A

End-Term Presentation

Of ELD 431: B.Tech. Project-1

on

Data based analysis of RNA folding

Presented By

Chenika Garg - 2019EE30877

Gautam - 2019EE30567

Under the Supervision of

Prof . Shaunak Sen

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY ,DELHI, INDIA

Background

- RNA is one of the key types of molecules found in living cells and is involved in important biological processes as a carrier of genetic information and also for cellular functions
- Similar to proteins, 3D structure of an RNA molecule determines its function. There are multiple experimental techniques to build a 3D model of RNA Structures.
- However, experimental techniques are sometimes tedious and time-consuming.
- Therefore, it is important to develop other computational methods to determine RNA structures.

Literature Review

- **Protein Folding**:
  - There are many similarities between protein folding and RNA folding.
  - In 2020, AlphaFold used deep learning to solve protein folding with remarkable accuracy
  - AlphaFold model was trained on over 17,000 known protein structures but a very few RNA template structures are available, making it unviable for RNA folding [2]
- **ARES:**
  - Atomically rotationally equivariant scorer (Ares) is a novel machine learning model to compute to determine 3D structure of RNA Structure.
  - It uses neural networks to to assess the structural motifs of RNA molecules [3]

Problem Statement

To determine the energetically favourable conformations of the RNA structures by applying a simplistic ball and stick model for the RNA molecules. We also aim to employ methods such as Newton Interval Method, Global minima method and Discrete Sets approach to find the structural motifs of RNA molecules.

Introduction

- **Ball Stick Model for RNA Molecules:**
  - We considered a very simple ball stick model for our analysis of RNA molecules.
  - The potential energy of these structures can be calculated to find the stability of these structures comparatively.

- **Energetical Favorable Confirmations of RNA molecules**:
  - The conformation of RNA molecules are dependent on the base-composition, sequence and chemical structure.
  - There are multiple roots for the potential function of RNA molecules.
  - The stable roots are considered as the Energetical Favorable Confirmations.

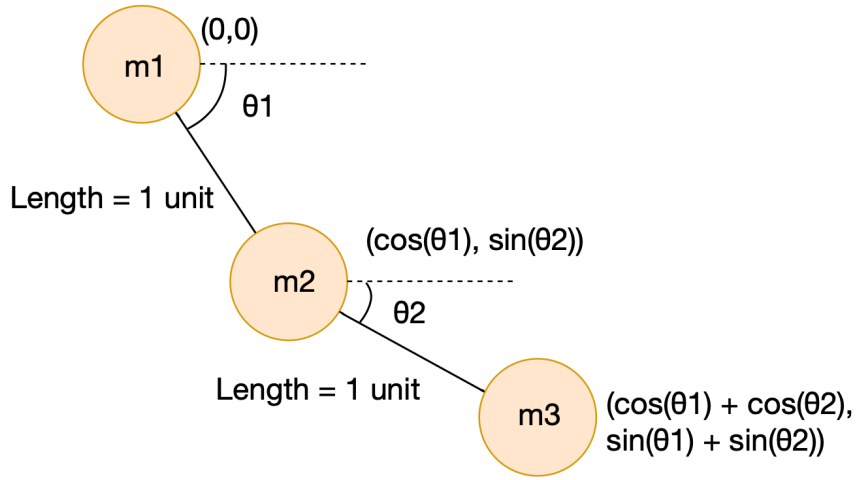


Fig1: Ball stick Model for RNA Molecules

Mathematical Model for the RNA Structure

- Position of the 1st molecule is taken as the origin. For beads i = 2,3....N, θ_{i-1} clockwise angle between the bead i and horizontal

- Co-ordinates of the relative position of each bead i w.r.t. the origin are:

$$[\sum_{j=1}^{j=i-1}\cos(\theta_j)\,,\ \sum_{j=1}^{j=i-1}\sin(\theta_j)]$$ in the X-Y plane

- The potential V_{ij} between any arbitrary beads i and j:
  - $V_{ij} = \frac{1}{2}d_{ij}^2$, where d_{ij} is the distance between bead i and bead j.

