} \in \mathbb{R}^{n \times r \times k},\ \boldsymbol{y} \in \mathbb{R}^m,$$

我们可以得到

$$\sup_{\boldsymbol{\mathcal{X}}_\star:\mathrm{rank}_t(\boldsymbol{\mathcal{X}}_\star) \leq r} \mathbb{P}\left(||\boldsymbol{\mathcal{X}}_{est} - \boldsymbol{\mathcal{X}}_\star||_F^2 \geq \frac{nrk\sigma^2}{2m(1+\delta)} \right) \geq 1 - e^{-nrk/16}, \tag{107}$$

由此完成第二个不等式的证明。

G 定理 4 的证明

引理 12。 假设验证测量$\boldsymbol{\mathcal{A}}_i$, $i \in \mathcal{I}_{\text{val}}$ 中的每个元素均从均值为零、方差为 *1* 的独立同分布次高斯分布中采样，且每个 e_i 均服从方差为 σ^2的零均值高斯分布，其中 c_1, $c_2 \geq 1$为一些绝对常数。我们还假设张量$\boldsymbol{\mathcal{D}}_1, \boldsymbol{\mathcal{D}}_2, ..., \boldsymbol{\mathcal{D}}_T$与 $\mathfrak{M}_{\text{val}}$和$\boldsymbol{e}_{\text{val}}$相互独立。那么，对于任意$\delta_{\text{val}} > 0$, 给定 $m_{\text{val}} \geq {}^{C_1 \log_{\delta_{\text{val}}^2} T}$, 以至少 $1 - 2T\exp\ -C_2 m_{\text{val}} \delta^2$的概率，

$$\left| ||\mathfrak{M}_{\text{val}}(\boldsymbol{\mathcal{D}}_t) + \boldsymbol{e}||_F^2 - m_{\text{val}}(||\boldsymbol{\mathcal{D}}_t||_F^2 + \sigma^2) \right| \leq \delta_{\text{val}} m_{\text{val}}(||\boldsymbol{\mathcal{D}}_t||_F^2 + \sigma^2), \forall t = 1, ..., T, \quad (108)$$

其中， C_1、 $C_2 \geq 0$ 是可能依赖于 c_1 和 c_2的常数。

证明。 该引理的证明直接由 (Ding 等，2025) 中的引理 D.1 得出，因为无论$\boldsymbol{\mathcal{A}}_i$ 和$\boldsymbol{\mathcal{D}}$ 是矩阵还是张量，$\langle \boldsymbol{\mathcal{A}}_i, \boldsymbol{\mathcal{D}} \rangle + e_i$ 都是均值为零、方差为 $||\boldsymbol{\mathcal{D}}||_F^2 + \sigma^2$的次高斯随机变量。因此，引理 D.1 的结论可直接应用于本引理。 □

引理 13。 设$\check{t} = \arg\min_{1 \leq t \leq T} ||\mathfrak{M}_{\text{val}}(\boldsymbol{\mathcal{D}}_t) + \boldsymbol{e}||_2$和$\hat{t} = \arg\min_{1 \leq t \leq T} ||\boldsymbol{\mathcal{D}}_t||_F$, 在引理 *12* 的假设下，我们有

$$||\boldsymbol{\mathcal{D}}_{\check{t}}||_F^2 \leq \frac{1 + \delta_{\text{val}}}{1 - \delta_{\text{val}}} ||\boldsymbol{\mathcal{D}}_{\hat{t}}||_F^2 + \frac{2\delta_{\text{val}}}{1 - \delta_{\text{val}}} \sigma^2. \quad (109)$$

证明。 在引理 12 的假设下，我们有

$$(1 - \delta_{\text{val}})(||\boldsymbol{\mathcal{D}}_t||_F^2 + \sigma^2) \leq \frac{1}{m_{\text{val}}} ||\mathfrak{M}_{\text{val}}(\boldsymbol{\mathcal{D}}_t) + \mathbf{e}||_F^2 \leq (1 + \delta_{\text{val}})(||\boldsymbol{\mathcal{D}}_t||_F^2 + \sigma^2), \forall t = 1, ..., T, \quad (110)$$

根据引理 12 的结果，我们有

$$\begin{aligned} ||\boldsymbol{\mathcal{D}}_{\hat{t}}||_F^2 + \sigma^2 &\leq \frac{1}{m_{\text{val}}(1 - \delta_{\text{val}})} ||\mathfrak{M}_{\text{val}}(\boldsymbol{\mathcal{D}}_{\hat{t}}) + \mathbf{e}||_F^2 \\ &\leq \frac{1}{m_{\text{val}}(1 - \delta_{\text{val}})} ||\mathfrak{M}_{\text{val}}(\boldsymbol{\mathcal{D}}_{\hat{t}}) + \mathbf{e}||_F^2 \leq \frac{1 + \delta_{\text{val}}}{1 - \delta_{\text{val}}} (||\boldsymbol{\mathcal{D}}_{\hat{t}}||_F^2 + \sigma^2), \end{aligned} \quad (111)$$

这表明

$$||\boldsymbol{\mathcal{D}}_{\hat{t}}||_F^2 \leq \frac{1 + \delta_{\text{val}}}{1 - \delta_{\text{val}}} ||\boldsymbol{\mathcal{D}}_{\hat{t}}||_F^2 + \frac{2\delta_{\text{val}}}{1 - \delta_{\text{val}}} \sigma^2. \quad (112)$$

因此，我们完成了引理 13 的证明。 □

利用这两个引理以及定理 2，我们接着证明定理 4。将引理 13 的结果替换为 $\boldsymbol{\mathcal{D}}_{\check{t}} = \boldsymbol{\mathcal{U}}_{\check{t}} * \boldsymbol{\mathcal{U}}_{\check{t}}^\top - \boldsymbol{\mathcal{X}}_\star$和 $\boldsymbol{\mathcal{D}}_{\hat{t}} = \boldsymbol{\mathcal{U}}_{\hat{t}} * \boldsymbol{\mathcal{U}}^\top{}_{\hat{t}} - \boldsymbol{\mathcal{X}}_\star$, 我们有

$$||\boldsymbol{\mathcal{U}}_{\check{t}} * \boldsymbol{\mathcal{U}}_{\check{t}}^\top - \boldsymbol{\mathcal{X}}_\star||_F^2 \leq \frac{1 + \delta_{\text{val}}}{1 - \delta_{\text{val}}} ||\boldsymbol{\mathcal{U}}_{\hat{t}} * \boldsymbol{\mathcal{U}}_{\hat{t}}^\top - \boldsymbol{\mathcal{X}}_\star||_F^2 + \frac{2\delta_{\text{val}}}{1 - \delta \text{val}} \sigma^2. \quad (113)$$

