Tensor Decompositions: Algorithms and Uniqueness

Benjamin Lovitz

NSF Postdoc, Department of Mathematics, Northeastern University

2022 SIAM Conference on Mathematics of Data Science

MS168: Multilinear Algebra and Tensors for Data Science

September 30, 2022

arXiv:2103.15633

What is a matrix?

A matrix is an element of $\mathbb{C}^{d_x} \otimes \mathbb{C}^{d_y}$ $\begin{bmatrix} 2 & 4 \\ 6 & 12 \end{bmatrix} \in \mathbb{C}^2 \otimes \mathbb{C}^2$

A rank-one matrix is a matrix of the form $x \otimes y = xy^T = \left(x_i y_j\right)_{(i,j)}$

What is a matrix?

A matrix is an element of $\mathbb{C}^{d_x} \otimes \mathbb{C}^{d_y}$ $\begin{bmatrix} 2 & 4 \\ 6 & 12 \end{bmatrix} = \binom{1}{3}(2\ 4)$

A rank-one matrix is a matrix of the form $x \otimes y = xy^T = \left(x_i y_j\right)_{(i,j)}$

What is a ~~matrix~~ tensor?

A ~~matrix~~ tensor is an element of $\mathbb{C}^{d_x} \otimes \mathbb{C}^{d_y} \otimes \mathbb{C}^{d_z}$ $\begin{bmatrix} 15 & 18 \\ 20 & 24 \end{bmatrix} \begin{bmatrix} 30 & 36 \\ 40 & 48 \end{bmatrix} \in \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$

A product tensor is a tensor of the form $x \otimes y \otimes z = \left(x_i y_j z_k\right)_{(i,j,k)}$

What is a ~~matrix~~ tensor?

A ~~matrix~~ tensor is an element of $\mathbb{C}^{d_x} \otimes \mathbb{C}^{d_y} \otimes \mathbb{C}^{d_z}$

$$\begin{bmatrix} 15 & 18 \\ 20 & 24 \end{bmatrix} \begin{bmatrix} 30 & 36 \\ 40 & 48 \end{bmatrix} = \binom{1}{2} \otimes \binom{3}{4} \otimes \binom{5}{6}$$

A product tensor is a tensor of the form $x \otimes y \otimes z = \left(x_i y_j z_k\right)_{(i,j,k)}$

Tensor decompositions

Definition: Let $n \in \mathbb{N}$ and $[n] \coloneqq \{1, \dots, n\}$.

For $T \in \mathbb{C}^{d_x} \otimes \mathbb{C}^{d_y} \otimes \mathbb{C}^{d_z}$, an expression $T = \sum_{a \in [n]} x_a \otimes y_a \otimes z_a \in \mathbb{C}^{d_x} \otimes \mathbb{C}^{d_y} \otimes \mathbb{C}^{d_z}$

is called a decomposition of T into product tensors

$$\mathrm{rank}(T) := \min\{ n : \text{there exists a decomposition of } T \text{ into } n \text{ product tensors}\}$$

Uniqueness of tensor decompositions

<u>Definition:</u> Let $n \in \mathbb{N}$ and $[n] := \{1, \ldots, n\}$.

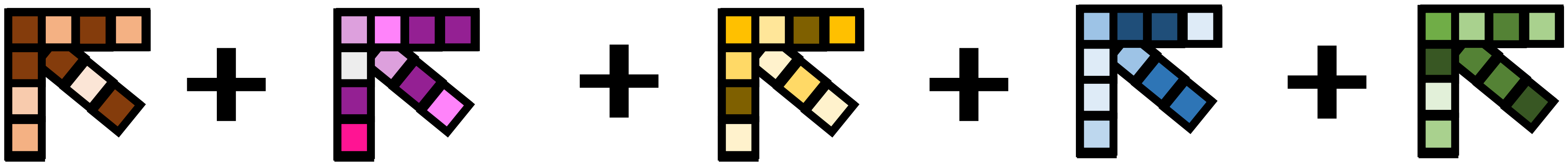
A rank decomposition

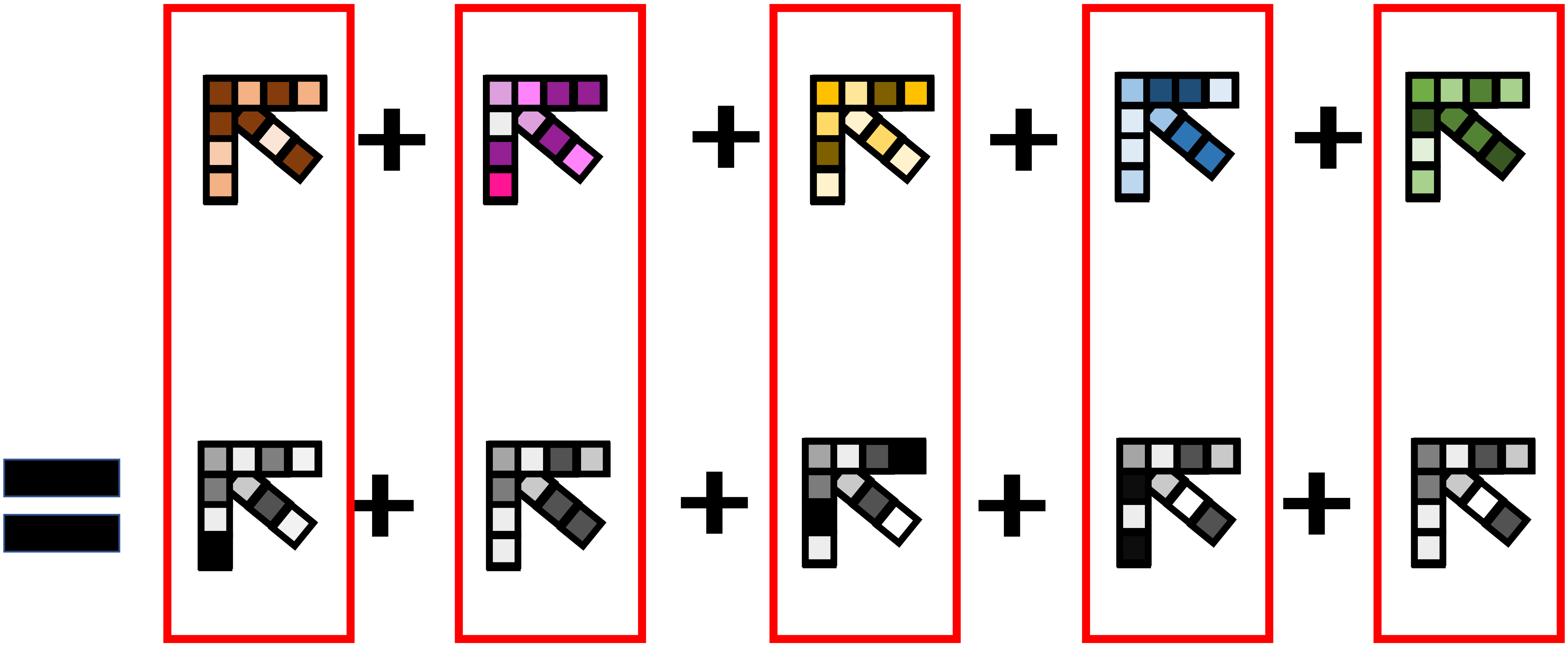
$$T = \sum_{a \in [n]} x_a \otimes y_a \otimes z_a \in \mathbb{C}^{d_x} \otimes \mathbb{C}^{d_y} \otimes \mathbb{C}^{d_z}$$

