when Randomized Algs meet TNs

Beheshteh T. Rakhshan

Agenda:

- Motivation
- why Tensor Decompositions ?
- why Randomization?
  - Tensor-Train (TT)
    - TT_SVD
    - Randomized TT_SVD
    - TT_ALS
    - Randomized TT-ALS

Motivation

- Randomized Algs and Tensor Network methods

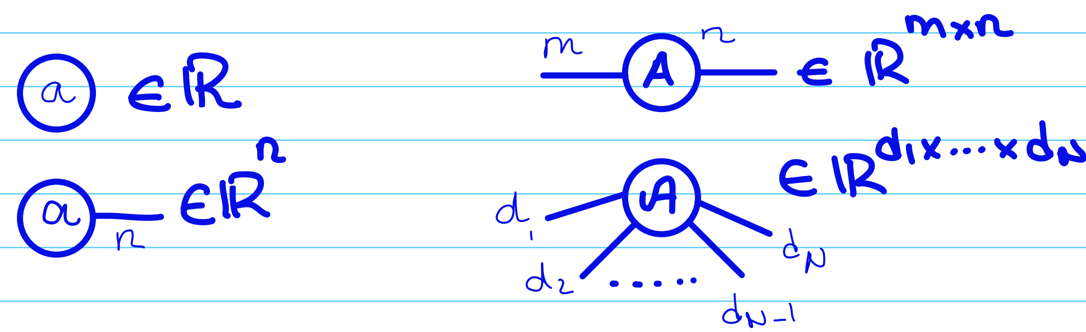
  two powerful tools

  - TN ⇝ parameterize linear models in exponentially large space (e.g. stoudenmire et al)
  - Randomized ⇝ speed up classical methods e.g., SVD

⚠ Classical Randomized Algs suffer Curse of dimensionality

Tensors and decompositions

_Notations

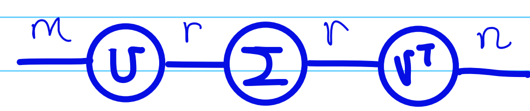


_Decompositions

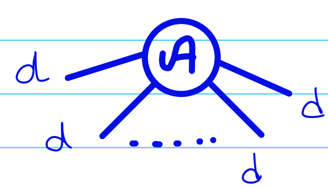
↳ Tensors are huge
  ↳ Decompose them into smaller objs
    ↳ Compressed representation
      ↳ Different ways doing it !

2D example:

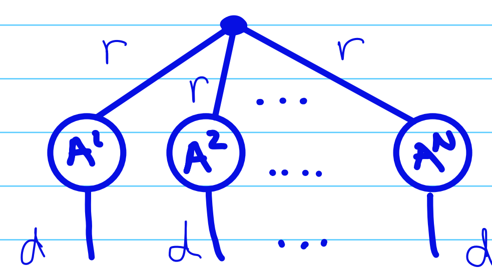
SVD ⇝ $A \in \mathbb{R}^{m\times n}$ =



CP Decomposition



~

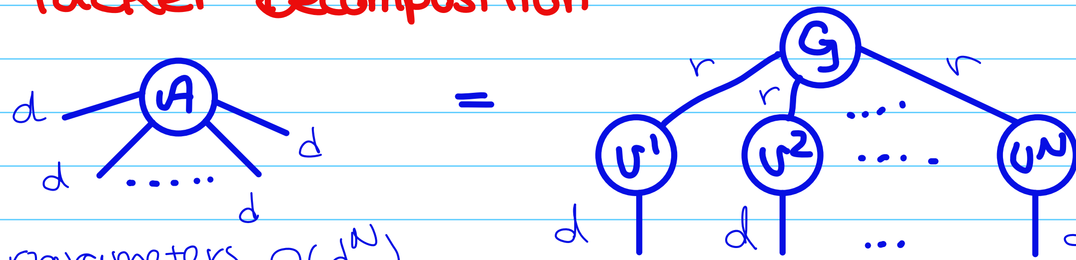


parameters $O(d^N)$

parameters $O(Ndr)$

⚠ Finding rank-r decomp. is NP-hard !

