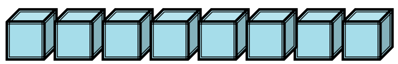
张量分解

彭毅

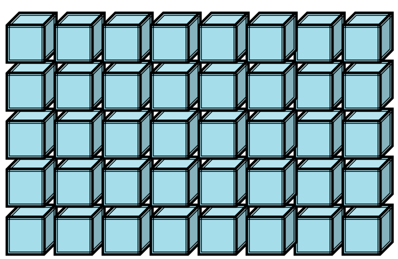
基本概念及记号

基本概念及记号

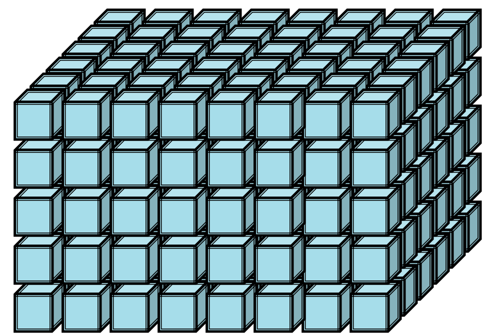
- 张量（tensor）
  - 多维数组



一阶张量
（向量）



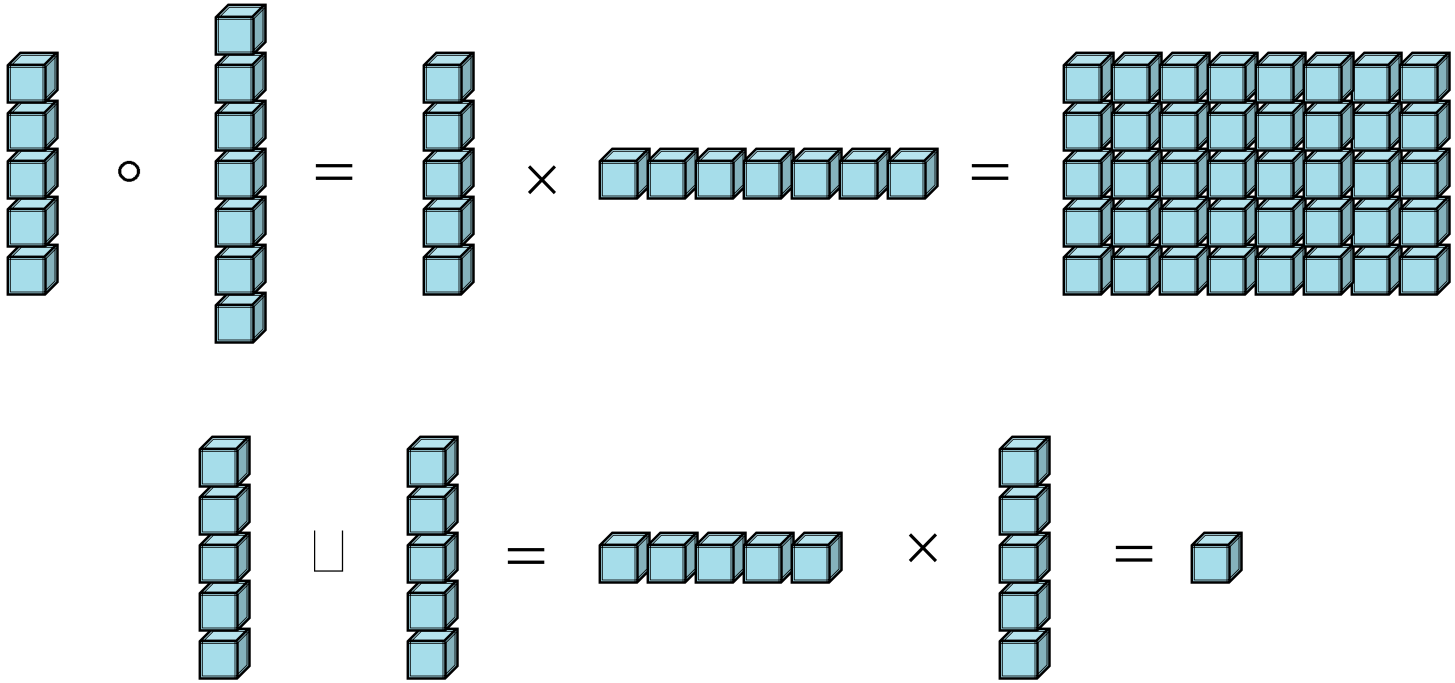
二阶张量
（矩阵）



三阶张量

基本概念及记号

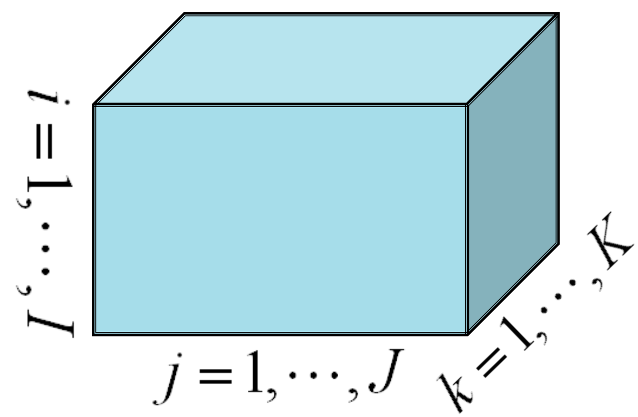
- 张量空间
  - 由若干个向量空间中的基底的外积张成的空间



向量的外积和内积

基本概念及记号

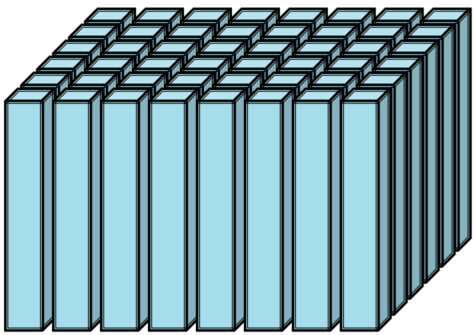
- 阶（order/ways/modes/rank）
  - 张成所属张量空间的向量空间的个数
    - 一阶张量（向量）：$\mathbf{x} = \{x_i\}$
    - 二阶张量（矩阵）：$\mathbf{X} = \{x_{ij}\}$
    - 三阶或更高阶张量：$\mathcal{X} = \{x_{ij\cdots k}\}$
    - 零阶张量（数量）：x



