# 面向时空交通数据修复及预测的低秩机器学习模型

陈新宇 (https://xinychen.github.io)
伍元凯 (https://kaimaoge.github.io)
孙立君 (https://lijunsun.github.io)

发布时间：2023 年 2 月
更新时间：2023 年 2 月

本文唯一下载网址：
https://xinychen.github.io/books/spatiotemporal_low_rank_models.pdf

# 目录

# 第一章　前言

## 1.1　基本代数结构

长期以来，线性代数一直作为机器学习中最为重要的数学工具之一，被人们广泛用于开发各类机器学习算法。线性代数本质上是以向量与矩阵为基本代数结构，本文要讨论的张量分解等模型则主要以张量为基本代数结构。在过去的数十年间，借助线性代数这一基本数学工具，机器学习中涌现出了很多经典的代数模型，这其中不乏矩阵分解、主成分分析，而张量分解在某种程度上可看作是矩阵分解的一种衍生物高维扩展。

近年来，张量分解在机器学习的众多问题中得到了很好的应用 [Kolda and Bader, 2009, Sidiropoulos et al., 2017]，但关于张量的一些计算与我们所熟悉的线性代数却大相径庭，同时，张量计算相比以矩阵计算为主导的线性代数更为抽象，这使得很多与张量分解相关的内容看起来晦涩难懂。实际上，向量与矩阵都是张量的特例，可以被定义为低阶张量。一般而言，向量是一阶张量，英文表述为 first-order tensor；矩阵是二阶张量，英文表述为 second-order tensor；三阶或者更高阶数的张量被称为高阶张量，英文表述为 higher-order tensor。在各类文献中，通常提到的张量都是特指高阶张量，当然，这在本文的叙述中也不例外。需要注意的是，在各类程序语言中，人们更愿意将张量称为多维数组。

在一个矩阵中，某一元素的位置可以说是“第 $i$ 行、第 $j$ 列”，即要描述某一元素的位置需用到行和列索引构成的组合 $(i,j)$。类似地，在一个三阶张量中，描述某一元素的位置需用到三个索引构成的组合，例如 $(i,j,k)$。在处理稀疏矩阵或稀疏张量时，用索引来标记元素的位置会节省下一些不必要的存储开支。

### 1.1.1　向量与矩阵

#### 向量

向量包括行向量与列向量。在写法上，为避免混淆，向量在没有特别申明的情况下是指列向量，给定任意向量 $\boldsymbol{x}\in\mathbb{R}^n$ 表示大小为 $n$ 的向量，写作

$$\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^\top \tag{1.1}$$

或

$$\boldsymbol{x}=\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix} \tag{1.2}$$

其中，符号 $\cdot^\top$ 表示转置 (transpose)。

### 矩阵

一般而言，给定任意矩阵 $\boldsymbol{X} \in \mathbb{R}^{m\times n}$，矩阵的行数为 $m$、列数为 $n$，其第 $(i,j)$ 个元素（即矩阵的第 $i$ 行、第 $j$ 列元素）为

$$x_{i,j} = \boldsymbol{X}_{i,j} \tag{1.3}$$

其中，$i = 1,2,\ldots,m$ 与 $j = 1,2,\ldots,n$。

单位矩阵一般记作 $\boldsymbol{I}_n$，大小为 $n\times n$，其对角线上的元素均为 1、其他位置上的元素均为 0。

### 矩阵向量化

给定任意矩阵 $\boldsymbol{X} \in \mathbb{R}^{m\times n}$，若矩阵的列向量为 $\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_n \in \mathbb{R}^m$，即

$$\boldsymbol{X} = \begin{bmatrix} | & | & & | \\ \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \\ | & | & & | \end{bmatrix} \tag{1.4}$$

则可对矩阵按列进行向量化，得到的向量为

$$\operatorname{vec}(\boldsymbol{X}) = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_n \end{bmatrix} \in \mathbb{R}^{mn} \tag{1.5}$$

其中，符号 $\operatorname{vec}(\cdot)$ 表示向量化操作。

与矩阵向量化相反，也可定义向量的矩阵化规则。

## 1.1.2 高阶张量

一般而言，高阶张量可写成 $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{m_1\times m_2\times\cdots\times m_d}$，张量的阶数为 $d$，大小为 $m_1\times m_2\times \cdots\times m_d$。

### 三阶张量中的元素

这里以三阶张量为例，给定任意三阶张量 $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{m\times n\times t}$，其第 $(i,j,k)$ 个元素可写作如下形式：

$$x_{i,j,k} = \boldsymbol{\mathcal{X}}_{i,j,k} \tag{1.6}$$

其中，$i = 1,2,\ldots,m$、$j = 1,2,\ldots,n$ 与 $k = 1,2,\ldots,t$。

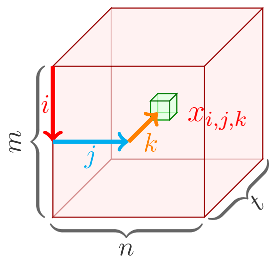
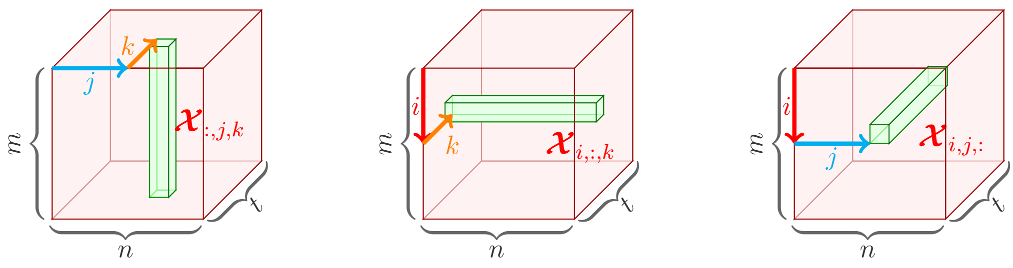



图 1.1: 三阶张量 $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{m\times n\times t}$ 及其第 $(i,j,k)$ 个元素 $x_{i,j,k}$

图1.1直观地展现了三阶张量元素的示意图，可以看出：描述三阶张量中的某一元素需用到三个索引构成的组合，例如 $(i,j,k)$。

### 三阶张量中的纤维

给定任意三阶张量 $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{m\times n\times t}$, 其各个方向的纤维 (fiber) 都是向量, 如图1.2所示, 这些纤维分别为向量 $\boldsymbol{\mathcal{X}}_{:,j,k} \in \mathbb{R}^m$、$\boldsymbol{\mathcal{X}}_{i,:,k} \in \mathbb{R}^n$ 与 $\boldsymbol{\mathcal{X}}_{i,j,:} \in \mathbb{R}^t$, 其中, $i = 1,2,\dots,m$、$j = 1,2,\dots,n$ 与 $k = 1,2,\dots,t$。与矩阵中的行向量、列向量类似，纤维是张量的基本组成部分。




(a) 向量 $\boldsymbol{\mathcal{X}}_{:,j,k} \in \mathbb{R}^m$ (b) 向量 $\boldsymbol{\mathcal{X}}_{i,:,k} \in \mathbb{R}^n$ (c) 向量 $\boldsymbol{\mathcal{X}}_{i,j,:} \in \mathbb{R}^t$

图 1.2: 三阶张量 $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{m\times n\times t}$ 自三个维度的纤维

### 三阶张量中的切片

对于任意三阶张量 $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{m\times n\times t}$，可用三个维度的切片 (slice) 书写该张量，其中，horizontal 切片共有 $m$ 个，分别为

$$\boldsymbol{\mathcal{X}}_{1,:,:}, \boldsymbol{\mathcal{X}}_{2,:,:}, \dots, \boldsymbol{\mathcal{X}}_{m,:,:} \in \mathbb{R}^{n\times t} \tag{1.7}$$

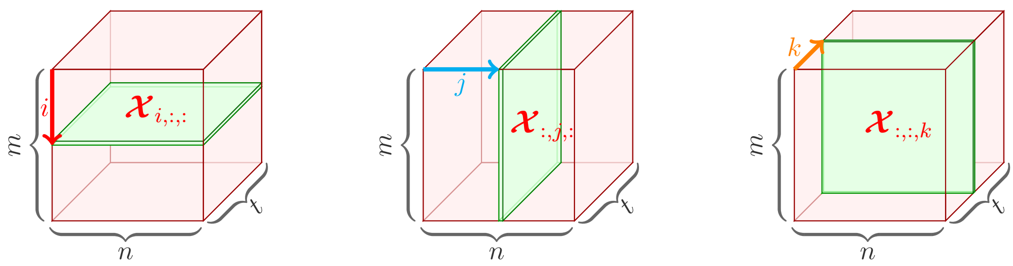
lateral 切片共有 $n$ 个，分别为

$$\boldsymbol{\mathcal{X}}_{:,1,:}, \boldsymbol{\mathcal{X}}_{:,2,:}, \dots, \boldsymbol{\mathcal{X}}_{:,n,:} \in \mathbb{R}^{m\times t} \tag{1.8}$$

frontal 切片共有 $t$ 个，分别为

$$\boldsymbol{\mathcal{X}}_{:,:,1}, \boldsymbol{\mathcal{X}}_{:,:,2}, \dots, \boldsymbol{\mathcal{X}}_{:,:,t} \in \mathbb{R}^{m\times n} \tag{1.9}$$

如图1.3所示，这些矩阵结构的切片是张量的基本组成部分。




(a) horizontal 切片 $\boldsymbol{\mathcal{X}}_{i,:,:}$ (b) lateral 切片 $\boldsymbol{\mathcal{X}}_{:,j,:}$ (c) frontal 切片 $\boldsymbol{\mathcal{X}}_{:,:,k} \in \mathbb{R}^t$

图 1.3: 三阶张量 $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{m\times n\times t}$ 自三个维度的切片

**例 1.** *给定张量* $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{2\times 2\times 2}$, *若其 frontal 切片为*

$$\boldsymbol{\mathcal{X}}_{:,:,1} = \begin{bmatrix} x_{111} & x_{121} \\ x_{211} & x_{221} \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \boldsymbol{\mathcal{X}}_{:,:,2} = \begin{bmatrix} x_{112} & x_{122} \\ x_{212} & x_{222} \end{bmatrix} = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \tag{1.10}$$

*试写出张量* $\boldsymbol{\mathcal{X}}$ *的 lateral 切片与 horizontal 切片。*

**解.** 张量 $\boldsymbol{\mathcal{X}}$ 的 *lateral* 切片为

$$\boldsymbol{\mathcal{X}}_{:,1,:}=\begin{bmatrix}x_{111} & x_{112}\\ x_{211} & x_{212}\end{bmatrix}=\begin{bmatrix}1 & 5\\ 3 & 7\end{bmatrix}\quad \boldsymbol{\mathcal{X}}_{:,2,:}=\begin{bmatrix}x_{121} & x_{122}\\ x_{221} & x_{222}\end{bmatrix}=\begin{bmatrix}2 & 6\\ 4 & 8\end{bmatrix} \tag{1.11}$$

张量 $\boldsymbol{\mathcal{X}}$ 的 *horizontal* 切片为

$$\boldsymbol{\mathcal{X}}_{1,:,:}=\begin{bmatrix}x_{111} & x_{112}\\ x_{121} & x_{122}\end{bmatrix}=\begin{bmatrix}1 & 5\\ 2 & 6\end{bmatrix}\quad \boldsymbol{\mathcal{X}}_{2,:,:}=\begin{bmatrix}x_{211} & x_{212}\\ x_{221} & x_{222}\end{bmatrix}=\begin{bmatrix}3 & 7\\ 4 & 8\end{bmatrix} \tag{1.12}$$

### 1.1.3 高阶张量矩阵化

给定任意张量 $\boldsymbol{\mathcal{X}}\in\mathbb{R}^{m_1\times m_2\times\cdots\times m_d}$，阶数为 $d$，若其自第 $k$ 维度展开得到的矩阵为 $\boldsymbol{X}_{(k)}$，大小为 $n_k\times\prod_{h\neq k}n_h$，则张量 $\boldsymbol{\mathcal{X}}$ 的第 $(i_1,i_2,\ldots,i_d)$ 个元素对应着矩阵 $\boldsymbol{X}_{(k)}$ 的第 $(i_k,j)$ 个元素：

$$j=1+\sum_{h=1,h\neq k}^{d}(i_h-1)J_h \tag{1.13}$$

其中，$J_h=\prod_{l=1,l\neq k}^{h-1}n_l$。张量展开的过程往往也被称为张量矩阵化 (matricization)。

为更容易理解张量展开的规则，不妨以大小为 $3\times4\times2$ 的张量 $\boldsymbol{\mathcal{X}}$ 为例 [Kolda and Bader, 2009]。

**例 2.** 给定张量 $\boldsymbol{\mathcal{X}}\in\mathbb{R}^{3\times4\times2}$，若其 *frontal* 切片为

$$\boldsymbol{\mathcal{X}}_{:,:,1}=\begin{bmatrix}1 & 4 & 7 & 10\\ 2 & 5 & 8 & 11\\ 3 & 6 & 9 & 12\end{bmatrix}\quad \boldsymbol{\mathcal{X}}_{:,:,2}=\begin{bmatrix}13 & 16 & 19 & 22\\ 14 & 17 & 20 & 23\\ 15 & 18 & 21 & 24\end{bmatrix} \tag{1.14}$$

试写出张量矩阵化的结果 $\boldsymbol{X}_{(1)}\in\mathbb{R}^{3\times8}$、$\boldsymbol{X}_{(2)}\in\mathbb{R}^{4\times6}$ 与 $\boldsymbol{X}_{(3)}\in\mathbb{R}^{2\times12}$。

**解.** 根据张量矩阵化规则，张量 $\boldsymbol{\mathcal{X}}$ 自第 1 维度展开得到的矩阵为 $\boldsymbol{X}_{(1)}$ 为

$$\boldsymbol{X}_{(1)}=\begin{bmatrix}\boldsymbol{\mathcal{X}}_{:,:,1} & \boldsymbol{\mathcal{X}}_{:,:,2}\end{bmatrix}=\begin{bmatrix}1 & 4 & 7 & 10 & 13 & 16 & 19 & 22\\ 2 & 5 & 8 & 11 & 14 & 17 & 20 & 23\\ 3 & 6 & 9 & 12 & 15 & 18 & 21 & 24\end{bmatrix} \tag{1.15}$$

张量 $\boldsymbol{\mathcal{X}}$ 自第 2 维度展开得到的矩阵为 $\boldsymbol{X}_{(2)}$ 为

$$\boldsymbol{X}_{(2)}=\begin{bmatrix}\boldsymbol{\mathcal{X}}_{:,:,1}^{\top} & \boldsymbol{\mathcal{X}}_{:,:,2}^{\top}\end{bmatrix}=\begin{bmatrix}1 & 2 & 3 & 13 & 14 & 15\\ 4 & 5 & 6 & 16 & 17 & 18\\ 7 & 8 & 9 & 19 & 20 & 21\\ 10 & 11 & 12 & 22 & 23 & 24\end{bmatrix} \tag{1.16}$$

张量 $\boldsymbol{\mathcal{X}}$ 自第 3 维度展开得到的矩阵为 $\boldsymbol{X}_{(3)}$ 为

$$\boldsymbol{X}_{(2)}=\begin{bmatrix}\boldsymbol{\mathcal{X}}_{:,1,:}^{\top} & \boldsymbol{\mathcal{X}}_{:,2,:}^{\top}\end{bmatrix}=\begin{bmatrix}1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12\\ 13 & 14 & 15 & 16 & 17 & 18 & 19 & 20 & 21 & 22 & 23 & 24\end{bmatrix} \tag{1.17}$$

对于任意三阶张量 $\boldsymbol{\mathcal{X}}\in\mathbb{R}^{m\times n\times t}$，其自三个维度展开得到的矩阵分别为

$$\begin{cases}\boldsymbol{X}_{(1)}=\begin{bmatrix}\boldsymbol{\mathcal{X}}_{:,:,1} & \boldsymbol{\mathcal{X}}_{:,:,2} & \cdots & \boldsymbol{\mathcal{X}}_{:,:,t}\end{bmatrix}\in\mathbb{R}^{m\times(nt)}\\ \boldsymbol{X}_{(2)}=\begin{bmatrix}\boldsymbol{\mathcal{X}}_{:,:,1}^{\top} & \boldsymbol{\mathcal{X}}_{:,:,2}^{\top} & \cdots & \boldsymbol{\mathcal{X}}_{:,:,t}^{\top}\end{bmatrix}\in\mathbb{R}^{n\times(mt)}\\ \boldsymbol{X}_{(3)}=\begin{bmatrix}\boldsymbol{\mathcal{X}}_{:,1,:} & \boldsymbol{\mathcal{X}}_{:,2,:} & \cdots & \boldsymbol{\mathcal{X}}_{:,t,:}\end{bmatrix}\in\mathbb{R}^{t\times(mn)}\end{cases} \tag{1.18}$$

### 1.1.4 高阶张量向量化

给定任意张量 $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{m_1 \times m_2 \times \cdots \times m_d}$, 阶数为 $d$, 若其以第一个维度展开得到的矩阵为 $\boldsymbol{X}_{(1)}$, 则张量向量化可写作如下形式:

$$\operatorname{vec}(\boldsymbol{\mathcal{X}}) = \operatorname{vec}(\boldsymbol{X}_{(1)}) \tag{1.19}$$

在张量 $\boldsymbol{\mathcal{X}}$ 中，第 $(i_1, i_2, \ldots, i_d)$ 个元素通过张量向量化之后，该元素在向量中的位置为

$$\Big(\sum_{k=1}^{d-1} m_k\Big) \cdot i_d + \Big(\sum_{k=1}^{d-2} m_k\Big) \cdot i_{d-1} + \cdots + m_1 \cdot i_2 + i_1 \tag{1.20}$$

**例 3.** 给定张量 $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{2\times 2\times 2}$，若其 *frontal* 切片为

$$\boldsymbol{\mathcal{X}}_{:,:,1} = \begin{bmatrix} x_{111} & x_{121} \\ x_{211} & x_{221} \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \boldsymbol{\mathcal{X}}_{:,:,2} = \begin{bmatrix} x_{112} & x_{122} \\ x_{212} & x_{222} \end{bmatrix} = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \tag{1.21}$$

试写出张量向量化的结果 $\operatorname{vec}(\boldsymbol{\mathcal{X}})$。

**解.** 根据张量向量化规则，有

$$\begin{aligned} \operatorname{vec}(\boldsymbol{\mathcal{X}}) &= \operatorname{vec}(\boldsymbol{X}_{(1)}) \\ &= \operatorname{vec}([\boldsymbol{\mathcal{X}}_{:,:,1} \quad \boldsymbol{\mathcal{X}}_{:,:,2}]) \\ &= (1,3,2,4,5,7,6,8)^\top \end{aligned} \tag{1.22}$$

## 1.2 时空交通数据类型

# 第二章　张量分解基础

本章将要介绍的内容包括:

- Kronecker 积的定义及性质
- 向量外积的定义及性质
- 经典的 CP 张量分解形式

## 2.1　Kronecker 积定义

### 2.1.1　基本定义

Kronecker 积是以德国数学家 Leopold Kronecker 的名字命名的运算规则，已广泛应用于各类矩阵计算以及张量计算算法中。从定义出发，给定任意矩阵 $\boldsymbol{X}\in\mathbb{R}^{m\times n}$ 与 $\boldsymbol{Y}\in\mathbb{R}^{p\times q}$，则两者之间的 Kronecker 积为

$$\boldsymbol{X}\otimes\boldsymbol{Y}=\begin{bmatrix}x_{11}\boldsymbol{Y} & x_{12}\boldsymbol{Y} & \cdots & x_{1n}\boldsymbol{Y}\\ x_{21}\boldsymbol{Y} & x_{22}\boldsymbol{Y} & \cdots & x_{2n}\boldsymbol{Y}\\ \vdots & \vdots & \ddots & \vdots\\ x_{m1}\boldsymbol{Y} & x_{m2}\boldsymbol{Y} & \cdots & x_{mn}\boldsymbol{Y}\end{bmatrix}\in\mathbb{R}^{(mp)\times(nq)} \tag{2.1}$$

