



## O problema da distância geométrica intervalar com dados esparsos via otimização global

**Luiz Leduino de Salles-Neto**

Universidade Federal de São Paulo

Av. Cesar Lattes, 1211, Parque Tecnológico, São José dos Campos

luiz.leduino@unifesp.br

**Weldon Lodwick**

University of Colorado Denver

Student Commons Building, Suite 4000, 1201 Larimer Street, Denver, CO, USA

Weldon.Lodwick@ucdenver.edu

### RESUMO

Neste trabalho apresentamos uma nova modelagem matemática, via otimização global e função intervalar, para o problema da distância geométrica intervalar com dados esparsos (PDGiE). O objetivo do problema da distância geométrica intervalar, no contexto da estrutura de proteínas, consiste em encontrar uma realização em  $R^3$  de um grafo  $G=(V,E)$ , onde as distâncias entre alguns vértices são dadas por intervalos, em conformidade com as medidas experimentais obtidas pela Ressonância Magnética Nuclear (RMN). Os dados esparsos também é uma realidade oriunda das características da RMN. Em particular, esse trabalho aborda o problema de encontrar a posição de cada átomo (vértice) de uma proteína, dado as distâncias, intervalares ou não, desse vértice a três vértices anteriores e/ou outros vértices. Os resultados demonstram que a abordagem é promissora.

**PALAVRAS CHAVE.** Problema da Distância Geométrica, Otimização , Incerteza.

**Tópicos Programação Matemática**

### ABSTRACT

In this work we present a new mathematical modeling, via optimization Global and interval function, for the problem of interval geometric distance (PDGi). The goal of PDGi, in the context of the protein structure, is to find a  $R^3$  of a graph  $G = (V, E)$ , where an interval distance is associated with each edge. This structure is in accordance with the experimental constraints of Resonance Nuclear Magnetics. In particular, this paper addresses the problem of finding Of each atom of a protein, the interval distances of this atom being known to three previous atoms (click). The results demonstrate that the approach is successful.

**KEYWORDS.** Geometry Distance Problem. Optimization. Uncertainty.



## 1. Introdução

O problema da distância geométrica (PDG) tem sido estudado por diversos pesquisadores, possuindo aplicações na determinação da geometria de proteínas, na localização de sensores, com aplicações na internet das coisas, robótica, entre outros [Carlile et al. 2012,2013].

Em especial é possível, por meio de técnicas como a Ressonância Magnética Nuclear (RMN), estimar distâncias entre pares de átomos de uma determinada molécula, e o problema se torna o de identificar a conformação tridimensional da molécula, isto é, as posições de todos os seus átomos. Neste campo, o principal interesse é sobre as proteínas, visto que a descoberta de sua conformação tridimensional nos permite obter informações sobre as funções que as proteínas são capazes de realizar. Logo, é possível concluir que o PDG tem implicações no desenvolvimento de novos fármacos e uma vasta aplicação em biotecnologia. Quando se trata de moléculas biológicas, o PDG é geralmente conhecido como PDG molecular (PDGM).

Os dados oriundos da RMN são também geralmente esparsos [Dong and Wa 2003, Bowers 2000] e as distâncias menores ou igual 5 Angstrom. Entretanto, esses trabalhos [Dong and Wa 2003, Bowers 2000] não utilizaram distâncias intervalares.

O PDG em  $R^3$  pode ser definido, de forma geral, da seguinte forma:

**Definição .** Dado um grafo não direcionado  $G = (V, E, d)$  com pesos  $d : E \rightarrow (0, \infty)$  em suas arestas o PDG consiste em encontrar uma função  $x : V \rightarrow R^3$  tal que  $\|x(u) - x(v)\|_2 = d(u, v)$  para todo  $u, v \in V$ . A solução é, dessa forma, associar a cada vértice de  $G$  um único ponto em  $R^3$  satisfazendo as equações acima.