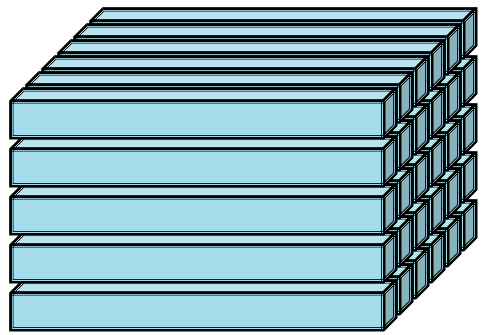
三阶张量：$\mathcal{X} \in \mathbb{R}^{I\times J\times K}$

基本概念及记号

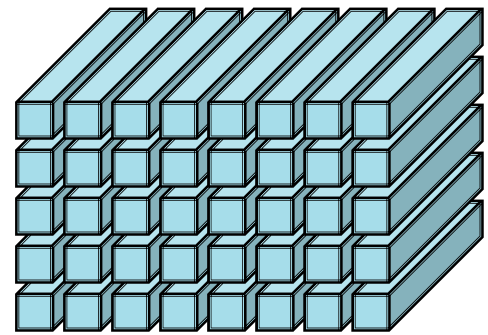
- 纤维（fiber）



mode-1 （列）
纤维：$\mathbf{x}_{:jk}$



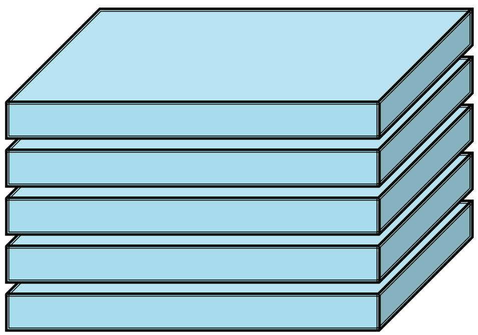
mode-2 （行）
纤维：$\mathbf{x}_{i:k}$



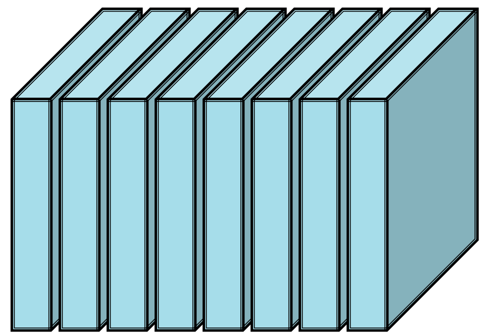
mode-3 （管）
纤维：$\mathbf{x}_{ij:}$

基本概念及记号

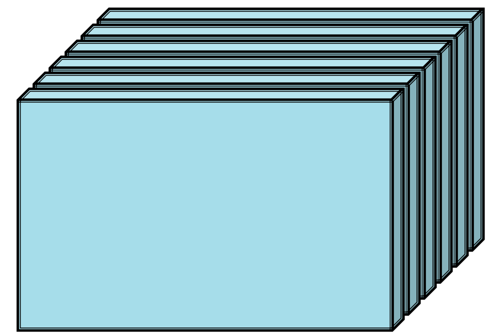
- 切片（slice）



水平切片：$\mathbf{X}_{i::}$



侧面切片：$\mathbf{X}_{:j:}$



正面切片：$\mathbf{X}_{::k}\,(\mathbf{X}_k)$

基本概念及记号

- 内积和范数
  - 设 $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$
    内积：

$$\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} x_{i_1 i_2 \cdots i_N} y_{i_1 i_2 \cdots i_N}$$

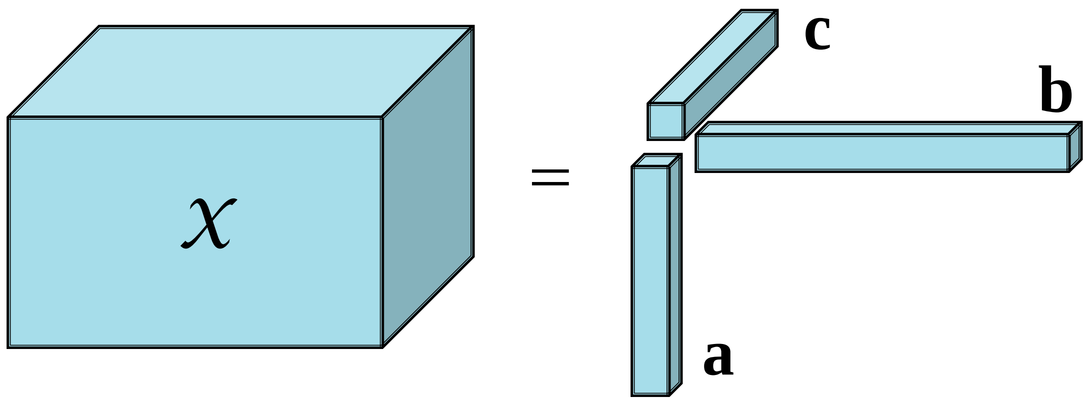
（Frobenius）范数：

$$\|\mathcal{X}\| = \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle} = \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} x_{i_1 i_2 \cdots i_N}^2}$$

基本概念及记号

- 秩一张量/可合张量
  - N阶张量 $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ 是一个秩一张量，如果它能被写成N个向量的外积，即

