

复杂缺失模式下基于自适应截断 Schatten 范数的交通数据补全方法

邓浩彭, 郑富城, 马楷祥, 夏新海*

(1. 广州航海学院 智能交通与工程学院 (未来交通学院), 广东 广州 510725)

摘要: 智能交通系统 (ITS) 的运行效能依赖高质量的交通数据。然而, 传感器故障与数据传输中断常导致复杂多变的数据缺失模式, 严重威胁系统可靠性。为此, 本研究提出一种基于自适应截断 Schatten 范数的低秩张量补全模型 (LRTC-ATSN)。相较于传统 LRTC 方法, 该模型采用 Schatten 范数替代核范数实现更灵活的低秩近似, 并引入自适应截断机制, 抑制小奇异值噪声的同时保留数据主成分, 有效提高了补全精度。针对 Schatten 范数松弛带来的非凸优化问题, 设计了 ADMM-TGST 求解框架: 结合交替方向乘子法与截断广义软阈值化, 将复杂优化问题分解为易于求解的子问题序列, 并运用集成 Nesterov 动量的 Adan 算法实现参数迭代寻优, 平衡了效率与精度。为全面评估模型性能, 构建了复杂缺失模式生成框架, 涵盖元素缺失、纤维缺失及其混合形态, 精准模拟现实交通场景中的多样化缺失特性。在广州与西雅图实测数据上的实验表明: LRTC-ATSN 在各类缺失场景下的效率、精度均优于现有方法。在 95.85% 的极端缺失条件下, 较最优基线模型 MAPE 与 RMSE 分别降低 10.6% 和 6.1%, 收敛效率提升超 20%。该模型为复杂缺失模式下的交通数据补全提供了高效、精准的解决方案, 并在金融时序预测、医疗影像补全等领域展现出泛化应用潜力。

关键词: 交通数据插补; 低秩张量补全; 数据缺失模式; 自适应更新; 非凸优化; 截断 Schatten 范数

1 引言

智能交通系统 (ITS) 在现代城市交通管理中发挥着不可或缺的作用, 通过流量调控、信号配时和紧急响应, 保障交通系统的高效与安全运行。然而, 随着城市交通规模化与复杂化, 数据缺失问题日益凸显。传感器故障、通信中断、设备老化、环境遮挡等因素交织, 导致缺失模式复杂多变。Yang 等^[1] 研究指出, 高峰时段或关键路段的交通数据缺失, 将严重威胁 ITS 决策可靠性, 加剧拥堵与事故风险。因此, 实现复杂缺失场景下精准、高效的交通数据补全, 已成为智慧交通发展的迫切需求。

当前, 主流交通数据补全方法可分为三类: 时序方法、空间方法和时空方法。时序方法通过挖掘交通参数的周期性特征实现数据修复, 如 Chen 等^[2] 提出的拉普拉斯卷积表示 (LCR) 模型, 通过时序正则化技术有效抑制了误差累积; 空间方法则侧重于路网拓扑关联性, 如 Chen 等^[3] 开发的动态图卷积网络 (RDGCNI) 模型, 捕捉站点间的交通状态传播规律, 以提升恢复精度。然而, 时序方法忽视了空间依赖, 空间方法则缺乏时间动态性。故此, 全面考虑数据多维特性的时空方法成为突破口。

近年来，低秩张量补全（Low-Rank Tensor Completion, LRTC）因其对时空关联性的显式建模能力备受关注。Liu 等人^[4]提出经典的 HaLRTC 模型，采用核范数松弛张量秩最小化问题，但其对各阶奇异值的等权处理方式难以区分数据主成分与噪声干扰，导致补全精度受限。对此，学者们相继提出截断核范数（TNN）^[5]和截断 Schatten 范数（TSpN）^[6]等改进方法，通过抑制次要奇异值提升模型性能。然而，现有方法仍面临两大挑战：其一，固定截断阈值和模式权重难以适应复杂多变的缺失场景；其二，非凸优化问题的求解效率低下，限制了模型在大规模交通数据中的应用。

为解决上述问题，本文提出基于自适应截断 Schatten 范数的低秩张量补全模型（LRTC-ATSN）。通过将 Adan 算法与 Nesterov 动量相结合^[7]，该模型在加速参数动态寻优的同时维持较高的计算精度。此外，与针对特定缺失模式的现有方法相比，LRTC-ATSN 能更好地适应反映真实世界的交通场景的复杂缺失模式。

本研究的贡献总结为下列三点：

1. 构建了混合缺失模式生成器，模拟真实交通场景中元素缺失、纤维缺失及其复合形态，为模型评估提供了标准化基准。
2. 设计了一种基于自适应截断 Schatten 范数的低秩张量补全模型（LRTC-ATSN），有效提高了复杂缺失场景下的补全精度。
3. 采用集成 Nesterov 动量的 Adan 优化算法，显著加速了模型收敛，使其在处理大规模交通数据张量时兼顾精度和效率。

2 符号说明

本文沿用 Kolda 和 Bader^[8]提出的张量符号体系。设 $\mathcal{X} \in \mathbb{R}^{d_1 \times d_2 \times \dots \times d_N}$ 为实数域 \mathbb{R} 上的 N 阶张量，其中 d_1, d_2, \dots, d_N 代表各维度长度。标量、向量、矩阵和张量分别用小写字母（如 a ）、粗体小写字母（如 $\mathbf{a} \in \mathbb{R}^M$ ）、粗体大写字母（如 $\mathbf{A} \in \mathbb{R}^{M \times N}$ ），及花体字母（如三阶张量 $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ ）表示。

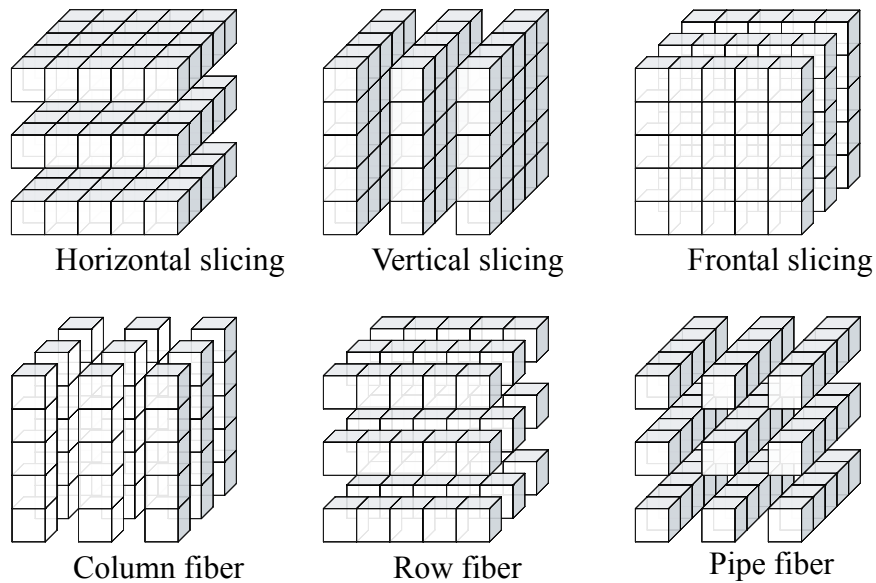


图 1. 张量切片和纤维类型说明

如图 1，进行切片操作可以提取张量局部信息。例如固定张量 $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ 的第三维度索引 k 可得 $I \times J$ 的矩阵切片 $\mathcal{X}_{::k}$ 。张量中的“纤维”则通过固定除一个以外的所有索引获得。在 $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ 中，列纤维 $\mathcal{X}_{:jk}$ 、行纤维 $\mathcal{X}_{i:k}$ 和管纤维 $\mathcal{X}_{ij:}$ 分别表示沿三个维度的纤维。张量中的单个元素则用 \mathcal{X}_{ijk} 表示。

为便于低秩优化中的张量操作，定义第 n 模下的展平（即矩阵化）与折叠操作：展平将张量 $\mathcal{X} \in \mathbb{R}^{d_1 \times d_2 \times \dots \times d_N}$ 转换为一个矩阵 $X^{(n)} \in \mathbb{R}^{d_n \times \prod_{i \neq n} d_i}$ ，而折叠则逆转该过程，将矩阵 $X^{(n)}$ 恢复至张量 \mathcal{X} 的原始结构。展平和折叠的具体映射关系如下：

$$j = 1 + \sum_{k=1, k \neq n}^N (i_k - 1)J_k, \quad \text{with } J_k = \prod_{m=1, m \neq n}^{k-1} d_m. \quad (1)$$

