CS 59300 – Algorithms for Data Science

Classical and Quantum approaches

Lecture 1 (08/28)

Tensor Methods (I)

https://ruizhezhang.com/course_fall_2025.html

Today's plan

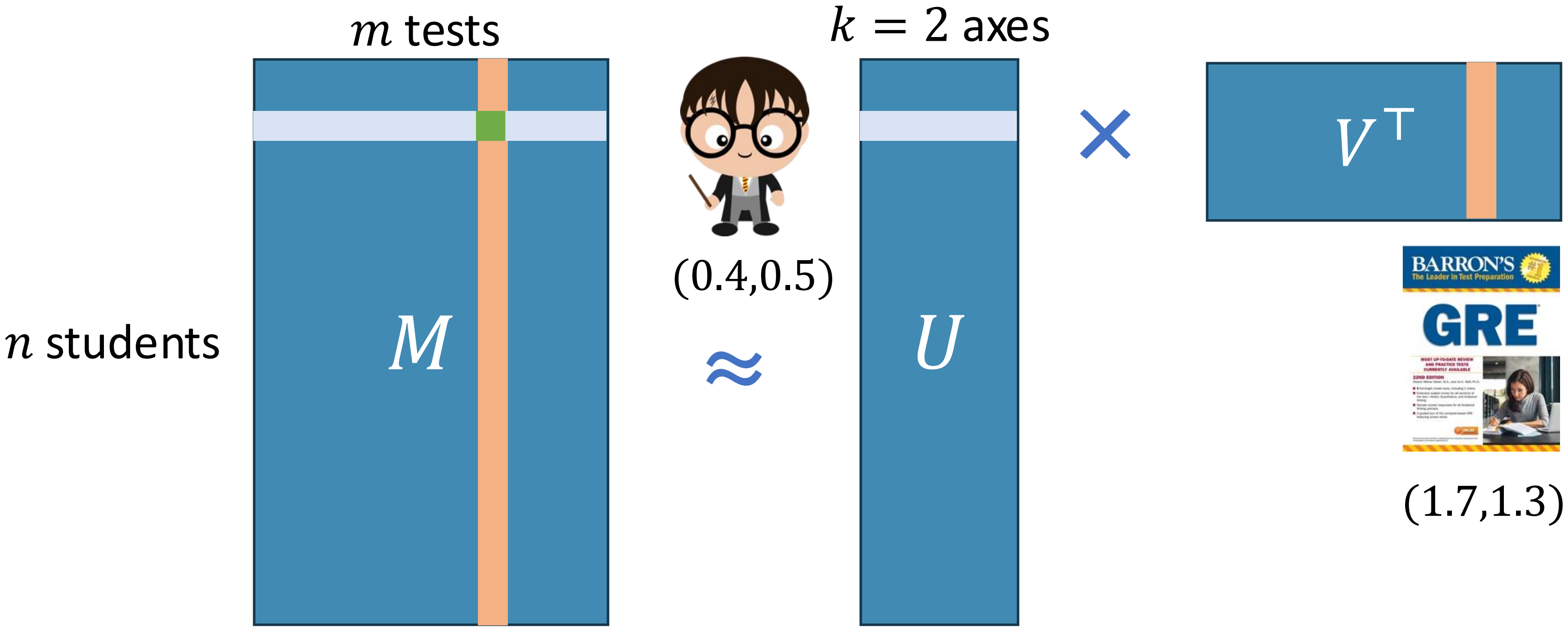
- Historical motivation
- Tensor decomposition algorithm (I): Jennrich's algorithm
- Applications of the tensor method

Slides is based on Ankur Moitra's notes and Sitan Chen's slides

Historical motivation

Factor analysis is a statistical method, pioneered by Charles Spearman, that explains observed correlations among many variables by modeling them as combinations of a few underlying latent factors.

Suppose a psychologist has the hypothesis that there are two kinds of intelligence, “verbal” and “mathematical”.

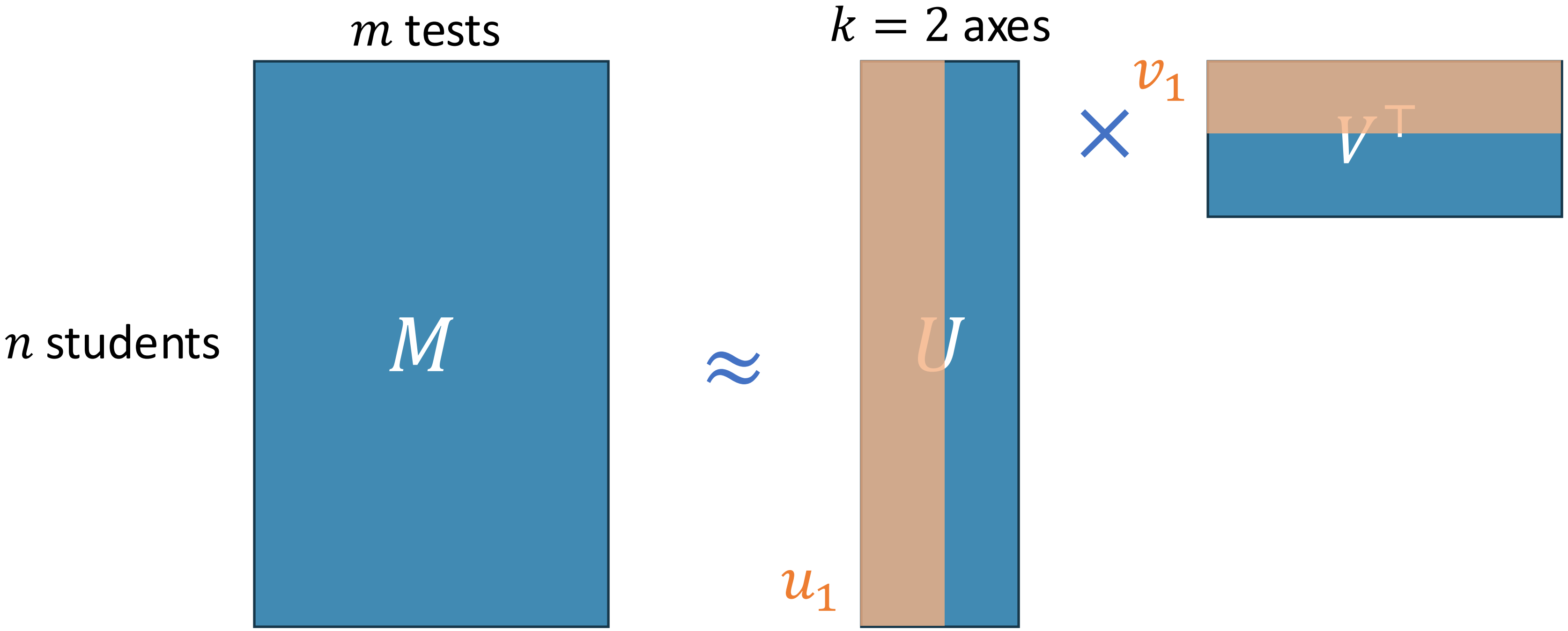


Charles Spearman
(1863-1945)