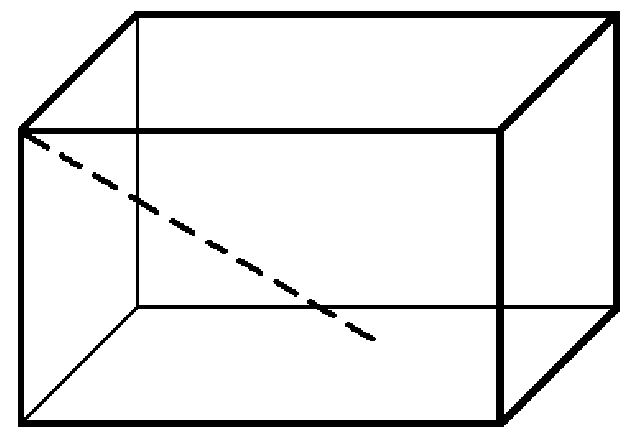
$$\mathcal{X} = \mathbf{a}^{(1)} \circ \mathbf{a}^{(2)} \circ \cdots \circ \mathbf{a}^{(N)}$$



三阶秩一张量：$\mathcal{X} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c}$

基本概念及记号

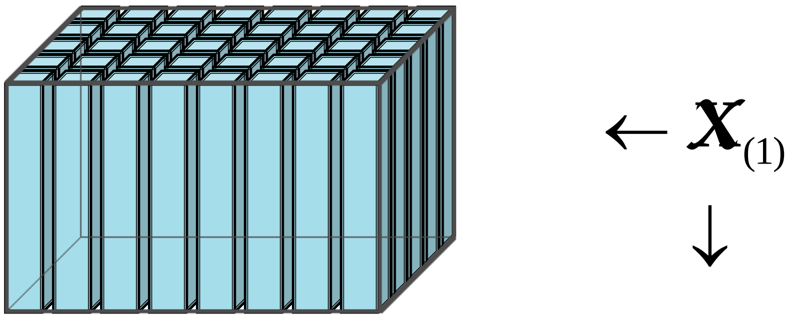
- （超）对称和（超）对角
  - 立方张量：各个mode的长度相等
  - 对称：一个立方张量是对称的，如果其元素在下标的任意排列下是常数。如一个三阶立方张量是超对称的，如果
    $$x_{ijk}=x_{ikj}=x_{jik}=x_{jki}=x_{kij}=x_{kji},\forall i,j,k$$
  - 对角：仅当 $i_1=i_2=\cdots=i_N$ 时，$x_{i_1 i_2 \cdots i_N}\neq 0$



张量的（超）对角线

基本概念及记号

- 展开（matricization/unfolding/flattening）
  - 将N阶张量 $\mathcal{X}$ 沿mode-n展开成一个矩阵 $\mathbf{X}_{(n)}$



三阶张量的mode-1展开

基本概念及记号

- n-mode（矩阵）乘积
  - 一个张量$\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$和一个矩阵$\mathbf{U} \in \mathbb{R}^{J \times I_n}$的n-mode乘积$\left(\mathcal{X} \times_n \mathbf{U}\right) \in \mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times J \times I_{n+1} \times \cdots \times I_N}$，其元素定义为

$$\left(\mathcal{X} \times_n \mathbf{U}\right)_{i_1 \cdots i_{n-1} j i_{n+1} \cdots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 i_2 \cdots i_N} u_{j i_n}$$

  - 这个定义可以写成沿mode-n展开的形式

$$\mathcal{Y} = \mathcal{X} \times_n \mathbf{U} \Leftrightarrow \mathbf{Y}_{(n)} = \mathbf{U}\mathbf{X}_{(n)}$$

  - 性质：$\mathcal{X} \times_m \mathbf{A} \times_n \mathbf{B} = \mathcal{X} \times_n \mathbf{B} \times_m \mathbf{A}, m \neq n$

$$\mathcal{X} \times_n \mathbf{A} \times_n \mathbf{B} = \mathcal{X} \times_n \left(\mathbf{B}\mathbf{A}\right)$$

基本概念及记号

- n-mode（向量）乘积
  - 一个张量$\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$和一个向量 $\mathbf{v} \in \mathbb{R}^{I_n}$ 的n-mode乘积$\left(\mathcal{X} \bar{\times}_n \mathbf{v}\right) \in \mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times I_{n+1} \times \cdots \times I_N}$，其元素定义为

$$\left(\mathcal{X} \bar{\times}_n \mathbf{v}\right)_{i_1 \cdots i_{n-1} i_{n+1} \cdots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 i_2 \cdots i_N} v_{i_n}$$

  - 性质：

$$\mathcal{X} \bar{\times}_m \mathbf{a} \bar{\times}_n \mathbf{b} = \left(\mathcal{X} \bar{\times}_m \mathbf{a}\right) \bar{\times}_{n-1} \mathbf{b} = \left(\mathcal{X} \bar{\times}_n \mathbf{b}\right) \bar{\times}_m \mathbf{a}, m < n$$

基本概念及记号

- 矩阵的Kronecker乘积
  - $\mathbf{A} \in \mathbb{R}^{I \times J}, \mathbf{B} \in \mathbb{R}^{K \times L}$，则

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1J}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2J}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1}\mathbf{B} & a_{I2}\mathbf{B} & \cdots & a_{IJ}\mathbf{B} \end{bmatrix} \in \mathbb{R}^{IK \times JL}$$

  - 性质：$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$

$$(\mathbf{A} \otimes \mathbf{B})^{+} = \mathbf{A}^{+} \otimes \mathbf{B}^{+}$$

基本概念及记号

- 矩阵的Kronecker乘积
  - 矩阵的Kronecker积还和张量和矩阵的n-mode乘积有如下关系

$$\mathcal{Y} = \mathcal{X} \times_1 \mathbf{A}^{(1)} \cdots \times_N \mathbf{A}^{(N)} \Leftrightarrow$$

$$\mathbf{Y}_{(n)} = \mathbf{A}^{(n)} \mathbf{X}_{(n)} \left(\mathbf{A}^{(N)} \otimes \cdots \otimes \mathbf{A}^{(n+1)} \otimes \mathbf{A}^{(n-1)} \otimes \cdots \otimes \mathbf{A}^{(1)} \right)^{\mathrm{T}}$$

