Grover's Algorithm applied to the Molecular Distance Geometry Problem

Carlile Lavor, Leo Liberti and Nelson Maculan

Abstract—Grover's algorithm is a quantum algorithm for searching in unstructured databases. Due to the properties of quantum mechanics, it provides a quadratic speedup over its classical counterparts. We present a new application of the algorithm to the Molecular Distance Geometry Problem. This problem is related to the determination of the tridimensional structure of a molecule based on the knowledge of some of the distances between pairs of atoms. This problem is NP-hard unless all possible inter-atomic distances are known.

Index Terms—Quantum Computation, Grover's algorithm, Molecular Distance Geometry Problem.

I. INTRODUCTION

$\mathbf{G}$ ROVER'S algorithm [2], [6], [7], [11], [14], [15] is a quantum algorithm for unstructured discrete search. Due to the properties of quantum mechanics, it provides a quadratic speedup over its classical counterparts.

We present a new application of Grover's algorithm to the Molecular Distance Geometry Problem (MDGP) [4], [5], [8], [9], [10]. This problem is related to the determination of the tridimensional structure of a molecule based on knowledge of some distances between pairs of atoms. Molecular conformation is very important because it is associated to the physical and chemical properties of the molecule.

The MDGP can be formalized as finding an immersion in $\mathbb{R}^3$ of a given undirected graph $G = (V, E)$, so it can be very naturally cast as a continuous search problem. Hence, to apply Grover's algorithm, we have to transform it to a discrete search problem. We show that this is indeed possible if all distances between atoms separated by three covalent bonds are known (using distance data obtained from the NMR experiments this assumption is very realistic [3], [17]). Naturally, distances between atoms separated by one and two covalent bonds can be obtained from the covalent bond lengths and bond angles. We shall call DMDGP (Discretizable Molecular Distance Geometry Problem) the class of all MDGP instances satisfying the above requirements.

In Section II, we show a discrete formulation for the DMDGP. Section III presents the unitary operators that will be used in Grover's algorithm. Section IV concludes the paper.

C. Lavor: Department of Applied Mathematics (IMECC-UNICAMP), State University of Campinas, CP 6065, 13081-970, Campinas-SP, Brazil. Email: `clavor@ime.unicamp.br`.

L. Liberti: DEI, Politecnico di Milano, Piazza L. da Vinci 32, 20133 Milano, Italy. Email: `liberti@elet.polimi.it`.

N. Maculan: COPPE, Universidade Federal do Rio de Janeiro – UFRJ, C.P. 68511, Rio de Janeiro 21945-970, Brazil. Email: `maculan@cos.ufrj.br`.

II. THE MOLECULAR DISTANCE GEOMETRY PROBLEM

Formally, the MDGP can be defined as the problem of finding Cartesian coordinates $x_1, \ldots, x_n \in \mathbb{R}^3$ of the atoms of a molecule such that for all $(i, j) \in S$,

$$||x_i - x_j|| = d_{ij},$$

where S is the set of pairs of atoms (i, j) whose Euclidean distances d_{ij} are known. If all distances are given, the problem can be solved in linear time [5]. Otherwise, the problem is NP-hard [16].

The MDGP is usually formulated as a continuous least-squares minimization problem, where the objective function is as follows:

$$g(x_1, \ldots, x_n) = \sum_{(i,j) \in S} (||x_i - x_j||^2 - d_{ij}^2)^2. \qquad (1)$$

Obviously, $(x_1^*, \ldots, x_n^*)$ solve the problem if and only if $g(x_1^*, \ldots, x_n^*) = 0$ (this will be very useful in defining Grover's oracle operator).

A. Discrete formulation of the DMDGP

In order to apply Grover's algorithm to the DMDGP, we need to exhibit a finite number of candidate solutions. In this section, we give a discrete formulation of the DMDGP which serves this purpose.

