BERKELEY LAB
Bringing Science Solutions to the World

U.S. DEPARTMENT OF ENERGY
Office of Science

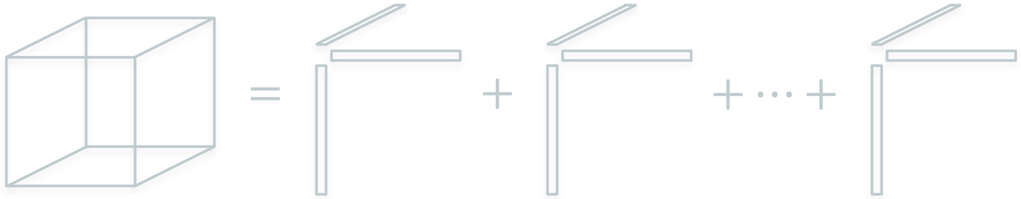
More Efficient Sampling for Tensor Decomposition With Worst-Case Guarantees

Osman Asif Malik
Applied Mathematics & Computational Research Division
Berkeley Lab

ICML 2022 · Baltimore, MD

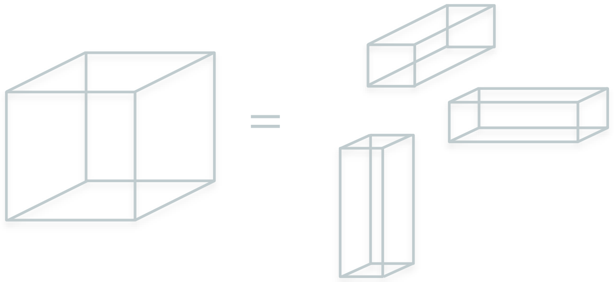
Tensors and Their Decompositions

- In this talk: Tensors = multidimensional arrays
- d-way tensor = array with d indices: $X(i_1, i_2, \ldots, i_d)$
- Decomposition breaks tensors into smaller pieces
  - CP decomposition:



$$X(i_1, i_2, i_3) = \sum_{r=1}^{R} A(i_1, r) B(i_2, r) C(i_3, r)$$

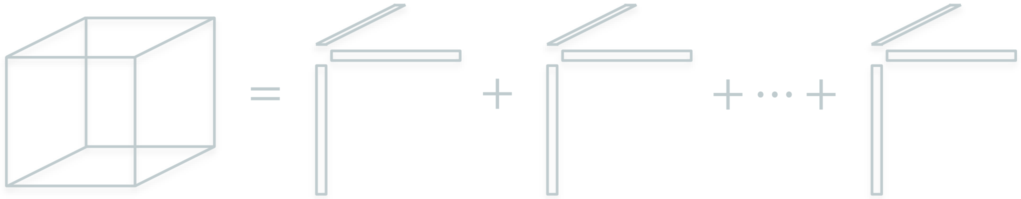
  - Tensor ring decomposition:



$$X(i_1, i_2, i_3) = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} G^{(1)}(r_1, i_1, r_2) G^{(2)}(r_2, i_2, r_3) G^{(3)}(r_3, i_3, r_1)$$

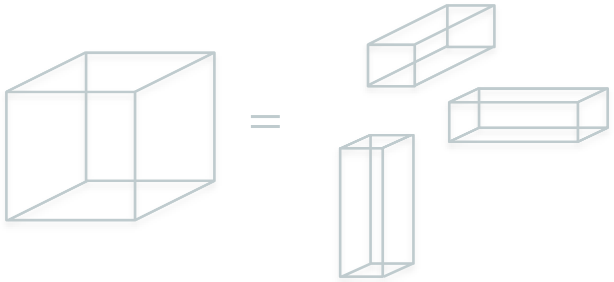
Tensors and Their Decompositions

- In this talk: Tensors = multidimensional arrays
- d-way tensor = array with d indices: $X(i_1, i_2, \dots, i_d)$
- Decomposition breaks tensors into smaller pieces
  - CP decomposition:



$$X(i_1, i_2, i_3) = \sum_{r=1}^{R} A(i_1, r) B(i_2, r) C(i_3, r)$$

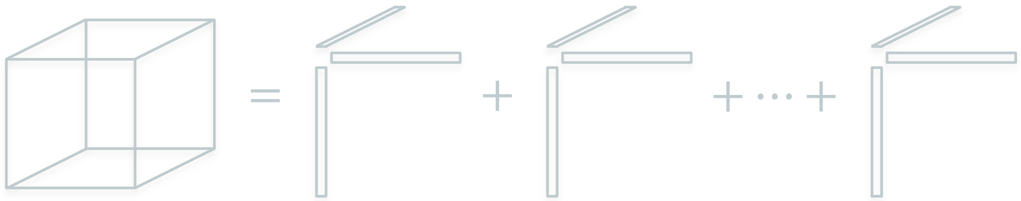
  - Tensor ring decomposition:



$$X(i_1, i_2, i_3) = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} G^{(1)}(r_1, i_1, r_2) G^{(2)}(r_2, i_2, r_3) G^{(3)}(r_3, i_3, r_1)$$