为得到误差 $C^{\frac{nkr\sigma^2\kappa^4}{m}}$, 我们需要界 $\frac{2\delta}{1-\delta}\sigma^2$, 这要求 $\delta \leq \frac{nkr\kappa^4}{3m_{\text{train}}}$. 取$\delta = \frac{nkr\kappa^4}{3m_{\text{train}}}$, 则可以验证 $\frac{1+\delta_{\text{val}}}{1-\delta_{\text{val}}} \leq 2$：

$$\begin{aligned} \frac{1 + \delta_{\text{val}}}{1 - \delta_{\text{val}}} = \frac{1 + \frac{nkr\kappa^4}{3m_{\text{train}}}}{1 - \frac{nkr\kappa^4}{3m_{\text{train}}}} = \frac{3m_{\text{train}} + nkr\kappa^4}{3m_{\text{train}} - nkr\kappa^4} &= 1 + \frac{2nkr\kappa^4}{3m_{\text{train}} - nkr\kappa^4} \\ &\overset{(a)}{\leq} 1 + \frac{2nkr\kappa^4}{nkr\kappa^4(3r\kappa^4 - 1)} \leq 2, \end{aligned} \quad (114)$$

其中，(a) 用到了如下假设： $m \gtrsim nkr^2\kappa^8$。因此，结合定理 2 的结果，我们有

$$||\boldsymbol{\mathcal{U}}_{\hat{t}} * \boldsymbol{\mathcal{U}}_{\hat{t}}^\top - \boldsymbol{\mathcal{X}}_\star||_F^2 \leq C \frac{nkr\sigma^2\kappa^4}{m_{\text{train}}}.$$

此外，结合 $\frac{nkr\kappa^4}{3m_{\text{train}}}$ 假设 $m_{\text{val}} \geq C_1 \log_{\delta_{\text{val}}^2} T$, 我们得到 $m_{\text{val}} \geq C_1{}^{m^2}$ 训练$\log T(rnk\kappa^4)^2$ ·

因此，定理 4 的证明完成。

H 技术引理

引理 14. 假设线性映射 $\mathfrak{M}: \mathbb{R}^{n\times n\times k} \to \mathbb{R}^m$ 满足 *(r+1, δ_1) t-RIP*，且 $\delta_1 \in (0,1)$，那么 $\mathfrak{M}$ 也满足 *(r, $\sqrt{kr}\delta_1$) S2S-t-RIP*。

证明。该引理的证明可以通过引入不等式 $||\mathcal{Z}|_F \leq \sqrt{r}||\mathcal{Z}||$ 和 $||\underline{\mathcal{Z}}||_F = \sqrt{1/k}||\mathcal{Z}||_F$，由 [(Karnik 等人，2025) 的引理 G.2] 的证明进行改编得到。 □

引理 15。假设线性映射 $\mathfrak{M}: \mathbb{R}^{n\times n\times k} \to \mathbb{R}^m$ 满足 *(2, δ_2) t-RIP*，且 $\delta_2 \in (0, 1)$，那么 $\mathfrak{M}$ 也满足 $\sqrt{k}\delta_2$*-S2N-t-RIP*。

证明。该引理的证明可以通过引入不等式 $||\mathcal{Z}|_F \leq \sqrt{r}||\mathcal{Z}||$ 和 $||\underline{\mathcal{Z}}||_F = \sqrt{1/k}||\mathcal{Z}||_F$，由 [(Karnik 等人，2025) 的引理 G.3] 的证明进行改编得到。 □

引理 16。对于管秩为 r 的张量 $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{n\times n\times k}$，有

$$||\boldsymbol{\mathcal{X}}|| \leq \sqrt{n_3}||\boldsymbol{\mathcal{X}}||_F,\ ||\boldsymbol{\mathcal{X}}||_F \leq \sqrt{r}||\boldsymbol{\mathcal{X}}||,\ \ ||\boldsymbol{\mathcal{X}}||_* \leq r||\boldsymbol{\mathcal{X}}||. \tag{115}$$

I 对一般张量的扩展

本节简要分析如何通过将非对称模型表述为对称模型，将其扩展到非对称情形。我们首先给出非对称张量感知模型:

$$\mathbf{y} = \mathfrak{M}_a(\boldsymbol{\mathcal{X}}_\star) + \mathbf{s}, \tag{116}$$

其中 $\mathrm{X}_\star \in \mathbb{R}^{n_1\times n_2\times k}$, $\mathfrak{M}_a(\mathrm{X}) = [\langle \mathrm{B}_1, \mathrm{X}_\star\rangle, \langle \mathrm{B}_2, \mathrm{X}_\star\rangle, ..., \langle \mathrm{B}_i, \mathrm{X}_\star\rangle]$。在此非对称模型下，我们采用非对称因子分解 $\mathrm{X} = \mathrm{L} * \mathrm{R}^\top$，$\mathrm{L} \in \mathbb{R}^{n_1\times r\times k}$，$\mathrm{R} \in \mathbb{R}^{n_2\times r\times k}$. 。接着定义对称测量张量 $\mathrm{C}_i \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)\times k}$:

$$\boldsymbol{\mathcal{C}}_i := \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & \boldsymbol{\mathcal{B}}_i \\ \boldsymbol{\mathcal{B}}_i^\top & 0 \end{pmatrix} \tag{117}$$

以及相应的线性映射 $\mathfrak{C}: \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)\times k} \to \mathbb{R}^m$，其定义如下:

$$(\mathfrak{C}(\boldsymbol{\mathcal{X}}))_i = \langle \boldsymbol{\mathcal{C}}_i, \boldsymbol{\mathcal{X}}\rangle.$$

定义

$$\mathrm{sym}(\boldsymbol{\mathcal{X}}) := \begin{bmatrix} 0 & \boldsymbol{\mathcal{X}} \\ \boldsymbol{\mathcal{X}}^\top & 0 \end{bmatrix}$$

and

$$\boldsymbol{\mathcal{Z}}_t := \frac{1}{\sqrt{2}}\begin{bmatrix} \boldsymbol{\mathcal{L}}_t \\ \boldsymbol{\mathcal{R}}_t \end{bmatrix} \text{ and } \tilde{\boldsymbol{\mathcal{Z}}}_t := \frac{1}{\sqrt{2}}\begin{bmatrix} \boldsymbol{\mathcal{L}}_t \\ -\boldsymbol{\mathcal{R}}_t \end{bmatrix}.$$

根据这些定义，我们将非对称感知模型转化为对称模型:

$$\frac{1}{\sqrt{2}}\mathfrak{C}(\mathrm{sym}(\boldsymbol{\mathcal{X}})) = \mathfrak{M}_a(\boldsymbol{\mathcal{X}}) \text{ and } \frac{1}{\sqrt{2}}\mathfrak{C}(\boldsymbol{\mathcal{Z}}_\mathfrak{t} * \boldsymbol{\mathcal{Z}}_\mathfrak{t}^\top - \tilde{\boldsymbol{\mathcal{Z}}}_\mathfrak{t} * \tilde{\boldsymbol{\mathcal{Z}}}_\mathfrak{t}^\top) = \mathfrak{M}_a(\boldsymbol{\mathcal{L}}_t * \boldsymbol{\mathcal{R}}_t^\top).$$

