arXiv:2601.00377v1 [math.ST] 1 Jan 2026

面向高维时间序列预测的稀疏 Tucker 分解与图正则化

Sijia Xia*、Michael K. Ng† 和 XiongjunZhang‡

2026年1月5日

摘要

现有的用于多元时间序列分析的向量自回归模型方法，采用低秩矩阵近似或Tucker分解来降低过参数化问题的维度。本文针对高维向量自回归时间序列，提出一种结合图正则化的稀疏 Tucker 分解方法。通过将时间序列转移矩阵堆叠为三阶张量，利用稀疏 Tucker 分解刻画转移三阶张量内部的重要交互，并减少参数数量。此外，在向量自回归模型中，采用图正则化来度量响应、预测变量和时间因子矩阵的局部一致性。研究表明，所提出的两种正则化技术能够实现更准确的参数估计。本文建立了所提方法估计量的非渐近误差界，该误差界低于现有矩阵或基于张量的方法。为求解所得模型，设计了一种近端交替线性化最小化算法，并在非常温和的条件下证明了其全局收敛性。在合成数据和真实世界数据集上开展了广泛的数值实验，结果验证了所提方法相较于现有最先进方法的优越性能。

关键词：高维向量自回归、稀疏 Tucker 分解、图正则化、近端交替线性化最小化算法

数学学科分类：62M10, 90C30

*华中师范大学数学与统计学院，中国武汉 430079（电子邮箱：si-jiax@mails.ccnu.edu.cn）。

†香港浸会大学数学系，中国香港九龙塘（电子邮箱：michael-ng@hkbu.edu.hk）。该作者的研究部分得到了广东省科学技术厅与香港高校“1+1+1”联合研究合作计划（UICR0800008-24）、中国国家重点研发计划项目（资助编号：2024YFE0202900）、研究资助局普通研究基金（资助编号：12300125）以及国家自然科学基金委员会与研究资助局联合项目（HKU769/21）的支持。

‡华中师范大学数学与统计学院及非线性分析与应用重点实验室（教育部），中国武汉 430079（电子邮箱：xjzhang@ccnu.edu.cn）。该作者的研究部分得到了国家自然科学基金委员会（资助编号：12171189）、湖北省自然科学基金（资助编号：2025AFB966）以及中央高校基本科研业务费（资助编号：CCNU24ai002）的支持。

1 引言

随着信息技术的快速发展，在当今数据丰富的环境中，各领域涌现出高维时间序列数据，例如金融 [43], 经济学 [41],生态学 [20], 和气象学 [16]。经典的向量自回归（VAR）模型因能够灵活地捕捉变量间的时间动态关系，被广泛用于对多元时间序列数据建模，并已应用于众多领域，例如神经影像学 [18], 信号处理 [4], 交通状态估计 [11]。考虑一个 m维零均值 p 阶 VAR 模型（记为 VAR(p)），其形式如下

$$\mathbf{y}_t = \mathbf{W}_1\mathbf{y}_{t-1} + \mathbf{W}_2\mathbf{y}_{t-2} + \cdots + \mathbf{W}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad 1 \le t \le T, \tag{1}$$

其中，$\{\mathbf{y}_t\}$ 是观测数据，其中 $\mathbf{y}_t = (y_{1t}, y_{2t}, \ldots, y_{mt})^H \in \mathbb{R}^m$，$\mathbf{W}_i \in \mathbb{R}^{m\times m}, i = 1, 2, \ldots, p$为转移矩阵，$p$ 是 VAR 模型的滞后阶数，T 表示样本量，$\boldsymbol{\varepsilon}_t$ 表示独立同分布（i.i.d.）误差，且 $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{mt})^H \in \mathbb{R}^m, \mathbb{E}(\boldsymbol{\varepsilon}_t) = 0$，$\mathrm{var}(\boldsymbol{\varepsilon}_t) < \infty$。

在实际应用中，m （1）中的维度通常非常大，这意味着系数参数的数量（即 m^2p）很大。因此，不受约束的 VAR(p) 模型很可能面临过参数化的难题，其中参数数量（例如，时变 VAR 中的系数矩阵）不可避免地超过观测值数量。因此，若无进一步限制，相应的 VAR 方法无法提供可靠的估计或准确的预测 [15]。在满足转移矩阵相关的某些正则条件时，可以实现高维VAR模型的估计一致性 [44]。例如，如果系数矩阵具有未观测到的低维结构，如稀疏性或低秩性，那么包括最小绝对收缩与选择算子（Lasso） [5]和核范数惩罚 [33]在内的结构诱导正则化方法，在时间序列满足高斯假设时能够给出一致估计。

另一方面，$\mathbf{W}_i$ 中的参数数量会随维度 m呈二次增长，这使得向量自回归模型难以应用于高维数据。为解决这一问题，一种常用方法是假设参数矩阵具有稀疏性，进而采用多种施加或诱导稀疏性的方法进行估计和变量选择，包括 ℓ_1 正则化 [5,21, 28, 51], 加权 ℓ_1 正则化 [50], 转移矩阵上的弱稀疏约束 [21],以及基于 ℓ_1 范数的非凹惩罚方法 [54]。然而，与 Lasso 的线性回归不同，时间序列数据具有时间与截面依赖性，而基于 ℓ_1 正则化的方法忽略了这一特性，这会严重影响基于稀疏正则化方法的估计量的准确性。

为解决 VAR 模型的过参数化问题以及时间序列数据中的数据依赖性，人们提出并研究了一些基于降秩回归的方法，参见 [4, 32, 33, 36]及其中的参考文献。这些论文假设转移矩阵为低秩，从而减少参数数量。例如，Basu 等人[4] 将转移矩阵组合成一个更大的矩阵，提出了一种适用于高维VAR模型的低秩加稀疏估计，但该模型仅讨论了一阶情形。Samadi等人 [40] 将包络模型融入降秩VAR模型，提出了降秩包络VAR模型，以高效地从复杂数据中提取相关信息。此外，Reinsel等人 [37]基于 $\mathbf{W}_i^H$堆叠矩阵的低秩性假设，提出了一种自回归指数模型。然而，基于降秩模型的矩阵方法只考虑特定堆叠矩阵的低秩性，因此只能沿一个方向降低这些参数矩阵的维度。

通过将转移矩阵重新表述为三阶张量，Wang 等人 [47] 开发了一种用于 VAR 时间序列建模的张量 Tucker分解方法，其中转移张量被分解为一个核心张量和三个稀疏因子矩阵，并可挖掘其低秩性

沿不同方向展开。然而，上述基于张量的方法未考虑时间序列数据的局部模式。Bahadori 等人 [3] 通过采用低 Tucker秩约束和空间拉普拉斯正则化，提出了一个统一的低秩张量学习框架，用于多变量时空分析，该框架可应用于 VAR 时间序列。此外，Harris 等人 [23] 提出了一种随时间变化的自回归模型，通过引入 CANDECOMP/PARAFAC (CP) [26] 分解以及时间上的平滑先验来处理多变量时间序列。然而，一般而言，CP 秩的确定是 NP 难的 [24]。此外，上述两项工作均没有理论保证。

1.1 我们的方案

本文提出一种采用图正则化的稀疏 Tucker 分解方法，用于高维 VAR 时间序列。具体而言，将（1）的转移矩阵堆叠成一个三阶张量，并利用稀疏 Tucker 分解探索转移张量沿不同维度的低秩性；当 Tucker 秩较小时，该方法可以大幅减少参数。此外，对核心张量施加稀疏性约束，可以进一步减少参数，并筛选出 VAR 时间序列在响应、预测变量和时间因子方面的重要变量。此外，利用因子矩阵上的图正则化刻画转移张量的局部模式，从而保留数据在响应、预测变量和时间因子方面的内在流形结构。随后，在一定条件下，我们建立了所提模型估计量的非渐近误差界，该误差界小于现有基于稀疏和低秩的方法所对应的误差界。我们开发了一种近端交替线性化最小化（PALM）算法来求解所得模型，并在非常宽松的条件下证明了其全局收敛性。模拟数据和真实世界数据集上的数值实验表明，与其他竞争方法相比，所提方法具有优越性。

本文其余部分安排如下。在下一节中，我们将介绍有关张量的一些符号和预备知识。在第 2 节中，我们提出一种用于高维 VAR 时间序列的稀疏 Tucker 分解与图正则化方法，并建立所提模型估计量的非渐近误差界。在第 3 节中，我们设计一种 PALM 算法来求解所得模型，并证明其全局收敛性。在第 4 节中，我们通过数值实验验证所提方法的有效性。最后，在第 5 节中给出总结性评述。所有引理和主要结果的技术性证明均列于附录。

1.2 预备知识

本文中使用的基本符号和记法汇总于表1。

对于任意矩阵 $\mathbf{A} \in \mathbb{C}^{n_1 \times n_2}$、$\mathbf{B} \in \mathbb{C}^{n_3 \times n_4}$，它们的克罗内克积 [17, 第 1.3 节。6] 定义为

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1n_2}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2n_2}\mathbf{B} \\ \vdots & \vdots & & \vdots \\ a_{n_1 1}\mathbf{B} & a_{n_1 2}\mathbf{B} & \cdots & a_{n_1 n_2}\mathbf{B} \end{pmatrix}.$$

对于任意两个张量 $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ 和 $\mathcal{Y} \in \mathbb{R}^{n_4 \times n_5 \times n_6}$，其张量外积记为 $\mathcal{X} \circ \mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times n_4 \times n_5 \times n_6}$ $\mathcal{X} \circ \mathcal{Y}_{i_1 i_2 i_3 i_4 i_5 i_6} = \mathcal{X}_{i_1 i_2 i_3} \mathcal{Y}_{i_4 i_5 i_6}$ $1 \le i_j \le$，定义为 ()，其中，，n_j $j = 1 \ldots 6$ k $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ $\in \mathbb{R}^{J \times n_k}$，。张量与矩阵 A 的 - 模乘积记为 $\mathcal{X} \times_k \mathbf{A}$，其大小为 $n_1 \times \cdots \times n_{k-1} \times J \times n_{k+1} \times \cdots \times n_d$ [29]。逐点方式

表 1：符号汇总。

符号	说明
a/a/A/$\mathcal{A}$	标量/向量/矩阵/张量
$\mathbf{I}_n$	大小为 $n \times n$ 的单位矩阵
$\mathcal{A}_{ijl}$	$\mathcal{A}$ 的 (i,j,l) 元素
$.^H$	共轭转置算子
$.^{-1}$	求逆算子
tr(·)	矩阵的迹
$\mathcal{A}^{\langle i \rangle}$	$\mathcal{A}$ 的第 i 个前向切片
$\mathcal{A}_{(i)}$	$\mathcal{A}$的第i 模展开
$\|\mathbf{a}\|_2$	向量 a 的 ℓ_2 范数
σ_j(A)	A 的第 j 大奇异值
$\|\cdot\|_0$	定义为非零元素个数的 ℓ_0 范数
$\|\mathbf{A}\|_*$	A 的核范数定义为 $\|\mathrm{A}\|_* := \sum_{j=1}^{\min\{n_1,n_2\}} \sigma_j(\mathrm{A})$
$\|\mathbf{A}\|$	A 的谱范数定义为 $\|\mathrm{A}\| := \sigma_1(\mathrm{A})$
$\|\mathbf{A}\|_F$	A 的弗罗贝尼乌斯范数定义为 $\|\mathrm{A}\|_F := \sqrt{\mathrm{Tr}(\mathrm{A}^H\mathrm{A})}$
$\langle \mathcal{A}, \mathcal{B} \rangle$	两个张量的内积定义为 $\langle \mathcal{A}, \mathcal{B} \rangle := \sum_{i=1}^{n_3} \mathrm{Tr}((\mathcal{A}^{\langle i \rangle})^H \mathcal{B}^{\langle i \rangle})$
vec($\mathcal{A}$)	按字典序将张量 $\mathcal{A}$向量化
$\|\mathcal{A}\|_\infty$	张量 ℓ_∞ 范数定义为 $\mathcal{A}$ $\|\mathcal{A}\|_\infty := \max_{ijl} \|\mathcal{A}_{ijl}\|$
$\|\mathcal{A}\|_1$	张量 ℓ_1 范数 $\mathcal{A}$ 定义为 $\|\mathcal{A}\|_1 := \sum_{i,j,l} \|\mathcal{A}_{ijl}\|$
$\|\mathcal{A}\|_F$	张量弗罗贝尼乌斯范数定义为 $\mathcal{A}$ $\|\mathcal{A}\|_F := \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle}$
$\delta_D(\cdot)$	集合 D 的指示函数为 $\delta_D(x) = 0$，当 $x \in D$时；否则为 $+\infty$

定义为

$$(\mathcal{X} \times_k \mathbf{A})_{i_1 \cdots i_{k-1} j i_{k+1} \cdots i_d} = \sum_{i_k=1}^{n_k} \mathcal{X}_{i_1 \cdots i_d} \mathbf{A}_{j i_k}.$$

张量 $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ 的模-k 矩阵化记为 $\mathcal{X}_{(k)}$，并将模-k纤维排列为所得矩阵的列。这里，张量元素 $(i_1, i_2, \ldots, i_d)$映射到矩阵元素 (i_k, j)，其中 $j = 1 + \sum_{m=1, m \neq k}^{d} (i_m - 1) L_m$ 满足 $L_m = \prod_{s=1, s \neq k}^{m-1} n_s$. 此外，$\mathcal{Y} = \mathcal{X} \times_k \mathrm{A}$等价于 $\mathcal{Y}_{(k)} = \mathrm{A}\mathcal{X}_{(k)}$。

接下来，我们回顾一般函数的次梯度定义，如下所示。

定义 **1** *[38,* 定义 8。*3]* 考虑一个函数 f： $\mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ 以及任意 $\mathbf{x} \in \mathrm{dom}(f)$。对于向量 $\mathbf{v} \in \mathbb{R}^n$，称

(a) v 是 f 在 x 处的正则次梯度，记作 $\hat{\partial} f(\mathrm{x})$，如果

$$\liminf_{\substack{\mathbf{y} \neq \mathbf{x} \\ \mathbf{y} \to \mathbf{x}}} \frac{1}{\|\mathbf{x} - \mathbf{y}\|_2} [f(\mathbf{y}) - f(\mathbf{x}) - \langle \mathbf{v}, \mathbf{y} - \mathbf{x} \rangle] \geq 0.$$

(b) 若存在序列 $\mathrm{x}^k \to \mathrm{x}$ $f(\mathrm{x}^k) \to f(\mathrm{x})$，$\mathrm{v}^k \in \hat{\partial} f(\mathrm{x}^k)$，使得 $\mathrm{v}^k \to \mathrm{v}$，则称 v 是 f 在 x 处的次梯度，记作 $\partial f(\mathrm{x})$。

2 稀疏 Tucker 分解与图正则化

在本节中，我们介绍所提出的方法。设 $\mathcal{W} \in \mathbb{R}^{m\times m\times p}$ 的每个前向切片为 $\mathcal{W}^{\langle i\rangle} = \mathbf{W}_i$、$i = 1$、2、……、$p$。则问题（1）可改写为

$$\begin{aligned}\mathbf{y}_t &= (\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_p)\begin{pmatrix}\mathbf{y}_{t-1}\\ \mathbf{y}_{t-2}\\ \vdots\\ \mathbf{y}_{t-p}\end{pmatrix} + \boldsymbol{\varepsilon}_t \\ &= \mathcal{W}_{(1)}\mathbf{x}_t + \boldsymbol{\varepsilon}_t,\end{aligned} \tag{2}$$

其中 $\mathbf{x}_t = (\mathbf{y}_{t-1}^H, \mathbf{y}_{t-2}^H, \ldots, \mathbf{y}_{t-p}^H)^H \in \mathbb{R}^{mp\times 1}$。因此，可以利用 $\mathcal{W}$ 的 Tucker分解来探究 VAR 模型不同模态的低秩性，即

$$\mathcal{W} = \mathcal{G} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \times_3 \mathbf{A}_3 =: [\![\mathcal{G}; \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3]\!], \tag{3}$$

其中 $\mathcal{G} \in \mathbb{R}^{r_1\times r_2\times r_3}$ 是核心张量，$\mathbf{A}_1 \in \mathbb{R}^{m\times r_1}$、$\mathbf{A}_2 \in \mathbb{R}^{m\times r_2}$ 和 $\mathbf{A}_3 \in \mathbb{R}^{p\times r_3}$ 分别是响应、预测变量和时间因子矩阵。因此，问题（2）可以表示为

$$\mathbf{y}_t = \mathcal{W}_{(1)}\mathbf{x}_t + \boldsymbol{\varepsilon}_t = (\mathcal{G} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \times_3 \mathbf{A}_3)_{(1)}\mathbf{x}_t + \boldsymbol{\varepsilon}_t.$$

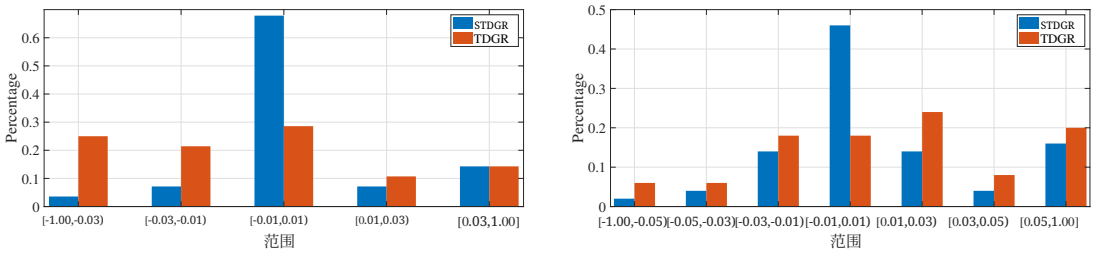
此外，为了获得 $\mathcal{W}$的唯一 Tucker分解，我们对因子矩阵 [29], 施加列正交约束，即

$$\mathbf{A}_i^H\mathbf{A}_i = \mathbf{I}_{r_i}, i = 1, 2, 3.$$

注意到

$$\mathbf{A}_1^H\mathbf{y}_t = \mathcal{G}_{(1)}(\mathbf{A}_3 \otimes \mathbf{A}_2)^H\mathbf{x}_t + \mathbf{A}_1^H\boldsymbol{\varepsilon}_t.$$

令 $\mathbf{A}_1^H\mathbf{y}_t := \mathbf{f}_t = (f_{1t}, f_{2t}, \ldots, f_{r_1t})^H \in \mathbb{R}^{r_1}$，表示 $\mathbf{y}_t$ 的 m 个变量上的 r_1 个响应因子。此外，$f_{it} = \sum_{j=1}^m (\mathbf{A}_1)_{ji}y_{jt}$ 是第 i 个响应因子。如果 $(\mathbf{A}_1)_{ji} = 0$，则 f_{it} 与y_{jt}无关。在这种情况下，$\mathbf{A}_1$ 可以解释为响应因子的载荷。类似地，$\mathbf{A}_2$ 和 $\mathbf{A}_3$ 可以分别解释为预测变量因子和时间因子的载荷，详见[47]以获取更多讨论。



(a) CCD　　(b) PeMS04

图1：非稀疏模型（TDGR）和稀疏模型（STDGR）分别在 CCD 和 PeMS04数据集上获得的 $\mathcal{G}$ 元素分布。

在 Tucker 分解中，核心张量 $\mathcal{G}$ 的元素反映了不同模态矩阵中各分量（列）之间的相互作用和连接，只保留不同模态（响应、预测变量和时间）分量之间最显著的连接。

因此，我们对核心张量施加稀疏性约束。在图1中，我们通过 CCD 和 PeMS04 数据集上的数值示例展示核心张量稀疏性约束的有效性（详见第4.4 节和4.5节）。图中展示了通过稀疏模型（STDGR）和非稀疏模型（TDGR）得到的 $\mathcal{G}$ 元素分布。显然，许多 $\mathcal{G}$ 元素都接近于零（在 $[-0.01, 0.01$中），并且识别出了不同模态（响应、预测变量和时间）分量之间的重要连接。

注意到，系数矩阵 $\mathbf{A}_i$ 是响应、预测变量或时间模式下采样数据的低维表示，能够反映采样数据在低维子空间中的相似性。因此，对于 $\mathbf{A}_i$ 的 Tucker分解，我们利用图先验构建其每个因子矩阵，以刻画 VAR 中采样数据沿响应、预测变量和时间模式的相似性。

本文提出一种结合因子矩阵图正则化的稀疏 Tucker 分解模型（简称 STDGR），如下所示:

$$\begin{aligned}\min_{\mathcal{G},\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3} & \frac{1}{2T}\sum_{t=1}^{T}\left\|\mathbf{y}_t-\left(\mathcal{G}\times_1\mathbf{A}_1\times_2\mathbf{A}_2\times_3\mathbf{A}_3\right)_{(1)}\mathbf{x}_t\right\|_2^2+\beta\|\mathcal{G}\|_1+\sum_{i=1}^{3}\alpha_i\operatorname{tr}\left(\mathbf{A}_i^H\mathbf{L}_i\mathbf{A}_i\right) \\ \text{s.t. } & \|\mathcal{G}\|_\infty\le c, \mathbf{A}_i^H\mathbf{A}_i=\mathbf{I}_{r_i}, i=1,2,3,\end{aligned} \tag{4}$$

其中，$\beta>0$、$\alpha_i>0$、$c>0$ 为给定参数，$\mathbf{L}_i$ 为图拉普拉斯矩阵，$i=1$、2、3。这里注意，对于任意矩阵 $\mathbf{A}\in\mathbb{R}^{m\times n}$，其图拉普拉斯正则化定义为 [35]

$$\sum_{i,j=1}^{m}\frac{1}{2}z_{ij}\left\|\mathbf{a}_i-\mathbf{a}_j\right\|_2^2=\operatorname{tr}\left(\mathbf{A}^H\mathbf{L}\mathbf{A}\right), \tag{5}$$

其中，$\mathbf{a}_i$ 表示 $\mathbf{A}$ 的第 i 行，

$$\mathbf{L}=\mathbf{D}-\mathbf{Z} \tag{6}$$

表示由 $\mathbf{Z}$ 的每个元素构成的拉普拉斯矩阵z_{ij}，并且 $\mathbf{D}\in\mathbb{R}^{m\times m}$ 是一个对角矩阵，其 i个对角元素为 $\sum_{j=1}^{m}z_{ij}$，$i=1$，.....，m [13]。实际上，$\mathbf{L}_i$ 是在数值示例第 4 节中预先给定或构造的，我们将讨论如何在实现中构造 $\mathbf{L}_i$。

为防止向量自回归过程中出现非平稳性，我们假设核心张量 $\mathcal{G}$ 有上界，即 $\|\mathcal{G}\|_\infty\le c$。模型（4）中对因子矩阵施加图拉普拉斯正则化，以探究系数张量低维子空间的局部一致性。因子矩阵 $\mathbf{A}_1$,$\mathbf{A}_2$, 和 $\mathbf{A}_3$ 分别表示 $\mathcal{W}$ 沿响应、预测变量和时间模式的低维子空间载荷 [47]。反过来，这也能反映向量自回归采样数据的相似性。以交通数据为例，相邻传感器呈现相似模式和时间相关性 [39], ，并且可能具有相似的低维载荷 [9], ；对响应载荷 $\mathbf{A}_1$ 和预测变量载荷 $\mathbf{A}_2$ 施加图拉普拉斯正则化，便能有效捕捉这些特征。对时间载荷矩阵 $\mathbf{A}_3$ 施加拉普拉斯正则化，可以捕捉 p 时间滞后之间的时间相似性，从而使相邻滞后具有相似的子空间载荷。此外，模型（4）中的稀疏 Tucker 分解用于降维，并能解决向量自回归中的过参数化问题。事实上，采用稀疏 Tucker 分解的向量自回归模型仅含有 $mr_1+mr_2+pr_3+\tilde{s}r_1r_2r_3$个参数，其中 $0<\tilde{s}<1$ 表示核心张量的稀疏率。这少于模型（1）的参数数量（即 m^2p），也少于降秩模型 [33] 的参数数量（即 $(mp+m-r_1)\,r_1$）。

注 1 与模型 *[47],* 中的模型（*4*）相比，该模型在 Tucker分解中使用了稀疏核心张量，能够进一步减少转移张量的参数。此外，图

正则化被应用于模型（4）中，以刻画因子矩阵的局部一致性，从而反映采样数据在响应、预测变量和时间模式上的相似性。

2.1 非渐近误差界

在本小节中，我们建立模型（4）的估计量的非渐近误差界。记 $\widehat{\mathcal{G}}, \widehat{\mathbf{A}}_1, \widehat{\mathbf{A}}_2, \widehat{\mathbf{A}}_3$（4）的最优解为 $(\mathcal{G}, \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3)$。令 $\mathcal{W} = \mathcal{G} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \times_3 \mathbf{A}_3$。为推导$\widehat{\mathcal{W}}$的非渐近误差界，我们首先作出以下假设。

假设1 矩阵多项式 $\mathbf{W}(z) = \mathbf{I}_m - \sum_{i=1}^{p} \mathbf{W}_i z^i, z \in \mathbb{C}$ 满足 $\det(\mathbf{W}(z)) \neq 0$ 对所有 $|z| = 1$均成立，其中 $\mathbb{C}$ 表示复数域，且 $\mathbf{W}_i$ 定义见（1）。

假设2 （ $\{\boldsymbol{\varepsilon}_t\}$ ）中的误差向量 1）是均值为零、协方差矩阵 $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$正定的独立同分布高斯随机向量。

假设 3 核心张量 $\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ 至多有 s 个非零元素。并定义 $\bar{s} = 1 - s/(r_1 r_2 r_3) \in [0, 1]$，其表示 $\mathcal{G}$ 的稀疏度，即零元素所占的比例。

假设 4 对于 $i = 1, 2, 3, ,$ $\mathcal{W}_{(i)}$ 在 (2) 中的非零奇异值满足 $\sigma_{j-1}^2(\mathcal{W}_{(i)}) - \sigma_j^2(\mathcal{W}_{(i)}) \geq \rho \sigma_{j-1}^2(\mathcal{W}_{(i)})$，$j = 2, \ldots, r_i$，其中 $\rho > 0$ 是一个常数。

假设5 记 $\Delta_{\mathbf{A}_i} := \widehat{\mathbf{A}}_i - \mathbf{A}_i$，对于 $i = 1, 2, 3$。假设 $\Delta_{\mathbf{A}_i}$ 满足 $\sum_{l=1}^{r_i} \sum_{j=1}^{n_i} (\Delta_{\mathbf{A}_i})_{jl}^2 \gg \sum_{l=1}^{r_i} \frac{1}{n_i} (\sum_{j=1}^{n_i} (\Delta_{\mathbf{A}_i})_{jl})^2$，其中 $n_1 = n_2 = m$ 和 $n_3 = p$。

假设 1 保证了向量自回归过程 [5]的稳定性。假设 2 用于对向量自回归模型 [5]应用集中不等式。假设 3 反映了核心张量 $\mathcal{G}$的稀疏度。假设 4进一步确保 $\mathcal{W}$ 的每个展开矩阵的奇异值充分分离，从而消除不可辨识性，并使我们能够根据引理 7确定扰动误差的上界。最后，假设 5 对于图正则化项的分析至关重要。

令 $\mathbf{W}^H(z)$ 表示 $\mathbf{W}(z)$的共轭转置。根据 [5, 47], , 我们定义

$$\mu_{\min}(\mathbf{W}) := \min_{|z|=1} \lambda_{\min}(\mathbf{W}^H(z)\mathbf{W}(z)), \ \mu_{\max}(\mathbf{W}) := \max_{|z|=1} \lambda_{\max}(\mathbf{W}^H(z)\mathbf{W}(z)),$$

其中 $\lambda_{\min}(\cdot)$ 和 $\lambda_{\max}(\cdot)$ 分别表示矩阵的最小和最大特征值。此外，我们假设模型（4）中的核心张量还满足约束条件 $\mathcal{U} := \{\mathcal{G} : \ \|\mathcal{G}\|_0 \leq s\}$。现在，我们在以下定理中给出（4）的估计量的误差界。

定理 1 假设假设 1-5 成立。若 VAR 模型的样本量满足 $T \geq \max\{a_1 \log(m^2 p), \frac{8c^2}{a_4} \bar{p} \min(\log(mp), \log(21emp/\bar{p}))\}$，且正则化参数满足 $\alpha_i \geq \lambda_2 \sqrt{\frac{6c^2}{s}} (\mathrm{L}_i) \log(m^2 p) / T, \beta \geq 2\pi\vartheta a_2 \sqrt{\log(m^2 p)/T}$，则以至少 $1 - 6\exp(-a\log(m^2 p)) - 2\exp(-\bar{p}\min(\log\ mp, \log(21emp/\bar{p})))$ 的概率，估计量 (4) 满足

$$\|\widehat{\mathcal{W}} - \mathcal{W}\|_F \leq \frac{4\beta(\sqrt{\tilde{s}} + b\sqrt{s}) + \hat{a}\kappa + \bar{c}\sum_{i=1}^{3} \alpha_i l_i \eta_i}{\varpi + 2\kappa\sqrt{s/(r_1 r_2 r_3)^2}}, \tag{7}$$

and

$$\frac{1}{T}\sum_{t=1}^{T} \|\widehat{\mathcal{W}}_{(1)}\mathbf{x}_t - \mathcal{W}_{(1)}\mathbf{x}_t\|_2^2 \leq \frac{\lambda_{\max}(\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) \left(4\beta(\sqrt{\tilde{s}} + b\sqrt{s}) + \hat{a}\kappa + \bar{c}\sum_{i=1}^{3} \alpha_i l_i \eta_i\right)^2}{\mu_{\min}(\mathbf{W}) \left(\varpi + 2\kappa\sqrt{s/(r_1 r_2 r_3)^2}\right)^2}, \tag{8}$$

其中 $a_1, a_2 \geq 1, a, \hat{a}, a_4, \bar{c}, b > 0$ 为绝对常数，$\tilde{s}, \bar{p}$ 依赖于 $\bar{s}, \kappa = \sqrt{(\log(m^2 p))/T}$,$\overline{l_i = \|\mathrm{L}_i\|_F}$, $\varpi = \lambda_{\min}(\mathbf{\Sigma}_\varepsilon)/\mu_{\max}(\mathrm{W})$, $\eta_i = \left(3\sigma_1(\mathcal{W}_{(i)}) + c\sqrt{r_1 r_2 r_3}\right)\sqrt{\sum_{j=1}^{r_i} \sigma_j^4}(\mathcal{W}_{(i)})$, $\overline{\varsigma = \lambda_{\min}(\mathbf{\Sigma}_\varepsilon)/\mu_{\max}(\mathrm{W})\ \lambda_{\max}(\mathbf{\Sigma}_\varepsilon)/\mu_{\min}(\mathrm{W})}$, $\vartheta = \lambda_{\max}(\mathbf{\Sigma}_\varepsilon)(1 + \mu_{\max}(\mathrm{W})/\mu_{\min}(\mathrm{W}))$，且 $\lambda_2(\mathrm{L}_i) > 0$ 表示拉普拉斯矩阵 L_i 的第二小特征值，$i = 1, 2, 3$。

在定理 1 中，我们为模型 (4) 建立了一个非渐近误差界。当 Tucker秩 $\widehat{(r_1, r_2, r_3)}$ 固定时，(7) 中的不等式表明，随着 $T \to \infty$，$\mathcal{W}$ 是一个相合估计量。事实上，(7) 和 (8) 中的估计误差界和预测误差界分别为 $O_p((\sqrt{s} + \sqrt{\tilde{s}})\sqrt{\log(m^2 p)/T})$ 和$O_p((\sqrt{s} + \sqrt{\tilde{s}})^2 \log(m^2 p)/T)$。

注 2 如果 $\tilde{s} \leq s$，其中 $\tilde{s}$ 定义见 (75)，则 (7) 中的误差界可简化为

$$O_p(\sqrt{s \log(m^2 p)/T}). \tag{9}$$

现在我们比较 定理 1 中的估计误差界与现有的向量自回归建模方法。

- Basu 等人 [5] 研究了具有稀疏转移矩阵的平稳高斯向量自回归(p)模型的估计。相应的 Lasso 估计量为

$$\widehat{\mathcal{W}}_{\text{Lasso}} = \arg\min_{\mathcal{W}} \frac{1}{T}\sum_{t=1}^{T} \|\mathbf{y}_t - \mathcal{W}_{(1)}\mathbf{x}_t\|_2^2 + \lambda_{\text{Lasso}}\|\mathcal{W}_{(1)}\|_1.$$

  已有研究表明，估计量 $\widehat{\mathcal{W}}_{\text{Lasso}}$ 满足 $\|\widehat{\mathcal{W}}_{\text{Lasso}} - \mathcal{W}\|_F = O_p(\sqrt{k_0 \log(m^2 p)/T})$[5, 命题 4。1], 其中 $k_0 = \|\mathcal{W}\|_0$。在这种情况下，可以看出，当 1 中的定理误差界小于 [5] 时 $s \leq k_0$，其小于 [5] 中的 $s \leq k_0$Lasso 估计量的误差界。通常，在实践中，核心张量中的非零元素数量远小于完整张量中的非零元素数量。

- Negahban 等人 [33] 提出使用核范数（NN）刻画 VAR 模型中转移矩阵的低秩性，但他们的分析仅限于 VAR(1) 情形。我们将 VAR(1) 模型扩展到 VAR(p) 情形，得到

$$\widehat{\mathcal{W}}_{\text{NN}} = \arg\min \frac{1}{T}\sum_{t=1}^{T} \|\mathbf{y}_t - \mathcal{W}_{(1)}\mathbf{x}_t\|_2^2 + \lambda_{\text{NN}}\|\mathcal{W}_{(1)}\|_*. \tag{10}$$

  此外，已有研究表明，$\|\widehat{\mathcal{W}}_{\text{NN}} - \mathcal{W}\|_F = O_p(\sqrt{r_1 mp/T})$ [47, 备注 10], 其中r_1 是 $\mathcal{W}$的模-1 展开矩阵的秩。因此，如果 $\frac{s}{r_1} \leq \frac{mp}{\log(m^2 p)}$，核范数方法的估计误差界大于（9）中的误差界。在现实场景中，这一条件很容易满足，因为 s 远小于 $r_1 r_2 r_3$，r_1，r_2 远小于 m，且 r_3 远小于 p。

- 最近，Wang 等人 [47] 提出了一种针对向量自回归（VAR）的 ℓ_1-惩罚稀疏高阶降秩估计量，其中系数张量 $\mathcal{W}$ 采用高阶奇异值分解（HOSVD）的形式，且因子矩阵是稀疏的。当 A_i 的每一列$\widehat{\text{至多有}}$ s_i个非零元素时，$i = 1$、2、3，所得估计量记为 $\mathcal{W}$SHORR，满足$\|\widehat{\mathcal{W}}\text{SHORR} - \mathcal{W}\|_F = O_p(\sqrt{(s_1 s_2 s_3)\log(m^2 p)/T})$ [47, 定理 2]。因此，若 $s \leq s_1 s_2 s_3$，则我们在（9）中的估计量的误差界小于 [47]中的误差界。这里的条件 $s \leq s_1 s_2 s_3$ 在实际应用中很容易满足，因为 s远小于 $r_1 r_2 r_3$、$1 \leq s_i \leq m$、$i = 1$、2，$1 \leq s_3 \leq p$远小于 m，$i = 1$、2，r_3远小于 p，且通常 s_i 不会太小，$i = 1$、2、3。

3 近端交替线性化最小化算法

令 $\mathbf{A}_i = \mathbf{U}_i$ 并将这些等式约束代入目标函数，模型（4）可改写如下：

$$\begin{aligned}\min_{\substack{\mathcal{G},\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3,\\ \mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3}} & \frac{1}{2T}\sum_{t=1}^{T}\left\|\mathbf{y}_t-\left(\mathcal{G}\times_1\mathbf{A}_1\times_2\mathbf{A}_2\times_3\mathbf{A}_3\right)_{(1)}\mathbf{x}_t\right\|_2^2+\beta\|\mathcal{G}\|_1\\ &+\sum_{i=1}^{3}\alpha_i\operatorname{tr}\left(\mathbf{U}_i^H\mathbf{L}_i\mathbf{U}_i\right)+\sum_{i=1}^{3}\frac{\gamma_i}{2}\|\mathbf{U}_i-\mathbf{A}_i\|_F^2\\ \text{s.t. } & \|\mathcal{G}\|_\infty\le c,\mathbf{A}_i^H\mathbf{A}_i=\mathbf{I}_{r_i},i=1,2,3,\end{aligned}\tag{11}$$

其中 $\gamma_i>0$ 为给定常数。根据指示函数的定义，模型（11）可改写为以下无约束优化问题：

$$\begin{aligned}\min_{\substack{\mathcal{G},\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3,\\ \mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3}} & \frac{1}{2T}\sum_{t=1}^{T}\left\|\mathbf{y}_t-\left(\mathcal{G}\times_1\mathbf{A}_1\times_2\mathbf{A}_2\times_3\mathbf{A}_3\right)_{(1)}\mathbf{x}_t\right\|_2^2+\beta\|\mathcal{G}\|_1\\ &+\sum_{i=1}^{3}\alpha_i\operatorname{tr}\left(\mathbf{U}_i^H\mathbf{L}_i\mathbf{U}_i\right)+\sum_{i=1}^{3}\frac{\gamma_i}{2}\|\mathbf{U}_i-\mathbf{A}_i\|_F^2+\sum_{i=1}^{3}\delta_{\mathfrak{B}_i}(\mathbf{A}_i)+\delta_{\mathcal{D}}(\mathcal{G}),\end{aligned}\tag{12}$$

其中 $\mathfrak{B}_i:=\{\mathrm{A}_i:\mathrm{A}_i^H\mathrm{A}_i=\mathrm{I}_{r_i}\},i=1,2,3,\mathcal{D}:=\{\mathcal{G}:\ \|\mathcal{G}\|_\infty\le c\}$。

现在，我们设计一种近端交替线性化最小化（PALM）算法 [8] 来求解模型（12）。在继续之前，为简便起见，我们记

$$\begin{aligned}F(\mathcal{G},\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3,\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3):=&\frac{1}{2T}\sum_{t=1}^{T}\left\|\mathbf{y}_t-\left(\mathcal{G}\times_1\mathbf{A}_1\times_2\mathbf{A}_2\times_3\mathbf{A}_3\right)_{(1)}\mathbf{x}_t\right\|_2^2\\ &+\beta\|\mathcal{G}\|_1+\sum_{i=1}^{3}\alpha_i\operatorname{tr}\left(\mathbf{U}_i^H\mathbf{L}_i\mathbf{U}_i\right)+\sum_{i=1}^{3}\frac{\gamma_i}{2}\|\mathbf{U}_i-\mathbf{A}_i\|_F^2+\sum_{i=1}^{3}\delta_{\mathfrak{B}_i}(\mathbf{A}_i)+\delta_{\mathcal{D}}(\mathcal{G}),\end{aligned}\tag{13}$$

$$Q([\![\mathcal{G};\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3]\!]):=\frac{1}{2T}\sum_{t=1}^{T}\left\|\mathbf{y}_t-\left(\mathcal{G}\times_1\mathbf{A}_1\times_2\mathbf{A}_2\times_3\mathbf{A}_3\right)_{(1)}\mathbf{x}_t\right\|_2^2,\tag{14}$$

and

$$\Psi(\mathcal{G},\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3,\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3):=Q([\![\mathcal{G};\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3]\!])+\sum_{i=1}^{3}\frac{\gamma_i}{2}\|\mathbf{U}_i-\mathbf{A}_i\|_F^2.\tag{15}$$

因此，我们得到

$$\begin{aligned}F(\mathcal{G},\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3,\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3)=&\Psi(\mathcal{G},\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3,\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3)+\beta\|\mathcal{G}\|_1\\ &+\sum_{i=1}^{3}\alpha_i\operatorname{tr}\left(\mathbf{U}_i^H\mathbf{L}_i\mathbf{U}_i\right)+\sum_{i=1}^{3}\delta_{\mathfrak{B}_i}\left(\mathbf{A}_i\right)+\delta_{\mathcal{D}}(\mathcal{G}).\end{aligned}$$

注意，损失函数的梯度 Q相对于系数张量 $\mathcal{W}$ 可表示为 [45,附录 D。2]

$$\nabla Q([\![\mathcal{G};\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3]\!])=\frac{1}{T}\sum_{t=1}^{T}\left(\mathcal{W}_{(1)}\mathbf{x}_t-\mathbf{y}_t\right)\circ\mathbf{X}_t=\frac{1}{T}\sum_{t=1}^{T}\left(-\boldsymbol{\varepsilon}_t\right)\circ\mathbf{X}_t,\tag{16}$$

