Maria Fernanda Guerra, Raquel Ovadia, Ernesto Mealla
April 19, 2024
ISYE 3133

Final Report: Implementation and Analysis

Optimal Solutions

Through the creation of different linear programming models and Gurobi we were able to find an optimal solution for each model. We used the scenario included in the project description to reach these solutions. Our string of nucleotide sequence has 12 characters and is equal to "ACGUGCCACGAU".

Model 1: The First Crude Model
The solution found is optimal.
The optimal objective value is 5.0
Number of AU pairs is 2.0
Number of GC pairs is 3.0

In Model 1 we obtained an optimal objective value of 5. This signifies that the most likely pairing consists of 5 pairs. Furthermore, the most stable pairing of this nucleotide sequence has 2 AU pairs and 3 GC pairs.

Model 2: Simple Biological Enhancements
The solution found is optimal.
The optimal objective value is 8.0
Number of AU pairs is 1.0
Number of GC pairs is 2.0
Number of AC pairs is 0.0
Number of GU pairs is 0.0

In Model 2 we obtained an optimal objective value of 8. This signifies that the stability of the pairing is equal to 8. Furthermore, the most stable/likely pairing of this nucleotide sequence, under the assumptions of Model 2, is made of 1 AU pair and 2 GC pairs.

Model 3: More Complex Biological Enhancements
The optimal objective value is 10.0
Number of AU pairs is 1.0
Number of GC pairs is 2.0
Number of AC pairs is 0.0
Number of GU pairs is 0.0

Number of stacked quartets is 2.0

In Model 3 we obtained an optimal objective value of 10. This signifies that the stability of the pairing is equal to 10. Furthermore, the most stable/likely pairing of this nucleotide sequence, under the assumptions of Model 3, is composed of 1 AU pair, 2 GC pairs and 2 stacked quartets.

Model 4: A Model with Crossing Pairs
The optimal objective value is 14.05
Number of AU pairs is 2.0
Number of GC pairs is 3.0
Number of AC pairs is 1.0
Number of GU pairs is 0.0
Number of stacked quartets is 1.0
Number of crossing pairs is 8.0

In Model 4 we obtained an optimal objective value of 14.5. This signifies that the stability of the pairing is equal to 14.5. Furthermore, the most stable/likely pairing of this nucleotide sequence, under the assumptions of Model 4, is composed of 2 AU pairs, 3 GC pairs, 1 AC pair, 1 stacked quartet and 8 crossing pairs.

Running Time Plots

As seen in *Appendix 1* through *Appendix 4*, we observe that as the string size increases the running time in seconds of all of our models also increases. In all of the plots, we can observe that running time has an exponential growth. We noticed that Model 4, as seen in *Appendix 4*, has the highest rate of exponential growth.

Model Scale

Our model also increases as the size of the input (size of the string = n) increases. The number of variables and constraints depending on the size of the string in each of the models can be seen from *Appendix 5* through *Appendix 8*. In these tables we can observe that as the size of the string increases, more variables and constraints are created.

Constraints

These models contain multiple constraints. Some of these constraints were harder to solve than others. The first constraint we encountered issues solving was the constraint “Each nucleotide in at most 1 pair”. This was hard to solve because we couldn’t find a way to cover all the required scenarios. We got this wrong in the initial model. We had a hard time updating it until we were able to combine the previous two constraints into a correct new one. The second constraint we encountered problems when solving, also the one we found the most complicated,

was the constraint for stacked quartets. In our initial model we included i<j in our indices, which didn't allow us to meet edge cases for the stacked quartets. It took us several hours and hundreds of attempts to come up with a solution that did include the edge cases. This was also a hard constraint to code. We also considered the constraints that enforce crossed pairs complicated. We found it extremely hard to find the indices for this constraint that only consider the pairs that actually cross.

