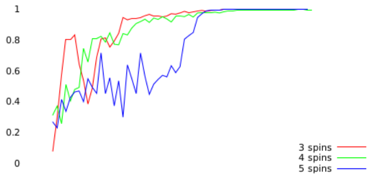
Many Body Gates: from small chains to networks

Nicola Pancotti
Max Planck Institute of Quantum Optics



April 22, 2015

COLLABORATIONS

Prof. Sougato Bose
University College of London

Dr. Leonardo Banchi
University College of London

&

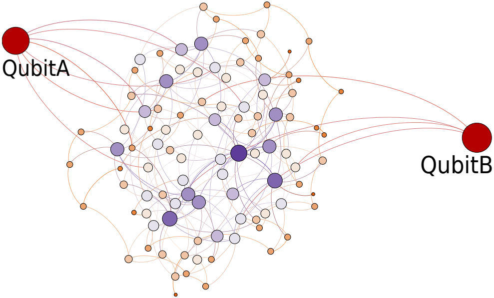
Prof. Paolo Mataloni
La Sapienza, Universit di Roma

Introduction

Interactions and Topology

Our Aim...

- We design the Hamiltonian
- We study new topologies



$\longrightarrow \hat{H}(\vec{\mathsf{J}})$

$$\hat{H}(\vec{\mathsf{J}}) = \sum_{ij} \sum_{\alpha\beta} \mathsf{J}_{ij}^{\alpha\beta} \sigma_i^{\alpha} \sigma_j^{\beta}$$

Note

The vector $\vec{\mathsf{J}}$ contains all the information we need.

In order to...

- Implement perfect gates between Qubits A and B
- find new Hamiltonian representations for quantum gates
- speed up the activation of the gates
- develop new algorithms capable of dealing with large systems
- ...

THE IMPORTANCE OF J

It gives us information about

- WHETHER OR NOT TWO NODES ARE CONNECTED
- THE STRENGTH AND THE DIRECTION OF ITS CONNECTIONS
- AND WHAT IT IS CONNECTED TO

Example (Spin Chain XX-YY)

Quantum Evolution

$$H(\vec{\mathsf{J}}) = \sum_{i=1}^{N} \mathsf{J}_i^x \sigma_i^x \sigma_{i+1}^x + \mathsf{J}_i^y \sigma_i^y \sigma_{i+1}^y$$

$$|\psi_{ch}(t)\rangle = \exp\{-iH(\vec{\mathsf{J}})t\} |\psi_{ch}(0)\rangle$$

Question

*Is there any time t^* and vector $\vec{J}$ such that a particular **entangling gate** between the first and the last qubit of the chain is implemented?*

Theory

Definition

G is an entangling gate iif there exists at least one separable density matrix ρ_{AB} such that $\rho^{G}_{AB} = \hat{G}\rho_{AB}\hat{G}^{\dagger}$ is not separable, i.e. entangled.

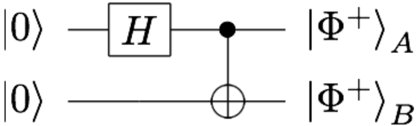
Example

$$CNOT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\hat{H} \otimes \mathbb{1} \left|00\right\rangle = \left|+\right\rangle \left|0\right\rangle$$

$$\downarrow$$

$$CNOT \left|+\right\rangle \left|0\right\rangle = \left|\phi^{+}\right\rangle$$



A VERY USEFUL OBSERVATION

CARTAN DECOMPOSITION *(any 2-qubit gate)*

$$U_{AB} = (U_A \otimes U_B) U_d (V_A \otimes V_B)$$

$$U_d = \exp\{i\, \vec{\sigma}_A^T\, \mathbf{d}\, \vec{\sigma}_B\}$$

$$\mathbf{d} = \begin{pmatrix} \alpha_x & 0 & 0 \\ 0 & \alpha_y & 0 \\ 0 & 0 & \alpha_z \end{pmatrix} \qquad \vec{\sigma} = \begin{pmatrix} \sigma_x \\ \sigma_y \\ \sigma_z \end{pmatrix}$$