其中，符号 $\otimes$ 表示 Kronecker 积。这里的 Kronecker 积得到的矩阵大小为 $(mp)\times(nq)$，在写法上符合线性代数中对分块矩阵 (block matrix) 的定义，其中，分块矩阵的子矩阵是由矩阵 $\boldsymbol{X}$ 的每个元素与矩阵 $\boldsymbol{Y}$ 相乘得到。

矩阵 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 之间的 Kronecker 积存在前后顺序，根据 Kronecker 积的定义，可得到矩阵 $\boldsymbol{Y}$ 与 $\boldsymbol{X}$ 之间的 Kronecker 积为

$$\boldsymbol{Y}\otimes\boldsymbol{X}=\begin{bmatrix}y_{11}\boldsymbol{X} & y_{12}\boldsymbol{X} & \cdots & y_{1q}\boldsymbol{X}\\ y_{21}\boldsymbol{X} & y_{22}\boldsymbol{X} & \cdots & y_{2q}\boldsymbol{X}\\ \vdots & \vdots & \ddots & \vdots\\ y_{p1}\boldsymbol{X} & y_{p2}\boldsymbol{X} & \cdots & y_{pq}\boldsymbol{X}\end{bmatrix}\in\mathbb{R}^{(mp)\times(nq)} \tag{2.2}$$

尽管矩阵 $\boldsymbol{X}\otimes\boldsymbol{Y}$ 与矩阵 $\boldsymbol{Y}\otimes\boldsymbol{X}$ 大小一致，但两者并不相等，因此，Kronecker 积不存在交换律。

**例 4.** *给定矩阵* $\boldsymbol{X}=\begin{bmatrix}1&2\\3&4\end{bmatrix}$ *与* $\boldsymbol{Y}=\begin{bmatrix}5&6&7\\8&9&10\end{bmatrix}$，*试写出两者之间的 Kronecker 积* $\boldsymbol{X}\otimes\boldsymbol{Y}$ *与* $\boldsymbol{Y}\otimes\boldsymbol{X}$。

**解.** 根据 *Kronecker* 积定义，有

$$\boldsymbol{X}\otimes\boldsymbol{Y}=\begin{bmatrix}1\times\begin{bmatrix}5&6&7\\8&9&10\end{bmatrix}&2\times\begin{bmatrix}5&6&7\\8&9&10\end{bmatrix}\\3\times\begin{bmatrix}5&6&7\\8&9&10\end{bmatrix}&4\times\begin{bmatrix}5&6&7\\8&9&10\end{bmatrix}\end{bmatrix}=\begin{bmatrix}5&6&7&10&12&14\\8&9&10&16&18&20\\15&18&21&20&24&28\\24&27&30&32&36&40\end{bmatrix}\tag{2.3}$$

$$\boldsymbol{Y}\otimes\boldsymbol{X}=\begin{bmatrix}5\times\begin{bmatrix}1&2\\3&4\end{bmatrix}&6\times\begin{bmatrix}1&2\\3&4\end{bmatrix}&7\times\begin{bmatrix}1&2\\3&4\end{bmatrix}\\8\times\begin{bmatrix}1&2\\3&4\end{bmatrix}&9\times\begin{bmatrix}1&2\\3&4\end{bmatrix}&10\times\begin{bmatrix}1&2\\3&4\end{bmatrix}\end{bmatrix}=\begin{bmatrix}5&10&6&12&7&14\\15&20&18&24&21&28\\8&16&9&18&10&20\\24&32&27&36&30&40\end{bmatrix}\tag{2.4}$$

**例 5.** 给定矩阵 $\boldsymbol{X}=\begin{bmatrix}1&2\\3&4\end{bmatrix}$ 与 $\boldsymbol{Y}=\begin{bmatrix}5&6&7\\8&9&10\end{bmatrix}$，试问等式 $(\boldsymbol{X}\otimes\boldsymbol{Y})^\top=\boldsymbol{X}^\top\otimes\boldsymbol{Y}^\top$ 是否成立。

**解.** 根据 *Kronecker* 积定义，有

$$\boldsymbol{X}^\top\otimes\boldsymbol{Y}^\top=\begin{bmatrix}1\times\begin{bmatrix}5&8\\6&9\\7&10\end{bmatrix}&3\times\begin{bmatrix}5&8\\6&9\\7&10\end{bmatrix}\\2\times\begin{bmatrix}5&8\\6&9\\7&10\end{bmatrix}&4\times\begin{bmatrix}5&8\\6&9\\7&10\end{bmatrix}\end{bmatrix}=\begin{bmatrix}5&8&15&24\\6&9&18&27\\7&10&21&30\\10&16&20&32\\12&18&24&36\\14&20&28&40\end{bmatrix}\tag{2.5}$$

在这里，等式 $(\boldsymbol{X}\otimes\boldsymbol{Y})^\top=\boldsymbol{X}^\top\otimes\boldsymbol{Y}^\top$ 是成立的。

**例 6.** 给定向量 $\boldsymbol{x}=(1,2)^\top$ 与 $\boldsymbol{y}=(3,4)^\top$，试写出 $\boldsymbol{x}\otimes\boldsymbol{y}$ 与 $\boldsymbol{x}\otimes\boldsymbol{y}^\top$。

**解.** 根据 *Kronecker* 积定义，有

$$\boldsymbol{x}\otimes\boldsymbol{y}=\begin{bmatrix}1\times\begin{bmatrix}3\\4\end{bmatrix}\\2\times\begin{bmatrix}3\\4\end{bmatrix}\end{bmatrix}=\begin{bmatrix}3\\4\\6\\8\end{bmatrix}\tag{2.6}$$

$$\boldsymbol{x}\otimes\boldsymbol{y}^\top=\begin{bmatrix}1\times\begin{bmatrix}3&4\end{bmatrix}\\2\times\begin{bmatrix}3&4\end{bmatrix}\end{bmatrix}=\begin{bmatrix}3&4\\6&8\end{bmatrix}\tag{2.7}$$

在这里，$\boldsymbol{x}\otimes\boldsymbol{y}^\top=\boldsymbol{x}\boldsymbol{y}^\top$，即向量外积。

**例 7** (向量自回归)**.** 对于多元时间序列，向量自回归可写作如下形式（参见例*3*）：

$$\boldsymbol{X}\boldsymbol{\Psi}_0^\top=\sum_{k=1}^{d}\boldsymbol{A}_k\boldsymbol{X}\boldsymbol{\Psi}_k^\top+\boldsymbol{E}\tag{2.8}$$

若令

$$\begin{aligned}\boldsymbol{A}&=\begin{bmatrix}\boldsymbol{A}_1&\boldsymbol{A}_2&\cdots&\boldsymbol{A}_d\end{bmatrix}\in\mathbb{R}^{N\times(dN)}\\\boldsymbol{\Psi}&=\begin{bmatrix}\boldsymbol{\Psi}_1&\boldsymbol{\Psi}_2&\cdots&\boldsymbol{\Psi}_d\end{bmatrix}\in\mathbb{R}^{(T-d)\times(dT)}\end{aligned}\tag{2.9}$$

则向量自回归可进一步写作如下形式：

$$\boldsymbol{X}\boldsymbol{\Psi}_0^\top=\boldsymbol{A}(\boldsymbol{I}_d\otimes\boldsymbol{X})\boldsymbol{\Psi}^\top+\boldsymbol{E}\tag{2.10}$$

### 2.1.2 Khatri-Rao 积

以 Kronecker 积为基础，可定义另一种十分重要的运算规则，即 Khatri-Rao 积。给定任意矩阵

$$\boldsymbol{X} = \begin{bmatrix} | & | & & | \\ \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_d \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{m\times d} \quad \boldsymbol{Y} = \begin{bmatrix} | & | & & | \\ \boldsymbol{y}_1 & \boldsymbol{y}_2 & \cdots & \boldsymbol{y}_d \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{n\times d} \tag{2.11}$$

若两个矩阵列数相同，则两者之间的 Khatri-Rao 积为

$$\boldsymbol{X} \odot \boldsymbol{Y} = \begin{bmatrix} | & | & & | \\ \boldsymbol{x}_1 \otimes \boldsymbol{y}_1 & \boldsymbol{x}_2 \otimes \boldsymbol{y}_2 & \cdots & \boldsymbol{x}_d \otimes \boldsymbol{y}_d \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{(mn)\times d} \tag{2.12}$$

其中，列向量是由 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 的列向量进行 Kronecker 积运算得到的。

**例 8.** 给定矩阵 $\boldsymbol{X} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ 与 $\boldsymbol{Y} = \begin{bmatrix} 5 & 6 \\ 7 & 8 \\ 9 & 10 \end{bmatrix}$，试写出 $\boldsymbol{X} \odot \boldsymbol{Y}$。

**解.** 根据 *Khatri-Rao* 积定义，有

$$\boldsymbol{X} \odot \boldsymbol{Y} = \begin{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} \otimes \begin{bmatrix} 5 \\ 7 \\ 9 \end{bmatrix} & \begin{bmatrix} 2 \\ 4 \end{bmatrix} \otimes \begin{bmatrix} 6 \\ 8 \\ 10 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 5 & 12 \\ 7 & 16 \\ 9 & 20 \\ 15 & 24 \\ 21 & 32 \\ 27 & 40 \end{bmatrix} \tag{2.13}$$

## 2.2 Kronecker 积基本性质

### 2.2.1 结合律与分配律

在小学数学中，我们学习了加减乘除的运算规则。以乘法为例，不妨重温一下烙印在我们脑海中的基本概念：

- 乘法结合律：$x \times y \times z = x \times (y \times z)$
- 乘法分配律：$x \times z + y \times z = (x + y) \times z$

由于 Kronecker 积本质上也是元素间相乘，所以同样存在结合律与分配律。对于任意矩阵 $\boldsymbol{X}$、$\boldsymbol{Y}$ 与 $\boldsymbol{Z}$，结合律可归纳为

$$\boldsymbol{X} \otimes \boldsymbol{Y} \otimes \boldsymbol{Z} = \boldsymbol{X} \otimes (\boldsymbol{Y} \otimes \boldsymbol{Z}) \tag{2.14}$$

分配律可归纳为

$$\boldsymbol{X} \otimes \boldsymbol{Z} + \boldsymbol{Y} \otimes \boldsymbol{Z} = (\boldsymbol{X} + \boldsymbol{Y}) \otimes \boldsymbol{Z} \tag{2.15}$$

**例 9.** 给定矩阵 $\boldsymbol{X} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$、$\boldsymbol{Y} = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$ 与 $\boldsymbol{Z} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$，试写出 $\boldsymbol{X} \otimes \boldsymbol{Y} \otimes \boldsymbol{Z}$ 与 $\boldsymbol{X} \otimes (\boldsymbol{Y} \otimes \boldsymbol{Z})$。

**解.** 根据 *Kronecker* 积定义，有

$$\boldsymbol{X}\otimes\boldsymbol{Y}=\begin{bmatrix}5&6&10&12\\7&8&14&16\\15&18&20&24\\21&24&28&32\end{bmatrix}\tag{2.16}$$

$$\boldsymbol{Y}\otimes\boldsymbol{Z}=\begin{bmatrix}5&5&6&6\\5&5&6&6\\7&7&8&8\\7&7&8&8\end{bmatrix}\tag{2.17}$$

从而，可得到

$$\boldsymbol{X}\otimes\boldsymbol{Y}\otimes\boldsymbol{Z}=\begin{bmatrix}5&5&6&6&10&10&12&12\\5&5&6&6&10&10&12&12\\7&7&8&8&14&14&16&16\\7&7&8&8&14&14&16&16\\15&15&18&18&20&20&24&24\\15&15&18&18&20&20&24&24\\21&21&24&24&28&28&32&32\\21&21&24&24&28&28&32&32\end{bmatrix}=\boldsymbol{X}\otimes(\boldsymbol{Y}\otimes\boldsymbol{Z})\tag{2.18}$$

**例 10.** 给定 $\boldsymbol{X}=\begin{bmatrix}1&2\\3&4\end{bmatrix}$、$\boldsymbol{Y}=\begin{bmatrix}5&6\\7&8\end{bmatrix}$ 与 $\boldsymbol{Z}=\begin{bmatrix}1&1\\1&1\end{bmatrix}$，试写出 $\boldsymbol{X}\otimes\boldsymbol{Z}+\boldsymbol{Y}\otimes\boldsymbol{Z}$ 与 $(\boldsymbol{X}+\boldsymbol{Y})\otimes\boldsymbol{Z}$。

**解.** 根据 *Kronecker* 积定义，有

$$\boldsymbol{X}\otimes\boldsymbol{Z}+\boldsymbol{Y}\otimes\boldsymbol{Z}=\begin{bmatrix}1&1&2&2\\1&1&2&2\\3&3&4&4\\3&3&4&4\end{bmatrix}+\begin{bmatrix}5&5&6&6\\5&5&6&6\\7&7&8&8\\7&7&8&8\end{bmatrix}=\begin{bmatrix}6&6&8&8\\6&6&8&8\\10&10&12&12\\10&10&12&12\end{bmatrix}\tag{2.19}$$

$$(\boldsymbol{X}+\boldsymbol{Y})\otimes\boldsymbol{Z}=\begin{bmatrix}6&8\\10&12\end{bmatrix}\otimes\begin{bmatrix}1&1\\1&1\end{bmatrix}=\begin{bmatrix}6&6&8&8\\6&6&8&8\\10&10&12&12\\10&10&12&12\end{bmatrix}\tag{2.20}$$

### 2.2.2 矩阵相乘

对于任意矩阵 $\boldsymbol{X}\in\mathbb{R}^{m\times n}$、$\boldsymbol{Y}\in\mathbb{R}^{s\times t}$、$\boldsymbol{U}\in\mathbb{R}^{n\times p}$ 与 $\boldsymbol{V}\in\mathbb{R}^{t\times q}$，则矩阵 $\boldsymbol{X}\otimes\boldsymbol{Y}\in\mathbb{R}^{(ms)\times(nt)}$ 的列数 $nt$ 与矩阵 $\boldsymbol{U}\otimes\boldsymbol{V}\in\mathbb{R}^{(nt)\times(pq)}$ 的行数 $nt$ 一致，可进行矩阵相乘，两者相乘得到的矩

阵满足：

$$\begin{aligned}(\boldsymbol{X}\otimes\boldsymbol{Y})(\boldsymbol{U}\otimes\boldsymbol{V}) &= \begin{bmatrix} x_{11}\boldsymbol{Y} & \cdots & x_{1n}\boldsymbol{Y} \\ \vdots & \ddots & \vdots \\ x_{m1}\boldsymbol{Y} & \cdots & x_{mn}\boldsymbol{Y}\end{bmatrix}\begin{bmatrix} u_{11}\boldsymbol{V} & \cdots & u_{1p}\boldsymbol{V} \\ \vdots & \ddots & \vdots \\ u_{n1}\boldsymbol{V} & \cdots & u_{np}\boldsymbol{V}\end{bmatrix} \\ &= \begin{bmatrix} \sum_{k=1}^{n} x_{1k}u_{k1}\boldsymbol{Y}\boldsymbol{V} & \cdots & \sum_{k=1}^{n} x_{1k}u_{kp}\boldsymbol{Y}\boldsymbol{V} \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^{n} x_{mk}u_{k1}\boldsymbol{Y}\boldsymbol{V} & \cdots & \sum_{k=1}^{n} x_{mk}u_{kp}\boldsymbol{Y}\boldsymbol{V}\end{bmatrix} \\ &= \begin{bmatrix} \sum_{k=1}^{n} x_{1k}u_{k1} & \cdots & \sum_{k=1}^{n} x_{1k}u_{kp} \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^{n} x_{mk}u_{k1} & \cdots & \sum_{k=1}^{n} x_{mk}u_{kp}\end{bmatrix}\otimes(\boldsymbol{Y}\boldsymbol{V}) \\ &=(\boldsymbol{X}\boldsymbol{U})\otimes(\boldsymbol{Y}\boldsymbol{V})\in\mathbb{R}^{(ms)\times(pq)}\end{aligned} \tag{2.21}$$

**例 11** (矩阵的奇异值分解)**.** 给定任意矩阵 $\boldsymbol{X}\in\mathbb{R}^{m\times n}$ 与 $\boldsymbol{Y}\in\mathbb{R}^{p\times q}$，若奇异值分解分别为

$$\boldsymbol{X}=\boldsymbol{W}\boldsymbol{S}\boldsymbol{Q}^\top \quad \boldsymbol{Y}=\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top \tag{2.22}$$

试证明矩阵 $\boldsymbol{X}\otimes\boldsymbol{Y}$ 的奇异值分解可由矩阵 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 的奇异值分解计算得到，即

$$\boldsymbol{X}\otimes\boldsymbol{Y}=(\boldsymbol{W}\otimes\boldsymbol{U})(\boldsymbol{S}\otimes\boldsymbol{D})(\boldsymbol{Q}\otimes\boldsymbol{V})^\top \tag{2.23}$$

**解.** 根据 *Kronecker* 积性质，有

$$\begin{aligned}\boldsymbol{X}\otimes\boldsymbol{Y} &=(\boldsymbol{W}\boldsymbol{S}\boldsymbol{Q}^\top)\otimes(\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top) \\ &=(\boldsymbol{W}\otimes\boldsymbol{U})((\boldsymbol{S}\boldsymbol{Q}^\top)\otimes(\boldsymbol{D}\boldsymbol{V}^\top)) \\ &=(\boldsymbol{W}\otimes\boldsymbol{U})(\boldsymbol{S}\otimes\boldsymbol{D})(\boldsymbol{Q}^\top\otimes\boldsymbol{V}^\top) \\ &=(\boldsymbol{W}\otimes\boldsymbol{U})(\boldsymbol{S}\otimes\boldsymbol{D})(\boldsymbol{Q}\otimes\boldsymbol{V})^\top\end{aligned} \tag{2.24}$$

### 2.2.3 求逆矩阵

对于任意可逆矩阵 $\boldsymbol{X}\in\mathbb{R}^{m\times m}$ 与 $\boldsymbol{Y}\in\mathbb{R}^{n\times n}$，由于

$$(\boldsymbol{X}\otimes\boldsymbol{Y})\left(\boldsymbol{X}^{-1}\otimes\boldsymbol{Y}^{-1}\right)=\left(\boldsymbol{X}\boldsymbol{X}^{-1}\right)\otimes\left(\boldsymbol{Y}\boldsymbol{Y}^{-1}\right)=\boldsymbol{I}_m\otimes\boldsymbol{I}_n=\boldsymbol{I}_{mn} \tag{2.25}$$

故有

$$(\boldsymbol{X}\otimes\boldsymbol{Y})^{-1}=\boldsymbol{X}^{-1}\otimes\boldsymbol{Y}^{-1} \tag{2.26}$$

恒成立。这意味着：若计算 $\boldsymbol{X}\otimes\boldsymbol{Y}$ 的逆矩阵，可先对 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 分别求逆矩阵，再对得到的逆矩阵进行 Kronecker 积运算。

**例 12.** 给定矩阵 $\boldsymbol{X}=\begin{bmatrix}1 & 2 \\ 3 & 4\end{bmatrix}$ 与 $\boldsymbol{Y}=\begin{bmatrix}5 & 6 \\ 7 & 8\end{bmatrix}$，试写出 $(\boldsymbol{X}\otimes\boldsymbol{Y})^{-1}$ 与 $\boldsymbol{X}^{-1}\otimes\boldsymbol{Y}^{-1}$。

