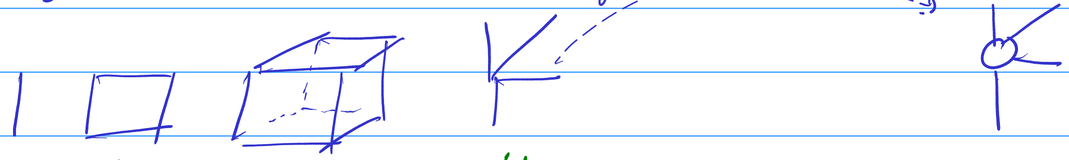
# Tensors

draw via tensor diagrams

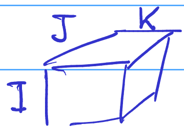


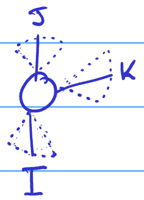
attributes

Order: 1 2 3 4

Dimensions: $\mathcal{T} \in \mathbb{R}^{I \otimes J \otimes K}$

Size: $IJK$



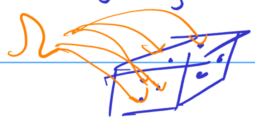



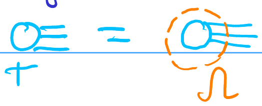

properties

Equidimensional: $I = J = K$

Symmetric: $t_{ijk} = t_{jik} = t_{ikj} = t_{jki} = t_{kij} = t_{kji}$ (partially-symmetric $t_{ijk} = t_{jik}$)

Sparse: $t_{ijk} = 0$ if $(i,j,k) \notin \Omega$

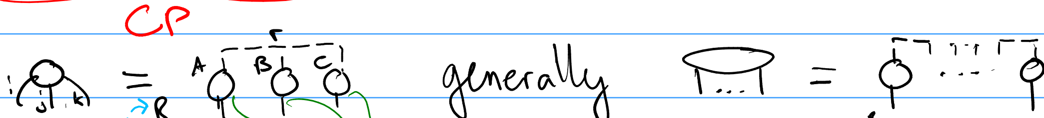





# Tensor Decompositions

## Canonical Polyadic (CANDECOMP/PARAFAC)

$\underbrace{\text{Canonical Polyadic}}_{\text{CP}}$




$t_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr}$

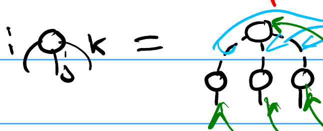
rank

factor matrices

generally

$t_{i_1 \cdots i_N} = \sum_{r=1}^{R} a^{(1)}_{i_1 r} \cdots a^{(N)}_{i_N r}$

## Tucker Decomposition

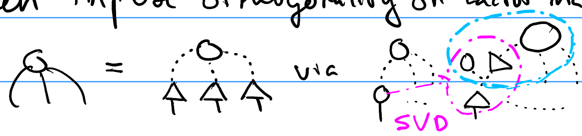


Tucker ranks

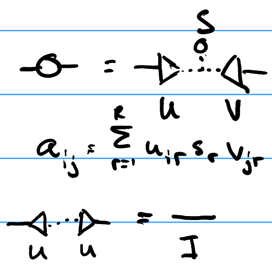
core tensor

factor matrices

can impose orthogonality on factor matrices

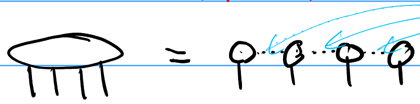



## Singular Value Decomposition (SVD)




$a_{ij} = \sum_{r=1}^{R} u_{ir} s_r v_{jr}$

## Tensor Train = Matrix Product State (MPS)

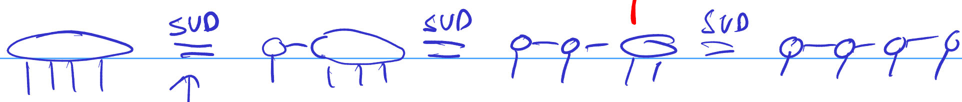


TT-ranks

$t_{ijkl} = \sum_{r_1 r_2 r_3} A^{(1)}_{i r_1} A^{(2)}_{r_1 j r_2} A^{(3)}_{r_2 k r_3} A^{(4)}_{r_3 l} = \langle a^{(1)}_{i}, A^{(2)}_{j}, A^{(3)}_{k}, a^{(4)}_{l} \rangle$