基本概念及记号

- 矩阵的Khatri-Rao乘积
  - $\mathbf{A} \in \mathbb{R}^{I \times K}, \mathbf{B} \in \mathbb{R}^{J \times K}$，则

$$\mathbf{A} \odot \mathbf{B} = \begin{bmatrix} \mathbf{a}_1 \otimes \mathbf{b}_1 & \mathbf{a}_2 \otimes \mathbf{b}_2 & \cdots & \mathbf{a}_K \otimes \mathbf{b}_K \end{bmatrix} \in \mathbb{R}^{IJ \times K}$$

  - 性质：$\mathbf{A} \odot \mathbf{B} \odot \mathbf{C} = (\mathbf{A} \odot \mathbf{B}) \odot \mathbf{C} = \mathbf{A} \odot (\mathbf{B} \odot \mathbf{C})$

基本概念及记号

- 矩阵的Hadamard乘积
  - $\mathbf{A} \in \mathbb{R}^{I \times J}, \mathbf{B} \in \mathbb{R}^{I \times J}$，则

$$\mathbf{A} * \mathbf{B} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots & a_{1J}b_{1J} \\ a_{21}b_{21} & a_{22}b_{22} & \cdots & a_{2J}b_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1}b_{I1} & a_{I2}b_{I2} & \cdots & a_{IJ}b_{IJ} \end{bmatrix} \in \mathbb{R}^{I \times J}$$

  - 性质：$(\mathbf{A} \odot \mathbf{B})^{\mathrm{T}} (\mathbf{A} \odot \mathbf{B}) = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right) * \left(\mathbf{B}^{\mathrm{T}}\mathbf{B}\right)$

$$(\mathbf{A} \odot \mathbf{B})^{+} = \left(\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right) * \left(\mathbf{B}^{\mathrm{T}}\mathbf{B}\right)\right)^{+} (\mathbf{A} \odot \mathbf{B})^{\mathrm{T}}$$

CP分解

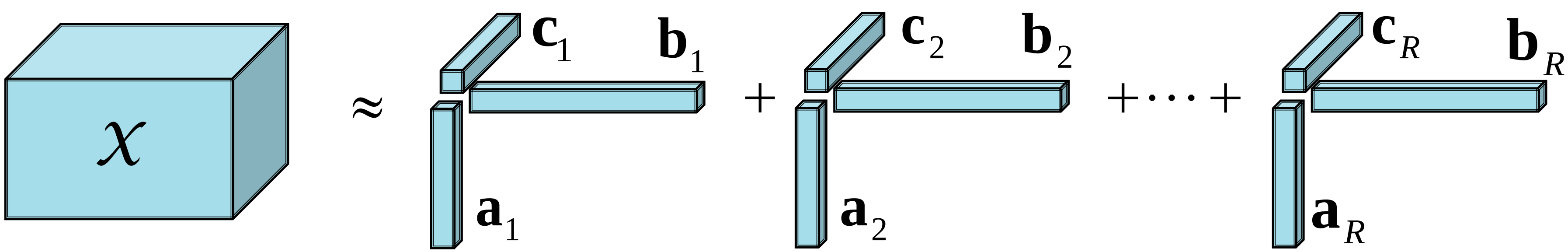
CP分解

- CP分解的其他名字
  - Polyadic Form of a Tensor, Hitchcock, 1927
  - PARAFAC(Parallel Factors), Harshman, 1970
  - CANDECOMP/CAND(Canonical decomposition), Carroll & Chang, 1970
  - Topographic Components Model, Möcks, 1988
  - CP(CANDECOMP/PARAFAC), Kiers, 2000

CP分解

- CP分解的张量形式
  - 将一个张量表示成有限个秩一张量之和，比如一个三阶张量可以分解为

$$\mathcal{X} \approx \mathbf{A},\mathbf{B},\mathbf{C} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$



三阶张量的CP分解

CP分解

- CP分解的矩阵形式
  - 因子矩阵：秩一张量中对应的向量组成的矩阵，如

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_R \end{bmatrix}$$

  - 利用因子矩阵，一个三阶张量的CP分解可以写成展开形式

$$\mathbf{X}_{(1)} \approx \mathbf{A}\left(\mathbf{C} \odot \mathbf{B}\right)^{\mathrm{T}}$$

$$\mathbf{X}_{(2)} \approx \mathbf{B}\left(\mathbf{C} \odot \mathbf{A}\right)^{\mathrm{T}}$$

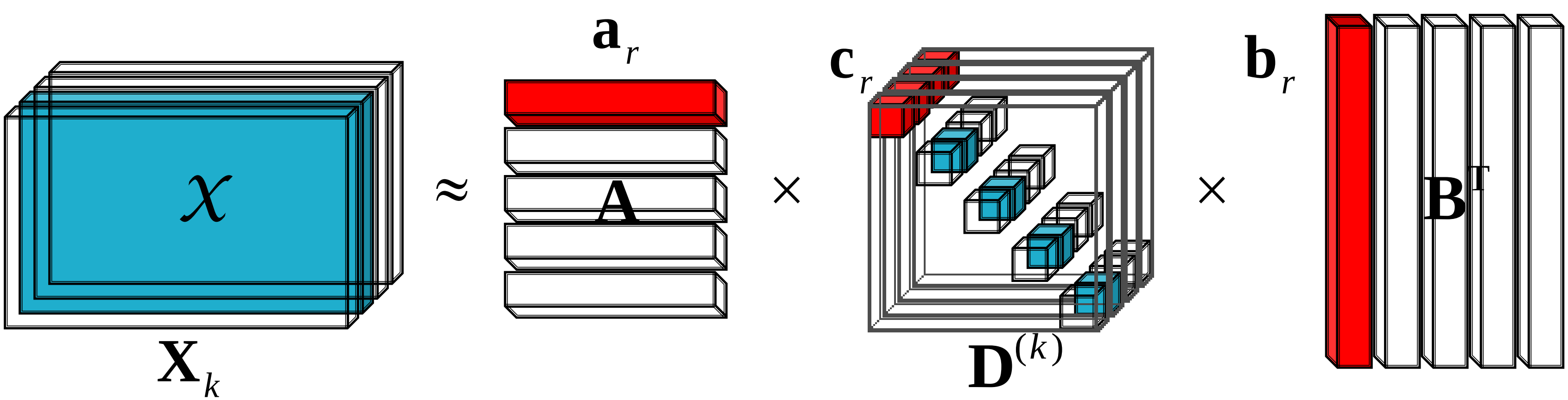
$$\mathbf{X}_{(3)} \approx \mathbf{C}\left(\mathbf{B} \odot \mathbf{A}\right)^{\mathrm{T}}$$