Historical motivation

Factor analysis is a statistical method, pioneered by Charles Spearman, that explains observed correlations among many variables by modeling them as combinations of a few underlying latent factors.

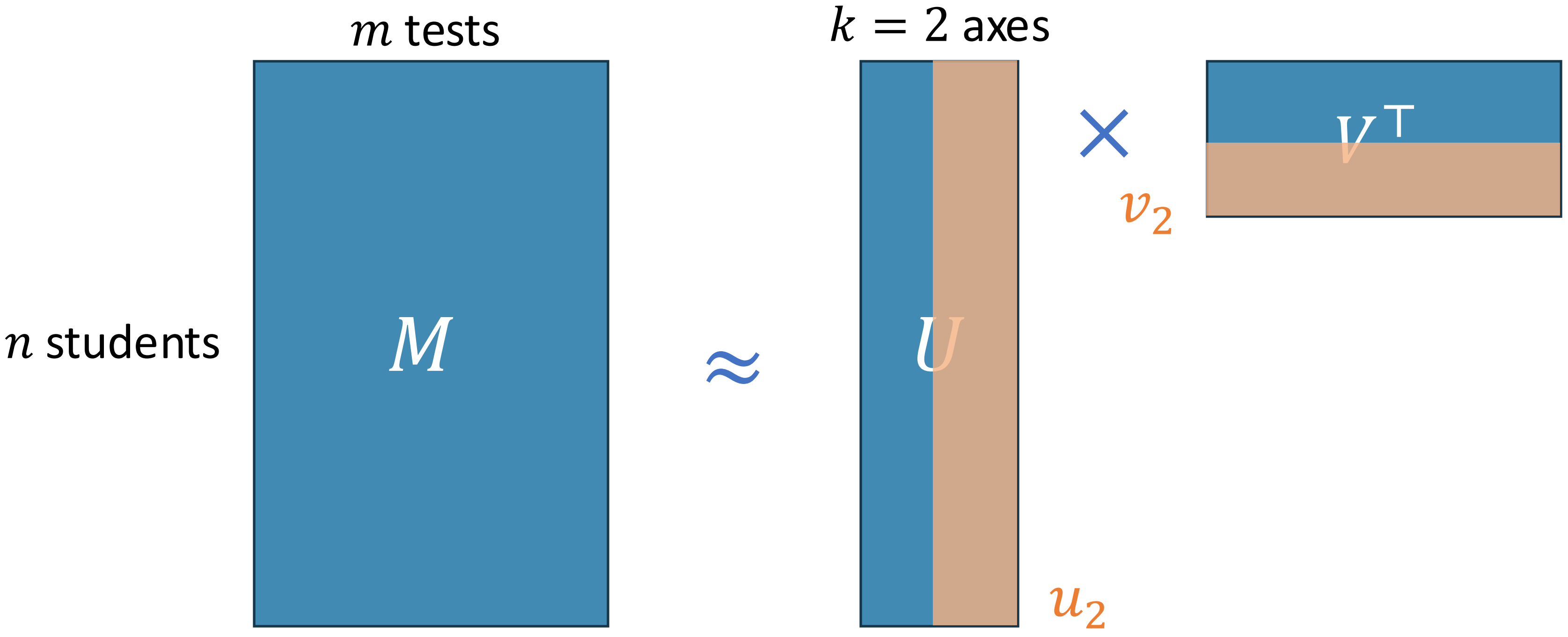
Suppose a psychologist has the hypothesis that there are two kinds of intelligence, “verbal” and “mathematical”.



Historical motivation

Factor analysis is a statistical method, pioneered by Charles Spearman, that explains observed correlations among many variables by modeling them as combinations of a few underlying latent factors.

Suppose a psychologist has the hypothesis that there are two kinds of intelligence, “verbal” and “mathematical”.



Historical motivation

Factor analysis is a statistical method, pioneered by Charles Spearman, that explains observed correlations among many variables by modeling them as combinations of a few underlying latent factors.

Suppose a psychologist has the hypothesis that there are two kinds of intelligence, "verbal" and "mathematical".

$$M = UV^\top = \sum_{i \in [k]} u_i v_i^\top$$

Issue: this factorization is not unique ("Rotation problem")

- Let $\tilde{U} \leftarrow UO, \tilde{V} \leftarrow VO$, where $O \in \mathbb{R}^{k \times k}$ is any orthogonal matrix
- $\tilde{U}\tilde{V}^\top = UOO^\top V^\top = UV^\top$

Charles Spearman
(1863-1945)

Historical motivation

Factor analysis is a statistical method, pioneered by Charles Spearman, that explains observed correlations among many variables by modeling them as combinations of a few underlying latent factors.