Model Updates

Data:

The first change we made was to our data. We changed the data value of r_{ij} because we got confused on the value of n when establishing the indices for this definition. The new definition implemented ensures the correct notation in distance indices. How we originally defined r_{ij} and the new definition can be observed below.

$$Original\colon\ r_{ij} = shortest\ distance\ between\ (S_i, S_j)\ calculated\ as\ min\{j - i, n + 1 - (j - i)\}$$

$$Updated\colon\ r_{ij} = shortest\ distance\ between\ (S_i, S_j)\ calculated\ as\ min\{j - i, n - j + i)\}$$

Model 1: The First Crude Model

In Model 1, we originally created for the first time two constraints to ensure that each nucleotide is in at most 1 pair. Nevertheless, we updated this constraint because the constraints only accounted for scenarios where i was less than j. The original and updated constraints can be observed below. Updating the constraint into one constraint allowed us to fix the indices.

$$Original\ -\ Each\ nucleotide\ in\ at\ most\ 1\ pair\colon\ \sum_{i=0}^{n-1} x_{ij} \le 1\ \forall j = 1,\dots, n-1$$

$$\sum_{j=1}^{n-1} x_{ij} \le 1\ \forall i = 0,\dots, n-1$$

$$Updated\ -\ Each\ nucleotide\ in\ at\ most\ 1\ pair\colon\ \sum_{i=0}^{k-1} x_{ik} + \sum_{j=k+1}^{n-1} x_{kj} \le 1\ \forall k = 0,\dots, n-1$$

Model 2: Simple Biological Enhancements

In Model 1, we updated the constraints that ensure that each nucleotide is in at most 1 pair. We must also implement this update in Model 2 to account for all the possible scenarios. In

our initial Model 2, we had an error in the distance constraint. By fixing the value of n in the indices of the definition of r_{ij} in our data, we were able to fix the error in the distance constraint.

Model 3: More Complex Biological Enhancements

In Model 1, we updated the constraints that ensure that each nucleotide is in at most 1 pair. We must also implement this update in Model 3 to account for all the possible scenarios. The error in the distance constraint was also fixed as mentioned above. In our initial Model 3, our stacked quartets constraint did not consider edge cases. For example, our initial constraint did not consider the stacked quartet that involves 0 in one pair and n in the other pair. In order to fix this error we had to remove one decision variable, add two decision variables, and update the objective and stacked quartets constraints accordingly. The original and updated decision variables, objective and constraints can be observed below.

$Original\ Decision\ Variables$:

$$x_{ij} = \{1\ if\ nucleotide\ in\ position\ i,j\ form\ a\ pair\ \forall\ i,j,\ i < j,\ i \neq j\}$$
$$\{0\ otherwise\}$$

$$y_{ij} = \left\{1\ if\ x_{ij} + x_{i-1,\,j+1} = 2\ \ (pairs\ are\ stacked)\ \ \forall\ i,j,\ i < j,\ i \neq j\right\}$$
$$\{0\ otherwise\}$$

$Updated\ Decision\ Variables$:

$$x_{ij} = \{1\ if\ nucleotide\ in\ position\ i,j\ form\ a\ pair\ \forall\ i,j,\ i < j,\ i \neq j\}$$
$$\{0\ otherwise\}$$

$$s_{ijkl} = \left\{1\ if\ x_{ij} = 1\ and\ xkl = 1\ for\ i = 0,\ldots,n-1,\ j = i+1,\ldots,\ n-1,\right\}$$
$$k = 0,\ldots,n-1,\ l = k+1,\ \ldots,\ n-1,\ k = (i+1) mod\ n,\ l = (j-1) mod\ n,\ i \neq k,\ j \neq l$$
$$\{0\ otherwise\}$$

$$t_{ijkl} = \left\{1\ if\ x_{ij} = 1\ and\ xkl = 1\ for\ i = 0,\ldots,n-1,\ j = i+1,\ldots,\ n-1,\right\}$$
$$k = 0,\ldots,n-1,\ l = k+1,\ \ldots,\ n-1,\ k = (j+1) mod\ n,\ l = (i-1) mod\ n,\ i \neq k,\ j \neq l$$
$$\{0\ otherwise\}$$

$Original\ Objective$:

$$Maximize\ Stability = max \sum_{(i,j)\epsilon I_{GC}} 3x_{ij} + \sum_{(i,j)\epsilon I_{AU}} 2x_{ij} + \sum_{(i,j)\epsilon I_{GU}} 0.1x_{ij} + \sum_{(i,j)\epsilon I_{AC}} 0.05x_{ij} + \sum_{i=0}^{n-1}\sum_{j=i+1}^{n-1} y_{ij}$$

