

低秩和联合平滑性约束下的时变海表面温度重构方法

李姣^① 万腾汶^① 邱伟^{*②}

^①(长沙理工大学数学与统计学院 长沙 410114)

^②(中国人民解放军国防科技大学气象与海洋学院 长沙 410073)

摘要: 海表面温度对于海洋动力过程及海气相互作用等具有重要意义, 是海洋环境关键要素之一。浮标是海表面温度观测的常用手段, 但由于浮标在空间的分布不规则, 浮标采集的海表面温度数据也呈现非规则性。另外, 浮标实际工作中难免存在故障, 致使采集的海表面温度数据存在缺失。因此对存在缺失的非规则海表面温度数据进行重构具有重要意义。该文通过将海表面温度数据建立为时变图信号, 利用图信号处理方法解决海表面温度数据缺失数据重构问题。首先, 利用数据的低秩性和时域-图域联合变差特性构建海表面温度重构模型; 其次, 基于交替方向乘法框架提出一种求解该优化模型的基于低秩和联合平滑性(LRJS)的时变图信号重构方法, 并分析该方法的计算复杂度和估计误差的理论极限; 最后, 采用南海和太平洋海域海表温度数据对方法的有效性进行了评估, 结果表明, 与现有缺失数据重构方法相比, 该文所提LRJS方法有更高的重建精度。

关键词: 海表面温度; 时变图信号; 信号重构; 低秩; 联合平滑性

中图分类号: TN911.7

文献标识码: A

文章编号: 1009-5896(2024)00-0001-09

DOI: 10.11999/JEIT240253

Time-varying Sea Surface Temperature Reconstruction Leveraging Low Rank and Joint Smoothness Constraints

LI Jiao^① WAN Tengwen^① QIU Wei^{*②}

^①(School of Mathematics and Statistics, Changsha University of Science and Technology, Changsha 410114, China)

^②(School of Meteorology and Oceanography, National University of Defense Technology, Changsha 410073, China)

Abstract: Sea surface temperature is one of the key elements of the marine environment, which is of great significance to the marine dynamic process and air-sea interaction. Buoy is a commonly used method of sea surface temperature observation. However, due to the irregular distribution of buoys in space, the sea surface temperature data collected by buoys also show irregularity. In addition, it is inevitable that sometimes the buoy is out of order, so that the sea surface temperature data collected is incomplete. Therefore, it is of great significance to reconstruct the incomplete irregular sea surface temperature data. In this paper, the sea surface temperature data is established as a time-varying graph signal, and the graph signal processing method is used to solve the problem of missing data reconstruction of sea surface temperature. Firstly, the sea surface temperature reconstruction model is constructed by using the low rank data and the joint variation characteristics of time-domain and graph-domain. Secondly, a time-varying graph signal reconstruction method based on Low Rank and Joint Smoothness (LRJS) constraints is proposed to solve the optimization problem by using the framework of alternating direction multiplier method, and the computational complexity and the theoretical limit of the estimation error of the method are analyzed. Finally, the sea surface temperature data of the South China Sea and the Pacific Ocean are used to evaluate the effectiveness of the method. The results show that the LRJS method proposed in this paper can improve the reconstruction accuracy compared with the

收稿日期: 2024-04-09; 改回日期: 2024-10-11; 网络出版: 2024-10-15

*通信作者: 邱伟 qiuwei08@nudt.edu.cn

基金项目: 国家自然科学基金(42176197), 湖南省自然科学基金(2022JJ40461), 湖南省教育厅优秀青年项目(21B0301)

Foundation Items: The National Natural Science Foundation of China (42176197), The Natural Science Foundation of Hunan Province (2022JJ40461), The Excellent Youth Foundation of Education Bureau of Hunan Province (21B0301)

existing missing data reconstruction methods.

Key words: Sea surface temperature; Time-varying graph signal; Signal reconstruction; Low rank; Joint smoothness

1 引言

海表面温度对于研究海洋动力过程、海气相互作用、海洋资源开发利用、地球气候变化检测等问题^[1,2]具有重要意义。浮标和卫星遥感是获取海表面温度数据的重要方法。但是,浮标在进行海表面温度监测时,由于数量有限,难以覆盖整个海洋区域,且由于传感器故障等原因,使得观测海域的海表面温度数据难免存在数据缺失问题;此外,由于海洋浮标放置的位置受到海水流动影响,导致收集到的浮标数据呈现空间上的不规则分布;而利用卫星遥感数据观测海表面温度时,会受到云层、降水和气溶胶等因素的影响,导致获取的数据呈现空间上的不规则分布以及一定的缺失。因此,研究存在缺失的不规则海表面温度数据重构问题具有重要应用价值。现有对于此类不完整数据进行重构的方法,可以分为基于统计的经验模型^[3]和插值方法^[4]两类。对于前者,需要特定区域大量的历史数据,而对于缺少历史数据的区域,重构精度有待提高;对于插值方法,存在缺失的测量数据不足以确定插值方法的回归模型,如果预先选择的基函数不能反映数据之间的空间相关性,则会导致重构性能退化。上述方法都没有利用海表面温度数据的时空特性,难以有效处理非规则海表面温度数据重构问题。本文引入图信号处理方法^[5]来重构这类不规则海洋观测数据,将海表面温度数据建模为时变图信号,通过图的结构来捕获时空数据的关联关系,建立时变图信号重构模型用于缺失数据重构,从而提高重构的准确性。

当前,已提出了多种时变图信号重构算法。为降低算法的计算复杂度,文献^[6]提出了一种采用截断Jacobi迭代的广义牛顿法求解图信号矩阵补全问题;文献^[7]提出一种基于笛卡尔乘积图上Sobolev平滑的分布式批量重构算法,有效提高大规模数据重构效率。为降低重构误差,学者们提出了一系列图信号重构算法,该类算法主要是基于数据的相关性,然后采用不同的先验信息进行重构。文献^[8]提出了基于1阶时差信号平滑性时变图信号重构算法,显著降低了重构误差。在文献^[8]的基础上,文献^[9]加入了低秩先验信息,提出了基于低秩和差分平滑的重构算法(Low Rank and Differential Smoothness, LRDS),从而降低了重构误差;文献^[10]为降低重构误差和提高重构效率,提出了基于

Sobolev差分平滑的重构算法(Graph Time-varying-signal Reconstruction via Sobolev Smoothness, GraphTRSS);文献^[11]考虑了时变图信号的时域-图域联合平滑性,提出基于联合平滑的重构算法(Joint Smooth Reconstruction of Time-Varying Graph Signals, JSR-TVGS),进一步提高了重构精度。较文献^[9]而言,文献^[12]考虑信号的空间结构和时间相关性,建立了一个1阶马尔可夫模型用于时变图信号重构,降低了重构误差;文献^[13]基于2阶时差信号平滑与低秩性提出了重构算法,提高了重构精度。此外,文献^[14-16]从图信号处理的角度重建海洋数据,结果表明利用图信号处理的方法较已有方法而言更具有优越性。

