

# Ajuste de Curvas de Energia Potencial de Interações Atômicas e Moleculares via Algoritmo Genético

João Augusto Sobral da Silva (IC)<sup>1\*</sup> e Luiz Antonio Ribeiro Junior (PQ)<sup>1</sup>

\*joaoaugustosds@gmail.com

<sup>1</sup>Instituto de Física, Universidade de Brasília

**Resumo:** Neste trabalho foram estudadas técnicas heurísticas de otimização global utilizando algoritmos genéticos no âmbito de interações atômicas e moleculares. Populações de indivíduos foram criadas e submetidas a operadores genéticos de forma a simular um processo de evolução natural, gerando indivíduos que se adequassem aos extremos da função de avaliação, sendo esta representada pela função analítica generalizada de Rydberg. Por fim, o algoritmo foi testado para o sistema diatômico simples  $H_2^+$ .

**Palavras-chave:** Algoritmos Genéticos; Otimização; Curva de Energia de Potencial (CEP).

## I. Introdução

Uma busca tem como principal objetivo encontrar a melhor solução dentre todas as soluções possíveis – o chamado espaço de soluções. Algoritmos genéticos são algoritmos de busca baseados nos mecanismos de seleção natural e genética, ou seja, em conceitos como de hereditariedade, reprodução e mutação [1]. Essa técnica quase não necessita de interferência humana e encontra soluções próximas das soluções ótimas, necessitando apenas de uma forma de avaliação do resultado. Além disso, esse tipo de algoritmo não possui a necessidade de se utilizar a derivada da função avaliada, o que seria extremamente complicado ao se tratar de funções descontínuas em determinados pontos.

A Figura 1 ilustra o funcionamento básico de um algoritmo genético. Inicialmente é gerada uma população com um determinado número de indivíduos e esses são avaliados através de uma função de avaliação. Em seguida, todos os indivíduos ganham uma probabilidade de serem selecionados para reproduzirem a partir de suas avaliações. Os indivíduos selecionados geram novos por um determinado método de reprodução. Visando uma maior variabilidade genética nas populações, operadores genéticos como os de mutação são aplicados em uma porcentagem da população. Após

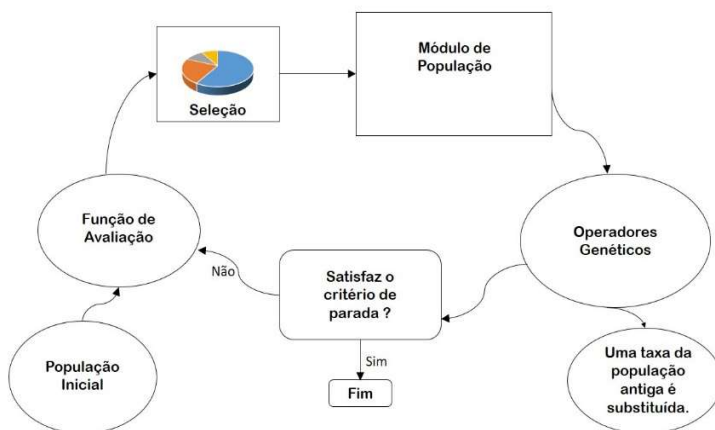


Figura 1 – Estrutura básica de um Algoritmo Genético (GA).

isso, uma taxa da população antiga é substituída pelos novos indivíduos, desses seleciona-se o melhor, que será utilizado como critério de parada caso seja encontrado, do contrário, todo o processo do loop é repetido.

Dentre várias aplicações possíveis utilizando algoritmos genéticos, interessa nesse trabalho o estudo de curvas de energia potencial da interação de sistemas atômicos para as quais as energias eletrônicas são obtidas por cálculos *ab initio*[4]. Em um sistema diatômico, por exemplo, o poço de

energia potencial, como mostrado na Figura 2, reúne diversas informações sobre o sistema atômico-molecular; dentre elas, importantes características espectroscópicas, cujos valores são de importância essencial em campos da química e física atômica e molecular como, por exemplo, a astroquímica.