is called the unique (rank) decomposition of T if for any other decomposition

$$T = \sum_{a \in [n]} x'_a \otimes y'_a \otimes z'_a \in \mathbb{C}^{d_x} \otimes \mathbb{C}^{d_y} \otimes \mathbb{C}^{d_z}$$

there is a permutation $\sigma \in S_n$ such that $x_a \otimes y_a \otimes z_a = x'_{\sigma(a)} \otimes y'_{\sigma(a)} \otimes z'_{\sigma(a)}$ for all $a \in [n]$.





Application: Latent parameter learning

L is for latent

- Let A, B, C, L be finite random variables such that A, B, C are conditionally independent, i.e.

$$\Pr(a, b, c|l) = \Pr(a|l)\Pr(b|l)\Pr(c|l) \quad \text{for all } a, b, c, l.$$

- Goal: Given the probability vector $\Pr(A, B, C)$, determine $\Pr(A, B, C, L)$.
- Method:

$$\Pr(A, B, C) = \sum_l \Pr(l)\Pr(A, B, C|l) = \sum_l \underbrace{\Pr(l)\Pr(A|l) \otimes \Pr(B|l) \otimes \Pr(C|l)}$$

… If $\Pr(A, B, C)$ has a unique decomposition, then we can recover $\Pr(A, B, C, l)$,

- Applications: Learning mixtures of spherical gaussians, phylogenetic tree reconstruction, hidden Markov models, orbit retrieval, blind signal separation, document topic models, …

Two goals

$$T = \sum_{a \in [n]} x_a \otimes y_a \otimes z_a \in \mathbb{C}^{d_x} \otimes \mathbb{C}^{d_y} \otimes \mathbb{C}^{d_z} \qquad (1)$$

1. Algorithms

Given a tensor $T \in \mathbb{C}^{d_x} \otimes \mathbb{C}^{d_y} \otimes \mathbb{C}^{d_z}$, find a rank decomposition (1).

2. Uniqueness

Given a rank decomposition (1), prove that it is the unique rank decomposition.

Two goals

$$T = \sum_{a \in [n]} x_a \otimes y_a \otimes z_a \in \mathbb{C}^{d_x} \otimes \mathbb{C}^{d_y} \otimes \mathbb{C}^{d_z} \qquad (1)$$

1. **Algorithms**

Given a tensor $T \in \mathbb{C}^{d_x} \otimes \mathbb{C}^{d_y} \otimes \mathbb{C}^{d_z}$, find a rank decomposition (1).

2. **Uniqueness**

Given a rank decomposition (1), prove that it is the unique rank decomposition.

Algorithms

$$T = \sum_{a \in [n]} x_a \otimes y_a \otimes z_a \in \mathbb{C}^{d_x} \otimes \mathbb{C}^{d_y} \otimes \mathbb{C}^{d_z} \quad (1)$$

Jennrich's Algorithm [1970]: Finds the decomposition (1) efficiently for generic tensors in $\mathbb{C}^2 \otimes \mathbb{C}^d \otimes \mathbb{C}^d$ of rank $n \leq O(d)$.

Image-Based Algorithm [Johnston-L-Vijayaraghavan 2022+?]:

Finds the decomposition (1) efficiently for generic tensors in $\mathbb{C}^{d_x} \otimes \mathbb{C}^{d_y} \otimes \mathbb{C}^{d_z}$ of rank $n < \min\{d_z + 1, (d_x - 1)(d_y - 1)/4\}$.

Example [Extends De Lathauwer et al 2007]: Finds (1) for generic tensors in $\mathbb{C}^d \otimes \mathbb{C}^d \otimes \mathbb{C}^{d^2}$ of rank $n \leq O(d^2)$.

Equal to generic rank up to constant

Algorithms

$$T = \sum_{a\in[n]} x_a \otimes y_a \otimes z_a \in \mathbb{C}^{d_x} \otimes \mathbb{C}^{d_y} \otimes \mathbb{C}^{d_z} \quad (1)$$

Proof sketch of Example:

1. View T as a map $T: \mathbb{C}^{d^2} \to \mathbb{C}^d \otimes \mathbb{C}^d$
2. For generic T of rank $n \leq O(d^2)$, the image of T will contain exactly n product tensors $\{x_a \otimes y_a : a \in [n]\}$. Find them, then find the z's.

<u>Example [Extends De Lathauwer et al 2007]:</u> Finds (1) for generic tensors in $\mathbb{C}^d \otimes \mathbb{C}^d \otimes \mathbb{C}^{d^2}$ of rank $n \leq O(d^2)$.