$Updated\ Objective$:

$$Maximize\ Stability = max \sum_{(i,j)\epsilon I_{GC}} 3x_{ij} + \sum_{(i,j)\epsilon I_{AU}} 2x_{ij} + \sum_{(i,j)\epsilon I_{GU}} 0.1x_{ij} + \sum_{(i,j)\epsilon I_{AC}} 0.05x_{ij}$$
$$+ \sum_{i=0}^{n-1}\sum_{j=i+1}^{n-1}\sum_{k=0}^{n-1}\sum_{l=k+1}^{n-1} (s_{ijkl} + 0.5t_{ijkl}) \quad i \neq k,\ j \neq l$$

$Original\ Stacked\ Quartets\ Constraint: 2y_{ij} \leq x_{ij} + x_{i+1,j-1} \forall\ l,\ j,\ i < j,\ i \neq j$

$Updated\ Stacked\ Quartets\ Constraint:$

$2s_{ijkl} \leq x_{ij} + x_{kl} \forall\ i = 0,\ldots, n-1,\ j = i+1,\ldots, n-1,\ k = 0,..,n-1,\ l = k+1,\ldots, n-1,$
$i \neq k,\ j \neq l$

$2t_{ijkl} \leq x_{ij} + x_{kl} \forall\ i = 0,\ldots, n-1,\ j = i+1,\ldots, n-1,\ k = 0,..,n-1,\ l = k+1,\ldots, n-1,$
$i \neq k,\ j \neq l$

$t_{ijkl} = t_{klij}\ \forall\ i = 0,\ldots, n-1,\ j = i+1,\ldots, n-1,\ k = 0,..,n-1,\ l = k+1,\ldots, n-1,$
$i \neq k,\ j \neq l$

$s_{ijkl} \epsilon \{0, 1\}$

$t_{ijkl} \epsilon \{0, 1\}$

Model 4: A Model with Crossing Pairs

In Model 1, we updated the constraints that ensure that each nucleotide is in at most 1 pair. We must also implement this update in Model 4 to account for all the possible scenarios. The error in the distance constraint was also fixed as mentioned above. We must also update Model 4 the same as in Model 3 to account for the new decision variables and correct the stacked quartet constraint. In Model 4, we also updated our crossed pairs constraints. In order for this model to allow up to ten crossing pairs we had to add a new constraint to the previous two. The indices of the previous two constraints did not guarantee that the pairs counted actually crossed. We updated the indices of the previous crossed pairs constraints and added a new one to make the constraint correct. The original and new constraints can be observed below.

$Original\ Constraints:$

$Crossed\ Pairs\ 1:\ 2z_{ijkl} \leq x_{ij} + x_{kl}\ \ \forall\ i < j < k < l$

$At\ Most\ 10\ Crossed\ Pairs:\ \sum_{(i,j),(k,l)} z_{ijkl} \leq 10$

$Updated\ Constraints:$

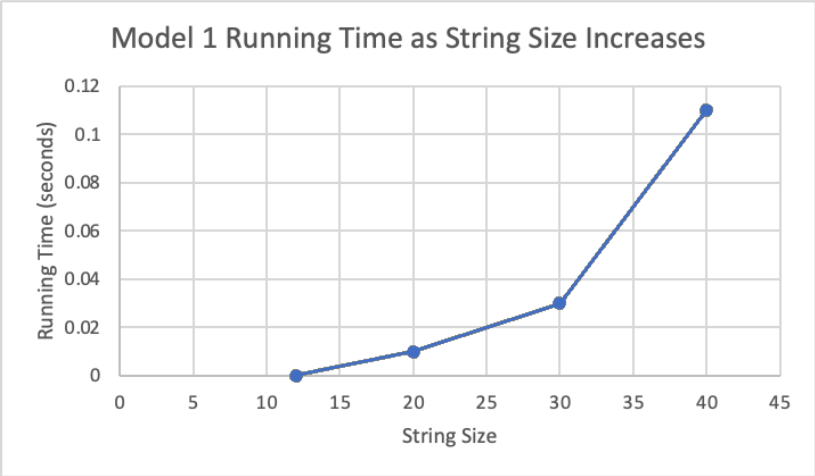
$Crossed\ Pairs\ 1:\ 2z_{ijkl} \leq x_{ij} + x_{kl}\ \ \forall\ i < k < j < l$