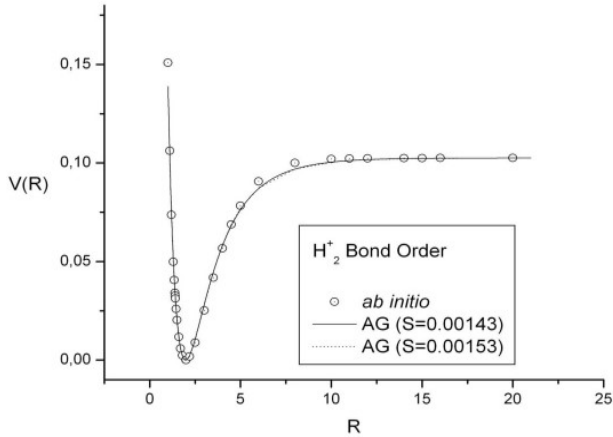


Figura 2 - Poço de energia potencial para o sistema diatômico  $H_2^+$ , utilizando a forma funcional Bond Order (BO)[5].  $V(R)$  representa a energia potencial no sistema dependente da distância  $R$  entre os átomos. Os valores em  $S$  indicam parâmetros de ajuste do método computacional utilizado. Fonte: [5], 2006, p.71.

No contexto de interações atômicas e moleculares, podemos possuir sistemas com estados ligados ou não ligados. Dado um potencial em uma molécula, um estado ligado é responsável por manter a ligação entre H e H estável em um sistema como o  $H_2^+$ , por exemplo. Diversos estudos à respeito de sistemas diatômicos tentam descrever o potencial, ou energia de um sistema, em termos da distância entre os átomos da molécula. Os mais utilizados na literatura são dados através das formas funcionais Bond Order (BO) e de Rydberg (Ryd)[4-5]:

$$V^{BO}([a], \rho, \beta) = a_1 e^{-\beta \rho} + a_2 e^{(-\beta)^2} + a_3 \exp(-\beta \rho)^3 \quad (1)$$

$$V^{Ryd}([a], \rho) = -De(1 + a_1 \rho + a_2 \rho^2 + a_3 \rho^3) e^{-a_1 \rho} \quad (2)$$

Onde para as expansões de terceira ordem (1) e (2) temos os parâmetros  $a$  e  $\beta$  à serem determi-

nados pelo AG,  $De$  a energia de dissociação do sistema e  $\rho = r - r_{eq}$ , onde  $r$  indica a distância entre os átomos e  $r_{eq}$  a distância de equilíbrio do sistema diatômico. Dentre o estudo dessas formas funcionais, existem os chamados sistemas estendidos, como o de Rydberg-Q (qRyd)[4]:

$$V^{qRyd}([a], \rho) = -De^{qRyd}(1 + a_1 \rho + a_2 \rho^2 + a_3 \rho^3) \exp_q(-a_1 \rho) \quad (3)$$

$$\exp_q(r) \equiv [1 + (1 - q)r]^{\frac{1}{1-q}} \quad (4)$$

A principal diferença de (3) para (2) está na definição de (4), para  $q$  vindo da estatística de Tsallis[9].

Sendo assim, esse projeto tem como objetivo o desenvolvimento de um algoritmo genético capaz de encontrar os extremos globais e locais de uma determinada função e de testar o mesmo em um sistema físico simples, como o sistema diatômico  $H_2^+$ .

## II. Seção Experimental

A primeira subrotina “População Inicial” foi criada através da função `ran1(idum)`[2], gerando uma certa quantidade de indivíduos em um intervalo específico. A função `ran1` é responsável por gerar um número entre 0 e 1. Para os efeitos do algoritmo a seguinte equação foi utilizada:

$$X(i) = \text{ran1}(\text{idum}) \cdot (\text{IntF} - \text{IntI}) + \text{IntI} \quad (1)$$

Onde `IntF` é último número e `IntI` o primeiro número do intervalo escolhido.