其中 $\mathbf{X}_t = (\mathbf{y}_{t-1},\ \mathbf{y}_{t-2},\ \ldots\ldots,\ \mathbf{y}^{t-p}) \in \mathbb{R}^{m\times p}$。根据 [22, 引理 2。1]， Q的偏梯度如下:

$$\begin{cases} \nabla_{\mathcal{G}} Q([\![\mathcal{G};\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3]\!]) = \nabla Q([\![\mathcal{G};\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3]\!]) \times_1 \mathbf{A}_1^H \times_2 \mathbf{A}_2^H \times_3 \mathbf{A}_3^H, \\ \nabla_{\mathbf{A}_1} Q([\![\mathcal{G};\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3]\!]) = (\nabla Q([\![\mathcal{G};\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3]\!]))_{(1)} (\mathbf{A}_3 \otimes \mathbf{A}_2)\, \mathcal{G}_{(1)}^H, \\ \nabla_{\mathbf{A}_2} Q([\![\mathcal{G};\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3]\!]) = (\nabla Q([\![\mathcal{G};\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3]\!]))_{(2)} (\mathbf{A}_1 \otimes \mathbf{A}_3)\, \mathcal{G}_{(2)}^H, \\ \nabla_{\mathbf{A}_3} Q([\![\mathcal{G};\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3]\!]) = (\nabla Q([\![\mathcal{G};\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3]\!]))_{(3)} (\mathbf{A}_2 \otimes \mathbf{A}_1)\, \mathcal{G}_{(3)}^H. \end{cases} \tag{17}$$

接下来，我们可以推导 Ψ （定义见（15））关于 $\mathcal{G}$、$\mathbf{A}_i$、$\mathbf{U}_i$ 的梯度如下:

$$\begin{cases} \nabla_{\mathcal{G}} \Psi(\mathcal{G},\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3,\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3) = \nabla_{\mathcal{G}} Q([\![\mathcal{G};\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3]\!]), \\ \nabla_{\mathbf{A}_1} \Psi(\mathcal{G},\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3,\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3) = \nabla_{\mathbf{A}_1} Q([\![\mathcal{G};\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3]\!]) - \gamma_1(\mathbf{U}_1 - \mathbf{A}_1), \\ \nabla_{\mathbf{A}_2} \Psi(\mathcal{G},\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3,\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3) = \nabla_{\mathbf{A}_2} Q([\![\mathcal{G};\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3]\!]) - \gamma_2(\mathbf{U}_2 - \mathbf{A}_2), \\ \nabla_{\mathbf{A}_3} \Psi(\mathcal{G},\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3,\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3) = \nabla_{\mathbf{A}_3} Q([\![\mathcal{G};\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3]\!]) - \gamma_3(\mathbf{U}_3 - \mathbf{A}_3), \\ \nabla_{\mathbf{U}_1} \Psi(\mathcal{G},\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3,\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3) = \gamma_1(\mathbf{U}_1 - \mathbf{A}_1), \\ \nabla_{\mathbf{U}_2} \Psi(\mathcal{G},\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3,\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3) = \gamma_2(\mathbf{U}_2 - \mathbf{A}_2), \\ \nabla_{\mathbf{U}_3} \Psi(\mathcal{G},\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3,\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3) = \gamma_3(\mathbf{U}_3 - \mathbf{A}_3). \end{cases} \tag{18}$$

然后，PALM 算法的迭代框架如下:

$$\begin{cases} \mathcal{G}^{k+1} \in \arg\min\limits_{\mathcal{G}} \left\langle \nabla_{\mathcal{G}} \Psi(\mathcal{G}^k,\mathbf{A}_1^k,\mathbf{A}_2^k,\mathbf{A}_3^k,\mathbf{U}_1^k,\mathbf{U}_2^k,\mathbf{U}_3^k), \mathcal{G} - \mathcal{G}^k \right\rangle + \beta\|\mathcal{G}\|_1 \\ \qquad\qquad + \delta_{\mathcal{D}}(\mathcal{G}) + \dfrac{\rho_1}{2}\|\mathcal{G} - \mathcal{G}^k\|_F^2, \\ \mathbf{A}_1^{k+1} \in \arg\min\limits_{\mathbf{A}_1} \left\langle \nabla_{\mathbf{A}_1} \Psi(\mathcal{G}^{k+1},\mathbf{A}_1^k,\mathbf{A}_2^k,\mathbf{A}_3^k,\mathbf{U}_1^k,\mathbf{U}_2^k,\mathbf{U}_3^k), \mathbf{A}_1 - \mathbf{A}_1^k \right\rangle \\ \qquad\qquad + \delta_{\mathfrak{B}_1}(\mathbf{A}_1) + \dfrac{\rho_2}{2}\|\mathbf{A}_1 - \mathbf{A}_1^k\|_F^2, \\ \mathbf{A}_2^{k+1} \in \arg\min\limits_{\mathbf{A}_2} \left\langle \nabla_{\mathbf{A}_2} \Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^k,\mathbf{A}_3^k,\mathbf{U}_1^k,\mathbf{U}_2^k,\mathbf{U}_3^k), \mathbf{A}_2 - \mathbf{A}_2^k \right\rangle \\ \qquad\qquad + \delta_{\mathfrak{B}_2}(\mathbf{A}_2) + \dfrac{\rho_3}{2}\|\mathbf{A}_2 - \mathbf{A}_2^k\|_F^2, \\ \mathbf{A}_3^{k+1} \in \arg\min\limits_{\mathbf{A}_3} \left\langle \nabla_{\mathbf{A}_3} \Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^k,\mathbf{U}_1^k,\mathbf{U}_2^k,\mathbf{U}_3^k), \mathbf{A}_3 - \mathbf{A}_3^k \right\rangle \\ \qquad\qquad + \delta_{\mathfrak{B}_3}(\mathbf{A}_3) + \dfrac{\rho_4}{2}\|\mathbf{A}_3 - \mathbf{A}_3^k\|_F^2, \\ \mathbf{U}_1^{k+1} \in \arg\min\limits_{\mathbf{U}_1} \left\langle \nabla_{\mathbf{U}_1} \Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^k,\mathbf{U}_2^k,\mathbf{U}_3^k), \mathbf{U}_1 - \mathbf{U}_1^k \right\rangle \\ \qquad\qquad + \alpha_1 \operatorname{tr}(\mathbf{U}_1^H \mathbf{L}_1 \mathbf{U}_1) + \dfrac{\rho_5}{2}\|\mathbf{U}_1 - \mathbf{U}_1^k\|_F^2, \\ \mathbf{U}_2^{k+1} \in \arg\min\limits_{\mathbf{U}_2} \left\langle \nabla_{\mathbf{U}_2} \Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^k,\mathbf{U}_3^k), \mathbf{U}_2 - \mathbf{U}_2^k \right\rangle \\ \qquad\qquad \alpha_2 \operatorname{tr}(\mathbf{U}_2^H \mathbf{L}_2 \mathbf{U}_2) + \dfrac{\rho_6}{2}\|\mathbf{U}_2 - \mathbf{U}_2^k\|_F^2, \\ \mathbf{U}_3^{k+1} \in \arg\min\limits_{\mathbf{U}_3} \left\langle \nabla_{\mathbf{U}_3} \Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^k), \mathbf{U}_3 - \mathbf{U}_3^k \right\rangle \\ \qquad\qquad + \alpha_3 \operatorname{tr}(\mathbf{U}_3^H \mathbf{L}_3 \mathbf{U}_3) + \dfrac{\rho_7}{2}\|\mathbf{U}_3 - \mathbf{U}_3^k\|_F^2, \end{cases} \tag{19}$$

其中 $\rho_i > 0$ 为给定常数， $i = 1$， ....., 7。现在，我们给出每个子问题的详细解法。式（ $\mathcal{G}$ （19）中的

$$\begin{aligned}\mathcal{G}^{k+1} &= \underset{\|\mathcal{G}\|_\infty \le c}{\arg\min}\, \beta\|\mathcal{G}\|_1 + \left\langle \nabla_{\mathcal{G}}\Psi(\mathcal{G}^k, \mathbf{A}_1^k, \mathbf{A}_2^k, \mathbf{A}_3^k, \mathbf{U}_1^k, \mathbf{U}_2^k, \mathbf{U}_3^k), \mathcal{G} - \mathcal{G}^k \right\rangle + \frac{\rho_1}{2}\|\mathcal{G} - \mathcal{G}^k\|_F^2 \\ &= \underset{\|\mathcal{G}\|_\infty \le c}{\arg\min}\, \beta\|\mathcal{G}\|_1 + \frac{\rho_1}{2}\left\| \mathcal{G} - \mathcal{G}^k + \frac{1}{\rho_1}\nabla_{\mathcal{G}}\Psi(\mathcal{G}^k, \mathbf{A}_1^k, \mathbf{A}_2^k, \mathbf{A}_3^k, \mathbf{U}_1^k, \mathbf{U}_2^k, \mathbf{U}_3^k)\right\|_F^2 \\ &= \underset{\|\mathcal{G}\|_\infty \le c}{\arg\min}\, \beta\|\mathcal{G}\|_1 + \frac{\rho_1}{2}\|\mathcal{G} - \mathcal{L}\|_F^2,\end{aligned}$$

其中 $\mathcal{L} = \mathcal{G}^k - \frac{1}{\rho_1}\nabla_{\mathcal{G}}\Psi(\mathcal{G}^k$、 $\mathbf{A}_1^k$、 $\mathbf{A}_2^k$、 $\mathbf{A}_3^k$、 $\mathbf{U}_1^k$、 $\mathbf{U}_2^k$、 $\mathbf{U}_3^k)$。简单计算可知，关于 $\mathcal{G}$的闭式解可表示为 [6, 例 6。34]

$$\mathcal{G}_{ijl}^{k+1} = \begin{cases} \operatorname{sign}(\mathcal{L}_{ijl}) \max\{|\mathcal{L}_{ijl}| - \beta/\rho_1, 0\}, & \text{if } |\mathcal{L}_{ijl}| \le c + \beta/\rho_1, \\ \operatorname{sign}(\mathcal{L}_{ijl})\, c, & \text{if } |\mathcal{L}_{ijl}| > c + \beta/\rho_1, \end{cases} \tag{20}$$

其中 sign(·) 表示符号函数。

关于 A_1 的子问题可等价地表示为

$$\begin{aligned}\mathbf{A}_1^{k+1} &= \underset{\mathbf{A}_1}{\arg\min}\, \delta_{\mathfrak{B}_1}(\mathbf{A}_1) + \left\langle \nabla_{\mathbf{A}_1}\Psi(\mathcal{G}^{k+1}, \mathbf{A}_1^k, \mathbf{A}_2^k, \mathbf{A}_3^k, \mathbf{U}_1^k, \mathbf{U}_2^k, \mathbf{U}_3^k), \mathbf{A}_1 - \mathbf{A}_1^k \right\rangle \\ &\quad + \frac{\rho_2}{2}\|\mathbf{A}_1 - \mathbf{A}_1^k\|_F^2 \\ &= \underset{\mathbf{A}_1}{\arg\min}\, \delta_{\mathfrak{B}_1}(\mathbf{A}_1) + \frac{\rho_2}{2}\left\| \mathbf{A}_1 - \left(\mathbf{A}_1^k - \frac{1}{\rho_2}\nabla_{\mathbf{A}_1}\Psi(\mathcal{G}^{k+1}, \mathbf{A}_1^k, \mathbf{A}_2^k, \mathbf{A}_3^k, \mathbf{U}_1^k, \mathbf{U}_2^k, \mathbf{U}_3^k)\right)\right\|_F^2.\end{aligned}$$

根据 [55, 定理 4]，上述问题的最优解可轻松计算为

$$\mathbf{A}_1^{k+1} = \bar{\mathbf{U}}_1\bar{\mathbf{V}}_1^H. \tag{21}$$

此处 $\mathrm{A}_1^k - \frac{1}{\rho_2}\nabla_{\mathbf{A}_1}\Psi(\mathcal{G}^{k+1}, \mathrm{A}_1^k, \mathrm{A}_2^k, \mathrm{A}_3^k, \mathrm{U}_1^k, \mathrm{U}_2^k, \mathrm{U}_3^k) = \bar{}\mathrm{U}_1\Sigma_1\bar{}\mathrm{V}_1^H$。

采用类似技术求解关于 A_2 和 A_3的子问题。对两个矩阵 $\mathrm{A}_2^k - \frac{1}{\rho_3}\nabla_{\mathbf{A}_2}\Psi(\mathcal{G}^{k+1}, \mathrm{A}_1^{k+1}, \mathrm{A}_2^k, \mathrm{A}_3^k, \mathrm{U}_1^k, \mathrm{U}_2^k, \mathrm{U}_3^k)$ 和 $\mathrm{A}_3^k - \frac{1}{\rho_4}\nabla_{\mathbf{A}_3}\Psi(\mathcal{G}^{k+1}, \mathrm{A}_1^{k+1}, \mathrm{A}_2^{k+1}, \mathrm{A}_3^k, \mathrm{U}_1^k, \mathrm{U}_2^k, \mathrm{U}_3^k)$ 进行奇异值分解，即:

$$\mathbf{A}_2^k - \frac{1}{\rho_3}\nabla_{\mathbf{A}_2}\Psi(\mathcal{G}^{k+1}, \mathbf{A}_1^{k+1}, \mathbf{A}_2^k, \mathbf{A}_3^k, \mathbf{U}_1^k, \mathbf{U}_2^k, \mathbf{U}_3^k) = \bar{\mathbf{U}}_2\Sigma_2\bar{\mathbf{V}}_2^H$$

and

$$\mathbf{A}_3^k - \frac{1}{\rho_4}\nabla_{\mathbf{A}_3}\Psi(\mathcal{G}^{k+1}, \mathbf{A}_1^{k+1}, \mathbf{A}_2^{k+1}, \mathbf{A}_3^k, \mathbf{U}_1^k, \mathbf{U}_2^k, \mathbf{U}_3^k) = \bar{\mathbf{U}}_3\Sigma_3\bar{\mathbf{V}}_3^H,$$

我们得到，关于$\mathbf{A}_2$ 和$\mathbf{A}_3$ 式（19）的闭式解为

$$\mathbf{A}_2^{k+1} = \bar{\mathbf{U}}_2\bar{\mathbf{V}}_2^H, \quad \mathbf{A}_3^{k+1} = \bar{\mathbf{U}}_3\bar{\mathbf{V}}_3^H. \tag{22}$$

关于$\mathbf{U}_1$ 在（19）中的子问题可等价地改写为

$$\begin{aligned}\mathbf{U}_1^{k+1} &= \underset{\mathbf{U}_1}{\arg\min}\, \alpha_1 \operatorname{tr}(\mathbf{U}_1^H\mathbf{L}_1\mathbf{U}_1) \\ &\quad + \frac{\rho_5}{2}\left\| \mathbf{U}_1 - \mathbf{U}_1^k + \frac{1}{\rho_5}\nabla_{\mathbf{U}_1}\Psi(\mathcal{G}^{k+1}, \mathbf{A}_1^{k+1}, \mathbf{A}_2^{k+1}, \mathbf{A}_3^{k+1}, \mathbf{U}_1^k, \mathbf{U}_2^k, \mathbf{U}_3^k)\right\|_F^2.\end{aligned} \tag{23}$$

根据（23）的最优性条件，关于$\mathbf{U}_1$ 的闭式解可表示为

$$\mathbf{U}_1^{k+1} = (2\alpha_1\mathbf{L}_1 + \rho_5\mathbf{I}_m)^{-1}(\rho_5\mathbf{U}_1^k - \nabla_{\mathbf{U}_1}\Psi(\mathcal{G}^{k+1}, \mathbf{A}_1^{k+1}, \mathbf{A}_2^{k+1}, \mathbf{A}_3^{k+1}, \mathbf{U}_1^k, \mathbf{U}_2^k, \mathbf{U}_3^k)). \tag{24}$$

与计算 U_1^{k+1} 类似，关于 U_2 和 U_3 的最优解可显式表示为

$$\mathbf{U}_2^{k+1} = (2\alpha_2\mathbf{L}_2 + \rho_6\mathbf{I}_m)^{-1}(\rho_6\mathbf{U}_2^k - \nabla_{\mathbf{U}_2}\Psi(\mathcal{G}^{k+1}, \mathbf{A}_1^{k+1}, \mathbf{A}_2^{k+1}, \mathbf{A}_3^{k+1}, \mathbf{U}_1^{k+1}, \mathbf{U}_2^k, \mathbf{U}_3^k)), \tag{25}$$

$$\mathbf{U}_3^{k+1} = (2\alpha_3\mathbf{L}_3 + \rho_7\mathbf{I}_p)^{-1}(\rho_7\mathbf{U}_3^k - \nabla_{\mathbf{U}_3}\Psi(\mathcal{G}^{k+1}, \mathbf{A}_1^{k+1}, \mathbf{A}_2^{k+1}, \mathbf{A}_3^{k+1}, \mathbf{U}_1^{k+1}, \mathbf{U}_2^{k+1}, \mathbf{U}_3^k)). \tag{26}$$

现在，我们总结用于求解问题（12）的 PALM 算法，见算法 1。

算法 1 用于求解问题（12）的 PALM 算法

1：输入：$\{\mathrm{y}_t\}, \mathcal{G}^0, \mathrm{A}_1^0, \mathrm{A}_2^0, \mathrm{A}_3^0, \mathrm{U}_1^0, \mathrm{U}_2^0, \mathrm{U}_3^0$，参数 $\beta, \alpha_1, \alpha_2, \alpha_3, \gamma_1, \gamma_2, \gamma_3, \rho_i, i = 1, \ldots, 7$。 2：重复执行 3：步骤 1。根据（20）计算 $\mathcal{G}^{k+1}$。 4：步骤 2。根据（21）和（22）计算 A_i^{k+1}。 5：步骤 3。根据（24）、（25）和（26）计算 U_i^{k+1}。 6：直到满足停止条件。 7：输出：$\mathcal{G}^{k+1} \times_1 \mathrm{A}_1^{k+1} \times_2 \mathrm{A}_2^{k+1} \times_3 \mathrm{A}_3^{k+1}$。

算法 1 的计算复杂度如下。$\mathcal{G}^{k+1}$ 的主要计算复杂度在于计算 $\nabla_{\mathcal{G}}\Psi$，其量级为 $O(m^2pT)$。A_1^{k+1}的计算成本为 $O(m^2pT + r_1m^2p + r_1r_2r_3mp)$。$\mathrm{A}_2^{k+1}$和 A_3^{k+1}的计算成本分别为$O(m^2pT + r_2m^2p + r_1r_2r_3mp)$和 $O(m^2pT + r_3m^2p + r_1r_2r_3m^2)$。$\mathrm{U}_1^{k+1}$和 U_2^{k+1}的计算复杂度均为 $O(m^3)$，而 U_3^{k+1}的计算复杂度为 $O(p^3)$。因此，算法 1 在每次迭代中的计算复杂度为 $O((r_1 + r_2 + r_3 + T)m^2p + r_1r_2r_3m(m + p) + m^3 + p^3)$。

3.1 收敛性分析

在本小节中，我们证明算法 1 的全局收敛性。首先，我们在以下引理中给出 F 的充分下降性质。

引理 1（充分下降）设序列 $\{(\mathcal{G}^k, \mathrm{A}_1^k, \mathrm{A}_2^k, \mathrm{A}_3^k, \mathrm{U}_1^k, \mathrm{U}_2^k, \mathrm{U}_3^k)\}$ 由算法 1 生成。假设 $\rho_1 > L_1, \rho_2 > L_2, \rho_3 > L_3, \rho_4 > L_4, \rho_5 > \gamma_1, \rho_6 > \gamma_2, \rho_7 > \gamma_3$。那么存在一个常数 $\bar{\rho} > 0$，使得以下不等式成立

$$\begin{aligned} &F\left(\mathcal{G}^{k+1}, \mathbf{A}_1^{k+1}, \mathbf{A}_2^{k+1}, \mathbf{A}_3^{k+1}, \mathbf{U}_1^{k+1}, \mathbf{U}_2^{k+1}, \mathbf{U}_3^{k+1}\right) + \frac{\bar{\rho}}{2}\left(\|\mathcal{G}^{k+1} - \mathcal{G}^k\|_F^2 + \|\mathbf{A}_1^{k+1} - \mathbf{A}_1^k\|_F^2\right. \\ &\left. + \|\mathbf{A}_2^{k+1} - \mathbf{A}_2^k\|_F^2 + \|\mathbf{A}_3^{k+1} - \mathbf{A}_3^k\|_F^2 + \|\mathbf{U}_1^{k+1} - \mathbf{U}_1^k\|_F^2 + \|\mathbf{U}_2^{k+1} - \mathbf{U}_2^k\|_F^2 + \|\mathbf{U}_3^{k+1} - \mathbf{U}_3^k\|_F^2\right) \\ &\leq F\left(\mathcal{G}^k, \mathbf{A}_1^k, \mathbf{A}_2^k, \mathbf{A}_3^k, \mathbf{U}_1^k, \mathbf{U}_2^k, \mathbf{U}_3^k\right), \end{aligned} \tag{27}$$

其中 F定义见（13），且 $\bar{\rho}$：＝最小值$\{\rho_1 - L_1, \rho_2 - L_2, \rho_3 - L_3, \rho_4 - L_4, \rho_5 - \gamma_1, \rho_6 - \gamma_2, \rho_7 - \gamma_3\}$。

接下来，我们证明以下引理中的相对误差条件成立。

引理 2（相对误差条件）设序列 $\{(\mathcal{G}^k, \mathrm{A}_1^k, \mathrm{A}_2^k, \mathrm{A}_3^k, \mathrm{U}_1^k, \mathrm{U}_2^k, \mathrm{U}_3^k)\}$ 由算法 1 生成。则存在 $\mathcal{N}^{k+1} \in \partial F(\mathcal{G}^{k+1}, \mathrm{A}_1^{k+1}, \mathrm{A}_2^{k+1}, \mathrm{A}_3^{k+1}, \mathrm{U}_1^{k+1}, \mathrm{U}_2^{k+1}, \mathrm{U}_3^{k+1})$ 和一个常数 $\delta > 0$，使得

$$\begin{aligned}\|\mathcal{N}^{k+1}\|_F \leq \bar{\delta}\Big(&\|\mathcal{G}^{k+1} - \mathcal{G}^k\|_F + \|\mathbf{A}_1^{k+1} - \mathbf{A}_1^k\|_F + \|\mathbf{A}_2^{k+1} - \mathbf{A}_2^k\|_F + \|\mathbf{A}_3^{k+1} - \mathbf{A}_3^k\|_F \\ &+ \|\mathbf{U}_1^{k+1} - \mathbf{U}_1^k\|_F + \|\mathbf{U}_2^{k+1} - \mathbf{U}_2^k\|_F + \|\mathbf{U}_3^{k+1} - \mathbf{U}_3^k\|_F\Big).\end{aligned} \tag{28}$$

现在，结合引理 1 和 2，我们在以下定理中建立算法 1的全局收敛性。

定理2 设序列 $\{(\mathcal{G}^k, \mathrm{A}_1^k, \mathrm{A}_2^k, \mathrm{A}_3^k, \mathrm{U}_1^k, \mathrm{U}_2^k, \mathrm{U}_3^k)\}$ 由算法 1 生成。假设 $\rho_1 > L_1, \rho_2 > L_2, \rho_3 > L_3, \rho_4 > L_4, \rho_5 > \gamma_1, \rho_6 > \gamma_2, \rho_7 > \gamma_3$。则序列$\{(\mathcal{G}^k, \mathrm{A}_1^k, \mathrm{A}_2^k, \mathrm{A}_3^k, \mathrm{U}_1^k, \mathrm{U}_2^k, \mathrm{U}_3^k)\}$收敛到（12）的一个临界点。

证明。由（5）容易看出，对于 $i = 1$、2、3，有 $\mathrm{tr}(\mathrm{U}_i^H \mathrm{L}_i \mathrm{U}_i) \geq 0$，结合（13）中 $F(\mathcal{G}, \mathrm{A}_1, \mathrm{A}_2, \mathrm{A}_3, \mathrm{U}_1, \mathrm{U}_2, \mathrm{U}_3)$ 的定义，可立即得到 $F \geq 0$。此外，注意到当 $\|\mathcal{G}\|_F$、 $\|\mathrm{A}_i\|_F$或 $\|\mathrm{U}_i\|_F$趋于无穷时，$F(\mathcal{G}, \mathrm{A}_1, \mathrm{A}_2, \mathrm{A}_3, \mathrm{U}_1, \mathrm{U}_2, \mathrm{U}_3)$ 趋于无穷，其中 $i = 1$、2、3。因此，我们得到 F 是强制的。由引理 1 可知， F 单调递减，从而可得序列 $\{(\mathcal{G}^k, \mathrm{A}_1^k, \mathrm{A}_2^k, \mathrm{A}_3^k, \mathrm{U}_1^k, \mathrm{U}_2^k, \mathrm{U}_3^k)\}$有界。因此，存在一个子序列 $\{(\mathcal{G}^{k_j}, \mathrm{A}_1^{k_j}, \mathrm{A}_2^{k_j}, \mathrm{A}_3^{k_j}, \mathrm{U}_1^{k_j}, \mathrm{U}_2^{k_j}, \mathrm{U}_3^{k_j})\}$，使得当 j趋于无穷时，序列 $\{(\mathcal{G}^{k_j}, \mathrm{A}_1^{k_j}, \mathrm{A}_2^{k_j}, \mathrm{A}_3^{k_j}, \mathrm{U}_1^{k_j}, \mathrm{U}_2^{k_j}, \mathrm{U}_3^{k_j})\}$ 收敛到 $\{\mathcal{G}^*, \mathrm{A}_1^*, \mathrm{A}_2^*, \mathrm{A}_3^*, \mathrm{U}_1^*, \mathrm{U}_2^*, \mathrm{U}_3^*\}$。值得注意的是， $\delta_{\mathfrak{B}_i}(\mathrm{A}_i)$ $i = 1$、2、3，以及 $\delta_{\mathcal{D}}(\mathcal{G})$ 中的 F 是不连续的，而 F 中的其他函数均连续。此外，A_i^k和 $\mathcal{G}^k$ 在每次迭代中满足（A_i^k）^{H} $\mathrm{A}_i^k = \mathrm{I}_{r_i}$、 i =1、2、3 和 $\|\mathcal{G}^k\|_\infty \leq c$，从而得到（$\mathrm{A}_i^*$）$^H\mathrm{A}_i^* = \mathrm{I}_{r_i}$、 $i = 1$、2、3、 $\|\mathcal{G}^*\|_\infty \leq c$。因此，当 j趋于无穷时， $\delta_{\mathfrak{B}_i}(\mathrm{A}_i^{k_j})$ 趋于 $\delta_{\mathfrak{B}_i}(\mathrm{A}_i^*)$，且 $\delta_{\mathcal{D}}(\mathcal{G}^{k_j})$ 趋于 $\delta_{\mathcal{D}}(\mathcal{G}^*)$。结果，当 j趋于无穷时，$F(\mathcal{G}^{k_j}, \mathrm{A}_1^{k_j}, \mathrm{A}_2^{k_j}, \mathrm{A}_3^{k_j}, \mathrm{U}_1^{k_j}, \mathrm{U}_2^{k_j}, \mathrm{U}_3^{k_j})$ 趋于 $F(\mathcal{G}^*, \mathrm{A}_1^*, \mathrm{A}_2^*, \mathrm{A}_3^*, \mathrm{U}_1^*, \mathrm{U}_2^*, \mathrm{U}_3^*)$。

注意到，(15) 中定义的 $\Psi(\mathcal{G}, \mathrm{A}_1, \mathrm{A}_2, \mathrm{A}_3, \mathrm{U}_1, \mathrm{U}_2, \mathrm{U}_3)$ 和 $\delta_{\mathcal{D}}(\mathcal{G})$ 是 Kurdyka-Łojasiewicz（KL）函数 [8, 例 2]。对于任意 $i = 1$，2，3，$\delta_{\mathfrak{B}_i}(\mathrm{A}_i)$ 是 KL 函数，因为 $\mathfrak{B}_i$ 是一个半代数集 [34]。此外，函数 $\mathrm{tr}(\mathrm{U}_i^H \mathrm{L}_i \mathrm{U}_i)$ 也是 KL 函数，因为它是一个实解析函数[8, 例 2]。因此可以推得， $F(\mathcal{G}, \mathrm{A}_1, \mathrm{A}_2, \mathrm{A}_3, \mathrm{U}_1, \mathrm{U}_2, \mathrm{U}_3)$ 是一个 KL 函数。结合引理 1、引理 2 和 [2,定理 3.9],，我们可以证明 $\{\mathcal{G}^k, \mathrm{A}_1^k, \mathrm{A}_2^k, \mathrm{A}_3^k, \mathrm{U}_1^k, \mathrm{U}_2^k, \mathrm{U}_3^k\}$收敛到 $F(\mathcal{G}, \mathrm{A}_1, \mathrm{A}_2, \mathrm{A}_3, \mathrm{U}_1, \mathrm{U}_2, \mathrm{U}_3)$ 的一个临界点。证明完毕。□

备注3 定理中的假设 2 非常宽松。我们只需要 $\rho_1 > L_1$， $\rho_2 > L_2$， $\rho_3 > L_3$， $\rho_4 > L_4$， $\rho_5 > \gamma_1$， $\rho_6 > \gamma_2$， $\rho_7 > \gamma_3$。在实验中，我们将详细给出 ρ_i， $i = 1$， ....., 7。

4 数值实验

本节通过一些数值实验，展示所提出的 STDGR 模型在高维时间序列预测中的有效性。我们将 STDGR 与以下五种模型进行比较：用于时间序列模型的 Lasso 型方法（ Lasso ） [5]，转移矩阵的核范数最小化（NNM）[33], 通过 Tucker 秩约束进行低秩张量学习（LRTLT） [3], 基于低秩张量的时变自回归（TVART） [23], 稀疏高阶降秩VAR（SHORR）[47]。所有实验均在 MATLAB R2020b 中进行，所用计算机配备 12 核、 2.6 GHz 的 Intel Core 处理器和 16 GB 内存。

均方误差（MSE）用于评估高维时间序列数据的预测准确度，其定义为

$$\text{MSE} := \frac{1}{mT_0}\sum_{t=1}^{T_0}\|\tilde{\mathbf{y}}_t - \hat{\mathbf{y}}_t\|_2^2,$$

其中， T_0 为测试样本的数量， $\tilde{\mathbf{y}}_t \in \mathbb{R}^m$ 表示真实值， $\hat{\mathbf{y}}_t \in \mathbb{R}^m$表示预测值。

对于任意 $i \in \{1,2,3\}$，记

$$\Lambda_1 = \frac{\|\mathcal{G}^{k+1} - \mathcal{G}^k\|_F}{\|\mathcal{G}^k\|_F},\ \Lambda_{i+1} = \frac{\|\mathbf{A}_i^{k+1} - \mathbf{A}_i^k\|_F}{\|\mathbf{A}_i^k\|_F},\ \Lambda_{i+4} = \frac{\|\mathbf{U}_i^{k+1} - \mathbf{U}_i^k\|_F}{\|\mathbf{U}_i^k\|_F}.$$

算法 1 将在 $\max_{1\le j\le 7}\{\Lambda_j\} \le 3\times 10^{-3}$ 或迭代次数达到 200时终止。

4.1 Tucker秩的选择

在本小节中，我们给出算法 1中 Tucker 秩的选择方法。(10)中的 NNM 用于获得模型的初始估计量，其中参数张量的模-1 展开用于 NNM 模型。(10)中的估计量是一个相合的初始估计量 [33, 47]。我们应用 [47,第 5] 节中的岭型比率估计量来估计算法 1中的 Tucker 秩，其定义如下:

$$\widehat{r}_i = \arg\min_{1\le j\le n_i-1}\frac{\sigma_{j+1}((\widehat{\mathcal{W}}_{\text{NN}})_{(i)}) + \bar{c}}{\sigma_j((\widehat{\mathcal{W}}_{\text{NN}})_{(i)}) + \bar{c}}, \tag{29}$$

其中 $i = 1$、2、3、 $n_1 = n_2 = m$、 $n_3 = p$和 $\bar{c} = o_p(\sqrt{r_1 mp/T})$。[47, 定理 3]表明，(29)中的秩选择是一致的，即 $\mathbb{P}(\widehat{r}_1 = r_1$、 $\widehat{r}_2 = r_2$、 $\widehat{r}_3 = r_3) \to 1$ ，当 $T \to \infty$时。在实验中，我们设定 $\bar{c} = \sqrt{mp}\log(T)/(50T)$。

4.2 拉普拉斯矩阵的构造

模型 (4) 中的拉普拉斯矩阵 $\mathbf{L}_i$ （ $i = 1$， 2， 3）构造如下：首先，使用式 (10) 中的 $\widehat{\text{NN}}$M 模型得到初始估计量 $\mathcal{W}_{\text{NN}}$ ，然后进行高阶奇异值分解（HOSVD） [14] ，得到秩为 $\widehat{r}_i$、 $i = 1$、2、3 的因子矩阵 $\mathbf{A}_i$ ，其中 $\widehat{r}_i$ 通过式 (29) 计算。令 $(\widehat{\mathbf{a}}_i)_l$ 表示 $\widehat{\mathbf{A}}_i$ 的第 l 行。那么，式 (5) 中 $(\widehat{\mathbf{a}}_i)_l$ 与 $(\widehat{\mathbf{a}}_i)_t$ 之间的权重 $(z_i)_{lt}$ 为 $(z_i)_{lt} = e^{-d_{lt}^2/(2\epsilon^2)}$ 和$d_{lt}^2 = \|(\widehat{\mathbf{a}}_i)_l - (\widehat{\mathbf{a}}_i)_t\|_2^2$， 为简化起见，所有实验中均令 $\epsilon = 0.2$ 。回顾一下， $\mathbf{L}_i$ 由 $\mathbf{L}_i = \mathbf{D}_i - \mathbf{Z}_i$ 构造，其中 $\mathbf{Z}_i$ 的第 (l,t) 个元素为 $(z_i)_{lt}$ ，而 $\mathbf{D}_i$ 是一个对角矩阵，其第 l 个对角元素等于 $\mathbf{Z}_i$ 的第 l 行元素之和。

4.3 合成数据

在本小节中，我们首先通过合成数据验证 STDGR 估计量的非渐近误差界。我们简单地设置 $\gamma_1 = \gamma_2 = \gamma_3$ ，并从集合 $\{0.001, 0.1, 10, 100\}$ 中进行选择，以获得最佳性能。根据定理2，对于 $i = 1$、2、3、4，我们简单地设置 $\rho_i = \bar{\alpha}_1 L_i$ ；对于 $j = 1$、2、3，设置 $\rho_{j+4} = \bar{\alpha}_2\gamma_j$，以保证算法 1 的收敛，其中 $\bar{\alpha}_1$ 设置为 1.1， $\bar{\alpha}_2$ 从集合 $\{1.1, 3, 5, 10\}$ 中选取。式（4）中定义的参数 c 从 $\{1,2\}$ 中选取，并将 β 设置为 0.001。为简单起见，我们在所有实验中设置 $\alpha_1 = \alpha_2 = \alpha_3 = \bar{\alpha}$ ，其中 $\bar{\alpha}$ 从集合 $\{0.001, 0.1, 1\}$ 中选取，以获得最佳性能。

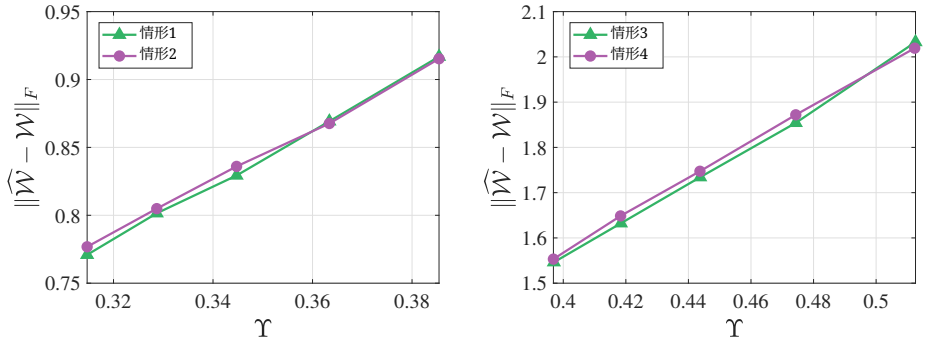


图2：不同合成数据的估计误差 $\|\widehat{\mathcal{W}}-\mathcal{W}\|_F$与 $\Upsilon=(\sqrt{s}+\sqrt{\tilde{s}})\sqrt{\log(m^2p)/T}$的关系。(a) (m,p) =(50, 8)。(b) (m,p) =(80, 12)。

针对合成数据讨论四种情形。具体而言，对于核心张量，我们选取一个超对角元素非零的对角张量，并将 $\mathbf{A}_i$ 生成为高斯随机矩阵的 r_i 个左奇异向量。维度和 Tucker 秩等其他参数设置如下：情形 1。(m,p) = (50, 8)，Tucker 秩 (r_1,r_2,r_3) = (3, 3, 3)，$\mathcal{G}_{111}=\mathcal{G}_{222}=\mathcal{G}_{333}=2$。情形 2。$(m,p)$ = (50, 8)，Tucker 秩 (r_1,r_2,r_3) = (6, 6, 6)，$\mathcal{G}_{111}=\mathcal{G}_{222}=\mathcal{G}_{333}=\mathcal{G}_{444}=\mathcal{G}_{555}=\mathcal{G}_{666}=2$。情形 3。$(m,p)$ = (80, 12)，Tucker 秩 (r_1,r_2,r_3) = (4, 4, 4)，$(\mathcal{G}_{111},\mathcal{G}_{222},\mathcal{G}_{333},\mathcal{G}_{444})$ = (2, 2, 2, 1)。情形 4。(m,p) = (80, 12)，Tucker 秩 (r_1,r_2,r_3) = (7, 7, 7)，且 $(\mathcal{G}_{111},\mathcal{G}_{222},\mathcal{G}_{333},\mathcal{G}_{444},\mathcal{G}_{555},\mathcal{G}_{666},\mathcal{G}_{777})$ = (2, 2, 2, 1, 1, 1, 0.5)。此外，ε_t 独立同分布于 $\sim N(0,\mathrm{I}_m)$。由于随机性，每种情形下的所有实验均重复进行 10 次，并取平均结果作为最终结果。由定理 1 可知，对于固定的 Tucker 秩，误差界 $\|\mathcal{W}-\mathcal{W}\|_F$ 的阶为 $O_p((\sqrt{s}+\sqrt{\tilde{s}})\sqrt{\log(m^2p)/T})$，其中我们设置 $\tilde{s}=s$。令 $\Upsilon=(\sqrt{s}+\sqrt{\tilde{s}})\sqrt{\log(m^2p)/T}$，并选取不同的样本量 T，例如情形 1 和情形 2 取 $\Upsilon=0.314$、0.328、0.344、0.363、0.385，情形 3 和情形 4 取 $\Upsilon=0.397$、0.418、0.443、0.474、0.512。在图2中，我们绘制了四种情形下 $\|\mathcal{W}-\mathcal{W}\|_F$ 相对于 Υ 的曲线。可以看出，估计误差 $\|\mathcal{W}-\mathcal{W}\|_F$ 总体上随 Υ 线性增加，并且每幅图中的两条曲线几乎重合，这验证了定理 1 中的理论结论。

接下来，我们比较 STDGR 与 Lasso、NNM、LRTLT、TVART 和 SHORR 在合成数据上的表现。对于噪声项，我们设定 ε_t 独立同分布于$\sim N(0, 0.25\mathrm{I}_m)$。此外，因子矩阵 $\mathbf{A}_i$ 按如下方式构造 [7]：我们随机生成权重矩阵 $\mathbf{Z}_i$，并根据式（6）得到相应的拉普拉斯矩阵。然后，使用与最小特征值对应的前 r_i 个特征向量来构造 $\mathbf{A}_i$。在谱聚类中，$\mathbf{A}_i$ 的元素衡量数据样本与聚类之间的相似度，因为这些元素表示数据样本属于特定聚类的可能性 [49]。因此，这种 $\mathbf{A}_i$ 的构造能够反映 $\mathbf{A}_i$各行之间的相似性。考虑两种情形，其中 $\mathcal{G}$是一个超对角元素非零的对角张量。情形a：(m,p) =(70, 12)，Tucker秩 (r_1,r_2,r_3) =(5, 5, 5)，且 $(\mathcal{G}_{111},\mathcal{G}_{222},\mathcal{G}_{333},\mathcal{G}_{444},\mathcal{G}_{555})$ =(0.2, 0.2, 0.2, 0.1, 0.1)。情形b：(m,p) =(100, 20)，Tucker秩 (r_1,r_2,r_3) =(7, 7, 7)，且 $(\mathcal{G}_{111},\mathcal{G}_{222},\mathcal{G}_{333},\mathcal{G}_{444},\mathcal{G}_{555},\mathcal{G}_{666},\mathcal{G}_{777})$= (0.1, 0.1, 0.1, 0.05, 0.05, 0.05, 0.05)。

在图3中，我们展示了在上述两种 $\|\mathcal{W}-\mathcal{W}\|_F$设置下，针对合成数据，不同方法的估计误差 T随不同样本量(m,p)的变化情况。可以观察到，随着 T规模增大，不同方法的估计误差均有所下降。此外，STDGR得到的误差低于其他方法，因为STDGR利用了$\mathbf{A}_i$ 的相似性。