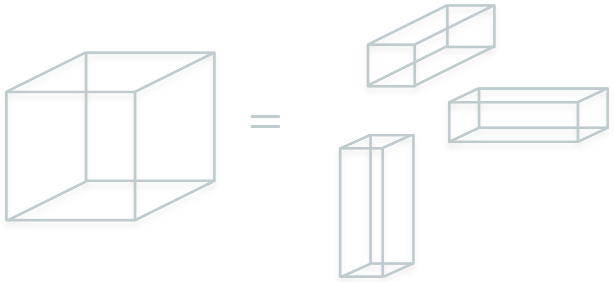
Tensors and Their Decompositions

- In this talk: Tensors = multidimensional arrays
- d-way tensor = array with d indices: $X(i_1, i_2, \dots, i_d)$
- Decomposition breaks tensors into smaller pieces
  - CP decomposition:



$$X(i_1, i_2, i_3) = \sum_{r=1}^{R} A(i_1, r) B(i_2, r) C(i_3, r)$$

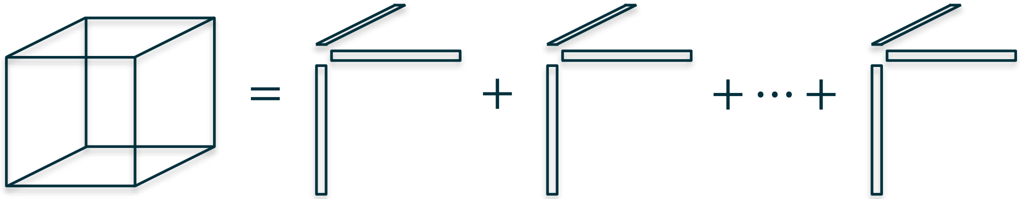
  - Tensor ring decomposition:



$$X(i_1, i_2, i_3) = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} G^{(1)}(r_1, i_1, r_2) G^{(2)}(r_2, i_2, r_3) G^{(3)}(r_3, i_3, r_1)$$

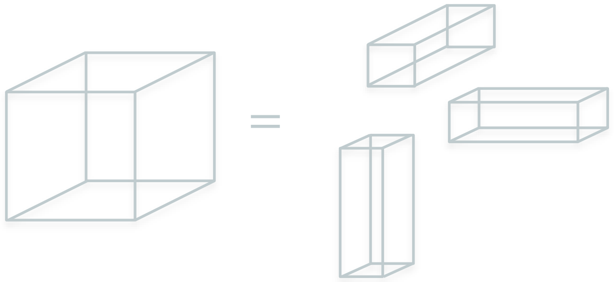
Tensors and Their Decompositions

- In this talk: Tensors = multidimensional arrays
- d-way tensor = array with d indices: $X(i_1, i_2, \dots, i_d)$
- Decomposition breaks tensors into smaller pieces
  - CP decomposition:



$$X(i_1, i_2, i_3) = \sum_{r=1}^{R} A(i_1, r) B(i_2, r) C(i_3, r)$$

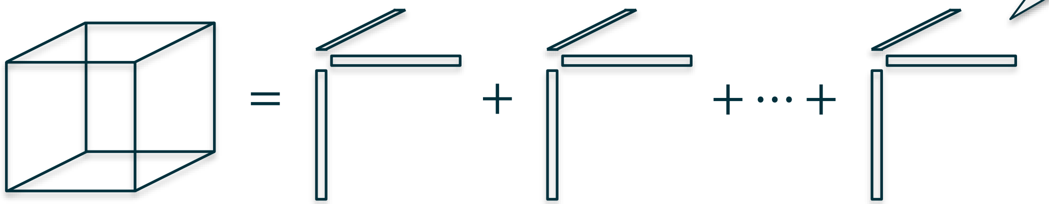
  - Tensor ring decomposition:



$$X(i_1, i_2, i_3) = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} G^{(1)}(r_1, i_1, r_2) G^{(2)}(r_2, i_2, r_3) G^{(3)}(r_3, i_3, r_1)$$

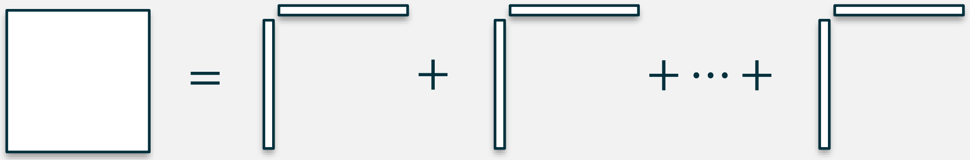
Tensors and Their Decompositions

- In this talk: Tensors = multidimensional arrays
- d-way tensor = array with d indices: $X(i_1, i_2, \dots, i_d)$
- Decomposition breaks tensors into smaller pieces
  - CP decomposition:

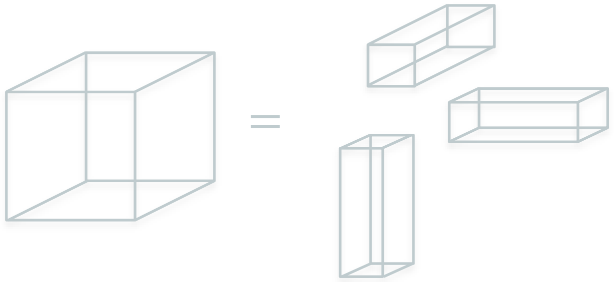


$$X(i_1, i_2, i_3) = \sum_{r=1}^{R} A(i_1, r) B(i_2, r) C(i_3, r)$$

Similar to the matrix SVD:



  - Tensor ring decomposition:

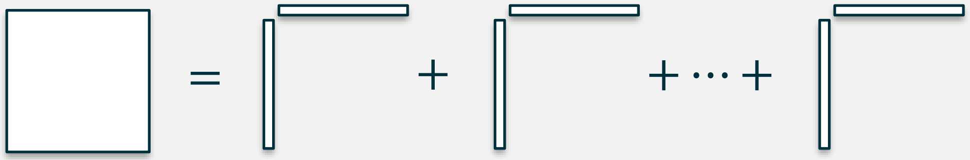


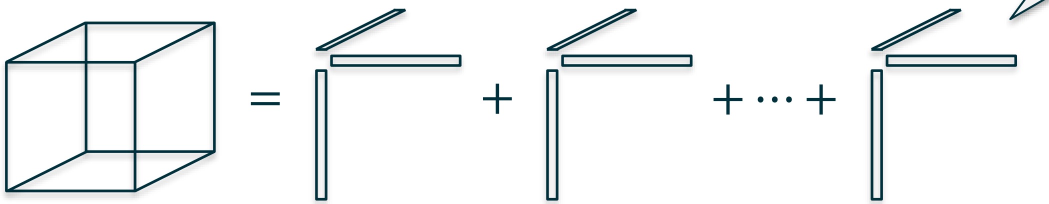
$$X(i_1, i_2, i_3) = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} G^{(1)}(r_1, i_1, r_2) G^{(2)}(r_2, i_2, r_3) G^{(3)}(r_3, i_3, r_1)$$

Tensors and Their Decompositions

- In this talk: Tensors = multidimensional arrays
- d-way tensor = array with d indices: $X(i_1, i_2, \dots, i_d)$
- Decomposition breaks tensors into smaller pieces
  - CP decomposition:

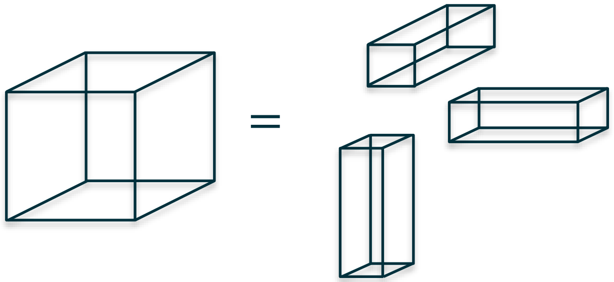
Similar to the matrix SVD:





$$X(i_1, i_2, i_3) = \sum_{r=1}^{R} A(i_1, r) B(i_2, r) C(i_3, r)$$

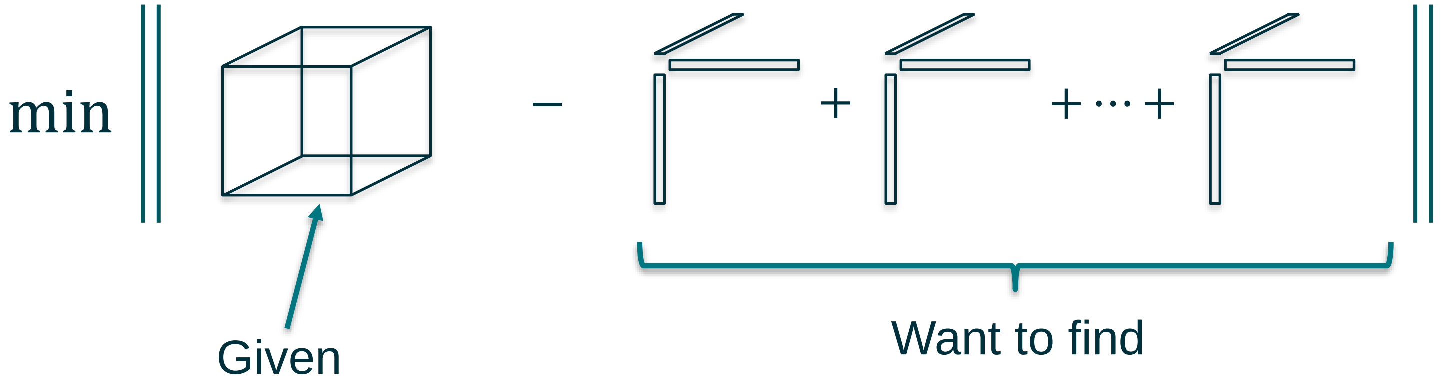
  - Tensor ring decomposition:



$$X(i_1, i_2, i_3) = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} G^{(1)}(r_1, i_1, r_2) G^{(2)}(r_2, i_2, r_3) G^{(3)}(r_3, i_3, r_1)$$