*Observação 1.* Nesse trabalho são utilizados os seguintes conceitos e notações :

- Utilizamos apenas a distância euclidiana. Assim, se  $u = (u_1, u_2, u_3)$  e  $v = (v_1, v_2, v_3) \in R^3$ , temos que:

$$\|u - v\|_2 = d(u, v) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + (u_3 - v_3)^2}$$

- Seja  $\gamma(v_k) = \{v_j \in V, j < k\}$  o conjunto de predecessores do vértice  $v_k$  e  $|\gamma(v_k)|$  a cardinalidade desse conjunto.
- Seja  $N(v_k)$  o conjunto das arestas adjacentes ao vértice  $v_k$ .
- Dado um grafo  $G(V, E)$ ,  $V$  é o seu conjunto de vértices,  $E$  o conjunto de suas arestas e  $n = |V|$  o número de vértices.

Uma forma de abordar o PDG é resolver o seguinte problema de otimização :

**Definição .** Dado um grafo  $G=(V,E,D)$  associamos o problema de otimização da distância geométrica:

$$(PODG) \quad \text{Min} \quad \sum_{x(u) \in R^3, u, v \in V} \|x(u) - x(v)\|_2^2 - d(u, v)^2$$

Em muitas situações, como no problema da geometria de proteínas, em virtude de limitações experimentais da Ressonância Magnética Nuclear (RMN) algumas distâncias são dadas como intervalos. Nesse caso definiremos o Problema da Distância Geométrica Discretizável Intervalar (PDGDi) da seguinte forma:

**Definição .** Dado um grafo  $G = (V, E, [di, ds])$  com uma distância-intervalar  $[di, ds]$  associada a cada uma de suas arestas, sendo  $di(v_l, v_k) \leq ds(v_l, v_k)$ , o PDGDi consiste em encontrar uma função  $x : V \rightarrow R^3$  tal que  $di(v_l, v_k) \leq \|x(v_l) - x(v_k)\| \leq ds(v_l, v_k)$  para todo  $v_l, v_k \in V$ . Nós assumimos que:



1. São dadas as coordenadas dos três primeiros vértices  $v_1, v_2, v_3$ ;
2. Para todo  $i = 4, 5, \dots, n$  os três vértices anteriores formam um clique com  $v_i$ , isto é,  $\{v_{i-3}, v_{i-2}, v_{i-1}, v_i\}$  é um clique onde:

$$d_{v_{i-3}v_{i-2}} + d_{v_{i-2}v_{i-1}} > d_{v_{i-3}v_{i-1}}$$

3. Para algum  $l = 5, \dots, n$  existe  $k \in \{1, 2, \dots, i - 4\}$  tal que:

$$[d(l, k)] = [di_{v_l, v_k}, ds_{v_l, v_k}], \quad 0 < di_{v_l, v_k} < ds_{v_l, v_k} \quad (1)$$

Em [D'Ambrosio et al 2014] os autores apresentam um conjunto de formulações de programação quadrática e de programação semidefinida bem como um conjunto de métodos novos e existentes para resolver estas formulações. Entretanto, a abordagem proposta nesse trabalho difere de todas essas formulações, e conseqüentemente, de seus métodos de resolução. Os testes realizados indicam que a abordagem proposta nesse trabalho é promissora quando comparado com os melhores resultados obtidos em [D'Ambrosio et al 2014].

Entre todas as formulações propostas em [D'Ambrosio et al 2014] a que obteve melhores resultados foi a seguinte:

$$\begin{aligned} & \text{Min } \sum_{v_k, v_l \in E} s_{v_k v_l} \\ \text{s. a: } & di_{v_k, v_l}^2 - \|x_{v_k} - x_{v_l}\|_2^2 \leq s_{v_k, v_l} \\ & \|x_{v_k} - x_{v_l}\|_2^2 - di_{v_k, v_l}^2 \leq s_{v_k, v_l} \\ & \sum_{v \in V} x_{vm} = 0 \quad \forall m \leq 3 \end{aligned}$$

Já o método que obteve melhores resultados em [D'Ambrosio et al 2014] foi a VNS - Variable Neighborhood Search. A VNS é uma metaheurística baseada em mudanças sistematicas na estrutura de vizinhança dentro de uma busca, visando escapar de ótimos locais [Mladenovic e Hansen 1997].

