

NQS - VMC

Trial wave function $|\psi_0\rangle$. G.S. $|\psi_n\rangle$

$$E_0 = \langle \psi_0 | H | \psi_0 \rangle = \sum_{n,n'} \psi_0^*(n) \psi_0(n) \langle \psi_n | H | \psi_n \rangle$$

$$= \sum_n |\psi_0(n)|^2 E_n \geq E_0 \sum_n |\psi_0(n)|^2$$

$E_0 \geq E_0$. minimize E_0 with θ .

Basis with $|\vec{\sigma}\rangle$ at 2^N . $\sum_{\vec{\sigma}} |\vec{\sigma}\rangle \langle \vec{\sigma}| = \hat{I}$

$$E_0 = \sum_{\vec{\sigma}, \vec{\sigma}'} \langle \psi_0 | \vec{\sigma} \rangle \langle \vec{\sigma} | H | \vec{\sigma}' \rangle \langle \vec{\sigma}' | \psi_0 \rangle$$

$$= \sum_{\vec{\sigma}, \vec{\sigma}'} \psi_0(\vec{\sigma}) H_{\vec{\sigma}\vec{\sigma}'} \psi_0(\vec{\sigma}')$$

$$= \sum_{\vec{\sigma}, \vec{\sigma}'} |\psi_0(\vec{\sigma})|^2 \sum_{\vec{\sigma}'} H_{\vec{\sigma}\vec{\sigma}'} \frac{\psi_0(\vec{\sigma}')}{\psi_0(\vec{\sigma})}$$

$$\sum_{\vec{\sigma}} P_0(\vec{\sigma}) E_{\text{loc}}(\vec{\sigma}) \quad \text{Monte-Carlo.}$$

For MC. VMC $\bar{E}_0 = \frac{1}{N_s} \sum_{\text{sample}} E_{\text{loc}}(\vec{\sigma}_i)$

We need sparse H so $\vec{\sigma}$ poly related to $\vec{\sigma}'$

minimize $\theta^{(k+1)} = \theta^{(k)} - \gamma \nabla_{\theta} \bar{E}_0$

$$\partial_{\theta} \bar{E}_0 = \frac{2}{N_s} \text{Re} \sum_{\text{sample}} \partial_{\theta} [\ln \psi_0^*(\vec{\sigma}_i)] E_{\text{loc}}(\vec{\sigma}_i)$$

Adding the term below doesn't bias the +h.c.
variance gradient but improve variance.

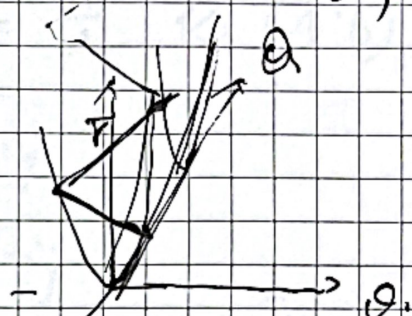
$\nabla_{\theta} \bar{E}_0$ New update scheme makes $\bar{E}_0 = \text{const} = E_0$. (G.S.)

$$\partial_{\theta} \bar{E}_0 = 0 = (E_{\text{loc}}(\vec{\sigma}_i) - \bar{E}_0)$$

We make sure when step towards E_0 nears E_0 the gradient also becomes small to improve variance. though new updating scheme.

$$\bar{E}_0 = \frac{2}{N_S} \text{Re} \sum_{\vec{\sigma}_i \sim \text{MPP}} \langle \psi_0 | \ln \psi_0^*(\vec{\sigma}_i) \cdot (E_{\text{loc}}(\vec{\sigma}_i) - E_0) | \psi_0 \rangle$$

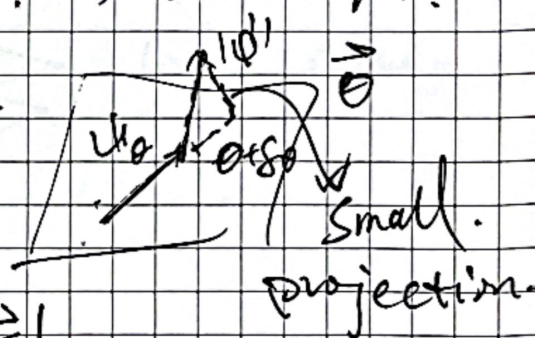
$$\theta^k = \theta^{k-1} - \lambda \nabla_{\theta} \bar{E}_0$$



small change in $\vec{\theta}$ can leads to large change in wave function.

