### 計算科学における情報圧縮

Information Compression in Computational Science 2018.12.13

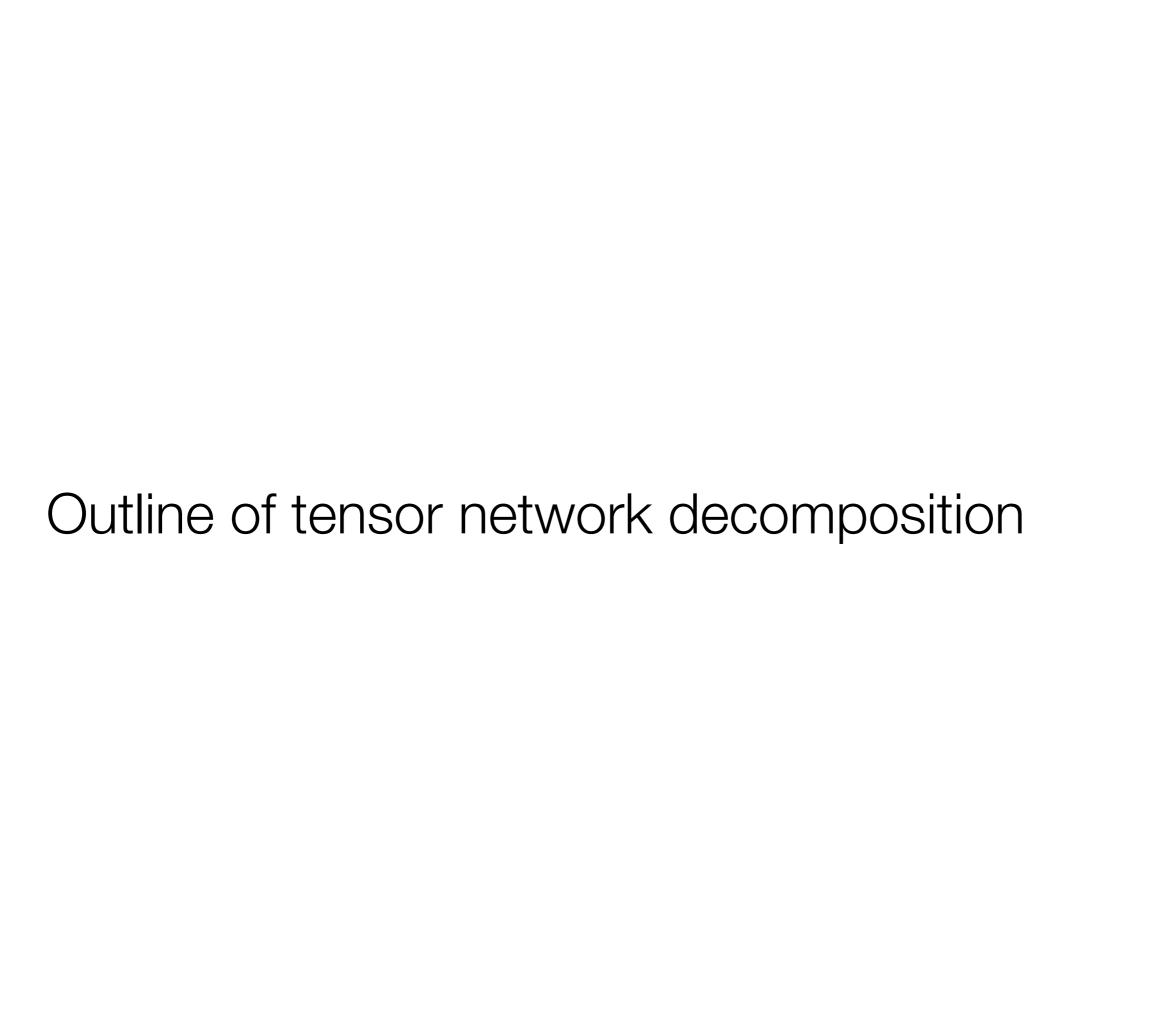
#10:高度なデータ圧縮:情報のエンタングルメントと行列積表現

Entanglement of information and matrix product state

理学系研究科 物理学専攻 大久保 毅 Department of Physics, **Tsuyoshi Okubo** 

### Outline

- Outline of tensor network decomposition
- Entanglement
  - Schmidt decomposition
  - Entanglement entropy and its area low
- Matrix product states
  - Matrix product states (MPS)
  - Canonical form
  - infinite MPS



# Classification of Information Compression by Memory Costs

Linear algebra for huge data:  $\vec{v} \in \mathbb{C}^M$ 

- (1) A matrix can be stored Required memory~  $O(M^2)$
- (2) Although a matrix cannot be stored, vectors can be stored Required memory  $\sim O(M)$
- (3) A vector cannot be stored

Required memory  $\ll O(M)$ 

We try to approximate a vector in a compact form.

$$M \sim a^N$$
 Memory ~  $O(N^x)$ 

**Exponential** 

**Polynomial** 

N:problem size (e.g. system size)

# When we efficiently compress a vector?

$$\vec{v} = \sum_{i=1}^{M} C_i \vec{e}_i \qquad \vec{v} \in \mathbb{C}^M$$

If we can find a basis where the coefficients have a structure (correlation).

(1) Almost all  $C_i$  are zero (or very small).



We store only a few finite elements  $\{(i,C_i)\}$ 

E.g. Fourier transformation 
$$\vec{v} = \sum_{k=1}^{M} D_k \vec{f}_k$$

If we can neglect larger wave numbers, we can efficiently approximate the vector with smaller number of coefficients.

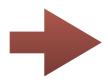
Classical state 
$$|\Psi\rangle=|01011\dots00\rangle$$

In this case, we know that only a specific  $C_i$  is non-zero. We need only an integer corresponding to the non-zero element.

# When we efficiently compress a vector?

$$\vec{v} = \sum_{i=1}^{M} C_i \vec{e}_i \qquad \vec{v} \in \mathbb{C}^M$$

(2) All of  $C_i$  are not necessarily independent.



