张量网络算法基础（一）

冉仕举

首都师范大学物理系

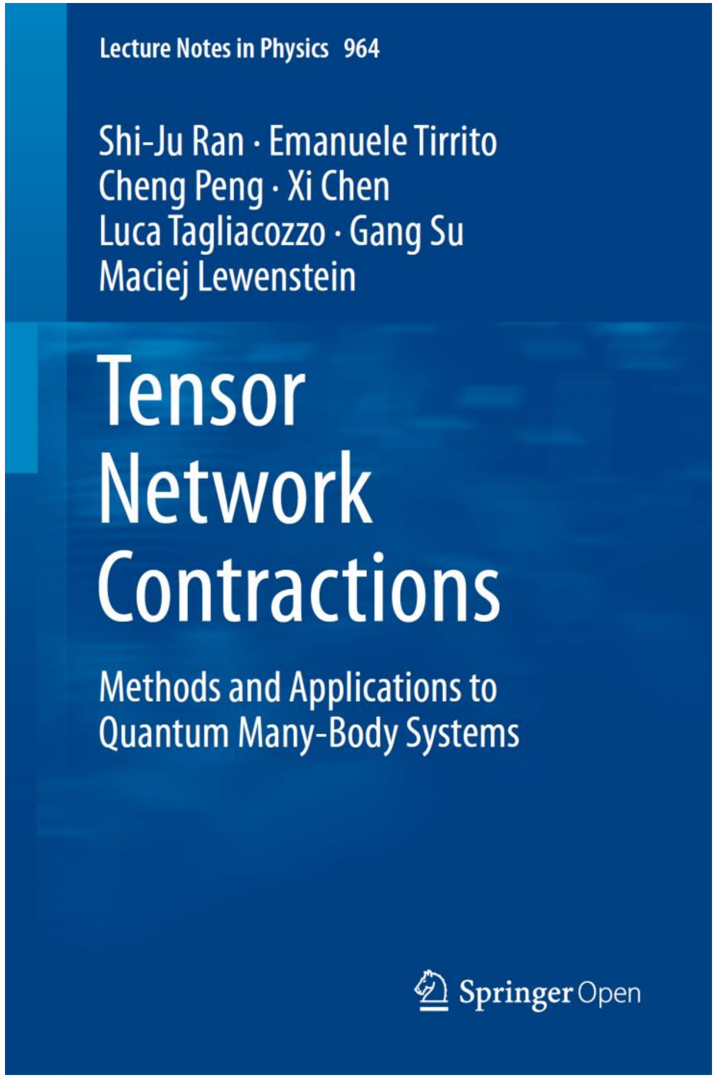
2020年春

提 纲

0. 张量网络与多体物理简介
1. 张量及Python基础
2. 线性代数及多线性代数基础
3. 格点模型基础
4. 矩阵乘积态方法
5. 张量网络计算方法
6. 张量网络机器学习

张量网络部分主要参考书：

- 《Tensor Network Contractions: Methods and Applications to Quantum Many-Body Systems》



下载地址：
https://link.springer.com/content/pdf/10.1007%2F978-3-030-34489-4.pdf

课前准备

课程资料下载地址:
https://icloud.cnu.edu.cn:443/link/7B1AC701FBCE6FB7F38604EA15A3844D
Jupyter notebook程序在线阅读地址：
https://nbviewer.jupyter.org/github/ranshiju/TN_tutorial/tree/master/
GitHub代码下载地址：
https://github.com/ranshiju/TN_tutorial
教学视频在线观看地址：
https://www.bilibili.com/video/BV17z411i7yM/

a. 安装Anaconda与python，并安装numpy与pytorch；
b. 安装编辑器Pycharm。

注：

- 可使用除Anaconda之外的方式安装与配置python
- 可使用其它编辑器软件
- 如果设备有英伟达独立显卡，安装GPU版本的pytorch；如果没有，则安装CPU版本的pytorch
- 课程配套程序建议使用Jupyter notebook打开；如果使用Anaconda安装，则自带了Jupyter notebook