Consider a molecule as being a sequence of n atoms with Cartesian coordinates given by $x_1, \ldots, x_n \in \mathbb{R}^3$ and such that there is a covalent bond between every pair of atoms $(i - 1, i)$, for $i = 2, \ldots, n$. For simplicity, we restrict our attention to the class of molecules where each atom has at most two covalent bonds. However, the same ideas can be applied to all molecules in the DMDGP class with small modifications. First, some definitions. The bond length $r_{i-1,i}$ is the Euclidean distance between atoms $i - 1$ and i (for all $i = 2, \ldots, n$). The bond angle $\theta_{i-2,i}$ is the angle between the segments joining atoms $i - 2, i - 1$ and $i - 1, i$ (for all $i = 3, \ldots, n$). The torsion angle $\omega_{i-3,i}$ is the angle between the normals through the planes defined by the atoms $i - 3, i - 2, i - 1$ and $i - 2, i - 1, i$ (for each $i = 4, \ldots, n$).

In most molecular conformation calculations, all covalent bond lengths and bond angles are assumed to be known *a priori*. Thus, the first three atoms in the chain can be fixed and the fourth atom can be determined by the torsion angle ω_{14}. The fifth atom can be determined by the torsion angles ω_{14} and ω_{25}, and so on. So, given all bond lengths $r_{12}, r_{23}, \ldots, r_{n-1,n}$, bond angles $\theta_{13}, \theta_{24}, \ldots, \theta_{n-2,n}$, and torsion angles $\omega_{14}, \omega_{25}, \ldots, \omega_{n-3,n}$ of a molecule with n atoms, the Cartesian coordinates $(x_{a_1}, x_{a_2}, x_{a_3})$ for each

atom a in the molecule can be obtained using the following formulae [12]:

$$\begin{bmatrix} x_{a_1} \\ x_{a_2} \\ x_{a_3} \\ 1 \end{bmatrix} = B_1 B_2 \cdots B_a \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad \forall\, a = 1, \ldots, n, \quad (2)$$

where

$$B_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad B_2 = \begin{bmatrix} -1 & 0 & 0 & -r_{12} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$B_3 = \begin{bmatrix} -\cos\theta_{13} & -\sin\theta_{13} & 0 & -r_{23}\cos\theta_{13} \\ \sin\theta_{13} & -\cos\theta_{13} & 0 & r_{23}\sin\theta_{13} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

and for $i = 4, \ldots, n$, B_i is as shown in Fig. 1.

For every four consecutive atoms x_i, x_j, x_k, x_l we can express the cosine of the torsion angle ω_{il} in terms of the distances $r_{ij}, r_{jk}, r_{kl}, r_{jl}, r_{il}$ and the bond angle θ_{ik} by using the cosine law for torsion angles [13] (p. 278), as follows:

$$\cos(\omega_{il}) = \frac{r_{ij}^2 + r_{jl}^2 - r_{ij}\left(\frac{r_{jl}^2 + r_{jk}^2 - r_{kl}^2}{r_{jk}}\right)\cos(\theta_{ik}) - r_{il}^2}{r_{ij}\left(\frac{\sqrt{4r_{jl}^2 r_{jk}^2 - (r_{jl}^2 + r_{jk}^2 - r_{kl}^2)^2}}{r_{jk}}\right)\sin(\theta_{ik})}. \quad (3)$$

Hence, if we know all the bond lengths ($r_{i-1,i}$), bond angles ($\theta_{i-2,i}$), and distances between atoms separated by three covalent bonds ($r_{i-3,i}$), we can calculate the cosine of the torsion angles defined by the atoms $i-3, i-2, i-1, i$, for $i = 4, \ldots, n$. We note in passing that in order for the above reasoning to hold, we obviously need the denominator of (3) to be nonzero.

Using the bond lengths r_{12}, r_{23} and the bond angle θ_{13}, we can determine the matrices B_2 and B_3 and obtain

$$x_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

$$x_2 = \begin{pmatrix} -r_{12} \\ 0 \\ 0 \end{pmatrix},$$

$$x_3 = \begin{pmatrix} r_{23}\cos(\theta_{13}) - r_{12} \\ r_{23}\sin(\theta_{13}) \\ 0 \end{pmatrix},$$

fixing the first three atoms of the molecule. Since we also know the distance r_{14}, by (3) we can obtain the value $\cos(\omega_{14})$. Thus, the sine of the torsion angle ω_{14} can have only two possible values: $\sin(\omega_{14}) = \pm\sqrt{1 - \cos^2(\omega_{14})}$. Consequently, we obtain only two possible positions x_4, x_4' for the fourth atom, as shown in Fig. 2. This dichotomy, by the way, is the basic reason why the DMDGP can be formulated combinatorially.

