

分类号：待学校审核确定

密级：公开

南方科技大学
博士后研究工作报告

模型和数据双驱动的低秩张量分解研究

Model- and Data-Driven Low-Rank Tensor Decomposition

博士后：龚文武

工作完成日期：2024 年 09 月–2026 年 09 月

报告提交日期：2026 年 09 月

南方科技大学（广东·深圳）

2026 年 09 月

模型和数据双驱动的低秩张量分解研究

MODEL- AND DATA-DRIVEN LOW-RANK TENSOR DECOMPOSITION

博士后姓名	龚文武
合作导师	杨丽丽教授
流动站（一级学科）	地球物理学博士后流动站
专业/研究领域	计算数学、低秩张量表示与可解释深度学习
研究工作起始时间	2024 年 09 月
研究工作期满时间	2026 年 09 月
单位	南方科技大学
报告提交日期	2026 年 09 月

摘 要

针对高光谱图像、交通数据等多维观测中的缺失、异常、噪声及跨模态耦合，本研究以“低秩表示如何参数化、表示变量及求解器参数如何由观测确定”为核心主线。第一项工作采用显式塔克参数化与固定目标求解：时空塔克分解插补方法（spatiotemporal Tucker decomposition approach for traffic imputation, TuckerAPP）在全尺寸塔克表示中联合核心稀疏、空间图拉普拉斯和时间差分约束；低秩塔克表示模型（low-rank Tucker representation model, LRTuckerRep）在降阶塔克表示中联合因子核范数、核心稀疏和图平滑，并以近端交替线性化极小化（proximal alternating linearized minimization, PALM）逐块求解固定目标；梯度正则张量分解模型（gradient-regularized tensor decomposition, GradientTD）进一步将局部先验移至重建张量的梯度域，并以变量分裂的交替方向乘子法（alternating direction method of multipliers, ADMM）求解。第二项工作把固定表示推进为可学习表示：三维深度低秩张量表示（three-dimensional deep low-rank tensor representation, 3DeepRep）参数化三个方向的低秩潜张量、深度变换与非线性聚合；低管秩连续表示（low-rank tubal representation, LRTubalRep）以低秩空间系数和共享连续基参数化第三模态，建立观测切片到整片缺失坐标的信息通路；通道感知张量核范数（channel-aware tensor nuclear norm, CA-TNN）学习潜在通道的正则分配。上述表示参数均通过观测区域自监督损失和反向传播求解，用以提高固定坐标对复杂退化的适应能力。第三项工作进一步参数化求解机制：可解释张量低秩近似网络（explainable tensor low-rank approximation network, ETLRA-Net）保留低秩因子、误差分裂、观测一致性和对偶状态，将解析更新、近端映射及可学习权重和逐层参数组织为有限层网络。由此，三项研究形成“显式表示与固定目标求解—可学习表示与观测自监督求解—显式状态与可学习优化求解”的递进关系，在提高表示适应性的同时保留求解过程的结构语义。拓展研究将这一“表示参数化—求解机制”思想延伸到时空预测、城市语义建模和鲁棒因果推荐相关任务中。

关键词：低秩张量；表示参数化；张量补全；连续坐标表示；观测自监督；算法展开；模型数据双驱动；可解释优化

Abstract

This report addresses missing entries, outliers, dense noise, and cross-mode coupling in multidimensional observations through a common question: how should the recovery tensor be parameterized, and how should the representation variables or solver parameters be determined from observations? The first line combines explicit Tucker parameterizations with fixed-objective optimization. TuckerAPP integrates core sparsity, graph-Laplacian spatial regularization, and temporal differences in a full-size Tucker representation. LRTuckerRep combines nuclear norms of reduced factors, core sparsity, and graph smoothness, and solves the resulting fixed objective by proximal alternating linearized minimization (PALM). GradientTD moves local regularization to the gradient domain of the reconstructed tensor and organizes its updates through a variable-splitting alternating direction method of multipliers (ADMM). The second line replaces fixed representations with learnable ones. 3DeepRep parameterizes three directional low-rank latent tensors, deep transforms, and nonlinear aggregation; LRTubalRep combines low-rank spatial coefficients with a shared continuous basis to transmit information from observed slices to wholly missing coordinates; and CA-TNN learns regularization allocation across latent channels. These representation parameters are fitted by observed-region self-supervision and backpropagation. The third line further parameterizes the optimization mechanism. The explainable tensor low-rank approximation network (ETLRA-Net) retains low-rank factors, error splitting, observation consistency, and dual states, while organizing analytic updates, proximal mappings, learnable weights, and layer-wise parameters into a finite-depth network. The three lines therefore progress from explicit representations with fixed-objective solvers, through learnable representations with observation-supervised optimization, to explicit optimization states with learnable update operators. Related extensions apply the same representation-parameterization and solution-mechanism perspective to spatiotemporal forecasting, urban semantic modeling, causal learning for robust recommendation.

KEY WORDS: low-rank tensor; representation parameterization; tensor completion; continuous coordinate representation; observation self-supervision; algorithm unrolling; model-data dual driving; interpretable optimization

目录

摘要	i
Abstract	iii
符号、缩略词和术语对照表	xiii
第一章 绪论	1
1.1 研究背景	1
1.2 研究问题与挑战	2
1.3 相关研究	2
1.3.1 模型驱动的张量建模	2
1.3.2 数据驱动的张量优化	3
1.3.3 基于算法展开的可解释性建模	3
1.4 研究思路与主要内容	4
1.4.1 三类驱动范式的统一关系	4
1.4.2 模型驱动的低秩表示	4
1.4.3 数据驱动的深度低秩表示	5
1.4.4 模型数据双驱动张量优化	6
1.5 主要贡献与研究价值	6
1.6 章节安排	6
第二章 低秩张量恢复理论基础与统一框架	9
2.1 多维观测与低秩逆问题	9
2.1.1 张量、展开与基本范数	9
2.1.2 观测算子与退化类型	10
2.1.3 可辨识性与模型失配	10
2.2 张量低秩表示与结构先验	11
2.2.1 多线性秩与降阶塔克表示	11
2.2.2 核心、因子与潜在通道正则	11
2.3 模型驱动的低秩表示	12
2.3.1 显式低秩参数化与领域先验	12
2.3.2 近端算子与块坐标更新	12

2.3.3	变量分裂与约束处理	13
2.4	数据驱动的深度低秩表示	13
2.4.1	从固定变换到可学习映射	13
2.4.2	连续坐标共享与整片缺失	14
2.4.3	通道感知的变换域正则	14
2.5	可解释模型数据双驱动优化	15
2.5.1	多先验变量分裂	15
2.5.2	有限层输出与迭代结论	15
2.6	评价原则与理论基础	15
2.6.1	评价指标与观测原则	15
2.6.2	理论基础与研究章节映射	16
2.7	本章小结	17
第三章	模型驱动的低秩表示	19
3.1	问题定义与统一表示	19
3.1.1	交通张量化与补全任务	19
3.1.2	塔克表示、展开与维度	20
3.2	时空正则的全尺寸塔克分解	20
3.2.1	建模动机与模型变量	20
3.2.2	空间图、时间约束与目标函数	21
3.2.3	逐块梯度、近端算子与步长	22
3.2.4	加速近端梯度算法	23
3.2.5	收敛条件与计算复杂度	24
3.2.6	实验结果	25
3.3	降阶低秩表示与梯度正则化	26
3.3.1	低秩塔克建模与初始化	27
3.3.2	互补正则与保真模型	28
3.3.3	近端交替线性化更新	29
3.3.4	收敛、辅助求解与计算代价	31
3.3.5	重建张量的离散梯度扩展	33
3.4	本章小结	35
第四章	数据驱动的深度低秩表示	37
4.1	数据驱动的低秩表示	37
4.2	多方向深度低秩表示	38
4.2.1	参数化与网络结构	41

4.2.2	联合目标与梯度求解	42
4.2.3	模型机制	43
4.3	多方向模型的实验结果	44
4.3.1	数据集、退化模式和基线	44
4.3.2	代表性数值结果	44
4.3.3	消融、稳定性与效率	45
4.3.4	真实退化实验与适用条件	46
4.4	面向整片缺失的连续坐标低秩表示	46
4.4.1	整片观测几何与零直接梯度	47
4.4.2	低秩空间系数、投影字典与张量形状	48
4.4.3	连续原子族与无偏置边界	49
4.4.4	目标函数与自适应矩估计求解	49
4.4.5	尺度规范化、收缩顺序与实现复杂度	50
4.4.6	训练与缺失切片推断	50
4.4.7	表示性质与秩界	51
4.5	面向潜在通道差异的自适应核范数	52
4.5.1	空间—光谱变换	53
4.5.2	联合目标与自监督求解	54
4.5.3	通道级加权实验	54
4.6	本章小结	55
第五章	模型数据双驱动张量优化	57
5.1	问题定义与求解器参数化动机	57
5.2	可解释张量低秩近似网络	58
5.2.1	观测模型与变量分裂	58
5.2.2	可逆通道变换与秩门控	59
5.2.3	四类先验与层级更新	59
5.2.4	有限层更新	60
5.3	本章小结	61
第六章	统一框架与拓展研究	63
6.1	三类驱动范式的技术递进	63
6.2	从离散低秩表示到连续结构表征	66
6.3	拓展研究：结构表征与可解释建模	66
6.3.1	多尺度时空预测	66
6.3.2	语义增强时空建模	67

6.3.3 因果稳健推荐	68
6.4 本章小结与后续项目衔接	71
第七章 结论与展望	73
7.1 主要结论	73
7.2 研究展望	75
参考文献	77
博士后个人简历	83
博士后期间科研成果	85
致谢	87
永久通信地址	89

插图

3.1	时空塔克分解插补模型的整体框架。	21
3.2	低秩塔克表示模型的因子、核心与图先验协同框架。	27
3.3	梯度正则张量分解模型的变量分裂与求解框架。	34
4.1	多方向深度低秩张量表示模型的结构框架。	40
4.2	低管秩连续坐标表示模型的共享参数与低秩空间系数框架。	47
4.3	通道感知张量核范数模型的空间-光谱变换与自适应加权框架。	53
5.1	可解释低秩优化展开网络的总体结构与单层流程。	58
6.1	三类低秩张量恢复范式的研究框架。灰色虚线表示统一模型在三类范式中的实例化，蓝色实线箭头依次表示可学习表示接口的扩展和受约束求解接口的参数化。	64

表格

1.1	三类驱动范式的表示、先验与求解关系。	4
2.1	共同数学基础在研究章节中的作用。	16
3.1	TuckerAPP 论文在 G/H 数据集上的代表性结果（原文表 3）。	26
3.2	LRTuckerRep 与 GradientTD 的模型构造对照。	27
4.1	数据驱动低秩表示	38
4.2	3DeepRep 的关键模型构造与实验结果。	43
4.3	3DeepRep 在 Pavia Centre 上的代表性结果（论文表 2）	45
4.4	3DeepRep 核心模块消融（Pavia Centre, Case 4）	45
4.5	3DeepRep 的方向和潜维度效率结果（Pavia Centre, Case 4）	46
4.6	LRTubalRep 所解决的问题与对应表示机制。	47
4.7	CA-TNN 论文的代表性结果（CoNoT 与 CA-TNN）。	55
4.8	CA-TNN 的通道加权与激励函数消融（Pavia, $MR = 0.95$ ）。	55
6.1	第 3–5 章低秩恢复方法的统一关系。	64

符号、缩略词和术语对照表

符号/缩写	中文名称	定义或说明
$\mathcal{X}$	待恢复张量	$\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$, 元素为 $x_{i_1 \dots i_N}$
$\mathcal{P}_\Omega$	观测投影	在观测索引集 Ω 上保留元素, 其余位置置零
$\times_n$	模态乘积	张量与第 n 模态矩阵的乘积
$\mathcal{G}, U_n$	塔克核心与因子	$\mathcal{X} = \mathcal{G} \times_1 U_1 \cdots \times_N U_N$
$\ \cdot\ _*$	核范数	矩阵奇异值之和; 用于因子或变换域切片
$\ \cdot\ _1, \ \cdot\ _F$	ℓ_1 范数、弗罗贝尼乌斯范数	分别刻画稀疏性和平方和能量
L_n	图拉普拉斯矩阵	$L_n = \Delta_n - W_n$, 其中 Δ_n 为度矩阵; 用于模态因子平滑
SNR	信噪比 (signal-to-noise ratio)	$20 \log_{10}(\ X\ _F / \ N\ _F)$, 单位为分贝 (dB)
PALM	近端交替线性化极小化	Proximal Alternating Linearized Minimization
D_n ; TV/QV	第 n 模有限差分; 总差分/二次变分	TV 为 total variation, QV 为 quadratic variation; 分别使用 $\ D_n X_{(n)}\ _1$ 与 $\ D_n X_{(n)}\ _F^2$
GradientTD	梯度正则塔克分解	在重建张量上施加离散梯度先验的低秩塔克模型
HPL-ADMM/FSL-ADMM	混合/完全分裂近端线性化交替方向乘子法	Hybrid proximal linearized ADMM / fully split linearized ADMM; GradientTD 的两种变量分裂求解构造
INR	隐式神经表示	Implicit Neural Representation
TNN/CA-TNN	变换域张量核范数/通道感知张量核范数	Transformed tensor nuclear norm / channel-aware tensor nuclear norm; 变换域切片核范数及其通道级数据依赖加权形式

第一章 绪论

1.1 研究背景

高光谱图像、交通流量、视频和网络服务记录都不是彼此独立的标量集合，而是沿空间、时间、光谱、传感器或用户等多个坐标轴组织起来的多维数据。以高光谱图像为例，同一空间位置上的光谱曲线具有连续性，不同空间位置又受到物质组成和场景结构的共同影响；以交通数据为例，传感器、日内时间和日期分别承载空间邻接、周期变化和长期趋势。若将这些数据直接展平为矩阵，虽然可以使用成熟的矩阵补全或去噪方法，但不同模态的语义关系会被压缩到同一坐标中，空间、时间和光谱结构难以分别建模。张量表示因此不仅是数据存储形式，也是区分多模态结构、设置先验和设计算法的基础。

设待恢复的 N 阶张量为

$$\mathcal{X}^* \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}, \quad \Omega \subseteq [I_1] \times \cdots \times [I_N],$$

其中 Ω 为可见条目索引集， $\mathcal{P}_\Omega$ 为观测投影。含有缺失、离群值和稠密噪声的观测可以统一写为

$$\mathcal{Y} = \mathcal{P}_\Omega(\mathcal{X}^* + \mathcal{E}^* + \mathcal{N}^*), \quad (1.1)$$

其中 $\mathcal{E}^*$ 表示少量但幅值可能较大的异常， $\mathcal{N}^*$ 表示较为稠密的随机扰动。低秩张量恢复的任务是在只知道 $(\mathcal{Y}, \Omega)$ 的条件下估计 $\hat{\mathcal{X}}$ ，使其既与可见观测相容，又能延续数据在未观测坐标上的全局关联和局部变化规律。

这类恢复问题具有明确的应用价值。高光谱图像中的坏列、缺失波段和随机采样会同时破坏空间纹理与光谱曲线，恢复质量直接影响后续分类、目标识别和异常检测；交通张量中的传感器失效、通信丢包和局部拥堵会造成随机缺失、连续时间段缺失甚至全局停摆，补全结果将影响交通状态估计和预测。医学影像、遥感和科学计算中的欠采样也具有类似特点：观测不足并不意味着真实对象没有结构，而是需要利用正确的结构坐标把有限观测传播到不可见位置。

本报告用表示参数化概括从一组有限维变量或函数参数生成恢复张量的映射，例如塔克核心与因子、低秩空间系数与连续坐标基。用求解机制概括由 $(\mathcal{Y}, \Omega)$ 确定这些表示变量或算法参数的更新规则，包括固定目标上的块近端迭代、观测自监督的梯度优化，以及保留显式分裂状态、解析子问题和对偶变量的有限层展开。低秩恢复的性能不仅取决于采用何种秩概念，也取决于表示参数化是否匹配观测几何、求解机制是否能够稳定地把观测信息传递到该表示中。

1.2 研究问题与挑战

式 (1.1) 所描述的困难不是单纯的缺失率问题。缺失模式、异常位置、噪声强度以及不同模态之间的耦合共同决定了恢复问题的病态程度。特别是连续条带、坏列和整片切片缺失会切断某些坐标与观测保真项之间的直接联系；若仍把每个条目作为独立参数，则缺失位置没有可用于更新的拟合梯度。因此，本报告围绕以下相互联系的技术挑战展开。

- (1) **低秩恢复应如何参数化。**低秩并不是一个与任务无关的单一数字。塔克核心与因子、变换域切片、连续坐标函数和分裂状态分别承载跨模态交互、方向相关性、跨坐标共享和误差分解。需要明确恢复映射的输入、参数和输出，并判断低秩、图平滑、时间差分或梯度能量作用于哪个参数化变量。
- (2) **表示参数如何由不完整观测求解。**显式分解和固定目标具有清楚的数学含义，但预设坐标、变换方向和正则形式可能无法覆盖异质结构及整片缺失。需要在保持显式低秩对象和观测投影的同时，让方向变换、连续坐标基、聚合规则或通道权重参与学习，使观测梯度能够沿共享参数传播到固定表示难以覆盖的位置。
- (3) **可学习求解机制如何保持算法语义。**当全部表示参数由通用梯度优化器联合更新时，状态、阈值、约束和先验作用位置不再天然对应可检查的子问题。在缺失、异常和噪声并存时，需要进一步参数化求解器本身：保留低秩因子、误差变量、近端更新、观测投影和对偶状态，只让权重、步长或近端势等接口由数据调节。

上述挑战形成一条由表示到求解的递进链：第 3 章在显式塔克参数化上建立固定目标及其块近端求解；第 4 章把深度变换、连续坐标基和通道权重纳入表示参数化，并以观测区域自监督梯度求解；第 5 章进一步把一次优化迭代参数化为有限层状态转移，在保留解析更新和约束语义的同时学习部分更新接口。模型驱动、数据驱动和模型数据双驱动由此成为这条“表示参数化-求解机制”主线的三个层次，而不是相互割裂的方法标签。

1.3 相关研究

1.3.1 模型驱动的张量建模

塔克分解以核心张量和模态因子表达多线性关系，是保留模态语义的经典张量模型^[1,2]。在张量补全中，低秩约束与观测一致性共同用于估计缺失元素，张量核范数等凸替代项为可计算的低秩优化提供了基础^[3]。对于三阶张量，张量奇异值分解（tensor singular value decomposition, t-SVD）在指定变换域的切片上进行矩阵分解，变换域张量核范数（transformed tensor nuclear norm, TNN）能够表达沿特定模态组织的

相关性^[4, 5]。

模型驱动方法的优势在于结构对象和求解变量明确：核心稀疏可以刻画有效的跨模态交互，因子核范数可以压缩潜在子空间，图拉普拉斯项可以约束相似节点的因子表示，差分或总变分（total variation, TV）项可以描述有序坐标上的局部变化。低秩与平滑联合建模已被用于张量恢复，并可在一定条件下建立恢复分析^[6]。然而，固定塔克秩、固定变换和固定图先验也带来适配性限制：当缺失几何改变、局部关系不再由预设图表示，或者平滑应直接作用于重建张量而非潜在因子时，单一模型难以同时保持表示容量和局部细节。第 3 章因此从全尺寸塔克模型出发，进一步研究降阶塔克表示及其梯度域扩展。

1.3.2 数据驱动的张量优化

深度图像先验（deep image prior, DIP）说明，未经监督训练的卷积网络参数化本身可以提供图像结构偏置^[7]；隐式神经表示（implicit neural representation, INR）则用坐标到信号值的连续函数表达图像、场或时空信号。正弦激活和傅里叶特征为坐标函数提供了高频表达能力^[8, 9]，层次低秩张量因子化、函数变换低秩分解和函数型张量奇异值分解进一步把连续函数与多维低秩结构结合起来^[10-12]。

数据驱动表示能够从观测中学习方向变换、连续坐标共享、非线性聚合或通道权重，从而缓解固定模型的表示失配。例如，多方向表示可以在一条模态受到条带或坏列破坏时利用其他方向的互补结构；共享连续参数可以使观测坐标的梯度传递到整片缺失坐标；通道级权重可以区分潜在通道的整体信息量。但这类模型通常含有非线性网络、多线性耦合或非凸正则，训练状态不再天然对应某个逐块优化步骤。第 4 章沿“方向协同一坐标共享一通道自适应”展开，保留显式低秩变量，并分别处理复杂退化、整片缺失和潜在通道差异。

1.3.3 基于算法展开的可解释性建模

近端方法能够将稀疏性、核范数和约束投影转化为软阈值、奇异值收缩或投影操作；近端交替线性化极小化（proximal alternating linearized minimization, PALM）为非凸、非光滑多块模型提供了条件性临界点收敛分析^[13]。交替方向乘子法（alternating direction method of multipliers, ADMM）通过副本变量、共识变量和对偶变量分离不同先验，适合处理观测一致性、低秩结构和局部平滑的耦合约束^[14]。算法展开把有限次迭代映射为有限层网络，在保留优化变量语义的同时学习步长、阈值或融合参数；其基本思想和 ADMM 展开实例分别可见相关综述与代表性工作^[15, 16]。

算法展开解决的是“如何学习求解路径”，而不是简单增加网络层数。若每一层仍保留干净张量、异常、噪声、先验副本和对偶状态，则网络可以说明观测误差如何被分解、不同先验如何形成共识以及参数如何影响更新；若完全取消这些变量，输出性能即使提高，也难以解释信息来自何处。第 5 章据此构造可解释张量低秩近似网

络 (explainable tensor low-rank approximation network, ETLRA-Net), 把模型驱动层确定的先验算子和数据驱动层提供的适应能力接入同一有限层求解框架。

1.4 研究思路与主要内容

1.4.1 三类驱动范式的统一关系

本报告以统一观测模型为入口, 把低秩恢复写成恢复映射 $\hat{\mathcal{X}} = \mathcal{R}_\theta(\mathcal{Y}, \Omega)$ 的参数化与求解问题。模型驱动、数据驱动和模型数据双驱动构成宏观研究路线; 在每一类范式内部, 则分别考察恢复张量如何参数化、结构先验如何约束参数, 以及表示变量或求解器参数如何由观测确定。三类范式的统一关系如表 1.1 所示。

表 1.1: 三类驱动范式的表示、先验与求解关系。

研究范式	表示参数化	先验及模型与数据的作用边界	求解机制
模型驱动	塔克核心、因子矩阵、重建张量或变换域切片	核心稀疏、因子低秩、图平滑、时间差分 and 梯度能量由任务模型预先确定	在固定目标上采用近端梯度、PALM 或变量分裂 ADMM
数据驱动	显式潜张量、深度方向变换、连续坐标场和潜在通道	保留低秩对象与观测投影, 由数据学习方向组织、坐标共享、聚合规则或通道权重	通过观测区域自监督损失和反向传播联合拟合表示参数
模型数据双驱动	干净张量、误差分量、低秩因子、先验副本和对偶状态	模型固定观测方程、状态语义和更新顺序, 数据调节置信权重、逐层参数或结构化近端映射	将解析、近端、投影和对偶更新组织为可学习的有限层求解器

三类范式沿两条相互配合的线索递进: 表示参数化由显式代数模型扩展到可学习函数族, 求解机制由固定目标块更新扩展到观测自监督梯度, 再发展为显式状态约束下的可学习有限层更新。围绕这一关系, 博士后阶段形成以下三项研究内容。

1.4.2 模型驱动的低秩表示

第 3 章研究显式低秩参数化及固定目标求解。已发表的时空塔克分解插补方法 (spatiotemporal Tucker decomposition approach for traffic imputation, TuckerAPP)^[17] 以交通张量为对象, 在全尺寸塔克坐标中用核心 ℓ_1 稀疏表达有效交互, 并通过空间图拉普拉斯和时间差分约束传感器与时间片的局部关系; 交替近端梯度、外推和观测校正将这些先验转化为可执行的补全算法。第 3 章对 TuckerAPP 按论文给出的数据集、缺失协议和基线重点呈现实验结果, 说明空间—时间先验在不同缺失几何下的作用范围。

低秩塔克表示模型 (low-rank Tucker representation model, LRTuckerRep)^[18] 将全尺寸坐标改为降阶塔克表示, 用因子核范数控制模态子空间的有效维数, 用核心稀

疏选择跨模态交互，并用固定图拉普拉斯表达潜在因子的邻域一致性。其研究重点是固定有限观测保真目标、核心软阈值与因子奇异值收缩的 PALM 更新，以及在块梯度 Lipschitz、迭代有界和 Kurdyka-Łojasiewicz (KL) 条件下的临界点收敛结论。梯度正则张量分解模型 (gradient-regularized tensor decomposition, GradientTD) 保留 LRTuckerRep 的全局低秩骨架，仅将局部先验从潜在因子图转移到重建张量的离散梯度，并通过变量分裂构造 TV/QV 近端更新和近端线性化 ADMM 求解路径。两者共同回答“低秩结构由哪些变量承载”，但分别强调潜在坐标邻域一致性和数据域相邻变化控制。

这一研究内容的必要性在于同时建立表示和求解基线：显式核心、因子与重建变量回答恢复张量由哪些有限变量生成，固定目标及其近端更新回答这些变量如何从观测中求得。没有这两个基线，后续方法便难以区分改进来自表示参数化的扩展，还是来自求解机制的改变。第 3 章由此为第 4 章和第 5 章提供参数化对象、正则接口和算法参照。

1.4.3 数据驱动的深度低秩表示

第 4 章研究可学习低秩参数化及观测自监督求解，沿“方向协同—坐标共享—通道自适应”逐层推进。三维深度低秩张量表示 (three-dimensional deep low-rank tensor representation, 3DeepRep)^[19] 为三个模态方向分别设置显式低秩潜张量和深度耦合变换，再以非线性聚合器融合候选重建。针对单一方向在连续坏列、条带或混合退化下信息覆盖不足的问题，该模型学习跨方向的信息组织方式；第 4 章按照论文设置报告其主实验、重复实验和消融结果，重点说明显式潜变量、三方向变换和聚合模块分别承担的作用。

低管秩连续表示 (low-rank tubal representation, LRTubalRep) 面向整片切片缺失时离散坐标没有直接训练梯度的问题，将一个模态的离散切片改写为共享参数的矩阵值连续场，并用连续原子、投影矩阵和低秩空间系数分别控制坐标函数、空间系数和组合容量。其重点是说明观测坐标如何通过共享参数影响缺失坐标，以及内部插值与边界外推在信息来源上的差异。通道感知张量核范数 (channel-aware tensor nuclear norm, CA-TNN)^[20] 将可学习正权作用于整个潜在通道的核范数，而不是仅对单个奇异值按序加权；第 4 章在已有随机缺失实验中报告其通道级加权和激励函数消融结果。三项工作都保留可正则化的低秩潜变量，但把数据适应能力分别放在方向变换、坐标函数和通道正则分配上。

这一研究内容的必要性在于揭示固定参数化难以覆盖的三种适配需求：方向受损时需要跨方向协同，整片坐标缺失时需要共享函数参数，潜在信息不均衡时需要通道级正则分配。数据驱动表示因此不是取消低秩结构，而是扩展恢复映射的参数化族，并用观测区域损失统一求解网络、坐标基和低秩系数；这也暴露出通用梯度求解难以逐层解释约束与误差分解的问题，为第 5 章参数化求解器提供直接动机。

1.4.4 模型数据双驱动张量优化

第 5 章研究显式状态约束下的求解机制参数化及可学习有限层更新。ETLRA-Net 在缺失、稀疏异常和稠密噪声共同存在时，显式区分干净张量、异常变量、噪声变量、低秩因子、先验副本和对偶状态。每层更新由观测投影、异常软阈值、噪声收缩、低秩因子更新、先验近端映射和对偶更新组成；模型固定状态含义和更新顺序，数据仅调节置信权重、步长、罚参数或具有结构限制的近端映射。该构造解决的是“如何把数据适应能力限制在可解释的求解接口内”，不把有限层网络直接解释为无限迭代算法的全局最优解。

ETLRA-Net 是前两项研究在求解参数化上的延伸：第 3 章给出显式表示与固定求解器，第 4 章扩展可学习表示族，第 5 章保持表示和约束状态可读，同时让求解器的部分接口随数据调整。由此，三项内容形成“参数化表示对象-扩大表示适应性-参数化求解机制”的完整递进。

1.5 主要贡献与研究价值

本报告的研究价值集中在表示参数化与求解机制的协同设计。第一，报告区分全尺寸与降阶塔克参数化、变换域切片参数化和连续坐标参数化，明确不同低秩对象如何生成恢复张量。第二，报告说明图、时间差分、数据域梯度、方向互补和通道权重如何进入相应参数化，而不是把不同秩概念和正则对象混为一谈。第三，报告把求解机制从固定目标上的近端梯度、PALM 和变量分裂 ADMM，推进到观测自监督反向传播，再推进到显式状态约束下的有限层展开，使表示能力的变化与算法更新的变化可以分别讨论。

三项研究由此构成一个闭环：模型驱动方法提供显式参数化与固定求解基线，数据驱动方法扩大低秩表示族并提高对复杂退化和非连续坐标的适应性，模型数据双驱动方法进一步参数化有限步更新而不放弃显式状态和约束。该闭环既解释了第 3-5 章为何递进安排，也为第 6 章的拓展研究以及后续连续低秩结构能量的表征、求解和医学图像应用提供共同方法基础。

1.6 章节安排

全文共七章，各章之间的关系如下：

- 第 1 章： 本章为绪论，说明多维数据恢复的应用背景、病态逆问题、技术挑战、相关研究和三项研究内容的递进关系。
- 第 2 章： 第二章介绍低秩张量恢复的理论基础与统一框架，定义张量运算、观测投影、主要低秩参数化、结构正则、近端方法、变量分裂及算法展开，并建立第 3-5 章“表示参数化-求解机制”的基础映射。
- 第 3 章： 第三章讨论模型驱动的低秩表示；其技术内核是显式塔克参数化与固定目

标求解，依次研究 TuckerAPP 的时空塔克分解、LRTuckerRep 的降阶塔克模型和 GradientTD 的梯度域扩展。

第 4 章：第四章讨论数据驱动的深度低秩表示；其技术内核是可学习低秩参数化与观测自监督求解，依次研究 3DeepRep 的三方向协同、LRTubalRep 的连续坐标共享和 CA-TNN 的通道自适应。

第 5 章：第五章讨论模型数据双驱动张量优化；其技术内核是显式状态约束下的求解器参数化与可学习有限层更新，介绍 ETLRA-Net 的误差分解、低秩因子、先验分裂、解析或近端更新、对偶状态和算法展开。

第 6 章：第六章总结统一框架与拓展研究，说明低秩建模如何从离散结构表示延伸到连续结构能量，并将结构表征和可解释建模迁移到多尺度时空预测、城市语义建模和因果稳健学习等方向。

第 7 章：第七章归纳主要结论，并围绕“连续低秩结构能量的表征—求解—医学图像应用”提出后续研究展望。

上述章节安排落实表 1.1 的统一关系：第 3–5 章分别回答“如何显式参数化并求解”“如何扩大表示族并由观测拟合”以及“如何参数化求解器且保留算法语义”，第 6 章和第 7 章再将这一主线延伸到拓展研究与后续项目。

第二章 低秩张量恢复理论基础与统一框架

第 3–5 章围绕同一个逆问题逐层展开：从不完整、含异常或含噪的多维观测中恢复具有低维结构的张量。低秩假设压缩全局自由度，空间、时间、光谱或上下文先验描述局部关系，优化算法则把这些结构转化为可计算的更新。三章之间的核心差异不在于是否使用低秩，而在于恢复张量采用何种表示参数化，以及表示变量或求解器参数通过何种机制由观测确定。