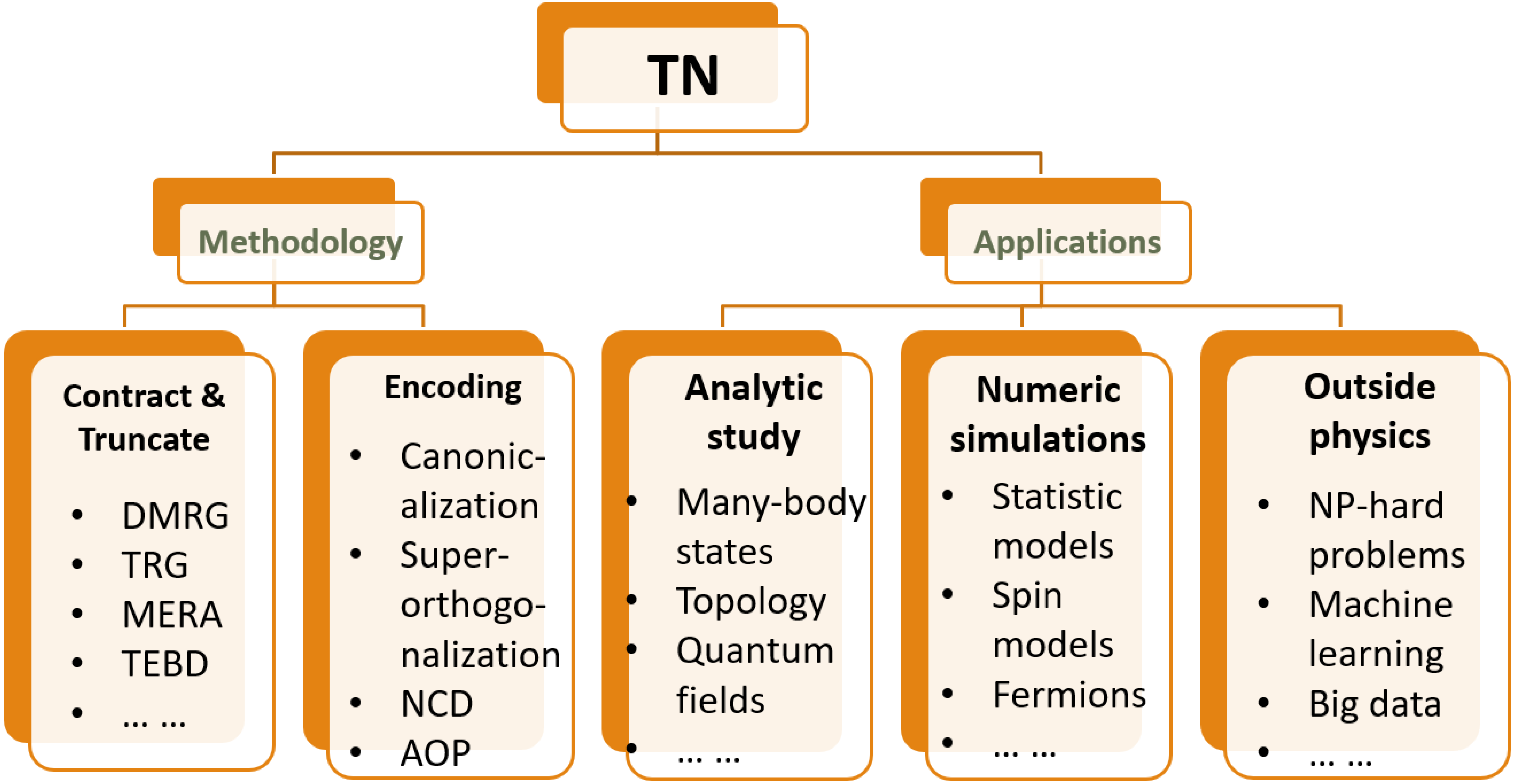
课程说明

- **目标：** 理解张量及张量网络基本思想；理解张量网络与物理学问题之间的联系；掌握编写张量网络算法程序基础
- 由于课时较短，不一定能够讲完所有内容。如果在课时内没有讲完，之后会抽时间继续在线上讲完全部内容，**但考核范围仅限于课时内讲完的部分**
- 不会系统讲解python，强调**边用边学**
- **受众：** 物理专业的本科生与研究生；量子力学基础较薄弱的非物理专业人士；编程基础较薄弱但对相关领域感兴趣的人士。