For the fifth atom, we will obtain four possible positions, one for each combination of $\pm\sqrt{1 - \cos^2(\omega_{14})}$ and

$\pm\sqrt{1 - \cos^2(\omega_{25})}$. By an easy induction argument, we can see that for the i-th atom we obtain 2^{i-3} possible positions. So, for a molecule shaped as a linear chain of n atoms, we get 2^{n-3} possible sequences of torsion angles $\omega_{14}, \omega_{25}, \ldots, \omega_{n-3,n}$ of length $n - 3$, each defining a different tridimensional structure. By using the matrices B_i defined above, we can convert a sequence of torsion angles into Cartesian coordinates $x_1, \ldots, x_n \in \mathbb{R}^3$ and, using the function g defined in (1), we can identify a solution for the problem (i.e. $x_1, \ldots, x_n$ solve the problem if and only if $g(x_1, \ldots, x_n) = 0$). In conclusion, based on the above arguments, the DMDGP can be considered as a discrete search problem.

III. Application of Grover's Algorithm to the DMDGP

Grover's algorithm uses two quantum registers and two unitary operators. The first one is defined by

$$U_f\left(|i\rangle\,|j\rangle\right) = |i\rangle\,|j \oplus f(i)\rangle,$$

where $|i\rangle$ is a state of the first register ($i \in \{0, \ldots, N-1\}$), $|j\rangle$ is a state of the second register ($j \in \{0,1\}$), $\oplus$ is the sum modulo 2, and f is an oracle function $f : \{0, \ldots, N-1\} \to \{0,1\}$ which "recognizes" the searched element i_0. f is defined by:

$$f(i) = \begin{cases} 1, & \text{if } i \text{ is the index } i_0 \text{ of the searched element} \\ 0, & \text{otherwise.} \end{cases}$$

By using g as given in (1), we define a function f that can be used in Grover's algorithm. First of all, we have to associate a state in the set $\{|0\rangle, \ldots, |N-1\rangle\}$ to each possible candidate solution for the DMDGP. Recall that we are considering a molecule with n atoms and, in this case, $N = 2^{n-3}$. Hence, for the i-th qubit ($i = 1, \ldots, n-3$) of the first register, the state $|0\rangle$ is associated to $\sin(\omega_{i,i+3}) = \sqrt{1 - \cos^2(\omega_{i,i+3})}$ and the state $|1\rangle$ to $\sin(\omega_{i,i+3}) = -\sqrt{1 - \cos^2(\omega_{i,i+3})}$. For example, for a molecule with 5 atoms, we need 2 qubits for the first register and we shall have 4 possible solutions:

$$|0\rangle = |00\rangle \quad (\sin(\omega_{14}) = +\sqrt{1 - \cos^2(\omega_{14})} \text{ and}$$
$$\sin(\omega_{25}) = +\sqrt{1 - \cos^2(\omega_{25})}),$$
$$|1\rangle = |01\rangle \quad (\sin(\omega_{14}) = +\sqrt{1 - \cos^2(\omega_{14})} \text{ and}$$
$$\sin(\omega_{25}) = -\sqrt{1 - \cos^2(\omega_{25})}),$$
$$|2\rangle = |10\rangle \quad (\sin(\omega_{14}) = -\sqrt{1 - \cos^2(\omega_{14})} \text{ and}$$
$$\sin(\omega_{25}) = +\sqrt{1 - \cos^2(\omega_{25})}),$$
$$|3\rangle = |11\rangle \quad (\sin(\omega_{14}) = -\sqrt{1 - \cos^2(\omega_{14})} \text{ and}$$
$$\sin(\omega_{25}) = -\sqrt{1 - \cos^2(\omega_{25})}).$$

Both g as well as the map given by (2), denoted by h, can be efficiently computed using classical circuits, so we can also obtain their respective quantum versions [1].