本章建立后续章节所需的共同数学符号。首先定义张量、展开、观测投影和低秩表示；其次区分核心、因子、变换域切片、重建张量及连续坐标函数上的参数化与先验；随后介绍固定目标近端求解、观测自监督梯度求解、变量分裂和算法展开；最后把上述概念组织为“显式表示与固定目标求解–可学习表示与观测自监督求解–显式状态与可学习有限层求解”的递进框架。

2.1 多维观测与低秩逆问题

2.1.1 张量、展开与基本范数

记 $[I] = \{1, \dots, I\}$ ， N 阶实张量写作

$$\mathcal{X} \in \mathbb{R}^{I_1 \times \dots \times I_N}, \quad x_{i_1 \dots i_N} = [\mathcal{X}]_{i_1 \dots i_N}.$$

第 n 模展开 $X_{(n)}$ 将第 n 个下标作为行，其余下标按照固定的字典序排列为列；向量化记为 $\text{vec}(\mathcal{X})$ 。对同型张量，内积和 Frobenius 范数分别为

$$\langle \mathcal{X}, \mathcal{Z} \rangle = \sum_{i_1=1}^{I_1} \dots \sum_{i_N=1}^{I_N} x_{i_1 \dots i_N} z_{i_1 \dots i_N}, \quad \|\mathcal{X}\|_F = \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle}.$$

矩阵核范数 $\|A\|_*$ 是矩阵奇异值之和，元素级 ℓ_1 范数为 $\|\mathcal{X}\|_1 = \sum_{i_1, \dots, i_N} |x_{i_1 \dots i_N}|$ 。前者常用来构造低秩替代项，后者常用来刻画稀疏异常或稀疏核心；它们分别作用在什么变量上，是后续模型含义的重要组成部分。

第 n 模态乘积定义为

$$(\mathcal{X} \times_n U)_{i_1 \dots i_{n-1} j i_{n+1} \dots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 \dots i_N} u_{j i_n},$$

其中 $U \in \mathbb{R}^{J \times I_n}$ ，所得张量的第 n 模长度变为 J 。相应的展开关系为 $(\mathcal{X} \times_n U)_{(n)} = U X_{(n)}$ 。Kronecker 积记为 $\otimes$ ，其作用是把多个模态上的线性变换写成一个矩阵乘积。

2.1.2 观测算子与退化类型

令 $\mathcal{X}^*$ 为真实张量, $\Omega \subseteq [I_1] \times \cdots \times [I_N]$ 为观测索引集。观测投影 $\mathcal{P}_\Omega$ 保留 Ω 上的元素, 并将其余元素置零。含缺失、稀疏异常和稠密噪声的统一观测模型为

$$\mathcal{Y} = \mathcal{P}_\Omega(\mathcal{X}^* + \mathcal{E}^* + \mathcal{N}^*), \quad [\mathcal{P}_\Omega(\mathcal{Z})]_{i_1 \dots i_N} = \begin{cases} z_{i_1 \dots i_N}, & (i_1, \dots, i_N) \in \Omega, \\ 0, & (i_1, \dots, i_N) \notin \Omega. \end{cases} \quad (2.1)$$

其中 $\mathcal{E}^*$ 表示少量但幅值可能较大的异常, $\mathcal{N}^*$ 表示分布较为稠密的随机扰动。仅有缺失时令两者均为零; 需要区分异常和噪声时, 三者分别承担信号、离群项和随机扰动的角色。经典张量补全以观测一致性和低秩约束共同估计缺失元素^[3]。

观测几何不仅由观测比例决定, 还由缺失位置的组织方式决定。随机条目缺失通常在各模态保留局部信息, 而连续条带、坏列和整片切片缺失会切断某些坐标与数据保真项之间的直接联系。因而, 随机缺失、结构化缺失和整片模态缺失需要使用不同的信息传播机制; 第 3 章以图和差分先验刻画已知邻接, 第 4 章进一步引入多方向表示与连续坐标共享。

在统一观测模型下, 一个一般的结构化恢复问题可写为

$$\min_{\mathcal{X}, \mathcal{E}, \mathcal{N}, \Theta} \frac{1}{2} \|\mathcal{P}_\Omega(\mathcal{X} + \mathcal{E} + \mathcal{N} - \mathcal{Y})\|_F^2 + \lambda_E \|\mathcal{E}\|_1 + \frac{\lambda_N}{2} \|\mathcal{N}\|_F^2 + \mathcal{R}(\mathcal{X}, \Theta), \quad (2.2)$$

其中 Θ 表示分解因子、变换参数、连续坐标网络或展开层参数, $\mathcal{R}$ 汇总低秩、图平滑、梯度平滑、连续性和上下文一致性等结构项。式 (2.2) 的三个基本接口是: 结构变量放在哪里, 先验作用于哪个变量, 以及观测如何约束这些变量。

2.1.3 可辨识性与模型失配

对给定参数域 $\mathcal{D}$, 由表示映射 $\mathcal{X}(\Theta)$ 生成的模型类定义为

$$\mathcal{M} = \{\mathcal{X}(\Theta) : \Theta \in \mathcal{D}\}.$$

无噪声时, 观测算子在 $\mathcal{M}$ 上可辨识, 至少需要满足

$$\mathcal{X}_1, \mathcal{X}_2 \in \mathcal{M}, \quad \mathcal{P}_\Omega(\mathcal{X}_1) = \mathcal{P}_\Omega(\mathcal{X}_2) \implies \mathcal{X}_1 = \mathcal{X}_2. \quad (2.3)$$

该条件把秩、采样拓扑、图连通性和参数化容量联系起来: 模型类过大而观测过少时, 可能存在多个观测一致的张量; 模型类过小而与真实结构不匹配时, 即使优化充分, 也可能产生表示偏差。用

$$\text{dist}(\mathcal{X}^*, \mathcal{M}) = \inf_{\mathcal{X} \in \mathcal{M}} \|\mathcal{X} - \mathcal{X}^*\|_F$$

表示模型失配程度, 可以分别讨论采样条件、表示逼近和噪声扰动对恢复结果的影响。第 3 章通过预设分解和先验缩小模型类, 第 4 章学习表示映射以提高适应性, 第 5 章则学习有限步求解轨迹; 三者对应的是不同的可学习接口, 而不是同一个误差结论的三种写法。

2.2 张量低秩表示与结构先验

2.2.1 多线性秩与降阶塔克表示

张量的低秩依赖于所选择的分解方式。对 N 阶张量，塔克多线性秩定义为

$$\text{rank}_T(\mathcal{X}) = (\text{rank}(X_{(1)}), \dots, \text{rank}(X_{(N)})). \quad (2.4)$$

它记录各模展开矩阵的列空间维数^[1,2]。给定秩向量 $(r_1, \dots, r_N)$ ，塔克表示为

$$\mathcal{X} = \mathcal{G} \times_{n=1}^N U_n, \quad \mathcal{G} \in \mathbb{R}^{r_1 \times \dots \times r_N}, \quad U_n \in \mathbb{R}^{I_n \times r_n}. \quad (2.5)$$

其中 $\mathcal{G}$ 为核心张量， U_n 为第 n 模因子矩阵；展开后有

$$X_{(n)} = U_n G_{(n)} V_n^\top, \quad V_n = U_N \otimes \dots \otimes U_{n+1} \otimes U_{n-1} \otimes \dots \otimes U_1.$$

当 $r_n \leq I_n$ 且至少一个模态满足 $r_n < I_n$ 时，本报告称式 (2.5) 为降阶塔克表示。该术语强调表示坐标相对于原始模态被压缩；若所有 $r_n = I_n$ ，则称为全尺寸塔克表示。核心稀疏性、因子矩阵的低秩性和预设秩分别控制跨模态交互数量、潜在子空间维数和参数容量，不能互换为同一个秩指标。

三阶张量还可以在指定变换域考察切片秩。经典张量奇异值分解 (tensor singular value decomposition, t-SVD) 通常沿第三模态使用离散傅里叶变换，并在变换域的正面切片上进行矩阵分解^[4]。在一般可逆变换 $\mathcal{T}$ 下，令 $\bar{\mathcal{Z}} = \mathcal{T}(\mathcal{Z})$ ，变换域张量核范数 (transformed tensor nuclear norm, TNN) 可写为

$$\|\mathcal{Z}\|_{\text{TNN}, \mathcal{T}} = \sum_{k=1}^{\tilde{I}_3} \|\bar{\mathcal{Z}}^{(k)}\|_*. \quad (2.6)$$

只有在变换、切片方向和秩定义均明确时，才能讨论该变换诱导的低秩；它与塔克多线性秩或经典傅里叶域管秩描述的是不同结构。

2.2.2 核心、因子与潜在通道正则

塔克表示的不同变量承载不同层次的结构。核心张量记录潜在模态子空间之间的交互，可用 $\|\mathcal{G}\|_1$ 选择少量有效耦合；因子矩阵的列空间描述每个模态的潜在子空间，可用 $\|U_n\|_*$ 压缩其有效维数。因此，常见的互补正则写为

$$\mathcal{R}_{\text{Tucker}}(\mathcal{G}, \{U_n\}) = \alpha \|\mathcal{G}\|_1 + (1 - \alpha) \sum_{n=1}^N \omega_n \|U_n\|_*, \quad 0 \leq \alpha \leq 1.$$

该组合不是塔克秩的凸包，而是针对塔克坐标的一种可计算替代。因子与核心同时带正权时，还可以缓解单独缩放某个因子、再反向缩放核心所造成的数值尺度逃逸，但分解仍可能存在旋转和置换不唯一性。

如果模态的样本之间存在相似度关系, 令 W_n 为相似度矩阵、 $\Delta_n = \text{diag}(W_n \mathbf{1})$ 为度矩阵, 图拉普拉斯矩阵为 $L_n = \Delta_n - W_n \succeq 0$, 则因子图平滑项满足

$$\text{tr}(U_n^\top L_n U_n) = \frac{1}{2} \sum_{i,j} W_{n,ij} \|u_{n,i} - u_{n,j}\|_2^2. \quad (2.7)$$

式 (2.7) 约束的是相似样本在潜在因子空间中的差异, 不等同于直接平滑原始张量。图的节点、边和权重应由可见观测或外部先验确定; 有序网格或时间轴则可以直接采用邻接差分。

对于具有可信顺序的模态, 记第 n 模有限差分矩阵为 D_n 。若使用非周期前向差分, 则 $(D_n x)_i = x_{i+1} - x_i$; 在周期边界下还需加入首尾差分。重建张量上的总变分 (total variation, TV) 和二次变分 (quadratic variation, QV) 可写为

$$\text{TV}_n(\mathcal{X}) = \|D_n X_{(n)}\|_1, \quad \text{QV}_n(\mathcal{X}) = \|D_n X_{(n)}\|_F^2. \quad (2.8)$$

TV 倾向于保留分段变化和突变边界, QV 对较大梯度施加二次惩罚, 更偏向连续平滑变化。低秩项控制全局自由度, 图或差分项控制局部变化; 低秩与平滑联合恢复的理论和模型构造可参见相关研究^[6]。第 3 章中的 LRTuckerRep 主要在潜在因子上使用图二次型, GradientTD 则把局部先验转移到重建张量的差分域。

2.3 模型驱动的低秩表示

2.3.1 显式低秩参数化与领域先验

模型驱动方法先规定表示坐标, 再把领域知识写成目标函数中的正则项或约束。以式 (2.5) 为基础, 观测一致性模型可以抽象为

$$\min_{\mathcal{G}, \{U_n\}} \frac{\lambda}{2} \|\mathcal{P}_\Omega(\mathcal{G} \times_{n=1}^N U_n - \mathcal{Y})\|_F^2 + \mathcal{R}_{\text{core}}(\mathcal{G}) + \mathcal{R}_{\text{factor}}(\{U_n\}) + \mathcal{R}_{\text{local}}. \quad (2.9)$$

不同研究的差异由三项正则的承载变量决定。已发表的时空塔克分解插补方法 (TuckerAPP) 采用全尺寸塔克表示, 以核心 ℓ_1 稀疏表达交通数据中的有效交互, 并在因子上加入空间图和时间差分约束^[17]; 低秩塔克表示模型 (low-rank Tucker representation model, LRTuckerRep) 采用降阶塔克表示, 同时约束因子核范数、核心稀疏和潜在因子图平滑^[18]; 梯度正则张量分解模型 (gradient-regularized tensor decomposition, GradientTD) 保留降阶塔克低秩骨架, 但把局部先验直接作用于重建张量的 TV 或 QV 能量。三者共享明确的表示变量, 却分别对应全尺寸容量控制、潜在子空间压缩和数据域局部变化控制。

2.3.2 近端算子与块坐标更新

对适当函数 g , 近端映射定义为

$$\text{prox}_{\tau g}(V) = \arg \min_Z \left\{ g(Z) + \frac{1}{2\tau} \|Z - V\|_F^2 \right\}. \quad (2.10)$$

元素软阈值和奇异值阈值 (singular value thresholding, SVT) 分别对应 ℓ_1 范数和核范数的典型近端步骤。若 $V = P \text{diag}(\sigma_i) Q^\top$, 则

$$\mathcal{S}_\tau(v) = \text{sign}(v)(|v| - \tau)_+, \quad \text{SVT}_\tau(V) = P \text{diag}((\sigma_i - \tau)_+) Q^\top. \quad (2.11)$$

非负约束则由逐元素投影 $\mathcal{P}_+(u) = \max\{u, 0\}$ 实现。对于多线性保真项, 固定其余变量后可以逐块计算核心或因子的梯度, 并用回溯主化系数保证当前块的二次上界; 这正是第 3 章交替近端梯度和近端交替线性化极小化 (proximal alternating linearized minimization, PALM) 更新的基础。

若固定目标函数由可微耦合项与非光滑块正则组成, 且满足块梯度局部 Lipschitz、迭代序列有界、每个主化系数严格大于相应 Lipschitz 上界, 并具有 Kurdyka-Łojasiewicz (KL) 性质, 则 PALM 可在这些条件下把充分下降和相对误差转化为临界点收敛结论^[13]。该结论针对固定目标和理想近端更新, 不自动覆盖在线重构图、在线改变权重、任意惯性外推或把近端算子替换成未受约束的网络模块。

2.3.3 变量分裂与约束处理

当观测一致性、塔克联结和局部平滑不能由同一个闭式更新处理时, 可为不同先验引入副本变量。例如令 $\mathcal{Z}_j$ 负责第 j 类结构, 令 $\mathcal{X}$ 汇聚各副本, 则共识形式为

$$\begin{aligned} \min_{\mathcal{X}, \mathcal{E}, \mathcal{N}, \{\mathcal{Z}_j\}} \quad & \lambda_E \|\mathcal{E}\|_1 + \frac{\lambda_N}{2} \|\mathcal{N}\|_F^2 + \sum_{j \in \mathcal{J}} \rho_j \Phi_j(\mathcal{Z}_j), \\ \text{s.t.} \quad & \mathcal{P}_\Omega(\mathcal{X} + \mathcal{E} + \mathcal{N} - \mathcal{Y}) = 0, \quad \mathcal{Z}_j = \mathcal{X}. \end{aligned} \quad (2.12)$$

交替方向乘子法 (alternating direction method of multipliers, ADMM) 通过增广拉格朗日项分别更新副本、共识变量和对偶变量^[14]。GradientTD 的梯度变量分裂以及第 5 章的多先验副本, 均可视为这一接口的不同实例。模型驱动层由此得到一条清晰的计算路径: 先固定低秩参数化、先验和目标函数, 再用近端、投影或乘子更新求解表示变量。

2.4 数据驱动的深度低秩表示

2.4.1 从固定变换到可学习映射

数据驱动方法保留显式低秩变量, 但允许表示映射、方向融合或正则权重由观测参与学习。设 $\mathcal{Z}$ 为潜在张量, g_θ 为从潜在域到数据域的映射, 则一般形式为

$$\hat{\mathcal{X}} = g_\theta(\mathcal{Z}), \quad \min_{\mathcal{Z}, \theta} \left\| \mathcal{P}_\Omega(\hat{\mathcal{X}} - \mathcal{Y}) \right\|_F^2 + \lambda \mathcal{R}(\mathcal{Z}; \theta). \quad (2.13)$$

当训练只使用观测张量和观测掩膜时, 这种设置称为单实例自监督: 缺失位置的填充值不构成训练标签, 完整参考张量只用于独立评价。与固定模型相比, 可学习映射提高表示适应性, 但也带来非凸训练、容量选择和有限训练稳定性问题。

对于三阶张量, 令 P_i 把第 i 模移到第三维, $g_{\theta_i}^{(i)}$ 为相应方向的变换, 则多方向候选为

$$\mathcal{X}_i = P_i^{-1} \left(g_{\theta_i}^{(i)} (P_i(\mathcal{Z}_i)) \right), \quad i = 1, 2, 3, \quad \hat{\mathcal{X}} = G_\phi(\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3). \quad (2.14)$$

三维深度低秩张量表示 (three-dimensional deep low-rank tensor representation, 3Deep-Rep) 把方向低秩潜张量、方向深度变换和非线性聚合器置于同一观测损失中^[19]。其低秩对象是各方向潜张量在变换域切片上的核范数, 而不是一个额外的塔克秩; 聚合器是学习到的非线性映射, 也不必解释为非负且和为一的凸权重。该构造为第 4 章分析“方向协同”提供基础。

2.4.2 连续坐标共享与整片缺失

若把每一个离散坐标都作为彼此独立的参数, 观测损失只在 $\mathcal{O}$ 上计算时, 缺失坐标 $k \notin \mathcal{O}$ 满足

$$\frac{\partial \mathcal{L}}{\partial d_k} = 0. \quad (2.15)$$

这说明整片模态缺失不仅是缺失率升高, 还会使该坐标失去直接拟合梯度。连续坐标表示以共享参数 θ 定义矩阵值函数

$$\mathbf{X}_\theta(z) : \mathcal{Z} \longrightarrow \mathbb{R}^{I_1 \times I_2}, \quad \mathcal{X}_{::,k} = \mathbf{X}_\theta(z_k),$$

观测坐标的损失通过共同参数影响任意查询坐标。正弦激活隐式表示和傅里叶特征为连续函数的高频坐标编码提供了基础^[8, 9], 函数型张量分解则把低秩系数与连续函数结合起来^[10-12]。第 4 章的低管秩连续表示 (low-rank tubal representation, LRTubalRep) 进一步把连续原子数、空间系数宽度和系数矩阵秩分开, 使坐标函数容量不必和空间组合容量同步增长。

需要区分观测坐标凸包内的连续插值和观测坐标凸包外的边界外推。共享参数可以解释观测信息如何传递到未观测坐标, 却不自动保证外推唯一或准确; 频率、参数范数、缺口长度和边界条件都会影响连续场的稳定性。

2.4.3 通道感知的变换域正则

若潜在通道的信息量不同, 把每个潜在通道的整体核范数赋予正权, 可写为

$$\mathcal{R}_{\text{ch}}(\mathcal{Z}; \theta_w) = \sum_{k=1}^{\tilde{I}_3} a_k(\mathcal{Z}; \theta_w) \|Z^{(k)}\|_*, \quad a_k > 0. \quad (2.16)$$

通道感知张量核范数 (channel-aware tensor nuclear norm, CA-TNN) 将权重作用对象从单个奇异值提升到潜在通道的整体核范数, 并通过通道统计生成正权^[20]。这一区别使模型能够学习通道之间的正则分配, 而不是仅按奇异值大小规律排序。

将式 (2.16) 代入观测保真项, 得到典型单实例自监督目标

$$\min_{\mathcal{Z}, \theta, \theta_w} \frac{1}{2} \|\mathcal{P}_\Omega(g_\theta(\mathcal{Z}) - \mathcal{Y})\|_F^2 + \lambda \sum_k a_k(\mathcal{Z}; \theta_w) \|Z^{(k)}\|_*.$$

方向变换、通道权重和连续坐标场都可以通过反向传播更新，但其输出是否可靠仍取决于观测几何和模型容量。第 4 章分别给出三方向深度表示、整片缺失连续表示和通道自适应正则的模型与结果。

2.5 可解释模型数据双驱动优化

2.5.1 多先验变量分裂

可解释模型数据双驱动优化把领域先验保留为显式算子，同时允许步长、阈值、罚参数或融合系数由数据学习。以低秩、空间、谱向和上下文四类先验为例，令 $\mathcal{Z}_j$ 为对应副本，观测和误差变量仍采用式 (2.12) 的共识结构。每个副本只承担一种先验，干净张量负责汇聚副本和观测关系，稀疏异常与稠密噪声则分别使用 ℓ_1 和二次惩罚。

算法展开把有限次迭代写成网络层。若 $\mathcal{D}_j^t$ 表示对偶或校正状态， $\mathcal{P}_{j,\vartheta_j^t}$ 为参数化先验算子， $\mathcal{C}_{\vartheta_c^t}$ 为观测校正和共识算子，则第 t 层可抽象为

$$\mathcal{Z}_j^{t+1} = \mathcal{P}_{j,\vartheta_j^t}(\mathcal{X}^t + \mathcal{D}_j^t), \quad \mathcal{X}^{t+1} = \mathcal{C}_{\vartheta_c^t}(\mathcal{Y}, \Omega, \{\mathcal{Z}_j^{t+1}, \mathcal{D}_j^t\}_j). \quad (2.17)$$

算法展开 (algorithm unrolling) 保留迭代算法的变量语义，同时允许部分更新参数从数据中学习^[15]。以 ADMM 为基础的展开网络还可以把原始变量、共识变量和对偶变量逐层传递^[16]。第 5 章的可解释张量低秩近似网络 (explainable tensor low-rank approximation network, ETLRA-Net) 正是在这一接口上显式保留干净张量、异常、噪声、四类先验副本和对偶状态。

2.5.2 有限层输出与迭代结论

有限层网络的输出 $\mathcal{X}^T$ 表示执行 T 次参数化更新。即使每层源自具有收敛分析的算法，当层间参数不共享、近端映射被学习模块替换或阈值随状态变化时，原算法的无限迭代结论也不能自动迁移。若要讨论 $t \rightarrow \infty$ 的固定点或收敛，需要另外验证算子非扩张性、参数有界性、误差可求和性和固定目标等条件。因而，本报告把第 5 章的主要问题表述为：如何在显式状态和先验接口中引入数据适应能力，并保持每层更新的数学含义。

2.6 评价原则与理论基础

基于第一章建立的三类驱动范式，本节集中给出各章共同遵循的评价原则，并说明本章数学基础在第 3–5 章中的具体用途。

2.6.1 评价指标与观测原则

若存在完整参考张量，可在隐藏测试位置或独立掩膜上计算均方误差 (mean squared error, MSE)、均方根误差 (root mean square error, RMSE)、峰值信噪比 (peak signal-to-noise ratio, PSNR) 和结构相似性指数 (structural similarity index measure,

SSIM):

$$\text{MSE} = \frac{1}{\prod_n I_n} \|\hat{\mathcal{X}} - \mathcal{X}^*\|_F^2, \quad \text{RMSE} = \sqrt{\text{MSE}}, \quad \text{PSNR} = 10 \log_{10} \frac{x_{\max}^2}{\text{MSE}}.$$

高光谱任务还常用光谱角映射 (spectral angle mapper, SAM) 评价光谱方向差异:

$$\text{SAM} = \frac{1}{I_1 I_2} \sum_{i,j} \arccos \frac{\langle \hat{\mathbf{x}}_{ij:}, \mathbf{x}_{ij:}^* \rangle}{(\|\hat{\mathbf{x}}_{ij:}\|_2 + \varepsilon)(\|\mathbf{x}_{ij:}^*\|_2 + \varepsilon)}.$$

PSNR、SSIM 越大越好, MSE、RMSE 和 SAM 越小越好。补全指标应在未用于训练的条目上计算; 若只有观测张量而没有完整参考, 则可以报告观测一致性、目标函数和训练状态, 但不能把观测损失直接当作缺失位置误差。

归一化、初始化、构图、参数选择和停止规则只能使用训练可见的观测或独立验证划分。对于随机缺失, 应采用多次独立掩膜并报告离散程度; 对于条带、连续坏列和整片切片缺失, 应分别说明缺失几何。这样的观测使用原则适用于三层方法, 也解释了为什么第 3 章的固定模型、第 4 章的单实例学习和第 5 章的有限层网络需要采用不同的验证方式。

2.6.2 理论基础与研究章节映射

表 2.1: 共同数学基础在研究章节中的作用。

基础模块	对应章节	在后续研究中的作用
张量代数与观测投影	第 3–5 章	统一数据布局、观测索引、模态展开和观测一致性, 使不同参数化与求解机制建立在同一逆问题上
显式低秩表示与结构先验	第 3 章	支撑全尺寸或降阶塔克参数化, 以及核心稀疏、因子核范数、图平滑、时间差分和重建梯度的作用位置
可学习表示与观测梯度	第 4 章	支撑多方向潜张量、连续坐标场和通道权重的参数化, 并解释观测损失如何通过共享参数传递到缺失位置
近端、变量分裂与算法展开	第 5 章	支撑解析更新、先验副本、误差分解和对偶状态的有限层传递, 并区分有限层输出与无限迭代收敛结论
评价指标与观测原则	第 3–5 章	统一隐藏位置评价、完整真值使用边界、缺失几何说明和多次随机实验要求

表 2.1 将共同数学基础分别映射到观测建模、显式低秩表示、可学习表示、有限层求解和实验评价; 第 3–5 章据此展开相应模型、算法与实验。

2.7 本章小结

本章从观测投影、异常与噪声分解出发，定义了张量展开、塔克多线性秩、降阶塔克表示、变换域张量核范数以及图和梯度先验，并给出近端更新、变量分裂、观测自监督梯度与有限层展开所需的共同接口。评价原则进一步限定了完整真值、隐藏位置指标和缺失几何的使用方式。上述内容分别支撑第 3 章的显式低秩建模、第 4 章的可学习表示和第 5 章的可解释求解，使三类研究共享数学定义与观测边界而保持各自技术职责。

第三章 模型驱动的低秩表示

本章总结博士后研究期间围绕模型驱动低秩张量补全形成的三项相互衔接的工作，核心是研究显式低秩参数化与固定目标求解如何协同设计。时空塔克分解插补方法（spatiotemporal Tucker decomposition approach for traffic imputation, TuckerAPP）^[17]以交通张量为对象，在全尺寸塔克坐标中组合核心稀疏、空间图正则与时间差分约束；低秩塔克表示模型（low-rank Tucker representation model, LRTuckerRep）^[18]面向一般多维数据，在降阶塔克表示中联合因子核范数、核心稀疏与因子图平滑，并以固定有限保真目标及近端交替线性化极小化（proximal alternating linearized minimization, PALM）理论作为本章的模型主线；在此基础上，梯度正则张量分解模型（gradient-regularized tensor decomposition, GradientTD）将局部先验的作用对象从潜在因子转移到重建张量本身，以离散梯度的总变分（total variation, TV）或二次变分（quadratic variation, QV）能量表达相邻位置的变化结构。三者共享“表示参数化与结构先验协同设计”的思想，但在分解尺度、局部正则的作用层次和求解机制方面并不相同。

本章依次给出任务定义、模型构造、算法更新、理论条件和计算代价。对时空塔克分解方法，重点陈述该论文的实验结果；对低秩塔克表示模型，重点说明其固定目标、近端块更新及条件性收敛结论；对梯度正则张量分解模型，仅展示相对于低秩塔克表示新增的梯度域构造及其变量分裂思路。

本文所称的模型驱动，是指在求解前明确给定张量表示、结构正则、观测保真项及其优化关系，再由观测数据估计核心张量、因子矩阵等模型变量。对应实现中，LRTuckerRep 始终计算同一个有限目标：塔克因子的核范数、核心张量的 ℓ_1 范数、固定图上的因子平滑项与观测区域均方误差共同组成目标函数，PALM 按核心块和因子块执行带回溯的近端更新。其优势在于变量含义、先验作用位置和每一步更新均可追踪，观测数据之外不需要额外训练样本，并可在明确条件下分析充分下降、平稳性与临界点收敛。与此同时，这类方法的表示坐标、图或差分算子及正则形式由建模者预先选定；当数据呈现方向相关、非平稳或整片缺失等复杂退化时，固定接口未必能够充分适配具体观测。这一局限构成第 4 章从模型驱动表示迈向数据驱动表示的直接动机。

3.1 问题定义与统一表示

3.1.1 交通张量化与补全任务

设原始交通观测以矩阵 $Y \in \mathbb{R}^{M \times (IJ)}$ 存储，其中 M 为传感器数， I 为每天的时间片数， J 为天数。按照“传感器–日内时间–日期”的方式堆叠后，张量化算子 $\mathcal{H}$ 将

其变为三阶张量

$$\mathcal{X} = \mathcal{H}(Y) \in \mathbb{R}^{I_1 \times I_2 \times I_3}, \quad (I_1, I_2, I_3) = (M, I, J).$$

记观测索引集为 Ω ，其补集为 $\bar{\Omega}$ ，观测投影定义为

$$[\mathcal{P}_\Omega(\mathcal{A})]_{i_1 i_2 i_3} = \begin{cases} a_{i_1 i_2 i_3}, & (i_1, i_2, i_3) \in \Omega, \\ 0, & (i_1, i_2, i_3) \notin \Omega. \end{cases} \quad \mathcal{X}_\Omega^0 = \mathcal{H}(Y_\Omega).$$

补全的要求是在观测位置保持原值，在未观测位置由低秩结构和领域先验给出估计

$$\hat{\mathcal{X}}_\Omega = \mathcal{X}_\Omega^0, \quad \hat{Y} = \mathcal{H}^{-1}(\hat{\mathcal{X}}).$$

这种张量化将长期趋势、日内相邻关系和传感器之间的空间关系分置于不同模态，为后续设计模态特定正则项提供了坐标基础。

3.1.2 塔克表示、展开与维度

对 N 阶张量 $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ ，其塔克表示为

$$\mathcal{X} = \mathcal{G} \times_{n=1}^N U_n, \quad \mathcal{G} \in \mathbb{R}^{r_1 \times \cdots \times r_N}, \quad U_n \in \mathbb{R}^{I_n \times r_n}. \quad (3.1)$$

第 n 模展开满足

$$X_{(n)} = U_n G_{(n)} V_n^\top, \quad V_n = U_N \otimes \cdots \otimes U_{n+1} \otimes U_{n-1} \otimes \cdots \otimes U_1.$$

其中 $\otimes$ 表示 Kronecker 积。若表示维数满足 $r_n \leq I_n$ ($n = 1, \dots, N$)，且至少一个模态严格满足 $r_n < I_n$ ，本文将式 (3.1) 称为降阶塔克表示 (reduced-order Tucker representation)。该术语强调表示坐标的维数相对于原始模态被压缩，核心张量的尺寸由 $(r_1, \dots, r_N)$ 决定；当所有 $r_n = I_n$ 时，则称为全尺寸塔克表示。后文分别使用 $\mathcal{G}, U_n$ 表示对应模型中的核心与因子； $\|\cdot\|_F$ 、 $\|\cdot\|_1$ 和 $\|\cdot\|_*$ 分别表示 Frobenius 范数、元素级 ℓ_1 范数和矩阵核范数。

3.2 时空正则的全尺寸塔克分解

3.2.1 建模动机与模型变量

交通张量通常同时具有两类结构。其一，多个模态之间存在长期、全局的低维交互；其二，相邻传感器和相邻时间片具有局部相似性。只对展开矩阵施加低秩约束难以保留这种多模态关系，而直接指定塔克秩又需要额外的先验或代价较高的秩搜索。TuckerAPP 的做法是采用全尺寸塔克分解

$$\hat{\mathcal{X}} = \mathcal{G} \times_{n=1}^N U_n, \quad \mathcal{G} \in \mathbb{R}^{I_1 \times \cdots \times I_N}, \quad U_n \in \mathbb{R}_+^{I_n \times I_n}, \quad (3.2)$$