We store "structure" and "independent elements".

$$\{(i,C_i)\}$$

E.g. Product state ("generalized" classical state)

A vector is decomposed into product of small vectors.

$$|\Psi
angle=|\phi_1
angle\otimes|\phi_2
angle\otimes\cdots$$
 e.g.  $|\phi_1
angle=lpha|0
angle+eta|1
angle$   $|\phi_1
angle=|01
angle-|10
angle$ 

structure: "product state"

independent elements: small vectors

# Tensor network decomposition of a vector

### Target:

Exponentially large Hilbert space

$$\vec{v} \in \mathbb{C}^M \quad \text{with} \ M \sim a^N$$

+

Total Hilbert space is decomposed as a product of "local" Hilbert space.

$$\mathbb{C}^M = \mathbb{C}^a \otimes \mathbb{C}^a \otimes \cdots \mathbb{C}^a$$



### Tensor network decomposition

$$v_i = v_{i_1, i_2, \dots, i_N} = \sum_{\{x\}} T^{(1)}[i_1]_{x_1, x_2, \dots} T^{(2)}[i_2]_{x_1, x_3, \dots} \cdots T^{(N)}[i_N]_{x_3, x_{100}, \dots}$$

 $i_n = 0, 1, \dots, a-1$ : index of local Hilbert space

 $T[i]_{x_1,x_2,...}$ : local tensor for "state" i

# Graphical representations for tensor network

Vector

$$ec{v}:v_i$$



Matrix

$$M$$
 :  $M_{i,j}$ 



Tensor

$$T:T_{i,j,k}$$

$$\frac{i}{k}$$

\* n-rank tensor = n-leg object

When indices are not presented in a graph, it represent a tensor itself.

$$\vec{v} =$$

$$T =$$

# Graphical representations for tensor network

### Matrix product

$$C_{i,j} = (AB)_{i,j} = \sum_{k} A_{i,k} B_{k,j}$$

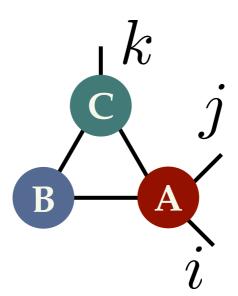
$$C = AB$$

$$\frac{i}{\mathbf{C}} = \frac{i}{\mathbf{A}} \frac{k}{\mathbf{B}} \frac{j}{j}$$

$$-C - = -A - B -$$

### Generalization to tensors

$$\sum_{\alpha,\beta,\gamma} A_{i,j,\alpha,\beta} B_{\beta,\gamma} C_{\gamma,k,\alpha}$$



Contraction of a network = Calculation of a lot of multiplications (縮約)

# Graph for a tensor network decomposition

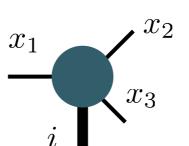
Vector

$$v_{i_1,i_2,i_3,i_4,i_5}$$

\*Vector looks like a tensor

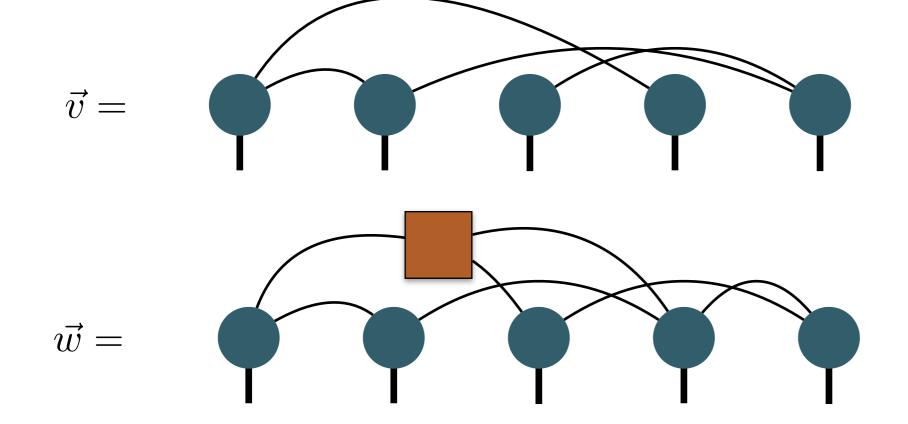
Tensor

$$T[i]_{x_1, x_2, x_3}$$



\*We treat *i* as an index of the tensor.

Tensor network decomposition



\*We can consider tensors independent on i.

Entanglement (エンタングルメント)

# N-qubit system (S=1/2 quantum spin system)

Example vector: Wave function of N-qubit systems

- - takes two states  $|0\rangle, |1\rangle \ (|\uparrow\rangle, |\downarrow\rangle)$

$$\begin{aligned} |\Psi\rangle &= \sum_{\{i_1,i_2,\dots i_N\}} \Psi_{i_1i_2\dots i_N} |i_1\rangle \otimes |i_2\rangle \otimes \dots \otimes |i_N\rangle \\ &= \sum_{\{i_1,i_2,\dots i_N\}} \Psi_{i_1i_2\dots i_N} |i_1i_2\dots i_N\rangle \end{aligned}$$

Coefficients = vector:  $\vec{\Psi} \in \mathbb{C}^{2^N}$ 

\* Inner product:  $\langle \Phi | \Psi \rangle = \vec{\Phi}^* \cdot \vec{\Psi}$ 

# Schmidt decomposition

General vector: 
$$\vec{x} \in V_1 \otimes V_2$$
 dim  $V_1 = n_1, \dim V_2 = n_2$ 

$$\dim \mathbf{V}_1 = n_1, \dim \mathbf{V}_2 = n_2$$
$$(n_1 \ge n_2)$$

### Schmidt decomposition

$$\vec{x} = \sum_{i=1}^{n_2} \lambda_i \vec{u}_i \otimes \vec{v}_i$$

Orthonormal vectors

$$\{\vec{u}_1, \vec{u}_2, \dots \vec{u}_{n_1}\} \in V_1$$
  
 $\{\vec{v}_1, \vec{v}_2, \dots \vec{v}_{n_2}\} \in V_2$ 

Schmidt coefficient  $\lambda_i > 0$ 

Schmidt decomposition is unique.