CP分解

- CP分解的切片形式
  - 三阶张量的CP分解有时按（正面）切片写成如下形式：

$$\mathbf{X}_k \approx \mathbf{A}\mathbf{D}^{(k)}\mathbf{B}^{\mathrm{T}}$$

其中 $\mathbf{D}^{(k)} \equiv \mathrm{diag}(\mathbf{c}_{k:})$



三阶张量CP分解的正面切片形式

CP分解

- 带权CP分解
  - 为了计算方便，通常假设因子矩阵的列是单位长度的，从而需要引入一个权重向量 $\boldsymbol{\lambda} \in \mathbb{R}^R$，使CP分解变为

$$\mathcal{X} \approx [\![\boldsymbol{\lambda}; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \equiv \sum_{r=1}^{R} \lambda_r \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

  - 对于高阶张量，有

$$\mathcal{X} \approx [\![\boldsymbol{\lambda}; \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \cdots, \mathbf{A}^{(N)}]\!] \equiv \sum_{r=1}^{R} \lambda_r \mathbf{a}_r^{(1)} \circ \mathbf{a}_r^{(2)} \circ \cdots \circ \mathbf{a}_r^{(N)}$$

其展开形式为

$$\mathbf{X}_{(n)} \approx \mathbf{A}^{(n)} \mathrm{diag}(\boldsymbol{\lambda}) \left(\mathbf{A}^{(N)} \odot \cdots \odot \mathbf{A}^{(n+1)} \odot \mathbf{A}^{(n-1)} \odot \cdots \odot \mathbf{A}^{(1)} \right)^{\mathrm{T}}$$

CP分解

- 张量的秩和秩分解
  - 张量 $\mathcal{X}$ 的秩定义为用秩一张量之和来精确表示 $\mathcal{X}$ 所需要的秩一张量的最少个数，记为 $\operatorname{rank}(\mathcal{X})$
  - 秩分解：

$$\mathcal{X} = \sum_{r=1}^{\operatorname{rank}(\mathcal{X})} \mathbf{a}_r^{(1)} \circ \mathbf{a}_r^{(2)} \circ \cdots \circ \mathbf{a}_r^{(N)}$$

    可见秩分解是一个特殊的CP分解，对应于矩阵的SVD
  - 目前还没有方法能够直接求解一个任意给定张量的秩，这被证明是一个NP-hard问题

CP分解

- 张量的秩
  - 不同于矩阵的秩，高阶张量的秩在实数域和复数域上不一定相同。例如一个三阶张量$\mathcal{X}$

$$\mathbf{X}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \mathbf{X}_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

在实数域内进行秩分解得到的因子矩阵为

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix}$$

而在复数域内进行分解得到的因子矩阵为

$$\mathbf{A} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix} \quad \mathbf{B} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}$$

CP分解

- 张量的低秩近似
  - 相对于矩阵的SVD来说，高阶张量的秩分解唯一性不需要正交性条件保证，只需满足：

$$\sum_{n=1}^{N} k_{\mathbf{A}^{(n)}} \geq 2R + N - 1$$

这里 $k_{\mathbf{A}}$ 表示矩阵 $\mathbf{A}$ 的k-秩：任意k列都线性无关的最大的k

CP分解

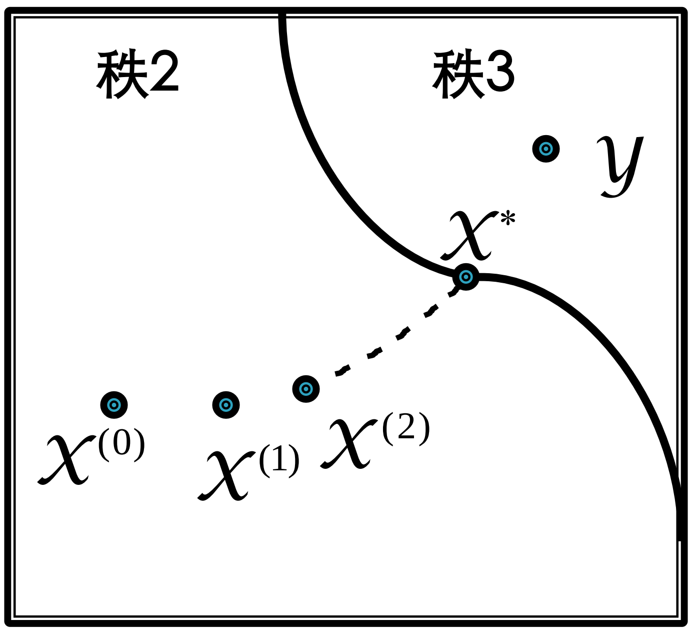
- 张量的低秩近似
  - 然而在低秩近似方面，高阶张量的性质比矩阵SVD差
    - Kolda给出了一个例子，一个立方张量的最佳秩–1近似并不包括在其最佳秩–2近似中，这说明张量的秩–k近似无法渐进地得到
    - 下面的例子说明，张量的“最佳”秩–k近似甚至不一定存在

$$\mathcal{X} = \mathbf{a}_1 \circ \mathbf{b}_1 \circ \mathbf{c}_2 + \mathbf{a}_1 \circ \mathbf{b}_2 \circ \mathbf{c}_1 + \mathbf{a}_2 \circ \mathbf{b}_1 \circ \mathbf{c}_1$$

$$\mathcal{Y} = \alpha \left(\mathbf{a}_1 + \frac{1}{\alpha} \mathbf{a}_2 \right) \circ \left(\mathbf{b}_1 + \frac{1}{\alpha} \mathbf{b}_2 \right) \circ \left(\mathbf{c}_1 + \frac{1}{\alpha} \mathbf{c}_2 \right) - \alpha \mathbf{a}_1 \circ \mathbf{b}_1 \circ \mathbf{c}_1$$