Tucker Decomposition

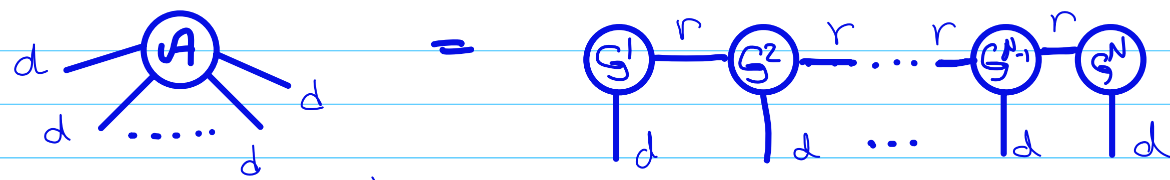


parameters $O(d^N)$

parameters $O(r^N + Nrd)$

⚠ **Exponential in N!**

Tensor Train (TT) Decomposition



parameters $O(d^N)$

parameters $O(Ndr^2)$

Randomized Methods

- speed up + provide accurate approximate solutions

Linear Least square $Ax \simeq b$, $A \in \mathbb{R}^{m \times d}$ $m \gg d$

↳ Construct S by

1) Sampling methods
2) Random projection

such that $SAx \simeq Sb$, $S \in \mathbb{R}^{s \times m}$

obtain $\tilde{x} \simeq x$

* Sampling e.g. Sample rows via uniform
Sample rows via leverage scores

$$\|A\tilde{x} - b\|_2^2 \leq (1+\epsilon)\|Ax_{opt} - b\| \quad \text{w.h.p}$$

* Random projection

$$f: \mathbb{R}^m \to \mathbb{R}^s \quad s \ll m \to \forall x \in \mathbb{R}^m \quad f(x) = \frac{1}{\sqrt{s}} \underbrace{R}_{s\times m} x$$

Define $S = \frac{1}{\sqrt{s}} R \longmapsto S$ here is not data dependent

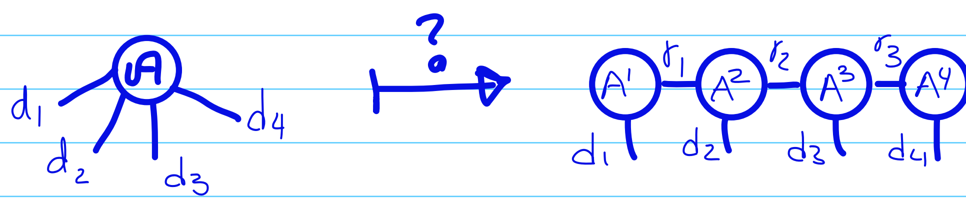
$R \overset{iid}{\sim} N(0,1)$ and dense

then $(1-\epsilon)\|AX\| \leq \|SAX\| \leq (1+\epsilon)\|AX\|$ w.h.p

---

How to Compute TT decomposition?

1) TT-SVD
2) Randomized TT-SVD
3) TT-ALS
4) Randomized TT-ALS



I) Matricization: Reshaping $\mathcal{A}$ such that resulting tensor is a 2D tensor

e.g., $\mathcal{A}_{(1)} \in \mathbb{R}^{d_1 \times d_2 d_3 d_4}$ $\qquad$ $\mathcal{A}_{(1,2)} \in \mathbb{R}^{d_1 d_2 \times d_3 d_4}$

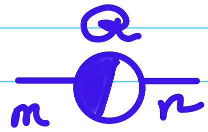
$\mathcal{A}_{(2)} \in \mathbb{R}^{d_2 \times d_1 d_3 d_4}$ $\qquad$ $\mathcal{A}_{(2,3)} \in \mathbb{R}^{d_2 d_3 \times d_1 d_4}$

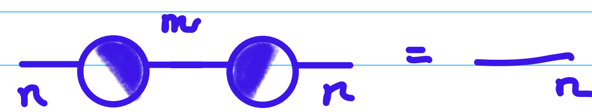
II) Left and right Orthogonal :

$Q \in \mathbb{R}^{m \times n}$ is left orthogonal $\iff Q^TQ = I_n$

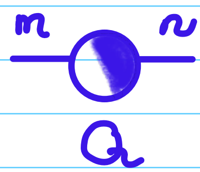
$Q \in \mathbb{R}^{m \times n}$ is right orthogonal $\iff QQ^T = I_m$