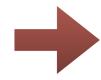
# Schmidt decomposition for wave function

Wave function: 
$$|\Psi\rangle = \sum_{\{i_1,i_2,...i_N\}} \Psi_{i_1i_2...i_N} |i_1i_2...i_N\rangle$$

### Schmidt decomposition

Divide system into two parts, A and B:



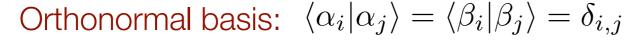


General wave function can be represented by a superposition of orthonormal basis set.

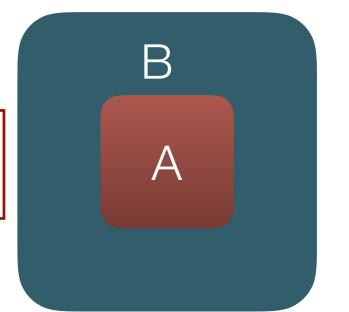
$$|\Psi\rangle = \sum_{i,j} M_{i,j} |A_i\rangle \otimes |B_j\rangle = \sum_i \lambda_i |\alpha_i\rangle \otimes |\beta_i\rangle$$

$$M_{i,j} \equiv \Psi_{(i_1,\dots),(\dots,i_N)} \quad |A_i\rangle = |i_1,i_2,\dots\rangle$$

$$A \quad B \quad |B_j\rangle = |\dots,i_{N-1},i_N\rangle$$



Schmidt coefficient:  $\lambda_i \geq 0$ 



# Relation between SVD and Schmidt decomposition

Singular value decomposition (SVD):

For a K × L matrix M,

$$M_{i,j} = \sum_{m} U_{i,m} \lambda_m V_{m,j}^{\dagger}$$

Singular values:  $\lambda_m \geq 0$ 

Singular vectors:  $egin{aligned} \sum_i U_{i,m} U_{m,j}^\dagger &= \delta_i, j \ \sum_i V_{i,m} V_{m,j}^\dagger &= \delta_i, j \end{aligned}$ 

Relation to the Schmidt decomposition:

$$|\Psi\rangle = \sum_{i,j} M_{i,j} |A_i\rangle \otimes |B_j\rangle = \sum_{m} \lambda_m |\alpha_m\rangle \otimes |\beta_m\rangle$$

$$|\alpha_m\rangle = \sum_{i} U_{i,m} |A_i\rangle$$

$$|\beta_m\rangle = \sum_{i} V_{m,j}^{\dagger} |B_j\rangle$$

$$\langle \alpha_i |\alpha_j\rangle = \langle \beta_i |\beta_j\rangle = \delta_{i,j}$$

By using SVD, we can perform Schmidt decomposition. (and can calculate entanglement entropy.)

# Partial trace and reduced density matrix

For 
$$\vec{x} \in V_1 \otimes V_2$$
 dim  $V_1 = n_1$ , dim  $V_2 = n_2$   $|\vec{x}| = 1$ 

Density matrix: 
$$\rho \equiv \vec{x}\vec{x}^{\dagger} \ (\rho_{ij} = x_i x_j^*)$$

(密度行列) 
$$(\rho = |x\rangle\langle x|)$$
 \*Note: rank  $\rho = 1$ 

Orthonormal basis:  $\{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_{n_1}\} \in V_1 \quad \{\vec{f}_1, \vec{f}_2, \dots, \vec{f}_{n_2}\} \in V_2$ 



Basis for  $\vec{x}$  :  $\vec{g}_{i_1,i_2} = \vec{e}_{i_1} \otimes \vec{f}_{i_2}$ 

 $V_2$ 

Index:  $i = (i_1, i_2)$ 

### **Reduced Density matrix:**

(縮約密度行列)

 $\rho_{V_1} \equiv \text{Tr}_{V_2} \ \rho$ : a positive-semidefinite square matrix in  $V_1$ 

$$(\rho_{\mathbf{V}_1})_{i_1,j_1} = \sum_{i_2} \rho_{(i_1,i_2),(j_1,i_2)} \qquad \qquad \mathbf{V}_1$$

# Entanglement entropy

### Entanglement entropy:

Reduced density matrix of a sub system (sub space):

Α

В

$$\rho_A = \text{Tr}_B |\Psi\rangle\langle\Psi|$$











Entanglement entropy = von Neumann entropy of  $\rho_A$ 

$$S = -\text{Tr}\left(\rho_A \log \rho_A\right)$$

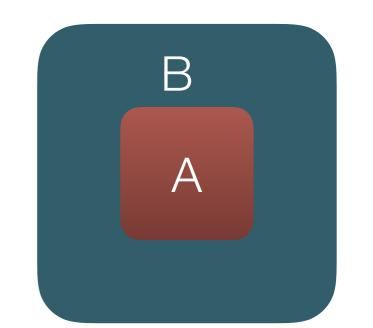
Schmidt decomposition  $|\Psi\rangle=\sum_{i}\lambda_{i}|\alpha_{i}\rangle\otimes|\beta_{i}\rangle$ 



$$\rho_A = \sum_i \lambda_i^2 |\alpha_i\rangle \langle \alpha_i| \qquad \text{(*Exercise)}$$



$$S = -\sum_{i} \lambda_i^2 \log \lambda_i^2$$



Entanglement entropy is calculated through the spectrum of Schmidt coefficients

# Intuition for EE: two s=1/2 spins

1. 
$$|\Psi\rangle = |\uparrow\rangle \otimes |\downarrow\rangle$$

A product state  $\lambda = 1$ , S = 0



$$\lambda = 1$$
,  $S = 0$ 

2. 
$$|\Psi\rangle = \frac{1}{2}(|\uparrow\rangle - |\downarrow\rangle) \otimes (|\uparrow\rangle - |\downarrow\rangle)$$