并用核心张量的稀疏性表达有效交互数量，而不是直接指定一个较小的多线性秩。因子矩阵的非负约束还与交通流量、速度等非负量的取值语义一致。

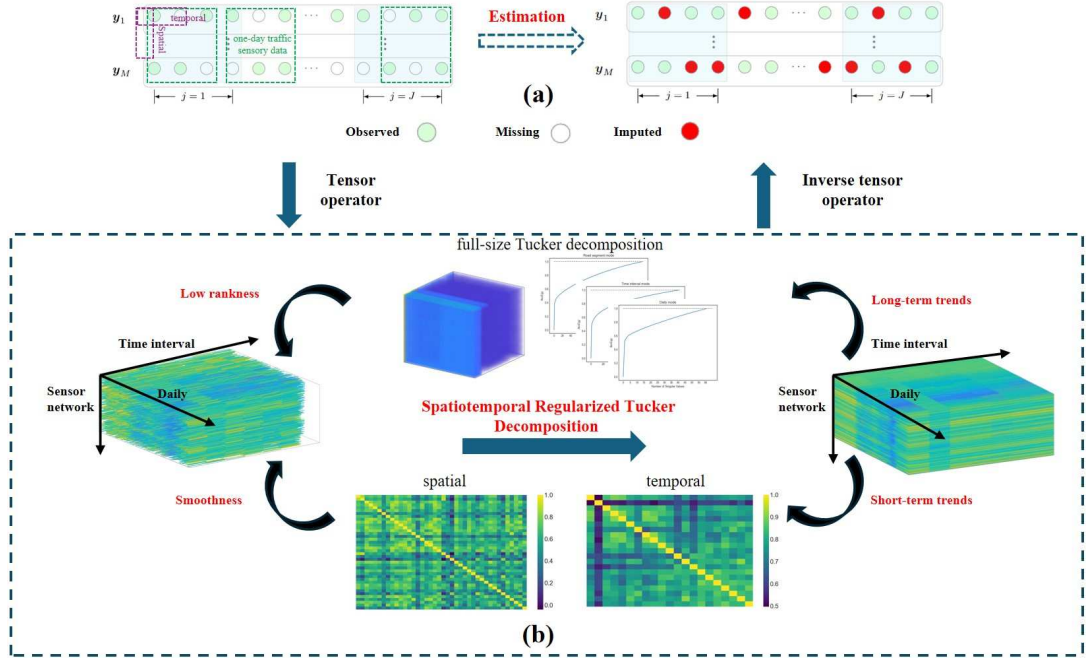


图 3.1: 时空塔克分解插补模型的整体框架。

图 3.1 展示了“观测-塔克表示-时空先验-插补结果”的信息流。首先输入是带缺失的交通观测及其空间邻接信息，核心模块以全尺寸塔克核心和非负模态因子形成低维交互表示，再分别通过空间图约束与时间变化约束注入局部先验，最后由观测投影得到插补张量。图中的核心、因子和重建输出分别对应式 (3.2) 的 $\mathcal{G}, U_n, \hat{\mathcal{X}}$ ；空间和时间分支对应后文的图拉普拉斯项与托普利兹/差分项，优化输出则由交替近端梯度算法更新。

3.2.2 空间图、时间约束与目标函数

空间图正则。 对传感器模态，论文根据传感器历史观测构造相似图。对第 i 个传感器的 p 个邻居，采用指数核

$$w_{ij} = \exp\left(-\frac{\|y_i - y_j\|_2^2}{\sigma^2}\right), \quad j \in \mathcal{N}_p(i), \quad w_{ij} = 0 \quad \text{其他情形}, \quad (3.3)$$

其中 y_i 是传感器 i 的历史序列，TuckerAPP 论文固定 $\sigma^2 = 1$ 。令 $W = [w_{ij}]$ 、 $D = \text{diag}(W\mathbf{1})$ ，图拉普拉斯矩阵为 $L = D - W$ 。对于因子矩阵的行向量 u_i ，有

$$\frac{1}{2} \sum_{i,j} w_{ij} \|u_i - u_j\|_2^2 = \text{tr}(U^\top L U).$$

因此，该项惩罚的是相似传感器在潜在因子空间中的不一致，而不是简单地把所有传感器行强制设成相同。

时间约束。 记时间模态上的 Toeplitz/差分约束矩阵为 T_n 。论文利用

$$\|\mathcal{X} \times_n T_n\|_F^2 = \|\mathcal{G} \times_1 U_1 \cdots \times_n (T_n U_n) \cdots \times_N U_N\|_F^2 \leq C \|T_n U_n\|_F^2$$

将时间变化约束转移到低维因子空间。这样可以用 $\|\mathbf{T}_n U_n\|_F^2$ 惩罚日内相邻时间片的剧烈变化，同时不把长期趋势与原始观测噪声混为一谈。

综合核心稀疏、空间图平滑和时间约束，TuckerAPP 的观测约束形式为

$$\begin{aligned} \min_{\mathcal{X}, \mathcal{G}, \{U_n\}} F(\mathcal{X}, \mathcal{G}, \{U_n\}) &= \frac{1}{2} \|\mathcal{X} - \mathcal{G} \times_{n=1}^N U_n\|_F^2 + \alpha \|\mathcal{G}\|_1 \\ &+ \sum_{n=1}^K \frac{\beta_n}{2} \text{tr}(U_n^\top L_n U_n) + \sum_{n=K+1}^N \frac{\gamma_n}{2} \|\mathbf{T}_n U_n\|_F^2, \quad (3.4) \\ \text{s.t. } U_n &\geq 0 \ (n = 1, \dots, N), \quad \mathcal{X}_\Omega = \mathcal{X}_\Omega^0. \end{aligned}$$

其中 K 是施加空间图正则的模态数， $\alpha, \beta_n, \gamma_n \geq 0$ 为罚参数。第一项使塔克重建贴近当前补全变量， ℓ_1 项选择少量核心交互，图项和时间项分别表达局部空间一致性与时间平滑性。该模型仍然存在核心与因子之间的尺度不唯一性；论文的收敛论证针对有界迭代序列和其采用的近端更新，并不声称非凸模型具有唯一全局最优解。

3.2.3 逐块梯度、近端算子与步长

固定其余变量，对第 n 模因子进行展开，令 $V_n = \bigotimes_{p \neq n} U_p$ ，则三类子问题统一写成

$$\ell_n(U_n) = \frac{1}{2} \|X_{(n)} - U_n G_{(n)} V_n^\top\|_F^2 + \begin{cases} \frac{\beta_n}{2} \text{tr}(U_n^\top L_n U_n), & n \leq K, \\ \frac{\gamma_n}{2} \|\mathbf{T}_n U_n\|_F^2, & n > K, \\ 0, & \text{无额外正则的模态.} \end{cases}$$

其梯度为

$$\nabla_{U_n} \ell_n = (U_n G_{(n)} V_n^\top - X_{(n)}) V_n G_{(n)}^\top + \begin{cases} \beta_n L_n U_n, & n \leq K, \\ \gamma_n \mathbf{T}_n^\top \mathbf{T}_n U_n, & n > K, \\ 0, & \text{其他情形.} \end{cases}$$

相应的块 Lipschitz 上界可取为

$$L_{U_n} = \begin{cases} \|G_{(n)} V_n^\top V_n G_{(n)}^\top\|_2 + \beta_n \|L_n\|_2, & n \leq K, \\ \|G_{(n)} V_n^\top V_n G_{(n)}^\top\|_2 + \gamma_n \|\mathbf{T}_n^\top \mathbf{T}_n\|_2, & n > K, \\ \|G_{(n)} V_n^\top V_n G_{(n)}^\top\|_2, & \text{其他情形.} \end{cases}$$

将核心张量向量化后，数据项为

$$f(\mathcal{G}) = \frac{1}{2} \left\| \text{vec}(\mathcal{X}) - \left(\bigotimes_{n=1}^N U_n \right) \text{vec}(\mathcal{G}) \right\|_2^2,$$

其梯度 Lipschitz 常数可以取

$$L_{\mathcal{G}} = \left\| \bigotimes_{n=1}^N U_n^\top U_n \right\|_2 = \prod_{n=1}^N \|U_n^\top U_n\|_2.$$

记逐元素软阈值算子和非负投影分别为

$$\mathcal{S}_\tau(z) = \text{sign}(z) \max\{|z| - \tau, 0\}, \quad \mathcal{P}_+(u) = \max\{u, 0\}.$$

在外推点 $\tilde{U}_n, \tilde{\mathcal{G}}$ 处, 近端梯度更新为

$$\hat{U}_n = \mathcal{P}_+ \left(\tilde{U}_n - \frac{1}{L_{U_n}} \nabla_{U_n} \ell_n(\tilde{U}_n) \right), \quad \hat{\mathcal{G}} = \mathcal{S}_{\alpha/L_{\mathcal{G}}} \left(\tilde{\mathcal{G}} - \frac{1}{L_{\mathcal{G}}} \nabla_{\mathcal{G}} f(\tilde{\mathcal{G}}) \right). \quad (3.5)$$

因子更新中的 $\mathcal{P}_+$ 同时处理非负约束, 核心更新中的 $\mathcal{S}$ 处理非光滑 ℓ_1 项, 因而不需要在每一块上调用通用二阶求解器。

3.2.4 加速近端梯度算法

TuckerAPP 论文在上述近端步外使用参数化惯性外推。记 $t^0 = 1$, 则

$$t^k = \frac{0.8 + \sqrt{4(t^{k-1})^2 + 0.8}}{2}, \quad \omega_k = \min \left\{ \frac{t^{k-1} - 1}{t^k}, 0.999 \sqrt{\frac{L^{k-1}}{L^k}} \right\}, \quad (3.6)$$

并对每个块使用

$$\tilde{\mathcal{G}}^k = \mathcal{G}^k + \omega_k(\mathcal{G}^k - \mathcal{G}^{k-1}), \quad \tilde{U}_n^k = U_n^k + \omega_k(U_n^k - U_n^{k-1}).$$

由于外推可能破坏观测一致性, 论文在每轮末进行控制校正。令 $\mathcal{Z}^0 = \mathcal{X}^0$, $\bar{\Omega}$ 为观测集补集, 则

$$\begin{aligned} \mathcal{Z}_\Omega^{k+1} &= \mathcal{X}_\Omega^0 + \phi \left(\mathcal{Z}_\Omega^{k-1} - \{\mathcal{G}^k \times_{n=1}^N U_n^k\}_\Omega \right), \\ \mathcal{Z}_{\bar{\Omega}}^{k+1} &= \{\mathcal{G}^k \times_{n=1}^N U_n^k\}_{\bar{\Omega}}, \quad 0 \leq \phi \leq 1. \end{aligned} \quad (3.7)$$

只有当校正后的目标值不增时才接受外推; 论文的停止准则为

$$\frac{\|\mathcal{P}_\Omega(\mathcal{Z}^{k+1} - \mathcal{X}^0)\|_F}{\|\mathcal{P}_\Omega(\mathcal{X}^0)\|_F} < \text{tol}, \quad \text{tol} = 10^{-4} \quad (\text{论文设置}). \quad (3.8)$$

最后输出 $\hat{\mathcal{X}}_\Omega = \mathcal{X}_\Omega^0$, $\hat{\mathcal{X}}_{\bar{\Omega}} = \mathcal{Z}_{\bar{\Omega}}^{K_\star}$, 其中 $K_\star$ 表示最后一次接受的迭代。

Algorithm 3.1 TuckerAPP 的交替近端梯度求解器**Require:** $\mathcal{X}^0$ 、观测集 Ω 、图矩阵 W_n 、约束矩阵 T_n 、 $\alpha, \{\beta_n\}, \{\gamma_n\}$ 和容差 tol **Ensure:** 补全张量 $\hat{\mathcal{X}}$

- 1: 由式 (3.3) 构造 W_n 、 $L_n = D_n - W_n$ ，并确定 T_n
- 2: 初始化 $\mathcal{G}^0, U_n^0 \geq 0$ ，计算 $\mathcal{Z}^0 = \mathcal{X}^0$ ，令 $t^0 = 1$ 、 $K_\star = 0$
- 3: **for** $k = 1, 2, \dots, K_{\max}$ **do**
- 4: 按式 (3.6) 计算外推参数，得到 $\tilde{\mathcal{G}}^k, \tilde{U}_n^k$
- 5: 计算 $L_{\mathcal{G}}^k$ ，按式 (3.5) 软阈值更新 $\mathcal{G}^k$
- 6: **for** $n = 1, 2, \dots, N$ **do**
- 7: 计算 $L_{U_n}^k$ ，按式 (3.5) 更新 U_n^k 并逐元素投影到非负正交象限
- 8: **end for**
- 9: 按式 (3.7) 更新 $\mathcal{Z}^k$ ；若目标值增加则回退外推并重新更新
- 10: 令 $K_\star \leftarrow k$
- 11: **if** 式 (3.8) 成立 **then**
- 12: **break**
- 13: **end if**
- 14: **end for**
- 15: **return** $\hat{\mathcal{X}}_\Omega = \mathcal{X}_\Omega^0$, $\hat{\mathcal{X}}_\Omega = \mathcal{Z}_\Omega^{K_\star}$

3.2.5 收敛条件与计算复杂度

论文将算法视为带非负投影和 ℓ_1 近端映射的非凸、多块交替优化。其收敛论证依次使用三点：在接受校正后目标函数充分下降；块梯度的 Lipschitz 性给出相对误差界；目标函数由多项半代数函数组成，满足 Kurdyka-Łojasiewicz (KL) 性质。在生成序列有界、每个步长主化相应块梯度且校正步骤保持下降的条件下，有

$$\sum_{k=1}^{\infty} \|\Theta^k - \Theta^{k-1}\|_F^2 < \infty, \quad \Theta^k = (\mathcal{G}^k, U_1^k, \dots, U_N^k), \quad (3.9)$$

并由 KL 论证得到整个序列收敛到一个临界点。这里的“收敛”不等于全局最优，也不排除不同初始化得到不同临界点。

若令 $P = \prod_{i=1}^N I_i$ ，则论文给出的主要单轮复杂度可概括为

$$\mathcal{O} \left(P \sum_{j=1}^N I_j^2 + \sum_{n=1}^N I_n \left(\prod_{i \neq n} I_i \right)^2 + \sum_{n=1}^K I_n^3 + \sum_{n=K+1}^N I_n^2 \log I_n \right).$$

第一项和第二项分别对应核心/重建与因子块的主要乘法，图拉普拉斯和 Toeplitz 约束带来后两项。实际瓶颈通常是大张量收缩和 Kronecker 结构计算，而核心软阈值本身仅需线性扫描核心元素。

3.2.6 实验结果

数据集、缺失机制与评价指标 论文使用两个真实交通数据集。Guangzhou (G) 为交通速度数据，原始矩阵大小为 214×8784 ，选取七天并重塑为 $214 \times 144 \times 7$ 的“传感器-日内时间-日期”张量；Hangzhou (H) 记录 80 个地铁站在 2019 年 1 月 1 日至 1 月 25 日、每 10 分钟的客流，张量大小为 $80 \times 108 \times 25$ 。缺失机制包括：随机缺失 (random missing, RM, 均匀随机屏蔽条目)、非随机缺失 (non-random missing, NM, 随机选择传感器并删除连续时间段) 以及全局停摆缺失 (blackout missing, BM, 所有传感器在一段时间内同时失效)。

交通补全在缺失位置计算平均绝对百分比误差 (mean absolute percentage error, MAPE) 和归一化平均绝对误差 (normalized mean absolute error, NMAE):

$$\text{MAPE} = \frac{100}{|\bar{\Omega}|} \sum_{i \in \bar{\Omega}, y_i \neq 0} \frac{|y_i - \hat{y}_i|}{|y_i|}, \quad \text{NMAE} = \frac{\sum_{i \in \bar{\Omega}} |y_i - \hat{y}_i|}{\sum_{i \in \bar{\Omega}} |y_i|}.$$

论文对 α 先在包含 99% RM、99% NM 和 95% BM 的复合场景 (采样比例 (sampling ratio, SR) 为 0.1) 中调节，并报告动态校正参数取 $\phi = 0.6$ 时在高缺失率下误差较低；图示迭代相对误差在约 100 轮后趋于稳定。该参数选择和收敛曲线是论文给定设置下的诊断，不应解释为对所有数据集的统一最优参数。

基线与数值结果 论文比较六类方法：低秩与稀疏增强塔克分解 (low-rank and sparse enhanced Tucker decomposition, LR-SETD)、平滑并行因子分析补全 (smooth parallel factor analysis completion, SPC; parallel factor analysis 简称 PARAFAC)、基于结构化核范数的汉克尔方法、低秩自回归张量补全 (low-rank autoregressive tensor completion, LATC)、时空张量列车补全 (spatio-temporal tensor train, stTT) 以及低管秩稀疏张量补全 (low-tubal-rank sparse tensor completion, LSTC)。TuckerAPP 论文在两类真实交通张量、RM/NM/BM 缺失协议和六个基线的给定实验中，验证了空间图与时间约束对极端随机缺失的帮助，并展示了约 100 轮后相对误差稳定的算法轨迹。表 3.1 转录论文表 3 的全部代表性行；数值越小越好，其中 G 报告 MAPE，H 报告 NMAE。

表 3.1: TuckerAPP 论文在 G/H 数据集上的代表性结果（原文表 3）。

数据集/指标	缺失场景	TuckerAPP	LR-SETD	LATC	LSTC	stTT	SPC	Hankel
G/MAPE	RM-30%	2.08	2.73	6.46	7.13	10.96	2.17	7.79
G/MAPE	RM-70%	5.61	6.58	6.93	6.89	7.88	10.95	11.62
G/MAPE	RM-90%	8.94	11.17	10.05	10.44	10.35	11.28	15.11
G/MAPE	RM-95%	11.02	13.55	12.60	16.22	12.25	12.84	17.69
G/MAPE	NM-70%	10.81	21.24	10.37	14.76	19.84	15.68	26.43
G/MAPE	NM-90%	17.48	57.33	20.89	25.16	22.23	28.97	49.35
G/MAPE	BM-30%	9.74	12.68	9.23	10.52	18.12	13.43	9.85
H/NMAE	RM-30%	0.0497	0.0501	0.1229	0.1159	0.1196	0.2120	0.3092
H/NMAE	RM-70%	0.1435	0.1753	0.1488	0.1935	0.1527	0.2251	0.3178
H/NMAE	RM-90%	0.2175	0.2274	0.2328	0.2356	0.2361	0.3764	0.3407
H/NMAE	RM-95%	0.2532	0.2752	0.4809	0.2601	0.3636	0.5124	0.3576
H/NMAE	NM-70%	0.1279	0.2722	0.1451	0.2017	0.1297	0.1388	0.6261
H/NMAE	NM-90%	0.2312	0.3761	0.2816	0.3051	0.2667	0.3361	0.6928
H/NMAE	BM-30%	0.1071	0.2024	0.1562	0.1785	0.0437	0.0422	0.1993

从表中可以作出三点有边界的判断。第一，在 G 的 RM-30% 至 RM-95% 以及 NM-90% 场景、H 的 RM 和 NM 场景中，TuckerAPP 的记录误差为该行最低；例如 G 的 MAPE 从 RM-30% 的 2.08 增加到 RM-95% 的 11.02，H 的 NMAE 从 RM-30% 的 0.0497 增加到 RM-95% 的 0.2532，但仍保持较低水平。第二，结果并非每一个结构化缺失场景都最优：G 的 NM-70% 和 BM-30% 分别由 LATC（10.37）和 LATC（9.23）取得更低 MAPE，H 的 BM-30% 则由 SPC（0.0422）和 stTT（0.0437）优于 TuckerAPP（0.1071）。第三，论文的消融结果显示空间和时间正则均带来改善；空间项对 G 的交通速度更重要，而时间项对 H 的网络客流更关键，时间正则图示设置下的贡献通常更明显。这些观察支持“全局核心稀疏与局部时空先验互补”的解释，但不支持“在任何缺失机制下均占优”的更强表述。

3.3 降阶低秩表示与梯度正则化

本节以 LRTuckerRep 为核心组织两种降阶塔克模型。LRTuckerRep 的关键在于对塔克坐标本身施加互补正则：因子核范数控制各模态子空间的有效维数，核心 ℓ_1 项选择跨模态交互，固定图拉普拉斯项约束潜在因子行的邻域一致性。GradientTD 保留前两类低秩正则，但把局部结构项改为直接作用于重建张量 $\mathcal{X}$ 的离散梯度能量。表 3.2 先给出两者的结构关系，随后详细展开 LRTuckerRep，并仅对 GradientTD 的差异化构造作专门说明。

表 3.2: LRTuckerRep 与 GradientTD 的模型构造对照。

比较维度	LRTuckerRep (本节核心)	GradientTD (差异化扩展)
全局表示	降阶塔克表示; 因子核范数与核心稀疏共同约束塔克坐标	沿用相同的降阶塔克低秩正则骨架
局部先验对象	潜在因子 U_n 的行向量	重建张量 $\mathcal{X}$ 的模态梯度
邻域结构	数据依赖的 k 近邻图或有序模态路径图	由有限差分矩阵给出的相邻位置关系
局部惩罚	$\text{tr}(U_n^\top L_n U_n)$	TV: $\ D_n X_{(n)}\ _1$; QV: $\ D_n X_{(n)}\ _F^2$
主要求解视角	固定有限保真目标上的 PALM; 另设硬约束近端线性化交替方向乘子法 (proximal-linearized alternating direction method of multipliers, ProADM)	对梯度变量和展开变量作分裂的近端线性化 ADMM
理论表述边界	在明确固定参数和步长条件下给出整序列临界点收敛	本报告仅概述附件中的算法构造, 不将其与 PALM 定理作同等级确认

3.3.1 低秩塔克建模与初始化

LRTuckerRep 面向一般多维张量补全^[18]。给定 $\mathcal{T} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ 和观测集 Ω , 采用降阶表示

$$\mathcal{Z} = \mathcal{G} \times_{n=1}^N U_n, \quad U_n \in \mathbb{R}^{I_n \times r_n}, \quad \mathcal{G} \in \mathbb{R}^{r_1 \times \cdots \times r_N}. \quad (3.10)$$

与 TuckerAPP 的全尺寸核心稀疏不同, LRTuckerRep 显式控制模态子空间的规模, 并同时约束因子和核心。

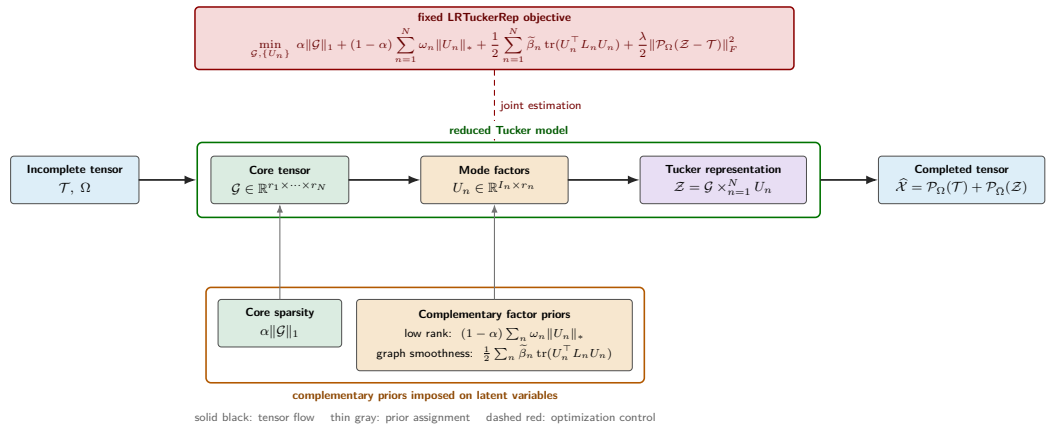


图 3.2: 低秩塔克表示模型的因子、核心与图先验协同框架。

图 3.2 的输入为观测张量和观测掩膜, 首先在可见数据上完成尺度归一化、预热初始化及一次性图构造; 随后将观测张量表示为核心张量与各模态因子的乘积, 并在因子、核心和选定模态图上分别施加核范数、稀疏和图平滑约束。输出端由塔克收缩形成潜在重建, 并在无噪声补全时覆盖观测位置。图中“因子低秩-核心稀疏-图平

滑”三个模块分别对应式 (3.11) 和式 (3.12)，观测一致性对应式 (3.13)，而固定目标的参数更新对应后文的近端交替线性化步骤。该框架图强调各先验的作用位置，不能将因子图平滑理解为对原始重建张量直接做差分。

为了避免把缺失位置的真实值泄漏到预处理，归一化、填充、子空间初始化和图构造均只使用 $\mathcal{T}_\Omega$ 。首先定义观测尺度

$$s = \begin{cases} \max_{i \in \Omega} |\mathcal{T}_i|, & \max_{i \in \Omega} |\mathcal{T}_i| > 0, \\ 1, & \text{其他情形,} \end{cases} \quad \tilde{\mathcal{T}} = \mathcal{T}/s,$$

并在归一化尺度上解释 λ 、图强度和停止容差；求解完成后再将核心与补全张量按 s 恢复量纲，因子保持不变。随后以可观测的模态纤维均值填充缺失条目；当一条纤维完全未观测时，回退到全局观测均值。基于掩膜展开执行高阶奇异值分解（higher-order singular value decomposition, HOSVD），并以有限步高阶正交迭代（higher-order orthogonal iteration, HOOI）型补全和观测核心细化形成预热初始化（warm start），得到 $\mathcal{G}^0$ 与 U_n^0 。对需要局部关系的模态，从初始填充张量 $\mathcal{X}^0$ 的展开行构造对称 k 近邻图；有明确顺序的时间模态也可直接采用一阶路径图。归一化和预热初始化只在优化前执行一次，图、权重、秩及全局系数在理论所对应的 PALM 运行中保持固定。若在迭代中重新构图或在线重加权，则目标函数随迭代改变，已经超出后述固定目标收敛结论的覆盖范围。

3.3.2 互补正则与保真模型

因子矩阵的列空间描述各模态的潜在子空间，核心元素描述这些子空间之间的交互。LRTuckerRep 据此定义

$$R_{\text{LR}}(\mathcal{G}, \{U_n\}) = (1 - \alpha) \sum_{n=1}^N \omega_n \|U_n\|_* + \alpha \|\mathcal{G}\|_1, \quad 0 < \alpha < 1. \quad (3.11)$$

因子核范数通过奇异值收缩压缩各模态坐标的有效维数，核心 ℓ_1 项则选择跨模态交互。二者同时存在还可抑制最直接的尺度逃逸：若放大某个因子并按比例缩小核心，至少有一个带正权的惩罚项随之增大。令 $R_n = \|U_n^0\|_*$ ，取 $\varepsilon_R > 0$ 为数值保护量，则初始化权重可写为

$$\omega_n = \frac{R_n + \varepsilon_R}{\sum_{j=1}^N (R_j + \varepsilon_R)}.$$

该表达式与对 $\prod_{i \neq n} (R_i + \varepsilon_R)^{-1}$ 归一化等价。权重仅用于减小不同模态初始化尺度造成的失衡，并非由数据端到端学习的参数；其在 PALM 迭代中保持不变。

对选定模态 $n \in \Gamma$ ，从 $\mathcal{X}_{(n)}^0$ 的行构造固定邻域图。若 x_i^0 表示第 i 行，则邻接边上的高斯权重及图拉普拉斯矩阵定义为

$$w_{ij}^{(n)} = \exp\left(-\frac{\|x_i^0 - x_j^0\|_2^2}{\sigma_n^2}\right), \quad (i, j) \in E_n, \quad L_n = \Delta_n - W_n \succeq 0,$$

其中 $\Delta_n = \text{diag}(W_n \mathbf{1})$ 为图的度矩阵， σ_n^2 可取正邻居平方距离的中位数。这里采用

Δ_n 而不使用 D_n 记度矩阵, 是为了与后文 GradientTD 的有限差分矩阵区分。图二次型满足

$$\text{tr}(U_n^\top L_n U_n) = \frac{1}{2} \sum_{i,j} w_{ij}^{(n)} \|u_{n,i} - u_{n,j}\|_2^2, \quad (3.12)$$

因而它约束的是邻接样本在潜在子空间中的表示差异, 而不是直接平滑原始观测。为平衡不同模态的图尺度, 令

$$\rho_n = \frac{\sigma_1(\mathcal{X}_{(n)}^0)}{2\|L_n\|_2 + \varepsilon_{\text{mach}}}, \quad \beta_n = \frac{\rho_n}{\sum_{j \in \Gamma} \rho_j} \quad (n \in \Gamma), \quad \beta_n = 0 \quad (n \notin \Gamma),$$

其中 $\sigma_1(\cdot)$ 为最大奇异值, $\eta \geq 0$ 进一步控制图正则的总体强度。

LRTuckerRep 的主模型采用有限观测保真项:

$$\begin{aligned} \min_{\mathcal{G}, \{U_n\}} \Phi(\mathcal{G}, \{U_n\}) &= (1 - \alpha) \sum_{n=1}^N \omega_n \|U_n\|_* + \alpha \|\mathcal{G}\|_1 \\ &+ \frac{\eta}{2} \sum_{n \in \Gamma} \beta_n \text{tr}(U_n^\top L_n U_n) + \frac{\lambda}{2} \|\mathcal{P}_\Omega(\mathcal{G} \times_{n=1}^N U_n - \mathcal{T})\|_F^2, \end{aligned} \quad (3.13)$$

其中 $\lambda > 0$ 调节潜在重建与观测值之间的有限保真。对于无噪声补全, 最终报告张量取为

$$\hat{\mathcal{X}} = \mathcal{P}_\Omega(\mathcal{T}) + \mathcal{P}_{\bar{\Omega}}(\hat{\mathcal{G}} \times_{n=1}^N \hat{U}_n). \quad (3.14)$$

观测覆盖是输出层的后处理, 并不把有限 λ 的目标函数变成硬约束问题; 在含噪恢复中, 则应保留潜在重建在全部位置上的估计。式 (3.13) 也是塔克坐标中的互补替代模型, 而非塔克秩的凸包或基不变刻画, 因此 α 、 λ 、 η 、预设秩和图参数仍需根据任务验证。

为区分有限保真与精确插值, 可另设硬约束模型

$$\begin{aligned} \min_{\mathcal{G}, \{U_n\}, \mathcal{X}} R_{\text{LR}}(\mathcal{G}, \{U_n\}) &+ \frac{\eta}{2} \sum_{n \in \Gamma} \beta_n \text{tr}(U_n^\top L_n U_n), \\ \text{s.t. } \mathcal{X} &= \mathcal{G} \times_{n=1}^N U_n, \quad \mathcal{P}_\Omega(\mathcal{X} - \mathcal{T}) = 0. \end{aligned} \quad (3.15)$$

该模型只有在给定塔克秩能够插值观测时才可行; 有限保真模型也不是其有限 λ 下的精确改写。后文分别用 PALM 与 ProADM 讨论这两个不同的优化对象。

3.3.3 近端交替线性化更新

将式 (3.13) 的可微部分记为

$$H(\mathcal{G}, \{U_n\}) = \frac{\lambda}{2} \|\mathcal{P}_\Omega(\mathcal{G} \times_{n=1}^N U_n - \mathcal{T})\|_F^2 + \frac{\eta}{2} \sum_{n \in \Gamma} \beta_n \text{tr}(U_n^\top L_n U_n).$$

于是

$$\Phi = H + (1 - \alpha) \sum_{n=1}^N \omega_n \|U_n\|_* + \alpha \|\mathcal{G}\|_1.$$

PALM 在同一固定目标上依次更新核心与各模态因子。输出覆盖式 (3.14) 只用于形成补全结果和诊断张量, $\mathcal{X}$ 本身不是该有限保真模型中的优化块; 因此, 观测覆盖不参

与后续梯度计算，也不会改变 Φ 的定义。

记当前重建为 $\mathcal{Z} = \mathcal{G} \times_{n=1}^N U_n$ ，则核心块梯度为

$$\nabla_{\mathcal{G}} H = \lambda \mathcal{P}_{\Omega}(\mathcal{Z} - \mathcal{T}) \times_{n=1}^N U_n^{\top},$$