CP分解

- 张量的低秩近似
  - 退化：如果一个张量能够被一系列的低秩张量任意逼近
  - 边缘秩（border rank）：能够任意逼近一个张量的最少的成分个数



一个秩为2的张量序列收敛到一个秩3张量

CP分解

- CP分解的计算
  - 分解成多少个秩一张量（成分）之和？
    - 通常的做法是从1开始尝试，知道碰到一个“好”的结果为止
    - 如果有较强的应用背景和先验信息，可以预先指定
  - 对于给定的成分数目，怎么求解CP分解？
    - 目前仍然没有一个完美的解决方案
    - 从效果来看，交替最小二乘（Alternating Least Square）是一类比较有效的算法

CP分解

- CP分解的计算
  - 以一个三阶张量 $\mathcal{X}$ 为例，假定成分个数 R 已知，目标为

$$\min_{\hat{\mathcal{X}}} \left\| \mathcal{X} - \hat{\mathcal{X}} \right\| \quad \text{s.t.} \quad \hat{\mathcal{X}} = \sum_{r=1}^{R} \lambda_r \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r = \ \boldsymbol{\lambda}; \mathbf{A}, \mathbf{B}, \mathbf{C}$$

  - 作为ALS的一个子问题，固定 **B** 和 **C** ，求解

$$\min_{\mathbf{A}} \left\| \mathbf{X}_{(1)} - \mathbf{A}\mathrm{diag}(\boldsymbol{\lambda})\left(\mathbf{C} \odot \mathbf{B}\right)^{\mathrm{T}} \right\|_F$$

得 $\mathbf{A}\mathrm{diag}(\boldsymbol{\lambda}) = \mathbf{X}_{(1)}\left[\left(\mathbf{C} \odot \mathbf{B}\right)^{\mathrm{T}}\right]^{+} = \mathbf{X}_{(1)}\left(\mathbf{C} \odot \mathbf{B}\right)\left(\mathbf{C}^{\mathrm{T}}\mathbf{C} * \mathbf{B}^{\mathrm{T}}\mathbf{B}\right)^{+}$

再通过归一化分别求出 **A** 和 $\boldsymbol{\lambda}$

CP分解

- CP分解的计算
  - ALS算法并不能保证收敛到一个极小点，甚至不一定能收敛到稳定点，它只能找到一个目标函数不再下降的点

  - 算法的初始化可以是随机的，也可以将因子矩阵初始化为对应展开的奇异向量，如将$\mathbf{A}$ 初始化为 $\mathcal{X}_{(1)}$ 的前 R 个左奇异向量

CP分解

- CP分解的应用
  - 计量心理学
  - 语音分析
  - 化学计量学
  - 独立成分分析
  - 神经科学
  - 数据挖掘
  - 高维算子近似
  - 随即偏微分方程
  - …………

Tucker分解

Tucker分解

- Tucker分解的其他名字
  - Three-mode factor analysis(3MFA/Tucker3), Tucker, 1966
  - Three-mode principal component analysis(3MPCA), Kroonenberg & De Leeuw, 1980
  - N-mode principal components analysis, Kapteyn et al., 1986
  - Higher-order SVD(HOSVD), De Lathauwer et al., 2000
  - N-mode SVD, Vasilescu and Terzopoulos, 2002

Tucker分解

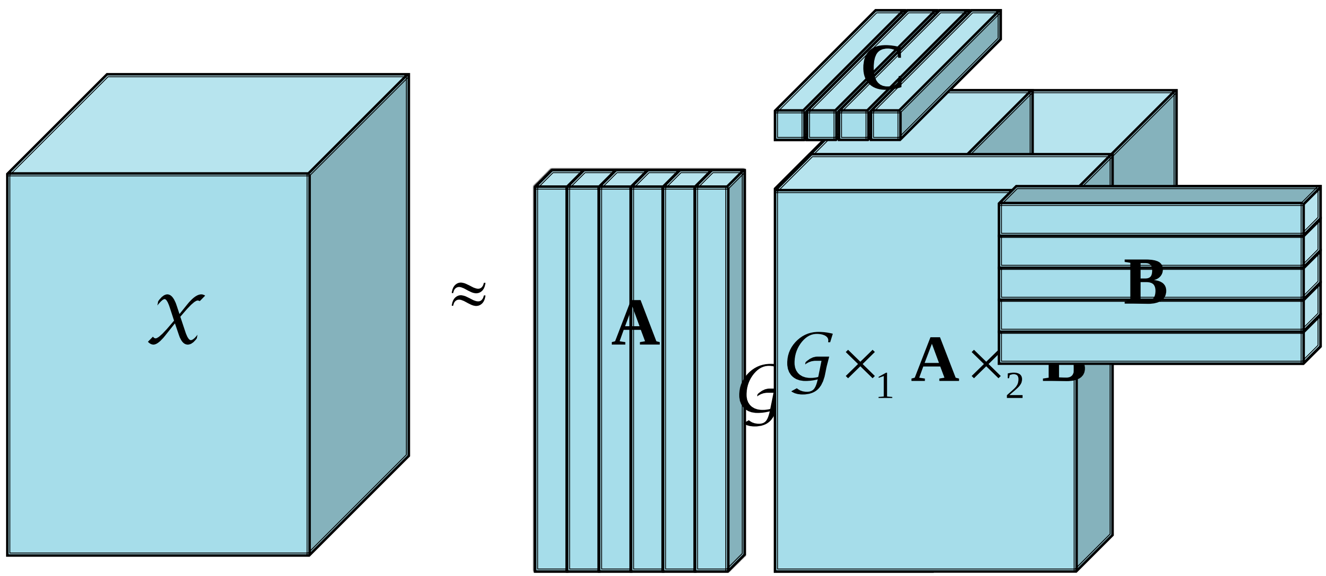
- Tucker分解
  - Tucker分解是一种高阶的主成分分析，它将一个张量表示成一个核心（core）张量沿每一个mode乘上一个矩阵。对于三阶张量$\mathcal{X} \in \mathbb{R}^{I\times J\times K}$来说，其Tucker分解为

$$\mathcal{X} \approx [\![\mathcal{G}; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] = \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{pqr} \mathbf{a}_p \circ \mathbf{b}_q \circ \mathbf{c}_r$$