3 问题描述

以广州城市路网数据集（记为 G ）为例，对问题进行具体描述^[9]。将数据集建模为三维张量 $\mathcal{T} \in \mathbb{R}^{214 \times 61 \times 144}$ ，其中三个维度分别对应 214 个路段、61 天及每天 144 个时间窗口。张量元素 \mathcal{T}_{ijk} 表示第 i 个路段在第 j 天第 k 个时间窗口的平均车速。为直观展示交通流速度的变化规律，沿张量的两个时间维度进行切片：

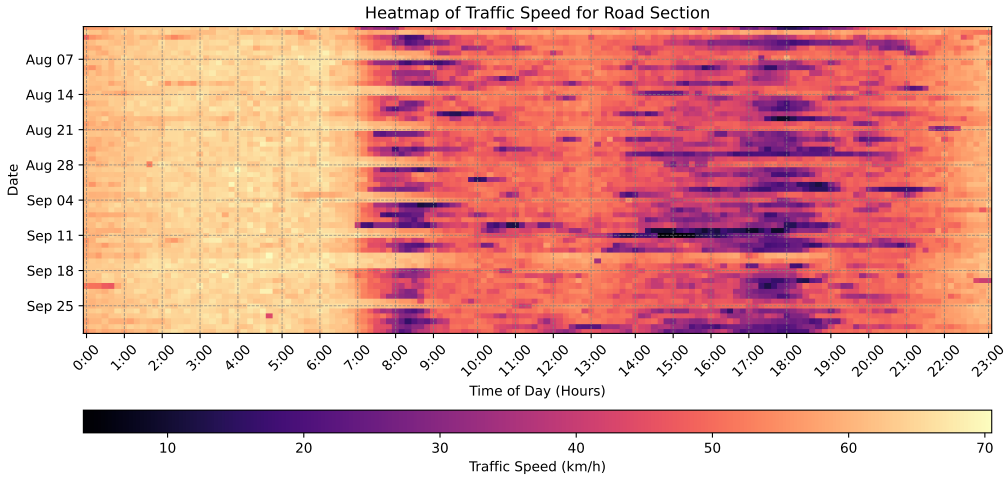


图 2. G 数据集的时间切片特征

由图 2 可见，交通流速度在不同时段与日期呈现明显的规律性变化。工作日早晚高峰（7:00-9:00 和 16:30-18:30）的车速明显下降，反映出通勤导致的交通拥堵。相比之下，周末（如 8 月 6 日、7 日与 13 日）和节假日（9 月 15 日至 17 日的中秋假期）的车速整体较高，即此时城市交通压力较小。此外，周末的交通高峰不如工作日明显，全天车速相对均衡，表明周末出行时间较为分散，路网负荷较低。

交通数据具有时空关联性，邻近传感器的记录通常呈现相似模式，特定时间窗口的数据在相近日期间也表现出相似趋势。这种局部至全局的相似性是交通缺失数据

补全的基本前提。张量作为多维数据结构的自然延伸，相较于传统的向量和矩阵，能够更全面地捕捉交通流数据中的时空关联性。

4 缺失类型

缺失数据的类型显著影响交通数据恢复的复杂性和适用性。根据其特性，我们将缺失数据分为三大类：元素缺失（Element Missing, EM）、纤维缺失（Fiber Missing, FM）和混合缺失（Mixed Missing, MM）。

4.1 元素缺失

元素缺失是指由传感器短暂故障或传输错误等原因，导致交通数据在特定路段、日期或时间区间内出现随机散布的数据丢失。由于元素缺失具有局部性特点，通常可以使用基于插值的方法进行恢复。

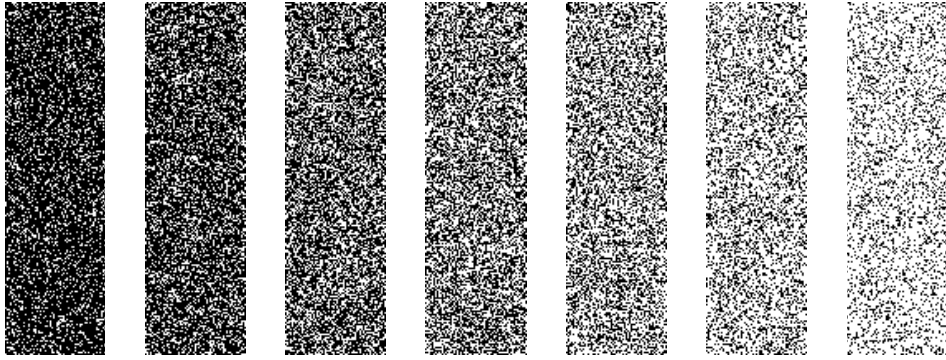


图 3. 缺失率在 20% 到 80% 之间，元素缺失场景下的张量切片变化

为模拟元素缺失，生成与交通数据 $\mathcal{T} \in \mathbb{R}^{I \times J \times K}$ 同维的均匀分布随机张量 $\mathcal{R} \sim U(0, 1)$ ，若 $\mathcal{R}_{ijk} \leq MR \in [0, 1]$ ，则将对对应元素标记为缺失值 NaN ，从而达成预设缺失率 $MR \times 100\%$ 。元素缺失场景下，张量切片形态随缺失率变化如图 3 所示。

4.2 纤维缺失

纤维缺失即交通数据中沿特定维度或多个维度上连续数据的缺失，通常由传感器长时间故障、遮挡或周期性系统问题引起。该类缺失涉及大量连续的数据点，信息损失的规模较大，恢复难度较高。

为模拟纤维缺失，在交通数据 $\mathcal{T} \in \mathbb{R}^{I \times J \times K}$ 中沿随机选定的维度 $d \in \{0, 1, 2\}$ （即路段、日期或时间区间），从随机起点 s_d 开始依次删除 l 个连续数据点，重复此过程直至达到目标缺失率。图 4 展示了纤维缺失场景下张量切片形态随缺失率变化。

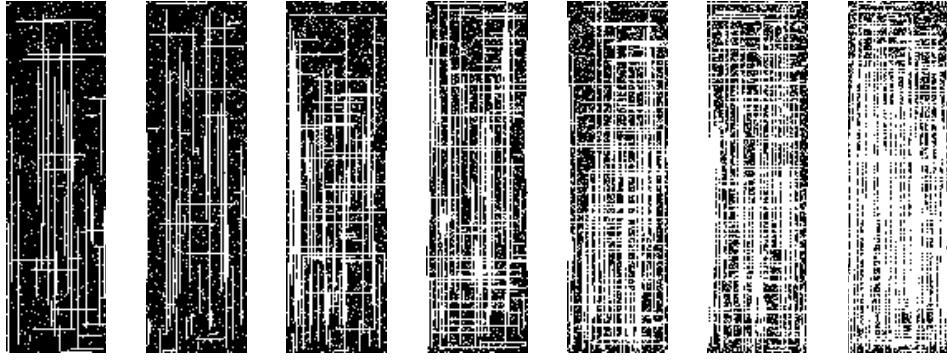
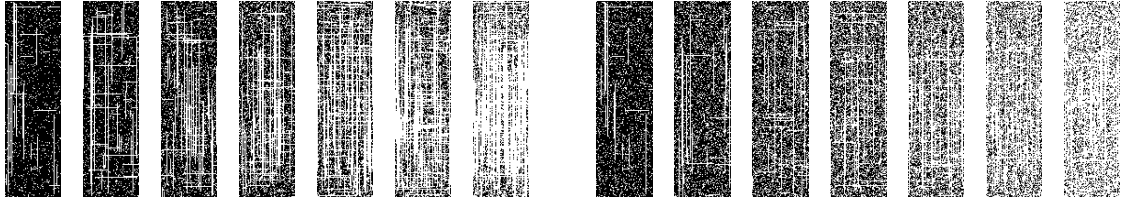


图 4. 缺失率在 20% 到 80% 之间，纤维缺失场景下的张量切片变化

4.3 混合缺失

混合缺失结合了元素缺失和纤维缺失，反映出如传感器不稳定、传输干扰及设备长时间故障等实际场景中更为复杂的情况。这种复合型缺失对数据补全提出了极大的挑战，需要同时采用局部插值与全局特征提取的恢复策略。