Equal to generic rank up to constant

Two goals

$$T = \sum_{a \in [n]} x_a \otimes y_a \otimes z_a \in \mathbb{C}^{d_x} \otimes \mathbb{C}^{d_y} \otimes \mathbb{C}^{d_z} \qquad (1)$$

1. <u>Algorithms</u>

   Given a tensor $T \in \mathbb{C}^{d_x} \otimes \mathbb{C}^{d_y} \otimes \mathbb{C}^{d_z}$, find a rank decomposition (1).

2. <u>Uniqueness</u>

   Given a rank decomposition (1), prove that it is the unique rank decomposition.

Uniqueness

Jennrich's Uniqueness Theorem: Given a rank decomposition

$$T = \sum_{a \in [d]} x_a \otimes y_a \otimes z_a \in \mathbb{C}^d \otimes \mathbb{C}^d \otimes \mathbb{C}^2 \qquad (1)$$

If it holds that

1. $\{x_1, \dots, x_d\} \subseteq \mathbb{C}^d$ is linearly independent,
2. $\{y_1, \dots, y_d\} \subseteq \mathbb{C}^d$ is linearly independent,
3. and $\{z_1, \dots, z_d\} \subseteq \mathbb{C}^2$ are non-parallel

then (1) is the unique rank decomposition of T.

Jennrich's Algorithm: Finds the decomposition (1) efficiently!

Matroid theory for product tensors

$$T = \sum_{a \in [n]} x_a \otimes y_a \otimes z_a \in \mathbb{C}^{d_x} \otimes \mathbb{C}^{d_y} \otimes \mathbb{C}^{d_z} \qquad (1)$$

- Recall the general setup: We are handed a set of product tensors $\{x_a \otimes y_a \otimes z_a : a \in [n]\}$, and want to determine if their sum (1) is a unique rank decomposition.

- Natural tool: Matroid theory (the study of finite sets of vectors).

- Line of attack: Determine matroidal properties of sets of product tensors.

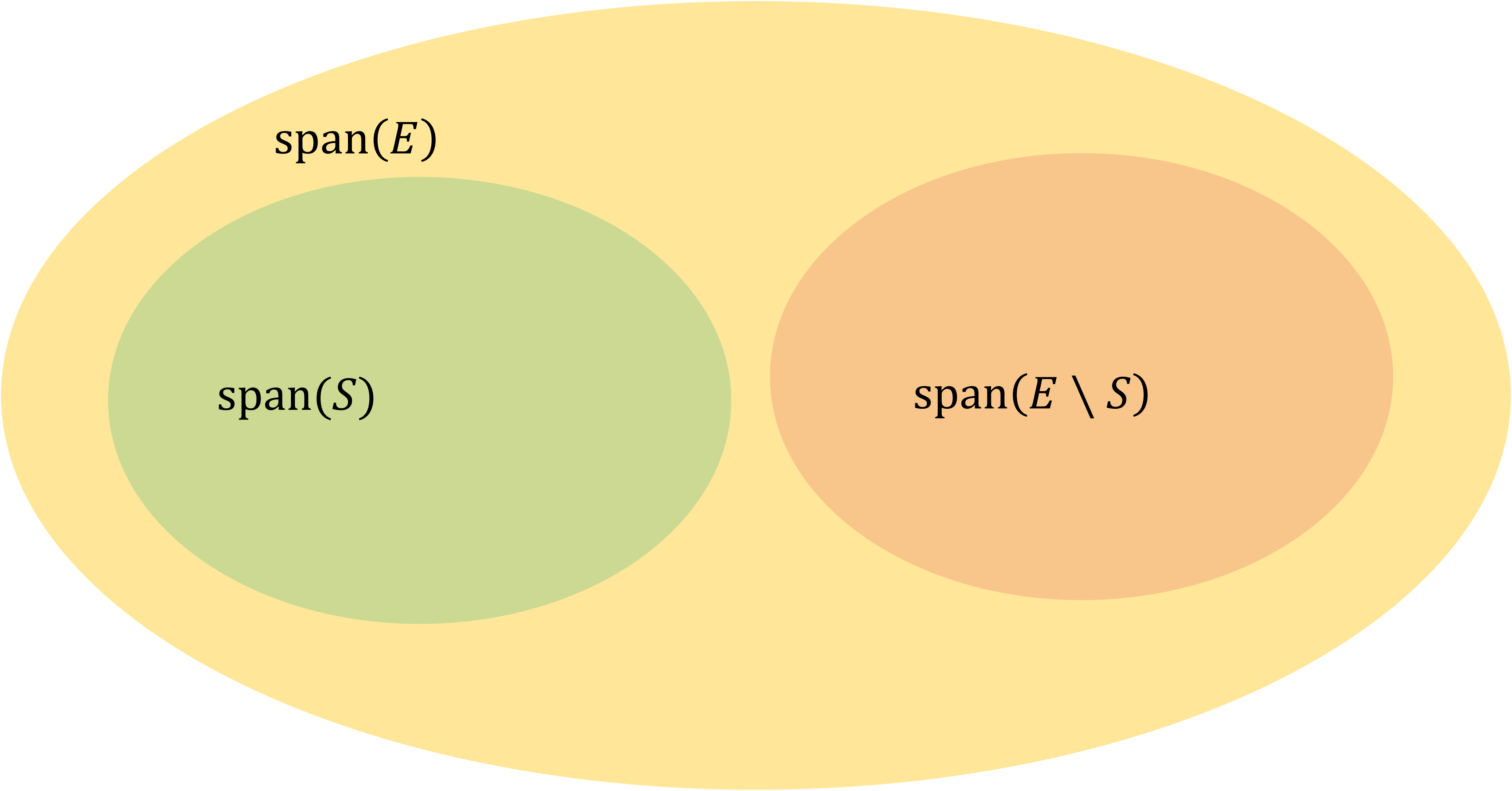
Rest of talk: A splitting theorem for product tensors