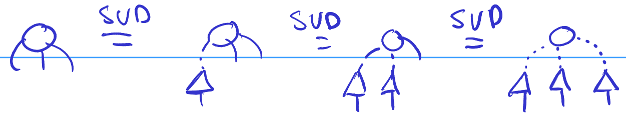
# Tensor Decomposition Algorithms

**Tensor Train** → Tensor Train (TT) decomposition




cost $O(s^N R)$

**Tucker** → High-order SVD (HoSVD)




exact if $\exists$ exact low-rank Tucker

near-optimal for approximation

QR + col. pivoting

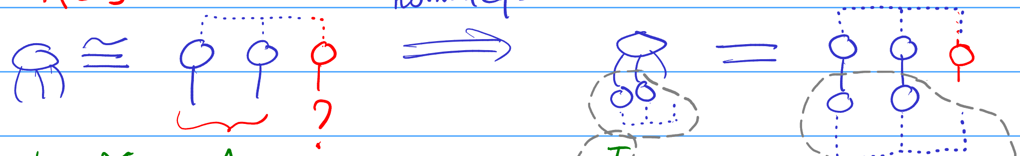
first SVD of size $s^{N-1}$-by-$s$ and rank $R$ $\Rightarrow$ $O(s^N R)$

$O(s^{N+1})$ naive full SVD

**CP decomposition** → Alternating Least Squares (ALS)

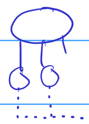
CP-ALS

normal eq's




$b \cong A \quad x \qquad\qquad A^T b = A^T A \quad x$

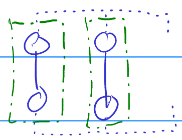
Matricized Tensor times Khatri-Rao Product (MTTKRP)



$a^{(N)}_{kr} = \sum_{ij} t_{ijk} a^{(1)}_{ir} a^{(2)}_{jr}$

$O(s^N R)$ cost

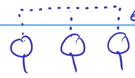
Matrix of equations relatively cheap to compute

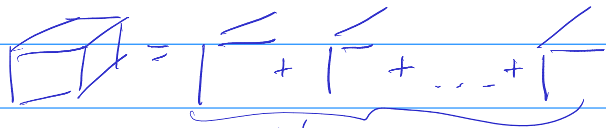


$O(sR^2)$

# Tensor rank and conditioning

Different decompositions provide different notions of rank

CP  CP rank



#times = CP rank

maximum rank: $R \leq s^{N-1}$

(for order $N=3$ tensor, factor matrices have $3sR$ degrees-of-freedom, maximum rank is $R \leq s^2$)

computing CP rank is NP hard

typical rank – for given dimensions ($s$), a rank that will occur in a randomly selected tensor with nonzero probability

for CP rank over $\mathbb{R}$, typical rank can be greater than 1

e.g. for $s=2$, $R=2$ 79% of time

$R=3$ 21% ($R=1$ 0% possible but)

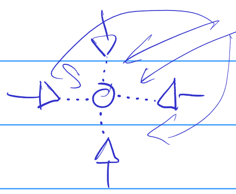
over $\mathbb{C}$, typical rank is unique

uniqueness of CP – modulo ordering of rank-1 components, scaling indeterminancy within each one, and assuming $A^{(1)} \dots A^{(N)}$ are full rank,

CP decomposition is unique if

$$R \leq \frac{sN}{2} - 1$$

## Tucker ranks



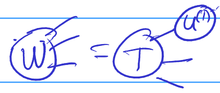
Tucker ranks

if $T \in \mathbb{R}^{s_1 \times \dots \times s_N}$, core tensor $Z \in \mathbb{R}^{R_1 \times \dots \times R_N}$ with $R_i \leq S_i$

HoSVD finds minimal ranks via SVD

Pf: we have that

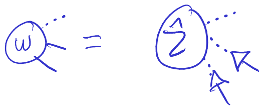
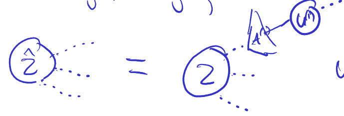
$$T_{i_1 \dots i_N} = \sum_{k_1 \dots k_N} Z_{k_1 \dots k_N} \prod_{j=1}^{N} A^{(j)}_{i_j k_j}$$