为生成混合缺失，定义总缺失率 $TMR = EMR + FMR$ ，其中 EMR 和 FMR 对应元素缺失率和纤维缺失率。基于历史数据参数估计两者占比，即可精准还原真实交通数据的多样化缺失特性。混合缺失的张量切片形态随缺失率变化见图 5。



(a) MM (20%EM, 80%FM)

(b) MM (40%EM, 60%FM)

图 5. 缺失率在 20% 到 80% 之间，混合缺失场景下的张量切片变化

5 模型构建

为应对复杂场景下的交通数据缺失修复问题，本研究提出了自适应截断 Schatten 范数低秩张量补全模型 (LRTC-ATSN, Adaptive and Truncated Schatten Norm)。LRTC-ATSN 通过结合非凸 Schatten 范数与截断机制，增强了对小奇异值的有效惩罚，从而实现更精确的稀疏性控制和数据恢复。在此基础上，采用 Adan 算法 (ADaptive Nesterov momentum algorithm) 的自适应更新策略，根据不同数据的特性动态调整模型参数，从而确保数据修复高效性与准确性。

5.1 传统 LRTC 模型

LRTC 的目标是最小化张量的秩，形式化为 $\min_{\mathcal{X}} \text{rank}(\mathcal{X})$ 。由于直接求解张量秩最小化是 NP-hard 问题，通常采用核范数进行凸松弛以获取近似解。对于张量 $\mathcal{X} \in \mathbb{R}^{d_1 \times d_2 \times \dots \times d_N}$ ，沿每个模式 k -mode 展平为矩阵 $\mathbf{X}_{(k)} = \text{Unfold}_k(\mathcal{X})$ ，其奇异值之和即定义为模式 k 的张量核范数： $\|\mathcal{X}\|_* = \sum_{k=1}^N \|\mathbf{X}_{(k)}\|_*$ 。为了在张量补全中仅对观测位置施加约束，定义观测掩码张量 \mathcal{P}_Ω ，保留在观测集 Ω 中的值，并将其他值设置为 0。因此，LRTC 的目标变为最小化核范数，并在观测位置 $(i_1, i_2, \dots, i_N) \in \Omega$ 约束补全值与观测值一致。设 \mathcal{T} 为包含观测值的原始张量，优化问题可表示为：

$$\min_{\mathcal{X}} \sum_{k=1}^N \|\mathbf{X}_{(k)}\|_* \quad \text{s.t.} \quad \mathcal{P}_\Omega(\mathcal{X}) = \mathcal{P}_\Omega(\mathcal{T}). \quad (2)$$

5.2 Schatten 范数

为提高 LRTC 的精度和灵活性，引入 Schatten 范数替代传统核范数松弛。根据 Gao 和 Fan^[10] 的研究，对于 N 阶张量 $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ ，Schatten 范数定义为：

$$\|\mathcal{X}\|_{S_p} = \left(\sum_{k=1}^N \alpha_k \|\mathbf{X}_{(k)}\|_{S_p}^p \right)^{\frac{1}{p}}, \quad (3)$$

其中， α_k 是模式权重，衡量模式 k 对张量整体低秩特性的贡献^[11]，满足 $\alpha_k \geq 0$ 且 $\sum_{k=1}^N \alpha_k = 1$ ；参数 p 控制 Schatten 范数对奇异值的敏感度。当 $p = 1$ 时，对应核范数 $\|\mathcal{X}\|_{S_1} = \sum_{i=1}^r \sigma_i$ ；在 $0 < p < 1$ 时，Schatten 范数呈非凸特性，能够更有效地惩罚小奇异值，从而实现更高的稀疏性和更紧凑的低秩结构。

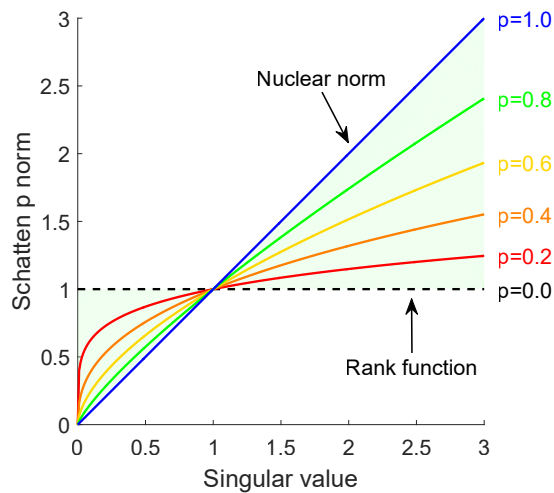


图 6. $0 < p < 1$ 时，Schatten 范数随奇异值变化曲线

如图 6 所示，随着 p 值趋近 0，Schatten 范数的曲线愈接近张量秩，此时非凸性增强，求解精度提高，但计算复杂度也随之增加。因此，根据数据集与缺失模式特性

动态调整 p 值，有助于在求解精度与计算效率之间取得平衡。

5.3 截断操作

奇异值衡量不同特征维度的重要性：较大的奇异值表示主要趋势和周期性成分，反映了数据的整体结构和低秩特性；而较小的奇异值则对应短期、偶发或无规律的波动。为了减少较小奇异值的干扰，在 Schatten 范数中加入了截断操作：

$$\|\mathcal{X}\|_{\theta, S_p} = \left(\sum_{k=1}^N \alpha_k \|\mathcal{X}_{(k)}\|_{\theta, S_p}^p \right)^{\frac{1}{p}} = \left(\sum_{k=1}^N \alpha_k \left(\sum_{i=r_k+1}^{\min\{I_k, \prod_{i \neq k} I_i\}} \sigma_i^p(\mathcal{X}_{(k)}) \right) \right)^{\frac{1}{p}}, \quad (4)$$

其中， r_k 表示每个展平矩阵的截断阈值，定义如下：

$$r_k = \left\lceil \theta \cdot \min\{I_k, \prod_{i \neq k} I_i\} \right\rceil, \quad \forall k \in \{1, 2, \dots, N\}, \quad (5)$$

其中， $\theta \in [0, 1]$ 为全局截断率，符号 $\lceil \cdot \rceil$ 表示向上取整，奇异值 σ_i 按非升序排列。通过截断，仅有最小的 $\left(\min I_k, \prod_{i \neq k} I_i - r \right)$ 个奇异值会对 Schatten 范数产生贡献，最大的 r 个奇异值保持不变，而较小的奇异值被缩减，以减少其对数据修复的噪声影响。在 LRTC 中应用带截断的 Schatten 范数后，优化问题变为：

$$\min_{\mathcal{X}} \|\mathcal{X}\|_{\theta, S_p}^p = \min_{\mathcal{X}} \sum_{k=1}^N \alpha_k \|\mathcal{X}_{(k)}\|_{\theta, S_p}^p, \quad \text{s.t.} \quad \mathcal{P}_{\Omega}(\mathcal{X}) = \mathcal{P}_{\Omega}(\mathcal{T}). \quad (6)$$

在该式中，参数 p 、截断率 θ 和模式权重 α_k 的选择至关重要。在不同的数据集、缺失模式和迭代阶段中，其最佳取值会有所不同。对此，需要实现这些参数的动态调整，以提高模型的适应性和整体性能。

5.4 自适应更新机制

为提升模型收敛速度和补全精度，本研究设计了一种自适应更新机制，通过动态调整参数 p 、截断率 θ 以及模式权重 α_k ，实现更优的稀疏控制和低秩约束。具体而言，我们引入 Adan 算法的动量更新策略，自适应调整 p 和 θ ，并结合正则化项，通过加权移动平均法对 α_k 进行平滑更新^[12]。

5.4.1 自适应更新 p 和 θ

在每轮迭代中，通过计算第 t 次迭代的误差梯度 $g^{(t)} = E^{(t)} - E^{(t-1)}$ ，更新一阶动量 $m(t)$ 和二阶动量 $v(t)$ ：

$$m^{(t)} = \beta_1 m^{(t-1)} + (1 - \beta_1) g^{(t)}, \quad v^{(t)} = \beta_2 v^{(t-1)} + (1 - \beta_2) (g^{(t)})^2, \quad (7)$$

其中, β_1 和 β_2 为一阶和二阶动量的衰减系数, 常取 $\beta_1 = 0.9$, $\beta_2 = 0.999$, 用以在模型更新中平衡历史和当前梯度的影响。