Suppose a psychologist has the hypothesis that there are two kinds of intelligence, "verbal" and "mathematical".

$$M = UV^\top = \sum_{i \in [k]} u_i v_i^\top$$

Issue: this factorization is not unique ("Rotation problem")

- Unless we put some additional assumptions, such as $\mathrm{rank}(M) = 1$, $\{u_k\}$ and $\{v_k\}$ are orthogonal, $\{u_k, v_k\}$ only have non-negative entries…

Charles Spearman
(1863-1945)

Tensor can help

Tensor product / Kronecker product

$$M = \sum_{i \in [k]} u_i v_i^\top = \sum_{i \in [k]} u_i \otimes v_i$$

If we can collect more data

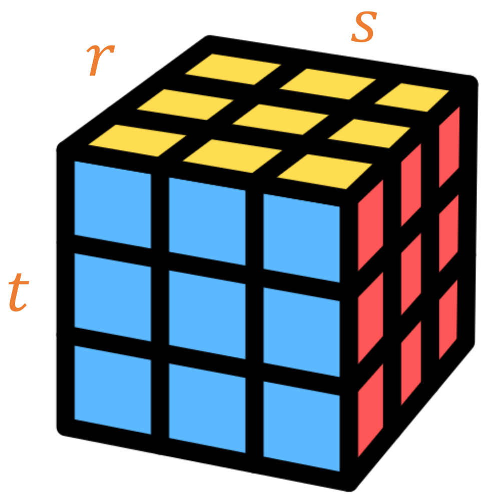
$$T = \sum_{i \in [k]} u_i \otimes v_i \otimes w_i$$

In this lecture, we'll see that there is no rotation problem for tensors.

Tensor basics

A third-order tensor $T \in \mathbb{R}^{r \times s \times t}$ is simply a three-dimensional array of numbers

Entries T_{abc} for $a \in [r],\ b \in [s],\ c \in [t]$



$$T = \sum_{i \in [k]} u_i \otimes v_i \otimes w_i$$

$u_i \otimes v_i \otimes w_i$ is a **rank-1 tensor** with entries given by

$$(u_i \otimes v_i \otimes w_i)_{abc} := (u_i)_a (v_i)_b (w_i)_c$$

The **rank** of T is the smallest number r such that T can be written as the sum of r rank-1 tensors.

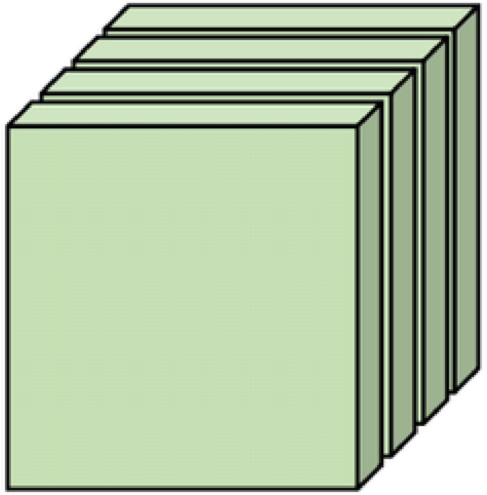
- For any $d \times d \times d$ tensor T, $\mathrm{rank}(T) \leq d^2$.

(Homework)

Tensor slicing

We can view tensor T as a stacked collection of matrices:

$$T_1 \coloneqq T(:,:,1), T_2 \coloneqq T(:,:,2), \text{etc}$$



Claim 1. If $\mathrm{rank}(T) \leq r$, then for all $a \in [t]$, $\mathrm{rank}(T_a) \leq r$.

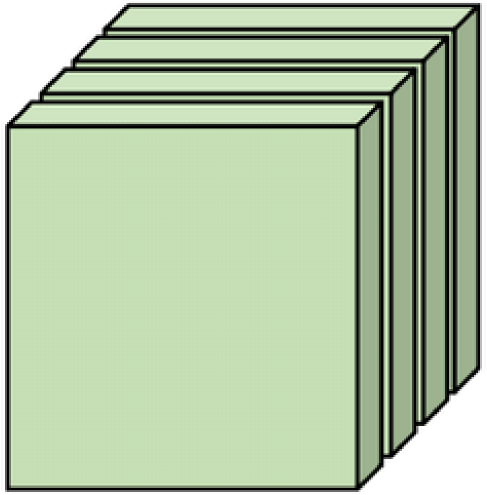
Proof.

$$T = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i \quad \Longrightarrow \quad T_a = \sum_{i=1}^{r} (w_i)_a \underbrace{(u_i \otimes v_i)}_{\text{rank-1 matrix}}$$

Tensor slicing

We can view tensor T as a stacked collection of matrices:

$$T_1 \coloneqq T(:,:,1), T_2 \coloneqq T(:,:,2), \text{etc}$$



Claim 1. If $\mathrm{rank}(T) \leq r$, then for all $a \in [t]$, $\mathrm{rank}(T_a) \leq r$.

However, a low-rank tensor is not just a collection of low-rank matrices!

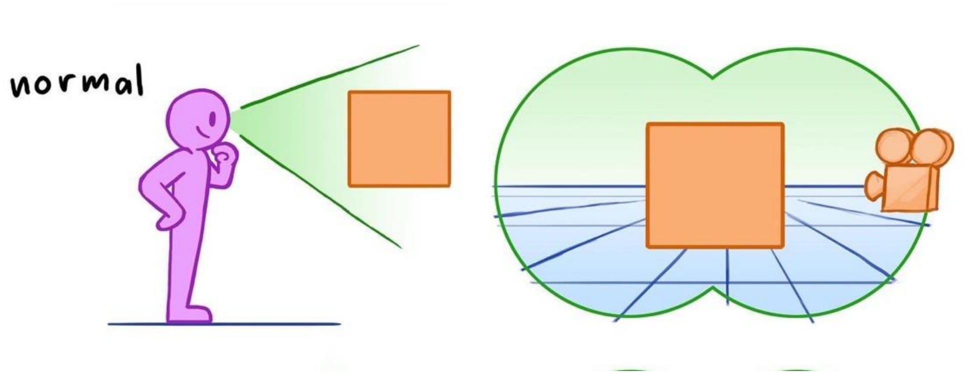
Claim 2. Consider a tensor $T = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i$. Then, for all $a \in [t]$, we have

- $\mathrm{colspan}(T_a) \subseteq \mathrm{span}(\{u_i\})$
- $\mathrm{rowspan}(T_a) \subseteq \mathrm{span}(\{v_i\})$

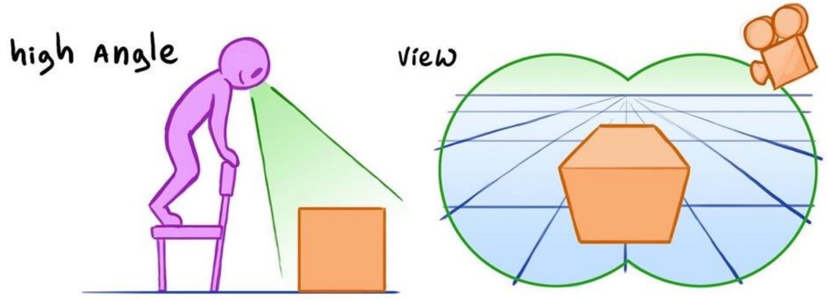
(Homework)

