Genome Rearrangement

↳ reversals

$d(H, M)$: distance between Human & Mouse Genome

H (reference) : $I = 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8$

Mouse : $\Pi = 3\ 6\ 8\ 5\ 4\ 7\ 1\ 2$

Given a reference genome (identity permutation)

$I = 1\ 2\ 3 \ldots n$

and its permutation Π, use reversals to transform Π into I (using the least # of reversals)

---

1) Prefix-based reversals

$[3\ 6\ 8\ 5\ 4\ 7\ 1]2$

⇓

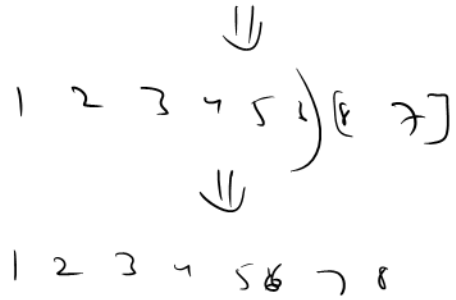
$1)[7\ 4\ 5\ 8\ 6\ 3\ 2]$

⇓

$1\ 2\ 3)[6\ 8\ 5\ 4]7$

⇓

$1\ 2\ 3\ 4\ 5)[8\ 6]7$

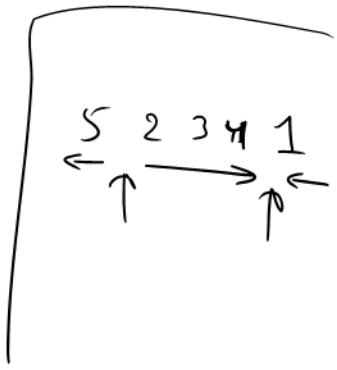