**解.** 根据 *Kronecker* 积定义，有

$$\boldsymbol{X}\otimes\boldsymbol{Y}=\begin{bmatrix}5 & 6 & 10 & 12 \\ 7 & 8 & 14 & 16 \\ 15 & 18 & 20 & 24 \\ 21 & 24 & 28 & 32\end{bmatrix} \tag{2.27}$$

对该矩阵求逆矩阵，得到

$$(\boldsymbol{X}\otimes\boldsymbol{Y})^{-1}=\begin{bmatrix}8&-6&-4&3\\-7&5&3.5&-2.5\\-6&4.5&2&-1.5\\5.25&-3.75&-1.75&1.25\end{bmatrix}\tag{2.28}$$

对矩阵 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 分别求逆矩阵：

$$\boldsymbol{X}^{-1}=\begin{bmatrix}-2&1\\1.5&-0.5\end{bmatrix}\quad\boldsymbol{Y}^{-1}=\begin{bmatrix}-4&3\\3.5&-2.5\end{bmatrix}\tag{2.29}$$

再对得到的逆矩阵进行 *Kronecker* 积运算，有

$$\boldsymbol{X}^{-1}\otimes\boldsymbol{Y}^{-1}=\begin{bmatrix}8&-6&-4&3\\-7&5&3.5&-2.5\\-6&4.5&2&-1.5\\5.25&-3.75&-1.75&1.25\end{bmatrix}\tag{2.30}$$

对于任意矩阵 $\boldsymbol{X}\in\mathbb{R}^{m\times m}$ 与 $\boldsymbol{Y}\in\mathbb{R}^{p\times q}$，由上述 Kronecker 积性质同样可得到如下性质：

$$(\boldsymbol{X}\otimes\boldsymbol{Y})^{\dagger}=\boldsymbol{X}^{\dagger}\otimes\boldsymbol{Y}^{\dagger}\tag{2.31}$$

其中，$\cdot^{\dagger}$ 表示伪逆 (Moore-Penrose pseudoinverse)。

### 2.2.4 向量化

对于任意矩阵 $\boldsymbol{A}\in\mathbb{R}^{m\times n}$、$\boldsymbol{X}\in\mathbb{R}^{n\times p}$ 与 $\boldsymbol{B}\in\mathbb{R}^{p\times q}$，三者相乘满足：

$$\operatorname{vec}(\boldsymbol{AXB})=(\boldsymbol{B}^{\top}\otimes\boldsymbol{A})\operatorname{vec}(\boldsymbol{X})\tag{2.32}$$

由此，也可得到

$$\begin{cases}\operatorname{vec}(\boldsymbol{AX})=(\boldsymbol{I}_p\otimes\boldsymbol{A})\operatorname{vec}(\boldsymbol{X})\\\operatorname{vec}(\boldsymbol{XB})=(\boldsymbol{B}^{\top}\otimes\boldsymbol{I}_n)\operatorname{vec}(\boldsymbol{X})\end{cases}\tag{2.33}$$

**例 13.** 试证明公式(2.32)。

**解.**

$$\begin{aligned}\operatorname{vec}(\boldsymbol{AXB})=&\boldsymbol{Ax}_1b_{11}+\boldsymbol{Ax}_2b_{21}+\cdots+\boldsymbol{Ax}_pb_{p1}\\&+\boldsymbol{Ax}_1b_{12}+\boldsymbol{Ax}_2b_{22}+\cdots+\boldsymbol{Ax}_pb_{p2}\\&+\cdots+\boldsymbol{Ax}_1b_{1q}+\boldsymbol{Ax}_2b_{2q}+\cdots+\boldsymbol{Ax}_pb_{pq}\\=&\begin{bmatrix}\boldsymbol{A}b_{11}&\boldsymbol{A}b_{21}&\cdots&\boldsymbol{A}b_{p1}\\\boldsymbol{A}b_{12}&\boldsymbol{A}b_{22}&\cdots&\boldsymbol{A}b_{p2}\\\vdots&\vdots&\ddots&\vdots\\\boldsymbol{A}b_{1q}&\boldsymbol{A}b_{2q}&\cdots&\boldsymbol{A}b_{pq}\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_1\\\boldsymbol{x}_2\\\vdots\\\boldsymbol{x}_p\end{bmatrix}\\=&(\boldsymbol{B}^{\top}\otimes\boldsymbol{A})\operatorname{vec}(\boldsymbol{X})\end{aligned}\tag{2.34}$$

其中，$\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_p\in\mathbb{R}^n$ 表示矩阵 $\boldsymbol{X}$ 的列向量。

**例 14.** 对于任意矩阵 $\boldsymbol{A}\in\mathbb{R}^{n\times n}$、$\boldsymbol{x}\in\mathbb{R}^n$ 与 $\boldsymbol{B}\in\mathbb{R}^{n\times n}$，试证明三者相乘满足*:*

$$\operatorname{vec}(\boldsymbol{A}\operatorname{diag}(\boldsymbol{x})\boldsymbol{B})=(\boldsymbol{B}^{\top}\odot\boldsymbol{A})\boldsymbol{x}\tag{2.35}$$

**解.** 根据 *Kronecker* 积与 *Khatri-Rao* 积性质，有

$$\begin{aligned}\operatorname{vec}(\boldsymbol{A}\operatorname{diag}(\boldsymbol{x})\boldsymbol{B}) &=(\boldsymbol{B}^\top\otimes\boldsymbol{A})\operatorname{vec}(\operatorname{diag}(\boldsymbol{x}))\\ &=(\boldsymbol{B}^\top\odot\boldsymbol{A})\boldsymbol{x}\end{aligned}\tag{2.36}$$

**例 15.** *Sylvester* 方程是一种著名的矩阵方程，由英国数学家 *James Joseph Sylvester* 于 *1884* 年提出。时至今日，*Sylvester* 方程已在控制理论中具有极为广泛的应用。具体而言，已知矩阵 $\boldsymbol{A}\in\mathbb{R}^{m\times m}$、$\boldsymbol{B}\in\mathbb{R}^{n\times n}$ 与 $\boldsymbol{C}\in\mathbb{R}^{m\times n}$，则 *Sylvester* 方程的一般形式为

$$\boldsymbol{A}\boldsymbol{X}+\boldsymbol{X}\boldsymbol{B}=\boldsymbol{C}\tag{2.37}$$

其中，$\boldsymbol{X}\in\mathbb{R}^{m\times n}$ 为待定参数。试根据 *Kronecker* 积性质写出 *Sylvester* 方程的解析解。

**解.** 首先将 *Sylvester* 方程写成

$$\boldsymbol{A}\boldsymbol{X}\boldsymbol{I}_n+\boldsymbol{I}_m\boldsymbol{X}\boldsymbol{B}=\boldsymbol{C}\tag{2.38}$$

根据 *Kronecker* 积性质，*Sylvester* 方程可写成如下形式：

$$\left(\boldsymbol{I}_n\otimes\boldsymbol{A}+\boldsymbol{B}^\top\otimes\boldsymbol{I}_m\right)\operatorname{vec}(\boldsymbol{X})=\operatorname{vec}(\boldsymbol{C})\tag{2.39}$$

因此，*Sylvester* 方程的解析解[1]为

$$\operatorname{vec}(\boldsymbol{X})=\left(\boldsymbol{I}_n\otimes\boldsymbol{A}+\boldsymbol{B}^\top\otimes\boldsymbol{I}_m\right)^{-1}\operatorname{vec}(\boldsymbol{C})\tag{2.40}$$

尽管该解析解形式简洁，但复杂度却很高。在实际问题中，往往需要借助更为高效的数值计算方法（如 *Bartels-Stewart* 算法）对 *Sylvester* 方程进行求解。

## 2.3 Kronecker 积特殊性质

### 2.3.1 矩阵的迹

在线性代数中，矩阵的迹 (trace) 表示方阵对角线元素之和，数学符号为 $\operatorname{tr}(\cdot)$。对于任意矩阵 $\boldsymbol{X}\in\mathbb{R}^{m\times m}$ 与 $\boldsymbol{Y}\in\mathbb{R}^{n\times n}$，矩阵 $\boldsymbol{X}\otimes\boldsymbol{Y}$ 的迹等于矩阵 $\boldsymbol{X}$ 的迹乘以矩阵 $\boldsymbol{Y}$ 的迹，即

$$\operatorname{tr}(\boldsymbol{X}\otimes\boldsymbol{Y})=\operatorname{tr}(\boldsymbol{X})\cdot\operatorname{tr}(\boldsymbol{Y})\tag{2.41}$$

恒成立。

**例 16.** 给定矩阵 $\boldsymbol{X}=\begin{bmatrix}1&2\\3&4\end{bmatrix}$ 与 $\boldsymbol{Y}=\begin{bmatrix}5&6\\7&8\end{bmatrix}$，试写出 $\operatorname{tr}(\boldsymbol{X})$、$\operatorname{tr}(\boldsymbol{Y})$ 与 $\operatorname{tr}(\boldsymbol{X}\otimes\boldsymbol{Y})$。

**解.** 根据定义，矩阵 $\boldsymbol{X}$ 的迹与矩阵 $\boldsymbol{Y}$ 的迹分别为

$$\operatorname{tr}(\boldsymbol{X})=1+4=5\quad\operatorname{tr}(\boldsymbol{Y})=5+8=13\tag{2.42}$$

由于

$$\boldsymbol{X}\otimes\boldsymbol{Y}=\begin{bmatrix}5&6&10&12\\7&8&14&16\\15&18&20&24\\21&24&28&32\end{bmatrix}\tag{2.43}$$

故 $\operatorname{tr}(\boldsymbol{X}\otimes\boldsymbol{Y})=5+8+20+32=65$。

[1]有时候，可定义 Kronecker 和 (Kronecker sum，数学符号通常为 $\oplus$) 令 $\boldsymbol{A}\oplus\boldsymbol{B}^\top=\boldsymbol{I}_n\otimes\boldsymbol{A}+\boldsymbol{B}^\top\otimes\boldsymbol{I}_m$，将该解析解简记为 $\operatorname{vec}(\boldsymbol{X})=\left(\boldsymbol{A}\oplus\boldsymbol{B}^\top\right)^{-1}\operatorname{vec}(\boldsymbol{C})$。

在矩阵计算中，矩阵的迹有两条重要性质，给定任意矩阵 $\boldsymbol{X} \in \mathbb{R}^{m \times n}$ 与 $\boldsymbol{Y} \in \mathbb{R}^{n \times m}$，满足

$$\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}) = \operatorname{tr}(\boldsymbol{B}\boldsymbol{A}) \tag{2.44}$$

及

$$\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}) = \operatorname{vec}(\boldsymbol{A}^\top)^\top \operatorname{vec}(\boldsymbol{B}) \tag{2.45}$$

**例 17.** 给定矩阵 $\boldsymbol{A} \in \mathbb{R}^{m \times n}$、$\boldsymbol{B} \in \mathbb{R}^{n \times p}$、$\boldsymbol{C} \in \mathbb{R}^{p \times q}$ 与 $\boldsymbol{D} \in \mathbb{R}^{q \times m}$，试证明

$$\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C}\boldsymbol{D}) = \operatorname{vec}(\boldsymbol{B})^\top (\boldsymbol{C} \otimes \boldsymbol{A}) \operatorname{vec}(\boldsymbol{D}^\top) \tag{2.46}$$

**解.** 根据矩阵的迹与 *Kronecker* 积性质，有

$$\begin{aligned} \operatorname{tr}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C}\boldsymbol{D}) &= \operatorname{tr}(\boldsymbol{D}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C})) \\ &= \operatorname{vec}(\boldsymbol{D}^\top)^\top \operatorname{vec}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C}) \\ &= \operatorname{vec}(\boldsymbol{D}^\top)^\top (\boldsymbol{C}^\top \otimes \boldsymbol{A}) \operatorname{vec}(\boldsymbol{B}) \\ &= \operatorname{vec}(\boldsymbol{B})^\top (\boldsymbol{C} \otimes \boldsymbol{A}) \operatorname{vec}(\boldsymbol{D}^\top) \end{aligned} \tag{2.47}$$

### 2.3.2　矩阵的 Frobenius 范数

从定义出发，矩阵的 Frobenius 范数表示矩阵元素的平方和开根号，一般用 $\|\cdot\|_F$ 表示。对于任意矩阵 $\boldsymbol{X} \in \mathbb{R}^{m \times n}$，其 Frobenius 范数为

$$\|\boldsymbol{X}\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}^2} \tag{2.48}$$

据此定义，给定任意矩阵 $\boldsymbol{X} \in \mathbb{R}^{m \times n}$ 与 $\boldsymbol{Y} \in \mathbb{R}^{p \times q}$，有

$$\|\boldsymbol{X} \otimes \boldsymbol{Y}\|_F = \|\boldsymbol{X}\|_F \cdot \|\boldsymbol{Y}\|_F \tag{2.49}$$

恒成立。

**例 18.** 给定矩阵 $\boldsymbol{X} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ 与 $\boldsymbol{Y} = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$，试写出 $\|\boldsymbol{X}\|_F$、$\|\boldsymbol{Y}\|_F$ 与 $\|\boldsymbol{X} \otimes \boldsymbol{Y}\|_F$。

**解.** 根据定义，矩阵 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 的 *Frobenius* 范数分别为

$$\|\boldsymbol{X}\|_F = \sqrt{1^2+2^2+3^2+4^2} = \sqrt{30} \quad \|\boldsymbol{Y}\|_F = \sqrt{5^2+6^2+7^2+8^2} = \sqrt{174} \tag{2.50}$$

由于

$$\boldsymbol{X} \otimes \boldsymbol{Y} = \begin{bmatrix} 5 & 6 & 10 & 12 \\ 7 & 8 & 14 & 16 \\ 15 & 18 & 20 & 24 \\ 21 & 24 & 28 & 32 \end{bmatrix} \tag{2.51}$$

故 $\|\boldsymbol{X} \otimes \boldsymbol{Y}\|_F = \sqrt{5220}$。

Frobenius 范数这一概念不适用于向量，对于任意向量 $\boldsymbol{x} \in \mathbb{R}^m$，其元素的平方和开根号是 $\ell_2$ 范数，即

$$\|\boldsymbol{x}\|_2 = \sqrt{\sum_{i=1}^{m} x_i^2} \tag{2.52}$$

**例 19.** 给定向量 $\boldsymbol{x} = (1,2)^\top$ 与 $\boldsymbol{y} = (3,4)^\top$，试写出 $\|\boldsymbol{x}\|_2$、$\|\boldsymbol{y}\|_2$ 与 $\|\boldsymbol{x} \otimes \boldsymbol{y}\|_2$。

**解.** 根据定义，向量 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的 $\ell_2$ 范数分别为

$$\|\boldsymbol{x}\|_2 = \sqrt{1^2+2^2} = \sqrt{5} \quad \|\boldsymbol{y}\|_2 = \sqrt{3^2+4^2} = 5 \tag{2.53}$$

由于 $\boldsymbol{x} \otimes \boldsymbol{y} = (3,4,6,8)^\top$，故 $\|\boldsymbol{x} \otimes \boldsymbol{y}\|_2 = \sqrt{3^2+4^2+6^2+8^2} = 5\sqrt{5}$。

### 2.3.3 矩阵的行列式

矩阵的行列式 (determinant) 是线性代数中非常重要的一个概念，贯穿线性代数的几乎所有内容，一般使用符号 $\det(\cdot)$ 表示。若给定矩阵 $\boldsymbol{X}\in\mathbb{R}^{m\times m}$ 与 $\boldsymbol{Y}\in\mathbb{R}^{n\times n}$，则

$$\det(\boldsymbol{X}\otimes\boldsymbol{Y})=\det(\boldsymbol{X})^n\cdot\det(\boldsymbol{Y})^m \tag{2.54}$$

恒成立。

**例 20.** 给定矩阵 $\boldsymbol{X}=\begin{bmatrix}1&2\\3&4\end{bmatrix}$ 与 $\boldsymbol{Y}=\begin{bmatrix}1&3&2\\4&1&3\\2&5&2\end{bmatrix}$，试写出矩阵的行列式 $\det(\boldsymbol{X})$、$\det(\boldsymbol{Y})$ 与 $\det(\boldsymbol{X}\otimes\boldsymbol{Y})$。

**解.** 矩阵 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 的行列式分别为

$$\det(\boldsymbol{X})=\begin{vmatrix}1&2\\3&4\end{vmatrix}=-2\quad\det(\boldsymbol{Y})=\begin{vmatrix}1&3&2\\4&1&3\\2&5&2\end{vmatrix}=17 \tag{2.55}$$

故 $\det(\boldsymbol{X})^3\cdot\det(\boldsymbol{Y})^2=-2312$。

矩阵 $\boldsymbol{X}\otimes\boldsymbol{Y}$ 的行列式为

$$\det(\boldsymbol{X}\otimes\boldsymbol{Y})=\begin{vmatrix}1&3&2&2&6&4\\4&1&3&8&2&6\\2&5&2&4&10&4\\3&9&6&4&12&8\\12&3&9&16&4&12\\6&15&6&8&20&8\end{vmatrix}=-2312 \tag{2.56}$$

### 2.3.4 矩阵的秩

矩阵的秩 (rank) 是线性代数中非常重要的一个概念，在信号处理、图像处理等领域中应用广泛，一般使用符号 $\mathrm{rank}(\cdot)$ 表示。若给定矩阵 $\boldsymbol{X}\in\mathbb{R}^{m\times n}$ 与 $\boldsymbol{Y}\in\mathbb{R}^{p\times q}$，则

$$\mathrm{rank}(\boldsymbol{X}\otimes\boldsymbol{Y})=\mathrm{rank}(\boldsymbol{X})\cdot\mathrm{rank}(\boldsymbol{Y}) \tag{2.57}$$

恒成立。

**例 21.** 给定矩阵 $\boldsymbol{X}=\begin{bmatrix}1&2\\2&4\end{bmatrix}$ 与 $\boldsymbol{Y}=\begin{bmatrix}5&6&7\\8&9&10\end{bmatrix}$，试写出 $\mathrm{rank}(\boldsymbol{X})$、$\mathrm{rank}(\boldsymbol{Y})$ 与 $\mathrm{rank}(\boldsymbol{X}\otimes\boldsymbol{Y})$。

**解.** 在这里，$\mathrm{rank}(\boldsymbol{X})=1$，$\mathrm{rank}(\boldsymbol{Y})=2$。

由于

$$\boldsymbol{X}\otimes\boldsymbol{Y}=\begin{bmatrix}5&6&7&10&12&14\\8&9&10&16&18&20\\10&12&14&20&24&28\\16&18&20&32&36&40\end{bmatrix} \tag{2.58}$$

故 $\mathrm{rank}(\boldsymbol{X}\otimes\boldsymbol{Y})=2$。

# 2.4 向量外积

## 2.4.1 定义

在线性代数中，两向量之间的外积可得到一个矩阵。对于任意向量 $\boldsymbol{a}\in\mathbb{R}^m$ 与 $\boldsymbol{b}\in\mathbb{R}^n$，则两者之间的外积为

$$\boldsymbol{c}=\boldsymbol{a}\otimes_{\text{outer}}\boldsymbol{b}=\boldsymbol{a}\boldsymbol{b}^\top\in\mathbb{R}^{m\times n} \tag{2.59}$$

其中，符号 $\otimes_{\text{outer}}$ 表示向量外积。

依此类推，对任意 $d$ 个向量 $\boldsymbol{x}^{(k)}\in\mathbb{R}^{n_k},k=1,2,\dots,d$，其外积 (outer product) 可定义为

$$\boldsymbol{\mathcal{Y}}=\boldsymbol{x}^{(1)}\otimes_{\text{outer}}\boldsymbol{x}^{(2)}\otimes_{\text{outer}}\cdots\otimes_{\text{outer}}\boldsymbol{x}^{(d)}\in\mathbb{R}^{n_1\times n_2\times\cdots\times n_d} \tag{2.60}$$