CP Decomposition Can Be Formulated as an Optimization Problem

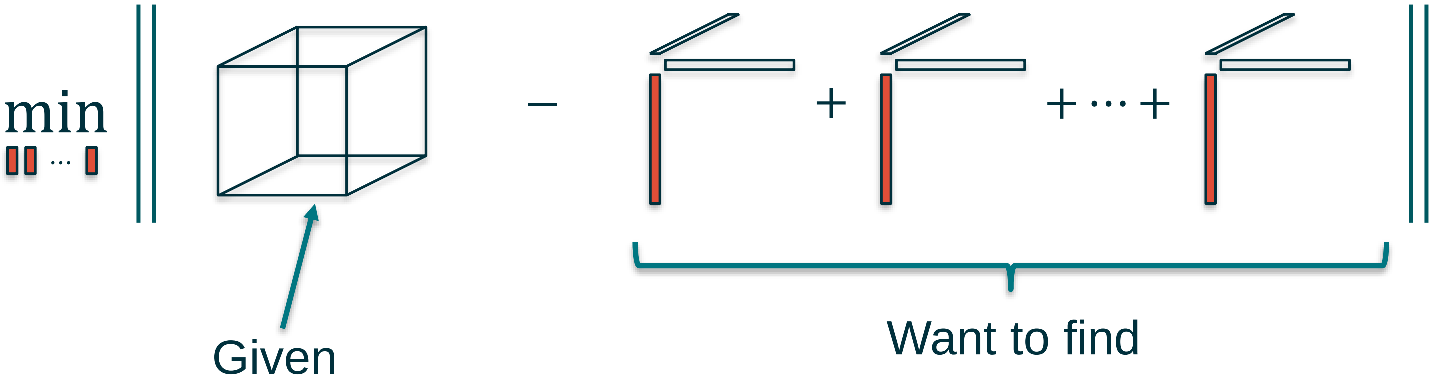
- Formulate as optimization problem:



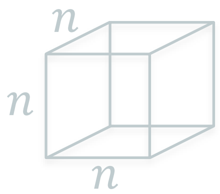
- Difficult non-convex optimization problem!

Alternating Minimization Is a Popular Approach for Computing CP Decomposition

- Solve via alternating minimization:

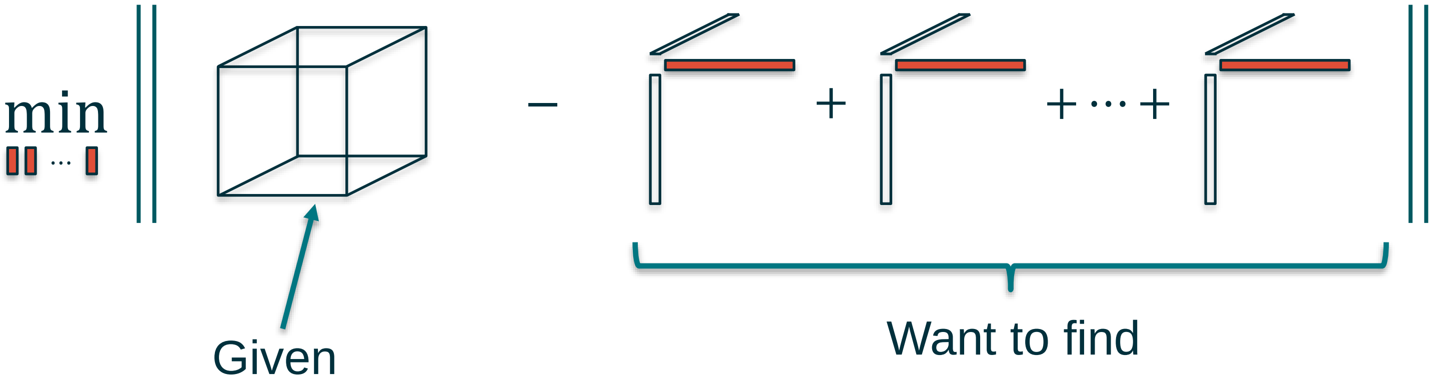


- Repeat!
- Curse of dimensionality: Cost of each step scales exponentially with number of modes
  - 3-way $n \times n \times n$ tensor: Cost $\gtrsim n^3$
  - d-way $n \times \cdots \times n$ tensor: Cost $\gtrsim n^d$

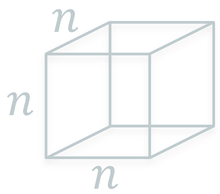


Alternating Minimization Is a Popular Approach for Computing CP Decomposition

- Solve via alternating minimization:

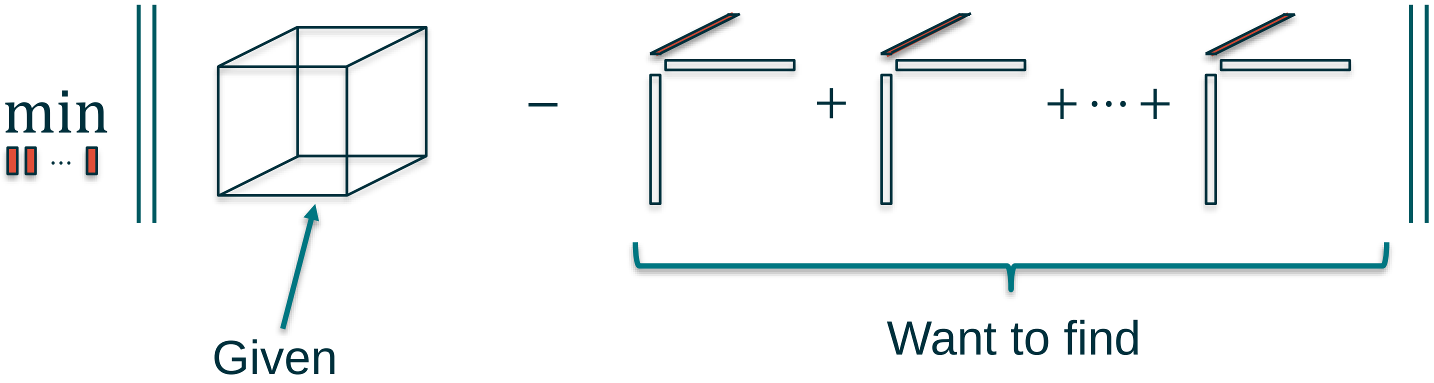


- Repeat!
- Curse of dimensionality: Cost of each step scales exponentially with number of modes
  - 3-way $n \times n \times n$ tensor: Cost $\gtrsim n^3$
  - d-way $n \times \cdots \times n$ tensor: Cost $\gtrsim n^d$

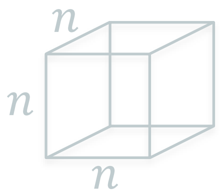


Alternating Minimization Is a Popular Approach for Computing CP Decomposition

- Solve via alternating minimization:

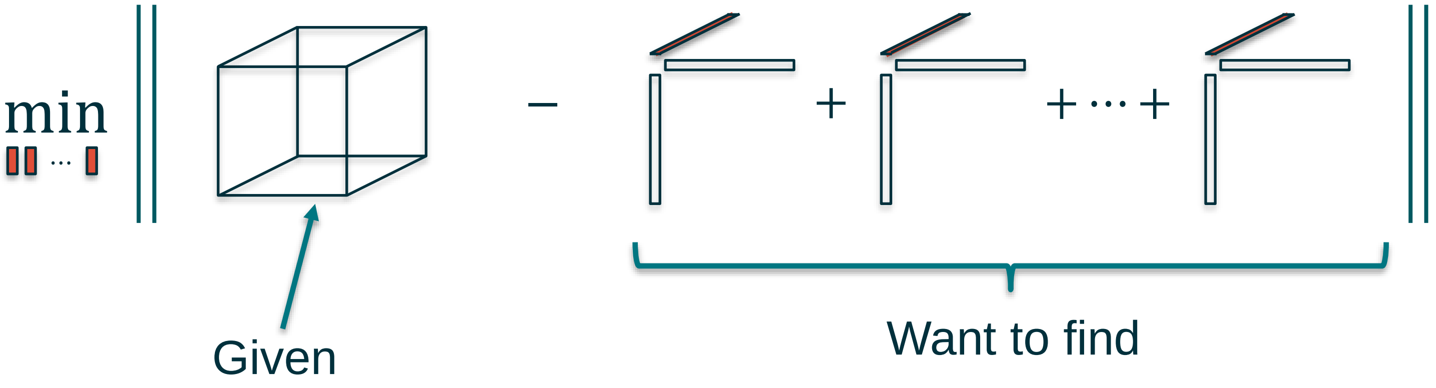


- Repeat!
- Curse of dimensionality: Cost of each step scales exponentially with number of modes
  - 3-way $n \times n \times n$ tensor: Cost $\gtrsim n^3$
  - d-way $n \times \cdots \times n$ tensor: Cost $\gtrsim n^d$

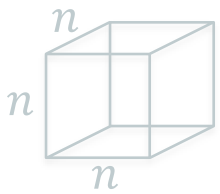


Alternating Minimization Is a Popular Approach for Computing CP Decomposition

- Solve via alternating minimization:

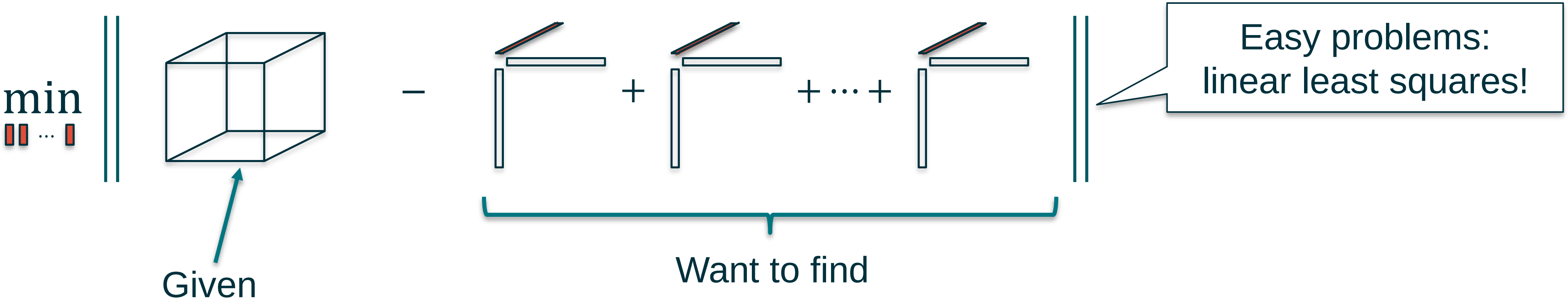


- Repeat!
- Curse of dimensionality: Cost of each step scales exponentially with number of modes
  - 3-way $n \times n \times n$ tensor: Cost $\gtrsim n^3$
  - d-way $n \times \cdots \times n$ tensor: Cost $\gtrsim n^d$