Finally, the function f that will be used in Grover's algorithm for identifying a solution for the DMDGP can be

$$B_i = \begin{bmatrix} -\cos\theta_{i-2,i} & -\sin\theta_{i-2,i} & 0 & -r_{(i-1),i}\cos\theta_{i-2,i} \\ \sin\theta_{i-2,i}\cos\omega_{i-3,i} & -\cos\theta_{i-2,i}\cos\omega_{i-3,i} & -\sin\omega_{i-3,i} & r_{(i-1),i}\sin\theta_{i-2,i}\cos\omega_{i-3,i} \\ \sin\theta_{i-2,i}\sin\omega_{i-3,i} & -\cos\theta_{i-2,i}\sin\omega_{i-3,i} & \cos\omega_{i-3,i} & r_{(i-1),i}\sin\theta_{i-2,i}\sin\omega_{i-3,i} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \forall\, i = 4,\ldots,n.$$

Fig. 1. The matrix B_i for $i = 4,\ldots,n$.

$$x_4 = \begin{bmatrix} -r_{12} + \cos(\theta_{23})r_{23} - \cos(\theta_{23})\cos(\theta_{34})r_{34} + \sin(\theta_{23})\sin(\theta_{34})r_{34}\cos(\omega_{14}) \\ \sin(\theta_{23})r_{23} - \sin(\theta_{23})\cos(\theta_{34})r_{34} - \cos(\theta_{23})\sin(\theta_{34})r_{34}\cos(\omega_{14}) \\ -\sin(\theta_{34})r_{34}\sqrt{1-\cos^2(\omega_{14})} \end{bmatrix},$$

$$x_4' = \begin{bmatrix} -r_{12} + \cos(\theta_{23})r_{23} - \cos(\theta_{23})\cos(\theta_{34})r_{34} + \sin(\theta_{23})\sin(\theta_{34})r_{34}\cos(\omega_{14}) \\ \sin(\theta_{23})r_{23} - \sin(\theta_{23})\cos(\theta_{34})r_{34} - \cos(\theta_{23})\sin(\theta_{34})r_{34}\cos(\omega_{14}) \\ \sin(\theta_{34})r_{34}\sqrt{1-\cos^2(\omega_{14})} \end{bmatrix}.$$

Fig. 2. The two possible positions for the 4th atom. The difference is in the sign of the third component.

given by

$$f(i) = 1 - \left\lfloor \left(\frac{g(h(i))}{p_1}\right)^{\frac{1}{p_2}} + 0.5 \right\rfloor,$$

where $i \in \{0,\ldots,N-1\}$ and p_1, p_2 are two parameters. For sufficiently large p_1 and p_2, which depend on the number of atoms n of the molecule, we have

$$\frac{g(h(i))}{p_1} \in [0,1],$$

and thus,

$$\left(\frac{g(h(i))}{p_1}\right)^{\frac{1}{p_2}}$$

will take values very close to 1, except when i is associated to a solution of the problem. In other words,

$$f(i) = \begin{cases} 1, & \text{if } i \text{ is the index } i_0 \text{ of the searched element} \\ 0, & \text{otherwise,} \end{cases}$$

as required in Grover's algorithm. The application of Grover's algorithm in this setting is now rather straightforward. We summarize it below.

The first step of the algorithm is to create a superposition of all N computational basis states $\{|0\rangle,\ldots,|N-1\rangle\}$ of the first register. This is achieved initializing the first register in the state $|0\rangle\cdots|0\rangle$ and applying the Hadamard operator H,

$$H = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

on each qubit $|0\rangle$. We obtain the state $|\psi\rangle$, given by

$$|\psi\rangle = \frac{1}{\sqrt{N}}\sum_{i=0}^{N-1}|i\rangle.$$

The second register begins with $|1\rangle$ and, after applying again the Hadamard operator, it will be in state $|-\rangle$, given by

$$|-\rangle = H|1\rangle = \frac{|0\rangle - |1\rangle}{\sqrt{2}}.$$

Using the fact that

$$1 \oplus f(i) = \begin{cases} 0 \text{ for } i = i_0 \\ 1 \text{ for } i \neq i_0, \end{cases}$$

it is easy to check that

$$U_f\left(|i\rangle\,|-\rangle\right) = (-1)^{f(i)}|i\rangle\,|-\rangle.$$

Using this result and applying now U_f to the superposition state coming from the first step, we obtain the state:

$$|\psi_1\rangle\,|-\rangle = U_f\left(|\psi\rangle\,|-\rangle\right)$$
$$= \left(\frac{1}{\sqrt{N}}\sum_{i=0}^{N-1}(-1)^{f(i)}|i\rangle\right)|-\rangle.$$

($|\psi_1\rangle$ is also a superposition of all basis elements, but the searched element has been marked with a minus sign).