其中，符号 $\otimes_{\text{outer}}$ 表示向量外积。在张量 $\boldsymbol{\mathcal{Y}}$ 中，任意第 $(i_1,i_2,\dots,i_d)$ 个元素为

$$y_{i_1,i_2,\dots,i_d}=\prod_{k=1}^{d}x_{i_k}^{(k)} \tag{2.61}$$

其中，$i_k=1,2,\dots,n_k$，$k=1,2,\dots,d$。

需要注意的是，由于张量 $\boldsymbol{\mathcal{Y}}$ 是由向量外积得到的，故常被称为秩一张量 (rank-one tensor)。当 $d=3$ 时，向量外积得到的三阶张量 $\boldsymbol{\mathcal{Y}}=\boldsymbol{x}^{(1)}\otimes_{\text{outer}}\boldsymbol{x}^{(2)}\otimes_{\text{outer}}\boldsymbol{x}^{(3)}\in\mathbb{R}^{n_1\times n_2\times n_3}$ 如图2.1所示，在这里，张量 $\boldsymbol{\mathcal{Y}}$ 的任意第 $(i_1,i_2,i_3)$ 个元素为

$$y_{i_1,i_2,i_3}=\prod_{k=1}^{3}x_{i_k}^{(k)} \tag{2.62}$$

其中，$i_k=1,2,\dots,n_k$，$k=1,2,3$。

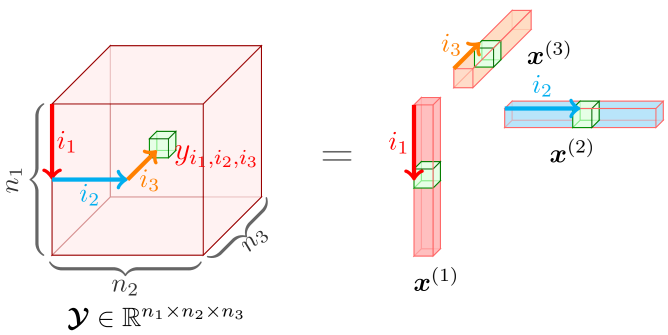



图 2.1: 向量外积得到的三阶张量 $\boldsymbol{\mathcal{Y}}\in\mathbb{R}^{n_1\times n_2\times n_3}$

当 $d=2$ 时，向量外积为

$$\boldsymbol{Y}=\boldsymbol{x}^{(1)}\otimes_{\text{outer}}\boldsymbol{x}^{(2)}=\boldsymbol{x}^{(1)}(\boldsymbol{x}^{(2)})^\top\in\mathbb{R}^{n_1\times n_2} \tag{2.63}$$

在矩阵 $\boldsymbol{Y}$ 中，任意第 $(i,j)$ 个元素为

$$y_{i,j}=x_i^{(1)}x_j^{(2)} \tag{2.64}$$

其中，$i=1,2,\dots,n_1$ 与 $j=1,2,\dots,n_2$。

**例 22.** 给定向量 $\boldsymbol{x}=(1,2)^\top$ 与 $\boldsymbol{y}=(3,4)^\top$，试写出 $\boldsymbol{x}\otimes_{\text{outer}}\boldsymbol{y}$。

**解.** 根据外积定义，有

$$\boldsymbol{x}\otimes_{\text{outer}}\boldsymbol{y}=\begin{bmatrix}3&4\\6&8\end{bmatrix} \tag{2.65}$$

**例 23.** 给定向量 $\boldsymbol{a}=(1,2)^\top$、$\boldsymbol{b}=(3,4,5)^\top$ 与 $\boldsymbol{c}=(6,7,8,9)^\top$，试写出 $\boldsymbol{a}\otimes_{\text{outer}}\boldsymbol{b}\otimes_{\text{outer}}\boldsymbol{c}$。

**解.** 令 $\boldsymbol{\mathcal{Y}}=\boldsymbol{a}\otimes_{\text{outer}}\boldsymbol{b}\otimes_{\text{outer}}\boldsymbol{c}\in\mathbb{R}^{2\times3\times4}$，根据外积定义，有

$$\boldsymbol{a}\otimes_{\text{outer}}\boldsymbol{b}=\begin{bmatrix}3&4&5\\6&8&10\end{bmatrix}\tag{2.66}$$

由此，可得张量 $\boldsymbol{\mathcal{Y}}$ 的 *frontal* 切片为

$$\begin{aligned}&\boldsymbol{\mathcal{Y}}_{:,:,1}=(\boldsymbol{a}\otimes_{\text{outer}}\boldsymbol{b})\cdot c_1=\begin{bmatrix}18&24&30\\36&48&60\end{bmatrix}\quad \boldsymbol{\mathcal{Y}}_{:,:,2}=(\boldsymbol{a}\otimes_{\text{outer}}\boldsymbol{b})\cdot c_2=\begin{bmatrix}21&28&35\\42&56&70\end{bmatrix}\\&\boldsymbol{\mathcal{Y}}_{:,:,3}=(\boldsymbol{a}\otimes_{\text{outer}}\boldsymbol{b})\cdot c_3=\begin{bmatrix}24&32&40\\48&64&80\end{bmatrix}\quad \boldsymbol{\mathcal{Y}}_{:,:,4}=(\boldsymbol{a}\otimes_{\text{outer}}\boldsymbol{b})\cdot c_4=\begin{bmatrix}27&36&45\\54&72&90\end{bmatrix}\end{aligned}\tag{2.67}$$

### 2.4.2 性质

**张量矩阵化**

根据 Khatri-Rao 积定义与张量矩阵化规则，由向量 $\boldsymbol{x}^{(k)}\in\mathbb{R}^{n_k},k=1,2,\ldots,d$ 的外积得到的张量 $\boldsymbol{\mathcal{Y}}\in\mathbb{R}^{n_1\times n_2\times\cdots\times n_d}$，其自第 $k$ 维度展开得到的矩阵可写作如下形式：

$$\begin{aligned}\boldsymbol{Y}_{(k)}=&\boldsymbol{x}^{(k)}\otimes_{\text{outer}}(\boldsymbol{x}^{(d)}\odot\cdots\odot\boldsymbol{x}^{(k+1)}\odot\boldsymbol{x}^{(k-1)}\odot\cdots\odot\boldsymbol{x}^{(1)})\\=&\boldsymbol{x}^{(k)}(\boldsymbol{x}^{(d)}\odot\cdots\odot\boldsymbol{x}^{(k+1)}\odot\boldsymbol{x}^{(k-1)}\odot\cdots\odot\boldsymbol{x}^{(1)})^\top\in\mathbb{R}^{n_k\times\prod_{h\neq k}n_h}\end{aligned}\tag{2.68}$$

其中，$\odot$ 表示 Khatri-Rao 积。

**例 24.** 给定向量 $\boldsymbol{a}=(1,2)^\top$、$\boldsymbol{b}=(3,4,5)^\top$ 与 $\boldsymbol{c}=(6,7,8,9)^\top$，若 $\boldsymbol{\mathcal{Y}}=\boldsymbol{a}\otimes_{\text{outer}}\boldsymbol{b}\otimes_{\text{outer}}\boldsymbol{c}$，试写出张量 $\boldsymbol{\mathcal{Y}}$ 的矩阵化形式 $\boldsymbol{Y}_{(1)}$、$\boldsymbol{Y}_{(2)}$ 与 $\boldsymbol{Y}_{(3)}$。

**解.** 根据 *Khatri-Rao* 积定义，有

$$\begin{cases}\boldsymbol{c}\odot\boldsymbol{b}=(18,24,30,21,28,35,24,32,40,27,36,45)^\top\\\boldsymbol{c}\odot\boldsymbol{a}=(6,12,7,14,8,16,9,18)^\top\\\boldsymbol{b}\odot\boldsymbol{a}=(3,6,4,8,5,10)^\top\end{cases}\tag{2.69}$$

从而，可得到

$$\boldsymbol{Y}_{(1)}=\boldsymbol{a}(\boldsymbol{c}\odot\boldsymbol{b})^\top=\begin{bmatrix}18&24&30&21&28&35&24&32&40&27&36&45\\36&48&60&42&56&70&48&64&80&54&72&90\end{bmatrix}\tag{2.70}$$

$$\boldsymbol{Y}_{(2)}=\boldsymbol{a}(\boldsymbol{c}\odot\boldsymbol{b})^\top=\begin{bmatrix}18&36&21&42&24&48&27&54\\24&48&28&56&32&64&36&72\\30&60&35&70&40&80&45&90\end{bmatrix}\tag{2.71}$$

$$\boldsymbol{Y}_{(3)}=\boldsymbol{c}(\boldsymbol{b}\odot\boldsymbol{a})^\top=\begin{bmatrix}18&36&24&48&30&60\\21&42&28&56&35&70\\24&48&32&64&40&80\\27&54&36&72&45&90\end{bmatrix}\tag{2.72}$$

**张量向量化**

根据 Khatri-Rao 积定义与张量向量化规则，由向量 $\boldsymbol{x}^{(k)} \in \mathbb{R}^{n_k}$, $k = 1, 2, \ldots, d$ 的外积得到的张量 $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$，其向量化形式为

$$\mathrm{vec}(\boldsymbol{\mathcal{Y}}) = \boldsymbol{x}^{(d)} \odot \boldsymbol{x}^{(d-1)} \odot \cdots \odot \boldsymbol{x}^{(2)} \odot \boldsymbol{x}^{(1)} \tag{2.73}$$

其中，$\odot$ 表示 Khatri-Rao 积；$\mathrm{vec}(\cdot)$ 表示向量化操作。

**例 25.** 给定向量 $\boldsymbol{a} = (1, 2)^\top$、$\boldsymbol{b} = (3, 4, 5)^\top$ 与 $\boldsymbol{c} = (6, 7, 8, 9)^\top$，若 $\boldsymbol{\mathcal{Y}} = \boldsymbol{a} \otimes_{\text{outer}} \boldsymbol{b} \otimes_{\text{outer}} \boldsymbol{c}$，试写出张量 $\boldsymbol{\mathcal{Y}}$ 的向量化形式 $\mathrm{vec}(\boldsymbol{\mathcal{Y}})$。

**解.** 根据 *Khatri-Rao* 积定义，有

$$\boldsymbol{c} \odot \boldsymbol{b} = (18, 24, 30, 21, 28, 35, 24, 32, 40, 27, 36, 45)^\top \tag{2.74}$$

从而，可得到

$$\begin{aligned} \mathrm{vec}(\boldsymbol{\mathcal{Y}}) = \boldsymbol{c} \odot \boldsymbol{b} \odot \boldsymbol{a} = (&18, 24, 30, 21, 28, 35, 24, 32, 40, 27, 36, 45, \\ &36, 48, 60, 42, 56, 70, 48, 64, 80, 54, 72, 90)^\top \end{aligned} \tag{2.75}$$

## 2.5 CP 张量分解

### 2.5.1 CP 分解形式

CP 分解全称为 CANDECOMP/PARAFAC 分解，它是以向量外积构成的分解形式，本质上是矩阵分解的高阶泛化 [Kolda and Bader, 2009]。给定任意张量 $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$，若令其秩为 $R$，则 CP 分解可写作如下形式：

$$\boldsymbol{\mathcal{Y}} = \sum_{r=1}^{R} \boldsymbol{u}_r^{(1)} \otimes_{\text{outer}} \boldsymbol{u}_r^{(2)} \otimes_{\text{outer}} \cdots \otimes_{\text{outer}} \boldsymbol{u}_r^{(d)} \tag{2.76}$$

其中，因子矩阵为

$$\boldsymbol{U}^{(k)} = \begin{bmatrix} | & | & & | \\ \boldsymbol{u}_1^{(k)} & \boldsymbol{u}_2^{(k)} & \cdots & \boldsymbol{u}_R^{(k)} \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{n_k \times R}, \, k = 1, 2, \ldots, d \tag{2.77}$$

在张量 $\boldsymbol{\mathcal{Y}}$ 中，任意第 $(i_1, i_2, \ldots, i_d)$ 个元素为

$$y_{i_1, i_2, \ldots, i_d} = \sum_{r=1}^{R} u_{i_1, r}^{(1)} \times u_{i_2, r}^{(2)} \times \cdots \times u_{i_d, r}^{(d)} = \sum_{r=1}^{R} \prod_{k=1}^{d} u_{i_k, r}^{(k)} \tag{2.78}$$

其中，$i_k = 1, 2, \ldots, n_k$, $k = 1, 2, \ldots, d$。

**解说 1** (三阶张量的 CP 分解). 给定三阶张量 $\boldsymbol{\mathcal{Y}}$，其 *CP* 分解可写作如下形式：

$$\boldsymbol{\mathcal{Y}} = \sum_{r=1}^{R} \boldsymbol{u}_r^{(1)} \otimes_{\text{outer}} \boldsymbol{u}_r^{(2)} \otimes_{\text{outer}} \boldsymbol{u}_r^{(3)} \tag{2.79}$$

任意第 $(i_1, i_2, \ldots, i_d)$ 个元素为

$$y_{i_1, i_2, i_3} = \sum_{r=1}^{R} u_{i_1, r}^{(1)} \times u_{i_2, r}^{(2)} \times u_{i_3, r}^{(3)} = \sum_{r=1}^{R} \prod_{k=1}^{3} u_{i_k, r}^{(k)} \tag{2.80}$$

图 *2.2* 给出了三阶张量的 *CP* 分解示意图。

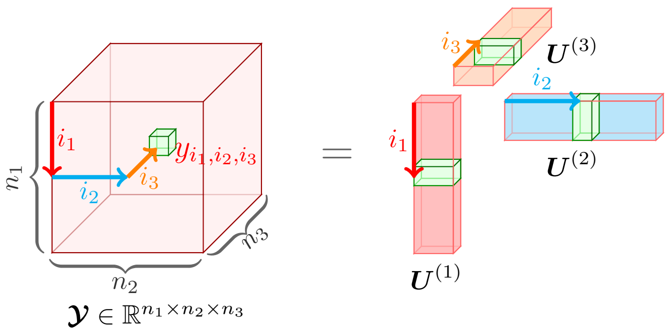



图 2.2: 三阶张量 $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ 的 CP 分解

由于 CP 分解可写成因子矩阵列向量外积的形式，因此，CP 分解具有以下性质：

- **张量矩阵化**。由因子矩阵 $\boldsymbol{U}^{(k)} \in \mathbb{R}^{n_k \times R}, k = 1, 2, \ldots, d$ 相乘得到的张量 $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$，其自第 $k$ 维度展开得到的矩阵可写作如下形式：

$$\boldsymbol{Y}_{(k)} = \boldsymbol{U}^{(k)}(\boldsymbol{U}^{(d)} \odot \cdots \odot \boldsymbol{U}^{(k+1)} \odot \boldsymbol{U}^{(k-1)} \odot \cdots \odot \boldsymbol{U}^{(1)})^\top \in \mathbb{R}^{n_k \times \prod_{h \neq k} n_h} \tag{2.81}$$

- **张量向量化**。由因子矩阵 $\boldsymbol{U}^{(k)} \in \mathbb{R}^{n_k \times R}, k = 1, 2, \ldots, d$ 相乘得到的张量 $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$，其向量化形式为

$$\operatorname{vec}(\boldsymbol{\mathcal{Y}}) = \sum_{r=1}^{R} \boldsymbol{u}_r^{(d)} \odot \boldsymbol{u}_r^{(d-1)} \odot \cdots \odot \boldsymbol{u}_r^{(2)} \odot \boldsymbol{u}_r^{(1)} \tag{2.82}$$

### 2.5.2　低秩逼近问题

# 第三章　时序矩阵分解

时序矩阵分解是矩阵分解中的一个重要模型，主要用于对时间序列数据进行建模 [Chen et al., 2022a]。当多元时间序列数据存在缺失值时，时序矩阵分解中的时序建模技术如向量自回归便会起到不可忽视的作用。在时序矩阵分解中，矩阵分解可从部分观测数据中学习出低秩模式，而时序建模则可刻画时序关联特征。本章分别介绍考虑时空平滑的矩阵分解与时序矩阵分解。

## 3.1　平滑矩阵分解

### 3.1.1　模型表达式

对于多元时间序列，若任意时刻 $t$ 对应的观测数据为向量 $\boldsymbol{y}_t \in \mathbb{R}^N$，则多元时间序列可写作矩阵形式：

$$\boldsymbol{Y} = \begin{bmatrix} | & | & & | \\ \boldsymbol{y}_1 & \boldsymbol{y}_2 & \cdots & \boldsymbol{y}_T \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{N \times T} \tag{3.1}$$

当矩阵中存在缺失值时，可用 $\Omega$ 表示被观测元素的索引集合。一般而言，可定义作用于集合 $\Omega$ 上的正交映射 (orthogonal projection) $\mathcal{P}_{\Omega}: \mathbb{R}^{N\times T} \to \mathbb{R}^{N\times T}$，对于矩阵 $\boldsymbol{Y}$ 任意第 $(i,t)$ 个元素，有

$$[\mathcal{P}_{\Omega}(\boldsymbol{Y})]_{i,t} = \begin{cases} \boldsymbol{y}_{i,t} & \text{if } (i,t) \in \Omega \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

同时，可定义作用于集合 $\Omega$ 补集上的正交映射 $\mathcal{P}_{\Omega}^{\perp}: \mathbb{R}^{N\times T} \to \mathbb{R}^{N\times T}$。

**例 26.** 给定矩阵 $\boldsymbol{X} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$，若 $\Omega = \{(1,1),(2,2)\}$，则

$$\mathcal{P}_{\Omega}(\boldsymbol{X}) = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} \quad \mathcal{P}_{\Omega}^{\perp}(\boldsymbol{X}) = \begin{bmatrix} 0 & 2 \\ 3 & 0 \end{bmatrix} \tag{3.3}$$

通常来说，对于矩阵 $\boldsymbol{Y}$，矩阵分解的优化问题为

$$\min_{\boldsymbol{W},\boldsymbol{X}} \frac{1}{2}\left\|\mathcal{P}_{\Omega}(\boldsymbol{Y} - \boldsymbol{W}^{\top}\boldsymbol{X})\right\|_F^2 + \frac{\rho}{2}\left(\|\boldsymbol{W}\|_F^2 + \|\boldsymbol{X}\|_F^2\right) \tag{3.4}$$

其中，$\boldsymbol{W} \in \mathbb{R}^{R\times N}$ 与 $\boldsymbol{X} \in \mathbb{R}^{R\times T}$ 为因子矩阵，在时空交通数据中，$\boldsymbol{W}$ 与 $\boldsymbol{X}$ 通常被分别称为空间因子矩阵 (spatial factor matrix) 与时序因子矩阵 (temporal factor matrix)；$\rho$ 为正则项的权重系数。

在公式(3.4)所示的矩阵分解中，低秩结构能捕捉数据的全局模式，但却往往忽略了局部的关联特征。为了分别对空间局部信息与时序局部信息进行建模，不妨引入如下形式的矩阵

分解，即平滑矩阵分解[1]：

$$\min_{\boldsymbol{W},\boldsymbol{X}}\ \frac{1}{2}\|\mathcal{P}_{\Omega}(\boldsymbol{Y}-\boldsymbol{W}^{\top}\boldsymbol{X})\|_F^2+\frac{\rho}{2}(\|\boldsymbol{W}\|_F^2+\|\boldsymbol{X}\|_F^2)+\frac{\lambda}{2}(\|\boldsymbol{W}\boldsymbol{\Psi}_1^{\top}\|_F^2+\|\boldsymbol{X}\boldsymbol{\Psi}_2^{\top}\|_F^2) \tag{3.5}$$

