計算科学における情報圧縮

Information Compression in Computational Science

2017.12.14

#10:行列積表現の基礎

Basics of matrix product state representation

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 state
  - Matrix product state (MPS)
  - Canonical form
  - infinite MPS

Outline of tensor network decomposition

Classification of Information Compression by Memory Costs (by Yamaji-san)

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~$O(M)$

(3) A vector cannot be stored

$\boxed{\text{Required memory} \ll O(M)}$

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 $|\Psi\rangle = |01011\ldots00\rangle$

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.

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

e.g.

$$|\phi_1\rangle = \alpha|0\rangle + \beta|1\rangle$$

$$|\phi_1\rangle = |01\rangle - |10\rangle$$

structure: **"product state"**

independent elements: **small vectors**

Tensor network decomposition of a vector

Target:

Exponentially large Hilbert space

$\vec{v} \in \mathbb{C}^M$ 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

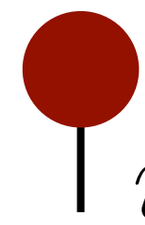
$$v_i = v_{i_1,i_2,\ldots,i_N} = \sum_{\{x\}} T^{(1)}[i_1]_{x_1,x_2,\ldots} T^{(2)}[i_2]_{x_1,x_3,\ldots} \cdots T^{(N)}[i_N]_{x_3,x_{100},\ldots}$$

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

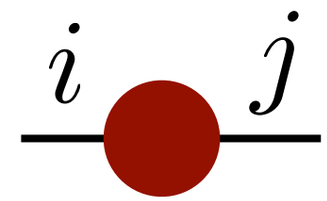
$T[i]_{x_1,x_2,\ldots}$: local tensor for "state" i

Graphical representations for tensor network

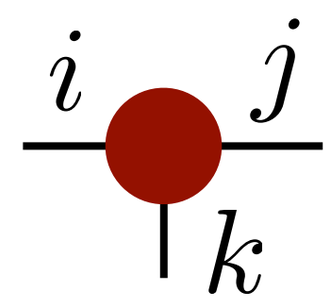
- Vector $\vec{v} : v_i$



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

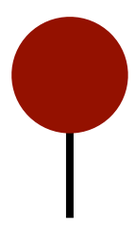


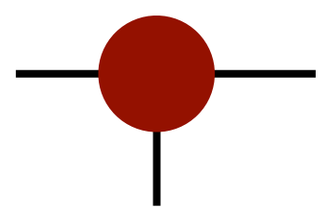
- Tensor $T : T_{i,j,k}$



*** n-rank tensor = n-leg object**

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

$\vec{v} =$ 

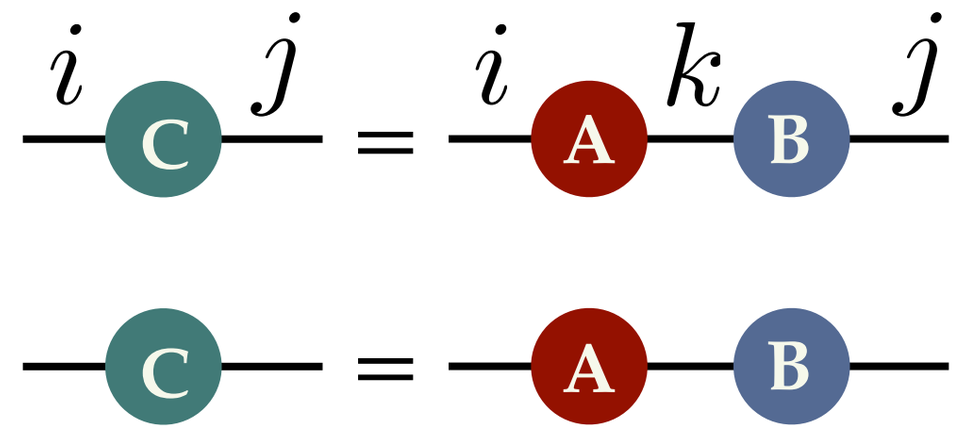
$T =$ 

Graphical representations for tensor network

Matrix product

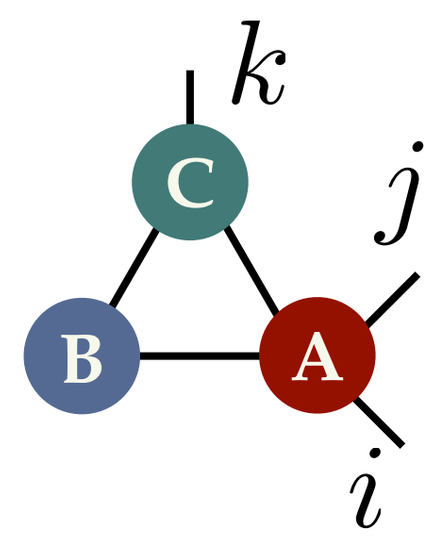
$$C_{i,j} = (AB)_{i,j} = \sum_k A_{i,k} B_{k,j}$$

$$C = AB$$



Generalization to tensors

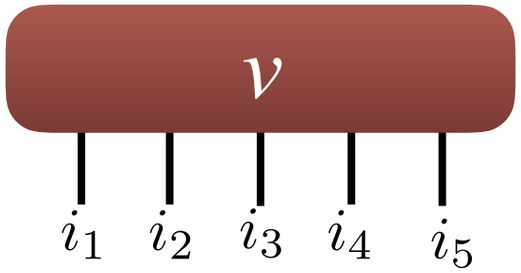
$$\sum_{\alpha,\beta,\gamma} A_{i,j,\alpha,\beta} B_{\beta,\gamma} C_{\gamma,k,\alpha}$$



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

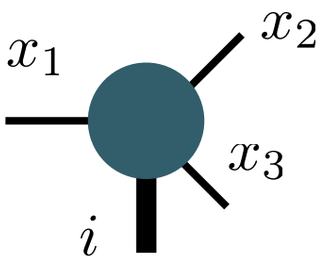
Graph for a tensor network decomposition