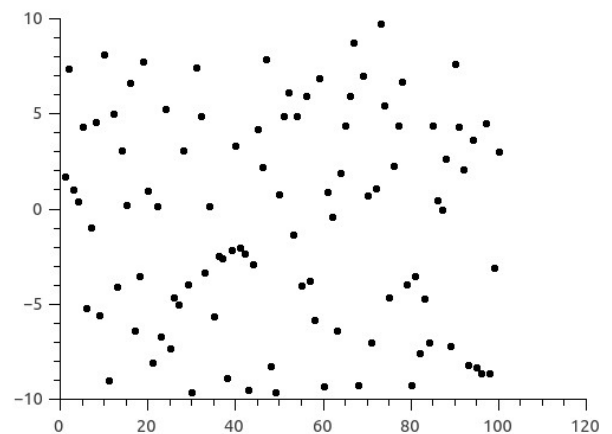


Figura 3 - 100 Indivíduos gerados de forma aleatória entre -10 e 10 em uma função teste.

Todos os indivíduos foram analisados pela função de avaliação. A subrotina “Ordenar” foi estruturada de modo a ordenar todas as avaliações e indivíduos de forma crescente utilizando o método de ordenação Bubblesort[2].

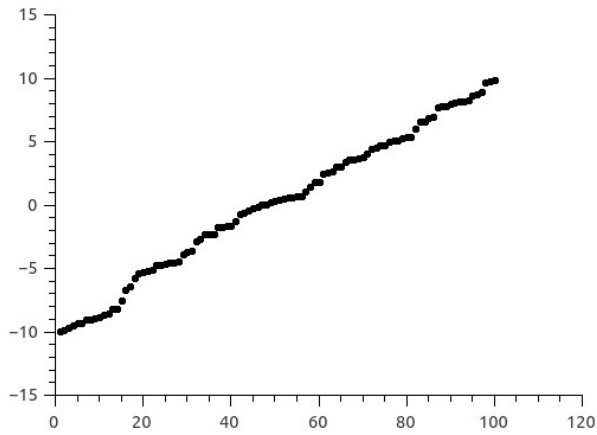


Figura 5 - Indivíduos ordenados pelo método Bubblesort.

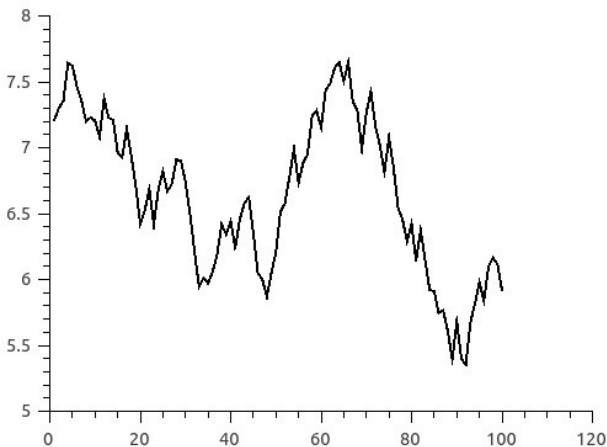


Figura 4 - Random Walk para 100 passos.

O método da roleta viciada atribui a cada indivíduo uma porcentagem de ser selecionado e parte de uma roleta virtual. Esta probabilidade pode ser obtida através da seguinte expressão:

$$prob(i) = \frac{Y(i)}{\sum_{i=1}^n Y(i)} \quad (2)$$

Onde  $prob(i)$  indica a probabilidade do indivíduo ser escolhido,  $Y(i)$  sua avaliação,  $\sum_{i=1}^n Y(i)$  a soma de todas as avaliações e “n” o número de indivíduos em cada geração.

Para indivíduos com uma avaliação negativa o termo no denominador da expressão em (2) não faria sentido. Uma subrotina chamada “Translação” foi criada com o intuito de evitar esse problema na seleção com a seguinte equação:

$$T(i) = Y(i) - Y(1) \quad (3)$$

Onde  $T(i)$  representa a avaliação do indivíduo transladada,  $Y(i)$  a avaliação do indivíduo e  $Y(1)$  a avaliação do primeiro indivíduo.