Orthogonal Tensors

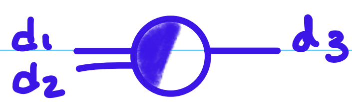


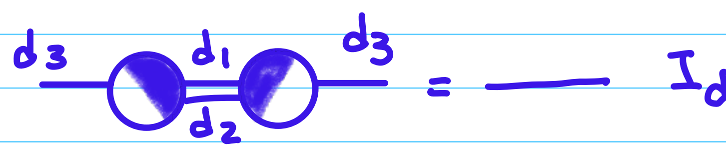


$$Q^TQ = I_n$$

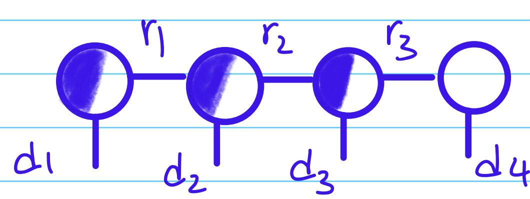


$$QQ^T = I_m$$

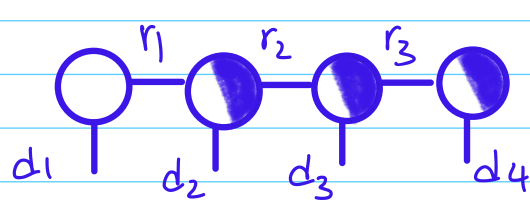




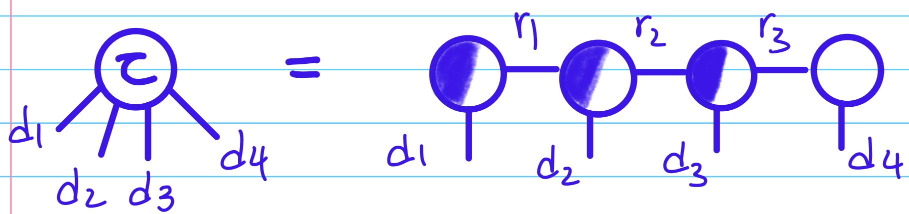
Left orthogonal form of a TT



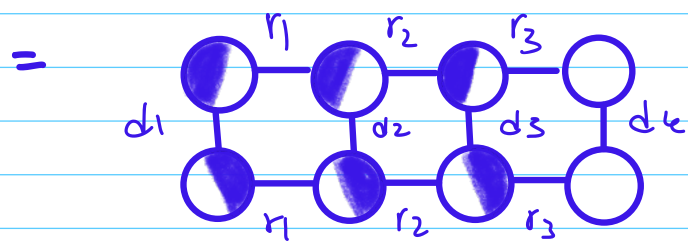
Right orthogonal form of a TT

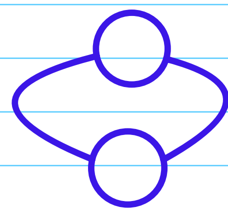


Frobenius norm of a Orthogonal TT



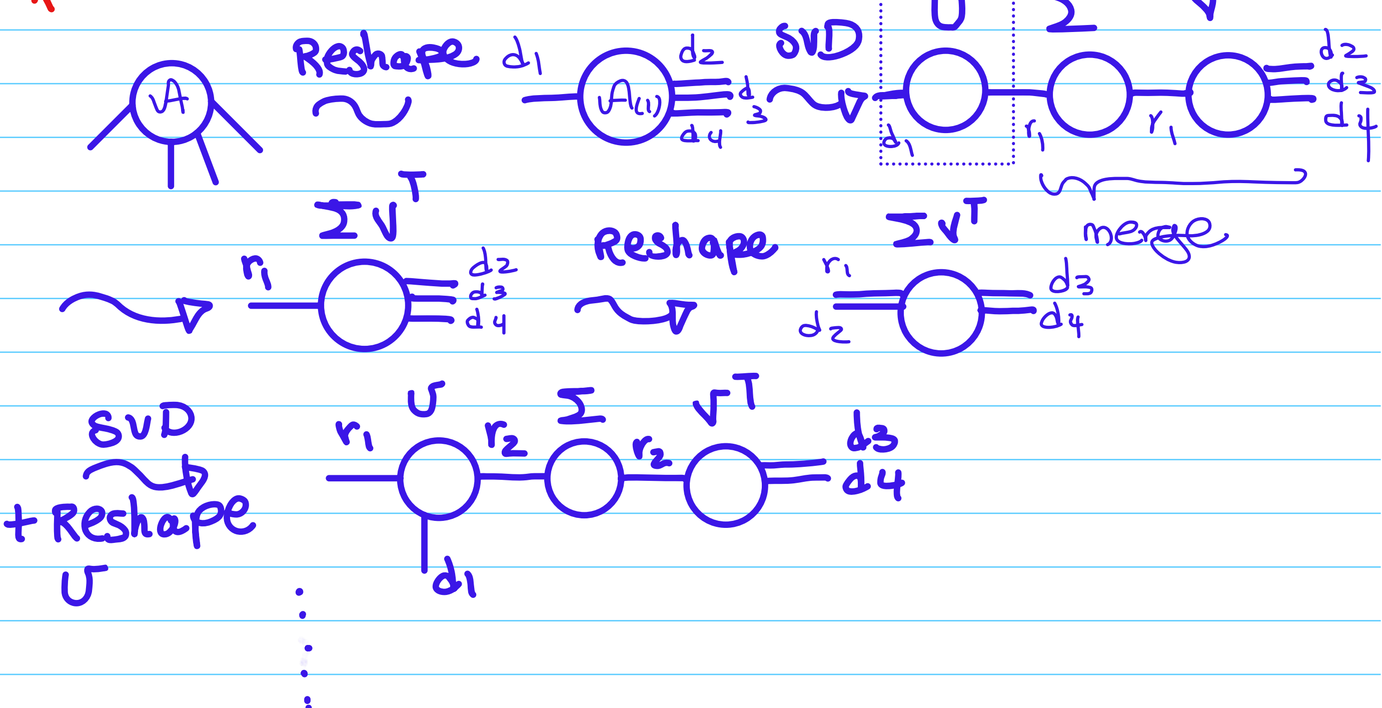
$$\|\tau\|_F^2 = Tr(\tau^T \tau)$$

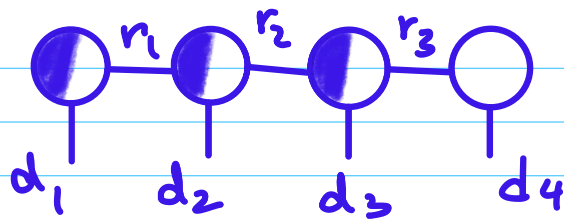




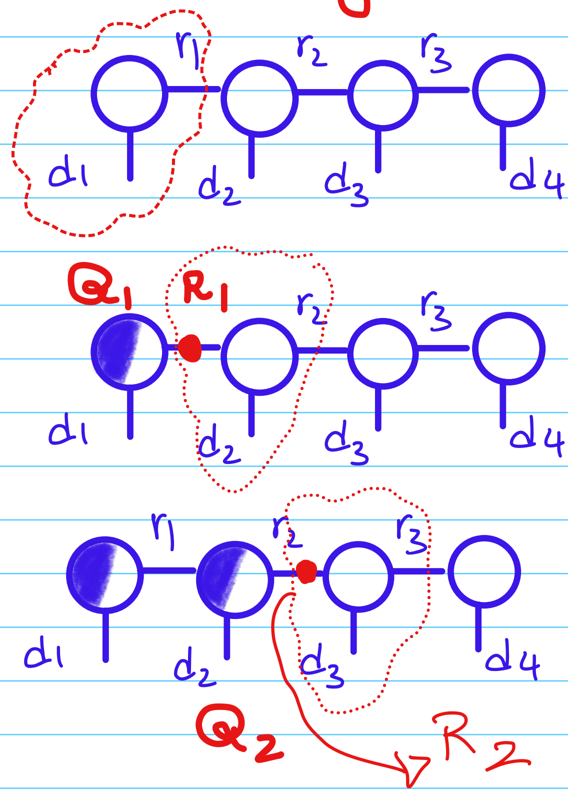
$= \|A^4\|_F^2$

* TT_SVD

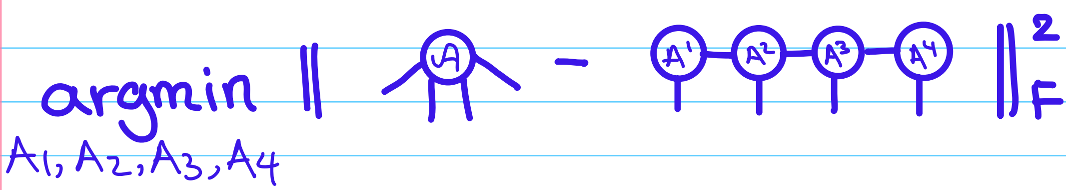




Efficient way to covert a TT

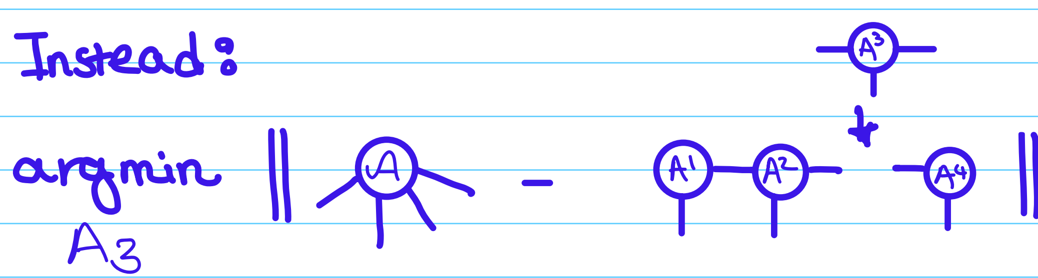


TT_ALS (Alternating Least Square)



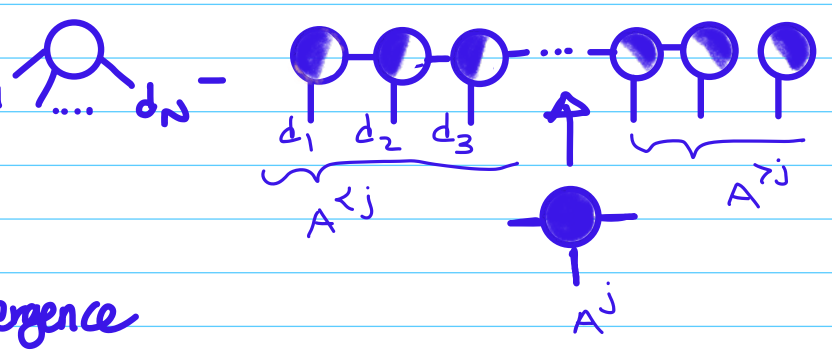
