Tutorial: Tensor Approximation in Scientific Visualization

Tensor Decomposition Models

Susanne K. Suter, Rafael Ballester-Ripoll, Renato Pajarola

University of Zurich[UZH]

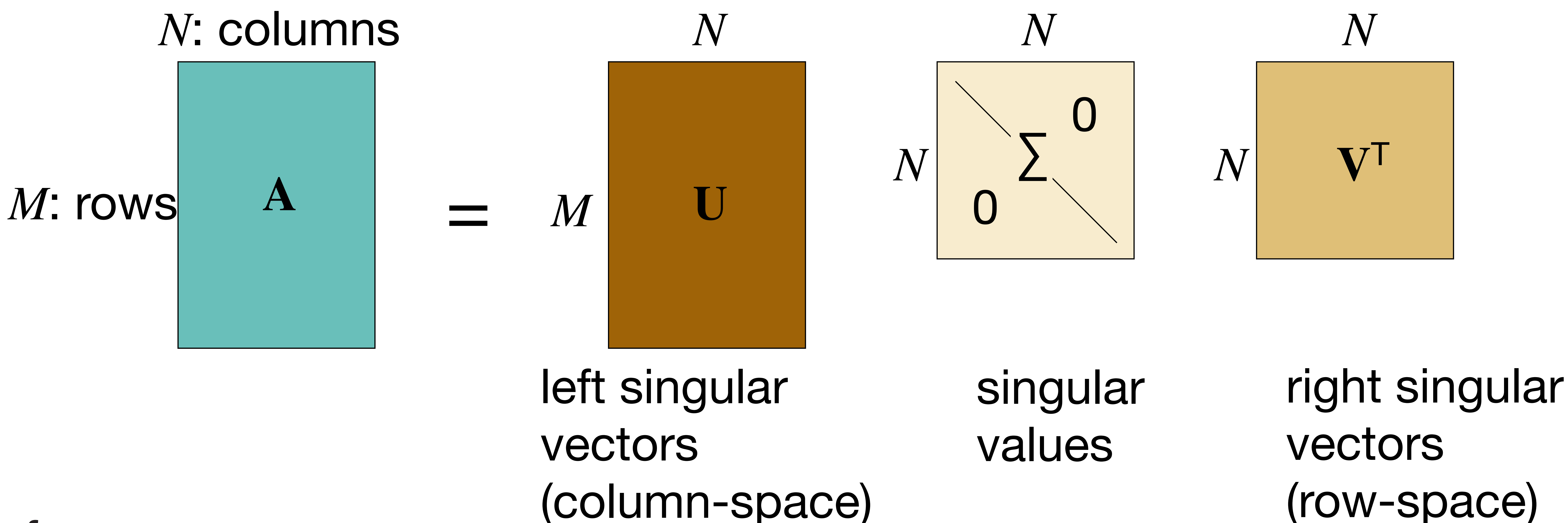
VISUALIZATIONAND MULTIMEDIALAB

Data Reduction and Approximation

- A fundamental concept of data reduction is to remove redundant and irrelevant information while preserving the relevant features
  - e.g. through frequency analysis by projection onto pre-defined bases, or extraction of data intrinsic principal components
    - identify spatio-temporal and frequency redundancies
  - maintain strongest and most significant signal components
- Data reduction linked to concepts and techniques of data compression, noise reduction as well as feature extraction and recognition/extraction

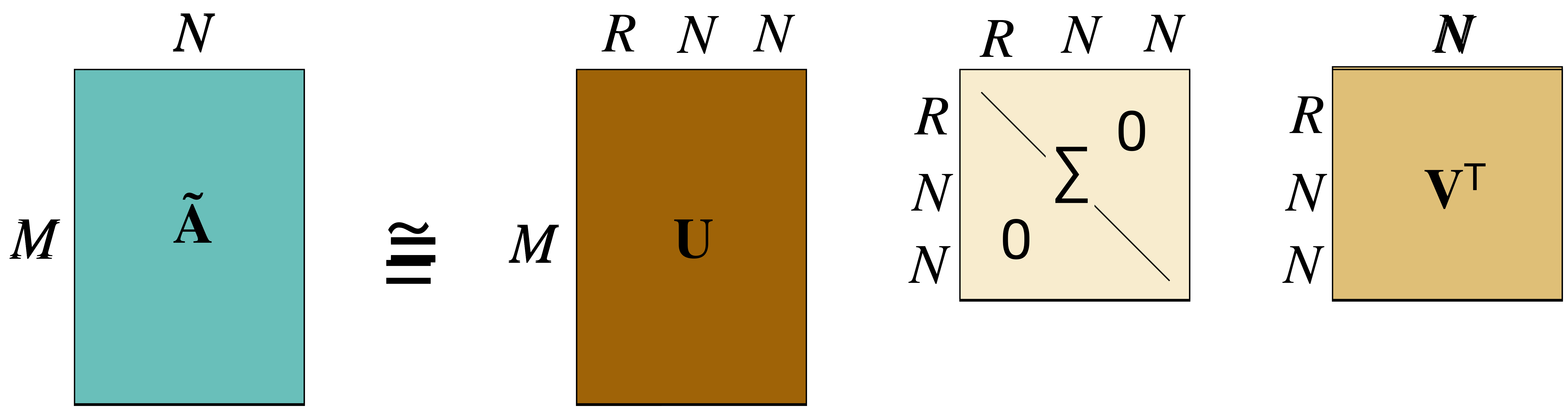
Data Approximation using SVD

- Singular Value Decomposition (SVD) standard tool for matrices, i.e., 2D input datasets
  - see also principal component analysis (PCA)



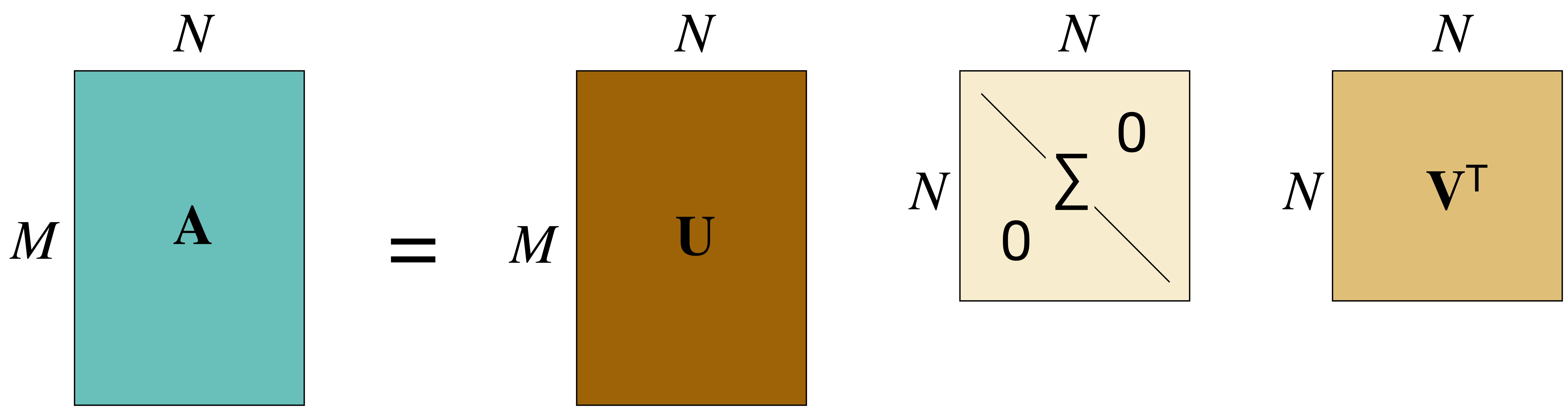
Low-rank Approximation

- Exploit ordered singular values: $s_1 \geq s_2 \geq \ldots \geq s_N$
- Select first r singular values (rank reduction)
    - use only bases (singular vectors) of corresponding subspace

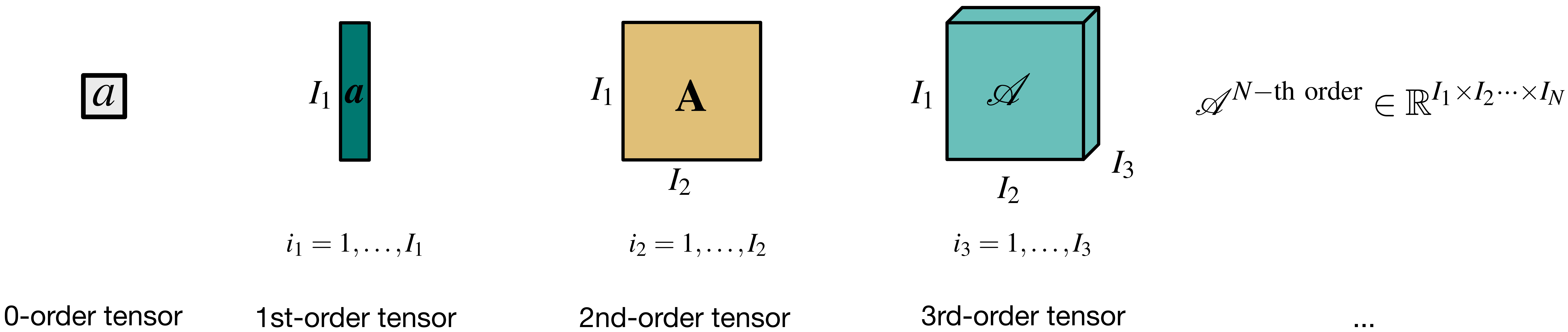


Matrix SVD Properties

- Matrix SVD
  - rank reducibility
  - orthonormal row/column matrices



What is a Tensor?

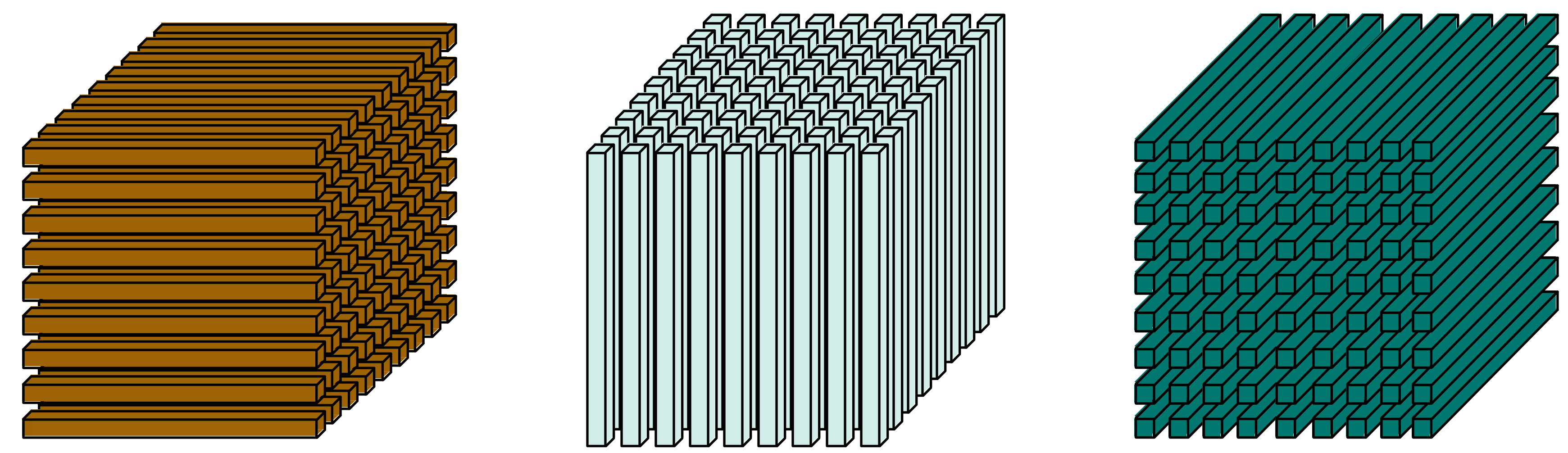


- Data sets are often multidimensional arrays (tensors)
  - images, image collections, video, volume data etc.

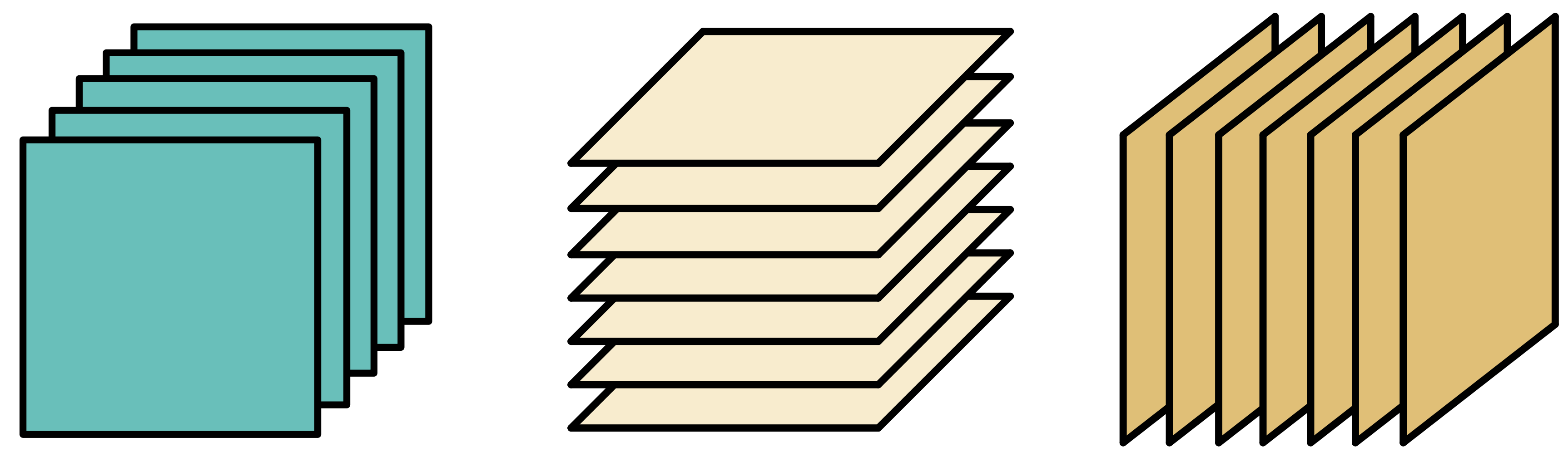
Fibers and Slices

- Individual elements of a vector $\boldsymbol{a}$ are given by a_{i1}, from a matrix $\mathbf{A}$ by $a_{i1,i2}$ and from a tensor $\mathscr{A}$ by $a_{i1,i2,i3}$

- The generalization of rows, columns (and tubes) is a *fiber* in a particular mode

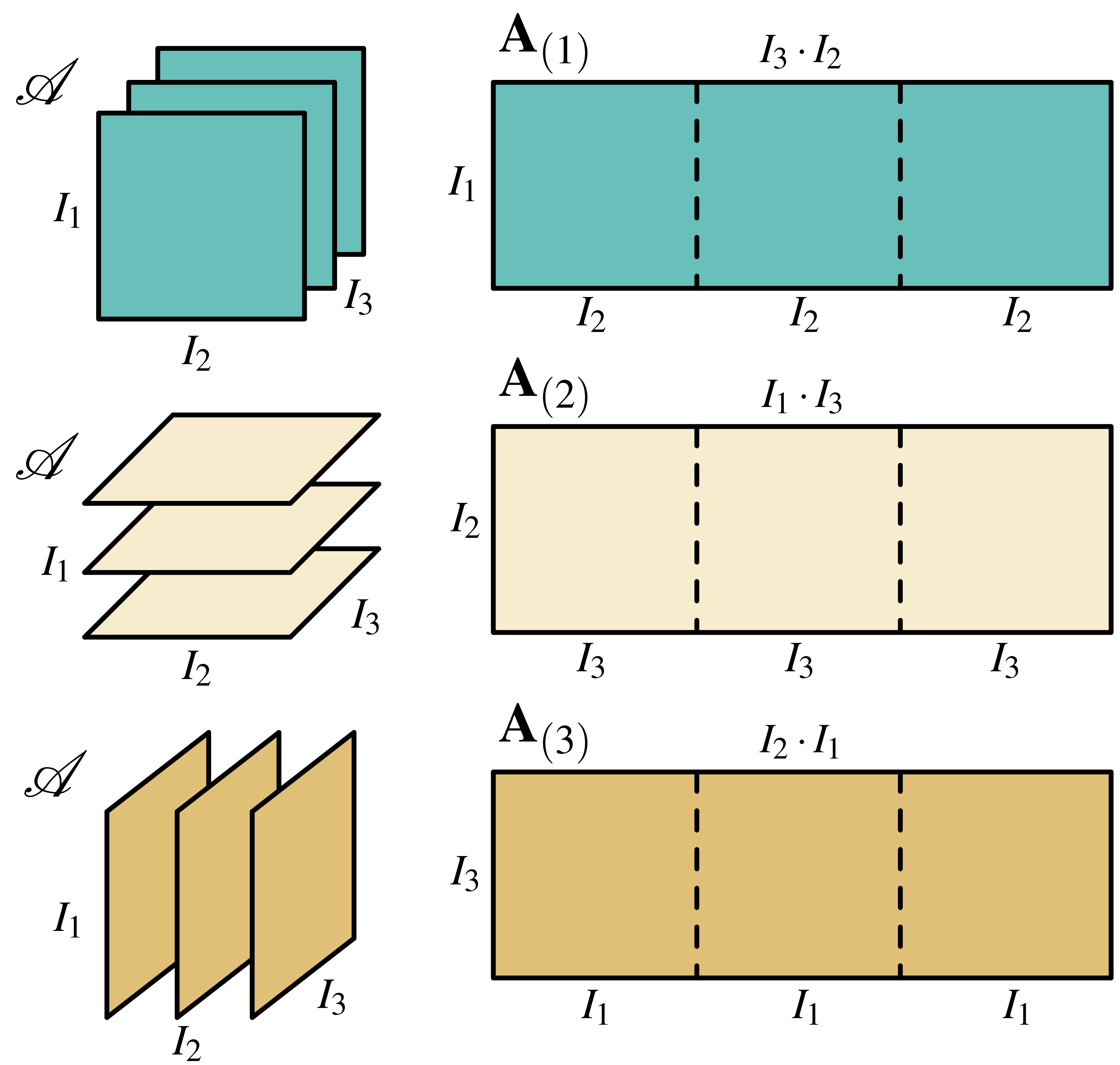


- Two dimensional sections of a tensor are called slices
  - frontal, horizontal and lateral for $\mathscr{A} \in \mathbb{R}^3$