基于该模型，我们得到以下目标函数:

$$h(\boldsymbol{\mathcal{L}}_t, \boldsymbol{\mathcal{R}}_t) = \frac{1}{2}||\mathfrak{M}(\boldsymbol{\mathcal{X}}_\star - \boldsymbol{\mathcal{L}}_t * \boldsymbol{\mathcal{R}}_t^\top) + \boldsymbol{s}||^2 \tag{118}$$

然后我们定义相应的对称损失函数:

$$h_{\mathrm{sym}}(\boldsymbol{\mathcal{Z}}_t, \tilde{\boldsymbol{\mathcal{Z}}}_t) = \frac{1}{4}||\mathfrak{C}(\mathrm{sym}(\boldsymbol{\mathcal{X}}_\star) - \boldsymbol{\mathcal{Z}}_t * \boldsymbol{\mathcal{Z}}_t^\top + \tilde{\boldsymbol{\mathcal{Z}}}_t * \tilde{\boldsymbol{\mathcal{Z}}}_t) + \sqrt{2}\boldsymbol{s}||^2.$$

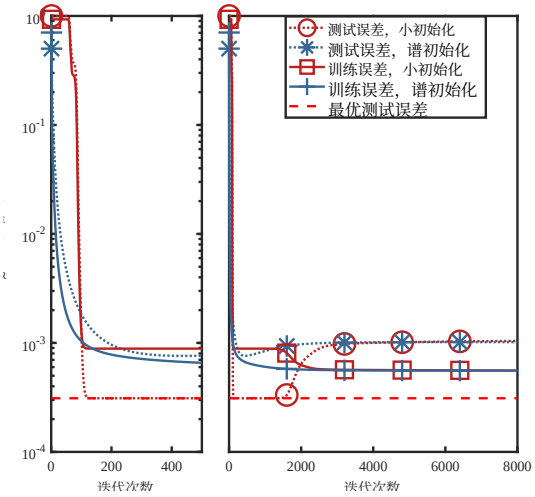


图 6：使用 FGD 比较问题 (118) 在谱初始化与小初始化下的训练误差和测试误差。真实张量的管秩为 $r = 2$，过估计秩为 $R = 4$，大小为 $n_1 = n_2 = 20, k = 3, m = 5kr$，包含 $2n_1 - r$ 个测量，噪声为 $\sigma = 10^{-3}$。谱初始化遵循 Liu等人 (2024b) 的方法，而小初始化使用近零起点。训练误差为 $\frac{1}{2}||\boldsymbol{y}-\mathfrak{M}(\boldsymbol{\mathcal{L}}*\boldsymbol{\mathcal{R}}^\top)||^2$，测试误差为 $||\boldsymbol{\mathcal{L}}*\boldsymbol{\mathcal{R}}^\top-\boldsymbol{\mathcal{X}}_\star||_F^2/||\boldsymbol{\mathcal{X}}_\star||_F^2$。“基线”表示在精确秩 $R = r$ 下的恢复。插图展示了早期（前 500 次迭代）与完整的误差曲线。

$\tilde{}$ h_{sym}（$\boldsymbol{\mathcal{Z}}_t$，$\boldsymbol{\mathcal{Z}}_t$）的梯度更新为

$$\begin{aligned}\boldsymbol{\mathcal{Z}}_{t+1} &= \boldsymbol{\mathcal{Z}}_t + \eta[(\boldsymbol{\mathfrak{C}}^*\boldsymbol{\mathfrak{C}})(\text{sym}(\boldsymbol{\mathcal{X}}_\star) - \boldsymbol{\mathcal{Z}}_t * \boldsymbol{\mathcal{Z}}_t^\top + \tilde{\boldsymbol{\mathcal{Z}}}_t * \tilde{\boldsymbol{\mathcal{Z}}}_t) + \boldsymbol{\mathfrak{C}}^*(\sqrt{2}\boldsymbol{s})] * \boldsymbol{\mathcal{Z}}_t \\ \tilde{\boldsymbol{\mathcal{Z}}}_{t+1} &= \tilde{\boldsymbol{\mathcal{Z}}}_t - \eta[(\boldsymbol{\mathfrak{C}}^*\boldsymbol{\mathfrak{C}})(\text{sym}(\boldsymbol{\mathcal{X}}_\star) - \boldsymbol{\mathcal{Z}}_t * \boldsymbol{\mathcal{Z}}_t^\top + \tilde{\boldsymbol{\mathcal{Z}}}_t * \tilde{\boldsymbol{\mathcal{Z}}}_t) + \boldsymbol{\mathfrak{C}}^*(\sqrt{2}\boldsymbol{s})] * \tilde{\boldsymbol{\mathcal{Z}}}_t.\end{aligned} \tag{119}$$

这一表述使我们能够借鉴对称情形中的一些证明技巧。然而，处理两个因子张量所引入的不平衡是一项重大挑战，我们目前正在积极研究这一问题。令人鼓舞的是，图 6 中的实验结果表明，本文描述的现象在非对称设定下同样存在。

J 附加实验

J.1 不同噪声分布下的仿真

我们进行仿真实验，以验证我们的理论结果在各种噪声分布下仍然成立，而不仅限于高斯噪声。实验设置与第6节中的设置相同，只是将高斯噪声替换为两种次指数噪声：拉普拉斯噪声和指数噪声。下面简要介绍实验中考虑的两种噪声模型：

- 拉普拉斯噪声：噪声向量服从拉普拉斯分布，

$$\boldsymbol{s} \sim \text{Laplace}(\mu, b), \quad f(s_i) = \frac{1}{2b}\exp\left(-\frac{|s_i-\mu|}{b}\right),$$

  这是一种均值为 μ、方差为 $2b^2$ 的对称次指数分布。
- 指数噪声：噪声向量服从指数分布，

$$\boldsymbol{s} \sim \text{Exp}(\lambda), \quad f(s_i) = \lambda\exp(-\lambda s_i), \quad x\boldsymbol{s}_i \geq 0,$$

  这是一个非对称次指数分布，其均值为 $1/\lambda$，方差为 $1/\lambda^2$。

图7和图8所示的结果表明，在两种噪声类型下，采用小初始化的FGD即使处于过参数化情形，也能达到与精确管秩情形相同的恢复误差。这证实了定理 2所提供的保证可以扩展到

高斯分布之外。此外，结合验证与提前停止的FGD所得到的误差与精确管秩设定下的误差非常接近，进一步验证了该方法的有效性，并表明定理 3中的结果同样可以扩展到次指数噪声。