↳ Non_Convex ⚠

Instead:



How?

Initialize : $A^1, \dots, A^N$ randomly (crude guess)

$\underset{A^j}{\operatorname{argmin}} \|$ 

$\|_F^2$

until convergence

$$\underset{A^j}{\operatorname{argmin}} \| (A^{<j} \otimes A^{>j^T}) A^j - A^T_{(j)} \|_F^2$$

cost $O(d^N)$ to solve LS

Randomized TT_ALS

$$\underset{X}{\operatorname{argmin}} \| AX - b \|_F^2 \xrightarrow{\text{instead}} \underset{X}{\operatorname{argmin}} \| SAX - Sb \|_F^2$$

$$\underset{A^j}{\operatorname{argmin}} \| S(A^{<j} \otimes A^{>j^T}) A^j - SA_{(j)} \|_F^2$$

Construct S

- $$p_i \propto \overbrace{A[i,:]\underbrace{(A^TA)^+}_{\text{orthogonal}} A[:,i]^T}^{\text{Leverage scores}}$$

$\rightsquigarrow$ $p_i \propto A[i,:]A[:,i]^T$ (squared norm of rows)

Let $L = A^{<j}$

Define
$$q = \frac{1}{R_j}\left(L[:,1]^2 + \dots + L[:,R_j]^2\right)$$