为了消除初始偏差, 对上述动量进行无偏校正, 使其更准确地反映当前梯度的真实值, 校正公式如下:

$$\hat{m}^{(t)} = \frac{m^{(t)}}{1 - \beta_1^t}, \quad \hat{v}^{(t)} = \frac{v^{(t)}}{1 - \beta_2^t}, \quad (8)$$

利用校正后的动量项, 对 p 和 θ 进行自适应更新:

$$p_{t+1} = \text{clip} \left(p_t - \eta \frac{\hat{m}^{(t)}}{\sqrt{\hat{v}^{(t)}} + \epsilon}, 0.1, 1 \right), \quad \theta_{t+1} = \text{clip} \left(\theta_t + \eta \frac{\hat{m}^{(t)}}{\sqrt{\hat{v}^{(t)}} + \epsilon}, 0, 1 \right), \quad (9)$$

其中, η 为学习率, 控制参数更新的步幅, ϵ 为防止除零的小量, $\text{clip}(\cdot)$ 函数用于确保参数 p 和 θ 在合理范围内变化, 避免出现无效值或不收敛现象。

5.4.2 平滑更新 α_k

模式权重 α_k 的更新通过计算每个模式的 Frobenius 范数 $s_k = \|\mathcal{M}_k\|_F$ 来实现, 并引入正则化项定义归一化权重候选值:

$$\tilde{\alpha}_k = \frac{s_k}{\sum_{j=1}^N s_j} + \lambda \left(\frac{1}{N} - \frac{s_k}{\sum_{j=1}^N s_j} \right), \quad (10)$$

其中, λ 为正则化系数, 控制权重向平均值的收缩程度, 以防止单个模式权重过度集中。

加权移动平均法是一种用于平滑数据更新的技术, 通过在更新中融合历史和当前信息, 防止权重更新剧烈波动。采用该方法更新权重:

$$\alpha_k^{(t+1)} = (1 - \gamma) \alpha_k^{(t)} + \gamma \tilde{\alpha}_k, \quad (11)$$

最后进行归一化 $\alpha_k^{(t+1)} = \alpha_k^{(t+1)} / \sum_{j=1}^N \alpha_j^{(t+1)}$, 式中 γ 为权重更新的学习率。

6 求解算法

为高效解决截断张量 Schatten 范数的最小化问题, 本研究采用交替方向乘子法^[13]与截断广义软阈值化^[14]相结合的方法 (ADMM-TGST), 将复杂的优化问题分解为易于求解的子问题, 并通过迭代逐步逼近最优解。模型求解流程见图 7。

在开始求解前, 重新审视原优化问题:

$$\min_{\mathcal{X}} \|\mathcal{X}\|_{\theta, S_p}^p = \min_{\mathcal{X}} \sum_{k=1}^N \alpha_k \|\mathcal{X}_{(k)}\|_{\theta, S_p}^p, \quad \text{s.t.} \quad \mathcal{P}_{\Omega}(\mathcal{X}) = \mathcal{P}_{\Omega}(\mathcal{T}). \quad (12)$$

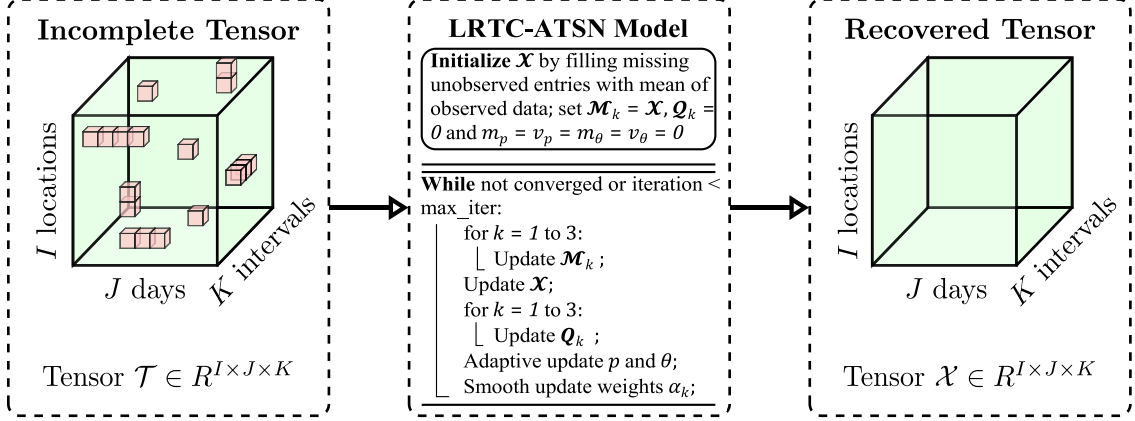


图 7. LRTC-ATSN 模型的求解流程

为方便求解，引入辅助变量 $\{\mathcal{M}_k\}_{k=1}^N$ ，将问题转化为约束优化形式：

$$\min_{\mathcal{X}, \{\mathcal{M}_k\}} \sum_{k=1}^N \alpha_k \|\mathcal{M}_k\|_{\theta, S_p}^p, \quad \text{s.t.} \quad \begin{cases} \mathcal{M}_k = \mathcal{X}, & \forall k = 1, 2, \dots, N, \\ \mathcal{P}_\Omega(\mathcal{X}) = \mathcal{P}_\Omega(\mathcal{T}). \end{cases} \quad (13)$$

在此基础上，引入拉格朗日乘子 $\{\mathcal{Q}_k\}_{k=1}^N$ 以构建增广拉格朗日函数：

$$\mathcal{L}(\mathcal{X}, \{\mathcal{M}_k\}, \{\mathcal{Q}_k\}) = \sum_{k=1}^N \left[\alpha_k \|\mathcal{M}_k\|_{\theta, S_p}^p + \langle \mathcal{Q}_k, \mathcal{X} - \mathcal{M}_k \rangle + \frac{\beta}{2} \|\mathcal{X} - \mathcal{M}_k\|_F^2 \right], \quad (14)$$

其中， $\beta > 0$ 是增广拉格朗日乘子的罚参数，用于加强等式约束。接下来，我们通过 ADMM 交替优化变量 \mathcal{X} 、 \mathcal{M}_k 和 \mathcal{Q}_k ，具体步骤如下：

6.1 更新 \mathcal{M}_k

为更新 \mathcal{M}_k ，ADMM 在第 t 次迭代中通过固定 $\mathcal{X}^{(t)}$ 和 $\mathbf{Q}_k^{(t)}$ 交替优化。目标函数最小化 Schatten 范数正则项和 \mathcal{M}_k 的惩罚项，其表达式为：

$$\mathcal{M}_k^{(t+1)} = \arg \min_{\mathcal{M}_k} \alpha_k \|\mathcal{M}_k\|_p^p + \left\langle \mathbf{Q}_k^{(t)}, \mathcal{X}^{(t)} - \mathcal{M}_k \right\rangle + \frac{\beta}{2} \|\mathcal{X}^{(t)} - \mathcal{M}_k\|_F^2. \quad (15)$$

通过展开线性项并重新整理，简化为：

$$\mathcal{M}_k^{(t+1)} = \arg \min_{\mathcal{M}_k} \alpha_k \|\mathcal{M}_k\|_p^p + \frac{\beta}{2} \left\| \mathcal{M}_k - \left(\mathcal{X}^{(t)} + \frac{1}{\beta} \mathbf{Q}_k^{(t)} \right) \right\|_F^2. \quad (16)$$

将张量 \mathcal{M}_k 和 $\mathcal{X}^{(t)} + \frac{1}{\beta} \mathbf{Q}_k^{(t)}$ 沿着第 k 模展开成矩阵 \mathbf{M}_k 和 \mathbf{Y}_k 。分别对 \mathbf{Y}_k 进行奇异值分解 (SVD)，即 $\mathbf{Y}_k = \mathbf{U}_k \Sigma_k \mathbf{V}_k^\top$ ，其中 $\Sigma_k = \text{diag}(\sigma_1, \dots, \sigma_R)$ 。

为增强低秩特性，选择截断阈值 $r_k = [0 : R]$ ，保留前 r_k 个显著奇异值^[15]。奇异

值小于 r_k 的部分通过广义软阈值 (GST) 函数调整：

$$\delta_i = \text{GST}(\sigma_i, \tau_k, p), \quad \tau_k = \frac{\alpha_k w_i}{\beta}, \quad w_i = \begin{cases} 1, & i \leq r_k, \\ 0, & i > r_k. \end{cases} \quad (17)$$