Unfolding and Ranks

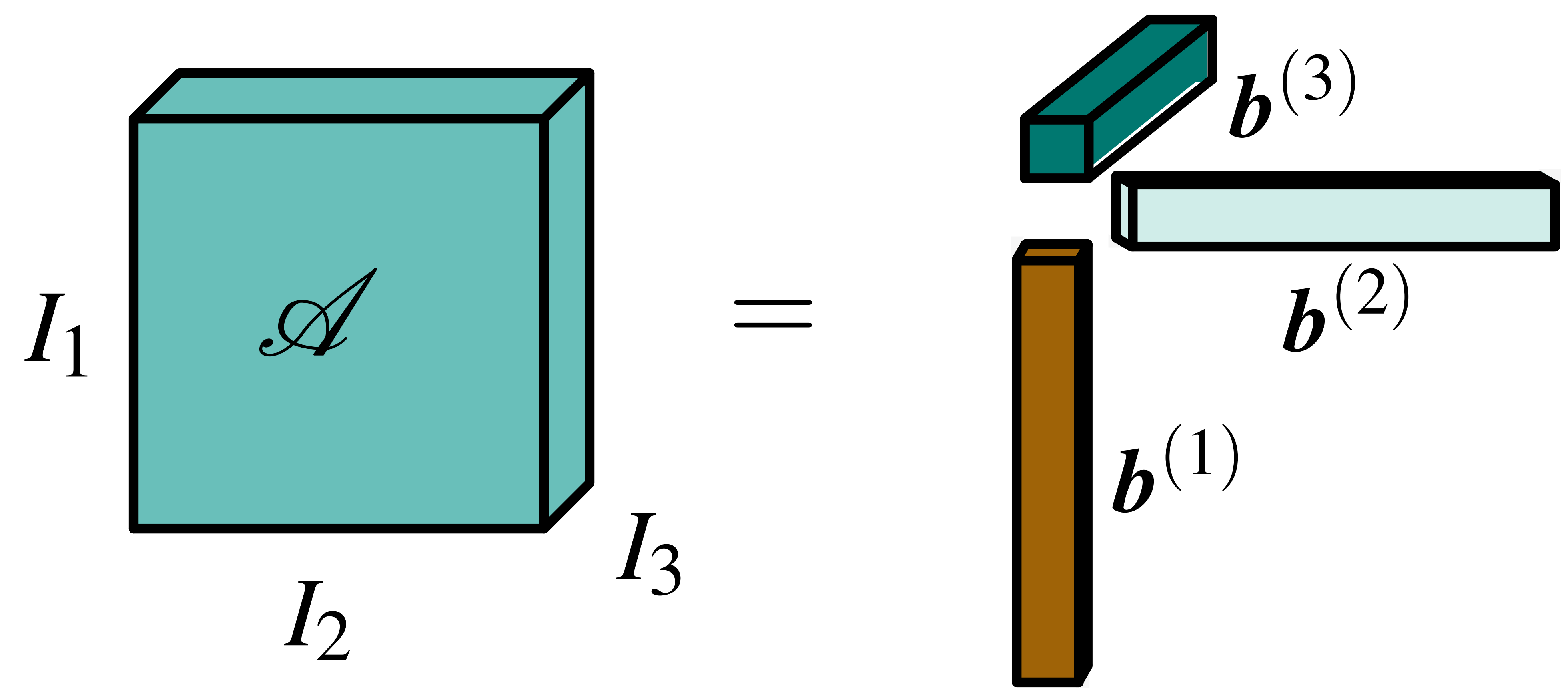
- Operations with tensors often performed as matrix operations using unfolded tensor representations
  - different tensor unfolding strategies possible
- Forward cyclic unfolding $\mathbf{A}_{(n)}$ of a 3rd order tensor $\mathscr{A}$ (or 3D volume)
- The n-rank of a tensor is typically defined on an unfolding
  - n-rank $R_n = \mathrm{rank}_n(\mathscr{A}) = \mathrm{rank}(\mathbf{A}_{(n)})$
  - multilinear rank-$(R_1, R_2, \ldots, R_N)$ of $\mathscr{A}$



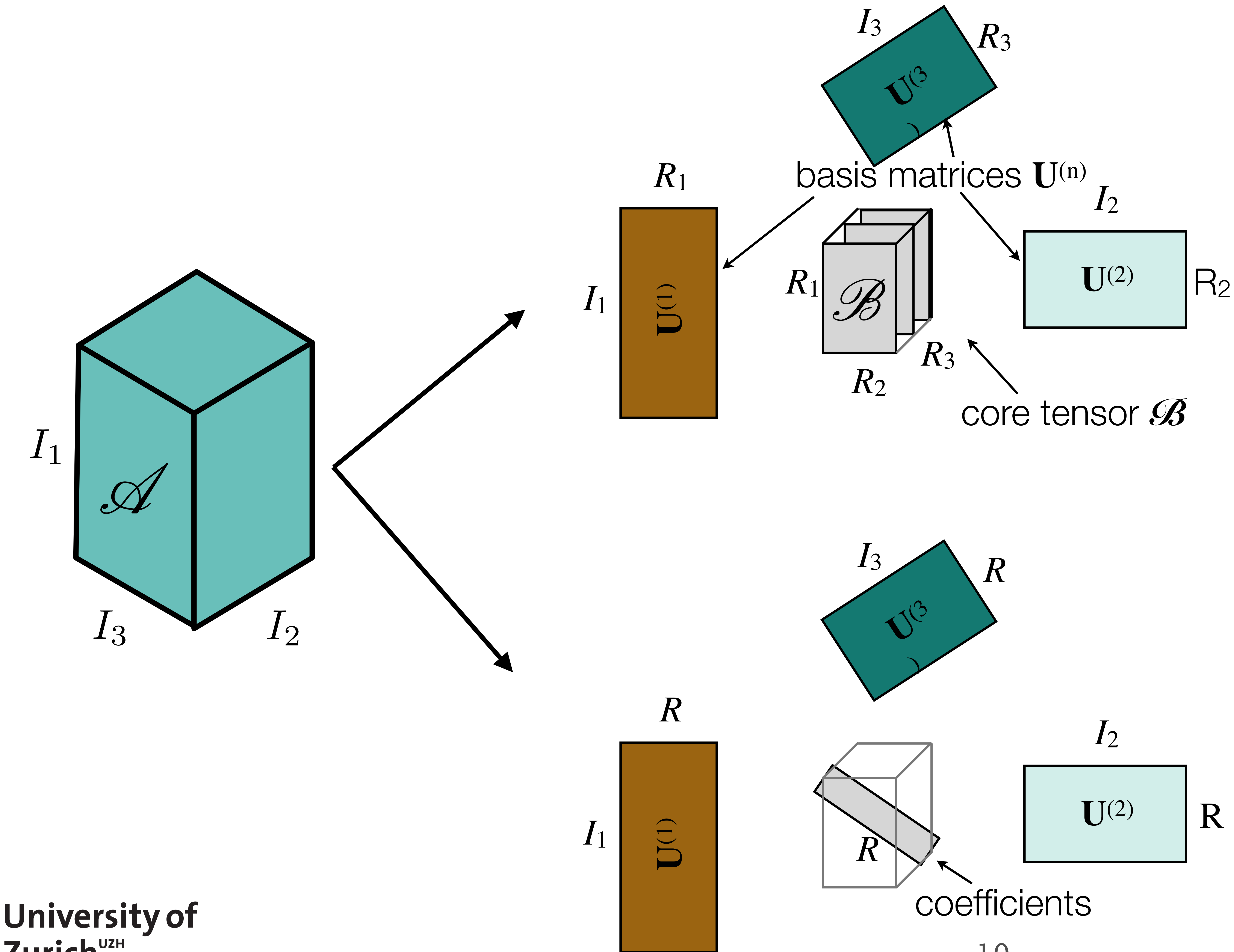
Rank-one Tensor

- N-mode tensor $\mathscr{A} \in \mathbb{R}^{I_1 \times \ldots \times I_N}$ that can be expressed as the outer product of N vectors
  - Kruskal tensor
- Useful to understand principles of rank-reduced tensor reconstruction
  - linear combination of rank-one tensors

$$\mathscr{A} = \boldsymbol{b}^{(1)} \circ \boldsymbol{b}^{(2)} \circ \cdots \circ \boldsymbol{b}^{(N)}$$



Tensor Decomposition Models



Tucker

- Three-mode factor analysis (**3MFA/Tucker3**)
  [Tucker, 1964+1966]
- Higher-order SVD (**HOSVD**)
  [De Lathauwer et al., 2000a]

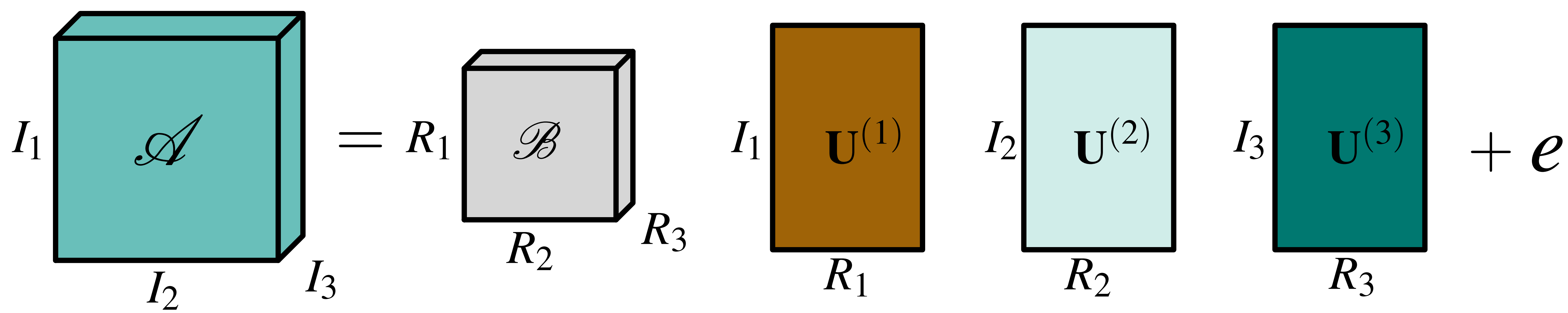
CP

- **PARAFAC** (parallel factors) [Harshman, 1970]
- **CANDECOMP** (CAND) (canonical decomposition)
  [Caroll & Chang, 1970]

Tucker Model

- Higher order tensor $\mathscr{A} \in \mathbb{R}^{I_1 \times \ldots \times I_N}$ represented as a product of a core tensor $\mathscr{B} \in \mathbb{R}^{R_1 \times \ldots \times R_N}$ and N factor matrices $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times R_n}$
  - using n-mode products $\times_n$

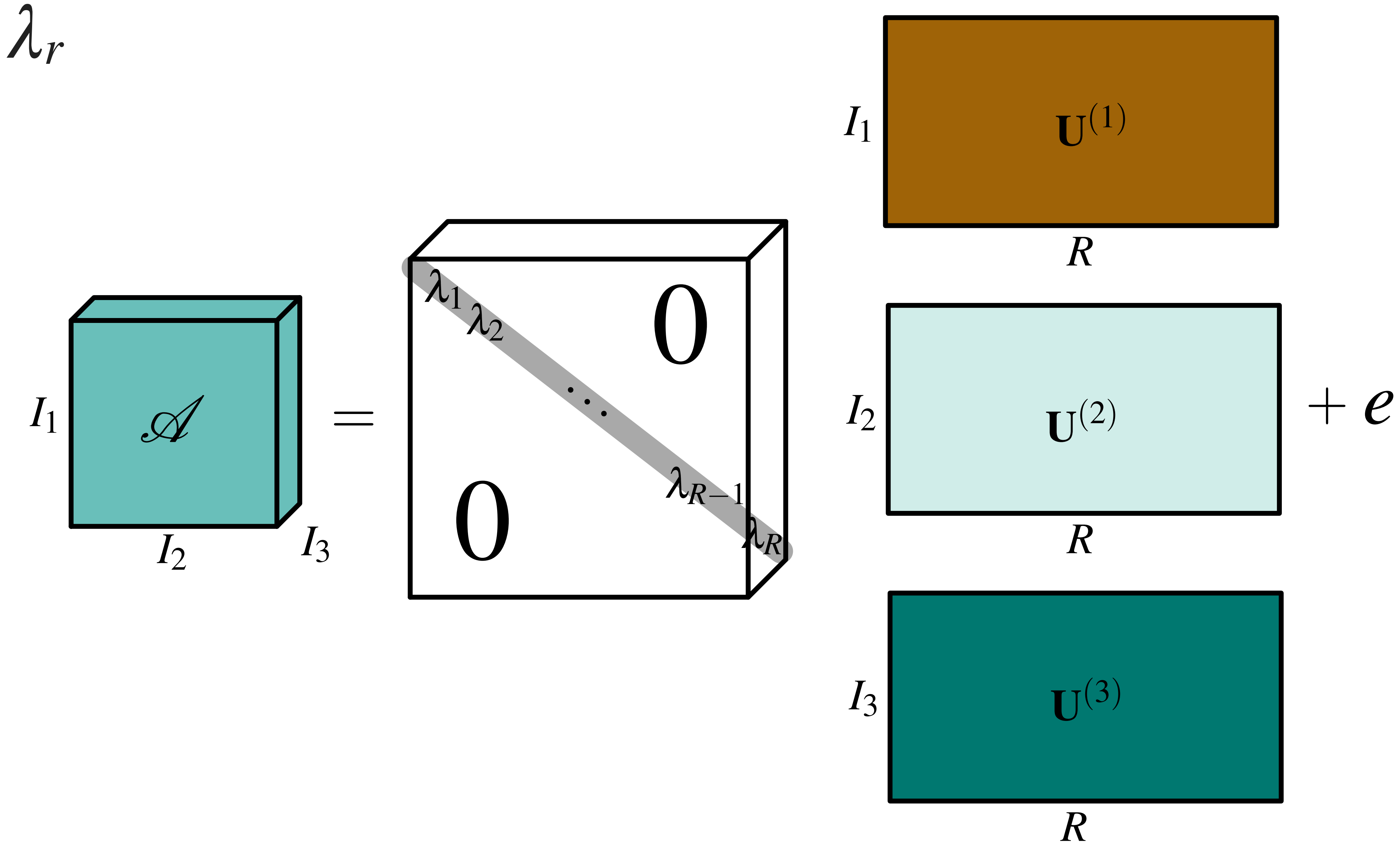
$$\mathscr{A} = \mathscr{B} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \cdots \times_N \mathbf{U}^{(N)} + \varepsilon$$



CANDECOMP-PARAFAC Model

- *Canonical decomposition* or *parallel factor analysis* model (CP)
- Higher order tensor $\mathscr{A}$ factorized into a sum of rank-one tensors
    - normalized column vectors $\boldsymbol{u}_r^{(n)}$ define factor matrices $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times R}$ and weighting factors λ_r

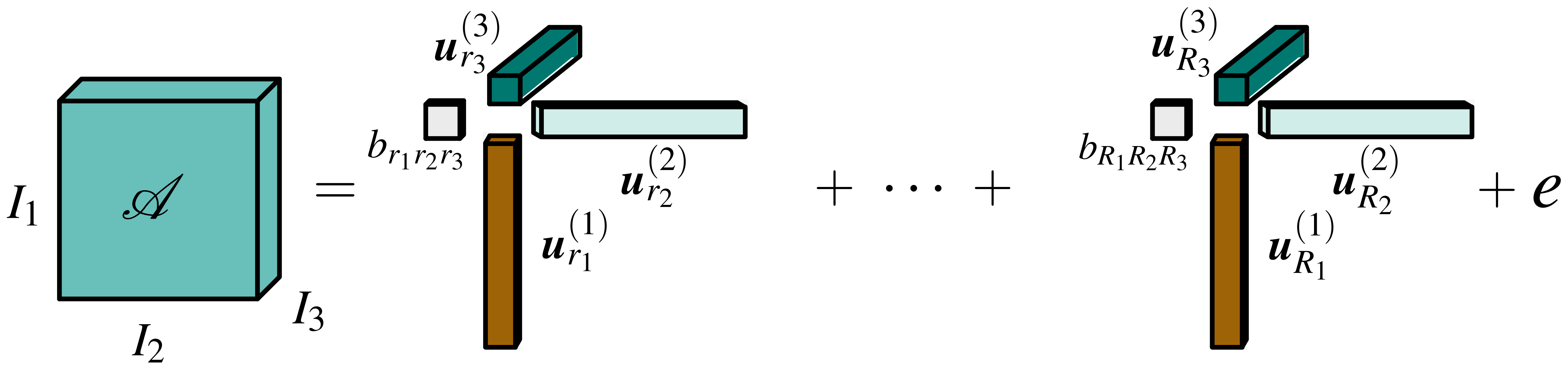
$$\mathscr{A} = \sum_{r=1}^{R} \lambda_r \cdot \boldsymbol{u}_r^{(1)} \circ \boldsymbol{u}_r^{(2)} \circ \ldots \boldsymbol{u}_r^{(N)} + \boldsymbol{\varepsilon}$$