1) Sample a column index uniformly $\hat{t} = t$

2) Sample a row index from $L[:,t]^2$

3) Define

$$h^{>k} := \Big(\underbrace{A^{k+1}[:,s_{k+1},:] \dots A^{j-2}[:,s_{j-2},:]\, A^{j-1}[:,s_{j-1},:]}_{\text{sample in reverse from } A^{j-1} \text{ upto } A^{k+1}}\Big)^T$$

Lemma:

$$\mathbb{P}(\hat{s}_k = s_k \mid \hat{s}^{>k} = s^{>k}, \hat{t} = t) \propto \|A^k[:,s_k,:] \cdot h^{>k}\|_2^2$$

$\rightsquigarrow$ $\exists$ a data structure; sample efficiently

$\hookrightarrow$ Vivek Bharadwaj

Cost of one sweep of Rand_TT_ALS

$$O\left(\frac{N}{\varepsilon_0 \delta}\left(NR^4 \log I + IR^4\right)\right)$$

such that

$$\|A\tilde{X} - B\|_F \leq (1+\varepsilon_0)\|AX - B\|_F \quad \text{where}$$

$$\tilde{X} = \underset{\tilde{X}}{\operatorname{argmin}} \|SA\tilde{X} - SB\|_F \qquad \text{w.h.p}$$