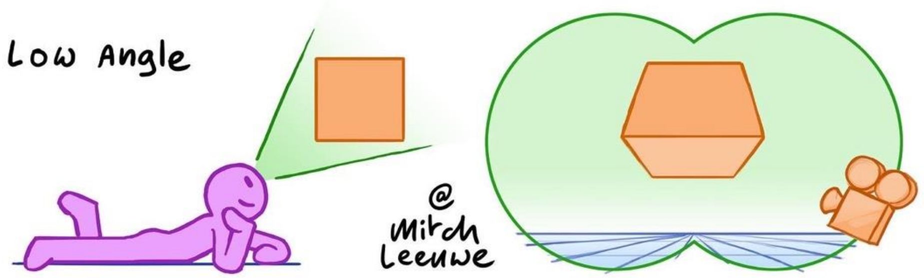
Intuition for why tensors do not suffer from the rotation problem

- **Matrix:** single "view" of $\{u_i\}$ and $\{v_i\}$



- **Tensor:** multiple "views"





The trouble with tensor

Many features of matrices that we take for granted simply do not hold for tensors

- The rank of a tensor depends on the field you are working over (i.e., $\mathrm{rank}_{\mathbb{R}} \neq \mathrm{rank}_{\mathbb{C}}$)

$$T = \left[\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\right]$$

$$= \frac{1}{2}\begin{bmatrix} 1 \\ -i \end{bmatrix} \otimes \begin{bmatrix} 1 \\ i \end{bmatrix} \otimes \begin{bmatrix} 1 \\ -i \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 1 \\ i \end{bmatrix} \otimes \begin{bmatrix} 1 \\ -i \end{bmatrix} \otimes \begin{bmatrix} 1 \\ i \end{bmatrix}$$

$\mathrm{rank}_{\mathbb{R}}(T) = 3$

$\mathrm{rank}_{\mathbb{C}}(T) = 2$

The trouble with tensor

Many features of matrices that we take for granted simply do not hold for tensors

- The rank of a tensor depends on the field you are working over (i.e., $\mathrm{rank}_{\mathbb{R}} \neq \mathrm{rank}_{\mathbb{C}}$)

- There are tensors of rank 3, but which are arbitrarily close to tensors of rank 2

$$T = \left[\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\right]$$

$$S = \left[\begin{bmatrix} n & 1 \\ 1 & 1/n \end{bmatrix}, \begin{bmatrix} 1 & 1/n \\ 1/n & 1/n^2 \end{bmatrix}\right] = \begin{bmatrix} 1 \\ 1/n \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 1/n \end{bmatrix} \otimes \begin{bmatrix} n \\ 1 \end{bmatrix}$$

$$R = \left[\begin{bmatrix} n & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}\right] = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} n \\ 0 \end{bmatrix}$$

$$S - R = \left[\begin{bmatrix} 0 & 1 \\ 1 & 1/n \end{bmatrix}, \begin{bmatrix} 1 & 1/n \\ 1/n & 1/n^2 \end{bmatrix}\right]$$

The **border rank** of T is the smallest number r such that $\forall \epsilon > 0$, $\exists\, T'$ of rank $\leq r$ such that T' is entry-wise ϵ close to T.

border rank $\neq$ rank for a tensor

The trouble with tensor

Many features of matrices that we take for granted simply do not hold for tensors

Computationally, basic linear algebraic primitives are intractable for tensors.

- **Hillar-Lim:** Most tensor problems are NP-hard

Eigenvalue over $\mathbb{R}$	NP-hard (Theorem 1.3)
Approximating Eigenvector over $\mathbb{R}$	NP-hard (Theorem 1.5)
Symmetric Eigenvalue over $\mathbb{R}$	NP-hard (Theorem 9.3)
Approximating Symmetric Eigenvalue over $\mathbb{R}$	NP-hard (Theorem 9.6)
Singular Value over $\mathbb{R}$, $\mathbb{C}$	NP-hard (Theorem 1.7)
Symmetric Singular Value over $\mathbb{R}$	NP-hard (Theorem 10.2)
Approximating Singular Vector over $\mathbb{R}$, $\mathbb{C}$	NP-hard (Theorem 6.3)
Spectral Norm over $\mathbb{R}$	NP-hard (Theorem 1.10)
Symmetric Spectral Norm over $\mathbb{R}$	NP-hard (Theorem 10.2)
Approximating Spectral Norm over $\mathbb{R}$	NP-hard (Theorem 1.11)
Nonnegative Definiteness	NP-hard (Theorem 11.2)
Best Rank-1 Approximation	NP-hard (Theorem 1.13)
Best Symmetric Rank-1 Approximation	NP-hard (Theorem 10.2)
Rank over $\mathbb{R}$ or $\mathbb{C}$	NP-hard (Theorem 8.2)

Tensor decomposition: Setup

Given a tensor $T \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ such that

$$T = \sum_{i=1}^{k} u_i \otimes v_i \otimes w_i$$

Our goal is to recover the set of factors $\{(u_i, v_i, w_i)\}$.

There are some symmetries in this decomposition.

- $\{(u_i, v_i, w_i)\}$ and $\{(\tilde{u}_i, \tilde{v}_i, \tilde{w}_i)\}$ are equivalent if there exists a permutation $\pi \in \mathcal{S}_k$ such that

$$u_i \otimes v_i \otimes w_i = \tilde{u}_{\pi(i)} \otimes \tilde{v}_{\pi(i)} \otimes \tilde{w}_{\pi(i)} \quad \forall\, i \in [k]$$

Main question: when are the factors of T are determined up to equivalence?

Tensor decomposition: theory

Theorem (Harshman, Jennrich).

Suppose the following conditions hold:

1) $\{u_i\}$ are linearly independent
2) $\{v_i\}$ are linearly independent
3) $d_3 \geq 2$ and no two w_i, w_j are collinear

Then the factors are uniquely determined up to equivalence, and there is a polynomial time algorithm to find them.