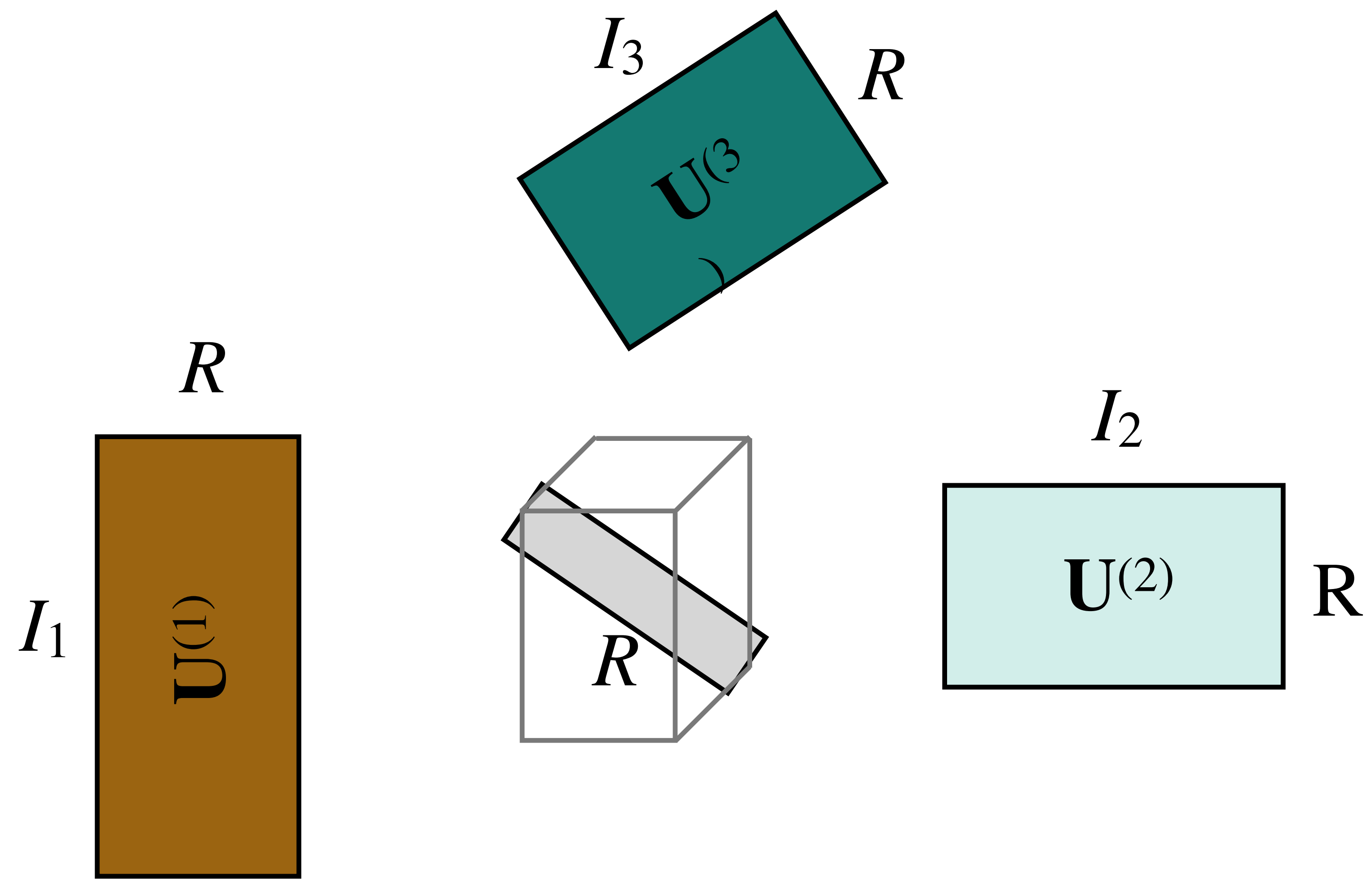
Linear Combination of Rank-one Tensors

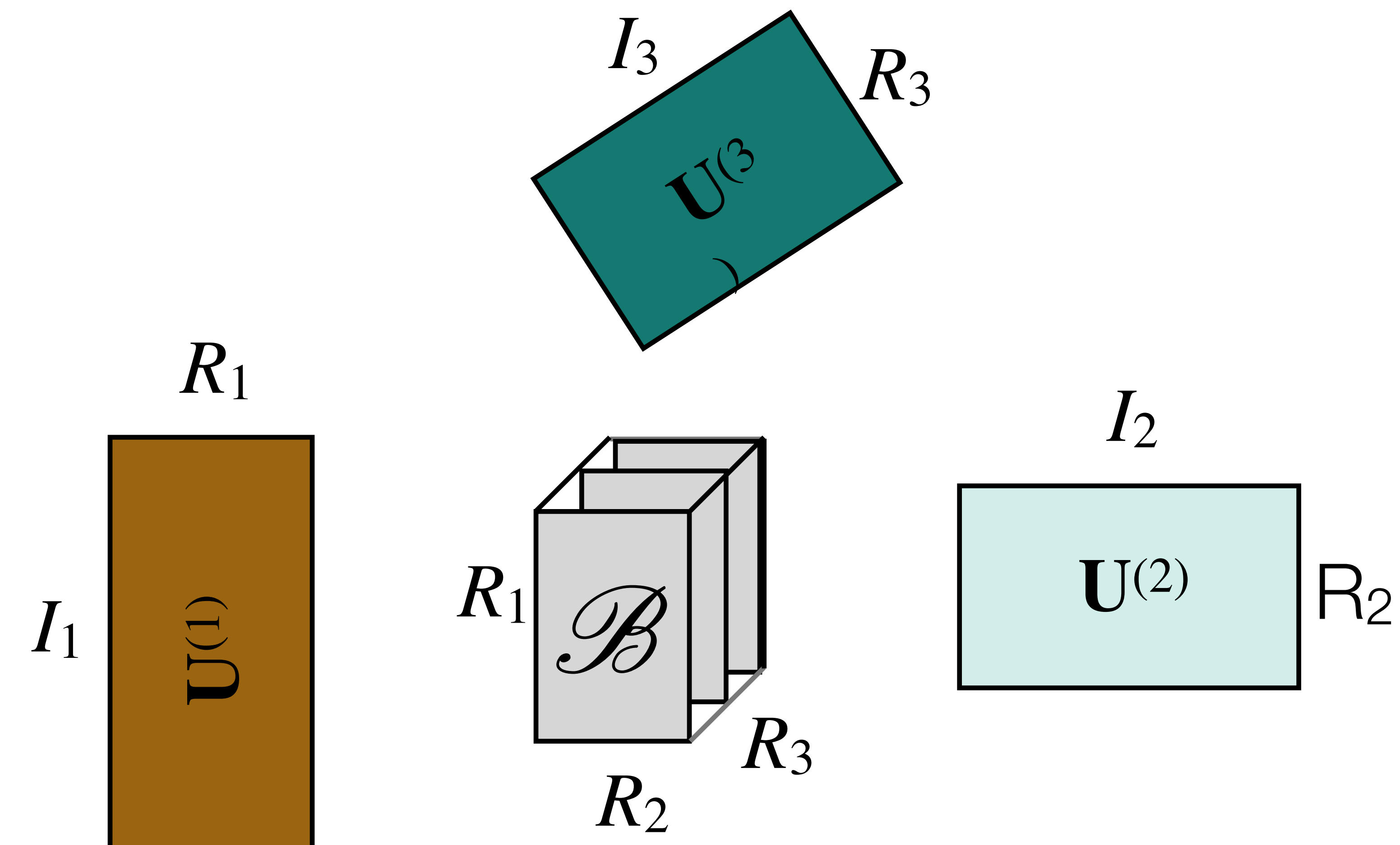
- The CP model is defined as a linear combination of rank-one tensors
- The Tucker model can be interpreted as linear combination of rank-one tensors

$$\mathscr{A} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \cdots \sum_{r_N=1}^{R_N} b_{r_1 r_2 \ldots r_N} \cdot \boldsymbol{u}_{r_1}^{(1)} \circ \boldsymbol{u}_{r_2}^{(2)} \circ \ldots \boldsymbol{u}_{r_N}^{(N)} + \varepsilon$$



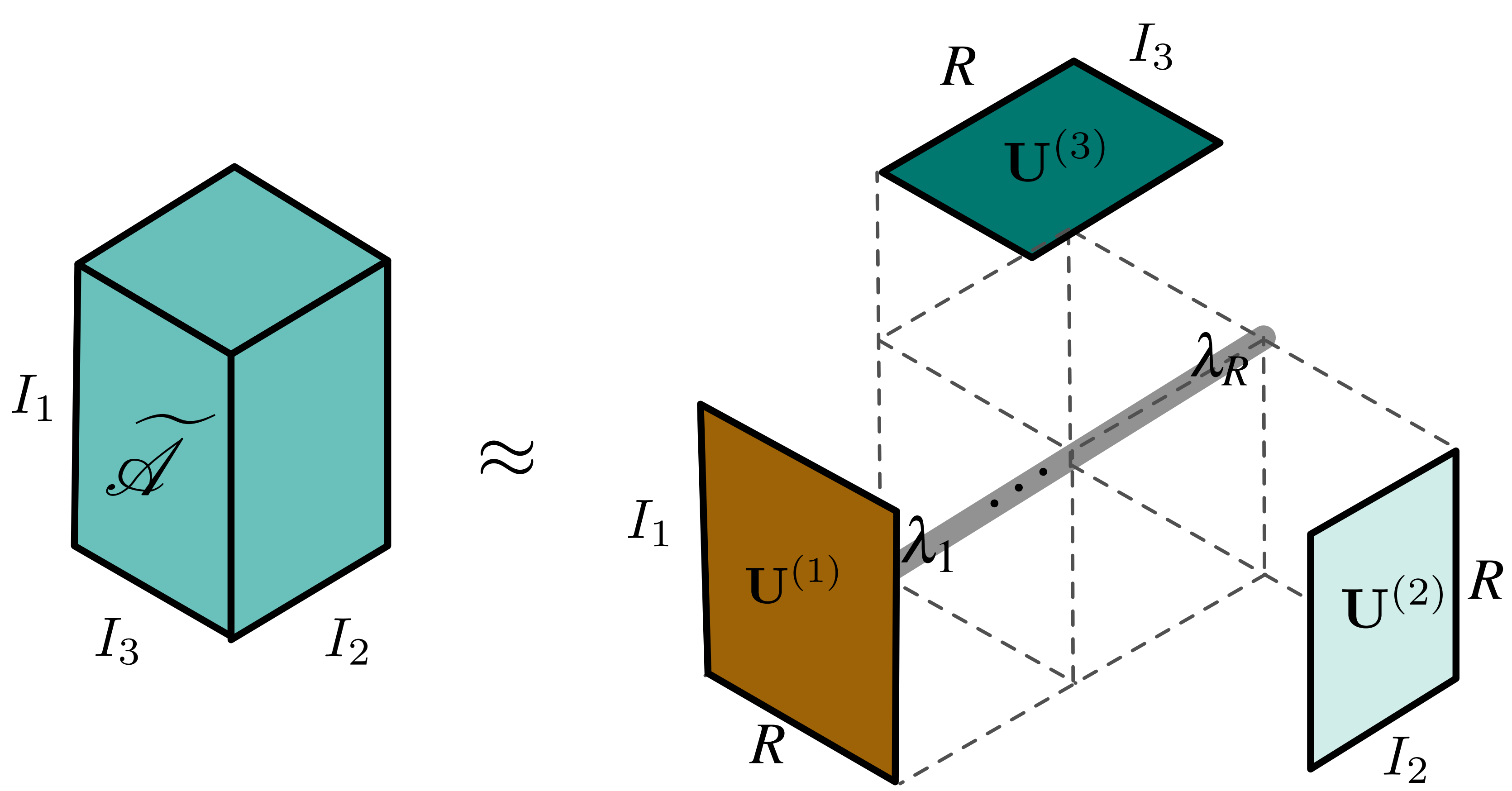
CP a Special Case of Tucker

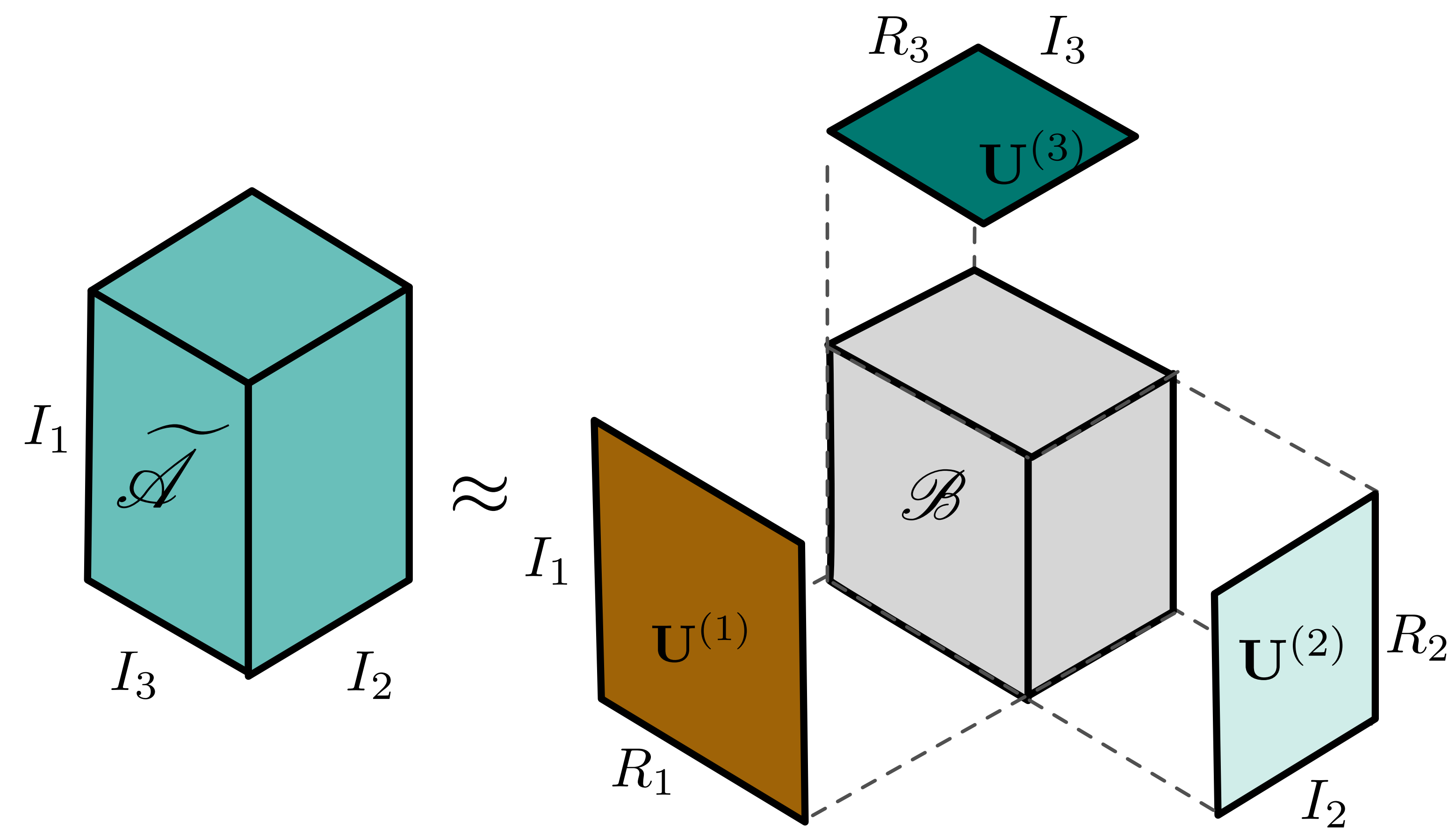




CP

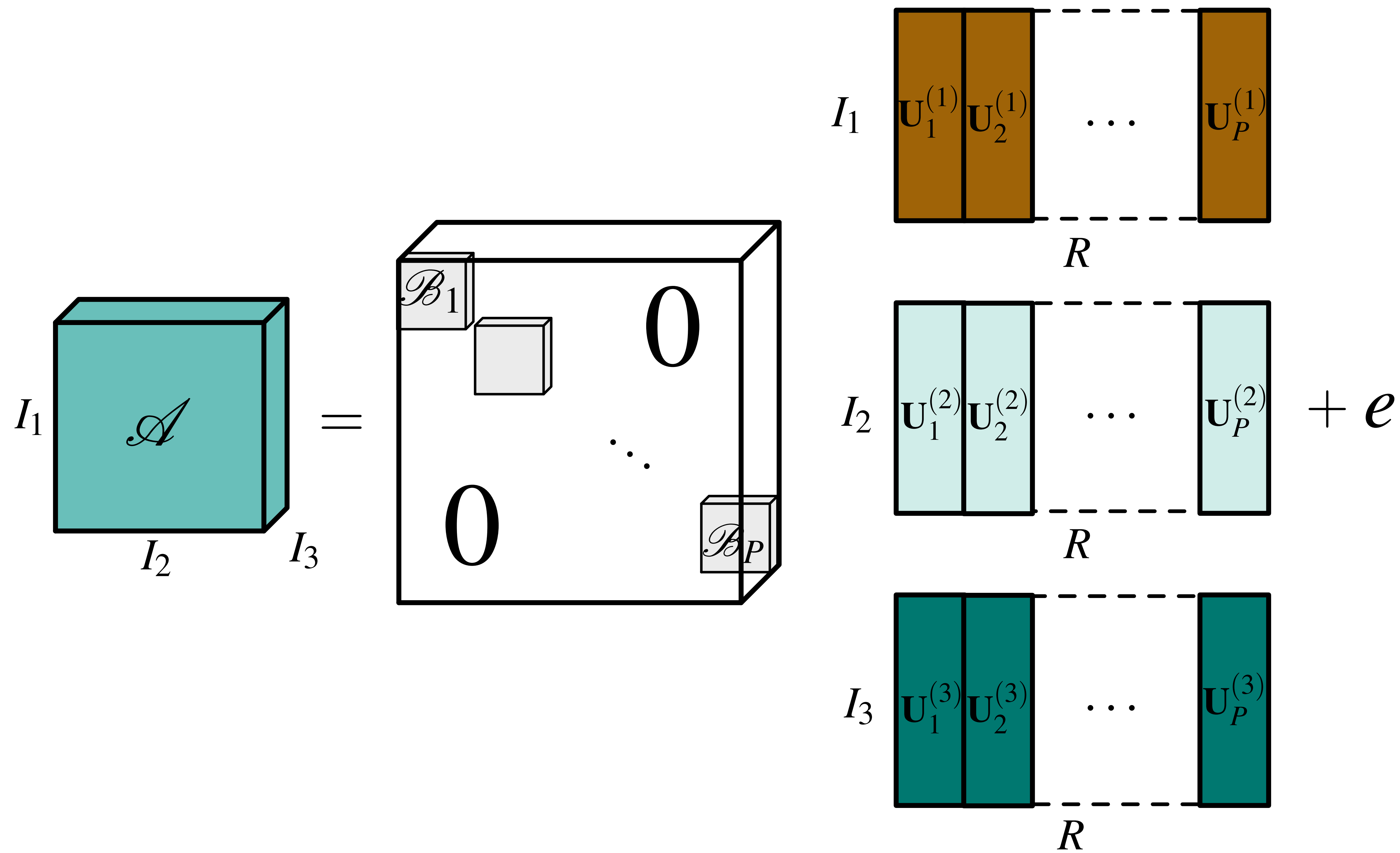
Tucker





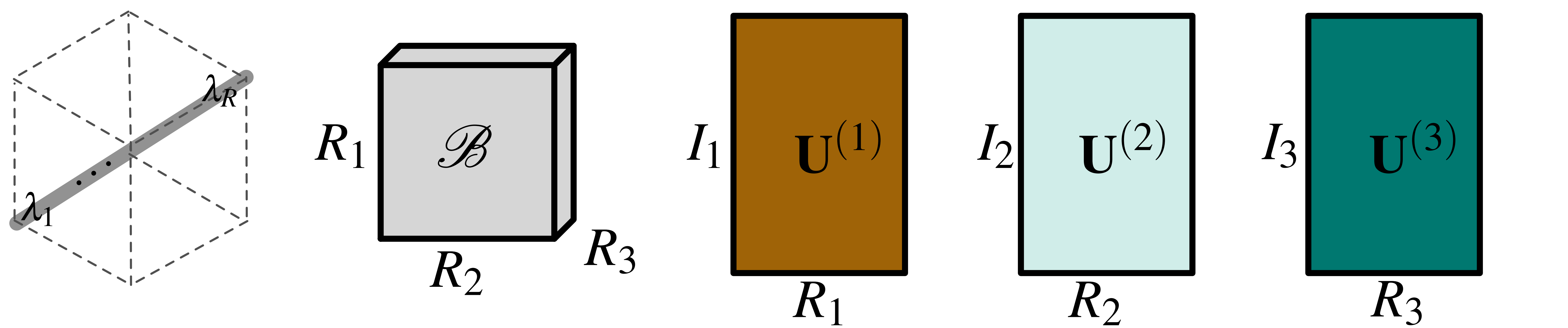
Generalizations

- Any special form of core and corresponding factor matrices
  - e.g. blocks along diagonal



Reduced Rank Approximation

- Full reconstruction using a Tucker or CP model may require excessively many coefficients and wide factor matrices
  - large rank values R (CP), or $R_1, R_2 \ldots R_N$ (Tucker)
- Quality of approximation increases with the rank, and number of column vectors of the factor matrices
  - best possible fit of these bases matrices discussed later

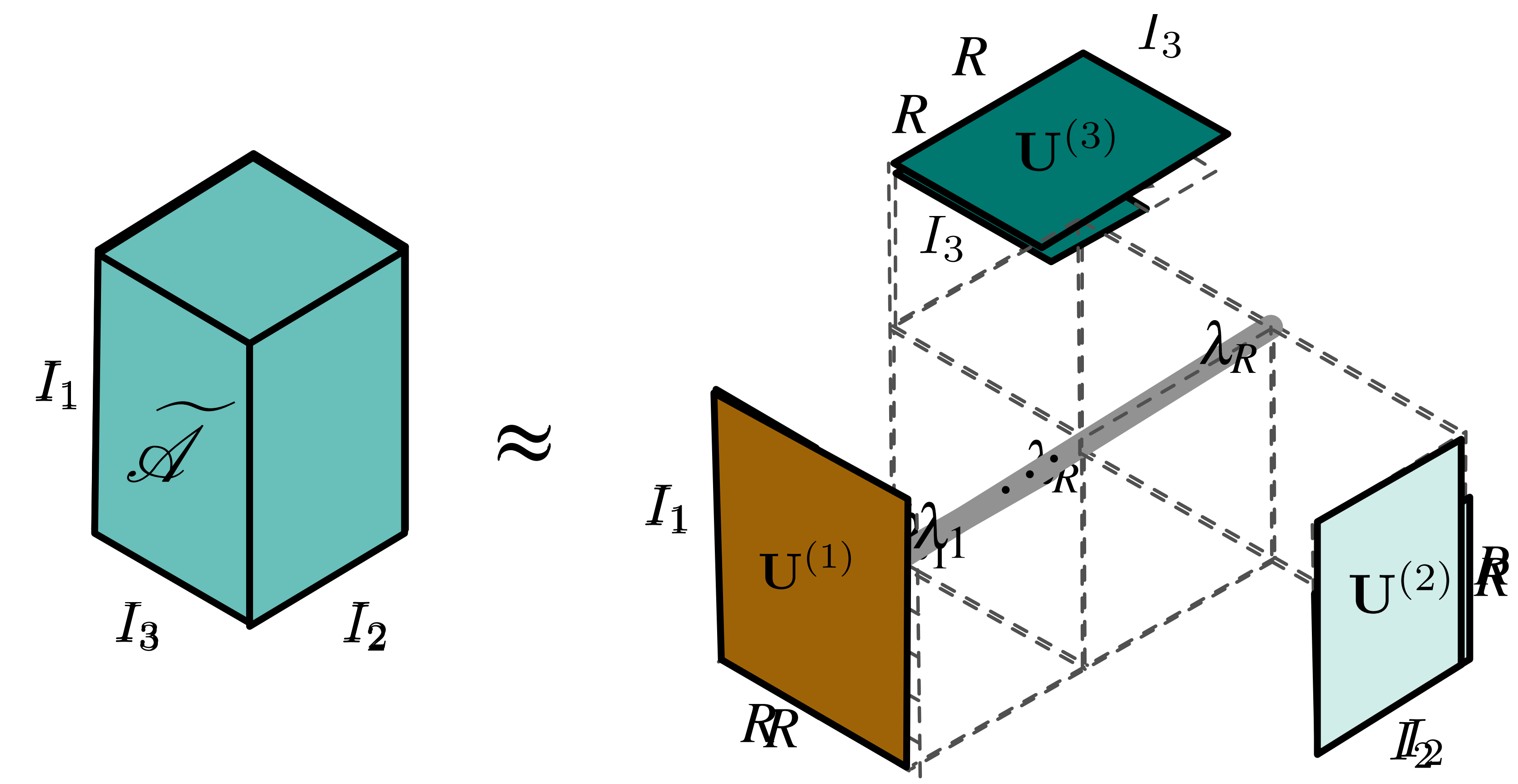


University of Zurich[UZH]