$Crossed\ Pairs\ 2:\ x_{ij} + x_{kl} - 1\ \leq\ z_{ijkl}\ \ \forall\ i < k < j < l$

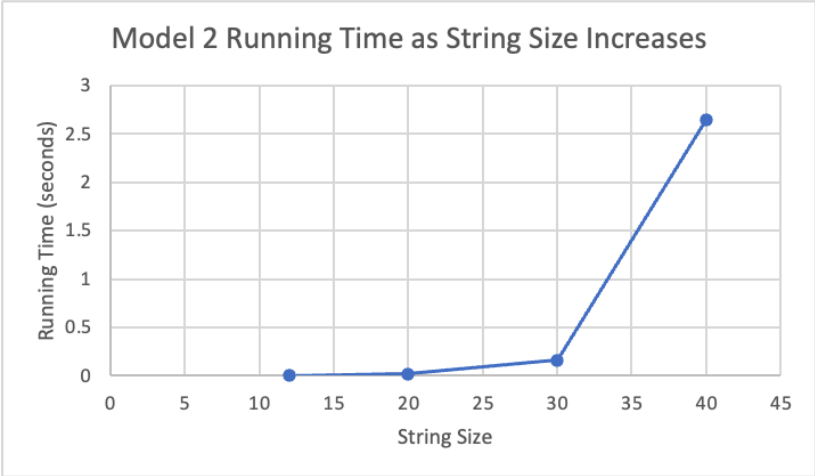
$At\ Most\ 10\ Crossed\ Pairs:\ \sum_{(i,j),(k,l)} z_{ijkl} \leq 10$