受上述相关研究的启发,为了以较低误差重构非规则海表面温度,本文以图信号处理为基础,引入2阶时差信号,提出了基于低秩和联合平滑性(Low Rank and Joint Smoothness, LRJS)的时变图信号重构模型,采用交替方向乘法求解,并进一步分析了所提算法的计算复杂度和重构误差的极限。最后在两个公开的真实数据集上进行实验,实验结果表明,本文所提方法较已有的方法而言,重构误差更小。

2 图信号处理基础理论

在图信号处理^[5]中,图可以表示为由节点和边组成的图结构,其中节点表示数据样本,边表示节点之间的关联关系。本文将海表面温度数据抽象为时变图信号,建立一个无向图 $G = (V, E, \mathbf{W})$ 来表示海域面积,其中 $V = \{v_1, v_2, \dots, v_N\}$ 是顶点集, N 为顶点数; $E = \{e_{ij}\}$ 是边集, e_{ij} 表示顶点 v_i 和 v_j 有边相连; $\mathbf{W} \in R^{N \times N}$ 是图的权矩阵, \mathbf{W}_{ij} 表示顶点 v_i 和 v_j 边的权重。若海表面温度数据的采样时刻总数为 T ,对于图中第 i 个顶点 v_i ,假设该点在 t 时刻的图信号为 x_{it} ,那么整个图上的 t 时刻的静态图信号可以表示为 $\mathbf{x}_t = [x_{1t}, x_{2t}, \dots, x_{Nt}]^T \in R^N$,对于时变图信号 \mathbf{X} ,可以将其表示为形如 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T] \in R^{N \times T}$ 的矩阵形式。

图拉普拉斯矩阵^[5]定义为 $\mathbf{L}_G = \mathbf{D} - \mathbf{W}$,式中矩阵 $\mathbf{D} = \text{diag}([d_1, d_2, \dots, d_N])$ 表示图 G 的度矩阵;拉普拉斯矩阵 \mathbf{L}_G 满足 $\mathbf{L}_G = \mathbf{U} \mathbf{\Lambda}_G \mathbf{U}^T$,其中 \mathbf{U} 为 \mathbf{L}_G 的正交特征向量矩阵, $\mathbf{\Lambda}_G$ 为 \mathbf{L}_G 的非负特征值对角矩阵。图傅里叶变换^[5]定义为 $\widehat{\mathbf{X}} = \mathbf{U}^T \mathbf{X}$ 。

$$\mathbf{J}_{vt} = \begin{cases} 1, & v \in \mathbf{S}_t \\ 0, & v \notin \mathbf{S}_t \end{cases} \quad (11)$$

其中, v 表示顶点, \mathbf{S}_t 表示在 t 时刻被采样的顶点的集合。

因为矩阵秩最小化是 NP-hard 问题, 为了有效地求解式(10), 将其进行凸松弛, 从而优化问题转化为

$$\begin{aligned} \min_{\mathbf{X}} \mu \|\mathbf{X}\|_* + \frac{1}{2} \|\mathbf{J} \circ \mathbf{X} - \mathbf{Y}\|_F^2 + \frac{\alpha}{2} \text{tr}((\mathbf{X}\mathbf{D}_h)^T \\ \cdot \mathbf{L}_G(\mathbf{X}\mathbf{D}_h)) + \frac{\beta}{2} \text{tr}(\mathbf{X}\mathbf{L}_T\mathbf{X}^T) \end{aligned} \quad (12)$$

其中, 核范数项 $\|\mathbf{X}\|_*$ 是指矩阵 \mathbf{X} 中非零奇异值之和, $\mu, \alpha, \beta > 0$ 为正则化参数。

上述目标函数中的4项在重构问题中作用如下所示:

- (1) 第1项 $\|\mathbf{X}\|_*$ 为估计矩阵 \mathbf{X} 的低秩性质, 保证了时空信号的全局相关性;
- (2) 第2项 $\|\mathbf{J} \circ \mathbf{X} - \mathbf{Y}\|_F^2$ 使估计值与观测值保持一致, 从而保证了重建结果的可靠性;
- (3) 第3项 $\text{tr}((\mathbf{X}\mathbf{D}_h)^T \mathbf{L}_G(\mathbf{X}\mathbf{D}_h))$ 保证了2阶时差信号在图域上的平滑性;
- (4) 第4项 $\text{tr}(\mathbf{X}\mathbf{L}_T\mathbf{X}^T)$ 保证了时变图信号在时域上的平滑性。

3.3 交替方向乘子法求解

通过观察上述优化问题模型式(12), 可以注意到上述式子的第1项是可逼近的, 后3项是可微的, 因此选择应用交替方向乘子法^[9]来求解该优化问题。

首先, 引入辅助变量 \mathbf{Z} 将上述优化模型式(12)的等价表达表示为

$$\begin{aligned} \min_{\mathbf{X}, \mathbf{Z}} \mu \|\mathbf{Z}\|_* + \frac{1}{2} \|\mathbf{J} \circ \mathbf{X} - \mathbf{Y}\|_F^2 \\ + \frac{\alpha}{2} \text{tr}((\mathbf{X}\mathbf{D}_h)^T \mathbf{L}_G(\mathbf{X}\mathbf{D}_h)) + \frac{\beta}{2} \text{tr}(\mathbf{X}\mathbf{L}_T\mathbf{X}^T), \\ \text{s.t. } \mathbf{X} = \mathbf{Z} \end{aligned} \quad (13)$$

然后, 将上述等价模型的增广拉格朗日函数写为

$$\begin{aligned} L_\rho(\mathbf{X}, \mathbf{Z}, \mathbf{P}) = \frac{1}{2} \|\mathbf{J} \circ \mathbf{X} - \mathbf{Y}\|_F^2 \\ + \frac{\alpha}{2} \text{tr}((\mathbf{X}\mathbf{D}_h)^T \mathbf{L}_G(\mathbf{X}\mathbf{D}_h)) \\ + \frac{\beta}{2} \text{tr}(\mathbf{X}\mathbf{L}_T\mathbf{X}^T) + \mu \|\mathbf{Z}\|_* \\ + \langle \mathbf{P}, \mathbf{X} - \mathbf{Z} \rangle + \frac{\rho}{2} \|\mathbf{X} - \mathbf{Z}\|_F^2 \end{aligned} \quad (14)$$