Example

$$U_d = \exp\{\frac{\pi i}{4} \sigma_x \otimes \sigma_x\},$$
$$U_A = \exp \frac{-\pi i \sigma_y}{4} \exp \frac{\pi i \sigma_x}{4},$$
$$U_B = \exp \frac{\pi i \sigma_x}{4} \exp \frac{\pi i \sigma_y}{2},$$
$$V_A = \exp \frac{\pi i \sigma_y}{4},\; V_B = \exp - \frac{\pi i \sigma_y}{2}$$

$$\Longrightarrow \mathbf{CNOT} = (\mathbf{U_A} \otimes \mathbf{U_B}) \mathbf{U_d} (\mathbf{V_A} \otimes \mathbf{V_B})$$

The entangling power of a gate isn't affected by local unitary transformations. That means that U_d and the CNOT Gate are LOCALLY EQUIVALENT, *they share the* SAME ENTANGLING POWER!

LIKELIHOOD

$$L_G\left(\vec{\mathsf{J}}, \rho_{ab}^k\right) = \mathrm{Tr}\left\{G\rho_{ab}^k G^\dagger \left[\mathrm{Tr}_{net}\left(e^{-iH(\vec{\mathsf{J}})}(\rho_{ab}^k \otimes \rho_{net})e^{iH(\vec{\mathsf{J}})}\right)\right]\right\}$$

- ρ_{ab}^k is the initial density matrix of the qubits A and B. It is pure and it can be either separable and entangled.
- ρ_{net} is the initial density matrix of the rest of the network.

FIDELITY

We want to measure how close is our dynamics to implement a given gate G. It comes to rescue the following function

$$\mathcal{F}_G\left(\vec{\mathrm{J}}\right) = \frac{1}{d+1}\left(1 + \frac{1}{d}\sum_{ijkl} G^*_{ik}\mathcal{E}_{ijkl}G_{jl}\right) \tag{1}$$

- d is the dimension of the gate
- $\mathcal{E}_{ijkl} = \langle i_{ab}| \operatorname{Tr}_{net}\left[e^{-iH(\vec{\mathrm{J}})}\left(|k_{ab}\rangle\langle l_{ab}| \otimes |0_{net}\rangle\langle 0_{net}|\right)e^{iH(\vec{\mathrm{J}})}\right]|j_{ab}\rangle$
  - $|s_{ab}\rangle$ is a basis

Advantages

FIDELITY:

- It is an average over all possible initial states
- It depends only on $\vec{\mathrm{J}}$ and t

LIKELIHOOD:

- It is quite fast to compute
- It gives a better control over the sample set of the initial states

Drawbacks

FIDELITY:

- It requires loads of time as the dimensions increase
- NO local control over the initial states

LIKELIHOOD:

- ONLY local information
- The initial state belongs to a set of infinitely many items

Note

$$\mathcal{F}_G\left(\vec{\mathrm{J}}\right) = \int d\rho_{ab}\, L_G\left(\vec{\mathrm{J}}, \rho_{ab}\right) \sim \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} L_G\left(\vec{\mathrm{J}}, \rho_{ab}^k\right)$$

Stochastic Gradient Descent

What is Stochastic Gradient Descent?

Machine Learning Setup

1. *A sample:* $z = (\mathbf{x}, y)$
2. *Loss function:* $\ell(\hat{y}, y)$
3. *Family of functions:* $f \in \mathcal{F}$
4. *we seek* $f_\omega(\boldsymbol{x})$ *that minimizes:*

$$\ell(f_\omega(\mathbf{x}), y)$$

Usually one must deal with averages, i.e. we aim to minimize

$$E_n(f_\omega) = \frac{1}{n}\sum_{i=1}^{n} \ell(f_\omega(\mathbf{x}_i), y_i)$$

$$\omega_{t+1} = \omega_t - \gamma \nabla_\omega \frac{1}{n}\sum_{i=1}^{n} \ell(f_{\omega_t}(\mathbf{x}_i), y_i)$$

$$\downarrow$$

$$\omega_{t+1} = \omega_t - \gamma_t \nabla_\omega \ell(f_{\omega_t}(\mathbf{x}_r), y_r)$$

The former equation is a Newton-like algorithm: *Gradient Descent*.