Splitting

Definition: A set of vectors $E = \{v_1, \ldots, v_n\}$ splits if there exists a non-trivial subset $S \subseteq E$ such that

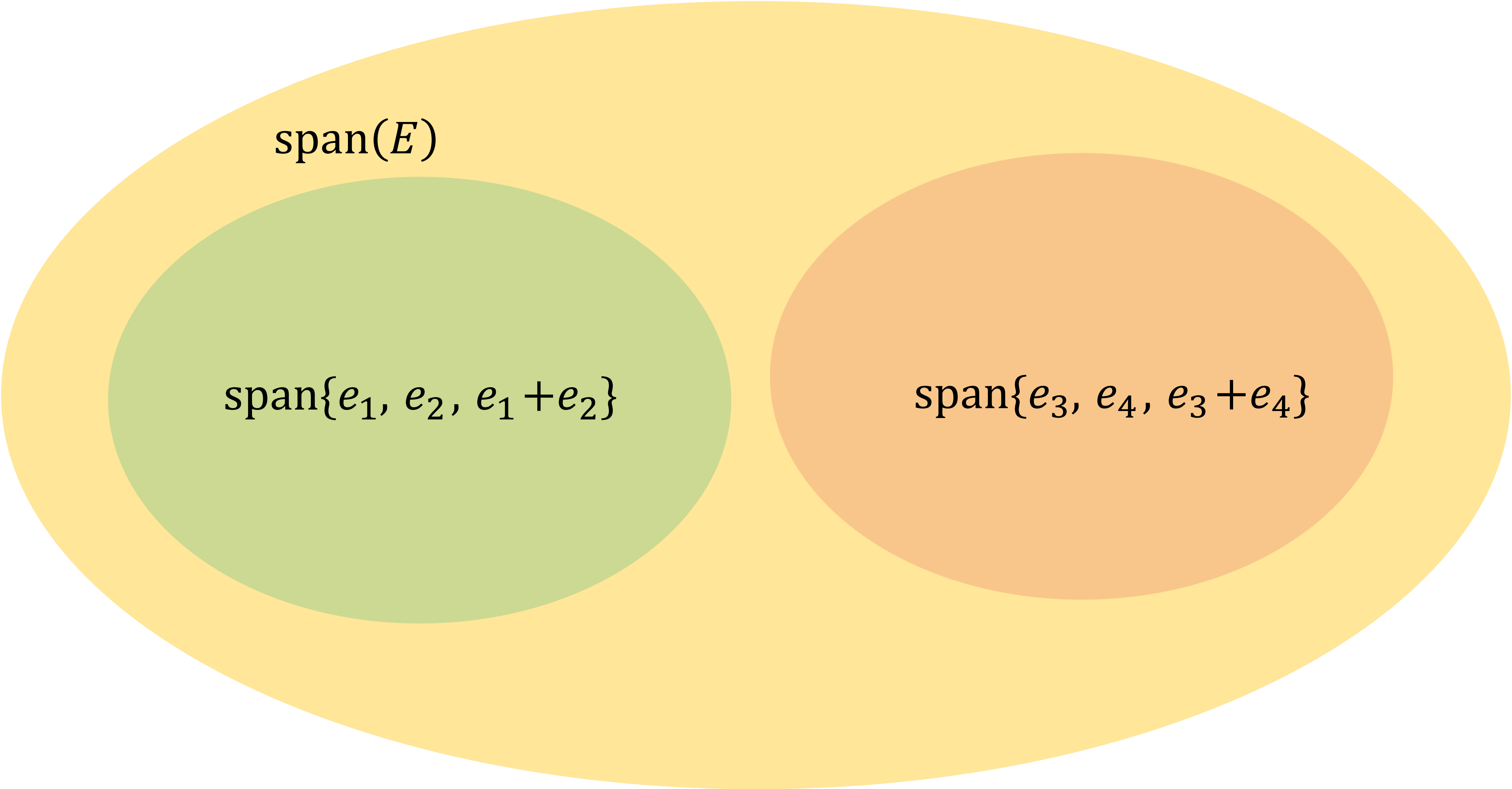
$$\mathrm{span}(S) \cap \mathrm{span}(E \setminus S) = \{0\} \tag{2}$$

E splits if there exists $S \subseteq E$ such that $\mathrm{span}(S) \cap \mathrm{span}(E \setminus S) = \{0\}$