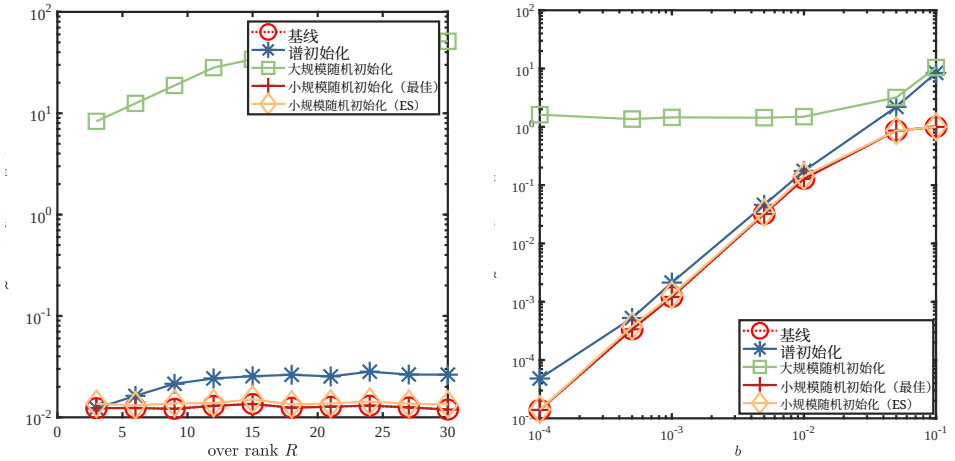


图 7：拉普拉斯噪声下，不同 R、b 的性能比较，噪声水平为 $\mu = 0$。子图 (a) 展示了所有方法在不同超秩值 R 下的恢复误差，参数设置为 $m = 5kr(2n - r)$、$n = 30$、$b = 10^{-3}$、$\eta = 0.1$ 和 $T = 5000$。子图 (b) 展示了不同噪声水平 b 下的误差，参数设置为 $m = 5kr(2n - r)$、$n = 30$、$R = 3r$、$\eta = 0.1$ 和 $T = 5000$。

J.2 真实数据实验

本节提供更多实验细节和结果。首先，我们介绍用于张量补全任务的算法，如算法3所示。随后，我们给出评估指标——峰值信噪比和相对误差的定义：

$$\text{PSNR} = 10\log_{10}\left(\frac{||\boldsymbol{\mathcal{X}}_\star||_\infty^2}{\frac{1}{n_1n_2n_3}||\hat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}_\star||_F^2}\right), \text{RE} = \frac{||\hat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}_\star||_F}{|\boldsymbol{\mathcal{X}}_\star|||_F},$$

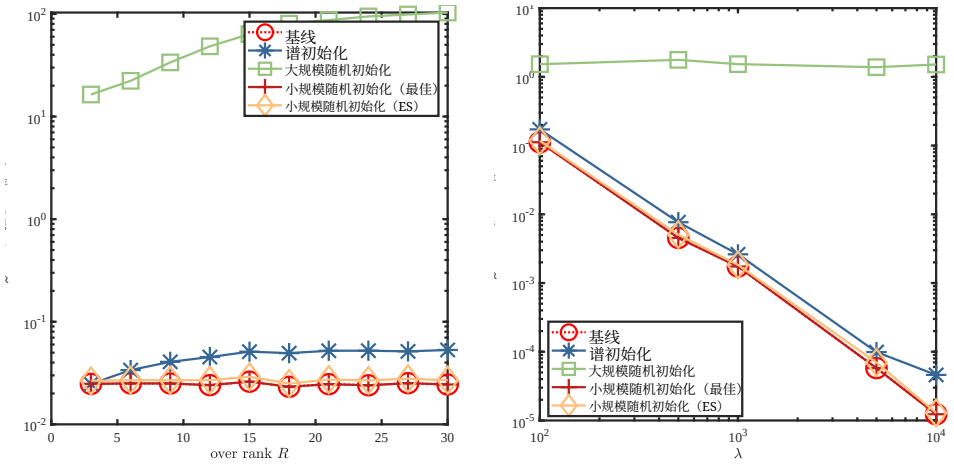
ˆ 其中$\boldsymbol{\mathcal{X}}_\star$ 是真实值，$\boldsymbol{\mathcal{X}}$ 是估计张量。接下来，我们简要介绍用于比较的基线方法：

- TNN（Lu等，2018年）：一种经典的凸方法，基于由()提出的管状张量核范数最小化，在张量补全中得到广泛应用。
- TCTF（Zhouet al., 2017）：一种基于张量分解的方法，具有管秩估计功能，旨在降低计算成本。
- UTF（Du 等人，2021）：另一种张量分解方法，它用施加于两个因子张量的 Frobenius 范数约束取代管状张量核范数约束。
- TC-RE（Shi等人，2021）：一种基于秩估计的方法，首先估计管秩，然后使用截断t-SVD进行张量补全。
- GTNN-HOP$_p$ （Wang等人，2024）：一种用混合普通-l_p 惩罚替代传统TNN软阈值处理的方法，以提升性能。

We 在彩色图像补全和视频补全任务上进行实验，并将该方法与 e our 上述方法进行比较。

J.2.1 彩色图像补全实验

我们在伯克利分割数据集（Martin等人，2001）上进行彩色图像补全实验。我们随机选取50张大小为 $481 \times 321 \times 3$ 的彩色图像，并设置采样



(a) (b)

图8：指数噪声下不同 R、 λ 设置的性能比较。子图 (a) 展示了所有方法在不同超秩值 R下的恢复误差，参数设置为$m = 5kr(2n - r)$、 $n = 30$、 $b = 10^{-3}$、 $\eta = 0.1$、 $\lambda = 1000$ 和 $T = 5000$。子图 (b) 展示了不同噪声水平 λ下的误差，其中 $m = 5kr(2n - r)$、 $n = 30$、 $R = 3r$、 $\eta = 0.1$和$T = 5000$。

算法3 使用提前停止的 FGD 求解张量补全 输入：训练数据 $\mathfrak{P}_{\Omega}^{\text{train}}(\mathrm{X}_{\star} + \mathrm{S}_n)$、验证数据 $\mathfrak{P}_{\Omega}^{\text{val}}(\mathrm{X}_{\star} + \mathrm{S}_n)$、初始化尺度 α、步长 η、估计秩 R、迭代次数 T 初始化：初始化 L_0,R_0, , 其中 $\mathrm{L}_0, \mathrm{R}_0$ 的每个元素均独立同分布地取自 $\mathcal{N}(0, \frac{\alpha^2}{R})$。1：对于 $t = 0$ 到 $T - 1$， 执行 2：$\mathrm{L}_{t+1} = \mathrm{L}_t - \frac{\eta}{p}\mathfrak{P}_{\Omega}^{\text{train}}(\mathrm{L}_t * \mathrm{R}_t^{\top} - \mathrm{X}_{\star} - \mathrm{S}_n) * \mathrm{R}_t$3：$\mathrm{R}_{t+1} = \mathrm{R}_t - \frac{\eta}{p}\mathfrak{P}_{\Omega}^{\text{train}}(\mathrm{L}_t * \mathrm{R}_t^{\top} - \mathrm{X}_{\star} - \mathrm{S}_n)^{\top} * \mathrm{L}_t$4：验证损失： $e_t = \frac{1}{2p}\left\|\mathfrak{P}_{\Omega}^{\text{val}}(\mathrm{L}_t * \mathrm{R}_t^{\top} - \mathrm{X}_{\star} - \mathrm{S}_n)\right\|_F^2$5：结束循环 6：输出：$\mathrm{L}_{\check{t}} * \mathrm{R}_{\check{t}}^{\top}$, 其中 $\check{t} = \operatorname{argmin}_{1 \le t \le T} e_t$。