目 录

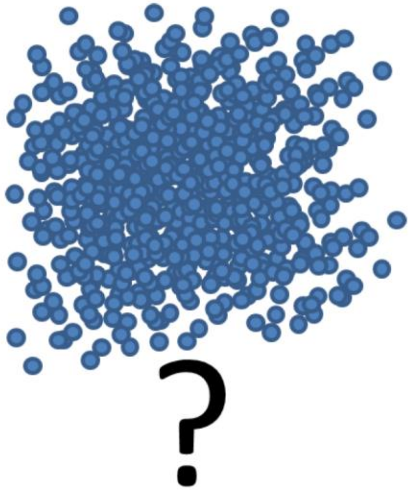
争取课内讲完这些部分，让大家具备自学剩余部分及阅读前沿文献的必要基础

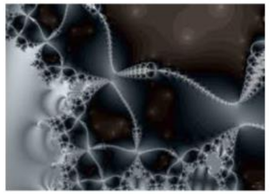
0. 张量网络与多体物理简介



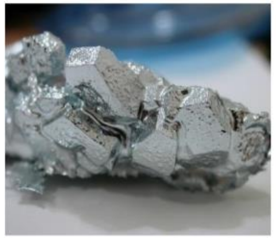
量子多体系统中出现的丰富物理现象

Emergent phenomena
in many-body systems

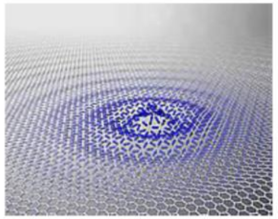




quantum criticality

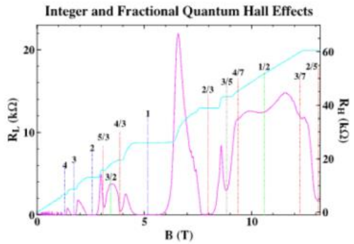


metal



topological order

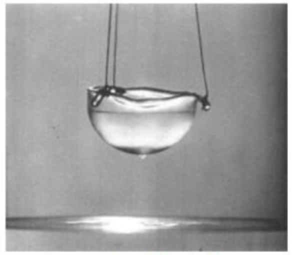
insulator



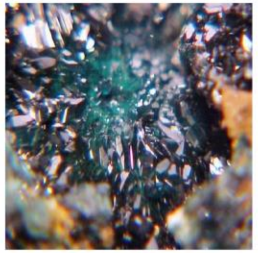
fractional quantum
Hall effect

A beautiful slide from
a talk of Guifre Vidal

superconductor



superfluid



spin liquid

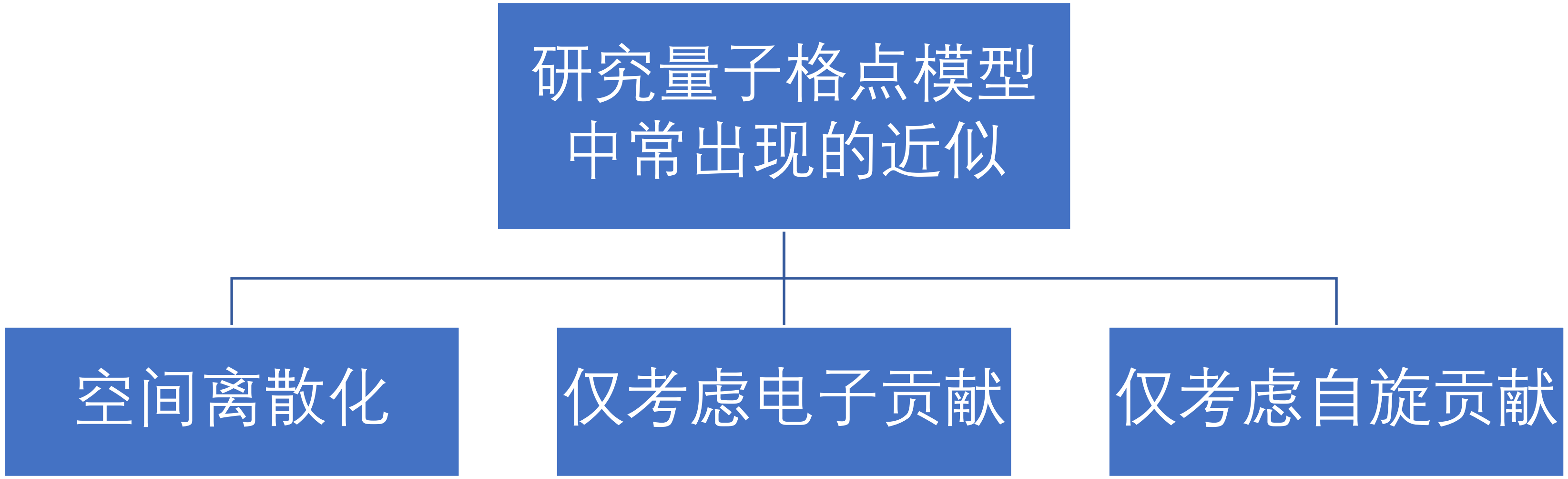
求解量子多体问题难度极高，需要采用一定的近似

- 第一性原理方法常用的近似
  - 周期势能
  - 平均场近似
  - 单粒子近似

优势：
- 可处理实际且较为复杂的系统，包括材料物理、量子化学等

劣势：
- 近似较大，忽略了多体相互作用

另一种倾向于考虑关联效应的近似思路：量子格点模型

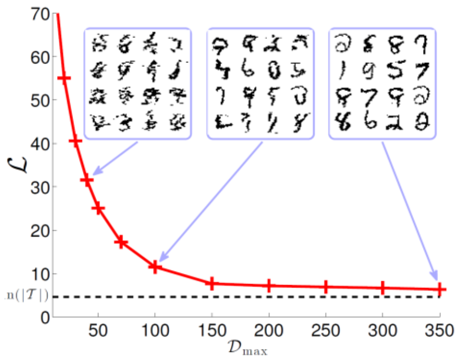


优势：
- 考虑相互作用及强关联效应

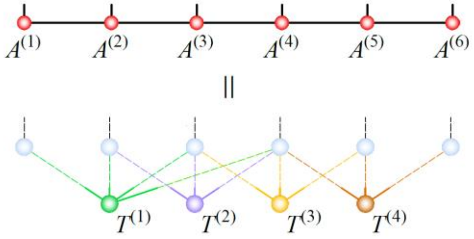
劣势：
- 对可描述的实际材料要求严格，多用于理论模型研究，目前实际应用范围较窄

目前的前沿问题之一：结合第一性原理计算方法与格点模型计算方法二者的优势，发现、解释强关联效应下的新奇物态，而**张量网络**是目前最为强大的理论、数值工具之一

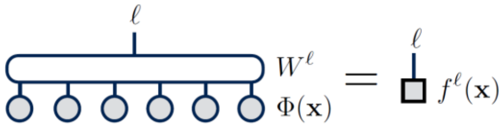
张量网络机器学习：一门本受关注的新交叉学科



TN and generative models
(Phys. Rev. X 8, 031012 (2018))

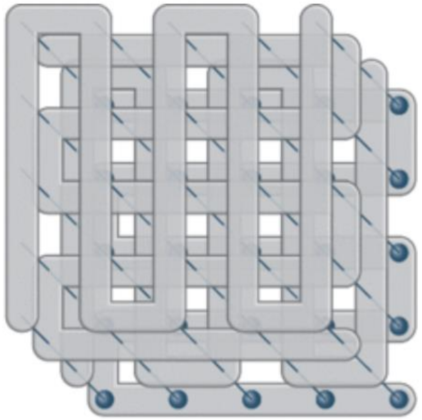


TN and Boltzmann machine
(Phys. Rev. B 97 085104 (2018))

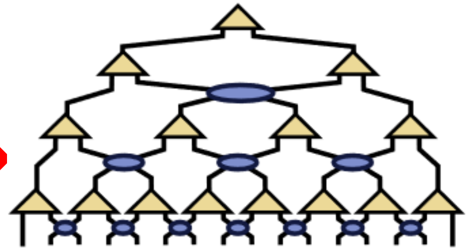


TN and supervised learning
(Adv. Neur. Inf. Proc. Sys. 29, 4799 (2016))

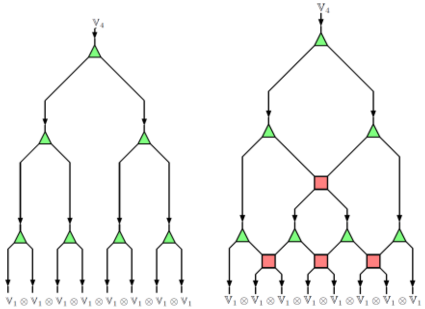
ML in many-body space by TN



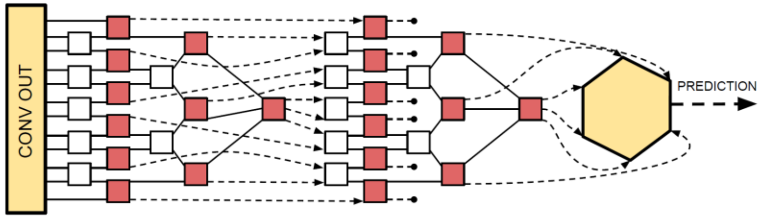
String-bond state and Boltzmann machine
(Phys. Rev. X 8, 011006 (2018))



Machine learning by MERA
(Quantum Sci. Technol. 3, 034003 (2018))



TN and language learning
(arXiv:1710.10248)



Simplify neural networks by quantum gates (arXiv:1711.03357)

Just Started!