Na próxima seção apresentamos os novos modelos matemáticos para dois casos do PDGDi via otimização global. Na seção 3 apresentamos alguns resultados dos testes computacionais. Por fim, na seção 4 apresentamos nossas conclusões e perspectivas.

## 2. Modelos matemáticos e algoritmos para o PDGi Esparso

Para a obtenção de uma solução para o PDGMi cada distância intervalar entre dois vértices foi considerada como uma função linear. Mais especificamente, para todo  $(v_k, v_l) \in E$  definimos uma função associada da seguinte forma:

**Definição (2.1).** A cada intervalo  $[di(v_k, v_l), ds(v_k, v_l)]$  associamos a função :

$$f_{kl}(\lambda) = di(v_k, v_l) + \lambda * (ds(v_k, v_l) - di(v_k, v_l)) \quad (2)$$

onde  $\lambda \in [0, 1]$ .

As propriedades dessa função, bem como os principais conceitos da aritmética intervalar podem ser encontrados em [More and Lodwick 1997].

Foi abordado nesse trabalho dois casos especiais do PDGi, conforme definições abaixo:

**Definição (2.2).** Considere um PDGMi onde o grau de cada vértice  $v_i$ ,  $i \geq 4$ , é igual a 6 da seguinte forma:  $v_i$  forma um clique com os vértices  $v_{i-1}$ ,  $v_{i-2}$  e  $v_{i-3}$ , e outro clique com os vértices  $v_{i+1}$ ,  $v_{i+2}$  e  $v_{i+3}$ . Temos assim um grafo com  $3 \times (n - 1)$  arestas. Chamaremos esse caso particular do PDGDi de PDGDi Esparso 1 (PDGDiE1).



**Definição (2.3).** Considere um PDGDi onde o grau de cada vértice  $v_i$ ,  $i \geq 4$ , é maior ou igual a 6 da seguinte forma:  $v_i$  forma um clique com os vértices  $v_{i-1}$ ,  $v_{i-2}$  e  $v_{i-3}$ , e outro clique com os vértices  $v_{i+1}$ ,  $v_{i+2}$  e  $v_{i+3}$ , e pode ser adjacente a outros vértices. Considere que a distância entre os vértices  $i$  e  $j$  seja intervalar se e somente se  $i = j - 3$ . Chamaremos esse caso especial do PDGDi de PDGDi Esparso 2 (PDGDiE2).

**Definição (2.4).** Considerando as hipóteses das definições 1.3 e 2.2, seja  $xk^*$  a solução ótima  $zk(xk^*) = 0$  do problema de otimização  $P_k$ ,  $k = 4, \dots, n - 3$ :

$$\begin{aligned} \text{Min } zk(x) \sum_{l=1}^3 [(x(1) - y(k-l, 1))^2 + (x(2) - y(k-l, 2))^2 + (x(3) - y(k-l, 3))^2 - f_{kl}(\lambda_l)^2] \\ \text{s. a: } 0 \leq \lambda_l \leq 1 \text{ para } l = 1, 2, 3. \\ x \in R^3 \end{aligned}$$

onde  $y$  é a matriz de coordenadas do grafo de  $i = 1$  até  $i = k - 1$ ,  $k = 4, \dots, n$ . Chamaremos essa abordagem de Algoritmo 1.

Assim, temos  $n - 3$  sub-problemas de otimização, com até 6 variáveis, onde em cada problema encontramos as coordenadas de um vértice/átomo, que serão utilizadas no sub-problema seguinte para a obtenção das coordenadas do vértice corrente, até a obtenção do posicionamento de todos os vértices e, assim, a solução do PDGDiE1.