其中, \mathbf{P} 表示拉格朗日乘子矩阵, ρ 表示罚参数, $\langle \cdot \rangle$ 表示两个矩阵的内积。

交替方向乘子法通过以下迭代方案找到了鞍点

$$\mathbf{X}^{k+1} = L_\rho(\mathbf{X}, \mathbf{Z}^k, \mathbf{P}^k) \quad (15)$$

$$\mathbf{Z}^{k+1} = L_\rho(\mathbf{X}^{k+1}, \mathbf{Z}, \mathbf{P}^k) \quad (16)$$

$$\mathbf{P}^{k+1} = \mathbf{P}^k + \rho(\mathbf{X}^{k+1} - \mathbf{Z}^{k+1}) \quad (17)$$

上述子问题式(15)相当于式(18)的形式

$$\begin{aligned} \mathbf{X}^{k+1} = \arg \min_{\mathbf{X}} \frac{1}{2} \|\mathbf{J} \circ \mathbf{X} - \mathbf{Y}\|_F^2 \\ + \frac{\alpha}{2} \text{tr}((\mathbf{X}\mathbf{D}_h)^T \mathbf{L}_G(\mathbf{X}\mathbf{D}_h)) \\ + \frac{\beta}{2} \text{tr}(\mathbf{X}\mathbf{L}_T\mathbf{X}^T) \\ + \frac{\rho}{2} \|\mathbf{X} - \mathbf{Z}^k + \frac{\mathbf{P}^k}{\rho}\|_F^2 \end{aligned} \quad (18)$$

采用算法1的共轭梯度法来迭代求解式(18), 将式(18)所示目标函数记为, 其梯度为

$$\begin{aligned} \nabla f(\mathbf{X}) = \mathbf{J} \circ \mathbf{X} - \mathbf{Y} + \alpha \mathbf{L}_G \mathbf{X} \mathbf{D}_h \mathbf{D}_h^T + \beta \mathbf{X} \mathbf{L}_T \\ + \rho \left(\mathbf{X} - \mathbf{Z}^k - \frac{\mathbf{P}^k}{\rho} \right) \end{aligned} \quad (19)$$

上述子问题式(16)相当于

$$\mathbf{Z}^{k+1} = \arg \min_{\mathbf{Z}} \frac{\mu}{\rho} \|\mathbf{Z}\|_* + \frac{1}{2} \|\mathbf{Z} - \mathbf{X}^{k+1} - \frac{\mathbf{P}^k}{\rho}\|_F^2 \quad (20)$$

式(20)的闭式解表示为

$$\mathbf{Z}^{k+1} = \text{SVT}_{\frac{\mu}{\rho}} \left(\mathbf{X}^{k+1} + \frac{\mathbf{P}^k}{\rho} \right) \quad (21)$$

其中, SVT为奇异值阈值算子, 定义为

$$\text{SVT}_\tau(\mathbf{X}) = \mathbf{U} \mathbf{A}_\tau(\boldsymbol{\Sigma}) \mathbf{V}^T \quad (22)$$

其中, \mathbf{U} , \mathbf{V} , $\boldsymbol{\Sigma}$ 来自于 \mathbf{X} 的奇异值分解, 即 $\mathbf{X} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^T$,

算法 1 子问题式(19)求解过程

输入: $\mathbf{Y}, \mathbf{J}, \mathbf{L}_G, \mathbf{L}_T, \mathbf{Z}^k, \mathbf{P}^k, \mathbf{D}, \alpha, \beta, \rho$, 终止迭代阈值 ε , 最大迭代次数 K

输出: 重构信号时变图信号 \mathbf{X}_i

初始化设置: $\mathbf{X}_i = 0, \Delta \mathbf{X}_i = 0, i = 0$

迭代:

(1) 确定步长:

$$\tau = - \frac{\langle \Delta \mathbf{X}_i, \nabla f(\mathbf{X}_i) \rangle}{\langle \Delta \mathbf{X}_i, \nabla f(\mathbf{X}_i) + \mathbf{Y} + \rho \mathbf{Z}_i - \mathbf{P}_i \rangle}$$

其中, $\Delta \mathbf{X}_i$ 为第 i 步的搜索方向, τ 为第 i 步的最优步长, 其由线性最小化步长准则 $\min_{\tau} f(\mathbf{X}_i + \tau \Delta \mathbf{X}_i)$ 决定。

确定步长:

(2) 更新搜索方向:

$$\mathbf{X}_{i+1} = \mathbf{X}_i + \tau \Delta \mathbf{X}_i; \xi = \frac{\|\nabla f(\mathbf{X}_{i+1})\|_F^2}{\|\nabla f(\mathbf{X}_i)\|_F^2};$$

$$\Delta \mathbf{X}_{i+1} = -\nabla f(\mathbf{X}_{i+1}) + \xi \Delta \mathbf{X}_i$$

终止条件: 如果 $i = K$ 或者 $\|\Delta \mathbf{X}_i\|_F \leq \varepsilon$, 则停止迭代; 否则令 $i = i + 1$, 继续迭代, 直至满足终止条件。

$\Lambda_\tau(x) := \text{sign}(x) \max(|x| - \tau, 0)$ 是 $\tau \in R^+$ 定义的软阈值算子。

通过上述分析, 本文提出了一种基于低秩和联合平滑性(LRJS)的时变图信号重构算法, 算法具体实现步骤如**算法2**所示。

算法 2 LRJS算法求解步骤

输入: $\mathbf{Y}, \mathbf{J}, \mathbf{L}_G, \mathbf{L}_T, \mathbf{Z}^k, \mathbf{P}^k, \mathbf{D}, \alpha, \beta, \rho, \mu$, 终止迭代阈值 ε , 最大迭代次数 K

输出: 重构海表面温度数据 \mathbf{X}^k

初始化设置: $\mathbf{X}^0 = \mathbf{Z}^0 = \mathbf{Y}, \mathbf{P}^0 = 0, k = 0$

迭代:

(1) 更新 \mathbf{X}^{k+1} : 利用上述表1的共轭梯度法;

(2) 更新 \mathbf{Z}^{k+1} : 利用式(22);

(3) 更新 \mathbf{P}^{k+1} : 利用式(18);

终止条件: 如果 $k = K$ 或者 $\|\Delta \mathbf{X}^k\|_F \leq \varepsilon$, 则停止迭代; 否则令 $k = k + 1$, 继续迭代, 直至满足终止条件。