VISUALIZATION AND MULTIMEDIA LAB

Rank-R Approximation

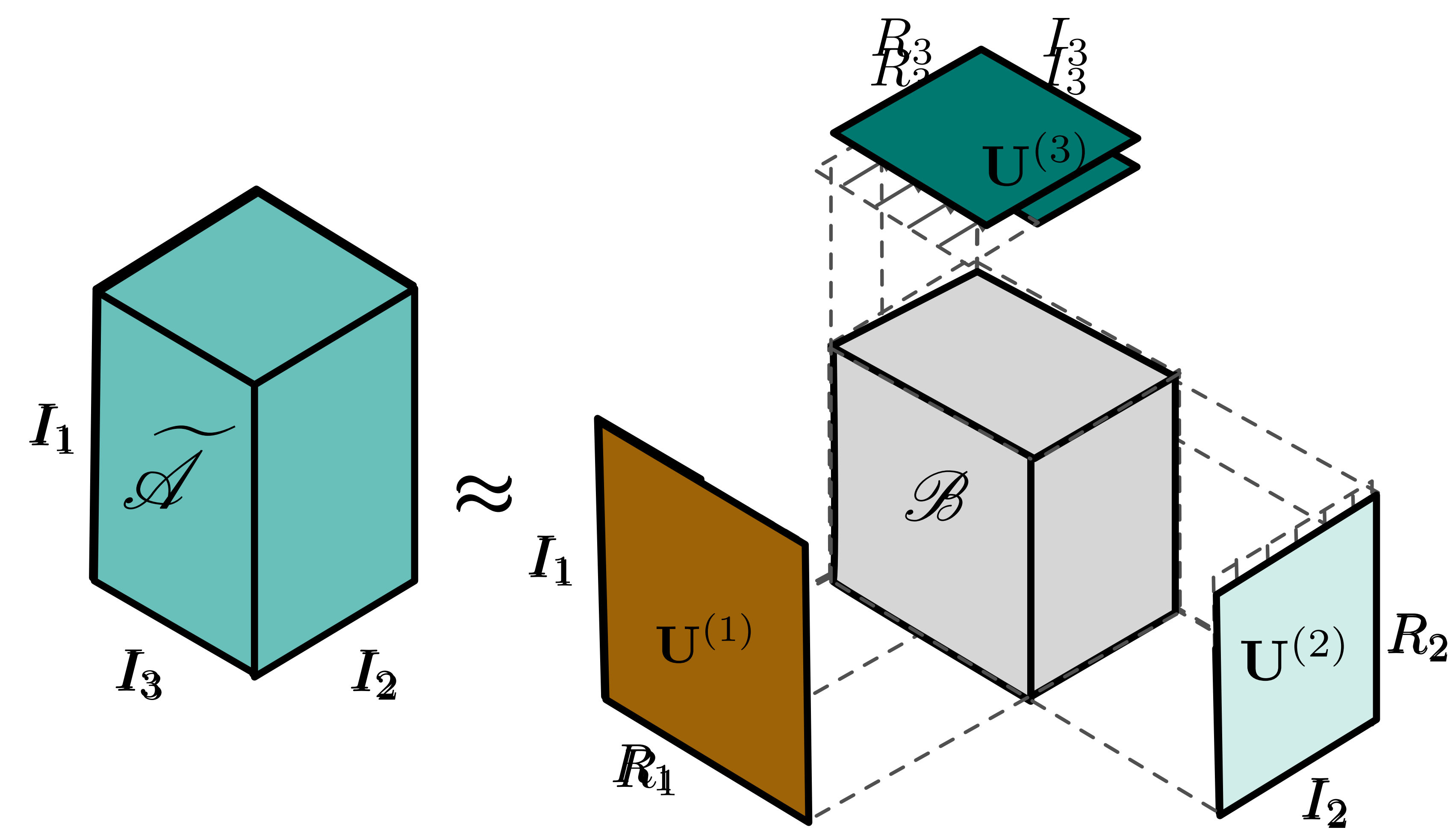
- Approximation of a tensor as a linear combination of ranke-one tensors using a limited number R of terms
  - CP model of limited rank R



$$\widetilde{\mathscr{A}} = \sum_{r=1}^{R} \lambda_r \cdot \boldsymbol{u}_r^{(1)} \circ \boldsymbol{u}_r^{(2)} \circ \ldots \boldsymbol{u}_r^{(N)}$$

Rank-$(R_1, R_2, \ldots, R_N)$ Approximation

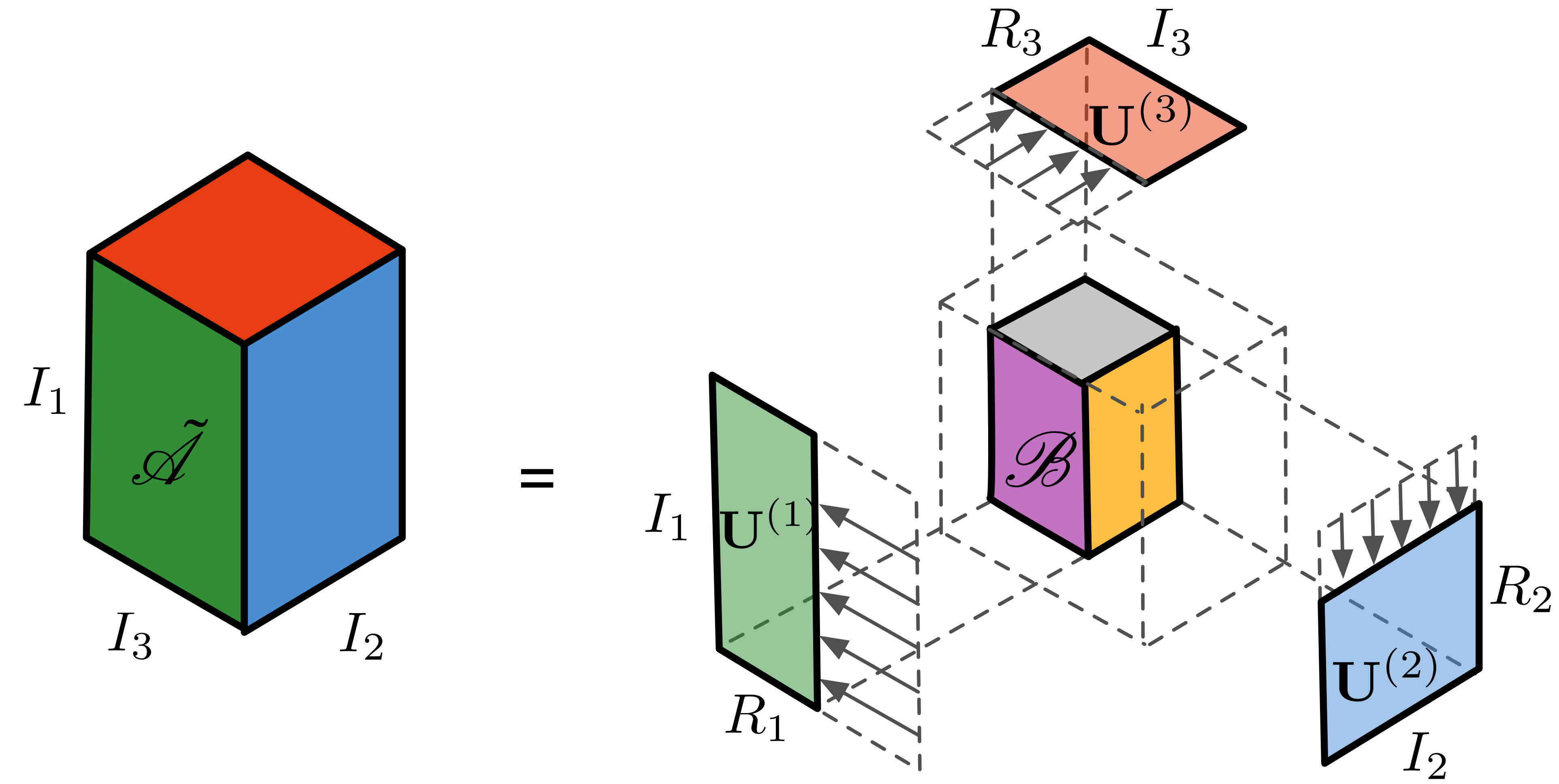
- Decomposition into a tensor with reduced, lower multilinear $\mathrm{rank}(R_1, R_2, \ldots, R_N)$
  - $\mathrm{rank}_n(\widetilde{\mathscr{A}}) = R_n \leq \mathrm{rank}_n(\mathscr{A}) = \mathrm{rank}(\mathbf{A}_{(n)})$
- $\boldsymbol{n}$-mode products of factor matrices and core tensor in a given reduced rank space
  - Tucker model with limited ranks R_i



$$\widetilde{\mathscr{A}} = \mathscr{B} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \cdots \times_N \mathbf{U}^{(N)}$$

Best Rank Approximation

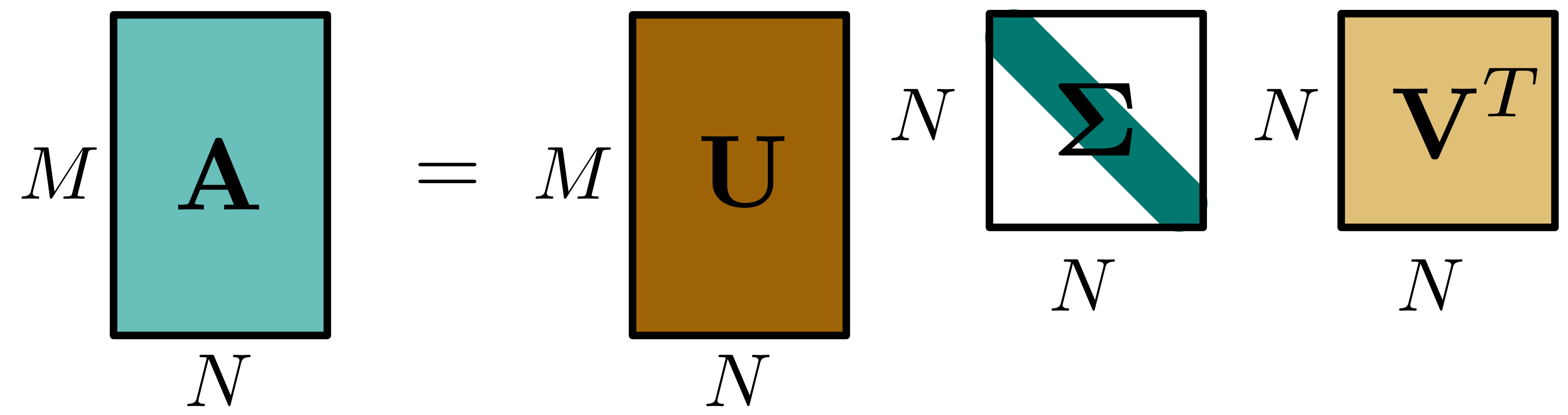
- Rank reduced approximation that minimizes least-squares cost
  - $$\widetilde{\mathscr{A}} = \arg\min(\widetilde{\mathscr{A}}) \left\| \mathscr{A} - \widetilde{\mathscr{A}} \right\|^2$$
- Alternating least squares (ALS) iterative algorithm that converges to a minimum approximation error based on the Frobenius norm $\|\ldots\|_F$
  - rotation of components in basis matrices



$$\widetilde{\mathscr{A}} = \mathscr{B} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)}$$

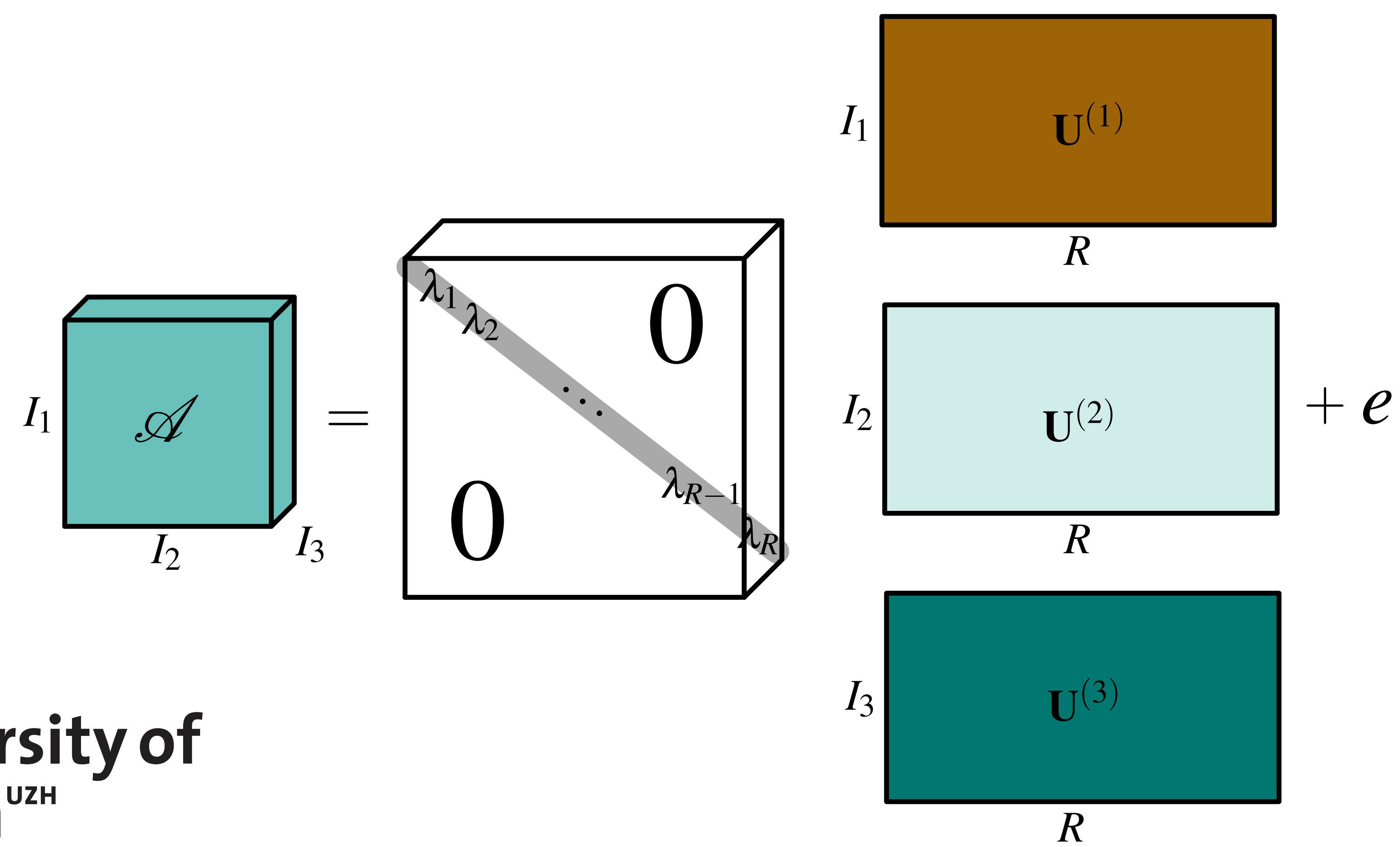
typical high-quality data reduction:
$R_k \leq I_k / 2$

Generalization of the Matrix SVD

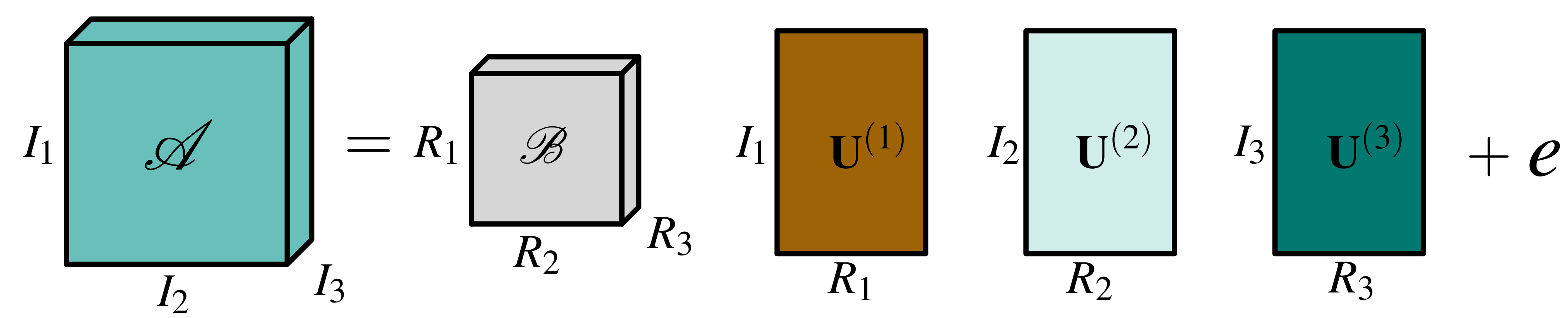


higher orders

CP model



Tucker model



Tutorial: Tensor Approximation in Scientific Visualization

TA Properties and Features

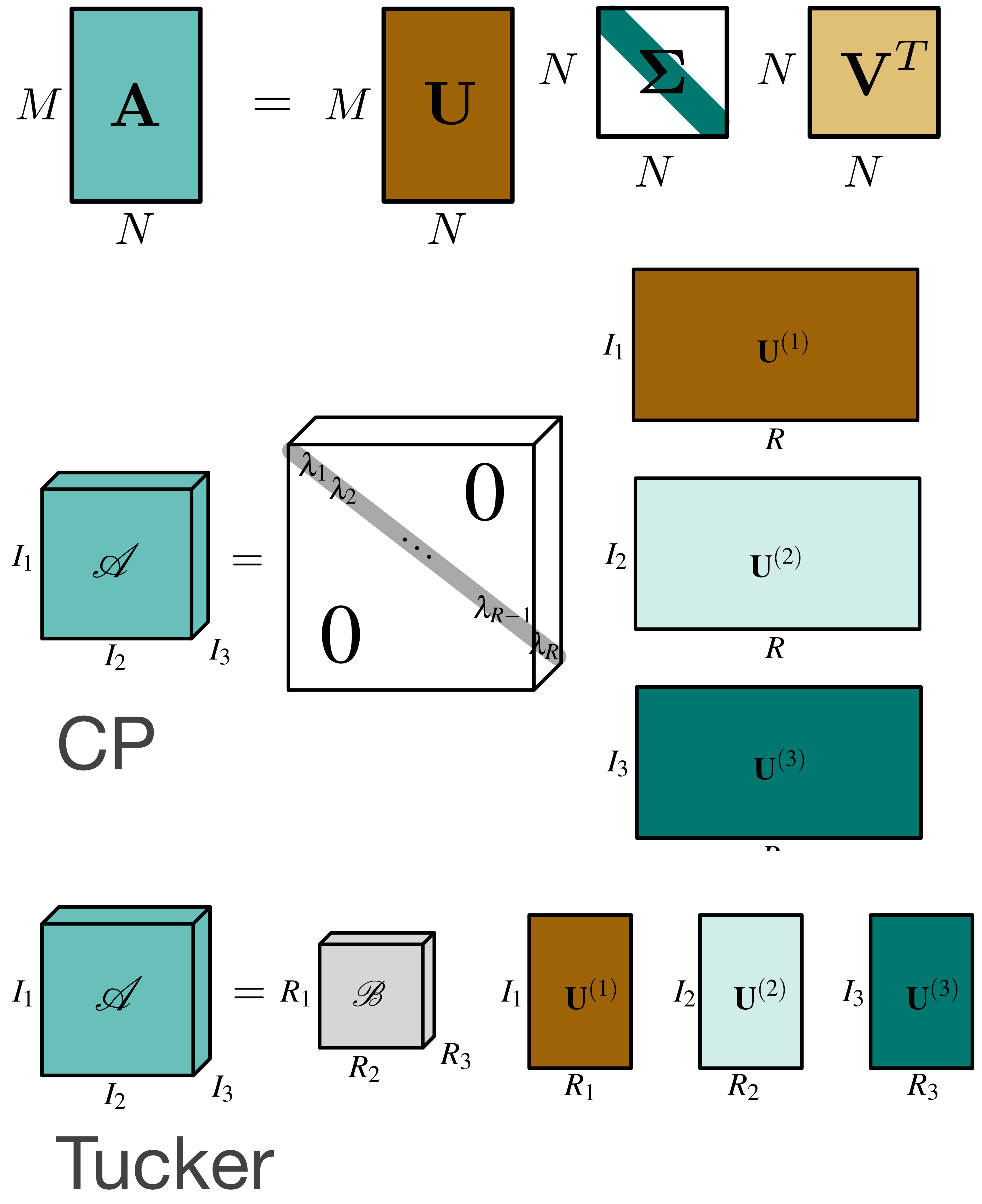
Susanne K. Suter, Rafael Ballester-Ripoll, Renato Pajarola