**Teorema 1.** Seja  $x^* = (x_1, x_2, x_3, x_4^*, x_5^*, \dots, x_{|V|}^*)$ , onde  $xk^*$  é solução ótima do PDGDiE1 conforme a definição 2.2. Temos que  $x^*$  é solução ótima do problema geral:

$$\text{(PODG) Min } \sum (||x_{v_k} - x_{v_l}||_2^2 - d_{v_k, v_l}^2)^2 \\ x(u) \in R^3, (u, v) \in E,$$

**Demonstração :** É fácil ver que para toda aresta  $(u, v) \in E$  nós temos  $||x^*(u) - x^*(v)||_2^2 - d(u, v)^2 = 0$ , sendo  $d(u, v)$  um intervalo degenerado ou não.

**Definição (2.5).** Considerando as hipóteses das definições 1.3 e 2.3, seja  $xk^*$  a solução ótima  $z_k = 0$  de cada um dos  $n - 3$   $P_k$  sub-problemas de otimização :

1. Seja  $\epsilon \geq 0$  dado e  $k=4$ . Enquanto  $k \leq n$  faça:
2. Seja  $xk1$  a solução ótima do problema de otimização  $Pk1$ :

$$\begin{aligned} \text{Min } z1k(x) = \\ \sum_{l=1}^2 [(x(1) - y(k-l, 1))^2 + (x(2) - y(k-l, 2))^2 + (x(3) - y(k-l, 3))^2 - d_{kl}^2]^2 + \\ + [(x(1) - y(k-3, 1))^2 + (x(2) - y(k-3, 2))^2 + (x(3) - y(k-3, 3))^2 - f_{k3}^2(\lambda_k)]^2 \\ \text{s. a: } 0 \leq \lambda_k \leq 1 \\ x \in R^3 \end{aligned}$$

onde  $y$  é a matriz de coordenadas do grafo de  $i=1$  até  $i=k-1$ .

3. Verifique se  $|N(v_k) \cap \gamma(v_k)| > 3$ . Se não, faça  $xk^*(i) = xk1(i)$ ,  $i=1, 2, 3$ ,  $k = k + 1$  e volte a 1. Se sim, verifique se existe  $j \in N(v_k) \cap \gamma(v_k)$  tal que  $||x^*(v_k) - x^*(v_j)||_2^2 - d(v_k, v_j)^2 > \epsilon$ . Se não, faça  $xk^*(i) = xk1(i)$ ,  $i=1, 2, 3$ ,  $k = k + 1$  e volte a 1. Se sim, resolva o problema de otimização  $Pk2$ :

$$\begin{aligned} \text{Pk2: Min } z2k(x) = \\ \sum_{j \in N(v_k) \cap \gamma(v_k), j \neq 3} [(x(1) - y(j, 1))^2 + (x(2) - y(j, 2))^2 + (x(3) - y(j, 3))^2 - d_{kl}^2]^2 + \\ + [(x(1) - y(k-3, 1))^2 + (x(2) - y(k-3, 2))^2 + (x(3) - y(k-3, 3))^2 - f_k^2(\lambda_k)]^2 \\ \text{s. a: } 0 \leq \lambda_k \leq 1 \\ x \in R^3 \end{aligned}$$



4. Seja  $xk2^*$  a solução ótima do problema  $Pk2$ . Faça  $xk^*(i) = xk2^*(i)$ , para  $i = 1, 2, 3$ ,  $k = k + 1$  e volte a 1.

O passo 3 é a principal diferença desse método em relação ao algoritmo 1. Trata-se de um re-otimização caso haja uma diferença maior do que a tolerada entre o vértice corrente e vértices antecessores que não formam um clique. Chamaremos essa abordagem de Algoritmo 2. Se é usado  $\epsilon = 0$  e existe uma realização para o grafo, o teorema a seguir demonstra que uma realização é encontrada por esse algoritmo.