表 3：在不同采样率和噪声水平下，各种方法在平均峰值信噪比（PSNR）和平均相对误差（RE）方面的比较。PSNR 越高且 RE 越小，表示重建质量越好。“FGD-ES”表示采用提前停止的 FGD，而“FGD-best”表示 FGD 在所有迭代次数中达到的最小误差。

方法	$p=0.3$				$p=0.4$			
	$\sigma=0.03$		$\sigma=0.05$		$\sigma=0.03$		$\sigma=0.05$	
	峰值信噪比↑	相对平方误差↓	PSNR ↑	RSE ↓	峰值信噪比↑	相对平方误差↓	峰值信噪比↑	相对平方误差↓
UTF	7.8242	0.276	7.3535	0.2884	6.8286	0.3112	5.5376	0.3559
TNN	20.211	0.0659	17.2288	0.0934	20.4965	0.0639	17.1281	0.0947
TC-RE	19.7102	0.0698	16.9971	0.096	19.7435	0.0698	16.4039	0.1029
GTNN-HOP$_{0.3}$	19.6553	0.0706	16.1069	0.107	20.0091	0.068	16.268	0.1051
GTNN-HOP$_{0.6}$	20.2583	0.0659	16.8056	0.0987	20.5764	0.0636	16.8933	0.0978
FGD-ES	<u>22.083</u>	<u>0.0529</u>	<u>21.02</u>	<u>0.0597</u>	<u>22.2876</u>	<u>0.0517</u>	<u>21.5831</u>	<u>0.0559</u>
FGD-best	**22.1411**	**0.0525**	**21.1517**	**0.0588**	**22.3001**	**0.0516**	**21.7213**	**0.0551**

率为 p，并添加高斯噪声 $\mathcal{N}(0,\ \sigma^2)$。对于 TNN、UTF、TCTF 和 GTNN-HOP，我们采用其原始论文中描述的初始化方案和超参数设置。条目“FGD-best”指采用小初始化的 FGD 所获得的最高峰值信噪比，而“FGD-ES”对应于基于验证进行提前停止所达到的峰值信噪比。对于这两种设置，初始化尺度设为 $\alpha = 10^{-5}$，步长设为 $\eta = 1e-3$。对于所有图像，FGD-ES、FGD-best 和 UTF 的管秩均设为 100。最大迭代次数为 2000。图 9–12 展示了不同方法在各种模型参数 ($(p,\ \sigma)$) 下对每幅图像得到的峰值信噪比和相对误差值。此外，图 13 展示了可视化重建结果。

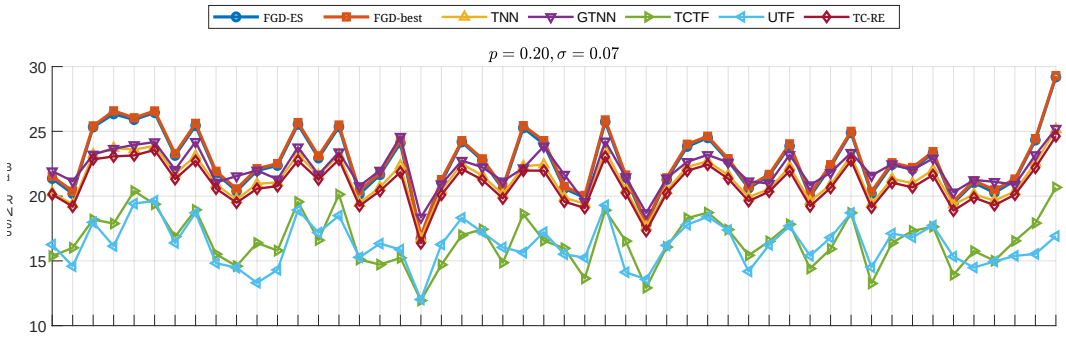
我们观察到，在大多数情况下，FGD-best 和 FGD-ES 的恢复性能最佳。此外，随着噪声水平增加，其他算法的性能显著下降，而采用小初始化的 FGD 受到的影响则小得多，这凸显了小初始化的优势。

J.2.2 视频补全实验

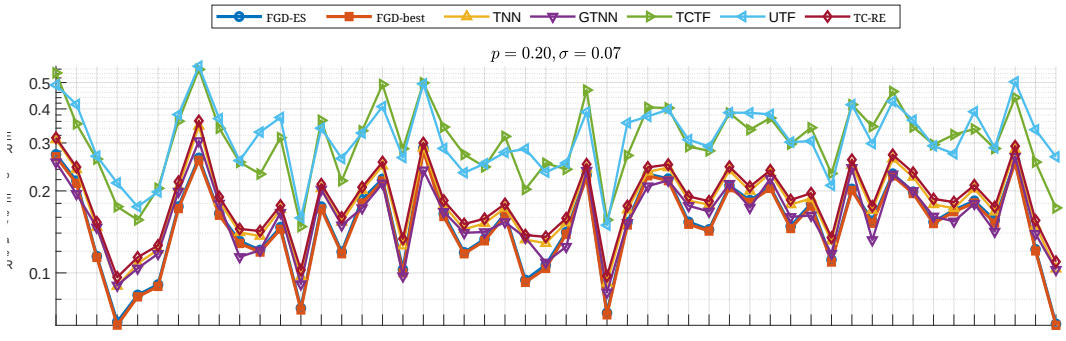
除了图像补全，我们还进行了含高斯噪声的视频补全实验。我们从 YUV视频序列数据集 [2] 中随机选取四段视频，提取每段视频的前 30 帧，组成尺寸为 $176 \times 144 \times 30$ 的张量，添加从 $\mathcal{N}(0,\ \sigma^2)$ 中抽取的高斯噪声，并再次采用采样率 p。由于 TCTF 在低采样率的视频补全场景下表现不佳，我们用 GTNN-HOP$_{0.6}$ 替代它；这是一种带有诱导稀疏正则项的非凸方法。对于 FGD-ES 和 FGD-best，初始化尺度设为 $\alpha = 10^{-5}$，步长设为 $\eta = 2e-4$。所有图像中，FGD-ES 、FGD-best 和 UTF 的管秩均设为 50。最大迭代次数为 4000。表 3-7 报告了所有方法在这四段视频上的峰值信噪比和相对误差，图14展示了各方法对 akiyo视频第一帧的重建结果。可以看出，我们的方法取得了最小的相对恢复误差和最高的峰值信噪比。此外，我们还评估了 FGD-best 和 FGD-ES 对管秩 R 选择的鲁棒性。图 15 所示结果表明，这两种方法在所有四段视频中对 R 的选择都具有很强的鲁棒性。