University of Zurich UZH

VISUALIZATION AND MULTIMEDIA LAB

Properties of Higher Order TA

- Matrix SVD (~PCA)
    - unique
    - rank-R decomposition
    - orthonormal row-space and column-space vectors
- Higher-order tensor decomposition
    - CP model preserves rank-R decomposition
    - all-orthogonal Tucker model preserves orthonormal row-space and column-space vectors



Matrix and Tensor Rank Definitions

- Matrix has unique equal column and row ranks
  - result of SVD
- The n-ranks $R_n = \mathrm{rank}_n(\mathscr{A})$ of a tensor $\mathscr{A}$ may all be different
  - different unfoldings $\mathbf{A}_{(n)}$ give rise to different n-ranks $\mathrm{rank}(\mathbf{A}_{(n)})$
- Matrix rank concept is not uniquely defined for higher order tensors
  - n-rank R_n
  - multilinear rank-$(R_1, R_2, \ldots, R_n)$
  - tensor rank $R = \mathrm{rank}(\mathscr{A})$

University of Zurich[UZH]

VISUALIZATION AND MULTIMEDIA LAB

Tensor Rank R

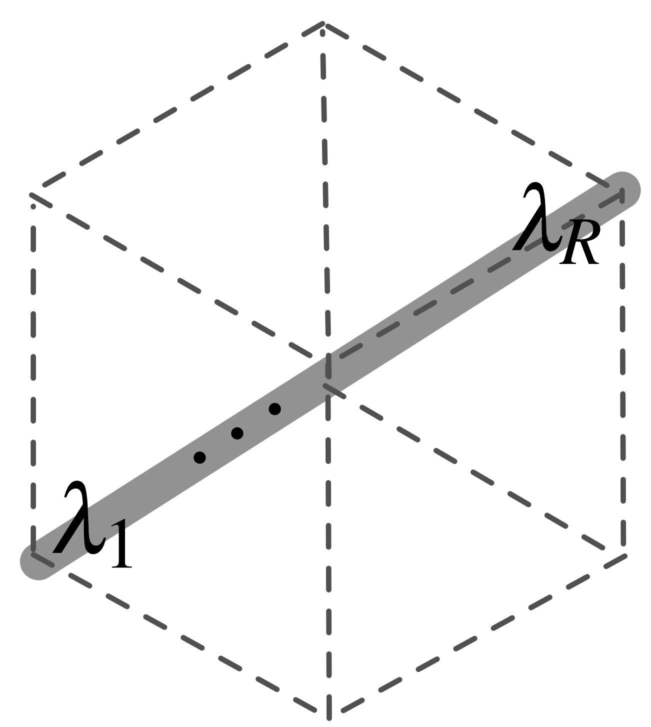
- The ***tensor rank*** $R = \mathrm{rank}(\mathscr{A})$ is the minimal number of rank-one tensors $\mathscr{A}_i$ that yield $\mathscr{A}$ in a linear combination
  - $\mathscr{A}_i$ are rank-one tensors, defined by outer product of N vectors
- Equal to the column and row rank for matrices
- Not necessarily equal to any n-rank R_n of a tensor
  - and it holds that $R \geq R_n$

Rank-R Decomposition

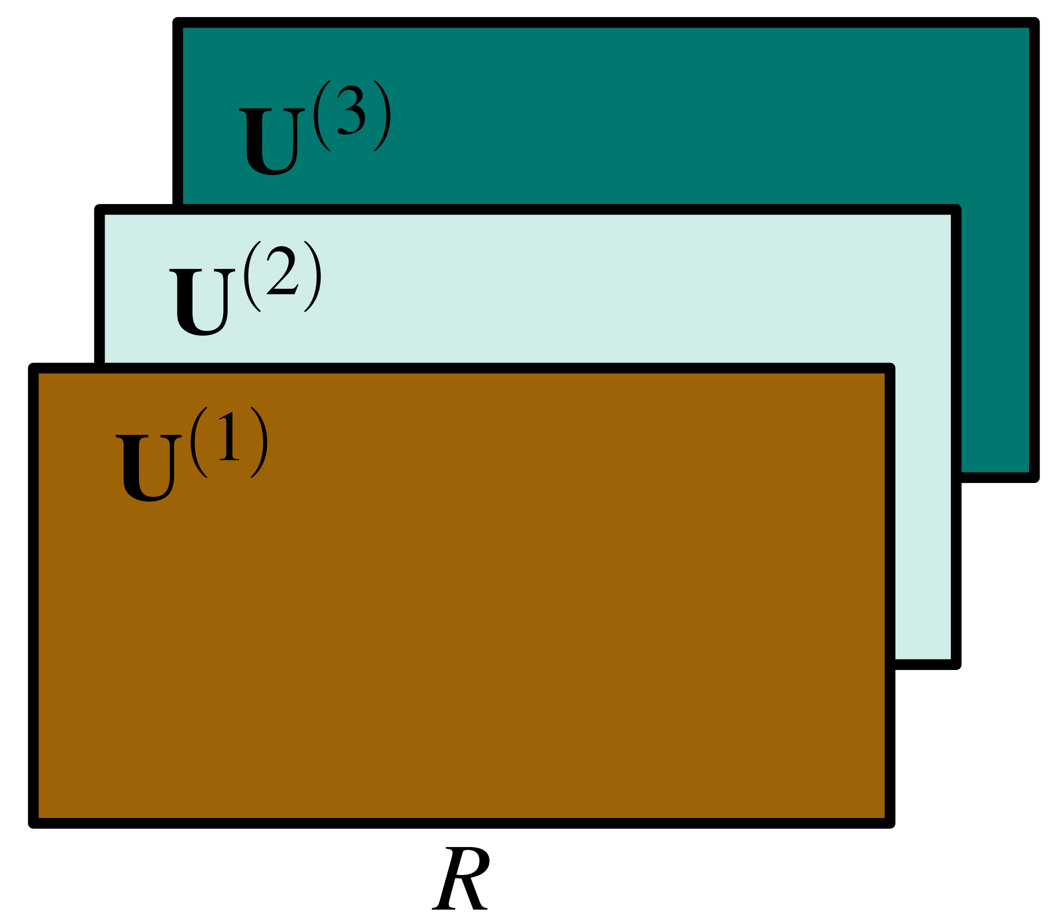
- Minimal number R of rank-one tensors $\mathscr{A}_i$ that yield $\mathscr{A}$ in a linear combination, $\mathscr{A} = \lambda_1 \mathscr{A}_1 + \lambda_2 \mathscr{A}_2 + \ldots + \lambda_R \mathscr{A}_R$
- CP model allows a direkt rank-R decomposition with respect to the *tensor rank R*

$$\mathscr{A} = \sum_{r=1}^{R} \lambda_r \cdot \boldsymbol{u}_r^{(1)} \circ \boldsymbol{u}_r^{(2)} \circ \cdots \circ \boldsymbol{u}_r^{(N)}$$

coefficients



factor matrices



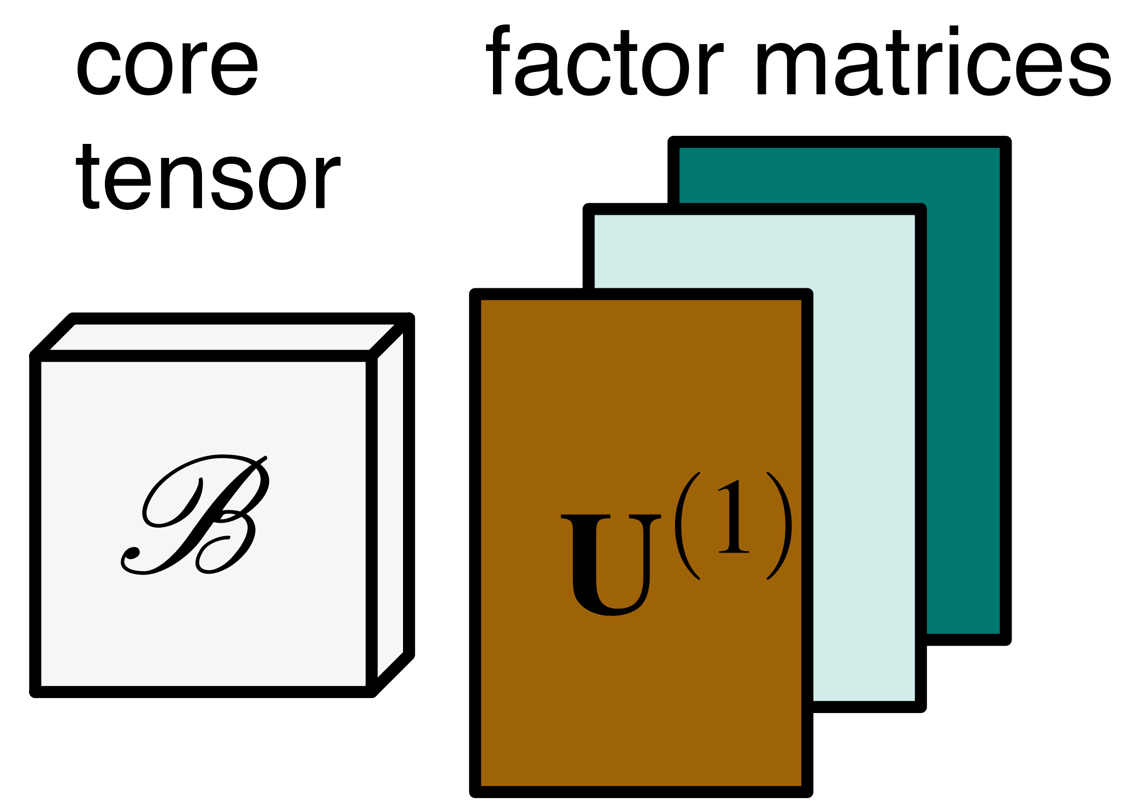
Uniqueness

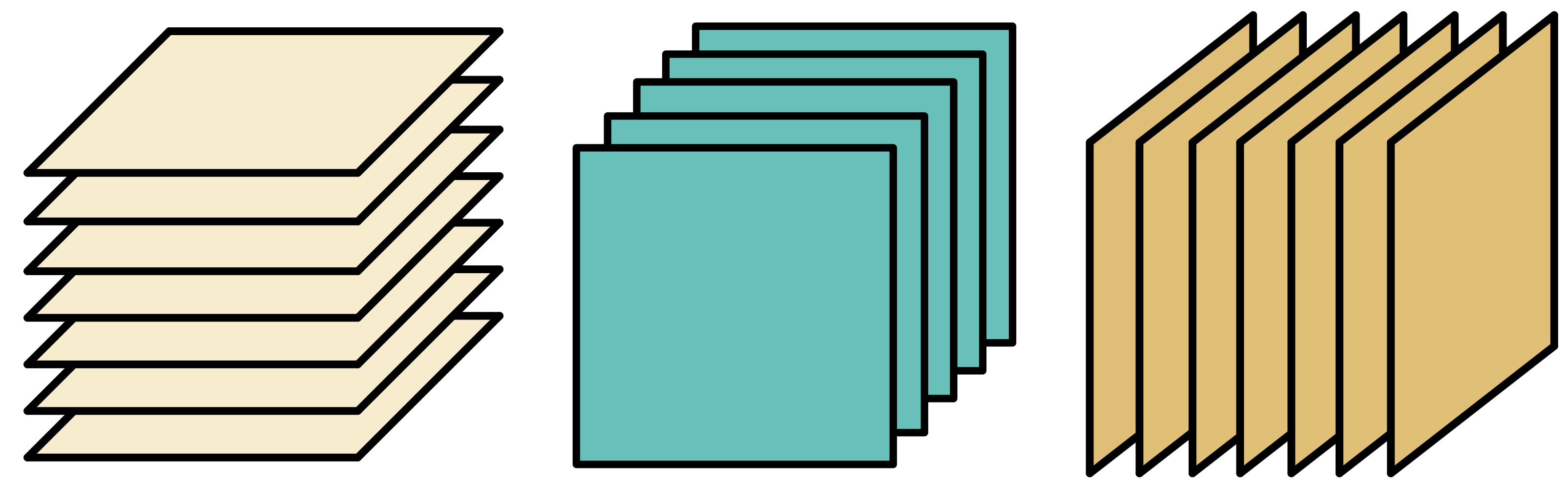
- Unique if it is the only possible decomposition
  - except for indeterminacies of scaling and permutations
- Rank-R decompositions of higher-order tensors are often unique
- Matrix rank decompositions are not generally unique,
  except e.g. for the SVD
  - due to the orthogonality constraints, and
  - the diagonal matrix of ordered singular values
- The CP decomposition is unique under weaker conditions (than the SVD)
  - non-orthogonal factor matrices
- The Tucker decomposition is not unique

[Kolda and Bader, 2009]

Orthonormality

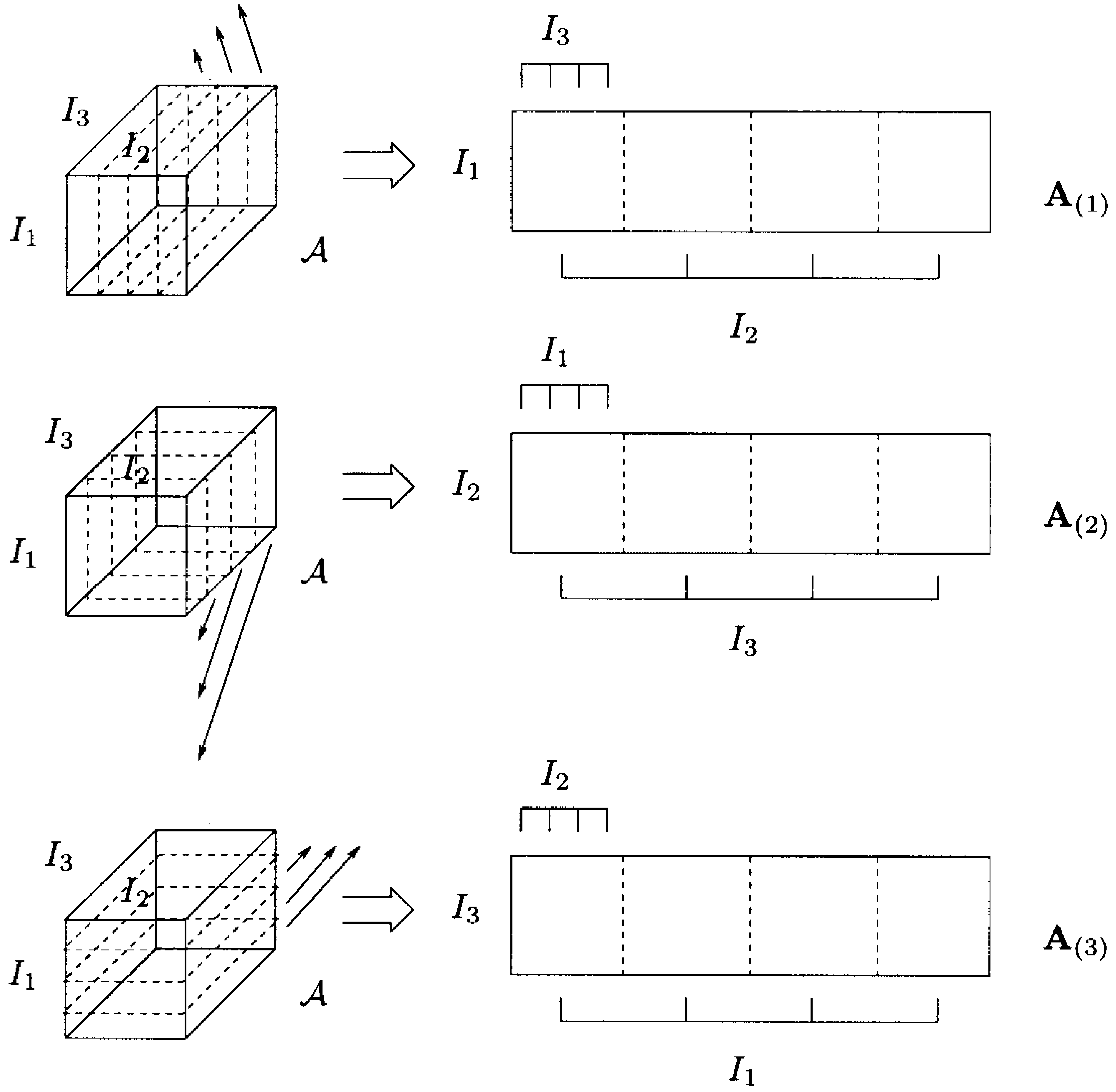
- Matrix SVD generates orthonormal bases $\mathbf{U}$ and $\mathbf{V}$
- A Tucker model can be formed with orthonormal factor matrices
  - all-orthogonal Tucker core tensor $\mathscr{B}$
- All-orthogonality example for third-order tensor:
  - horizontal matrices are mutually orthogonal with respect to the scalar product of matrices
    - the sum of the products of the corresponding entries vanishes
  - the same holds for all frontal slices and lateral slices
    - see De Lathauwer et al., 2000a





Higher-Order SVD (HOSVD)