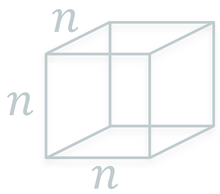


Alternating Minimization Is a Popular Approach for Computing CP Decomposition

- Solve via alternating minimization:

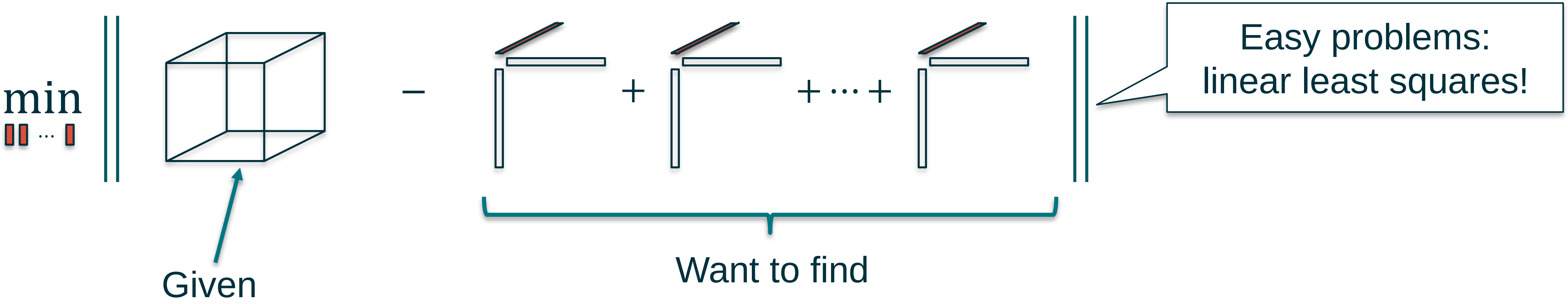


- Repeat!
- Curse of dimensionality: Cost of each step scales exponentially with number of modes
  - 3-way $n \times n \times n$ tensor: Cost $\gtrsim n^3$
  - d-way $n \times \cdots \times n$ tensor: Cost $\gtrsim n^d$

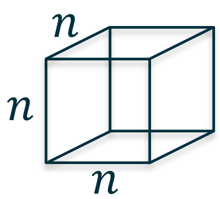


Alternating Minimization Is a Popular Approach for Computing CP Decomposition

- Solve via alternating minimization:

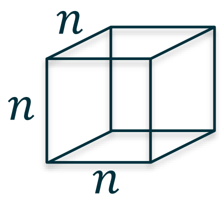


- Repeat!
- Curse of dimensionality: Cost of each step scales exponentially with number of modes
  - 3-way $n \times n \times n$ tensor: Cost $\gtrsim n^3$
  - d-way $n \times \cdots \times n$ tensor: Cost $\gtrsim n^d$



Sampling-Based Techniques Can Yield Input Sublinear Per-Iteration Cost

- Approach: Sample the optimization problems
- We improve on previous efforts:



Method	Complexity*	d dependence
CP-ALS	$\#\text{it} \cdot d(d+n)n^{d-1}R$	Exponential
SPALS [CP+'16]	$\#\text{it} \cdot d(d+n)R^{d+1}$	Exponential
CP-ARLS-LEV [LK'20]	$\#\text{it} \cdot d(R+n)R^{d}$	Exponential
CP-ALS-ES – our proposal	$\#\text{it} \cdot d^2R^3(R+dn)$	Polynomial

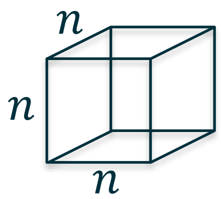
[CP+'16] Cheng, Peng, Liu, Perros. NeurIPS, 2016.
[LK'20] Larsen, Kolda. arXiv:2006.16438v3, 2020.
* Leading order complexity required for per-iteration relative-error guarantees. Ignores log factors and assumes fixed accuracy and failure probability.