Tensor decomposition: Jennrich's algorithm

- Choose $a, b \in \mathbb{S}^{d_3}$ uniformly at random
- Set

$$M_a \coloneqq \sum_{i \in [d_3]} a_i T(:,:,i) \quad \text{and} \quad M_b \coloneqq \sum_{i \in [d_3]} b_i \underbrace{T(:,:,i)}_{d_1 \times d_2}$$

tensor contraction

- Compute $A \coloneqq M_a M_b^+$ and $B \coloneqq (M_a^+ M_b)^\top$
- Let $\hat{u}_1, \ldots, \hat{u}_k$ be eigenvectors of A with eigenvalues $\lambda_1, \ldots, \lambda_k$
- Let $\hat{v}_1, \ldots, \hat{v}_k$ be eigenvectors of B with eigenvalues $\lambda_1^{-1}, \ldots, \lambda_k^{-1}$
- Solve linear system to recover $\widehat{w}_1, \ldots, \widehat{w}_k$:

$$T = \sum_{i=1}^{k} \hat{u}_i \otimes \hat{v}_i \otimes \widehat{w}_i$$

Analysis of Jennrich's algorithm

Let $D_a \coloneqq \mathrm{diag}(\{\langle a, w_i\rangle\})$ and $D_b \coloneqq \mathrm{diag}(\{\langle b, w_i\rangle\})$

Lemma. We have that

$$M_a = U D_a V^\top \quad and \quad M_b = U D_b V^\top$$

$$U \coloneqq [u_1 \quad \cdots \quad u_k]$$
$$V \coloneqq [v_1 \quad \cdots \quad v_k]$$

Proof.

$$M_a \coloneqq \sum_{i \in [d_3]} a_i T(:,:,i) = \sum_{i \in [d_3]} a_i \sum_{j \in [k]} (u_j \otimes v_j)(w_j)_i$$

$$= \sum_{j \in [k]} u_j \otimes v_j \langle a, w_j \rangle = U D_a V^\top$$

Analysis of Jennrich's algorithm

Let $D_a := \operatorname{diag}(\{\langle a, w_i \rangle\})$ and $D_b := \operatorname{diag}(\{\langle b, w_i \rangle\})$

Lemma. We have that

$$M_a = U D_a V^\top \quad and \quad M_b = U D_b V^\top$$

$$U := [u_1 \quad \cdots \quad u_k]$$
$$V := [v_1 \quad \cdots \quad v_k]$$

Using the lemma, we have

$$\begin{aligned} A = M_a M_b^+ &= U D_a V^\top (U D_b V^\top)^+ \\ &= U D_a V^\top (V^\top)^+ D_b^{-1} U^+ \\ &= U D_a D_b^{-1} U^+ \end{aligned}$$

eigendecompositions

Similarly, we have

$$B = (M_a^+ M_b)^\top = \left((V^\top)^+ D_a^{-1} U^+ U D_b V^\top\right)^\top = \left((V^\top)^+ D_a^{-1} D_b V^\top\right)^\top = V D_b D_a^{-1} V^+$$

Analysis of Jennrich's algorithm: recover u and v

$$A = UD_aD_b^{-1}U^+ \quad \text{and} \quad B = VD_aD_b^{-1}V^+$$

Eigenvectors: $u_1, \dots, u_k$ $\qquad$ $v_1, \dots, v_k$ $\qquad$ (up to rescaling and permutation)

Eigenvalues: $\frac{\langle a, w_1 \rangle}{\langle b, w_1 \rangle}, \dots, \frac{\langle a, w_k \rangle}{\langle b, w_k \rangle}$ $\qquad$ $\frac{\langle b, w_1 \rangle}{\langle a, w_1 \rangle}, \dots, \frac{\langle b, w_k \rangle}{\langle a, w_k \rangle}$

1-1 correspondence

Loophole: what if the eigendecompositions of A and B are not unique?

By the randomness of a and b, and the condition 3) that no two w_i, w_j are collinear, we can guarantee that all the eigenvalues are non-zero and distinct.

(Homework)

Analysis of Jennrich's algorithm: recover w

$$\underbrace{T}_{\text{known}} = \sum_{i=1}^{r} \underbrace{\hat{u}_i \otimes \hat{v}_i}_{\text{known}} \otimes \underbrace{\hat{w}_i}_{\text{unknown}}$$

- #var $= r \times d_3$ and #eqs $= d_1 d_2 d_3$
- Need to show that this linear system has a unique solution

$$T_{abc} = \sum_i (u_i)_a (v_i)_b (w_i)_c = \langle \lambda^{ab}, W_{c,:} \rangle \qquad \lambda^{ab} := \begin{bmatrix} (u_1)_a (v_1)_b \\ \vdots \\ (u_k)_a (v_k)_b \end{bmatrix} \in \mathbb{R}^k$$

- Each $c \in [d_3]$ corresponds to an independent linear system (#var $= r$, #eqs $= d_1 d_2$)

Lemma. For any $c \in [d_3]$, $\{\lambda^{ab}\}_{a \in [d_1], b \in [d_2]}$ spans $\mathbb{R}^k$.

Analysis of Jennrich's algorithm: recover w

Lemma. For any $c \in [d_3]$, $\{\lambda^{ab}\}_{a\in[d_1],b\in[d_2]}$ spans $\mathbb{R}^r$.

Proof.