- Vector v_{i_1,i_2,i_3,i_4,i_5}



*Vector looks like a tensor

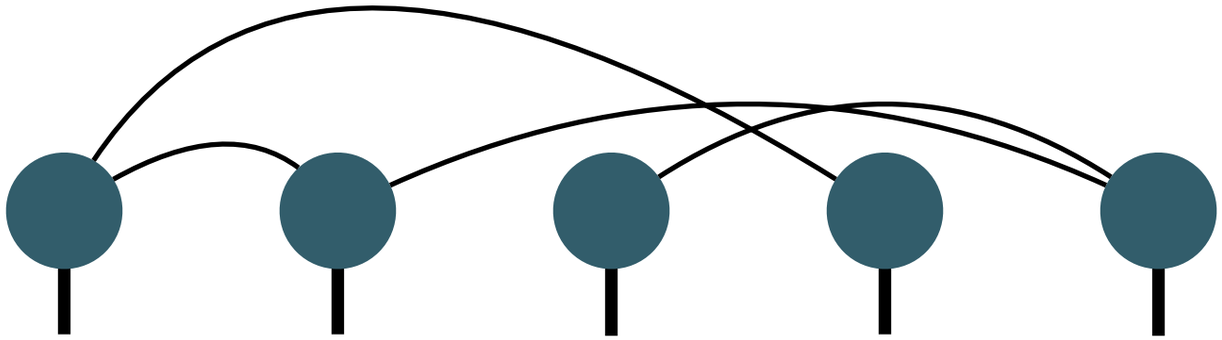
- Tensor $T[i]_{x_1,x_2,x_3}$



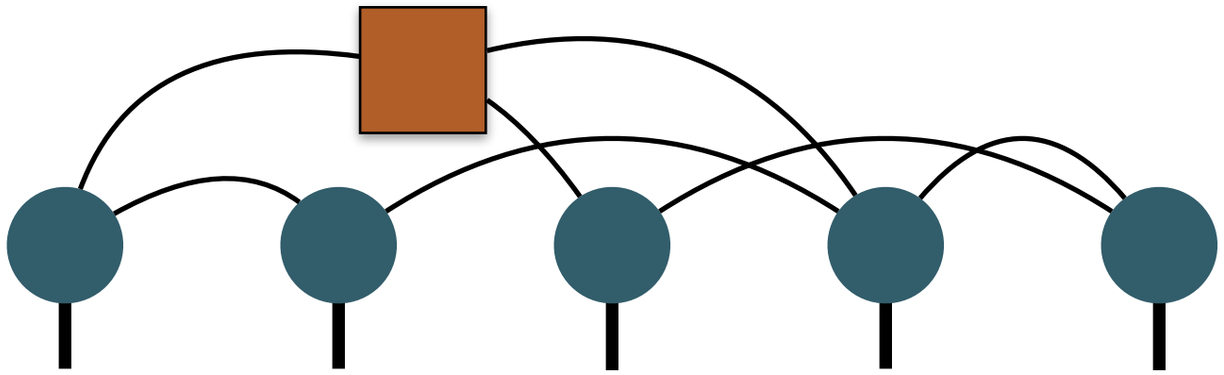
*We treat i as an index of the tensor.

Tensor network decomposition

$\vec{v} =$



$\vec{w} =$



*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 $|0\rangle, |1\rangle$

$(|\uparrow\rangle, |\downarrow\rangle)$

$$|\Psi\rangle = \sum_{\{i_1, i_2, \ldots i_N\}} \Psi_{i_1 i_2 \ldots i_N} |i_1\rangle \otimes |i_2\rangle \otimes \cdots \otimes |i_N\rangle$$

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

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

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

Schmidt decomposition

General vector: $\vec{x} \in \boldsymbol{V}_1 \otimes \boldsymbol{V}_2$

$$\dim \boldsymbol{V}_1 = n_1, \dim \boldsymbol{V}_2 = n_2$$
$$(n_1 \geq n_2)$$

Schmidt decomposition

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

Orthonormal vectors

$$\{\vec{u}_1, \vec{u}_2, \ldots \vec{u}_{n_1}\} \in \boldsymbol{V}_1$$
$$\{\vec{v}_1, \vec{v}_2, \ldots \vec{v}_{n_2}\} \in \boldsymbol{V}_2$$

Schmidt coefficient $\lambda_i \geq 0$

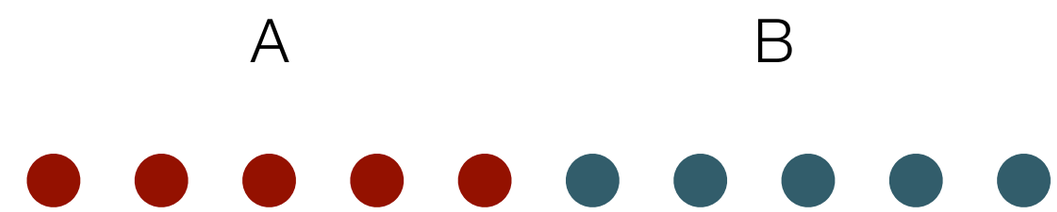
Schmidt decomposition is unique.

Schmidt decomposition for wave function

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

Schmidt decomposition

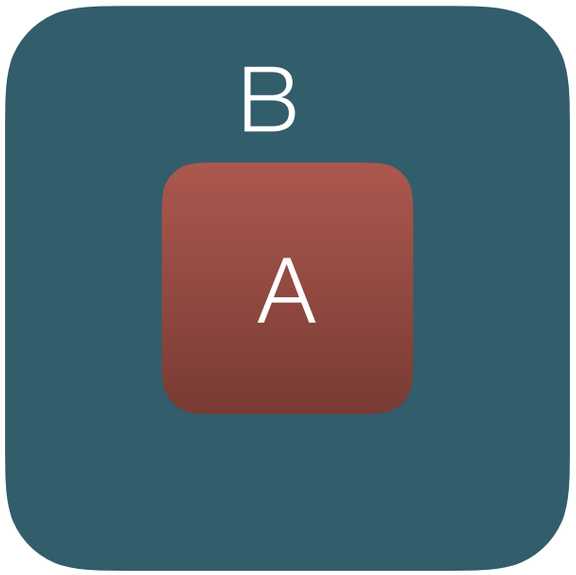
Divide system into two parts, A and B:



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

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

$$M_{i,j} \equiv \Psi_{\underbrace{(i_1, \ldots)}_{A}, \underbrace{(\ldots, i_N)}_{B}}$$



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

Schmidt coefficient: $\lambda_i \geq 0$

Partial trace and reduced density matrix

For $\vec{x} \in \boldsymbol{V}_1 \otimes \boldsymbol{V}_2$ $\quad \dim \boldsymbol{V}_1 = n_1, \dim \boldsymbol{V}_2 = n_2 \quad |\vec{x}| = 1$

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

*Note: $\mathrm{rank}\ \rho = 1$

Orthonormal basis: $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_{n_1}\} \in \boldsymbol{V}_1 \quad \{\vec{f}_1, \vec{f}_2, \ldots, \vec{f}_{n_2}\} \in \boldsymbol{V}_2$

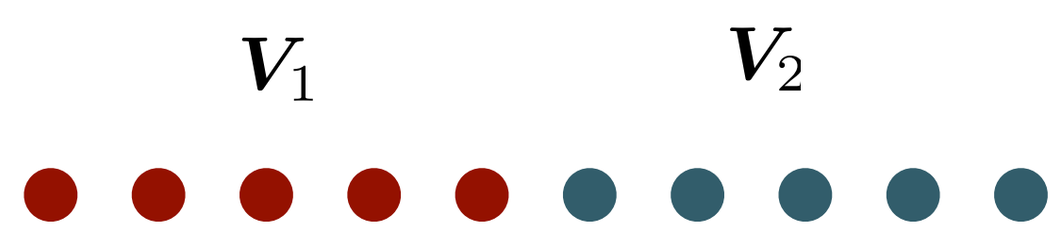
Basis for $\vec{x}$: $\vec{g}_{i_1,i_2} = \vec{e}_{i_1} \otimes \vec{f}_{i_2}$

Index: $i = (i_1, i_2)$

Reduced Density matrix:
(縮約密度行列)

$\rho_{\boldsymbol{V}_1} \equiv \mathrm{Tr}_{\boldsymbol{V}_2}\ \rho$: a positive-semidefinite square matrix in $\boldsymbol{V}_1$

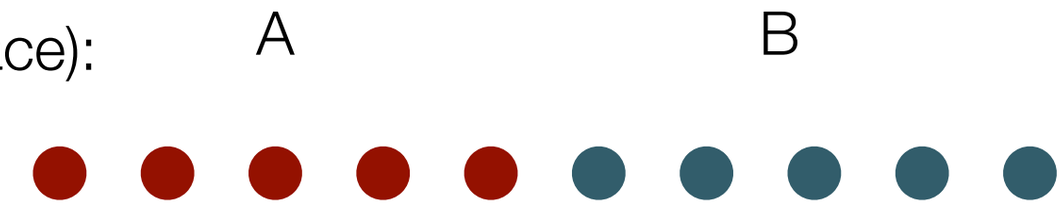
$$(\rho_{\boldsymbol{V}_1})_{i_1,j_1} = \sum_{\underline{i_2}} \rho_{(i_1,\underline{i_2}),(j_1,\underline{i_2})}$$



Entanglement entropy

Entanglement entropy:

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



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

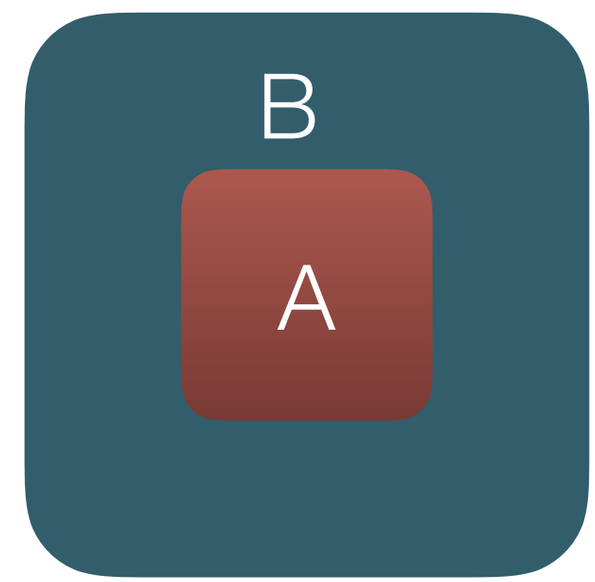
Entanglement entropy = von Neumann entropy of ρ_A

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

Schmidt decomposition $|\Psi\rangle = \sum_i \lambda_i |\alpha_i\rangle \otimes |\beta_i\rangle$

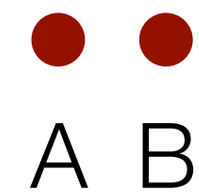
$$\rho_A = \sum_i \lambda_i^2 |\alpha_i\rangle\langle\alpha_i| \qquad \text{(*Exercise)}$$

$$S = -\sum_i \lambda_i^2 \log \lambda_i^2$$



Entanglement entropy is calculated through
the spectrum of Schmidt coefficients

Intuition for EE: two s=1/2 spins



Product state : S=0

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

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

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

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

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

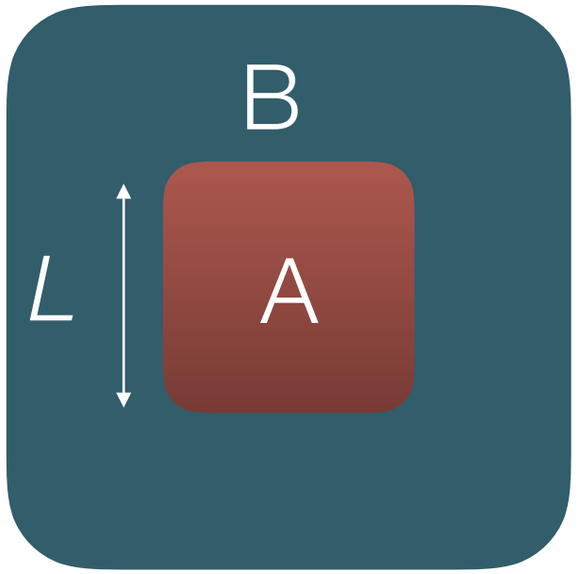
Area law of the entanglement entropy in physics

General wave functions:

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

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

(c.f. random vector)



Ground state wave functions:

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

J. Eisert, M. Cramer, and M. B. Plenio, Rev. Mod. Phys, 277, **82** (2010)

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

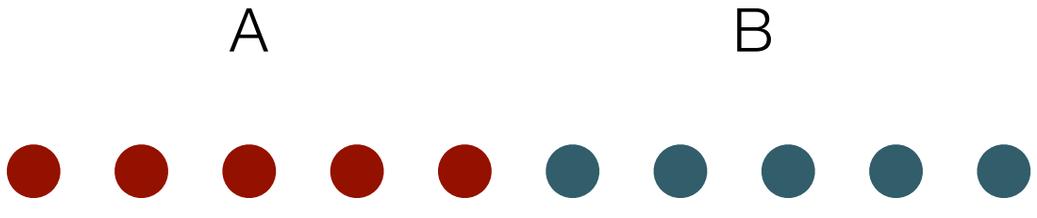
In the case of one-dimensional system:

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



Relation between SVD and Schmidt decomposition

Singular value decomposition (SVD):

For a K × L matrix M,

$$M_{i,j} = \sum_m U_{i,m} \lambda_m V^\dagger_{m,j}$$

Singular values: $\lambda_m \geq 0$

Singular vectors:

$$\sum_i U_{i,m} U^\dagger_{m,j} = \delta_i, j$$

$$\sum_i V_{i,m} V^\dagger_{m,j} = \delta_i, j$$

Relation to the Schmidt decomposition:

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

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

$$|\beta_m\rangle = \sum_j V^\dagger_{m,j} |B_j\rangle$$

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

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

Examples of Schmidt decomposition (SVD)

(Exercise)

1-1: Random wave function (Sample code: Ex1-1.py)

- 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)

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

- 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"**

1-1: Ex1-1.py

```
import numpy as np
import scipy.linalg as linalg
from matplotlib import pyplot

N=6## Chain length
m = 3          ## m = 2S + 1, e.g. m=3 for S=1
vec = (np.random.rand(m**N)-0.5) + 1.0j * (np.random.rand(m**N)-0.5)

## Make matrix from wave function
Mat = vec[:].reshape(m**(N/2),m**(N-N/2))

## SVD
U,s,VT = linalg.svd(Mat,full_matrices=False)

## Entanglement entropy
EE = -np.sum(s**2*np.log(s**2))
print "normalization=",np.sum(s**2)
s /=np.sqrt(np.sum(s**2))

EE = -np.sum(s**2*np.log(s**2))
print "EE=",EE

## plot singular values
pyplot.title("N sites random vector")
pyplot.plot(np.arange(m**(N/2)),s,"o")
pyplot.xlabel("index")
pyplot.ylabel("sigular value")
pyplot.yscale("log")
pyplot.show()
```

Make random vector corresponds to N-site S=1 spin chain

Singular value decomposition by "scipy.linalg.svd"

Output entanglement entropy

Plot singular values by matplotlib

1-2: Ex1-2.py

```
import numpy as np
import scipy.linalg as linalg
import ED
from matplotlib import pyplot

N=6             ## Chain length
m = 3           ## m = 2S + 1, e.g. m=3 for S=1
Delta = 1.0     ## Delta for XXZ
hx = 0.0        ## external field along x direction
D = 0.0         ## single ion anisotropy

eig_val,eig_vec = ED.Calc_GS(m,Delta,hx,D,N,k=1)


print "S=1 N-site open Heisenberg chain"
print "N=",N
print "Ground state energy per bond=", eig_val[0]/(N-1)

## Make matrix from wave function
Mat = eig_vec[:,0].reshape(m**(N/2),m**(N-N/2))

## SVD
U,s,VT = linalg.svd(Mat,full_matrices=False)

## Entanglement entropy
print "normalization=",np.sum(s**2)
s /= np.sqrt(np.sum(s**2))
EE = -np.sum(s**2*np.log(s**2))
print "EE=",EE

## plot singular values
pyplot.title("N sites S=1 Heisenberg chain")
```

import "ED.py" for exact diagonalization

Obtain ground state of S=1 Heisenberg chain
by exact diagoanlization

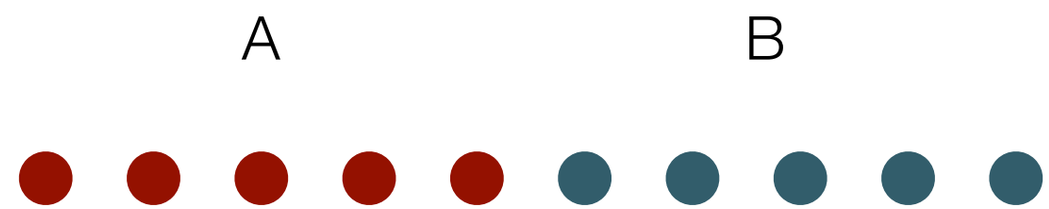
Singular value decomposition
by "scipy.linalg.svd"