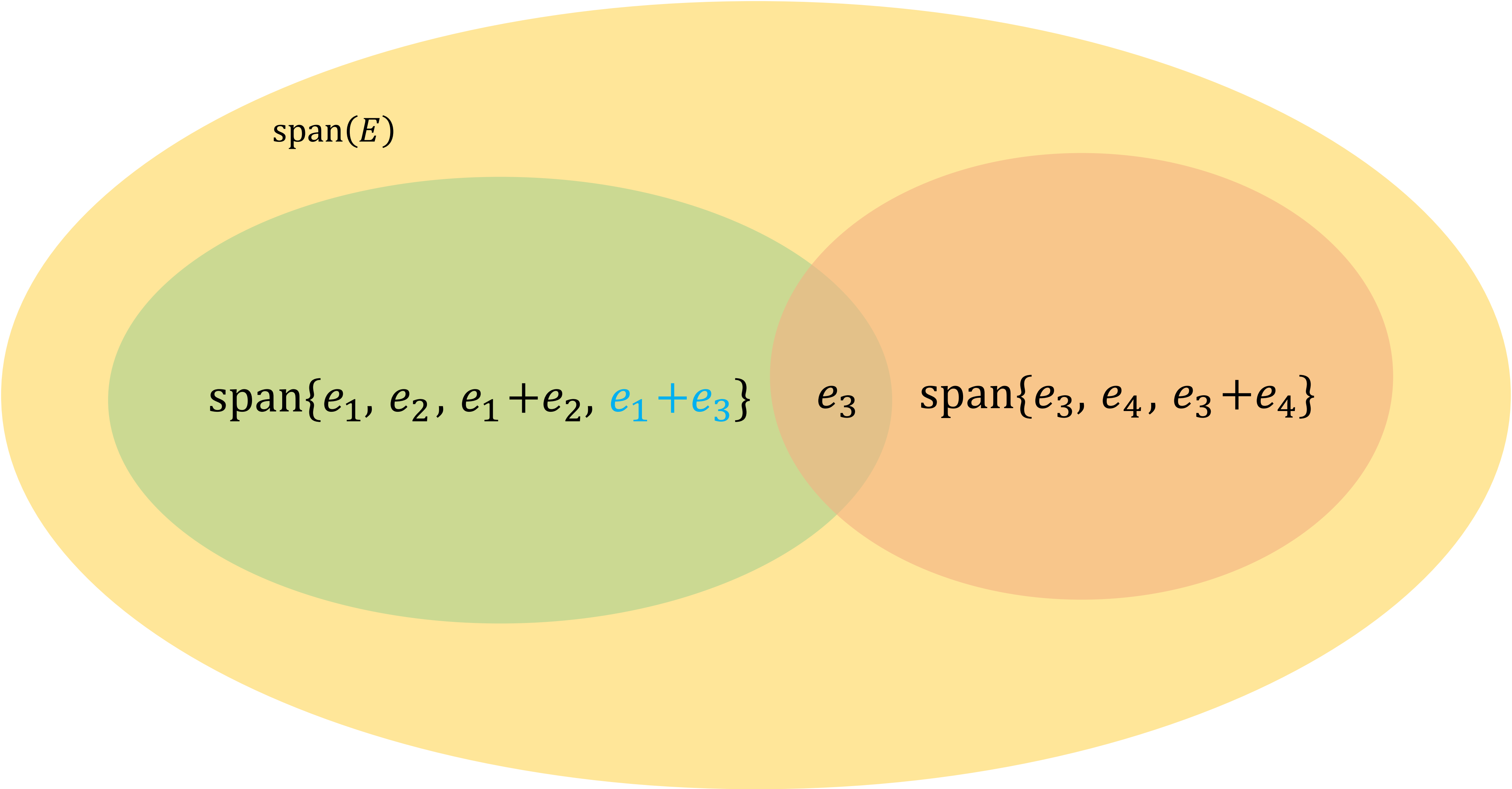
E splits if there exists $S \subseteq E$ such that $\mathrm{span}(S) \cap \mathrm{span}(E \setminus S) = \{0\}$

$E = \{e_1, e_2, e_1 + e_2, e_3, e_4, e_3 + e_4\}$



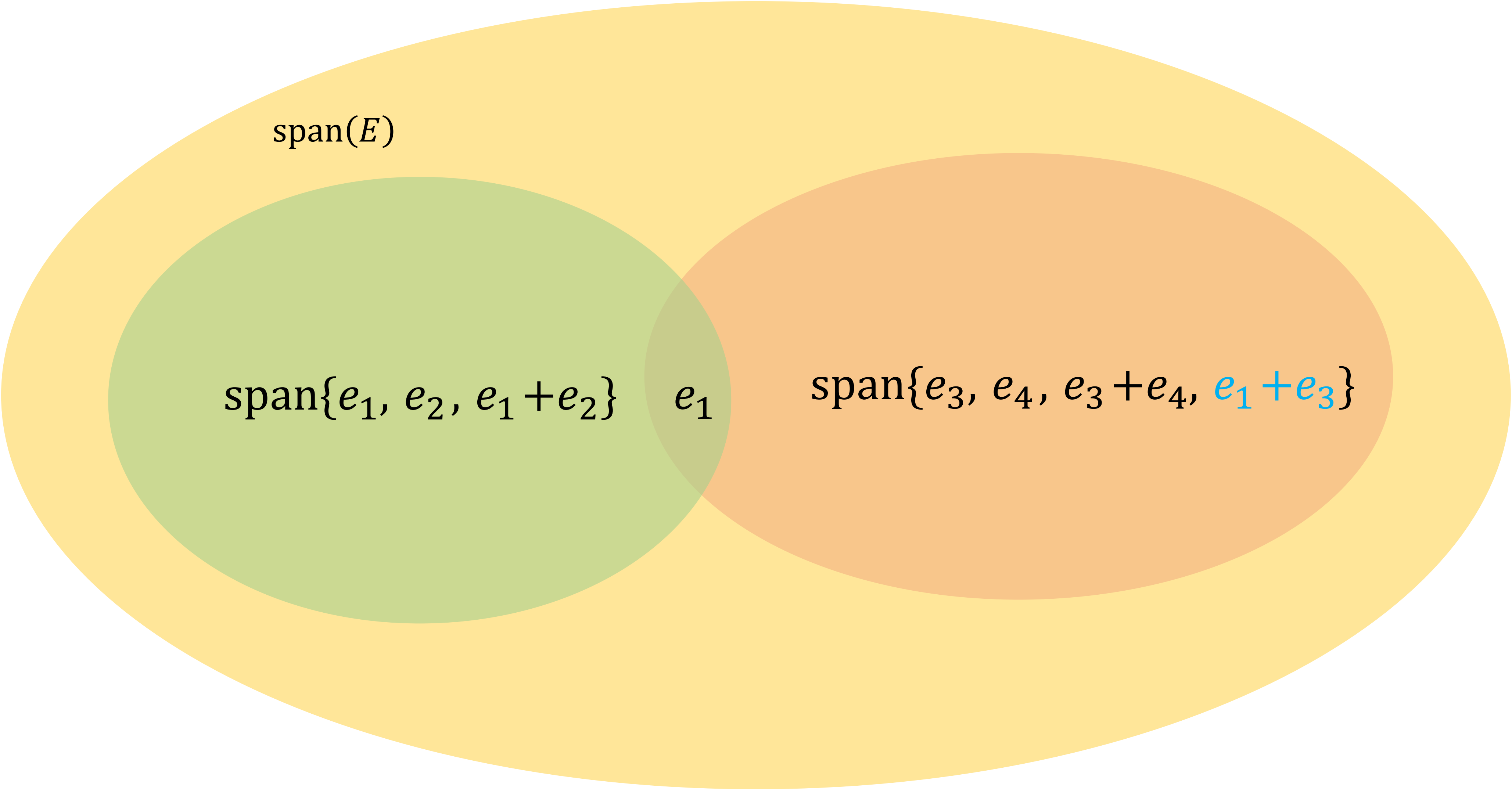
E splits if there exists $S \subseteq E$ such that $\mathrm{span}(S) \cap \mathrm{span}(E \setminus S) = \{0\}$