3.4 算法复杂度分析

为了更好地理解LRJS算法的可拓展性, 现在对LRJS算法的计算复杂度进行分析。

在每次迭代当中, 矩阵 \mathbf{X} 通过**算法1**的共轭梯度法进行更新, 其中主要计算为式(19)的梯度计算中的矩阵乘法 $\mathbf{L}_G \mathbf{X} \mathbf{D}_h \mathbf{D}_h^T$, 其计算消耗为 $O(N^2T + NT^2)$ 。在利用共轭梯度法进行迭代时, 需要的迭代次数^[19]为 $O(1/\ln((\sqrt{\kappa(\mathbf{A})} + 1)/(\sqrt{\kappa(\mathbf{A})} - 1)))$, 其中 $\mathbf{A} = \hat{\mathbf{J}} + \alpha \hat{\mathbf{L}}_G \hat{\mathbf{D}} + \beta \hat{\mathbf{L}}_T + \rho \hat{\mathbf{I}}$, $\hat{\mathbf{J}} = \text{diag}(\text{vec}(\mathbf{J}))$, $\hat{\mathbf{L}}_G = (\mathbf{I}_T \otimes \mathbf{L}_G)$, $\hat{\mathbf{I}} = \mathbf{I}_{NT}$, $\hat{\mathbf{D}} = (\mathbf{D}_h \mathbf{D}_h^T \otimes \mathbf{I}_N)$, $\hat{\mathbf{L}}_T = (\mathbf{L}_T \otimes \mathbf{I}_N)$ 。

上述 $\kappa(\mathbf{A}) = \sigma_{\max}(\mathbf{A})/\sigma_{\min}(\mathbf{A})$ 为矩阵 \mathbf{A} 的条件数, $\sigma_{\max}(\mathbf{A})$, $\sigma_{\min}(\mathbf{A})$ 分别为矩阵 \mathbf{A} 的最大特征值和最小特征值, 所以矩阵 \mathbf{X} 更新的计算复杂度为

$$O\left(1/\ln \frac{\sqrt{\kappa(\mathbf{A})} + 1}{\sqrt{\kappa(\mathbf{A})} - 1} \cdot (N^2T + NT^2)\right) \quad (23)$$

下面分析 $\kappa(\mathbf{A})$ 的上界:

命题1 若 d_{\max} 是图 $G = (V, E, \mathbf{W})$ 的所有顶点 N 的度集合 $\{d_1, d_2, \dots, d_N\}$ 的最大度, 则它满足

$$\kappa(\mathbf{A}) \leq 1 + (1 + 32\alpha d_{\max} + 4\beta)/\rho \quad (24)$$

证明 估算矩阵 \mathbf{A} 的最大特征值和最小特征值, 其中矩阵 $\hat{\mathbf{J}}$, $\hat{\mathbf{L}}_G$, $\hat{\mathbf{I}}$, $\hat{\mathbf{D}}$, $\hat{\mathbf{L}}_T$ 和 $\hat{\mathbf{L}}_T$ 都为半正定矩阵, 并且 $\alpha, \beta > 0$ 。现在有 $\sigma_{\min}(\mathbf{A}) \geq \rho$, 对于矩阵 \mathbf{A} 最

$$\|\hat{\mathbf{X}} - \mathbf{X}\|_F \leq \frac{\|\mathbf{J} \circ \mathbf{X} - \mathbf{Y}\|_F + \mu \sqrt{2\text{rank}(\mathbf{X})} + \alpha \|\mathbf{L}_G \mathbf{X} \mathbf{D}_h \mathbf{D}_h^T\|_F + \beta \|\mathbf{X} \mathbf{L}_T\|_F}{16(1 - \gamma)} \quad (28)$$

证明 对于误差上界, 有

大特征值的范围, 有

$$\sigma_{\max}(\mathbf{A}) \leq \sigma_{\max}(\hat{\mathbf{J}}) + \alpha \sigma_{\max}(\hat{\mathbf{L}}_G) \sigma_{\max}(\hat{\mathbf{D}}) + \beta \sigma_{\max}(\hat{\mathbf{L}}_T) + \rho \quad (25)$$

因为 $\hat{\mathbf{J}}$ 为一个主对角线上元素只有0或者1的对角矩阵, 所以 $\sigma_{\max}(\hat{\mathbf{J}}) \leq 1$; 此外, 由式(3)对矩阵 \mathbf{D}_h 的定义, 可知矩阵 $\mathbf{D}_h \mathbf{D}_h^T$ 为一个方阵, 利用 Gershgorin circle定理^[20], 可计算得到 $\sigma_{\max}(\hat{\mathbf{D}}) \leq 16$; 同理, 对于图域拉普拉斯矩阵 \mathbf{L}_G , 可计算得到 $\sigma_{\max}(\hat{\mathbf{L}}_G) \leq 2d_{\max}$; 对于式(8)中的时域拉普拉斯矩阵 \mathbf{L}_T , 可计算得到 $\sigma_{\max}(\hat{\mathbf{L}}_T) \leq 4$; 将 $\sigma_{\max}(\hat{\mathbf{J}})$, $\sigma_{\max}(\hat{\mathbf{L}}_G)$, $\sigma_{\max}(\hat{\mathbf{D}})$, $\sigma_{\max}(\hat{\mathbf{L}}_T)$ 的上界代入式(26), 命题1得证。

应用基本不等式 $\ln(1+x) \geq x/(1+x)$, $\forall x \geq 0$ 和命题1, 则有

$$\begin{aligned} 1/\ln \frac{\sqrt{\kappa(\mathbf{A})} + 1}{\sqrt{\kappa(\mathbf{A})} - 1} &\leq \frac{\kappa(\mathbf{A}) + 1}{2} \\ &\leq 1 + \frac{\sqrt{1 + 32\alpha d_{\max} + 4\beta}}{2\sqrt{\rho}} \end{aligned} \quad (26)$$

令 $K = 1 + \frac{\sqrt{1 + 32\alpha d_{\max} + 4\beta}}{2\sqrt{\rho}}$, 则矩阵 \mathbf{X} 的更新需消耗 $O(K \cdot (N^2T + NT^2))$;

矩阵 \mathbf{Z} 的更新中, SVT计算占主要地位, 式(22)中的矩阵乘积计算复杂度为 $O(\min(N^2T, NT^2))$, 而矩阵进行奇异值分解(Singular Value Decomposition, SVD)计算的复杂度也是 $O(\min(N^2T, NT^2))$, 则矩阵 \mathbf{Z} 的更新计算复杂度为 $O(\min(N^2T, NT^2))$; 对于矩阵 \mathbf{P} , 其每次更新只涉及到矩阵求和与矩阵标量积, 因此其计算复杂度为 $O(NT)$ 。

综上所述, 本文所提LRJS算法的计算消耗由矩阵 \mathbf{X} 的更新所主导, 总的计算复杂度为 $O(K \cdot (N^2T + NT^2))$ 。

3.5 算法误差上界分析

令 \mathbf{X} 是真实的时变图信号, $\hat{\mathbf{X}}$ 是重构的时变图信号, 由下述命题2可以得到LRJS算法的误差上界。

命题2 若对于任意矩阵 $\mathbf{M} \in R^{N \times T}$, 存在常数 $\gamma \in [0, 1)$, 有

$$\left\| \mathbf{M} - \frac{1}{16} \mathbf{J} \circ \mathbf{M} - \frac{\alpha}{16} \mathbf{L}_G \mathbf{M} \mathbf{D}_h \mathbf{D}_h^T - \frac{\beta}{16} \mathbf{M} \mathbf{L}_T \right\|_F \leq \gamma \|\mathbf{M}\|_F \quad (27)$$