关于张量网络的发展历史，为了避免一开始就引入太多的专有名词，我们在这里就不作过多的介绍了。本课程从设置上也是希望，在不了解相关专有名词的情况下，先从数学上理解张量网络基础，再结合具体的问题学习张量网络及其算法

1. 张量及Python基础

注：关于代码的学习，提倡边用边学，一定要善于利用网络资源来解决问题

1.1 什么是张量

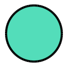
	张量阶数	例子
标量	0	能量、温度、一门课的总成绩
向量	1	速度、力、多门课的总成绩构成的数列
矩阵	2	算符的系数矩阵、黑白图片
张量	N	彩色图片（三阶张量）

- N阶张量又称**N维数列**
- 张量**阶数**即为**张量指标的个数**
- 向量、矩阵也被称为1阶张量、2阶张量
- 每个指标可取的值的个数，被称为指标的**维数**
- **张量的“直观”定义：多个指标标记下的一堆数**

Jupyter notebook: sec1_1_basic

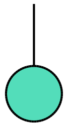
1.2 张量的图形表示

- 张量用连接着N个腿（bond）的圆圈或方块表示，N为张量的阶数



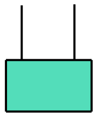
标量

c



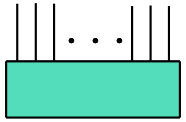
向量
（一阶张量）

v_a



矩阵
（二阶张量）

m_{ab}



N阶张量

$T_{s_1 s_2 \ldots s_N}$

1.3 张量的基本操作

- 张量可进行**切片**操作，提取相关的元素
- 张量指标的顺序可通过**transpose**命令进行改变
- 张量的阶数可通过**reshape**命令相互转换
- 使用transpose与reshape命令转换张量前后，其所包含的元素总个数不变，意义仅在于**采用不同了的index标记各个元素**

练习：建立任意随机三阶张量，设其形状（shape）为(a, b, c)，先将其reshape成$(a \times b, c)$的矩阵，再直接将该矩阵reshape成$(a, b \times c)$的矩阵，最后将所得矩阵reshape成(a, b, c)的三阶张量，检查最后所得张量与所建立的张量是否相等。

Jupyter Notebook: sec1_1_basic

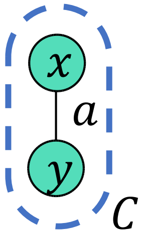
1.4 张量的基本运算

练习：1. 画出矩阵求迹trace(M)的图形表示；
2. 画出如下张量缩并式的图形表示

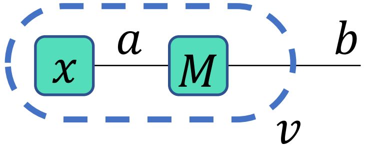
$$T_{ace} = \sum_{bd} A_{ab} B_{bcd} C_{ade}$$

- 向量内积

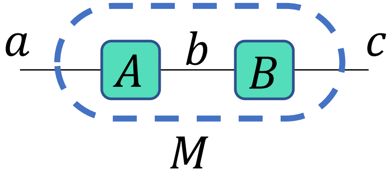
$$C = \sum_{a} x_a y_a$$



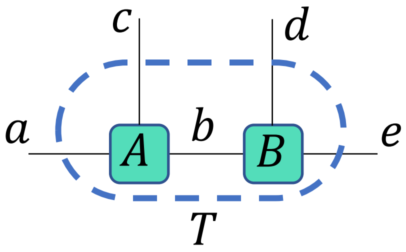
- 向量乘矩阵 $v_b = \sum_{a} x_a M_{ab} = xM$



- 矩阵乘矩阵 $M_{ac} = \sum_{b} A_{ab} B_{bc} = AB$



- 张量缩并 $T_{acde} = \sum_{b} A_{abc} B_{bde}$



任何可写成多张量求和的式子都可用程序实现
练习：用numpy中的einsum函数实现（1）矩阵求迹；（2）矩阵乘；（3）下式

$$T_{ace} = \sum_{bd} A_{abd} B_{bcd} C_{cde}$$

Jupyter Notebook: sec1_4_computation

2. 线性代数及多线性代数基础

2.1 本征值分解与最大本征值问题

a. **本征向量与本征值**（又称特征向量与特征值）：给定$D \times D$的矩阵M，设D维归一向量v与标量γ，当其满足$\boldsymbol{Mv} = \boldsymbol{\gamma v}$时，称$v$与$\gamma$分别为$M$的本征向量与本征值

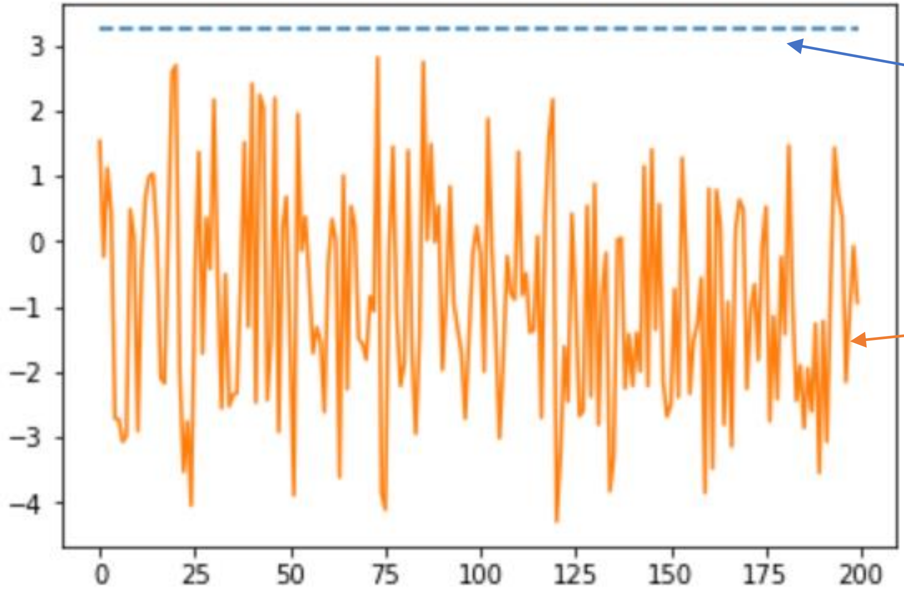
b. **本征值分解**（又称特征分解，eigenvalue decomposition）：$M = U\Gamma U^{\dagger}$，其中 U被称为**变换矩阵**，其每一列为M的本征向量，且满足正交归一性$UU^{\dagger} = I$；Γ为对角矩阵，称为**本征谱**（请证明U中的每一列向量满足定义1，为本征向量，且对应的本征值为Γ中对应的元素）。
（注：厄密矩阵的条件$M = M^{\dagger}$；不是任何矩阵都存在本征值分解，但是厄密矩阵一定存在本征值分解，且本征谱为实）