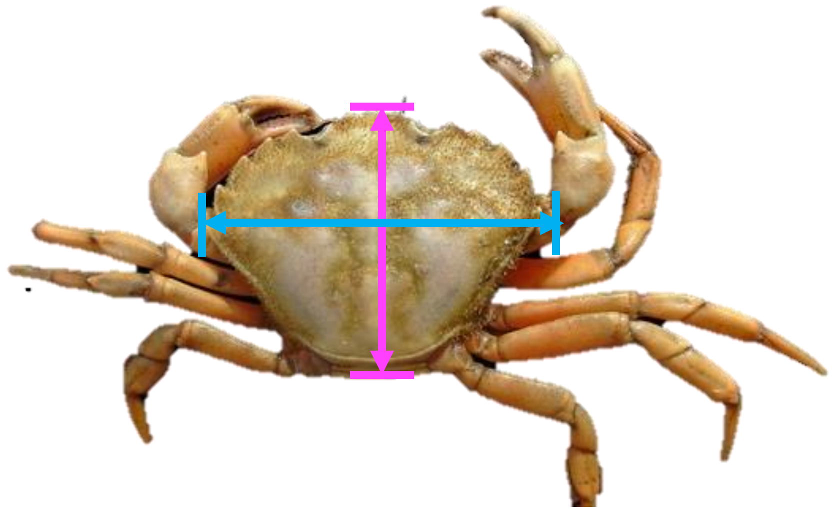
$$\Lambda \coloneqq \begin{bmatrix} (u_1)_1(v_1)_1 & \cdots & (u_k)_1(v_k)_1 \\ (u_1)_1(v_1)_2 & \cdots & (u_k)_1(v_k)_2 \\ \vdots & \ddots & \vdots \\ (u_1)_{d_1}(v_1)_{d_2} & \cdots & (u_k)_{d_1}(v_k)_{d_2} \end{bmatrix} \in \mathbb{R}^{d_1 d_2 \times k} \qquad \Lambda W_{c,:}^\top = T(:,:,c)$$

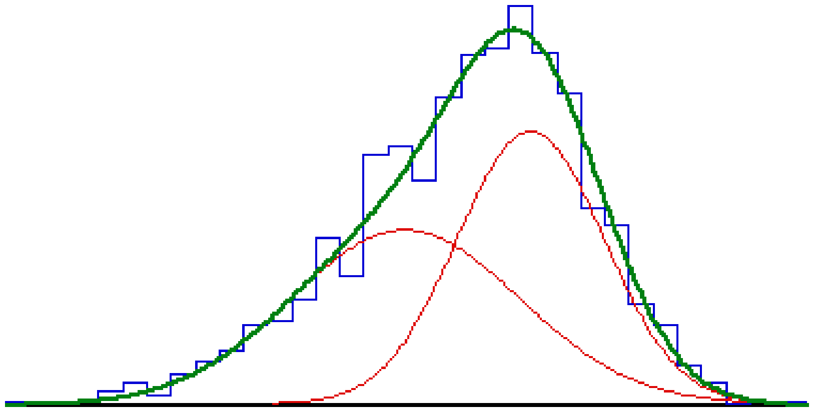
- Suppose $\exists\, c \in \mathbb{R}^k$ such that $\sum_{i\in[k]} c_i \Lambda_i = 0$. Wlog, assume $c_1 \neq 0$.
- Note that $\Lambda_i = \mathrm{vec}(u_i \otimes v_i)$. So $\sum_{i\in[k]} c_i u_i \otimes v_i = 0$
- Let $x \in \mathbb{R}^k$ such that $\langle x, u_1\rangle \neq 0$ while $\langle x, u_i\rangle = 0$ for all $i > 1$. (Why?)
- Then $$x^\top \sum_{i\in[k]} c_i\, u_i v_i^\top = c_1 \langle x, u_1\rangle v_1^\top + 0 = 0 \implies v_1 = 0$$ Contradiction!
- Thus, the solution of $W_{c,:}$ is unique

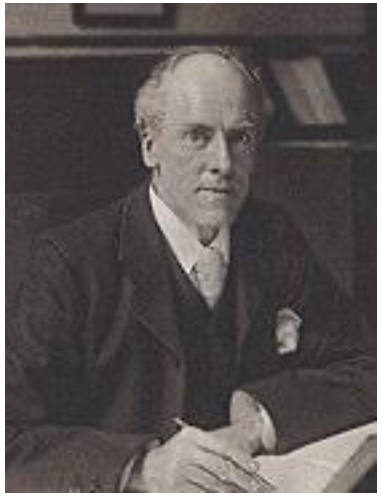
Application: Mixtures of Gaussians

Method of Moments

- Suppose we want to learn an unknown distribution q with parameters θ. But we can only draw samples from it
- We can use samples to estimate the moments $\mathbb{E}_{x\sim q}[p(x)]$ for some polynomials p
- The moments may contain enough information that allow us to “reverse-engineer” θ







Karl Pearson (1857-1936)

Application: Mixtures of Gaussians

Setup:

- Unknown distribution in $\mathbb{R}^d$:

$$q = \sum_{i \in [k]} \lambda_i \cdot \mathcal{N}(\mu_i, \mathrm{Id})$$

- Given i.i.d. samples from q, estimate $\{\mu_i\}$ and $\{\lambda_i\}$ up to small error
    1. Sample $i \in [k]$ with probability λ_i
    2. Sample from a Gaussian distribution $\mathcal{N}(\mu_i, \mathrm{Id})$

Application: Mixtures of Gaussians

$$q = \sum_{i \in [k]} \lambda_i \cdot \mathcal{N}(\mu_i, \mathrm{Id})$$

- First moment:

$$\mathbb{E}_{x \sim q}[x] = \sum_i \lambda_i \mu_i$$

- Third moment:

$$\mathbb{E}_{x \sim q}\left[x^{\otimes 3}\right] = \sum_i \lambda_i \mathbb{E}_{g \sim \mathcal{N}(0,\mathrm{Id})}\left[(\mu_i + g)^{\otimes 3}\right]$$

$$= \sum_i \lambda_i \mathbb{E}_{g \sim \mathcal{N}(0,\mathrm{Id})}\left[\begin{array}{c} \mu_i^{\otimes 3} + \cancel{\mu_i^{\otimes 2} \otimes g} + \cancel{\mu_i \otimes g \otimes \mu_i} + \mu_i \otimes g^{\otimes 2} + \cancel{g \otimes \mu_i^{\otimes 2}} \\ + g \otimes \mu_i \otimes g + g^{\otimes 2} \otimes \mu_i + \cancel{g^{\otimes 3}} \end{array}\right]$$

Application: Mixtures of Gaussians

- Third moment:

$$\mathbb{E}_{x\sim q}\left[x^{\otimes 3}\right] = \sum_i \lambda_i \mu_i^{\otimes 3} + \sum_i \lambda_i \mathbb{E}_{g\sim\mathcal{N}(0,\mathrm{Id})}\left[\mu_i \otimes g^{\otimes 2} + g \otimes \mu_i \otimes g + g^{\otimes 2} \otimes \mu_i\right]$$