Product state: S=0

Another product state  $\lambda = 1$  , S = 0



$$\lambda = 1$$
 ,  $S = 0$ 

3. 
$$|\Psi\rangle = \frac{1}{\sqrt{2}} (|\uparrow\rangle \otimes |\downarrow\rangle - |\downarrow\rangle \otimes |\uparrow\rangle)$$



Spin singlet 
$$\lambda_1 = \lambda_2 = \frac{1}{\sqrt{2}} , S = \log 2$$

Maximally entangled State

4. 
$$|\Psi\rangle = \left(x|\uparrow\rangle\otimes|\downarrow\rangle + \sqrt{1-x^2}|\downarrow\rangle\otimes|\uparrow\rangle\right)$$



Complicated state 
$$\lambda_1 = |x|, \lambda_2 = \sqrt{1 - x^2}$$

 $S = x^2 \log x^2 + \sqrt{1 - x^2} \log(1 - x^2)$ 

Large entanglement entropy ~ Large correlation between two parts

# Area law of the entanglement entropy in physics

### General wave functions:

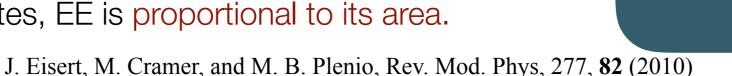
EE is proportional to its **volume** (# **of qubits**).

$$S = -\text{Tr}\left(\rho_A \log \rho_A\right) \propto L^d$$

(c.f. random vector)

Ground state wave functions:

For a lot of ground states, EE is proportional to its area.



$$S = -\text{Tr} \left( \rho_A \log \rho_A \right) \propto L^{d-1}$$

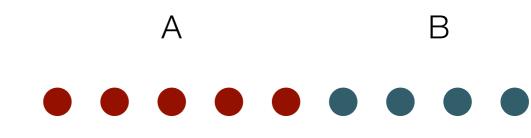
In the case of one-dimensional system:

Gapped ground state for local Hamiltonian

M.B. Hastings, J. Stat. Mech.: Theory Exp. P08024 (2007)

$$S = O(1)$$

Ground state are in a small part of the huge Hilbert space



# Exercise: examples of Schmidt decomposition

- 1-1: Random wave function (Sample code: Ex1-1.py or Ex1-1.ipynb)
  - Make a random vector
  - SVD it and see singular value spectrum and EE
- 1-2: Ground state of S=1 Heisenberg chain (Sample code: Ex1-2.py or Ex1-2.ipynb)

$$\mathcal{H} = \sum_{i} \vec{S}_{i} \cdot \vec{S}_{i+1}$$

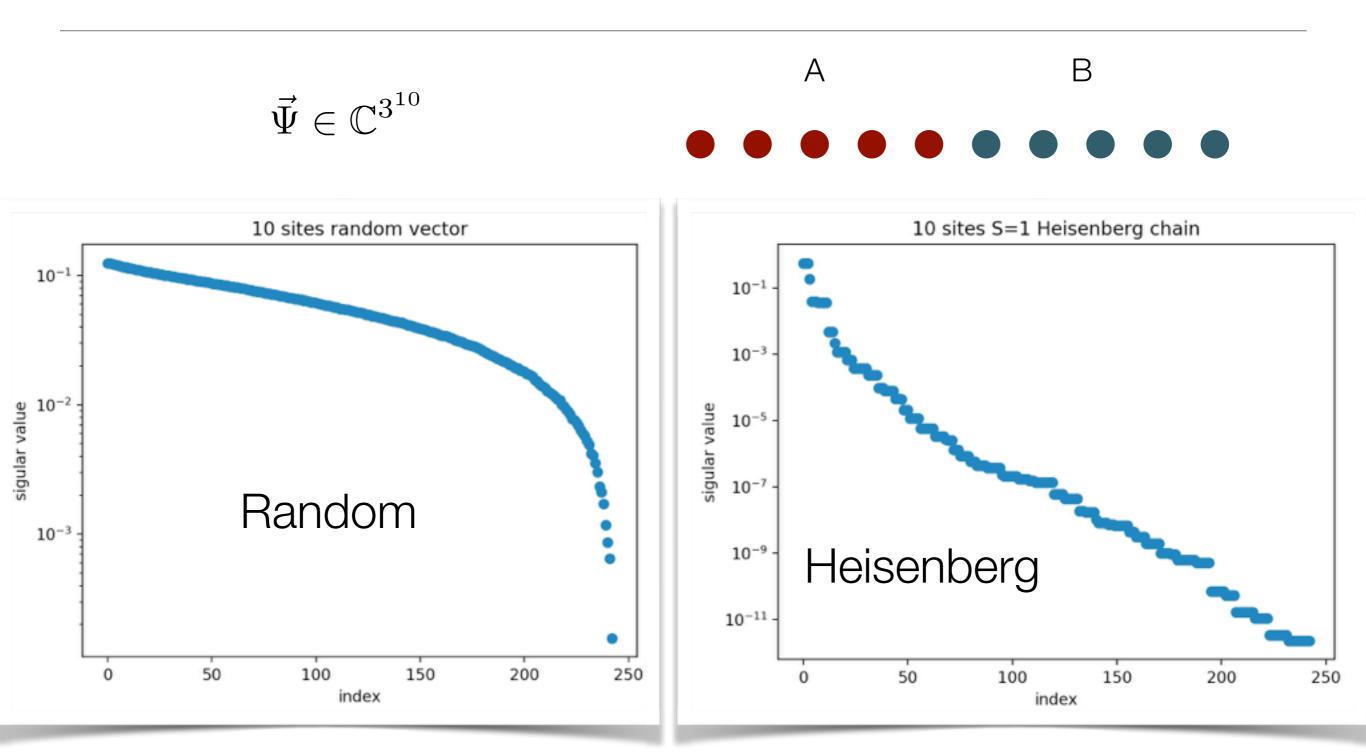
- Calculate GS by diagonalizing Hamiltonian
- SVD it and see singular value spectrum and EE

\*Note: the ground state of this model is gapped

\* Try to simulate different system size "N"

\* You can simulate other S by changing "m"

# Result: N=10 spectrum



Ground state wave function has lower entanglement!