$$V_{ij} = \frac{1}{2}[(\sum_{k=1}^{k=i-1}\cos(\theta_k) - \sum_{k=1}^{k=j-1}\cos(\theta_k))^2 + ((\sum_{k=1}^{k=i-1}\sin(\theta_k) - \sum_{k=1}^{k=j-1}\sin(\theta_k))^2]$$

$$V_{ij} = \frac{1}{2}[(i\text{-}j) + \sum\sum_{j\le q\le p\le i-1}\cos(\theta_p - \theta_q)]$$

- Tools Used: MATLAB, Julia, Visual Studio Code, Jupyter Notebook

Newton Interval Method: 1D and Higher Dimensions

- Objective is to apply Newton Interval Method for determining the stationary points, especially the minima, of the derived potential function V_{ij}

- **Newton Interval Method in 1D**:
  - m(X) is the midpoint taken as initial guess
  - The iteration of Newton's method on X till it converges to the root of f(x) is given by:

$$X_{k+1} = N_f(X_k) \cap X_k \qquad \text{where } N_f(X_k) = (m(X_k)) - \frac{f(m(X_k))}{f'(X_k)})$$

  - If $0 \notin f'(X)$, then the intervals X_k form a nested sequence squeezing quadratically to the root of f

- **Newton Method for higher dimensions:**
  - The derivative term in the Newton operator gets replaced by the Jacobian. To solve the system of equations F**(x)** = 0, where $F: \mathcal{R}^n \rightarrow \mathcal{R}^n$

$$X_{k+1} = m(X_k)) - J(X_k)^{-1}\mathrm{F}(\mathrm{m}(X_k))$$

Newton Method Implementation using Julia's Inbuilt functions

- Julia provides inbuilt functions to calculate roots of a gradient function using Newton method

```
In [4]: rts = roots(Vf, (0..6.28) × (0..6.28), Newton, 1e-2)
```

- The mid points are evaluated for each of the roots and corresponding graphs are obtained

```
In [5]: midpoints = mid.([root.interval for root in rts])
```

```
In [6]: xs = first.(midpoints)
        ys = last.(midpoints)
        using Plots; plotlyjs()
        surface(0:0.1:6.28, 0:0.1:6.28, (x,y)->f([x,y]))
        scatter!(xs, ys, f.(midpoints))
```

- The inbuilt Newton method for Julia gives output for N=3 and N=4 but fails for higher dimensions

Results for N=3,4 Newton Interval Method

- $V(\theta_1, \theta_2) = 1 + \cos(\theta_1 - \theta_2)$
- Ran the inbuilt Julia function for Newton method on the interval $[0,2\pi]^2$ and obtained the root for $V'(\theta_1, \theta_2) = 0$
- The 2 axes represent the values of θ_1, θ_2 while the vertical axes represents the potential function's value

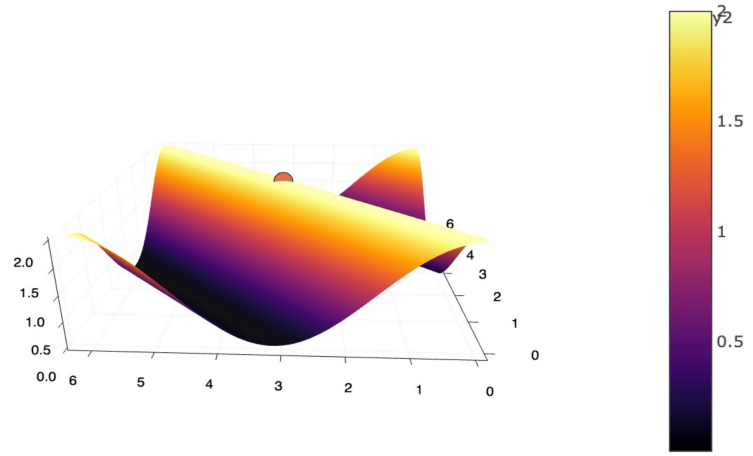


Fig. 3: Potential Function Plot for N=3 using Newton Interval Method in Julia

- $V(\theta_1, \theta_2, \theta_3) = \frac{3}{2} + \cos(\theta_1 - \theta_2) + \cos(\theta_2 - \theta_3) + \cos(\theta_3 - \theta_1)$
- Implemented the inbuilt Julia Newton method for N=4 by reducing the three variables θ_1, θ_2, θ_3 to 2 variables by setting $\theta_1 = 0$ and found the roots in the 2-D box $[0,2\pi]^2$
- $V(x, y) = \frac{3}{2} + \cos(x) + \cos(y) + \cos(x + y)$