Como a subrotina “Ordenar” organizou todos os indivíduos e avaliações de forma crescente, o primeiro indivíduo possui a menor avaliação, portanto, cada indivíduo possui uma avaliação maior ou igual a 0 na subrotina “translação”.

Fenômenos de convergência genética, ou convergência prematura, ocorrem quando o algoritmo converge de forma muito rápida para um máximo ou mínimo local. Para se evitar problemas como este, após transladadas, todas as avaliações foram enviadas para uma subrotina chamada “Linearização”, onde uma nova espécie de função de avaliação foi criada de modo a evitar que os melhores indivíduos fossem sempre selecionados.

Os indivíduos foram então enviados para a subrotina “seleção” onde o método da roleta viciada [1] foi utilizado.

Os pais escolhidos para reprodução foram enviados para a subrotina “Reprodução” onde o método Gaussian Random Walk[3] foi utilizado. O método random walk consiste em uma caminhada aleatória dentre um intervalo especificado - no caso os dois pais escolhidos - o último indivíduo da caminhada é sempre o escolhido como um dos filhos. Para um conjunto de passos e de indivíduos muito grande na caminhada, o gráfico deste método se aproxima ao de uma distribuição normal. Quanto maior a caminhada, maior a largura da distribuição.

Na subrotina “Atribuição” os filhos gerados substituíram metade dos indivíduos da geração de seus pais, os indivíduos com a avaliação mais baixa. Utilizando a técnica de elitismo[1], o melhor indivíduo de cada geração permanece nas próximas gerações, tendo mais chances de ser escolhido para reproduzir.

Por fim, todos estes indivíduos sofreram uma espécie de mutação: 5 por cento dos indivíduos foram selecionados para sofrerem um acréscimo ou decréscimo infinitesimal utilizando o método random walk[3].

A Figura 5 ilustra o funcionamento do programa como um todo, os loops foram efetuados até o critério de parada ser alcançado.

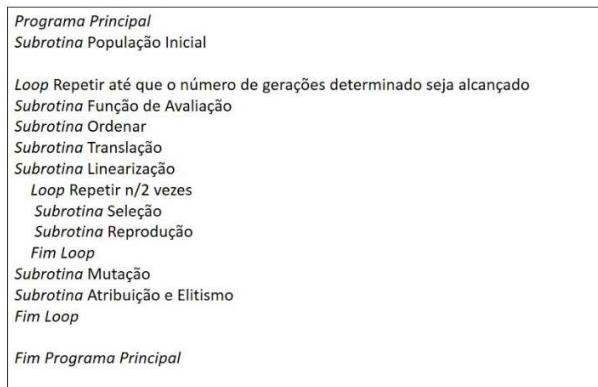


Figura 6 - Estrutura Principal do Algoritmo.

### III. Resultados e Discussão

O algoritmo teve resultados satisfatórios para as 3 funções teste, como demonstra a Tabela 1. Os valores encontrados para os extremos nas funções teste 2 e 3 tiveram uma precisão de pelo menos 7 casas decimais. O tempo de execução do pior caso foi extremamente baixo, portanto, o algoritmo consegue desempenhar o papel pelo qual foi desenvolvido de forma tangível.

A função teste 2 possui um máximo em -1,73 e um mínimo em 1,73. Após este ponto a função cresce para infinito positivo. Ao encontrar o ponto de máximo da função o algoritmo continuou procurando por outros candidatos a máximo, ou seja, os valores de x cresceram cada vez mais, tendendo ao infinito. Em casos de convergência genética o algoritmo ficaria preso nesse máximo local.

A função teste 3 apresenta um comportamento periódico. Os valores apresentados foram calculados somente para valores maiores que 0. Em outros testes, máximos e mínimos para x negativoS foram encontrados da mesma forma.

Com o intuito de se testar uma das formas extendidas de Rydberg em um sistema físico, o algoritmo genético foi executado para o sistema  $H_2^+$  - Figura 8.

