Bilkent University

Department of Industrial Engineering

2022-2023 Spring Semester

IE 400: Principles of Engineering Management

Project Report

Group 2

Member	Student Number
Efe Erkan	21902248
Recep Uysal	21803637
Uygar Onat Erol	21901908

Instructor: Arnab Basu

Teaching Assistant: Büşra Bayrak

Table of Content

Integer Programming Models

Part A

Aim: Finding a nested pairing that achieves the maximum number of pairs.

Notation:

Suppose the number of nucleotides in a RNA sequence is n and N_i is the i^{th} nucleotide in the sequence.

Decision variables:

$x_{ij} \in \{0, 1\}$ for $1 \leq i < j \leq n$: Pairing between i^{th} and j^{th} nucleotide

Objective Function:

maximize $\sum_{i<j} x_{ij}$

Constraint 1: Each nucleotide must be paired with at most 1 nucleotide.

$\sum_{j<i} x_{ji} + \sum_{i<j} x_{ij} \leq 1$ for $i = \{1, 2, ..., n\}$

Constraint 2: Nucleotide A must be paired with only U in which C must be paired with only G. There is a distance limitation ($minD$) that close nucleotides cannot be paired.

$x_{ij} = 0$ if i^{th} and j^{th} nucleotides are not complementary and $j - i < minD$

Note: For this part, $minD = 4$.

Constraint 3: The pairings are not allowed to cross each other.

$x_{ij} + x_{kl} \leq 1$ for every choice of four valid positions $1 \leq i < k < j < l \leq n$

Part B

Aim: Finding a nested pairing that gives the minimum total free energy associated with pairs in the sequence.

Notation:

Suppose the number of nucleotides in a RNA sequence is n and N_i is the i^{th} nucleotide in the sequence.

Decision variables:

$x_{ij} \in \{0,1\}$ for $1 \leq i < j \leq n$: Pairing between i^{th} and j^{th} nucleotide

Objective Function:

minimize $\sum_{i<j} c_{ij} \cdot x_{ij}$ where $c_{ij} = \begin{cases} -1.33 & (N_i, N_j) = (A, U) \vee (N_i, N_j) = (U, A) \\ -1.45 & (N_i, N_j) = (G, C) \vee (N_i, N_j) = (C, G) \end{cases}$

<u>Constraint 1</u>: Each nucleotide must be paired with at most 1 nucleotide.

$$\sum_{j<i} x_{ji} + \sum_{i<j} x_{ij} \leq 1 \text{ for } i = \{1, 2, ..., n\}$$

<u>Constraint 2:</u> Nucleotide A must be paired with only U in which C must be paired with only G. There is a distance limitation ($minD$) that close nucleotides cannot be paired.

$x_{ij} = 0$ if i^{th} and j^{th} nucleotides are not complementary and $j - i < minD$

Note: For this part, $minD = 4$.

<u>Constraint 3</u>: The pairings are not allowed to cross each other.

$x_{ij} + x_{kl} \leq 1$ for every choice of four valid positions $1 \leq i < k < j < l \leq n$

Part C

Aim: Finding a nested pairing that gives the minimum total free energy associated with pairs in the sequence.

Notation:

Suppose the number of nucleotides in a RNA sequence is n and N_i is the i^{th} nucleotide in the sequence.

Decision variables:

$x_{ij} \in \{0, 1\}$ for $1 \leq i < j \leq n$: Pairing between i^{th} and j^{th} nucleotide

Objective Function:

minimize $\sum_{i<j} c_{ij} \cdot x_{ij}$ where $c_{ij} = \begin{cases} -1.33 & (N_i, N_j) = (A, U) \vee (N_i, N_j) = (U, A) \\ -1.45 & (N_i, N_j) = (G, C) \vee (N_i, N_j) = (C, G) \end{cases}$

Constraint 1: Each nucleotide must be paired with at most 1 nucleotide.

$$\sum_{j<i} x_{ji} + \sum_{i<j} x_{ij} \leq 1 \text{ for } \quad i = \{1, 2, ..., n\}$$

Constraint 2: Nucleotide A must be paired with only U in which C must be paired with only G. There is a distance limitation ($minD$) that close nucleotides cannot be paired.

$x_{ij} = 0$ if i^{th} and j^{th} nucleotides are not complementary and $j - i < minD$

Note: For this part, $minD = 7$.

Constraint 3: The pairings are not allowed to cross each other.

$x_{ij} + x_{kl} \leq 1$ for every choice of four valid positions $1 \leq i < k < j < l \leq n$

Part D

Aim: Finding the minimum total free energy corresponding to stacked pairs.

Notation:

Suppose the number of nucleotides in a RNA sequence is n and N_i is the i^{th} nucleotide in the sequence.