Performing SVD truncation over 1 mode yields 

projection

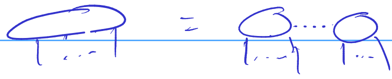
The projected $\hat{Z}$ has the desired remaining ranks,

$$W_{k_1 i_2 \dots i_N} = \sum_{k_2 \dots k_N} \hat{Z}_{k_1 \dots k_N} \prod_{j=2}^{N} A^{(j)}_{i_j k_j} \quad \text{where } \hat{Z}_{k_1 \dots k_N} = \sum_{\ell m} Z_{\ell k_2 \dots k_N} A^{(1)}_{\ell m} U^{(1)}_{m k_1}$$

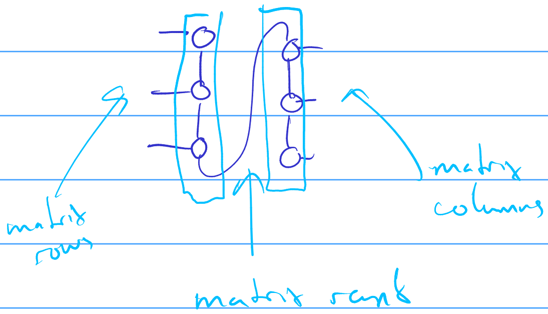
 where  w. 

# Tensor train ranks

TT ranks

SVD  can always be used to find min TT rank

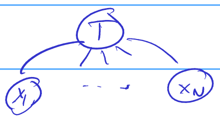
TT ranks behave like matrix ranks




# Tensor conditioning

For matrix $\kappa(A) = \frac{\underbrace{\max_{\|x\|_2=1} \|Ax\|_2}_{\sigma_{max}(A)}}{\underbrace{\min_{\|x\|_2=1} \|Ax\|_2}_{\sigma_{min}(A)}}$

Tensor spectral norm

$\|T\|_2 = \max_{\substack{x_1 \dots x_N \\ \|x_1\| = \dots = \|x_N\| = 1}} \Big|$  $\Big|$

for $N=2$, $\|T\|_2 = \max_{\substack{x, y \\ \|x\|=\|y\|=1}} |\langle x, Ay \rangle| = \sigma_{max}(A)$

$\|T\|_{max} \le \|T\|_2 \le \|T\|_F$

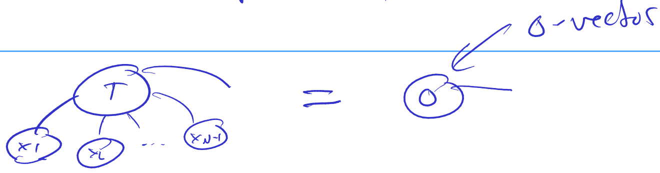
MTTKRP conditioning

$$\kappa(T, x_1, \dots x_N) = \max_{\delta x_1, \dots \delta x_{N-1}} \frac{\|MTTKRP(T, x_1 + \delta x_1, x_2 + \delta x_2, \dots, x_{N-1} + \delta x_{N-1}) - MTTKRP(T, x_1, \dots, x_{N-1})\|}{\|MTTKRP(T, x_1 \dots x_N)\|}$$

Thm: $\max_{\substack{x_1 \dots x_N \\ \|x_1\| = \dots = \|x_N\|}} \kappa(T, x_1, \dots x_N) = \infty$ for $s \notin \{1, 2, 4, 8\}$

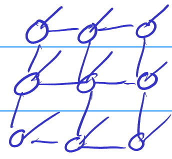
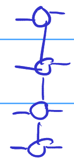
$\kappa \sim \frac{1}{MTTKRP(T, x_1, \dots x_{N-1})} \Rightarrow \min_{x_1 \dots x_{N-1}} \|MTTKRP(T, x_1, \dots x_{N-1})\|^2 = \sum_{j_N} \Big( \sum_{j_1 \dots j_{N-1}} T_{j_1 \dots j_N} x_{j_1} \cdots x_{j_{N-1}} \Big)^2$