Matrix product states(行列積状態)

### Data compression of wave functions (vectors)

General wave function:

$$|\Psi\rangle = \sum_{\{i_1, i_2, \dots i_N\}} \Psi_{i_1 i_2 \dots i_N} |i_1 i_2 \dots i_N\rangle$$

Coefficient vector can represent any points in the Hilbert space.



Ground states satisfy the area law.



In order to represent the ground state accurately, we might not need all of a<sup>N</sup> elements.



Data compression by tensor decomposition:

Tensor network states

Hilbert space



### Tensor network state

G.S. wave function: 
$$|\Psi\rangle=\sum_{\{i_1,i_2,...i_N\}}\Psi_{i_1i_2...i_N}|i_1i_2...i_N\rangle$$
 Vector (or N-rank tensor):  $\Psi_{i_1i_2...i_N}$  =  $\Psi_{i_1i_2...i_N}$  Tensor network" decomposition





Matrix Product State (MPS)

$$A_1[i_1]A_2[i_2]\cdots A_N[i_N] =$$

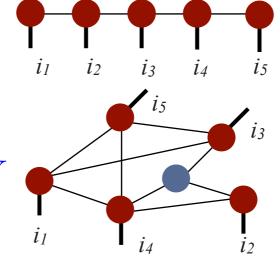
 $A[m]\,$  : Matrix for state m

General network

$$\mathrm{Tr} X_{1}[i_{1}] X_{2}[i_{2}] X_{3}[i_{3}] X_{4}[i_{4}] X_{5}[i_{5}] Y$$

X,Y: Tensors

Tr: Tensor network contraction



By choosing a "good" network, we can express G.S. wave function efficiently.

ex. MPS: # of elements  $=2ND^2$ 

D: dimension of the matrix A

Exponential → Linear

\*If D does not depend on N...

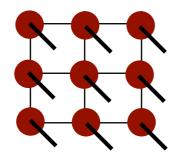
# Examples of TNS

MPS:



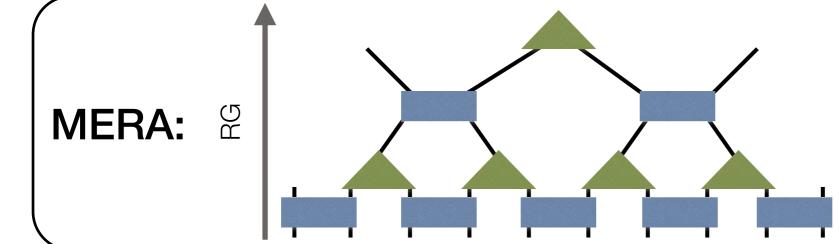
Good for 1-d gapped systems

PEPS, TPS:



For higher dimensional systems

Extension of MPS



Scale invariant systems

#### Good reviews:

### Matrix product state (MPS)

(U. Schollwöck, Annals. of Physics **326**, 96 (2011))

(R. Orús, Annals. of Physics **349**, 117 (2014))

$$|\Psi\rangle = \sum_{\{i_1,i_2,\ldots i_N\}} \Psi_{i_1i_2\ldots i_N} |i_1i_2\ldots i_N\rangle$$
 
$$MPS$$
 
$$\Psi_{i_1i_2\ldots i_N} \simeq A_1[i_1]A_2[i_2]\cdots A_N[i_N]$$
 
$$A[i]: \text{Matrix for state } i$$

### Note:

- MPS is called as "tensor train decomposition" in applied mathematics
   (I. V. Oseledets, SIAM J. Sci. Comput. 33, 2295 (2011))
- A product state is represented by MPS with 1×1 "Matrix" (scalar)

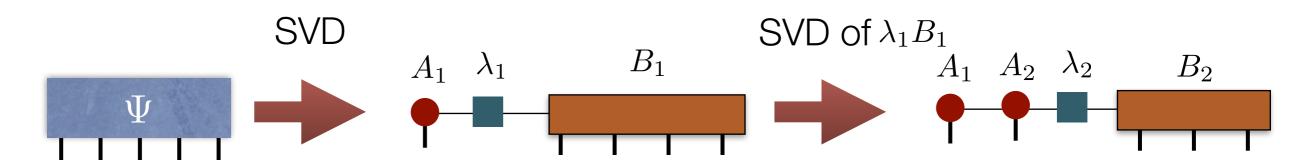
$$|\Psi\rangle = |\phi_1\rangle \otimes |\phi_2\rangle \otimes \cdots$$

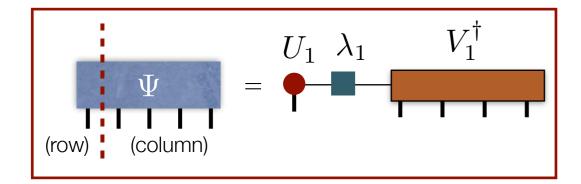
$$\Psi_{i_1 i_2 \dots i_N} = \phi_1[i_1]\phi_2[i_2] \cdots \phi_N[i_N]$$

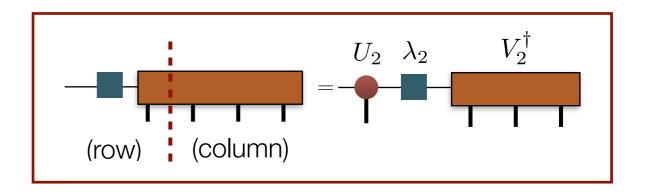
$$\phi_n[i] \equiv \langle i|\phi_i\rangle$$

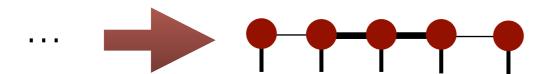
### Matrix product state without approximation

General wave function (or vector) can be represented by MPS exactly through successive Schmidt decompositions