  - 因子矩阵 $\mathbf{A} \in \mathbb{R}^{I\times P}, \mathbf{B} \in \mathbb{R}^{J\times Q}, \mathbf{C} \in \mathbb{R}^{K\times R}$ 通常是正交的，可以视为沿相应mode的主成分

Tucker分解

- Tucker分解
  - 容易看出，CP分解是Tucker分解的一种特殊形式：如果核心张量 $\mathcal{G}$ 是对角的，且 $P = Q = R$，则Tucker分解就退化成了CP分解



三阶张量的Tucker分解

Tucker分解

- Tucker分解的矩阵形式
  - 三阶Tucker分解的展开形式为

$$\mathbf{X}_{(1)} \approx \mathbf{A}\mathbf{G}_{(1)}\left(\mathbf{C}\otimes\mathbf{B}\right)^{\mathrm{T}}$$

$$\mathbf{X}_{(2)} \approx \mathbf{B}\mathbf{G}_{(2)}\left(\mathbf{C}\otimes\mathbf{A}\right)^{\mathrm{T}}$$

$$\mathbf{X}_{(3)} \approx \mathbf{C}\mathbf{G}_{(3)}\left(\mathbf{B}\otimes\mathbf{A}\right)^{\mathrm{T}}$$

  - Tucker分解可以推广到高阶张量

$$\mathcal{X} \approx \mathcal{G}\times_1 \mathbf{A}^{(1)}\times_2 \mathbf{A}^{(2)}\cdots\times_N \mathbf{A}^{(N)} = [\![\mathcal{G};\mathbf{A}^{(1)},\mathbf{A}^{(2)},\cdots,\mathbf{A}^{(N)}]\!]$$

$$\mathbf{X}_{(n)} \approx \mathbf{A}^{(n)}\mathbf{G}_{(n)}\left(\mathbf{A}^{(N)}\otimes\cdots\otimes\mathbf{A}^{(n+1)}\otimes\mathbf{A}^{(n-1)}\otimes\cdots\otimes\mathbf{A}^{(1)}\right)^{\mathrm{T}}$$

Tucker分解

- Tucker2和Tucker1
  - 对于三阶张量固定一个因子矩阵为单位阵，就得到Tucker分解一个重要的特例：Tucker2。例如固定$\mathbf{C} = \mathbf{I}$，则

$$\mathcal{X} \approx \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B} = \mathcal{G}; \mathbf{A}, \mathbf{B}, \mathbf{I}$$

  - 进一步，固定两个因子矩阵，就得到了Tucker1，例如令第二、三个因子矩阵为单位阵，则Tucker分解就退化成了普通的PCA

$$\mathcal{X} \approx \mathcal{G} \times_1 \mathbf{A} = \mathcal{G}; \mathbf{A}, \mathbf{I}, \mathbf{I} \Leftrightarrow \mathbf{X}_{(1)} = \mathbf{A}\mathbf{G}_{(1)}$$

Tucker分解

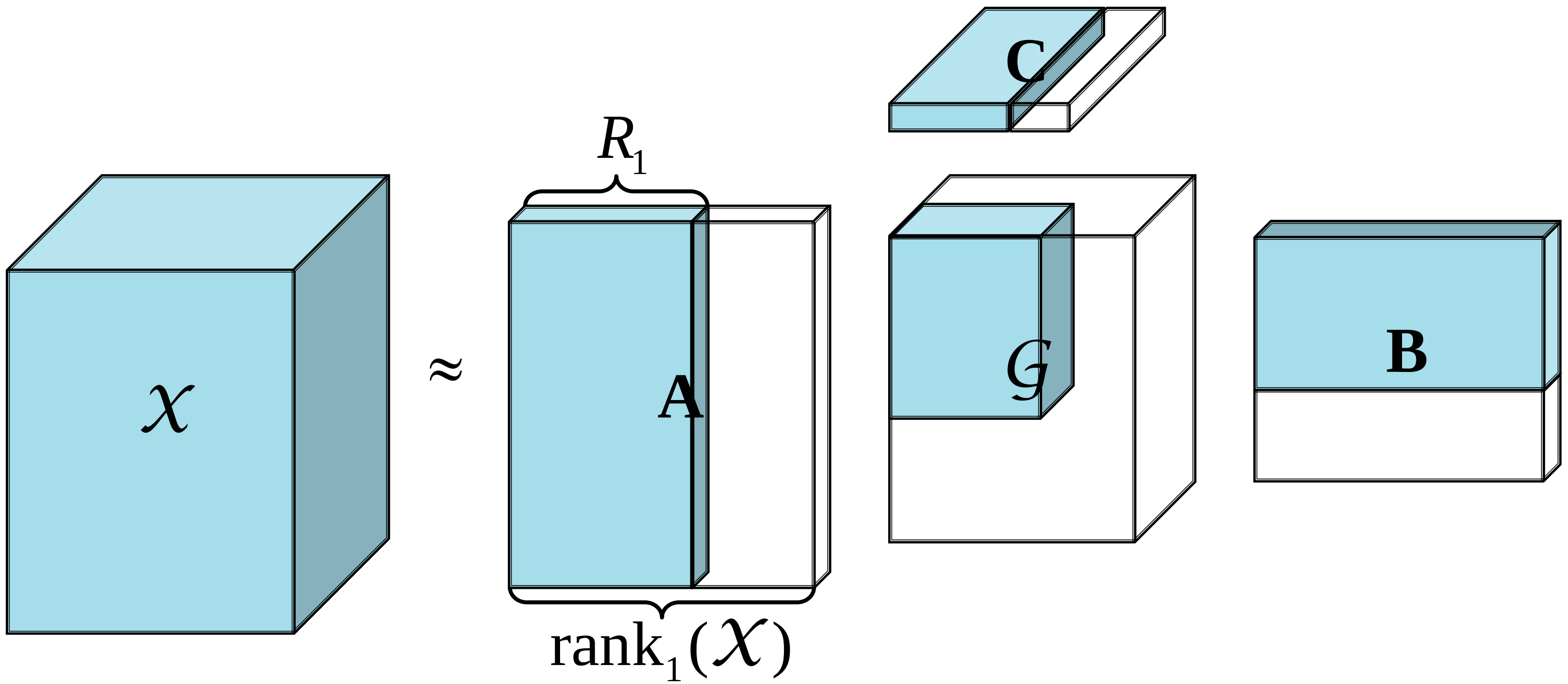
- 张量的n-秩近似
  - 一个N阶张量$\mathcal{X}$的n-秩定义为