$E = \{e_1, e_2, e_1+e_2, e_3, e_4, e_3+e_4\} \cup \{e_1+e_3\}$



E splits if there exists $S \subseteq E$ such that $\text{span}(S) \cap \text{span}(E \setminus S) = \{0\}$

$$E = \{e_1, e_2, e_1+e_2, e_3, e_4, e_3+e_4\} \cup \{e_1+e_3\}$$



Splitting theorem

E splits if there exists $S \subseteq E$ such that $\mathrm{span}(S) \cap \mathrm{span}(E \setminus S) = \{0\}$

Splitting theorem [L-Petrov 2021]: Let $E = \{x_a \otimes y_a : a \in [n]\}$.

If

$$\dim\mathrm{span}(E) \leq d_x^{[n]} + d_y^{[n]} - 2$$

$d_x^{[n]} = \dim\mathrm{span}\{x_1, \ldots, x_n\}$

then E splits.

Cornell University

arXiv > math > arXiv:2103.15633v2

Mathematics > Combinatorics

[Submitted on 29 Mar 2021 (v1), last revised 15 Sep 2021 (this version, v2)]

A generalization of Kruskal's theorem on tensor decomposition

Benjamin Lovitz, Fedor Petrov

Splitting theorem

E splits if there exists $S \subseteq E$ such that $\mathrm{span}(S) \cap \mathrm{span}(E \setminus S) = \{0\}$

Splitting theorem [L-Petrov 2021]: Let $E = \{x_a \otimes y_a : a \in [n]\}$.

If

$$\mathrm{dimspan}(E) \leq d_x^{[n]} + d_y^{[n]} - 2$$

then E splits.

$d_x^{[n]} = \mathrm{dimspan}\{x_1, \dots, x_n\}$

Corollary: A uniqueness result that is stronger than Jennrich's (and Kruskal's!)

More matroid theory for product tensors?

Splitting theorem

E splits if there exists $S \subseteq E$ such that $\mathrm{span}(S) \cap \mathrm{span}(E \setminus S) = \{0\}$

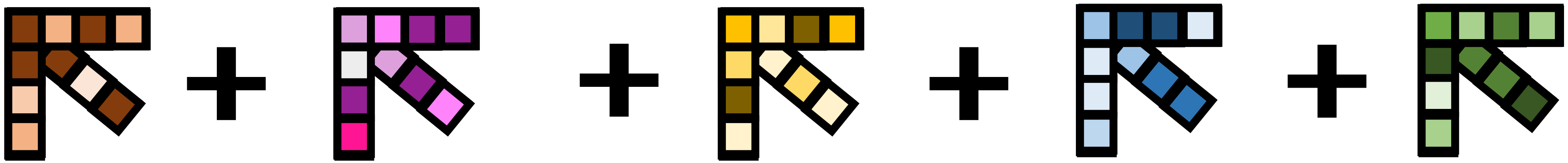
Splitting theorem [L-Petrov 2021]: Let $E = \{x_a \otimes y_a \otimes z_a : a \in [n]\}$.