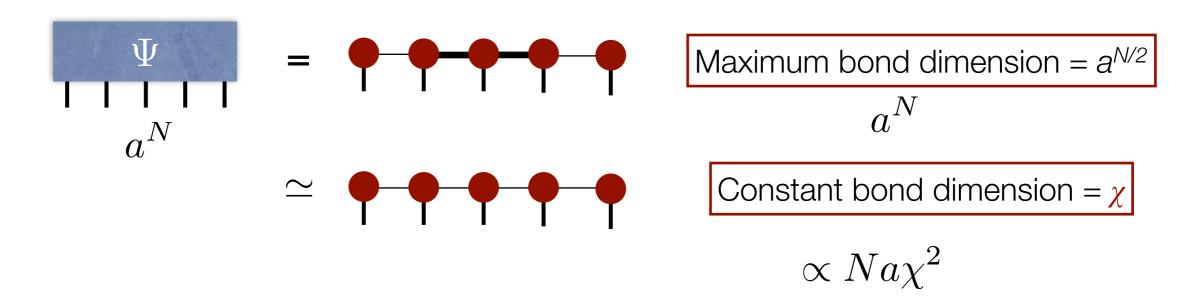


In this construction, the sizes of matrices depend on the position.

Maximum bond dimension =  $a^{N/2}$ 

At this stage, no data compression.

### Matrix product state: Low rank approximation

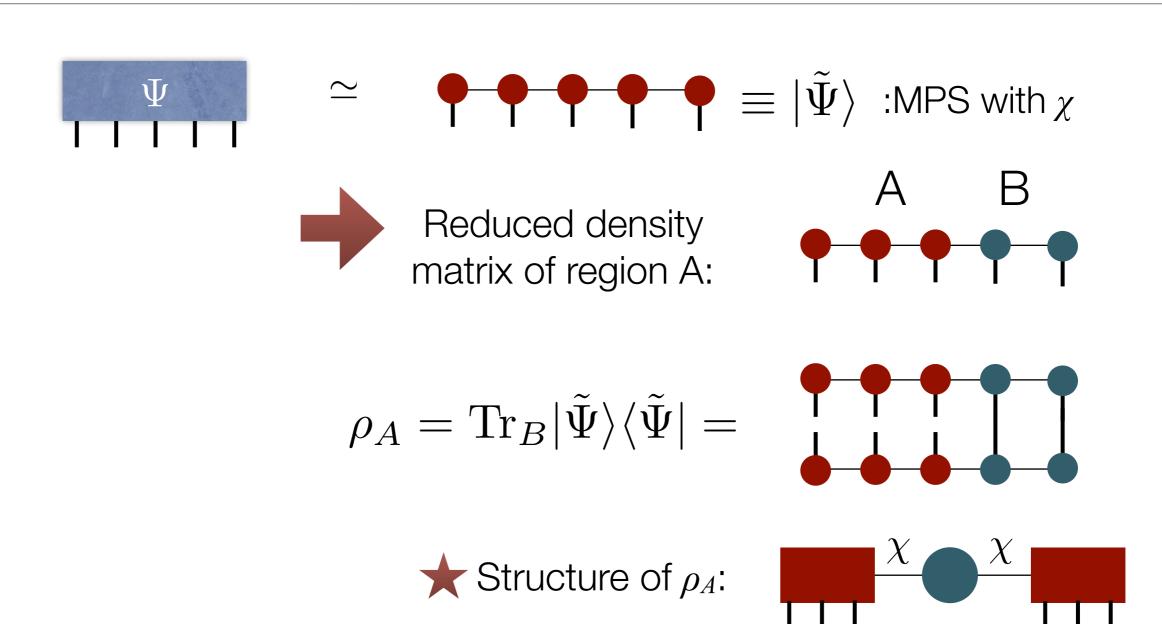


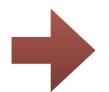
If the entanglement entropy of the system is O(1) (independent of N), matrix size " $\chi$ " can be small for accurate approximation.



On the other hand, if the EE increases as increase N, " $\chi$ " must be increased to keep the same accuracy.

# Upper bound of Entanglement entropy







$$S_A = -\text{Tr } \rho_A \log \rho_A \leq \log \chi$$

# Required bond dimension in MPS representation

$$S_A = -\text{Tr } \rho_A \log \rho_A \le \log \chi$$



The upper bound is independent of the "length".

length of MPS  $\Leftrightarrow$  size of the problem n



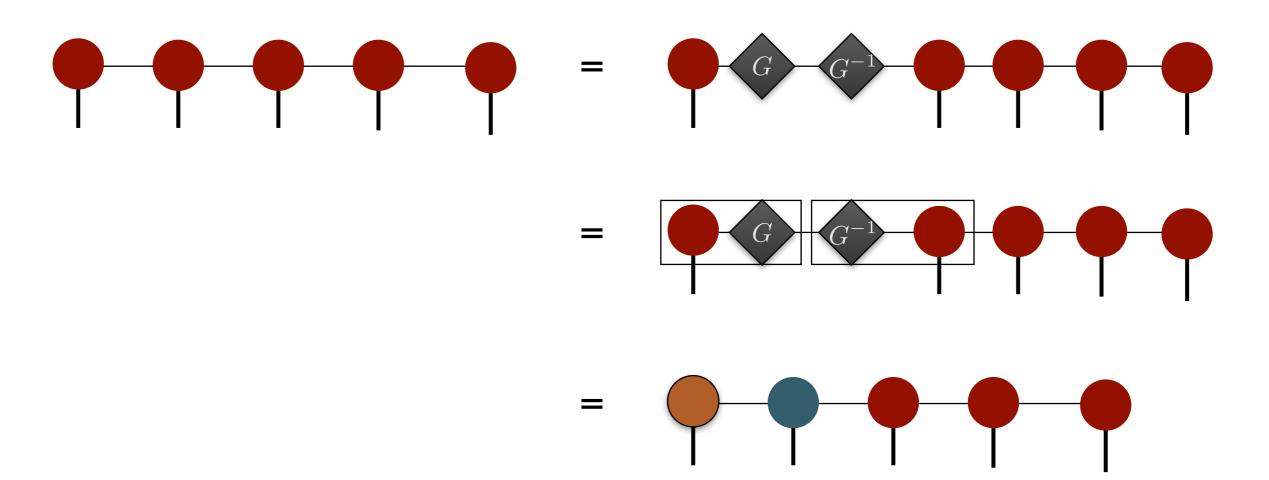
EE of the original vector	Required bond dimension in MPS representation
$S_A = O(1)$	$\chi = O(1)$
$S_A = O(\log N)$	$\chi = O(N^{\alpha})$
$S_A = O(N^{\alpha})$	$\chi = O(c^{N^{\alpha}})$