其中 $\times_{n=1}^N U_n^{\top}$ 表示依次乘以各因子的转置。对第 n 个因子，在 Gauss–Seidel 顺序中使用最新的其他块，令 $V_n = \bigotimes_{p \neq n} U_p$ ，有

$$\nabla_{U_n} H = \lambda \mathcal{P}_{\Omega}(\mathcal{Z} - \mathcal{T})_{(n)} V_n G_{(n)}^{\top} + \eta \beta_n L_n U_n. \quad (3.16)$$

由于 $\mathcal{P}_{\Omega}$ 是非扩张投影，相应块梯度的有效 Lipschitz 上界可取

$$\ell_{\mathcal{G}}^k = \lambda \prod_{n=1}^N \|U_n^k\|_2^2, \quad \ell_{U_n}^k = \lambda \|B_n^k\|_2^2 + \eta \beta_n \|L_n\|_2,$$

其中 $B_n^k = (\mathcal{G}^{k+1} \times_{p < n} U_p^{k+1} \times_{p > n} U_p^k)_{(n)}$ 为 Gauss–Seidel 混合点上的设计矩阵。设主化系数 $L_{\mathcal{G}}^k > \ell_{\mathcal{G}}^k$ 、 $L_{U_n}^k > \ell_{U_n}^k$ ，核心与因子的 PALM 更新分别为

$$\mathcal{G}^{k+1} = \mathcal{S}_{\alpha/L_{\mathcal{G}}} \left(\mathcal{G}^k - \frac{1}{L_{\mathcal{G}}^k} \nabla_{\mathcal{G}} H(\mathcal{G}^k, \{U_n^k\}) \right), \quad (3.17)$$

$$U_n^{k+1} = \text{SVT}_{(1-\alpha)\omega_n/L_{U_n}^k} \left(U_n^k - \frac{1}{L_{U_n}^k} \nabla_{U_n} H \right), \quad n = 1, \dots, N, \quad (3.18)$$

其中若 $A = P \text{diag}(\sigma_i) Q^{\top}$ ，则

$$\text{SVT}_{\tau}(A) = P \text{diag}((\sigma_i - \tau)_+) Q^{\top}.$$

每个块采用回溯主化：若试验步长未满足

$$H(B^+) \leq H(B) + \langle \nabla H(B), B^+ - B \rangle + \frac{L_B}{2} \|B^+ - B\|_F^2, \quad (3.19)$$

则增大 L_B 并重算该块。每轮结束后执行

$$\mathcal{X}^{k+1} = \mathcal{T}_{\Omega} + \mathcal{P}_{\bar{\Omega}}(\mathcal{G}^{k+1} \times_{n=1}^N U_n^{k+1}). \quad (3.20)$$

Algorithm 3.2 LRTuckerRep 的固定目标 PALM 算法**Require:** $\mathcal{T}_\Omega, \Omega$, 秩 $(r_1, \dots, r_N)$, 固定 $\alpha, \lambda, \eta, \{\omega_n, \beta_n, L_n\}$ **Ensure:** $\hat{\mathcal{X}}$

- 1: 仅用观测值完成纤维均值初始化, 得到 $\mathcal{X}^0, \mathcal{G}^0, \{U_n^0\}$
- 2: 固定图和权重, 计算 $\mathcal{Z}^0 = \mathcal{G}^0 \times_{n=1}^N U_n^0$, 令 $K_\star = 0$
- 3: **for** $k = 0, 1, \dots, K_{\max} - 1$ **do**
- 4: 计算 $\nabla_{\mathcal{G}} H$, 回溯增大 $L_{\mathcal{G}}^k$ 直至式 (3.19) 成立
- 5: 按式 (3.17) 对核心作软阈值更新
- 6: **for** $n = 1, 2, \dots, N$ **do**
- 7: 用最新块计算式 (3.16), 回溯并按式 (3.18) 作 SVT
- 8: **end for**
- 9: 按式 (3.20) 形成报告与诊断张量, 不将其作为新的优化块
- 10: 令 $K_\star \leftarrow k + 1$; 计算相对近端残差, 若小于容差则停止
- 11: **end for**
- 12: **return** 最后一次接受迭代 $K_\star$ 对应的 $\hat{\mathcal{X}} = \mathcal{X}^{K_\star}$

3.3.4 收敛、辅助求解与计算代价

固定目标的收敛范围。 令 $z^k = (\mathcal{G}^k, U_1^k, \dots, U_N^k)$ 。在 $0 < \alpha < 1$ 、 $\omega_n > 0$ 、 $\lambda > 0$ 、 $\eta\beta_n \geq 0$ 和 $L_n \geq 0$ 时, 目标的每个有限下水平集均为紧集, 因为

$$\|\mathcal{G}\|_F \leq \frac{\Phi}{\alpha}, \quad \|U_n\|_F \leq \frac{\Phi}{(1-\alpha)\omega_n}.$$

这一定量界排除了最直接的无界尺度逃逸, 但不意味着分解或补全结果唯一。进一步假设近端映射精确, 所有接受的主化系数一致有界, 且与块 Lipschitz 上界之间存在统一正间隔 $\underline{\gamma} > 0$, 则一次完整 Gauss–Seidel 块扫描满足

$$\Phi(z^k) - \Phi(z^{k+1}) \geq \frac{\underline{\gamma}}{2} \|z^{k+1} - z^k\|^2. \quad (3.21)$$

近端最优性还给出相对误差界: 存在 $v^{k+1} \in \partial\Phi(z^{k+1})$ 和常数 $C > 0$, 使 $\|v^{k+1}\| \leq C\|z^{k+1} - z^k\|$ 。由于多线性保真项和图二次项是多项式, ℓ_1 范数和核范数是半代数函数, 故 Φ 具有 Kurdyka–Łojasiewicz 性质。由充分下降、相对误差、紧水平集和 KL 不等式可得

$$\sum_{k=0}^{\infty} \|z^{k+1} - z^k\| < \infty, \quad z^k \longrightarrow z^\star, \quad 0 \in \partial\Phi(z^\star). \quad (3.22)$$

同时, 最优迭代点的一阶平稳性满足

$$\min_{0 \leq k < K} \text{dist}(0, \partial\Phi(z^{k+1})) = \mathcal{O}(K^{-1/2}), \quad (3.23)$$

对应 ε 平稳点的最坏迭代复杂度为 $\mathcal{O}(\varepsilon^{-2})$ 。这些结论是固定目标、理想精确近端序列的临界点收敛结论, 不等价于全局最优, 也不覆盖在线构图、在线重加权、惯性外

推或非可求和数值误差。

硬约束辅助求解路径的作用与边界。 针对式 (3.15)，ProADM 为塔克联结约束和观测约束分别引入乘子 $\mathcal{Y}, \mathcal{Q}$ 及罚参数 $\mu > 0$ 。核心和因子仍以软阈值和 SVT 更新；在得到新重建 $\mathcal{Z}^{k+1}$ 后， $\mathcal{X}$ 块具有闭式解

$$\mathcal{X}_{\Omega}^{k+1} = (\mathcal{Z}^{k+1} - \mathcal{Y}^k / \mu)_{\bar{\Omega}}, \quad \mathcal{X}_{\Omega}^{k+1} = \frac{1}{2}(\mathcal{Z}^{k+1} - \mathcal{Y}^k / \mu + \mathcal{T} - \mathcal{Q}^k / \mu)_{\Omega}.$$

冻结乘子时，主变量扫描具有充分下降；但乘子上升会使普通增广拉格朗日量增加

$$\mu (\|\mathcal{X}^{k+1} - \mathcal{Z}^{k+1}\|_F^2 + \|\mathcal{P}_{\Omega}(\mathcal{X}^{k+1} - \mathcal{T})\|_F^2),$$

所以不能仅凭主变量下降推出整个原始-对偶序列收敛。当前可确认的结论是：若某收敛子序列的两类可行性残差和各块近端 KKT 映射均趋于零，则该聚点满足硬约束模型的 Karush–Kuhn–Tucker (KKT) 条件。本报告据此将 ProADM 作为精确约束的辅助数值路径，而不把它与 PALM 的整序列定理等同。

计算代价。 令 $P = \prod_n I_n$ 、 $R = \prod_n r_n$ 、图最大度为 k ，单块回溯次数至多为 B 。稠密 HOSVD 初始化代价为

$$\mathcal{O} \left(\sum_{n=1}^N \min \{ I_n^2(P/I_n), I_n(P/I_n)^2 \} \right),$$

直接构造近邻图的代价为 $\mathcal{O}(\sum_n I_n P)$ 。若顺序塔克收缩代价记为

$$C_{\text{Tucker}} = \sum_{n=1}^N I_n r_n \left(\prod_{p < n} I_p \right) \left(\prod_{p > n} r_p \right),$$

则一次 PALM 扫描的主要工作量可概括为

$$\mathcal{O} \left(BC_{\text{Tucker}} + P \sum_{n=1}^N r_n + \sum_{n=1}^N (I_n r_n^2 + r_n^3) + k \sum_{n=1}^N I_n r_n \right).$$

峰值内存量级约为 $\mathcal{O}(P + R + \sum_n I_n r_n + k \sum_n I_n)$ 。因此主要瓶颈是稠密张量收缩、初始化谱分解及因子 SVT；核心软阈值只需对核心元素作线性扫描。ProADM 还需同时保存 $\mathcal{X}$ 、 $\mathcal{Z}$ 、乘子和残差，具有更大的全张量常数项。

数值结果衡量标准。 对于归一化到 $[0, 1]$ 的数据，令 $N = n_1 n_2 n_3$ ，峰值信噪比 (peak signal-to-noise ratio, PSNR) 和光谱角映射 (spectral angle mapper, SAM) 分别为

$$\text{PSNR} = 10 \log_{10} \frac{1}{N^{-1} \|\hat{\mathcal{X}} - \mathcal{X}^*\|_F^2},$$

$$\text{SAM} = \frac{1}{n_1 n_2} \sum_{l,j} \arccos \left(\frac{\langle \hat{\mathbf{x}}_{l,j}, \mathbf{x}_{l,j}^* \rangle}{(\|\hat{\mathbf{x}}_{l,j}\|_2 + \varepsilon)(\|\mathbf{x}_{l,j}^*\|_2 + \varepsilon)} \right).$$

PSNR、结构相似性指数 (structural similarity index measure, SSIM) 越大越好，SAM 越小越好。若实际任务没有 $\mathcal{X}^*$ ，则只能报告观测一致性、结构正则、残差和训练轨迹，不能把观测损失直接解释为缺失位置误差。

3.3.5 重建张量的离散梯度扩展

LRTuckerRep 的图正则作用于潜在因子行，其局部性取决于初始化数据所构造的邻域图。对于高光谱图像、视频和规则采样交通张量，局部连续性通常还可以直接表示为重建张量相邻位置之间的有限差分；低秩性与差分平滑的联合建模可参见张量补全和低秩-平滑融合工作的代表性构造^[3, 6]。GradientTD 保留式 (3.11) 的低秩塔克骨架，但用作用于 $\mathcal{X}$ 的离散梯度先验替换因子图二次项。为避免与图拉普拉斯 L_n 混淆，记第 n 模有限差分矩阵为 D_n ，并定义

$$\nabla_n \mathcal{X} \cong D_n X_{(n)}, \quad \phi_p(\nabla_n \mathcal{X}) = \begin{cases} \|D_n X_{(n)}\|_1, & p = 1, \\ \|D_n X_{(n)}\|_F^2, & p = 2. \end{cases} \quad (3.24)$$

其中，若第 n 模长度为 I_n ，非周期前向差分可取 $D_n \in \mathbb{R}^{(I_n-1) \times I_n}$ ，并满足 $(D_n x)_i = x_{i+1} - x_i$ ；若数据沿该模态具有周期边界，则首尾差分还应被加入 D_n 。因此，差分矩阵的尺寸、边界条件和作用模态集合 Γ 是模型的一部分，不能在不同实验之间隐式更换。与 LRTuckerRep 的因子图正则相比，GradientTD 的梯度项直接作用于重建张量的展开矩阵，因而其局部先验不依赖于潜在因子行所构造的近邻图；代价是它要求所选模态具有可信的顺序或网格邻接关系。由此得到 GradientTD 的约束形式

$$\begin{aligned} \min_{\mathcal{G}, \{U_n\}, \mathcal{X}} \quad & R_{\text{LR}}(\mathcal{G}, \{U_n\}) + \sum_{n \in \Gamma} \beta_n \phi_p(\nabla_n \mathcal{X}), \\ \text{s.t.} \quad & \mathcal{X} = \mathcal{G} \times_{n=1}^N U_n, \quad \mathcal{P}_\Omega(\mathcal{X}) = \mathcal{P}_\Omega(\mathcal{T}). \end{aligned} \quad (3.25)$$

当 $p = 1$ 时，TV 惩罚使梯度稀疏，适合保留分段常值结构和突变边界；当 $p = 2$ 时，QV 惩罚对应离散 Dirichlet 能量，对较大梯度施加二次代价，更适合表达连续过渡。这里的核心差异不在于重新定义全局低秩，而在于改变局部先验的作用层次：LRTuckerRep 在塔克潜在坐标中约束邻域表示，GradientTD 则在数据域直接约束重建张量的相邻变化。前者可以使用非局部、数据依赖的图邻域，后者依赖给定模态上的有序网格或可解释差分边界条件。

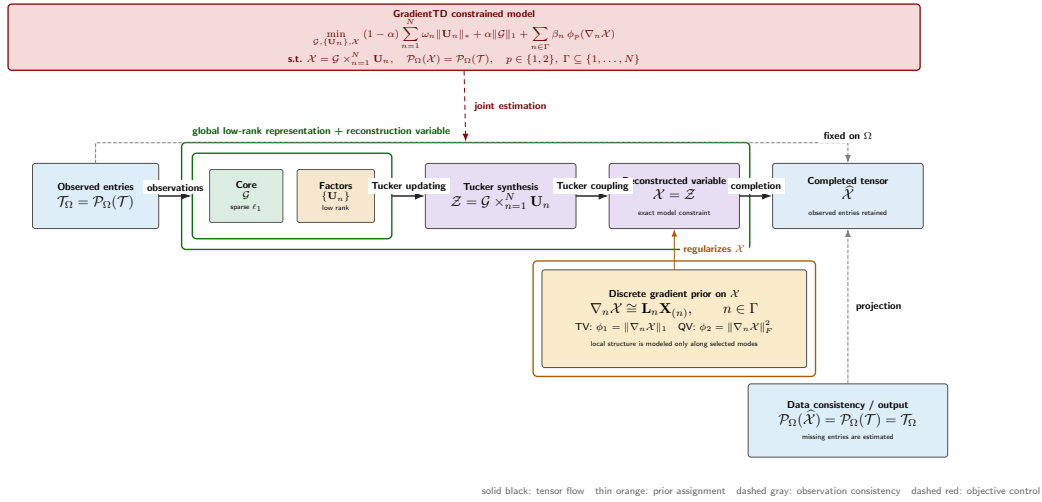


图 3.3: 梯度正则张量分解模型的变量分裂与求解框架。

图 3.3 的输入是观测张量、掩膜和各模态有限差分矩阵；模型先由塔克因子与核心产生重建变量，再把重建变量复制到各梯度分支，分别计算 TV 或 QV 惩罚，最后通过观测一致性和共识约束汇聚输出。图中的梯度变量、重建变量与塔克变量分别对应式 (3.26)、式 (3.25) 的约束关系以及后文的近端线性化交替方向乘子法更新； $p = 1$ 分支对应软阈值， $p = 2$ 分支对应线性收缩。该图突出 GradientTD 相对 LRTuckerRep 的唯一核心变化：局部结构从潜在因子图转移到数据域差分，而全局低秩骨架保持不变。

式 (3.25) 同时包含非光滑低秩正则、梯度算子和多仿射塔克等式。附件采用变量分裂；这一求解思路与交替方向乘子法及正则化多凸块坐标优化的基本框架一致^[14, 21]：

$$Q_n = D_n Y_n, \quad Y_n = X_{(n)}, \quad (3.26)$$

把梯度块与塔克块分开。对 Q_n 的近端更新具有明确形式：若 $p = 1$ ，则为逐元素软阈值；若 $p = 2$ ，则在采用 $\beta_n \|Q_n\|_F^2$ 的约定下为线性收缩

$$Q_n^{k+1} = \frac{\rho_1}{2\beta_n + \rho_1} \left(D_n Y_n^k - \frac{\Lambda_n^{Q,k}}{\rho_1} \right).$$

Y_n 子问题对应系数矩阵 $\rho_1 D_n^\top D_n + \rho_2 I$ 的线性方程；在周期边界下，该算子具有循环结构，可在傅里叶域求解。核心与因子仍分别通过软阈值和 SVT 处理，从而保留 LRTuckerRep 互补正则的近端结构。

基于优化问题的差异性，可以进一步给出两种约束处理方式。混合近端线性化交替方向乘子法（hybrid proximal-linearized alternating direction method of multipliers, HPL-ADMM）将 $\mathcal{X} = \mathcal{G} \times_{n=1}^N \mathbf{U}_n$ 写成有限二次罚项，并通过 $\mathcal{X}$ 汇聚低秩重建与各模态平滑变量；完全分裂线性化交替方向乘子法（fully-splitting linearized alternating direction method of multipliers, FSL-ADMM）则为该塔克等式增加独立乘子，把它保留为显式约束。二者共同体现的新增机制是“重建变量-梯度变量-塔克变量”的三层

分裂，而非对 LRTuckerRep 图权重的简单替换。GradientTD 在本章仅保留相对于低秩塔克表示模型新增的梯度域构造和变量分裂求解思路，不展开其完整实验结果。

3.4 本章小结

本章形成了由特定领域模型到一般低秩表示、再到梯度域扩展的递进关系。TuckerAPP 在全尺寸塔克分解中以核心 ℓ_1 稀疏避免直接搜索秩，并通过图拉普拉斯和 Toeplitz/差分约束刻画交通数据的空间与时间局部性；其交替近端梯度、外推校正、条件性 KL 收敛分析和论文实验共同构成完整的模型—算法—结果链条。

本章的降阶模型部分以 LRTuckerRep 为核心。该模型把因子核范数、核心稀疏和固定因子图平滑统一到有限观测保真目标中，并用观测数据完成归一化、初始化和一次性构图。固定目标 PALM 在紧水平集、充分下降、相对误差及 KL 条件下收敛到单一临界点，并具有 $\mathcal{O}(K^{-1/2})$ 的最优迭代平稳性界；硬约束 ProADM 则只在残差消失的聚点层面确认 KKT 一致性，二者的理论结论不能互换。

GradientTD 沿用低秩塔克表示模型的全局低秩正则骨架，但把局部先验从潜在因子图平滑改为重建张量上的有限差分 TV/QV，并通过梯度变量、展开变量与塔克变量的分裂组织近端线性化 ADMM。

因此，本章在全文主线中给出第一个基准组合：用核心、因子和重建变量显式参数化恢复张量，用与固定目标一致的近端、PALM 或变量分裂更新求解这些变量。其核心优势是结构明确、求解可审计、理论条件可陈述；研究者可以辨认性能来自低秩、稀疏、图平滑还是梯度正则，并能将算法更新对应到具体子问题。但其适应性受预设参数化接口限制。下一章不再只更换一个手工正则项，而是在保留显式低秩潜变量和观测投影的前提下，让方向变换、坐标基、聚合规则或通道权重由当前观测参与学习，从而处理固定塔克坐标难以覆盖的复杂退化以及整片缺失坐标缺少直接梯度的关键问题。

第四章 数据驱动的深度低秩表示

4.1 数据驱动的低秩表示

第3章通过预先选定的塔克坐标、图或差分算子表达低秩与局部结构，其突出优势是目标、变量和更新之间具有清楚的对应关系。然而，固定表示把“哪些方向应协同、相隔坐标如何共享信息、不同潜在通道应受到多强约束”等选择预先固化。当退化模式随样本改变，或者整个波段、时间帧均不可见时，仅调节固定正则系数难以建立观测位置到缺失结构的有效信息通路。因此，从模型驱动迈向数据驱动的动机不是取消低秩先验，而是把难以手工指定且依赖具体观测的表示接口转化为可学习映射。

本文所称的数据驱动低秩表示，是指保留低秩潜变量、观测投影和张量形状约束，同时让变换、连续坐标函数、聚合规则或正则权重由观测数据参与确定。对应核心实现中，LRTubalRep 将恢复张量写成低秩空间系数与连续第三模态基的收缩：每个空间系数矩阵由两个因子相乘而具有显式秩上界，坐标字典 $\Phi(z)$ 与投影矩阵 C 形成可查询的共享基 $W(z) = \Phi(z)C$ ；训练循环只最小化观测区域均方误差及可选的结构正则，并以自适应矩估计更新表示参数。由此，数据驱动方式能够学习跨方向的信息组织、跨坐标的参数共享和跨通道的正则分配，重点解决固定表示对复杂退化适配不足以及整片缺失位置没有直接数据梯度的问题。

高光谱图像是本章的主要载体，可表示为理想张量 $\mathcal{X}^* \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ ，其中前两维为空间坐标，第三维为光谱通道；本章的建模框架也可用于视频帧或按时间组织的多维观测。

为避免符号混用，以下将退化观测记为 $\mathcal{O}$ ，观测索引集记为 Ω 。基本观测模型为

$$\mathcal{O} = \mathcal{P}_\Omega(\mathcal{X}^* + \mathcal{N}), \quad [\mathcal{P}_\Omega(\mathcal{A})]_{ljs} = \begin{cases} a_{ljs}, & (l, j, s) \in \Omega, \\ 0, & (l, j, s) \notin \Omega, \end{cases} \quad (4.1)$$

其中 $\mathcal{N}$ 为观测位置上的加性扰动。本章三项工作均采用单实例自监督：训练只使用 $\mathcal{O}$ 、 Ω 和由观测可计算的统计量，数据项写成 $\mathcal{P}_\Omega(\hat{\mathcal{X}} - \mathcal{O})$ ；缺失位置的零填充值不作为标签，完整参考张量只在测试阶段计算恢复指标。

本章围绕三个逐层推进的问题组织。第一，**单一变换方向不足**：当某一模态的连续性被坏列、条带或混合缺失破坏时，需要利用其他模态的互补结构，并让融合规则由数据学习。第二，**整片缺失坐标没有直接梯度**：当整个光谱波段或时间帧未观测时，离散坐标参数无法从该位置获得训练信号，需要通过共享连续参数建立跨坐标信息通路。第三，**潜在通道并非同等重要**：统一惩罚或只按奇异值次序加权不能显式表达通道级差异，需要让正则权重依赖潜在通道本身。对应的研究路线为“方向协同—坐

标共享-通道自适应”。

表 4.1: 数据驱动低秩表示

工作	所解决的问题	表示与信息通路	研究重点
三维深度低秩张量表 示 (three-dimensional deep low-rank tensor representation, 3DeepRep)	单一方向难以覆盖复杂空间-光谱退化, 生成器式潜表示增加参数路径, 固定融合难以适应局部伪影	显式潜张量经三个模态方向的深度耦合变换形成候选重建, 再由拼接后的非线性聚合器融合	重点分析多方向变换、非线性聚合及其主实验和消融结果 ^[19]
低管秩连续表示 (low-rank tubal representation, LRTubalRep)	第三模态整片缺失时没有直接数据拟合梯度, 连续函数容量与空间系数容量易被同一宽度绑定	将第三模态写成共享参数的矩阵值连续场, 用 M 个坐标原子经投影组合 d 个低秩空间系数	重点说明零直接梯度、共享参数迁移和 (M, d, r) 容量解耦
通道感知张量核范数 (channel-aware tensor nuclear norm, CA-TNN)	传统加权张量核范数 (weighted tensor nuclear norm, WTNN) 的奇异值级权重难以显式反映潜在通道的信息量与退化差异	对每个潜在通道的整个核范数赋可学习正权, 并与空间-光谱变换和潜变量联合优化	重点分析通道级加权、指数激励及其随机缺失实验结果 ^[20]

三项工作形成由方向协同到坐标共享、再到通道自适应的递进关系。3DeepRep 学习跨方向的信息组织方式; LRTubalRep 把离散索引提升为可查询的连续坐标场, 学习跨坐标的信息传递方式; CA-TNN 则在潜在变换域学习跨通道的正则分配方式。以下各节均按照“待解决问题-表示构造-信息通路-优化接口”的顺序展开, 并在具有完整实验设置的部分给出相应数值结果。

4.2 多方向深度低秩表示

作为主线的第一步, 3DeepRep 针对三个相互关联的问题进行联合设计: 生成器式潜表示引入额外参数路径, 单一光谱方向在结构化缺失下可能失效, 固定线性权重又难以融合质量不均的方向候选。模型因此不只是增加三个网络分支, 而是让每个分支在相应模态方向承担显式低秩约束, 再由可学习的非线性空间-光谱映射完成融合。

设 $\mathcal{Z}_i$ 为第 i 个方向的显式潜张量, $\mathbf{P}_i$ 将第 i 模移动到第三维, g_{θ_i} 为置换坐标中

的耦合变换，则候选重建为

$$\mathcal{X}_i = \mathbf{P}_i^{-1}(g_{\theta_i}(\mathbf{P}_i(\mathcal{Z}_i))), \quad i = 1, 2, 3. \quad (4.2)$$

令 $\tilde{\mathcal{Z}}_i = \mathbf{P}_i(\mathcal{Z}_i)$ ，其第三维长度为 $\tilde{n}_i$ 。方向低秩项不是对一个模态展开矩阵取核范数，而是对置换后各正面切片的矩阵核范数求和：

$$\mathcal{R}_i(\mathcal{Z}_i) = \sum_{q=1}^{\tilde{n}_i} \|\tilde{\mathcal{Z}}_i(:, :, q)\|_*, \quad \mathcal{R}_{3D} = \sum_{i=1}^3 \alpha_i \mathcal{R}_i. \quad (4.3)$$

三个候选沿光谱方向拼接后进入聚合网络 a_ϕ ：

$$\mathcal{X}_{\text{cat}} = \text{concat}_3(\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3), \quad \hat{\mathcal{X}} = a_\phi(\mathcal{X}_{\text{cat}}). \quad (4.4)$$

该聚合器由切片卷积和第三模全连接映射构成，不是逐元素非负且和为一的凸权重。因此，其作用应解释为学习方向候选之间的非线性局部修正与通道混合，而不能直接解释为概率门控。

每个方向同时保持对观测数据的中间一致性，完整目标为

$$\begin{aligned} \min_{\{\mathcal{Z}_i, \theta_i\}_{i=1}^3, \phi} \quad & \sum_{i=1}^3 \alpha_i [\mathcal{R}_i(\mathcal{Z}_i) + \lambda_i \|\mathcal{P}_\Omega(\mathcal{X}_i - \mathcal{O})\|_F^2] \\ & + \|\mathcal{P}_\Omega(\hat{\mathcal{X}} - \mathcal{O})\|_F^2. \end{aligned} \quad (4.5)$$

显式潜变量缩短了观测损失到低秩变量的梯度路径；三个方向的中间保真项防止候选分支完全依赖最终聚合器；最终保真项则把方向变换和聚合网络置于同一自监督目标中。由此，模型的表示、信息融合和观测约束形成闭环。

为直观说明多方向表示与非线性融合的关系，图 4.1 给出 3DeepRep 的模型框架。

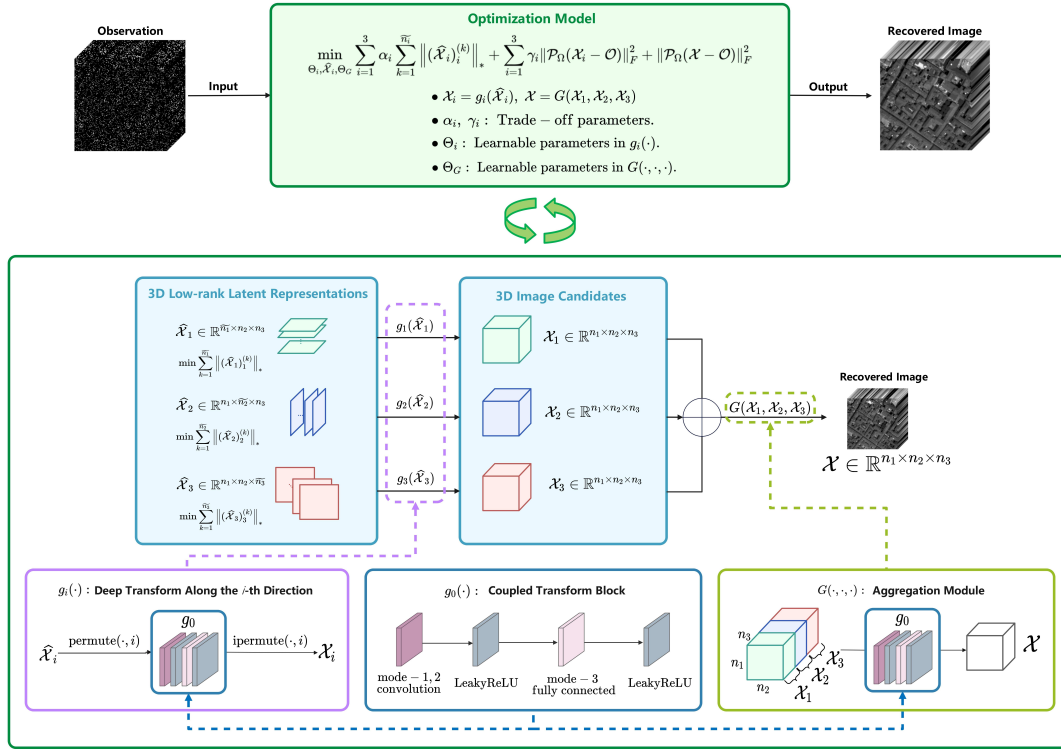


图 4.1: 多方向深度低秩张量表示模型的结构框架。

图 4.1 的输入是带缺失或退化的高光谱张量及其观测掩膜；第一层为三个方向的显式低秩潜表示 $\hat{\mathcal{Z}}_i$ ，分别将不同模态置于变换轴并送入方向耦合变换；第二层的 g_i 和耦合变换块在对应方向上提取局部空间结构与跨切片关联，得到候选张量 $\mathcal{X}_i$ ；第三层把候选张量沿预定模态拼接，经聚合模块 G 完成非线性空间-光谱融合，输出 $\hat{\mathcal{X}}$ 。因此，图中三个潜表示、三个候选分支和最终聚合输出分别对应式 (4.2)、式 (4.4) 及式 (4.5)；图底部的方向变换、耦合块和聚合模块与后文的具体卷积、管线映射及 1×1 融合参数化相对应。图示强调的是信息通路和模块职责，不将聚合器解释为非负概率权重，也不把三个方向的正则化秩混同为一个塔克秩。

Algorithm 4.3 3DeepRep 统一梯度驱动优化**Require:** 退化高光谱张量 $\mathcal{O}$ 和掩膜 Ω **Ensure:** 修复结果 $\hat{\mathcal{X}}$

```

1: 初始化显式潜张量  $\mathcal{Z}_1, \mathcal{Z}_2, \mathcal{Z}_3$ 、变换参数  $\theta_i$  和聚合器参数  $\phi$ 
2: for  $t = 0, 1, \dots, T - 1$  do
3:   由式 (4.2) 计算  $\mathcal{X}_i$ 
4:   拼接三个候选，并由式 (4.4) 执行非线性聚合
5:   计算观测一致性项和式 (4.3) 的低秩项
6:   对  $\mathcal{Z}_i, \theta_i, \phi$  反向传播并更新
7:   if 保留观测损失达到停止准则 then
8:     break
9:   end if
10: end for
11: return  $\hat{\mathcal{X}}$ 