其中，目标函数的最后两项为平滑正则项 (smoothing regularization)，用于平滑处理的矩阵被定义为

$$\begin{aligned}\boldsymbol{\Psi}_1&=\begin{bmatrix}\boldsymbol{0}_{N\times 1} & \boldsymbol{I}_{N-1}\end{bmatrix}-\begin{bmatrix}\boldsymbol{I}_{N-1} & \boldsymbol{0}_{N\times 1}\end{bmatrix}\in\mathbb{R}^{(N-1)\times N}\\ \boldsymbol{\Psi}_2&=\begin{bmatrix}\boldsymbol{0}_{T\times 1} & \boldsymbol{I}_{T-1}\end{bmatrix}-\begin{bmatrix}\boldsymbol{I}_{T-1} & \boldsymbol{0}_{T\times 1}\end{bmatrix}\in\mathbb{R}^{(T-1)\times T}\end{aligned} \tag{3.6}$$

根据定义，恒有

$$\begin{aligned}\boldsymbol{W}\boldsymbol{\Psi}_1^{\top}&=\begin{bmatrix}| & | & & |\\ \boldsymbol{w}_2 & \boldsymbol{w}_3 & \cdots & \boldsymbol{w}_N\\ | & | & & |\end{bmatrix}-\begin{bmatrix}| & | & & |\\ \boldsymbol{w}_1 & \boldsymbol{w}_2 & \cdots & \boldsymbol{w}_{N-1}\\ | & | & & |\end{bmatrix}\\ \boldsymbol{X}\boldsymbol{\Psi}_2^{\top}&=\begin{bmatrix}| & | & & |\\ \boldsymbol{x}_2 & \boldsymbol{x}_3 & \cdots & \boldsymbol{x}_T\\ | & | & & |\end{bmatrix}-\begin{bmatrix}| & | & & |\\ \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_{T-1}\\ | & | & & |\end{bmatrix}\end{aligned} \tag{3.7}$$

### 3.1.2 求解过程

为了估计优化问题中的待定参数，即 $\boldsymbol{W}$ 与 $\boldsymbol{X}$，可采用交替优化算法 (alternating minimization algorithm)。交替优化算法（如交替最小二乘法）是求解矩阵分解中非凸优化问题的常用方法，该方法采用迭代过程，可通过交替更新待估计变量的最优解（如最小二乘解）最终达到收敛。

在平滑矩阵分解的优化问题中，令目标函数为

$$f=\frac{1}{2}\|\mathcal{P}_{\Omega}(\boldsymbol{Y}-\boldsymbol{W}^{\top}\boldsymbol{X})\|_F^2+\frac{\rho}{2}(\|\boldsymbol{W}\|_F^2+\|\boldsymbol{X}\|_F^2)+\frac{\lambda}{2}(\|\boldsymbol{W}\boldsymbol{\Psi}_1^{\top}\|_F^2+\|\boldsymbol{X}\boldsymbol{\Psi}_2^{\top}\|_F^2) \tag{3.8}$$

**更新变量 $\boldsymbol{W}$**

对变量 $\boldsymbol{W}$ 求偏导数，有

$$\frac{\partial f}{\partial \boldsymbol{W}}=-\boldsymbol{X}\mathcal{P}_{\Omega}(\boldsymbol{W}^{\top}\boldsymbol{X})+\rho\boldsymbol{W}+\lambda\boldsymbol{W}\boldsymbol{\Psi}_1^{\top}\boldsymbol{\Psi}_1 \tag{3.9}$$

此时，令 $\frac{\partial f}{\partial \boldsymbol{W}}=\boldsymbol{0}$，则可得到如下矩阵方程：

$$\boldsymbol{X}\mathcal{P}_{\Omega}(\boldsymbol{W}^{\top}\boldsymbol{X})+\rho\boldsymbol{W}+\lambda\boldsymbol{W}\boldsymbol{\Psi}_1^{\top}\boldsymbol{\Psi}_1=\boldsymbol{X}\mathcal{P}_{\Omega}^{\top}(\boldsymbol{Y}) \tag{3.10}$$

在这里，可采用共轭梯度法对该矩阵方程进行求解。

**解说 2** (共轭梯度法求解线性方程组)**.** 共轭梯度法是一种经典的数值计算方法，主要用于求解线性方程组，在机器学习中有诸多应用。从算法结构上来看，共轭梯度法是一种迭代算法，当线性方程组的规模较大且存在稀疏性问题时，共轭梯度法便可派上用场。

在线性代数中，线性方程组的表达式为

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b} \tag{3.11}$$

其中，矩阵 $\boldsymbol{A}\in\mathbb{R}^{n\times n}$,$\boldsymbol{b}\in\mathbb{R}^{n}$ 已知，$\boldsymbol{x}\in\mathbb{R}^{n}$ 为待求解变量。

不妨采用共轭梯度法对该线性方程组进行求解：

---

[1] 当 $\lambda=0$ 时，则考虑平滑处理的矩阵分解变为标准的矩阵分解。

1. 明确线性方程组中的矩阵 $\boldsymbol{A}$ 为实对称矩阵和正定矩阵，对变量 $\boldsymbol{x}$ 进行初始化（可令向量的所有元素均为 0），记作 $\boldsymbol{x}_0$；

2. 计算残差向量 $\boldsymbol{r}_0 := \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_0$；

3. 令 $\boldsymbol{q}_0 := \boldsymbol{r}_0$；

4. 定义迭代过程，令 $\ell = 0$：

    - 计算系数
    $$\alpha_\ell := \frac{\boldsymbol{r}_\ell^\top \boldsymbol{r}_\ell}{\boldsymbol{q}_\ell^\top \boldsymbol{A}\boldsymbol{q}_\ell} \tag{3.12}$$

    - 更新变量
    $$\boldsymbol{x}_{\ell+1} := \boldsymbol{x}_\ell + \alpha_\ell \boldsymbol{q}_\ell \tag{3.13}$$

    - 更新变量
    $$\boldsymbol{r}_{\ell+1} := \boldsymbol{r}_\ell - \alpha_\ell \boldsymbol{A}\boldsymbol{q}_\ell \tag{3.14}$$

    - 判断：若此时残差 $\boldsymbol{r}_{\ell+1}$ 足够小，可终止循环；

    - 计算系数
    $$\beta_\ell := \frac{\boldsymbol{r}_{\ell+1}^\top \boldsymbol{r}_{\ell+1}}{\boldsymbol{r}_\ell^\top \boldsymbol{r}_\ell} \tag{3.15}$$

    - 更新变量
    $$\boldsymbol{q}_{\ell+1} := \boldsymbol{r}_{\ell+1} + \beta_\ell \boldsymbol{q}_\ell \tag{3.16}$$

    - $\ell := \ell + 1$

5. 输出最终迭代结果 $\boldsymbol{x}_{\ell+1}$ 作为线性方程组的近似解。

假设线性方程组 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ 中已知 $\boldsymbol{A} = \begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix}$ 与 $\boldsymbol{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$，即如下二元一次方程组：

$$\begin{cases} 4x_1 + x_2 = 1 \\ x_1 + 3x_2 = 2 \end{cases} \tag{3.17}$$

图3.1给出了共轭梯度法求解二元一次方程组的迭代过程，其中，两条直线的交点 $(\frac{1}{11}, \frac{7}{11})$ 为方程组的解。在迭代过程中，变量 $\mathbf{0}$ 的初始值为坐标原点，第二次迭代结果恰好落在交点上，结果为

$$\boldsymbol{x} = \begin{bmatrix} 0.09090909 \\ 0.63636364 \end{bmatrix} \approx \begin{bmatrix} \frac{1}{11} \\ \frac{7}{11} \end{bmatrix} \tag{3.18}$$

需要注意的是，共轭梯度法具有快速收敛的性质，一般仅需少量迭代次数就可得到精准的近似解。

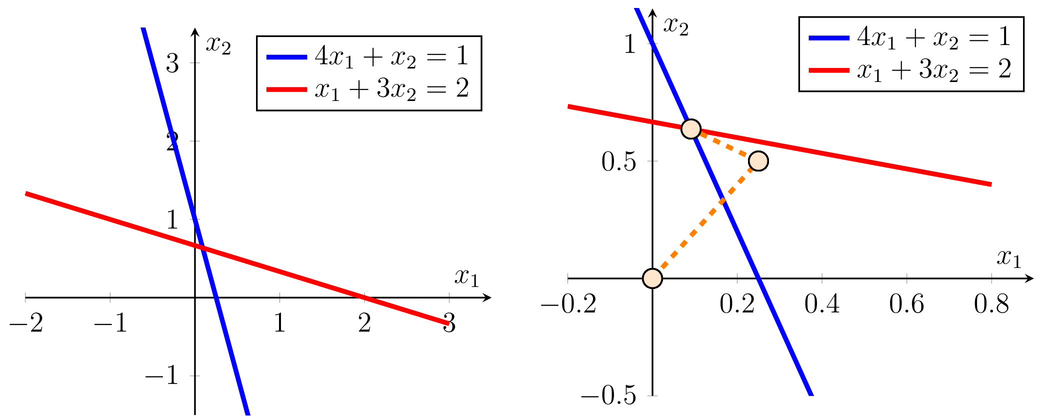
**例 27.** 试写出求解公式(3.10)所示矩阵方程的共轭梯度法。

**解.** 对公式(3.10)左边进行向量化操作，并记作

$$g(\boldsymbol{W}) = \operatorname{vec}(\boldsymbol{X}\mathcal{P}_\Omega(\boldsymbol{W}^\top \boldsymbol{X}) + \rho \boldsymbol{W} + \lambda \boldsymbol{W}\boldsymbol{\Psi}_1^\top \boldsymbol{\Psi}_1) \tag{3.19}$$

共轭梯度法的具体过程如下：

1. 对变量 $\boldsymbol{W}$ 进行向量化操作，并初始化为 $\boldsymbol{w}_0$；




(a) 二元一次方程组 (b) 共轭梯度法迭代过程

图 3.1: 使用共轭梯度法求解二元一次方程组

*2.* 计算残差向量 $\boldsymbol{r}_0 := \text{vec}(\boldsymbol{X}\mathcal{P}_\Omega^\top(\boldsymbol{Y})) - g(\boldsymbol{W}_0)$;

*3.* 令 $\boldsymbol{q}_0 := \boldsymbol{r}_0$;

*4.* 定义迭代过程，令 $\ell = 0$:

- 计算系数
$$\alpha_\ell := \frac{\boldsymbol{r}_\ell^\top \boldsymbol{r}_\ell}{\boldsymbol{q}_\ell^\top g(\boldsymbol{Q}_\ell)} \tag{3.20}$$
其中，$\boldsymbol{Q}_\ell$ 是 $\boldsymbol{q}_\ell$ 的矩阵化结果;
- 更新变量
$$\boldsymbol{w}_{\ell+1} := \boldsymbol{w}_\ell + \alpha_\ell \boldsymbol{q}_\ell \tag{3.21}$$
- 更新变量
$$\boldsymbol{r}_{\ell+1} := \boldsymbol{r}_\ell - \alpha_\ell g(\boldsymbol{Q}_\ell) \tag{3.22}$$
- 判断：若此时残差 $\boldsymbol{r}_{\ell+1}$ 足够小，可终止循环;
- 计算系数
$$\beta_\ell := \frac{\boldsymbol{r}_{\ell+1}^\top \boldsymbol{r}_{\ell+1}}{\boldsymbol{r}_\ell^\top \boldsymbol{r}_\ell} \tag{3.23}$$
- 更新变量
$$\boldsymbol{q}_{\ell+1} := \boldsymbol{r}_{\ell+1} + \beta_\ell \boldsymbol{q}_\ell \tag{3.24}$$
- $\ell := \ell + 1$

*5.* 输出最终迭代结果 $\boldsymbol{w}_{\ell+1}$ 并进行矩阵化，得到的矩阵 $\boldsymbol{W}_{\ell+1}$ 即作为矩阵方程的近似解。

**更新变量 $\boldsymbol{X}$**

对变量 $\boldsymbol{X}$ 求偏导数，有
$$\frac{\partial f}{\partial \boldsymbol{X}} = -\boldsymbol{W}\mathcal{P}_\Omega^\top(\boldsymbol{Y} - \boldsymbol{W}^\top\boldsymbol{X}) + \rho\boldsymbol{X} + \lambda\boldsymbol{X}\boldsymbol{\Psi}_2^\top\boldsymbol{\Psi}_2 \tag{3.25}$$
此时，令 $\frac{\partial f}{\partial \boldsymbol{X}} = \boldsymbol{0}$，则可得到如下矩阵方程:
$$\boldsymbol{W}\mathcal{P}_\Omega^\top(\boldsymbol{W}^\top\boldsymbol{X}) + \rho\boldsymbol{X} + \lambda\boldsymbol{X}\boldsymbol{\Psi}_2^\top\boldsymbol{\Psi}_2 = \boldsymbol{W}\mathcal{P}_\Omega(\boldsymbol{Y}) \tag{3.26}$$
在这里，同样可采用共轭梯度法对该矩阵方程进行求解。

**例 28.** 试写出求解公式(3.26)所示矩阵方程的共轭梯度法。

**解.** 对公式(3.26)左边进行向量化操作，并记作

$$g(\boldsymbol{X})=\operatorname{vec}(\boldsymbol{W}\mathcal{P}_{\Omega}^{\top}(\boldsymbol{W}^{\top}\boldsymbol{X})+\rho\boldsymbol{X}+\lambda\boldsymbol{X}\boldsymbol{\Psi}_2^{\top}\boldsymbol{\Psi}_2) \tag{3.27}$$

共轭梯度法的具体过程如下：

1. 对变量 $\boldsymbol{X}$ 进行向量化操作，并初始化为 $\boldsymbol{x}_0$；
2. 计算残差向量 $\boldsymbol{r}_0:=\operatorname{vec}(\boldsymbol{W}\mathcal{P}_{\Omega}(\boldsymbol{Y}))-g(\boldsymbol{X}_0)$；
3. 令 $\boldsymbol{q}_0:=\boldsymbol{r}_0$；
4. 定义迭代过程，令 $\ell=0$：
   - 计算系数
   $$\alpha_{\ell}:=\frac{\boldsymbol{r}_{\ell}^{\top}\boldsymbol{r}_{\ell}}{\boldsymbol{q}_{\ell}^{\top}g(\boldsymbol{Q}_{\ell})} \tag{3.28}$$
   其中，$\boldsymbol{Q}_{\ell}$ 是 $\boldsymbol{q}_{\ell}$ 的矩阵化结果；
   - 更新变量
   $$\boldsymbol{x}_{\ell+1}:=\boldsymbol{x}_{\ell}+\alpha_{\ell}\boldsymbol{q}_{\ell} \tag{3.29}$$
   - 更新变量
   $$\boldsymbol{r}_{\ell+1}:=\boldsymbol{r}_{\ell}-\alpha_{\ell}g(\boldsymbol{Q}_{\ell}) \tag{3.30}$$
   - 判断：若此时残差 $\boldsymbol{r}_{\ell+1}$ 足够小，可终止循环；
   - 计算系数
   $$\beta_{\ell}:=\frac{\boldsymbol{r}_{\ell+1}^{\top}\boldsymbol{r}_{\ell+1}}{\boldsymbol{r}_{\ell}^{\top}\boldsymbol{r}_{\ell}} \tag{3.31}$$
   - 更新变量
   $$\boldsymbol{q}_{\ell+1}:=\boldsymbol{r}_{\ell+1}+\beta_{\ell}\boldsymbol{q}_{\ell} \tag{3.32}$$
   - $\ell:=\ell+1$
5. 输出最终迭代结果 $\boldsymbol{x}_{\ell+1}$ 并进行矩阵化，得到的矩阵 $\boldsymbol{X}_{\ell+1}$ 即作为矩阵方程的近似解。

算法1给出了考虑平滑处理的矩阵分解算法的具体实现过程。

**Algorithm 1** 考虑平滑处理的矩阵分解算法

**Input:** 观测矩阵 $\boldsymbol{Y}\in\mathbb{R}^{N\times T}$，被观测元素的索引集合 $\Omega$，超参数 $\{\rho,\lambda\}$。
**Output:** 重构出来的矩阵 $\hat{\boldsymbol{Y}}\in\mathbb{R}^{N\times T}$。

```
1: 对变量 {W, X} 进行初始化；
2: for i = 0 to 最大迭代次数 do
3:     使用共轭梯度法对公式(3.10)中的变量 W 进行求解；
4:     使用共轭梯度法对公式(3.26)中的变量 X 进行求解；
5:     计算 Ŷ = W^⊤ X；
6: end for
```

## 3.2 时序矩阵分解

### 3.2.1 模型表达式

在该矩阵分解的优化问题中，可通过增加向量自回归过程使得模型具备时序建模能力。不妨对时序因子矩阵构造向量自回归过程，可得到时序矩阵分解的优化问题为

$$\min_{\boldsymbol{W},\boldsymbol{X},\{\boldsymbol{A}_k\}}\ \frac{1}{2}\left\|\mathcal{P}_{\Omega}(\boldsymbol{Y}-\boldsymbol{W}^\top\boldsymbol{X})\right\|_F^2+\frac{\rho}{2}\left(\|\boldsymbol{W}\|_F^2+\|\boldsymbol{X}\|_F^2\right)+\frac{\lambda}{2}\sum_{t=d+1}^{T}\left\|\boldsymbol{x}_t-\sum_{k=1}^{d}\boldsymbol{A}_k\boldsymbol{x}_{t-k}\right\|_2^2 \tag{3.33}$$

其中，$\boldsymbol{A}_1,\boldsymbol{A}_2,\ldots,\boldsymbol{A}_d\in\mathbb{R}^{R\times R}$ 为向量自回归过程的系数矩阵；$\lambda$ 为正则项的权重系数。

根据向量自回归的定义，时序矩阵分解的优化问题可写作如下形式：

$$\min_{\boldsymbol{W},\boldsymbol{X},\{\boldsymbol{A}_k\}}\ \frac{1}{2}\left\|\mathcal{P}_{\Omega}(\boldsymbol{Y}-\boldsymbol{W}^\top\boldsymbol{X})\right\|_F^2+\frac{\rho}{2}\left(\|\boldsymbol{W}\|_F^2+\|\boldsymbol{X}\|_F^2\right)+\frac{\lambda}{2}\left\|\boldsymbol{X}\boldsymbol{\Psi}_0^\top-\sum_{k=1}^{d}\boldsymbol{A}_k\boldsymbol{X}\boldsymbol{\Psi}_k^\top\right\|_F^2 \tag{3.34}$$

其中，$\boldsymbol{\Psi}_k\in\mathbb{R}^{(T-d)\times T},\,k=0,1,\ldots,d$ 为构造出来的矩阵（参见公式(3.37)）。

**解说 3** (向量自回归)**.** *对于多元时间序列，若任意时刻 $t$ 对应的观测数据为向量 $\boldsymbol{x}_t\in\mathbb{R}^N$，则向量自回归的表达式为*

$$\boldsymbol{x}_t=\sum_{k=1}^{d}\boldsymbol{A}_k\boldsymbol{x}_{t-k}+\boldsymbol{\epsilon}_t,\,t=d+1,d+2,\ldots,T \tag{3.35}$$

*其中，$\boldsymbol{A}_1,\boldsymbol{A}_2,\ldots,\boldsymbol{A}_d\in\mathbb{R}^{N\times N}$ 为自回归过程的系数矩阵 (coefficient matrix)；$d$ 为自回归过程的阶数 (order)；$\boldsymbol{\epsilon}_t\in\mathbb{R}^N$ 为残差向量。*

*令*

$$\boldsymbol{X}=\begin{bmatrix}| & | & & |\\ \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_T\\ | & | & & |\end{bmatrix}\in\mathbb{R}^{N\times T} \tag{3.36}$$