# Gauge redundancy of MPS

MPS is not unique: gauge degree of freedom

$$I = GG^{-1} \quad --- \quad = \quad -G$$

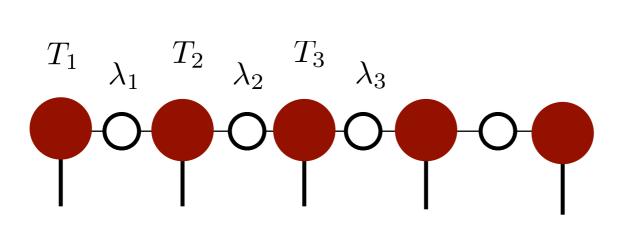
We can insert a pair of matrices GG<sup>-1</sup> to MPS



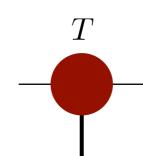
# Gauge fix: Canonical form of MPS

Ref. U. Schollwöck, Annals. of Physics 326, 96 (2011)

Canonical form of MPS: (Vidal canonical form)

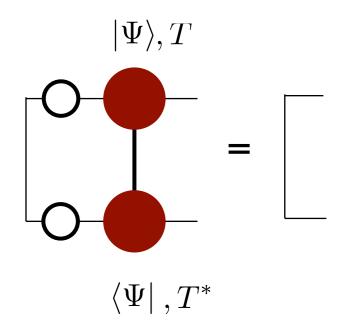


(G. Vidal, Phys. Rev. Lett. 91, 147902 (2003)

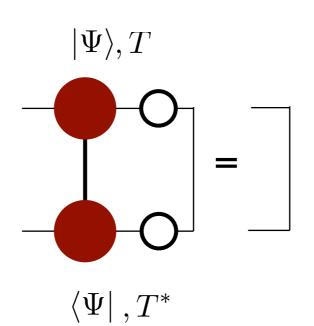


:Virtual indices corresponding to Schmidt basis

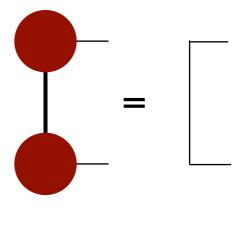
Left canonical condition:



Right canonical condition:



(Boundary)

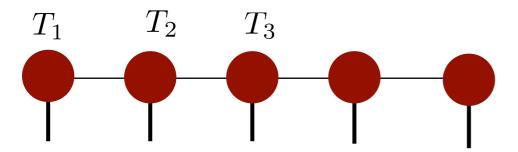


# Canonical forms: Left and Right canonical forms

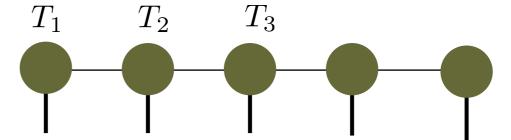
Ref. U. Schollwöck, Annals. of Physics 326, 96 (2011)

### Other "canonical" forms of MPS

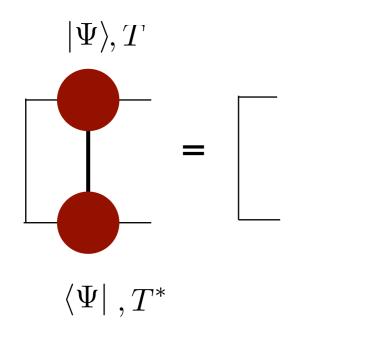
### Left canonical form:

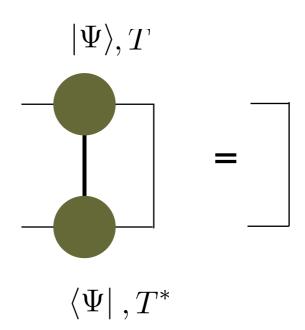


### Right canonical form:



Satisfies (at least) left or canonical condition:

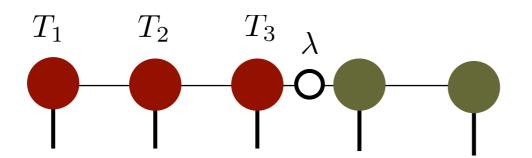




### Canonical forms: Mixed canonical forms

Ref. U. Schollwöck, Annals. of Physics 326, 96 (2011)

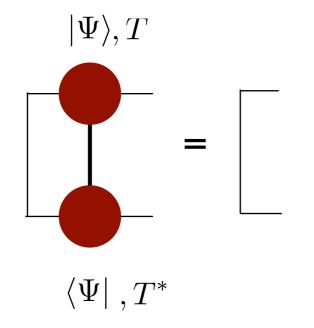
### Mixed canonical form:



 $\lambda$  is identical with the Schmidt coefficient.