```

4.2.1 参数化与网络结构

论文中的“方向”不是简单地把同一个变换重复三次，而是先把第 i 模态移到第三维，再在置换后的切片和管线方向上执行同一类耦合变换。设 P_i 为模态置换， $\tilde{\mathcal{Z}}_i = P_i(\mathcal{Z}_i)$ ，则

$$\tilde{\mathcal{Z}}_i \in \mathbb{R}^{p_1^{(i)} \times p_2^{(i)} \times \tilde{n}_i}, \quad p_1^{(i)} p_2^{(i)} = \prod_{r \neq i} n_r, \quad \tilde{n}_i = k n_i.$$

默认实验使用 $k = 1$ 。耦合变换块（coupled transformation block, CTB）的第一步对每个切片使用 3×3 卷积、零填充、步长 1 和带泄漏修正线性单元（leaky rectified linear unit, Leaky ReLU）：

$$\mathcal{U}_i(:, :, q) = \text{LReLU}\left(K_{i,q} *_{2D} \tilde{\mathcal{Z}}_i(:, :, q)\right), \quad q = 1, \dots, \tilde{n}_i, \quad (4.6)$$

再用 $W_{2,i} \in \mathbb{R}^{n_i \times \tilde{n}_i}$ 沿第 i 模态做管线映射：

$$\mathcal{V}_i = \text{LReLU}(\mathcal{U}_i \times_3 W_{2,i}), \quad \mathcal{X}_i = P_i^{-1}(\mathcal{V}_i). \quad (4.7)$$

因此 $K_{i,q}$ 负责局部空间结构， $W_{2,i}$ 负责被置换模态上的全局耦合，三个 $\mathcal{X}_i$ 最后才进入聚合器。该结构是多方向设计区别于仅沿光谱模态变换的关键。

聚合器沿第三模态连接三个候选结果：

$$\mathcal{X}_{\text{cat}} = \text{concat}_3(\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3) \in \mathbb{R}^{n_1 \times n_2 \times 3n_3}, \quad (4.8)$$

$$\mathcal{X}_{\text{sp}} = \text{LReLU}(w_{1,G}(\mathcal{X}_{\text{cat}})), \quad (4.9)$$

$$\hat{\mathcal{X}} = \text{LReLU}(\mathcal{X}_{\text{sp}} \times_3 W_{\text{agg}}), \quad W_{\text{agg}} \in \mathbb{R}^{n_3 \times 3n_3}. \quad (4.10)$$

这里的 1×1 映射相当于对每个空间位置的三个方向候选做光谱通道融合，因而聚合

方向选择会影响像素对齐。论文消融实验将 Mode-1、Mode-2 和 Mode-3 分别作为聚合方向，验证了 Mode-3 在给定实验设置下更稳定。

4.2.2 联合目标与梯度求解

令 $\mathcal{R}_i = \sum_{q=1}^{\tilde{n}_i} \|\tilde{Z}_i(:, :, q)\|_*$ ，三个方向的中间损失和最终损失可统一写为

$$\mathcal{L}_i = \mathcal{R}_i + \lambda_i \|\mathcal{P}_\Omega(\mathcal{X}_i - \mathcal{O})\|_F^2, \quad \mathcal{L} = \sum_{i=1}^3 \alpha_i \mathcal{L}_i + \|\mathcal{P}_\Omega(\hat{\mathcal{X}} - \mathcal{O})\|_F^2.$$

为减少方向权重的独立搜索，论文采用

$$(\alpha_1, \alpha_2, \alpha_3) = \frac{\beta}{2 + \vartheta} (1, 1, \vartheta), \quad \alpha_i \lambda_i = \gamma, \quad (4.11)$$

其中 β 、 ϑ 和 γ 分别调节低秩强度、空间—光谱平衡和中间分支观测一致性。实现时以 $\{\mathcal{Z}_i\}$ 、卷积与全连接参数以及聚合器参数为联合优化变量。对 $M = U_r \Sigma_r V_r^\top$ ，核范数次梯度可用

$$G_M = U_r V_r^\top + W, \quad U_r^\top W = 0, \quad W V_r = 0, \quad \|W\|_2 \leq 1,$$

并由链式法则传回潜张量；网络参数则直接使用自动微分。该求解路径没有交替方向乘子法（alternating direction method of multipliers, ADMM）的变量分裂，但仍是含核范数和非线性网络的非凸、非光滑优化问题，实验中的收敛应理解为达到损失平台的数值结果，而不是全局最优证明。

Algorithm 4.4 3DeepRep 的联合训练与停止规则**Require:** $\mathcal{O}, \Omega$; 潜维度比例 k ; β, ϑ, γ ; 最大迭代 $T_{\max}$; 停止阈值 τ_L, τ_X, q **Ensure:** $\hat{\mathcal{X}}$

- 1: 将像素归一化到 $[0, 1]$, 建立投影算子 $\mathcal{P}_\Omega$
- 2: 初始化 $\mathcal{Z}_1, \mathcal{Z}_2, \mathcal{Z}_3$; 论文默认使用变换域张量核范数 (transformed tensor nuclear norm, TNN) 结果预热初始化, 并记录随机、零填充或 TNN 初始化类型
- 3: 初始化 $\Theta_1, \Theta_2, \Theta_3, \Theta_G$, 并由式 (4.11) 计算系数
- 4: **for** $t = 1, \dots, T_{\max}$ **do**
- 5: **for** $i = 1, 2, 3$ **do**
- 6: 置换 $\tilde{\mathcal{Z}}_i = \mathcal{P}_i(\mathcal{Z}_i)$, 执行式 (4.6)–(4.7)
- 7: **end for**
- 8: 按式 (4.10) 聚合, 得到 $\hat{\mathcal{X}}$
- 9: 计算 $\mathcal{L}$ 、投影残差 $r_{\Omega, t}$ 和输出相对变化 η_t
- 10: 对 $\{\mathcal{Z}_i, \Theta_i, \Theta_G\}$ 反向传播, 并执行一次自适应矩估计 (adaptive moment estimation, Adam) 更新
- 11: **if** $|\mathcal{L}_t - \mathcal{L}_{t-1}| / (|\mathcal{L}_{t-1}| + \varepsilon) < \tau_L$ 且 $\eta_t < \tau_X$ 连续 q 次 **then**
- 12: **break**
- 13: **end if**
- 14: **end for**
- 15: 保存迭代数、随机种子、初始化、最终损失和 $r_{\Omega, t}$, 并返回 $\hat{\mathcal{X}}$

4.2.3 模型机制

3DeepRep^[19] 的主实验、消融和资源记录分别对应显式潜变量、多方向变换与聚合结构。表 4.2 汇总这些构造及其代表性结果。

表 4.2: 3DeepRep 的关键模型构造与实验结果。

模型构造	关键实验结果	作用分析
显式潜变量替代生成器	Pavia Centre/Case 4 中, Generated 配置为 36.823 dB、202.8 k 网络参数; 显式 Learnable 配置为 39.514 dB、112.1 k	显式潜变量缩短参数路径, 并在该设置下同时降低网络参数量、提高恢复指标
多方向深度变换	同一设置中, Mode-1、Mode-2、Mode-3 单方向分别为 30.838、21.907、33.939 dB, 多方向模型为 39.514 dB	三个方向为结构化退化提供互补信息; 显存和时间分别为 1.01 GB 和 1065 s
第三模非线性聚合	聚合方向消融中 Mode-1、Mode-2、Mode-3 分别为 33.912、22.623、39.514 dB	第三模聚合在该设置下保留空间位置对齐, 并取得最高 PSNR
复杂退化下的整体效果	Pavia Centre 连续坏列场景中 3DeepRep 与 CoNoT 分别为 51.164 dB 和 43.499 dB	多方向变换和非线性聚合共同改善连续坏列条件下的恢复质量
紧凑容量配置	$k = 0.3$ 时为 38.657 dB、48.6 k 参数、357 s; CoNoT 为 33.279 dB、54.8 k、633 s	该配置体现了显式潜变量在精度和计算资源之间的折中

上述结果表明, 显式潜变量提供了可直接正则化的低秩对象, 多方向变换在某一

方向受损时提供互补结构，非线性聚合将三个候选置于同一观测域目标中；参数量、运行时间和显存结果同时揭示了多方向结构的计算代价。

4.3 多方向模型的实验结果

本节介绍 3DeepRep 的实验设置和数值结果。模拟实验的像素值均归一化到 $[0, 1]$ ，恢复质量使用 PSNR、SSIM 和 SAM 评价；表中数值为 10 次重复实验的均值 $\pm$ 标准差，星号表示相对于一维方法 CoNoT 的配对 t 检验显著 ($p < 0.05$)。

4.3.1 数据集、退化模式和基线

论文使用三个标准高光谱数据集：Pavia Centre ($200 \times 200 \times 80$)、Pavia University ($256 \times 256 \times 103$) 和 Washington DC Mall ($256 \times 256 \times 191$)。四类模拟退化分别为：

Case 1: 随机像素缺失。四档采样比例 (sampling ratio, SR) 及对应缺失比例 (missing ratio, MR) 为

$$SR \in \{0.15, 0.10, 0.05, 0.01\}, \quad MR \in \{0.85, 0.90, 0.95, 0.99\}.$$

Case 2: 条带切片缺失。删除完整垂直空间条带，采用 $MR \in \{0.10, 0.30, 0.50\}$ 。

Case 3: 连续坏列缺失。在四组连续波段中设置 dead columns，宽度比取

$$\rho_{\text{dead}} \in \{0.1, 0.2, 0.3, 0.4\},$$

每个模式随机选取 6–10 条 deadline。

Case 4: 混合缺失。组合随机缺失 ($MR = 0.90$)、条带缺失 ($MR = 0.10$) 和上述连续坏列模式。

对比方法包括经典变换域张量核范数 (TNN)、多方向张量核范数 (three-directional tensor nuclear norm, 3DTNN)、加权和张量核范数 (weighted sum of tensor nuclear norm, WSTNN)、张量相关总变分 (tensor correlated total variation, t-CTV)、自监督非线性变换张量核范数 (self-supervised nonlinear transform-based tensor nuclear norm, S2NTNN)、耦合非线性变换低秩张量表示 (coupled nonlinear transform-based low-rank tensor representation, CoNoT)、多方向自监督非线性变换张量核范数 (three-directional self-supervised nonlinear transform-based tensor nuclear norm, 3DSTNN) 和基于改进扩散模型的无监督高光谱恢复 (hyperspectral image restoration via improved diffusion, HIR-Diff)，覆盖固定变换低秩模型、学习变换模型、多方向模型及扩散模型。

4.3.2 代表性数值结果

表 4.3 选取论文中的代表性实验场景，列出观测数据、CoNoT 以及 3DeepRep，覆盖随机极端缺失、条带缺失、deadline 缺失和混合退化。

表 4.3: 3DeepRep 在 Pavia Centre 上的代表性结果（论文表 2）

退化	MR	指标	观测	CoNoT	3DeepRep	单位	趋势
Case 1	0.99	PSNR	12.013	29.523	30.799 ± 0.369	dB	↑
		SSIM	0.004	0.885	0.915 ± 0.001	—	↑
		SAM	2.207	0.075	0.056 ± 0.001	rad	↓
Case 2	0.50	PSNR	14.940	25.963	30.237 ± 0.176	dB	↑
		SSIM	0.216	0.766	0.905 ± 0.001	—	↑
		SAM	1.571	0.043	0.036 ± 0.001	rad	↓
Case 3	0.10	PSNR	18.415	43.499	51.164 ± 0.213	dB	↑
		SSIM	0.648	0.995	0.999 ± 0.001	—	↑
		SAM	0.470	0.012	0.007 ± 0.001	rad	↓
Case 4	—	PSNR	12.274	33.279	39.648 ± 0.109	dB	↑
		SSIM	0.024	0.958	0.987 ± 0.001	—	↑
		SAM	1.512	0.042	0.021 ± 0.000	rad	↓

Pavia Centre 的 Case 3 体现了多方向设计的动机：deadline 破坏光谱连续性后，单一光谱方向先验难以工作，空间行和列方向的低秩约束仍可从未受损区域传播结构信息。

在另外两个数据集上，Pavia University 的 Case 4 中 3DeepRep 的 PSNR、SSIM、SAM 为 38.060 ± 0.288 dB、 0.959 ± 0.000 、 0.054 ± 0.001 ；Washington DC Mall 的 Case 4 中对应数值为 46.369 ± 0.170 dB、 0.977 ± 0.001 、 0.036 ± 0.001 ，展示了模型在不同空间尺寸和光谱波段数下的恢复效果。

4.3.3 消融、稳定性与效率

论文在 Pavia Centre 的 Case 4 上验证了显式潜张量、聚合方向和初始化的作用：

表 4.4: 3DeepRep 核心模块消融（Pavia Centre, Case 4）

模块	设置	PSNR/dB	SSIM	SAM	网络参数
潜变量策略	Generated	36.823	0.980	0.025	202.8 k
	Learnable（本文）	39.514	0.987	0.021	112.1 k
聚合方向	Mode-1	33.912	0.952	0.054	216.1 k
	Mode-2	22.623	0.684	0.241	216.1 k
	Mode-3（本文）	39.514	0.987	0.021	112.1 k
初始化	Random $U(0, 1)$	39.078	0.986	0.022	112.1 k
	Observed（零填充）	39.302	0.987	0.022	112.1 k
	TNN 结果	39.514	0.987	0.021	112.1 k

直接优化潜张量使网络参数由 202.8 k 减少到 112.1 k，PSNR 从 36.823 dB 提高到 39.514 dB；Mode-3 聚合的 PSNR 高于 Mode-1 和 Mode-2；三种初始化的 PSNR 均在 39.078–39.514 dB 之间。

方向、潜维度和实际运行成本列于表 4.5。完整多方向模型使用约 20 k 次迭代，运行时间为 1065 s（约 18 min）；增加 k 超过 1 后精度没有提高，却明显增加显存和时间。

表 4.5: 3DeepRep 的方向和潜维度效率结果 (Pavia Centre, Case 4)

设置	PSNR/dB	SSIM	SAM	参数	图形处理器 (graphics processing unit, GPU) 内存	时间/s
Mode-1 only	30.838	0.925	0.077	41.8 k	0.30 GB	278
Mode-2 only	21.907	0.647	0.266	41.8 k	0.30 GB	273
Mode-3 only	33.939	0.961	0.036	7.1 k	0.30 GB	493
3D (本文)	39.514	0.987	0.021	112.1 k	1.01 GB	1065
$k = 0.5$	39.079	0.987	0.022	66.7 k	0.85 GB	533
$k = 2.0$	38.913	0.986	0.022	202.8 k	1.31 GB	2140
$k = 3.0$	38.601	0.986	0.022	293.5 k	1.62 GB	2975
$k = 4.0$	38.310	0.984	0.022	384.2 k	1.96 GB	3600

论文还报告 $k = 0.3$ 的紧凑配置: 3DeepRep 参数量 48.6 k、时间 357 s、PSNR 38.657 dB; CoNoT 参数量 54.8 k、时间 633 s、PSNR 33.279 dB, 体现显式潜变量在精度和计算资源之间的折中。

4.3.4 真实退化实验与适用条件

中分辨率成像光谱仪 (Moderate Resolution Imaging Spectroradiometer, MODIS) 一级 B 产品和陆地卫星 7 增强型专题制图仪 (Landsat 7 Enhanced Thematic Mapper Plus, Landsat 7 ETM+) 用于检验真实退化下的结构恢复。MODIS 裁剪为 $256 \times 256 \times 10$, 第 27 波段有明显条带噪声; Landsat 7 ETM+ 裁剪为 $256 \times 256 \times 7$, 存在跨波段一致的斜条带缺失。可视化结果显示, 3DeepRep 能减弱条带伪影并保留局部结构。

极端复合退化会使问题病态; 若同一空间位置在所有波段都缺失, 光谱相关性无法提供有效信息。因此, 3DeepRep 适用于至少一个模态保留足够结构相关性的情形, 观测率也需要支持潜在低秩结构的辨识。

4.4 面向整片缺失的连续坐标低秩表示

3DeepRep 解决的是离散张量中多个变换方向如何协同; LRTubalRep 进一步处理一种不同的观测困难: 完整第三模态切片 (例如整个光谱波段、视频帧或时间截面) 同时缺失时, 该坐标没有任何像素残差; 这一问题与嵌入空间低秩模型所讨论的缺失切片恢复情形相对应^[22]。若仍为每个坐标设置彼此独立的离散参数, 缺失坐标的梯度恒为零; 若用一个向量输出网络同时承担坐标变化和空间系数组合, 其函数容量与空间容量又被同一输出宽度绑定。LRTubalRep 将第三模态改写为共享参数的矩阵值连续场, 并通过显式投影解耦连续原子数 M 、空间系数宽度 d 和单个系数矩阵秩 r 。这一表示吸收了层次低秩分解与函数变换张量方法关于“低秩系数-可学习函数”的基本思想^[10-12], 并重新定义坐标共享和容量解耦方式。

本节围绕三个问题展开: 整片缺失为什么不同于条目缺失, 观测坐标如何间接更新缺失坐标的输出, 以及 (M, d, r) 为何需要独立控制。其问题-构造对应关系见表 4.6。

表 4.6: LRTubalRep 所解决的问题与对应表示机制。

问题	构造与信息通路	结构作用
缺失切片没有直接数据梯度	用共享参数 Θ 定义连续矩阵值场 $\hat{X}_\Theta(z)$; 观测坐标更新同一组参数, 查询时在缺失坐标复用	观测坐标的误差通过共享参数传递到缺失坐标
坐标函数容量与空间组合容量耦合	M 个标量连续原子先经 $C \in \mathbb{R}^{M \times d}$ 投影, 再组合 d 个秩不超过 r 的空间系数矩阵	(M, d, r) 分别控制原子数量、系数宽度和空间因子秩
内部插值与边界外推混杂	显式区分观测坐标凸包内部缺口与首端/尾端缺口	内部缺口属于插值, 端点缺口属于相对于观测凸包的外推

为使整片缺失场景下的参数迁移路径可视化, 图 4.2 给出 LRTubalRep 的连续坐标表示框架。

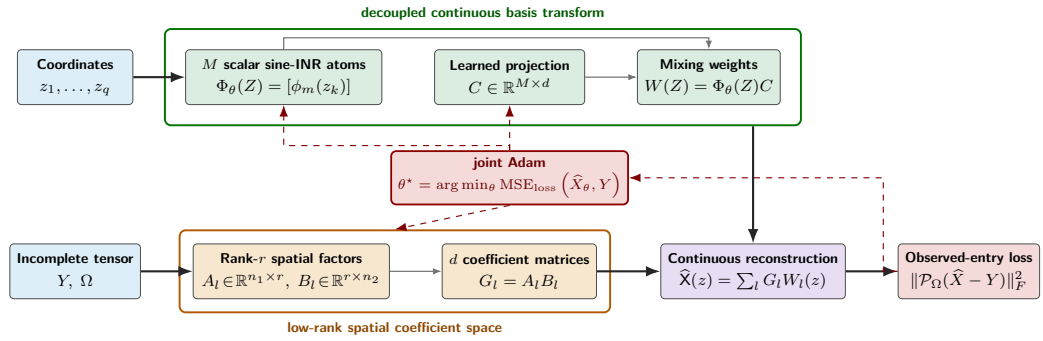


图 4.2: 低管秩连续坐标表示模型的共享参数与低秩空间系数框架。

图 4.2 的输入是观测切片及其归一化坐标；共享的连续原子网络先将第三模坐标映射为 M 个标量响应，再由投影矩阵压缩到宽度为 d 的系数通道，并与前两模态上的低秩空间系数相乘，输出查询坐标处的矩阵切片。观测切片通过均方误差更新共享参数，缺失切片在推断阶段复用同一连续场；图中的原子数、投影宽度和空间秩分别对应式 (4.15)、式 (4.16) 和式 (4.24)，训练入口对应式 (4.20)。该图刻画了“观测坐标—共享参数—缺失坐标”的信息传递机制，缺失切片由共享场生成而不引入额外监督信号。

4.4.1 整片观测几何与零直接梯度

设高光谱（或时间序列）数据为 $\mathcal{X}^* \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ 。将前两模态合并为矩阵值函数，把第三模态的采样索引映射到坐标 $z_k = k/n_3$ ：

$$\mathbf{X} : \mathcal{Z} \rightarrow \mathbb{R}^{n_1 \times n_2}, \quad \mathcal{X}_{:::,k}^* = \mathbf{X}(z_k), \quad \mathcal{Z} = \left[\frac{1}{n_3}, 1 \right]. \quad (4.12)$$

代码中的正弦原子不含偏置，因而 $\phi_m(0) = 0$ ；为避免把实现外推到一个具有结构性零值的未训练端点，当前模型的声明查询域是 $[1/n_3, 1]$ ，而不是 $[0, 1]$ 。令 $\mathcal{O}$ 为可观

测切片索引, $\Gamma = \{1, \dots, n_3\} \setminus \mathcal{O}$ 为整片缺失索引, 则

$$\mathcal{M}_{ijk} = \begin{cases} 1, & k \in \mathcal{O}, \\ 0, & k \in \Gamma, \end{cases} \quad \mathcal{Y} = \mathcal{M} \odot \mathcal{X}^*.$$

因此 $k \in \Gamma$ 时 $\mathcal{M}_{::k} = 0$, 训练阶段不计算该切片的像素残差。若缺失坐标落在观测坐标的凸包内, 它相对于已有切片属于插值; 首端或尾端连续缺失虽然仍在 $\mathcal{Z}$ 内, 却属于相对于观测凸包的边界外推。这两种几何位置对应不同的连续场查询方式。

离散的、每个切片独立的表项 $D_{k:}$ 在观测均方误差 $|\mathcal{O}|^{-1} \sum_{k \in \mathcal{O}} \ell_k(D_{k:})$ 下满足

$$\frac{\partial \mathcal{L}}{\partial D_{k:}} = 0, \quad k \in \Gamma. \quad (4.13)$$

LRTubalRep 的信息通路不是给缺失切片伪造标签, 而是让观测坐标更新在所有坐标复用的参数上发生。对 Θ 的一次梯度更新, 缺失坐标 z_* 的一阶变化可写成

$$\Delta x_* \simeq -\frac{\eta}{|\Omega|} \sum_{k \in \mathcal{O}} J_{\Theta} x_* J_{\Theta} x_k^\top \nabla_{x_k} \ell_k(x_k), \quad x_k = \text{vec}(\hat{X}(z_k)). \quad (4.14)$$

式 (4.14) 刻画了共享参数更新对缺失坐标输出的链式影响; 该影响由底层场、缺口位置和观测量共同决定, 多个连续场也可能在观测坐标上取得相同的损失。

4.4.2 低秩空间系数、投影字典与张量形状

令系数宽度

$$d = \lfloor n_3 \rho \rfloor, \quad \rho > 0, \quad (4.15)$$

并学习两个因子张量

$$A \in \mathbb{R}^{n_1 \times r \times d}, \quad B \in \mathbb{R}^{r \times n_2 \times d}.$$

对 $l = 1, \dots, d$ 定义

$$G_l = A_{::,l} B_{::,l} \in \mathbb{R}^{n_1 \times n_2}, \quad (G_l)_{ij} = \sum_{s=1}^r A_{isl} B_{sjl}, \quad \text{rank}(G_l) \leq r.$$

这里的 r 是每一个空间系数矩阵的因子秩, 而不是重建张量的经典 tubal rank。对坐标 z , 先评估 M 个标量连续原子

$$\Phi(z) = \begin{bmatrix} \phi_1(z) & \cdots & \phi_M(z) \end{bmatrix} \in \mathbb{R}^{1 \times M}, \quad W(z) = \Phi(z)C \in \mathbb{R}^{1 \times d}, \quad C \in \mathbb{R}^{M \times d}, \quad (4.16)$$

再合成

$$\hat{X}(z) = \sum_{l=1}^d W_l(z) G_l = \sum_{l=1}^d W_l(z) A_{::,l} B_{::,l}. \quad (4.17)$$

对查询网格 $Z = (z_1, \dots, z_q)$, 令 $\Phi(Z) \in \mathbb{R}^{q \times M}$, 并按行堆叠 $G_{(3)} \in \mathbb{R}^{d \times n_1 n_2}$, 则前向过程可压缩为

$$\hat{\mathcal{X}}_{(3)}(Z) = \Phi(Z) C G_{(3)} \in \mathbb{R}^{q \times n_1 n_2}. \quad (4.18)$$

上述分解有三个互相独立的容量旋钮： M 控制连续原子个数， d 控制可复用空间系数的宽度， r 控制每个空间系数的矩阵因子秩。 M 与 d 解耦后，增加连续函数的数量不必同步改变空间分支的宽度；但每个独立原子仍有自己的参数，所以 M 增大时函数分支参数量近似线性增加。

4.4.3 连续原子族与无偏置边界

模型使用独立参数化的正弦表示网络（sinusoidal representation network, SIREN）风格的正弦原子；周期激活隐式表示与高频坐标特征的设计依据分别来自相关工作^[8,9]。对第 m 个原子，令 $h_m^{(0)}(z) = z$ ，则一个输入正弦层、两个宽度为 64 的隐藏正弦层和无偏置标量线性输出可写为

$$\begin{aligned} h_m^{(t)}(z) &= \sin(\omega Q_m^{(t)} h_m^{(t-1)}(z)), \quad t = 1, 2, 3, \\ \phi_m(z) &= u_m^\top h_m^{(3)}(z), \quad \omega = 2, \end{aligned} \quad (4.19)$$

其中 $Q_m^{(1)} \in \mathbb{R}^{64 \times 1}$, $Q_m^{(2)}, Q_m^{(3)} \in \mathbb{R}^{64 \times 64}$, $u_m \in \mathbb{R}^{64}$ 。不同 m 的 $Q_m^{(t)}$ 和 u_m 不共享，模型将它们沿原子维度堆叠后用批量矩阵乘法计算。由于没有偏置，

$$\phi_m(0) = 0, \quad \hat{X}(0) = 0 \quad (\text{对所有参数值成立}).$$

因此模型的默认查询域为 $[1/n_3, 1]$ ；若需覆盖 $z = 0$ ，应采用坐标平移或带偏置原子，使端点值由模型参数决定。

模型还提供可选的一维高斯原子：

$$\phi_m^G(z) = \exp\left[-\frac{(z - \mu_m)^2}{2\sigma_m^2}\right], \quad \mu_m \in (0, 1), \quad \sigma_m \geq 10^{-4}.$$

高斯中心采用有界参数化，宽度带有下限；该原子族可替换正弦族而不改变投影、合成和观测损失的计算形式。

4.4.4 目标函数与自适应矩估计求解

模型只在观测切片计算均方误差。令 $\Omega = \{(i, j, k) : k \in \mathcal{O}\}$ ，则代码的核心目标为

$$\min_{\Theta} \mathcal{L}_{\text{obs}}(\Theta) = \frac{1}{|\Omega|} \left\| \mathcal{P}_{\Omega} \left(\hat{\mathcal{X}}_{\Theta} - \mathcal{Y} \right) \right\|_F^2, \quad \Theta = (A, B, C, \{Q_m^{(t)}, u_m\}_{m,t}). \quad (4.20)$$

主配置为 Adam、学习率 10^{-3} 、权重衰减 10^{-7} 、最大迭代 6001，不启用余弦学习率。训练结束后先用最终参数重算全部查询坐标，再将观测位置硬拷贝为 $\mathcal{Y}$ ，以保证返回结果在已观测条目上严格一致。

训练代码还保留一个显式的可选因子总变差项。当且仅当配置 $\gamma > 0$ 时才计算

$$\mathcal{L}_{\text{TV}} = \gamma \left(\|A_{1,:,:,} - A_{0,-1,:,:,}\|_1 + \|B_{:,1,:,:,} - B_{:,0,-1,:,:,}\|_1 \right), \quad \mathcal{L} = \mathcal{L}_{\text{obs}} + \mathcal{L}_{\text{TV}}. \quad (4.21)$$

主配置取 $\gamma = 0$ ，所以目标函数不包含该项，也没有额外的正交、Lipschitz 或显式因

子核范数约束。对所有可训练参数统一记为 ϑ ，Adam 的逐参数更新可写成

$$\begin{aligned} g_t &= \nabla_{\vartheta} \mathcal{L}_t, & m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t, & v_t &= \beta_2 v_{t-1} + (1 - \beta_2) g_t^2, \\ \hat{m}_t &= m_t / (1 - \beta_1^t), & \hat{v}_t &= v_t / (1 - \beta_2^t), \\ \vartheta_{t+1} &= \vartheta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \quad (\text{并按实现的 weight decay 更新}). \end{aligned}$$

迭代由预设的最大迭代次数控制，达到 `max_iter` 后结束。

4.4.5 尺度规范化、收缩顺序与实现复杂度

因子化存在规范自由度 $A_{:::,l} B_{:::,l} = (s_l A_{:::,l})(B_{:::,l}/s_l)$ 。当 $\gamma = 0$ 时，代码默认每 25 个优化步执行一次不改变 G_l 的平衡：

$$s_l = \sqrt{\frac{\|B_{:::,l}\|_F}{\|A_{:::,l}\|_F + \epsilon}}, \quad A_{:::,l} \leftarrow s_l A_{:::,l}, \quad B_{:::,l} \leftarrow B_{:::,l}/s_l, \quad (4.22)$$

实际实现对比值做 ϵ 保护和 10^6 的缩放上限，并同步变换 Adam 状态；这使表示本身不变，但会影响后续浮点优化轨迹。空间分支使用 Xavier 初始化，正弦网络和 C 分别按代码中的均匀/Xavier 规则初始化。

矩阵结合律允许两种完全等价的收缩顺序。令 $N_p = n_1 n_2$ ，查询数为 q ，则不计公共的原子评估成本时，直接先形成 $W = \Phi C$ 的乘法量代理为

$$c_{\text{dir}} = qMd + N_p qd, \quad c_{\text{dict}} = N_p Md + N_p qM.$$

自动模式仅当 $c_{\text{dict}} \leq 0.9c_{\text{dir}}$ 时采用 $\Phi(CG_{(3)})$ ，否则采用 $(\Phi C)G_{(3)}$ ；因此内存和时间记录必须同时报告查询数、 M, d 以及实际收缩顺序。训练主项不需要每一步 SVD，主要代价来自批量因子乘法、原子前向和收缩。

若每个正弦原子有三层宽度 64 和无偏置输出，则其函数参数量为 $64 + 2 \times 64^2 + 64 = 8320$ ；故忽略少量缓冲区时，总参数量可审计为

$$P_{\text{sin}} \simeq rd(n_1 + n_2) + Md + 8320M, \quad (4.23)$$

高斯原子分支将 $8320M$ 替换为约 $2M$ 。式 (4.23) 是结构计数，不等于某次运行日志中的所有优化器状态内存。

4.4.6 训练与缺失切片推断

算法 4.5 给出输入检查、联合反向传播、自适应收缩、因子平衡和观测值回填顺序。算法中的 $\gamma = 0$ 对应主配置；只有显式打开时才加入式 (4.21)。

Algorithm 4.5 LRTubalRep 观测坐标拟合、平衡与缺失切片推断

Require: 观测张量 $\mathcal{Y}$ 、掩膜 $\mathcal{M}$ 、坐标 $Z = (z_1, \dots, z_{n_3})$ ；容量 (r, M, ρ) ；步长 η ；最大迭代 T ；平衡周期 τ ；收缩阈值 κ