Before we can perform a measure, the next step should be to increase the amplitude of the searched element while decreasing the amplitude of the others. This is obtained using the second unitary operator defined by $(2|\psi\rangle\langle\psi| - I)$. The state resulting from the application of the operator $(2|\psi\rangle\langle\psi| - I)$ to $|\psi_1\rangle$ is given by:

$$|\psi_G\rangle = (2|\psi\rangle\langle\psi| - I)|\psi_1\rangle$$
$$= \sum_{i=0,i\neq i_0}^{N-1}\left(\frac{N-4}{N\sqrt{N}}|i\rangle\right) + \frac{3N-4}{N\sqrt{N}}|i_0\rangle.$$

Repeating this procedure, i.e., applying the operator $((2|\psi\rangle\langle\psi| - I)\otimes I)\,U_f$ exactly $\left\lfloor \frac{\arccos\left(\frac{1}{\sqrt{N}}\right)}{\arccos\left(\frac{N-2}{N}\right)} \right\rfloor$ times and measuring the first register, we will obtain the desired solution with the highest probability in $O(\sqrt{N})$ steps [6], [7], [14].

IV. Final Remarks

We have presented a new application of Grover's algorithm to the Molecular Distance Geometry Problem. In order to achieve this, we first provided a discrete formulation of the problem and then defined Grover's oracle function.

It is known that for an unstructured search, the complexity $O(\sqrt{N})$ of Grover's algorithm cannot be improved [2]. Research is ongoing towards finding further structural properties of the DMDGP which might help design a new quantum algorithm more efficient than Grover's.

Acknowledgments

References

[1] Bennett, C.H, Bernstein, E., Brassard, G., and Vazirani, U. (1997), Strengths and weaknesses of quantum computing, *SIAM Journal on Computing* **26**:1510-1523.

[2] Boyer, M., Brassard, G., Hoyer, P., and Tapp, A. (1998), Tight bounds on quantum search, *Forstschritte Der Physik* **4**:820-831, `quant-ph/9605034`.

[3] Creighton, T.E. (1993), *Proteins: Structures and Molecular Properties*. W.H. Freeman and Company, New York.

[4] Crippen, G.M. and Havel, T.F. (1988), *Distance Geometry and Molecular Conformation*, John Wiley & Sons, New York.

[5] Dong, Q. and Wu, Z. (2002), A linear-time algorithm for solving the molecular distance geometry problem with exact interatomic distances, *Journal of Global Optimization*, **22**:365-375.

[6] Grover, L.K. (1996), A fast quantum mechanical algorithm for database search, *Proc. of 28th Annual ACM Symposium on the Theory of Computing (STOC)*, 212-219.

[7] Grover, L.K. (1997), Quantum mechanics helps in searching for a needle in a haystack, *Physical Review Letters* **79**:325-328.

[8] Hendrickson, B.A. (1995), The molecule problem: exploiting structure in global optimization, *SIAM Journal on Optimization*, **5**:835-857.

[9] Moré, J.J. and Wu, Z. (1997), Global continuation for distance geometry problems, *SIAM Journal on Optimization*, **7**:814-836.

[10] Moré, J.J. and Wu, Z. (1999), Distance geometry optimization for protein structures, *Journal of Global Optimization*, **15**:219-234.

[11] Nielsen, M.A. and Chuang, I.L. (2000), *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge.

[12] Phillips, A.T., Rosen, J.B., and Walke, V.H. (1996), Molecular structure determination by convex underestimation of local energy minima, *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, **23**:181-198, American Mathematical Society, Providence.

[13] Pogorelov, A. (1987), *Geometry*, Mir Publishers, Moscow.

[14] Portugal, R., Lavor, C., Carvalho, L.M., and Maculan, N. (2004), *Uma Introdução à Computação Quântica*, Notas em Matemática Aplicada, vol. 8, SBMAC, São Carlos – SP.

[15] Preskill, J. (1998), Quantum Information and Computation, Lecture Notes, California Institute of Technology, `http://www.theory.caltech.edu/people/preskill/ph229`.

[16] Saxe, J.B. (1979), *Embeddability of weighted graphs in k-space is strongly NP-hard*, Proc. of 17th Allerton Conference in Communications, Control, and Computing, Monticello, IL, 480-489.

[17] Schlick, T. (2002), *Molecular modelling and simulation: an interdisciplinary guide*, Springer, New York.