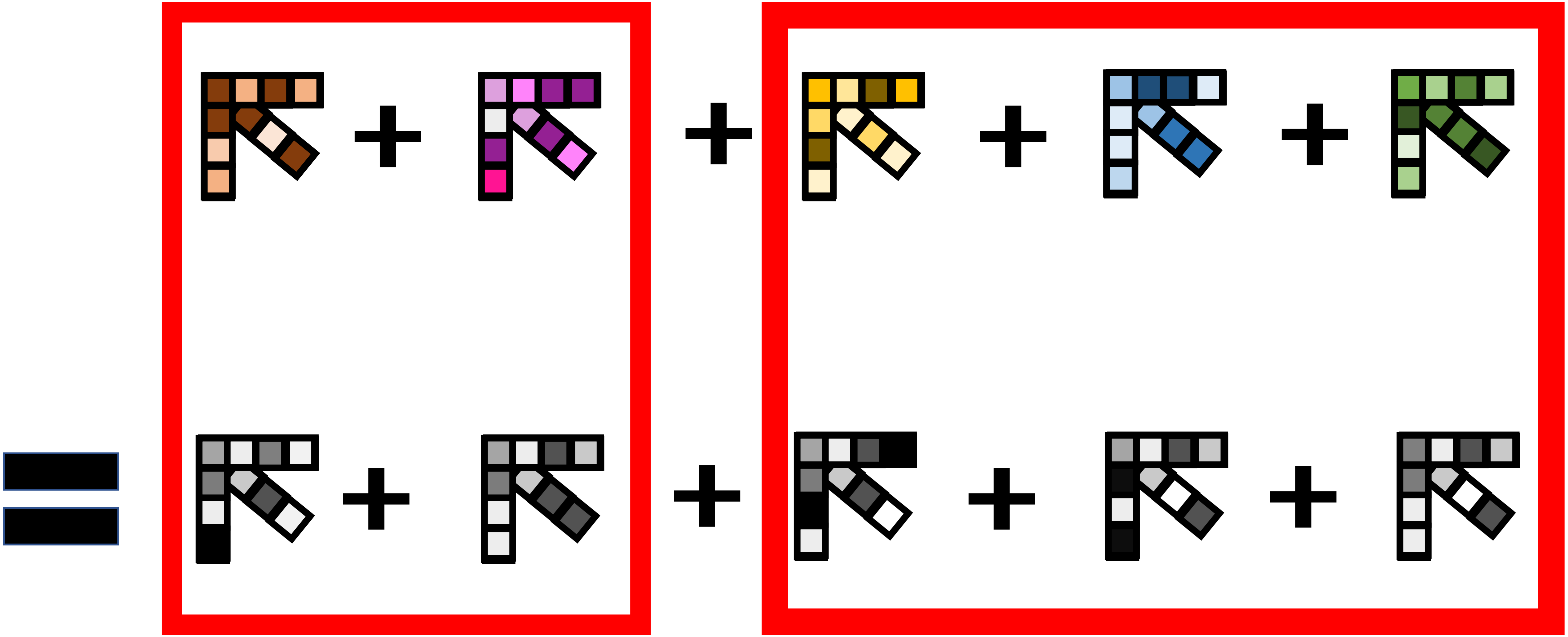
If

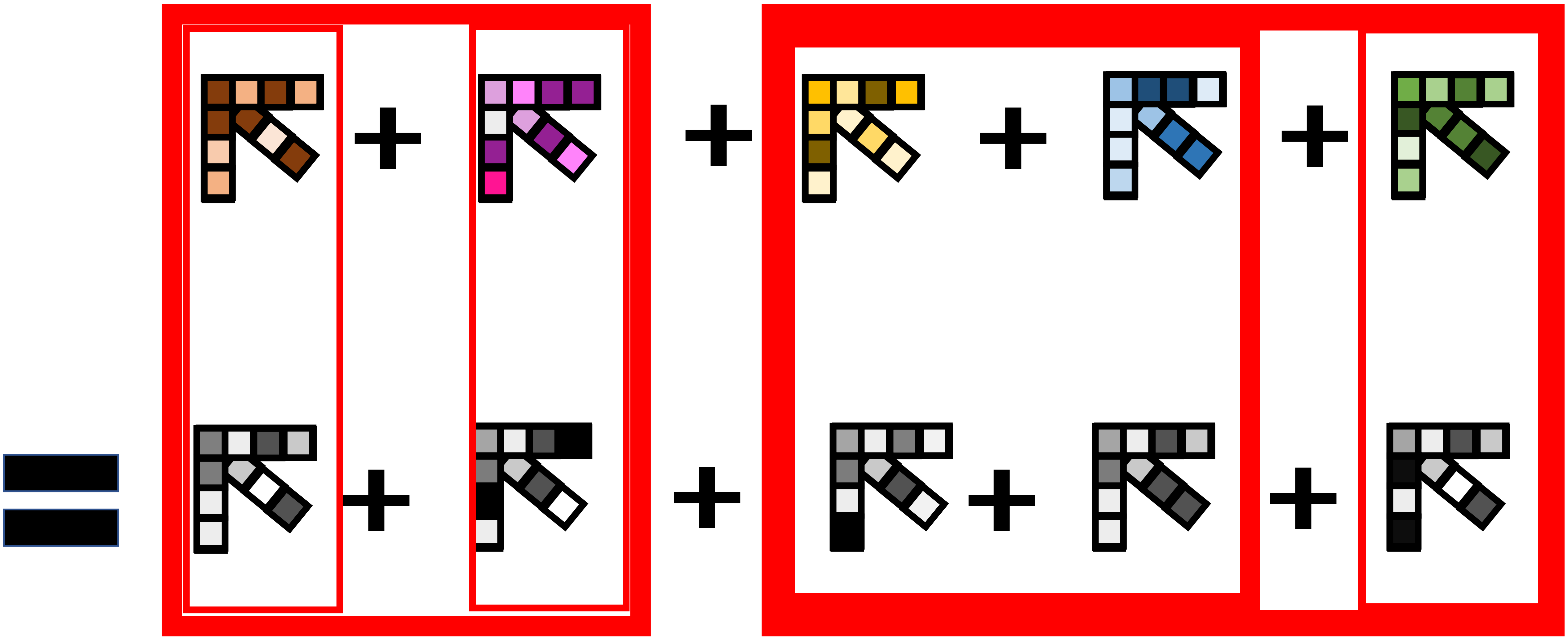
$$\mathrm{dimspan}(E) \leq d_x^{[n]} + d_y^{[n]} + d_z^{[n]} - 3$$

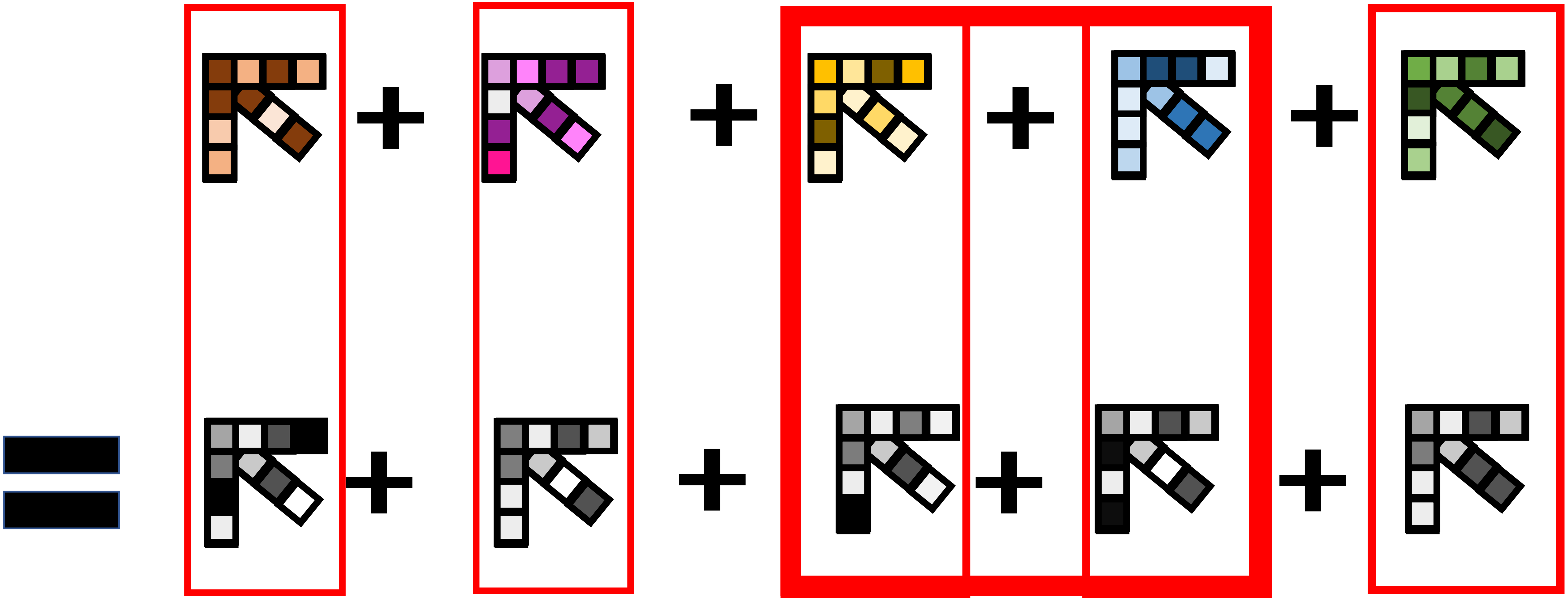
then E splits.