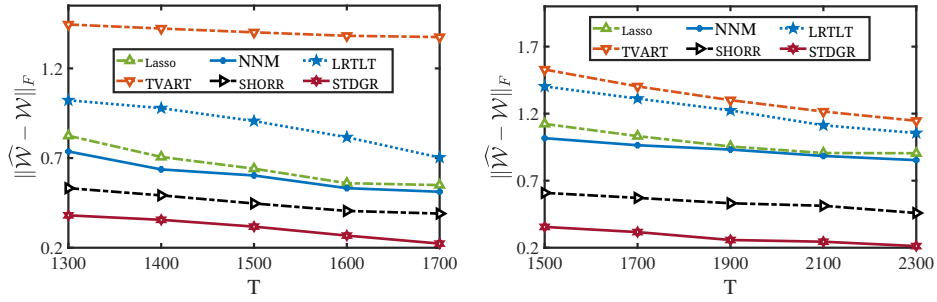


图3：在两种设置（m、p）下，合成数据中不同方法的估计误差 $\|\widehat{\mathcal{W}} - \mathcal{W}\|_F$ 与不同样本量 T 的关系。（a）情形a。（b）情形b。

此外，在估计误差方面，SHORR优于Lasso、NNM、LRTLT和TVART，这表明基于Tucker分解的方法优于基于矩阵或CP的方法。

4.4 综合气候数据集

综合气候数据集（CCD）[1] 包含 1990 年至 2002 年北美洲的 17 个气候变量。数据由五个联邦机构收集并预处理，然后插值到 2.5×2.5度网格上，因此每个变量有 125 个观测地点。每个地点记录了 156 个时间点（每年 12 个月，持续 $\times$ 13 年）。所有序列均通过去除季节平均值进行去季节化，随后标准化为零均值和单位方差。对于 CCD，我们在实验中选取了八个变量，包括二氧化碳（CO2）、甲烷（CH4）、一氧化碳（CO）、氢气（H2）、湿日、最低气温（TMN）、平均气温（TMP）和全球水平辐射（GLO）。

对于向量自回归（p）模型，除 TVART 外，所有方法的阶数均设为 $p = 2$；对于 TVART，则采用 $p = 1$。我们在每种情形下设定 $\gamma_1 = \gamma_2 = \gamma_3$，并从集合 $\{50, 80, 90, 100, 140, 150, 180, 300, 350, 1000, 2000, 3000\}$ 中选取，使预测性能达到最佳。根据定理2中的条件，我们令 $\rho_i = \bar{a}_1 L_i$取 $i = 1$、2、3、4，并令 $\rho_{j+4} = \bar{a}_2\gamma_j$取 $j = 1$、2、3，以保证算法 1 收敛；其中 $\bar{a}_1$ 从集合 $\{1.001, 1.1\}$ 中选取，$\bar{a}_2$从集合 $\{10, 50\}$中选取。式（4）中定义的参数 c 设为 1，β 设为 0.001。为简便起见，所有实验中均设定 $\alpha_1 = \alpha_2 = \alpha_3 = \bar{\alpha}$，其中 $\bar{\alpha}$ 从集合$\{0.001, 0.4, 1, 2, 4, 5, 6, 10, 12, 14, 100\}$ 中选取，以获得最佳预测性能。

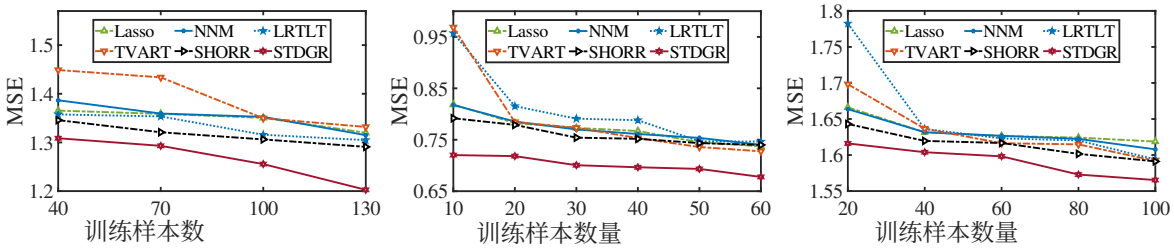
在表 2中，我们报告了不同方法在 CO2、CH4、CO、H2、湿日、TMN、TMP 和 GLO 这些变量上的均方误差，其中每个变量前 60% 的时间序列用于训练，其余 40% 的时间序列用于测试。最佳结果以粗体标出。从表中可以看出，STDGR 的均方误差低于其他方法。此外，在均方误差方面，SHORR 优于 Lasso、NNM、LRTLT 和 TVART。与 SHORR 相比，STDGR 在 CH4、CO 和 H2 上的提升十分显著。

在图 4中，我们展示了 CO2、CH4 和 H2 的不同方法的均方误差与训练样本数量之间的关系，其中 CO2 的测试样本数量为 20，CH4 和 H2 的测试样本数量分别为 10。我们可以观察到，对于这三种变量，随着训练样本数量增加，不同方法的均方误差均有所下降，这表明增加训练样本数量能够提高

[1] https://viterbi-web.usc.edu/~liu32/data/NA-1990-2002-Monthly.csv

表2：CCD 中八个变量采用不同方法的均方误差。

变量	Lasso	NNM	LRTLT	TVART	SHORR	STDGR
CO2	0.9787	0.9775	1.0161	1.1477	0.9774	**0.9620**
CH4	1.1942	1.1977	1.1619	1.2653	1.1655	**1.1107**
CO	0.8832	1.0420	0.8972	0.9267	0.8374	**0.7986**
H2	0.5687	0.5652	0.5575	0.6191	0.5299	**0.4730**
Wet	0.8850	0.8788	1.1835	1.0781	0.8747	**0.8645**
TMN	0.7620	0.7741	0.9386	0.8626	0.7416	**0.7332**
TMP	0.7720	0.7711	0.9829	0.8553	0.7584	**0.7388**
GLO	0.9603	0.9424	1.1006	0.9956	0.9236	**0.9101**



(a) 二氧化碳　(b) 甲烷　(c) 氢气

图4：CCD 的不同方法的均方误差与训练样本数量的关系。

向量自回归的预测准确度。此外，STDGR 得到的均方误差低于其他方法，表明 STDGR 在预测方面性能更优。

4.5 交通流数据集

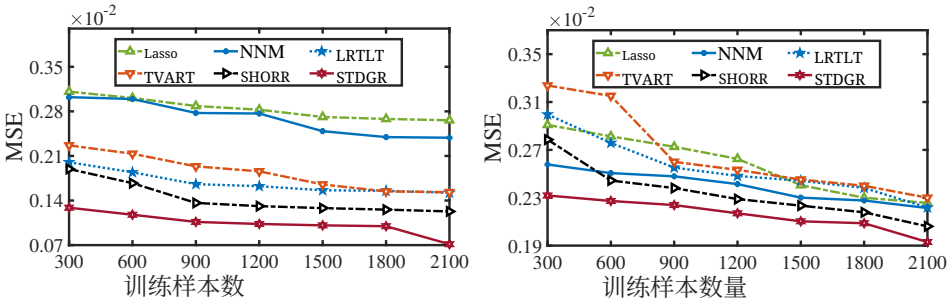
在本小节中，我们对两个交通流数据集进行测试，以验证 STDGR 对向量自回归的有效性。第一个是杭州地铁客流数据集[2]，，该数据集记录了中国杭州 80 个地铁站在 2019 年 1 月 1 日至 1 月 25 日期间的进站客流，时间分辨率为 10 分钟。由于地铁在 00:00 至 06:00 期间不运营，我们剔除了该时段的数据，保留每天其余 108 个时间间隔。我们从该数据集中选取了 2500 个时间间隔。第二个是 PeMS04数据集[3]，，这是由美国旧金山湾区的加利福尼亚州交通运输机构性能测量系统（PeMS）采集的公开交通流数据集。PeMS04数据集包含 307 个传感器在 2018 年 1 月 1 日至 2 月 28 日期间的流量数据，记录间隔为 5 分钟。因此，该数据集共有 16992 个时间间隔，我们在后续实验中选取了 7500 个时间间隔。

对于交通流数据集，除 TVART 外，所有方法的 VAR(p) 模型阶数均设为 $p = 7$；对于 TVART，我们设为 $p = 1$。参数 γ_1、γ_2、γ_3 在每种情形下均设为相同的值，并从集合 $\{0.001, 0.01, 0.1, 1, 100\}$ 中选取，以获得最佳预测性能。我们固定式（4）中定义的 $c = 1$，并设定 $\beta = 0.001$。在所有实验中，我们设定 $\alpha_1 = \alpha_2 = \alpha_3 = \bar{\alpha}$，其中 $\bar{\alpha}$ 从集合 $\{0.001, 0.1\}$ 中选取。根据备注3，为确保算法 1 收敛，我们对 $i = 1$、2、3、4 设定 $\rho_i = \bar{a}_1 L_i$，对 $j = 1$、2、3 设定 $\rho_{j+4} = \bar{a}_2 \gamma_j$。这里，$\bar{a}_1$ 和 $\bar{a}_2$ 从集合中选取

[2]https://tianchi.aliyun.com/competition/entrance/231708/information[3]
https://ieee-dataport.org/documents/pems03-and-pems04

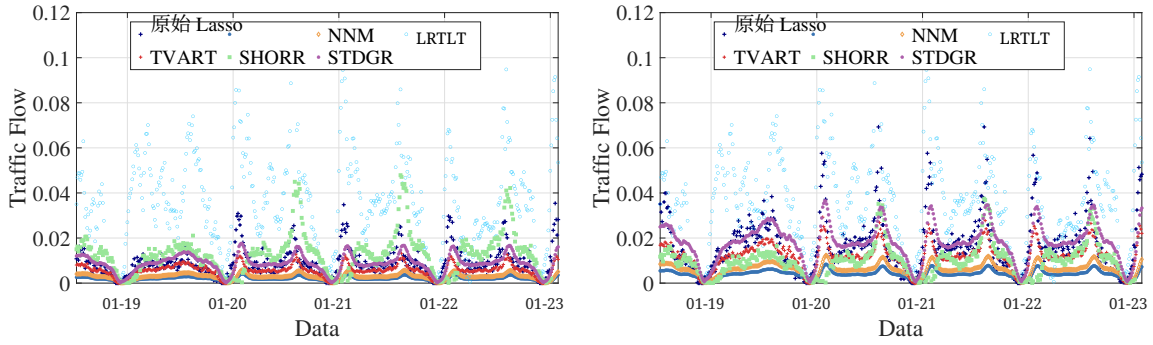
表3：交通流数据集上不同方法的预测结果均方误差（$\times 10^{-3}$）。

数据集	Lasso	NNM	LRTLT	TVART	SHORR	STDGR
杭州	1.9612	1.7411	1.4653	4.2263	1.2394	**0.5961**
PeMS04	3.3012	2.7345	5.4871	5.2348	2.5538	**2.0632**



(a) 杭州　　(b) PeMS04

图 5：均方误差 两个交通流数据集上不同方法的均方误差与训练样本数量的关系。



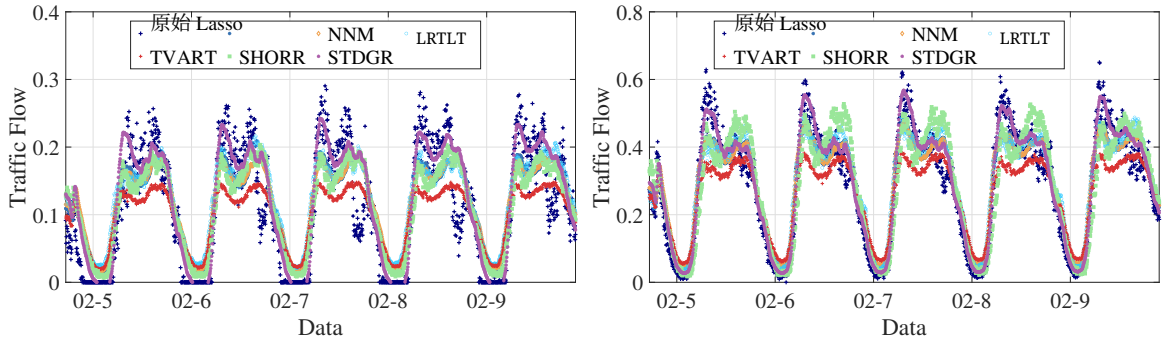
(a) 74号站　　(b) 第79号车站

图6：杭州数据集中第74和第79个地铁站在不同时间点采用不同方法进行交通流量预测的结果。

$\{1.1, 1.3, 10\}$，以获得最佳预测性能。

在表3中，我们报告了不同方法在杭州数据集和PeMS04数据集上的预测误差。每个数据集中，我们使用70%的样本进行训练，其余30%的样本用于测试。最佳结果以粗体突出显示。从表中可以看出，与其他方法相比，STDGR在这两个交通流数据集上取得了最小的预测误差，这体现了STDGR优越的预测性能。此外，在均方误差方面，SHORR优于Lasso、NNM、LRTLT和TVART。

在图5中，我们展示了杭州和 PeMS04 数据集上不同方法的均方误差（$\times 10^{-2}$）与训练样本数量的关系，其中训练样本数量从 300 到 2100，以 300 为步长变化，测试样本数量为 300。可以看出，随着训练样本数量增加，所有方法的均方误差总体上均有所下降。对于所有测试情形，STDGR 始终取得最低的均方误差，并且随着训练样本数量增加，性能提升显著。此外，在大多数情况下，尤其是训练样本数量较大时，SHORR 的表现优于 Lasso、NNM、LRTLT 和



(a) 传感器71　　(b) 传感器74

图7：PeMS04数据集的第71和第74个传感器在不同时间点的不同方法交通流量预测结果。

TVART。结果表明，即使训练样本数量相对较少，STDGR 的预测准确度也高于竞争方法。

在图6 和7中，我们分别展示了杭州和PeMS04数据集在不同时间点的不同方法交通流量预测结果，其中杭州数据集的第74和第79个地铁站如图6 所示，PeMS04数据集的第71和第74个传感器如图7所示。对于这两个数据集，80%的样本用于训练，其余20%的样本用于测试。可以观察到，与其他方法相比，STDGR对不同传感器或地铁站的原始交通流量拟合得更好。此外，STDGR能够更好地预测这两个数据集的交通流量趋势，展现出其卓越的预测性能。

4.6 宏观经济数据集

在本小节中，我们测试STDGR在宏观经济数据集上的性能。该数据集包含美国1959年至2007年的40个季度宏观经济变量，每个变量有194个时间点[4] [30]。除金融序列外，这些变量均经过季节性调整，并转换为平稳序列，使其具有零均值和单位方差 [47]。

我们在 $p = 4$ VAR(p)模型中设定，并且为简便起见设 $\gamma_1 = \gamma_2 = \gamma_3 = 0.001$ 。如备注3所建议，惩罚参数简单地设为 $\rho_i = \bar{a}_1 L_i$，对于 $i = 1$、.....、4，以及 $\rho_{j+4} = \bar{a}_2 \gamma_j$ 对于 $j = 1$、2、3，其中 $\bar{a}_1 = 1.001$ 且 $\bar{a}_2 = 250$。此外，为简便起见， c 设为1 。对于STDGR模型中的参数β、 α_i ，我们设 $\beta = 200$、 $\alpha_1 = \alpha_2 = \alpha_3 = 0.001$。此外，对于该数据集，SHORR和STDGR的Tucker秩选取为相同值。

表 4 报告了 Lasso、NNM、LRTLT、TVART、SHORR 和 STDGR 的均方误差和训练时间（秒）。我们分别使用 85% 和 90% 的样本进行训练，其余样本用于测试。从表中可以看出，在两种测试情形下，STDGR 的均方误差均低于其他方法，表明 STDGR 的预测准确度优于其他对比方法。此外，SHORR 的预测性能位居第二，但其运行时间比其他方法更长。尽管 STDGR 的运行时间长于 Lasso、NNM、LRTLT 和 TVART，但其均方误差低于这些方法。

[4] https://www.stlouisfed.org/research/economists/mccracken/fred-databases

表4：宏观经济数据集中不同方法的均方误差和运行时间（单位：秒）比较。

指标	训练规模	Lasso	NNM	LRTLT	TVART	SHORR	STDGR
MSE	85%	5.3213	5.4821	5.7142	6.1278	5.0664	**4.3838**
	90%	4.6321	4.6500	4.2347	4.8468	3.8021	**3.1845**
Time	85%	0.02	3.00	0.05	0.26	537.14	**17.08**
	90%	0.02	2.88	0.07	0.25	188.51	**20.15**

表5：比较　　杭州数据集中 SHORR、TDSC、TDGR、STDGR 的均方误差（$\times 10^{-3}$）。

训练集大小	SHORR	TDSC	TDGR	STDGR
75%	0.4475	0.3941	0.4782	0.3532
85%	0.3714	0.3370	0.3921	0.2743

4.7 消融分析

在本小节中，我们展示 STDGR 在稀疏 Tucker 分解和图正则化方面的有效性。具体而言，我们将其与采用稀疏核心张量的 Tucker 分解模型（简称 TDSC）进行比较。与[47] 相比，该模型对核心张量而非因子矩阵施加稀疏性约束，且不采用图正则化。此外，我们还将其与不施加稀疏核心张量约束的 Tucker 分解和图正则化模型（简称 TDGR）进行比较。我们也将 SHORR 纳入比较，以展示在 Tucker 分解中对核心张量而非因子矩阵施加稀疏性约束的有效性。

在表 5 中，我们报告了 SHORR、TDSC、TDGR 和 STDGR 在杭州数据集上的均方误差（$\times 10^{-3}$）。其中，75% 和 85% 的样本分别用于训练，其余样本用于测试。从表中可以看出，TDSC 的均方误差低于 SHORR，这表明对核心张量施加稀疏性约束优于对因子矩阵施加该约束。此外，在测试情况下，SHORR 和 TDSC 的表现优于 TDGR，这意味着对核心张量或因子矩阵施加稀疏性约束比图正则化更有效。此外，STDGR 的表现显著优于其他三种方法。这表明，对核心张量施加稀疏性约束以及对因子矩阵进行图正则化，均有助于 VAR 建模。

5 结语

本文提出了一种结合图拉普拉斯正则化的稀疏 Tucker 分解模型，用于高维 VAR 时间序列预测。通过将转移矩阵堆叠成三阶张量，采用 Tucker分解对转移张量进行分解，以解决过度参数化问题，同时从三个方向对参数空间进行约束。此外，通过对核心张量施加稀疏性约束，进一步减少参数数量并增强可解释性。同时，采用图正则化刻画因子矩阵的局部一致性，从而表征 VAR 时间序列采样数据之间的相似性。针对所得模型设计了 PALM 算法，并保证其全局收敛性。在模拟数据和真实数据上的数值实验表明，与多个基线模型相比，所提出方法具有更优的预测性能。

未来，我们将基于其他类型的张量分解，探索张量值自回归时间序列，从而更好地挖掘数据的相关性（参见 [10, 46]）。此外，还将利用 Tucker分解的二阶信息以及核心张量上的 ℓ_1 范数，设计一种更高效的算法，以处理高维VAR建模。

附录A. 辅助引理

P命题 1 对于任意矩阵 $\mathrm{A} \in \mathbb{C}^{n_1 \times n_2}$,$\mathrm{B} \in \mathbb{C}^{n_3 \times n_4}$，以下断言成立 :

(1) $(\mathrm{A} \otimes \mathrm{B})^H = \mathrm{A}^H \otimes \mathrm{B}^H$ *[17,* 公式 (1.3.1) 4.2.*4].* (2) 给定矩阵 $\mathrm{D} \in \mathbb{C}^{n_2 \times n_5}$, $\mathrm{E} \in \mathbb{C}^{n_4 \times n_6}$, $(\mathrm{A} \otimes \mathrm{B})(\mathrm{D} \otimes \mathrm{E}) = (\mathrm{AD}) \otimes (\mathrm{BE})$ *[17,* 公式 (1.3.2) 4.2.*4].* (3) $\|\mathrm{A} \otimes \mathrm{B}\| = \|\mathrm{A}\|\|\mathrm{B}\|$ *[19,* 公式 (1.10)*].* (4) $\|\mathrm{A} \otimes \mathrm{B}\|_F = \|\mathrm{A}\|_F\|\mathrm{B}\|_F$ *[19,* 公式 (1.11)*].*

命题 2 对于任意满足 $\mathrm{A}^H\mathrm{A} = \mathrm{I}_{n_2}$ 的矩阵 $\mathrm{A} \in \mathbb{C}^{n_1 \times n_2}$，以及满足 $\mathrm{B}^H\mathrm{B} = \mathrm{I}_{n_4}$ 的矩阵 $\mathrm{B} \in \mathbb{C}^{n_3 \times n_4}$，有

$$\|\mathbf{A} \otimes \mathbf{B}\| \leq 1.$$

证明。对于任意满足$\mathbf{A}$ 的矩阵$\mathbf{A}^H\mathbf{A} = \mathbf{I}_{n_2}$，有 $\|\mathbf{A}\| \leq 1$ [42, 定理 3。9]。结合命题1(4)可得

$$\|\mathbf{A} \otimes \mathbf{B}\| = \|\mathbf{A}\|\|\mathbf{B}\| \leq 1,$$

至此，证明完成。 □

命题 3 对于任意向量 $\mathrm{a} \in \mathbb{R}^{n_1}$ 和矩阵 $\mathrm{X} \in \mathbb{R}^{n_2 \times n_3}$，有

$$(\mathbf{a} \circ \mathbf{X})_{(2)} = \mathbf{X} \otimes \mathbf{a}^H, \ (\mathbf{a} \circ \mathbf{X})_{(3)} = \mathbf{X}^H \otimes \mathbf{a}^H.$$

证明。对于向量 $\mathrm{a} = (a_1, \ldots, a_{n_1})^H \in \mathbb{R}^{n_1}$ 以及矩阵 $\mathrm{X} \in \mathbb{R}^{n_2 \times n_3}$，其 (j,l)-th 元素为x_{jl}（$1 \leq j \leq n_2, 1 \leq l \leq n_3$），a 和 X 的外积记为 $\mathcal{T} = \mathrm{a} \circ \mathrm{X}$，是一个大小为 $n_1 \times n_2 \times n_3$ 的三阶张量，其每个元素由下式给出

$$\mathcal{T}_{ijl} = a_i x_{jl}.$$

对于 $\mathcal{T}$ 的模态2展开，记为 $\mathcal{T}_{(2)} \in \mathbb{R}^{n_2 \times n_1 n_3}$，其相应的逐元素表示为

$$(\mathcal{T}_{(2)})_{j,(l-1)n_3+i} = \mathcal{T}_{ijl} = a_i x_{jl}. \tag{30}$$

令 $\mathrm{B} = \mathrm{X} \otimes \mathrm{a}^H \in \mathbb{R}^{n_2 \times n_1 n_3}$。则 B 的第 $(j, (l-1)n_3 + i)$ 个元素，记为 $\mathrm{B}_{j,\ (l-1)n_3+i}$，可表示为

$$\mathbf{B}_{j,(l-1)n_3+i} = x_{jl} a_i,$$

结合（30）可得 $\mathcal{T}_{(2)} = (\mathbf{a} \circ \mathbf{X})_{(2)} = \mathbf{X} \otimes \mathbf{a}^H$. 类似地，我们可以推得$(\mathbf{a} \circ \mathbf{X})_{(3)} = \mathbf{X}^H \otimes \mathbf{a}^H$。 □

命题 4 对任意向量 $\mathrm{a} \in \mathbb{R}^{n_1}$ 和矩阵 $\mathrm{X} \in \mathbb{R}^{n_2 \times n_3}$，都有

$$\|\mathbf{a} \circ \mathbf{X}\|_F = \|\mathbf{a}\|_2 \|\mathbf{X}\|_F.$$

证明。结合命题 3 与命题 1，可得

$$\|\mathbf{a}\circ\mathbf{X}\|_F=\|(\mathbf{a}\circ\mathbf{X})_{(2)}\|_F=\|\mathbf{X}\otimes\mathbf{a}^H\|_F=\|\mathbf{X}\|_F\|\mathbf{a}^H\|_2=\|\mathbf{X}\|_F\|\mathbf{a}\|_2.$$

证明完成。 □

考虑加权无向图 $G_i=(V_i,E_i,\mathrm{W}_i)$ $i=1,2,3$，其中 V_i 是基于 A_i 的行所构成的有限顶点集合，而 E_i 是有限边集合，W_i 表示加权邻接矩阵。假设图 G_i 是连通的。则关于图正则化项有如下不等式。

引理 3 记 $\Delta_{\mathbf{A}_i}:=\mathbf{A}_i-\mathbf{A}_{i0}$。假设条件 5 成立且图 G_i 是连通的。则矩阵 $\mathbf{A}_i\in\mathbb{R}^{m\times r_i}$，$i=1,2$，$\mathbf{A}_3\in\mathbb{R}^{p\times r_3}$ 满足以下不等式:

$$\mathrm{tr}(\Delta_{\mathbf{A}_i}^H\mathbf{L}_i\Delta_{\mathbf{A}_i})\geq\frac{\lambda_2(\mathbf{L}_i)}{2}\|\widehat{\mathbf{A}}_i-\mathbf{A}_i\|_F^2,$$

其中 $\lambda_2(\mathrm{L}_i)>0$ 表示 L_i 的第二小特征值 $i=1,i=1$，2，3。

证明。设 $\mathrm{L}_1\in\mathbb{R}^{m\times m}$ 的特征值分解为 $\mathrm{L}_1=\mathrm{Q}\Lambda\mathrm{Q}^H$。定义矩阵 $\mathrm{M}:=\mathrm{Q}^H\Delta_{\mathbf{A}_1}\in\mathbb{R}^{m\times r_1}$。则 M 的第 l 列为 $\mathrm{M}_{:,l}=\mathrm{Q}^H(\Delta_{\mathbf{A}_1})_{:,l}$，其中 $(\Delta_{\mathbf{A}_1})_{:,l}$ 表示 l 的第 $\Delta_{\mathbf{A}_1}$ 列。注意，L_1 是一个对称半正定矩阵，其特征值均为实数且非负。令 $\lambda_j(\mathrm{L}_1)$ 表示 L_1 的第 j 小特征值，其中 $\lambda_1(\mathrm{L}_1)\leq\lambda_2(\mathrm{L}_1)\leq\cdots\leq\lambda_m(\mathrm{L}_1)$。很容易证明

$$\begin{aligned}\mathrm{tr}\left(\Delta_{\mathbf{A}_1}^H\mathbf{L}_1\Delta_{\mathbf{A}_1}\right)&=\sum_{l=1}^{r_1}(\Delta_{\mathbf{A}_1})_{:,l}^H\mathbf{L}_1(\Delta_{\mathbf{A}_1})_{:,l}\\&=\sum_{l=1}^{r_1}(\Delta_{\mathbf{A}_1})_{:,l}^H\mathbf{Q}\Lambda\mathbf{Q}^H(\Delta_{\mathbf{A}_1})_{:,l}=\sum_{l=1}^{r_1}\sum_{j=1}^{m}\lambda_j(\mathbf{L}_1)\mathbf{M}_{jl}^2,\end{aligned}\tag{31}$$

其中，$\mathbf{M}_{jl}$ 表示(j,l)处的元素$\mathbf{M}$。根据 [52, 第 3 节。1] 和 [25, 定理 5], 可得 $\lambda_1(\mathbf{L}_1)=0$ 和 $\lambda_2(\mathbf{L}_1)>0$。于是，我们有

$$\begin{aligned}\sum_{j=1}^{m}\lambda_j(\mathbf{L}_1)\mathbf{M}_{jl}^2=\sum_{j=2}^{m}\lambda_j(\mathbf{L}_1)\mathbf{M}_{jl}^2\geq\sum_{j=2}^{m}\lambda_2(\mathbf{L}_1)\mathbf{M}_{jl}^2&=\sum_{j=1}^{m}\lambda_2(\mathbf{L}_1)\mathbf{M}_{jl}^2-\lambda_2(\mathbf{L}_1)\mathbf{M}_{1l}^2\\&=\lambda_2(\mathbf{L}_1)\left\|(\Delta_{\mathbf{A}_1})_{:,l}\right\|_2^2-\lambda_2(\mathbf{L}_1)\mathbf{M}_{1l}^2.\end{aligned}\tag{32}$$

将（32）代入（31）可得

$$\begin{aligned}\mathrm{tr}\left(\Delta_{\mathbf{A}_1}^H\mathbf{L}_1\Delta_{\mathbf{A}_1}\right)&\geq\sum_{l=1}^{r_1}\left(\lambda_2(\mathbf{L}_1)\left\|(\Delta_{\mathbf{A}_1})_{:,l}\right\|_2^2-\lambda_2(\mathbf{L}_1)\mathbf{M}_{1l}^2\right)\\&=\lambda_2(\mathbf{L}_1)\|\Delta_{\mathbf{A}_1}\|_F^2-\lambda_2(\mathbf{L}_1)\left\|(\mathbf{Q}_{:,1})^H\Delta_{\mathbf{A}_1}\right\|_2^2,\end{aligned}\tag{33}$$

其中，最后一个等式由事实 $\mathrm{M}=\mathrm{Q}^H\Delta_{\mathbf{A}_1}$ 得出。注意，Q 的第一列满足 $\mathrm{Q}_{:,1}=(1/\sqrt{m},1/\sqrt{m},\ldots,1/\sqrt{m})^H$ [25, 第 7] 节。因此，我们得到 $\left\|(\mathrm{Q}_{:,1})^H\Delta_{\mathbf{A}_1}\right\|_2^2=\sum_{l=1}^{r_1}\frac{1}{m}\left(\sum_{j=1}^{m}(\Delta_{\mathbf{A}_1})_{jl}\right)^2$。将 $\|\Delta_{\mathbf{A}_1}\|_F^2=\sum_{l=1}^{r_1}\sum_{j=1}^{m}(\Delta_{\mathbf{A}_1})_{jl}^2$ 与 (33) 结合可得

$$\mathrm{tr}\left(\Delta_{\mathbf{A}_1}^H\mathbf{L}_1\Delta_{\mathbf{A}_1}\right)\geq\lambda_2(\mathbf{L}_1)\sum_{l=1}^{r_1}\sum_{j=1}^{m}(\Delta_{\mathbf{A}_1})_{jl}^2-\lambda_2(\mathbf{L}_1)\sum_{l=1}^{r_1}\frac{1}{m}\Big(\sum_{j=1}^{m}(\Delta_{\mathbf{A}_1})_{jl}\Big)^2.$$

根据假设 5，我们可以得出 $\sum_{l=1}^{r_1}\frac{1}{m}(\sum_{j=1}^{m}(\Delta_{\mathbf{A}_1})_{jl})^2 \le \frac{1}{2}\sum_{l=1}^{r_1}\sum_{j=1}^{m}(\Delta_{\mathbf{A}_1})_{jl}^2$。因此，我们得到

$$\begin{aligned}\operatorname{tr}\left(\Delta_{\mathbf{A}_1}^H \mathbf{L}_1 \Delta_{\mathbf{A}_1}\right) &\ge \lambda_2(\mathbf{L}_1)\sum_{l=1}^{r_1}\sum_{j=1}^{m}(\Delta_{\mathbf{A}_1})_{jl}^2 - \frac{\lambda_2(\mathbf{L}_1)}{2}\sum_{l=1}^{r_1}\sum_{j=1}^{m}(\Delta_{\mathbf{A}_1})_{jl}^2 \\ &= \frac{\lambda_2(\mathbf{L}_1)}{2}\|\Delta_{\mathbf{A}_1}\|_F^2 = \frac{\lambda_2(\mathbf{L}_1)}{2}\|\widehat{\mathbf{A}}_1 - \mathbf{A}_1\|_F^2.\end{aligned}$$

根据假设 5，通过类似的论证可得

$$\begin{aligned}\operatorname{tr}(\Delta_{\mathbf{A}_2}^H \mathbf{L}_2 \Delta_{\mathbf{A}_2}) &\ge \frac{\lambda_2(\mathbf{L}_2)}{2}\|\widehat{\mathbf{A}}_2 - \mathbf{A}_2\|_F^2, \\ \operatorname{tr}(\Delta_{\mathbf{A}_3}^H \mathbf{L}_3 \Delta_{\mathbf{A}_3}) &\ge \frac{\lambda_2(\mathbf{L}_3)}{2}\|\widehat{\mathbf{A}}_3 - \mathbf{A}_3\|_F^2,\end{aligned} \tag{34}$$

其中，$\lambda_2(\mathbf{L}_2)$ $\lambda_2(\mathbf{L}_3) > 0$ 分别表示 $\mathbf{L}_2$ 和 $\mathbf{L}_3$，的第二小特征值。证明完毕。□

^引理 4 令 Δ 在（*71*）中定义，$\mathbf{Z}$ 和 $\mathbf{e}$ 在（*62*）中定义。假设 *1* 和 *2*成立，且样本量 $T \ge a_1\log(m^2p)$。则以下不等式以至少 $1 - 6\exp(-a\log(m^2p))$ 的概率成立：

$$\frac{1}{T}\langle \mathbf{Z}\widehat{\Delta}, \mathbf{e}\rangle \le 2\pi\vartheta a_2\sqrt{\log(m^2p)/T}\|\widehat{\Delta}\|_1,$$

其中 $\vartheta := \lambda_{\max}(\mathbf{\Sigma}_{\boldsymbol{\varepsilon}})(1 + \mu_{\max}(\mathbf{W})/\mu_{\min}(\mathbf{W}))$，常数为 $a > 0$、a_1、$a_2 \ge$ 1.

证明。根据 H¨older 不等式，有

$$\frac{1}{T}\langle \mathbf{Z}\widehat{\Delta}, \mathbf{e}\rangle = \frac{1}{T}\langle \mathbf{Z}^H\mathbf{e}, \widehat{\Delta}\rangle \le \frac{1}{T}\|\mathbf{Z}^H\mathbf{e}\|_\infty\|\widehat{\Delta}\|_1. \tag{35}$$

通过简单计算可得 $\|\mathbf{Z}^H\mathbf{e}\|_\infty = \|\mathbf{X}^H\mathbf{E}\|_\infty = \max_{1\le i\le mp, 1\le j\le m}\left|\mathbf{e}_i^H\mathbf{X}^H\mathbf{E}\mathbf{e}_j\right|$，其中 $\mathbf{e}_i$ 表示标准基向量，其第 i 个分量为 1，其余分量为 0。根据 [5, 命题 2。4] 在假设 1 和 2 下，对于任意满足 $\|\mathbf{u}\|_2 \le 1$、$\|\mathbf{v}\|_2 \le 1$ 的向量 $\mathbf{u}$ 和 $\mathbf{v}$，以及任意 $\eta > 0$，存在一个常数 $a_3 > 0$，使得

$$\begin{aligned}&\mathbb{P}\left[\left|\mathbf{u}^H\left(\mathbf{X}^H\mathbf{E}/T\right)\mathbf{v}\right| > 2\pi\left(\lambda_{\max}\left(\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}\right)\left(1 + \frac{\mu_{\max}(\mathbf{W})}{\mu_{\min}(\mathbf{W})}\right)\right)\eta\right] \\ &\le 6\exp(-a_3T\min(\eta, \eta^2)).\end{aligned}$$

D注意 $\vartheta := \lambda_{\max}\left(\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}\right)[1 + \mu_{\max}(\mathbf{W})/\mu_{\min}(\mathbf{W})]$。通过使用并集界，我们得到 t

$$\begin{aligned}&\mathbb{P}\left[\max_{1\le i\le mp, 1\le j\le m}\left|\mathbf{e}_i^H\mathbf{X}^H\mathbf{E}\mathbf{e}_j/T\right| > 2\pi\vartheta\eta\right] \\ &\le \sum_{1\le i\le mp, 1\le j\le m}\mathbb{P}\left[\left|\mathbf{e}_i^H\mathbf{X}^H\mathbf{E}\mathbf{e}_j/T\right| > 2\pi\vartheta\eta\right] \\ &\le 6m^2p\exp(-a_3T\min(\eta, \eta^2)).\end{aligned} \tag{36}$$

假设 $a_3 > 1$，并令 $\eta = \sqrt{(\log(m^2p))/T}$，其中 $T \ge \log(m^2p)$。那么，直接可得 $\min(\eta, \eta^2) = (\log(m^2p))/T$。因此，我们得到

$$\begin{aligned}6m^2p\exp(-a_3T\min(\eta, \eta^2)) &= 6\exp(\log(m^2p))\exp(-a_3\log(m^2p)) \\ &= 6\exp(-(a_3 - 1)\log(m^2p)).\end{aligned} \tag{37}$$

假设 $a_3 \in (0,1]$ 且 $T \geq \frac{2}{a_3}\log(m^2p)$。令 $\eta = \sqrt{\frac{2}{a_3}}\sqrt{\log(m^2p)/T}$，其中 $\eta \leq 1$。那么，我们有

$$\begin{aligned} 6m^2p\exp(-a_3T\min(\eta,\eta^2)) &= 6m^2p\exp(-2\frac{a_3}{2}T\min(\eta,\eta^2)) \\ &= 6m^2p\exp(-2T\frac{a_3}{2}\eta^2) \\ &= 6m^2p\exp(-2\log\left(m^2p\right)) \\ &= 6\exp(-\log\left(m^2p\right)). \end{aligned} \tag{38}$$

通过结合（37）、（38）以及条件 $T \geq a_1\log(m^2p)$，可得以下不等式:

$$\mathbb{P}\left[\max_{1\leq i\leq mp, 1\leq j\leq m}\left|\mathbf{e}_i^H\mathbf{X}^H\mathbf{E}\mathbf{e}_j/T\right| > 2\pi\vartheta a_2\sqrt{\log\left(m^2p\right)/T}\right] \leq 6\exp(-a\log(m^2p)),$$

其中 $a_1, a_2 \geq 1$ 和 $a > 0$。结合（35）可知

$$\frac{1}{T}\langle\mathbf{Z}\widehat{\Delta},\mathbf{e}\rangle \leq \frac{1}{T}\|\mathbf{Z}^H\mathbf{e}\|_\infty\|\widehat{\Delta}\|_1 \leq 2\pi\vartheta a_2\sqrt{\log(m^2p)/T}\|\widehat{\Delta}\|_1$$

以至少 $1-6\exp(-a\log(m^2p))$ 的概率。 □

引理 5 令 Δ 和 X 分别由（71）和（61）定义。假设 1、2 和3成立，且样本量 $T \geq \frac{8\varsigma^2}{a_4}\bar{p}\min(\log(mp), \log(21emp/\bar{p}))$，其中 $\varsigma := \lambda_{\min}(\mathbf{\Sigma_\varepsilon})/\mu_{\max}(\mathbf{W})\,\lambda_{\max}(\mathbf{\Sigma_\varepsilon})/\mu_{\min}(\mathbf{W})$，$a_4 > 0$ 是一个常数，且 $\bar{p}$ 取决于 $\bar{s}$。则

$$T^{-1}\left\|(\mathbf{I}_m\otimes\mathbf{X})\,\widehat{\Delta}\right\|_2^2 \geq \varpi\|\widehat{\Delta}\|_2^2/2$$

以至少 $1-2\exp(-\bar{p}\min(\log\, mp, \log(21emp/\bar{p})))$的概率成立，其中 $\varpi := \lambda_{\min}\left(\mathbf{\Sigma_\varepsilon}\right)/\mu_{\max}(\mathbf{W})$。证明。对于 $\widehat{\Delta}$ 在（71）中，将其划分为 m 个部分，即 $\widehat{\Delta} = (\widehat{\Delta}_1^H, \ldots, \widehat{\Delta}_m^H)^H$，其中 $\widehat{\Delta}_i \in \mathbb{R}^{mp}$。于是得到

$$\left\|(\mathbf{I}_m\otimes\mathbf{X})\,\widehat{\Delta}\right\|_2^2 = \sum_{i=1}^m\left\|\mathbf{X}\widehat{\Delta}_i\right\|_2^2.$$

记 $\widehat{\mathbf{\Gamma}} = \mathrm{X}^H\mathrm{X}/T$ 和 $\mathbf{\Gamma} = \mathbb{E}\widehat{\mathbf{\Gamma}}$。可以看出,

$$\begin{aligned} \widehat{\Delta}^H\left(\mathbf{I}_m\otimes\mathbf{\Gamma}\right)\widehat{\Delta} = \widehat{\Delta}^H\left(\mathbf{I}_m\otimes\mathbb{E}\widehat{\mathbf{\Gamma}}\right)\widehat{\Delta} &= \widehat{\Delta}^H\Big(\mathbf{I}_m\otimes\mathbb{E}\big(\mathbf{X}^H\mathbf{X}/T\big)\Big)\widehat{\Delta} \\ &= \widehat{\Delta}^H\mathbb{E}\Big(\mathbf{I}_m\otimes\big(\mathbf{X}^H\mathbf{X}/T\big)\Big)\widehat{\Delta} \\ &= \mathbb{E}\Big(\widehat{\Delta}^H\Big(\mathbf{I}_m\otimes\big(\mathbf{X}^H\mathbf{X}/T\big)\Big)\widehat{\Delta}\Big) \\ &= T^{-1}\mathbb{E}\Big(\widehat{\Delta}^H\Big(\mathbf{I}_m\otimes\big(\mathbf{X}^H\mathbf{X}\big)\Big)\widehat{\Delta}\Big) \\ &= T^{-1}\mathbb{E}\left(\|(\mathbf{I}_m\otimes\mathbf{X})\widehat{\Delta}\|_2^2\right). \end{aligned} \tag{39}$$