5 reversal operations

$$\left\lceil \frac{b}{2} \right\rceil \leq d(G_1, G_2) \leq n - 1$$

lower bound

7 break points



a single reversal can eliminate at most 2 breakpoints

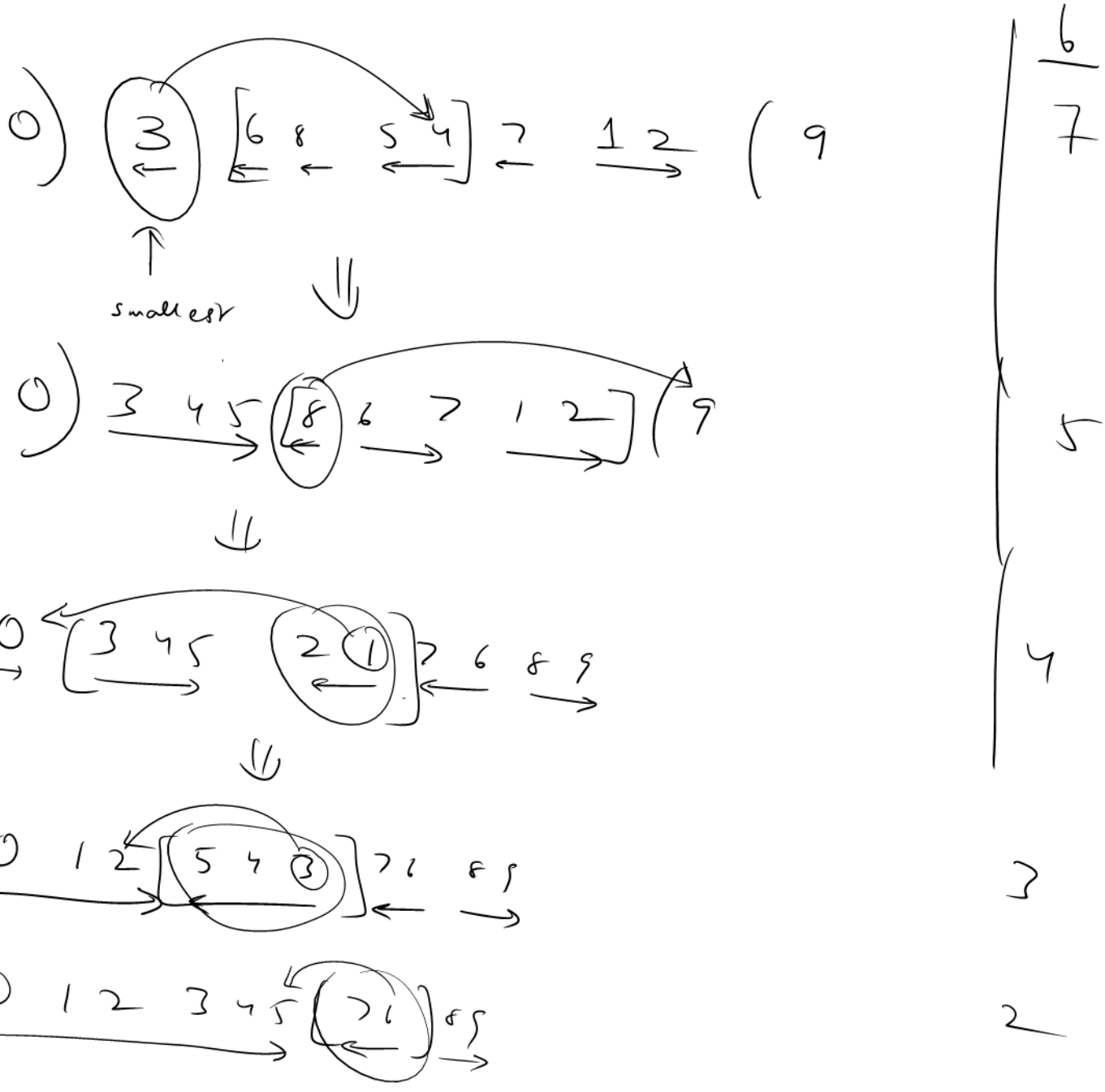
2) 4-opt solution

$$opt \geq \left\lceil \frac{6}{2} \right\rceil$$

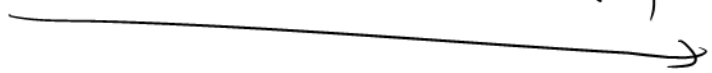
?

will take at most $\left\lceil \frac{6}{2} \right\rceil \times 4$ reversal

$= 12$



0 1 2 3 4 5 6 7 8 9



0

Breakpoint reversal

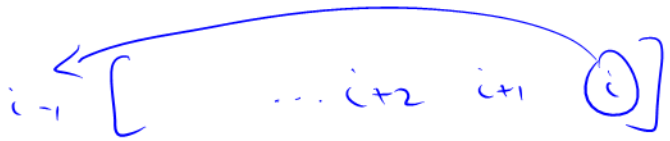
1) check if there is a decreasing block
   if not flip one of the increasing ones