Ensure: 完整估计 $\tilde{\mathcal{X}}$

- 1: 检查 $\mathcal{Y}$ 、 $\mathcal{M}$ 形状相同，建立 $\Omega = \{(i, j, k) : \mathcal{M}_{ijk} = 1\}$
- 2: 计算 $d = \lfloor n_3 \rho \rfloor$ ，以 Xavier 方式初始化 A, B, C 和 M 个正弦原子参数
- 3: 形成 $G_{(3)}$ ；对每个 k 计算 $\Phi(Z)_{km} = \phi_m(z_k)$
- 4: **for** $t = 1, \dots, T$ **do**
- 5: 计算 $c_{\text{dir}}, c_{\text{dict}}$ ，按 $c_{\text{dict}} \leq \kappa c_{\text{dir}}$ 选择收缩顺序
- 6: 用式 (4.18) 生成所有查询切片 $\hat{\mathcal{X}}^t$
- 7: 仅在 Ω 上计算 $\mathcal{L}_{\text{obs}}$ ；若 $\gamma > 0$ ，加上 $\mathcal{L}_{\text{TV}}$
- 8: 反向传播并执行一次 Adam 更新；记录损失、梯度范数和参数范数
- 9: **if** $t \bmod \tau = 0$ 且 $\gamma = 0$ **then**
- 10: 对每个 l 按式 (4.22) 平衡 A_l, B_l ，并同步变换对应 Adam 状态
- 11: **end if**
- 12: **end for**
- 13: 用最终参数重算 $\hat{\mathcal{X}}$ ，并令 $\tilde{\mathcal{X}}_\Omega \leftarrow \mathcal{Y}_\Omega$
- 14: 在 $k \in \Gamma$ 上读取 $\tilde{\mathcal{X}}_{:,k} = \hat{\mathcal{X}}(z_k)$ ；按 Γ 的内部/首端/尾端类别报告
- 15: **return** $\tilde{\mathcal{X}}$

4.4.7 表示性质与秩界

由矩阵乘积的秩不增性质，对任意 q 个查询坐标有

$$\text{rank}(W(Z)) \leq \min\{M, d, q\}, \quad \text{rank}(\hat{\mathcal{X}}_{(3)}(Z)) \leq \min\{M, d, q, n_1 n_2\}. \quad (4.24)$$

固定坐标的切片是 d 个秩至多为 r 的矩阵之和，因此

$$\text{rank}(\hat{\mathcal{X}}(z)) \leq \min\{n_1, n_2, rd\}.$$

该上界在 $rd \geq \min(n_1, n_2)$ 时可能为空；它控制的是系数表示，不能改写为“重建张量 tubal rank 不超过 r ”。将投影展开还可得到完全等价的连续字典形式

$$\hat{\mathcal{X}}(z) = \sum_{m=1}^M H_m \phi_m(z), \quad H_m = \sum_{l=1}^d C_{ml} G_l, \quad \text{rank}(H_m) \leq \min\{n_1, n_2, rd\}. \quad (4.25)$$

正弦网络在有界参数和紧坐标区间上是连续且局部 Lipschitz 的；若进一步给定参数范数上界 L_ϕ 、系数矩阵范数上界 L_G ，则

$$\|\hat{\mathcal{X}}(z) - \hat{\mathcal{X}}(z')\|_F \leq \left(\sum_{l=1}^d \|G_l\|_F \right) \|C\|_2 \sqrt{M} L_\phi |z - z'|. \quad (4.26)$$

这是条件性表示界，不是训练后的泛化误差界。特别地，边界整片缺失属于外推；越过最后一个观测坐标的距离、正弦频率和参数范数均会影响稳定性。由于训练目标只

在观测坐标上计算，式 (4.26) 仅刻画共享连续场的局部变化上界，不能直接推出跨数据集或跨时间域的精确恢复。

模型适用性。 LRTubalRep 的训练损失只在观测切片上计算，缺失坐标通过共享连续场获得输出。内部缺口对应观测坐标凸包内的插值，首端或尾端缺口对应观测凸包外的外推；两类查询对连续原子的频率、参数范数和缺口长度具有不同敏感性。该模型的分析重点是共享坐标参数和低秩空间系数的构造，以及由此形成的观测到缺失坐标的信息通路。

LRTubalRep 与本章前面的 3DeepRep 形成离散—连续的递进关系：3DeepRep 在三个离散方向上联合变换并聚合，LRTubalRep 把一个坐标方向变成共享连续场；其后 CA-TNN 仍在离散变换域工作，但学习通道级正则权重。三者分别回答“方向如何协同”“整片缺失如何通过坐标共享传递”“不同通道如何自适应加权”三个层次的问题。

4.5 面向潜在通道差异的自适应核范数

CA-TNN^[20] 处理的不是“是否使用加权核范数”，而是权重应作用于什么对象。传统 WTNN 通常按奇异值的大小或次序赋权；在可学习变换中，奇异值分布会随网络参数改变，而且同一次序的奇异值未必代表不同潜在通道中相同的信息成分。该模型由此把加权对象从单个奇异值提升为整个潜在通道的核范数，使每个通道获得统一、数据依赖的正权。

令 $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \times \tilde{n}_3}$ 为显式潜张量， $\hat{\mathcal{X}} = g(\mathcal{Z}; \Theta_g)$ 为映射到观测空间的恢复结果。对第 k 个潜在通道先执行空间平均池化

$$s_k = \frac{1}{n_1 n_2} \sum_{l=1}^{n_1} \sum_{j=1}^{n_2} Z(l, j, k), \quad k = 1, \dots, \tilde{n}_3, \quad (4.27)$$

再由压缩—激励网络生成正权向量

$$\mathbf{a} = \exp(\mathbf{b}_2) + \exp(W_2) \sigma(W_1 \mathbf{s}) \in (0, +\infty)^{\tilde{n}_3}, \quad (4.28)$$

其中指数按元素计算， σ 为 sigmoid。相应的通道感知核范数为

$$\|\mathcal{Z}\|_{\text{CA-TNN}} = \sum_{k=1}^{\tilde{n}_3} a_k \|Z^{(k)}\|_*$$

于是缺失观测上的自监督目标写成

$$\min_{\mathcal{Z}, \Theta_g, \Theta_w} \|\mathcal{P}_\Omega(g(\mathcal{Z}; \Theta_g) - \mathcal{O})\|_F^2 + \lambda \sum_{k=1}^{\tilde{n}_3} a_k(\mathcal{Z}; \Theta_w) \|Z^{(k)}\|_*, \quad (4.29)$$

其中 $\Theta_w = \{W_1, W_2, \mathbf{b}_2\}$ 。与逐奇异值加权相比，该构造显式建模潜在通道之间的整体差异；正权没有 softmax 归一化和预设上界，因而同时调节通道相对重要性和正则项相对于数据保真项的尺度。

为直观呈现通道权重如何嵌入低秩恢复流程，图 4.3 给出 CA-TNN 的空间-光谱变换、通道统计和自监督优化框架。

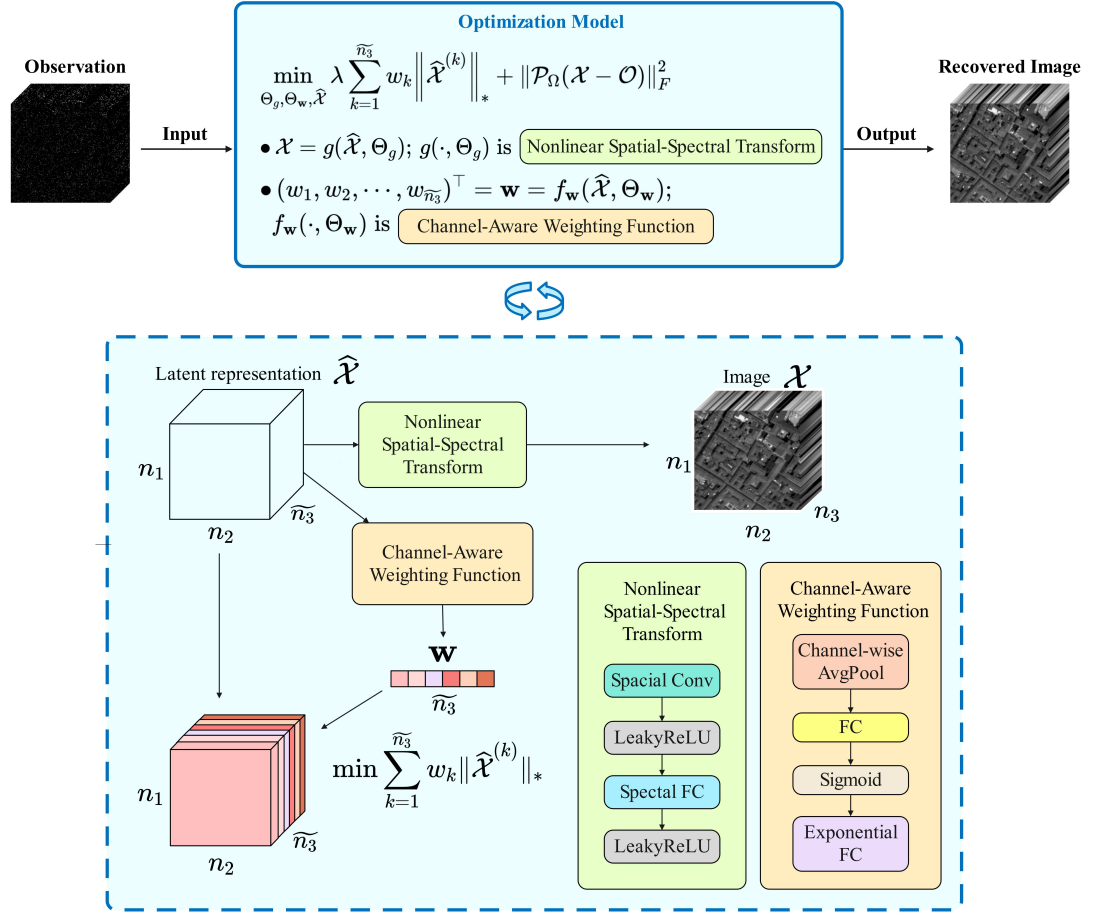


图 4.3: 通道感知张量核范数模型的空间-光谱变换与自适应加权框架。

图 4.3 的输入是退化观测和显式潜张量；空间变换先对各潜在通道提取局部结构，光谱变换再把潜通道映射回观测波段，同时由全局平均池化得到通道描述向量并通过压缩-激励模块产生正权。最后，通道核范数作为正则项与观测域保真项共同优化，输出恢复张量。图中的通道统计、权重生成、空间-光谱变换和观测损失分别对应式 (4.27)、式 (4.28)、式 (4.31) 和式 (4.29)；图后续的链式梯度与训练算法则对应权重对潜变量的反向传播。该图说明 CA-TNN 学习的是“通道级正则分配”，并不等同于对每个奇异值独立排序加权。

4.5.1 空间-光谱变换

为明确各变量的尺寸，令潜张量 $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \times \tilde{n}_3}$ ，空间变换对每个潜通道使用 3×3 卷积、零填充、步长 1 和 LeakyReLU：

$$g_1(\mathcal{Z})(:, :, k) = \text{LReLU}(W_{3,k} *_{2D} \mathcal{Z}(:, :, k)), \quad W_{3,k} \in \mathbb{R}^{3 \times 3}. \quad (4.30)$$

令 $\mathcal{U} = g_1(\mathcal{Z})$ ，光谱变换用 $W_4 \in \mathbb{R}^{n_3 \times \tilde{n}_3}$ 沿第三模态映射回原始波段数：

$$g(\mathcal{Z}; \Theta_g) = g_2(\mathcal{U}) = \text{LReLU}(\mathcal{U} \times_3 W_4). \quad (4.31)$$

通道描述向量由空间平均池化产生, $\mathbf{s} \in \mathbb{R}^{\tilde{n}_3}$; 指数激励模块可更完整地写为

$$\mathbf{a} = \exp(\mathbf{b}_2) + \exp(W_2)\sigma(W_1\mathbf{s}), \quad \mathbf{a} \in (0, +\infty)^{\tilde{n}_3}, \quad (4.32)$$

其中 σ 为 sigmoid。由于 $\mathbf{a}$ 不经过 softmax 归一化, 权重网络可以在优化中整体降低或提高正则化规模, 故它同时改变通道相对重要性和正则项相对于数据保真项的尺度。

4.5.2 联合目标与自监督求解

将通道权重代入观测投影损失, CA-TNN 的求解问题为

$$\min_{\mathcal{Z}, \Theta_g, \Theta_w} \mathcal{L}_{\text{CA}} = \frac{1}{2} \|\mathcal{P}_\Omega(g(\mathcal{Z}; \Theta_g) - \mathcal{O})\|_F^2 + \lambda \sum_{k=1}^{\tilde{n}_3} a_k(\mathcal{Z}; \Theta_w) \|Z(:, :, k)\|_*. \quad (4.33)$$

令 $\hat{\mathcal{X}} = g(\mathcal{Z}; \Theta_g)$ 。潜张量、空间卷积与光谱映射参数、通道权重参数在一个反向传播循环中联合更新。对 $\xi \in \{\mathcal{Z}, \Theta_g\}$, 数据项的梯度为

$$\nabla_\xi \frac{1}{2} \|\mathcal{P}_\Omega(\hat{\mathcal{X}} - \mathcal{O})\|_F^2 = J_\xi(\hat{\mathcal{X}})^\top \mathcal{P}_\Omega(\hat{\mathcal{X}} - \mathcal{O}),$$

而 Θ_w 不直接进入 g , 其梯度仅由加权核范数产生。由于权重没有归一化或下界, 若所有 a_k 同时趋近于零, 正则项会发生尺度退化; 因此模型训练需要同时控制权重范围和正则项与保真项的比例。

Algorithm 4.6 CA-TNN 的详细自监督训练流程

Require: 观测张量 $\mathcal{O}$ 、掩膜 Ω 、正则化系数 λ 、最大迭代 $T_{\max}$

Ensure: 恢复结果 $\hat{\mathcal{X}}$

- 1: 初始化潜张量 $\mathcal{Z}$ 、 $\Theta_g = \{W_{3,k}, W_4\}$ 和 $\Theta_w = \{W_1, W_2, \mathbf{b}_2\}$
 - 2: **for** $t = 1, \dots, T_{\max}$ **do**
 - 3: 由式 (4.27) 计算通道描述 $\mathbf{s}$
 - 4: 由式 (4.32) 计算正权重 $\mathbf{a}$
 - 5: 由式 (4.30)–(4.31) 得到 $\hat{\mathcal{X}}$
 - 6: 计算式 (4.33), 并记录投影残差和通道权重范围
 - 7: 对 $\mathcal{Z}, \Theta_g, \Theta_w$ 反向传播并执行一次 Adam 更新
 - 8: **if** 损失在后半段迭代中不再明显下降, 或达到最大迭代数 **then**
 - 9: **break**
 - 10: **end if**
 - 11: **end for**
 - 12: **return** $\hat{\mathcal{X}}$
-

4.5.3 通道级加权实验

CA-TNN^[20] 在 Pavia Centre 和 Washington DC Mall 两个高光谱数据集上采用随机像素缺失, 缺失率为 0.90、0.95 和 0.99, 并与 TNN、WSTNN、S2NTNN、CoNoT 和

HIR-Diff 比较。表 4.7 给出代表性条件下的结果：CA-TNN 同时提高 PSNR、SSIM 并降低 SAM。

表 4.7: CA-TNN 论文的代表性结果（CoNoT 与 CA-TNN）。

数据集	MR	PSNR/dB $\uparrow$		SSIM $\uparrow$		SAM $\downarrow$	
		CoNoT	CA-TNN	CoNoT	CA-TNN	CoNoT	CA-TNN
Pavia	0.99	29.523	29.815	0.885	0.895	0.075	0.074
Pavia	0.95	40.456	42.023	0.989	0.993	0.026	0.022
DC	0.99	43.150	44.161	0.940	0.958	0.090	0.072
DC	0.90	57.666	60.694	0.998	0.999	0.019	0.014

Pavia、MR = 0.95 的消融结果见表 4.8。通道级加权 ChannelW 高于不加权 WoW 和传统奇异值加权 SingularW；指数激励 ExpSE 也高于 sigmoid、softmax 和归一化指数变体。该结果对应模型的两个设计：将加权对象由单个奇异值转向潜在通道，并使用非归一化正权扩大正则分配范围。

表 4.8: CA-TNN 的通道加权与激励函数消融（Pavia, MR = 0.95）。

消融类别	设置	PSNR/dB $\uparrow$	SSIM $\uparrow$	SAM $\downarrow$
加权对象	WoW	40.661	0.990	0.026
	SingularW	39.270	0.986	0.032
	ChannelW + SingularW	41.173	0.991	0.023
	ChannelW	42.023	0.993	0.022
激励函数	SigmoidSE	37.443	0.977	0.040
	SoftmaxSE	40.878	0.990	0.025
	N-ExpSE	40.127	0.989	0.028
	ExpSE	42.023	0.993	0.022

上述结果表明，通道级核范数和指数激励能够在随机像素缺失条件下改善恢复质量；结构化缺失和真实传感器退化需要结合相应观测模型重新设置正则化对象与权重机制。

4.6 本章小结

3DeepRep、LRTubalRep 和 CA-TNN 均采用观测域自监督。对训练状态 Θ_t ，观测残差及三类结构量可写为

$$\begin{aligned}
 r_{\Omega}(t) &= \frac{\|\mathcal{P}_{\Omega}(\hat{\mathcal{X}}_t - \mathcal{O})\|_F}{\|\mathcal{P}_{\Omega}\mathcal{O}\|_F + \varepsilon}, \\
 r_{3D}(t) &= \sum_{i=1}^3 \sum_{q=1}^{\tilde{n}_i} \|\tilde{Z}_i^t(:, :, q)\|_*, \\
 r_{\text{bal}}(t) &= \max_{1 \leq l \leq d} \max \left\{ \frac{\|A_l^t\|_F}{\|B_l^t\|_F + \varepsilon}, \frac{\|B_l^t\|_F}{\|A_l^t\|_F + \varepsilon} \right\}, \\
 r_{\text{ch}}(t) &= \left(\min_k a_k^t, \frac{1}{\tilde{n}_3} \sum_k a_k^t, \max_k a_k^t \right).
 \end{aligned}$$

其中, r_{3D} 表示 3DeepRep 三个方向的低秩代价, r_{bal} 表示 LRTubalRep 的因子尺度平衡, r_{ch} 表示 CA-TNN 权重范围及整体尺度。LRTubalRep 的缺失切片可按内部插值和边界外推区分; CA-TNN 则同时使用加权核范数和观测保真项描述训练状态。当缺失位置没有参考真值时, 上述量描述观测一致性和结构正则, 而不是缺失位置的误差。三类目标均含非凸网络或多线性耦合, 损失平台表示训练过程趋于稳定, 不等同于全局最优。

本章以“固定表示不足时, 哪些接口应由数据学习”为主线, 形成方向协同、坐标共享和通道自适应三个递进层次。3DeepRep 通过显式潜张量、多方向耦合变换和非线性聚合解决单一方向在复杂空间-光谱退化下覆盖不足的问题; 论文实验在三个数据集、四类退化下给出主结果, 并以显式潜变量、方向和聚合消融说明模型效果来自哪些构造, 同时报告相应的显存与时间代价。

LRTubalRep 面向整个第三模态切片缺失时“该坐标没有直接训练梯度”的问题, 把离散表项改为共享参数的矩阵值连续场, 并用投影矩阵解耦连续原子数 M 、空间系数宽度 d 和单系数秩 r 。共享连续场使观测坐标的误差能够通过网络参数传递到缺失坐标, 三个容量参数则分别控制坐标函数、空间系数和单系数矩阵的复杂度。

CA-TNN 则解决传统 WTNN 只按奇异值加权、难以显式反映潜在通道差异的问题。其通道级核范数和非归一化指数激励在随机像素缺失实验中改善了恢复结果。三项工作共同说明, 数据驱动低秩表示并非取消结构先验, 而是保留可追踪的低秩潜变量, 同时学习方向组织、跨坐标共享或通道正则分配; 观测投影是所有学习接口的共同约束。

相较于第 3 章, 本章在全文主线中改变了表示参数化及其求解机制: 低秩结构不再局限于单一预设坐标, 而可随观测学习互补方向、连续坐标基和通道权重; 这些表示参数统一由观测区域自监督损失和反向传播求解。其关键优势是表示适应性, 其代价则是表示参数与优化过程更加耦合。例如, LRTubalRep 的核心训练循环由通用自适应矩估计器直接更新全部参数; 虽然观测区域损失和显式低秩因子仍然可辨认, 但每一步不再像固定目标 PALM 那样对应一个可单独检查的解析子问题, 有限次训练也不能直接解释为原优化模型的收敛迭代。当任务还要求分离稀疏异常与稠密噪声、维持显式约束并解释每一层状态时, 仅学习表示仍不充分。第 5 章据此迈向模型数据双驱动: 固定观测一致性、变量分裂、低秩因子和对偶更新的优化骨架, 只把置信权重、逐层参数与近端先验等难以手工设定的接口交给数据学习, 从而同时保留结构可解释性与观测适应性。

第五章 模型数据双驱动张量优化

5.1 问题定义与求解器参数化动机

第3章的模型驱动方法预先给定表示、正则与求解规则，因而具有清楚的变量语义和分析基础，但对样本相关结构的适应能力受固定接口限制；第4章进一步让方向变换、连续坐标基、聚合规则或通道权重从观测中学习，提高了表示复杂退化和跨坐标共享信息的能力，但通用梯度优化会将表示学习与求解过程耦合，难以逐层辨认约束是否满足、误差如何分解以及参数变化对应哪个优化子问题。由数据驱动迈向模型数据双驱动的动机，正是要在提高适应性的同时恢复求解过程的结构可解释性。

本文将模型数据双驱动定义为：由模型给定观测方程、低秩因子化、变量分裂、更新顺序和对偶约束等求解骨架，由数据学习置信权重、逐层罚参数、松弛系数或近端势等难以准确手工指定的接口。在这一边界下，状态显式保存干净张量、稀疏误差、有界噪声、重建与误差分裂变量以及对偶变量；低秩部分通过管乘积因子更新；固定 Gauss-Seidel 路径对同一增广目标逐块更新并检查充分下降，有限层展开路径则可选择学习注意力权重、正的逐层参数或由强凸势梯度给出的近端映射。学习模块不能改变掩膜定义、观测一致性和变量语义，测试阶段也不使用缺失位置真值更新模型。

从第4章的可学习表示转向本章的求解器参数化，技术推进发生在求解器内部。第4章主要扩大恢复张量的可学习表示族，并用观测区域损失联合拟合表示参数；当任务进一步要求区分稀疏异常与稠密噪声、维持显式约束并解释有限步计算时，仅参数化表示仍不足以说明每一步如何满足模型条件。为此，本章固定观测方程、变量语义和约束传递，把一次结构化优化迭代参数化为有限层状态转移，只将难以手工设定的更新接口交由数据调节。该转变的核心不是再次扩大表示族，而是使求解机制同时具备数据适应性、状态可读性和约束可检查性。

本章聚焦可解释张量低秩近似网络 (explainable tensor low-rank approximation network, ETLRA-Net)：在缺失、稀疏异常和稠密噪声同时存在时，把低秩、空间平滑、谱向平滑和上下文一致性写成具有明确变量语义的分裂模型，并展开为有限层优化网络。该构造以算法展开将迭代优化映射为有限层网络的基本范式^[15]、交替方向乘子法展开网络的显式状态设计^[16]以及张量补全的低秩建模^[3]为主要参考。本章重点说明 ETLRA-Net 的观测模型、先验接口、近端更新和层间状态传递。

5.2 可解释张量低秩近似网络

5.2.1 观测模型与变量分裂

ETLRA-Net 面向

$$\mathcal{Y} = \mathcal{M} \odot (\mathcal{X} + \mathcal{E} + \mathcal{N}), \quad \mathcal{M} \in \{0, 1\}^{I_1 \times I_2 \times I_3}. \quad (5.1)$$

为让每类先验可单独检查，参照交替方向乘子法的共识变量分裂思想^[14]，引入四个副本 $\mathcal{Z}_{lr}, \mathcal{Z}_{sp}, \mathcal{Z}_{spec}, \mathcal{Z}_{ctx}$ ，并令其与干净张量共识：

$$\begin{aligned} \min_{\mathcal{X}, \mathcal{E}, \mathcal{N}, \{\mathcal{Z}_j\}} \quad & \lambda_E \|\mathcal{E}\|_1 + \frac{\lambda_N}{2} \|\mathcal{N}\|_F^2 + \sum_{j \in \mathcal{J}} \rho_j \Phi_j(\mathcal{Z}_j) \\ \text{s.t.} \quad & \mathcal{M} \odot (\mathcal{X} + \mathcal{E} + \mathcal{N} - \mathcal{Y}) = 0, \quad \mathcal{Z}_j = \mathcal{X}. \end{aligned} \quad (5.2)$$

为直观呈现有限层展开如何保留原优化变量的语义，图 5.1 展示 ETLRA-Net 的总体结构及其单层更新过程。

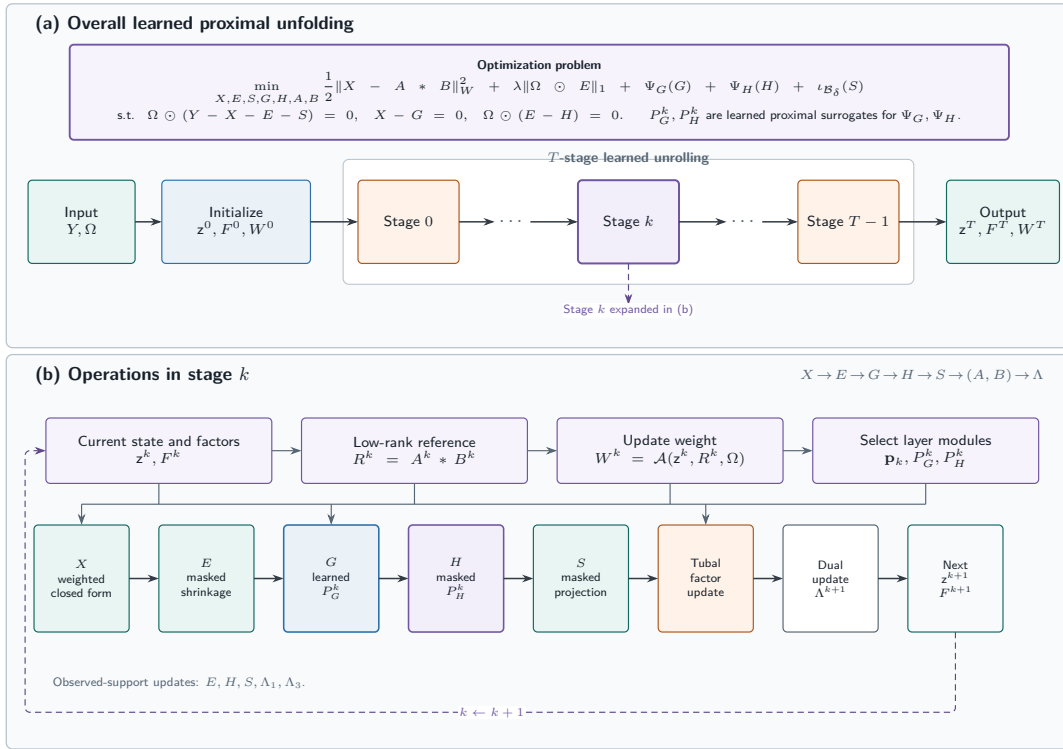


图 5.1: 可解释低秩优化展开网络的总体结构与单层流程。

图 5.1 的输入是观测张量 $\mathcal{Y}$ 、掩膜 $\mathcal{M}$ 和上一层状态；总体结构沿深度方向重复有限次更新，每一层由低秩、空间、谱向和上下文四个先验分支，以及稀疏异常、稠密噪声和观测一致性分支组成。各先验副本通过共识变量汇聚为新的干净张量，并由对偶变量传递跨层约束信息，最终输出 X^T 、 E^T 和 N^T 。图中先验模块、误差分解、观测投影、共识更新和对偶更新分别对应式 (5.2) 以及式 (5.6)–(5.7)；因此，该图强调的是“优化步骤–网络层–状态语义”的对应关系。

5.2.2 可逆通道变换与秩门控

令 $A^\top = -A$, Cayley 参数化为

$$Q(A) = (I - A)(I + A)^{-1}, \quad Q(A)^\top Q(A) = I. \quad (5.3)$$

与可逆带泄漏修正线性单元 (leaky rectified linear unit, Leaky ReLU) 组合得 $\mathcal{T}_\theta = \sigma_a \circ Q(A)$, 其中 $\sigma_a(x) = \max(x, ax)$ 、 $a > 0$ 。对变换域切片的奇异值分解

$$\overline{Z}^{(k)} = U_k \text{diag}(\sigma_{k,1}, \dots) V_k^\top,$$

使用硬门控

$$g_{k,i} = \mathbf{1}(\sigma_{k,i} > \tau_k), \quad \tilde{Z}^{(k)} = U_k \text{diag}(g_{k,i} \sigma_{k,i}) V_k^\top. \quad (5.4)$$

训练时可用直通估计器近似门控梯度；重复或近重复奇异值应分组处理或不作单值梯度解释。式 (5.4) 给出的是学习变换切片的矩阵有效秩。

5.2.3 四类先验与层级更新

空间先验采用边界零延拓的五点平均，谱向先验采用非周期三点平均：

$$\begin{aligned} \mathcal{P}_{\text{sp}}(X)_{i,j,k} &= \frac{1}{5}(X_{i,j,k} + X_{i-1,j,k} + X_{i+1,j,k} + X_{i,j-1,k} + X_{i,j+1,k}), \\ \mathcal{P}_{\text{spec}}(X)_{i,j,k} &= \frac{1}{3}(X_{i,j,k-1} + X_{i,j,k} + X_{i,j,k+1}). \end{aligned} \quad (5.5)$$

上下文先验为

$$\mathcal{P}_{\text{ctx}}(X) = \gamma X + (1 - \gamma) \text{Broadcast} \left(\frac{1}{I_1 I_2} \sum_{i,j} X_{i,j,:} \right), \quad \gamma \in [0, 1].$$

第 t 层的通用更新为

$$Z_j^{t+1} = \mathcal{P}_j(X^t + D_j^t; \theta_j^t), \quad (5.6)$$

$$E^{t+1} = \mathcal{S}_{\lambda_E^t} (M \odot (Y - X^t - N^t)),$$

$$N^{t+1} = \frac{\mu^t}{\mu^t + \lambda_N^t} M \odot (Y - X^t - E^{t+1}),$$

$$X_{\text{obs}}^{t+1} = M \odot (Y - E^{t+1} - N^{t+1}) + (1 - M) \odot X^t,$$

$$\tilde{X}^{t+1} = \frac{\mu^t X_{\text{obs}}^{t+1} + \sum_j \rho_j^t (Z_j^{t+1} - D_j^t)}{\mu^t + \sum_j \rho_j^t},$$

$$X^{t+1} = (1 - \eta^t) X^t + \eta^t \tilde{X}^{t+1}, \quad D_j^{t+1} = D_j^t + X^{t+1} - Z_j^{t+1}. \quad (5.7)$$

控制器以尺度不变残差为输入，例如

$$\begin{aligned} r_{\text{obs}}^t &= \frac{\|M \odot (X^t + E^t + N^t - Y)\|_F}{\|M \odot Y\|_F + \varepsilon}, \\ r_{\text{pri}}^t &= \frac{1}{4} \sum_j \frac{\|X^t - Z_j^t\|_F}{\|X^t\|_F + \varepsilon}, \\ r_{\text{dual}}^t &= \frac{1}{4} \sum_j \frac{\|Z_j^t - Z_j^{t-1}\|_F}{\|X^t\|_F + \varepsilon}. \end{aligned} \quad (5.8)$$

由残差控制器输出有界的 $\mu^t, \rho_j^t, \lambda_E^t, \lambda_N^t, \eta^t$ 。

稀疏异常和稠密噪声的更新分别来自两个具有闭式解的近端子问题。令 $R_E^t = Y - X^t - N^t$ ，则稀疏异常块为

$$\begin{aligned} E^{t+1} &= \arg \min_E \lambda_E^t \|E\|_1 + \frac{\mu^t}{2} \|M \odot (E - R_E^t)\|_F^2 \\ &= M \odot \mathcal{S}_{\lambda_E^t / \mu^t}(R_E^t), \end{aligned} \quad (5.9)$$