Imaginary time $e^{-\delta \tau H} |\psi_0\rangle = (1 - \delta \tau H) |\psi_0\rangle = |\psi'\rangle$

$$|\psi_{\vec{\theta} + \delta \vec{\theta}}\rangle \simeq |\psi'\rangle$$



$$|\psi_{\vec{\theta} + \delta \vec{\theta}}\rangle \simeq (1 + \delta \vec{\theta} \cdot \vec{\nabla}) |\psi_{\vec{\theta}}\rangle$$

$$\vec{\nabla} = \sum_{\vec{\sigma}} \nabla_{\vec{\sigma}} \ln \psi_0(\vec{\sigma}) |\vec{\sigma}\rangle \langle \vec{\sigma}|$$

$$= \sum_{\vec{\sigma}} O(\vec{\sigma}) |\vec{\sigma}\rangle \langle \vec{\sigma}|$$

$$G = \langle \theta \theta \rangle - \langle \theta \rangle \langle \theta \rangle$$

$$\begin{aligned} F(\delta \vec{\theta}, \delta \vec{\theta}) &= |\langle \psi_{\vec{\theta} + \delta \vec{\theta}} | \psi' \rangle|^2 \\ &= 1 + \delta \vec{\theta}^T (\langle H^2 \rangle - E_0^2) \end{aligned}$$

$$- \delta \vec{\theta}^T \nabla_{\vec{\theta}} E_0 + \delta \vec{\theta}^T G \cdot \delta \vec{\theta} \dots$$

$$\theta^{k+1} = \theta^k - \frac{1}{G} \nabla_{\theta} E_0$$

QBIT.

Fisher information

In imaginary time. $G \Delta \theta = dz/2 \nabla_{\theta} E_0$

$$S_{\theta^k \theta^{k'}} = \sum_{\vec{\sigma}} P_{\vec{\sigma}}(\vec{\sigma}) \partial_{\theta^k} \ln(P_{\vec{\sigma}}(\vec{\sigma})) \partial_{\theta^{k'}} \ln(P_{\vec{\sigma}}(\vec{\sigma}))$$

For VMC. we know $\vec{\theta}$. we have $E_{loc}(\theta)$.

we have $e_k = \sum_{\delta} H \frac{\psi_0(\delta)}{\psi_g(\delta)}$ and $Q = \partial \theta_i \ln \psi_0(\delta)$.

we have $b_k = e_k - \bar{e}$. $A_{ki} = \partial_{\theta_i} b_k - \bar{\partial}_{\theta_i}$

we define. $S = \frac{1}{n} \text{Re}(A^T A)$ $F = \frac{1}{n} \text{Re}(A^T b)$

we proof: $\langle \Delta \theta_i^* \Delta \theta_j \rangle = Q_{ij}$

$$S_{ij} = \text{Re } Q_{ij}$$

$$F_i = \text{Re} \langle \Delta \theta_i^* \Delta E_{loc} \rangle = \frac{1}{2} \frac{\partial E}{\partial \theta_i}$$

we need Imaginary time evolution to get G.S.

we want it fit the gradient $\Rightarrow$ lower energy

One step after Imaginary time.

$$- \delta_L \Delta E_{loc}(\theta).$$

One step after gradient.

$$\sum_i \Delta \theta_i \Delta Q_i(\delta).$$

we want the effect of these two similar. so define the

overlap. $L = \langle |\sum_i \Delta \theta_i \Delta Q_i + \delta_L \Delta E_{loc}|^2 \rangle$
Maximize.

$$L = \Delta \theta^T S \Delta \theta + 2 \delta_L F^T \Delta \theta + \text{const.}$$

$$\frac{\partial L}{\partial \theta} = 0 = 2 S \Delta \theta + 2 \delta_L F \Rightarrow \delta \Delta \theta = - \delta_L F$$

So. For Imaginary time. $\theta_{\text{new}} = \theta - \frac{\text{Imaginary}}{\delta_L \Delta E_{loc}}$

$$= \theta - \delta_L G^{-1} F$$

$$= \theta - \frac{\delta_L}{2} (\text{Re } Q)^{-1} \nabla_{\theta} E$$

Actually, calculation Θ^{-1} is dangerous numerically. So:
add small $\lambda > 0$ $S \rightarrow S + \lambda I$.

$$L = \Delta \theta^T (S + \lambda I) \Delta \theta + 2 \delta \tau F^T \Delta \theta + \text{const}$$

$$\Theta_{\text{new}} = \Theta - \delta \tau (S + \lambda I)^{-1} F$$

we define $v = (S + \lambda I)^{-1} F$, that is equal to
solve $(S + \lambda I) v = F$ simpler!!

we solve v . $\boxed{\Theta_{\text{new}} = \Theta - \delta \tau v}$ updating.