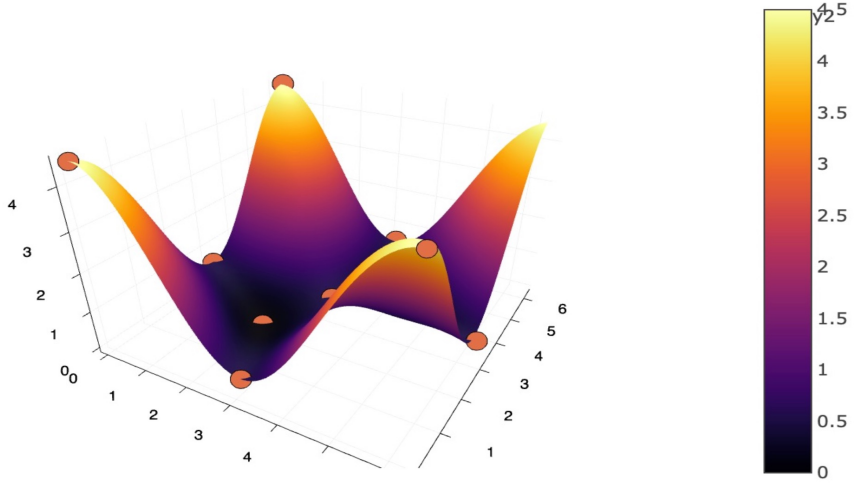


Fig. 4: Potential Function Plot for N=4 using Newton Interval Method in Julia

Newton Method from scratch in Julia

- Implemented Newton method using formulae and calculations in a step-wise manner
- For N = 4, $V(x, y, z) = \frac{3}{2} + \cos(x - y) + \cos(y - z) + \cos(z - x)$
- Newton method code for $\nabla V = 0$ failed at points where the Jacobian term becomes non-invertible, i.e. |J(∇V)| = 0
- An approach to exclude the points from the n-D box $[0,2\pi]^n$ (here, n=3), where the determinant of the Jacobian goes to zero is followed ahead

Points where Jacobian of Gradient function is non-invertible

- For a valid solution by the Newton method, an invertible Jacobian matrix is required, i.e. only points where det(Jacobian(∇V_{ij})) ≠ 0 are considered.

- All other points should be eliminated from the set of points in consideration for RNA structure to achieve all the possible roots for the structure.

N=3, 4: Non-Invertible Jacobian points for the Gradient function

- $V(\theta_1, \theta_2) = 1 + \cos(\theta_1 - \theta_2)$
- $V(x) = 1 + \cos(x)$ keeping $\theta_1 = 0$
- $V'(x) = -\sin(x)$
- $J(V'(x)) = -\cos(x)$
- Thus, $|J(V'(x))| = 0$ for $x = \frac{\pi}{2}, \frac{3\pi}{2}$