**Teorema 2.** Seja  $x^* = (x_1, x_2, x_3, x_4^*, x_5^*, \dots, x_n^*)$ , onde  $xk^*$  é solução ótima do PDGDiE2, com  $z1k(xk^*) = 0$  para todo  $k = 4, \dots, n$  and  $\epsilon = 0$ , conforme a definição 2.4. Temos que  $x^*$  é solução ótima do problema geral:

$$(PDG) \quad \text{Min} \quad \sum (||x_{v_k} - x_{v_l}||_2^2 - di_{v_k, v_l}^2)^2 \\ x(u) \in R^3, (u, v) \in E,$$

Demonstração : Dada a escolha de  $\epsilon = 0$ , se existe  $xk^*$  é solução ótima do PDGDiE2, com  $z1k(xk^*) = 0$ , é fácil ver que para toda aresta  $(u, v) \in E$  nós temos  $||x^*(u) - x^*(v)||_2^2 - d(u, v)^2 = 0$ .

### 3. Implementação e testes computacionais

Para os testes computacionais das abordagens propostas nesse trabalho foram implementados os dois métodos descritos na seção anterior. Para resolver os problemas de otimização foi utilizada a função *optim* do software livre *scilab* [Baudin et al 2012]. Dentre as opções de método dessa função nesse trabalho foi utilizada o método quase newton [Baudin et al 2013].

Os testes foram feitos em um laptop HP, 4GB RAM, processador intel celeron 1.6 GHz, 64 bit, sistema operacional Windows Home 10.

Para aferir a qualidade da solução foram utilizadas as métricas  $\psi(x, G)$  e  $\phi(x, G)$  proposta em [D'Ambrosio et al 2014] e a seguir detalhadas. Sejam:

- 

$$\alpha_{v_l v_k}(x) = \max(0, di(v_l, v_k) - ||x_{v_l} - x_{v_k}||_2) + \max(0, ||x_{v_l} - x_{v_k}||_2 - ds(v_l, v_k)) \quad (3)$$

se  $di(v_l, v_k) \neq ds(v_l, v_k)$ .

- 

$$\alpha_{v_l v_k}(x) = ||x_{v_l} - x_{v_k}||_2 - d(v_l, v_k) \quad (4)$$

se  $di(v_l, v_k) = ds(v_l, v_k) = d(v_l, v_k)$ .

- 

$$\psi(x, G) = \max\{\alpha_{v_l v_k}(x)\} \quad (5)$$

- 

$$\phi(x, G) = \frac{\sum \alpha_{v_l v_k}(x)}{n} \quad (6)$$

Para realizar os testes computacionais foram utilizados os dados de proteínas presentes no banco de dados PDB - *Protein Data Bank* [PDB 2017], o mais importante banco de dados de proteínas utilizado por pesquisadores de diversas áreas em todo mundo. Nesse banco de dados é possível obter as coordenadas de diversas proteínas. Para obter as distâncias entre os átomos foi realizado um pré-processamento.

Visando simular os dados obtidos por RMN foram utilizadas apenas as distâncias entre átomos, iguais, de hidrogênio ou carbono com distância menor ou igual 5.



- **Proteína 1BRV** - proteína imunodominante do vírus respiratório sincicial bovino, possui 130 átomos de hidrogênio.
- **Proteína 100D** - estrutura de cristal, possui 200 átomos de carbono.
- **Proteína 1PPT** - polipeptídeo pancreático, possui 192 átomos de carbono.

Descrevemos a seguir as características e resultados obtidos para cada um dos problemas estudados nesse trabalho.