$d_x^{[n]} = \mathrm{dimspan}\{x_1, \dots, x_n\}$









Our generalization

$$T = \sum_{a \in [n]} x_a \otimes y_a \otimes z_a \in \mathbb{C}^{d_x} \otimes \mathbb{C}^{d_y} \otimes \mathbb{C}^{d_z} \quad (1)$$

<u>Theorem [L-Petrov]:</u> If for every subset $S \subseteq [n]$ of size $|S| \geq 2$, it holds that

$$2|S| \leq d_x^S + d_y^S + d_z^S - 2,$$

$d_x^S = \mathrm{dimspan}\{x_a : a \in S\}$

then (1) is the unique rank decomposition of T.

Outline

$$T = \sum_{a \in [n]} x_a \otimes y_a \otimes z_a \in \mathbb{C}^{d_x} \otimes \mathbb{C}^{d_y} \otimes \mathbb{C}^{d_z} \quad (1)$$

1. <u>Algorithms</u>

Image-based algorithm [Johnston-L-Vijayaraghavan 2022+] based on [De Lathauwer et al 2007]

Jennrich's algorithm [Harshmann 1970]

2. <u>Uniqueness</u>

Kruskal's theorem [Kruskal 1977]

Splitting theorem [L-Petrov 2021]

Example

$$T = \sum_{a \in [n]} x_a \otimes y_a \otimes z_a \in X \otimes Y \otimes Z \qquad (1)$$

$$T = e_1^{\otimes 3} + e_2^{\otimes 3} + e_3^{\otimes 3} + e_4^{\otimes 3} + (e_1 + e_2) \otimes (e_2 + e_3) \otimes (e_1 + e_4)$$

$x_1 \otimes y_1 \otimes z_1$ … x_5 $\otimes$ y_5 $\otimes$ z_5

Example

$$T = \sum_{a \in [n]} x_a \otimes y_a \otimes z_a \in X \otimes Y \otimes Z \qquad (1)$$

$$T = e_1^{\otimes 3} + e_2^{\otimes 3} + e_3^{\otimes 3} + e_4^{\otimes 3} + (e_1 + e_2) \otimes (e_2 + e_3) \otimes (e_1 + e_4)$$

$x_1 \otimes y_1 \otimes z_1$ … x_5 $\otimes$ y_5 $\otimes$ z_5

Kruskal's theorem does not certify uniqueness

$$10 = 2n \nleq k_x + k_y + k_z - 2 = 2 + 2 + 2 - 2 = 4$$

Example

$$T = \sum_{a \in [n]} x_a \otimes y_a \otimes z_a \in X \otimes Y \otimes Z \qquad (1)$$

$$T = e_1^{\otimes 3} + e_2^{\otimes 3} + e_3^{\otimes 3} + e_4^{\otimes 3} + (e_1 + e_2) \otimes (e_2 + e_3) \otimes (e_1 + e_4)$$

$x_1 \otimes y_1 \otimes z_1$ … x_5 $\otimes$ y_5 $\otimes$ z_5

<u>Theorem [L-Petrov]:</u> If for every subset $S \subseteq [n]$ of size $|S| \geq 2$, it holds that

$2|S| \leq d_x^S + d_y^S + d_z^S - 2$, then (1) is the unique decomposition of T.

Example

$$T = \sum_{a \in [n]} x_a \otimes y_a \otimes z_a \in X \otimes Y \otimes Z \qquad (1)$$

$$T = e_1^{\otimes 3} + e_2^{\otimes 3} + e_3^{\otimes 3} + e_4^{\otimes 3} + (e_1 + e_2) \otimes (e_2 + e_3) \otimes (e_1 + e_4)$$

$x_1 \otimes y_1 \otimes z_1$ ... x_5 $\otimes$ y_5 $\otimes$ z_5

Theorem [L-Petrov]: If for every subset $S \subseteq [n]$ of size $|S| \geq 2$, it holds that

$2|S| \leq d_x^S + d_y^S + d_z^S - 2$, then (1) is the unique decomposition of T.

For $S = \{1,2\}$, $4 = 2|S| \leq d_x^S + d_y^S + d_z^S - 2 = 2 + 2 + 2 - 2 = 4$ ✓

For $S = \{1,2,5\}$, $6 = 2|S| \leq d_x^S + d_y^S + d_z^S - 2 = 2 + 3 + 3 - 2 = 6$ ✓

For $S = \{1,2,3,5\}$, $8 = 2|S| \leq d_x^S + d_y^S + d_z^S - 2 = 3 + 3 + 4 - 2 = 8$ ✓

For $S = [5]$, $10 = 2|[n]| \leq d_x^{[n]} + d_y^{[n]} + d_z^{[n]} - 2 = 4 + 4 + 4 - 2 = 10$ ✓