重构 $\Delta_k = \text{diag}(\delta_1, \dots, \delta_R)$, 更新 $\mathbf{M}_k^{(t+1)} = \mathbf{U}_k \Delta_k \mathbf{V}_k^\top$, 并将其折叠回张量 $\mathcal{M}_k^{(t+1)}$, 完成更新步骤。

6.2 更新 \mathcal{X}

在保持 $\mathcal{M}_k^{(t+1)}$ 和 $\mathbf{Q}_k^{(t)}$ 固定的情况下, 为更新 \mathcal{X} , 构造以下优化问题：

$$\mathcal{X}^{(t+1)} = \arg \min_{\mathcal{X}} \sum_{k=1}^N \left\langle \mathbf{Q}_k^{(t)}, \mathcal{X} - \mathcal{M}_k^{(t+1)} \right\rangle + \frac{\beta}{2} \left\| \mathcal{X} - \mathcal{M}_k^{(t+1)} \right\|_F^2. \quad (18)$$

约束条件可以通过以下解析来满足：

$$\mathcal{X}^{(t+1)} = \mathcal{P}_{\Omega^c}(\mathcal{T}) + \mathcal{P}_{\Omega} \left(\frac{1}{N} \sum_{k=1}^N \left(\mathcal{M}_k^{(t+1)} - \frac{1}{\beta} \mathbf{Q}_k^{(t)} \right) \right), \quad (19)$$

其中, $\mathcal{P}_{\Omega^c} = \mathbf{I} - \mathcal{P}_{\Omega}$ 是未观测条目的投影算子, \mathbf{I} 是单位算子。此步骤通过整合所有模式信息更新 \mathcal{X} , 并保持观测数据不变。

6.3 更新 \mathcal{Q}_k

为逐步逼近等式约束 $\mathcal{X} = \mathcal{M}_k$ 的理想解, 调整拉格朗日乘子 \mathcal{Q}_k , 确保 \mathcal{X} 与各模式辅助变量 \mathcal{M}_k 的一致性：

$$\mathcal{Q}_k^{(t+1)} = \mathcal{Q}_k^{(t)} + \beta \left(\mathcal{X}^{(t+1)} - \mathcal{M}_k^{(t+1)} \right). \quad (20)$$

对各模式 k 的拉格朗日乘子进行更新, 该调整逐步提升了 \mathcal{X} 和 \mathcal{M}_k 的相似度, 使得 \mathcal{X} 的解稳定在一个满足约束的状态。

6.4 迭代与收敛

持续更新主变量 \mathcal{X} 、辅助变量 \mathcal{M}_k 和拉格朗日乘子 \mathcal{Q}_k , 直到满足以下收敛条件：

$$\frac{\left\| \mathcal{X}^{(t+1)} - \mathcal{X}^{(t)} \right\|_F}{\left\| \mathcal{X}^{(t)} \right\|_F} < \epsilon, \quad (21)$$

其中, $\epsilon > 0$ 为预设精度阈值。

7 实验

7.1 数据集

为验证 LRTC-ATSN 模型在真实场景中的性能, 选取了两个来自实际交通系统的数据集。除尺寸不同外, 均遵循相同的张量结构: 位置 \times 天数 \times 间隔。

a) 中国广州城市路网车速数据 (简记为 G): 记录 2016 年 8 月 1 日至 9 月 30 日的实测数据, 采样间隔为 10 分钟 (每天有 144 个时间窗口), 涵盖 214 条城市快速路和主干道。自带 1.29% 的缺失值。张量尺寸为 $214 \times 61 \times 144$ 。

b) 美国西雅图高速公路车速数据 (简记为 S)^[16]: 记录 2015 年 1 月 1 日至 1 月 28 日的实测数据, 采样间隔为 5 分钟 (每天有 288 个时间窗口), 数据来自西雅图的主要高速公路 323 个环路检测器。张量尺寸为 $323 \times 28 \times 288$ 。

7.2 基线模型

本研究选取了五个具代表性的基线模型用于对比: HaLRTC、LRTC-TNN、LRTC-TSpN、BGCP^[17] 以及 BPMF^[18]。其中, HaLRTC、LRTC-TNN 和 LRTC-TSpN 是采用不同低秩近似方法的 LRTC 模型, 分别通过核范数、截断核范数和截断 Schatten 范数实现低秩逼近。而 BGCP 和 BPMF 则基于贝叶斯推断, 前者采用高斯 CP 分解进行时间依赖数据恢复, 后者则通过概率矩阵分解描述数据潜在结构。

7.3 评价指标

本研究采用平均绝对百分比误差 (MAPE) 和均方根误差 (RMSE) 作为数据修复精度的评价指标 (式 22)。MAPE 衡量预测值与实际值的平均相对误差, RMSE 衡量预测值与实际值之间误差的均方根。两者的数值越小, 模型补全精度越高。

$$MAPE = \frac{1}{|\Omega|} \sum_{(i,j,k) \in \Omega} \left| \frac{x_{i,j,k} - \hat{x}_{i,j,k}}{x_{i,j,k}} \right| \times 100\%, \quad RMSE = \sqrt{\frac{1}{|\Omega|} \sum_{(i,j,k) \in \Omega} (x_{i,j,k} - \hat{x}_{i,j,k})^2}. \quad (22)$$

7.4 实验结果

7.4.1 收敛效率分析

计算复杂度是非凸优化的首要挑战。为测试 LRTC-ATSN 求解算法的效率, 对数据集 G 和 S 在三种混合缺失条件 (EM-FM 组合: 0.1/0.3, 0.2/0.2, 0.3/0.1) 下进行了收敛效率分析。如图 8 所示, MAPE 和 RMSE 在最初 5 次迭代中迅速下降, 在 30 次迭代左右稳定至可接受水平。

如下表 1 所示, LRTC-ATSN 的效率超越所有基线模型, 较最优基线提升 20% 以上, 展现了 ADMM-TGST 求解算法与自适应更新机制在不同场景中的有效性。

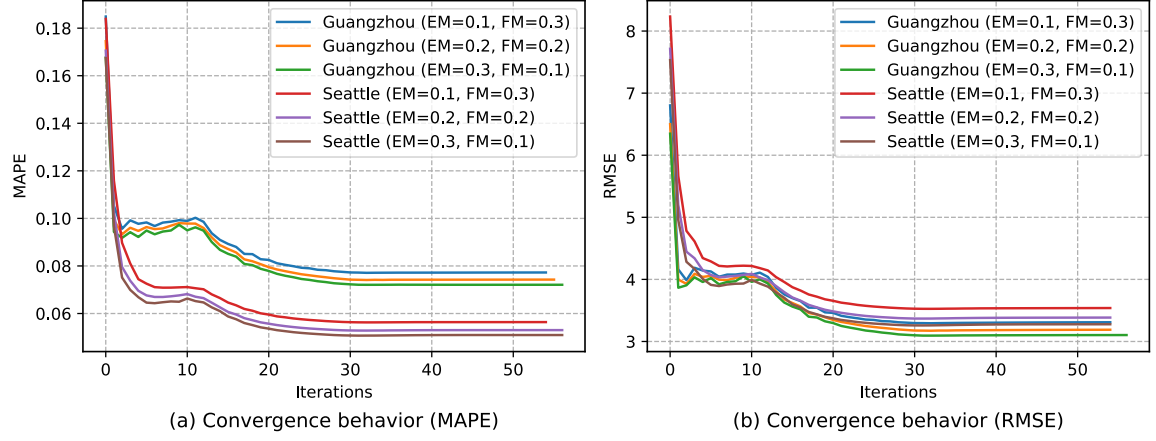


图 8. 不同情景下模型的收敛性能

表 1. 不同模型在不同缺失场景中的收敛效率

Datasets	Missing Pattern	TMR	BPMF	BGCP	HaLRTC	LRTC-TNN	LRTC-TSpN	LRTC-ATSN
G	EM 0.1 FM 0.1	19.60%	91.691	57.914	30.564	24.129	20.912	15.915
	EM 0.3 FM 0.3	48.62%	207.556	131.088	69.185	54.622	47.337	37.654
	EM 0.5 FM 0.5	69.75%	291.021	183.803	97.007	76.584	66.373	53.378
	EM 0.7 FM 0.7	85.00%	383.564	242.251	127.855	100.938	87.479	64.401
	EM 0.9 FM 0.9	95.85%	417.859	263.911	139.286	109.963	95.301	70.047
S	EM 0.1 FM 0.1	18.52%	127.204	85.339	42.401	33.475	29.011	22.079
	EM 0.3 FM 0.3	47.91%	272.064	171.827	90.687	71.595	62.049	49.356
	EM 0.5 FM 0.5	69.37%	389.622	246.077	129.874	102.532	88.861	71.463
	EM 0.7 FM 0.7	84.80%	520.298	328.617	173.433	136.921	118.665	87.359
	EM 0.9 FM 0.9	95.78%	592.801	374.536	197.697	156.312	135.258	99.373

Best results are marked in bold font.