Appendix

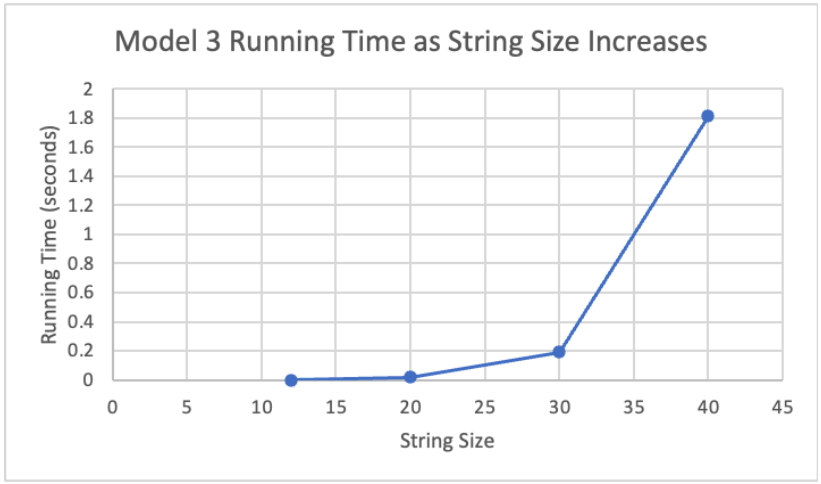
Appendix 1: Model 1 Running Time



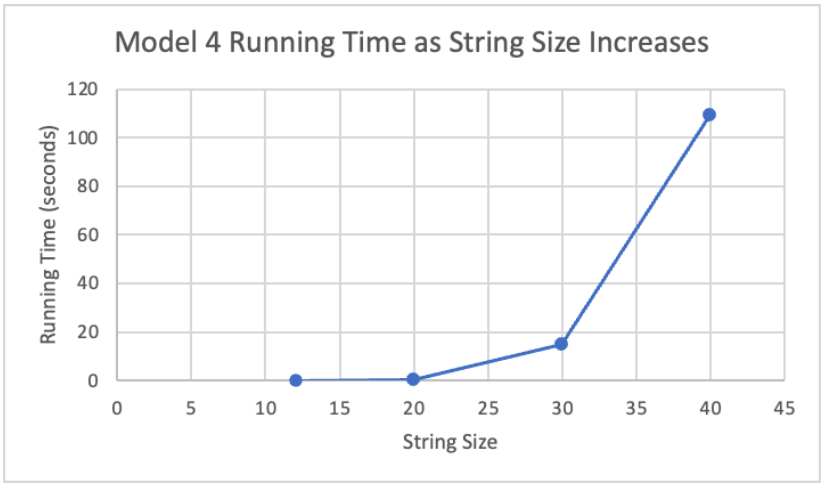
Appendix 2: Model 2 Running Time



Appendix 3: Model 3 Running Time



Appendix 4: Model 4 Running Time



Appendix 5: Number of Variables and Constraints in Model 1

Size of String (n)	**Decision Variable**	**Number of Decision Variables**
12	x_{ij}	$O(12^2) = O(144)$
20	x_{ij}	$O(20^2) = O(400)$
30	x_{ij}	$O(30^2) = O(900)$
40	x_{ij}	$O(40^2) = O(1600)$

Size of String (n)	**Constraint**	**Number of Constraints**
12	Complementarity	$O(12^2) = O(144)$
	Non-Crossing	$O(12^4) = O(20736)$
	Each nucleotide in at most 1 pair	12
20	Complementarity	$O(20^2) = O(400)$
	Non-Crossing	$O(20^4) = O(160000)$
	Each nucleotide in at most 1 pair	20
30	Complementarity	$O(30^2) = O(900)$
	Non-Crossing	$O(30^4) = O(810000)$
	Each nucleotide in at most 1 pair	30

40	Complementarity	$O(40^2) = O(1600)$
	Non-Crossing	$O(40^4) = O(256000)$
	Each nucleotide in at most 1 pair	40

Appendix 6: Number of Variables and Constraints in Model 2

Size of String (n)	**Decision Variable**	**Number of Decision Variables**
12	x_{ij}	$O(12^2) = O(144)$
20	x_{ij}	$O(20^2) = O(400)$
30	x_{ij}	$O(30^2) = O(900)$
40	x_{ij}	$O(40^2) = O(1600)$

Size of String (n)	**Constraint**	**Number of Constraints**
12	Complementarity	$O(12^2) = O(144)$
	Non-Crossing	$O(12^4) = O(20736)$
	Each nucleotide in at most 1 pair	12
	Distance	$O(12^2) = O(144)$

20	Complementarity	$O(20^2) = O(400)$
	Non-Crossing	$O(20^4) = O(160000)$
	Each nucleotide in at most 1 pair	20
	Distance	$O(20^2) = O(400)$
30	Complementarity	$O(30^2) = O(900)$
	Non-Crossing	$O(30^4) = O(810000)$
	Each nucleotide in at most 1 pair	30
	Distance	$O(30^2) = O(900)$
40	Complementarity	$O(40^2) = O(1600)$
	Non-Crossing	$O(40^4) = O(256000)$
	Distance	$O(40^2) = O(1600)$
	Each nucleotide in at most 1 pair	40

Appendix 7: Number of Variables and Constraints in Model 3

Size of String (n)	**Decision Variable**	**Number of Decision Variables**
12	x_{ij}	$O(12^2) = O(144)$

20	x_{ij}	$O(20^2) = O(400)$
30	x_{ij}	$O(30^2) = O(900)$
40	x_{ij}	$O(40^2) = O(1600)$
12	s_{ijkl}	$O(12^2) = O(144)$
20	s_{ijkl}	$O(20^2) = O(400)$
30	s_{ijkl}	$O(30^2) = O(900)$
40	s_{ijkl}	$O(40^2) = O(1600)$
12	t_{ijkl}	$O(12^2) = O(144)$
20	t_{ijkl}	$O(20^2) = O(400)$
30	t_{ijkl}	$O(30^2) = O(900)$
40	t_{ijkl}	$O(40^2) = O(1600)$

Size of String (n)	**Constraint**	**Number of Constraints**
12	Complementarity	$O(12^2) = O(144)$
	Non-Crossing	$O(12^4) = O(20736)$
	Each nucleotide in at most 1 pair	12