练习：画出本征值分解的式子对应的图形表示

c. **最大本征值问题**：假设矩阵本征值为实数，求解给定矩阵的最大本征值及其本征态

- 最大本征值问题对应于如下**优化问题**：
  给定矩阵M，求解归一化向量v，使得函数
  $$f = |v^T M v|$$
  的值极大化。

- 该最优化问题的解为M的（绝对值）最大本征态，对应的f值为最大本征值



练习：尝试数学上证明最大本征值问题与上述优化问题的等价性

Jupyter Notebook: sec2_1_eig_opt

c. 最大本征值问题的幂级数求解法

- 考虑实对称矩阵M，设Γ_0和$u^{(0)}$为其绝对值最大的唯一本征值及本征向量，则

$$\lim_{K\to\infty} M^K = \Gamma_0^K u^{(0)} u^{(0)T}$$

Jupyter Notebook: sec2_1_eig_opt

证明:

设M的本征值分解为$M = U\Gamma U^T$，有$M^K = U\Gamma U^T U\Gamma U^T \dots U\Gamma U^T$;

由$UU^T = U^T U = I$得， $M^K = U\Gamma^K U^T = \Gamma_0^K\ U(\frac{\Gamma}{\Gamma_0})^K U^T$;

其中， $\lim_{K\to\infty} (\frac{\Gamma}{\Gamma_0})^K = \lim_{K\to\infty} \begin{bmatrix} \Gamma_0/\Gamma_0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Gamma_{D-1}/\Gamma_0 \end{bmatrix}^K = \mathrm{diag}([1 \quad 0 \quad \dots \quad 0])$

得: $\lim_{K\to\infty} M^K = U \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} U^T = \Gamma_0^K u^{(0)} u^{(0)T}$，其中$u^{(0)}$为$U$的第0列。

证毕

2.2 奇异值分解与最优低秩近似问题

a. **奇异值分解**（singular value decomposition）：给定$D \times D'$的矩阵M，有$\boldsymbol{M} = \boldsymbol{U\Lambda V}^{\dagger}$，其中 U与V每一列分别被称为M的**左、右奇异向量**，且满足正交归一性$UU^{\dagger} = I$，$VV^{\dagger} = I$；Λ为非负定实对角矩阵，称为**奇异谱**，其元素称为**奇异值**，按非升序排列。

注：任何矩阵都存在奇异值分解，且极大化U第一个元素实部时，分解唯一
思考：去掉极大化实部的条件后，分解为什么不唯一？

b. **秩**：定义为矩阵的非零奇异值个数。

c. **矩阵的低秩近似问题**：给定$D \times D'$的矩阵M，设其秩为R，求解秩为R'的矩阵M'，有$R > R' > 0$，且极小化二矩阵间的范数

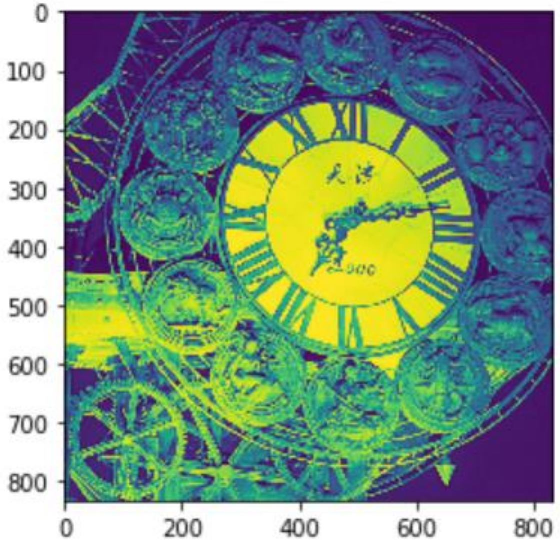
$$\varepsilon = ||M - M'|| = \sqrt{\sum_{ij} (M_{ij} - M'_{ij})^2} \sim ||\Lambda_{R':R-1}||$$（裁剪误差）