Two-body wavefunction. $\Theta_i: \{a_i, V_{ij}\}$.

$$\psi_{\Theta}(\vec{\sigma}) = \exp \left[\sum_i a_i \sigma_i + \sum_{i < j} V_{ij} \sigma_i \sigma_j \right]$$

$$\partial a_i \ln \psi_{\Theta}(\vec{\sigma}) = \sigma_i$$

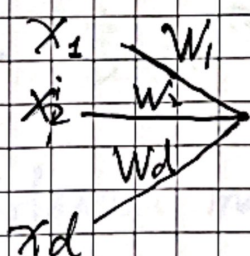
$$\partial V_{ij} \ln \psi_{\Theta}(\vec{\sigma}) = \sigma_i \sigma_j$$

MF two-body correlation.

$$|\psi_{\Theta}(\vec{\sigma})|^2 = P_{\Theta}(\vec{\sigma})$$

NQS

parameters



pre-activation.

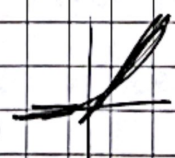
$$z = \vec{W} \cdot \vec{X} + b$$

Neuron: it's a function can be non-linear.

input: $(x_1, \dots, x_d)$. output $\delta(z)$ function non-linear.

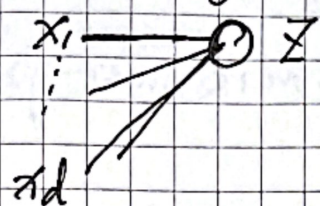
free parameters: $(\vec{W}, \vec{b})$.

$\text{Relu}(x) = \max(0, x)$

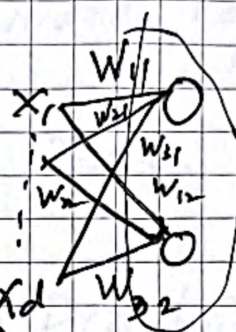


input $\vec{x}$. output $\delta(\vec{W} \cdot \vec{x} + b)$. neuron

Add a layer..



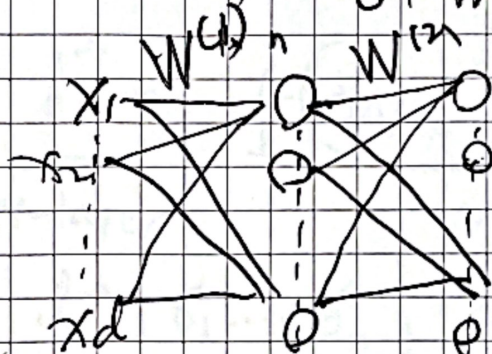
Now



Matrix of weights

$W_{d \times n}$ layers.

Now: $\delta(z) = \delta(\vec{W} \cdot \vec{x} + b)$ hidden layer



$W_{ij}^{(n)}$ $1 \times n$

Neuron-Network

input d Hidden Output k layer.

d -dim input to k -dim vector.

input $\vec{\sigma}$ basis. output $\psi_0(\vec{\sigma})$. $K=1$

Notation: $W_{ij}^{(1)} b_i^{(1)} \vec{\sigma} : \{W^{(1)} b^{(1)}\}$

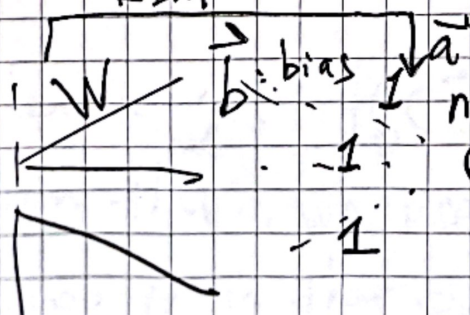
$$\delta(\vec{z}_1) = \delta(W^{(1)} \vec{\sigma} + b^{(1)})$$

$$\vec{z}_2 = W^{(2)} \delta(\vec{z}_1) + b^{(2)}$$

$$\psi_0(\vec{\sigma}) = \delta(W^{(2)} \delta(W^{(1)} \vec{\sigma} + b^{(1)}) + b^{(2)})$$