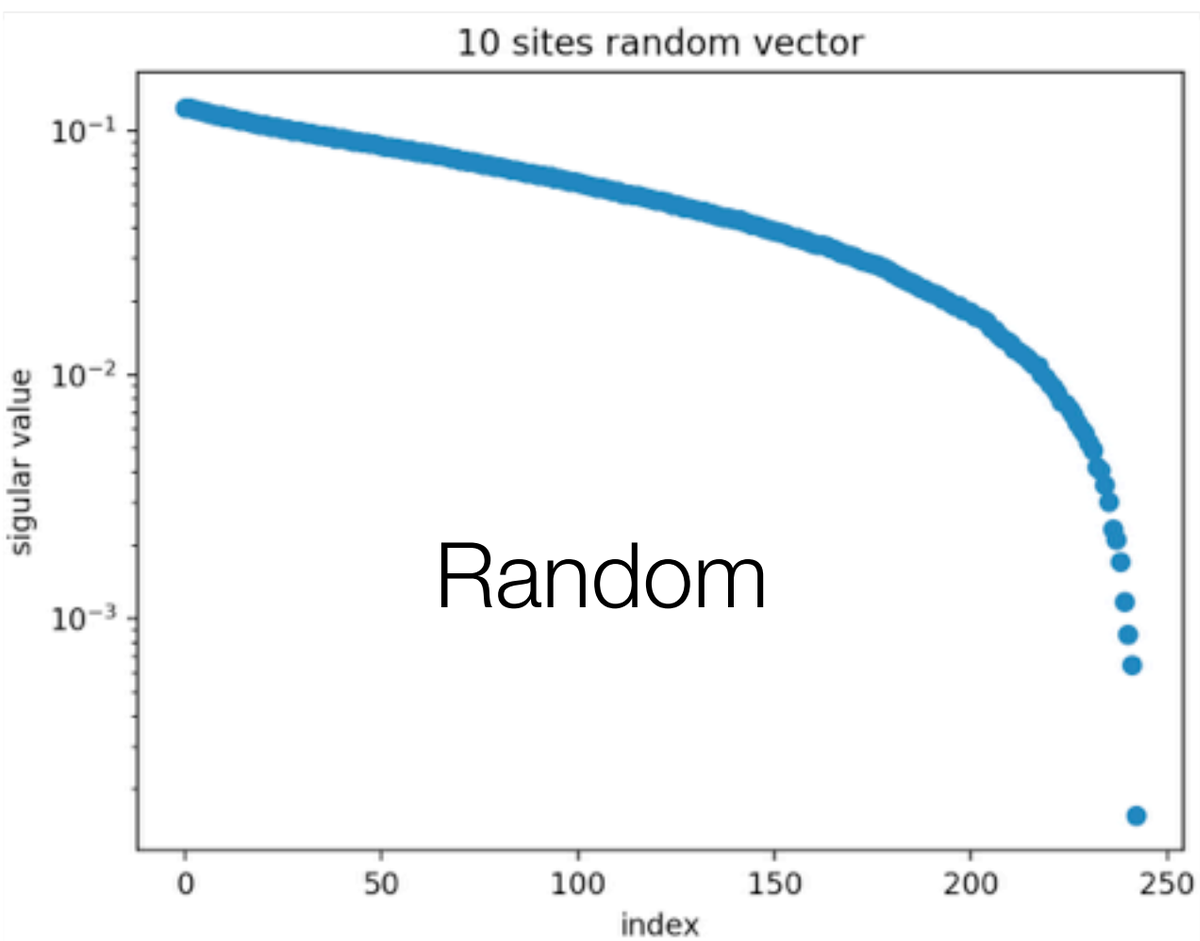
Output entanglement entropy

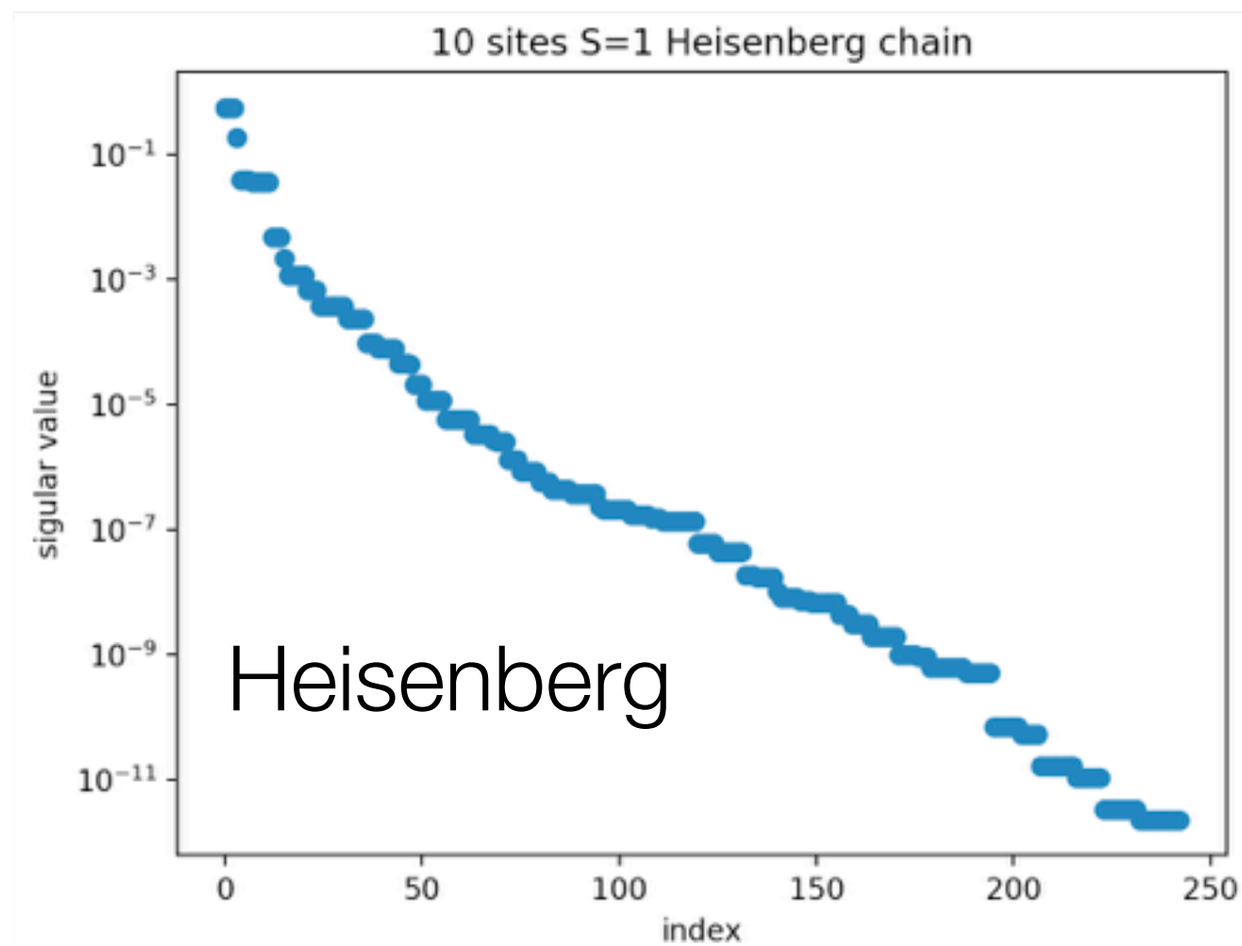
Plot singular values
by matplotlib

Result: N=10 spectrum

$$\vec{\Psi} \in \mathbb{C}^{3^{10}}$$







Ground state wave function has lower entanglement!