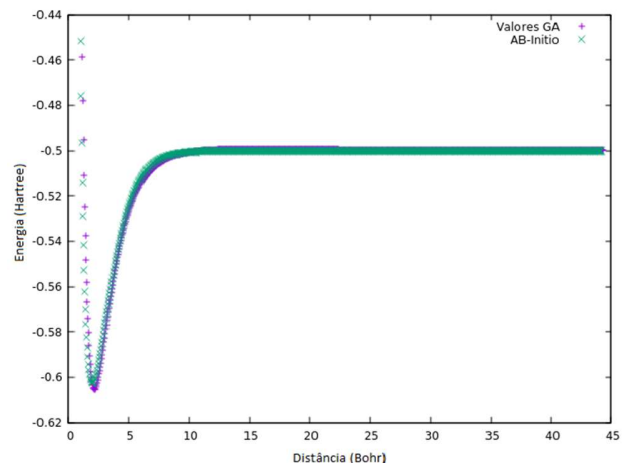


Figura 7- Ajuste para o potencial do sistema  $H_2^+$  através da forma extendida de q-Ryd utilizando-se o algoritmo genético. O traço em roxo indica a curva esperada através de cálculos AB-Initio, e a curva em verde representa os resultados encontrados pelo algoritmo.

### IV. Conclusão

Todas as soluções encontradas são satisfatórias ou estão muito perto das soluções ótimas. No caso do sistema físico  $H_2^+$ , o desvio quadrático mínimo encontrado entre as duas curvas foi de  $5,42 \times 10^{-4} \text{ Kcal/Mol}$ . Para resultados mais precisos, esse resultado deve estar na ordem de  $10^{-5}$  ou ainda menores, como indicam resultados mais recentes na literatura [4].

Função Teste	Ponto de Máximo Valor Esperado	Ponto de Máximo Valor Encontrado	Ponto de Mínimo Valor Esperado	Ponto de Mínimo Valor Encontrado
$x^2$	-	-	0	0
$x^3 - 9x$	-1,73205080756	-1,73205081221	1,73205080756	1,73205080605
$e^{\cos(x)}$	6,28318530717	6,28318530307	3,14159265358	3,14159265207

Tabela 1 - Resultados do Algoritmo para algumas funções teste

Apesar disso, em todos os resultados o tempo de execução se encontrou dentro do intervalo esperado.

Os objetivos do projeto foram alcançados tendo em vista que o algoritmo respondeu bem a diversas funções teste e a um sistema físico simples. Possíveis aperfeiçoamentos para o algoritmo poderiam ser feitos através de técnicas de programação paralela, com o intuito de se diminuir o tempo de execução, e para uma maior precisão os operadores genéticos do algoritmo poderiam ser melhorados, utilizando métodos mais recentes da literatura.

## V. Agradecimentos

Os autores são gratos pelo suporte dado pela Fundação Universidade de Brasília (FUB) e pela ajuda e apoio do Prof. Dr. Ricardo Gargano.

## VI. Referências

[1] R. Linden; *Algoritmos Genéticos*, **2006**.

[2] S. Luke; *Essentials Of Metaheuristics*, **2015**.

[3] S. Teukolsky; W. Vetterling; W. Press, B. Flannery; *Numerical Recipes: Example Book (FORTRAN)*, **1986**.

[4] C.S. Esteves; H.C.B. de Oliveira; L. Ribeiro; R. Gargano; K.C. Mundim; *Chem. Phy. Lett.* **2006**, 427, 10.

[5] L. F. Roncaratti Jr: *Um Algoritmo Genético para o Ajuste de Superfícies de Energia Potencial de Sistemas Moleculares*, **2006**.

[6] E. Herbst; J. T.Yates; *Chem. Rev.* **2013**, 113, 12.

[7] L. R. Nyhoff; S. Leestma; *Introduction to Fortran 90 for Engineers and Scientists*, **1997**

[8] Universidade Estadual de Campinas; *Introdução ao OpenMp*, **2014**.

[9] C. Tsallis, J. State Phys. **1988**, 479, 52.

[10] M. B. Arouxét; N. Echebest; E. A. Pilotta; *Comput. Appl. Math.* **2011**, 30, 1.

[11] Gargano, R. (2017). *Modelagem Molecular Sob o Olhar da Física*.

---