此外,

$$\begin{aligned} T^{-1}\|\left(\mathbf{I}_m\otimes\mathbf{X}\right)\widehat{\Delta}\|_2^2 &= \widehat{\Delta}^H(\mathbf{I}_m\otimes\widehat{\mathbf{\Gamma}})\widehat{\Delta} \\ &= \widehat{\Delta}^H\left(\mathbf{I}_m\otimes\mathbf{\Gamma}\right)\widehat{\Delta} + \widehat{\Delta}^H\left(\mathbf{I}_m\otimes\left(\widehat{\mathbf{\Gamma}}-\mathbf{\Gamma}\right)\right)\widehat{\Delta} \\ &= \widehat{\Delta}^H\left(\mathbf{I}_m\otimes\mathbf{\Gamma}\right)\widehat{\Delta} + \sum_{i=1}^m\widehat{\Delta}_i^H(\widehat{\mathbf{\Gamma}}-\mathbf{\Gamma})\widehat{\Delta}_i \\ &= T^{-1}\mathbb{E}\left(\|(\mathbf{I}_m\otimes\mathbf{X})\widehat{\Delta}\|_2^2\right) + \sum_{i=1}^m\widehat{\Delta}_i^H(\widehat{\mathbf{\Gamma}}-\mathbf{\Gamma})\widehat{\Delta}_i, \end{aligned} \tag{40}$$

其中最后一个等式由式（39）得出。具体而言，根据[5] 中命题 4.2 的证明以及假设 1、2 可知

$$\lambda_{\min}(\mathbf{\Gamma}) \geq \lambda_{\min}\left(\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}\right) / \mu_{\max}(\mathbf{W}). \tag{41}$$

于是，我们有

$$\begin{aligned} T^{-1}\mathbb{E}\left(\|(\mathbf{I}_m \otimes \mathbf{X})\widehat{\Delta}\|_2^2\right) = \widehat{\Delta}^H\left(\mathbf{I}_m \otimes \mathbf{\Gamma}\right)\widehat{\Delta} &\geq \lambda_{\min}(\mathbf{\Gamma})\|\widehat{\Delta}\|_2^2 \\ &\geq \lambda_{\min}\left(\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}\right)/\mu_{\max}(\mathbf{W})\|\widehat{\Delta}\|_2^2 \\ &= \varpi\|\widehat{\Delta}\|_2^2, \end{aligned} \tag{42}$$

其中，第一个不等式由 Rayleigh-Ritz 定理 [27,定理4.2.2] 和 $\lambda_{\min}(\mathbf{\Gamma}) = \lambda_{\min}(\mathbf{I}_m \otimes \mathbf{\Gamma})$ 得出，等式由 $\varpi := {}^{\lambda_{\min}}(\mathbf{\Sigma}_{\boldsymbol{\varepsilon}})\,\mu_{\max}(\mathbf{W})$ 得出。根据引理 8，对~~于任意~~ $i = 1$、2、.....、m，我们有 $\|\Delta_i\|_0 \leq q_i mp$，其中 $q_i \in [0,1]$ 与 $\bar{s}$有关。令 $\bar{p} = \max\{q_1 mp, q_2 mp, \ldots, q_m mp\}$。我们将 $\mathcal{K}(\bar{p}) := \{\boldsymbol{v} \in \mathbb{R}^{mp} : \|\boldsymbol{v}\|_0 \leq \bar{p}, \|\boldsymbol{v}\|_2 \leq 1\}$ 定义为 $\bar{p}$-稀疏向量的集合。注意，

$$\begin{aligned} & T^{-1}\|\left(\mathbf{I}_m \otimes \mathbf{X}\right)\widehat{\Delta}\|_2^2 \\ \geq\ & T^{-1}\mathbb{E}\left(\|(\mathbf{I}_m \otimes \mathbf{X})\widehat{\Delta}\|_2^2\right) - \sup_{\widehat{\Delta}_i \in \mathcal{K}(\bar{p})}\left|T^{-1}\|\left(\mathbf{I}_m \otimes \mathbf{X}\right)\widehat{\Delta}\|_2^2 - T^{-1}\mathbb{E}\left(\|(\mathbf{I}_m \otimes \mathbf{X})\widehat{\Delta}\|_2^2\right)\right| \\ =\ & T^{-1}\mathbb{E}\left(\|(\mathbf{I}_m \otimes \mathbf{X})\widehat{\Delta}\|_2^2\right) - \sup_{\widehat{\Delta}_i \in \mathcal{K}(\bar{p})}\left|\sum_{i=1}^m \widehat{\Delta}_i^H(\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma})\widehat{\Delta}_i\right| \\ \geq\ & \varpi\|\widehat{\Delta}\|_2^2 - \sup_{\widehat{\Delta}_i \in \mathcal{K}(\bar{p})}\left|\sum_{i=1}^m \widehat{\Delta}_i^H(\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma})\widehat{\Delta}_i\right|, \end{aligned}$$

其中 t 等式由 (40) 得出，第二个不等式由 (42) 得出。于是，我们有 e

$$T^{-1}\|\left(\mathbf{I}_m \otimes \mathbf{X}\right)\widehat{\Delta}\|_2^2 - \frac{\varpi}{2}\|\widehat{\Delta}\|_2^2 \geq \frac{\varpi}{2}\|\widehat{\Delta}\|_2^2 - \sup_{\widehat{\Delta}_i \in \mathcal{K}(\bar{p})}\left|\sum_{i=1}^m \widehat{\Delta}_i^H(\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma})\widehat{\Delta}_i\right|,$$

这很容易推出

$$\begin{aligned} & \mathbb{P}\left[T^{-1}\|\left(\mathbf{I}_m \otimes \mathbf{X}\right)\widehat{\Delta}\|_2^2 - \frac{\varpi}{2}\|\widehat{\Delta}\|_2^2 \geq 0\right] \\ \geq\ & \mathbb{P}\left[\frac{\varpi}{2}\|\widehat{\Delta}\|_2^2 - \sup_{\widehat{\Delta}_i \in \mathcal{K}(\bar{p})}\left|\sum_{i=1}^m \widehat{\Delta}_i^H(\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma})\widehat{\Delta}_i\right| \geq 0\right] \\ =\ & \mathbb{P}\left[\frac{\varpi}{2}\sum_{i=1}^m\|\widehat{\Delta}_i\|_2^2 - \sup_{\widehat{\Delta}_i \in \mathcal{K}(\bar{p})}\left|\sum_{i=1}^m \widehat{\Delta}_i^H(\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma})\widehat{\Delta}_i\right| \geq 0\right]. \end{aligned} \tag{43}$$

对于稳定的向量自回归(p)过程，如[5, 第 4], 节所述，谱密度函数 $f_{\boldsymbol{X}}(\theta)$定义为 $f_{\boldsymbol{X}}(\theta) := \frac{1}{2\pi}\left(\mathbf{W}^{-1}\left(e^{-i\theta}\right)\right)\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}\left(\mathbf{W}^{-1}\left(e^{-i\theta}\right)\right)^H$，其中 $\theta \in [-\pi, \pi]$。接着，我们定义 $\mathcal{M}\left(f_{\boldsymbol{X}}\right) := \text{ess sup}\,\lambda_{\max}\left(f_{\boldsymbol{X}}(\theta)\right)$ [5,假设21]。根据 [5,命题 24],，对于任意 $\eta > 0$, $\|\mathbf{u}\|_2 \leq 1$,$^{\theta \in [-\pi,\pi]}$，存在一个常数 $a_4 > 0$ 使得

$$\mathbb{P}\left[\left|\mathbf{u}^H(\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma})\mathbf{u}\right| > 2\pi\mathcal{M}\left(f_{\boldsymbol{X}}\right)\eta\right] \leq 2\exp\left(-a_4 T\min(\eta, \eta^2)\right). \tag{44}$$

令 $\eta := \frac{\lambda_{\min}(\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})/\mu_{\max}(\mathbf{W})}{4\pi\mathcal{M}(f_{\boldsymbol{X}})}$。因此，我们得到

$$\mathbb{P}\left[\left|\mathbf{u}^H(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\mathbf{u}\right| > \frac{\varpi}{2}\right] \leq 2\exp\left(-a_4 T\min(\eta,\eta^2)\right). \tag{45}$$

根据 [5, 命题 2.3]，我们有 $2\pi\mathcal{M}(f_{\boldsymbol{X}}) \geq \lambda_{\min}(\boldsymbol{\Gamma})$。结合 (41) 可得 $2\pi\mathcal{M}(f_{\boldsymbol{X}}) \geq \lambda_{\min}(\boldsymbol{\Gamma}) \geq \lambda_{\min}(\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})/\mu_{\max}(\mathbf{W})$。因此，$\eta \leq 1/2$。可以证明，(43) 与 (44) 一起蕴含

$$\begin{aligned}
&\mathbb{P}\left[T^{-1}\left\|(\mathbf{I}_m\otimes\mathbf{X})\,\widehat{\Delta}\right\|_2^2 \geq \frac{\varpi}{2}\|\widehat{\Delta}\|_2^2\right]\\
\geq\ &\mathbb{P}\left[\sup_{\widehat{\Delta}_i\in\mathcal{K}(\bar{p})}\left|\sum_{i=1}^m \widehat{\Delta}_i^H(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\widehat{\Delta}_i\right| \leq \frac{\varpi}{2}\sum_{i=1}^m\|\widehat{\Delta}_i\|_2^2\right]\\
\geq\ &\mathbb{P}\left[\sup_{\widehat{\Delta}_i\in\mathcal{K}(\bar{p})}\left|\widehat{\Delta}_i^H(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\widehat{\Delta}_i\right| \leq \frac{\varpi}{2}\|\widehat{\Delta}_i\|_2^2\right]\\
=\ &\mathbb{P}\left[\sup_{\widehat{\Delta}_i\in\mathcal{K}(\bar{p})}\left|\frac{\widehat{\Delta}_i^H}{\|\widehat{\Delta}_i\|_2}(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\frac{\widehat{\Delta}_i}{\|\widehat{\Delta}_i\|_2}\right| \leq \frac{\varpi}{2}\right]\\
\geq\ &1-2\exp\left(-a_4 T\frac{\varpi^2}{(4\pi\mathcal{M}(f_{\boldsymbol{X}}))^2}+\bar{p}\min(\log(mp),\log(21emp/\bar{p}))\right)\\
\geq\ &1-2\exp\left(-a_4 T\frac{\varpi^2}{4(\lambda_{\max}(\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})/\mu_{\min}(\mathbf{W}))^2}+\bar{p}\min(\log(mp),\log(21emp/\bar{p}))\right),
\end{aligned} \tag{46}$$

其中，第三个不等式由 [5, 引理 F.2] 以及 $\eta = \frac{\lambda_{\min}(\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})/\mu_{\max}(\mathbf{W})}{4\pi\mathcal{M}(f_{\boldsymbol{X}})} = \frac{\varpi}{4\pi\mathcal{M}(f_{\boldsymbol{X}})}$ 得到，最后一个不等式则由 $2\pi\mathcal{M}(f_{\boldsymbol{X}}) \leq \lambda_{\max}(\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})/\mu_{\min}(\mathbf{W})$ [5, 公式 (4.1)]得到。

假设 $T \geq \frac{8}{a_4\varpi^2}(\lambda_{\max}(\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})/\mu_{\min}(\mathbf{W}))^2\bar{p}\min(\log(mp),\log(21emp/\bar{p}))$，则

$$\begin{aligned}
&1-2\exp\left(-a_4 T\frac{\varpi^2}{4(\lambda_{\max}(\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})/\mu_{\min}(\mathbf{W}))^2}+\bar{p}\min(\log(mp),\log(21emp/\bar{p}))\right)\\
\geq\ &1-2\exp\left(-\bar{p}\min(\log(mp),\log(21emp/\bar{p}))\right).
\end{aligned}$$

结合式（46）可得

$$\mathbb{P}\left[T^{-1}\left\|(\mathbf{I}_m\otimes\mathbf{X})\,\widehat{\Delta}\right\|_2^2 \geq \frac{\varpi}{2}\|\widehat{\Delta}\|_2^2\right] \geq 1-2\exp(-\bar{p}\min(\log mp,\log(21emp/\bar{p}))).$$

证明至此完成。 □

引理 6 设 Δ定义见式（71）。则

$$\|\widehat{\Delta}\|_2^2 \leq 2\|\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H)-\operatorname{vec}(\mathcal{G}_{(1)}^H)\|_2^2 + 6c^2 r_1 r_2 r_3\left(\|\widehat{\mathbf{A}}_3-\mathbf{A}_3\|_F^2+\|\widehat{\mathbf{A}}_2-\mathbf{A}_2\|_F^2+\|\widehat{\mathbf{A}}_1-\mathbf{A}_1\|_F^2\right),$$

其中 $\widehat{\mathcal{G}}$、$\mathcal{G}\in\mathbb{R}^{r_1\times r_2\times r_3}$，且 $\|\mathcal{G}\|_\infty \leq c$、$\|\widehat{\mathcal{G}}\|_\infty \leq c$、$c$ 在（4）中定义，$\mathbf{A}_i$、$\widehat{\mathbf{A}}_i \in \mathbb{R}^{n_i\times r_i}$以及 $\widehat{\mathbf{A}}_i^H\widehat{\mathbf{A}}_i = \mathbf{I}_{r_i}$、$\mathbf{A}_i^H\mathbf{A}_i = \mathbf{I}_{r_i}$、$i = 1、2、3$。

^ ^证明。根据 A、A $\mathcal{G}$、$\mathcal{G}$ 的定义，很容易看出

$$\begin{aligned}
0 \leq \|\widehat{\Delta}\|_2 &= \left\|(\widehat{\mathbf{A}}_1 \otimes \widehat{\mathbf{A}}_3 \otimes \widehat{\mathbf{A}}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - (\mathbf{A}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\mathcal{G}_{(1)}^H)\right\|_2 \\
&= \Big\|(\widehat{\mathbf{A}}_1 \otimes \widehat{\mathbf{A}}_3 \otimes \widehat{\mathbf{A}}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - (\mathbf{A}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) \\
&\quad + (\mathbf{A}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - (\mathbf{A}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\mathcal{G}_{(1)}^H)\Big\|_2 \\
&\leq \left\|(\widehat{\mathbf{A}}_1 \otimes \widehat{\mathbf{A}}_3 \otimes \widehat{\mathbf{A}}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - (\mathbf{A}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H)\right\|_2 \\
&\quad + \left\|(\mathbf{A}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - (\mathbf{A}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\mathcal{G}_{(1)}^H)\right\|_2 .
\end{aligned}$$

于是，我们推得

$$\begin{aligned}
\|\widehat{\Delta}\|_2^2 &\leq \Big(\left\|(\widehat{\mathbf{A}}_1 \otimes \widehat{\mathbf{A}}_3 \otimes \widehat{\mathbf{A}}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - (\mathbf{A}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H)\right\|_2 \\
&\quad + \left\|(\mathbf{A}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - (\mathbf{A}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\mathcal{G}_{(1)}^H)\right\|_2\Big)^2 \\
&\leq 2\left\|(\widehat{\mathbf{A}}_1 \otimes \widehat{\mathbf{A}}_3 \otimes \widehat{\mathbf{A}}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - (\mathbf{A}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H)\right\|_2^2 \\
&\quad + 2\left\|(\mathbf{A}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - (\mathbf{A}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\mathcal{G}_{(1)}^H)\right\|_2^2 ,
\end{aligned} \tag{47}$$

其中，最后一个不等式源于以下事实：对于任意 a，$b \geq 0$，均有 $\frac{a+b}{2} \leq \sqrt{\frac{a^2+b^2}{2}}$。注意，

$$\begin{aligned}
&\left\|(\mathbf{A}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - (\mathbf{A}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\mathcal{G}_{(1)}^H)\right\|_2 \\
&= \left\|(\mathbf{A}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)(\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - \operatorname{vec}(\mathcal{G}_{(1)}^H))\right\|_2 \\
&\leq \|(\mathbf{A}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\| \left\|\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - \operatorname{vec}(\mathcal{G}_{(1)}^H)\right\|_2 \\
&= \|\mathbf{A}_3\| \, \|\mathbf{A}_2\| \, \|\mathbf{A}_1\| \left\|\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - \operatorname{vec}(\mathcal{G}_{(1)}^H)\right\|_2 \\
&\leq \left\|\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - \operatorname{vec}(\mathcal{G}_{(1)}^H)\right\|_2 ,
\end{aligned} \tag{48}$$

其中，第一个不等式由 [12, 引理 2。1], 第二个等式由命题1得出，第二个不等式由命题2得出。

另一方面 , 注意到

$$\begin{aligned}
&\left\|(\widehat{\mathbf{A}}_1 \otimes \widehat{\mathbf{A}}_3 \otimes \widehat{\mathbf{A}}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - (\mathbf{A}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H)\right\|_2 \\
&= \Big\|(\widehat{\mathbf{A}}_1 \otimes \widehat{\mathbf{A}}_3 \otimes \widehat{\mathbf{A}}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - (\widehat{\mathbf{A}}_1 \otimes \widehat{\mathbf{A}}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) \\
&\quad + (\widehat{\mathbf{A}}_1 \otimes \widehat{\mathbf{A}}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - (\widehat{\mathbf{A}}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) \\
&\quad + (\widehat{\mathbf{A}}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - (\mathbf{A}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H)\Big\|_2 \\
&\leq \left\|(\widehat{\mathbf{A}}_1 \otimes \widehat{\mathbf{A}}_3 \otimes \widehat{\mathbf{A}}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - (\widehat{\mathbf{A}}_1 \otimes \widehat{\mathbf{A}}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H)\right\|_2 \\
&\quad + \left\|(\widehat{\mathbf{A}}_1 \otimes \widehat{\mathbf{A}}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - (\widehat{\mathbf{A}}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H)\right\|_2 \\
&\quad + \left\|(\widehat{\mathbf{A}}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - (\mathbf{A}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H)\right\|_2 .
\end{aligned} \tag{49}$$

对于上述不等式右侧的第一项，可以看出

$$
\begin{aligned}
&\left\|(\widehat{\mathbf{A}}_1 \otimes \widehat{\mathbf{A}}_3 \otimes \widehat{\mathbf{A}}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - (\widehat{\mathbf{A}}_1 \otimes \widehat{\mathbf{A}}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H)\right\|_2 \\
=&\left\|(\widehat{\mathbf{A}}_1 \otimes \widehat{\mathbf{A}}_3 \otimes (\widehat{\mathbf{A}}_2 - \mathbf{A}_2))\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H)\right\|_2 \\
\leq&\left\|\widehat{\mathbf{A}}_1 \otimes \widehat{\mathbf{A}}_3 \otimes (\widehat{\mathbf{A}}_2 - \mathbf{A}_2)\right\| \|\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H)\|_2 \\
=&\|\widehat{\mathbf{A}}_1\|\|\widehat{\mathbf{A}}_3\|\|\widehat{\mathbf{A}}_2 - \mathbf{A}_2\|\|\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H)\|_2 \\
\leq&\|\widehat{\mathbf{A}}_2 - \mathbf{A}_2\|\|\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H)\|_2 \\
=&\|\widehat{\mathbf{A}}_2 - \mathbf{A}_2\|\|\widehat{\mathcal{G}}\|_F \\
\leq&c\sqrt{r_1r_2r_3}\|\widehat{\mathbf{A}}_2 - \mathbf{A}_2\| \\
\leq&c\sqrt{r_1r_2r_3}\|\widehat{\mathbf{A}}_2 - \mathbf{A}_2\|_F,
\end{aligned}
\tag{50}
$$

其中，第一个不等式由 [12, 引理 2。1], 第二个等式由命题 ^11，第二个不等式由命题 2，$\|\mathcal{G}\|_F \leq$ 第三个不等式由 $\sqrt{r_1r_2r_3}\|\mathcal{G}\|_\infty \leq \sqrt{r_1r_2r_3}c$，最后一个不等式由 $\|\mathbf{A}\| \leq \|\mathbf{A}\|_F$ 对任意矩阵$\mathbf{A}$。类似地，

$$
\begin{aligned}
&\left\|(\widehat{\mathbf{A}}_1 \otimes \widehat{\mathbf{A}}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - (\widehat{\mathbf{A}}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H)\right\|_2 \leq c\sqrt{r_1r_2r_3}\|\widehat{\mathbf{A}}_3 - \mathbf{A}_3\|_F, \\
&\left\|(\widehat{\mathbf{A}}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - (\mathbf{A}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H)\right\|_2 \leq c\sqrt{r_1r_2r_3}\|\widehat{\mathbf{A}}_1 - \mathbf{A}_1\|_F.
\end{aligned}
\tag{51}
$$

将 (50)、(51) 代入 (49，得到

$$
\begin{aligned}
0 \leq&\left\|(\widehat{\mathbf{A}}_1 \otimes \widehat{\mathbf{A}}_3 \otimes \widehat{\mathbf{A}}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - (\mathbf{A}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H)\right\|_2 \\
\leq&c\sqrt{r_1r_2r_3}(\|\widehat{\mathbf{A}}_3 - \mathbf{A}_3\|_F + \|\widehat{\mathbf{A}}_2 - \mathbf{A}_2\|_F + \|\widehat{\mathbf{A}}_1 - \mathbf{A}_1\|_F).
\end{aligned}
$$

此外，我们有

$$
\begin{aligned}
&\left\|(\widehat{\mathbf{A}}_1 \otimes \widehat{\mathbf{A}}_3 \otimes \widehat{\mathbf{A}}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - (\mathbf{A}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H)\right\|_2^2 \\
\leq& c^2r_1r_2r_3\left(\|\widehat{\mathbf{A}}_3 - \mathbf{A}_3\|_F + \|\widehat{\mathbf{A}}_2 - \mathbf{A}_2\|_F + \|\widehat{\mathbf{A}}_1 - \mathbf{A}_1\|_F\right)^2 \\
=& c^2r_1r_2r_3\left(\|\widehat{\mathbf{A}}_3 - \mathbf{A}_3\|_F^2 + \|\widehat{\mathbf{A}}_2 - \mathbf{A}_2\|_F^2 + \|\widehat{\mathbf{A}}_1 - \mathbf{A}_1\|_F^2 + 2\|\widehat{\mathbf{A}}_3 - \mathbf{A}_3\|_F\|\widehat{\mathbf{A}}_2 - \mathbf{A}_2\|_F\right. \\
&\left. + 2\|\widehat{\mathbf{A}}_3 - \mathbf{A}_3\|_F\|\widehat{\mathbf{A}}_1 - \mathbf{A}_1\|_F + 2\|\widehat{\mathbf{A}}_1 - \mathbf{A}_1\|_F\|\widehat{\mathbf{A}}_2 - \mathbf{A}_2\|_F\right).
\end{aligned}
\tag{52}
$$

对于任意非负数 a、b、c，已知 $a^2 + b^2 + c^2 \geq ab + ac + bc$。因此，我们可以得出

$$
\begin{aligned}
&\|\widehat{\mathbf{A}}_3 - \mathbf{A}_3\|_F\|\widehat{\mathbf{A}}_2 - \mathbf{A}_2\|_F + \|\widehat{\mathbf{A}}_3 - \mathbf{A}_3\|_F\|\widehat{\mathbf{A}}_1 - \mathbf{A}_1\|_F + \|\widehat{\mathbf{A}}_1 - \mathbf{A}_1\|_F\|\widehat{\mathbf{A}}_2 - \mathbf{A}_2\|_F \\
\leq&\|\widehat{\mathbf{A}}_3 - \mathbf{A}_3\|_F^2 + \|\widehat{\mathbf{A}}_2 - \mathbf{A}_2\|_F^2 + \|\widehat{\mathbf{A}}_1 - \mathbf{A}_1\|_F^2,
\end{aligned}
$$

将其与（52）结合即可得到

$$
\begin{aligned}
&\left\|(\widehat{\mathbf{A}}_3 \otimes \widehat{\mathbf{A}}_2 \otimes \widehat{\mathbf{A}}_1)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - (\mathbf{A}_3 \otimes \mathbf{A}_2 \otimes \mathbf{A}_1)\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H)\right\|_2^2 \\
\leq& c^2r_1r_2r_3\left(3\|\widehat{\mathbf{A}}_3 - \mathbf{A}_3\|_F^2 + 3\|\widehat{\mathbf{A}}_2 - \mathbf{A}_2\|_F^2 + 3\|\widehat{\mathbf{A}}_1 - \mathbf{A}_1\|_F^2\right).
\end{aligned}
\tag{53}
$$

将（53）、（48）代入（47）可得

$$\|\widehat{\Delta}\|_2^2 \leq 2\|\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H) - \operatorname{vec}(\mathcal{G}_{(1)}^H)\|_2^2 + 6c^2 r_1 r_2 r_3 \left(\|\widehat{\mathbf{A}}_3 - \mathbf{A}_3\|_F^2 + \|\widehat{\mathbf{A}}_2 - \mathbf{A}_2\|_F^2 + \|\widehat{\mathbf{A}}_1 - \mathbf{A}_1\|_F^2\right).$$

至此，证明完毕。 □

引理 7 设 $\mathcal{W} = [\![\mathcal{G}; \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3]\!]$ 和 $\mathcal{W} = [\![\mathcal{G}; \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3]\!]$ 分别为 $\mathcal{W} \in \mathbb{R}^{m\times m\times p}$ 和 $\mathcal{W} \in \mathbb{R}^{m\times m\times p}$ 的高阶奇异值分解，且具有相同的多线性秩（r_1、r_2、r_3）。假设 $\mathcal{W}$ 满足假设 4。此外，假设 $\|\mathcal{G}\|_\infty \leq c$ 和 $\|\mathcal{G}\|_\infty \leq c$。则有

$$\|\widetilde{\mathcal{G}} - \mathcal{G}\|_F \leq b\|\widetilde{\mathcal{W}} - \mathcal{W}\|_F,$$

and

$$\|\widetilde{\mathbf{A}}_i - \mathbf{A}_i\|_F \leq 2^{\frac{3}{2}}(\eta_i/\rho)\,\|\widetilde{\mathcal{W}} - \mathcal{W}\|_F,$$

其中 $\eta_i = \left(3\sigma_1(\mathcal{W}_{(i)}) + c\sqrt{r_1 r_2 r_3}\right)\sqrt{\sum_{j=1}^{r_i} \frac{1}{\sigma_j^4(\mathcal{W}_{(i)})}}$，$b = 1 + \frac{2^{\frac{3}{2}}\sigma_1(\mathcal{W}_{(1)})}{\rho}\sum_{i=1}^3 \eta_i$ 和 ρ 定义于假设 4。

证明。 根据[53, 定理 3], 在假设 4下，对于 j-th 奇异向量 $\mathcal{W}_{(i)}$,

$$\begin{aligned}\left\|(\widetilde{\mathbf{A}}_i)_{:,j} - (\mathbf{A}_i)_{:,j}\right\|_2 &\leq \frac{2^{\frac{3}{2}}\left(2\sigma_1(\mathcal{W}_{(i)}) + \|\widetilde{\mathcal{W}}_{(i)} - \mathcal{W}_{(i)}\|\right)\|\widetilde{\mathcal{W}}_{(i)} - \mathcal{W}_{(i)}\|_F}{\min\left[\sigma_{j-1}^2(\mathcal{W}_{(i)}) - \sigma_j^2(\mathcal{W}_{(i)}), \sigma_j^2(\mathcal{W}_{(i)}) - \sigma_{j+1}^2(\mathcal{W}_{(i)})\right]} \\ &\leq \frac{2^{\frac{3}{2}}\left(2\sigma_1(\mathcal{W}_{(i)}) + \|\widetilde{\mathcal{W}}_{(i)} - \mathcal{W}_{(i)}\|\right)\|\widetilde{\mathcal{W}}_{(i)} - \mathcal{W}_{(i)}\|_F}{\rho\sigma_j^2(\mathcal{W}_{(i)})},\end{aligned} \tag{54}$$

其中，$(\mathbf{A}_i)_{:,j}$ 表示奇异值分解中 j的第 -th 左奇异向量，$(\mathbf{A}_i)_{:,j}$ 表示奇异值分解中 $\mathcal{W}_{(i)}$ 的第 j个左奇异向量，且 $\sigma_1(\mathcal{W}_{(i)}) = \|\mathcal{W}_{(i)}\|$。此外，设 $\widetilde{\mathcal{G}}_{(1)}$ 的奇异值分解为 $\widetilde{\mathcal{G}}_{(1)} = \bar{\mathbf{U}}_1\bar{\mathbf{S}}_1\bar{\mathbf{V}}_1^H$，其中 $\bar{\mathbf{U}}_1 \in \mathbb{R}^{r_1\times r_1}$ 和 $\bar{\mathbf{V}}_1 \in \mathbb{R}^{r_2 r_3\times r_2 r_3}$ 为正交矩阵，$\mathbf{S}_1 \in \mathbb{R}^{r_1\times r_2 r_3}$ 为对角矩阵。由命题 1可得

$$(\widetilde{\mathbf{A}}_3 \otimes \widetilde{\mathbf{A}}_2)^H(\widetilde{\mathbf{A}}_3 \otimes \widetilde{\mathbf{A}}_2) = (\widetilde{\mathbf{A}}_3^H \otimes \widetilde{\mathbf{A}}_2^H)(\widetilde{\mathbf{A}}_3 \otimes \widetilde{\mathbf{A}}_2) = (\widetilde{\mathbf{A}}_3^H\widetilde{\mathbf{A}}_3) \otimes (\widetilde{\mathbf{A}}_2^H\widetilde{\mathbf{A}}_2) = \mathbf{I}_{r_3} \otimes \mathbf{I}_{r_2} = \mathbf{I}_{r_2 r_3},$$

这意味着 $\mathbf{A}_3 \otimes \mathbf{A}_2$ 具有正交列。此外，注意到

$$\begin{aligned}\widetilde{\mathcal{W}}_{(1)} = \widetilde{\mathbf{A}}_1\widetilde{\mathcal{G}}_{(1)}(\widetilde{\mathbf{A}}_3 \otimes \widetilde{\mathbf{A}}_2)^H &= \widetilde{\mathbf{A}}_1\bar{\mathbf{U}}_1\bar{\mathbf{S}}_1\bar{\mathbf{V}}_1^H(\widetilde{\mathbf{A}}_3 \otimes \widetilde{\mathbf{A}}_2)^H \\ &= (\widetilde{\mathbf{A}}_1\bar{\mathbf{U}}_1)\bar{\mathbf{S}}_1((\widetilde{\mathbf{A}}_3 \otimes \widetilde{\mathbf{A}}_2)\bar{\mathbf{V}}_1)^H.\end{aligned} \tag{55}$$

这里，$\mathbf{A}_1\mathbf{U}_1$ 和$(\mathbf{A}_3 \otimes \mathbf{A}_2)\mathbf{V}_1$ 具有正交列，且（55）是 $\mathcal{W}_{(1)}$ 的奇异值分解。因此，这意味着 $\|\mathcal{W}_{(1)}\| = \|\mathcal{G}_{(1)}\|$。类似地，我们可以得到 $\|\mathcal{W}_{(2)}\| = \|\mathcal{G}_{(2)}\|$和 $\|\mathcal{W}_{(3)}\| = \|\mathcal{G}_{(3)}\|$。于是，对于任意 $i = 1, 2, 3$，有

$$\|\widetilde{\mathcal{W}}_{(i)} - \mathcal{W}_{(i)}\| \leq \|\widetilde{\mathcal{W}}_{(i)}\| + \|\mathcal{W}_{(i)}\| = \|\widetilde{\mathcal{G}}_{(i)}\| + \sigma_1(\mathcal{W}_{(i)}) \leq c\sqrt{r_1 r_2 r_3} + \sigma_1(\mathcal{W}_{(i)}),$$

其中，最后一个不等式由 $\|\mathcal{G}_{(i)}\| \leq \|\mathcal{G}_{(i)}\|_F = \|\mathcal{G}\|_F \leq c\sqrt{r_1 r_2 r_3}$得出。将上述不等式与式(54)结合，立即可得

$$\left\|(\widetilde{\mathbf{A}}_i)_{:,j} - (\mathbf{A}_i)_{:,j}\right\|_2 \leq \frac{2^{\frac{3}{2}}\left(3\sigma_1(\mathcal{W}_{(i)}) + c\sqrt{r_1 r_2 r_3}\right)\|\widetilde{\mathcal{W}}_{(i)} - \mathcal{W}_{(i)}\|_F}{\rho\sigma_j^2(\mathcal{W}_{(i)})}.$$

因此，

$$\left\|\widetilde{\mathbf{A}}_i-\mathbf{A}_i\right\|_F^2=\sum_{j=1}^{r_i}\left\|(\widetilde{\mathbf{A}}_i)_{:,j}-(\mathbf{A}_i)_{:,j}\right\|_2^2\leq\sum_{j=1}^{r_i}\frac{2^3\left(3\sigma_1\left(\mathcal{W}_{(i)}\right)+c\sqrt{r_1r_2r_3}\right)^2\|\widetilde{\mathcal{W}}_{(i)}-\mathcal{W}_{(i)}\|_F^2}{\rho^2\sigma_j^4\left(\mathcal{W}_{(i)}\right)}.$$

然后，我们可以进一步证明

$$\left\|\widetilde{\mathbf{A}}_i-\mathbf{A}_i\right\|_F\leq 2^{\frac{3}{2}}(\eta_i/\rho)\,\|\widetilde{\mathcal{W}}-\mathcal{W}\|_F,\tag{56}$$

其中 $\eta_i=\left(3\sigma_1\left(\mathcal{W}_{(i)}\right)+c\sqrt{r_1r_2r_3}\right)\sqrt{\sum_{j=1}^{r_i}\frac{1}{\sigma_j^4\left(\mathcal{W}_{(i)}\right)}}$。

另一方面，我们注意到

$$\begin{aligned}
\|\widetilde{\mathcal{G}}-\mathcal{G}\|_F=&\left\|\widetilde{\mathcal{W}}\times_1\widetilde{\mathbf{A}}_1^H\times_2\widetilde{\mathbf{A}}_2^H\times_3\widetilde{\mathbf{A}}_3^H-\mathcal{W}\times_1\mathbf{A}_1^H\times_2\mathbf{A}_2^H\times_3\mathbf{A}_3^H\right\|_F\\
=&\Big\|\widetilde{\mathcal{W}}\times_1\widetilde{\mathbf{A}}_1^H\times_2\widetilde{\mathbf{A}}_2^H\times_3\widetilde{\mathbf{A}}_3^H-\mathcal{W}\times_1\widetilde{\mathbf{A}}_1^H\times_2\widetilde{\mathbf{A}}_2^H\times_3\widetilde{\mathbf{A}}_3^H\\
&+\mathcal{W}\times_1\widetilde{\mathbf{A}}_1^H\times_2\widetilde{\mathbf{A}}_2^H\times_3\widetilde{\mathbf{A}}_3^H-\mathcal{W}\times_1\mathbf{A}_1^H\times_2\widetilde{\mathbf{A}}_2^H\times_3\widetilde{\mathbf{A}}_3^H\\
&+\mathcal{W}\times_1\mathbf{A}_1^H\times_2\widetilde{\mathbf{A}}_2^H\times_3\widetilde{\mathbf{A}}_3^H-\mathcal{W}\times_1\mathbf{A}_1^H\times_2\mathbf{A}_2^H\times_3\widetilde{\mathbf{A}}_3^H\\
&+\mathcal{W}\times_1\mathbf{A}_1^H\times_2\mathbf{A}_2^H\times_3\widetilde{\mathbf{A}}_3^H-\mathcal{W}\times_1\mathbf{A}_1^H\times_2\mathbf{A}_2^H\times_3\mathbf{A}_3^H\Big\|_F\\
\leq&\left\|\widetilde{\mathcal{W}}\times_1\widetilde{\mathbf{A}}_1^H\times_2\widetilde{\mathbf{A}}_2^H\times_3\widetilde{\mathbf{A}}_3^H-\mathcal{W}\times_1\widetilde{\mathbf{A}}_1^H\times_2\widetilde{\mathbf{A}}_2^H\times_3\widetilde{\mathbf{A}}_3^H\right\|_F\\
&+\left\|\mathcal{W}\times_1\widetilde{\mathbf{A}}_1^H\times_2\widetilde{\mathbf{A}}_2^H\times_3\widetilde{\mathbf{A}}_3^H-\mathcal{W}\times_1\mathbf{A}_1^H\times_2\widetilde{\mathbf{A}}_2^H\times_3\widetilde{\mathbf{A}}_3^H\right\|_F\\
&+\left\|\mathcal{W}\times_1\mathbf{A}_1^H\times_2\widetilde{\mathbf{A}}_2^H\times_3\widetilde{\mathbf{A}}_3^H-\mathcal{W}\times_1\mathbf{A}_1^H\times_2\mathbf{A}_2^H\times_3\widetilde{\mathbf{A}}_3^H\right\|_F\\
&+\left\|\mathcal{W}\times_1\mathbf{A}_1^H\times_2\mathbf{A}_2^H\times_3\widetilde{\mathbf{A}}_3^H-\mathcal{W}\times_1\mathbf{A}_1^H\times_2\mathbf{A}_2^H\times_3\mathbf{A}_3^H\right\|_F\\
\leq&\|\widetilde{\mathcal{W}}-\mathcal{W}\|_F+\left\|\mathcal{W}\times_1\widetilde{\mathbf{A}}_1^H\times_2\widetilde{\mathbf{A}}_2^H\times_3\widetilde{\mathbf{A}}_3^H-\mathcal{W}\times_1\mathbf{A}_1^H\times_2\widetilde{\mathbf{A}}_2^H\times_3\widetilde{\mathbf{A}}_3^H\right\|_F\\
&+\left\|\mathcal{W}\times_1\mathbf{A}_1^H\times_2\widetilde{\mathbf{A}}_2^H\times_3\widetilde{\mathbf{A}}_3^H-\mathcal{W}\times_1\mathbf{A}_1^H\times_2\mathbf{A}_2^H\times_3\widetilde{\mathbf{A}}_3^H\right\|_F\\
&+\left\|\mathcal{W}\times_1\mathbf{A}_1^H\times_2\mathbf{A}_2^H\times_3\widetilde{\mathbf{A}}_3^H-\mathcal{W}\times_1\mathbf{A}_1^H\times_2\mathbf{A}_2^H\times_3\mathbf{A}_3^H\right\|_F.
\end{aligned}\tag{57}$$

与（50）类似，我们有

$$\begin{aligned}
&\left\|\mathcal{W}\times_1\widetilde{\mathbf{A}}_1^H\times_2\widetilde{\mathbf{A}}_2^H\times_3\widetilde{\mathbf{A}}_3^H-\mathcal{W}\times_1\mathbf{A}_1^H\times_2\widetilde{\mathbf{A}}_2^H\times_3\widetilde{\mathbf{A}}_3^H\right\|_F\\
=&\left\|\widetilde{\mathbf{A}}_1^H\mathcal{W}_{(1)}(\widetilde{\mathbf{A}}_3\otimes\widetilde{\mathbf{A}}_2)-\mathbf{A}_1^H\mathcal{W}_{(1)}(\widetilde{\mathbf{A}}_3\otimes\widetilde{\mathbf{A}}_2)\right\|_F\\
=&\left\|(\widetilde{\mathbf{A}}_1^H-\mathbf{A}_1^H)\mathcal{W}_{(1)}(\widetilde{\mathbf{A}}_3\otimes\widetilde{\mathbf{A}}_2)\right\|_F\\
\leq&\|\widetilde{\mathbf{A}}_1^H-\mathbf{A}_1^H\|_F\|\mathcal{W}_{(1)}(\widetilde{\mathbf{A}}_3\otimes\widetilde{\mathbf{A}}_2)\|\\
\leq&\|\widetilde{\mathbf{A}}_1^H-\mathbf{A}_1^H\|_F\|\mathcal{W}_{(1)}\|\|\widetilde{\mathbf{A}}_3\otimes\widetilde{\mathbf{A}}_2\|\\
\leq&\|\widetilde{\mathbf{A}}_1^H-\mathbf{A}_1^H\|_F\|\mathcal{W}_{(1)}\|\\
=&\sigma_1(\mathcal{W}_{(1)})\|\widetilde{\mathbf{A}}_1^H-\mathbf{A}_1^H\|_F.
\end{aligned}$$

将上述不等式与（56）结合可得

$$\left\|\mathcal{W}\times_1\widetilde{\mathbf{A}}_1^H\times_2\widetilde{\mathbf{A}}_2^H\times_3\widetilde{\mathbf{A}}_3^H-\mathcal{W}\times_1\mathbf{A}_1^H\times_2\widetilde{\mathbf{A}}_2^H\times_3\widetilde{\mathbf{A}}_3^H\right\|_F\leq\frac{2^{\frac{3}{2}}\eta_1\sigma_1(\mathcal{W}_{(1)})}{\rho}\|\widetilde{\mathcal{W}}-\mathcal{W}\|_F.\tag{58}$$