Matrix product state（行列積状態）

Data compression of wave functions (vectors)

General wave function:

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

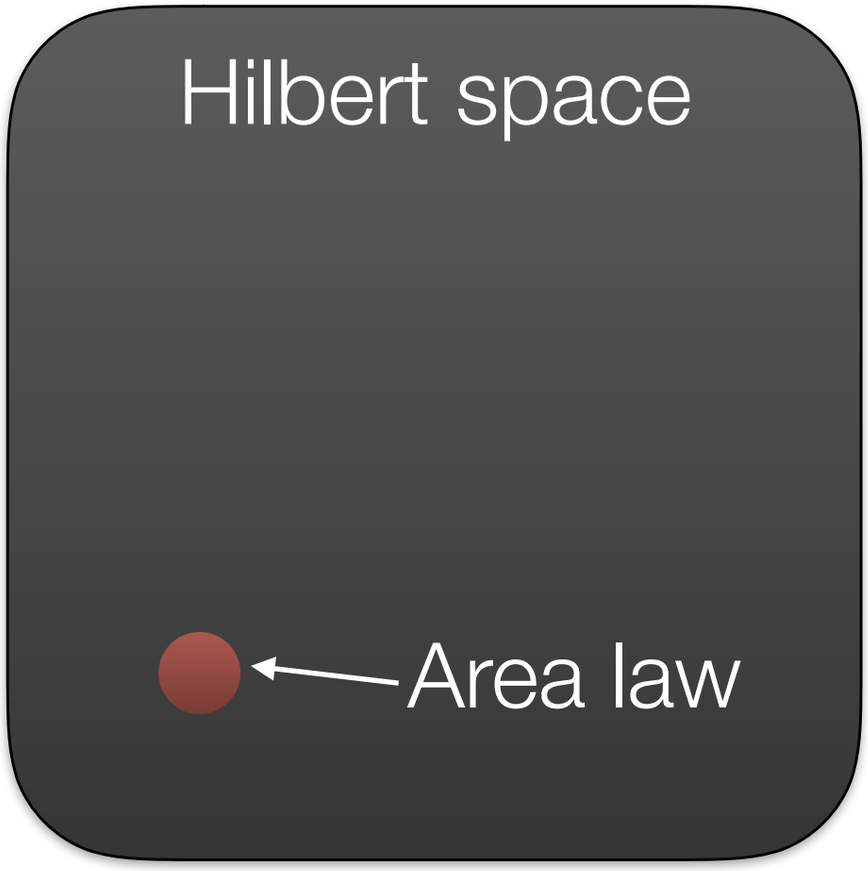
Coefficient vector can represent any points in the Hilbert space.

Ground states satisfy the area law.

In order to represent the ground state, we do not need all of a^N elements.

Data compression by tensor decomposition:

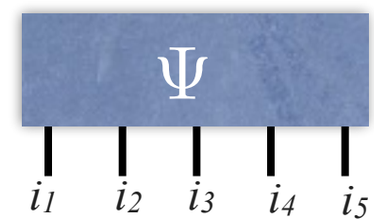
Tensor network states



Tensor network state

G.S. wave function: $|\Psi\rangle = \sum_{\{i_1, i_2, \ldots i_N\}} \Psi_{i_1 i_2 \ldots i_N} |i_1 i_2 \ldots i_N\rangle$

Vector (or N-rank tensor): $\Psi_{i_1 i_2 \ldots i_N}$ =

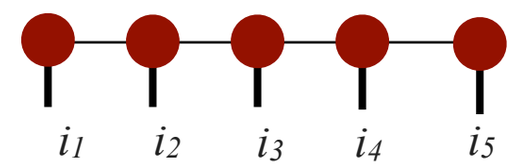


# of Elements = a^N

"Tensor network" decomposition

- Matrix Product State (MPS)

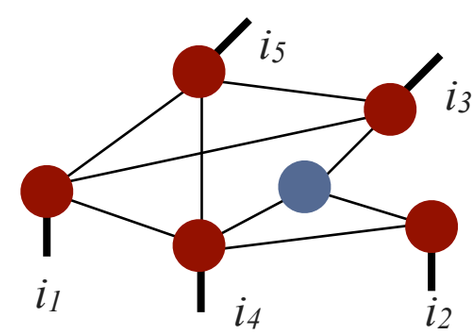
$A_1[i_1] A_2[i_2] \cdots A_N[i_N]$ =



$A[m]$: 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

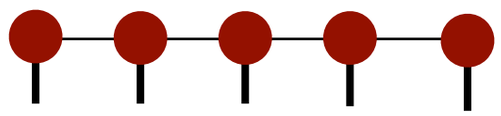
By choosing a "good" network, we can express G.S. wave function efficiently.

ex. MPS: # of elements $= 2ND^2$ D: dimension of the matrix A

Exponential → Linear *If D does not depend on N…

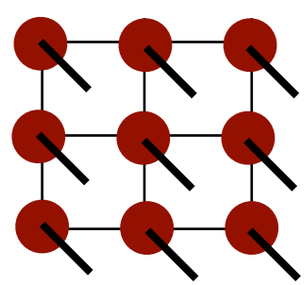
Examples of TNS

MPS:



Good for 1-d gapped systems

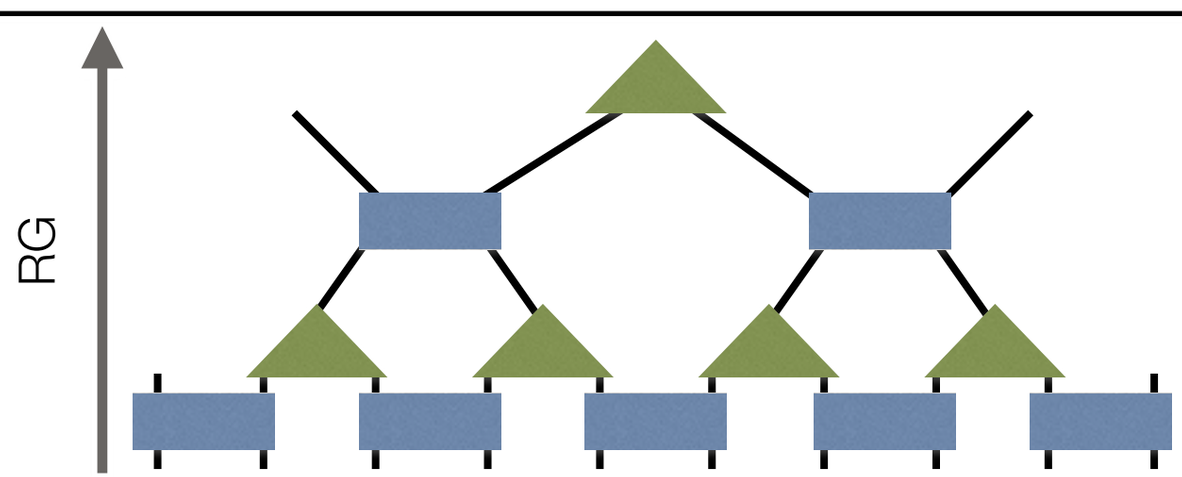
PEPS, TPS:



For higher dimensional systems

Extension of MPS

MERA:



Scale invariant systems

Matrix product state (MPS)