Id

$$= \sum_i \lambda_i \mu_i^{\otimes 3} + \sum_i \lambda_i \left(\mu_i \otimes \mathrm{Id} + \mathrm{Id} \otimes \mu_i + \sum_{a\in[d]} e_a \otimes \mu_i \otimes e_a\right)$$

$$= \sum_i \lambda_i \mu_i^{\otimes 3} + \mathbb{E}_{x\sim q}[x] \otimes \mathrm{Id} + \mathrm{Id} \otimes \mathbb{E}_{x\sim q}[x] + \sum_{a\in[d]} e_a \otimes \mathbb{E}_{x\sim q}[x] \otimes e_a$$

Thus, we get that

$$\sum_i \lambda_i \mu_i^{\otimes 3} = \mathbb{E}_{x\sim q}\left[x^{\otimes 3} + x \otimes \mathrm{Id} + \mathrm{Id} \otimes x + \sum_{a\in[d]} e_a \otimes x \otimes e_a\right]$$

Application: Mixtures of Gaussians

Algorithm:

- Use samples to estimate $T = \mathbb{E}_{x\sim q}\left[x^{\otimes 3} + x \otimes \mathrm{Id} + \mathrm{Id} \otimes x + \sum_{a\in[d]} e_a \otimes x \otimes e_a\right]$
- Run Jennrich's algorithm to recover $\left\{\lambda_i^{1/3}\mu_i\right\}_{i\in[k]}$
- Solve a linear system to recover λ_i:

$$\sum_{i\in[k]} \left(\lambda_i^{1/3}\mu_i\right) \cdot \lambda_i^{2/3} = \mathbb{E}_{x\sim q}[x]$$

Bonus: Perturbation analysis for Jennrich's algorithm

- Choose $a, b \in \mathbb{S}^{d_3}$ uniformly at random
- Set

$$\widetilde{M_a} \coloneqq \sum_{i \in [d_3]} a_i \widetilde{T}(:,:,i) \quad \text{and} \quad \widetilde{M_b} \coloneqq \sum_{i \in [d_3]} b_i \widetilde{T}(:,:,i)$$

- Compute $A \coloneqq M_a M_b^+$ and $B \coloneqq (M_a^+ M_b)^\top$
- Let $\hat{u}_1, \dots, \hat{u}_r$ be eigenvectors of A with eigenvalues $\lambda_1, \dots, \lambda_r$
- Let $\hat{v}_1, \dots, \hat{v}_r$ be eigenvectors of B with eigenvalues $\lambda_1^{-1}, \dots, \lambda_r^{-1}$
- Solve linear system to recover $\widehat{w}_1, \dots, \widehat{w}_r$:

$$T = \sum_{i=1}^{r} \hat{u}_i \otimes \hat{v}_i \otimes \widehat{w}_i$$

$$\tilde{A} = \widetilde{M}_a \widetilde{M}_b^+ = A + E$$

How does the error affect the eigenvectors of A?

Perturbation analysis

The condition number of a matrix A is defined as

$$\kappa(A) \coloneqq \sigma_{\max}(A)/\sigma_{\min}(A) = \kappa(A^{-1})$$

- Consider a linear system $Ax = b$
- Let $\tilde{x}$ be the perturbed solution of $Ax = \tilde{b} = b + e$
- $\tilde{x} - x = A^{-1}(\tilde{b} - b) = A^{-1}e$
- So the relative error is:

$$\frac{\|\tilde{x} - x\|}{\|x\|} = \frac{\|A^{-1}e\|}{\|A^{-1}b\|} \leq \frac{\sigma_{\max}(A^{-1})\|e\|}{\sigma_{\min}(A^{-1})\|b\|} = \kappa(A)\frac{\|\tilde{b} - b\|}{\|b\|}$$

Perturbation analysis

$$\tilde{A} = A + E = UDU^{-1} + E$$

1. Show that $\tilde{A}$ is diagonalizable
2. Show that the matrix that diagonalizes $\tilde{A}$ is close to U

We first consider the second part

- Let $\tilde{A} = \widetilde{U}\widetilde{D}\widetilde{U}^{-1}$. How close is $(\tilde{u}_i, \tilde{\lambda}_i)$ to (u_i, λ_i)?
- Let's assume that $\tilde{\lambda}_i \approx \lambda_i$, and the λ_i's are well-separated
- We can expand $\tilde{u}_i$ in the basis of $\{u_i\}$ as $\tilde{u}_i = \sum_j c_j u_j$
- Multiplying $\tilde{A}$ gives

$$\tilde{\lambda}_i \tilde{u}_i = \sum_j c_j \tilde{A} u_j = \sum_j c_j \lambda_j u_j + \sum_j c_j E u_j = \sum_j c_j \lambda_j u_j + E\tilde{u}_i$$

Perturbation analysis

$$\tilde{\lambda}_i \tilde{u}_i = \sum_j c_j \tilde{A} u_j = \sum_j c_j \lambda_j u_j + \sum_j c_j E u_j = \sum_j c_j \lambda_j u_j + E \tilde{u}_i$$

$$\sum_j c_j (\lambda_j - \tilde{\lambda}_i) u_j = -E \tilde{u}_i$$

- For any $\ell \in [d]$, let $U^{-1}_{\ell,:}$ be the ℓ-th row of U^{-1}.
- Multiplying $U^{-1}_{\ell,:}$ on both sides, we get:

$$U^{-1}_{\ell,:} \sum_j c_j (\lambda_j - \tilde{\lambda}_i) u_j = \sum_j c_j (\lambda_j - \tilde{\lambda}_i) \delta_{\ell j} = c_\ell (\lambda_\ell - \tilde{\lambda}_i) = -U^{-1}_{\ell,:} E \tilde{u}_i$$

$\Rightarrow$ $$|c_\ell| = \frac{|U^{-1}_{\ell,:} E \tilde{u}_i|}{|\lambda_\ell - \tilde{\lambda}_i|} \le \frac{\|U^{-1}\| \cdot \|E\| \cdot \|\tilde{u}_i\|}{\Delta} = \frac{\|U^{-1}\| \cdot \|E\|}{\Delta} \qquad \forall\, \ell \neq i$$

$\Rightarrow$ $|c_i|$ is large since $\|c\| = 1$, which means $\tilde{u}_i \approx u_i$