*若构造分块矩阵*

$$\begin{aligned}\boldsymbol{\Psi}_k&=\begin{bmatrix}\underbrace{\begin{matrix}0&\cdots&0\\ \vdots&\ddots&\vdots\\ 0&\cdots&0\end{matrix}}_{d-k}&\underbrace{\begin{matrix}1&\cdots&0\\ \vdots&\ddots&\vdots\\ 0&\cdots&1\end{matrix}}_{T-d}&\underbrace{\begin{matrix}0&\cdots&0\\ \vdots&\ddots&\vdots\\ 0&\cdots&0\end{matrix}}_{k}\end{bmatrix}\\ &=\begin{bmatrix}\boldsymbol{0}_{(T-d)\times(d-k)} & \boldsymbol{I}_{T-d} & \boldsymbol{0}_{(T-d)\times k}\end{bmatrix}\in\mathbb{R}^{(T-d)\times T},\,k=0,1,\ldots,d\end{aligned} \tag{3.37}$$

*则向量自回归可写作如下形式：*

$$\boldsymbol{X}\boldsymbol{\Psi}_0^\top=\sum_{k=1}^{d}\boldsymbol{A}_k\boldsymbol{X}\boldsymbol{\Psi}_k^\top+\boldsymbol{E} \tag{3.38}$$

*其中，$\boldsymbol{E}\in\mathbb{R}^{N\times(T-d)}$ 为残差矩阵。*

### 3.2.2 求解过程

为了估计优化问题中的待定参数，即变量 $\boldsymbol{W}$、$\boldsymbol{X}$ 以及 $\boldsymbol{A}_1,\boldsymbol{A}_2,\ldots,\boldsymbol{A}_d$，可采用交替优化算法。在时序矩阵分解中，每次更新特定变量时，可令其他变量固定不变，仅求解当前变量的最优解（如最小二乘解）或近似解。

**更新变量 $\boldsymbol{W}$**

不妨将时序矩阵分解优化问题的目标函数记作 $f$，对变量 $\boldsymbol{W}$ 求偏导数，有

$$\frac{\partial f}{\partial \boldsymbol{W}}=-\boldsymbol{X}\mathcal{P}_{\Omega}^{\top}(\boldsymbol{Y}-\boldsymbol{W}^{\top}\boldsymbol{X})+\rho\boldsymbol{W} \tag{3.39}$$

此时，令 $\frac{\partial f}{\partial \boldsymbol{W}}=\boldsymbol{0}$，则矩阵方程为

$$\boldsymbol{X}\mathcal{P}_{\Omega}^{\top}(\boldsymbol{W}^{\top}\boldsymbol{X})+\rho\boldsymbol{W}=\boldsymbol{X}\mathcal{P}_{\Omega}^{\top}(\boldsymbol{Y}) \tag{3.40}$$

对于变量 $\boldsymbol{W}$，该矩阵方程的最小二乘解为

$$\boldsymbol{w}_i=\Big(\sum_{t:(i,t)\in\Omega}\boldsymbol{x}_t\boldsymbol{x}_t^{\top}+\rho\boldsymbol{I}_R\Big)^{-1}\sum_{t:(i,t)\in\Omega}\boldsymbol{x}_t y_{i,t} \tag{3.41}$$

其中，$i=1,2,\ldots,N$；符号 $\sum\limits_{t:(i,t)\in\Omega}$ 表示固定索引 $i$ 后对索引集合 $\Omega$ 内所有索引 $t$ 进行求和。

**更新变量 $\boldsymbol{X}$**

令 $\boldsymbol{A}_0=-\boldsymbol{I}_R$，则时序矩阵分解优化问题的目标函数可写作如下形式：

$$f=\frac{1}{2}\left\|\mathcal{P}_{\Omega}(\boldsymbol{Y}-\boldsymbol{W}^{\top}\boldsymbol{X})\right\|_F^2+\frac{\rho}{2}\left(\|\boldsymbol{W}\|_F^2+\|\boldsymbol{X}\|_F^2\right)+\frac{\lambda}{2}\Big\|\sum_{k=0}^{d}\boldsymbol{A}_k\boldsymbol{X}\boldsymbol{\Psi}_k^{\top}\Big\|_F^2 \tag{3.42}$$

对变量 $\boldsymbol{X}$ 求偏导数，有

$$\frac{\partial f}{\partial \boldsymbol{X}}=-\boldsymbol{W}\mathcal{P}_{\Omega}(\boldsymbol{Y}-\boldsymbol{W}^{\top}\boldsymbol{X})+\rho\boldsymbol{X}+\lambda\sum_{k=0}^{d}\boldsymbol{A}_k^{\top}\Big(\sum_{h=0}^{d}\boldsymbol{A}_h\boldsymbol{X}\boldsymbol{\Psi}_h^{\top}\Big)\boldsymbol{\Psi}_k \tag{3.43}$$

此时，令 $\frac{\partial f}{\partial \boldsymbol{X}}=\boldsymbol{0}$，关于变量 $\boldsymbol{X}$ 的矩阵方程为

$$\boldsymbol{W}\mathcal{P}_{\Omega}(\boldsymbol{W}^{\top}\boldsymbol{X})+\rho\boldsymbol{X}+\lambda\sum_{k=0}^{d}\boldsymbol{A}_k^{\top}\Big(\sum_{h=0}^{d}\boldsymbol{A}_h\boldsymbol{X}\boldsymbol{\Psi}_h^{\top}\Big)\boldsymbol{\Psi}_k=\boldsymbol{W}\mathcal{P}_{\Omega}(\boldsymbol{Y}) \tag{3.44}$$

**例 29.** 试写出求解公式(3.44)所示矩阵方程的共轭梯度法。

**解.** 对公式(3.26)左边进行向量化操作，并记作

$$g(\boldsymbol{X})=\operatorname{vec}\left(\boldsymbol{W}\mathcal{P}_{\Omega}(\boldsymbol{W}^{\top}\boldsymbol{X})+\rho\boldsymbol{X}+\lambda\sum_{k=0}^{d}\boldsymbol{A}_k^{\top}\Big(\sum_{h=0}^{d}\boldsymbol{A}_h\boldsymbol{X}\boldsymbol{\Psi}_h^{\top}\Big)\boldsymbol{\Psi}_k\right) \tag{3.45}$$

共轭梯度法的具体过程可参考例28。

**更新系数矩阵 $\boldsymbol{A}_1,\boldsymbol{A}_2,\ldots,\boldsymbol{A}_d$**

若令

$$\begin{aligned}\boldsymbol{A}&=\begin{bmatrix}\boldsymbol{A}_1 & \boldsymbol{A}_2 & \cdots & \boldsymbol{A}_d\end{bmatrix}\in\mathbb{R}^{R\times(dR)}\\ \boldsymbol{\Psi}&=\begin{bmatrix}\boldsymbol{\Psi}_1 & \boldsymbol{\Psi}_2 & \cdots & \boldsymbol{\Psi}_d\end{bmatrix}\in\mathbb{R}^{(T-d)\times(dT)}\end{aligned} \tag{3.46}$$

则对于系数矩阵 $\boldsymbol{A}$ 的子问题为

$$\min_{\boldsymbol{A}}\ \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\Psi}_0^{\top}-\boldsymbol{A}(\boldsymbol{I}_d\otimes\boldsymbol{X})\boldsymbol{\Psi}^{\top}\|_F^2 \tag{3.47}$$

因此，系数矩阵 $\boldsymbol{A}$ 的最小二乘解为

$$\boldsymbol{A}=\boldsymbol{X}\boldsymbol{\Psi}_0^{\top}\left((\boldsymbol{I}_d\otimes\boldsymbol{X})\boldsymbol{\Psi}^{\top}\right)^{\dagger} \tag{3.48}$$

算法2给出了时序矩阵分解算法的具体实现过程。在这里，不妨使用共轭梯度法对变量 $\boldsymbol{W}$ 与 $\boldsymbol{X}$ 进行求解。

**Algorithm 2** 时序矩阵分解算法

**Input:** 观测矩阵 $\boldsymbol{Y} \in \mathbb{R}^{N \times T}$, 被观测元素的索引集合 $\Omega$, 向量自回归的阶数 $d$, 超参数 $\{\rho, \lambda\}$。

**Output:** 重构出来的矩阵 $\hat{\boldsymbol{Y}} \in \mathbb{R}^{N \times T}$。

```
对变量 {W, X, A} 进行初始化；
for i = 0 to 最大迭代次数 do
    使用共轭梯度法对公式(3.40)中的变量 W 进行求解；
    使用共轭梯度法对公式(3.44)中的变量 X 进行求解；
    根据公式(3.48)计算系数矩阵 A；
    计算 Ŷ = W^⊤ X；
end for
```

### 3.2.3　时间序列预测

# 第四章　贝叶斯张量分解

# 第五章　低秩张量填充

## 5.1　基于多重核范数的张量填充

## 5.2　基于多重截断核范数的张量填充

## 5.3　基于张量核范数的张量填充

引入多维傅立叶变换

# 第六章　低秩拉普拉斯卷积模型

## 6.1　离散傅立叶变换与循环卷积

离散傅立叶变换 (discrete Fourier transform) 是数学中非常重要的一个概念，被应用于众多领域，如信号处理与图像处理。由于离散傅立叶变换通常采用快速傅立叶变换 (fast Fourier transform) 进行高效求解，所以两者经常出现在一起。

### 6.1.1　一维卷积定理

实际上，循环卷积与离散傅立叶变换联系紧密。卷积定理 (convolution theorem) 可用于描述两者之间的这种关系，给定任意向量 $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^{T}$，有

$$\boldsymbol{x}\star\boldsymbol{y}=\mathcal{F}^{-1}(\mathcal{F}(\boldsymbol{x})\circ\mathcal{F}(\boldsymbol{y})) \tag{6.1}$$

恒成立。在这里，$\mathcal{F}(\cdot)$ 表示离散傅立叶变换；$\mathcal{F}^{-1}(\cdot)$ 表示离散傅立叶逆变换；符号 $\circ$ 表示元素间的点乘 (Hadamard product)。将向量做离散傅立叶变换的意义在于利用快速傅立叶变换进行高效计算。

假设向量 $\boldsymbol{x}\in\mathbb{R}^{T}$，给定向量 $\boldsymbol{y}\in\mathbb{R}^{\tau},\tau<T$，使用离散傅立叶变换计算循环卷积时需首先令

$$\boldsymbol{y}=(y_1,y_2,\cdots,y_{\tau},\underbrace{0,\cdots,0}_{T-\tau})^{\top}\in\mathbb{R}^{T} \tag{6.2}$$

然后分别对向量 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 进行离散傅立叶变换。

**例 30.** 给定向量 $\boldsymbol{x}=(0,1,2,3,4)^{\top}$ 与 $\boldsymbol{y}=(2,-1,3)^{\top}$，试根据卷积定理计算循环卷积 $\boldsymbol{z}=\boldsymbol{x}\star\boldsymbol{y}$。

**解.** 分别对向量 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 进行离散傅立叶变换，有

$$\begin{cases}\mathcal{F}(\boldsymbol{x})=(10,-2.5+3.44i,-2.5+0.81i,-2.5-0.81i,-2.5-3.44i)^{\top}\\ \mathcal{F}(\boldsymbol{y})=(4,-0.74-0.81i,3.74+3.44i,3.74-3.44i,-0.74+0.81i)^{\top}\end{cases} \tag{6.3}$$

其中，$i=\sqrt{-1}$ 表示复数的虚部。

根据卷积定理，有

$$\boldsymbol{z}=\mathcal{F}^{-1}(\mathcal{F}(\boldsymbol{x})\circ\mathcal{F}(\boldsymbol{y}))=(5,14,3,7,11)^{\top} \tag{6.4}$$

根据卷积定理，若 $\boldsymbol{z}=\boldsymbol{x}\star\boldsymbol{y}$，则

$$\boldsymbol{x}=\mathcal{F}^{-1}(\mathcal{F}(\boldsymbol{z})\oslash\mathcal{F}(\boldsymbol{y}))\qquad\boldsymbol{y}=\mathcal{F}^{-1}(\mathcal{F}(\boldsymbol{z})\oslash\mathcal{F}(\boldsymbol{x})) \tag{6.5}$$

其中，符号 $\oslash$ 表示元素间的点除 (Hadamard division)。

**例 31.** 给定向量 $\boldsymbol{x}=(0,1,2,3,4)^\top$ 与 $\boldsymbol{z}=(5,14,3,7,11)^\top$，若 $\boldsymbol{z}=\boldsymbol{x}\star\boldsymbol{y}$，试根据卷积定理计算 $\boldsymbol{y}$。

**解.** 根据卷积定理，有

$$\boldsymbol{y}=\mathcal{F}^{-1}(\mathcal{F}(\boldsymbol{z})\oslash\mathcal{F}(\boldsymbol{x}))=(2,-1,3,0,0)^\top \tag{6.6}$$

### 6.1.2 二维卷积定理

对于任意矩阵 $\boldsymbol{X}\in\mathbb{R}^{M\times N}$ 与 $\boldsymbol{K}\in\mathbb{R}^{\nu_1\times\nu_2}$，其中，$\nu_1\leq M,\nu_2\leq N$，若两者之间的循环卷积为 $\boldsymbol{Y}=\boldsymbol{K}\star\boldsymbol{X}\in\mathbb{R}^{M\times N}$，则矩阵 $\boldsymbol{Y}$ 的任意元素为

$$y_{m,n}=\sum_{i=1}^{\nu_1}\sum_{j=1}^{\nu_2}\kappa_{i,j}x_{m-i+1,n-j+1},\forall(m,n) \tag{6.7}$$

其中，$\kappa_{i,j}$ 为矩阵 $\boldsymbol{K}$ 的第 $(i,j)$ 个元素。

根据循环卷积定义，矩阵 $\boldsymbol{Y}$ 的第 $m$ 行为

$$\begin{aligned}\boldsymbol{y}_{m,:}&=\sum_{i=1}^{\nu_1}\boldsymbol{\kappa}_{i,:}\star\boldsymbol{x}_{m-i+1,:}\\&=\sum_{i=1}^{\nu_1}\mathcal{F}^{-1}(\mathcal{F}(\boldsymbol{\kappa}_{i,:})\circ\mathcal{F}(\boldsymbol{x}_{m-i+1,:}))\end{aligned} \tag{6.8}$$

其中，$\boldsymbol{\kappa}_{i,:}$ 为矩阵 $\boldsymbol{K}$ 的第 $i$ 行。

矩阵 $\boldsymbol{Y}$ 的第 $n$ 列为

$$\begin{aligned}\boldsymbol{y}_{:,n}&=\sum_{j=1}^{\nu_2}\boldsymbol{\kappa}_{:,j}\star\boldsymbol{x}_{:,n-j+1}\\&=\sum_{j=1}^{\nu_2}\mathcal{F}^{-1}(\mathcal{F}(\boldsymbol{\kappa}_{:,j})\circ\mathcal{F}(\boldsymbol{x}_{:,n-j+1}))\end{aligned} \tag{6.9}$$

其中，$\boldsymbol{\kappa}_{:,j}$ 为矩阵 $\boldsymbol{K}$ 的第 $j$ 列。

在这里，二维循环卷积与离散傅立叶变换之间的卷积定理也可写作如下形式：

$$\boldsymbol{Y}=\boldsymbol{K}\star\boldsymbol{X}=\mathcal{F}^{-1}(\mathcal{F}(\boldsymbol{K})\circ\mathcal{F}(\boldsymbol{X})) \tag{6.10}$$

其中，$\mathcal{F}(\cdot)$ 表示二维离散傅立叶变换；$\mathcal{F}^{-1}(\cdot)$ 表示二维离散傅立叶逆变换。

**例 32.** 给定矩阵 $\boldsymbol{X}=\begin{bmatrix}1&2&3&4\\4&5&6&7\\7&8&9&10\\10&11&12&13\end{bmatrix}$ 与 $\boldsymbol{K}=\begin{bmatrix}1&2&3\\4&5&6\\7&8&9\end{bmatrix}$，试根据卷积定理计算二维循环卷积 $\boldsymbol{Y}=\boldsymbol{K}\star\boldsymbol{X}$。

**解.** 根据卷积定理，二维循环卷积为

$$\boldsymbol{Y}=\mathcal{F}^{-1}(\mathcal{F}(\boldsymbol{K})\circ\mathcal{F}(\boldsymbol{X}))=\begin{bmatrix}405&390&363&408\\360&345&318&363\\207&192&165&210\\342&327&300&345\end{bmatrix} \tag{6.11}$$

### 6.1.3 Parseval 定理

Parseval 定理 (Parseval's thorem) 表明信号的能量在时域与频域相等。在离散傅立叶变换中，对于任意向量 $\boldsymbol{x}=(x_1,x_2,\cdots,x_T)\in\mathbb{R}^T$，离散形式的 Parseval 定理为

$$\|\boldsymbol{x}\|_2^2=\frac{1}{T}\|\mathcal{F}(\boldsymbol{x})\|_2^2 \tag{6.12}$$

**例 33.** 给定向量 $\boldsymbol{x}=(0,1,2,3,4)^\top$，试写出 $\|\boldsymbol{x}\|_2^2$ 与 $\|\mathcal{F}(\boldsymbol{x})\|_2^2$。

**解.** 根据 $\ell_2$ 范数定义，有

$$\|\boldsymbol{x}\|_2^2=0^2+1^2+2^2+3^2+4^2=30 \tag{6.13}$$

由于

$$\mathcal{F}(\boldsymbol{x})=(10,-2.5+3.44i,-2.5+0.81i,-2.5-0.81i,-2.5-3.44i)^\top \tag{6.14}$$

其中，$i=\sqrt{-1}$ 表示复数的虚部。故 $\|\mathcal{F}(\boldsymbol{x})\|_2^2=150$。

针对二维离散傅立叶变量，给定任意矩阵 $\boldsymbol{X}\in\mathbb{R}^{M\times N}$，Parseval 定理为

$$\|\boldsymbol{X}\|_F^2=\frac{1}{MN}\|\mathcal{F}(\boldsymbol{X})\|_F^2 \tag{6.15}$$

**例 34.** 给定矩阵 $\boldsymbol{X}=\begin{bmatrix}5 & 6 & 7\\ 8 & 9 & 10\end{bmatrix}$，试写出 $\|\boldsymbol{X}\|_F^2$ 与 $\|\mathcal{F}(\boldsymbol{X})\|_F^2$。

**解.** 根据 $F$ 范数定义，有

$$\|\boldsymbol{X}\|_F^2=5^2+6^2+7^2+8^2+9^2+10^2=355 \tag{6.16}$$

由于

$$\mathcal{F}(\boldsymbol{X})=\begin{bmatrix}45 & -3+1.73i & -3-1.73i\\ -9 & 0 & 0\end{bmatrix} \tag{6.17}$$

其中，$i=\sqrt{-1}$ 表示复数的虚部。故 $\|\mathcal{F}(\boldsymbol{X})\|_F^2=2130$。

## 6.2 离散傅立叶变换与循环矩阵核范数

### 6.2.1 循环矩阵定义

循环矩阵 (circulant matrix) 是一种特殊的代数结构，广泛应用于信号处理等。从定义出发，给定任意向量 $\boldsymbol{x}=(x_1,x_2,\cdots,x_T)^\top\in\mathbb{R}^T$，其对应的循环矩阵可写作如下形式：

$$\mathcal{C}(\boldsymbol{x})=\begin{bmatrix}x_1 & x_T & x_{T-1} & \cdots & x_2\\ x_2 & x_1 & x_T & \cdots & x_3\\ x_3 & x_2 & x_1 & \cdots & x_4\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ x_T & x_{T-1} & x_{T-2} & \cdots & x_1\end{bmatrix}\in\mathbb{R}^{T\times T} \tag{6.18}$$

其中，$\mathcal{C}:\mathbb{R}^T\to\mathbb{R}^{T\times T}$ 表示循环算子 (circulant operator)。该循环矩阵的第一列为向量 $\boldsymbol{x}$ 本身，对角线元素均为 $x_1$。