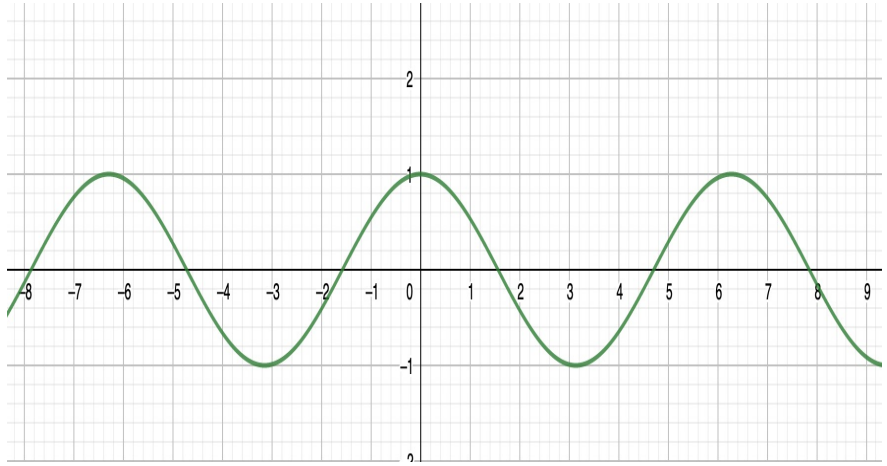
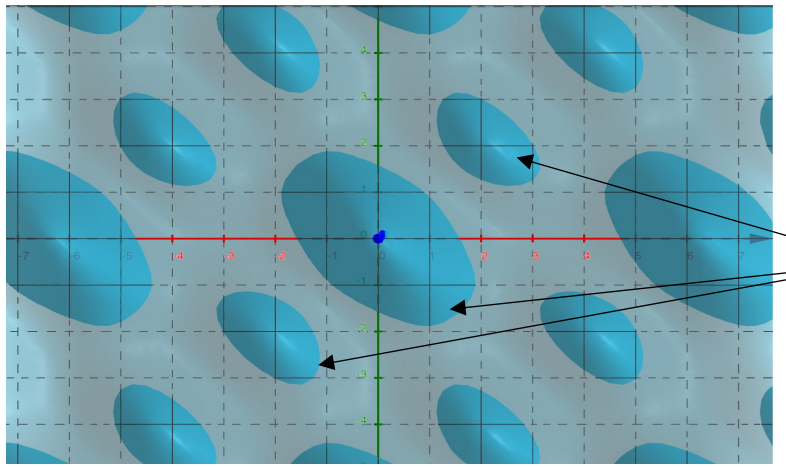


Fig. 5: Plot of det(Jacobian) function for N=3

- $V(x,y) = 3/2 + \cos(x) + \cos(y) + \cos(x+y)$
- $\nabla V(x,y) = \begin{bmatrix} -\sin(x) - \sin(x+y) \\ -\sin(y) - \sin(x+y) \end{bmatrix}$
- $J(\nabla V) = \begin{bmatrix} -\cos(x) - \cos(x+y) & -\cos(x+y) \\ -\cos(x+y) & -\cos(y) - \cos(x+y) \end{bmatrix}$
- Thus, $|J(\nabla V)| = \cos(x+y) * (\cos(x) + \cos(y)) + \cos(x) * \cos(y) = 0$



The boundary of these 3 ellipses are the excluded region of points, when taken in $[-\pi, \pi]^2$, which is equivalent to $[0, 2\pi]^2$ due to the function's periodicity

Fig. 6: Plot of det(Jacobian) function for N=4

Global Minima Approach

- Calculated the global minima value of the potential function and the respective angle values
- Utilised the inbuilt Julia function for finding the most stable conformation for the RNA molecules
- Varied the tolerance or maximum width of the interval to obtain precise points of global minima
- Obtained global minima for both low and high tolerance for N=3 and for high tolerance for N=4

```
julia> using IntervalArithmetic, IntervalOptimisation

julia> global_min, minimisers = minimise(  X -> ( (x,y) = X; 1.5 +  cos(x)
  + cos(y) + cos(x+y) ), (0..6.28) × (0..6.28));

julia> global_min
[-0.00139583, 9.49031e-09]

julia> minimisers
32560-element Vector{IntervalBox{2, Float64}}:
 [4.18627, 4.18709]²
 [4.18548, 4.18628] × [4.18627, 4.18709]
 [4.18468, 4.18549] × [4.18627, 4.18709]
 [4.1831, 4.18391] × [4.18627, 4.18709]
 [4.18627, 4.18709] × [4.1831, 4.18391]
 [4.18468, 4.18549] × [4.18548, 4.18628]
 [4.18548, 4.18628] × [4.18468, 4.18549]
 [4.1839, 4.18469] × [4.18627, 4.18709]
 [4.18627, 4.18709] × [4.1839, 4.18469]
 [4.19235, 4.19313] × [4.18627, 4.18709]
 ⋮
 [4.2, 4.20077] × [4.13218, 4.13296]
 [4.24109, 4.24188] × [4.13685, 4.13766]
 [4.13685, 4.13766] × [4.24109, 4.24188]
 [2.03519, 2.03595] × [2.12882, 2.12958]
 [2.12882, 2.12958] × [2.03519, 2.03595]
 [4.2473, 4.24809] × [4.14981, 4.15057]
 [4.14981, 4.15057] × [4.2473, 4.24809]
 [4.24574, 4.24652] × [4.14519, 4.14596]
 [4.14519, 4.14596] × [4.24574, 4.24652]

julia>
```