同理,

$$\begin{aligned}\left\|\mathcal{W}\times_1\mathbf{A}_1^H\times_2\widetilde{\mathbf{A}}_2^H\times_3\widetilde{\mathbf{A}}_3^H-\mathcal{W}\times_1\mathbf{A}_1^H\times_2\mathbf{A}_2^H\times_3\widetilde{\mathbf{A}}_3^H\right\|_F&\leq\frac{2^{\frac{3}{2}}\eta_2\sigma_1(\mathcal{W}_{(1)})}{\rho}\|\widetilde{\mathcal{W}}-\mathcal{W}\|_F,\\ \left\|\mathcal{W}\times_1\mathbf{A}_1^H\times_2\mathbf{A}_2^H\times_3\widetilde{\mathbf{A}}_3^H-\mathcal{W}\times_1\mathbf{A}_1^H\times_2\mathbf{A}_2^H\times_3\mathbf{A}_3^H\right\|_F&\leq\frac{2^{\frac{3}{2}}\eta_3\sigma_1(\mathcal{W}_{(1)})}{\rho}\|\widetilde{\mathcal{W}}-\mathcal{W}\|_F.\end{aligned}\tag{59}$$

将（58）、（59）代入（57）即可得到

$$\begin{aligned}\|\widetilde{\mathcal{G}}-\mathcal{G}\|_F&\leq\|\widetilde{\mathcal{W}}-\mathcal{W}\|_F+\sum_{i=1}^{3}\frac{2^{\frac{3}{2}}\eta_i\sigma_1(\mathcal{W}_{(1)})}{\rho}\|\widetilde{\mathcal{W}}-\mathcal{W}\|_F\\&=b\|\widetilde{\mathcal{W}}-\mathcal{W}\|_F,\end{aligned}$$

其中 $b:=1+\frac{2^{\frac{3}{2}}\sigma_1(\mathcal{W}_{(1)})\sum_{i=1}^{3}\eta_i}{\rho}>0$。 □

引理 8 假设 3 成立，则 $\|\mathbf{Ag}\|_0\leq s_1m^2p$，其中 $\mathbf{A}$ 和 $\mathbf{g}$ 定义见（62），且 $s_1\in[0,1]$ 是与 $\bar{s}$相关的稀疏度阈值。

证明。回顾一下，$\mathbf{g}=\mathrm{vec}(\mathcal{G}_{(1)}^H)\in\mathbb{R}^{r_1r_2r_3}$。向量或矩阵中零元素的比例记为$\bar{z}(\cdot)$ [48]。由 $\mathcal{G}\in\mathbb{R}^{r_1\times r_2\times r_3}$ 属于集合 $\mathcal{U}$ 可知，向量 $\mathbf{g}$ 的稀疏性等于 $\bar{s}$，即 $\bar{z}(\mathbf{g})=\bar{s}$。矩阵 $\mathbf{G}\in\mathbb{R}^{r_1r_2r_3\times m^2p}$ 的构造方式使得，对于每个$j=1,2,\ldots,m^2p$，都有 $\mathbf{G}_{:,j}=\mathbf{g}$，其中 $\mathbf{G}_{:,j}$ 表示 $\mathbf{G}$ 的第 j 列。$\mathbf{G}$ 的每一行的稀疏性记为 $\bar{z}(\mathbf{G}_{i,:})$，其中 $\mathbf{G}_{i,:}$ 表示 $\mathbf{G}$ 的第 i 行，且对于 $i=1,2,\ldots,r_1r_2r_3$，有 $\bar{z}(\mathbf{G}_{i,:})\in\{0,1\}$。类似地，$\mathbf{A}\in\mathbb{R}^{m^2p\times r_1r_2r_3}$ 的第 i 列的稀疏性定义为 $\bar{z}(\mathbf{A}_{:,i})$。根据 [48,命题 1]，，我们可以推得

$$\bar{z}(\mathbf{AG})\geq\max\left\{0,1+\sum_{i=1}^{r_1r_2r_3}\left(\bar{z}(\mathbf{G}_{i,:})+\bar{z}\left(\mathbf{A}_{:,i}\right)-\bar{z}(\mathbf{G}_{i,:})\bar{z}\left(\mathbf{A}_{:,i}\right)\right)-r_1r_2r_3\right\}:=z_1.\tag{60}$$

很容易验证 $\bar{z}(\mathbf{Ag})=\bar{z}(\mathbf{AG})\geq z_1$.因此，我们有 $\|\mathbf{Ag}\|_0\leq(1-z_1)m^2p:=s_1m^2p$。注意，$\sum_{i=1}^{r_1r_2r_3}\bar{z}(\mathbf{G}_{i,:})/(r_1r_2r_3)=\bar{s}$，这意味着 s_1 取决于 $\bar{s}$。□

利用类似技术，由于集合 $\mathcal{U}$ 可知，$\|\widehat{\mathbf{A}}\widehat{\mathbf{g}}\|_0\leq s_2m^2p$，且 $s_2\in[0,1]$ 与 $\bar{s}$相关。

附录 B. 定理 1 的证明

(1) 中的 VAR 模型也可等价地表示为

$$\underbrace{\begin{bmatrix}\mathbf{y}_1^H\\ \mathbf{y}_2^H\\ \vdots\\ \mathbf{y}_T^H\end{bmatrix}}_{\mathbf{Y}}=\underbrace{\begin{bmatrix}\mathbf{y}_0^H & \mathbf{y}_{-1}^H & \cdots & \mathbf{y}_{-p+1}^H\\ \mathbf{y}_1^H & \mathbf{y}_0^H & \cdots & \mathbf{y}_{-p+2}^H\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{y}_{T-1}^H & \mathbf{y}_{T-2}^H & \cdots & \mathbf{y}_{T-p}^H\end{bmatrix}}_{\mathbf{X}}\underbrace{\begin{bmatrix}\mathbf{W}_1^H\\ \mathbf{W}_2^H\\ \vdots\\ \mathbf{W}_p^H\end{bmatrix}}_{\mathcal{W}_{(1)}^H}+\underbrace{\begin{bmatrix}\boldsymbol{\varepsilon}_1^H\\ \boldsymbol{\varepsilon}_2^H\\ \vdots\\ \boldsymbol{\varepsilon}_T^H\end{bmatrix}}_{\mathbf{E}}.\tag{61}$$

注意，$\mathcal{W}=\mathcal{G}\times_1\mathbf{A}_1\times_2\mathbf{A}_2\times_3\mathbf{A}_3$表示 $\mathcal{W}$的高阶奇异值分解，并且 $\mathcal{W}_{(1)}=\mathbf{A}_1\mathcal{G}_{(1)}(\mathbf{A}_3\otimes\mathbf{A}_2)^H$ [29]。根据（61）中的表述，我们可以将（1）重写为

$$\mathbf{Y}=\mathbf{X}(\mathbf{A}_3\otimes\mathbf{A}_2)\mathcal{G}_{(1)}^H\mathbf{A}_1^H+\mathbf{E}.$$

上述等式等价于

$$\begin{aligned}\mathbf{y} := \mathrm{vec}(\mathbf{Y}) &= \mathrm{vec}\left(\mathbf{X}(\mathbf{A}_3 \otimes \mathbf{A}_2)\mathcal{G}_{(1)}^H \mathbf{A}_1^H\right) + \mathrm{vec}(\mathbf{E}) \\ &= \mathrm{vec}\left(\mathbf{X}(\mathbf{A}_3 \otimes \mathbf{A}_2)\mathcal{G}_{(1)}^H \mathbf{A}_1^H \mathbf{I}_m\right) + \mathrm{vec}(\mathbf{E}) \\ &= \left(\mathbf{I}_m^H \otimes \mathbf{X}\right) \mathrm{vec}\left((\mathbf{A}_3 \otimes \mathbf{A}_2)\mathcal{G}_{(1)}^H \mathbf{A}_1^H\right) + \mathrm{vec}(\mathbf{E}) \\ &= \left(\mathbf{I}_m \otimes \mathbf{X}\right)\left(\mathbf{A}_1 \otimes (\mathbf{A}_3 \otimes \mathbf{A}_2)\right) \mathrm{vec}(\mathcal{G}_{(1)}^H) + \mathrm{vec}(\mathbf{E}) \\ &= \underbrace{\left(\mathbf{I}_m \otimes \mathbf{X}\right)}_{\mathbf{Z}} \underbrace{\left(\mathbf{A}_1 \otimes \mathbf{A}_3 \otimes \mathbf{A}_2\right)}_{\mathbf{A}} \underbrace{\mathrm{vec}(\mathcal{G}_{(1)}^H)}_{\mathbf{g}} + \underbrace{\mathrm{vec}(\mathbf{E})}_{\mathbf{e}},\end{aligned} \tag{62}$$

其中第三个和第四个等式由[17,式 12.39]得出。注意$\mathcal{W} = \mathcal{G} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \times_3 \mathbf{A}_3$。记作

$$\widehat{\mathbf{A}} = \widehat{\mathbf{A}}_1 \otimes \widehat{\mathbf{A}}_3 \otimes \widehat{\mathbf{A}}_2,\ \widehat{\mathbf{g}} = \mathrm{vec}(\widehat{\mathcal{G}}_{(1)}^H). \tag{63}$$

然后，可以很容易地验证

$$\begin{aligned}\frac{1}{2T}\sum_{t=1}^{T}\left\|\mathbf{y}_t - (\mathcal{G} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \times_3 \mathbf{A}_3)_{(1)}\,\mathbf{x}_t\right\|_2^2 &= \frac{1}{2T}\|\mathbf{y} - \mathbf{Z}\mathbf{A}\mathbf{g}\|_2^2, \\ \frac{1}{2T}\sum_{t=1}^{T}\left\|\mathbf{y}_t - (\widehat{\mathcal{G}} \times_1 \widehat{\mathbf{A}}_1 \times_2 \widehat{\mathbf{A}}_2 \times_3 \widehat{\mathbf{A}}_3)_{(1)}\mathbf{x}_t\right\|_2^2 &= \frac{1}{2T}\|\mathbf{y} - \mathbf{Z}\widehat{\mathbf{A}}\widehat{\mathbf{g}}\|_2^2.\end{aligned} \tag{64}$$

由于 $\mathcal{W}$ 在（4）中的最优性，我们有

$$\begin{aligned}&\frac{1}{2T}\sum_{t=1}^{T}\left\|\mathbf{y}_t - (\widehat{\mathcal{G}} \times_1 \widehat{\mathbf{A}}_1 \times_2 \widehat{\mathbf{A}}_2 \times_3 \widehat{\mathbf{A}}_3)_{(1)}\mathbf{x}_t\right\|_2^2 + \beta\|\widehat{\mathbf{g}}\|_1 + \sum_{i=1}^{3}\alpha_i \,\mathrm{tr}(\widehat{\mathbf{A}}_i^H \mathbf{L}_i \widehat{\mathbf{A}}_i) \\ \le\ &\frac{1}{2T}\sum_{t=1}^{T}\left\|\mathbf{y}_t - (\mathcal{G} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \times_3 \mathbf{A}_3)_{(1)}\,\mathbf{x}_t\right\|_2^2 + \beta\|\mathbf{g}\|_1 + \sum_{i=1}^{3}\alpha_i \,\mathrm{tr}(\mathbf{A}_i^H \mathbf{L}_i \mathbf{A}_i).\end{aligned}$$

进一步利用（64）可得

$$\begin{aligned}&\frac{1}{2T}\|\mathbf{y} - \mathbf{Z}\widehat{\mathbf{A}}\widehat{\mathbf{g}}\|_2^2 + \beta\|\widehat{\mathbf{g}}\|_1 + \sum_{i=1}^{3}\alpha_i\left(\mathrm{tr}(\widehat{\mathbf{A}}_i^H \mathbf{L}_i \widehat{\mathbf{A}}_i) - \mathrm{tr}(\mathbf{A}_i^H \mathbf{L}_i \mathbf{A}_i)\right) \\ \le\ &\frac{1}{2T}\|\mathbf{y} - \mathbf{Z}\mathbf{A}\mathbf{g}\|_2^2 + \beta\|\mathbf{g}\|_1.\end{aligned} \tag{65}$$

记 $\Delta_{\mathbf{A}_i} = \mathrm{A}_i - \mathrm{A}_i, i = 1, 2, 3$。于是我们得到

$$\begin{aligned}&\mathrm{tr}(\widehat{\mathbf{A}}_1^H \mathbf{L}_1 \widehat{\mathbf{A}}_1) - \mathrm{tr}(\mathbf{A}_1^H \mathbf{L}_1 \mathbf{A}_1) \\ =\ &\mathrm{tr}((\mathbf{A}_1 + \Delta_{\mathbf{A}_1})^H \mathbf{L}_1 (\mathbf{A}_1 + \Delta_{\mathbf{A}_1})) - \mathrm{tr}(\mathbf{A}_1^H \mathbf{L}_1 \mathbf{A}_1) \\ =\ &\mathrm{tr}(\mathbf{A}_1^H \mathbf{L}_1 \mathbf{A}_1 + \mathbf{A}_1^H \mathbf{L}_1 \Delta_{\mathbf{A}_1} + \Delta_{\mathbf{A}_1}^H \mathbf{L}_1 \mathbf{A}_1 + \Delta_{\mathbf{A}_1}^H \mathbf{L}_1 \Delta_{\mathbf{A}_1}) - \mathrm{tr}(\mathbf{A}_1^H \mathbf{L}_1 \mathbf{A}_1) \\ =\ &2\,\mathrm{tr}(\mathbf{A}_1^H \mathbf{L}_1 \Delta_{\mathbf{A}_1}) + \mathrm{tr}(\Delta_{\mathbf{A}_1}^H \mathbf{L}_1 \Delta_{\mathbf{A}_1}),\end{aligned} \tag{66}$$

其中，最后一个等式源于 L_1 具有对称性。注意，

$$\mathrm{tr}(\mathbf{A}_1^H \mathbf{L}_1 \Delta_{\mathbf{A}_1}) = \langle \mathbf{L}_1 \mathbf{A}_1, \Delta_{\mathbf{A}_1} \rangle \ge -\|\mathbf{L}_1 \mathbf{A}_1\|_F \|\Delta_{\mathbf{A}_1}\|_F, \tag{67}$$

其中，不等式根据柯西-施瓦茨不等式成立。结合上述不等式、引理 3和（66）可得

$$\mathrm{tr}(\widehat{\mathbf{A}}_1^H\mathbf{L}_1\widehat{\mathbf{A}}_1)-\mathrm{tr}(\mathbf{A}_1^H\mathbf{L}_1\mathbf{A}_1)\geq\frac{\lambda_2(\mathbf{L}_1)}{2}\|\widehat{\mathbf{A}}_1-\mathbf{A}_1\|_F^2-2\|\mathbf{L}_1\mathbf{A}_1\|_F\|\Delta_{\mathbf{A}_1}\|_F. \tag{68}$$

采用类似技术，我们可以轻松得到

$$\begin{aligned}\mathrm{tr}(\widehat{\mathbf{A}}_2^H\mathbf{L}_2\widehat{\mathbf{A}}_2)-\mathrm{tr}(\mathbf{A}_2^H\mathbf{L}_2\mathbf{A}_2)&\geq\frac{\lambda_2(\mathbf{L}_2)}{2}\|\widehat{\mathbf{A}}_2-\mathbf{A}_2\|_F^2-2\|\mathbf{L}_2\mathbf{A}_2\|_F\|\Delta_{\mathbf{A}_2}\|_F,\\ \mathrm{tr}(\widehat{\mathbf{A}}_3^H\mathbf{L}_3\widehat{\mathbf{A}}_3)-\mathrm{tr}(\mathbf{A}_3^H\mathbf{L}_3\mathbf{A}_3)&\geq\frac{\lambda_2(\mathbf{L}_3)}{2}\|\widehat{\mathbf{A}}_3-\mathbf{A}_3\|_F^2-2\|\mathbf{L}_3\mathbf{A}_3\|_F\|\Delta_{\mathbf{A}_3}\|_F.\end{aligned} \tag{69}$$

将（68）、（69）代入（65）可得

$$\begin{aligned}&\frac{1}{2T}\|\mathbf{y}-\mathbf{Z}\widehat{\mathbf{A}}\widehat{\mathbf{g}}\|_2^2+\beta\|\widehat{\mathbf{g}}\|_1+\sum_{i=1}^{3}\alpha_i\left(\frac{\lambda_2(\mathbf{L}_i)}{2}\|\widehat{\mathbf{A}}_i-\mathbf{A}_i\|_F^2-2\|\mathbf{L}_i\mathbf{A}_i\|_F\|\Delta_{\mathbf{A}_i}\|_F\right)\\ &\leq\frac{1}{2T}\|\mathbf{y}-\mathbf{Z}\mathbf{A}\mathbf{g}\|_2^2+\beta\|\mathbf{g}\|_1.\end{aligned} \tag{70}$$

记

$$\widehat{\Delta}:=\widehat{\mathbf{A}}\widehat{\mathbf{g}}-\mathbf{A}\mathbf{g}, \tag{71}$$

其中 $\mathbf{A}$、$\mathbf{g}$ 和 $\mathbf{A}\widehat{\ }\mathbf{g}$ 分别定义于（62）和（63）。容易得到

$$\begin{aligned}\frac{1}{2T}\|\mathbf{y}-\mathbf{Z}\widehat{\mathbf{A}}\widehat{\mathbf{g}}\|_2^2=\frac{1}{2T}\|\mathbf{y}-\mathbf{Z}(\widehat{\Delta}+\mathbf{A}\mathbf{g})\|_2^2&=\frac{1}{2T}\|-\mathbf{Z}\widehat{\Delta}+\mathbf{e}\|_2^2\\ &=\frac{1}{2T}\|\mathbf{Z}\widehat{\Delta}\|_2^2+\frac{1}{2T}\|\mathbf{e}\|_2^2-\frac{1}{T}\langle\mathbf{Z}\widehat{\Delta},\mathbf{e}\rangle,\end{aligned}$$

其中第二个等号由（62）可得。结合上述结果与（62）和（70）即可得到

$$\begin{aligned}&\frac{1}{2T}\|\mathbf{Z}\widehat{\Delta}\|_2^2+\frac{1}{2T}\|\mathbf{e}\|_2^2-\frac{1}{T}\langle\mathbf{Z}\widehat{\Delta},\mathbf{e}\rangle+\beta\|\widehat{\mathbf{g}}\|_1\\ &+\sum_{i=1}^{3}\alpha_i\left(\frac{\lambda_2(\mathbf{L}_i)}{2}\|\widehat{\mathbf{A}}_i-\mathbf{A}_i\|_F^2-2\|\mathbf{L}_i\mathbf{A}_i\|_F\|\Delta_{\mathbf{A}_i}\|_F\right)\leq\frac{1}{2T}\|\mathbf{e}\|_2^2+\beta\|\mathbf{g}\|_1,\end{aligned}$$

这等价于

$$\begin{aligned}&\frac{1}{2T}\|\mathbf{Z}\widehat{\Delta}\|_2^2+\beta\|\widehat{\mathbf{g}}\|_1+\sum_{i=1}^{3}\frac{\alpha_i\lambda_2(\mathbf{L}_i)}{2}\|\widehat{\mathbf{A}}_i-\mathbf{A}_i\|_F^2\\ &\leq\beta\|\mathbf{g}\|_1+\frac{1}{T}\langle\mathbf{Z}\widehat{\Delta},\mathbf{e}\rangle+\sum_{i=1}^{3}2\alpha_i\|\mathbf{L}_i\mathbf{A}_i\|_F\|\Delta_{\mathbf{A}_i}\|_F.\end{aligned}$$

然后，根据引理 4 和假设 $\beta\geq2\pi\vartheta a_2\sqrt{\log(m^2p)/T}$，以至少 $1-6\exp(-a\log(m^2p))$的概率，以下不等式成立：

$$\begin{aligned}&\frac{1}{2T}\|\mathbf{Z}\widehat{\Delta}\|_2^2+\beta\|\widehat{\mathbf{g}}\|_1+\sum_{i=1}^{3}\frac{\alpha_i\lambda_2(\mathbf{L}_i)}{2}\|\widehat{\mathbf{A}}_i-\mathbf{A}_i\|_F^2\\ &\leq\beta\|\mathbf{g}\|_1+\beta\|\widehat{\Delta}\|_1+\sum_{i=1}^{3}2\alpha_i\|\mathbf{L}_i\mathbf{A}_i\|_F\|\Delta_{\mathbf{A}_i}\|_F.\end{aligned} \tag{72}$$

令 $\Delta_{\mathbf{g}} := \widehat{\mathbf{g}} - \mathbf{g}$。用 $\mathbb{S}_g$ 表示 $\mathrm{vec}(\mathcal{G}^H)$ 的非零索引集合，并令 $\mathbb{S}_{g^c}$ 表示其补集。那么，wehave

$$\begin{aligned}\|\widehat{\mathbf{g}}\|_1 = \left\|\widehat{\mathbf{g}}_{\mathbb{S}_g}\right\|_1 + \left\|\widehat{\mathbf{g}}_{\mathbb{S}_{g^c}}\right\|_1 &= \left\|(\mathbf{g}+\Delta_{\mathbf{g}})_{\mathbb{S}_g}\right\|_1 + \left\|\widehat{\mathbf{g}}_{\mathbb{S}_{g^c}}\right\|_1 \\ &\geq \left\|\mathbf{g}_{\mathbb{S}_g}\right\|_1 - \left\|(\Delta_{\mathbf{g}})_{\mathbb{S}_g}\right\|_1 + \left\|\widehat{\mathbf{g}}_{\mathbb{S}_{g^c}}\right\|_1,\end{aligned}$$

将其与 (72) 合并可得

$$\begin{aligned}&\frac{1}{2T}\|\mathbf{Z}\widehat{\Delta}\|_2^2 + \beta\left\|\mathbf{g}_{\mathbb{S}_g}\right\|_1 - \beta\left\|(\Delta_{\mathbf{g}})_{\mathbb{S}_g}\right\|_1 + \beta\left\|\widehat{\mathbf{g}}_{\mathbb{S}_{g^c}}\right\|_1 + \sum_{i=1}^{3}\frac{\alpha_i\lambda_2(\mathbf{L}_i)}{2}\|\widehat{\mathbf{A}}_i - \mathbf{A}_i\|_F^2 \\ \leq\ &\beta\|\mathbf{g}\|_1 + \beta\|\widehat{\Delta}\|_1 + \sum_{i=1}^{3}2\alpha_i\|\mathbf{L}_i\mathbf{A}_i\|_F\|\Delta_{\mathbf{A}_i}\|_F.\end{aligned}$$

上述不等式可改写为

$$\begin{aligned}&\frac{1}{2T}\|\mathbf{Z}\widehat{\Delta}\|_2^2 + \beta\left\|\widehat{\mathbf{g}}_{\mathbb{S}_{g^c}}\right\|_1 + \sum_{i=1}^{3}\frac{\alpha_i\lambda_2(\mathbf{L}_i)}{2}\|\widehat{\mathbf{A}}_i - \mathbf{A}_i\|_F^2 \\ \leq\ &\beta\|\widehat{\Delta}\|_1 + \sum_{i=1}^{3}2\alpha_i\|\mathbf{L}_i\mathbf{A}_i\|_F\|\Delta_{\mathbf{A}_i}\|_F + \beta\left\|(\Delta_{\mathbf{g}})_{\mathbb{S}_g}\right\|_1.\end{aligned}$$

由 $\beta\left\|\widehat{\mathbf{g}}_{\mathbb{S}_{g^c}}\right\|_1 \geq 0$ 可知

$$\frac{1}{2T}\|\mathbf{Z}\widehat{\Delta}\|_2^2 + \sum_{i=1}^{3}\frac{\alpha_i\lambda_2(\mathbf{L}_i)}{2}\|\widehat{\mathbf{A}}_i - \mathbf{A}_i\|_F^2 \leq \beta\|\widehat{\Delta}\|_1 + \sum_{i=1}^{3}2\alpha_i\|\mathbf{L}_i\mathbf{A}_i\|_F\|\Delta_{\mathbf{A}_i}\|_F + \beta\left\|(\Delta_{\mathbf{g}})_{\mathbb{S}_g}\right\|_1.$$

注意 $T \geq \frac{8c^2}{a_4}\bar{p}\min(\log(mp),\ \log(21emp/\bar{p}))$。引理5 表明以下不等式成立

$$\frac{\varpi}{4}\|\widehat{\Delta}\|_2^2 + \sum_{i=1}^{3}\frac{\alpha_i\lambda_2(\mathbf{L}_i)}{2}\|\widehat{\mathbf{A}}_i - \mathbf{A}_i\|_F^2 \leq \beta\|\widehat{\Delta}\|_1 + \sum_{i=1}^{3}2\alpha_i\|\mathbf{L}_i\mathbf{A}_i\|_F\|\Delta_{\mathbf{A}_i}\|_F + \beta\left\|(\Delta_{\mathbf{g}})_{\mathbb{S}_g}\right\|_1 \quad (73)$$

with 概率至少为 $1 - 2\exp(-\bar{p}\min(\log mp, \log(21emp/\bar{p})))$，其中 $\bar{p}$ 与以下内容有关 $\bar{s}$.

注意，$\alpha_i \geq \lambda_2(\mathbf{L}_i)\sqrt{\frac{6c^2}{T}\log(m^2p)} = \frac{6c^2r_1r_2r_3}{\lambda_2(\mathbf{L}_i)}\sqrt{\frac{\log(m^2p)}{(r_1r_2r_3)^2T}}$。于是，我们有

$$\begin{aligned}&\sum_{i=1}^{3}\alpha_i\frac{\lambda_2(\mathbf{L}_i)}{2}\|\widehat{\mathbf{A}}_i - \mathbf{A}_i\|_F^2 + \sqrt{\frac{s\log(m^2p)}{(r_1r_2r_3)^2T}}\|\mathrm{vec}(\widehat{\mathcal{G}}_{(1)}^H) - \mathrm{vec}(\mathcal{G}_{(1)}^H)\|_2^2 \\ \geq\ &3c^2r_1r_2r_3\sqrt{\frac{s\log(m^2p)}{(r_1r_2r_3)^2T}}\left(\|\widehat{\mathbf{A}}_3 - \mathbf{A}_3\|_F^2 + \|\widehat{\mathbf{A}}_2 - \mathbf{A}_2\|_F^2 + \|\widehat{\mathbf{A}}_1 - \mathbf{A}_1\|_F^2\right) \\ &+ \sqrt{\frac{s\log(m^2p)}{(r_1r_2r_3)^2T}}\|\mathrm{vec}(\widehat{\mathcal{G}}_{(1)}^H) - \mathrm{vec}(\mathcal{G}_{(1)}^H)\|_2^2 \\ \geq\ &\frac{1}{2}\sqrt{\frac{s\log(m^2p)}{(r_1r_2r_3)^2T}}\|\widehat{\Delta}\|_2^2,\end{aligned}$$

其中，最后一个不等式由引理6得到。结合（73），可得

$$\begin{aligned}
&\frac{\varpi}{4}\|\widehat{\Delta}\|_2^2+\frac{1}{2}\sqrt{\frac{s\log(m^2p)}{(r_1r_2r_3)^2T}}\|\widehat{\Delta}\|_2^2\\
&\leq\beta\|\widehat{\Delta}\|_1+\sum_{i=1}^{3}2\alpha_i\|\mathbf{L}_i\mathbf{A}_i\|_F\|\Delta_{\mathbf{A}_i}\|_F+\beta\left\|(\Delta_{\mathbf{g}})_{\mathbb{S}_g}\right\|_1+\sqrt{\frac{s\log(m^2p)}{(r_1r_2r_3)^2T}}\|\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H)-\operatorname{vec}(\mathcal{G}_{(1)}^H)\|_2^2
\end{aligned}\tag{74}$$

其概率至少为 $1-2\exp(-\bar{p}\min(\log\ mp,\log(21emp/\bar{p})))$.

可知 $\widehat{\mathbf{A}}\widehat{\mathbf{g}}-\mathbf{A}\mathbf{g}\in\mathcal{U}$，由引理8可知 $\|\mathbf{A}\mathbf{g}\|_0\leq s_1m^2p$ 和 $\|\widehat{\mathbf{A}}\widehat{\mathbf{g}}\|_0\leq s_2m^2p$，其中 s_1, $s_2\in[0,1]$ 是与 $\bar{s}$相关的稀疏度阈值。于是，我们有

$$\begin{aligned}
\|\widehat{\Delta}\|_0=\|\widehat{\mathbf{A}}\widehat{\mathbf{g}}-\mathbf{A}\mathbf{g}\|_0&\leq\|\widehat{\mathbf{A}}\widehat{\mathbf{g}}\|_0+\|\mathbf{A}\mathbf{g}\|_0\\
&\leq s_2m^2p+s_1m^2p:=\tilde{s}.
\end{aligned}\tag{75}$$

因此，我们得到

$$\|\widehat{\Delta}\|_1\leq\sqrt{\tilde{s}}\|\widehat{\Delta}\|_2.\tag{76}$$

此外，

$$\left\|(\Delta_{\mathbf{g}})_{\mathbb{S}_g}\right\|_1\leq\sqrt{s}\|\Delta_{\mathbf{g}}\|_2\leq b\sqrt{s}\|\widehat{\Delta}\|_2,\tag{77}$$

其中，最后一个不等式由引理7得出。根据 H¨older 不等式，我们有

$$\begin{aligned}
\|\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H)-\operatorname{vec}(\mathcal{G}_{(1)}^H)\|_2^2&\leq\|\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H)-\operatorname{vec}(\mathcal{G}_{(1)}^H)\|_\infty\|\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H)-\operatorname{vec}(\mathcal{G}_{(1)}^H)\|_1\\
&\leq 2c\|\operatorname{vec}(\widehat{\mathcal{G}}_{(1)}^H)-\operatorname{vec}(\mathcal{G}_{(1)}^H)\|_1\\
&=2c\|\Delta_{\mathbf{g}}\|_1\\
&\leq 2c\sqrt{2s}\,\|\Delta_{\mathbf{g}}\|_2\\
&\leq 2bc\sqrt{2s}\|\widehat{\Delta}\|_2,
\end{aligned}\tag{78}$$

其中，第三个不等式是由以下事实得出的：两者 $\mathcal{G}$ 和 $\mathcal{G}$ 都属于集合 $\mathcal{U}$，最后一个不等式则由引理7得出。将（76）、（77）和（78）代入（74），即可得到

$$\begin{aligned}
&\left(\frac{\varpi}{4}+\frac{1}{2}\sqrt{\frac{s\log(m^2p)}{(r_1r_2r_3)^2T}}\right)\|\widehat{\Delta}\|_2^2\\
&\leq\beta\sqrt{\tilde{s}}\|\widehat{\Delta}\|_2+\sum_{i=1}^{3}2\alpha_i\|\mathbf{L}_i\mathbf{A}_i\|_F\|\Delta_{\mathbf{A}_i}\|_F+b\beta\sqrt{s}\|\widehat{\Delta}\|_2+2\sqrt{2}bc\sqrt{\frac{s^2\log(m^2p)}{(r_1r_2r_3)^2T}}\|\widehat{\Delta}\|_2\\
&\leq\beta\sqrt{\tilde{s}}\|\widehat{\Delta}\|_2+\sum_{i=1}^{3}2\alpha_i\|\mathbf{L}_i\mathbf{A}_i\|_F\|\Delta_{\mathbf{A}_i}\|_F+b\beta\sqrt{s}\|\widehat{\Delta}\|_2+2\sqrt{2}bc\sqrt{\frac{\log(m^2p)}{T}}\|\widehat{\Delta}\|_2
\end{aligned}\tag{79}$$

其概率至少为 $1-2\exp(-\bar{p}\min(\log\ mp,\log(21emp/\bar{p}))$，其中最后一个不等式由 $s\leq r_1r_2r_3$ 得出。注意，

$$2\alpha_i\|\mathbf{L}_i\mathbf{A}_i\|_F\|\Delta_{\mathbf{A}_i}\|_F\leq 2\alpha_i\|\mathbf{L}_i\|_F\|\mathbf{A}_i\|\|\Delta_{\mathbf{A}_i}\|_F\leq 2\alpha_il_i\|\Delta_{\mathbf{A}_i}\|_F,$$

其中 $l_i:=\|\mathbf{L}_i\|_F$。结合引理 7 可进一步得到

$$2\alpha_il_i\|\Delta_{\mathbf{A}_i}\|_F\leq 2\alpha_il_i\frac{2^{\frac{3}{2}}\eta_i}{\rho}\|\widehat{\Delta}\|_2,i=1,2,3.$$

将上述不等式与（79）结合可得

$$\begin{aligned}&\left(\frac{\varpi}{4}+\frac{1}{2}\sqrt{\frac{s\log(m^2p)}{(r_1r_2r_3)^2T}}\right)\|\widehat{\Delta}\|_2^2\\&\leq\beta\sqrt{\tilde{s}}\|\widehat{\Delta}\|_2+\sum_{i=1}^{3}\alpha_i l_i\frac{2^{\frac{5}{2}}\eta_i}{\rho}\|\widehat{\Delta}\|_2+b\beta\sqrt{s}\|\widehat{\Delta}\|_2+2\sqrt{2}bc\sqrt{\frac{\log(m^2p)}{T}}\|\widehat{\Delta}\|_2\end{aligned}$$

以至少 $1-2\exp(-\bar{p}\min(\log\ mp,\log(21emp/\bar{p}))$ 的概率。因此，我们得到

$$\begin{aligned}\|\widehat{\Delta}\|_2&\leq\frac{4\beta\sqrt{\tilde{s}}+\sum_{i=1}^{3}2^{\frac{9}{2}}\alpha_i l_i\frac{\eta_i}{\rho}+4b\beta\sqrt{s}+8\sqrt{2}bc\sqrt{(\log(m^2p))/T}}{\varpi+2\sqrt{(s\log(m^2p))/((r_1r_2r_3)^2T)}}\\&=\frac{4\beta(\sqrt{\tilde{s}}+b\sqrt{s})+\hat{a}\kappa+\bar{c}\sum_{i=1}^{3}\alpha_i l_i\eta_i}{\varpi+2\kappa\sqrt{s/(r_1r_2r_3)^2}}\end{aligned}\tag{80}$$

以至少 $1-2\exp(-\bar{p}\min(\log\ mp,\log(21emp/\bar{p})))$ 的概率，其中 $\hat{a}=8\sqrt{2}bc,\kappa=\sqrt{(\log(m^2p))/T}$ 且 $\bar{c}=\frac{2^{\frac{9}{2}}}{\rho}$。

此外，以至少 $1-2\exp(-\bar{p}\min(\log\ mp,\log(21emp/\bar{p})))$ 的概率，

$$\begin{aligned}T^{-1}\sum_{t=1}^{T}\|\widehat{\mathcal{W}}_{(1)}\mathbf{x}_t-\mathcal{W}_{(1)}\mathbf{x}_t\|_2^2&=T^{-1}\|\mathbf{X}\widehat{\mathcal{W}}_{(1)}^H-\mathbf{X}\mathcal{W}_{(1)}^H\|_F^2\\&=T^{-1}\|\operatorname{vec}(\mathbf{X}\widehat{\mathcal{W}}_{(1)}^H)-\operatorname{vec}(\mathbf{X}\mathcal{W}_{(1)}^H)\|_2^2\\&=T^{-1}\|\left(\mathbf{I}_m\otimes\mathbf{X}\right)\widehat{\mathbf{A}}\widehat{\mathbf{g}}-\left(\mathbf{I}_m\otimes\mathbf{X}\right)\mathbf{A}\mathbf{g}\|_2^2\\&=T^{-1}\|\left(\mathbf{I}_m\otimes\mathbf{X}\right)\widehat{\Delta}\|_2^2\\&=\widehat{\Delta}^H\left(\mathbf{I}_m\otimes(\mathbf{X}^H\mathbf{X}/T)\right)\widehat{\Delta}\\&\leq\lambda_{\max}(\mathbf{X}^H\mathbf{X}/T)\|\widehat{\Delta}\|_2^2\\&\leq\lambda_{\max}\left(\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}\right)/\mu_{\min}(\mathbf{W})\|\widehat{\Delta}\|_2^2\\&\leq\frac{\lambda_{\max}\left(\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}\right)(4\beta(\sqrt{\tilde{s}}+b\sqrt{s})+\hat{a}\kappa+\bar{c}\sum_{i=1}^{3}\alpha_i l_i\eta_i)^2}{\mu_{\min}(\mathbf{W})(\varpi+2\kappa\sqrt{s/(r_1r_2r_3)^2})^2},\end{aligned}$$

其中，第一个等式由（61）得出，第三个等式由（62）得出，第四个等式由 $\widehat{\ }\quad\widehat{\ }$ 得出，依据 $\Delta=\mathbf{A}\widehat{\ }\mathbf{g}-\mathbf{A}\mathbf{g}$，第一个不等式由 Rayleigh-Ritz 定理 [27,定理4.2。2]和 $\lambda_{\max}\left(\mathbf{I}_m\otimes(\mathbf{X}^H\mathbf{X}/T)\right)=\lambda_{\max}((\mathbf{X}^H\mathbf{X}/T))$得出，第二个不等式由 [31] 中的方程（6）得出，第三个不等式由（80）得出。证明完毕。□

附录 C. Ψ 在（15）中的梯度 Lipschitz 连续性

¯引理 9 设 $f(\mathcal{X})$: $\mathbb{R}^{n_1\times n_2\times n_3}\to\mathbb{R}$ 是可微函数，且 $f(\mathcal{X}_{(1)})$: $\mathbb{R}^{n_1\times n_2n_3}\to\mathbb{R}$ 是关于 $\mathcal{X}$ 的模-1 展开的¯可微函数，其中，对于任意 $\mathcal{X}$，$f(\mathcal{X})=f(\mathcal{X}_{(1)})$。则有

$$(\nabla f(\mathcal{X}))_{(1)}=\nabla\bar{f}(\mathcal{X}_{(1)}).$$

证明。对于任意三阶张量 $\mathcal{X}\in\mathbb{R}^{n_1\times n_2\times n_3}$，我们得到 $\mathcal{X}_{(1)}$ 的（i，n）元素为

$$(\mathcal{X}_{(1)})_{i,n}=\mathcal{X}_{ijl},\quad\text{where } n=(l-1)n_2+j,\quad i\in[n_1],j\in[n_2],l\in[n_3].$$

注意到 $(\nabla f(\mathcal{X}))_{ijl} = \frac{\partial f}{\partial \mathcal{X}_{ijl}}$，以及 $((\nabla f(\mathcal{X}))_{(1)})_{i,n} = (\nabla f(\mathcal{X}))_{ijl}$，其中 $n = (l-1)n_2 + j$.对于 f 关于 $\mathcal{X}_{(1)}$ 的梯度，记为 $\nabla f(\mathcal{X}_{(1)}) \in \mathbb{R}^{n_1 \times n_2 n_3}$，有

$$(\nabla \bar{f}(\mathcal{X}_{(1)}))_{i,n} = \frac{\partial \bar{f}}{\partial (\mathcal{X}_{(1)})_{i,n}}.$$

由于 $(\mathcal{X}_{(1)})_{i,n} = \mathcal{X}_{ijl}$ 且 $n = (l-1)n_2 + j$ 和 $f(\mathcal{X}) = f(\mathcal{X}_{(1)})$，我们得到

$$\frac{\partial \bar{f}}{\partial (\mathcal{X}_{(1)})_{i,n}} = \frac{\partial f}{\partial \mathcal{X}_{ijl}} = (\nabla f(\mathcal{X}))_{ijl}.$$

因此，我们得到 $(\nabla f(\mathcal{X}))_{(1)} = \nabla f(\mathcal{X}_{(1)})$。 □

注释4 设 $\bar{Q}(\mathcal{W}_{(1)}) := \frac{1}{2T}\sum_{t=1}^{T} \left\|\mathbf{y}_t - \mathcal{W}_{(1)}\mathbf{x}_t\right\|_2^2$ 是关于 $\mathcal{W}_{(1)}$的函数。记 $\bar{Q}$ 关于 Q 的梯度 $\mathcal{W}_{(1)}$ 为 $\nabla Q(\mathcal{W}_{(1)})$。根据引理9 和 (14)，可得

$$\nabla \bar{Q}(\mathcal{W}_{(1)}) = \frac{1}{T}\sum_{t=1}^{T} \left(\mathcal{W}_{(1)}\mathbf{x}_t - \mathbf{y}_t\right)\mathbf{x}_t^H = (\nabla Q([\![\mathcal{G}; \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3]\!]))_{(1)}, \tag{81}$$

其中 $\nabla Q([\![\mathcal{G};\ \mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3]\!])$ 定义见 (16)。

引理 10 假设 $\mathbf{A}_1 \in \mathbb{R}^{n_i \times r_i}$ 且 $\mathbf{A}_2 \in \mathfrak{B}_i$，其中 $n_1 = n_2 = m, n_3 = p$。因此，$\nabla_{\mathcal{G}}\Psi(\mathcal{G}, \mathbf{A}$ $\frac{1}{T}\sum_{t=1}^{T}\| \ _t\|_2^2$ $L_1 > 0$ $\mathcal{G}^1\ \mathcal{G}^2 \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ A A U U U) 是