**例 35.** 给定任意向量 $\boldsymbol{x}=(x_1,x_2,x_3,x_4,x_5)^\top\in\mathbb{R}^5$，试写出其对应的循环矩阵。

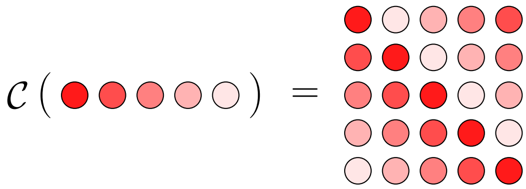


图 6.1: 循环矩阵示意图

**解.** 向量 $\boldsymbol{x}$ 对应的循环矩阵为

$$\mathcal{C}(\boldsymbol{x})=\begin{bmatrix}x_1 & x_5 & x_4 & x_3 & x_2\\ x_2 & x_1 & x_5 & x_4 & x_3\\ x_3 & x_2 & x_1 & x_5 & x_4\\ x_4 & x_3 & x_2 & x_1 & x_5\\ x_5 & x_4 & x_3 & x_2 & x_1\end{bmatrix}\in\mathbb{R}^{5\times 5} \tag{6.19}$$

图*6.1*直观描述了循环矩阵的构造规则。

**例 36.** 给定任意向量 $\boldsymbol{x}=(x_1,x_2,\cdots,x_T)^\top\in\mathbb{R}^T$ 与 $\boldsymbol{y}=(y_1,y_2,\cdots,y_T)^\top\in\mathbb{R}^T$，若两者之间的循环卷积 *(circular convolution)* 为 $\boldsymbol{z}=\boldsymbol{x}\star\boldsymbol{y}\in\mathbb{R}^T$，其中，符号 $\star$ 表示卷积运算，则向量 $\boldsymbol{z}$ 的任意元素为

$$z_t=\sum_{k=1}^{T}x_{t-k+1}y_k,\ \forall t\in\{1,2,\ldots,T\} \tag{6.20}$$

其中，当 $t+1\leq k$ 时，则令 $x_{t-k+1}=x_{t-k+1+T}$。试根据循环矩阵的定义写出循环卷积。

**解.** 在这里，循环卷积可写作如下形式：

$$\boldsymbol{z}=\boldsymbol{x}\star\boldsymbol{y}=\begin{bmatrix}x_1y_1+x_Ty_2+\cdots+x_2y_T\\ x_2y_1+x_1y_2+\cdots+x_3y_T\\ \vdots\\ x_Ty_1+x_{T-1}y_2+\cdots+x_1y_T\end{bmatrix}=\begin{bmatrix}x_1 & x_T & \cdots & x_2\\ x_2 & x_1 & \cdots & x_3\\ \vdots & \vdots & \ddots & \vdots\\ x_T & x_{T-1} & \cdots & x_1\end{bmatrix}\begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_T\end{bmatrix}=\mathcal{C}(\boldsymbol{x})\boldsymbol{y} \tag{6.21}$$

### 6.2.2 循环矩阵核范数

在线性代数中，矩阵的核范数为奇异值之和。对于任意矩阵 $\boldsymbol{X}\in\mathbb{R}^{M\times N}$，其奇异值分解为

$$\boldsymbol{X}=\sum_{r=1}^{\min\{M,N\}}\sigma_r\boldsymbol{u}_r\boldsymbol{v}_r^\top \tag{6.22}$$

其中，奇异值为 $\sigma_1\geq\sigma_2\geq\cdots\geq\sigma_{\min\{M,N\}}$；矩阵的核范数为

$$\|\boldsymbol{X}\|_*=\sum_{r=1}^{\min\{M,N\}}\sigma_r \tag{6.23}$$

给定向量 $\boldsymbol{x}\in\mathbb{R}^T$，其循环矩阵为

$$\mathcal{C}(\boldsymbol{x})=\begin{bmatrix}x_1 & x_T & x_{T-1} & \cdots & x_2\\ x_2 & x_1 & x_T & \cdots & x_3\\ x_3 & x_2 & x_1 & \cdots & x_4\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ x_T & x_{T-1} & x_{T-2} & \cdots & x_1\end{bmatrix}\in\mathbb{R}^{T\times T} \tag{6.24}$$

对该循环矩阵进行特征值分解，有

$$\mathcal{C}(\boldsymbol{x})=\boldsymbol{U}\operatorname{diag}(\mathcal{F}(\boldsymbol{x}))\boldsymbol{U}^H \tag{6.25}$$

其中，$\boldsymbol{U}\in\mathbb{C}^{T\times T}$ 为酉矩阵 (unitary matrix)；$\cdot^H$ 表示共轭转置 (conjugate transpose)。

因此，循环矩阵的核范数可写作如下形式：

$$\begin{aligned}\|\mathcal{C}(\boldsymbol{x})\|_* =&\|\boldsymbol{U}\operatorname{diag}(\mathcal{F}(\boldsymbol{x}))\boldsymbol{U}^H\|_* \\ =&\|\operatorname{diag}(\mathcal{F}(\boldsymbol{x}))\|_* \\ =&\|\mathcal{F}(\boldsymbol{x})\|_1\end{aligned} \tag{6.26}$$

由此可见，循环矩阵的核范数可转化为离散傅立叶变换的 $\ell_1$ 范数。在这里，快速傅立叶变换的计算复杂度为 $\mathcal{O}(T\log T)$，可大大提高求解循环矩阵的核范数最小化问题的计算效率。

**例 37.** 给定向量 $\boldsymbol{x}=(0,1,2,3,4)^\top$，试写出循环矩阵 $\mathcal{C}(\boldsymbol{x})$ 的奇异值与 $\|\mathcal{F}(\boldsymbol{x})\|_1$。

**解.** 根据循环矩阵定义，有

$$\mathcal{C}(\boldsymbol{x})=\begin{bmatrix}0&4&3&2&1\\1&0&4&3&2\\2&1&0&4&3\\3&2&1&0&4\\4&3&2&1&0\end{bmatrix} \tag{6.27}$$

对其进行奇异值分解，奇异值构成的向量为

$$\boldsymbol{\sigma}=(10,4.25,4.25,2.63,2.63)^\top \tag{6.28}$$

另外，直接对向量 $\boldsymbol{x}$ 作离散傅立叶变换，有

$$\mathcal{F}(\boldsymbol{x})=(10,-2.5+3.44i,-2.5+0.81i,-2.5-0.81i,-2.5-3.44i)^\top \tag{6.29}$$

其中，$i=\sqrt{-1}$ 表示复数的虚部。

由此，可得到

$$\|\mathcal{F}(\boldsymbol{x})\|_1=10+2\sqrt{2.5^2+3.44^2}+2\sqrt{2.5^2+0.81^2}=10+8.50+5.26=23.76 \tag{6.30}$$

### 6.2.3 $\ell_1$ 范数最小化问题

一般而言，假设图优化问题的目标函数为 $f(x)=g(x)+h(x)$ 是由 $g(x)$ 与 $h(x)$ 叠加而成的，其中，限定 $g(x)$ 是不可微的凸函数、$h(x)$ 是可微的凸函数，则这类优化问题可通过近端梯度下降法 (proximal gradient descent) 进行求解。

对于向量 $\boldsymbol{x}$，令 $g(\boldsymbol{x})=\|\boldsymbol{x}\|_1$ 表示不可微函数[1]、$h(\boldsymbol{x})=\frac{\lambda}{2}\|\boldsymbol{x}-\boldsymbol{w}\|_2^2$（$\boldsymbol{w}$ 已知）表示可微函数，则 $\ell_1$ 范数最小化问题可归纳为

$$\min_{\boldsymbol{x}}\ \|\boldsymbol{x}\|_1+\frac{\lambda}{2}\|\boldsymbol{x}-\boldsymbol{w}\|_2^2 \tag{6.31}$$

在 $\ell_1$ 范数最小化问题中，若 $\boldsymbol{w}\in\mathbb{R}^T$ 已知、$\boldsymbol{x}\in\mathbb{R}^T$ 未知，则近端算子 (proximal operator) 可写作如下的软阈值函数 (soft thresholding)：

$$\mathcal{S}_{1/\lambda}(\boldsymbol{w})=\arg\min_{\boldsymbol{x}}\ \|\boldsymbol{x}\|_1+\frac{\lambda}{2}\|\boldsymbol{x}-\boldsymbol{w}\|_2^2 \tag{6.32}$$

[1] 对于任意向量 $\boldsymbol{x}\in\mathbb{R}^T$，其 $\ell_1$ 范数为 $\|\boldsymbol{x}\|_1=\sum_{t=1}^{T}|x_t|$，即元素绝对值之和。

对于向量 $\boldsymbol{x}$ 中的任意元素 $x_t$，有

$$x_t := \mathcal{S}_{1/\lambda}(w_i) = \begin{cases} w_i - 1/\lambda & \text{if } w_i > 1/\lambda \\ w_i + 1/\lambda & \text{if } w_i \leq -1/\lambda \\ 0 & \text{otherwise} \end{cases} \tag{6.33}$$

在写法上，软阈值函数可进一步写作如下形式[2]：

$$x_t := \frac{w_t}{|w_t|} \cdot \max\{0, |w_t| - 1/\lambda\},\, t = 1, 2, \ldots, T \tag{6.34}$$

故向量 $\boldsymbol{x}$ 的解析解为

$$\boldsymbol{x} := \mathcal{S}_{1/\lambda}(\boldsymbol{w}) \tag{6.35}$$

其中，$\mathcal{S}_{1/\lambda}(\cdot)$ 表示超参数为 $\lambda$ 得软阈值函数；$\max\{x, y\}$ 表示取 $x$ 与 $y$ 之间较大的数值。

### 6.2.4 循环矩阵核范数最小化问题

对于任意观测向量 $\boldsymbol{y} \in \mathbb{R}^T$，若被观测元素的索引集合为 $\Omega$，则循环矩阵核范数最小化问题 [Liu and Zhang, 2022] 可描述为

$$\begin{aligned} \min_{\boldsymbol{x}} \;\; & \|\mathcal{C}(\boldsymbol{x})\|_* \\ \text{s.t.} \;\; & \|\mathcal{P}_\Omega(\boldsymbol{x} - \boldsymbol{y})\|_2 \leq \epsilon \end{aligned} \tag{6.36}$$

其中，约束条件中的 $\epsilon$ 表示容许误差。为便于求解，可将上述核范数最小化问题中的约束条件进行改写，令约束条件作为目标函数的正则项，则构造出来的优化问题为

$$\begin{aligned} \min_{\boldsymbol{x}, \boldsymbol{z}} \;\; & \|\mathcal{C}(\boldsymbol{x})\|_* + \frac{\eta}{2}\|\mathcal{P}_\Omega(\boldsymbol{z} - \boldsymbol{y})\|_2^2 \\ \text{s.t.} \;\; & \boldsymbol{x} = \boldsymbol{z} \end{aligned} \tag{6.37}$$

其中，$\eta$ 为正则项的权重系数。由于观测向量存在缺失值，为便于求解优化问题，可令中间变量的等价关系作为约束条件。

通常来说，求解上述优化问题可采用 ADMM 算法。使用 ADMM 求解过程中，需首先构造增广拉格朗日函数，即

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{w}) = \|\mathcal{C}(\boldsymbol{x})\|_* + \frac{\lambda}{2}\|\boldsymbol{x} - \boldsymbol{z}\|_2^2 + \langle \boldsymbol{w}, \boldsymbol{x} - \boldsymbol{z} \rangle + \frac{\eta}{2}\|\mathcal{P}_\Omega(\boldsymbol{z} - \boldsymbol{y})\|_2^2 \tag{6.38}$$

其中，$\boldsymbol{w} \in \mathbb{R}^T$ 为拉格朗日乘子；$\lambda$ 为权重系数。符号 $\langle :, : \rangle$ 表示内积 (inner product)，满足如下关系：

$$\langle \boldsymbol{w}, \boldsymbol{x} - \boldsymbol{z} \rangle = \boldsymbol{w}^\top (\boldsymbol{x} - \boldsymbol{z}) \in \mathbb{R} \tag{6.39}$$

然后，可采用如下 ADMM 算法：

$$\begin{cases} \boldsymbol{x} := \arg\min\limits_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{w}) \\ \boldsymbol{z} := \arg\min\limits_{\boldsymbol{z}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{w}) \\ \boldsymbol{w} := \boldsymbol{w} + \lambda(\boldsymbol{x} - \boldsymbol{z}) \end{cases} \tag{6.40}$$

[2]这种写法可确保当 $\boldsymbol{w}, \boldsymbol{x} \in \mathbb{C}^T$ 时，软阈值函数依然适用。

对于变量 $\boldsymbol{x}$，有

$$\begin{aligned}
\boldsymbol{x} :=& \arg\min_{\boldsymbol{x}} \ \|\mathcal{C}(\boldsymbol{x})\|_* + \frac{\lambda}{2}\|\boldsymbol{x}-\boldsymbol{z}\|_2^2 + \langle \boldsymbol{w}, \boldsymbol{x}-\boldsymbol{z}\rangle \\
=& \arg\min_{\boldsymbol{x}} \ \|\mathcal{C}(\boldsymbol{x})\|_* + \frac{\lambda}{2}\boldsymbol{x}^\top\boldsymbol{x} - \lambda\langle \boldsymbol{z}, \boldsymbol{x}\rangle + \langle \boldsymbol{w}, \boldsymbol{x}\rangle \\
=& \arg\min_{\boldsymbol{x}} \ \|\mathcal{C}(\boldsymbol{x})\|_* + \frac{\lambda}{2}\boldsymbol{x}^\top\boldsymbol{x} - \lambda\langle \boldsymbol{z}-\boldsymbol{w}/\lambda, \boldsymbol{x}\rangle \\
=& \arg\min_{\boldsymbol{x}} \ \|\mathcal{C}(\boldsymbol{x})\|_* + \frac{\lambda}{2}\|\boldsymbol{x}-\boldsymbol{z}+\boldsymbol{w}/\lambda\|_2^2 \\
=& \arg\min_{\boldsymbol{x}} \ \|\mathcal{F}(\boldsymbol{x})\|_1 + \frac{\lambda}{2T}\|\mathcal{F}(\boldsymbol{x}-\boldsymbol{z}+\boldsymbol{w}/\lambda)\|_2^2
\end{aligned} \tag{6.41}$$

令

$$\boldsymbol{h} = \boldsymbol{z} - \boldsymbol{w}/\lambda \tag{6.42}$$

若 $\{\hat{\boldsymbol{x}}, \hat{\boldsymbol{h}}\} = \{\mathcal{F}(\boldsymbol{x}), \mathcal{F}(\boldsymbol{h})\}$ 记作离散傅立叶变换之后的变量，根据公式(6.32)中给出的软阈值过程，则变量 $\hat{\boldsymbol{x}}$ 的解析解为

$$\hat{x}_t = \frac{\hat{h}_t}{|\hat{h}_t|}\cdot\max\{0, |\hat{h}_t| - T/\lambda\}, t = 1, 2, \ldots, T \tag{6.43}$$

因此，通过离散傅立叶逆变换，则变量 $\boldsymbol{x}$ 的更新公式为

$$\boldsymbol{x} := \mathcal{F}^{-1}(\hat{\boldsymbol{x}}) \tag{6.44}$$

**例 38.** 现有循环矩阵核范数最小化问题为

$$\min_{\boldsymbol{x}} \ \|\mathcal{C}(\boldsymbol{x})\|_* + \frac{\lambda}{2}\|\boldsymbol{x}-\boldsymbol{z}\|_2^2 \tag{6.45}$$

其中，$\boldsymbol{z} \in \mathbb{R}^T$ 为已知变量；$\boldsymbol{x} \in \mathbb{R}^T$ 为待优化变量。

通常来说，可将公式(6.45)中的优化问题写作如下形式：

$$\min_{\boldsymbol{x}} \ \|\mathcal{F}(\boldsymbol{x})\|_1 + \frac{\lambda}{2T}\|\mathcal{F}(\boldsymbol{x}-\boldsymbol{z})\|_2^2 \tag{6.46}$$

其中，$\mathcal{F}(\cdot)$ 表示离散傅立叶变换。令

$$\hat{\boldsymbol{h}} = \mathcal{F}(\boldsymbol{z}) \tag{6.47}$$

则该 $\ell_1$ 范数最小化问题的解析解为

$$\boldsymbol{x} := \mathcal{F}^{-1}(\hat{\boldsymbol{x}}) \qquad \hat{x}_t = \frac{\hat{h}_t}{|\hat{h}_t|}\cdot\max\{0, |\hat{h}_t| - T/\lambda\}, t = 1, 2, \ldots, T \tag{6.48}$$

其中，$\mathcal{F}^{-1}(\cdot)$ 表示离散傅立叶逆变换。

不妨令 $\boldsymbol{z} = (0, 1, 2, 3, 4)^\top$ 与 $\lambda = 2$，试写出公式(6.45)的最优解。

**解.** 对已知变量 $\boldsymbol{z}$ 进行离散傅立叶变换，有

$$\hat{\boldsymbol{h}} = \mathcal{F}(\boldsymbol{z}) = (10, -2.5+3.44i, -2.5+0.81i, -2.5-0.81i, -2.5-3.44i)^\top \tag{6.49}$$

其中，$i = \sqrt{-1}$ 表示复数的虚部。向量 $\hat{\boldsymbol{h}}$ 的绝对值为

$$|\hat{\boldsymbol{h}}| = (10, 4.25, 2.63, 2.63, 4.25)^\top \tag{6.50}$$

根据公式(6.48)，可得到

$$\boldsymbol{x} = (1.04, 0.86, 1.5, 2.14, 1.96)^\top \tag{6.51}$$

在这里，公式(6.45)中的优化问题目标函数为

$$\|\mathcal{C}(\boldsymbol{x})\|_* + \frac{\lambda}{2}\|\boldsymbol{x}-\boldsymbol{z}\|_2^2 = 17.51 \tag{6.52}$$

对于变量 $\boldsymbol{z}$，可分别对 $\mathcal{P}_{\Omega}(\boldsymbol{z})$ 与 $\mathcal{P}_{\Omega}^{\perp}(\boldsymbol{z})$ 求偏导数：

$$\begin{cases}\dfrac{\partial\mathcal{L}(\boldsymbol{x},\boldsymbol{z},\boldsymbol{w})}{\partial\mathcal{P}_{\Omega}(\boldsymbol{z})} = \lambda\mathcal{P}_{\Omega}(\boldsymbol{z}-\boldsymbol{x}-\boldsymbol{w}/\lambda)+\eta\mathcal{P}_{\Omega}(\boldsymbol{z}-\boldsymbol{y}) \\ \dfrac{\partial\mathcal{L}(\boldsymbol{x},\boldsymbol{z},\boldsymbol{w})}{\partial\mathcal{P}_{\Omega}^{\perp}(\boldsymbol{z})} = \lambda\mathcal{P}_{\Omega}^{\perp}(\boldsymbol{z}-\boldsymbol{x}-\boldsymbol{w}/\lambda)\end{cases} \tag{6.53}$$

令偏导数为 $\mathbf{0}$，则变量 $\boldsymbol{z}$ 的解析解为

$$\boldsymbol{z} := \frac{1}{\lambda+\eta}\mathcal{P}_{\Omega}(\lambda\boldsymbol{x}+\boldsymbol{w}+\eta\boldsymbol{y})+\mathcal{P}_{\Omega}^{\perp}(\boldsymbol{x}+\boldsymbol{w}/\lambda) \tag{6.54}$$