Decision variables:

$x_{ij} \in \{0,1\}$ for $1 \leq i < j \leq n$: Pairing between i^{th} and j^{th} nucleotide

$s_{ij} \in \{0,1\}$ for $1 \leq i < j \leq n$: Pairing between i^{th} and j^{th} nucleotide is the first pairing in a stack pair, i.e. whether (i,j) and ($i+1$,$j-1$) are both in the nested pairing

Objective Function:

minimize $\sum_{i<j} e_{ij} \cdot s_{ij}$ where e_{ij} is weight of the stacked pair from Table 1

Constraint 1: Each nucleotide must be paired with at most 1 nucleotide.

$$\sum_{j<i} x_{ji} + \sum_{i<j} x_{ij} \leq 1 \text{ for } i = \{1, 2, ..., n\}$$

Constraint 2: Nucleotide A must be paired with only U in which C must be paired with only G. There is a distance limitation ($minD$) that close nucleotides cannot be paired.

$x_{ij} = 0$ if i^{th} and j^{th} nucleotides are not complementary and $j - i < minD$

Note: For this part, $minD = 4$.

Constraint 3: The pairings are not allowed to cross each other.

$x_{ij} + x_{kl} \leq 1$ for every choice of four valid positions $1 \leq i < k < j < l \leq n$

Constraint 4: If both (i, j) and $(i+1, j-1)$ are in the nested pairing, then $s_{ij} = 1$

$$x_{ij} + x_{(i+1)(j-1)} - s_{ij} \leq 1$$

Constraint 5: If $s_{ij} = 1$, then both (i, j) and $(i+1, j-1)$ are in the nested pairing.

$$2s_{ij} - x_{ij} - x_{(i+1)(j-1)} \leq 0$$

Part E

Aim: Finding the minimum total free energy corresponding to stacked pairs by incorporating pseudoknots.

Notation:

Suppose the number of nucleotides in a RNA sequence is n and N_i is the i^{th} nucleotide in the sequence.

Decision Variables:

$x_{ij} \in \{0, 1\}$ for $1 \leq i < j \leq n$: Upward pairing between i^{th} and j^{th} nucleotide

$s_{ij} \in \{0, 1\}$ for $1 \leq i < j \leq n$: Upward pairing between i^{th} and j^{th} nucleotide is the first pairing in a stack pair, i.e. whether (i, j) and $(i+1, j-1)$ are both in the upward nested pairing

$y_{ij} \in \{0, 1\}$ for $1 \leq i < j \leq n$: Downward pairing between i^{th} and j^{th} nucleotide

$q_{ij} \in \{0, 1\}$ for $1 \leq i < j \leq n$: Downward pairing between i^{th} and j^{th} nucleotide is the first pairing in a stack pair, i.e. whether (i, j) and $(i+1, j-1)$ are both in the downward nested pairing

Objective Function:

minimize $\sum_{i<j} e_{ij} \cdot s_{ij} + \sum_{i<j} e_{ij} \cdot q_{ij}$ where e_{ij} is weight of the stacked pair from Table 1

Constraint 1: Each nucleotide must be paired with at most 1 nucleotide.

$$\sum_{j<i} x_{ji} + \sum_{j<i} y_{ji} + \sum_{i<j} x_{ij} + \sum_{i<j} y_{ij} \leq 1 \quad \text{for} \quad i = \{1, 2, ..., n\}$$

Constraint 2: Nucleotide A must be paired with only U in which C must be paired with only G. There is a distance limitation ($minD$) that close nucleotides cannot be paired.

$x_{ij} = 0$ if i^{th} and j^{th} nucleotides are not complementary and $j - i < minD$

$y_{ij} = 0$ if i^{th} and j^{th} nucleotides are not complementary and $j - i < minD$

Note: For this part, $minD = 4$.

Constraint 3: The pairings are not allowed to cross each other.

$x_{ij} + x_{kl} \leq 1$ for every choice of four valid positions $1 \leq i < k < j < l \leq n$
$y_{ij} + y_{kl} \leq 1$ for every choice of four valid positions $1 \leq i < k < j < l \leq n$

Constraint 4: If both (i, j) and $(i+1, j-1)$ are in the upward nested pairing, then $s_{ij} = 1$

$$x_{ij} + x_{(i+1)(j-1)} - s_{ij} \leq 1$$

Constraint 5: If $s_{ij} = 1$, then both (i, j) and $(i+1, j-1)$ are in the upward nested pairing.

$$2s_{ij} - x_{ij} - x_{(i+1)(j-1)} \leq 0$$

Constraint 6: If both (i, j) and $(i+1, j-1)$ are in the downward nested pairing, then $q_{ij} = 1$

$$y_{ij} + y_{(i+1)(j-1)} - q_{ij} \leq 1$$

Constraint 7: If $q_{ij} = 1$, then both (i, j) and $(i+1, j-1)$ are in the downward nested pairing.

$$2q_{ij} - y_{ij} - y_{(i+1)(j-1)} \leq 0$$

Dynamic Programming Model for Part F

Aim: Minimizing the total energy using the information given in Part B and including the energy released from the formation of stacking pairs.