7.4.2 补全精度比较

表 2 展示了不同模型在 G 和 S 数据集上的混合缺失模式补全性能。实验涵盖了 EM-FM 按不同比例（10%、30%、50%、70% 和 90%）组合形成的缺失场景，对应总缺失率大致为 20%、50%、70%、85% 及 95%。所使用的评估指标是 MAPE / RMSE。

结果表明，LRTC-ATSN 始终领先其他模型，即便在极端缺失率（95.85%）下，MAPE 和 RMSE 较最优基线仍分别降低 10.6% 和 6.1%。

表 2. 不同模型在不同缺失场景中的补全精度

Datasets	Missing Pattern	TMR	BPMF	BGCP	HaLRTC	LRTC-TNN	LRTC-TSpN	LRTC-ATSN
G	EM 0.1 FM 0.1	19.60%	0.106/4.412	0.104/4.321	0.088/3.561	0.073/3.160	0.073/3.132	0.072/3.097
	EM 0.3 FM 0.3	48.62%	0.108/4.491	0.104/4.318	0.098/3.929	0.082/3.530	0.081/3.423	0.078/3.360
	EM 0.5 FM 0.5	69.75%	0.112/4.665	0.103/4.316	0.109/4.301	0.089/3.849	0.087/3.689	0.087/3.676
	EM 0.7 FM 0.7	85.00%	0.119/4.951	0.104/4.320	0.123/4.800	0.098/4.195	0.093/3.943	0.093/3.947
	EM 0.9 FM 0.9	95.85%	0.139/5.675	0.107/4.473	0.819/33.808	0.111/4.627	0.105/4.409	0.105/4.355
S	EM 0.1 FM 0.1	18.52%	0.098/5.526	0.100/5.613	0.071/3.952	0.052/3.329	0.052/3.318	0.050/3.233
	EM 0.3 FM 0.3	47.91%	0.099/5.616	0.101/5.615	0.082/4.440	0.060/3.761	0.058/3.583	0.057/3.582
	EM 0.5 FM 0.5	69.37%	0.103/5.818	0.099/5.555	0.098/5.078	0.072/4.359	0.065/3.966	0.066/4.028
	EM 0.7 FM 0.7	84.80%	0.114/6.327	0.101/5.640	0.121/5.959	0.090/5.274	0.077/4.531	0.076/4.590
	EM 0.9 FM 0.9	95.78%	0.169/8.529	0.105/5.900	0.200/9.356	0.360/23.742	0.113/6.135	0.101/5.758

Best results are marked in bold font.

7.4.3 插值可视化

LRTC-ATSN 与 LRTC-TSpN 均采用截断操作与非凸优化插补缺失数据，但实现过程却有所不同。为直观展现 LRTC-ATSN 在复杂缺失模式下的稳健性，我们设计了一系列对比实验： G 数据集在场景 α ($EM = 0.2$, $FM = 0.3$) 产生约 41.28% 的 MM 数据； S 数据集在场景 β ($EM = 0.5$, $FM = 0.6$) 产生约 72.21% 的 MM 数据。

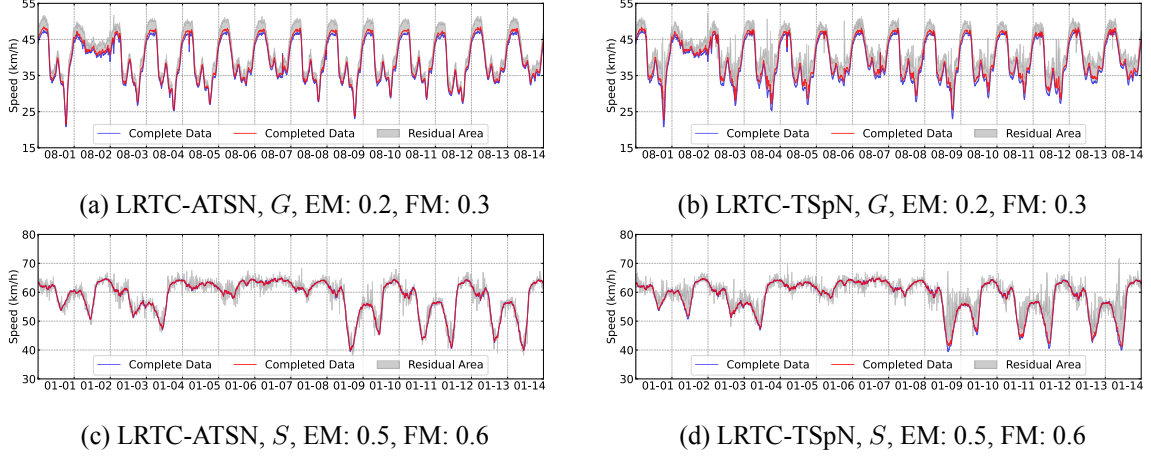


图 9. LRTC-ATSN 与 LRTC-TSpN 的插补性能比较

图 9 对比了原始完整数据与插补后的数据，为清晰起见，放大了残差区域。在场景 α 中，LRTC-ATSN 的残差面积较 LRTC-TSpN 减少 28.3% (1463.12 vs. 2040.98)。在数据规模更大、混合缺失率更高的场景 β 中，LRTC-ATSN 进一步缩减了 36.1% 的残差面积 (320.33 vs. 501.29)。得益于其灵活的参数与权重更新机制，LRTC-ATSN 在大规模数据集上表现更优，且在高缺失率下仍能保持真实的补全曲线。

8 结论

本研究提出了自适应截断 Schatten 范数低秩张量补全模型 (LRTC-ATSN)，通过动态调参与自适应截断，模型为复杂缺失场景下的交通数据修复提供了高效且精准的解决方案。为全面评估模型性能，设计了生成复杂缺失模式的框架，用以模拟真实交通场景的多样化缺失特性。实验结果表明，LRTC-ATSN 在各类缺失模式下均表现出优越的综合性能：效率较最优基线模型提升 20% 以上，且在极端缺失率 (95.78%) 下，MAPE 和 RMSE 分别降低 10.6% 和 6.1%。该模型为 ITS 数据可靠性保障提供了理论工具，并具备跨领域泛化潜力。未来工作将聚焦轻量化自适应机制以降低工程部署成本，并探索流形学习与注意力机制在高维数据特征提取中的应用。

9 致谢

本研究得到了广东省教育厅重点科研平台与项目(编号:2022ZDZX1021、2024ZDZX1033)，广州市教育局科研项目(编号:2024312023)以及广东省科技创新战略专项资金资助项目(编号:PDJH2024A289)的支持。

10 参考文献

- [1] Yang L and Wu H. A brief review on missing traffic data imputation methods for intelligent transportation systems. In *2024 7th International Symposium on Autonomous Systems (ISAS)*, pages 1–6. IEEE, 2024. 1
- [2] Chen X, Cheng Z, Cai H Q, Saunier N, and Sun L. Laplacian convolutional representation for traffic time series imputation. *IEEE Transactions on Knowledge and Data Engineering*, 2024. 1
- [3] Chen Y and Chen X M. A novel reinforced dynamic graph convolutional network model with data imputation for network-wide traffic flow prediction. *Transportation Research Part C: Emerging Technologies*, 143:103820, 2022. 1
- [4] Liu J, Musialski P, Wonka P, and Ye J. Tensor completion for estimating missing values in visual data. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):208–220, 2012. 1
- [5] Chen X, Yang J, and Sun L. A nonconvex low-rank tensor completion model for spatiotemporal traffic data imputation. *Transportation Research Part C: Emerging Technologies*, 117:102673, 2020. 1
- [6] Nie T, Qin G, and Sun J. Truncated tensor Schatten p -norm based approach for spatiotemporal traffic data imputation with complicated missing patterns. *Transportation research part C: emerging technologies*, 141:103737, 2022. 1
- [7] Xie X, Zhou P, Li H, Lin Z, and Yan S. Adan: Adaptive nesterov momentum algorithm for faster optimizing deep models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 1
- [8] Kolda T G and Bader B W. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009. 2
- [9] OpenITS Org. Opendata v12.0—large-scale traffic speed data set, 2021. [Online]. Available: <https://www.openits.cn/openData2/792.jhtml>. 3
- [10] Gao S and Fan Q. Robust Schatten- p norm based approach for tensor completion. *Journal of Scientific Computing*, 82:1–23, 2020. 5.2
- [11] Wang S, Zhao Y, Zhang Y, Hu Y, and Yin B. Spatiotemporal traffic data imputation via tensorial weighted Schatten- p norm minimization. *IET Intelligent Transport Systems*, 16(7):926–939, 2022. 5.2