### 3.1. PDGDi Esparsos 1

Para buscar o ótimo global de cada um dos sub-problemas de otimização foram geradas 50 soluções iniciais aleatórias se o sub-problema possuía um intervalo não degenerado, ou seja, onde  $d_i < d_s$  para algum par de vértices presentes no sub-problema corrente. Caso o sub-problema não possuísse intervalos não degenerados, ou seja, todos os intervalos seriam números reais ( $d_i = d_s$ ), foram geradas 20 soluções iniciais aleatórias.

Para a escolha de quais distâncias  $d(v_l, v_k)$  seriam, de fato, intervalares foi gerado um número pseudo-aleatório  $\alpha$  para cada aresta do grafo, utilizando-se a função  $ran()$  do *scilab*. Verifica-se, então, se  $\delta < 0.1$ . Se sim, foi define-se  $d_i(v_l, v_k) = d(v_l, v_k) - 0.1$  e  $d_s(v_l, v_k) = d(v_l, v_k) + 0.1$ . Caso contrário a distância associada a aresta não é modificada.

As proteínas, suas características e os resultados obtidos estão descritos abaixo:

- **Proteína 1BRV** - Entre todas as distâncias calculadas 417 se constituíram como intervalares pela metodologia acima exposta. Para os cálculos do posicionamento dos 127 átomos, do total de 130, já que o posicionamento dos 3 primeiros átomos e um parâmetro de entrada, por meio de 127 sub-problemas de otimização, em 22 deles havia pelo menos uma distância intervalar.

Melhor solução geral encontrada:  $\psi(x, G) = 0.546$ ,  $\phi(x, g) = 0.01$ . Tempo computacional = 145.42 segundos.

- **Proteína 100D** - Entre todas as distâncias calculadas 960 se constituíram como intervalares pela metodologia exposta acima. Para os cálculos do posicionamento dos 197 átomos, do total de 200, por meio de 197 sub-problemas de otimização, em 24 deles havia pelo menos uma distância intervalar.

Melhor solução geral encontrada:  $\psi(x, G) = 0.615$ ,  $\phi(x, g) = 0.00104$ . Tempo computacional = 1313.12 segundos.

- **Proteína 1PPT** - Entre todas as distâncias calculadas 869 se constituíram como intervalares pela metodologia exposta acima. Para os cálculos do posicionamento dos 189 átomos, do total de 192, por meio de 189 sub-problemas de otimização, em 28 deles havia pelo menos uma distância intervalar.

Melhor solução geral encontrada:  $\psi(x, G) = 1.2$ ,  $\phi(x, g) = 0.1152$ . Tempo computacional = 151.19 segundos.

Vale observar que em [D'Ambrosio et al. 2014] o melhor resultado obtido entre as diversas combinações de formulações matemáticas e métodos computacionais para a Proteína 100D foi  $\psi(x, G) = 2.05$ , e para a proteína 1PPT foi  $\psi(x, G) = 1.93$ . Contudo, é importante ressaltar que não é possível dizer qual abordagem é melhor, se a abordagem proposta nesse trabalho ou as melhores combinações (formulação + método) propostas em [D'Ambrosio et al 2014], pois as distâncias intervalares não são as mesmas, logo, não é possível comparar os métodos. De qualquer forma, é possível concluir que a abordagem proposta nesse trabalho é, pelo menos, promissora, como confirmam os testes com o PDGDi Molecular descritos a seguir.



### 3.2. PDGDi Esparso 2

Diferentemente do PDGDiE1, o PDGDiE2 possui uma e somente uma distância intervalar em cada sub-problema de otimização  $P_k$ . Cada uma destes intervalos tem comprimento igual a  $[0.2]$  nestes testes computacionais. Contudo, existem muito mais arestas. Para  $k > 4$  um número aleatório  $\rho_i$  é gerado para definir se haverá uma aresta entre o vértice  $i$  e o vértice  $k$ ,  $i = i - 5, i - 4, i + 4, i + 5$ , além das arestas que formam os dois cliques previstos na definição do problema (ver definição 2.3). Nos testes computacionais foi utilizado  $\epsilon = 0.5$  como tolerância (ver algoritmo 2 na definição 2.5), pois nem sempre é possível garantir a obtenção do ótimo global.