Tensor unfolding



HOSVD algorithm

- SVD on every mode's tensor unfolding $\mathbf{A}_{(n)}$
  - set basis factor matrices $\mathbf{U}^{(n)}$ as R leading left singular vectors of $\mathbf{A}_{(n)}$
- Derive core $\mathscr{B}$ from original data and inverse factor matrices
  - defines a Tucker model with $\mathscr{B}$, $\mathbf{U}^{(n)}$

$$\widetilde{\mathscr{A}} = \mathscr{B} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)}$$
$$\mathscr{B} = \mathscr{A} \times_1 \mathbf{U}^{(1)^{-1}} \times_2 \mathbf{U}^{(2)^{-1}} \times_3 \mathbf{U}^{(3)^{-1}}$$

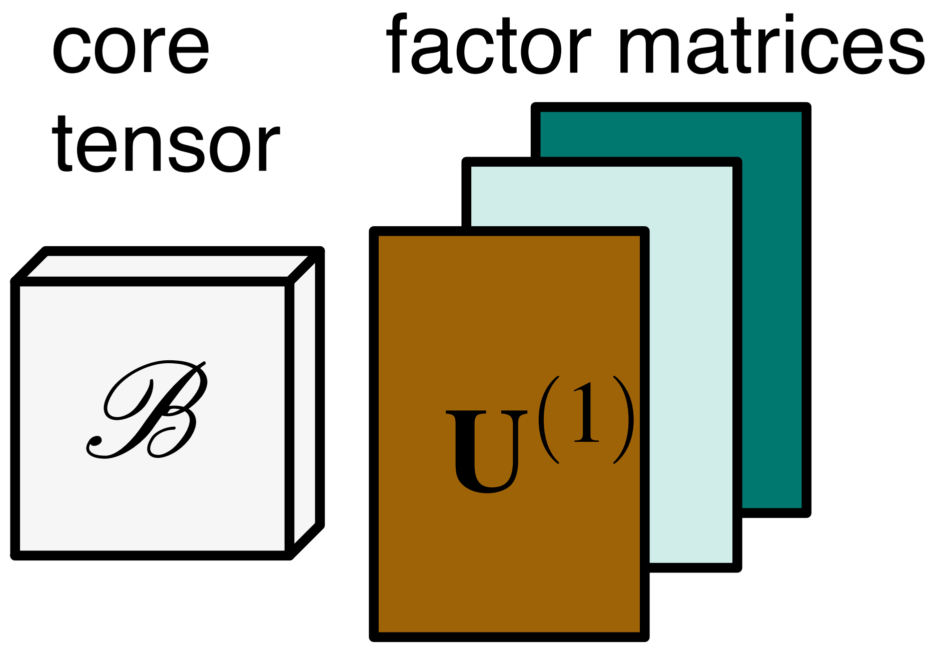
HOSVD: Lathauwer et al., 2000a

VISUALIZATIONAND MULTIMEDIALAB

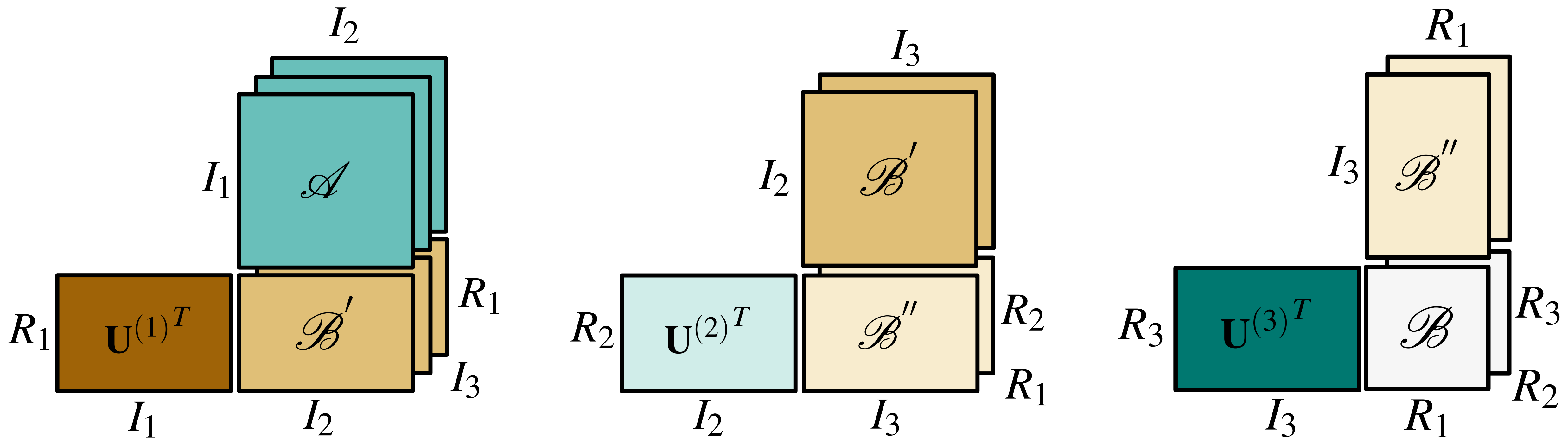
Tucker Core

- Tucker column vectors of factor matrices $\mathbf{U}^{(n)}$ are often defined to be orthonormal
- Core tensor $\mathscr{B}$ represents projection of data $\mathbf{A}$ onto its factor matrices $\mathbf{U}^{(n)}$, thus is a representation in new bases
  - computed using transposes for orthogonal factor matrices

$$\mathscr{B} = \mathscr{A} \times_1 \mathbf{U}^{(1)^{-1}} \times_2 \mathbf{U}^{(2)^{-1}} \times_3 \mathbf{U}^{(3)^{-1}}$$

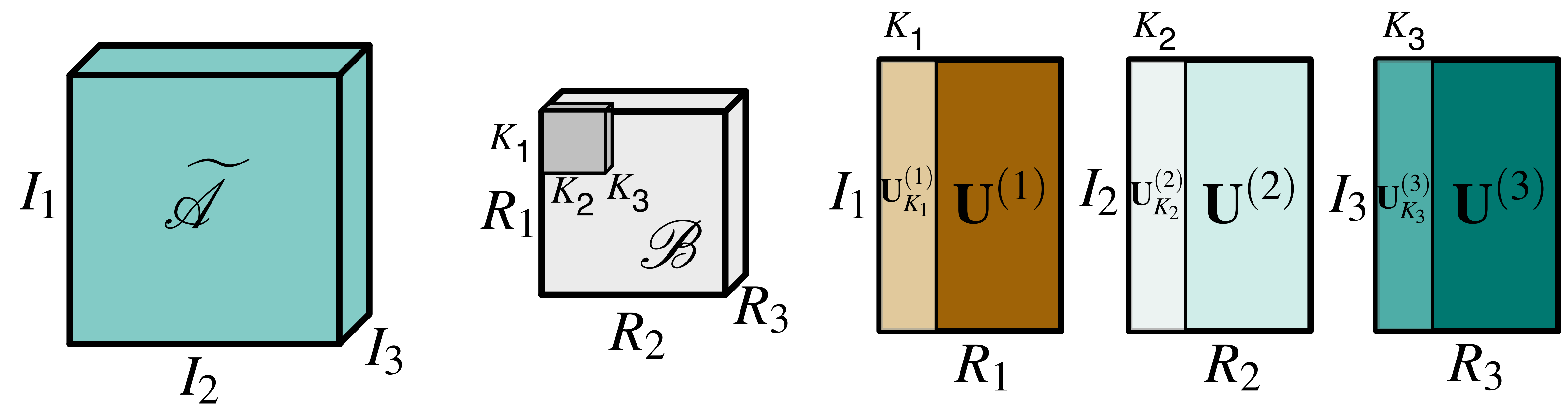


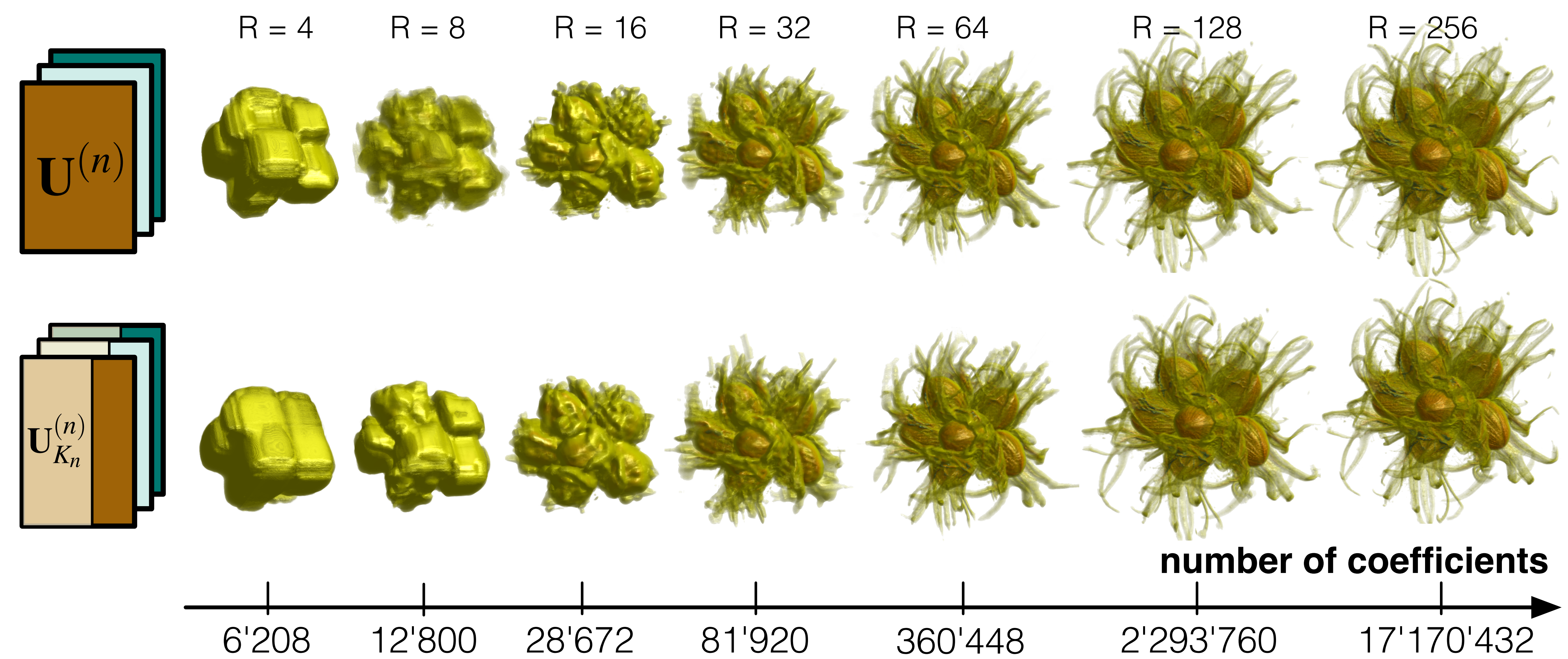
- Optimized order of computation



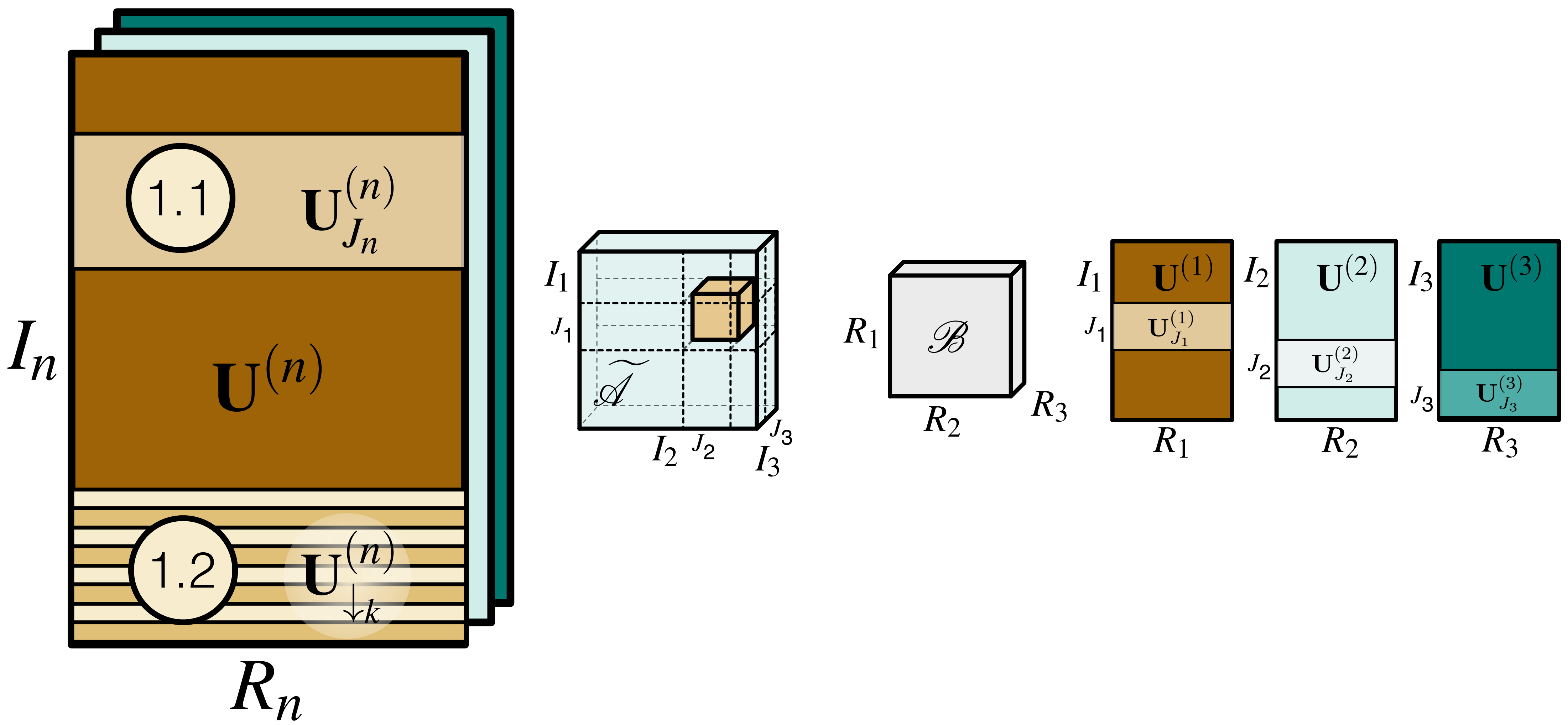
Rank Truncation

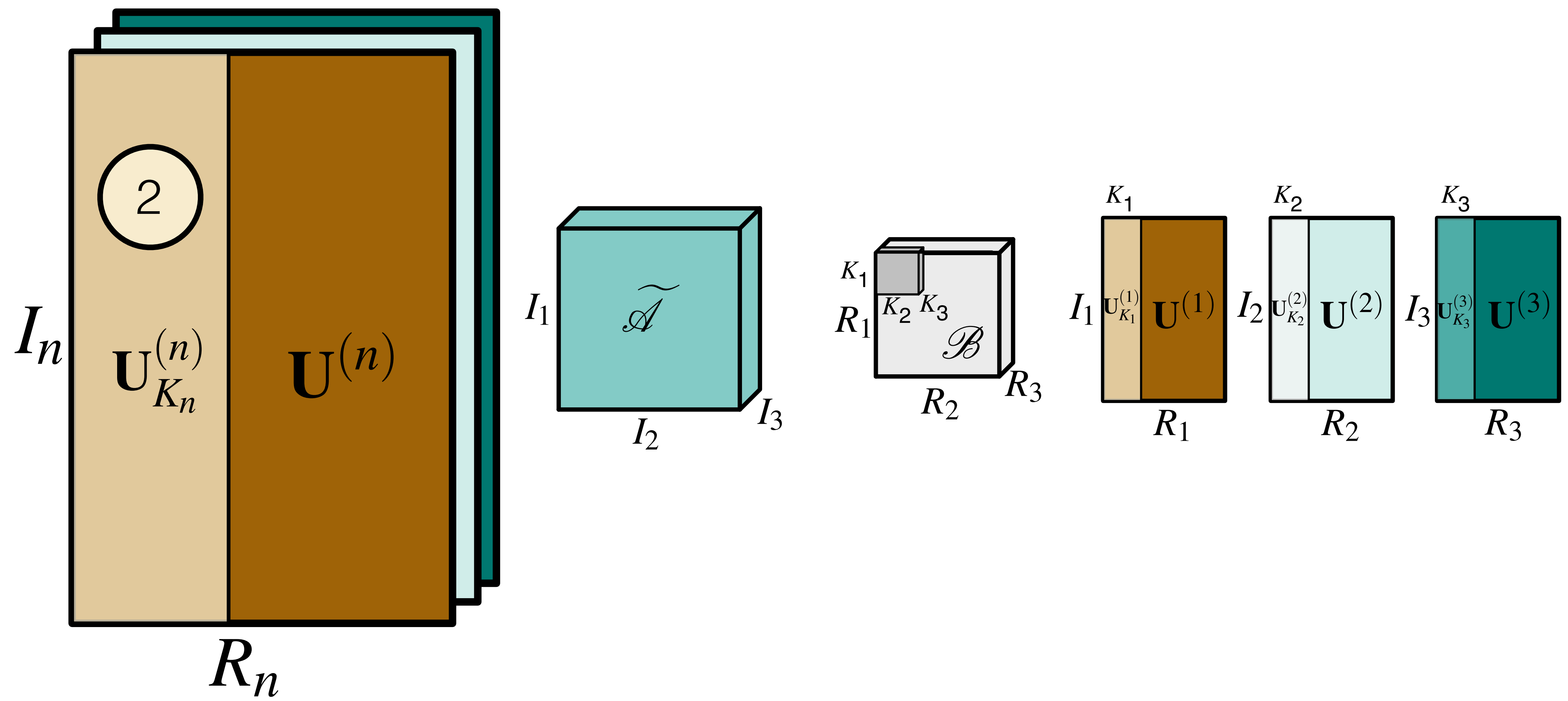
- SVD allows for progressive rank truncation
  - orthogonality of singular vectors
  - order of increasing singular values
- CP does not exhibit good progressive truncation behavior
  - non-orthogonal factor matrices
- All-orthogonal Tucker model supports progressive truncation
  - does not necessarily give best possible progression