### Left canonical condition:

### Right canonical condition:



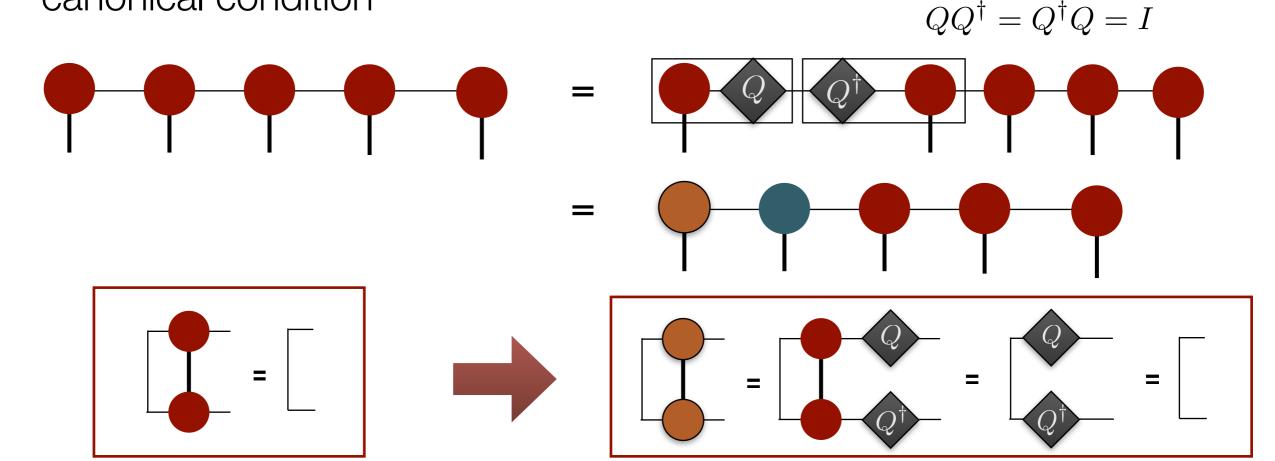
$$|\Psi\rangle, T$$

$$=$$

$$\langle\Psi|, T^*$$

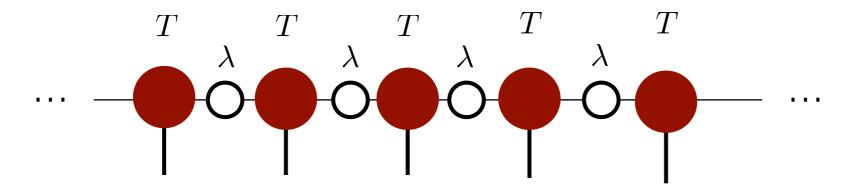
### Canonical forms: Note

- Vidal canonical form is unique, up to trivial unitary transformation to virtual indices which keep the same diagonal matrix structure (Schmidt coefficients).
- Left, right and mixed canonical form is "not unique". Under general unitary transformation to virtual indices, it remains to satisfy the canonical condition



### MPS for infinite chains

By using MPS, we can write the wave function of a translationally invariant **infinite chain** 



Infinite MPS (iMPS) is made by repeating T and  $\lambda$  infinitely.

Translationally invariant system



T and  $\lambda$  are independent of positions!

Infinite MPS can be accurate when the EE satisfies the 1d area low  $(S\sim O(1))$ .

If the EE increases as increase the system size, we may need infinitely large  $\chi$  for infinite system.

# Calculation of expectation value

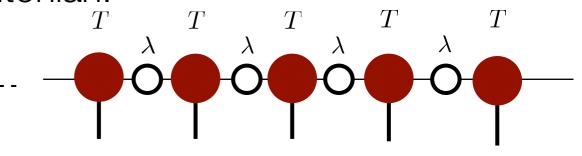
For iMPS, if it is in the (Vidal) canonical form, the final graph is identical to the above finite system.

### Example of iMPS: AKLT state

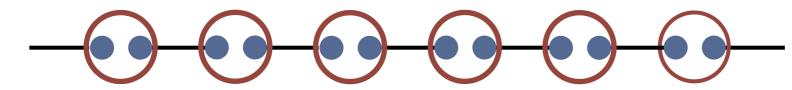
S=1 Affleck-Kennedy-Lieb-Tasaki (AKLT) Hamiltonian:

$$\mathcal{H} = J \sum_{\langle i,j \rangle} \vec{S}_i \cdot \vec{S}_j + \frac{J}{3} \sum_{\langle i,j \rangle} \left( \vec{S}_i \cdot \vec{S}_j \right)^2$$

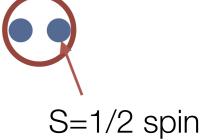
$$(J > 0)$$



The ground state of AKLT model:



S=1 spin:



 $\chi=2$  iMPS: (U. Schollwock, Annals. of Physics **326**, 96 (2011))

$$T[S_z = 1] = \sqrt{\frac{4}{3}} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

$$T[S_z = 0] = \sqrt{\frac{2}{3}} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} , \lambda = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$T[S_z = -1] = \sqrt{\frac{4}{3}} \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix}$$

Spin singlet



# Exercise 2: Make MPS and approximate it

### 2-1: Make exact MPS from GS wave function obtained by ED

(We can easily check that the MPS obtained by successive SVD satisfy the canonical condition.)

Sample code: Ex2-1.py or Ex2-1.ipynb

show help: python Ex2-1.py -h

### 2-2: Approximate the MPS by truncating singular values

- Calculate approximate GS energy and compare it with ED
- Change chi\_max and see energies

Sample code: Ex2-2.py or Ex2-1.ipynb

show help: python Ex2-2.py -h

# Requirement for running sample scripts

Python environment: python2.7 or python3

Modules: numpy, scipy and matplotlib

### Usage:

For jupyter notebook, type

jupyter notebook

and select Ex?-?.ipynb.

For python (command line), type

python Ex?-?.py -h

, then you can know how to change the parameters.

# Next week (by Yamaji sensei)

第1回: 現代物理学における巨大なデータ

第2回: 現代物理学と情報圧縮

第3回: 情報圧縮の数理1 (線形代数の復習)

第4回: 情報圧縮の数理2 (特異値分解と低ランク近似)

第5回: 情報圧縮の数理3 (スパース・モデリングの基礎)

第6回: 情報圧縮の数理4 (クリロフ部分空間法の基礎)

第7回: 物質科学における情報圧縮

第8回: データ解析の高速化:スパース・モデリングの物質科学への応用

第9回: データ空間の圧縮:クリロフ部分空間法の物質科学への応用

第10回: 高度なデータ圧縮:情報のエンタングルメントと行列積表現

第11回: 行列積表現の固有値問題への応用

**Application of MPS to eigenvalue problems** 

第12回: テンソルネットワーク表現への発展

第13回: テンソルネットワーク繰り込みによる情報圧縮