	Distance	$O(12^2) = O(144)$
	Stacked Quartets	$O(12^4) = O(20736)$
20	Complementarity	$O(20^2) = O(400)$
	Non-Crossing	$O(20^4) = O(160000)$
	Each nucleotide in at most 1 pair	20
	Distance	$O(20^2) = O(400)$
	Stacked Quartets	$O(20^4) = O(160000)$
30	Complementarity	$O(30^2) = O(900)$
	Non-Crossing	$O(30^4) = O(810000)$
	Each nucleotide in at most 1 pair	30
	Distance	$O(30^2) = O(900)$
	Stacked Quartets	$O(30^4) = O(810000)$
40	Complementarity	$O(40^2) = O(1600)$
	Non-Crossing	$O(40^4) = O(256000)$
	Each nucleotide in at most 1 pair	$O(40^2) = O(1600)$

	Distance	40
	Stacked Quartets	$O(40^4) = O(256000)$

Appendix 8: Number of Variables and Constraints in Model 4

Size of String (n)	**Decision Variable**	**Number of Decision Variables**
12	x_{ij}	$O(12^2) = O(144)$
20	x_{ij}	$O(20^2) = O(400)$
30	x_{ij}	$O(30^2) = O(900)$
40	x_{ij}	$O(40^2) = O(1600)$
12	s_{ijkl}	$O(12^2) = O(144)$
20	s_{ijkl}	$O(20^2) = O(400)$
30	s_{ijkl}	$O(30^2) = O(900)$
40	s_{ijkl}	$O(40^2) = O(1600)$
12	t_{ijkl}	$O(12^2) = O(144)$
20	t_{ijkl}	$O(20^2) = O(400)$
30	t_{ijkl}	$O(30^2) = O(900)$

40	t_{ijkl}	$O(40^2) = O(1600)$
12	z_{ijkl}	$O(12^2) = O(144)$
20	z_{ijkl}	$O(20^2) = O(400)$
30	z_{ijkl}	$O(30^2) = O(900)$
40	z_{ijkl}	$O(40^2) = O(1600)$

Size of String (n)	**Constraint**	**Number of Constraints**
12	Complementarity	$O(12^2) = O(144)$
	Non-Crossing	$O(12^4) = O(20736)$
	Each nucleotide in at most 1 pair	12
	Distance	$O(12^2) = O(144)$
	Stacked Quartets	$O(12^4) = O(20736)$
	Crossed Pairs 1	$O(12^4) = O(20736)$
	Crossed Pairs 2	$O(12^4) = O(20736)$
	At most 10 crossing pairs	$O(12^4) = O(20736)$
20	Complementarity	$O(20^2) = O(400)$

	Non-Crossing	$O(20^4) = O(160000)$
	Each nucleotide in at most 1 pair	20
	Distance	$O(20^2) = O(400)$
	Stacked Quartets	$O(20^4) = O(160000)$
	Crossed Pairs 1	$O(20^4) = O(160000)$
	Crossed Pairs 2	$O(20^4) = O(160000)$
	At most 10 crossing pairs	$O(20^4) = O(160000)$
30	Complementarity	$O(30^2) = O(900)$
	Non-Crossing	$O(30^4) = O(810000)$
	Each nucleotide in at most 1 pair	30
	Distance	$O(30^2) = O(900)$
	Stacked Quartets	$O(30^4) = O(810000)$
	Crossed Pairs 1	$O(30^4) = O(810000)$
	Crossed Pairs 2	$O(30^4) = O(810000)$
	At most 10 crossing pairs	$O(30^4) = O(810000)$

40	Complementarity	$O(40^2) = O(1600)$
	Non-Crossing	$O(40^4) = O(256000)$
	Each nucleotide in at most 1 pair	$O(40^2) = O(1600)$
	Distance	40
	Stacked Quartets	$O(40^4) = O(256000)$
	Crossed Pairs 1	$O(40^4) = O(256000)$
	Crossed Pairs 2	$O(40^4) = O(256000)$
	At most 10 crossing pairs	$O(40^4) = O(256000)$