Fig. 7: Code for Global Minima for N=4

```
julia> using IntervalArithmetic, IntervalOptimisation

julia> global_min, minimisers = minimise(  X -> ( (x,y) = X; 1 + cos(x-y)
 ), (0..6.28) × (0..6.28));

julia> global_min
[0, 1.11023e-16]

julia> minimisers
16376-element Vector{IntervalBox{2, Float64}}:
 [6.25305, 6.2539] × [3.11135, 3.11219]
 [6.25223, 6.25306] × [3.11135, 3.11219]
 [6.25223, 6.25306] × [3.11054, 3.11136]
 [6.24978, 6.25061] × [3.10812, 3.10895]
 [6.25305, 6.2539] × [3.11218, 3.11299]
 [6.15066, 6.15149] × [3.0094, 3.01021]
 [6.15382, 6.15463] × [3.0118, 3.01262]
 [6.1554, 6.15621] × [3.01342, 3.01425]
 [6.15462, 6.15541] × [3.01342, 3.01425]
 [6.15303, 6.15383] × [3.0118, 3.01262]
 ⋮
 [1.28445, 1.28523] × [4.42559, 4.42638]
 [1.24423, 1.24501] × [4.38558, 4.38634]
 [1.24577, 1.24657] × [4.38784, 4.3886]
 [1.24347, 1.24424] × [4.38558, 4.38634]
 [1.24347, 1.24424] × [4.38484, 4.38559]
 [1.24423, 1.24501] × [4.38633, 4.38709]
 [1.28292, 1.28369] × [4.42403, 4.42483]
 [1.24268, 1.24348] × [4.38484, 4.38559]
 [3.33685, 3.3376] × [0.195176, 0.195903]

julia>
```

Fig. 8: Code for Global Minima for N=3

Discrete Sets Approach

- Observed that energetically favourable conformations occurred as polygon-shaped structures for N=3, 4; tried to extend the approach for higher dimensions.
- Defined a discrete set S_n for a n-bead molecule chain as the union of S_{n-1} and the angle values for a regular polygon of dimension (n-1), with the base case being $S_3 = \{0, \pi, 2\pi\}$.

$$S_n = S_{n-1} \cup \{\frac{2\pi}{n-1}, \frac{2*2\pi}{n-1}, \frac{3*2\pi}{n-1}, \ldots, \frac{(n-2)*2\pi}{n-1}\}$$

- Computed the value of the potential function over all (n-2)-sized tuples belonging to ${S_n}^{n-2}$ and selected the tuples with minimum value of potential function to sketch the energetically favourable conformations
- We observed that we did obtain one of the stable conformations for N=5, 6 as regular polygon structures, a square for N=5 and a pentagon for N=6, with potential energy = 0, confirming our hypothesis.
- We can generalise that for a n-bead RNA molecule chain, one of the energetically favourable conformation would be a regular polygon of dimension **n-1**.

N =3,4: Results of Discrete Sets Approach

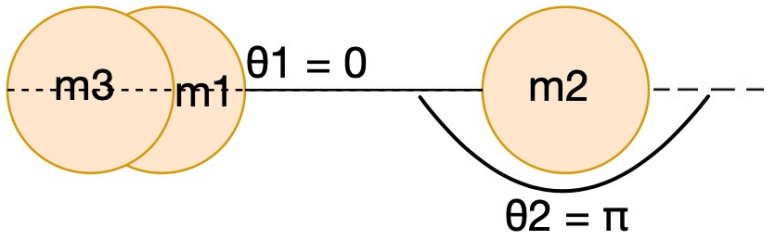
- V (θ1, θ2), θ1 = 0 for rotational invariance, reduces to:

  $V(\theta 2) = 1 + \cos(\theta 2)$

- For N = 3, beads (1 variable) $S_3 = \{0, \pi, 2\pi\}$
- Code for N = 3:

```
syms x
f(x) = 1 + cos(x)

a = [0, pi, 2*pi]

minimisers = []

for i = 1:3
    if(f(a(i))==0)
        minimisers = [minimisers, a(i)]
    end
end
```

- Stable structure obtained for N=3 (Minima of potential function):