Notation:

Let $W(i,j)$ denote the total energy level of the subsequence between i^{th} and j^{th} nucleotides.

Let N_i denote the i^{th} nucleotide in the RNA sequence.

Let $c_{i,j}$ denote the energy value between i^{th} and j^{th} nucleotides:

$$c_{ij} = \begin{cases} -1.33 & (N_i, N_j) = (A, U) \vee (N_i, N_j) = (U, A) \\ -1.45 & (N_i, N_j) = (G, C) \vee (N_i, N_j) = (C, G) \\ \end{cases}$$

Let $S(i,j)$ denote the energy value for stacking pairs:

$S(i,j) = 0$ if i^{th} and j^{th} nucleotides are not the first pair of a stack.

Otherwise, $S(i,j)$is calculated using Table 1.

Dynamic Programming Recursion:

$$W(i,j) = \min \begin{cases} W(i,j-1) \\ \min_{i \le t < j}\{W(i,t-1) + c_{t,j} + W(t+1,j-1) + S(t,j)\} \end{cases}$$

Results

The RNA sequence used for the solutions is (Group 2 RNA sequence):

5'-AAGCUCCUUGGUCCGGGCCAUUAAAGCCCGGAGCAGGAACAGAUGCUAU CCUAUUAAAGUAGGGUUACCC-3'

The execution environment is Lenovo Legion Y540, Intel Core i7 9750H, 16GB RAM, Ubuntu 22.04.

Part A

Maximum number of pairs: 25

Pairs: (1, 9), (2, 8), (3, 7), (10, 70), (11, 34), (12, 32), (13, 31), (14, 30), (15, 29), (16, 28), (17, 27), (19, 26), (21, 25), (36, 69), (37, 68), (38, 66), (39, 65), (40, 64), (41, 49) or (43, 49), (44, 48), (50, 63), (51, 62), (52, 61), (53, 60), (54, 58)

Execution time: 8826 ms

Explanation: In the resulting pairs, it seems that one solution contains the pair (41, 49) and the other solution contains (43, 49). Therefore, alternate solutions exist for the RNA sequence.

Part B

Minimum Energy: -35.05

Number of Pairs: 25

Pairs: (1, 9), (2, 8), (3, 7), (10, 70), (11, 34), (12, 32), (13, 31), (14, 30), (15, 29), (16, 28), (17, 27), (18, 26), (21, 25), (36, 69), (37, 68), (38, 66), (39, 65), (40, 64), (41, 49), (42, 46), (50, 63), (51, 62), (52, 61), (53, 60), (54, 58)

Execution Time: 13769 ms

Explanation: Compared to Part A, the number of pairs is identical even though the total energy is minimized. In terms of execution time, the difference is approximately 3 seconds.

Part C

Minimum Energy: -29.61

Number of Pairs: 21

Pairs: (3, 70), (4, 42), (5, 38), (6, 37), (7, 36), (8, 35), (10, 34), (12, 32), (13, 31), (14, 30), (15, 29), (16, 28), (17, 27), (19, 26), (45, 68), (47, 67), (48, 65), (50, 63), (51, 62), (52, 61), (53, 60)

Execution Time: 13556 ms

Explanation: Compared to Part B, the number of pairs is less because the minimum distance for pairing becomes 4 to 7. Therefore, less number of pairs is expected. In addition, there is no significant difference in execution time.

Part D

Minimum Energy: -67.5

Number of Pairs: 20

Pairs: (1, 55), (2, 54), (3, 34), (4, 33), (5, 32), (13, 31), (14, 30), (15, 29), (16, 28), (17, 27), (18, 26), (35, 52), (36, 51), (37, 50), (38, 49), (41, 47), (42, 46), (62, 70), (63, 69), (64, 68)

Execution Time: 1596 ms

Part E

Minimum Energy: -103.6

Number of Pairs: 15 + 15 = 30

Upward Pairs: (5, 38), (6, 37), (7, 36), (13, 31), (14, 30), (15, 29), (16, 28), (17, 27), (39, 60), (40, 59), (42, 50), (43, 49), (62, 70), (63, 69), (64, 68)

Downward Pairs: (1, 22), (2, 21), (9, 20), (10, 19), (11, 18), (23, 55), (24, 54), (25, 52), (26, 51), (32, 47), (33, 46), (34, 45), (35, 44), (56, 66), (57, 65)

Execution Time: 3671 ms

Explanation: In this part, upward and downward pairs are exchangable.

Part F

Minimum Energy: -71.22

Number of Pairs: 24

Pairs: (1, 9), (2, 8), (3, 7), (10, 51), (11, 50), (12, 48), (13, 31), (14, 30), (15, 29), (16, 28), (17, 27), (18, 26), (21, 25), (32, 47), (33, 46), (34, 45), (35, 44), (36, 40), (52, 61), (53, 60), (54, 58), (62, 70), (63, 69), (64, 68)

Execution Time: 80 ms