d. **低秩近似问题的最优解为**：$M' = U[:,0:R']\ \Lambda[0:R',0:R']\ V[:,0:R']^{\dagger}$

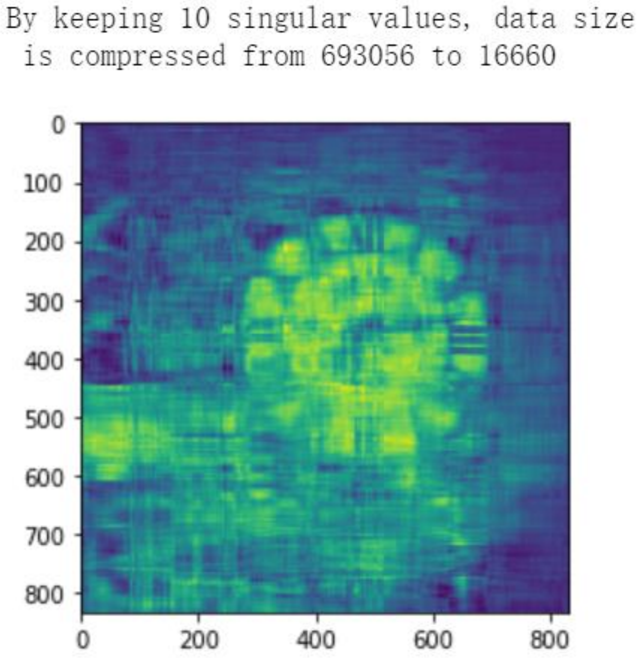
2.2 奇异值分解与最优低秩近似问题

Jupyter Notebook: sec2_2_svd_img

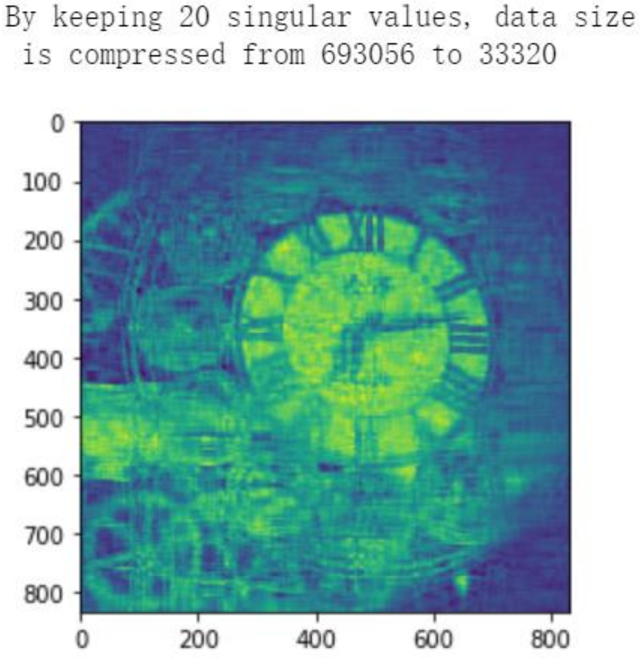
- 例：基于奇异值分解低秩近似的图形压缩



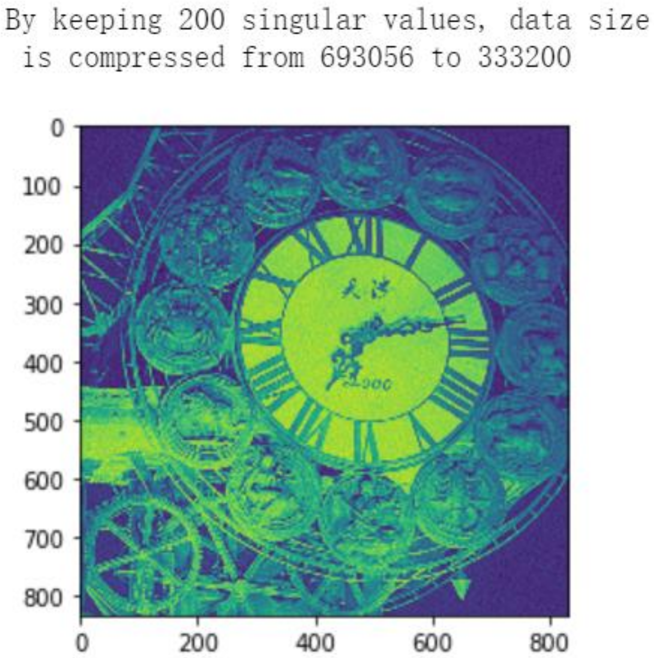
原图



error = 0.324876



error = 0.285488



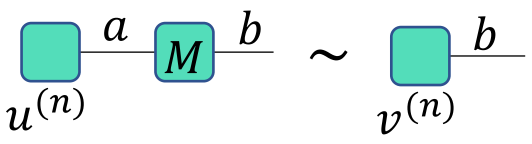
error = 0.086106

练习：1. 对保留的奇异值数量k取适当的间隔，完成绘制误差关于k的曲线的函数；
2. 证明：对于矩阵M有奇异值分解$M = U\Lambda V^\dagger$时，其左约化矩阵$L = MM^\dagger$的本征值分解满足$L = U\Gamma U^\dagger$，其右约化矩阵$R = M^\dagger M$的本征值分解满足$R = V\Gamma V^\dagger$，且有$\Gamma = \Lambda^2$。

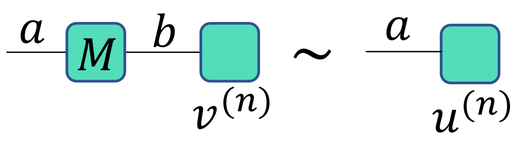
2.3 多线性代数中的张量单秩分解

- 思考：矩阵（二阶张量）存在本征值分解、奇异值分解等，对于更高阶的张量，是否存在类似的分解？
- 考虑矩阵M（简要起见设M为实矩阵）的奇异值分解，记第n个左、右奇异向量分别为$u^{(n)}$和$v^{(n)}$，证明如下等式：

$$\sum_a u_a^{(n)} M_{ab} = \Lambda^{(n)} v_b^{(n)}$$



$$\sum_b M_{ab} v_b^{(n)} = \Lambda^{(n)} u_a^{(n)}$$



其中，$\Lambda^{(n)}$为第n个奇异值，且有$\Lambda^{(n)} = \sum_{ab} u_a^{(n)} M_{ab} v_b^{(n)}$。

（提示：利用变换矩阵的正交性）

2.3 多线性代数中的张量单秩分解

Jupyter Notebook: sec2_3_svd_eqs

- **计算最大奇异值及奇异向量的迭代算法**

（a）随机初始化归一向量u和v;