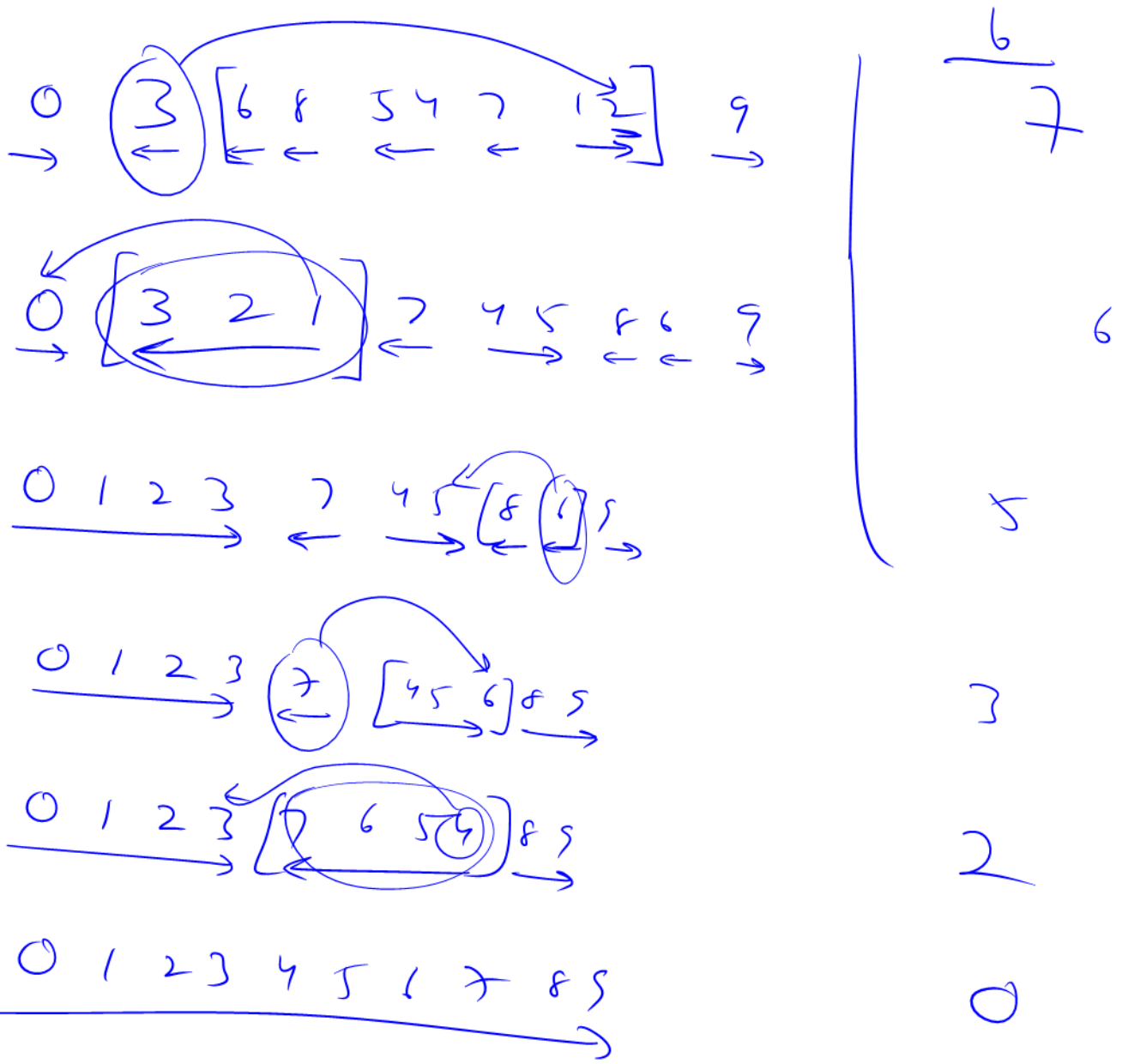
2) find the smallest decreasing block

3) identify the next smaller block (which must be increasing)

4) flip the range

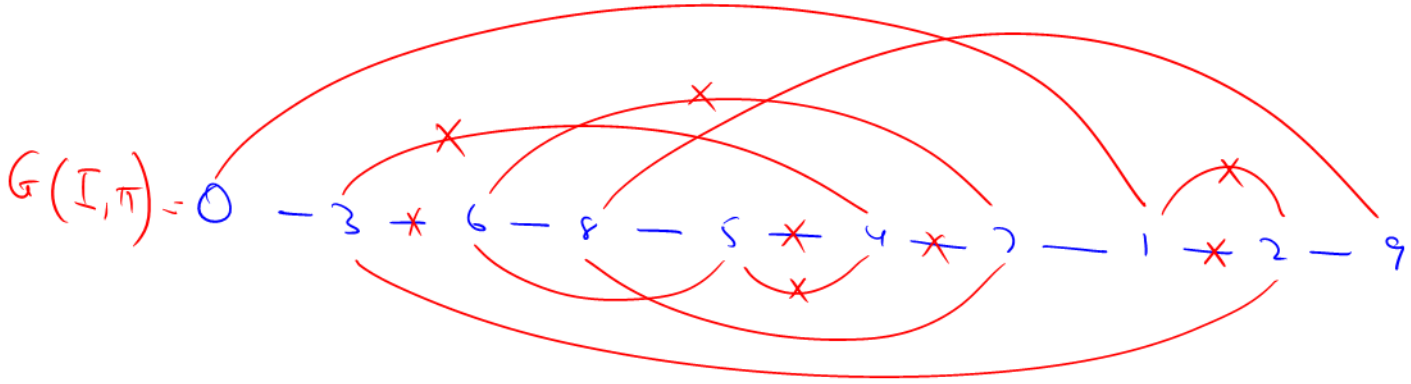
guarantee that 1 breakpoint is removed





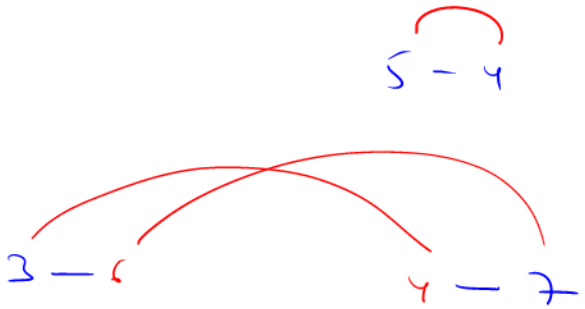
I : 0 1 2 3 4 5 6 7 8 9

π : 0 3 6 8 5 4 7 1 2 9



Task: find the maximum number of edge-disjoint alternating colored cycles in $G(I, \pi)$