Foram realizados testes com as mesmas proteínas descritas anteriormente, conforme detalhado a seguir. Para cada problema de otimização resolvido são geradas 10 soluções iniciais aleatórias visando escapar de ótimos locais.

- **Proteína 1BRV** - Foram geradas 139 arestas além das arestas que formam os cliques. A fase de re-otimização foi chamada 27 vezes.

Melhor solução geral encontrada:  $\psi(x, G) = 3.57 \times 10^{-7}$ ,  $\phi(x, G) = 2.46 \times 10^{-9}$ . Tempo computacional = 94.58 segundos.

- **Proteína 100D** - Foram geradas 215 arestas além das arestas que formam os cliques. A fase de re-otimização foi chamada 86 vezes.

Melhor solução geral encontrada:  $\psi(x, G) = 0.00301$ ,  $\phi(x, G) = 1.27 \times 10^{-5}$ . Tempo computacional = 175.01 segundos.

- **Proteína 1PPT** - Foram geradas 203 arestas além das arestas que formam os cliques. A fase de re-otimização foi chamada 43 vezes.

Melhor solução geral encontrada:  $\psi(x, G) = 0.0167$ ,  $\phi(x, G) = 7.18 \times 10^{-5}$ . Tempo computacional = 160.23 segundos.

Podemos concluir que esses testes indicam que a abordagem é exitosa, tendo obtido soluções ao menos muito próximas dos ótimos globais.

### 4. Conclusões

Nesse trabalho são apresentadas duas novas abordagens do problema da distância geométrica intervalar (PDGi). Para modelar o PDGi como um problema de otimização global irrestrita foi utilizado o conceito de função intervalar nos dois *approach*. Os testes realizados com dados reais de proteínas demonstram que as abordagens propostas são promissoras, e que seus estudos e testes devem ser aprofundados.

### Agradecimentos

O primeiro autor agradece o apoio da FAPESP, Projeto 2016/21459-0. O segundo autor agradece o apoio do CNPq, Projeto 400754/2014-2.

### Referências

C. D'Ambrosio, V. Ky, C. Lavor, L. Liberti, N. Maculan, *Solving distance geometry problems with interval data using formulation-based methods*. Technical report, LIX Ecole Polytechnique, 2014.

M. Baudin, S. Steer, Optimization with scilab, present and future. In *Open-source Software for Scientific Computation (OSSC)*, IEEE International Workshop, pages 99-106, 2009.





M. Baudin, V. Couvert, S. Steer, Optimization in scilab. Scilab Consortium, INRIA Paris-Rocquencourt, July 2010.

P.M. Bowers, C.E. Strauss, D. Baker. De novo protein structure determination using sparse NMR data. *Journal of biomolecular NMR*. 18:311-318, 2000.

Q. Dong, Z. Wu, A geometric build-up algorithm for solving the molecular distance geometry problem with sparse distance data. *Journal of Global Optimization*.26:321-33, 2003.

C. Lavor, L. Liberti, A. Mucherino, The interval Branch-and-Prune algorithm for the discretizable molecular distance geometry problem with inexact distances, *Journal of Global Optimization* 56: 855-871, 2013.

C. Lavor, L. Liberti, N. Maculan, A. Mucherino, Recent advances on the discretizable molecular distance geometry problem, *European Journal of Operational Research*, 219: 698-706, 2012.

M. Mladenovic, P. Hansen, Variable neighborhood search. *Computers & operations research*, 24: 1097-1100, 1997.

R. Moore, W. Lodwick, Interval analysis and fuzzy set theory. *Fuzzy sets and systems*, 135: 5-9, 2003.

[www.rcsb.org/pdb](http://www.rcsb.org/pdb) Acessado em 27 de marco de 2017.