Restricted Boltzmann machine: RBM

$$\psi_{\text{RBM}}(\vec{\sigma}) = \exp \left[\sum_i a_i \sigma_i + \sum_k \log(2 \cdot \cosh(\vec{W} \cdot \vec{\sigma} + b_k)) \right]$$



We choose σ function to make this possible ψ_{RBM} .

$$\sigma^{(1)}(x) = \log[2 \cosh(x)] \leftarrow \sum_j h_j u_j, h_j = \pm 1 \text{ 隐藏层}$$

$$\sigma^{(2)}(x) = \exp(x) \quad 1 \text{ hidden layer. 双向传递}$$

skip-connection: brings MF term.

RNN.

probability $P(\vec{\sigma}) = P(\sigma_1) \underbrace{P(\sigma_2 | \sigma_1)}_{\text{through neural network}} \cdots P(\sigma_n | \sigma_1, \sigma_2, \dots, \sigma_{n-1})$

$$\vec{h}_n = f\left(\underbrace{W \left[\vec{h}_{n-1}, \vec{\sigma}_{n-1} \right]}_{N-1\text{-layer}} + \vec{b}\right) \text{ similar to MPS.}$$

N-layer

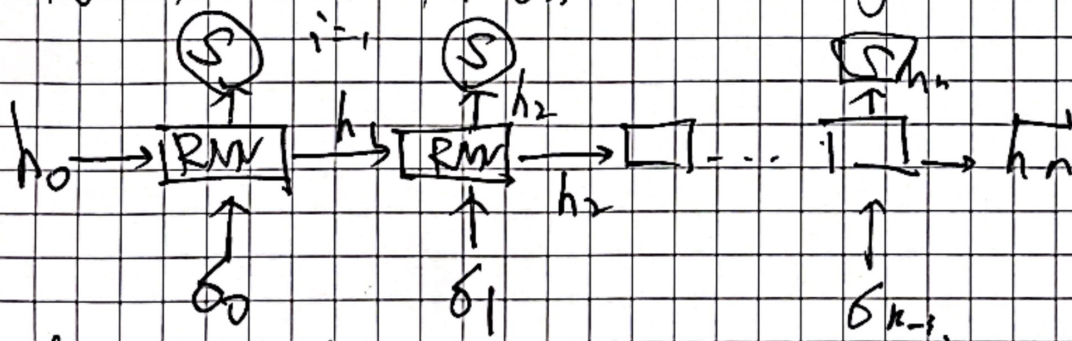
$$\vec{h}_n(\sigma_1, \dots, \sigma_{n-1}, \vec{W})$$

$$P_\theta(\sigma_n | \sigma_{n-1}) = S(u \cdot \vec{h}_n + c) \cdot \vec{\sigma}_n$$

u_h

$$S(V_k) = e^{V_k} / \sum_i e^{V_i} \text{ Gives a normalized probability distribution}$$

$$P_\theta(\vec{\sigma}) = \prod_{i=1}^n \psi_i \cdot \vec{\sigma}_i \text{ def over } \sigma_n.$$



direct sampling.

RNN: softmax function $(h_0, \sigma_0) \xrightarrow{\text{gives}} P_\theta(\sigma_1)$

$\longrightarrow \dots \longrightarrow P_\theta(\vec{\sigma})$

Then we get $\psi_\theta(\vec{\sigma}) = \sqrt{P_\theta(\vec{\sigma})}$. Also we get $E_{\text{loc}}(\vec{\sigma})$ from that. we use that to minimize energy and change θ in that process. New θ in RNN gives

new $\mathbb{P}(h_0, s_0) \xrightarrow{\text{RMN}} \mathbb{P}_\theta(s_1 | s_0)$

$$\text{Softmax}(u_{h_i} + c) = \mathbb{P}_\theta(s_i | s_{<i})$$

$$h_i = f_\theta(h_{i-1}, e_{i-1})$$

Softmax ^{shifts} gives $\forall x \in \mathbb{R}$ to function non-negative.

and total equal to 1 probability. $P_i = \frac{e^{z_j^*}}{\sum_{j=1}^k e^{z_j^*}}$

f is RMN function: (δ) .

based on parameters: (θ)

NAS is a method to form a $\mathcal{H}$ for a large given θ parameters. when $\mathcal{P}$ can also be formed in the same time.