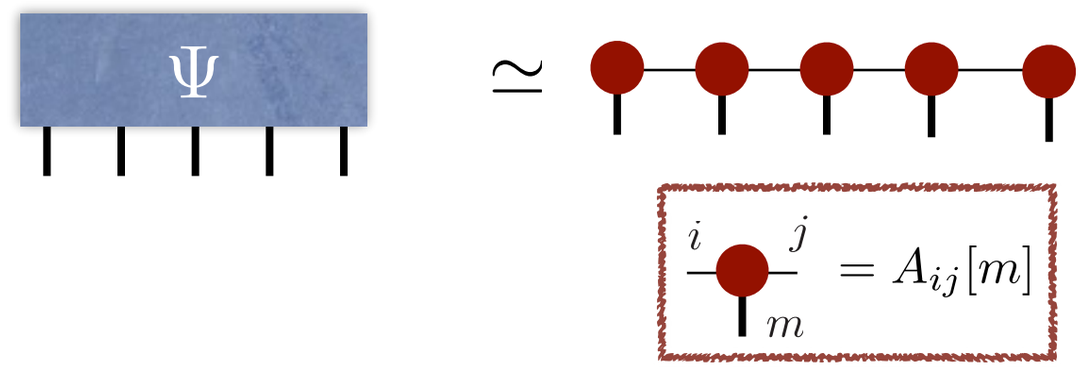
Good reviews:

(U. Schollwöck, Annals. of Physics **326**, 96 (2011))
(R. Orús, Annals. of Physics **349**, 117 (2014))

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

$$\Psi_{i_1 i_2 \ldots i_N} \simeq A_1[i_1] A_2[i_2] \cdots A_N[i_N]$$

$A[i]$: 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)

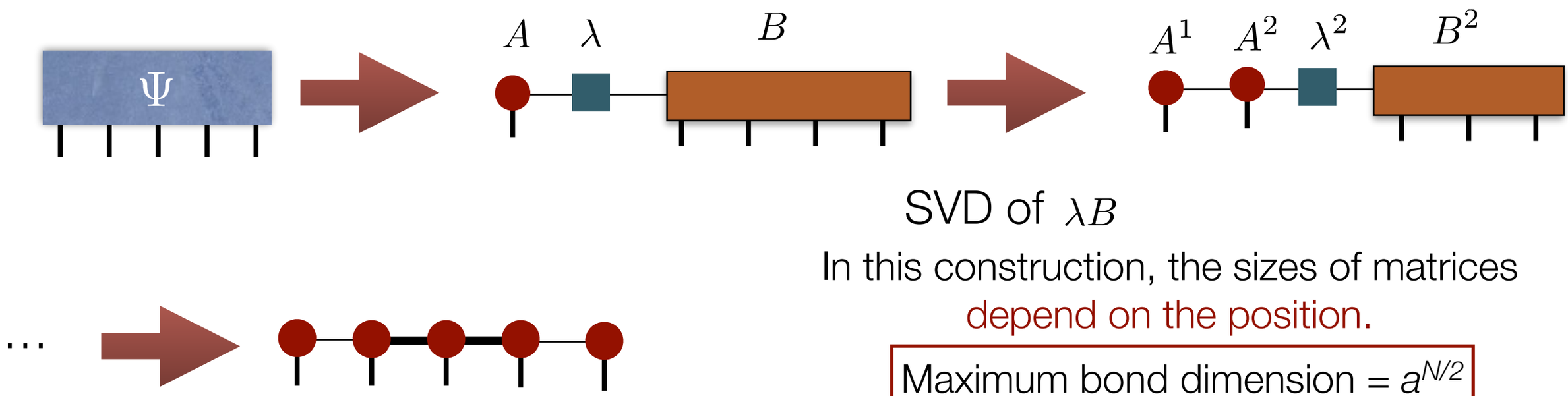
$$|\Psi\rangle = |\phi_1\rangle \otimes |\phi_2\rangle \otimes \cdots$$

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

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

Matrix product state without approximation

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



SVD of λB

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

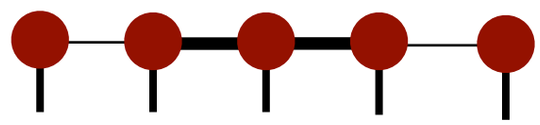
Maximum bond dimension = $a^{N/2}$

At this stage, **no data compression.**

Matrix product state: Low rank approximation

a^N

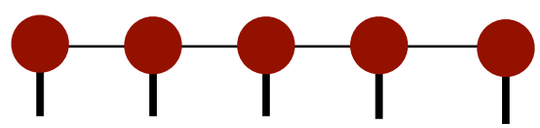
$=$



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

a^N

$\simeq$



Constant bond dimension = χ

$\propto Na\chi^2$

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

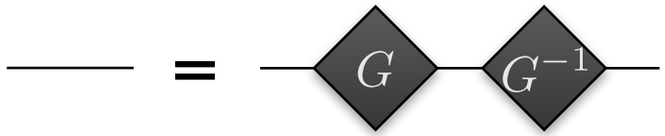
➡ MPS is good for gapped 1d systems.