Notation:
d: Number of modes/indices
n: Dimension
R: CP rank
$\#\text{it}$: Number of iterations

R can exceed n in tensor decomposition

Sampling-Based Techniques Can Yield Input Sublinear Per-Iteration Cost

- Approach: Sample the optimization problems
- We improve on previous efforts:



Method	Complexity*	d dependence
CP-ALS	$\#\mathrm{it} \cdot d(d+n)n^{d-1}R$	Exponential
SPALS [CP+'16]	$\#\mathrm{it} \cdot d(d+n)R^{d+1}$	Exponential
CP-ARLS-LEV [LK'20]	$\#\mathrm{it} \cdot d(R+n)R^{d}$	Exponential
CP-ALS-ES – our proposal	$\#\mathrm{it} \cdot d^2R^3(R+dn)$	Polynomial

[CP+'16] Cheng, Peng, Liu, Perros. NeurIPS, 2016.
[LK'20] Larsen, Kolda. arXiv:2006.16438v3, 2020.
* Leading order complexity required for per-iteration relative-error guarantees. Ignores log factors and assumes fixed accuracy and failure probability.

Notation:
d: Number of modes/indices
n: Dimension
R: CP rank
$\#\mathrm{it}$: Number of iterations

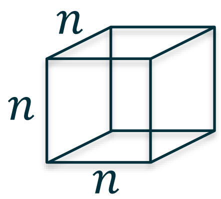
R can exceed n in tensor decomposition

Sampling-Based Techniques Can Yield Input Sublinear Per-Iteration Cost

- Approach: Sample the optimization problems
- We improve on previous efforts:

Method	Complexity*	d dependence
CP-ALS	$\#\mathrm{it} \cdot d(d+n)n^{d-1}R$	Exponential
SPALS [CP+'16]	$\#\mathrm{it} \cdot d(d+n)R^{d+1}$	Exponential
CP-ARLS-LEV [LK'20]	$\#\mathrm{it} \cdot d(R+n)R^{d}$	Exponential
CP-ALS-ES – our proposal	$\#\mathrm{it} \cdot d^2R^3(R+dn)$	Polynomial

[CP+'16] Cheng, Peng, Liu, Perros. NeurIPS, 2016.
[LK'20] Larsen, Kolda. arXiv:2006.16438v3, 2020.
* Leading order complexity required for per-iteration relative-error guarantees. Ignores log factors and assumes fixed accuracy and failure probability.



Notation:
d: Number of modes/indices
n: Dimension
R: CP rank
$\#\mathrm{it}$: Number of iterations

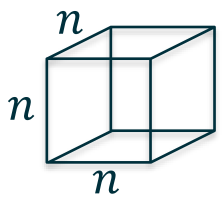
R can exceed n in tensor decomposition

Sampling-Based Techniques Can Yield Input Sublinear Per-Iteration Cost

- Approach: Sample the optimization problems
- We improve on previous efforts:

Method	Complexity*	d dependence
CP-ALS	$\#\text{it} \cdot d(d+n)n^{d-1}R$	Exponential
SPALS [CP+'16]	$\#\text{it} \cdot d(d+n)R^{d+1}$	Exponential
CP-ARLS-LEV [LK'20]	$\#\text{it} \cdot d(R+n)R^{d}$	Exponential
CP-ALS-ES – our proposal	$\#\text{it} \cdot d^2R^3(R+dn)$	Polynomial

[CP+'16] Cheng, Peng, Liu, Perros. NeurIPS, 2016.
[LK'20] Larsen, Kolda. arXiv:2006.16438v3, 2020.
* Leading order complexity required for per-iteration relative-error guarantees. Ignores log factors and assumes fixed accuracy and failure probability.



Notation:
d: Number of modes/indices
n: Dimension
R: CP rank
$\#\text{it}$: Number of iterations

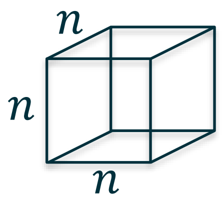
R can exceed n in tensor decomposition

Sampling-Based Techniques Can Yield Input Sublinear Per-Iteration Cost

- Approach: Sample the optimization problems
- We improve on previous efforts:

Method	Complexity*	d dependence
CP-ALS	$\#\text{it} \cdot d(d+n)n^{d-1}R$	Exponential
SPALS [CP+'16]	$\#\text{it} \cdot d(d+n)R^{d+1}$	Exponential
CP-ARLS-LEV [LK'20]	$\#\text{it} \cdot d(R+n)R^{d}$	Exponential
CP-ALS-ES – our proposal	$\#\text{it} \cdot d^2R^3(R+dn)$	Polynomial

[CP+'16] Cheng, Peng, Liu, Perros. NeurIPS, 2016.
[LK'20] Larsen, Kolda. arXiv:2006.16438v3, 2020.
* Leading order complexity required for per-iteration relative-error guarantees. Ignores log factors and assumes fixed accuracy and failure probability.



Notation:
d: Number of modes/indices
n: Dimension
R: CP rank
$\#\text{it}$: Number of iterations

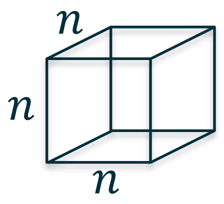
R can exceed n in tensor decomposition

Sampling-Based Techniques Can Yield Input Sublinear Per-Iteration Cost

- Approach: Sample the optimization problems
- We improve on previous efforts:

Method	Complexity*	d dependence
CP-ALS	$\#\mathrm{it} \cdot d(d+n)n^{d-1}R$	Exponential
SPALS [CP+'16]	$\#\mathrm{it} \cdot d(d+n)R^{d+1}$	Exponential
CP-ARLS-LEV [LK'20]	$\#\mathrm{it} \cdot d(R+n)R^{d}$	Exponential
CP-ALS-ES – our proposal	$\#\mathrm{it} \cdot d^2R^3(R+dn)$	Polynomial

[CP+'16] Cheng, Peng, Liu, Perros. NeurIPS, 2016.
[LK'20] Larsen, Kolda. arXiv:2006.16438v3, 2020.
* Leading order complexity required for per-iteration relative-error guarantees. Ignores log factors and assumes fixed accuracy and failure probability.