算法3给出了循环矩阵核范数最小化算法的具体实现过程。

**Algorithm 3** 循环矩阵核范数最小化算法

**Input:** 观测向量 $\boldsymbol{y}\in\mathbb{R}^{T}$，被观测元素的索引集合 $\Omega$，超参数 $\{\lambda,\eta\}$。

**Output:** 重构出来的向量 $\boldsymbol{x}\in\mathbb{R}^{T}$。

对变量 $\{\boldsymbol{x},\boldsymbol{z},\boldsymbol{w}\}$ 进行初始化；
**for** $i=0$ to 最大迭代次数 **do**
  对变量 $\{\boldsymbol{z},\boldsymbol{w}\}$ 进行快速傅立叶变换；
  根据公式(6.42)计算 $\boldsymbol{h}$；
  根据公式(6.43)计算 $\hat{\boldsymbol{x}}$；
  令 $\boldsymbol{x}:=\mathcal{F}^{-1}(\hat{\boldsymbol{x}})$；
  根据公式(6.54)计算 $\boldsymbol{z}$；
  计算 $\boldsymbol{w}:=\boldsymbol{w}+\lambda(\boldsymbol{x}-\boldsymbol{z})$；
**end for**

## 6.3 低秩拉普拉斯卷积模型

### 6.3.1 拉普拉斯卷积核

一般而言，对关系型数据进行建模时，可对数据之间的关联构造拉普拉斯矩阵。若 $\boldsymbol{D}\in\mathbb{R}^{T\times T}$ 表示度矩阵 (degree matrix)、$\boldsymbol{A}\in\mathbb{R}^{T\times T}$ 表示邻接矩阵 (adjacency matrix)，则对应的拉普拉斯矩阵为

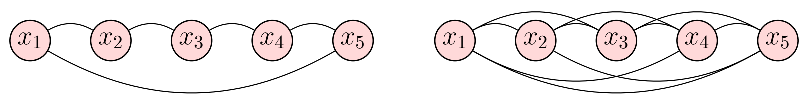
$$\boldsymbol{L}=\boldsymbol{D}-\boldsymbol{A} \tag{6.55}$$

举例来说，图6.2(a)中的图模型为两两相结、首尾相连的循环图，对应的图拉普拉斯矩阵为

$$\boldsymbol{L}=\begin{bmatrix}2 & -1 & 0 & 0 & -1\\ -1 & 2 & -1 & 0 & 0\\ 0 & -1 & 2 & -1 & 0\\ 0 & 0 & -1 & 2 & -1\\ -1 & 0 & 0 & -1 & 2\end{bmatrix} \tag{6.56}$$

图6.2(b)对应的图拉普拉斯矩阵为

$$\boldsymbol{L}=\begin{bmatrix}4 & -1 & -1 & -1 & -1\\ -1 & 4 & -1 & -1 & -1\\ -1 & -1 & 4 & -1 & -1\\ -1 & -1 & -1 & 4 & -1\\ -1 & -1 & -1 & -1 & 4\end{bmatrix} \tag{6.57}$$




(a) 度为 2 的循环图 (b) 度为 4 的循环图

图 6.2: 基于关系型数据 $\{x_1, x_2, \ldots, x_5\}$ 的无向循环图

由于这两个拉普拉斯矩阵均为循环矩阵，不妨令其第一列为拉普拉斯核 (Laplacian kernel)，则两者的拉普拉斯核分别为

$$\boldsymbol{\ell} = (2, -1, 0, 0, -1)^\top \qquad \boldsymbol{\ell} = (4, -1, -1, -1, -1)^\top \tag{6.58}$$

由此，对于任意向量 $\boldsymbol{x} = (x_1, x_2, \cdots, x_T)^\top \in \mathbb{R}^T$，令 $\tau \in \mathbb{N}^+$ 为拉普拉斯核的超参数，其中，$\tau \leq \frac{1}{2}(T-1)$，则拉普拉斯核可被定义 [Chen et al., 2022b] 为

$$\boldsymbol{\ell} = (2\tau, \underbrace{-1, \cdots, -1}_{\tau}, 0, \cdots, 0, \underbrace{-1, \cdots, -1}_{\tau})^\top \in \mathbb{R}^T \tag{6.59}$$

该拉普拉斯核同时也是拉普拉斯矩阵的第一列。拉普拉斯核的第一个元素为 $2\tau$，表示拉普拉斯矩阵的度。

### 6.3.2 拉普拉斯时序正则

在时间序列中，若数据向量为 $\boldsymbol{x} = (x_1, x_2, \cdots, x_T)^\top \in \mathbb{R}^T$，则基于拉普拉斯矩阵的时序正则项为

$$\mathcal{R}_\tau(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{L}\boldsymbol{x}\|_2^2 \tag{6.60}$$

根据 Parseval 定理，可得到基于拉普拉斯核的时序正则项为

$$\mathcal{R}_\tau(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{\ell} \star \boldsymbol{x}\|_2^2 = \frac{1}{2T}\|\mathcal{F}(\boldsymbol{\ell}) \circ \mathcal{F}(\boldsymbol{x})\|_2^2 \tag{6.61}$$

**例 39.** *试根据 Parseval 定理证明公式(6.61)。*

**解.** 在公式(6.61)中，不妨令

$$\begin{cases} \boldsymbol{\alpha} = \boldsymbol{\ell} \star \boldsymbol{x} \\ \boldsymbol{\beta} = \mathcal{F}(\boldsymbol{\ell}) \circ \mathcal{F}(\boldsymbol{x}) \end{cases} \tag{6.62}$$

根据卷积定理，有

$$\mathcal{F}(\boldsymbol{\alpha}) = \boldsymbol{\beta} \tag{6.63}$$

再根据 *Parseval* 定理，则

$$\|\boldsymbol{\alpha}\|_2^2 = \frac{1}{T}\|\mathcal{F}(\boldsymbol{\alpha})\|_2^2 = \frac{1}{T}\|\boldsymbol{\beta}\|_2^2 \tag{6.64}$$

由此，公式(6.61)得证。

**例 40.** *对于时间序列 $\boldsymbol{x} = (x_1, x_2, \cdots, x_T)^\top \in \mathbb{R}^T$ 而言，基于拉普拉斯核的时序正则项可对时间序列的局部趋势 (local trend) 进行建模，以拉普拉斯核 $\boldsymbol{\ell} = (2, -1, 0, \cdots, 0\ \ -1)^\top \in \mathbb{R}^T$ 为例，试解释基于拉普拉斯核的时序正则项在局部趋势建模中所起的作用。*

**解.** 根据定义，时序正则项 $\mathcal{R}_\tau(\boldsymbol{x})$ 可写作如下形式：

$$\mathcal{R}_\tau(\boldsymbol{x}) = \frac{1}{2}(2x_1 - (x_2 + x_T))^2 + \frac{1}{2}(2x_2 - (x_3 + x_1))^2 + \cdots + \frac{1}{2}(2x_T - (x_1 + x_{T-1}))^2 \tag{6.65}$$

由此可见，该正则项实际上对时间序列进行了平滑处理。

上述定义的拉普拉斯卷积核对于有向图依然适用,若令拉普拉斯卷积核为 $\boldsymbol{\ell}=(1,-1,0,\cdots,0)^\top$,则相应的拉普拉斯矩阵 [Takayama and Yokota, 2022] 为

$$\boldsymbol{L}=\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & -1 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \tag{6.66}$$

除去拉普拉斯矩阵的第一行,可构造如下矩阵:

$$\boldsymbol{\Psi}=\begin{bmatrix}\boldsymbol{0}_{(T-1)\times 1} & \boldsymbol{I}_{T-1}\end{bmatrix}-\begin{bmatrix}\boldsymbol{I}_{T-1} & \boldsymbol{0}_{(T-1)\times 1}\end{bmatrix} \tag{6.67}$$

由此,相应的时序正则项为 $\frac{1}{2}\|\boldsymbol{\Psi}\boldsymbol{x}\|_2^2$,该正则项对时间序列建模时可起到平滑的作用。

### 6.3.3 一维低秩拉普拉斯卷积模型

在时间序列缺失值重构任务中,对全局趋势 (global trend) 与局部趋势建模往往缺一不可。对于任意时间序列 $\boldsymbol{x}=(x_1,x_2,\cdots,x_T)^\top\in\mathbb{R}^T$ 而言,可采用循环矩阵核范数捕捉低秩信息、借助拉普拉斯时序正则项刻画局部趋势(如图6.3所示),由此,得到的低秩拉普拉斯卷积模型的目标函数兼具循环矩阵核范数与循环卷积 [Chen et al., 2022b],即

$$\begin{aligned} \min_{\boldsymbol{x}}\ & \|\mathcal{C}(\boldsymbol{x})\|_*+\frac{\gamma}{2}\|\boldsymbol{\ell}\star\boldsymbol{x}\|_2^2 \\ \text{s.t. } & \|\mathcal{P}_\Omega(\boldsymbol{x}-\boldsymbol{y})\|_2\leq\epsilon \end{aligned} \tag{6.68}$$

其中,$\boldsymbol{\ell}\in\mathbb{R}^T$ 为表征时序关联的拉普拉斯卷积核;$\gamma$ 为拉普拉斯时序正则项的权重系数;约束条件中的 $\epsilon$ 表示容许误差。

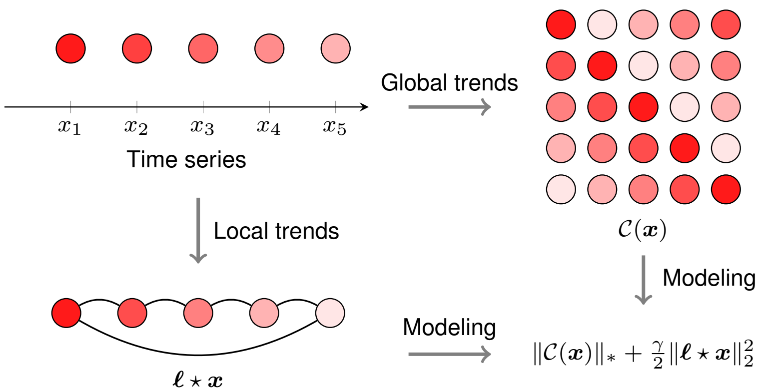



图 6.3: 低秩拉普拉斯卷积模型的示意图

为便于求解,可将上述优化问题中的约束条件进行改写,令约束条件作为目标函数的正则项,则构造出来的优化问题为

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{z}}\ & \|\mathcal{C}(\boldsymbol{x})\|_*+\frac{\gamma}{2}\|\boldsymbol{\ell}\star\boldsymbol{x}\|_2^2+\frac{\eta}{2}\|\mathcal{P}_\Omega(\boldsymbol{z}-\boldsymbol{y})\|_2^2 \\ \text{s.t. } & \boldsymbol{x}=\boldsymbol{z} \end{aligned} \tag{6.69}$$

其中,$\eta$ 为正则项的权重系数。由于观测向量存在缺失值,为便于求解优化问题,可令中间变量的等价关系作为约束条件。

使用 ADMM 算法求解时，需首先构造增广拉格朗日函数，即

$$\mathcal{L}(\boldsymbol{x},\boldsymbol{z},\boldsymbol{w})=\|\mathcal{C}(\boldsymbol{x})\|_*+\frac{\gamma}{2}\|\boldsymbol{\ell}\star\boldsymbol{x}\|_2^2+\frac{\lambda}{2}\|\boldsymbol{x}-\boldsymbol{z}\|_2^2+\langle\boldsymbol{w},\boldsymbol{x}-\boldsymbol{z}\rangle+\frac{\eta}{2}\|\mathcal{P}_\Omega(\boldsymbol{z}-\boldsymbol{y})\|_2^2 \tag{6.70}$$

对于变量 $\boldsymbol{x}$，子问题为

$$\begin{aligned}\boldsymbol{x}:=&\arg\min_{\boldsymbol{x}}\ \|\mathcal{C}(\boldsymbol{x})\|_*+\frac{\gamma}{2}\|\boldsymbol{\ell}\star\boldsymbol{x}\|_2^2+\frac{\lambda}{2}\|\boldsymbol{x}-\boldsymbol{z}+\boldsymbol{w}/\lambda\|_2^2\\=&\arg\min_{\boldsymbol{x}}\ \|\mathcal{F}(\boldsymbol{x})\|_1+\frac{\gamma}{2T}\|\mathcal{F}(\boldsymbol{\ell})\circ\mathcal{F}(\boldsymbol{x})\|_2^2+\frac{\lambda}{2T}\|\mathcal{F}(\boldsymbol{x}-\boldsymbol{z}+\boldsymbol{w}/\lambda)\|_2^2\end{aligned} \tag{6.71}$$

若 $\{\hat{\boldsymbol{\ell}},\hat{\boldsymbol{x}},\hat{\boldsymbol{z}},\hat{\boldsymbol{w}}\}=\{\mathcal{F}(\boldsymbol{\ell}),\mathcal{F}(\boldsymbol{x}),\mathcal{F}(\boldsymbol{z}),\mathcal{F}(\boldsymbol{w})\}$，令子问题的正则项为

$$f=\frac{\gamma}{2T}\|\mathcal{F}(\boldsymbol{\ell})\circ\mathcal{F}(\boldsymbol{x})\|_2^2+\frac{\lambda}{2T}\|\mathcal{F}(\boldsymbol{x}-\boldsymbol{z}+\boldsymbol{w}/\lambda)\|_2^2 \tag{6.72}$$

相应地，对于变量 $\hat{\boldsymbol{x}}$，函数 $f$ 的偏导数为

$$\begin{aligned}\frac{\partial f}{\partial\hat{\boldsymbol{x}}}=&\frac{\gamma}{T}\hat{\boldsymbol{\ell}}\circ\hat{\boldsymbol{\ell}}\circ\hat{\boldsymbol{x}}+\frac{\lambda}{T}(\hat{\boldsymbol{x}}-\hat{\boldsymbol{z}}+\hat{\boldsymbol{w}}/\lambda)\\=&\frac{1}{T}(\gamma\hat{\boldsymbol{\ell}}\circ\hat{\boldsymbol{\ell}}+\lambda\mathbb{1}_T)\circ\hat{\boldsymbol{x}}-\frac{1}{T}(\lambda\hat{\boldsymbol{z}}-\hat{\boldsymbol{w}})\end{aligned} \tag{6.73}$$

其中，向量 $\mathbb{1}_T\in\mathbb{R}^T$ 的所有元素均为 1。

不妨定义

$$\hat{\boldsymbol{h}}=(\lambda\hat{\boldsymbol{z}}-\hat{\boldsymbol{w}})\oslash(\gamma\hat{\boldsymbol{\ell}}\circ\hat{\boldsymbol{\ell}}+\lambda\mathbb{1}_T) \tag{6.74}$$

对应于 $\frac{\partial f}{\partial\hat{\boldsymbol{x}}}=\mathbf{0}$。

由此，变量 $\hat{\boldsymbol{x}}$ 的解析解可根据公式(6.43)计算得到，通过离散傅立叶逆变换，变量 $\boldsymbol{x}$ 的更新公式为 $\boldsymbol{x}:=\mathcal{F}^{-1}(\hat{\boldsymbol{x}})$。

在 ADMM 算法中，变量 $\boldsymbol{z}$ 的解析解与公式(6.54)一致；变量 $\boldsymbol{w}$ 的更新公式参见公式(6.40)。算法4给出了一维低秩拉普拉斯卷积模型的具体实现过程。

**Algorithm 4** 一维低秩拉普拉斯卷积模型

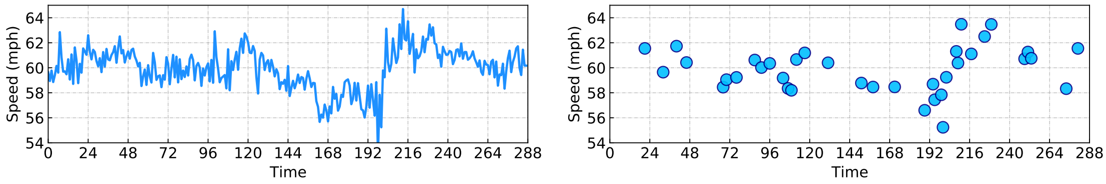
**Input:** 观测向量 $\boldsymbol{y}\in\mathbb{R}^T$，被观测元素的索引集合 $\Omega$，超参数 $\{\gamma,\lambda,\eta,\tau\}$。

**Output:** 重构出来的向量 $\boldsymbol{x}\in\mathbb{R}^T$。

```
1: 对变量 {x, z, w} 进行初始化；
2: for i = 0 to 最大迭代次数 do
3:    对变量 {z, w} 进行快速傅立叶变换；
4:    根据公式(6.74)计算 ĥ；
5:    根据公式(6.43)计算 x̂；
6:    令 x := F^{-1}(x̂)；
7:    根据公式(6.54)计算 z；
8:    计算 w := w + λ(x − z)；
9: end for
```

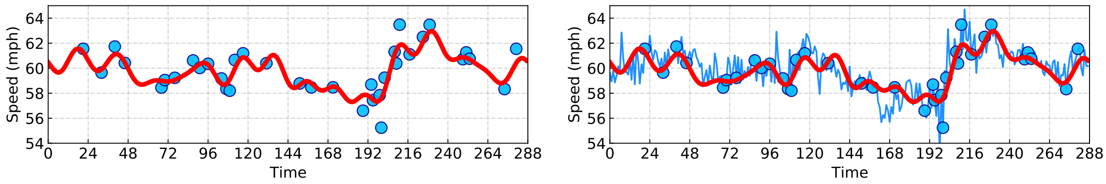
**例 41.** *给定某高速公路断面交通流的车速时间序列如图6.4(a)所示，采集数据的时间粒度为 15 分钟，即每天预期可获取 96 个观测值；采集时长为 3 天，预期产生 288 个观测值，即 $T=288$。现假设该车速时间序列存在 90% 的缺失值，如图6.4(b)所示，试使用一维低秩拉普拉斯卷积模型对部分观测的车速时间序列进行重构、修复缺失值。*

**解.** *在图6.5所示的重构时间序列中，将一维低秩拉普拉斯卷积模型中的超参数设置为 $\lambda=5\times10^{-3}T,\gamma=2\lambda,\eta=100\lambda,\tau=2$。从中不难发现，该模型重构出来的时间序列与真实时间序列趋势吻合。*




(a) 原始车速时间序列 (b) 随机抽取的 36 个观测值

图 6.4: 某高速公路断面交通流的车速时间序列，其中，蓝色曲线表示车速时间序列；蓝色圆圈表示抽取的部分观测值。




(a) 重构出来的车速时间序列 (b) 重构出来的车速时间序列与原始车速时间序列对比

图 6.5: 基于一维低秩拉普拉斯卷积模型的 90% 缺失率的车速时间序列重构，其中，红色曲线表示重构出来的车速时间序列。

### 6.3.4 二维低秩拉普拉斯卷积模型

## 6.4 Python 实现代码

Table of functions

# 第七章　基于延迟嵌套的张量分解

# 第八章　低秩深度学习时空预测模型

# 附录 A　公开交通数据集

## A.1　波特兰高速公路交通流量数据集

## A.2　西雅图高速公路交通速度数据集

# 参考文献


Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

Nicholas D Sidiropoulos, Lieven De Lathauwer, Xiao Fu, Kejun Huang, Evangelos E Papalexakis, and Christos Faloutsos. Tensor decomposition for signal processing and machine learning. *IEEE Transactions on Signal Processing*, 65(13):3551–3582, 2017.

Xinyu Chen, Chengyuan Zhang, Xi-Le Zhao, Nicolas Saunier, and Lijun Sun. Nonstationary temporal matrix factorization for multivariate time series forecasting. *arXiv preprint arXiv:2203.10651*, 2022a.

Guangcan Liu and Wayne Zhang. Recovery of future data via convolution nuclear norm minimization. *IEEE Transactions on Information Theory*, 69(1):650–665, 2022.

Xinyu Chen, Zhanhong Cheng, Nicolas Saunier, and Lijun Sun. Laplacian convolutional representation for traffic time series imputation. *arXiv preprint arXiv:2212.01529*, 2022b.

Hiromu Takayama and Tatsuya Yokota. Fast signal completion algorithm with cyclic convolutional smoothing. In *2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 364–371. IEEE, 2022.