$$\operatorname{rank}_n(\mathcal{X}) = \operatorname{rank}(\mathbf{X}_{(n)})$$

  - 若设$\operatorname{rank}_n(\mathcal{X}) = R_n, n = 1, \cdots, N$，则$\mathcal{X}$叫做一个秩-$(R_1, R_2, \cdots, R_N)$张量
  - 如果 $\operatorname{rank}_n(\mathcal{X}) = R_n, n = 1, \cdots, N$，则很容易得到$\mathcal{X}$的一个精确秩-$(R_1, R_2, \cdots, R_N)$Tucker分解；然而如果至少有一个$n$使得$\operatorname{rank}_n(\mathcal{X}) > R_n$，则通过Tucker分解得到的就是$\mathcal{X}$的一个秩-$(R_1, R_2, \cdots, R_N)$近似

Tucker分解

- 张量的n-秩近似



截断的Tucker分解：秩-$\left(R_1, R_2, R_3\right)$近似

Tucker分解

- 张量的n-秩近似
  - 对于固定的n-秩，Tucker分解的唯一性不能保证，所以需要添加其他的约束
  - 通常要求核心张量是“简单”的，如各个mode的主成分之间尽量不发生相互作用（稀疏性），或者其他的“简单性”约束

Tucker分解

- Tucker分解的计算
  - HOSVD：利用SVD对每个mode做一次Tucker1分解（截断或者不截断）
  - HOSVD不能保证得到一个较好的近似，但HOSVD的结果可以作为一个其他迭代算法（如HOOI）的很好的初始解

Tucker分解

- Tucker分解的计算
  - 为了导出HOOI迭代算法，先考虑目标函数

$$\left\| \mathcal{X} - [\![\mathcal{G}; \mathbf{A}^{(1)}, \cdots, \mathbf{A}^{(N)}]\!] \right\|$$

$$= \left\| \mathrm{vec}(\mathcal{X}) - \left(\mathbf{A}^{(N)} \otimes \cdots \otimes \mathbf{A}^{(1)} \right) \mathrm{vec}(\mathcal{G}) \right\|$$

  - 从而 $\mathcal{G}$ 应该满足

$$\mathcal{G} = \mathcal{X} \times_1 \mathbf{A}^{(1)\mathrm{T}} \cdots \times_N \mathbf{A}^{(N)\mathrm{T}}$$

Tucker分解

- Tucker分解的计算
  - 目标函数的平方变为

$$\left\| \mathcal{X} - [\![\mathcal{G}; \mathbf{A}^{(1)}, \cdots, \mathbf{A}^{(N)}]\!] \right\|^2$$

$$= \|\mathcal{X}\|^2 - 2\left\langle \mathcal{X}, [\![\mathcal{G}; \mathbf{A}^{(1)}, \cdots, \mathbf{A}^{(N)}]\!] \right\rangle + \left\| [\![\mathcal{G}; \mathbf{A}^{(1)}, \cdots, \mathbf{A}^{(N)}]\!] \right\|^2$$

$$= \|\mathcal{X}\|^2 - 2\left\langle \mathcal{X} \times_1 \mathbf{A}^{(1)\mathrm{T}} \cdots \times_N \mathbf{A}^{(N)\mathrm{T}}, \mathcal{G} \right\rangle + \|\mathcal{G}\|^2$$

$$= \|\mathcal{X}\|^2 - 2\langle \mathcal{G}, \mathcal{G} \rangle + \|\mathcal{G}\|^2$$

$$= \|\mathcal{X}\|^2 - \left\| \mathcal{X} \times_1 \mathbf{A}^{(1)\mathrm{T}} \cdots \times_N \mathbf{A}^{(N)\mathrm{T}} \right\|^2$$

Tucker分解

- Tucker分解的计算
  - 所以问题可以进行如下转化

$$\min \left\| \mathcal{X} - [\![\mathcal{G}; \mathbf{A}^{(1)}, \cdots, \mathbf{A}^{(N)}]\!] \right\|$$

$$\Leftrightarrow \max \left\| \mathcal{X} \times_1 \mathbf{A}^{(1)\mathrm{T}} \cdots \times_N \mathbf{A}^{(N)\mathrm{T}} \right\|$$

  - 利用交替求解的思想，问题变为解如下子问题

$$\max \left\| \mathbf{A}^{(n)\mathrm{T}} \mathbf{W} \right\|$$

$$\text{s.t.} \quad \mathbf{W} = \mathbf{X}_{(n)} \left(\mathbf{A}^{(N)} \otimes \cdots \otimes \mathbf{A}^{(n+1)} \otimes \mathbf{A}^{(n-1)} \cdots \otimes \mathbf{A}^{(1)} \right)$$

这个问题可以通过令 $\mathbf{A}^{(n)}$ 为 $\mathbf{W}$ 的前 R_n 个左奇异值向量来解决

Tucker分解

- Tucker分解的应用
  - 化学分析
  - 计量心理学
  - 信号处理
  - 机器视觉（面部、动作）
  - 数据压缩
  - 纹理生成
  - 数据挖掘
  - 环境和网络建模
  - …………

欢迎大家提出宝贵建议