- [12] Zhao M, Gahrooei M R, and Ilbeigi M. Change detection in partially observed large-scale traffic network data. *IEEE Transactions on Intelligent Transportation Systems*, 2024. 5.4
- [13] Zeng Z, Liu B, Feng J, and Yang X. Low-rank tensor and hybrid smoothness regularization-based approach for traffic data imputation with multimodal missing. *IEEE Transactions on Intelligent Transportation Systems*, 2024. 6
- [14] Goulart J H M, A Y Kibangou, and Favier G. Traffic data imputation via tensor completion based on soft thresholding of tucker core. *Transportation Research Part C: Emerging Technologies*, 85:348–362, 2017. 6
- [15] Bayati F and Trad D. 3-d data interpolation and denoising by an adaptive weighting rank-reduction method using multichannel singular spectrum analysis algorithm. *Sensors*, 23(2):577, 2023. 6.1
- [16] Cui Z, Henrickson K, Ke R, and Wang Y. Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting. *IEEE Transactions on Intelligent Transportation Systems*, 21(11):4883–4894, 2019. 7.1
- [17] Chen X, He Z, and Sun L. A bayesian tensor decomposition approach for spatiotemporal traffic data imputation. *Transportation research part C: emerging technologies*, 98:73–84, 2019. 7.2
- [18] Vander Aa T, Chakroun I, and Haber T. Distributed bayesian probabilistic matrix factorization. *Procedia Computer Science*, 108:1030–1039, 2017. 7.2

1 附录 A 参数表

参数	说明
α_k	模式权重, 衡量每个模式在张量整体低秩特性中的贡献。满足 $\alpha_k \geq 0$ 且 $\sum_{k=1}^N \alpha_k = 1$ 。初始值设为均匀分布 $\alpha_k = 1/N$ 。
p	Schatten 范数的指数参数, $0 < p < 1$ 时引入非凸性, 用于增强低秩性, 精确惩罚小奇异值。初始值设为 0.7。
θ	截断率, 控制奇异值截断的程度, 以减少小奇异值对恢复张量的影响, 降低噪声干扰。初始值设为 0.1。
r_k	模式 k 维度的截断阈值, 确定保留的奇异值数量。计算公式为: $r_k = \lceil \theta \cdot \min\{I_k, \prod_{i \neq k} I_i\} \rceil$ 。
β	增广拉格朗日乘子的罚参数, 确保 \mathcal{X} 和 \mathcal{M}_k 之间的一致性, 从而满足约束条件。初始值设为 1×10^{-5} 。
ϵ	收敛阈值, 控制迭代终止条件。当满足以下收敛准则时停止迭代: $\frac{\ \mathcal{X}^{(t+1)} - \mathcal{X}^{(t)}\ _F}{\ \mathcal{X}^{(t)}\ _F} < \epsilon$, 初始值设为 1×10^{-5} 。
η	学习率, 控制自适应更新过程中 p 和 θ 的调整步长。初始值设为 0.01。
β_1, β_2	动量参数, 控制 Adan 优化算法的动量更新, 分别影响一阶和二阶动量的衰减速率。推荐初始值 $\beta_1 = 0.9, \beta_2 = 0.999$ 。
λ	正则化系数, 用于平滑更新模式权重 α_k , 调节权重向均匀分布的收缩程度。推荐初始值 $\lambda = 0.01$ 。
γ	平滑因子, 控制模式权重 α_k 在更新时的平滑程度, 以避免剧烈波动。初始值设为 0.02。

表 1. LRTC-ATSN 模型参数说明

2 附录 B 引理证明及公式推导

B.1 Adan 动量更新的全局收敛性证明（针对公式 9）

本节证明 Adan 优化算法的参数 p_t 和 t 在迭代过程中收敛至最优值，即：

$$\lim_{t \rightarrow \infty} p_t = p^*, \quad \lim_{t \rightarrow \infty} \theta_t = \theta^*.$$

(1) Adan 优化算法的更新规则

Adan 使用一阶和二阶动量进行梯度更新：

$$m(t) = \beta_1 m^{(t-1)} + (1 - \beta_1) g(t).$$

$$v(t) = \beta_2 v^{(t-1)} + (1 - \beta_2) g(t)^2.$$

对动量进行无偏修正：

$$\hat{m}^{(t)} = \frac{m^{(t)}}{1 - \beta_1^t}, \quad \hat{v}^{(t)} = \frac{v^{(t)}}{1 - \beta_2^t}.$$

最终的参数更新形式：

$$p_{t+1} = p_t - \eta \frac{\hat{m}^{(t)}}{\sqrt{\hat{v}^{(t)}} + \epsilon}.$$

(2) Adan 的收敛性的数学保证

由 Adan 优化算法的理论收敛性条件：

$$\sum_{t=1}^{\infty} \eta \frac{\hat{m}^{(t)}}{\sqrt{\hat{v}^{(t)}} + \epsilon} < \infty.$$

可推出：

$$\lim_{t \rightarrow \infty} \|p^{(t+1)} - p^{(t)}\| = 0.$$

即参数更新步长逐渐变小，最终收敛至最优值 p^* 和 θ^* 。

(3) clip 限制的作用

为了防止梯度爆炸或过小导致数值不稳定，参数更新时引入裁剪（clip）机制：

$$p_{t+1} = \max(0.1, \min(p_{t+1}, 1)).$$

因此，clip 只在极端情况下生效，不影响 Adan 的单调下降性，最终确保参数稳定收敛。

B.2 ADMM 优化变量 \mathcal{X} 的封闭解推导（针对公式 18）

本节推导 ADMM 方法中优化变量 \mathcal{X} 的封闭解，即：

$$\mathcal{X}^{(t+1)} = \arg \min_{\mathcal{X}} \sum_{k=1}^N \left\langle Q_k^{(t)}, \mathcal{X} - M_k^{(t+1)} \right\rangle + \frac{\beta}{2} \sum_{k=1}^N \|\mathcal{X} - M_k^{(t+1)}\|_F^2.$$

(1) 目标函数展开

第一项展开：

$$\sum_{k=1}^N \left\langle Q_k^{(t)}, \mathcal{X} - M_k^{(t+1)} \right\rangle = \sum_{k=1}^N \text{Tr} \left(Q_k^{(t)T} (\mathcal{X} - M_k^{(t+1)}) \right).$$

第二项展开：

$$\frac{\beta}{2} \sum_{k=1}^N \|\mathcal{X} - M_k^{(t+1)}\|_F^2 = \frac{\beta}{2} \sum_{k=1}^N \text{Tr} \left((\mathcal{X} - M_k^{(t+1)})^T (\mathcal{X} - M_k^{(t+1)}) \right).$$

(2) 对 \mathcal{X} 求导并令其为 0

$$\sum_{k=1}^N Q_k^{(t)} + \beta \sum_{k=1}^N (\mathcal{X} - M_k^{(t+1)}) = 0.$$

整理得：

$$\mathcal{X} = \frac{1}{N} \sum_{k=1}^N \left(M_k^{(t+1)} - \frac{1}{\beta} Q_k^{(t)} \right).$$

(3) 考虑观测约束

在观测位置 Ω , \mathcal{X} 应保持等于 T , 未观测位置 Ω^c 由上式确定, 因此：

$$\mathcal{X}^{(t+1)} = P_{\Omega^c}(T) + P_{\Omega} \left(\frac{1}{N} \sum_{k=1}^N \left(M_k^{(t+1)} - \frac{1}{\beta} Q_k^{(t)} \right) \right).$$