say $C(G)$



$$d(G_i, G_j) \geq \underbrace{n+1 - C(G)}_{\text{lower-bound}} \geq \left\lceil \frac{b}{2} \right\rceil$$

finding the optimal / least # of reversals is NP-hard!

Add biology to make the problem easier!

signed permutations

$+$

$\xrightarrow{A\ C\ G\ T}$ top strand

$-$

$\xleftarrow{A\ C\ G\ T}$ bottom strand

π : Signed permutation

goal is to transform into signed identity permutation

I : +1 +2 −3 +4 +5 ← reference signed permutation

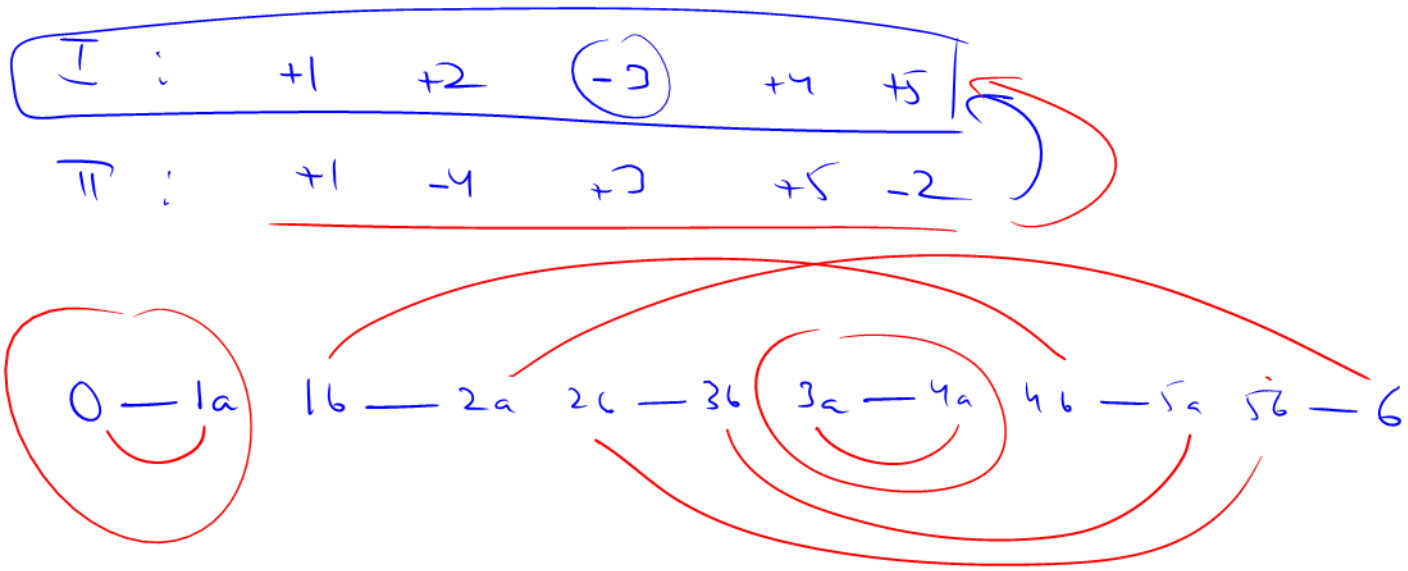
H : +1 +5 −2 +3 +4

M : +1 −3 +2 +4 −5

π (+1 −4 +3 +5 −2)

Create a new graph $G(I, \pi)$ ← finding the cycle number corresponds to finding the connected components

$$d(H, M) \geq n + 1 - c$$



C : 3

n + 1 − 3

n = 6 optimal steps : 6 − 3 = 3

1 [-4 3 5] -2

⇓

1 -5 -3 +4 -2