（b）利用左边第一个式子，计算u和M的收缩并归一化，更新v，归一化因子记为Λ;

（c）利用左边第二个式子，计算v和M的收缩并归一化，更新u，归一化因子记为Λ;

（d）如果u和v（以及Λ）收敛，则返回u、v和Λ；否则，返回执行步骤（b）；

$$\sum_a u_a M_{ab} = \Lambda v_b$$

$$\sum_b M_{ab} v_b = \Lambda u_a$$

练习：

1. 证明或数值验证，仅最大奇异向量为上述迭代过程的稳定不动点；
2. 证明当矩阵为实对称实，上述算法可获得绝对值最大的本征值及对应的本征向量；
3. 写出程序数值验证第2条。

2.3 多线性代数中的张量单秩分解

Jupyter Notebook: sec2_3_rank1

- 自然地将上述自洽方程组推广到高阶张量的**单秩分解**（**rank-1分解**；以三阶张量T为例）$T=\Lambda u\otimes v\otimes w$，满足如下自洽方程组：

$$\sum_{ab} T_{abc}u_a v_b = \Lambda w_c$$

$$\sum_{ac} T_{abc}u_a w_c = \Lambda v_b$$

$$\sum_{bc} T_{abc}v_b w_c = \Lambda u_a$$

练习：画出方程组对应的图形表示

其中，u、v、w为归一化向量，方程组成立时有$\Lambda=\sum_{abc} T_{abc}u_a v_b w_c$，被称为rank-1分解系数。

- 类似于奇异值分解的迭代算法，上述方程组可迭代求解

2.3 多线性代数中的张量单秩分解

- Rank-1分解对应于如下**优化问题**：**给定张量$\boldsymbol{T}$，求解张量$\tilde{\boldsymbol{T}} = \boldsymbol{\Lambda} \prod_{\otimes n} \boldsymbol{v}^{(n)}$，使得$\boldsymbol{T}$和$\tilde{\boldsymbol{T}}$之间的范数$\boldsymbol{f} = |\boldsymbol{T} - \tilde{\boldsymbol{T}}|$极小。**
  （注：张量的范数这里定义为张量元平方和之后开方）
- 可以证明：rank-1分解得到的向量与系数构成的$\tilde{T}$，使范数f极小。
- 思考：当T为二阶张量时，rank-1分解与矩阵奇异值分解对应的最优化问题之间的关系。

练习：

请写出对任意四阶张量进行rank-1分解的程序，并在随机张量上测试程序，与Jupyter notebook中给出的任意阶数张量的rank-1分解程序比较计算结果。

2.4 高阶奇异值分解

- 在rank-1分解中，我们将最大奇异值及奇异向量的方法推广到了高阶张量，下面我们考虑**将完整的奇异值分解进行推广**
- **高阶奇异值分解**（Higher-order singular value decomposition，简称HOSVD），又称**Tucker分解**，其定义如下（以三阶实张量T为例）：

$$T_{abc} = \sum_{ijk} G_{ijk} U_{ai} V_{bj} W_{ck}$$

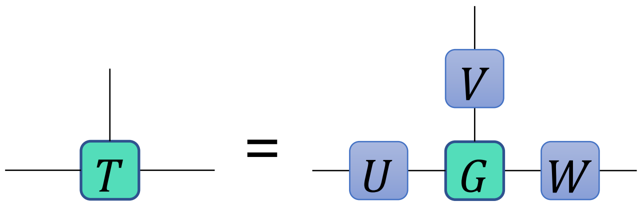
该式也被称为Tucker积

其中，变换矩阵满足正交性$UU^T = VV^T = WW^T = I$，张量G被称为核张量（core tensor），满足如下性质：

- 定义G的**键约化矩阵**（bond reduced matrix），以指标i为例，其键约化矩阵定义为

$$J_{ii'} = \sum_{jk} G_{ijk} G_{i'jk}$$

核张量的任意G的任意键约化矩阵为非负定对角阵，且元素按非升序排列（$J_{00} \geq J_{11} \geq \cdots \geq 0$）



2.4 高阶奇异值分解

Jupyter Notebook: sec2_4_hosvd

- 高阶奇异值分解的**HOSVD算法**（以三阶实张量T为例）

(a) 计算各个指标的键约化矩阵（如右式）；
(b) 计算每个键约化矩阵的本征值分解：

$$I = U\Omega U^T$$
$$J = V\Pi V^T$$
$$K = W\Upsilon W^T$$

(c) 计算核张量： $G_{ijk} = \sum_{abc} T_{abc} U_{ai} V_{bj} W_{ck}$

$$I_{aa'} = \sum_{jk} T_{abc} T_{a'bc}$$
$$J_{bb'} = \sum_{ik} T_{abc} T_{ab'c}$$
$$K_{cc'} = \sum_{ij} T_{abc} T_{abc'}$$

三个键约化矩阵的定义

最终得到高阶奇异值分解： $T_{abc} = \sum_{ijk} G_{ijk} U_{ai} V_{bj} W_{ck}$

* 各个本征谱中，非零值得个数被称为张量的**Tucker秩**
* 张量的Tucker低秩近似可通过HOSVD，HOOI等算法实现

练习：

1. 编写任意给定四阶张量HOSVD的程序，并于jupyter notebook中任意阶张量的分解结果对比；
2. 编写程序验证核张量满足的性质。

2.5 扩展：张量的CP分解与CP秩

- 回顾：单秩张量满足的形式为$\tilde{T} = \Lambda \prod_{\otimes n} v^{(n)}$，易得，单秩张量的Tucker秩为(1, 1, ⋯, 1)。
- 将单秩张量的形式稍作扩展，定义**CP积**，设N阶张量T为R个单秩张量的求和：

$$T = \sum_{r=0}^{R-1} \tilde{T}^{(r)} = \sum_{r=0}^{R-1} \Lambda^{(r)} \prod_{\otimes n=0}^{N-1} v^{(r,n)}$$