Properties of Tucker Factor Matrices

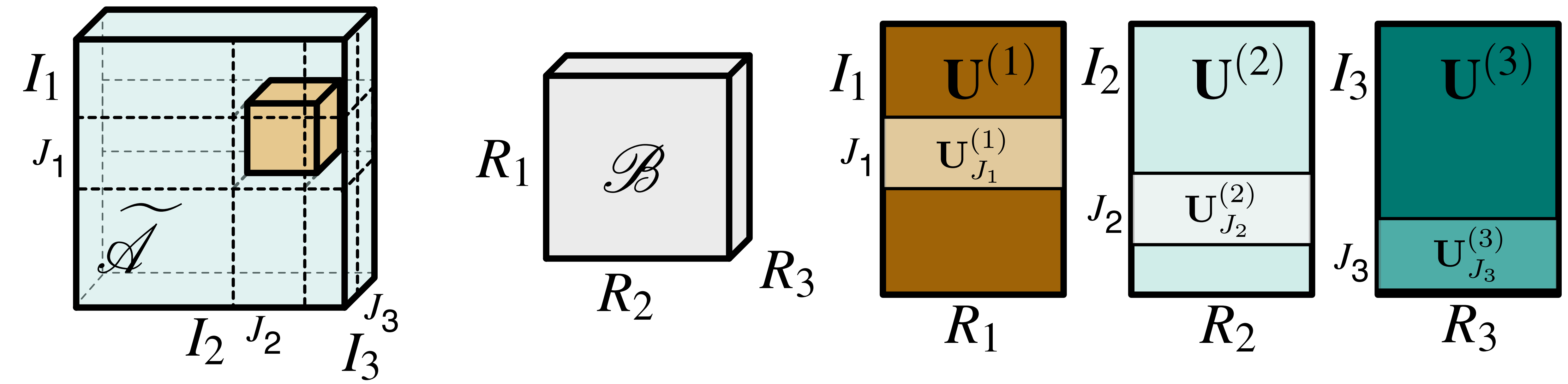




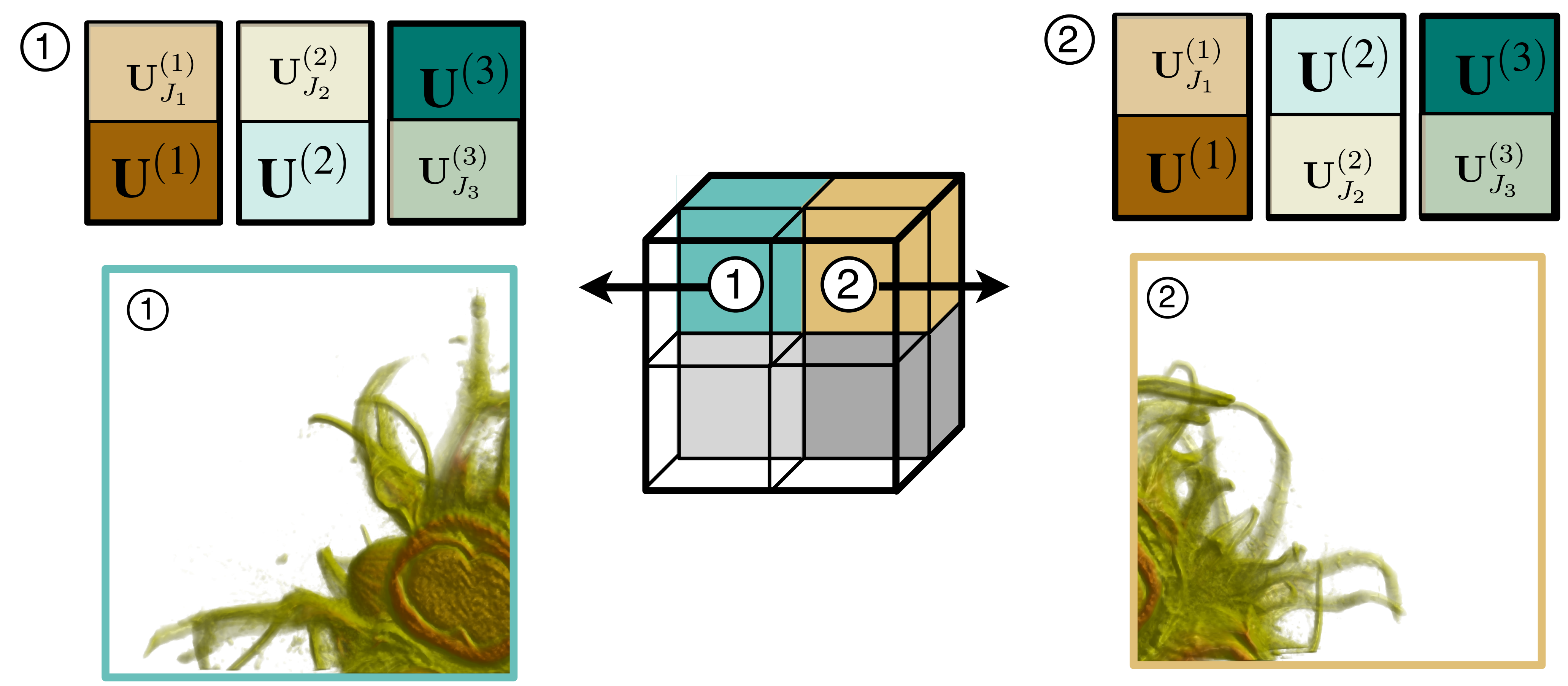
- Vectors along horizontal axis (rows)
  - 1.1; spatial selectivity
  - 1.2; spatial subsampling
- Vectors along vertical axis (columns)
  - 2: rank reduction

Spatial Selection in Factor Matrices

- Select submatrices $\mathbf{U}^{(n)}{}_{J_n}$ (a selection of J_n row vectors)
  - reconstruct only from submatrices and core tensor
- Core tensor stays unchanged
- Potential applications
  - view-frustum culling
  - adaptive spatial selection (multiresolution DVR)

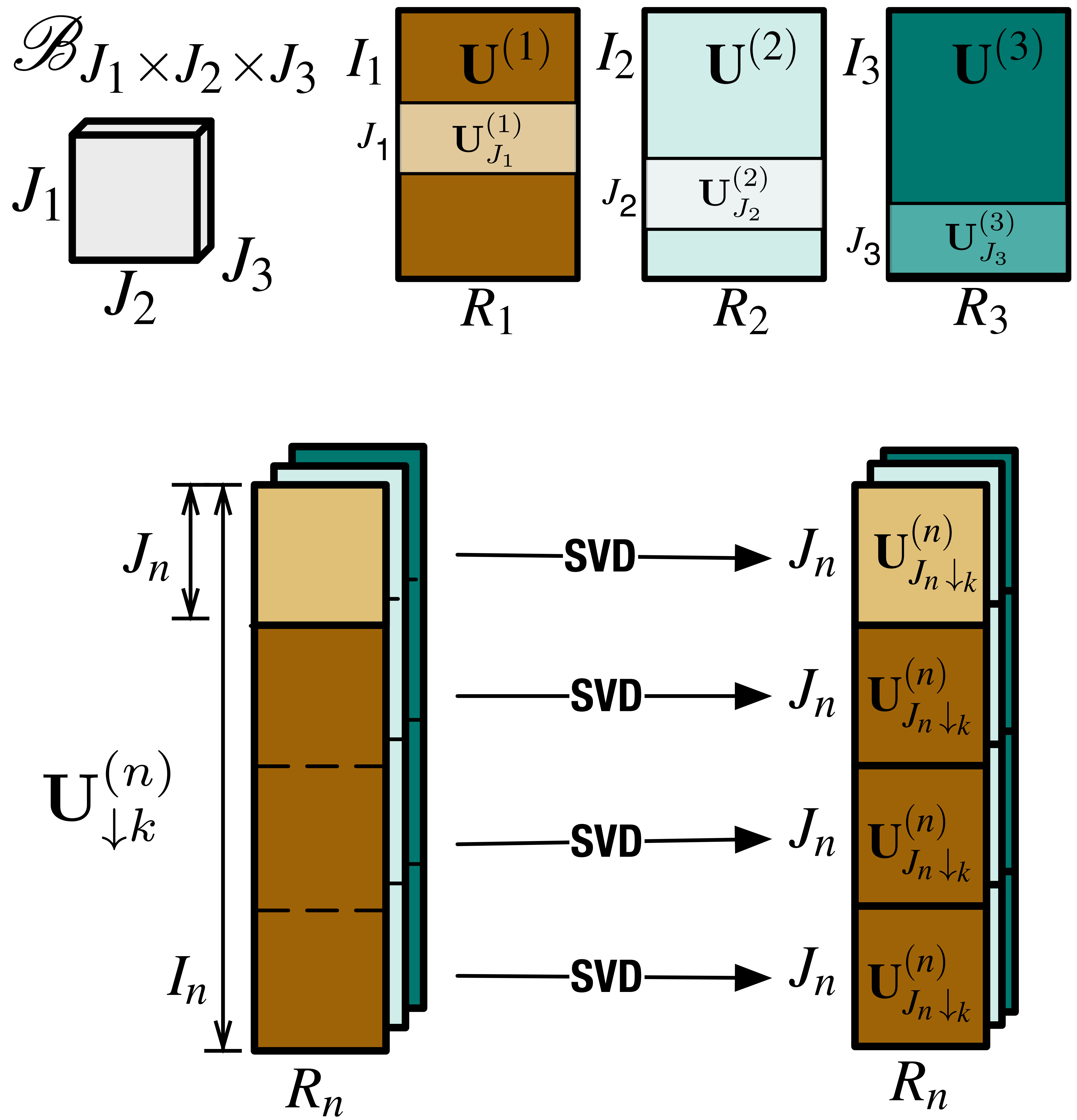


$$\widetilde{\mathscr{A}}_{J_1 \times J_2 \times J_3} = \mathscr{B} \times_1 \mathbf{U}^{(1)}_{J_1} \times_2 \mathbf{U}^{(2)}_{J_2} \times_3 \mathbf{U}^{(3)}_{J_3}$$



Orthogonality Issues and Truncation

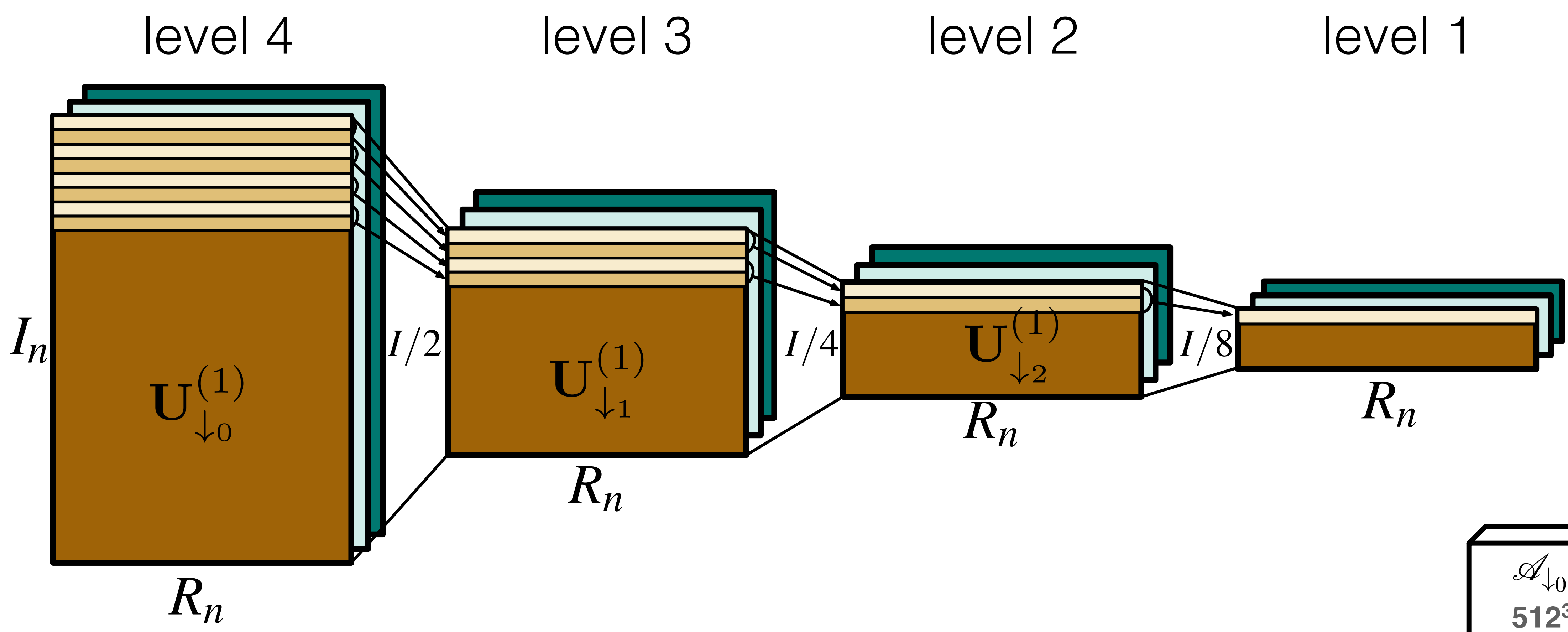
- Spatial selection of factor matrix row ranges destroys the orthogonality property
- Newly derived, spatially local tensor cores from non-orthogonal factor submatrices are not all-orthogonal
  - but only the all-orthogonality makes core tensors rank-reducible
- In order to achieve rank-reducible core tensors, another SVD is applied to spatially selective or averaged submatrices
  - see Tsai and Shih, 2012; Suter, 2013



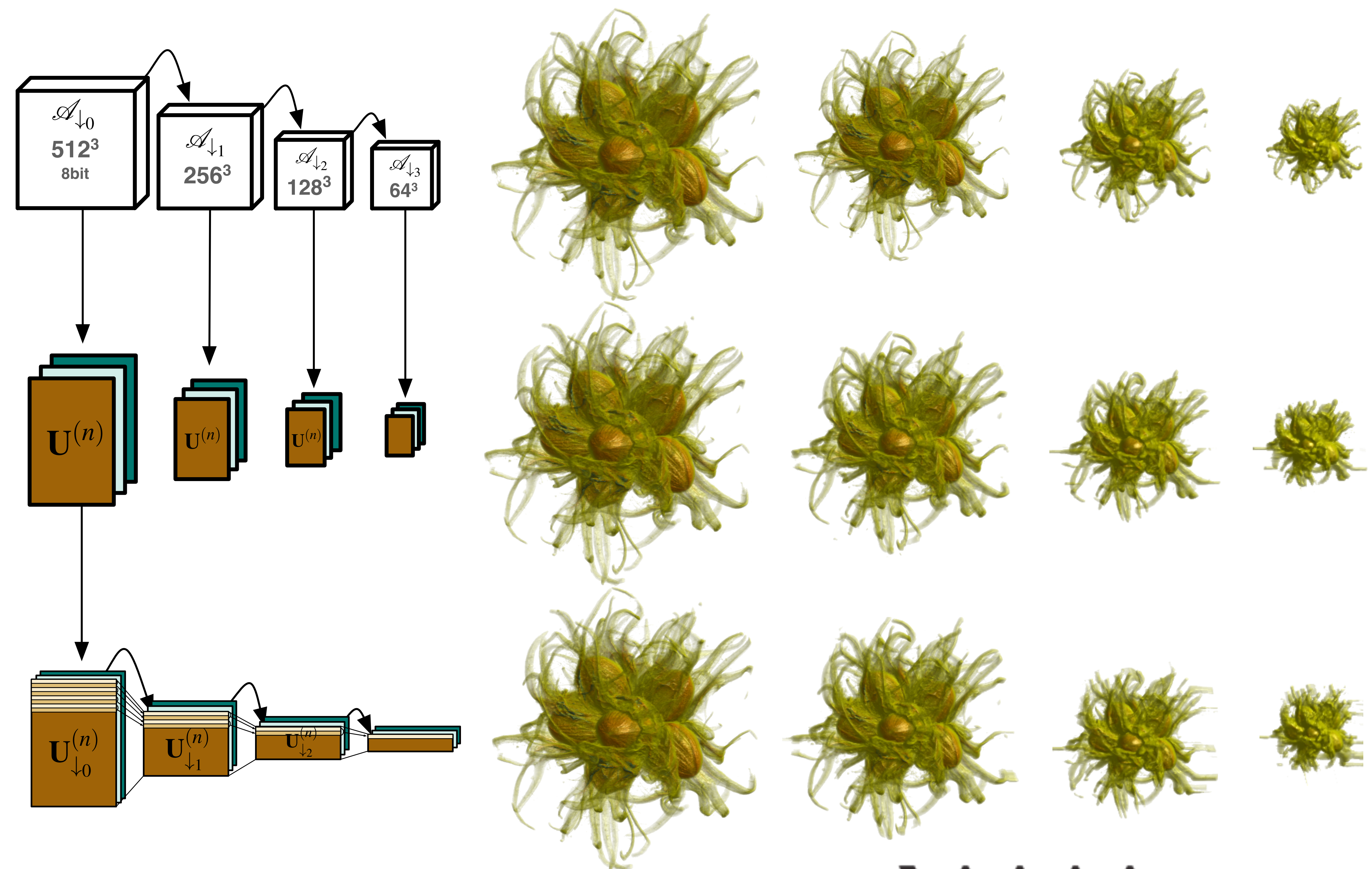
University of Zurich UZH

VISUALIZATION AND MULTIMEDIA LAB

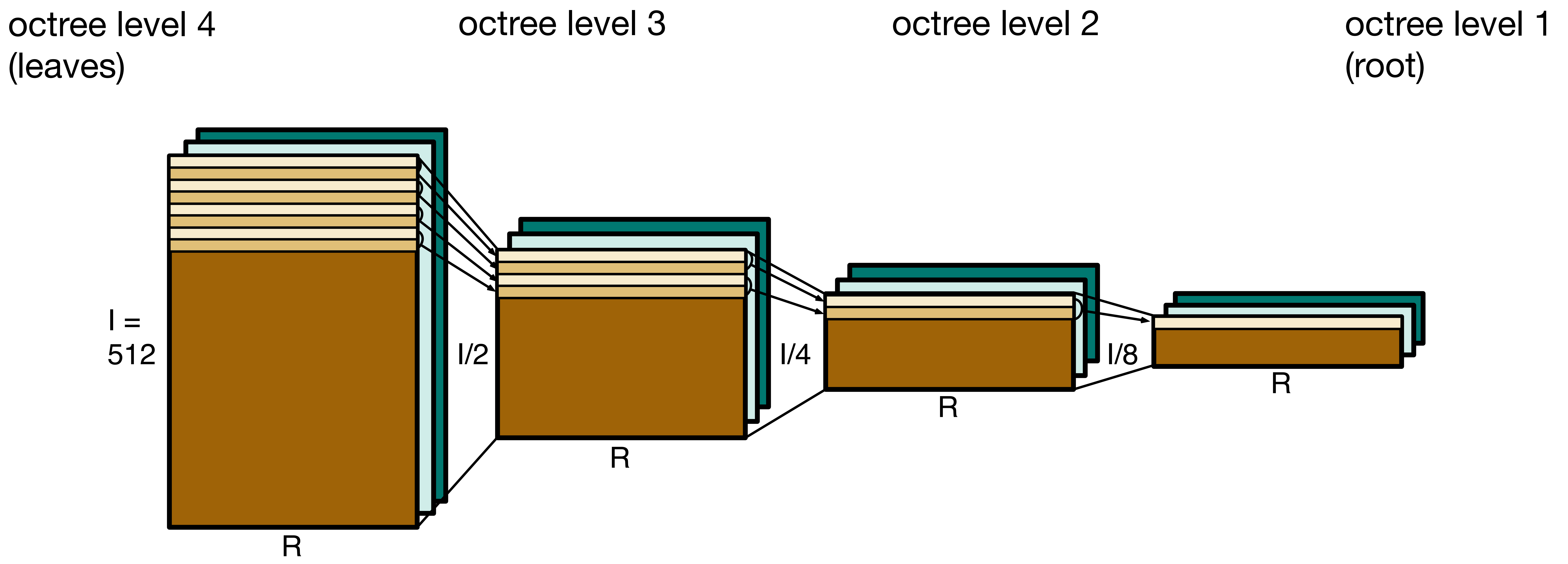
Spatial Subsampling in Factor Matrices



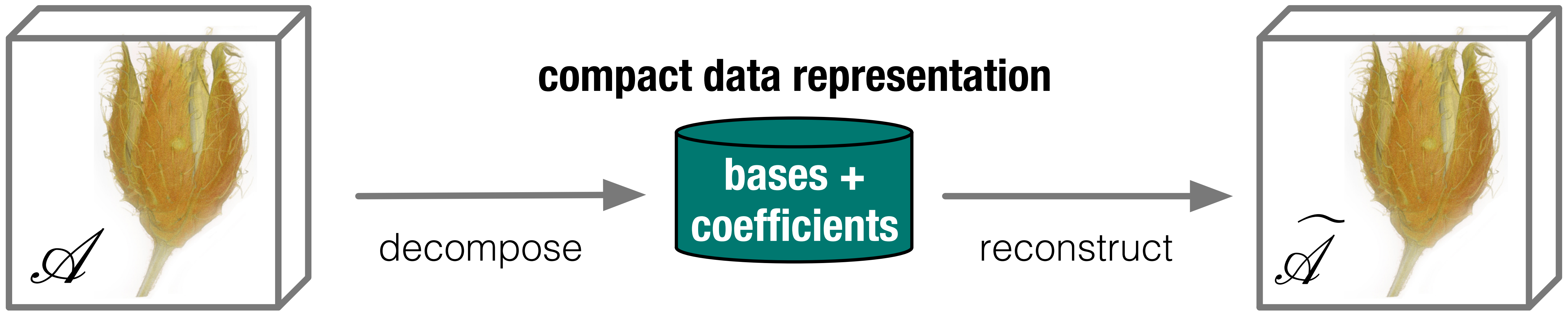
- Spatial correspondence of rows allows for averaging or subsampling of factor matrix row vectors
- Potential applications
  - multiresolution modeling