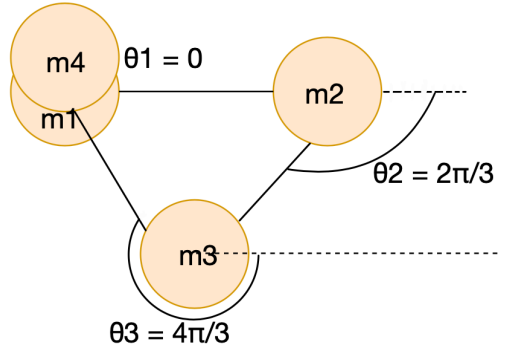
- V (θ1, θ2, θ3), θ1 = 0 for rotational invariance, reduces to:

  $V(\theta 2, \theta 3) = 3/2 + \cos(\theta 2) + \cos(\theta 2 - \theta 3) + \cos(\theta 3)$

- For N = 4, beads (2 variables) $S_4 = \{0, \pi, 2\pi, 2\pi/3, 4\pi/3\}$
- Code for N = 4:

```
syms x
syms y
f(x,y) = 3/2 + cos(x)+cos(y)+cos(x-y)

a = [0, pi, 2*pi, 2*pi/3, 4*pi/3]

minimisers = []

for i = 1:5
    for j = i:5
        if(f(a(i),a(j))==0)
            minimisers = [minimisers; [a(i),a(j)]]
        end
    end
end
```

- Stable structure obtained for N=4 (Minima of potential function):



N =5,6: Results of Discrete Sets Approach

- V (θ1, θ2, θ3, θ4), θ1 = 0 for rotational invariance, reduces to:

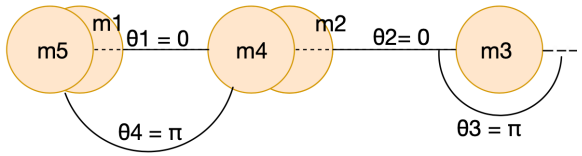
  V (θ2, θ3, θ4) = 2 + cos(θ2) + cos(θ3) + cos(θ4) + cos(θ2 − θ3) + cos(θ3 − θ4) + cos(θ2 − θ4)

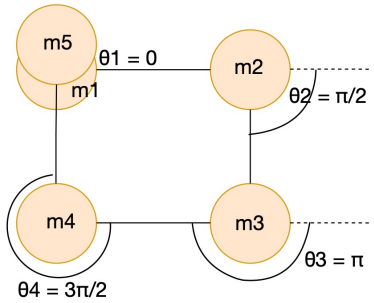
- For N = 5, beads (3 variables) S_5 = {0, π, 2π, 2 π/3, 4 π/3, π/2, 3 π/2}

- Code for N = 5:

```
syms x
syms y
syms z
f(x,y,z) = 2 + cos(x)+cos(y)+cos(z)+cos(x-y)+cos(y-z)+cos(x-z)

a = [0, pi, 2*pi, 2*pi/3, 4*pi/3, pi/2, 3*pi/2]

minimisers = []

for i = 1:7
    for j = i:7
        for k = j:7
            if(f(a(i),a(j),a(k))==0)
                ans = f(a(i),a(j),a(k))
                minimisers = [minimisers; [a(i),a(j),a(k)]]
            end
        end
    end
end
```

- Stable structures obtained for N=5 (Minima of potential function):





- V (θ1, θ2, θ3, θ4, θ5), θ1 = 0 for rotational invariance, reduces to:

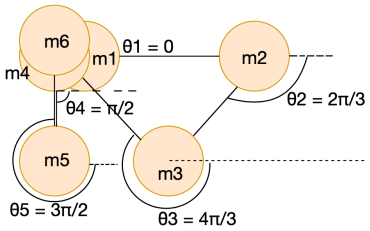
  V (θ2, θ3, θ4, θ5) = 5/2 + cos(θ2) + cos(θ3) + cos(θ4) + cos(θ5)+ $\sum\sum_{2\le q\le p\le 5}\cos(\theta_p-\theta_q)$