Lipschitz连续的，其利普希茨常数为 L_1: x，也就是说，存在一个常数，使得对于任意的，，以下不等式成立:

$$\left\|\nabla_{\mathcal{G}}\Psi(\mathcal{G}^1, \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3) - \nabla_{\mathcal{G}}\Psi(\mathcal{G}^2, \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3)\right\|_F \le L_1 \left\|\mathcal{G}^1 - \mathcal{G}^2\right\|_F.$$

证明。设 $\mathcal{G}^1$，$\mathcal{G}^2 \in \mathbb{R}^{r_1 \times r_2 \times r_3}$，$\mathbf{A}_i \in \mathfrak{B}_i$，$i = 1$，$2$，3分别表示 $\mathcal{W}^i = \mathcal{G}^i \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \times_3 \mathbf{A}_3$，$i = 1$，$2$。将式（16）和式（17）结合起来，可得

$$\begin{aligned}
&\left\|\nabla_{\mathcal{G}}Q([\![\mathcal{G}^1; \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3]\!]) - \nabla_{\mathcal{G}}Q([\![\mathcal{G}^2; \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3]\!])\right\|_F \\
&= \left\| \left(\frac{1}{T}\sum_{t=1}^{T}(\mathcal{W}_{(1)}^1\mathbf{x}_t - \mathbf{y}_t) \circ \mathbf{X}_t\right) \times \mathbf{A}_1^H \times \mathbf{A}_2^H \times \mathbf{A}_3^H \right. \\
&\quad \left. - \left(\frac{1}{T}\sum_{t=1}^{T}(\mathcal{W}_{(1)}^2\mathbf{x}_t - \mathbf{y}_t) \circ \mathbf{X}_t\right) \times \mathbf{A}_1^H \times \mathbf{A}_2^H \times \mathbf{A}_3^H \right\|_F \\
&= \left\| \left(\frac{1}{T}\sum_{t=1}^{T}\left((\mathcal{W}_{(1)}^1\mathbf{x}_t - \mathbf{y}_t) \circ \mathbf{X}_t - (\mathcal{W}_{(1)}^2\mathbf{x}_t - \mathbf{y}_t) \circ \mathbf{X}_t\right)\right) \times \mathbf{A}_1^H \times \mathbf{A}_2^H \times \mathbf{A}_3^H \right\|_F \\
&= \left\| \frac{1}{T}\sum_{t=1}^{T}(\mathcal{W}_{(1)}^1\mathbf{x}_t - \mathbf{y}_t) \circ \mathbf{X}_t - \frac{1}{T}\sum_{t=1}^{T}(\mathcal{W}_{(1)}^2\mathbf{x}_t - \mathbf{y}_t) \circ \mathbf{X}_t \right\|_F \\
&= \left\| \left(\frac{1}{T}\sum_{t=1}^{T}(\mathcal{W}_{(1)}^1\mathbf{x}_t - \mathbf{y}_t) \circ \mathbf{X}_t\right)_{(1)} - \left(\frac{1}{T}\sum_{t=1}^{T}(\mathcal{W}_{(1)}^2\mathbf{x}_t - \mathbf{y}_t) \circ \mathbf{X}_t\right)_{(1)} \right\|_F,
\end{aligned} \tag{82}$$

其中第三个等号成立，是因为弗罗贝尼乌斯范数是酉不变的。由式（81）和式（16），可得

$$\begin{aligned}
&\left\| \left(\frac{1}{T}\sum_{t=1}^{T}(\mathcal{W}_{(1)}^{1}\mathbf{x}_t - \mathbf{y}_t)\circ\mathbf{X}_t \right)_{(1)} - \left(\frac{1}{T}\sum_{t=1}^{T}(\mathcal{W}_{(1)}^{2}\mathbf{x}_t - \mathbf{y}_t)\circ\mathbf{X}_t \right)_{(1)} \right\|_F \\
=& \left\| \frac{1}{T}\sum_{t=1}^{T}(\mathcal{W}_{(1)}^{1}\mathbf{x}_t - \mathbf{y}_t)\mathbf{x}_t^H - \frac{1}{T}\sum_{t=1}^{T}(\mathcal{W}_{(1)}^{2}\mathbf{x}_t - \mathbf{y}_t)\mathbf{x}_t^H \right\|_F \\
=& \left\| \frac{1}{T}\sum_{t=1}^{T}(\mathcal{W}_{(1)}^{1} - \mathcal{W}_{(1)}^{2})\mathbf{x}_t\mathbf{x}_t^H \right\|_F \\
\leq& \left\| \mathcal{W}_{(1)}^{1} - \mathcal{W}_{(1)}^{2} \right\|_F \left\| \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t^H \right\|_F \\
\leq& \left\| \mathcal{W}_{(1)}^{1} - \mathcal{W}_{(1)}^{2} \right\|_F \frac{1}{T}\sum_{t=1}^{T}\|\mathbf{x}_t\mathbf{x}_t^H\|_F \\
=& L_1 \left\| \mathcal{W}_{(1)}^{1} - \mathcal{W}_{(1)}^{2} \right\|_F,
\end{aligned} \tag{83}$$

其中，第一个不等式由柯西-施瓦茨不等式得出，最后一个等式由 $L_1 := \frac{1}{T}\sum_{t=1}^{T}\|\mathbf{x}_t\|_2^2 = \frac{1}{T}\sum_{t=1}^{T}\|\mathbf{x}_t\mathbf{x}_t^H\|_F$ 得出。注意到

$$\begin{aligned}
\|\mathcal{W}_{(1)}^{1} - \mathcal{W}_{(1)}^{2}\|_F &= \|\mathcal{W}^1 - \mathcal{W}^2\|_F \\
&= \|\mathcal{G}^1 \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \times_3 \mathbf{A}_3 - \mathcal{G}^2 \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \times_3 \mathbf{A}_3\|_F \\
&= \|(\mathcal{G}^1 - \mathcal{G}^2) \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \times_3 \mathbf{A}_3\|_F \\
&= \|\mathcal{G}^1 - \mathcal{G}^2\|_F.
\end{aligned}$$

结合上述结果以及（83）、（82），可得

$$\|\nabla_{\mathcal{G}} Q(\llbracket \mathcal{G}^1; \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3 \rrbracket) - \nabla_{\mathcal{G}} Q(\llbracket \mathcal{G}^2; \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3 \rrbracket)\|_F \leq L_1\|\mathcal{G}^1 - \mathcal{G}^2\|_F.$$

由（18），进一步得到

$$\|\nabla_{\mathcal{G}}\Psi(\mathcal{G}^1, \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3) - \nabla_{\mathcal{G}}\Psi(\mathcal{G}^2, \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3)\|_F \leq L_1\|\mathcal{G}^1 - \mathcal{G}^2\|_F$$

证明至此完成。 □

引理 11 假设 $\mathcal{G} \in \mathbb{R}^{r_1\times r_2\times r_3}$ 满足 $\mathcal{G} \in \mathcal{D}$，且 $\mathbf{A}_i \in \mathbb{R}^{n_i\times r_i}, i = 2, 3$ 满足 $\mathbf{A}_i \in \mathfrak{B}_i$，其中$n_2 = m, n_3 = p$。那么 $\nabla_{\mathbf{A}_1}\Psi(\mathcal{G}, \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3)$ 是 Lipschitz连续的，其利普希茨常数为 $L_2 := \nu^2 c_1 + \gamma_1$，其中 $\nu := \sqrt{r_1 r_2 r_3}c$ 且 $c_1 := \frac{1}{T}\sum_{t=1}^{T}\|\mathbf{x}_t\|_2^2$，也就是说，存在一个常数 $L_2 > 0$，使得对于任意 $\mathbf{A}_1^1, \mathbf{A}_1^2 \in \mathbb{R}^{m\times r_1}$，以下不等式成立：

$$\|\nabla_{\mathbf{A}_1}\Psi(\mathcal{G}, \mathbf{A}_1^1, \mathbf{A}_2, \mathbf{A}_3, \mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3) - \nabla_{\mathbf{A}_1}\Psi(\mathcal{G}, \mathbf{A}_1^2, \mathbf{A}_2, \mathbf{A}_3, \mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3)\|_F \leq L_2\|\mathbf{A}_1^1 - \mathbf{A}_1^2\|_F.$$

证明。假设 $\mathcal{G}\in\mathcal{D}$，$\mathbf{A}_i\in\mathfrak{B}_i$，$i=2,\ 3$。由（16）和（17），对于任意 $\mathbf{A}_1^1,\mathbf{A}_1^2\in\mathbb{R}^{m\times r_1}$，有

$$
\begin{aligned}
&\left\|\nabla_{\mathbf{A}_1}Q([\![\mathcal{G};\mathbf{A}_1^1,\mathbf{A}_2,\mathbf{A}_3]\!])-\nabla_{\mathbf{A}_1}Q([\![\mathcal{G};\mathbf{A}_1^2,\mathbf{A}_2,\mathbf{A}_3]\!])\right\|_F\\
=&\left\|\left(\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{A}_1^1\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2\right)^H\mathbf{x}_t-\mathbf{y}_t\right)\circ\mathbf{X}_t\right)_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2\right)\mathcal{G}_{(1)}^H\right.\\
&\left.-\left(\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{A}_1^2\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2\right)^H\mathbf{x}_t-\mathbf{y}_t\right)\circ\mathbf{X}_t\right)_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2\right)\mathcal{G}_{(1)}^H\right\|_F\\
\le&\left\|\left(\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{A}_1^1\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2\right)^H\mathbf{x}_t-\mathbf{y}_t\right)\circ\mathbf{X}_t\right)_{(1)}\right.\\
&\left.-\left(\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{A}_1^2\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2\right)^H\mathbf{x}_t-\mathbf{y}_t\right)\circ\mathbf{X}_t\right)_{(1)}\right\|_F\left\|\left(\mathbf{A}_3\otimes\mathbf{A}_2\right)\mathcal{G}_{(1)}^H\right\|_F\\
=&\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{A}_1^1\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2\right)^H\mathbf{x}_t-\mathbf{y}_t\right)\circ\mathbf{X}_t\right.\\
&\left.-\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{A}_1^2\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2\right)^H\mathbf{x}_t-\mathbf{y}_t\right)\circ\mathbf{X}_t\right\|_F\left\|\left(\mathbf{A}_3\otimes\mathbf{A}_2\right)\mathcal{G}_{(1)}^H\right\|_F,
\end{aligned}
\tag{84}
$$

其中第一个不等式由柯西-施瓦茨不等式得出。命题2结合[12, 引理2。1] 给出

$$
\left\|\left(\mathbf{A}_3\otimes\mathbf{A}_2\right)\mathcal{G}_{(1)}^H\right\|_F\le\|\mathbf{A}_3\otimes\mathbf{A}_2\|\left\|\mathcal{G}_{(1)}^H\right\|_F\le\left\|\mathcal{G}_{(1)}^H\right\|_F=\|\mathcal{G}_{(1)}\|_F=\|\mathcal{G}\|_F\le\nu,
\tag{85}
$$

其中，最后一个不等式由$\|\mathcal{G}\|_F\le\sqrt{r_1r_2r_3}\|\mathcal{G}\|_\infty\le\sqrt{r_1r_2r_3}c:=\nu$ 对任意 $\mathcal{G}\in\mathcal{D}$成立可得。此外，

$$
\begin{aligned}
&\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{A}_1^1\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2\right)^H\mathbf{x}_t-\mathbf{y}_t\right)\circ\mathbf{X}_t-\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{A}_1^2\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2\right)^H\mathbf{x}_t-\mathbf{y}_t\right)\circ\mathbf{X}_t\right\|_F\\
=&\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{A}_1^1\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2\right)^H\mathbf{x}_t-\mathbf{A}_1^2\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2\right)^H\mathbf{x}_t\right)\circ\mathbf{X}_t\right\|_F\\
=&\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\left(\mathbf{A}_1^1-\mathbf{A}_1^2\right)\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2\right)^H\mathbf{x}_t\right)\circ\mathbf{X}_t\right\|_F\\
\le&\frac{1}{T}\sum_{t=1}^{T}\left\|\left(\left(\mathbf{A}_1^1-\mathbf{A}_1^2\right)\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2\right)^H\mathbf{x}_t\right)\circ\mathbf{X}_t\right\|_F\\
=&\frac{1}{T}\sum_{t=1}^{T}\left\|\left(\mathbf{A}_1^1-\mathbf{A}_1^2\right)\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2\right)^H\mathbf{x}_t\right\|_2\|\mathbf{X}_t\|_F\\
\le&\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{A}_1^1-\mathbf{A}_1^2\|_F\left\|\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2\right)^H\right\|_F\|\mathbf{x}_t\|_2\|\mathbf{X}_t\|_F\\
\le&\frac{1}{T}\sum_{t=1}^{T}\nu\|\mathbf{A}_1^1-\mathbf{A}_1^2\|_F\|\mathbf{x}_t\|_2\|\mathbf{X}_t\|_F,
\end{aligned}
\tag{86}
$$

其中，最后一个等式由命题 4得出，第二个不等式由柯西-施瓦茨不等式得出，最后一个不等式由（85）以及 $\|\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2\right)^H\|_F=\|\left(\mathbf{A}_3\otimes\mathbf{A}_2\right)\mathcal{G}_{(1)}^H\|_F$。根据 $\mathbf{x}_t$ 和 $\mathbf{X}_t$的定义，不难看出

$$\|\mathbf{x}_t\|_2=\|\mathbf{X}_t\|_F. \tag{87}$$

结合（84）、（85）和（86），我们很容易得到

$$\begin{aligned}\|\nabla_{\mathbf{A}_1}Q([\![\mathcal{G};\mathbf{A}_1^1,\mathbf{A}_2,\mathbf{A}_3]\!])-\nabla_{\mathbf{A}_1}Q([\![\mathcal{G};\mathbf{A}_1^2,\mathbf{A}_2,\mathbf{A}_3]\!])\|_F&\le\nu^2\frac{1}{T}\sum_{t=1}^T\|\mathbf{x}_t\|_2^2\,\|\mathbf{A}_1^1-\mathbf{A}_1^2\|_F\\&=\nu^2c_1\|\mathbf{A}_1^1-\mathbf{A}_1^2\|_F,\end{aligned}$$

其中 $c_1:=\frac{1}{T}\sum_{t=1}^T\|\mathbf{x}_t\|_2^2$。结合（18）可得

$$\begin{aligned}&\|\nabla_{\mathbf{A}_1}\Psi(\mathcal{G},\mathbf{A}_1^1,\mathbf{A}_2,\mathbf{A}_3,\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3)-\nabla_{\mathbf{A}_1}\Psi(\mathcal{G},\mathbf{A}_1^2,\mathbf{A}_2,\mathbf{A}_3,\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3)\|_F\\&=\|\nabla_{\mathbf{A}_1}Q([\![\mathcal{G};\mathbf{A}_1^1,\mathbf{A}_2,\mathbf{A}_3]\!])-\nabla_{\mathbf{A}_1}Q([\![\mathcal{G};\mathbf{A}_1^2,\mathbf{A}_2,\mathbf{A}_3]\!])+\gamma_1(\mathbf{A}_1^1-\mathbf{A}_1^2)\|_F\\&\le\|\nabla_{\mathbf{A}_1}Q([\![\mathcal{G};\mathbf{A}_1^1,\mathbf{A}_2,\mathbf{A}_3]\!])-\nabla_{\mathbf{A}_1}Q([\![\mathcal{G};\mathbf{A}_1^2,\mathbf{A}_2,\mathbf{A}_3]\!])\|_F+\gamma_1\|\mathbf{A}_1^1-\mathbf{A}_1^2\|_F\\&\le\nu^2c_1\|\mathbf{A}_1^1-\mathbf{A}_1^2\|_F+\gamma_1\|\mathbf{A}_1^1-\mathbf{A}_1^2\|_F\\&=(\nu^2c_1+\gamma_1)\|\mathbf{A}_1^1-\mathbf{A}_1^2\|_F=L_2\|\mathbf{A}_1^1-\mathbf{A}_1^2\|_F,\end{aligned}\tag{88}$$

其中 $L_2:=\nu^2c_1+\gamma_1$。证毕。 □

引理 12 设 $\mathcal{G}\in\mathbb{R}^{r_1\times r_2\times r_3}$ with $\mathcal{G}\in\mathcal{D}$, $\mathrm{A}_i\in\mathbb{R}^{n_i\times r_i}$ with $\mathrm{A}_i\in\mathfrak{B}_i, i=1,3$, where $n_1=m, n_3=p$。则 $\nabla_{\mathbf{A}_2}\Psi(\mathcal{G},\mathrm{A}_1,\mathrm{A}_2,\mathrm{A}_3,\mathrm{U}_1,\mathrm{U}_2,\mathrm{U}_3)$ 是 Lipschitz 连续的，其利普希茨常数为 $L_3:=\nu^2c_1+\gamma_2$，其中 $\nu:=\sqrt{r_1r_2r_3}c$ and $c_1:=\frac{1}{T}\sum_{t=1}^T\|\mathrm{x}_t\|_2^2$，即存在一个常数 $L_3>0$，使得对于任意 $\mathrm{A}_2^1,\mathrm{A}_2^2\in\mathbb{R}^{m\times r_2}$，以下不等式成立：

$$\left\|\nabla_{\mathbf{A}_2}\Psi(\mathcal{G},\mathbf{A}_1,\mathbf{A}_2^1,\mathbf{A}_3,\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3)-\nabla_{\mathbf{A}_2}\Psi(\mathcal{G},\mathbf{A}_1,\mathbf{A}_2^2,\mathbf{A}_3,\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3)\right\|_F\le L_3\left\|\mathbf{A}_2^1-\mathbf{A}_2^2\right\|_F.$$

证明。类似于对 $\nabla_{\mathbf{A}_1}Q$ 的分析，我们有

$$\begin{aligned}&\left\|\nabla_{\mathbf{A}_2}Q([\![\mathcal{G};\mathbf{A}_1,\mathbf{A}_2^1,\mathbf{A}_3]\!])-\nabla_{\mathbf{A}_2}Q([\![\mathcal{G};\mathbf{A}_1,\mathbf{A}_2^2,\mathbf{A}_3]\!])\right\|_F\\&=\left\|\left(\frac{1}{T}\sum_{t=1}^T\left(\mathbf{A}_1\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2^1\right)^H\mathbf{x}_t-\mathbf{y}_t\right)\circ\mathbf{X}_t\right)_{(2)}\left(\mathbf{A}_1\otimes\mathbf{A}_3\right)\mathcal{G}_{(2)}^H\right.\\&\quad\left.-\left(\frac{1}{T}\sum_{t=1}^T\left(\mathbf{A}_1\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2^2\right)^H\mathbf{x}_t-\mathbf{y}_t\right)\circ\mathbf{X}_t\right)_{(2)}\left(\mathbf{A}_1\otimes\mathbf{A}_3\right)\mathcal{G}_{(2)}^H\right\|_F\\&\le\left\|\left(\frac{1}{T}\sum_{t=1}^T\left(\mathbf{A}_1\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2^1\right)^H\mathbf{x}_t-\mathbf{y}_t\right)\circ\mathbf{X}_t\right)_{(2)}\right.\\&\quad\left.-\left(\frac{1}{T}\sum_{t=1}^T\left(\mathbf{A}_1\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2^2\right)^H\mathbf{x}_t-\mathbf{y}_t\right)\circ\mathbf{X}_t\right)_{(2)}\right\|_F\nu\\&=\left\|\frac{1}{T}\sum_{t=1}^T\mathbf{X}_t\otimes\left(\mathbf{A}_1\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2^1\right)^H\mathbf{x}_t-\mathbf{y}_t\right)^H\right.\\&\quad\left.-\frac{1}{T}\sum_{t=1}^T\mathbf{X}_t\otimes\left(\mathbf{A}_1\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2^2\right)^H\mathbf{x}_t-\mathbf{y}_t\right)^H\right\|_F\nu,\end{aligned}\tag{89}$$

其中 $\nu:=\sqrt{r_1r_2r_3}c$ 最后一个等式由命题 3 得出。注意到

$$\begin{aligned}
&\left\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t\otimes\left(\mathbf{A}_1\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2^1\right)^H\mathbf{x}_t-\mathbf{y}_t\right)^H\right.\\
&\left.-\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t\otimes\left(\mathbf{A}_1\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2^2\right)^H\mathbf{x}_t-\mathbf{y}_t\right)^H\right\|_F\\
=&\left\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t\otimes\left(\mathbf{A}_1\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2^1\right)^H\mathbf{x}_t-\mathbf{A}_1\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2^2\right)^H\mathbf{x}_t\right)^H\right\|_F\\
=&\left\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t\otimes\left(\mathbf{A}_1\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes(\mathbf{A}_2^1-\mathbf{A}_2^2)\right)^H\mathbf{x}_t\right)^H\right\|_F\\
\le&\frac{1}{T}\sum_{t=1}^{T}\left\|\mathbf{X}_t\otimes\left(\mathbf{A}_1\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes(\mathbf{A}_2^1-\mathbf{A}_2^2)\right)^H\mathbf{x}_t\right)^H\right\|_F\\
=&\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{X}_t\|_F\left\|\mathbf{A}_1\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes(\mathbf{A}_2^1-\mathbf{A}_2^2)\right)^H\mathbf{x}_t\right\|_2,
\end{aligned}\tag{90}$$

其中第三个等式由命题 1 得出。根据 [27,例 5.66] 可知

$$\begin{aligned}
&\left\|\mathbf{A}_1\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes(\mathbf{A}_2^1-\mathbf{A}_2^2)\right)^H\mathbf{x}_t\right\|_2\\
\le&\|\mathbf{A}_1\|\left\|\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes(\mathbf{A}_2^1-\mathbf{A}_2^2)\right)^H\mathbf{x}_t\right\|_2\\
\le&\left\|\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes(\mathbf{A}_2^1-\mathbf{A}_2^2)\right)^H\mathbf{x}_t\right\|_2\\
\le&\left\|\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes(\mathbf{A}_2^1-\mathbf{A}_2^2)\right)^H\right\|_F\|\mathbf{x}_t\|_2\\
\le&\left\|\mathcal{G}_{(1)}\right\|_F\left\|\left(\mathbf{A}_3\otimes(\mathbf{A}_2^1-\mathbf{A}_2^2)\right)^H\right\|\|\mathbf{x}_t\|_2\\
=&\|\mathcal{G}\|_F\|\mathbf{A}_3\|\left\|(\mathbf{A}_2^1-\mathbf{A}_2^2)^H\right\|\|\mathbf{x}_t\|_2\\
\le&\nu\|\mathbf{A}_3\|\left\|(\mathbf{A}_2^1-\mathbf{A}_2^2)^H\right\|\|\mathbf{x}_t\|_2\\
\le&\nu\left\|(\mathbf{A}_2^1-\mathbf{A}_2^2)^H\right\|_F\|\mathbf{x}_t\|_2,
\end{aligned}\tag{91}$$

其中，第二个不等式由$\mathbf{A}_1\in\mathfrak{B}_1$ 与 $\|\mathbf{A}_1\|\le 1$成立，第三个不等式由柯西-施瓦茨不等式成立，第四个不等式由 [12, 引理 2。1], 第五个不等式由 $\mathcal{G}\in\mathcal{D}$ 与 $\|\mathcal{G}\|_F\le\nu$ 成立，最后一个不等式由 $\|\mathbf{A}_3\|\le 1$以及任意矩阵$\mathbf{X}$ 满足$\|\mathbf{X}\|\le\|\mathbf{X}\|_F$成立。结合式（91）和（90），可得

$$\begin{aligned}
&\left\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t\otimes\left(\mathbf{A}_1\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2^1\right)^H\mathbf{x}_t-\mathbf{y}_t\right)^H\right.\\
&\left.-\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t\otimes\left(\mathbf{A}_1\mathcal{G}_{(1)}\left(\mathbf{A}_3\otimes\mathbf{A}_2^2\right)^H\mathbf{x}_t-\mathbf{y}_t\right)^H\right\|_F\\
\le&\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{X}_t\|_F\nu\left\|(\mathbf{A}_2^1-\mathbf{A}_2^2)^H\right\|_F\|\mathbf{x}_t\|_2\\
=&\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{x}_t\|_2^2\nu\left\|\mathbf{A}_2^1-\mathbf{A}_2^2\right\|_F=\nu c_1\left\|\mathbf{A}_2^1-\mathbf{A}_2^2\right\|_F,
\end{aligned}\tag{92}$$

其中，第一个等式由式（87）和 $c_1 := \frac{1}{T}\sum_{t=1}^{T}\|\mathbf{x}_t\|_2^2$得到。结合式（89），可得

$$\|\nabla_{\mathbf{A}_2} Q(\llbracket \mathcal{G}; \mathbf{A}_1, \mathbf{A}_2^1, \mathbf{A}_3 \rrbracket) - \nabla_{\mathbf{A}_2} Q(\llbracket \mathcal{G}; \mathbf{A}_1, \mathbf{A}_2^2, \mathbf{A}_3 \rrbracket)\|_F \leq \nu^2 c_1 \|\mathbf{A}_2^1 - \mathbf{A}_2^2\|_F.$$

类似于（88），我们有

$$\|\nabla_{\mathbf{A}_2}\Psi(\mathcal{G}, \mathbf{A}_1, \mathbf{A}_2^1, \mathbf{A}_3, \mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3) - \nabla_{\mathbf{A}_2}\Psi(\mathcal{G}, \mathbf{A}_1, \mathbf{A}_2^2, \mathbf{A}_3, \mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3)\|_F \leq L_3 \|\mathbf{A}_2^1 - \mathbf{A}_2^2\|_F,$$

其中 $L_3 := \nu^2 c_1 + \gamma_2$。证明完成。 □

引理 13 设 $\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$，且 $\mathcal{G} \in \mathcal{D}$，$\mathbf{A}_i \in \mathbb{R}^{n_i \times r_i}$，$\mathbf{A}_i \in \mathfrak{B}_i$，$i = 1, 2$，其中$n_1 = n_2 = m$。则 $\nabla_{\mathbf{A}_3}\Psi(\mathcal{G}, \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3)$ 是 Lipschitz连续的，其 Lipschitz常数为 $L_4 := \nu^2 c_1 + \gamma_3$，其中 $\nu := \sqrt{r_1 r_2 r_3}c$，且 $c_1 := \frac{1}{T}\sum_{t=1}^{T}\|\mathbf{x}_t\|_2^2$，也就是说，存在一个常数$L_4 > 0$，使得对于任意 $\mathbf{A}_3^1$,$\mathbf{A}_3^2 \in \mathbb{R}^{p \times r_3}$，以下不等式成立：

$$\|\nabla_{\mathbf{A}_3}\Psi(\mathcal{G}, \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3^1, \mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3) - \nabla_{\mathbf{A}_3}\Psi(\mathcal{G}, \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3^2, \mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3)\|_F \leq L_4 \|\mathbf{A}_3^1 - \mathbf{A}_3^2\|_F.$$

证明。类似于对 $\nabla_{\mathbf{A}_1} Q$ 的分析，根据（17），我们有

$$\begin{aligned}
&\left\|\nabla_{\mathbf{A}_3} Q(\llbracket \mathcal{G}; \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3^1 \rrbracket) - \nabla_{\mathbf{A}_3} Q(\llbracket \mathcal{G}; \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3^2 \rrbracket)\right\|_F \\
=&\left\| \left(\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{A}_1 \mathcal{G}_{(1)} \left(\mathbf{A}_3^1 \otimes \mathbf{A}_2\right)^H \mathbf{x}_t - \mathbf{y}_t\right) \circ \mathbf{X}_t\right)_{(3)} (\mathbf{A}_2 \otimes \mathbf{A}_1)\, \mathcal{G}_{(3)}^H \right. \\
&\left. - \left(\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{A}_1 \mathcal{G}_{(1)} \left(\mathbf{A}_3^2 \otimes \mathbf{A}_2\right)^H \mathbf{x}_t - \mathbf{y}_t\right) \circ \mathbf{X}_t\right)_{(3)} (\mathbf{A}_2 \otimes \mathbf{A}_1)\, \mathcal{G}_{(3)}^H \right\|_F \\
\leq&\left\| \left(\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{A}_1 \mathcal{G}_{(1)} \left(\mathbf{A}_3^1 \otimes \mathbf{A}_2\right)^H \mathbf{x}_t - \mathbf{y}_t\right) \circ \mathbf{X}_t\right)_{(3)} \right. \\
&\left. - \left(\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{A}_1 \mathcal{G}_{(1)} \left(\mathbf{A}_3^2 \otimes \mathbf{A}_2\right)^H \mathbf{x}_t - \mathbf{y}_t\right) \circ \mathbf{X}_t\right)_{(3)} \right\|_F \nu \\
=&\left\| \frac{1}{T}\sum_{t=1}^{T} \mathbf{X}_t^H \otimes \left(\mathbf{A}_1 \mathcal{G}_{(1)} \left(\mathbf{A}_3^1 \otimes \mathbf{A}_2\right)^H \mathbf{x}_t - \mathbf{y}_t\right)^H \right. \\
&\left. - \frac{1}{T}\sum_{t=1}^{T} \mathbf{X}_t^H \otimes \left(\mathbf{A}_1 \mathcal{G}_{(1)} \left(\mathbf{A}_3^2 \otimes \mathbf{A}_2\right)^H \mathbf{x}_t - \mathbf{y}_t\right)^H \right\|_F \nu
\end{aligned} \tag{93}$$

其中 $\nu := \sqrt{r_1 r_2 r_3}c$ 最后一个等式由命题 3 得出。类似于（92），我们有 e

$$\begin{aligned}
&\left\| \frac{1}{T}\sum_{t=1}^{T} \mathbf{X}_t^H \otimes \left(\mathbf{A}_1 \mathcal{G}_{(1)} \left(\mathbf{A}_3^1 \otimes \mathbf{A}_2\right)^H \mathbf{x}_t - \mathbf{y}_t\right)^H \right. \\
&\left. - \frac{1}{T}\sum_{t=1}^{T} \mathbf{X}_t^H \otimes \left(\mathbf{A}_1 \mathcal{G}_{(1)} \left(\mathbf{A}_3^2 \otimes \mathbf{A}_2\right)^H \mathbf{x}_t - \mathbf{y}_t\right)^H \right\|_F \\
\leq&\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{X}_t\|_F \left\| \left(\mathbf{A}_1 \mathcal{G}_{(1)} \left((\mathbf{A}_3^1 - \mathbf{A}_3^2) \otimes \mathbf{A}_2\right)^H \mathbf{x}_t\right)^H \right\|_F \leq \nu c_1 \|\mathbf{A}_3^1 - \mathbf{A}_3^2\|_F,
\end{aligned} \tag{94}$$

其中 $c_1 := \frac{1}{T}\sum_{t=1}^{T}\|\mathbf{x}_t\|_2^2$。结合（93）和（94），可得

$$\|\nabla_{\mathbf{A}_3} Q(\llbracket \mathcal{G}; \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3^1 \rrbracket) - \nabla_{\mathbf{A}_3} Q(\llbracket \mathcal{G}; \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3^2 \rrbracket)\|_F \leq \nu^2 c_1 \|\mathbf{A}_3^1 - \mathbf{A}_3^2\|_F.$$

此外，类似于（88），我们有

$$\|\nabla_{\mathbf{A}_3}\Psi(\mathcal{G}, \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3^1, \mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3) - \nabla_{\mathbf{A}_3}\Psi(\mathcal{G}, \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3^2, \mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3)\|_F \leq L_4 \left\|\mathbf{A}_3^1 - \mathbf{A}_3^2\right\|_F,$$

其中 $L_4 := \nu^2 c_1 + \gamma_3$。证明完毕。 □

由（18）可 쉽게验证

$$\begin{aligned}&\|\nabla_{\mathbf{U}_1}\Psi(\mathcal{G}, \mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3,\mathbf{U}_1^1,\mathbf{U}_2,\mathbf{U}_3) - \nabla_{\mathbf{U}_1}\Psi(\mathcal{G}, \mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3,\mathbf{U}_1^2,\mathbf{U}_2,\mathbf{U}_3)\|_F \leq \gamma_1\|\mathbf{U}_1^1 - \mathbf{U}_1^2\|_F,\|\nabla_{\mathbf{U}_2}\Psi\\ &(\mathcal{G}, \mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3,\mathbf{U}_1,\mathbf{U}_2^1,\mathbf{U}_3) - \nabla_{\mathbf{U}_2}\Psi(\mathcal{G}, \mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3,\mathbf{U}_1,\mathbf{U}_2^2,\mathbf{U}_3)\|_F \leq \gamma_2\|\mathbf{U}_2^1 - \mathbf{U}_2^2\|_F,\|\nabla_{\mathbf{U}_3}\Psi(\mathcal{G}, \mathbf{A}_1,\\ &\mathbf{A}_2,\mathbf{A}_3,\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3^1) - \nabla_{\mathbf{U}_3}\Psi(\mathcal{G}, \mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3,\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3^2)\|_F \leq \gamma_3\|\mathbf{U}_3^1 - \mathbf{U}_3^2\|_F.\end{aligned} \tag{95}$$

附录 D。引理证明1

由（19）可知

$$\begin{aligned}&\beta\|\mathcal{G}^{k+1}\|_1 + \delta_{\mathcal{D}}(\mathcal{G}^{k+1}) + \left\langle \nabla_{\mathcal{G}}\Psi(\mathcal{G}^k, \mathbf{A}_1^k, \mathbf{A}_2^k, \mathbf{A}_3^k, \mathbf{U}_1^k, \mathbf{U}_2^k, \mathbf{U}_3^k), \mathcal{G}^{k+1} - \mathcal{G}^k \right\rangle + \frac{\rho_1}{2}\|\mathcal{G}^{k+1} - \mathcal{G}^k\|_F^2\\ &\leq \beta\|\mathcal{G}^k\|_1 + \delta_{\mathcal{D}}(\mathcal{G}^k).\end{aligned} \tag{96}$$

此外，应用引理 10 和 [8,引理 1], 可得

$$\begin{aligned}&\Psi\left(\mathcal{G}^{k+1}, \mathbf{A}_1^k, \mathbf{A}_2^k, \mathbf{A}_3^k, \mathbf{U}_1^k, \mathbf{U}_2^k, \mathbf{U}_3^k\right) \leq \Psi\left(\mathcal{G}^k, \mathbf{A}_1^k, \mathbf{A}_2^k, \mathbf{A}_3^k, \mathbf{U}_1^k, \mathbf{U}_2^k, \mathbf{U}_3^k\right)\\ &+ \left\langle \nabla_{\mathcal{G}}\Psi(\mathcal{G}^k, \mathbf{A}_1^k, \mathbf{A}_2^k, \mathbf{A}_3^k, \mathbf{U}_1^k, \mathbf{U}_2^k, \mathbf{U}_3^k), \mathcal{G}^{k+1} - \mathcal{G}^k \right\rangle + \frac{L_1}{2}\|\mathcal{G}^{k+1} - \mathcal{G}^k\|_F^2.\end{aligned} \tag{97}$$

结合（96）和（97），可得

$$\begin{aligned}&\Psi\left(\mathcal{G}^{k+1}, \mathbf{A}_1^k, \mathbf{A}_2^k, \mathbf{A}_3^k, \mathbf{U}_1^k, \mathbf{U}_2^k, \mathbf{U}_3^k\right) + \beta\|\mathcal{G}^{k+1}\|_1 + \delta_{\mathcal{D}}(\mathcal{G}^{k+1})\\ &+ \left\langle \nabla_{\mathcal{G}}\Psi(\mathcal{G}^k, \mathbf{A}_1^k, \mathbf{A}_2^k, \mathbf{A}_3^k, \mathbf{U}_1^k, \mathbf{U}_2^k, \mathbf{U}_3^k), \mathcal{G}^{k+1} - \mathcal{G}^k \right\rangle + \frac{\rho_1}{2}\|\mathcal{G}^{k+1} - \mathcal{G}^k\|_F^2\\ &\leq \Psi\left(\mathcal{G}^k, \mathbf{A}_1^k, \mathbf{A}_2^k, \mathbf{A}_3^k, \mathbf{U}_1^k, \mathbf{U}_2^k, \mathbf{U}_3^k\right) + \left\langle \nabla_{\mathcal{G}}\Psi(\mathcal{G}^k, \mathbf{A}_1^k, \mathbf{A}_2^k, \mathbf{A}_3^k, \mathbf{U}_1^k, \mathbf{U}_2^k, \mathbf{U}_3^k), \mathcal{G}^{k+1} - \mathcal{G}^k \right\rangle\\ &+ \frac{L_1}{2}\|\mathcal{G}^{k+1} - \mathcal{G}^k\|_F^2 + \beta\|\mathcal{G}^k\|_1 + \delta_{\mathcal{D}}(\mathcal{G}^k),\end{aligned}$$

由此可得

$$\begin{aligned}&\Psi\left(\mathcal{G}^{k+1}, \mathbf{A}_1^k, \mathbf{A}_2^k, \mathbf{A}_3^k, \mathbf{U}_1^k, \mathbf{U}_2^k, \mathbf{U}_3^k\right) + \beta\|\mathcal{G}^{k+1}\|_1 + \delta_{\mathcal{D}}(\mathcal{G}^{k+1}) + \frac{\rho_1 - L_1}{2}\|\mathcal{G}^{k+1} - \mathcal{G}^k\|_F^2\\ &\leq \Psi\left(\mathcal{G}^k, \mathbf{A}_1^k, \mathbf{A}_2^k, \mathbf{A}_3^k, \mathbf{U}_1^k, \mathbf{U}_2^k, \mathbf{U}_3^k\right) + \beta\|\mathcal{G}^k\|_1 + \delta_{\mathcal{D}}(\mathcal{G}^k).\end{aligned} \tag{98}$$

重复类似的思路和技术，并结合引理 11-13 与（95）和（19），我们可以得到

$$\delta_{\mathfrak{B}_1}(\mathrm{A}_1^{k+1}) + \Psi\left(\mathcal{G}^{k+1}, \mathrm{A}_1^{k+1}, \mathrm{A}_2^{k}, \mathrm{A}_3^{k}, \mathrm{U}_1^{k}, \mathrm{U}_2^{k}, \mathrm{U}_3^{k}\right) + \frac{\rho_2 - L_2}{2}\|\mathrm{A}_1^{k+1} - \mathrm{A}_1^{k}\|_F^2 \leq \delta_{\mathfrak{B}_1}(\mathrm{A}_1^{k}) + \Psi\left(\mathcal{G}^{k+1}, \mathrm{A}_1^{k}, \mathrm{A}_2^{k}, \mathrm{A}_3^{k}, \mathrm{U}_1^{k}, \mathrm{U}_2^{k}, \mathrm{U}_3^{k}\right), \delta_{\mathfrak{B}_2}(\mathrm{A}_2^{k+1}) + \Psi\left(\mathcal{G}^{k+1}, \mathrm{A}_1^{k+1}, \mathrm{A}_2^{k+1}, \mathrm{A}_3^{k}, \mathrm{U}_1^{k}, \mathrm{U}_2^{k}, \mathrm{U}_3^{k}\right) + \frac{\rho_3 - L_3}{2}\|\mathrm{A}_2^{k+1} - \mathrm{A}_2^{k}\|_F^2 \leq \delta_{\mathfrak{B}_2}(\mathrm{A}_2^{k}) + \Psi\left(\mathcal{G}^{k+1}, \mathrm{A}_1^{k+1}, \mathrm{A}_2^{k}, \mathrm{A}_3^{k}, \mathrm{U}_1^{k}, \mathrm{U}_2^{k}, \mathrm{U}_3^{k}\right), \delta_{\mathfrak{B}_3}(\mathrm{A}_3^{k+1}) + \Psi\left(\mathcal{G}^{k+1}, \mathrm{A}_1^{k+1}, \mathrm{A}_2^{k+1}, \mathrm{A}_3^{k+1}, \mathrm{U}_1^{k}, \mathrm{U}_2^{k}, \mathrm{U}_3^{k}\right) + \frac{\rho_4 - L_4}{2}\|\mathrm{A}_3^{k+1} - \mathrm{A}_3^{k}\|_F^2 \leq \delta_{\mathfrak{B}_3}(\mathrm{A}_3^{k}) + \Psi\left(\mathcal{G}^{k+1}, \mathrm{A}_1^{k+1}, \mathrm{A}_2^{k+1}, \mathrm{A}_3^{k}, \mathrm{U}_1^{k}, \mathrm{U}_2^{k}, \mathrm{U}_3^{k}\right), \alpha_1 迹\left((\mathrm{U}_1^{k+1})^H \mathrm{L}_1 \mathrm{U}_1^{k+1}\right) + \Psi\left(\mathcal{G}^{k+1}, \mathrm{A}_1^{k+1}, \mathrm{A}_2^{k+1}, \mathrm{A}_3^{k+1}, \mathrm{U}_1^{k+1}, \mathrm{U}_2^{k}, \mathrm{U}_3^{k}\right) + \frac{\rho_5 - \gamma_1}{2}\|\mathrm{U}_1^{k+1} - \mathrm{U}_1^{k}\|_F^2 \leq \alpha_1 迹\left((\mathrm{U}_1^{k})^H \mathrm{L}_1 \mathrm{U}_1^{k}\right) + \Psi\left(\mathcal{G}^{k+1}, \mathrm{A}_1^{k+1}, \mathrm{A}_2^{k+1}, \mathrm{A}_3^{k+1}, \mathrm{U}_1^{k}, \mathrm{U}_2^{k}, \mathrm{U}_3^{k}\right), \alpha_2 迹\left((\mathrm{U}_2^{k+1})^H \mathrm{L}_2 \mathrm{U}_2^{k+1}\right) + \Psi\left(\mathcal{G}^{k+1}, \mathrm{A}_1^{k+1}, \mathrm{A}_2^{k+1}, \mathrm{A}_3^{k+1}, \mathrm{U}_1^{k+1}, \mathrm{U}_2^{k+1}, \mathrm{U}_3^{k}\right) + \frac{\rho_6 - \gamma_2}{2}\|\mathrm{U}_2^{k+1} - \mathrm{U}_2^{k}\|_F^2 \leq \alpha_2 迹\left((\mathrm{U}_2^{k})^H \mathrm{L}_2 \mathrm{U}_2^{k}\right) + \Psi\left(\mathcal{G}^{k+1}, \mathrm{A}_1^{k+1}, \mathrm{A}_2^{k+1}, \mathrm{A}_3^{k+1}, \mathrm{U}_1^{k+1}, \mathrm{U}_2^{k+1}, \mathrm{U}_3^{k}\right), \alpha_3 迹\left((\mathrm{U}_3^{k+1})^H \mathrm{L}_3 \mathrm{U}_3^{k+1}\right) + \Psi\left(\mathcal{G}^{k+1}, \mathrm{A}_1^{k+1}, \mathrm{A}_2^{k+1}, \mathrm{A}_3^{k+1}, \mathrm{U}_1^{k+1}, \mathrm{U}_2^{k+1}, \mathrm{U}_3^{k+1}\right) + \frac{\rho_7 - \gamma_3}{2}\|\mathrm{U}_3^{k+1} - \mathrm{U}_3^{k}\|_F^2 \leq \alpha_3 迹\left((\mathrm{U}_3^{k})^H \mathrm{L}_3 \mathrm{U}_3^{k}\right) + \Psi\left(\mathcal{G}^{k+1}, \mathrm{A}_1^{k+1}, \mathrm{A}_2^{k+1}, \mathrm{A}_3^{k+1}, \mathrm{U}_1^{k+1}, \mathrm{U}_2^{k+1}, \mathrm{U}_3^{k}\right).$$