一个潜在问题是，基于梯度的方法对底层矩阵或张量的条件数较为敏感；当条件数较大时，收敛速度会变慢。因此，如何在控制噪声放大的同时加速 FGD，仍是未来研究的一个有趣方向。

[2] https://www.cnets.io/traces.cnets.io/trace.eas.asu.edu/YUV/index.html



(a)

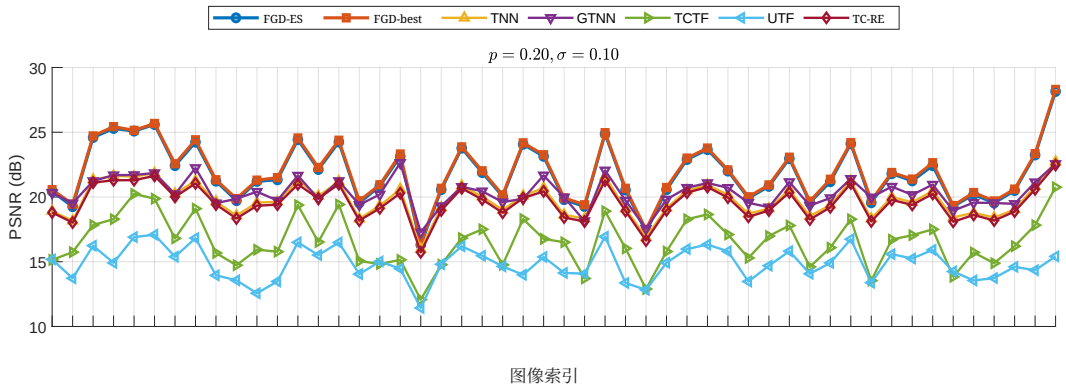


(b)

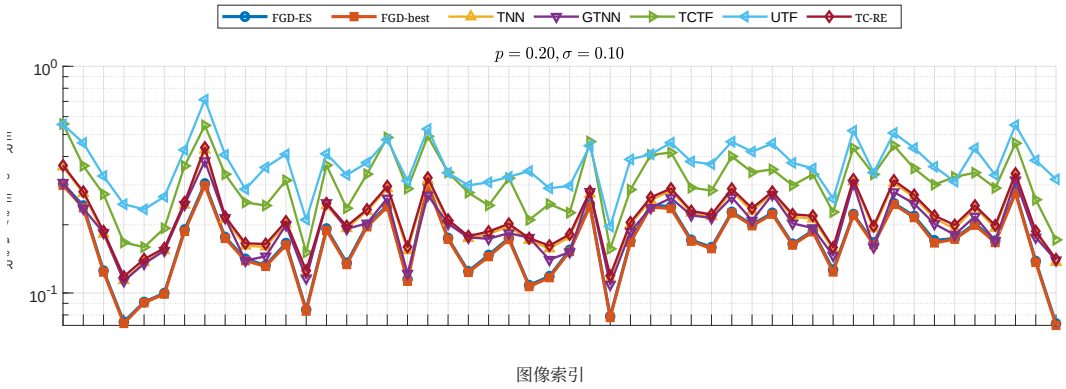
图 9：在采样率为 $p = 0.2$、噪声标准差为 $\sigma = 0.07$ 时，不同方法在 50 张图像上的峰值信噪比和相对误差对比。

表 4：在不同采样率和噪声水平下，各方法的平均峰值信噪比（PSNR）和平均相对误差（RE）对比。峰值信噪比越高、相对误差越小，重建质量越好。“FGD-ES”表示采用早停的 FGD，而 “FGD-best” 表示 FGD 在所有迭代次数中达到的最小误差。

方法	$p = 0.3$				$p = 0.4$			
	$\sigma = 0.03$		$\sigma = 0.05$		$\sigma = 0.03$		$\sigma = 0.05$	
	峰值信噪比↑	相对平方误差↓	峰值信噪比↑	相对平方误差↓	峰值信噪比↑	相对平方误差↓	峰值信噪比↑	相对平方误差↓
TNN	20.7258	0.0613	17.4371	0.0895	21.0157	0.0592	17.3267	0.0906
TC-RE	19.928	0.0671	17.1895	0.0920	20.2059	0.0650	16.5464	0.0991
UTF	6.3467	0.3207	6.4038	0.3186	5.9414	0.3360	4.7523	0.3853
GTNN-HOP$_{0.3}$	19.9438	0.0670	16.1328	0.1039	20.2975	0.0643	16.2845	0.1021
GTNN-HOP$_{0.6}$	20.5461	0.0625	16.8126	0.0961	20.8648	0.0603	16.9083	0.0951
FGD-ES	23.2899	0.0456	21.6321	0.0552	23.8244	0.0429	22.3928	0.0501
FGD-best	**23.4511**	**0.0448**	**21.8002**	**0.0541**	**23.8539**	**0.0427**	**22.5611**	**0.0496**



(a)

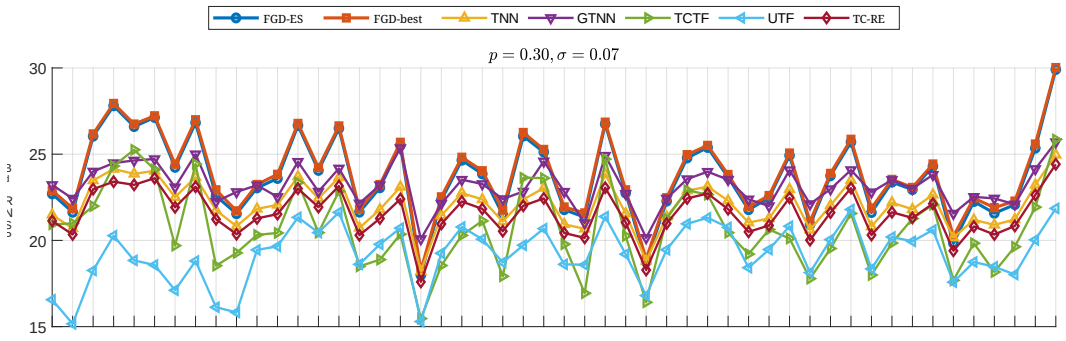


(b)