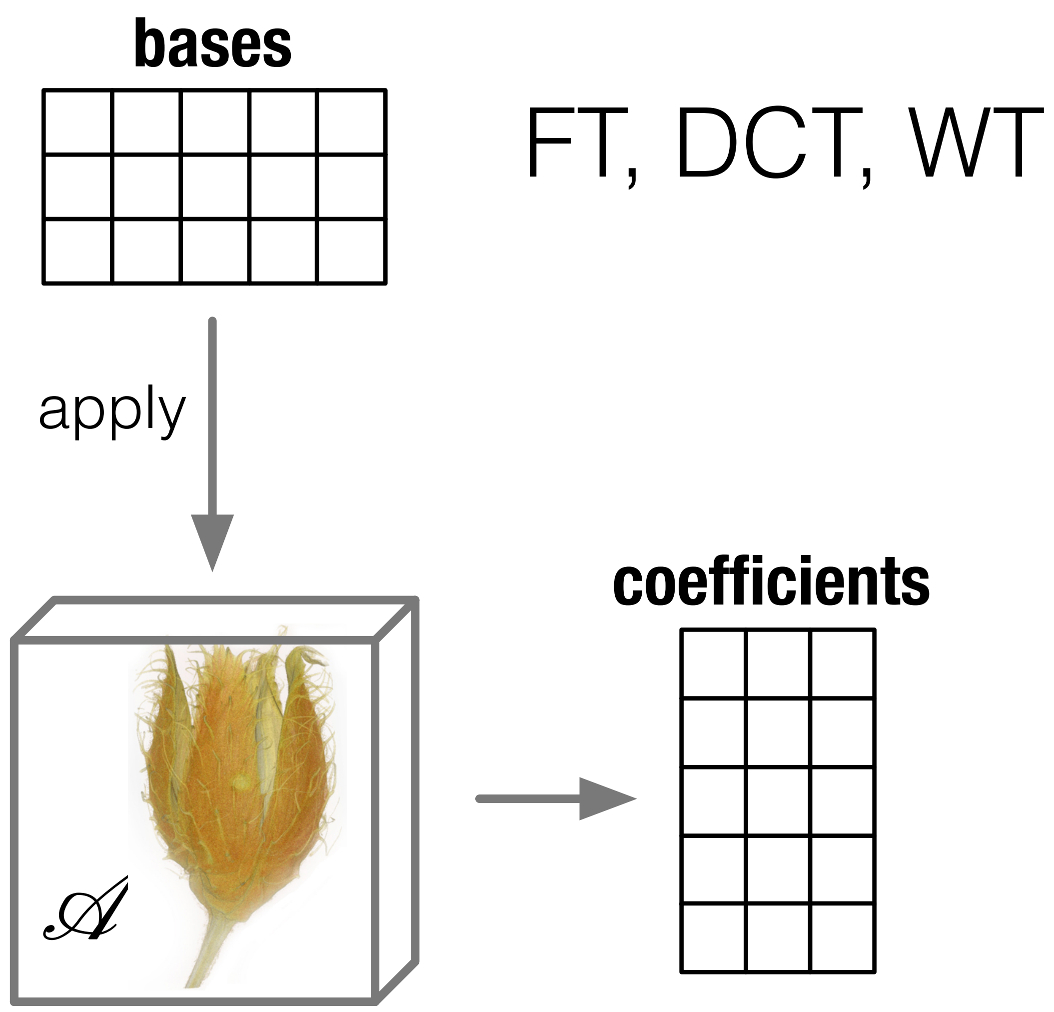
Global Factor Matrices Octree Hierarchy



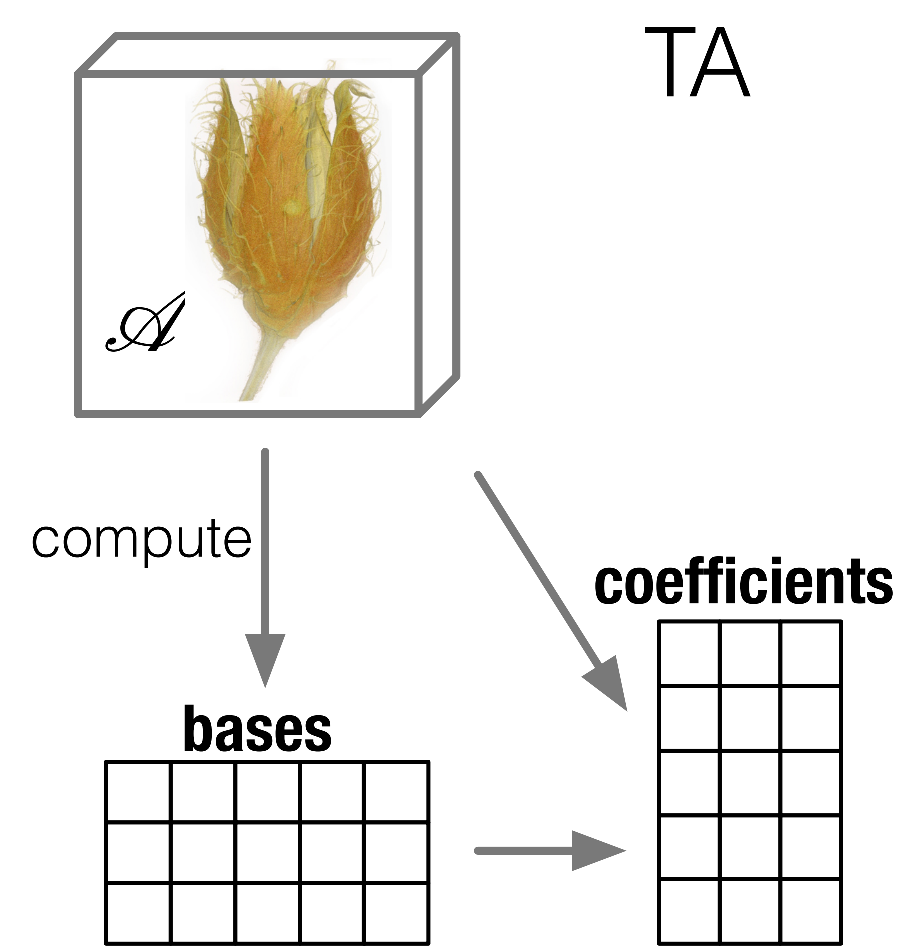
Pre-Defined vs. Learned Bases



Pre-defined bases



Learned bases

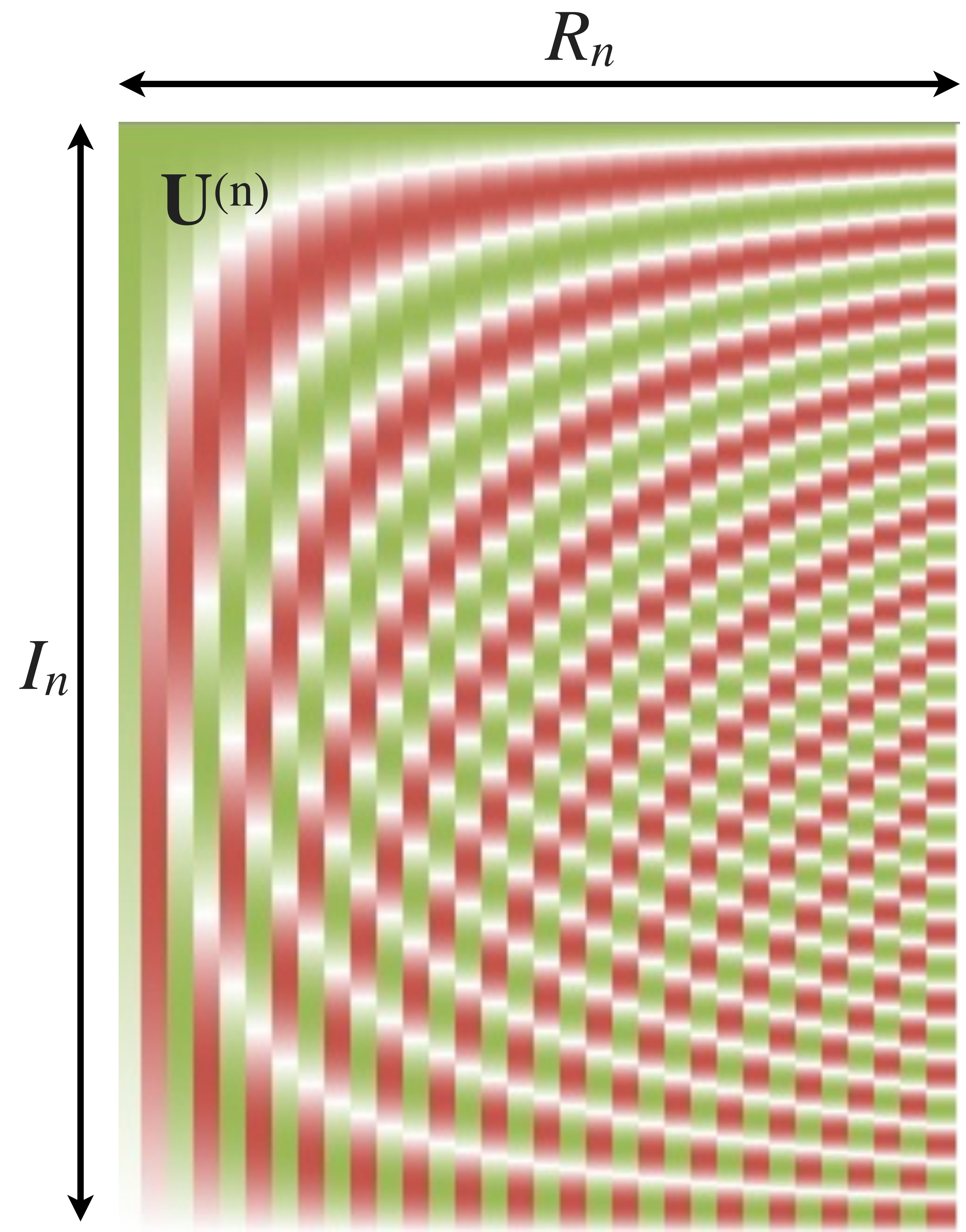


DCT as Tucker Decomposition

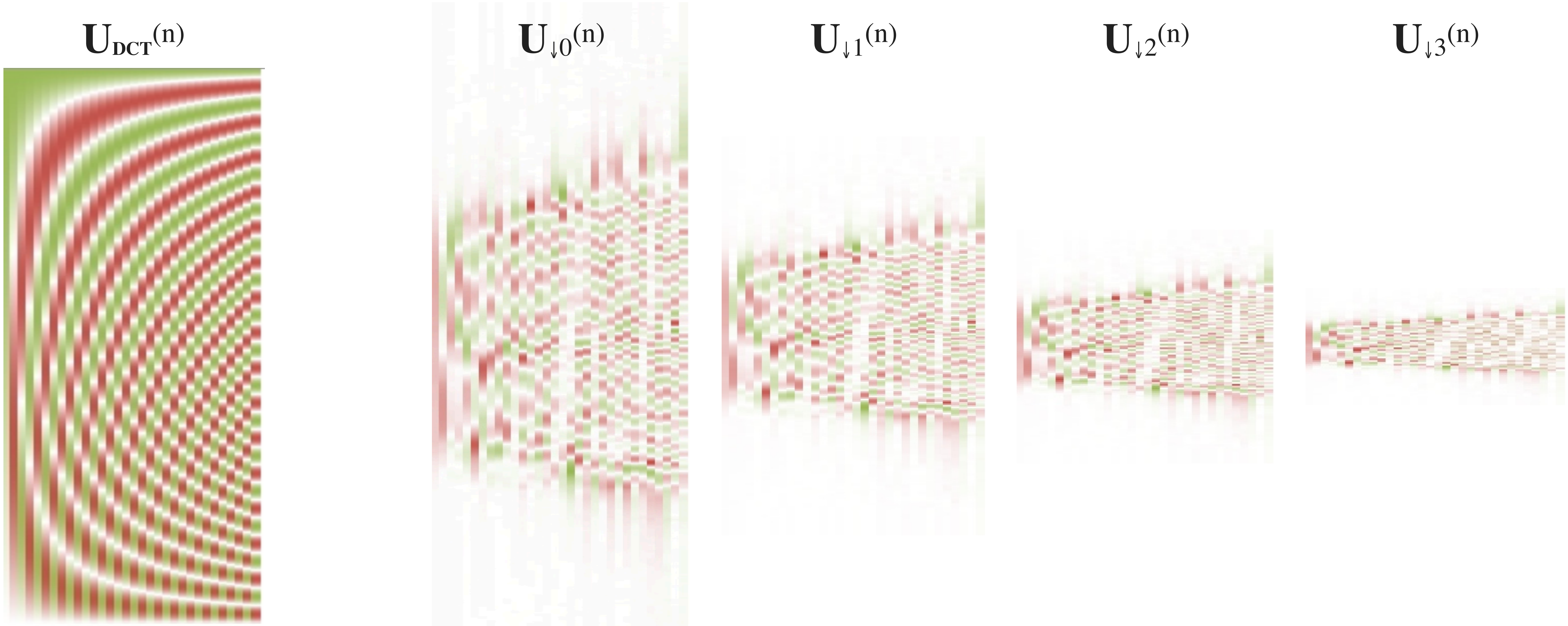
- The DCT of a 2D image or higher order tensor directly maps to the Tucker tensor decomposition model
  - tensor decomposition using pre-defined basis factor matrices
- Using the DCT type-II formulation, the basis matrices $\mathbf{U}^{(n)}$ entries can be formed by:

$$u_{ij}^{(n)} = C_i \cos\left(\frac{(2(j-1)+1)(i-1)\pi}{2I_n}\right)$$

  - where $i \in \{1, \ldots, I_n\}$ and $\mathrm{j} \in \{1, \ldots, R_n\}$



Example of Subsampled TA Factor Matrices



DCT factor matrix

Tucker factor matrices

VISUALIZATIONAND
MULTIMEDIALAB

Tucker Reconstruction

$$\widetilde{\mathscr{A}} = \mathscr{B} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)}$$

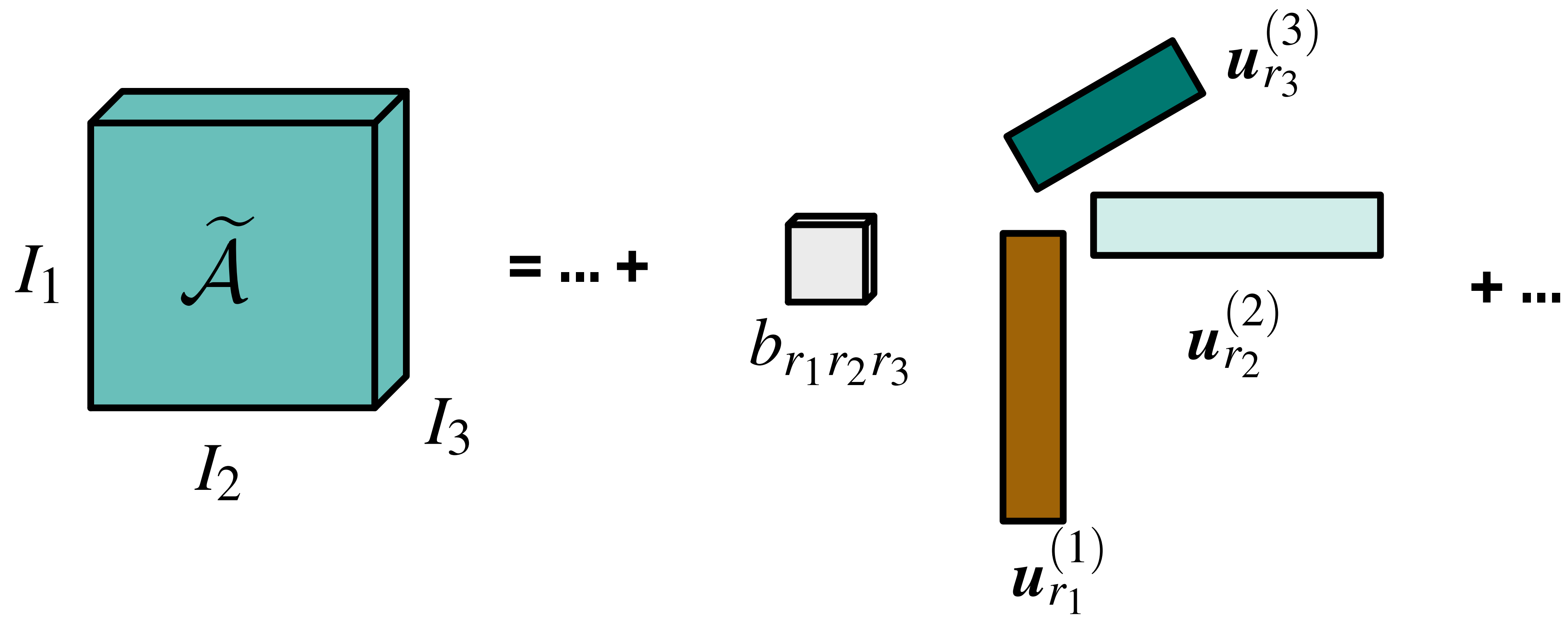
- Reconstruction from rank-one tensors

$$\widetilde{\mathscr{A}} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} b_{r_1,r_2,r_3} \cdot \boldsymbol{u}_{r_1}^{(1)} \circ \boldsymbol{u}_{r_2}^{(2)} \circ \boldsymbol{u}_{r_3}^{(3)}$$

-> progressive reconstruction

- Element-wise reconstruction

$$\widetilde{a}_{i_1 i_2 i_3} = \sum_{r_1} \sum_{r_2} \sum_{r_3} b_{r_1 r_2 r_3} \cdot u_{i_1 r_1}^{(1)} \cdot u_{i_2 r_2}^{(2)} \cdot u_{i_3 r_3}^{(3)}$$



Reconstruction Complexity

(1)

$$\widetilde{\mathcal{A}} = \sum_{r_k} \mathcal{B}[r_1, \ldots, r_N] \cdot \mathbf{U}^{(1)}_{(r_1)} \otimes \mathbf{U}^{(2)}_{(r_2)} \cdots \otimes \mathbf{U}^{(N)}_{(r_N)}$$

$$\mathrm{O}(R_1 \cdot R_2 \cdot R_3 \cdot I_1 \cdot I_2 \cdot I_3)$$

(2)

$$\widetilde{\mathcal{A}} = \sum_{r} \mathbf{U}^{(1)}_{(r)} \otimes \sum_{s} \mathbf{U}^{(2)}_{(s)} \otimes \sum_{t} \mathcal{B}[r, s, t] \cdot \mathbf{U}^{(3)}_{(t)}$$

$$\mathrm{O}(R_1 \cdot I_1 \cdot I_2 \cdot I_3)$$

University of Zurich[UZH]

VISUALIZATION AND MULTIMEDIA LAB