则LRJS算法有误差上界

$$\begin{aligned} \|\widehat{\mathbf{X}} - \mathbf{X}\|_F \leq & \underbrace{\|\widehat{\mathbf{X}} - \text{SVT}_{\frac{\mu}{16}}\left(\mathbf{X} - \frac{\alpha}{16}\mathbf{L}_G\mathbf{X}\mathbf{D}_h\mathbf{D}_h^T - \frac{\beta}{16}\mathbf{X}\mathbf{L}_T\right)\|_F}_A \\ & + \underbrace{\|\text{SVT}_{\frac{\mu}{16}}\left(\mathbf{X} - \frac{\alpha}{16}\mathbf{L}_G\mathbf{X}\mathbf{D}_h\mathbf{D}_h^T - \frac{\beta}{16}\mathbf{X}\mathbf{L}_T\right) - \mathbf{X}\|_F}_B \end{aligned} \quad (29)$$

首先, 为了处理A项, 讨论收敛点 $\widehat{\mathbf{X}}$, 利用Karush-Kuhn-Tucker条件, 有

$$\left. \begin{aligned} \widehat{\mathbf{X}} &= \widehat{\mathbf{Z}} \\ 0 &= \mathbf{J} \circ \widehat{\mathbf{X}} - \mathbf{Y} + \alpha\mathbf{L}_G\widehat{\mathbf{X}}\mathbf{D}_h\mathbf{D}_h^T + \beta\widehat{\mathbf{X}}\mathbf{L}_T + \widehat{\mathbf{P}} \\ 0 &\in \mu\partial_{\|\cdot\|_*}(\widehat{\mathbf{Z}}) - \widehat{\mathbf{P}} \end{aligned} \right\} \quad (30)$$

共同应用上述3个条件, 则有

$$\mathbf{Y} - \mathbf{J} \circ \widehat{\mathbf{X}} - \alpha\mathbf{L}_G\widehat{\mathbf{X}}\mathbf{D}_h\mathbf{D}_h^T - \beta\widehat{\mathbf{X}}\mathbf{L}_T \in \mu\partial_{\|\cdot\|_*}(\widehat{\mathbf{X}}) \quad (31)$$

通过式(31), 有

$$\begin{aligned} \widehat{\mathbf{X}} + \frac{1}{16}(\mathbf{Y} - \mathbf{J} \circ \widehat{\mathbf{X}} - \alpha\mathbf{L}_G\widehat{\mathbf{X}}\mathbf{D}_h\mathbf{D}_h^T - \beta\widehat{\mathbf{X}}\mathbf{L}_T) \\ \in \frac{\mu}{16}\partial_{\|\cdot\|_*}(\widehat{\mathbf{X}}) + \widehat{\mathbf{X}} \end{aligned} \quad (32)$$

根据文献[21]的定理3, 对于任何 $\tau \in R^+$, 有 $\text{SVT}_\tau = (I + \partial_{\|\cdot\|_*})^{-1}$, 在式(32)中, 通过令 $\tau = 1/16$, 有

$$\begin{aligned} \widehat{\mathbf{X}} &= \text{SVT}_{\frac{\mu}{16}}\left(\widehat{\mathbf{X}} - \frac{1}{16}(\mathbf{J} \circ \widehat{\mathbf{X}} - \mathbf{Y}) \right. \\ &\quad \left. - \frac{1}{16}\alpha\mathbf{L}_G\widehat{\mathbf{X}}\mathbf{D}_h\mathbf{D}_h^T - \frac{1}{16}\beta\widehat{\mathbf{X}}\mathbf{L}_T\right) \end{aligned} \quad (33)$$

根据文献[21]引理1中关于近端算子的性质, 即对于任意两个矩阵 $\mathbf{Y}_1, \mathbf{Y}_2$, 以及任意 $\tau \in R^+$, 有

$$\|\text{SVT}_\tau(\mathbf{Y}_1) - \text{SVT}_\tau(\mathbf{Y}_2)\|_F \leq \|\mathbf{Y}_1 - \mathbf{Y}_2\|_F \quad (34)$$

现在将式(33)代入式(29)中的A项, 并应用式(34), 有

$$\begin{aligned} A \leq & \|\widehat{\mathbf{X}} - \mathbf{X} - \frac{1}{16}\alpha\mathbf{L}_G(\widehat{\mathbf{X}} - \mathbf{X})\mathbf{D}_h\mathbf{D}_h^T \\ & - \frac{1}{16}\beta(\widehat{\mathbf{X}} - \mathbf{X})\mathbf{L}_T - \frac{1}{16}(\mathbf{J} \circ \widehat{\mathbf{X}} - \mathbf{Y})\|_F \end{aligned} \quad (35)$$

进一步考虑式(27), 并且令 $\mathbf{M} = \widehat{\mathbf{X}} - \mathbf{X}$ 有

$$A \leq \gamma\|\widehat{\mathbf{X}} - \mathbf{X}\|_F + \frac{1}{16}\|\mathbf{J} \circ \widehat{\mathbf{X}} - \mathbf{Y}\|_F \quad (36)$$

接下来, 对于B项

$$\begin{aligned} B \leq & \|\text{SVT}_{\frac{\mu}{16}}\left(\mathbf{X} - \frac{\alpha}{16}\mathbf{L}_G\mathbf{X}\mathbf{D}_h\mathbf{D}_h^T - \frac{\beta}{16}\mathbf{X}\mathbf{L}_T\right) \\ & - \left(\mathbf{X} - \frac{\alpha}{16}\mathbf{L}_G\mathbf{X}\mathbf{D}_h\mathbf{D}_h^T - \frac{\beta}{16}\mathbf{X}\mathbf{L}_T\right)\|_F \\ & + \frac{\alpha}{16}\|\mathbf{L}_G\mathbf{X}\mathbf{D}_h\mathbf{D}_h^T\|_F + \frac{\beta}{16}\|\mathbf{X}\mathbf{L}_T\|_F \\ \leq & \|\mathbf{A}_{\frac{\mu}{16}}(\boldsymbol{\Sigma}) - \boldsymbol{\Sigma}\|_F + \frac{\alpha}{16}\|\mathbf{L}_G\mathbf{X}\mathbf{D}_h\mathbf{D}_h^T\|_F \\ & + \frac{\beta}{16}\|\mathbf{X}\mathbf{L}_T\|_F \\ \leq & \frac{\mu}{16}\sqrt{2\text{rank}(\mathbf{X})} + \frac{\alpha}{16}\|\mathbf{L}_G\mathbf{X}\mathbf{D}_h\mathbf{D}_h^T\|_F \\ & + \frac{\beta}{16}\|\mathbf{X}\mathbf{L}_T\|_F \end{aligned} \quad (37)$$