其中 $\mathcal{S}_\tau(a) = \text{sign}(a)(|a| - \tau)_+$ 为逐元素软阈值算子。掩膜外不引入异常变量，因而该更新只在有观测的位置估计稀疏偏差。令 $R_N^t = Y - X^t - E^{t+1}$ ，稠密噪声块为

$$\begin{aligned} N^{t+1} &= \arg \min_N \frac{\lambda_N^t}{2} \|N\|_F^2 + \frac{\mu^t}{2} \|M \odot (N - R_N^t)\|_F^2 \\ &= \frac{\mu^t}{\lambda_N^t + \mu^t}, M \odot R_N^t. \end{aligned} \quad (5.10)$$

这两个闭式映射分别把稀疏性和能量惩罚转化为网络层中的阈值与收缩系数；当 λ_E^t 、 λ_N^t 随层变化时，网络能够根据当前状态调整异常保留和噪声收缩的强度。

Algorithm 5.7 ETLRA-Net 第 t 层

Require: 当前状态 $X^t, E^t, N^t, \{Z_j^t, D_j^t\}$ ，观测 Y 和掩膜 M

Ensure: 下一层状态 $X^{t+1}, E^{t+1}, N^{t+1}, \{Z_j^{t+1}, D_j^{t+1}\}$

- 1: 计算四类先验副本 Z_j^{t+1}
 - 2: 更新稀疏异常 E^{t+1} 与稠密噪声 N^{t+1}
 - 3: 执行观测一致性投影，得到 X_{obs}^{t+1}
 - 4: 计算观测、先验、对偶和状态变化残差
 - 5: 残差控制器输出有界罚参数、阈值和松弛系数
 - 6: 计算共识候选并松弛更新 X^{t+1}
 - 7: 更新四个对偶变量，并记录有效秩、支持集和逆变换误差
-

5.2.4 有限层更新

从核心求解代码看，有限层展开由同一组可命名状态连接，而不是把原迭代替换为无结构的端到端映射。解析 Gauss–Seidel 路径采用 $X \rightarrow E \rightarrow G \rightarrow H \rightarrow S \rightarrow$

$(A, B) \rightarrow \Lambda$ 的固定顺序, 对原变量块执行充分下降验收; 有限层 ADMM 路径保留解析的 X 、稀疏误差、噪声球投影、低秩因子和对偶更新, 同时允许步长跨层共享或逐层变化; 学习近端路径进一步用强凸势函数的梯度参数化重建与误差近端, 并可输出势函数和正则值以检查学习先验。由此, 数据学习被限制在权重、参数和近端接口内, 模型定义的观测支撑、低秩因子化与约束传递保持不变。

ETLRA-Net 的有限深度输出 X^T 表示执行了 T 次参数化更新, 不能据此等同于 $t \rightarrow \infty$ 的迭代收敛。只有在先验算子非扩张、参数有界、误差可求和等附加条件下, 才可讨论条件性固定点收敛。硬门控的奇异值在重复或近重复时还可能不可微, 直通估计器只能作为训练近似; 可逆变换的数值稳定性也依赖 $(I + A)$ 的条件数。因此, 网络层的步长、阈值、共识权重和松弛系数应与相应的优化子问题保持一致, 并在有限层状态传递中保留其变量语义。

5.3 本章小结

本章围绕 ETLRA-Net 建立了“观测一致性–四类结构先验–共识变量–显式对偶状态”的有限深度展开框架。通过 Cayley 可逆通道变换、奇异值秩门控、空间与谱向平滑、上下文广播以及稀疏/稠密误差分解, 模型把每一层的中间状态和近端映射保留为具有明确含义的更新。稀疏异常软阈值和稠密噪声线性收缩进一步给出了误差分解分支的闭式求解基础, 从而形成结构先验、优化算子和网络参数之间的可解释接口。

相较于纯模型驱动方法, 本章的优势在于置信权重、逐层参数和近端势可以随数据调整, 减轻固定正则强度与固定近端形式对复杂退化的限制; 相较于纯数据驱动表示, 其优势在于学习接口被嵌入显式变量分裂和有限层更新, 干净张量、异常、噪声、低秩因子及对偶状态均可逐层读取, 观测残差、共识残差和低秩状态也可直接诊断。因而, 第 5 章在全文主线中不再主要扩大恢复张量的参数化族, 而是把求解机制本身参数化为具有明确状态语义的有限层映射。模型数据双驱动不是模型项与神经网络的简单叠加, 而是以模型限定可学习模块的职责和取值边界, 再利用数据提高这些接口对具体退化的适应性; 它解决的关键问题是在结构可解释性、约束可检查性与表示适应性之间建立统一接口。

第六章 统一框架与拓展研究

本章在第 3–5 章研究内容的基础上进行总结，并将博士后期间探索的时空建模和因果推荐工作放回“表示参数化—求解机制”的脉络中进行审视。章节前半按照所解决的问题、表示参数化创新和求解机制归纳三类驱动范式；章节后半说明这一协同设计思想如何迁移到相邻应用，并给出面向后续研究的具体推进方向。

6.1 三类驱动范式的技术递进

第 3–5 章围绕同一类不完整多维观测恢复问题，依次研究模型驱动、数据驱动和模型数据双驱动三类范式。为在同一数学框架下描述这些方法，记 $\mathcal{R}_\phi$ 为从表示变量或优化状态生成恢复张量的参数化映射， s^t 为第 t 步状态， Π 为结构先验与观测约束， $\mathcal{S}_{\psi_t}$ 为相应的求解或更新机制。三类研究均可写为

$$s^{t+1} = \mathcal{S}_{\psi_t}(s^t; \mathcal{Y}, \Omega, \Pi), \quad \hat{\mathcal{X}} = \mathcal{R}_{\phi_T}(s^T),$$

其区别在于 $\mathcal{R}_\phi$ 、 Π 和 $\mathcal{S}_{\psi_t}$ 中哪些部分由模型固定，哪些部分由观测估计或学习。表 6.1 比较了三类范式中表示对象、模型固定部分、数据学习部分与求解机制的具体对应关系。图 6.1 从同一观测方程出发，将三类研究放在相同的输入和输出之间：第 3 章固定表示形式、先验与更新规则并求解显式变量；第 4 章保留低秩对象和观测约束，同时学习表示变换与信息通路；第 5 章进一步固定状态语义和约束传递，只让求解器中的部分接口由数据调节。三条路径因此不是针对不同任务提出的并列方案，而是统一恢复模型在表示参数化和求解机制两个维度上的递进实例。

表 6.1: 第 3–5 章低秩恢复方法的统一关系。

层面	表示参数化	模型固定与数据学习的边界	求解机制及其解释
模型驱动	塔克核心 $\mathcal{G}$ 、模态因子 U_n 和重建张量	核心稀疏、因子核范数、图拉普拉斯、时间差分或重建梯度由任务模型确定；观测用于估计显式表示变量	采用近端梯度、PALM 或变量分裂 ADMM 求解固定目标；变量和更新步骤具有明确含义
数据驱动	显式潜张量、可学习变换、连续坐标场和潜在通道	保留低秩对象与观测投影，由观测学习方向组织、连续坐标基、聚合规则或通道权重	采用观测区域自监督损失和反向传播联合拟合表示参数；共享参数建立跨方向、跨坐标或跨通道的信息通路
模型数据双驱动	干净张量、异常、噪声、先验副本和对偶状态	模型固定观测方程、变量语义、更新顺序和约束传递；数据调节逐层参数、权重或结构化近端接口	将解析、近端、投影和对偶更新组织为有限层网络；状态可读，约束可检查，部分求解接口可随数据适配

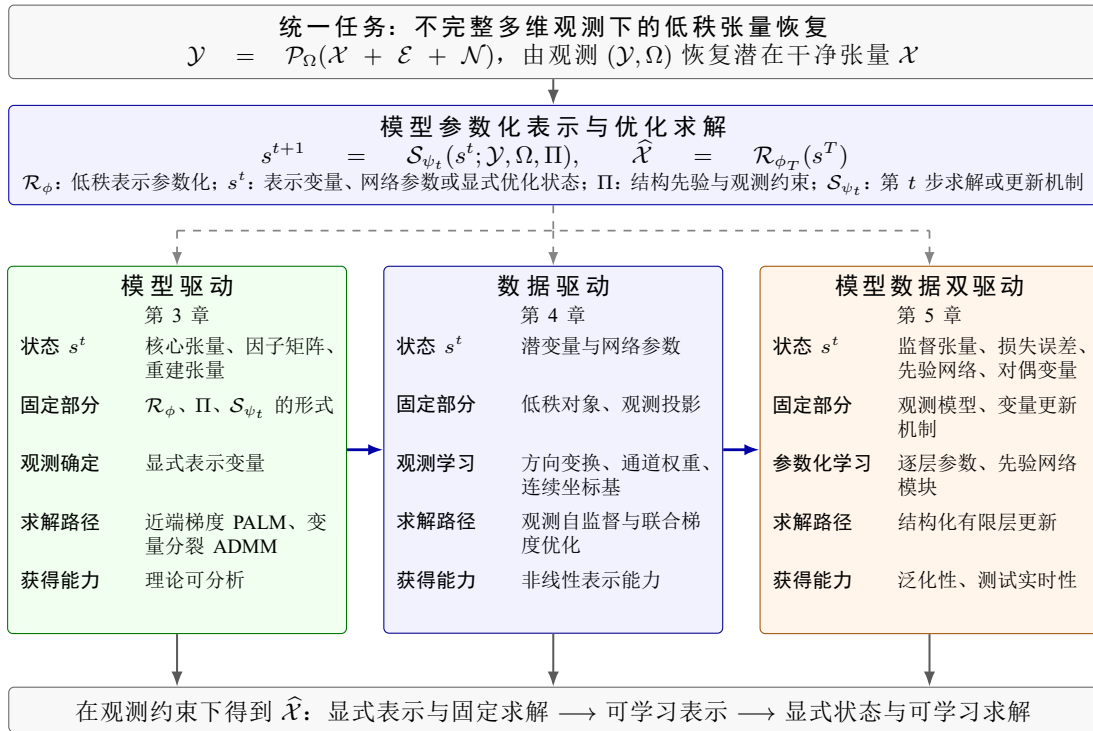


图 6.1: 三类低秩张量恢复范式的研究框架。灰色虚线表示统一模型在三类范式中的实例化，蓝色实线箭头依次表示可学习表示接口的扩展和受约束求解接口的参数化。

模型驱动：建立显式表示与固定目标求解基线。 第 3 章首先解决低秩结构如何落到具体优化变量以及结构先验应作用于何处的问題。TuckerAPP^[17] 在全尺寸塔克坐标

中联合核心稀疏、空间图拉普拉斯和时间差分，使跨模态交互、空间邻域和时间变化分别对应可辨认的结构项。LRTuckerRep^[18] 进一步采用降阶塔克表示，以因子核范数控制模态子空间，以核心稀疏选择跨模态交互，并围绕固定观测目标构造近端块更新及条件性收敛分析。GradientTD 保留这一低秩骨架，只把局部先验从潜在因子图转移到重建张量的梯度域，并通过梯度变量和重建变量分裂组织近端线性化 ADMM。三项工作的共同创新是把表示对象、正则位置和更新子问题一一对应，从而建立可解释、可分析的模型驱动基线；其限制在于坐标系统、变换形式和正则分配主要依赖预先设计，难以随具体退化几何调整。

数据驱动：扩展可由观测适配的低秩表示族。 第 4 章针对固定表示难以覆盖方向受损、整片坐标缺失和潜在通道贡献不均的问题，将数据学习引入表示参数化。3DeepRep^[19] 用三个方向的显式低秩潜张量生成候选重建，并学习方向耦合变换与非线性聚合，使互补信息能够跨方向传播；LRTubalRep 用共享参数的连续坐标场生成离散切片，使观测坐标的梯度能够通过共同函数参数传递到缺失坐标；CA-TNN^[20] 将正则分配提升到潜在通道层面，使不同通道的整体低秩贡献能够自适应调节。这一阶段的核心创新不是取消低秩先验，而是将方向组织、坐标共享和通道权重纳入可学习表示，并以观测区域自监督损失统一求解。由此，研究主线从“在固定参数化中估计变量”推进到“由观测共同确定参数化形式及其变量”；与此同时，通用反向传播使表示学习与求解过程耦合，逐步状态和约束满足不再天然对应明确的优化子问题。

模型数据双驱动：在显式优化骨架中参数化求解机制。 第 5 章进一步解决数据适应性与算法可解释性之间的矛盾。ETLRA-Net 显式区分干净张量、稀疏异常、稠密噪声、低秩状态、先验副本和对偶变量，并把观测投影、异常软阈值、噪声收缩、低秩更新、先验近端和对偶传递组织为有限层状态转移。模型负责固定观测方程、变量语义、更新顺序和约束关系，数据只调节置信权重、逐层罚参数、松弛系数或满足结构限制的近端映射。其核心创新是把可学习性从表示参数扩展到职责明确的求解接口，使每层计算仍能追溯到特定变量和优化作用；有限层输出因此应解释为参数化求解过程的结果，而不等同于无限迭代的全局最优解。

三类范式形成两条相互配合的递进线索。在表示参数化上，研究由显式塔克变量发展到可学习变换与连续坐标表示，再发展到包含误差分量、先验副本和对偶变量的显式状态；在求解机制上，研究由固定目标的块近端迭代发展到观测自监督的联合梯度优化，再发展到结构约束下的可学习有限层更新。因此，第 3 章回答“低秩结构和先验如何写成可求解模型”，第 4 章回答“哪些表示接口需要随观测自适应”，第 5 章回答“如何把这种自适应能力嵌入可解释的求解过程”。这一递进把模型可分析性、表示适应性和求解可解释性连接为统一的低秩恢复研究主线。

6.2 从离散低秩表示到连续结构表征

第 3 章的核心、因子和差分算子给出离散坐标上的显式参数化及固定求解基线；第 4 章的连续坐标场把光谱、时间或空间索引提升为可查询变量，并以观测自监督学习函数参数；第 5 章进一步把显式状态更新参数化为可学习求解器。三者共同提供了后续“连续低秩结构表征—求解—医学图像应用”项目的方法基础。

若以连续坐标 $\xi = (\xi_1, \dots, \xi_N)$ 表示物理位置、时间或光谱变量，可考虑低秩函数族

$$\mathcal{X}(\xi) = \sum_{r=1}^R c_r \prod_{n=1}^N f_{nr}(\xi_n), \quad (6.1)$$

并在观测保真、低秩表示和连续平滑之间构造结构能量

$$\mathcal{E}(\mathcal{X}) = \mathcal{D}_\Omega(\mathcal{X}, \mathcal{Y}) + \lambda_{\text{lr}} \mathcal{R}_{\text{lr}}(\mathcal{X}) + \lambda_{\text{sm}} \sum_{n=1}^N \|\partial_{\xi_n} \mathcal{X}\|_{\mathcal{H}_n}^2. \quad (6.2)$$

式 (6.2) 把已有离散低秩、连续坐标和梯度正则构造抽象到连续能量视角下的研究框架。 $\mathcal{R}_{\text{lr}}$ 可继承塔克、管状或通道级低秩结构， ∂_{ξ_n} 可在连续函数空间中表达平滑、边缘和时间变化， $\mathcal{D}_\Omega$ 则保持对实际观测算子的约束。未来可围绕该框架研究连续函数表示的逼近性质、低秩与平滑的耦合条件以及离散采样误差如何影响求解。

6.3 拓展研究：结构表征与可解释建模

拓展研究没有脱离主线，而是把“先识别结构，再配置算子，最后保留机制语义”的方法迁移到时空建模和因果推荐。

6.3.1 多尺度时空预测

长时序同时含有慢变化趋势、周期成分、突发事件和高频噪声。自适应小波分解双路径变换器 (adaptive wavelet decomposition-based dual-path transformer, AWD-DPT)^[23] 先将序列分为低频近似和高频细节，再分别建立趋势路径与脉冲路径。设输入为 $\mathbf{X} \in \mathbb{R}^{L \times C}$ ，小波分解和高频软阈值可写为

$$[\mathbf{X}_{A_m}, \mathbf{X}_{D_m}, \dots, \mathbf{X}_{D_1}] = \text{DWT}(\mathbf{X}, \psi, m), \quad \tilde{\mathbf{X}}_{D_i} = \text{sign}(\mathbf{X}_{D_i}) (|\mathbf{X}_{D_i}| - \sigma(\theta_i))_+.$$

低频分支学习趋势和周期，高频分支利用卷积重要性分数形成门控，再经逆小波变换得到预测。其最直接的强基线 WPMixer^[24] 同样使用多分辨率小波混合，但 AWD-DPT 进一步把高频随机扰动抑制与趋势/脉冲专门路径结合；PatchTST、TimesNet 和 TimeMixer^[25-27] 则分别代表片段化、二维时间结构和多尺度混合的比较视角。已有基准结果显示，AWD-DPT 在 ETT、Weather 和 Electricity 等数据上改善了预测误差，并在 ETTh1、预测长度 336 的效率实验中将 GFLOPs 从 0.481 降至 0.260；这些结果说明频率选择与路径分工能够缓解趋势和突发模式的相互干扰。

时空张量融合网络 (spatio-temporal tensor fusion network, STTFN)^[28] 则把交通历史窗口组织为时间超平面和空间超平面, 并用自动相关提取周期依赖、用静态图、自适应图和频域图表达互补空间关系:

$$\mathcal{Y}_T = f_T(\mathcal{X}_T), \quad \mathcal{Y}_S = f_S(\mathcal{X}_S), \quad \hat{\mathcal{Y}} = \text{MLP}(\mathcal{Y}_T \odot \mathcal{Y}_S).$$

与耦合传播的 STGCN、统一注意力的 STTN^[29, 30] 和最接近的显式解耦基线 D2STGNN^[31] 相比, STTFN 的区别在于同时进行时间/空间超平面建模, 并增加频域图和张量融合。PeMS04 和 PeMS08 的交通预测结果及去除自动相关、频域图块的消融表明, 周期提取和频域空间关系在其协议内具有互补作用。

可推进方向。 对该类任务, 后续研究可以从四个角度展开: 一是将小波尺度、周期集合和图结构改为随时间或节点状态变化的可学习对象; 二是把缺失、异步采样和传感器失效直接写进观测算子, 而不是先行插值; 三是用连续时间动力系统连接自动相关和图传播, 分析长预测滚动误差; 四是以频带、图边和路径权重为解释变量, 研究不同结构对异常事件预测的实际贡献。这样可将时空预测从“提高指标”推进为“结构选择—机制解释—稳定外推”的完整问题。

6.3.2 语义增强时空建模

城市交通流量还受到土地利用、道路可达性和区域功能影响。兴趣点增强大语言模型 (point-of-interest-enhanced large language model, POI-LLM)^[32] 将道路地块、兴趣点类别、日/周时间编码和交通词元组织为可对齐的输入框架。若 $\mathbf{e}_r$ 是地块 i 内兴趣点短语的向量, w_r 是其权重, 则地块语义可写为

$$\mathbf{E}_{\text{POI}}(i) = \frac{\sum_{r \in \text{POI}(i)} w_r \mathbf{e}_r}{\sum_{r \in \text{POI}(i)} w_r}, \quad \mathbf{E}_S = \text{Norm}(\text{Proj}[\mathbf{E}_{\text{parcel}}; \mathbf{E}_{\text{POI}}]).$$

以交通词元为查询、以空间语义和时间编码为键值的交叉注意力为

$$\mathbf{Z}_i = \text{Concat}_h \left(\text{softmax} \left(\frac{Q_i W_Q^{(h)} (K_i W_K^{(h)})^\top}{\sqrt{d_k}} \right) V_i W_V^{(h)} \right) W_O.$$

随后只开放部分变换器参数, 以控制城市语义框架接入大语言模型的成本。ST-LLM、UrbanGPT 和 Time-LLM^[33-35] 提供了空间—时间大语言模型或时间序列重编程的强比较视角; GMAN^[36] 则代表非大语言模型的图注意力模型。POI-LLM 的已有结果表明, 显式地块与兴趣点语义可以补充单纯历史序列, 尤其适用于解释同样的拓扑连接为何在不同功能区域产生不同流量。

可推进方向。 该方向可以围绕“语义是否真正改善了结构推断”推进: 其一, 建立地块—道路—兴趣点的多尺度图文本对齐, 并对地图版本变化进行增量更新; 其二,

区分稳定城市功能与短期事件语义，构造时间变化的语义权重；其三，采用部分冻结、低秩适配或提示学习降低跨城市迁移成本；其四，将预测不确定性、语义来源和注意力路径同时输出，使城市规划者能够追踪模型使用了哪些结构信息。

6.3.3 因果稳健推荐

推荐交互同时受到用户稳定偏好、未观测环境、曝光机制和物品流行度影响。仅由交互似然拟合潜变量时，不同的潜在环境重参数化以及稳定偏好–环境敏感偏好的线性混合，可能产生相同的观测分布，却给出不同的机制解释。该问题与可辨识因果表示学习^[37-40]以及分布外推荐^[41, 42]中的表示歧义相联系。

条件可辨识潜在环境推荐（Conditionally Identifiable Latent-Environment Recommendation, CILER）^[43]将分布外推荐表述为条件可辨识风险推荐（conditionally identifiable risk-aware recommendation, CI-RR）问题。设 $\mathcal{M}$ 为潜在环境推荐模型， $\mathcal{F}$ 为候选模型类， $\mathcal{E}_{\mathcal{F}}(\mathcal{M})$ 为与 $\mathcal{M}$ 产生相同 $p(\mathbf{y} | \mathbf{x}_1)$ 的观测等价类，并以 $\approx$ 表示环境敏感坐标的置换和逐分量可逆变换，则 CI-RR 的形式为

$$\min_{\mathcal{M} \in \mathcal{F}} \mathcal{R}_{\text{te}}(f_{\mathcal{M}}) \quad \text{s.t.} \quad \mathcal{E}_{\mathcal{F}}(\mathcal{M}) \subseteq \left\{ \widetilde{\mathcal{M}} \in \mathcal{F} : \widetilde{\mathcal{M}} \approx \mathcal{M} \right\}. \quad (6.3)$$

这个约束把“预测风险最小化”和“环境敏感表示可辨识”放在同一模型选择问题中，而不是把可辨识性当作训练后的附加解释。

模型构造。 CILER 用观测用户特征 $\mathbf{x}_1$ 参数化潜在环境分布。其变分环境分布属于带特征依赖自然参数的指数族：

$$q_{\varphi_1}(\boldsymbol{\varepsilon} | \mathbf{x}_1) \propto b(\boldsymbol{\varepsilon}) \exp\{\boldsymbol{\eta}_{\varphi_1}(\mathbf{x}_1)^\top \mathbf{T}(\boldsymbol{\varepsilon})\}. \quad (6.4)$$

其中 b 是基测度， $\mathbf{T}$ 是充分统计量， $\boldsymbol{\eta}_{\varphi_1}(\mathbf{x}_1)$ 把可观测用户上下文映射为潜在环境的有限维参数。交互历史和用户上下文进一步用于推断未观测用户差异 $\mathbf{x}_2$ ，并分别生成稳定表示 $\mathbf{z}_1$ 与环境敏感表示 $\mathbf{z}_2$ 。对环境敏感坐标，CILER 使用特征索引的多项式结构方程

$$z'_{2,i} = g_i(\text{pa}(z_{2,i}); \boldsymbol{\Lambda}_i(\mathbf{x}_1)) + \varepsilon_i, \quad i = 1, \dots, K, \quad (6.5)$$

其中 $\text{pa}(z_{2,i})$ 是预设参考顺序下的父坐标， g_i 对父坐标为至多 p 次多项式， $\boldsymbol{\Lambda}_i(\mathbf{x}_1)$ 的系数随观测特征变化；下三角参数化用于保证参考顺序下的无环性。指数族限制环境分布的变化方式，多项式结构方程限制环境如何改变偏好机制，两者由同一个 $\mathbf{x}_1$ 索引，因而共同排除仅凭观测似然无法区分的重参数化。

在训练阶段，CILER 将交互重构、结构方程拟合、环境分布对齐、解码门控和权重衰减合并为

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{recon}} + \lambda \mathcal{L}_{\text{causal}} + \alpha \|\mathbf{W}_z\|_1 + \beta \|\boldsymbol{\Theta}\|_2^2. \quad (6.6)$$

这里 $\mathcal{L}_{\text{causal}}$ 包含结构方程损失、 $q_{\varphi_2}(\mathbf{x}_2 | \mathbf{y}, \mathbf{x}_1)$ 与其先验的 Kullback–Leibler (KL) 散度、条件环境分布对齐项和系数稀疏项。部署时不把环境固定为单一状态，而是对条件环境分布积分；用 S 个环境样本近似时，物品概率为

$$\hat{\pi}(\mathbf{y}, \mathbf{x}_1) = \frac{1}{S} \sum_{s=1}^S \text{softmax} \left(f_{\theta_3} \left[\mathbf{W}_z \odot (\mathbf{z}_1, \mathbf{z}_2^{(s)}) \right] \right). \quad (6.7)$$

因此，环境边缘化直接保留了给定用户特征下的环境不确定性，并且不需要环境标签或测试时梯度更新。

可辨识性、部署风险与实验结果。 在条件环境为坐标可分解的正则连续指数族、观测特征诱导足够多的独立自然参数变化、偏好机制的多项式阶数和因果顺序设定正确、解码器在表示支集上光滑且单射的条件下，CILER 的环境敏感表示可辨识到坐标置换和逐分量可逆变换，相关有向关系也在同一等价意义下确定。若训练与部署仅在观测用户特征边缘分布上发生变化，并满足部署分布绝对连续于训练分布且密度比有界，则环境边缘化的对数风险满足

$$\begin{aligned} \mathcal{R}_{\text{te}}(p_q) - \mathcal{R}_{\text{te}}(p^*) &= \mathbb{E}_{\mathbf{x}_1 \sim P_{\text{te}}} \text{KL}(p^*(\mathbf{y} | \mathbf{x}_1) \| p_q(\mathbf{y} | \mathbf{x}_1)) \\ &\leq \mathbb{E}_{\mathbf{x}_1 \sim P_{\text{te}}} \text{KL}(p^*(\boldsymbol{\varepsilon} | \mathbf{x}_1) \| q(\boldsymbol{\varepsilon} | \mathbf{x}_1)). \end{aligned} \quad (6.8)$$

该不等式由共享条件预测通道下的 Kullback–Leibler 数据处理关系给出；有界密度比还可将右端转化为训练特征分布上的误差控制。CILER 在 Synthetic、Meituan 和 Yelp 三个数据集上考察特征、时间和地理变化，并与 13 个基线比较。在 12 个分布外排序指标上均取得第一，最高相对提升为 25.6%；去除用户条件环境限制或特征索引多项式机制会使三个数据集的每项排序指标下降，单一环境状态替代边缘化也会降低排序和概率校准指标。上述结果把“条件环境建模—偏好机制约束—环境边缘化”三个组件与分布外排序效果联系起来。

原型间隔环境与修复感知选择群体稳健学习 (ProME: Prototype-Margin Environments with Repair-Aware Selection for Group-Robust Learning)^[44] 研究的是无训练组标签的群体稳健学习，不是推荐模型。其出发点是：由独立参考编码器得到的环境划分可能与后续优化表示不一致，而在分类器修复前选择编码器又可能与最终部署分类器不一致。ProME 将这两个问题合并为“内生环境与修复感知选择” (endogenous environments with repair-aware selection, ERAS) 问题。若 $\tilde{\phi}$ 是同一表示学习轨迹上的检查点、 $\Pi(\tilde{\phi})$ 是其诱导的二元环境划分、 $\hat{h}_t$ 是对冻结编码器 θ_t 重新拟合的群组平衡线性头，则其选择逻辑可写为

$$\begin{aligned} \hat{\mathcal{E}} &= \Pi(\tilde{\phi}), \\ \phi^* &\in \arg \min_{\phi} \left\{ \mathcal{R}_{\text{tr}}(\phi) + \lambda \mathcal{P}_{\text{inv}}(\phi, \hat{\mathcal{E}}) \right\}, \\ t^* &\in \arg \max_{t: \theta_t \in \mathcal{C}} \text{WGA}_{\text{val}}(\hat{h}_t \circ \theta_t). \end{aligned} \quad (6.9)$$

原型间隔环境构造。 令 b_ξ 为骨干网络, p_ω 为投影层, ProME 使用球面归一化表示

$$\phi_{\xi,\omega}(x) = \text{normalize}(p_\omega(b_\xi(x))), \quad \mu_c = \text{normalize}\left(\frac{1}{|\mathcal{D}_c|} \sum_{(x_i, y_i) \in \mathcal{D}_c} \phi_{\xi,\omega}(x_i)\right). \quad (6.10)$$

对二分类任务, 观测类别 y 的原型间隔为

$$s(x; y) = \cos(\phi(x), \mu_y) - \cos(\phi(x), \mu_{1-y}), \quad m = \text{median}_{\mathcal{D}_r} s, \quad (6.11)$$

并按全局中位数构造低、高间隔环境

$$\hat{\mathcal{E}}_0 = \{(x, y) : s(x; y) \leq m\}, \quad \hat{\mathcal{E}}_1 = \{(x, y) : s(x; y) > m\}. \quad (6.12)$$

中位数使两个单元在样本量上近似平衡; 低间隔单元包含对观测类别支持较弱、可能与主导捷径相关的样本, 高间隔单元包含类别原型支持较强的样本。原型每隔若干训练步刷新, 标准方案在预热后固定划分, 刷新方案则同步更新划分, 使环境信号来自同一表示学习轨迹。

对诱导环境, ProME 同时使用不变风险最小化 (invariant risk minimization, IRM) 的分类器最优性梯度惩罚和风险外推 (risk extrapolation, REx) 的环境风险离散度惩罚:

$$\mathcal{L}_{\text{ProME}} = \overline{\mathcal{R}}(f_\phi) + \lambda_t \mathcal{P}_{\text{IRM}}(f_\phi) + \mathcal{P}_{\text{REx}}(f_\phi), \quad \mathcal{P}_{\text{REx}} = \frac{1}{K} \sum_{e=1}^K \left(\hat{\mathcal{R}}_e - \overline{\mathcal{R}}\right)^2. \quad (6.13)$$

分类器修复与选择。 阶段二冻结候选编码器, 使用带群组标注的验证集拟合相同的群组平衡线性头。对正则化参数 γ , 其修复目标为

$$\begin{aligned} \mathcal{L}_{\text{DFR}}^{(t)}(h; \gamma) &= \frac{1}{|\mathcal{G}|} \sum_{g \in \mathcal{G}} \frac{1}{|\mathcal{D}_{\text{val}}^g|} \sum_{(x, y) \in \mathcal{D}_{\text{val}}^g} \ell(h(\theta_t(x)), y) + \gamma \|h\|_2^2, \\ t^* &= \arg \max_t \text{WGA}_{\text{val}}(h^{(t)} \circ \theta_t). \end{aligned} \quad (6.14)$$

训练集在第一阶段只使用输入和类别标签; 验证组标签用于候选跟踪、分类器修复、正则参数选择和最终排序, 不回写第一阶段表示。ProME 的理论分析表明, 在因果成分与捷径成分足够可分、投影噪声受控时, 低原型间隔能够富集捷径冲突样本; 在同一训练样本上, 表示变化引起的重分配比例受间隔中位数邻域内样本比例控制。对固定预测器和二元诱导环境, REx 项给出

$$\mathcal{R}_{\text{wg}}^{(\mathcal{E})}(f) \leq \overline{\mathcal{R}}(f) + \sqrt{(K-1)\mathcal{P}_{\text{REx}}(f)}, \quad (6.15)$$