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

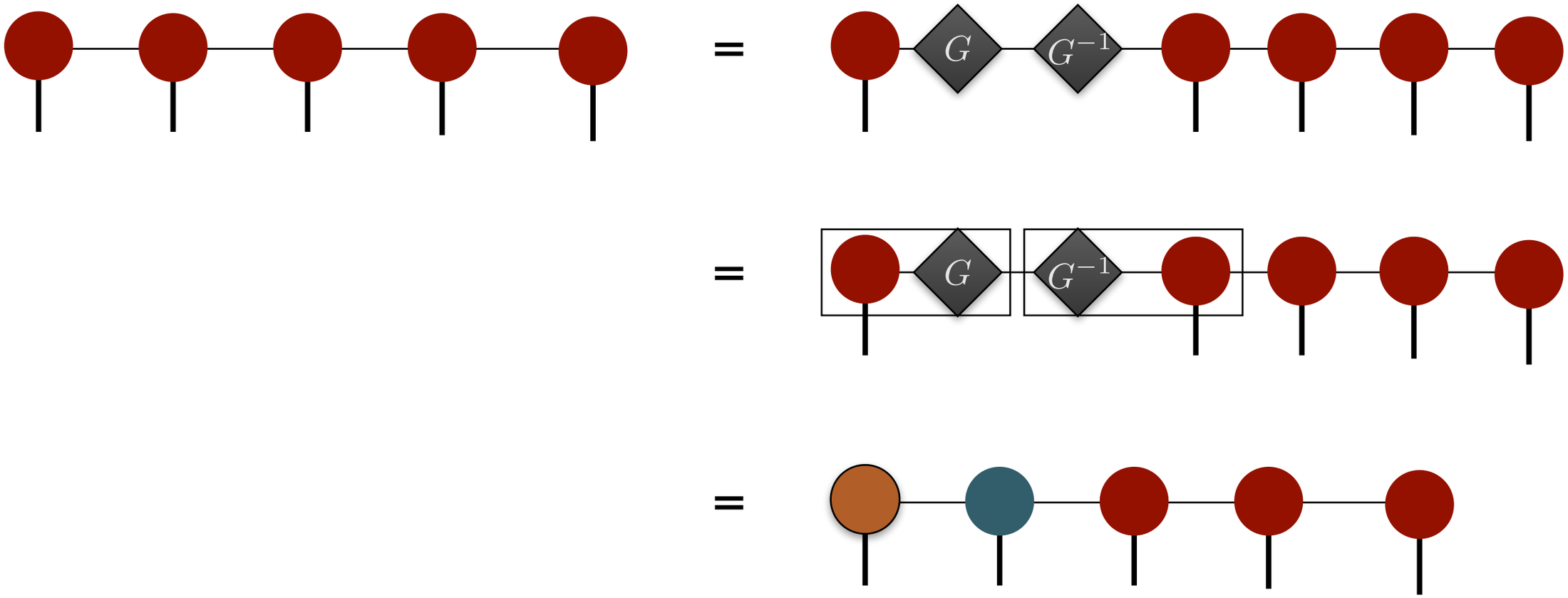
Gauge redundancy of MPS

MPS is not unique: gauge degree of freedom

$$I = GG^{-1}$$



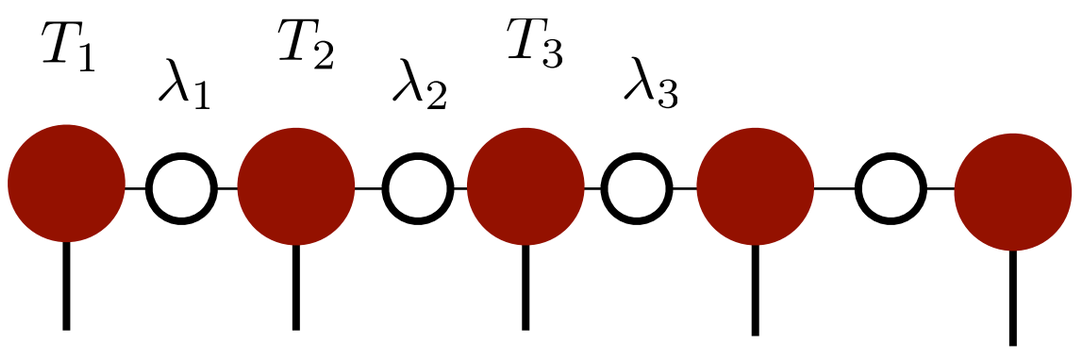
We can insert a pair of matrices GG^{-1} to MPS

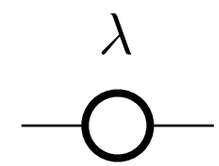


Gauge fix: Canonical form of MPS

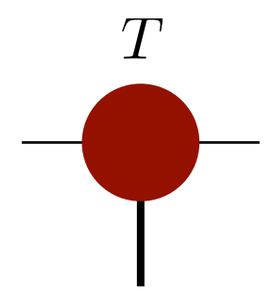
Canonical form of MPS: (Convenient for TEBD algorithm)

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



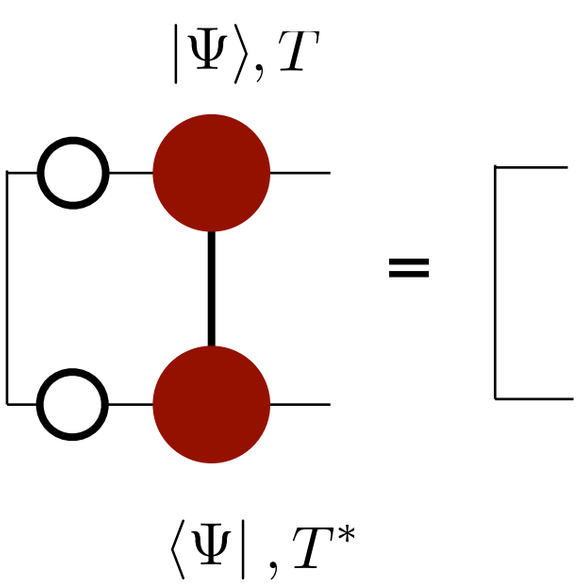


:Diagonal matrix corresponding to Schmidt coefficient

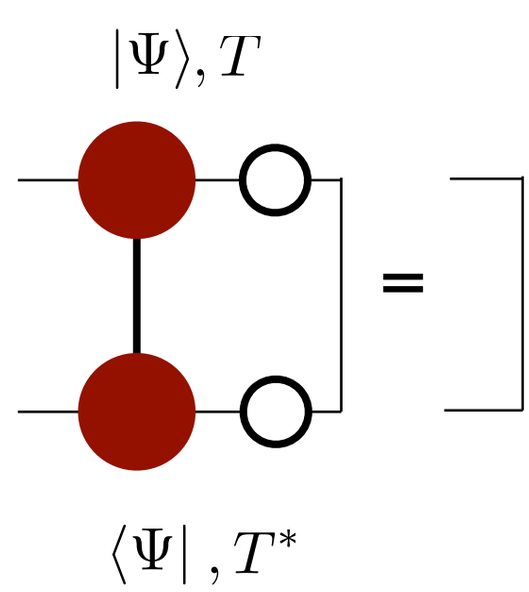


:Virtual indices corresponding to Schmidt basis

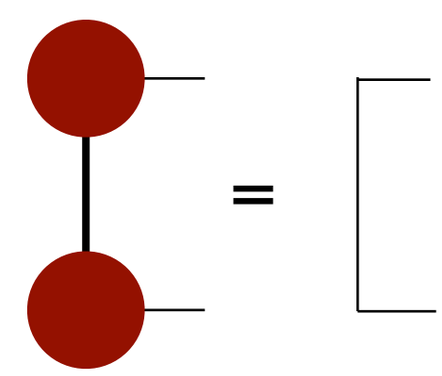
Left canonical condition:



Right canonical condition:

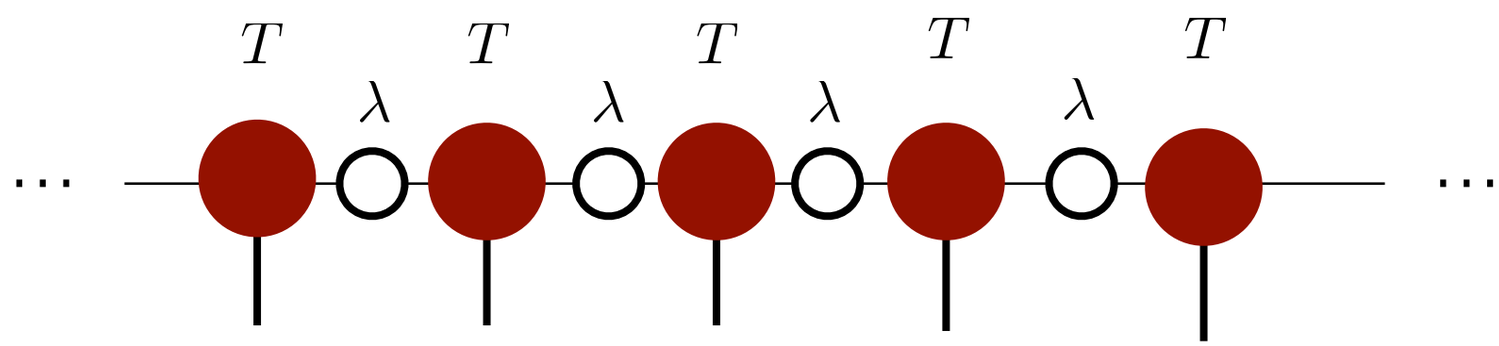


(Boundary)



MPS for infinite chains

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



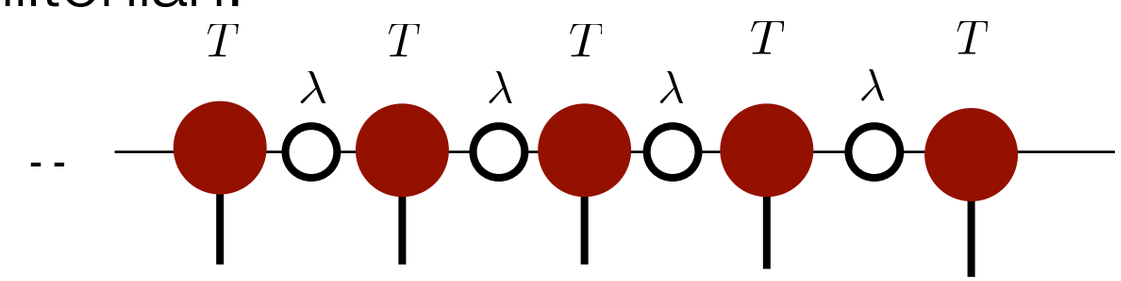
Infinite MPS (iMPS) is made by repeating T and λ infinitely.