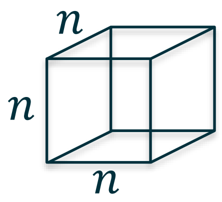


Notation:
d: Number of modes/indices
n: Dimension
R: CP rank
$\#\mathrm{it}$: Number of iterations

R can exceed n in tensor decomposition

Sampling-Based Techniques Can Yield Input Sublinear Per-Iteration Cost

- Approach: Sample the optimization problems
- We improve on previous efforts:



Method	Complexity*	d dependence
CP-ALS	$\#\text{it} \cdot d(d+n)n^{d-1}R$	Exponential
SPALS [CP+'16]	$\#\text{it} \cdot d(d+n)R^{d+1}$	Exponential
CP-ARLS-LEV [LK'20]	$\#\text{it} \cdot d(R+n)R^{d}$	Exponential
CP-ALS-ES – our proposal	$\#\text{it} \cdot d^2R^3(R+dn)$	Polynomial

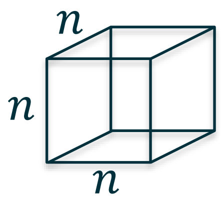
[CP+'16] Cheng, Peng, Liu, Perros. NeurIPS, 2016.
[LK'20] Larsen, Kolda. arXiv:2006.16438v3, 2020.
* Leading order complexity required for per-iteration relative-error guarantees. Ignores log factors and assumes fixed accuracy and failure probability.

Notation:
d: Number of modes/indices
n: Dimension
R: CP rank
$\#\text{it}$: Number of iterations

R can exceed n in tensor decomposition

Sampling-Based Techniques Can Yield Input Sublinear Per-Iteration Cost

- Approach: Sample the optimization problems
- We improve on previous efforts:



Method	Complexity*	d dependence
CP-ALS	$\#\mathrm{it} \cdot d(d+n)n^{d-1}R$	Exponential
SPALS [CP+'16]	$\#\mathrm{it} \cdot d(d+n)R^{d+1}$	Exponential
CP-ARLS-LEV [LK'20]	$\#\mathrm{it} \cdot d(R+n)R^{d}$	Exponential
CP-ALS-ES – our proposal	$\#\mathrm{it} \cdot d^2R^3(R+dn)$	Polynomial

[CP+'16] Cheng, Peng, Liu, Perros. NeurIPS, 2016.
[LK'20] Larsen, Kolda. arXiv:2006.16438v3, 2020.
* Leading order complexity required for per-iteration relative-error guarantees. Ignores log factors and assumes fixed accuracy and failure probability.

Notation:
d: Number of modes/indices
n: Dimension
R: CP rank
$\#\mathrm{it}$: Number of iterations

R can exceed n in tensor decomposition

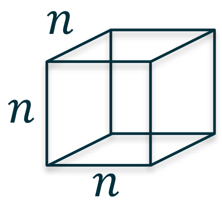
We Achieve Similar Improvements for the Tensor Ring Decomposition

- We improve on previous efforts:

Method	Complexity*	d dependence
TR-ALS [ZZ+'16]	$\#\mathrm{it} \cdot dn^{d} R^2$	Exponential
rTR-ALS [YL+'19]	$dn^{d} K + \#\mathrm{it} \cdot dK^{d} R^2$	Exponential
TR-SVD [ZZ+'16] [MK'20]	$n^{d+1} + n^{d} R^3$	Exponential
TR-SVD-Rand [AAC+'20]	$n^{d} R^2$	Exponential
TR-ALS-Sampled [MB'21]	$\#\mathrm{it} \cdot dnR^{2d+2}$	Exponential
TR-ALS-ES – our proposal	$\#\mathrm{it} \cdot d^3 R^8 (R + n)$	Polynomial

[ZZ+'16] Zhao, Zhou, Xie, Zhang, Cichocki. arXiv:1606.05535, 2016.
[YL+'19] Yuan, Li, Cao, Zhao. ICASSP, 2019.
[MK'20] Mickelin, Karaman. Numer Linear Algebra Appl 27(3):e2289, 2020.
[AAC+'20] Ahmadi-Asl, Cichocki, Phan, Asante-Mensah, Mousavi, Oseledets, Tanaka. Mach learn: sci technol, 2020.
[MB'21] Malik, Becker. ICML, 2021.

* Leading order complexity required for per-iteration relative-error guarantees. Ignores log factors and assumes fixed accuracy and failure probability.



Notation:
d: Number of modes/indices
n: Dimension
R: Tensor ring rank
$\#\mathrm{it}$: Number of iterations

R can exceed n in tensor decomposition

Poster #614, Hall E