B.3 ADMM-TGST 迭代求解的收敛性证明（针对公式 21）

本节证明 ADMM-TGST 迭代求解方法的收敛性，即：

$$\|\mathcal{X}(t+1) - \mathcal{X}(t)\|_F < \epsilon, t \rightarrow \infty$$

(1) 证明误差递减

构造 Lyapunov 递减函数：

$$L^{(t)} = f(\mathcal{X}^{(t)}) + \sum_{k=1}^N \|Q_k^{(t)}\|_F^2.$$

如果：

1. $L^{(t)}$ 单调递减：

$$L^{(t+1)} \leq L^{(t)} - C \|\mathcal{X}^{(t+1)} - \mathcal{X}^{(t)}\|_F^2, \quad C > 0.$$

2. $L^{(t)}$ 有下界, 则由递减有界定理, 得到:

$$\lim_{t \rightarrow \infty} \|\mathcal{X}^{(t+1)} - \mathcal{X}^{(t)}\|_F = 0.$$

(2) 证明 ADMM 是收缩映射

ADMM-TGST 的更新公式:

$$\mathcal{X}^{(t+1)} = \frac{1}{N} \sum_{k=1}^N \left(M_k^{(t+1)} - \frac{1}{\beta} Q_k^{(t)} \right).$$

定义映射 $T(X)$ 为:

$$\mathcal{X}^{t+1} = T(\mathcal{X}^{(t)})$$

如果能找到 $0 < \lambda < 1$, 使得:

$$\|T(X1) - T(X2)\|_F \leq \lambda \|X1 - X2\|_F,$$

则由 Banach 不动点定理, 序列 $\mathcal{X}^{(t)}$ 必然收敛到某个唯一的不动点 X^* 。

综上, ADMM-TGST 迭代过程是收敛的。

3 附录 C 算法伪代码

Algorithm 1 Adaptive Update Mechanism

```

1: Input:  $p, \theta, E, m_p, v_p, m_\theta, v_\theta, \alpha, \{M_k\}_{k=1}^N, \beta_1, \beta_2, \eta, \gamma, \lambda, \epsilon$ 
2: Output:  $p, \theta, \alpha, m_p, v_p, m_\theta, v_\theta$ 
3: Initialize  $m_p, v_p, m_\theta, v_\theta \leftarrow 0$ 
4: while not converged do
5:   Compute gradient from consecutive iteration errors:
6:    $g \leftarrow E(t) - E(t-1)$ 
7:   Momentum-based update for  $p$ :
8:    $m_p \leftarrow \beta_1 m_p + (1 - \beta_1) g$ 
9:    $v_p \leftarrow \beta_2 v_p + (1 - \beta_2) g^2$ 
10:   $\hat{m}_p \leftarrow \frac{m_p}{1 - \beta_1^{t+1}}, \quad \hat{v}_p \leftarrow \frac{v_p}{1 - \beta_2^{t+1}}$ 
11:   $p \leftarrow \text{clip}\left(p + \eta \frac{\hat{m}_p}{\sqrt{\hat{v}_p} + \epsilon}, 0.1, 1\right)$ 
12:  Momentum-based update for  $\theta$ :
13:   $m_\theta \leftarrow \beta_1 m_\theta + (1 - \beta_1) (-g)$ 
14:   $v_\theta \leftarrow \beta_2 v_\theta + (1 - \beta_2) g^2$ 
15:   $\hat{m}_\theta \leftarrow \frac{m_\theta}{1 - \beta_1^{t+1}}, \quad \hat{v}_\theta \leftarrow \frac{v_\theta}{1 - \beta_2^{t+1}}$ 
16:   $\theta \leftarrow \text{clip}\left(\theta + \eta \frac{\hat{m}_\theta}{\sqrt{\hat{v}_\theta} + \epsilon}, 0, 1\right)$ 
17:  Weighted moving average for  $\alpha$ :
18:  for  $k = 1$  to  $N$  do
19:     $s_k \leftarrow \|M_k\|_F$ 
20:  end for
21:   $S \leftarrow \sum_{k=1}^N s_k + \epsilon$ 
22:  for  $k = 1$  to  $N$  do
23:     $\tilde{\alpha}_k \leftarrow \frac{s_k}{S} + \lambda \left( \frac{1}{N} - \frac{s_k}{S} \right)$ 
24:     $\alpha_k \leftarrow (1 - \gamma) \alpha_k + \gamma \tilde{\alpha}_k$ 
25:  end for
26:   $\alpha \leftarrow \frac{\alpha}{\sum_{k=1}^N \alpha_k}$ 
27: end while

```

Algorithm 2 ADMM-TGST for LRTC-ATSN

```
1: Input:  $\mathcal{T}$ ,  $\mathcal{P}_\Omega$ ,  $\alpha_k$ ,  $p$ ,  $\theta$ ,  $\beta$ ,  $\epsilon$ ,  $\text{max\_iter}$ 
2: Output:  $\mathcal{X}$ ,  $\text{MAPE\_list}$ ,  $\text{RMSE\_list}$ ,  $T_{\text{iter}}$ 
3: Initialize:
4:    $\mathcal{X}^{(0)} \leftarrow \mathcal{P}_\Omega(\mathcal{T}) + \mathcal{P}_{\bar{\Omega}}(\text{mean}(\mathcal{P}_\Omega(\mathcal{T})))$ 
5:    $\mathcal{M}_k^{(0)} \leftarrow 0$ ,  $\mathcal{Q}_k^{(0)} \leftarrow 0$ ,  $k = 1, 2, \dots, N$ 
6:    $m_p \leftarrow 0$ ,  $v_p \leftarrow 0$ ,  $m_\theta \leftarrow 0$ ,  $v_\theta \leftarrow 0$ 
7:    $\text{err\_list} \leftarrow \emptyset$ ,  $\text{MAPE\_list} \leftarrow \emptyset$ ,  $\text{RMSE\_list} \leftarrow \emptyset$ 
8: for  $t = 1$  to  $\text{max\_iter}$  do
9:    $\beta \leftarrow \beta \cdot (1 + \text{incre})$ 
10:  for  $k = 1$  to  $N$  do
11:     $\mathbf{Y}_k \leftarrow \mathcal{X}_{(k)}^{(t)} + \frac{1}{\beta} \mathcal{Q}_k^{(t)}$ 
12:    Compute SVD:  $\mathbf{Y}_k = \mathbf{U}_k \square_k \mathbf{V}_k^\top$ 
13:    Compute  $r_k \leftarrow \lceil \theta \cdot \min\{I_k, \prod_{i \neq k} I_i\} \rceil$ 
14:    Compute  $\delta_i = \text{GST}(\sigma_i, \frac{\alpha_k w_{k,i}}{\beta}, p)$ 
15:     $\mathcal{M}_k^{(t+1)} \leftarrow \text{Fold}_k(\mathbf{U}_k \square_k \mathbf{V}_k^\top)$ 
16:  end for
17:   $\mathcal{X}^{(t+1)} \leftarrow \mathcal{P}_\Omega(\mathcal{T}) + \mathcal{P}_{\bar{\Omega}}\left(\frac{1}{N} \sum_{k=1}^N (\mathcal{M}_k^{(t+1)} - \frac{1}{\beta} \mathcal{Q}_k^{(t)})\right)$ 
18:  for  $k = 1$  to  $N$  do
19:     $\mathcal{Q}_k^{(t+1)} \leftarrow \mathcal{Q}_k^{(t)} + \beta(\mathcal{X}^{(t+1)} - \mathcal{M}_k^{(t+1)})$ 
20:  end for
21:  Compute error:  $e^{(t+1)} \leftarrow \frac{\|\mathcal{X}^{(t+1)} - \mathcal{X}^{(t)}\|_F}{\|\mathcal{X}^{(t)}\|_F}$ 
22:  Append  $e^{(t+1)}$ ,  $\text{MAPE}(\mathcal{T}, \mathcal{X}^{(t+1)})$ ,  $\text{RMSE}(\mathcal{T}, \mathcal{X}^{(t+1)})$  to lists
23:  if  $e^{(t+1)} < \epsilon$  then
24:    break
25:  end if
26:  // Adaptive update for  $p$  and  $\theta$ 
27:   $(p, \theta, m_p, v_p, m_\theta, v_\theta) \leftarrow \text{adaptive\_update}(p, \theta, m_p, v_p, m_\theta, v_\theta, \eta)$ 
28:  // Smooth update for  $\alpha_k$ 
29:   $\alpha_k \leftarrow \text{update\_alpha}(\mathcal{M}, \alpha, \gamma, \lambda)$ 
30: end for
```