The latter represents the ***Stochastic Gradient Descent***.

Instead of computing the gradient of $E_n(f)$ exactly, each iteration estimates this gradient on the basis of a single randomly picked example z_t

TWO (QUANTUM) INTERPRETATIONS

While it does not converge close to one:

- Generate a random pure density matrix ρ_{ab}
- Update: $\mathbf{J}_t \leftarrow \mathbf{J}_t + \gamma_t \nabla_{\mathbf{J}_t} L(\mathbf{J}_t, \rho_{ab})$

- Generate a random pure density matrix ρ_{ab}
- Generate an integer random number i between 1 and length of $\mathbf{J}$
- Update: $J_t[i] \leftarrow J_t[i] + \gamma_t \dfrac{\partial L(\mathbf{J}_t, \rho_{ab})}{\partial J_t[i]}$

```python
while Likelihood(J, rho_0) < 0.99  :

        rho_0 = rand_ket(N = 4, dims = [[2,2], [1,1]], pure = True)
        grad = Gradient_Likelihood(J,rho_0)
        J = [J[k] + grad[k]/sqrt(time) for k in range(len(J))]
```

THE CONVERGENCE OF SGD

CLASSICALLY

Convergence is guaranteed for learning rates satisfying

$$\sum_t \gamma_t^2 < \infty$$

$$\sum_t \gamma_t = \infty$$

Its speed is limited by the noisy approximation of the true gradient.

QUANTUM MECHANICALLY

1. The sample set is infinite and uncountable! We pick random samples.
2. The learning rate strongly depends on the size of the system
3. The optimal we found is

$$\gamma_t = t^{-1/2}$$

OPEN PROBLEM

Ideally t is a continuum variable. However, in order to implement an algorithm, it must be discretized.

We found out that

- The length of the step influences the convergence
- For different sizes one needs different lengths!

Question

How to pick the learning rates?
How to choose the step length?

WE ARE TESTING THE HESSIAN

$$\omega_{t+1} = \omega_t - \gamma_t \Gamma_t \nabla_\omega Q(\mathbf{z}_t, \omega_t)$$

Γ_t approaches the inverse of the Hessian. This is known as the *Second Order SGD*

Numerical Results

Our Starting point[1]

The Model

$H_M = \mathrm{J}\sum_{n=1}^{N} \sigma_n^x\sigma_{n+1}^x + \sigma_n^y\sigma_{n+1}^y + \lambda\sigma_n^z\sigma_{n+1}^z$
$H_I = \mathrm{J}_0\sum_{n=1,N} \sigma_n^x\sigma_{n+1}^x + \sigma_n^y\sigma_{n+1}^y + \lambda\sigma_n^z\sigma_{n+1}^z$

In the case of our interest $\lambda = 0$.

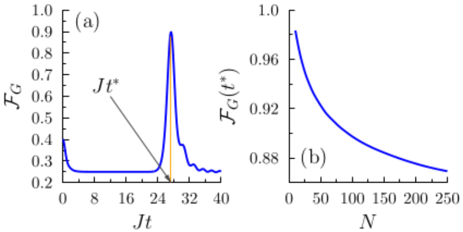
They found an optimal time t^* for a spin chain for which the Fidelity is maximized.

$$G_{AB} =$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & i & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

N ODD *N EVEN*

$$\mathcal{F}_G(t) = \int d\psi\, \langle\psi|\, G_{AB}^\dagger \mathcal{E}_t\, [|\psi\rangle\langle\psi|]\, G_{AB}\, |\psi\rangle$$