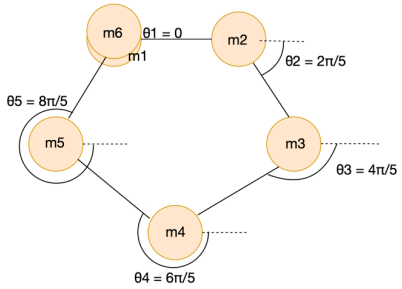
- For N = 6, beads (4 variables) S_3 = {0, π, 2π, 2 π/3, 4 π/3, π/2, 3 π/2, 2π/5, 4 π/5, 6 π/5, 8π/5}

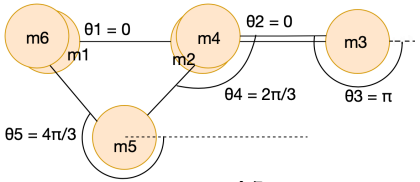
- Code N = 6:

```
syms x
syms y
syms z
syms w
f(w,x,y,z) = 5/2 + cos(w)+cos(x)+cos(y)+cos(z)+cos(w-x)+cos(w-y)+cos(w-z)+cos(x-y)+cos(y-z)+cos(x-z)

a = [0, pi, 2*pi, 2*pi/3, 4*pi/3, pi/2, 3*pi/2, 2*pi/5, 4*pi/5, 6*pi/5, 8*pi/5]

minimisers = []

for i = 1:11
    for j = i:11
        for k = j:11
            for l = k:11
                if(f(a(i),a(j),a(k),a(l))==0)
                    minimisers = [minimisers; [a(i),a(j),a(k),a(l)]]
                end
            end
        end
    end
end
```

- Stable structure obtained for N=3 (Minima of potential function):







Summary

- Julia's inbuilt Newton interval method implemented for N=3,4; graphical results were obtained
- Implemented step-wise Newton Interval function using Julia. It was observed that at certain values in the 3-D interval box $[0,2\pi]^3$, the Jacobian becomes non-invertible and the Newton's method is rendered ineffective
- Determined points in the 3-D interval box $[0,2\pi]^3$ where the Jacobian is not invertible
- Applied global minima approach for N =3, 4 for high tolerance
- Applied the discrete set approach for N=3, 4, 5, 6; verified that one of the stable conformations for an N-bead molecule chain is a (N-1)-sided regular polygon

Future Work

- To find the roots after the removal of non-invertible points of the jacobian function
- Global minima approach with closer tolerance to be applied in Julia to find the most stable solution for the RNA structures for higher dimensions
- Improve the time complexity of the Discrete Set Approach from exponential $O(|S_n|^{n-2})$ through algorithmic modifications to enable accurate structure determination for higher dimension sizes of RNA molecule chain.
- Employ a dynamic programming approach to improve and greatly reduce the time complexity of calculating solutions to the gradient of the potential function, V_{ij} to linear time complexity.

References

- [1] John M. Jumper et al. “Highly accurate protein structure prediction with AlphaFold”. In: Nature 596 (2021), pp. 583 –589. URL: https://doi.org/10.1038/s41586-021-03819-2.
- [2] R. Toews. “Alphafold is the most important achievement in ai-ever”. In: Forbes (2021). url: https://www.forbes.com/sites/robtoews/2021/10/03/ alphafold-is-the-most-important-achievement-in-ai-ever/?sh=25eedaf46e0a
- [3] Raphael Townshend, Stephan Eismann, Andrew Watkins, Ramya Rangan, Maria Karelina, Rhiju Das, and Ron Dror. “Geometric deep learning of RNA structure”. In: Science 373 (Aug. 2021), pp. 1047–1051. URL https://www.science.org/doi/abs/10.1126/science. abe5650.
- [4] Franklin.jl. “Newton method for interval root finding”. In: JuliaIntervals (2021). URL: https://juliaintervals.github.io/pages/explanations/explanationNewton/
- [5] David P. Sanders. “Julia Intervals/ Interval Optimization”. In: (2021). URL: https://github.com/JuliaIntervals/IntervalOptimisation.jl

Acknowledgement

- We take this opportunity to express our gratitude to our mentor and advisor Dr. Shaunak Sen, whose support and guidance have been invaluable to us during this whole project

- Our gratitude is also due to Mr. Vinod Kumar (PhD Scholar under Dr. Shaunak Sen) for his guidance and the Department of Electrical Engineering, for facilities to carry out this project

- In the end, we would like to thank our teachers, parents and friends for their constant support and encouragement

Thank You