其中, $\boldsymbol{\Sigma}$ 是以矩阵 $\mathbf{X} - \frac{\alpha}{16}\mathbf{L}_G\mathbf{X}\mathbf{D}_h\mathbf{D}_h^T - \frac{\beta}{16}\mathbf{X}\mathbf{L}_T$ 的奇异值为对角线元素的对角矩阵。

最后将式(36)和式(37)的上界代入式(29), 则可证得命题2。

4 仿真实验

为测试本文所提LRJS方法的重构性能, 将本文所提方法与GraphTRSS算法^[8]、LRDS算法^[13]、JSR-TVGS算法^[14]、近似SVD分解的固定点迭代法(Fixed Point Continuation with Approximate SVD, FPCA)^[21]、自然邻点插值法(Natural Neighbor Interpolation, NNI)^[22]5种方法进行比较。

仿真实验中, 利用图信号处理工具箱GSP-BOX^[23]构建图 $G = (V, E, \mathbf{W})$ 。利用 k 最近邻法构造一个无向加权图来表征观测点的地理邻接关系, 为了得到一个连接的连通图, 取 $k = 10$ 。设置随机采样比例为10%~90%, 对应数据缺失率为90%~10%。在每次测试中, 采样数据的节点均是随机的, 并在不同采样率情况下分别进行100次蒙特卡洛测试, 最大迭代次数为 $K = 2 \times 10^3$, 终止迭代阈值 $\varepsilon = 10^{-6}$ 。

采用均方根误差(Root Mean Square Error, RMSE)和平均绝对误差(Mean Absolute Error, MAE)两个指标来评价方法的重构性能, 计算公式为

$$\text{MAE} = \left(\sum_{i=1}^{N_x} |\hat{x}_i - x_i| \right) / N_x,$$

$$\text{RMSE} = \sqrt{\left(\sum_{i=1}^{N_x} (\hat{x}_i - x_i)^2 \right) / N_x} \quad (38)$$

其中, \hat{x} 表示重构信号, x 为真实信号, N_x 是信号长度。MAE和RMSE越小, 表示重构数据与真实数据的匹配度越高, 即重构算法的重构精度越高。

本文所有仿真实验的硬件环境均为配置2.50 GHz Intel i5-7200U处理器、8 GB内存的计算机, 软件环境为Matlab R2022a。

4.1 实验结果

(1) 太平洋海平面温度数据集。该实验数据集来自美国NOAA地球系统研究实验室发布的海面温度数据^[24], 该数据集覆盖从1870年-2014年采集的月平均海表面温度数据, 空间分辨率为 $1^\circ \times 1^\circ$ 全球网格。在太平洋海域上随机选择100个位置采集的数据进行实验, 选取观测时间为200个月, 因此数据大小为 100×200 。

算法中设置的参数为 $\mu = 0.5, \alpha = 0.01, \beta = 0.001$,

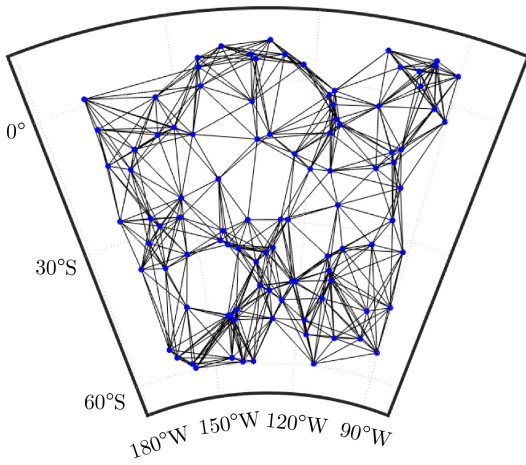
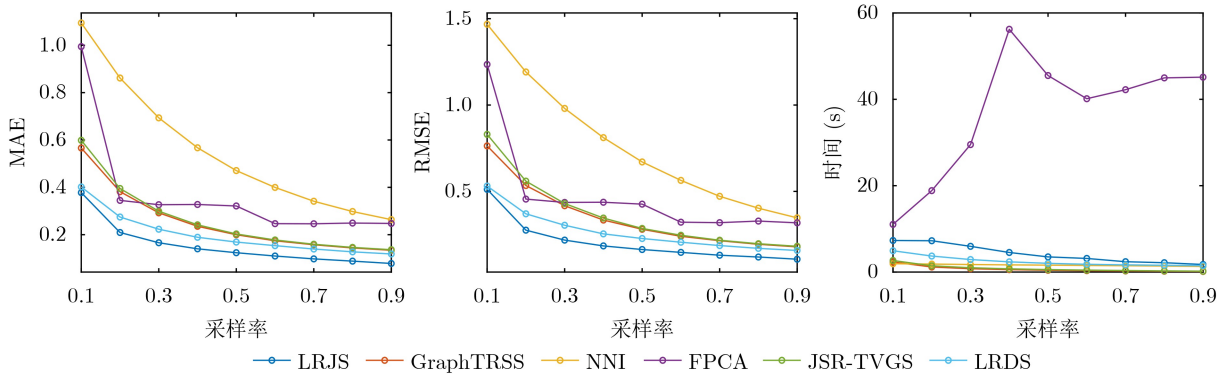


图1 太平洋海域海表面温度数据采集连接图



(a) 不同采样率下MAE

(b) 不同采样率下RMSE

(c) 不同采样率下算法运行时间

图2 所有方法在太平洋海域海表面温度数据集上的性能

$\rho = 1$ 。利用GSPBOX构建的太平洋海域海表面温度数据采集连接图如图1所示。对于该数据集, 不同方法在不同采样率条件下海表面温度重构的MAE, RMSE和运行时间结果如图2所示。

(2) 南海海平面温度数据集。该实验使用ARGO南海海平面温度数据集^[25]作为实验数据集。该数据集的空间分辨率为 $1^\circ \times 1^\circ$ 全球网格, 实验中, 在南海海域随机选择66个地点进行模拟, 选取的观测时间为2010年到2019年共计120个月, 因此数据的大小为 66×120 。

算法中设置的参数为 $\mu = 0.5, \alpha = 0.01, \beta = 0.001, \rho = 0.5$ 。太平洋海域海表面温度数据采集连接图如图3所示。对于该数据集, 不同方法在不同采样率条件下海表面温度重构的MAE, RMSE和运行时间结果如图4所示。

4.2 实验结果分析

上述仿真实验主要从MAE, RMSE以及算法运行时间3个方面比较了各算法的性能, 具体分析如下:

(1) 从图2和图4可以看出, 随着采样率的提高, LRJS算法的MAE和RMSE逐渐降低, 这是因为随着采样率的增加提供了更充分的图信号先验信息, 从而减小了重构误差, 这验证了本文算法的合理性。在相同的采样率条件下, 与其他5种算法相比, 本文所提LRJS算法表现出更小的MAE和RMSE, 表明其重构精度更高, 这是因为LRJS算法充分利用了更多的先验知识。

(2) 从图2(a)和图2(b)可知, 在太平洋海域海平面温度数据集的实验中, 随着采样率的增加, 本文所提LRJS算法相较于LRDS算法而言, MAE下降了6%~34%, RMSE下降了4%~32%, 相较于JSR-TVGS算法而言, MAE下降了37%~45%, RMSE下降了39%~51%。

(3) 从图4(a)和图4(b)可知, 随着采样率的增加, 在南海海平面温度数据集的实验中, 本文所提

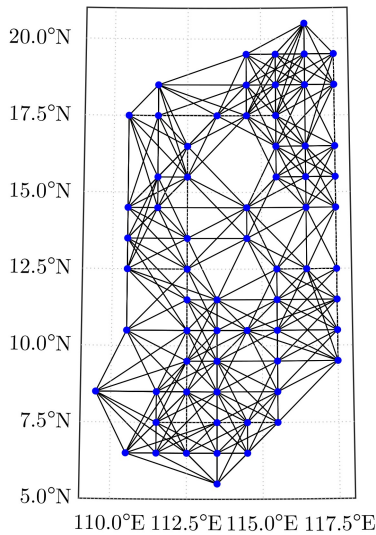


图3 南海海域海表温度数据采集连接图

LRJS算法相较于LRDS算法而言, MAE下降了5%~19%, RMSE下降了4%~17%; 本文所提LRJS算法相较于JSR-TVGS算法而言, MAE下降了20%~31%, RMSE下降了17%~29%。

(4)但是从图2(c)和图4(c)可知, 在算法重构效率方面, 在相同的采样率下, LRJS算法的运行时

间仅仅低于FPCA算法, 这是因为在每次迭代过程中, 核范数都需要进行SVD分解和利用共轭梯度法进行子问题的更新, 导致计算成本较高。

上述仿真实验结果表明, 本文所提LRJS算法尽管在重构效率方面, 未能表现出太大优势, 但是提高了海表温度的重构精度, 综合考虑所有进行比较的性能指标, 本文所提LRJS算法仍占有优势。

5 结束语

本文研究了基于图信号处理的海表温度数据重构问题。通过将海表面温度数据建模为时变图信号, 并将时变图信号重构问题转化为凸优化问题, 利用海表面温度的低秩性, 以及该信号在图域和时域联合平滑性, 提出了LRJS算法, 并采用交替方向乘法来进行求解, 最终实现时变图信号重构, 解决海表温度数据重构问题。在两种不同的真实数据集上进行了仿真实验, 结果表明, 与其他相关的时变图信号重构算法相比, 本文所提LRJS算法能够获得更高的重构精度, 表明了LRJS算法在时变图信号重构方面的有效性。下一步的研究将探索在确保较小重构误差的基础上, 改进优化算法的重构效率, 以更好地满足实际应用需求。

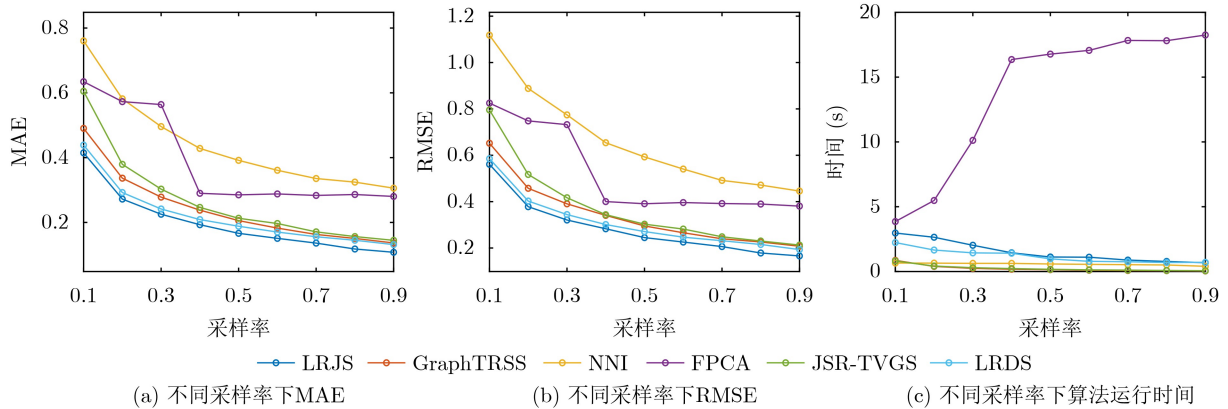


图4 所有方法在南海海域海表温度数据集上的性能

参考文献

- [1] DONEY S C, RUCKELSHAUS M, EMMETT DUFFY J, et al. Climate change impacts on marine ecosystems[J]. *Annual Review of Marine Science*, 2012, 4: 11–37. doi: 10.1146/annurev-marine-041911-111611.
- [2] LODER J W, VAN DER BAAREN A, and YASHAYAEV I. Climate comparisons and change projections for the Northwest Atlantic from six CMIP5 models[J]. *Atmosphere-Ocean*, 2015, 53(5): 529–555. doi: 10.1080/07055900.2015.1087836.
- [3] BARREIRO M, CHANG Ping, and SARAVANAN R. Variability of the South Atlantic convergence zone simulated by an atmospheric general circulation model[J]. *Journal of Climate*, 2002, 15(7): 745–763. doi: 10.1175/1520-0442(2002)015<0745:VOTSAC>2.0.CO;2.
- [4] 王奎民. 主要海洋环境因素对水下航行器航行影响分析[J]. *智能系统学报*, 2015, 10(2): 316–323. doi: 10.3969/j.issn.1673-4785.201503006.
- [5] WANG Kuimin. Influence of main ocean environments on the navigation of underwater vehicles[J]. *CAA Transactions on Intelligent Systems*, 2015, 10(2): 316–323. doi: 10.3969/j.issn.1673-4785.201503006.
- [5] SHUMAN D I, NARANG S K, FROSSARD P, et al. The emerging field of signal processing on graphs: Extending