(a) Evolution of an average gate Fidelity for a chain of $N = 100$ and $\mathrm{J}_0 = 0.5\,\mathrm{J}$

(b) $\mathcal{F}_G(t^*)$ as a function of N. Insets shows the optimal time versus N

[1] Banchi, Bayat, Verrucchi, Bose, PRL (2011)

How to drop the Even/Odd concern

We hoped

$$G_{Even} = (U_A \otimes U_B) G_{Odd} (U_A^T \otimes U_B^T)$$

$\rightarrow$ they differ only by a change of basis.

Remind the Cartan Decomposition

$$U_{AB} = (U_A \otimes U_B) U_d (V_A \otimes V_B)$$

G_{Odd} and G_{Even} are *locally equivalent*; i.e. they share the same kernel U_d.

Since $\rho(t) = U(t)\rho_0 U^\dagger(t)$, the generalization of the Hamiltonian

$$H_{gen} = \sum_{ij} \overset{\downarrow}{\sum_{\alpha\beta}} \mathsf{J}_{ij}^{\alpha\beta} \sigma_i^\alpha \sigma_j^\beta$$

Allows any rotation: the implementation of both the gates

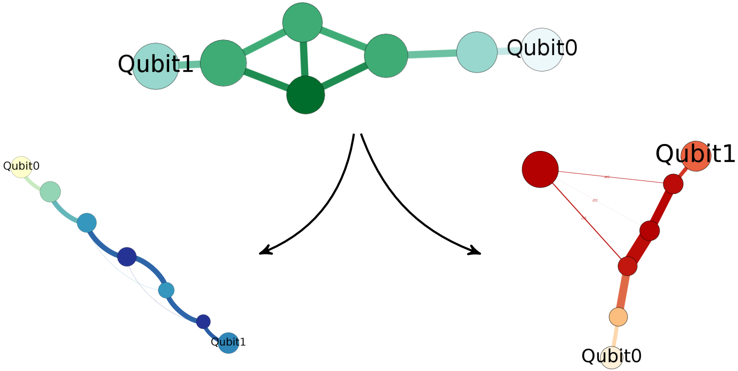
However[a]

$$G_{Even} = \left(\begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \right) \cdot G_{Odd} \cdot \mathbb{1}^{\otimes 2}$$

$\rightarrow$ that is not a change of basis.

[a] Zhang, Vala, Whaley, Sastry, Phys.Rev.A (2003)

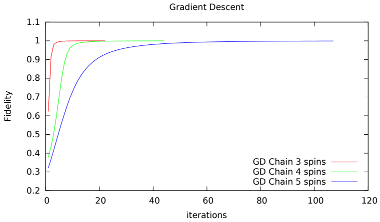
Introducing the Network

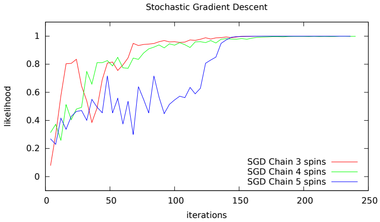


Odd Gate

Even Gate

SGD AT WORK (LIKELIHOOD VS FIDELITY)





	3 spins		4 spins		5 spins	
	time	iterations	time	iterations	time	iterations
GD Fidelity	01′ 09.13″	22	03′ 35.20″	44	11′ 41.15″	107
SGD Likelihood	00′ 12.10″	45	00′ 20.99″	36	00′ 54.14″	60

Note

1. The Gradient Descent is a *deterministic* algorithm. It means that, given an initial guess, its final destination is completely determined; i.e. it is more likely to converge towards a relative maximum.
2. We have tried several times to find the absolute maximum for 5 spins using random initial guess. We finally surrendered and used something close to the output of the SGD

SO FAR, SO GOOD!