Perturbation analysis

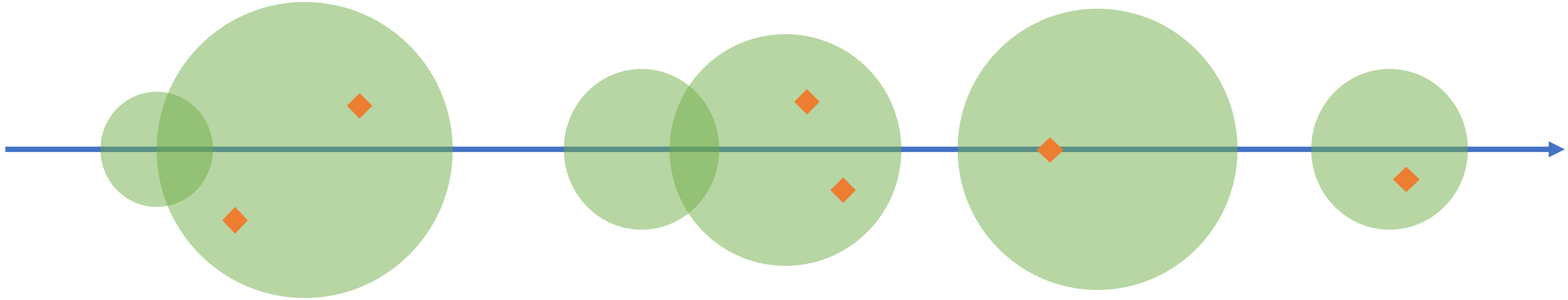
Theorem (Gershgorin's disk theorem).

The eigenvalues of A are contained in in the following union of disks in the complex plane:

$$\bigcup_i \mathcal{D}(A_{ii}, R_i)$$

where $\mathcal{D}(a,b) \coloneqq \{z \in \mathbb{C} \,|\, |z-a| \le b\}$ and $R_i \coloneqq \sum_{j \ne i} |A_{ij}|$.

Moreover, if one disk is disjoint from others, then there must be one eigenvalue in it.



Perturbation analysis

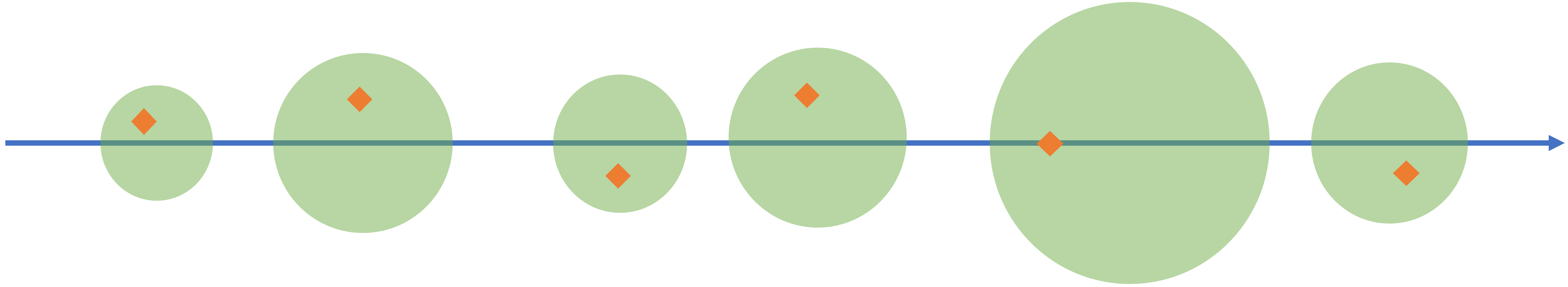
Theorem (Gershgorin's disk theorem).

The eigenvalues of A are contained in in the following union of disks in the complex plane:

$$\bigcup_i \mathcal{D}(A_{ii}, R_i)$$

where $\mathcal{D}(a,b) \coloneqq \{z \in \mathbb{C} \,|\, |z-a| \le b\}$ and $R_i \coloneqq \sum_{j \ne i} |A_{ij}|$.

Moreover, if one disk is disjoint from others, then there must be one eigenvalue in it.



Perturbation analysis

Now we use Gershgorin's disk theorem to prove that $\tilde{A}$ is diagonalizable.

- Recall $\tilde{A} = A + E = UDU^{-1} + E$.
- We'll show that $\tilde{A}$ has distinct eigenvalues
- Consider $U^{-1}\tilde{A}U = D + U^{-1}EU$, which has the same spectrum as $\tilde{A}$
- By Gershgorin's disk theorem, all the eigenvalues are contained in

$$\bigcup_i \mathcal{D}(\lambda_i, R_i)$$

- If $\max R_i \le \frac{1}{2} \cdot \underbrace{\min_{i \ne j} |\lambda_i - \lambda_j|}_{\Delta}$, then we are done
- $\|U^{-1}EU\|_{\max} \le \|U^{-1}EU\| \le \kappa(U)\|E\|$ $\Rightarrow$ $R_i \le (n-1)\kappa(U)\|E\|$
- As long as $\|E\| \lesssim \frac{\Delta}{n\kappa(U)}$, $\{\tilde{\lambda}_i\}$ are disjoint and close to $\{\lambda_i\}$

Back to Jennrich's algorithm

- Set

$$M_a \coloneqq \sum_{i \in [d_3]} a_i T(:,:,i) \quad \text{and} \quad M_b \coloneqq \sum_{i \in [d_3]} b_i T(:,:,i)$$

- Compute $A \coloneqq M_a M_b^+$

$$\tilde{A} = \widetilde{M}_a \widetilde{M}_b^+ = A + E$$

- Let $\hat{u}_1, \dots, \hat{u}_r$ be eigenvectors of A with eigenvalues $\lambda_1, \dots, \lambda_r$

We need to guarantee that:

$$E = \widetilde{M}_a \widetilde{M}_b^+ - M_a M_b^+ \text{ is small,}$$

provided $\tilde{T} \approx T$.

- Follows from more tedious perturbation bounds

Recap

- In the factor analysis, matrix suffers from the Rotation Problem. And we understand when and why tensor does not suffer.

- We introduce the Jennrich's algorithm (or simultaneous diagonalization), which is a rigorous approach to decompose low-rank tensors

- We also discuss an application of learning mixture of Gaussians using the method of moments

- In the next lecture, we will talk about more practical tensor decomposition algorithms based on optimization