Translationally invariant system → T and λ are independent of positions!

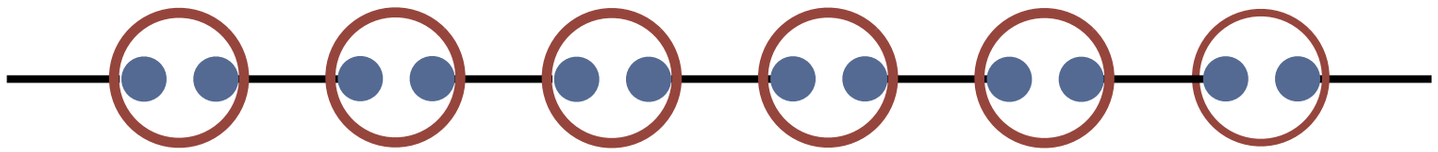
Example of iMPS: AKLT state

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

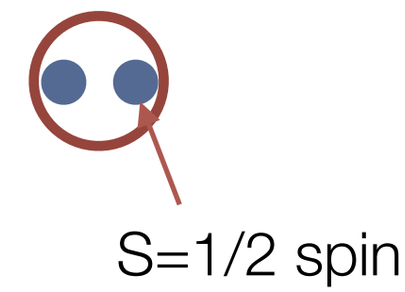
$$\mathcal{H} = J \sum_{\langle i,j \rangle} \vec{S}_i \cdot \vec{S}_j + \frac{J}{3} \sum_{\langle i,j \rangle} \left(\vec{S}_i \cdot \vec{S}_j \right)^2$$



The ground state of AKLT model:



S=1 spin:



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

Spin singlet

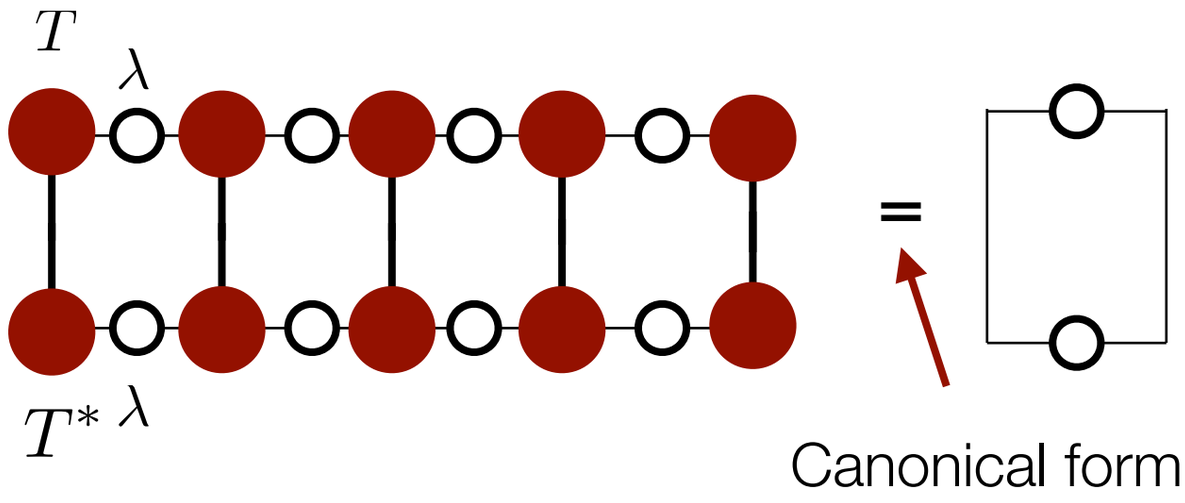
$$T[S_z = 1] = \sqrt{\frac{4}{3}} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

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

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

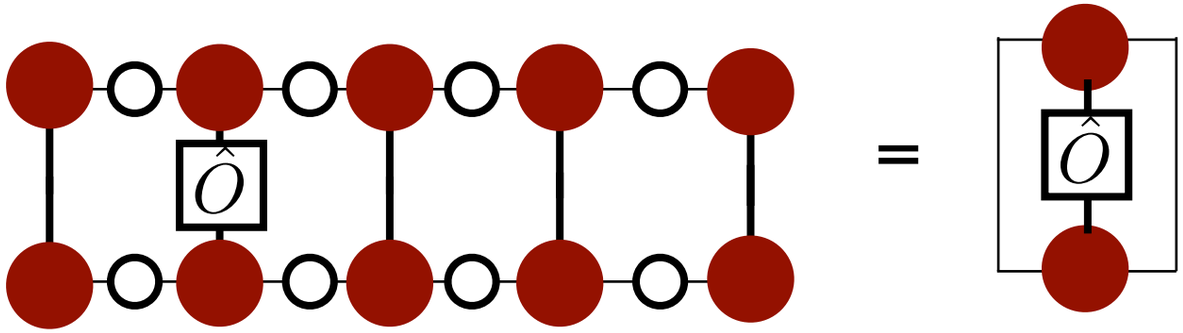
Calculation of expectation value

$\langle \Psi | \Psi \rangle =$



$= \sum_i \lambda_i^2 = 1$

$\langle \Psi | \hat{O} | \Psi \rangle =$



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

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

python Ex2-1.py

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

python Ex2-2.py

2-2: Ex2-2.py

```
import numpy as np
import scipy.linalg as linalg
import ED
import TEBD
from matplotlib import pyplot

N=6             ## Chain length
m = 3           ## m = 2S + 1, e.g. m=3 for S=1
Delta = 1.0     ## Delta for XXZ
hx = 0.0        ## external field along x direction
D = 0.0         ## single ion anisotropy

chi_max = 10      ## maxmum bond dimension at truncation


eig_val,eig_vec = ED.Calc_GS(m,Delta,hx,D,N,k=1)
```

import "TEBD.py" for energy calculation of MPS

Obtain ground state of S=1 Heisenberg chain by exact diagoanlization

```
## Make exact MPS (from "left")
Tn = []
lam = [np.ones((1,))]
lam_inv = 1.0/lam[0]
R_mat = eig_vec[:,0].reshape(m,m**(N-1))

chi_l=1
for i in range(N-1):
    U,s,VT = linalg.svd(R_mat,full_matrices=False)
    chi_r = s.size

    Tn.append(np.tensordot(np.diag(lam_inv),U.reshape(chi_l,m,chi_r),(1,0)).transpose(1,0,2))
    lam.append(s)
    lam_inv = 1.0/s
    R_mat = np.dot(np.diag(s),VT).reshape(chi_r*m,m**(N-i-2))
    chi_l = chi_r
Tn.append(VT.reshape(m,m,1).transpose(1,0,2))
lam.append(np.ones((1,)))

## Truncation
for i in range(N-1):
    chi = min(chi_max,lam[i+1].shape[0])
    lam[i+1]=lam[i+1][:chi]
    Tn[i]=Tn[i][:,:,:chi]
    Tn[i+1]=Tn[i+1][:,:chi,:]
```

Successive SVD to make MPS

Truncate singular values (Data compression)

Next week

第１回：　現代物理学における巨大なデータ
第２回：　情報圧縮と繰り込み
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：　スパース・モデリングの物質科学への応用
第９回：　クリロフ部分空間法の物質科学への応用
第１０回：　行列積表現の基礎
第１１回：　行列積表現の応用

Application of matrix product state representation

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と低ランク近似の応用