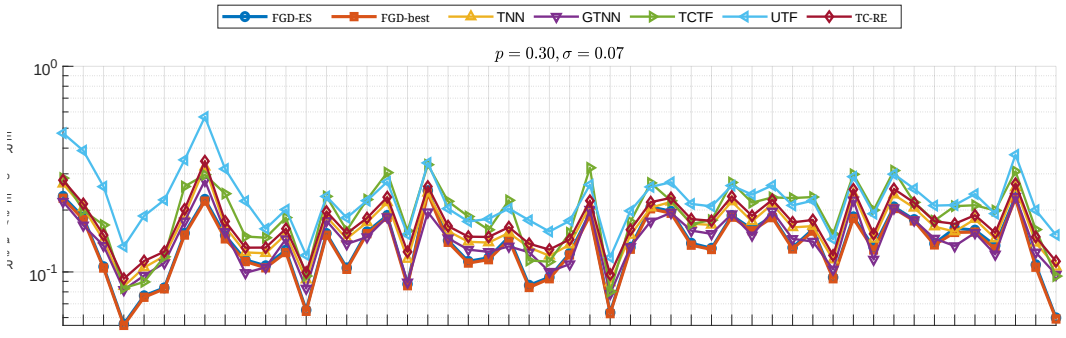
图 10：不同方法在 50 张图像上的峰值信噪比和相对误差对比，采样率为 $p = 0.2$，噪声标准差为 $\sigma = 0.1$。

表 5：“highway”视频在不同采样率和噪声水平下，各方法的平均峰值信噪比（PSNR）和平均相对误差（RE）对比。峰值信噪比越高、相对误差越小，表示重建质量越好。“FGD-ES”表示带提前停止的 FGD，“FGD-best”表示 FGD 在所有迭代次数中达到的最小误差。

方法	$p = 0.3$				$p = 0.4$			
	$\sigma = 0.03$		$\sigma = 0.05$		$\sigma = 0.03$		$\sigma = 0.05$	
	峰值信噪比 ↑	相对平方误差 ↓	峰值信噪比 ↑	相对平方误差 ↓	峰值信噪比 ↑	相对平方误差 ↓	峰值信噪比 ↑	相对平方误差 ↓
TNN	19.4802	0.0474	16.9433	0.0635	19.8369	0.0455	16.8423	0.0642
TC-RE	19.0227	0.0499	16.6549	0.0656	19.1568	0.0492	16.1336	0.0697
UTF	4.0060	0.2814	3.7117	0.2911	1.1904	0.3891	0.7714	0.4084
$\text{GTNN-HOP}_{0.3}$	19.3115	0.0483	16.1687	0.0694	19.7269	0.0461	16.3390	0.0680
$\text{GTNN-HOP}_{0.6}$	19.8894	0.0452	16.8541	0.0641	20.2802	0.0432	16.9568	0.0634
FGD-ES	20.6667	0.0413	20.0041	0.0446	20.7364	0.0410	20.3462	0.0429
FGD-best	**20.7000**	**0.0412**	**20.1063**	**0.0441**	**20.7278**	**0.0410**	**20.4994**	**0.0421**



(a)

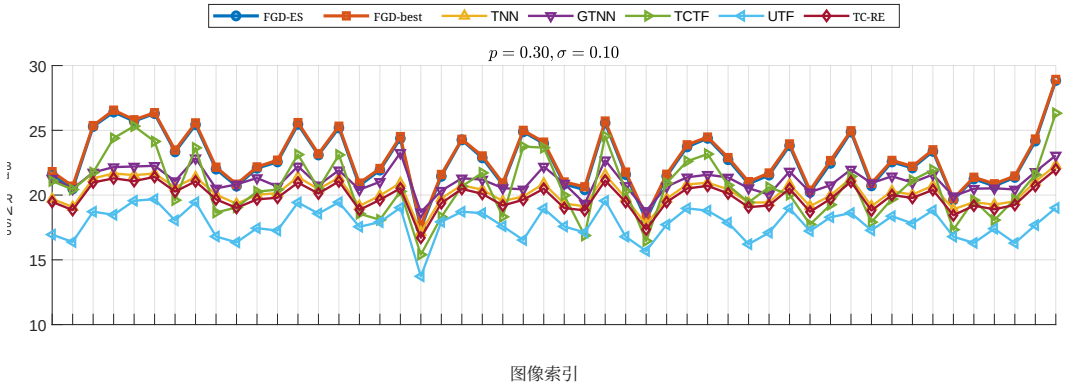


(b)

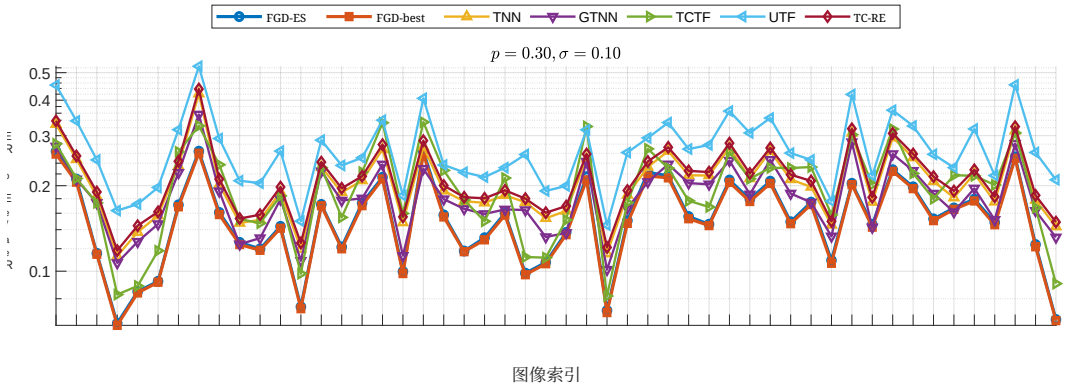
图 11：不同方法在采样率 $p = 0.3$和噪声标准差 $\sigma = 0.07$下，对 50 张图像的峰值信噪比和相对误差进行比较。

表6：“suzie”视频在不同采样率和噪声水平下，各方法的平均峰值信噪比（PSNR）和平均相对误差（RE）比较。峰值信噪比越高、相对误差越小，表示重建质量越好。“FGD-ES”表示采用提前停止的 FGD，而“FGD-best”表示 FGD 在所有迭代中达到的最小误差。

方法	$p = 0.3$				$p = 0.4$			
	$\sigma = 0.03$		$\sigma = 0.05$		$\sigma = 0.03$		$\sigma = 0.05$	
	峰值信噪比 ↑	相对平方误差 ↓	峰值信噪比 ↑	相对平方误差 ↓	峰值信噪比 ↑	相对平方误差 ↓	峰值信噪比 ↑	相对平方误差 ↓
TNN	19.3458	0.0717	16.6844	0.0974	19.8125	0.0679	16.7441	0.0967
TC-RE	19.2177	0.0728	16.5983	0.0984	19.2186	0.0728	16.0643	0.1046
UTF	11.4879	0.1772	10.2100	0.2053	10.8456	0.1908	8.6211	0.2465
GTNN-HOP$_{0.3}$	19.0898	0.0738	15.7370	0.1086	19.5613	0.0699	15.9544	0.1059
GTNN-HOP$_{0.6}$	19.6531	0.0692	16.4167	0.1005	20.0997	0.0657	16.5620	0.0988
FGD-ES	<u>20.5670</u>	<u>0.0623</u>	<u>19.4874</u>	<u>0.0705</u>	<u>20.7175</u>	<u>0.0612</u>	<u>20.1540</u>	<u>0.0653</u>
FGD-best	**20.5947**	**0.0621**	**19.6263**	**0.0694**	**20.7445**	**0.0610**	**20.2629**	**0.0645**