若每个真实组分布与某个诱导环境的总变差距离不超过 ρ , 且损失取值在 $[0, B]$, 则该界可进一步加上 $B\rho$ 后转移到真实组风险。这个条件明确了原型环境与真实组之间仍需分布对齐, 不能将诱导单元直接等同于真实组。

ProME 在 Waterbirds、CelebA 和 CivilComments 三个真实基准上进行主要评估,并以 ColoredMNIST 进行可控实验。在具有相同验证组标签使用权限的比较中,ProME 的平均最差组准确率为 87.0%,高于最佳对比方法 GSR 的 83.9%;三个真实数据集分别为 $93.1\% \pm 0.3\%$ 、 $89.3\% \pm 0.5\%$ 和 $78.7\% \pm 0.2\%$ 。Waterbirds 的诊断中,低间隔环境的捷径冲突样本比例为 0.68,高间隔环境为 0.01;多候选修复选择在 CelebA 上优于单检查点修复。上述实验支持原型间隔的环境构造和修复后的候选选择分别承担不同作用。与群体分布鲁棒优化^[45]、最后一层重训练^[46,47]、失败样本去偏^[48]和基于影响函数的群体重加权^[49]等最接近的基线相比,ProME 的区别在于环境划分和部署分类器选择都与同一表示轨迹及最终修复头相耦合。

与其他因果稳健工作的互补关系。 CILER 通过用户条件指数族和特征索引多项式确定潜在环境及其影响机制,重点是环境敏感表示的条件可辨识性与环境边缘化预测;ProME 不估计连续潜在环境,而是从表示轨迹的原型间隔构造近似平衡的内生训练环境,再用少量验证组标签进行修复感知选择,重点是群体稳健学习中的部署对齐。GenCAR^[50]则在稳定偏好表示上进行离线反事实候选生成、信任半径过滤和选择性校准。对候选物品 $\tilde{j}$,其信任集合可表示为

$$\mathcal{Y}_u = \left\{ \tilde{j} : d(\tilde{j}, \hat{z}_c) \leq \delta \right\}, \quad d(\tilde{j}, \hat{z}_c) = 1 - \cos(e_{\tilde{j}}, \hat{z}_c). \quad (6.16)$$

在满足可交换性条件的校准集上,还可结合共形 p 值和 Benjamini–Hochberg 程序控制选择集合的错误发现率^[51–55]。三者分别从“潜在环境建模”“稳定机制与群体划分”“候选选择与风险控制”切入,形成互补的因果稳健学习路线。

可推进方向。 后续可沿三条线推进:在 CILER 方向研究连续环境、长期偏好漂移和多目标推荐中的可辨识机制;在 ProME 方向研究原型间隔与真实群体之间的可检验对齐、少量验证标注下的修复稳定性以及多类别扩展;在 GenCAR 方向研究反事实生成的语义可靠性、信任半径的有限样本校准和在线曝光反馈。这样能够分别回答“环境如何变化”“哪些表示机制稳定”以及“在不确定性下何时采取选择性决策”。

6.4 本章小结与后续项目衔接

本章得到三点综合认识。第一,低秩建模的核心不是固定某个秩数,而是选择与观测机制相匹配的结构坐标:第 3 章在塔克坐标中显式约束核心、因子和局部关系,第 4 章让方向、坐标和通道框架参与学习,第 5 章把先验和求解状态展开为可解释更新。第二,拓展研究表明,同一方法论可以迁移到频率—图结构、城市语义和因果环境等对象;迁移的关键在于明确结构变量、信息来源和可干预的求解框架。第三,连续坐标表示、梯度能量和可解释求解轨迹为后续“连续低秩结构表征—求解—医学图像应用”提供了自然衔接:连续低秩函数负责表达物理空间中的跨尺度结构,能量

和变分算子负责约束平滑与观测一致性，结构化展开或连续动力系统负责稳定计算，最终可服务于动态 MRI 重建以及脑部 MRI 异常检测与分割等医学图像任务。

沿着这一衔接关系，下一阶段可在连续函数空间中研究低秩与平滑耦合的表示能力，在可解释优化中研究能量下降、轨迹稳定性和跨域自适应，在医学图像中研究动态 MRI 的时空恢复、脑部 MRI 的异常结构识别以及物理先验与低维流形的一致性。由此，博士后阶段形成的离散模型、深度表示和算法展开将共同构成“表征—求解—应用”研究路线的理论与方法起点。

第七章 结论与展望

7.1 主要结论

本报告围绕“模型和数据双驱动的低秩张量分解”回答了两个相互依赖的问题：在观测缺失、异常、噪声和多模态耦合同同时存在时，恢复张量应采用何种低秩参数化，以及这些表示变量或求解器参数应如何由不完整观测确定。围绕这一核心，报告形成了“显式低秩参数化与固定目标求解—可学习低秩参数化与观测自监督求解—显式优化状态与可学习有限层求解”的方法链，并将同一“表示参数化—求解机制”视角延伸到时空预测、城市语义建模和因果稳健推荐。

结论一：模型驱动方法建立了显式低秩参数化与固定目标求解基线。 TuckerAPP^[17] 在全尺寸塔克坐标中联合核心稀疏、空间图拉普拉斯和时间差分约束，以交替近端梯度、外推和观测校正完成交通张量补全。论文在时空数据及多种缺失协议下给出的结果表明，空间和时间先验能够在多数随机、非随机缺失场景中互补提升恢复质量；同时，部分停摆场景由 LATC、SPC 或 stTT 取得更低误差，说明先验必须与缺失几何相匹配，不能把单一实验协议下的优势推广为所有场景的普遍最优。

低秩塔克表示模型 LRTuckerRep^[18] 将表示降阶，并在固定有限观测保真目标中联合因子核范数、核心稀疏和固定因子图平滑。其 PALM 更新给出核心软阈值、因子奇异值收缩和回溯主化步骤；在迭代有界、块梯度满足 Lipschitz 条件、主化系数有统一正间隔且目标满足 Kurdyka-Łojasiewicz 性质时，能够得到整序列收敛到临界点及相应的平稳性迭代界。GradientTD 在此基础上将局部先验从潜在因子图转移到重建张量的离散梯度，用总变分或二次变分表达相邻位置变化，并以变量分裂组织近端线性化 ADMM。三项方法虽然参数化尺度和正则作用位置不同，但都先固定“恢复张量如何由核心、因子或重建变量生成”，再由与目标一致的块更新求解这些变量；其共同优势是表示含义、先验位置和算法步骤可以逐项对应。

结论二：数据驱动方法通过可学习低秩参数化提高了表示适应性。 3DeepRep^[19] 以三个方向的显式潜张量、方向耦合变换和非线性聚合器构造深度低秩表示。在三个高光谱数据集和四类退化协议下，论文的主结果、重复实验、初始化比较和模块消融共同表明，三方向互补能够缓解单一方向在连续坏列或混合退化下的信息不足；Pavia Centre 连续坏列条件下的代表性 PSNR 为 51.164 dB，高于 CoNoT 的 43.499 dB。CATNN^[20] 将正则权重从单个奇异值提升到潜在通道的整体核范数；在已有随机像素缺失实验中，Pavia 的 95% 缺失率条件下 PSNR 为 42.023 dB，Washington DC Mall 的

90% 缺失率条件下 PSNR 为 60.694 dB，且通道加权和指数激励消融与该构造一致。

LRTubalRep 面向整片切片缺失，采用共享参数的连续坐标场，使观测坐标的损失可以通过共同参数传递到缺失坐标，并以连续原子数、空间系数宽度和单系数秩分别控制表示容量。第二项研究内容由此说明，数据驱动低秩表示的核心变化发生在参数化层：深度变换、连续坐标基和通道权重扩大了固定塔克表示能够表达的恢复映射族；观测区域自监督损失与反向传播则提供统一求解机制。其关键优势是可以随退化几何调整信息通路，而不是放弃低秩结构。

结论三：模型数据双驱动方法实现了可解释的求解机制参数化。 ETLRA-Net 将观测一致性、干净张量、稀疏异常、稠密噪声、低秩因子、先验副本和对偶变量组织为具有明确语义的状态，并把解析更新、近端映射、投影和对偶更新映射为有限层网络。与第 4 章直接用通用优化器联合拟合表示参数不同，第 5 章固定状态含义、更新顺序和约束传递，只让置信权重、逐层参数或结构化近端势由数据学习。其关键优势是数据适应性发生在职责明确的求解接口内，因而可以逐层读取误差分解、低秩状态和约束残差。现有结论属于模型和算法设计层面的认识，有限层网络不等同于无限迭代算法，也不能由结构设计直接推出全局最优或普适恢复保证。

结论四：拓展研究展示了三类驱动范式共同方法的迁移价值。 AWD-DPT^[23] 和 STTFN^[28] 分别从频率子带、时间/空间超平面和多视角图中分解时空依赖；POILLM^[32] 将道路地块与兴趣点语义转化为可对齐词元；CILER^[43] 通过用户条件潜在环境、特征索引多项式机制和环境边缘化处理分布外推荐中的偏好变化，在三个数据集的 12 个分布外排序指标上均取得第一，最高相对提升为 25.6%；ProME^[44] 则从原型间隔构造无训练组标签的内生环境，并通过分类器修复后的验证最差组准确率选择部署模型，在相同验证组标签使用权限的比较中取得 87.0% 的平均最差组准确率，高于最佳对比方法的 83.9%；GenCAR^[50] 在稳定偏好表示上进行反事实候选生成、信任半径过滤和选择性风险控制。这些工作共同支持一个方法论判断：面对新任务时，应先确定对象如何被参数化，再选择与该参数化相匹配的学习或优化机制；显式区分结构变量、信息通路与求解约束，有利于分析信息来源、迁移条件和失效原因。

综合以上结论，本报告形成的研究链条为

显式低秩参数化与固定目标求解 → 可学习低秩参数化与观测自监督求解
 → 显式优化状态与可学习有限层求解 (7.1)
 → 面向任务的结构化恢复.

该链条表明，三类方法的本质差异同时位于表示参数化和求解机制两个维度：第 3 章固定参数化族及目标后求解显式变量，第 4 章扩展可学习参数化族并由观测梯度拟

合，第5章进一步参数化有限步求解器而保留显式状态。它既保留张量代数和非凸优化的数学结构，也吸收深度表示对复杂退化的适应能力。

7.2 研究展望

下一阶段的研究将以“连续低秩结构的表征—求解—医学图像应用”展开，并继续沿表示参数化与求解机制两条线协同推进：先建立连续函数空间中的低秩参数化，再构造与连续结构能量相匹配的稳定求解机制，最后在医学成像的物理观测约束下联合验证表示能力、数值稳定性与应用边界。

一、连续低秩结构参数化。 将空间、时间和光谱索引提升为连续变量 ξ ，研究函数空间中的低秩分解、平滑性和观测算子之间的关系。可从

$$\mathcal{X}(\xi) = \sum_{r=1}^R c_r \prod_{n=1}^N f_{nr}(\xi_n), \quad \mathcal{E}_{\text{cont}} = \mathcal{D}_{\Omega} + \lambda_{\text{lr}} \mathcal{R}_{\text{lr}} + \lambda_{\text{sm}} \sum_n \|\partial_{\xi_n} \mathcal{X}\|_{\mathcal{H}_n}^2 \quad (7.2)$$

出发，结合函数型张量分解、正弦或傅里叶特征表示^[8, 9, 12]，研究连续低秩表示的逼近能力、采样误差和多尺度细节保持。重点问题包括：连续函数的低秩性如何与离散张量秩对应，平滑能量如何避免过度抹平边缘，以及连续原子和潜在通道如何表达异质局部流形。

二、连续结构能量的可解释稳定求解。 将观测保真、低秩能量和连续平滑项的变分更新写成可控动力系统，研究

$$\dot{s}(t) = -\nabla \mathcal{E}_{\text{cont}}(s(t)), \quad s^{k+1} = \mathcal{U}_{\theta_k}(s^k, \mathcal{P}_{\Omega}(\mathcal{Y})) \quad (7.3)$$

之间的离散—连续对应。算法展开、交替方向乘子法和近端优化的相关研究^[14-16]可作为构造基础；未来需要进一步研究能量下降、连续动力系统稳定性、步长和阈值的可学习范围，以及不同采样密度和退化分布下的跨域适应。目标是使每个网络层或时间步都对应可说明的能量变化和状态转移，而不是只以输出指标解释模型。

三、医学图像应用。 在动态 MRI 中，将时间维度与空间维度共同作为连续低秩结构的坐标，研究欠采样、运动和噪声下的动态重建；在脑部 MRI 异常检测和分割中，将显式物理观测、健康组织参考、低维流形和病灶差异放入同一连续能量，研究结构先验如何帮助区分个体解剖差异与异常改变。可进一步探索健康参考配准、低秩潜空间约束和生成式重建之间的可微接口，但任何临床性能、跨设备泛化和诊断价值都必须在未来的独立数据与临床协议中验证。该方向的推进重点是：保持观测物理一致性，控制连续模型的外推风险，给出能够被医学专家理解的结构变化解释。

三层展望延续本报告的核心主线：连续低秩参数化回答恢复对象如何由函数型低秩因子生成，稳定可解释求解回答这些参数如何在连续结构能量和物理观测约束

下计算，医学图像应用则检验参数化与求解机制是否共同满足真实性、稳定性和可解释性要求。博士后阶段形成的显式塔克参数化、深度连续表示和有限层可学习求解器分别提供离散结构基线、可学习表示接口和算法状态语义；未来项目将在连续函数空间、稳定动力学和医学影像验证中进一步统一这三部分。

参考文献

- [1] Tucker L R. Some mathematical notes on three-mode factor analysis[J]. *Psychometrika*, 1966, 31(3): 279-311.
- [2] Kolda T G, Bader B W. Tensor decompositions and applications[J]. *SIAM Review*, 2009, 51(3): 455-500. DOI: 10.1137/07070111X.
- [3] Liu J, Musialski P, Wonka P, Ye J. Tensor Completion for Estimating Missing Values in Visual Data[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(1): 208-220. DOI: 10.1109/TPAMI.2012.39.
- [4] Kilmer M E, Martin C D. Factorization strategies for third-order tensors[J]. *Linear Algebra and its Applications*, 2011, 435(3): 641-658. DOI: 10.1016/j.laa.2010.11.023.
- [5] Lu C, Feng J, Chen Y, Liu W, Lin Z, Yan S. Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization[J]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 5249-5257.
- [6] Wang H, Peng J, Qin W, Wang J, Meng D. Guaranteed Tensor Recovery Fused Low-Rankness and Smoothness[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(9): 10990-11007. DOI: 10.1109/TPAMI.2023.3259640.
- [7] Ulyanov D, Vedaldi A, Lempitsky V. Deep Image Prior[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l. : s.n.], 2018: 9446-9454.
- [8] Sitzmann V, Martel J N P, Bergman A W, Lindell D B, Wetzstein G. Implicit Neural Representations with Periodic Activation Functions[C]//*Advances in Neural Information Processing Systems*: vol. 33. [S.l. : s.n.], 2020: 7462-7473.
- [9] Tancik M, Srinivasan P P, Mildenhall B, Fridovich-Keil S, Raghavan N, Singhal U, Ramamoorthi R, Barron J T, Ng R. Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains[C]//*Advances in Neural Information Processing Systems*: vol. 33. [S.l. : s.n.], 2020: 7537-7547.

- [10] Luo Y S, Zhao X L, Meng D, Jiang T X. HLRTF: Hierarchical Low-Rank Tensor Factorization for Inverse Problems in Multi-Dimensional Imaging[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l. : s.n.], 2022: 19281-19290. DOI: 10.1109/CVPR52688.2022.01870.
- [11] Wang J, Zhao X. Functional Transform-Based Low-Rank Tensor Factorization for Multi-Dimensional Data Recovery[C]//Lecture Notes in Computer Science: Computer Vision – ECCV 2024: vol. 15089. [S.l.]: Springer, 2024: 39-56. DOI: 10.1007/978-3-031-72751-1_3.
- [12] Wang C, Zhao X L, Zheng Y B, Li B Z, Ng M K. Functional Tensor Singular Value Decomposition[J]. SIAM Journal on Scientific Computing, 2025, 47(4): A2180-A2204. DOI: 10.1137/24M1644687.
- [13] Bolte J, Sabach S, Teboulle M. Proximal alternating linearized minimization for non-convex and nonsmooth problems[J]. Mathematical Programming, 2014, 146(1–2): 459-494. DOI: 10.1007/s10107-013-0701-9.
- [14] Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers[J]. Foundations and Trends in Machine Learning, 2011, 3(1): 1-122. DOI: 10.1561/22000000016.
- [15] Monga V, Li Y, Eldar Y C. Algorithm Unrolling: Interpretable, Efficient Deep Learning for Signal and Image Processing[J]. IEEE Signal Processing Magazine, 2021, 38(2): 18-44. DOI: 10.1109/MSP.2020.3016905.
- [16] Yang Y, Sun J, Li H, Xu Z. ADMM-CSNet: A Deep Learning Approach for Image Compressive Sensing[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(3): 521-538. DOI: 10.1109/TPAMI.2018.2883941.
- [17] Gong W, Huang Z, Lu J, Yang L. TuckerAPP: A novel spatiotemporal Tucker decomposition approach for traffic imputation[J]. Signal Processing, 2026, 240: 110383. DOI: 10.1016/j.sigpro.2025.110383.
- [18] Gong W, Yang L. LRTuckerRep: Low-rank Tucker Representation Model for Multi-dimensional Data Completion[Z]. 2025. arXiv: 2508.03755 [cs.LG]. DOI: 10.48550/arXiv.2508.03755.
- [19] Li Y, Gong W, Wang Q, Wang C, Yang L. 3DeepRep: 3D deep low-rank tensor representation for hyperspectral image inpainting[J]. Neurocomputing, 2026, 678: 133160. DOI: 10.1016/j.neucom.2026.133160.

- [20] Li Y, Wang Q, Gong W, Yang L. Channel-Aware Tensor Nuclear Norm: A Self-Supervised Approach for HSI Inpainting[C]//2026 IEEE International Conference on Image Processing (ICIP). Tampere, Finland: [s.n.], 2026. DOI: 10.1109/ICIP61757.2026.11630166.
- [21] Xu Y, Yin W. A Block Coordinate Descent Method for Regularized Multiconvex Optimization with Applications to Nonnegative Tensor Factorization and Completion[J]. SIAM Journal on Imaging Sciences, 2013, 6(3): 1758-1789. DOI: 10.1137/120887795.
- [22] Yokota T, Erem B, Guler S, Warfield S K, Hontani H. Missing Slice Recovery for Tensors Using a Low-Rank Model in Embedded Space[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l. : s.n.], 2018: 8251-8259. DOI: 10.1109/CVPR.2018.00861.
- [23] Xia X, Gong W, Yang L. Adaptive Wavelet Decomposition-Based Dual-Path Transformer for Long-Term Time Series Forecasting[C]//Lecture Notes in Computer Science: Advanced Intelligent Computing Technology and Applications: vol. 16646. [S.l.]: Springer, 2026: 14-25. DOI: 10.1007/978-981-92-3394-6_2.
- [24] Murad M M N, Aktukmak M, Yilmaz Y. WPMixer: Efficient Multi-Resolution Mixing for Long-Term Time Series Forecasting[C]//Proceedings of the AAAI Conference on Artificial Intelligence: vol. 39: 18. [S.l. : s.n.], 2025: 19581-19588. DOI: 10.1609/aaai.v39i18.34156.
- [25] Nie Y, Nguyen N H, Sinthong P, Kalagnanam J. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers[C]//International Conference on Learning Representations. [S.l. : s.n.], 2023. arXiv: 2211.14730.
- [26] Wu H, Hu T, Liu Y, Zhou H, Wang J, Long M. TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis[C]//International Conference on Learning Representations. [S.l. : s.n.], 2023.
- [27] Wang S, Wu H, Shi X, Hu T, Luo H, Ma L, Zhang J, Zhou J. TimeMixer: Decomposable Multiscale Mixing for Time Series Forecasting[C]//International Conference on Learning Representations. [S.l. : s.n.], 2024.
- [28] Shi X, Gong W, Yang L. Spatio-Temporal Decoupling Network with Frequency Domain Enhancement for Traffic Forecasting[C]//Lecture Notes in Computer Science: Advanced Intelligent Computing Technology and Applications: vol. 16647. [S.l.]: Springer, 2026: 14-26. DOI: 10.1007/978-981-92-3397-7_2.

- [29] Yu B, Yin H, Zhu Z. Spatio-temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting[J]. ArXiv preprint arXiv:1709.04875, 2017.
- [30] Xu M, Dai W, Liu C, Gao X, Lin W, Qi G J, Xiong H. Spatial-temporal Transformer Networks for Traffic Flow Forecasting[J]. ArXiv preprint arXiv:2001.02908, 2020.
- [31] Shao Z, Zhang Z, Wei W, Wang F, Xu Y, Cao X, Jensen C S. Decoupled Dynamic Spatial-temporal Graph Neural Network for Traffic Forecasting[J]. Proceedings of the VLDB Endowment, 2022, 15(11): 2733-2746. DOI: 10.14778/3551793.3551827.
- [32] Shi X, Gong W, Xu X, Yang L. POI-LLM: Leveraging POI Enhanced Large Language Models for Spatial-Temporal Modeling[C]//2025 IEEE 6th International Conference on Computer, Big Data and Artificial Intelligence (ICCBD+AI). Xiamen, China: [s.n.], 2025. DOI: 10.1109/ICCBDAI66607.2025.11388271.
- [33] Liu C, Yang S, Xu Q, Li Z, Long C, Li Z, Zhao R. Spatial-Temporal Large Language Model for Traffic Prediction[Z]. 2024. arXiv: 2401.10134. DOI: 10.48550/arXiv.2401.10134.
- [34] Li Z, Xia L, Tang J, Xu Y, Shi L, Xia L, Yin D, Huang C. UrbanGPT: Spatio-Temporal Large Language Models[C]//Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining. [S.l. : s.n.], 2024: 5351-5362.
- [35] Jin M, Wang S, Ma L, Chu Z, Zhang J, Shi X, Chen P, Liang Y, Li Y F, Pan S, et al. Time-LLM: Time Series Forecasting by Reprogramming Large Language Models[C]//International Conference on Learning Representations. [S.l. : s.n.], 2024.
- [36] Zheng C, Fan X, Wang C, Qi J. GMAN: A Graph Multi-Attention Network for Traffic Prediction[C]//Proceedings of the AAAI Conference on Artificial Intelligence: vol. 34: 1. [S.l. : s.n.], 2020: 1234-1241.
- [37] Schölkopf B, Locatello F, Bauer S, Ke N R, Kalchbrenner N, Goyal A, Bengio Y. Toward Causal Representation Learning[J]. Proceedings of the IEEE, 2021, 109(5): 612-634. DOI: 10.1109/JPROC.2021.3058954.
- [38] Khemakhem I, Kingma D P, Monti R P, Hyvarinen A. Variational Autoencoders and Nonlinear ICA: A Unifying Framework[C]//Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics. [S.l. : s.n.], 2020: 2207-2217.

- [39] Yang M, Liu F, Chen Z, Shen X, Hao J, Wang J. CausalVAE: Disentangled Representation Learning via Neural Structural Causal Models[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l. : s.n.], 2021: 9593-9602.
- [40] Liu Y, Zhang Z, Gong D, Gong M, Huang B, van den Hengel A, Zhang K, Shi J Q. Identifiable Latent Polynomial Causal Models through the Lens of Change[C]//The Twelfth International Conference on Learning Representations. [S.l. : s.n.], 2024.
- [41] Wang W, Lin X, Feng F, He X, Lin M, Chua T S. Causal Representation Learning for Out-of-Distribution Recommendation[C]//Proceedings of the ACM Web Conference 2022. [S.l. : s.n.], 2022: 3562-3571. DOI: 10.1145/3485447.3512251.
- [42] Wang W, Lin X, Wang L, Feng F, Ma Y, Chua T S. Causal Disentangled Recommendation against User Preference Shifts[J]. ACM Transactions on Information Systems, 2023, 42(1): 1-27.
- [43] Wang Q, Gong W, Li Y, Wu Z, Wang R, Yang L. Conditionally Identifiable Latent-Environment Modeling for Out-of-Distribution Recommendation[EB/OL]. 2026. <https://arxiv.org/abs/2608.03647>. arXiv: 2608.03647 [cs.IR].
- [44] Wang Q, Li Y, Huang D, Gong W, Yang L. ProME: Prototype-Margin Environments with Repair-Aware Selection for Group-Robust Learning[EB/OL]. 2026. <https://arxiv.org/abs/2608.13190>. arXiv: 2608.13190 [cs.LG].
- [45] Sagawa S, Koh P W, Hashimoto T B, Liang P. Distributionally Robust Neural Networks[C/OL]//International Conference on Learning Representations. [S.l. : s.n.], 2020. <https://openreview.net/forum?id=ryxGuJrFvS>.
- [46] Kirichenko P, Izmailov P, Wilson A G. Last Layer Retraining Is Sufficient for Robustness to Spurious Correlations[C/OL]//ICML 2022: Workshop on Spurious Correlations, Invariance, and Stability. [S.l. : s.n.], 2022. <https://openreview.net/forum?id=THOObYluWV>.
- [47] LaBonte T, Muthukumar V, Kumar A. Towards Last-Layer Retraining for Group Robustness with Fewer Annotations[C/OL]//Advances in Neural Information Processing Systems: vol. 36. [S.l. : s.n.], 2023: 11552-11579. https://proceedings.neurips.cc/paper_files/paper/2023/file/265bee74aee86df77e8e36d25e786abf-Paper-Conference.pdf.

- [48] Nam J, Cha H, Ahn S, Lee J, Shin J. Learning from Failure: De-biasing Classifier from Biased Classifier[C/OL]//Advances in Neural Information Processing Systems: vol. 33. [S.l. : s.n.], 2020: 20673-20684. https://proceedings.neurips.cc/paper_files/paper/2020/file/eddc3427c5d77843c2253f1e799fe933-Paper.pdf.
- [49] Qiao R, Wu Z, Wang J, Koh P W, Low B K H. Group-Robust Sample Reweighting for Subpopulation Shifts via Influence Functions[C/OL]//International Conference on Learning Representations. [S.l. : s.n.], 2025: 36959-36980. https://proceedings.iclr.cc/paper_files/paper/2025/file/5b6c1a77153c9d11861e26f92780cf2f-Paper-Conference.pdf.
- [50] Anonymous authors. GenCAR: Counterfactual Generation and Causal Alignment for Out-of-Distribution Recommendation[Z]. Unpublished manuscript. 2026.
- [51] Rendle S, Freudenthaler C, Gantner Z, Schmidt-Thieme L. BPR: Bayesian Personalized Ranking from Implicit Feedback[C]//Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence. [S.l. : s.n.], 2009: 452-461.
- [52] Schnabel T, Swaminathan A, Singh A, Chandak N, Joachims T. Recommendations as Treatments: Debiasing Learning and Evaluation[C]//Proceedings of the 33rd International Conference on Machine Learning. [S.l. : s.n.], 2016: 1670-1679.
- [53] Vovk V, Gammerman A, Shafer G. Algorithmic Learning in a Random World[M]. [S.l.]: Springer, 2005. DOI: 10.1007/b106715.
- [54] Bates S, Candes E, Lei L, Romano Y, Sesia M. Testing for Outliers with Conformal p -Values[J]. The Annals of Statistics, 2023, 51(1): 149-178. DOI: 10.1214/22-AOS2244.
- [55] Benjamini Y, Yekutieli D. The Control of the False Discovery Rate in Multiple Testing under Dependency[J]. The Annals of Statistics, 2001, 29(4): 1165-1188. DOI: 10.1214/aos/1013699998.

博士后个人简历

龚文武，数学博士，研究方向包括模型驱动可解释人工智能、低秩张量表示、计算成像、时空数据挖掘和跨学科数学建模。2024 年 9 月起在南方科技大学开展博士后研究，合作导师为杨丽丽教授。博士后阶段主持中国博士后科学基金面上项目“模型和数据双驱动的低秩张量分解研究”（批准号：2025M773057），并围绕计算成像、时空建模和因果稳健推荐开展研究。

2014 年 9 月至 2018 年 6 月就读于南昌大学数学学院，获数学学士学位；2018 年 9 月至 2020 年 6 月就读于哈尔滨工业大学数学学院，获数学硕士学位；2020 年 9 月至 2024 年 6 月就读于南方科技大学统计与数据科学系，获数学博士学位。

主持与参与的科研项目

项目名称	起止时间	项目类别、本人角色及编号
模型和数据双驱动的低秩张量分解研究	2025.03– 2026.09	中国博士后科学基金面上项目； 主持；2025M773057
基于多模式混合出行的交通事故风险预测分析技术研究	2021.07– 2025.07	广东省高等学校自然科学研究重点项目；课题骨干

博士后期间与本报告相关的科研成果

序号	成果名称（作者顺序按书目信息）	本人贡献	发表或公开信息
1	Gong, W.; Huang, Z.; Lu, J.; Yang, L. TuckerAPP: A Novel Spatiotemporal Tucker Decomposition Approach for Traffic Imputation	第一作者	<i>Signal Processing</i> , 240, 110383, 2026; DOI
2	Li, Y.; Gong, W.; Wang, Q.; Wang, C.; Yang, L. 3DeepRep: 3D Deep Low-Rank Tensor Representation for Hyperspectral Image Inpainting	共同第一作者	<i>Neurocomputing</i> , 678, 133160, 2026; DOI
3	Li, Y.; Wang, Q.; Gong, W.; Yang, L. Channel-Aware Tensor Nuclear Norm: A Self-Supervised Approach for HSI Inpainting	通讯作者	<i>IEEE International Conference on Image Processing (ICIP)</i> , 2026; DOI
4	Gong, W.; Yang, L. LRTuckerRep: Low-Rank Tucker Representation Model for Multi-Dimensional Data Completion	第一作者	arXiv:2508.03755, 2025; arXiv
5	Wang, Q.; Gong, W.; Li, Y.; Wu, Z.; Wang, R.; Yang, L. Conditionally Identifiable Latent-Environment Modeling for Out-of-Distribution Recommendation	通讯作者	arXiv:2608.03647, 2026; arXiv
6	Wang, Q.; Li, Y.; Huang, D.; Gong, W.; Yang, L. ProME: Prototype-Margin Environments with Repair-Aware Selection for Group-Robust Learning	通讯作者	arXiv:2608.13190, 2026; arXiv
7	Xia, X.; Gong, W.; Yang, L. Adaptive Wavelet Decomposition-Based Dual-Path Transformer for Long-Term Time Series Forecasting	通讯作者	<i>Advanced Intelligent Computing Technology and Applications</i> , LNCS 16646, 14–25, 2026; DOI
8	Shi, X.; Gong, W.; Yang, L. Spatio-Temporal Decoupling Network with Frequency Domain Enhancement for Traffic Forecasting	通讯作者	<i>Advanced Intelligent Computing Technology and Applications</i> , LNCS 16647, 14–26, 2026; DOI
9	Shi, X.; Gong, W.; Xu, X.; Yang, L. POI-LLM: Leveraging POI Enhanced Large Language Models for Spatial-Temporal Modeling	通讯作者	<i>IEEE International Conference on Computer, Big Data and Artificial Intelligence (ICCBDAI)</i> , 2025; DOI

致 谢

感谢合作导师杨丽丽教授在研究方向和科研组织方面的指导，感谢南方科技大学及相关科研平台提供的研究条件，感谢中国博士后科学基金第 77 批面上项目“模型和数据双驱动的低秩张量分解研究”（项目编号：2025M773057）以及南方科技大学校长卓越博士后计划的支持。感谢合作作者在实验验证与论文写作中的贡献。

永久通信地址

通信地址：正式提交前按学校要求填写

邮政编码：正式提交前填写

E-mail：正式提交前填写