What if we try

- ✓ Three body gates:
  - Toffoli
  - CCZ
- ✓ Four body gates:
  - ZZZZ
  - ...

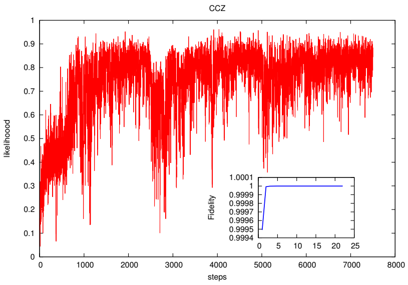
Possible Issues

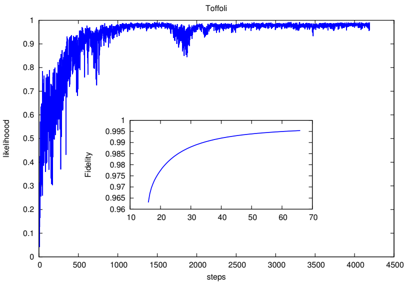
- ✗ Add complexity to the network
  - add nodes
  - add interactions
- ✗ Toffoli & CCZ involve 3-body interactions
- ✗ We have many choices for designing the network

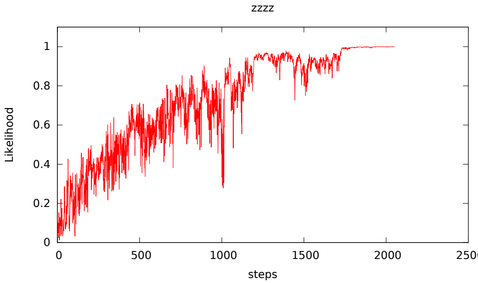
Note that one can always decompose any 3-body gate G by checking

$$i\,\mathrm{Tr}\left[(\sigma_i \otimes \sigma_j \otimes \sigma_k)\log G\right] \neq 0 \;\; \forall\, i,j,k \in (0,x,y,z)$$

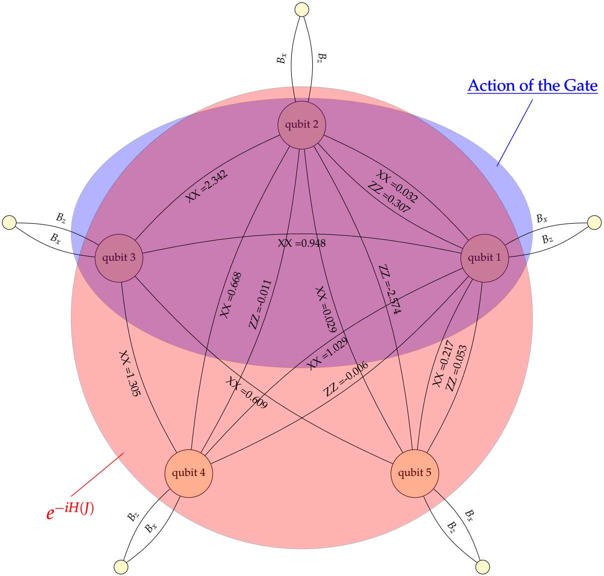
$$G = e^{-iH'}$$


CCZ
likelihood
steps
Fidelity


Toffoli
likelihood
steps
Fidelity


zzzz
Likelihood
steps

3-body gates: Optimal Toffoli



Summary & Conclusions

Targets

1. Universal Quantum Computation with Local Hamiltonians.
2. Entanglement creation over long distances.
3. Contact point between Quantum Mechanics and Machine Learning.

Methods

- *ML: Stochastic Gradient Descent*
- *QM: Many Body spin 1/2 systems*
- *Optimization over Hamiltonian parameters*
- *Quantum Dynamics*

POSSIBLE DEVELOPMENTS

- *Study larger networks and more complicated topologies*
- *Find the connection between the learning rates and the scaling of the degrees of freedom*
- *We accept suggestions!*

VIELEN DANK!