By reduction from Hurwitz problem, if $s \in \{1, 2, 4, 8\}$, $\exists x_1, \dots x_{N-1}$ such that

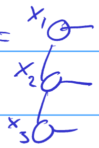
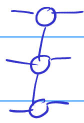
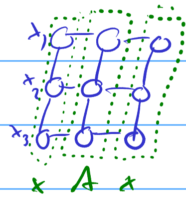


# Tensor Networks

MPS (TT) | PEPS (2D) | MPO (1D)

Example Application: Density Matrix Renormalization Group (DMRG)

Let $x =$  , $A =$  and $\rho_A(x) =$  $\sim x^T A x$

A represents system Hamiltonian:

$A = A^H$, for simplicity also assume $A$ is pos. def.

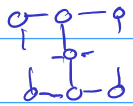
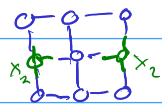
$\min \lambda(A) = \min_{\|x\|=1} \rho_A(x)$ is ground-state energy

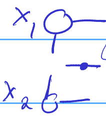
DMRG Algorithm

alternate among $i = 1 \dots N$

fix $x_j \ \forall j \neq i$, find $\underset{\|x_i\|=1}{\operatorname{argmin}} \rho_A(x)$

reduced problem $(i=2)$

$H =$  $\rho_H(x_2) =$ 

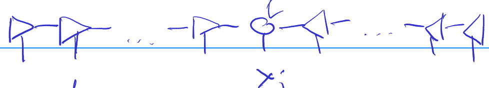
$= G^T A G$ where $G =$ 

identity

to ensure stability and avoid generalized eigenvalue problem,

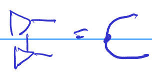
want orthogonal basis $Q$, e.g. $G = QR$, use

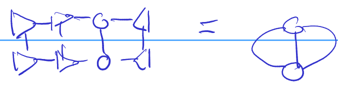
canonical form: center:

 (for MPS)

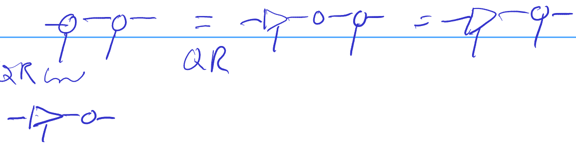
$x_i$

property:

 implies $\langle x, x \rangle = \langle x_i, x_i \rangle$

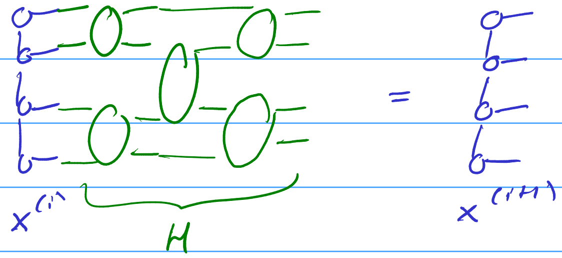


can canonicalize any MPS via QR

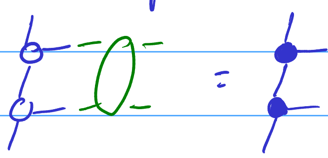


QR

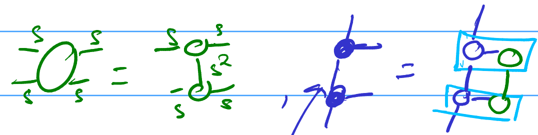
# Time Evolution Block Decimation (TEBD)




2-site operator

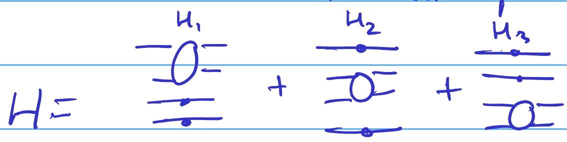


since

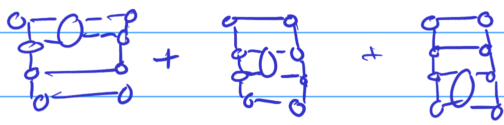



many quantum Hamiltonians can be written as sum of local operators, e.g.