- 将Λ看成向量，$v^{(r,n)}$看成矩阵，则上式的CP形式也可写成如下的形式：

$$T_{a_1 a_2 \ldots} = \sum_{r=0}^{R-1} \Lambda_r \prod_{\otimes n=0}^{N-1} V_{r a_n}^{(n)}$$

其中，Λ为R维向量，$V^{(n)}$为对应于第n个指标的矩阵，由R个列向量组成。

2.5 扩展：张量的CP分解与CP秩

- 进一步改写CP积的形式

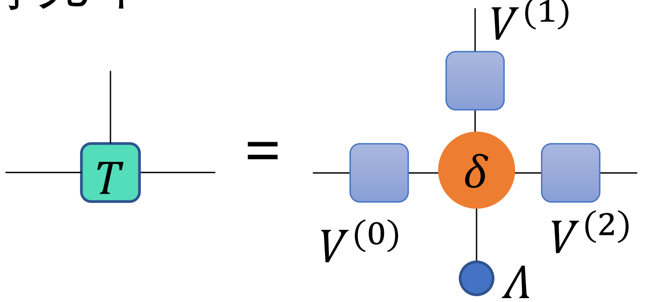
$$T_{a_1a_2\ldots}=\sum_{r=0}^{R-1}\Lambda_r\prod_{\otimes n=0}^{N-1}V_{ra_n}^{(n)}=\sum_{r=0}^{R-1}\sum_{r_1r_2\ldots}\delta_{r,r_1r_2\ldots}\Lambda_r\prod_{\otimes n=0}^{N-1}V_{ra_n}^{(n)}$$

其中，δ为(N+1)阶**推广的超对角张量**（N为张量T的阶数），第一个指标的维数为R，其余指标的维数与T指标的维数相等。

- 推广的超对角张量δ满足：对于任意r（逗号之前的指标），张量$\delta_{r,r_1r_2\ldots}$满足

$$\delta_{r,r_1r_2\ldots}=\begin{cases}1, & 当r_1=r_2=\cdots r_N \\ 0, & 其它情况下\end{cases},\quad \forall\, r$$

三阶CP积的图形表示:

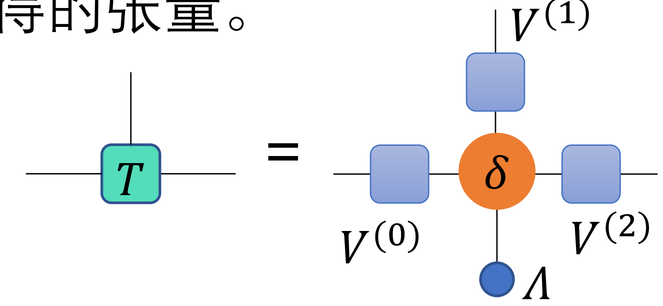


2.5 扩展：张量的CP分解与CP秩

- 定义**CP分解**：给定张量T，求解Λ 和V，满足$T_{a_1 a_2 \ldots} = \sum_{r=0}^{R-1} \Lambda_r \prod_{\otimes n=0}^{N-1} V_{r a_n}^{(n)}$
- 定义**CP秩**：使得CP分解严格成立的最小的R取值
- 定义**CP低秩近似**：给定R，使得CP分解式中左右之间的范数极小

CP分解还存在尚未解决的关键科学问题：

- 计算任意给定张量的CP秩是一个NP难问题；
- 目前存在多种CP低秩近似的算法，但是都无法保证获得的近似是最优的；
- 当取R=1时，CP分解退化成rank-1分解；但CP分解中贡献最大的项（如Λ 绝对值最大的元素对应的rank-1张量）不一定对应于rank-1分解所得的张量。



推荐一篇多线性代数的参考文献

Tensor Decompositions and Applications*

Tamara G. Kolda†
Brett W. Bader‡

Abstract. This survey provides an overview of higher-order tensor decompositions, their applications, and available software. A tensor is a multidimensional or N-way array. Decompositions of higher-order tensors (i.e., N-way arrays with $N \geq 3$) have applications in psychometrics, chemometrics, signal processing, numerical linear algebra, computer vision, numerical analysis, data mining, neuroscience, graph analysis, and elsewhere. Two particular tensor decompositions can be considered to be higher-order extensions of the matrix singular value decomposition: CANDECOMP/PARAFAC (CP) decomposes a tensor as a sum of rank-one tensors, and the Tucker decomposition is a higher-order form of principal component analysis. There are many other tensor decompositions, including INDSCAL, PARAFAC2, CANDELINC, DEDICOM, and PARATUCK2 as well as nonnegative variants of all of the above. The N-way Toolbox, Tensor Toolbox, and Multilinear Engine are examples of software packages for working with tensors.

Key words. tensor decompositions, multiway arrays, multilinear algebra, parallel factors (PARAFAC), canonical decomposition (CANDECOMP), higher-order principal components analysis (Tucker), higher-order singular value decomposition (HOSVD)

AMS subject classifications. 15A69, 65F99

重要内容总结

1. **张量的基本定义**：是用指标标记的一堆数
2. **不同的标记方式由reshape、transpose等命令相互转换**；
3. **张量的图形表示**：张量本身由节点（圆圈或方块等）表示，指标由连接节点的边表示；不同张量的共有指标由连接对应节点的边表示，默认进行求和；
4. **本征值分解**的形式与性质；
5. **最大本征问题对应的最优化问题**；
6. 最大本征问题的**幂级数求解法**；
7. **奇异值分解**的形式与性质，与矩阵**秩**的定义；
8. 基于奇异值分解的**矩阵低秩近似问题**，与**裁剪误差**；
9. 最大奇异值/向量的**迭代算法**；
10. 张量的**单秩分解**与**最优单秩近似问题**；
11. **高阶奇异值分解**与**Tucker秩**；
12. 高阶奇异值分解的**HOSVD算法**；
13. 张量的**Tucker低秩近似**问题（HOOI）；
14. **CP分解与CP秩**