- high-dimensional data analysis to networks and other irregular domains[J]. *IEEE Signal Processing Magazine*, 2013, 30(3): 83–98. doi: [10.1109/MSP.2012.2235192](https://doi.org/10.1109/MSP.2012.2235192).
- [6] LIU Jinling, JIANG Junzheng, LIN Jiming, *et al.* Generalized Newton methods for graph signal matrix completion[J]. *Digital Signal Processing*, 2021, 112: 103009. doi: [10.1016/j.dsp.2021.103009](https://doi.org/10.1016/j.dsp.2021.103009).
- [7] 张彦海, 蒋俊正. 基于笛卡尔乘积图上Sobolev平滑的时变图信号分布式批量重构[J]. *电子与信息学报*, 2023, 45(5): 1585–1592. doi: [10.11999/JEIT221194](https://doi.org/10.11999/JEIT221194).
ZHANG Yanhai and JIANG Junzheng. Distributed batch reconstruction of time-varying graph signals via Sobolev smoothness on Cartesian product graph[J]. *Journal of Electronics & Information Technology*, 2023, 45(5): 1585–1592. doi: [10.11999/JEIT221194](https://doi.org/10.11999/JEIT221194).
- [8] QIU Kai, MAO Xianghui, SHEN Xinyue, *et al.* Time-varying graph signal reconstruction[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2017, 11(6): 870–883. doi: [10.1109/JSTSP.2017.2726969](https://doi.org/10.1109/JSTSP.2017.2726969).
- [9] MAO Xianghui, QIU Kai, LI Tiejian, *et al.* Spatio-temporal signal recovery based on low rank and differential smoothness[J]. *IEEE Transactions on Signal Processing*, 2018, 66(23): 6281–6296. doi: [10.1109/TSP.2018.2875886](https://doi.org/10.1109/TSP.2018.2875886).
- [10] GIRALDO J H, MAHMOOD A, GARCIA-GARCIA B, *et al.* Reconstruction of time-varying graph signals via Sobolev smoothness[J]. *IEEE Transactions on Signal and Information Processing over Networks*, 2022, 8: 201–214. doi: [10.1109/TSIPN.2022.3156886](https://doi.org/10.1109/TSIPN.2022.3156886).
- [11] 索秋月, 林基明, 王俊义. 利用联合平滑性的时变图信号重构算法[J]. *计算机应用与软件*, 2022, 39(4): 275–280. doi: [10.3969/j.issn.1000-386x.2022.04.043](https://doi.org/10.3969/j.issn.1000-386x.2022.04.043).
SUO Qiuyue, LIN Jiming, and WANG Junyi. A time-varying graph signal reconstruction algorithm based on joint smoothness[J]. *Computer Applications and Software*, 2022, 39(4): 275–280. doi: [10.3969/j.issn.1000-386x.2022.04.043](https://doi.org/10.3969/j.issn.1000-386x.2022.04.043).
- [12] ZHAI Shiyu, LI Guobing, ZHANG Guomei, *et al.* Spatio-temporal signal recovery under diffusion-induced smoothness and temporal correlation priors[J]. *IET Signal Processing*, 2022, 16(2): 157–169. doi: [10.1049/sil2.12082](https://doi.org/10.1049/sil2.12082).
- [13] LIU Jinling, LIN Jiming, QIU Hongbing, *et al.* Time-varying signal recovery based on low rank and graph-time smoothness[J]. *Digital Signal Processing*, 2023, 133: 103821. doi: [10.1016/j.dsp.2022.103821](https://doi.org/10.1016/j.dsp.2022.103821).
- [14] QI Zefeng, LIAO Xuewen, LI Ang, *et al.* Signal recovery for incomplete ocean data via graph signal processing[C]. Proceedings of GLOBECOM 2022-2022 IEEE Global Communications Conference, Rio de Janeiro, Brazil, 2022: 4274–4279. doi: [10.1109/GLOBECOM48099.2022.10000902](https://doi.org/10.1109/GLOBECOM48099.2022.10000902).
- [15] LI Siyuan, CHENG Lei, ZHANG Ting, *et al.* Graph-guided Bayesian matrix completion for ocean sound speed field reconstruction[J]. *The Journal of the Acoustical Society of America*, 2023, 153(1): 689–710. doi: [10.1121/10.0017064](https://doi.org/10.1121/10.0017064).
- [16] CHEN Yangge, CHENG Lei, and WU Y C. Bayesian low-rank matrix completion with dual-graph embedding: Prior analysis and tuning-free inference[J]. *Signal Processing*, 2023, 204: 108826. doi: [10.1016/j.sigpro.2022.108826](https://doi.org/10.1016/j.sigpro.2022.108826).
- [17] LOUKAS A and FOUCARD D. Frequency analysis of time-varying graph signals[C]. Proceedings of 2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Washington, USA, 2016: 346–350. doi: [10.1109/GlobalSIP.2016.7905861](https://doi.org/10.1109/GlobalSIP.2016.7905861).
- [18] GRASSI F, LOUKAS A, PERRAUDIN N, *et al.* A time-vertex signal processing framework: Scalable processing and meaningful representations for time-series on graphs[J]. *IEEE Transactions on Signal Processing*, 2018, 66(3): 817–829. doi: [10.1109/TSP.2017.2775589](https://doi.org/10.1109/TSP.2017.2775589).
- [19] HACKBUSCH W. Iterative Solution of Large Sparse Systems of Equations[M]. New York, USA: Springer, 1994: 248–295. doi: [10.1007/978-1-4612-4288-8](https://doi.org/10.1007/978-1-4612-4288-8).
- [20] BELL H E. Gershgorin's theorem and the zeros of polynomials[J]. *The American Mathematical Monthly*, 1965, 72(3): 292–295. doi: [10.2307/2313703](https://doi.org/10.2307/2313703).
- [21] MA Shiqian, GOLDFARB D, and CHEN Lifeng. Fixed point and Bregman iterative methods for matrix rank minimization[J]. *Mathematical Programming*, 2011, 128(1/2): 321–353. doi: [10.1007/s10107-009-0306-5](https://doi.org/10.1007/s10107-009-0306-5).
- [22] SIBSON R. A brief description of natural neighbour interpolation[M]. BARNETT V. Interpreting Multivariate Data. Hoboken: Wiley, 1981.
- [23] PERRAUDIN N, PARATTE J, SHUMAN D, *et al.* GSPBOX: A toolbox for signal processing on graphs[J]. Eprint Arxiv, 2016, 61(7): 1644–1656. (查阅网上资料, 未找到本条文献刊名和卷期页码信息, 请确认).
- [24] Physical Sciences Division, Earth System Research Laboratory, National Oceanic and Atmospheric Administration. Sea surface temperature (SST) V2[EB/OL]. <http://www.esrl.noaa.gov/psd/data/gridded/data.noaa.oisst.v2.html>, 2023.
- [25] China Argo Real-Time Data Center. Argo real-time data[EB/OL]. <ftp://data.argo.org.cn/pub/ARGO/>, 2023.
- 李 姣: 女, 副教授, 研究方向为最优化方法及应用、生物分子计算。
万腾汶: 女, 硕士生, 研究方向为最优化方法及应用。
邱 伟: 男, 副教授, 研究方向为稀疏信号重构、海洋信息处理。