(99)

结合式（98），可得

$$\begin{aligned}
&\beta\|\mathcal{G}^{k+1}\|_1 + \delta_{\mathcal{D}}(\mathcal{G}^{k+1}) + \Psi\left(\mathcal{G}^{k+1}, \mathbf{A}_1^{k+1}, \mathbf{A}_2^{k+1}, \mathbf{A}_3^{k+1}, \mathbf{U}_1^{k+1}, \mathbf{U}_2^{k+1}, \mathbf{U}_3^{k+1}\right) + \sum_{i=1}^{3}\delta_{\mathfrak{B}_i}(\mathbf{A}_i^{k+1}) \\
&+\frac{\rho_1 - L_1}{2}\|\mathcal{G}^{k+1} - \mathcal{G}^{k}\|_F^2 + \frac{\rho_2 - L_2}{2}\|\mathbf{A}_1^{k+1} - \mathbf{A}_1^{k}\|_F^2 + \frac{\rho_3 - L_3}{2}\|\mathbf{A}_2^{k+1} - \mathbf{A}_2^{k}\|_F^2 \\
&+\frac{\rho_4 - L_4}{2}\|\mathbf{A}_3^{k+1} - \mathbf{A}_3^{k}\|_F^2 + \frac{\rho_5 - \gamma_1}{2}\|\mathbf{U}_1^{k+1} - \mathbf{U}_1^{k}\|_F^2 + \frac{\rho_6 - \gamma_2}{2}\|\mathbf{U}_2^{k+1} - \mathbf{U}_2^{k}\|_F^2 \\
&+\frac{\rho_7 - \gamma_3}{2}\|\mathbf{U}_3^{k+1} - \mathbf{U}_3^{k}\|_F^2 + \sum_{i=1}^{3}\alpha_i \operatorname{tr}\left((\mathbf{U}_i^{k+1})^H \mathbf{L}_i \mathbf{U}_i^{k+1}\right) \\
&\leq \beta\|\mathcal{G}^{k}\|_1 + \delta_{\mathcal{D}}(\mathcal{G}^{k}) + \Psi\left(\mathcal{G}^{k}, \mathbf{A}_1^{k}, \mathbf{A}_2^{k}, \mathbf{A}_3^{k}, \mathbf{U}_1^{k}, \mathbf{U}_2^{k}, \mathbf{U}_3^{k}\right) + \sum_{i=1}^{3}\delta_{\mathfrak{B}_i}(\mathbf{A}_i^{k}) + \sum_{i=1}^{3}\alpha_i \operatorname{tr}\left((\mathbf{U}_i^{k})^H \mathbf{L}_i \mathbf{U}_i^{k}\right).
\end{aligned}$$

记 $\bar{\rho}:=\min\{\rho_1-L_1,\rho_2-L_2,\rho_3-L_3,\rho_4-L_4,\rho_5-\gamma_1,\rho_6-\gamma_2,\rho_7-\gamma_3\}$，则有

$$
\begin{aligned}
&\beta\|\mathcal{G}^{k+1}\|_1+\delta_{\mathcal{D}}(\mathcal{G}^{k+1})+\Psi\left(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1}\right)+\sum_{i=1}^{3}\delta_{\mathfrak{B}_i}(\mathbf{A}_i^{k+1})\\
&+\frac{\bar{\rho}}{2}\left(\|\mathcal{G}^{k+1}-\mathcal{G}^{k}\|_F^2+\|\mathbf{A}_1^{k+1}-\mathbf{A}_1^{k}\|_F^2+\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F^2+\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F^2\right.\\
&\left.+\|\mathbf{U}_1^{k+1}-\mathbf{U}_1^{k}\|_F^2+\|\mathbf{U}_2^{k+1}-\mathbf{U}_2^{k}\|_F^2+\|\mathbf{U}_3^{k+1}-\mathbf{U}_3^{k}\|_F^2\right)+\sum_{i=1}^{3}\alpha_i\operatorname{tr}\left((\mathbf{U}_i^{k+1})^H\mathbf{L}_i\mathbf{U}_i^{k+1}\right)\\
&\leq\beta\|\mathcal{G}^{k}\|_1+\delta_{\mathcal{D}}(\mathcal{G}^{k})+\Psi\left(\mathcal{G}^{k},\mathbf{A}_1^{k},\mathbf{A}_2^{k},\mathbf{A}_3^{k},\mathbf{U}_1^{k},\mathbf{U}_2^{k},\mathbf{U}_3^{k}\right)+\sum_{i=1}^{3}\delta_{\mathfrak{B}_i}(\mathbf{A}_i^{k})+\sum_{i=1}^{3}\alpha_i\operatorname{tr}\left((\mathbf{U}_i^{k})^H\mathbf{L}_i\mathbf{U}_i^{k}\right).
\end{aligned}
$$

回顾 F 在（13）中的定义，以及 Ψ 在（15）中的定义，上述不等式等价于

$$
\begin{aligned}
&F\left(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1}\right)+\frac{\bar{\rho}}{2}\left(\|\mathcal{G}^{k+1}-\mathcal{G}^{k}\|_F^2+\|\mathbf{A}_1^{k+1}-\mathbf{A}_1^{k}\|_F^2\right.\\
&\left.+\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F^2+\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F^2+\|\mathbf{U}_1^{k+1}-\mathbf{U}_1^{k}\|_F^2+\|\mathbf{U}_2^{k+1}-\mathbf{U}_2^{k}\|_F^2+\|\mathbf{U}_3^{k+1}-\mathbf{U}_3^{k}\|_F^2\right)\\
&\leq F\left(\mathcal{G}^{k},\mathbf{A}_1^{k},\mathbf{A}_2^{k},\mathbf{A}_3^{k},\mathbf{U}_1^{k},\mathbf{U}_2^{k},\mathbf{U}_3^{k}\right).
\end{aligned}
\tag{100}
$$

至此，证明完成。

附录 E. 引理2的证明

首先，我们给出式（19）中 PALM 算法的一阶最优性条件：

$$
\begin{cases}
0\in\partial(\beta\|\mathcal{G}^{k+1}\|_1+\delta_{\mathcal{D}}(\mathcal{G}^{k+1}))+\nabla_{\mathcal{G}}\Psi(\mathcal{G}^{k},\mathbf{A}_1^{k},\mathbf{A}_2^{k},\mathbf{A}_3^{k},\mathbf{U}_1^{k},\mathbf{U}_2^{k},\mathbf{U}_3^{k})+\rho_1(\mathcal{G}^{k+1}-\mathcal{G}^{k}),\\
0\in\partial\delta_{\mathfrak{B}_1}(\mathbf{A}_1^{k+1})+\nabla_{\mathbf{A}_1}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k},\mathbf{A}_2^{k},\mathbf{A}_3^{k},\mathbf{U}_1^{k},\mathbf{U}_2^{k},\mathbf{U}_3^{k})+\rho_2(\mathbf{A}_1^{k+1}-\mathbf{A}_1^{k}),\\
0\in\partial\delta_{\mathfrak{B}_2}(\mathbf{A}_2^{k+1})+\nabla_{\mathbf{A}_2}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k},\mathbf{A}_3^{k},\mathbf{U}_1^{k},\mathbf{U}_2^{k},\mathbf{U}_3^{k})+\rho_3(\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}),\\
0\in\partial\delta_{\mathfrak{B}_3}(\mathbf{A}_3^{k+1})+\nabla_{\mathbf{A}_3}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k},\mathbf{U}_1^{k},\mathbf{U}_2^{k},\mathbf{U}_3^{k})+\rho_4(\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}),\\
0=2\alpha_1\mathbf{L}_1\mathbf{U}_1^{k+1}+\nabla_{\mathbf{U}_1}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k},\mathbf{U}_2^{k},\mathbf{U}_3^{k})+\rho_5(\mathbf{U}_1^{k+1}-\mathbf{U}_1^{k}),\\
0=2\alpha_2\mathbf{L}_2\mathbf{U}_2^{k+1}+\nabla_{\mathbf{U}_2}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k},\mathbf{U}_3^{k})+\rho_6(\mathbf{U}_2^{k+1}-\mathbf{U}_2^{k}),\\
0=2\alpha_3\mathbf{L}_3\mathbf{U}_3^{k+1}+\nabla_{\mathbf{U}_3}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k})+\rho_7(\mathbf{U}_3^{k+1}-\mathbf{U}_3^{k}).
\end{cases}
\tag{101}
$$

记

$$
\begin{aligned}
N_1 &= \nabla_{\mathcal{G}}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1})\\
&\quad-\nabla_{\mathcal{G}}\Psi(\mathcal{G}^{k},\mathbf{A}_1^{k},\mathbf{A}_2^{k},\mathbf{A}_3^{k},\mathbf{U}_1^{k},\mathbf{U}_2^{k},\mathbf{U}_3^{k})-\rho_1(\mathcal{G}^{k+1}-\mathcal{G}^{k}),\\
N_2 &= \nabla_{\mathbf{A}_1}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1})\\
&\quad-\nabla_{\mathbf{A}_1}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k},\mathbf{A}_2^{k},\mathbf{A}_3^{k},\mathbf{U}_1^{k},\mathbf{U}_2^{k},\mathbf{U}_3^{k})-\rho_2(\mathbf{A}_1^{k+1}-\mathbf{A}_1^{k}),\\
N_3 &= \nabla_{\mathbf{A}_2}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1})\\
&\quad-\nabla_{\mathbf{A}_2}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k},\mathbf{A}_3^{k},\mathbf{U}_1^{k},\mathbf{U}_2^{k},\mathbf{U}_3^{k})-\rho_3(\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}),\\
N_4 &= \nabla_{\mathbf{A}_3}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1})\\
&\quad-\nabla_{\mathbf{A}_3}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k},\mathbf{U}_1^{k},\mathbf{U}_2^{k},\mathbf{U}_3^{k})-\rho_4(\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}),\\
N_5 &= \nabla_{\mathbf{U}_1}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1})\\
&\quad-\nabla_{\mathbf{U}_1}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k},\mathbf{U}_2^{k},\mathbf{U}_3^{k})-\rho_5(\mathbf{U}_1^{k+1}-\mathbf{U}_1^{k}),\\
N_6 &= \nabla_{\mathbf{U}_2}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1})\\
&\quad-\nabla_{\mathbf{U}_2}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k},\mathbf{U}_3^{k})-\rho_6(\mathbf{U}_2^{k+1}-\mathbf{U}_2^{k}),\\
N_7 &= \nabla_{\mathbf{U}_3}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1})\\
&\quad-\nabla_{\mathbf{U}_3}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k})-\rho_7(\mathbf{U}_3^{k+1}-\mathbf{U}_3^{k}).
\end{aligned}
\tag{102}
$$

于是我们得到

$$
\begin{aligned}
&N_1 \in \partial_{\mathcal{G}}F(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1}),\\
&N_2 \in \partial_{\mathbf{A}_1}F(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1}),\\
&N_3 \in \partial_{\mathbf{A}_2}F(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1}),\\
&N_4 \in \partial_{\mathbf{A}_3}F(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1}),\\
&N_5 \in \partial_{\mathbf{U}_1}F(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1}),\\
&N_6 \in \partial_{\mathbf{U}_2}F(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1}),\\
&N_7 \in \partial_{\mathbf{U}_3}F(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1}).
\end{aligned}
$$

根据 [1, 命题 2。1], 我们可以得到

$$
N^{k+1} := (N_1,N_2,N_3,N_4,N_5,N_6,N_7) \in \partial F(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1}).
$$

请注意

$$
\begin{aligned}
&\|N_1\|_F\\
=&\Big\|\nabla_{\mathcal{G}}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1})\\
&-\nabla_{\mathcal{G}}\Psi(\mathcal{G}^{k},\mathbf{A}_1^{k},\mathbf{A}_2^{k},\mathbf{A}_3^{k},\mathbf{U}_1^{k},\mathbf{U}_2^{k},\mathbf{U}_3^{k})-\rho_1(\mathcal{G}^{k+1}-\mathcal{G}^{k})\Big\|_F\\
\le&\Big\|\nabla_{\mathcal{G}}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1})\\
&-\nabla_{\mathcal{G}}\Psi(\mathcal{G}^{k},\mathbf{A}_1^{k},\mathbf{A}_2^{k},\mathbf{A}_3^{k},\mathbf{U}_1^{k},\mathbf{U}_2^{k},\mathbf{U}_3^{k})\Big\|_F+\rho_1\|\mathcal{G}^{k+1}-\mathcal{G}^{k}\|_F\\
=&\Big\|\nabla_{\mathcal{G}}Q(\llbracket\mathcal{G}^{k+1};\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1}\rrbracket)-\nabla_{\mathcal{G}}Q(\llbracket\mathcal{G}^{k};\mathbf{A}_1^{k},\mathbf{A}_2^{k},\mathbf{A}_3^{k}\rrbracket)\Big\|_F+\rho_1\|\mathcal{G}^{k+1}-\mathcal{G}^{k}\|_F\\
=&\Big\|\Big(\nabla_{\mathcal{G}}Q(\llbracket\mathcal{G}^{k+1};\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1}\rrbracket)\Big)_{(1)}-\Big(\nabla_{\mathcal{G}}Q(\llbracket\mathcal{G}^{k};\mathbf{A}_1^{k},\mathbf{A}_2^{k},\mathbf{A}_3^{k}\rrbracket)\Big)_{(1)}\Big\|_F+\rho_1\|\mathcal{G}^{k+1}-\mathcal{G}^{k}\|_F,
\end{aligned}
\tag{103}
$$

其中第二个等式由（18）得出。注意，（14）可以改写为

$$Q(\llbracket \mathcal{G}; \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3 \rrbracket) = \frac{1}{2T} \sum_{t=1}^{T} \left\| \mathbf{y}_t - \mathbf{A}_1 \mathcal{G}_{(1)} \left(\mathbf{A}_3 \otimes \mathbf{A}_2 \right)^H \mathbf{x}_t \right\|_2^2 . \tag{104}$$

结合式（104）与引理 9，经简单计算可得

$$\begin{aligned}
&\left\| \left(\nabla_{\mathcal{G}} Q(\llbracket \mathcal{G}^{k+1}; \mathbf{A}_1^{k+1}, \mathbf{A}_2^{k+1}, \mathbf{A}_3^{k+1} \rrbracket) \right)_{(1)} - \left(\nabla_{\mathcal{G}} Q(\llbracket \mathcal{G}^{k}; \mathbf{A}_1^{k}, \mathbf{A}_2^{k}, \mathbf{A}_3^{k} \rrbracket) \right)_{(1)} \right\|_F \\
=& \Bigg\| \frac{1}{T} \sum_{t=1}^{T} (\mathbf{A}_1^{k+1})^H \left(\mathbf{A}_1^{k+1} \mathcal{G}_{(1)}^{k+1} (\mathbf{A}_3^{k+1} \otimes \mathbf{A}_2^{k+1})^H \mathbf{x}_t - \mathbf{y}_t \right) \left((\mathbf{A}_3^{k+1} \otimes \mathbf{A}_2^{k+1})^H \mathbf{x}_t \right)^H \\
&\quad - \frac{1}{T} \sum_{t=1}^{T} (\mathbf{A}_1^{k})^H \left(\mathbf{A}_1^{k} \mathcal{G}_{(1)}^{k} (\mathbf{A}_3^{k} \otimes \mathbf{A}_2^{k})^H \mathbf{x}_t - \mathbf{y}_t \right) \left((\mathbf{A}_3^{k} \otimes \mathbf{A}_2^{k})^H \mathbf{x}_t \right)^H \Bigg\|_F \\
=& \Bigg\| \frac{1}{T} \sum_{t=1}^{T} \left(\mathcal{G}_{(1)}^{k+1} (\mathbf{A}_3^{k+1} \otimes \mathbf{A}_2^{k+1})^H \mathbf{x}_t - (\mathbf{A}_1^{k+1})^H \mathbf{y}_t \right) \left((\mathbf{A}_3^{k+1} \otimes \mathbf{A}_2^{k+1})^H \mathbf{x}_t \right)^H \\
&- \frac{1}{T} \sum_{t=1}^{T} \left(\mathcal{G}_{(1)}^{k} (\mathbf{A}_3^{k} \otimes \mathbf{A}_2^{k})^H \mathbf{x}_t - (\mathbf{A}_1^{k})^H \mathbf{y}_t \right) \left((\mathbf{A}_3^{k} \otimes \mathbf{A}_2^{k})^H \mathbf{x}_t \right)^H \Bigg\|_F \\
=& \Bigg\| \frac{1}{T} \sum_{t=1}^{T} \mathcal{G}_{(1)}^{k+1} (\mathbf{A}_3^{k+1} \otimes \mathbf{A}_2^{k+1})^H \mathbf{x}_t \mathbf{x}_t^H (\mathbf{A}_3^{k+1} \otimes \mathbf{A}_2^{k+1}) \\
&\quad - \frac{1}{T} \sum_{t=1}^{T} (\mathbf{A}_1^{k+1})^H \mathbf{y}_t \mathbf{x}_t^H (\mathbf{A}_3^{k+1} \otimes \mathbf{A}_2^{k+1}) + \frac{1}{T} \sum_{t=1}^{T} (\mathbf{A}_1^{k})^H \mathbf{y}_t \mathbf{x}_t^H (\mathbf{A}_3^{k} \otimes \mathbf{A}_2^{k}) \\
&\quad - \frac{1}{T} \sum_{t=1}^{T} \mathcal{G}_{(1)}^{k} (\mathbf{A}_3^{k} \otimes \mathbf{A}_2^{k})^H \mathbf{x}_t \mathbf{x}_t^H (\mathbf{A}_3^{k} \otimes \mathbf{A}_2^{k}) \Bigg\|_F \\
\le& \Bigg\| \frac{1}{T} \sum_{t=1}^{T} \mathcal{G}_{(1)}^{k+1} (\mathbf{A}_3^{k+1} \otimes \mathbf{A}_2^{k+1})^H \mathbf{x}_t \mathbf{x}_t^H (\mathbf{A}_3^{k+1} \otimes \mathbf{A}_2^{k+1}) \\
&- \frac{1}{T} \sum_{t=1}^{T} \mathcal{G}_{(1)}^{k} (\mathbf{A}_3^{k} \otimes \mathbf{A}_2^{k})^H \mathbf{x}_t \mathbf{x}_t^H (\mathbf{A}_3^{k} \otimes \mathbf{A}_2^{k}) \Bigg\|_F \\
&+ \Bigg\| \frac{1}{T} \sum_{t=1}^{T} (\mathbf{A}_1^{k})^H \mathbf{y}_t \mathbf{x}_t^H (\mathbf{A}_3^{k} \otimes \mathbf{A}_2^{k}) - \frac{1}{T} \sum_{t=1}^{T} (\mathbf{A}_1^{k+1})^H \mathbf{y}_t \mathbf{x}_t^H (\mathbf{A}_3^{k+1} \otimes \mathbf{A}_2^{k+1}) \Bigg\|_F \\
\le& \frac{1}{T} \sum_{t=1}^{T} \left\| \mathcal{G}_{(1)}^{k+1} (\mathbf{A}_3^{k+1} \otimes \mathbf{A}_2^{k+1})^H \mathbf{x}_t \mathbf{x}_t^H (\mathbf{A}_3^{k+1} \otimes \mathbf{A}_2^{k+1}) - \mathcal{G}_{(1)}^{k} (\mathbf{A}_3^{k} \otimes \mathbf{A}_2^{k})^H \mathbf{x}_t \mathbf{x}_t^H (\mathbf{A}_3^{k} \otimes \mathbf{A}_2^{k}) \right\|_F \\
&\quad + \frac{1}{T} \sum_{t=1}^{T} \left\| (\mathbf{A}_1^{k})^H \mathbf{y}_t \mathbf{x}_t^H (\mathbf{A}_3^{k} \otimes \mathbf{A}_2^{k}) - (\mathbf{A}_1^{k+1})^H \mathbf{y}_t \mathbf{x}_t^H (\mathbf{A}_3^{k+1} \otimes \mathbf{A}_2^{k+1}) \right\|_F ,
\end{aligned} \tag{105}$$

其中，第二个等式由 $(\mathbf{A}_1^{k+1})^H \mathbf{A}_1^{k+1} = \mathbf{I}_{r_i}$ 得出。接下来，我们首先分析第二个

项，即（105）右侧的项。注意，

$$
\begin{aligned}
&\left\|(\mathbf{A}_1^k)^H\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^k\otimes\mathbf{A}_2^k)-(\mathbf{A}_1^{k+1})^H\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})\right\|_F\\
\le&\left\|(\mathbf{A}_1^k)^H\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^k\otimes\mathbf{A}_2^k)-(\mathbf{A}_1^{k+1})^H\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^k\otimes\mathbf{A}_2^k)\right.\\
&+(\mathbf{A}_1^{k+1})^H\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^k\otimes\mathbf{A}_2^k)-(\mathbf{A}_1^{k+1})^H\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^k\otimes\mathbf{A}_2^{k+1})\\
&\left.+(\mathbf{A}_1^{k+1})^H\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^k\otimes\mathbf{A}_2^{k+1})-(\mathbf{A}_1^{k+1})^H\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})\right\|_F\\
\le&\left\|(\mathbf{A}_1^k)^H\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^k\otimes\mathbf{A}_2^k)-(\mathbf{A}_1^{k+1})^H\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^k\otimes\mathbf{A}_2^k)\right\|_F\\
&+\left\|(\mathbf{A}_1^{k+1})^H\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^k\otimes\mathbf{A}_2^k)-(\mathbf{A}_1^{k+1})^H\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^k\otimes\mathbf{A}_2^{k+1})\right\|_F\\
&+\left\|(\mathbf{A}_1^{k+1})^H\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^k\otimes\mathbf{A}_2^{k+1})-(\mathbf{A}_1^{k+1})^H\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})\right\|_F.
\end{aligned}
\tag{106}
$$

根据柯西-施瓦茨不等式，可得

$$
\begin{aligned}
&\left\|(\mathbf{A}_1^k)^H\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^k\otimes\mathbf{A}_2^k)-(\mathbf{A}_1^{k+1})^H\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^k\otimes\mathbf{A}_2^k)\right\|_F\\
=&\left\|(\mathbf{A}_1^k-\mathbf{A}_1^{k+1})^H\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^k\otimes\mathbf{A}_2^k)\right\|_F\\
\le&\left\|(\mathbf{A}_1^k-\mathbf{A}_1^{k+1})^H\right\|_F\left\|\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^k\otimes\mathbf{A}_2^k)\right\|_F\\
\le&\|\mathbf{A}_1^k-\mathbf{A}_1^{k+1}\|_F\|\mathbf{y}_t\mathbf{x}_t^H\|_F\|\mathbf{A}_3^k\otimes\mathbf{A}_2^k\|\\
\le&\|\mathbf{A}_1^k-\mathbf{A}_1^{k+1}\|_F\|\mathbf{y}_t\mathbf{x}_t^H\|_F,
\end{aligned}
\tag{107}
$$

其中第二个不等式由 [12, 引理 2.1] 得出，最后一个不等式则由命题 2得出。类似地可得

$$
\begin{aligned}
&\left\|(\mathbf{A}_1^{k+1})^H\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^k\otimes\mathbf{A}_2^k)-(\mathbf{A}_1^{k+1})^H\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^k\otimes\mathbf{A}_2^{k+1})\right\|_F\le\|\mathbf{y}_t\mathbf{x}_t^H\|_F\|\mathbf{A}_2^k-\mathbf{A}_2^{k+1}\|_F,\\
&\left\|(\mathbf{A}_1^{k+1})^H\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^k\otimes\mathbf{A}_2^{k+1})-(\mathbf{A}_1^{k+1})^H\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})\right\|_F\le\|\mathbf{y}_t\mathbf{x}_t^H\|_F\|\mathbf{A}_3^k-\mathbf{A}_3^{k+1}\|_F.
\end{aligned}
$$

结合（106）和（107），可推得

$$
\begin{aligned}
&\left\|(\mathbf{A}_1^k)^H\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^k\otimes\mathbf{A}_2^k)-(\mathbf{A}_1^{k+1})^H\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})\right\|_F\\
\le&\|\mathbf{y}_t\mathbf{x}_t^H\|_F\left(\|\mathbf{A}_1^k-\mathbf{A}_1^{k+1}\|_F+\|\mathbf{A}_2^k-\mathbf{A}_2^{k+1}\|_F+\|\mathbf{A}_3^k-\mathbf{A}_3^{k+1}\|_F\right).
\end{aligned}
\tag{108}
$$

将上述不等式对 $t=1,\dots,T$ 求和，可得

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^T\left\|(\mathbf{A}_1^k)^H\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^k\otimes\mathbf{A}_2^k)-(\mathbf{A}_1^{k+1})^H\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})\right\|_F\\
\le&\frac{1}{T}\sum_{t=1}^T\|\mathbf{y}_t\mathbf{x}_t^H\|_F\left(\|\mathbf{A}_1^k-\mathbf{A}_1^{k+1}\|_F+\|\mathbf{A}_2^k-\mathbf{A}_2^{k+1}\|_F+\|\mathbf{A}_3^k-\mathbf{A}_3^{k+1}\|_F\right)\\
=&\,c_2\left(\|\mathbf{A}_1^k-\mathbf{A}_1^{k+1}\|_F+\|\mathbf{A}_2^k-\mathbf{A}_2^{k+1}\|_F+\|\mathbf{A}_3^k-\mathbf{A}_3^{k+1}\|_F\right),
\end{aligned}
\tag{109}
$$

其中

$$
c_2:=\frac{1}{T}\sum_{t=1}^T\|\mathbf{y}_t\mathbf{x}_t^H\|_F.
\tag{110}
$$

现在我们分析式（105）右侧的第一项。不难看出，

$$
\begin{aligned}
&\left\|\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})-\mathcal{G}_{(1)}^{k}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})\right\|_F\\
=&\left\|\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})-\mathcal{G}_{(1)}^{k}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})\right.\\
&\left.+\mathcal{G}_{(1)}^{k}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})-\mathcal{G}_{(1)}^{k}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})\right\|_F\\
\le&\left\|\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})-\mathcal{G}_{(1)}^{k}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})\right\|_F\\
&+\left\|\mathcal{G}_{(1)}^{k}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})-\mathcal{G}_{(1)}^{k}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})\right\|_F.
\end{aligned}
\tag{111}
$$

对（111）应用柯西-施瓦茨不等式可得

$$
\begin{aligned}
&\left\|\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})-\mathcal{G}_{(1)}^{k}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})\right\|_F\\
=&\left\|(\mathcal{G}_{(1)}^{k+1}-\mathcal{G}_{(1)}^{k})(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})\right\|_F\\
\le&\|\mathcal{G}_{(1)}^{k+1}-\mathcal{G}_{(1)}^{k}\|_F\left\|(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})\right\|_F\\
=&\|\mathcal{G}^{k+1}-\mathcal{G}^{k}\|_F\left\|(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\right\|_2^2\\
\le&\|\mathcal{G}^{k+1}-\mathcal{G}^{k}\|_F\left(\|\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1}\|\,\|\mathbf{x}_t\|_2\right)^2\\
\le&\|\mathcal{G}^{k+1}-\mathcal{G}^{k}\|_F\,\|\mathbf{x}_t\|_2^2,
\end{aligned}
\tag{112}
$$

其中，第二个不等式由[27,例 5.66]得到。

另一方面，对于（111）中右侧不等式的第二项，我们得到

$$
\begin{aligned}
&\left\|\mathcal{G}_{(1)}^{k}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})-\mathcal{G}_{(1)}^{k}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})\right\|_F\\
\le&\|\mathcal{G}_{(1)}^{k}\|_F\left\|(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})-(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})\right\|_F\\
=&\|\mathcal{G}_{(1)}^{k}\|_F\left\|(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})-(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k})\right.\\
&+(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k})-(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})\\
&+(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})-(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})\\
&\left.+(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})-(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})\right\|_F\\
\le&\nu\left\|(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})-(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k})\right\|_F\\
&+\nu\left\|(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k})-(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})\right\|_F\\
&+\nu\left\|(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})-(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})\right\|_F\\
&+\nu\left\|(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})-(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})\right\|_F,
\end{aligned}
\tag{113}
$$

其中，第二个 d 不等式满足三角不等式，且 $\|\mathcal{G}_{(1)}^{k}\|_F\le\nu:=\sqrt{r_1r_2r_3}c_\circ$ 与

(91)，我们有

$$\begin{aligned}
&\left\|(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})-(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k})\right\|_F\\
=&\left\|(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H\left(\mathbf{A}_3^{k+1}\otimes(\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k})\right)\right\|_F\\
\le&\left\|\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1}\right\|\left\|\mathbf{x}_t\mathbf{x}_t^H\left(\mathbf{A}_3^{k+1}\otimes(\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k})\right)\right\|_F\\
\le&\left\|\mathbf{x}_t\mathbf{x}_t^H\left(\mathbf{A}_3^{k+1}\otimes(\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k})\right)\right\|_F\\
\le&\left\|\mathbf{x}_t\mathbf{x}_t^H\right\|_F\|\mathbf{A}_3^{k+1}\otimes(\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k})\|\\
=&\left\|\mathbf{x}_t\mathbf{x}_t^H\right\|_F\|\mathbf{A}_3^{k+1}\|\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|\\
\le&\left\|\mathbf{x}_t\mathbf{x}_t^H\right\|_F\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F.
\end{aligned}\tag{114}$$

同理，我们可以得到

$$\begin{aligned}
&\left\|(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k})-(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})\right\|_F\\
\le&\left\|\mathbf{x}_t\mathbf{x}_t^H\right\|_F\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F,\\
&\left\|(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})-(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})\right\|_F\\
\le&\left\|\mathbf{x}_t\mathbf{x}_t^H\right\|_F\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F,\\
&\left\|(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})-(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})\right\|_F\\
\le&\left\|\mathbf{x}_t\mathbf{x}_t^H\right\|_F\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F.
\end{aligned}\tag{115}$$

将（114）和（115）代入（113），即可得到

$$\begin{aligned}
&\left\|\mathcal{G}_{(1)}^k(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})-\mathcal{G}_{(1)}^k(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})\right\|_F\\
\le&2\nu\left\|\mathbf{x}_t\mathbf{x}_t^H\right\|_F\left(\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F+\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F\right).
\end{aligned}\tag{116}$$

将（116）和（112）代入（111），可得

$$\begin{aligned}
&\left\|\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})-\mathcal{G}_{(1)}^k(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})\right\|_F\\
\le&\|\mathcal{G}^{k+1}-\mathcal{G}^k\|_F\|\mathbf{x}_t\|_2^2+2\nu\left\|\mathbf{x}_t\mathbf{x}_t^H\right\|_F\left(\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F+\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F\right).
\end{aligned}\tag{117}$$

将上述不等式对 $t=1,\ldots,T$ 求和，可得

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\Big\|\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})\\
&\qquad-\mathcal{G}_{(1)}^k(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})\Big\|_F\\
\le&\frac{1}{T}\sum_{t=1}^{T}\|\mathcal{G}^{k+1}-\mathcal{G}^k\|_F\|\mathbf{x}_t\|_2^2+2\nu\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{x}_t\|_2^2\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F\\
&+2\nu\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{x}_t\|_2^2\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F\\
=&c_1\|\mathcal{G}^{k+1}-\mathcal{G}^k\|_F+2\nu c_1\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F+2\nu c_1\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F,
\end{aligned}\tag{118}$$

其中

$$c_1 := \frac{1}{T}\sum_{t=1}^{T}\|\mathbf{x}_t\|_2^2. \tag{119}$$

将（118）和（109）代入（105）即可得

$$\begin{aligned}
&\left\|\left(\nabla_{\mathcal{G}} Q([\![\mathcal{G}^{k+1};\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1}]\!])\right)_{(1)} - \left(\nabla_{\mathcal{G}} Q([\![\mathcal{G}^{k};\mathbf{A}_1^{k},\mathbf{A}_2^{k},\mathbf{A}_3^{k}]\!])\right)_{(1)}\right\|_F \\
\le\ & c_2\left(\|\mathbf{A}_1^k-\mathbf{A}_1^{k+1}\|_F+\|\mathbf{A}_2^k-\mathbf{A}_2^{k+1}\|_F+\|\mathbf{A}_3^k-\mathbf{A}_3^{k+1}\|_F\right)+c_1\|\mathcal{G}^{k+1}-\mathcal{G}^k\|_F \\
&+2\nu c_1\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^k\|_F+2\nu c_1\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^k\|_F.
\end{aligned}$$

结合（103）可得

$$\begin{aligned}
\|N_1\|_F &\le c_2\left(\|\mathbf{A}_1^k-\mathbf{A}_1^{k+1}\|_F+\|\mathbf{A}_2^k-\mathbf{A}_2^{k+1}\|_F+\|\mathbf{A}_3^k-\mathbf{A}_3^{k+1}\|_F\right)+c_1\|\mathcal{G}^{k+1}-\mathcal{G}^k\|_F \\
&\quad+2\nu c_1\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^k\|_F+2\nu c_1\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^k\|_F+\rho_1\|\mathcal{G}^{k+1}-\mathcal{G}^k\|_F \\
&= (\rho_1+c_1)\|\mathcal{G}^{k+1}-\mathcal{G}^k\|_F+c_2\|\mathbf{A}_1^{k+1}-\mathbf{A}_1^k\|_F+(c_2+2\nu c_1)\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^k\|_F \\
&\quad+(c_2+2\nu c_1)\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^k\|_F.
\end{aligned} \tag{120}$$

与上述关于 N_1 上界的分析类似，我们可以很容易地得到

$$\begin{aligned}
\|N_2\|_F &= \Big\|\nabla_{\mathbf{A}_1}\Psi\left(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1}\right) \\
&\quad-\nabla_{\mathbf{A}_1}\Psi\left(\mathcal{G}^{k+1},\mathbf{A}_1^{k},\mathbf{A}_2^{k},\mathbf{A}_3^{k},\mathbf{U}_1^{k},\mathbf{U}_2^{k},\mathbf{U}_3^{k}\right)-\rho_2(\mathbf{A}_1^{k+1}-\mathbf{A}_1^k)\Big\|_F \\
&\le \Big\|\nabla_{\mathbf{A}_1}\Psi\left(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1}\right) \\
&\quad-\nabla_{\mathbf{A}_1}\Psi\left(\mathcal{G}^{k+1},\mathbf{A}_1^{k},\mathbf{A}_2^{k},\mathbf{A}_3^{k},\mathbf{U}_1^{k},\mathbf{U}_2^{k},\mathbf{U}_3^{k}\right)\Big\|_F+\rho_2\|\mathbf{A}_1^{k+1}-\mathbf{A}_1^k\|_F \\
&= \Big\|\nabla_{\mathbf{A}_1}Q([\![\mathcal{G}^{k+1};\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1}]\!])-\gamma_1(\mathbf{U}_1^{k+1}-\mathbf{A}_1^{k+1}) \\
&\quad-\nabla_{\mathbf{A}_1}Q([\![\mathcal{G}^{k+1};\mathbf{A}_1^{k},\mathbf{A}_2^{k},\mathbf{A}_3^{k}]\!])+\gamma_1(\mathbf{U}_1^{k}-\mathbf{A}_1^{k})\Big\|_F+\rho_2\|\mathbf{A}_1^{k+1}-\mathbf{A}_1^k\|_F \\
&\le \Big\|\nabla_{\mathbf{A}_1}Q([\![\mathcal{G}^{k+1};\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1}]\!])-\nabla_{\mathbf{A}_1}Q([\![\mathcal{G}^{k+1};\mathbf{A}_1^{k},\mathbf{A}_2^{k},\mathbf{A}_3^{k}]\!])\Big\|_F \\
&\quad+\gamma_1\|\mathbf{U}_1^{k+1}-\mathbf{U}_1^k\|_F+(\gamma_1+\rho_2)\|\mathbf{A}_1^{k+1}-\mathbf{A}_1^k\|_F.
\end{aligned} \tag{121}$$

其中，第二个等式由式（18）得出，最后一个不等式由三角不等式得出。

结合式（17）和（81）可得

$$
\begin{aligned}
&\left\|\nabla_{\mathbf{A}_1} Q([\![\mathcal{G}^{k+1};\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1}]\!])-\nabla_{\mathbf{A}_1} Q([\![\mathcal{G}^{k+1};\mathbf{A}_1^{k},\mathbf{A}_2^{k},\mathbf{A}_3^{k}]\!])\right\|_F\\
=&\Big\|(\nabla Q([\![\mathcal{G}^{k+1};\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1}]\!]))_{(1)}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})(\mathcal{G}_{(1)}^{k+1})^H\\
&-(\nabla Q([\![\mathcal{G}^{k+1};\mathbf{A}_1^{k},\mathbf{A}_2^{k},\mathbf{A}_3^{k}]\!]))_{(1)}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})(\mathcal{G}_{(1)}^{k+1})^H\Big\|_F\\
=&\Big\|\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t-\mathbf{y}_t\right)\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})(\mathcal{G}_{(1)}^{k+1})^H\\
&-\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{A}_1^{k}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t-\mathbf{y}_t\right)\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})(\mathcal{G}_{(1)}^{k+1})^H\Big\|_F\\
\le&\frac{1}{T}\sum_{t=1}^{T}\Big\|\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})(\mathcal{G}_{(1)}^{k+1})^H\\
&-\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})(\mathcal{G}_{(1)}^{k+1})^H-\mathbf{A}_1^{k}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})(\mathcal{G}_{(1)}^{k+1})^H\\
&+\mathbf{y}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})(\mathcal{G}_{(1)}^{k+1})^H\Big\|_F\\
\le&\frac{1}{T}\sum_{t=1}^{T}\Big\|\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})(\mathcal{G}_{(1)}^{k+1})^H\\
&-\mathbf{A}_1^{k}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})(\mathcal{G}_{(1)}^{k+1})^H\Big\|_F\\
&+\frac{1}{T}\sum_{t=1}^{T}\left\|\mathbf{y}_t\mathbf{x}_t^H\left((\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})(\mathcal{G}_{(1)}^{k+1})^H-(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})(\mathcal{G}_{(1)}^{k+1})^H\right)\right\|_F,
\end{aligned}
\tag{122}
$$