(a)



(b)

图 12：不同方法在 50 张图像上的峰值信噪比和相对误差对比，采样率为 $p = 0.3$，噪声标准差为 $\sigma = 0.1$。

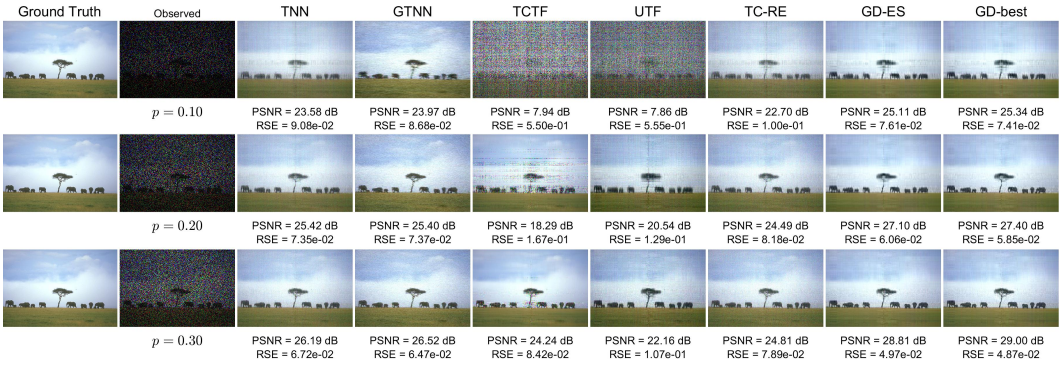


图 13：不同方法在不同采样率 p 下的图像恢复性能对比。噪声标准差为 $\sigma = 0.05$。观测条目中的 5% 用于验证。

表 7：在不同采样率和噪声水平下，“miss-america”视频各方法的平均峰值信噪比（PSNR）和平均相对误差（RE）对比。PSNR 越高、RE 越小，表示重建质量越好。“FGD-ES”表示采用提前停止的 FGD，而“FGD-best”表示 FGD 在所有迭代中达到的最小误差。

方法	$p=0.3$				$p=0.4$			
	$\sigma=0.03$		$\sigma=0.05$		$\sigma=0.03$		$\sigma=0.05$	
	峰值信噪比↑	相对平方误差↓	峰值信噪比↑	相对平方误差↓	峰值信噪比↑	相对平方误差↓	峰值信噪比↑	相对平方误差↓
TNN	21.2922	0.0831	17.8503	0.1235	21.3210	0.0828	17.5991	0.1271
TC-RE	20.6724	0.0892	17.5455	0.1279	20.3926	0.0921	16.8711	0.1382
UTF	9.4560	0.3245	9.0886	0.3386	9.3371	0.3290	8.0055	0.3835
GTNN-HOP$_{0.3}$	20.2760	0.0934	16.3891	0.1461	20.4505	0.0915	16.4938	0.1443
GTNN-HOP$_{0.3}$	20.9446	0.0865	17.1388	0.1340	21.0611	0.0853	17.1462	0.1339
FGD-ES	<u>23.8085</u>	<u>0.0622</u>	<u>22.9563</u>	<u>0.0686</u>	<u>23.8721</u>	<u>0.0617</u>	<u>23.4395</u>	<u>0.0649</u>
FGD-best	**23.8186**	**0.0621**	**23.0738**	**0.0677**	**23.8743**	**0.0617**	**23.5618**	**0.0640**

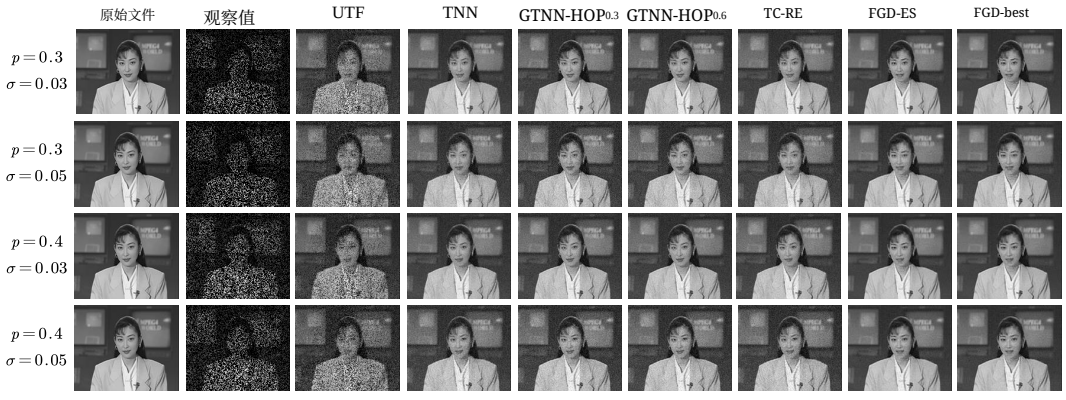


图14：视频“akiyo”在不同采样率 p 和噪声标准差 σ 下，不同方法的视频恢复性能对比。对于 FGD-ES，使用 5% 的观测条目进行验证。

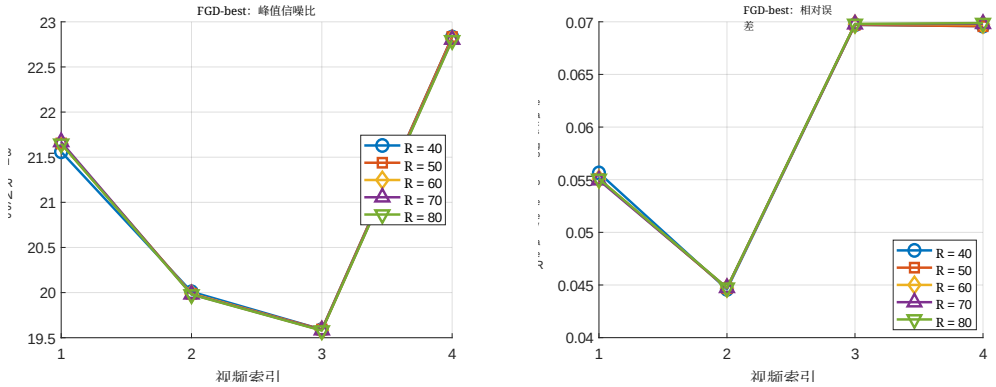


图15：评估不同管秩 R 对视频补全性能的影响。子图 (a) 展示了 FGD-best 在四个视频上的峰值信噪比，子图 (b) 展示了相应的相对误差。子图 (c) 报告了 FGD-ES 在四个视频上的峰值信噪比，子图 (d) 则展示了相应的相对误差。