rank increases by $s^2$
could truncate back to s via SVD




so $\langle x, Hx \rangle =$



## (Imaginary) Time Evolution

For time-independent Hamiltonians, state $\psi(t) = e^{-iHt}\psi(0)$
↑
time

Moreover $\tilde{\psi}(t) = e^{-Ht}\psi(0) \xrightarrow{t\to\infty}$ smallest eigenvector of $H$

since $\lambda_{max}(e^{-Ht}) = (e^{-\lambda_{min}(H)})^t$

$e^{-Ht} \neq e^{-H_1 t} e^{-H_2 t} e^{-H_3 t}$ unless $[H_i, H_j] = 0$ for all $i \neq j$

Trotterization

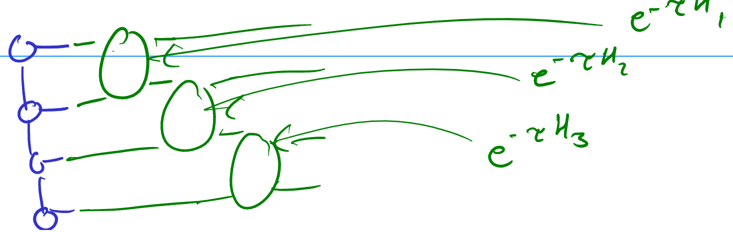
pick $\tau \ll 1$ then apply

$$e^{-Ht} = \prod_{i=1}^{i=t/\tau} e^{-H\tau}$$

where each

$$e^{-H\tau} = \underbrace{e^{-H_1\tau} e^{-H_2\tau}}\, e^{-H_3\tau} + O(\tau^2)$$

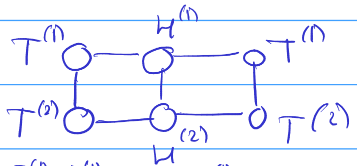
approximation accurate for small $\tau$



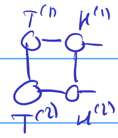
# Conditioning in Tensor Network Computations

## Tensor Network Optimization

$$\min_{\{T^{(1)},\dots,T^{(N)}\}} f(T^{(1)},\dots,T^{(N)})$$

e.g. $f(T^{(1)}, T^{(2)}) =$  (2-site DMRG)



or $f(T^{(1)}, T^{(2)}) = \Big\|$  $-$  $\Big\|_F^2$ (TT-structured linear system of equations)





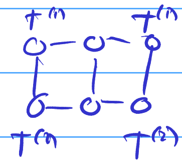
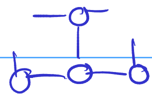
## Alternating Optimization Methods

can derive ALS for CP or Tucker, as well as DMRG

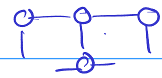
Let $g_i(T^{(i)}) = f(T^{(1)},\dots,T^{(i)},\dots,T^{(N)})$

Newton's method on $g_i$ gives

$$T^{(i)}_{new} = T^{(i)} - H_{g_i}^{-1}\,\mathrm{vec}(T^{(i)})$$

e.g. for $f(T^{(1)}, T^{(2)}) =$  , $H_{g_1} =$ 

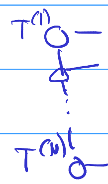


$H_{g_2} =$ 

## Perturbation Analysis

Consider a more general tensor-valued function

$$T^{(out)} = F(T^{(1)},\dots,T^{(N)})$$

e.g. $F^{MPS}(T^{(1)},\dots,T^{(N)}) =$ 



Let $\hat{f} = \mathrm{vec}(F)$

Given a perturbation $\delta T^{(i)}$ to site $i$, measure

$$\delta F = F(T^{(1)},\dots,T^{(i)}+\delta T^{(i)},\dots,T^{(N)}) - F(T^{(1)},\dots,T^{(N)})$$

$$\lim_{\frac{\|\delta T^{(i)}\|}{\|T^{(i)}\|}\to 0} \frac{\|\delta F\|/\|F\|}{\|\delta T^{(i)}\|/\|T^{(i)}\|} \le \underbrace{\|J_{\hat{f}}(T^{(i)})\|_2}_{\text{Jacobian w.r.t. } T_i} \Big/ \|J_{\hat{f}}(T^{(i)})\,\mathrm{vec}(T^{(i)})\|_2$$

e.g. for $J_{F^{MPS}}(T^{(1)}) =$  , 2-norm minimized by canonical form