其中，第二个等式由引理9得出。首先，我们考虑上述不等式右侧的第二项。注意到

$$
\begin{aligned}
&\left\|\mathbf{y}_t\mathbf{x}_t^H\left((\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})(\mathcal{G}_{(1)}^{k+1})^H-(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})(\mathcal{G}_{(1)}^{k+1})^H\right)\right\|_F\\
=&\left\|\mathbf{y}_t\mathbf{x}_t^H\left((\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})-(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k})+(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k})-(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})\right)\mathcal{G}_{(1)}^{k+1})^H\right\|_F\\
\le&\left\|\mathbf{y}_t\mathbf{x}_t^H\left((\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})-(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k})\right)\mathcal{G}_{(1)}^{k+1})^H\right\|_F\\
&+\left\|\mathbf{y}_t\mathbf{x}_t^H\left((\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k})-(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})\right)\mathcal{G}_{(1)}^{k+1})^H\right\|_F\\
=&\left\|\mathbf{y}_t\mathbf{x}_t^H\left(\mathbf{A}_3^{k+1}\otimes(\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k})\right)(\mathcal{G}_{(1)}^{k+1})^H\right\|_F+\left\|\mathbf{y}_t\mathbf{x}_t^H\left((\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k})\otimes\mathbf{A}_2^{k}\right)(\mathcal{G}_{(1)}^{k+1})^H\right\|_F.
\end{aligned}
\tag{123}
$$

采用式（91）中的相同推导，可得

$$
\begin{aligned}
\left\|\mathbf{y}_t\mathbf{x}_t^H\left(\mathbf{A}_3^{k+1}\otimes(\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k})\right)(\mathcal{G}_{(1)}^{k+1})^H\right\|_F&\le\nu\left\|\mathbf{y}_t\mathbf{x}_t^H\right\|_F\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F,\\
\left\|\mathbf{y}_t\mathbf{x}_t^H\left((\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k})\otimes\mathbf{A}_2^{k}\right)(\mathcal{G}_{(1)}^{k+1})^H\right\|_F&\le\nu\left\|\mathbf{y}_t\mathbf{x}_t^H\right\|_F\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F.
\end{aligned}
\tag{124}
$$

将（124）代入（123）可得

$$
\begin{aligned}
&\left\|\mathbf{y}_t\mathbf{x}_t^H\left((\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})(\mathcal{G}_{(1)}^{k+1})^H-(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})(\mathcal{G}_{(1)}^{k+1})^H\right)\right\|_F\\
\le&\nu\left\|\mathbf{y}_t\mathbf{x}_t^H\right\|_F\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F+\nu\left\|\mathbf{y}_t\mathbf{x}_t^H\right\|_F\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F,
\end{aligned}
$$

这表明

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\left\|\mathbf{y}_t\mathbf{x}_t^H\left((\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})(\mathcal{G}_{(1)}^{k+1})^H-(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})(\mathcal{G}_{(1)}^{k+1})^H\right)\right\|_F\\
&\le\frac{1}{T}\sum_{t=1}^{T}\nu\left\|\mathbf{y}_t\mathbf{x}_t^H\right\|_F\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F+\frac{1}{T}\sum_{t=1}^{T}\nu\left\|\mathbf{y}_t\mathbf{x}_t^H\right\|_F\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F\\
&=c_2\nu\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F+c_2\nu\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F,
\end{aligned}
\tag{125}
$$

其中 c_2 定义见（110）。

For 对于（122）右端的第一项，采用与（111）类似的技术可得 s

$$
\begin{aligned}
&\left\|\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})(\mathcal{G}_{(1)}^{k+1})^H\right.\\
&\quad\left.-\mathbf{A}_1^{k}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})(\mathcal{G}_{(1)}^{k+1})^H\right\|_F\\
&\le\|\mathbf{A}_1^{k+1}-\mathbf{A}_1^{k}\|_F\nu^2\left\|\mathbf{x}_t\mathbf{x}_t^H\right\|_F+\nu^2\left\|\mathbf{x}_t\mathbf{x}_t^H\right\|_F\left(2\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F+2\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F\right).
\end{aligned}
$$

将上述不等式对 $t=1$、...、T 求和，可容易得到

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\left\|\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})(\mathcal{G}_{(1)}^{k+1})^H\right.\\
&\quad\left.-\mathbf{A}_1^{k}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t\mathbf{x}_t^H(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})(\mathcal{G}_{(1)}^{k+1})^H\right\|_F\\
&\le\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{A}_1^{k+1}-\mathbf{A}_1^{k}\|_F\nu^2\left\|\mathbf{x}_t\mathbf{x}_t^H\right\|_F+\frac{1}{T}\sum_{t=1}^{T}2\nu^2\left\|\mathbf{x}_t\mathbf{x}_t^H\right\|_F\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F\\
&\quad+\frac{1}{T}\sum_{t=1}^{T}2\nu^2\left\|\mathbf{x}_t\mathbf{x}_t^H\right\|_F\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F\\
&=\nu^2c_1\|\mathbf{A}_1^{k+1}-\mathbf{A}_1^{k}\|_F+2\nu^2c_1\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F+2\nu^2c_1\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F,
\end{aligned}
\tag{126}
$$

其中 c_1 定义见（119）。将（126）和（125）代入（122），我们得出

$$
\begin{aligned}
&\left\|\nabla_{\mathbf{A}_1}Q([\![\mathcal{G}^{k+1};\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1}]\!])-\nabla_{\mathbf{A}_1}Q([\![\mathcal{G}^{k+1};\mathbf{A}_1^{k},\mathbf{A}_2^{k},\mathbf{A}_3^{k}]\!])\right\|_F\\
&\le\nu c_2\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F+\nu c_2\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F+\nu^2c_1\|\mathbf{A}_1^{k+1}-\mathbf{A}_1^{k}\|_F\\
&\quad+2\nu^2c_1\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F+2\nu^2c_1\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F.
\end{aligned}
$$

结合上述不等式和（121），可得

$$
\begin{aligned}
\|N_2\|_F=&\left\|\nabla_{\mathbf{A}_1}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1})\right.\\
&\left.-\nabla_{\mathbf{A}_1}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k},\mathbf{A}_2^{k},\mathbf{A}_3^{k},\mathbf{U}_1^{k},\mathbf{U}_2^{k},\mathbf{U}_3^{k})-\rho_2(\mathbf{A}_1^{k+1}-\mathbf{A}_1^{k})\right\|_F\\
&\le\nu c_2\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F+\nu c_2\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F+\nu^2c_1\|\mathbf{A}_1^{k+1}-\mathbf{A}_1^{k}\|_F\\
&\quad+2\nu^2c_1\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F+2\nu^2c_1\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F+\gamma_1\|\mathbf{U}_1^{k+1}-\mathbf{U}_1^{k}\|_F\\
&\quad+(\rho_2+\gamma_1)\|\mathbf{A}_1^{k+1}-\mathbf{A}_1^{k}\|_F\\
&=(\nu^2c_1+\rho_2+\gamma_1)\|\mathbf{A}_1^{k+1}-\mathbf{A}_1^{k}\|_F+(\nu c_2+2\nu^2c_1)\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F\\
&\quad+(\nu c_2+2\nu^2c_1)\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F+\gamma_1\|\mathbf{U}_1^{k+1}-\mathbf{U}_1^{k}\|_F.
\end{aligned}
\tag{127}
$$

与对 N_2 的分析类似，我们有

$$\begin{aligned}\|N_3\|_F = &\Big\|\nabla_{\mathbf{A}_2}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1})\\ &-\nabla_{\mathbf{A}_2}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k},\mathbf{A}_3^{k},\mathbf{U}_1^{k},\mathbf{U}_2^{k},\mathbf{U}_3^{k})-\rho_3(\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k})\Big\|_F\\ \le &\Big\|\nabla_{\mathbf{A}_2}Q([\![\mathcal{G}^{k+1};\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1}]\!])-\nabla_{\mathbf{A}_2}Q([\![\mathcal{G}^{k+1};\mathbf{A}_1^{k+1},\mathbf{A}_2^{k},\mathbf{A}_3^{k}]\!])\Big\|_F\\ &+\gamma_2\|\mathbf{U}_2^{k+1}-\mathbf{U}_2^{k}\|_F+(\gamma_2+\rho_3)\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F.\end{aligned}\tag{128}$$

注意到（17）与（16）可得

$$\begin{aligned}&\Big\|\nabla_{\mathbf{A}_2}Q([\![\mathcal{G}^{k+1};\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1}]\!])-\nabla_{\mathbf{A}_2}Q([\![\mathcal{G}^{k+1};\mathbf{A}_1^{k+1},\mathbf{A}_2^{k},\mathbf{A}_3^{k}]\!])\Big\|_F\\ =&\Big\|\Big(\frac{1}{T}\sum_{t=1}^{T}(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t-\mathbf{y}_t)\circ\mathbf{X}_t\Big)_{(2)}(\mathbf{A}_1^{k+1}\otimes\mathbf{A}_3^{k+1})(\mathcal{G}_{(2)}^{k+1})^H\\ &-\Big(\frac{1}{T}\sum_{t=1}^{T}(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t-\mathbf{y}_t)\circ\mathbf{X}_t\Big)_{(2)}(\mathbf{A}_1^{k+1}\otimes\mathbf{A}_3^{k})(\mathcal{G}_{(2)}^{k+1})^H\Big\|_F\\ =&\Big\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t\otimes\Big(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t-\mathbf{y}_t\Big)^H(\mathbf{A}_1^{k+1}\otimes\mathbf{A}_3^{k+1})(\mathcal{G}_{(2)}^{k+1})^H\\ &-\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t\otimes\Big(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t-\mathbf{y}_t\Big)^H(\mathbf{A}_1^{k+1}\otimes\mathbf{A}_3^{k})(\mathcal{G}_{(2)}^{k+1})^H\Big\|_F\\ \le&\frac{1}{T}\sum_{t=1}^{T}\Big\|\mathbf{X}_t\otimes\Big(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t-\mathbf{y}_t\Big)^H(\mathbf{A}_1^{k+1}\otimes\mathbf{A}_3^{k+1})(\mathcal{G}_{(2)}^{k+1})^H\\ &-\mathbf{X}_t\otimes\Big(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t-\mathbf{y}_t\Big)^H(\mathbf{A}_1^{k+1}\otimes\mathbf{A}_3^{k})(\mathcal{G}_{(2)}^{k+1})^H\Big\|_F\\ =&\frac{1}{T}\sum_{t=1}^{T}\Big\|\mathbf{X}_t\otimes\Big(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t-\mathbf{y}_t\Big)^H\mathbf{A}_1^{k+1}\otimes\mathbf{A}_3^{k+1})(\mathcal{G}_{(2)}^{k+1})^H\\ &-\mathbf{X}_t\otimes\Big(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t-\mathbf{y}_t\Big)^H(\mathbf{A}_1^{k+1}\otimes\mathbf{A}_3^{k+1})(\mathcal{G}_{(2)}^{k+1})^H\\ &+\mathbf{X}_t\otimes\Big(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t-\mathbf{y}_t\Big)^H(\mathbf{A}_1^{k+1}\otimes\mathbf{A}_3^{k+1})(\mathcal{G}_{(2)}^{k+1})^H\\ &-\mathbf{X}_t\otimes\Big(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t-\mathbf{y}_t\Big)^H(\mathbf{A}_1^{k+1}\otimes\mathbf{A}_3^{k+1})(\mathcal{G}_{(2)}^{k+1})^H\\ &+\mathbf{X}_t\otimes\Big(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t-\mathbf{y}_t\Big)^H(\mathbf{A}_1^{k+1}\otimes\mathbf{A}_3^{k+1})(\mathcal{G}_{(2)}^{k+1})^H\\ &-\mathbf{X}_t\otimes\Big(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t-\mathbf{y}_t\Big)^H(\mathbf{A}_1^{k+1}\otimes\mathbf{A}_3^{k})(\mathcal{G}_{(2)}^{k+1})^H\Big\|_F,\end{aligned}\tag{129}$$

其中第二个等式由命题 3 得出。因此，我们得到

$$
\begin{aligned}
&\left\|\nabla_{\mathbf{A}_2} Q(\llbracket \mathcal{G}^{k+1}; \mathbf{A}_1^{k+1}, \mathbf{A}_2^{k+1}, \mathbf{A}_3^{k+1} \rrbracket) - \nabla_{\mathbf{A}_2} Q(\llbracket \mathcal{G}^{k+1}; \mathbf{A}_1^{k+1}, \mathbf{A}_2^{k}, \mathbf{A}_3^{k} \rrbracket)\right\|_F \\
\leq & \frac{1}{T}\sum_{t=1}^{T} \Big\| \mathbf{X}_t \otimes \left(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t - \mathbf{y}_t\right)^H \mathbf{A}_1^{k+1}\otimes\mathbf{A}_3^{k+1})(\mathcal{G}_{(2)}^{k+1})^H \\
& - \mathbf{X}_t \otimes \left(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t - \mathbf{y}_t\right)^H (\mathbf{A}_1^{k+1}\otimes\mathbf{A}_3^{k+1})(\mathcal{G}_{(2)}^{k+1})^H \Big\|_F \\
& + \frac{1}{T}\sum_{t=1}^{T} \Big\| \mathbf{X}_t \otimes \left(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t - \mathbf{y}_t\right)^H (\mathbf{A}_1^{k+1}\otimes\mathbf{A}_3^{k+1})(\mathcal{G}_{(2)}^{k+1})^H \\
& - \mathbf{X}_t \otimes \left(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t - \mathbf{y}_t\right)^H (\mathbf{A}_1^{k+1}\otimes\mathbf{A}_3^{k+1})(\mathcal{G}_{(2)}^{k+1})^H \Big\|_F \\
& + \frac{1}{T}\sum_{t=1}^{T} \Big\| \mathbf{X}_t \otimes \left(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t - \mathbf{y}_t\right)^H (\mathbf{A}_1^{k+1}\otimes\mathbf{A}_3^{k+1})(\mathcal{G}_{(2)}^{k+1})^H \\
& - \mathbf{X}_t \otimes \left(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t - \mathbf{y}_t\right)^H (\mathbf{A}_1^{k+1}\otimes\mathbf{A}_3^{k})(\mathcal{G}_{(2)}^{k+1})^H \Big\|_F .
\end{aligned} \tag{130}
$$

现在我们关注上述不等式右侧的第一项。

$$
\begin{aligned}
&\Big\| \mathbf{X}_t \otimes \left(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t - \mathbf{y}_t\right)^H (\mathbf{A}_1^{k+1}\otimes\mathbf{A}_3^{k+1})(\mathcal{G}_{(2)}^{k+1})^H \\
&\quad - \mathbf{X}_t \otimes \left(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t - \mathbf{y}_t\right)^H (\mathbf{A}_1^{k+1}\otimes\mathbf{A}_3^{k+1})(\mathcal{G}_{(2)}^{k+1})^H \Big\|_F \\
\leq & \Big\| \mathbf{X}_t \otimes \left(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t - \mathbf{y}_t\right)^H \\
&\quad - \mathbf{X}_t \otimes \left(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t - \mathbf{y}_t\right)^H \Big\|_F \left\|(\mathbf{A}_1^{k+1}\otimes\mathbf{A}_3^{k+1})(\mathcal{G}_{(2)}^{k+1})^H\right\|_F \\
\leq & \Big\| \mathbf{X}_t \otimes \left(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t - \mathbf{y}_t\right)^H \\
&\quad - \mathbf{X}_t \otimes \left(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k})^H\mathbf{x}_t - \mathbf{y}_t\right)^H \Big\|_F \nu \\
= & \nu \left\| \mathbf{X}_t \otimes \left(\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes(\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}))(\mathcal{G}_{(1)}^{k+1})^H(\mathbf{A}_1^{k+1})^H\right)\right\|_F ,
\end{aligned}
$$

其中第二个不等式由（85）得出。进一步利用（91）可得

$$
\begin{aligned}
& \nu \left\| \mathbf{X}_t \otimes \left(\mathbf{x}_t^H \left(\mathbf{A}_3^{k+1}\otimes(\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k})\right)(\mathcal{G}_{(1)}^{k+1})^H(\mathbf{A}_1^{k+1})^H\right)\right\|_F \\
= & \nu \left\|\mathbf{X}_t\right\|_F \left\| \left(\mathbf{x}_t^H(\mathbf{A}_3^{k+1}\otimes(\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}))(\mathcal{G}_{(1)}^{k+1})^H(\mathbf{A}_1^{k+1})^H\right)\right\|_F \\
\leq & \nu \left\|\mathbf{X}_t\right\|_F \left\|\mathbf{x}_t\right\|_2 \nu \|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F \\
= & \nu^2 \left\|\mathbf{x}_t\right\|_2^2 \|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F ,
\end{aligned}
$$

第一个等式由命题1得出，第二个等式则由以下事实得出：

$\|\mathbf{x}_t\|_2 = \|\mathbf{X}_t\|_F$。将上述不等式从 $t+1$ 到 T 求和，得到

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\nu\left\|\mathbf{X}_t\otimes\left(\mathbf{x}_t^H\left(\mathbf{A}_3^{k+1}\otimes(\mathbf{A}_2^{k+1}-\mathbf{A}_2^k)\right)(\mathcal{G}_{(1)}^{k+1})^H(\mathbf{A}_1^{k+1})^H\right)\right\|_F \\
\le\ &\frac{1}{T}\sum_{t=1}^{T}\nu^2\|\mathbf{x}_t\|_2^2\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^k\|_F = \nu^2 c_1\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^k\|_F,
\end{aligned} \tag{131}$$

其中 c_1 在（119）中定义。同理，可得

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\Bigg\|\mathbf{X}_t\otimes\left(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^k)^H\mathbf{x}_t-\mathbf{y}_t\right)^H(\mathbf{A}_1^{k+1}\otimes\mathbf{A}_3^{k+1})(\mathcal{G}_{(2)}^{k+1})^H \\
&-\mathbf{X}_t\otimes\left(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^k)^H\mathbf{x}_t-\mathbf{y}_t\right)^H(\mathbf{A}_1^{k+1}\otimes\mathbf{A}_3^{k+1})(\mathcal{G}_{(2)}^{k+1})^H\Bigg\|_F \le \nu^2 c_1\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^k\|_F, \\
&\frac{1}{T}\sum_{t=1}^{T}\Bigg\|\mathbf{X}_t\otimes\left(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^k)^H\mathbf{x}_t-\mathbf{y}_t\right)^H(\mathbf{A}_1^{k+1}\otimes\mathbf{A}_3^{k+1})(\mathcal{G}_{(2)}^{k+1})^H \\
&-\mathbf{X}_t\otimes\left(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^k)^H\mathbf{x}_t-\mathbf{y}_t\right)^H(\mathbf{A}_1^{k+1}\otimes\mathbf{A}_3^{k})(\mathcal{G}_{(2)}^{k+1})^H\Bigg\|_F \le \nu^2 c_1\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^k\|_F.
\end{aligned} \tag{132}$$

将（132）和（131）代入（130），可得

$$\begin{aligned}
&\left\|\nabla_{\mathbf{A}_2}Q([\![\mathcal{G}^{k+1};\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1}]\!])-\nabla_{\mathbf{A}_2}Q([\![\mathcal{G}^{k+1};\mathbf{A}_1^{k+1},\mathbf{A}_2^{k},\mathbf{A}_3^{k}]\!])\right\|_F \\
\le\ &\nu^2 c_1\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^k\|_F+2\nu^2 c_1\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^k\|_F.
\end{aligned} \tag{133}$$

结合（133）和（128），可得

$$\begin{aligned}
\|N_3\|_F &\le \nu^2 c_1\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^k\|_F+2\nu^2 c_1\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^k\|_F+\gamma_2\|\mathbf{U}_2^{k+1}-\mathbf{U}_2^k\|_F \\
&\quad+(\gamma_2+\rho_3)\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^k\|_F \\
&= (\nu^2 c_1+\gamma_2+\rho_3)\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^k\|_F+2\nu^2 c_1\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^k\|_F+\gamma_2\|\mathbf{U}_2^{k+1}-\mathbf{U}_2^k\|_F.
\end{aligned} \tag{134}$$

关于 N_4 在（102）中定义，我们有

$$\begin{aligned}
\|N_4\|_F &= \Big\|\nabla_{\mathbf{A}_3}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1}) \\
&\quad-\nabla_{\mathbf{A}_3}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k},\mathbf{U}_1^{k},\mathbf{U}_2^{k},\mathbf{U}_3^{k})-\rho_4(\mathbf{A}_3^{k+1}-\mathbf{A}_3^k)\Big\|_F \\
&\le \left\|\nabla_{\mathbf{A}_3}Q([\![\mathcal{G}^{k+1};\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1}]\!])-\nabla_{\mathbf{A}_3}Q([\![\mathcal{G}^{k+1};\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k}]\!])\right\|_F \\
&\quad+\gamma_3\|\mathbf{U}_3^{k+1}-\mathbf{U}_3^k\|_F+(\gamma_3+\rho_4)\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^k\|_F.
\end{aligned} \tag{135}$$

根据（17）、（16）和命题 3，可得

$$
\begin{aligned}
&\left\|\nabla_{\mathbf{A}_3} Q([\![\mathcal{G}^{k+1};\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1}]\!])-\nabla_{\mathbf{A}_3} Q([\![\mathcal{G}^{k+1};\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k}]\!])\right\|_F\\
=&\left\|\left(\frac{1}{T}\sum_{t=1}^{T}(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t-\mathbf{y}_t)\circ\mathbf{X}_t\right)_{(3)}(\mathbf{A}_2^{k+1}\otimes\mathbf{A}_1^{k+1})(\mathcal{G}_{(3)}^{k+1})^H\right.\\
&\left.-\left(\frac{1}{T}\sum_{t=1}^{T}(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t-\mathbf{y}_t)\circ\mathbf{X}_t\right)_{(3)}(\mathbf{A}_2^{k+1}\otimes\mathbf{A}_1^{k+1})(\mathcal{G}_{(3)}^{k+1})^H\right\|_F\\
=&\left\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t^H\otimes\left(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t-\mathbf{y}_t\right)^H(\mathbf{A}_2^{k+1}\otimes\mathbf{A}_1^{k+1})(\mathcal{G}_{(3)}^{k+1})^H\right.\\
&\left.-\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t^H\otimes\left(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t-\mathbf{y}_t\right)^H(\mathbf{A}_2^{k+1}\otimes\mathbf{A}_1^{k+1})(\mathcal{G}_{(3)}^{k+1})^H\right\|_F\\
\le&\frac{1}{T}\sum_{t=1}^{T}\left\|\mathbf{X}_t^H\otimes\left(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t-\mathbf{y}_t\right)^H(\mathbf{A}_2^{k+1}\otimes\mathbf{A}_1^{k+1})(\mathcal{G}_{(3)}^{k+1})^H\right.\\
&\left.-\mathbf{X}_t^H\otimes\left(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t-\mathbf{y}_t\right)^H(\mathbf{A}_2^{k+1}\otimes\mathbf{A}_1^{k+1})(\mathcal{G}_{(3)}^{k+1})^H\right\|_F\\
\le&\frac{1}{T}\sum_{t=1}^{T}\left\|\mathbf{X}_t^H\otimes\left(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t-\mathbf{y}_t\right)^H\right.\\
&\left.-\mathbf{X}_t^H\otimes\left(\mathbf{A}_1^{k}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t-\mathbf{y}_t\right)^H\right\|_F\left\|(\mathbf{A}_2^{k+1}\otimes\mathbf{A}_1^{k+1})(\mathcal{G}_{(3)}^{k+1})^H\right\|_F\\
\le&\frac{1}{T}\sum_{t=1}^{T}\left\|\mathbf{X}_t^H\otimes\left(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t-\mathbf{y}_t\right)^H\right.\\
&\left.-\mathbf{X}_t^H\otimes\left(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t-\mathbf{y}_t\right)^H\right\|_F\nu,
\end{aligned}
\tag{136}
$$

其中最后一个等式由（85）得出。采用与（91）中类似的技术可得

$$
\begin{aligned}
&\left\|\mathbf{X}_t^H\otimes(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k+1}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t-\mathbf{y}_t)^H-\mathbf{X}_t^H\otimes(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}(\mathbf{A}_3^{k}\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t-\mathbf{y}_t)^H\right\|_F\\
=&\left\|\mathbf{X}_t^H\otimes(\mathbf{A}_1^{k+1}\mathcal{G}_{(1)}^{k+1}((\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k})\otimes\mathbf{A}_2^{k+1})^H\mathbf{x}_t)^H\right\|_F\\
\le&\|\mathbf{x}_t\|_2^2\,\nu\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F.
\end{aligned}
$$

结合上述不等式与（136）可得

$$
\begin{aligned}
&\left\|\nabla_{\mathbf{A}_3} Q([\![\mathcal{G}^{k+1};\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1}]\!])-\nabla_{\mathbf{A}_3} Q([\![\mathcal{G}^{k+1};\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k}]\!])\right\|_F\\
\le&\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{x}_t\|_2^2\,\nu^2\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F=\nu^2c_1\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F.
\end{aligned}
\tag{137}
$$

将（137）代入（135）可得

$$
\begin{aligned}
\|N_4\|_F&\le\nu^2c_1\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F+\gamma_3\|\mathbf{U}_3^{k+1}-\mathbf{U}_3^{k}\|_F+(\gamma_3+\rho_4)\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F\\
&=(\nu^2c_1+\gamma_3+\rho_4)\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F+\gamma_3\|\mathbf{U}_3^{k+1}-\mathbf{U}_3^{k}\|_F.
\end{aligned}
\tag{138}
$$

对于 N_5 在 (102) 中，由 (18) 可得

$$\begin{aligned}\|N_5\|_F =& \Big\|\nabla_{\mathbf{U}_1}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1})\\ &-\nabla_{\mathbf{U}_1}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k},\mathbf{U}_2^{k},\mathbf{U}_3^{k})-\rho_5(\mathbf{U}_1^{k+1}-\mathbf{U}_1^{k})\Big\|_F\\ =& \Big\|\gamma_1(\mathbf{U}_1^{k+1}-\mathbf{A}_1^{k+1})-\gamma_1(\mathbf{U}_1^{k}-\mathbf{A}_1^{k+1})-\rho_5(\mathbf{U}_1^{k+1}-\mathbf{U}_1^{k})\Big\|_F\\ =& \Big\|\gamma_1(\mathbf{U}_1^{k+1}-\mathbf{U}_1^{k})-\rho_5(\mathbf{U}_1^{k+1}-\mathbf{U}_1^{k})\Big\|_F\\ \le& \|\gamma_1(\mathbf{U}_1^{k+1}-\mathbf{U}_1^{k})\|_F+\|\rho_5(\mathbf{U}_1^{k+1}-\mathbf{U}_1^{k})\|_F\\ =&(\gamma_1+\rho_5)\|\mathbf{U}_1^{k+1}-\mathbf{U}_1^{k}\|_F.\end{aligned} \tag{139}$$

同理，我们可以很容易地得到

$$\begin{aligned}\|N_6\|_F =& \Big\|\nabla_{\mathbf{U}_2}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1})\\ &-\nabla_{\mathbf{U}_2}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k},\mathbf{U}_3^{k})-\rho_6(\mathbf{U}_2^{k+1}-\mathbf{U}_2^{k})\Big\|_F\\ \le&(\gamma_2+\rho_6)\|\mathbf{U}_2^{k+1}-\mathbf{U}_2^{k}\|_F,\\ \|N_7\|_F =& \Big\|\nabla_{\mathbf{U}_3}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k+1})\\ &-\nabla_{\mathbf{U}_3}\Psi(\mathcal{G}^{k+1},\mathbf{A}_1^{k+1},\mathbf{A}_2^{k+1},\mathbf{A}_3^{k+1},\mathbf{U}_1^{k+1},\mathbf{U}_2^{k+1},\mathbf{U}_3^{k})-\rho_7(\mathbf{U}_3^{k+1}-\mathbf{U}_3^{k})\Big\|_F\\ \le&(\gamma_3+\rho_7)\|\mathbf{U}_3^{k+1}-\mathbf{U}_3^{k}\|_F.\end{aligned} \tag{140}$$

将 (120) 与 (127)、(134)、(138)、(139) 及 (140) 结合起来，可得

$$\begin{aligned}&\|N^{k+1}\|_F=\|\,(N_1,N_2,N_3,N_4,N_5,N_6,N_7)\,\|_F\\ \le&\|N_1\|_F+\|N_2\|_F+\|N_3\|_F+\|N_4\|_F+\|N_5\|_F+\|N_6\|_F+\|N_7\|_F\\ \le&(\rho_1+c_1)\|\mathcal{G}^{k+1}-\mathcal{G}^{k}\|_F+c_2\|\mathbf{A}_1^{k}-\mathbf{A}_1^{k+1}\|_F+(c_2+2\nu c_1)\|\mathbf{A}_2^{k}-\mathbf{A}_2^{k+1}\|_F\\ &+(c_2+2\nu c_1)\|\mathbf{A}_3^{k}-\mathbf{A}_3^{k+1}\|_F+(\nu^2c_1+\rho_2+\gamma_1)\|\mathbf{A}_1^{k+1}-\mathbf{A}_1^{k}\|_F\\ &+(\nu c_2+2\nu^2c_1)\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F+(\nu c_2+2\nu^2c_1)\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F+\gamma_1\|\mathbf{U}_1^{k+1}-\mathbf{U}_1^{k}\|_F\\ &+(\nu^2c_1+\gamma_2+\rho_3)\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F+2\nu^2c_1\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F+\gamma_2\|\mathbf{U}_2^{k+1}-\mathbf{U}_2^{k}\|_F\\ &+(\nu^2c_1+\gamma_3+\rho_4)\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F+\gamma_3\|\mathbf{U}_3^{k+1}-\mathbf{U}_3^{k}\|_F\\ &+(\gamma_1+\rho_5)\|\mathbf{U}_1^{k+1}-\mathbf{U}_1^{k}\|_F+(\gamma_2+\rho_6)\|\mathbf{U}_2^{k+1}-\mathbf{U}_2^{k}\|_F+(\gamma_3+\rho_7)\|\mathbf{U}_3^{k+1}-\mathbf{U}_3^{k}\|_F\\ =&(\rho_1+c_1)\|\mathcal{G}^{k+1}-\mathcal{G}^{k}\|_F+(c_2+\nu^2c_1+\rho_2+\gamma_1)\|\mathbf{A}_1^{k}-\mathbf{A}_1^{k+1}\|_F\\ &+(c_1+2\nu c_1+\nu c_2+3\nu^2c_1+\gamma_2+\rho_3)\|\mathbf{A}_2^{k}-\mathbf{A}_2^{k+1}\|_F\\ &+(c_2+2\nu c_1+\nu c_2+5\nu^2c_1+\gamma_3+\rho_4)\|\mathbf{A}_3^{k}-\mathbf{A}_3^{k+1}\|_F\\ &+(2\gamma_1+\rho_5)\|\mathbf{U}_1^{k+1}-\mathbf{U}_1^{k}\|_F+(2\gamma_2+\rho_6)\|\mathbf{U}_2^{k+1}-\mathbf{U}_2^{k}\|_F+(2\gamma_3+\rho_7)\|\mathbf{U}_3^{k+1}-\mathbf{U}_3^{k}\|_F\\ \le&\bar{\delta}\Big(\|\mathcal{G}^{k+1}-\mathcal{G}^{k}\|_F+\|\mathbf{A}_1^{k+1}-\mathbf{A}_1^{k}\|_F+\|\mathbf{A}_2^{k+1}-\mathbf{A}_2^{k}\|_F\\ &+\|\mathbf{A}_3^{k+1}-\mathbf{A}_3^{k}\|_F+\|\mathbf{U}_1^{k+1}-\mathbf{U}_1^{k}\|_F+\|\mathbf{U}_2^{k+1}-\mathbf{U}_2^{k}\|_F+\|\mathbf{U}_3^{k+1}-\mathbf{U}_3^{k}\|_F\Big),\end{aligned}$$

其中 $\bar{\delta}:=\max\{\rho_1+c_1,c_2+\nu^2c_1+\rho_2+\gamma_1,c_1+2\nu c_1+\nu c_2+3\nu^2c_1+\gamma_2+\rho_3,c_2+\nu c_2+2\nu c_1+5\nu^2c_1+\gamma_3+\rho_4,2\gamma_1+\rho_5,2\gamma_2+\rho_6,2\gamma_3+\rho_7\}$。证明完毕。

参考文献

[1] H. Attouch、J. Bolte、P. Redont 和 A. Soubeyran。针对非凸问题的近端交替最小化与投影方法：一种基于Kurdyka-Łojasiewicz不等式的方法。《数学运筹研究》，35(2):438–457，2010。[2] H. Attouch、J. Bolte 和 B. F. Svaiter。半代数问题与驯 tame问题的下降方法收敛性：近端算法、前向–后向分裂和正则化Gauss-Seidel方法。《数学规划》，137(1-2):91–129，2013。[3] M. T. Bahadori、Q. R. Yu 和 Y. Liu。通过低秩张量学习进行快速多变量时空分析。神经信息处理系统进展，27，2014。[4] S. Basu、X. Li 和 G. Michailidis。高维向量自回归的低秩与结构化建模。《IEEE信号处理汇刊》，67(5):1207–1222，2019。[5] S. Basu 和 G. Michailidis。稀疏高维时间序列模型中的正则化估计。《统计年刊》，43(4):1535–1567，2015。[6] A. Beck。优化中的一阶方法。工业与应用数学学会，费城，2017。[7] M. Belkin 和 P. Niyogi。用于嵌入与聚类的拉普拉斯特征映射和谱技术。载于：第15届神经信息处理系统国际会议论文集，第585–591页，2001。[8] J. Bolte、S. Sabach 和 M. Teboulle。用于非凸与非光滑问题的近端交替线性化最小化。《数学规划》，146(1-2):459–494，2014。[9] D. Cai、X. He、J. Han 和 T. S. Huang。用于数据表示的图正则化非负矩阵分解。《IEEE模式分析与机器智能汇刊》，33(8):1548–1560，2011。[10] R. Chen、D. Yang 和 C.-H. Zhang。高维张量时间序列的因子模型。《美国统计学会会刊》，117(537):94–116，2022。[11] X. Chen、C. Zhang、X.-L. Zhao、N. Saunier 和 L. Sun。利用非平稳时间矩阵分解预测城市道路网络中的稀疏移动速度。《交通科学》，59(3):670–687，2025。[12] Y. Chen、Y. Chi、J. Fan 和 C. Ma。面向数据科学的谱方法：统计学视角。《机器学习基础与趋势》，14(5):566–806，2021。[13] F. R. Chung。《谱图论》，第92卷。普罗维登斯：美国数学学会，1997。[14] L. De Lathauwer、B. De Moor 和 J. Vandewalle。多线性奇异值分解。《SIAM矩阵分析与应用杂志》，21(4):1253–1278，2000。[15] C. De Mol、D. Giannone 和 L. Reichlin。使用大量预测变量进行预测：贝叶斯收缩是否是主成分的有效替代方案? 《计量经济学杂志》，146(2):318–328，2008。[16] C. Duchon 和 R. Hale。《气象学与气候学中的时间序列分析：导论》。Wiley-Blackwell，牛津，2012。[17] G. H. Golub 和 C. F. Van Loan。《矩阵计算》，第4版。约翰斯·霍普金斯大学出版社，费城，宾夕法尼亚州，2013。[18] C. Gorrostieta、H. Ombao、P. B´edard 和 J. N. Sanes。利用混合效应向量自回归模型研究脑连接性。《NeuroImage》，59(4):3347–3355，2012。[19] D. Gueridi 和 F. Kittaneh。矩阵克罗内克积的不等式。《泛函分析年刊》，13(3):50，2022。[20] S. E. Hampton、E. E. Holmes、L. P. Scheef、M. D. Scheuerell、S. L. Katz、D. E. Pendleton 和 E. J. Ward。利用多变量方法量化非生物与生物驱动因素对群落动态的影响

自回归（mar）模型。《生态学》，94(12):2663–2669, 2013.[21] F. Han、H. Lu 和 H. Liu。高维平稳向量自回归的直接估计。《机器学习研究杂志》，16(1):3115–3150, 2015.[22] R. Han、R. Willett 和 A. R. Zhang。广义张量估计的最优统计与计算框架。《统计年刊》，50(1):1–29, 2022.[23] K. D. Harris、A. Aravkin、R. Rao 和 B. W. Brunton。低秩张量时变自回归。《SIAM应用动力系统杂志》，20(4):2335–2358, 2021.[24] C. J. Hillar 和 L.-H. Lim。大多数张量问题都是 NP 难。《ACM杂志》，60(6):45, 2013.[25] M. Hirn。讲座 01：谱图论导论。技术报告，2021.[26] F. L. Hitchcock。张量或多元项表示为乘积之和。《数学物理杂志》，6(1-4):164–189, 1927.[27] R. A. Horn 和 C. R. Johnson。《矩阵分析》。剑桥大学出版社，纽约，第 2 版，2013.[28] A. B. Kock、R. S. Pedersen 和 J. R.-V. Sørensen。高维向量自回归的数据驱动调优参数选择。《美国统计学会会刊》，第 1–19 页，2025.[29] T. G. Kolda 和 B. W. Bader。张量分解及其应用。《SIAM评论》，51(3):455–500, 2009.[30] G. M. Koop。利用中、大型贝叶斯向量自回归模型进行预测。《应用计量经济学杂志》，28(2):177–203, 2013.[31] R. P. Masini、M. C. Medeiros 和 E. F. Mendes。弱相关创新项下高维向量自回归的正则化估计。《时间序列分析杂志》，43(4):532–557, 2022.[32] K. Miao、P. C. Phillips 和 L. Su。具有共同因子的高维向量自回归模型。《计量经济学杂志》，233(1):155–183, 2023.[33] S. Negahban 和 M. J. Wainwright。含噪声及高维尺度下（近似）低秩矩阵的估计。《统计年刊》，39(2):1069–1097, 2011.[34] D. Qiu、B. Yang 和 X. Zhang。通过字典学习和广义非凸正则化实现稳健张量补全，以恢复视觉数据。《IEEE电路与系统视频技术汇刊》，34(11):11026–11039, 2024.[35] Y. Qiu、G. Zhou、Y. Wang、Y. Zhang 和 S. Xie。用于张量数据表示的广义图正则化非负 Tucker 分解框架。《IEEE控制论汇刊》，52(1):594–607, 2022.[36] G. Raskutti、M. Yuan 和 H. Chen。高维多响应张量回归的凸正则化。《统计年刊》，47(3):1554–1584, 2019.[37] G. Reinsel。多元自回归指数模型的一些结果。《生物统计学》，70(1):145–156, 1983.[38] R. T. Rockafellar 和 R. J. B. Wets。《变分分析》。施普林格，纽约，1998.[39] M. Roughan、Y. Zhang、W. Willinger 和 L. Qiu。时空压缩感知与互联网流量矩阵（扩展版）。《IEEE/ACM网络汇刊》，20(3):662–676, 2012.[40] S. Y. Samadi 和 H. W. B. Herath。降秩包络向量自回归模型。《商业与经济统计学杂志》，42(3):918–932, 2024.[41] C. A. Sims。宏观经济学与现实。《计量经济学》，48(1):1–48, 1980.[42] G. W. Stewart 和 J.-g. Sun。《矩阵扰动理论》。学术出版社，纽约，1990.[43] R. S. Tsay。《金融时间序列分析》。约翰·威立父子出版公司，纽约，2005.[44] D. Wang 和 R. S. Tsay。高维向量自回归模型的速率最优稳健估计。《统计年刊》，51(2):846–877, 2023.[45] D. Wang、X. Zhang、G. Li 和 R. Tsay。具有共同

响应因子和预测因子的高维向量自回归。arXiv:2203.15170, 2022.[46] D. Wang、Y. Zheng 和 G. Li。高维低秩张量自回归时间序列建模。《计量经济学杂志》，238(1):105544, 2024.[47] D. Wang、Y. Zheng、H. Lian 和 G. Li。通过张量分解进行高维向量自回归时间序列建模。《美国统计学会会刊》，117(539):1338–1356, 2022.[48] H. Wang、T. Liu、R. Li、M. X. Cheng、T. Zhao 和 J. Gao。Roselora：用于知识编辑与微调的预训练语言模型行列级稀疏低秩适配。载于《2024年自然语言处理实证方法会议论文集》，第 996–1008 页。计算语言学协会，2024.[49] H. Wang、C. Weng 和 J. Yuan。通过极小极大优化进行多特征谱聚类。载于《IEEE计算机视觉与模式识别会议论文集》，第 4106–4113 页，2014.[50] Z. Wang、A. Safikhani、Z. Zhu 和 D. S. Matteson。利用时空信息对高维向量自回归模型进行正则化估计。《统计学报》，33:1271–1294, 2023.[51] K. C. Wong、Z. Li 和 A. Tewari。β-混合重尾时间序列的 Lasso保证。《统计年刊》，48(2):1124–1142, 2020.[52] K. Yang、X. Dong 和 L. Toni。图拉普拉斯正则化估计量的误差分析。arXiv:1902.03720, 2019.[53] Y. Yu、T. Wang 和 R. J. Samworth。适用于统计学家的Davis–Kahan定理一个实用变体。《生物统计学》，102(2):315–323, 2015.[54] X. Zhu。稀疏向量自回归模型中的非凹惩罚估计。《电子统计学杂志》，14(1):1413–1448, 2020.[55] H. Zou、T. Hastie 和 R. Tibshirani。稀疏主成分分析。《计算与图形统计学杂志》，15(2):265–286, 2006.