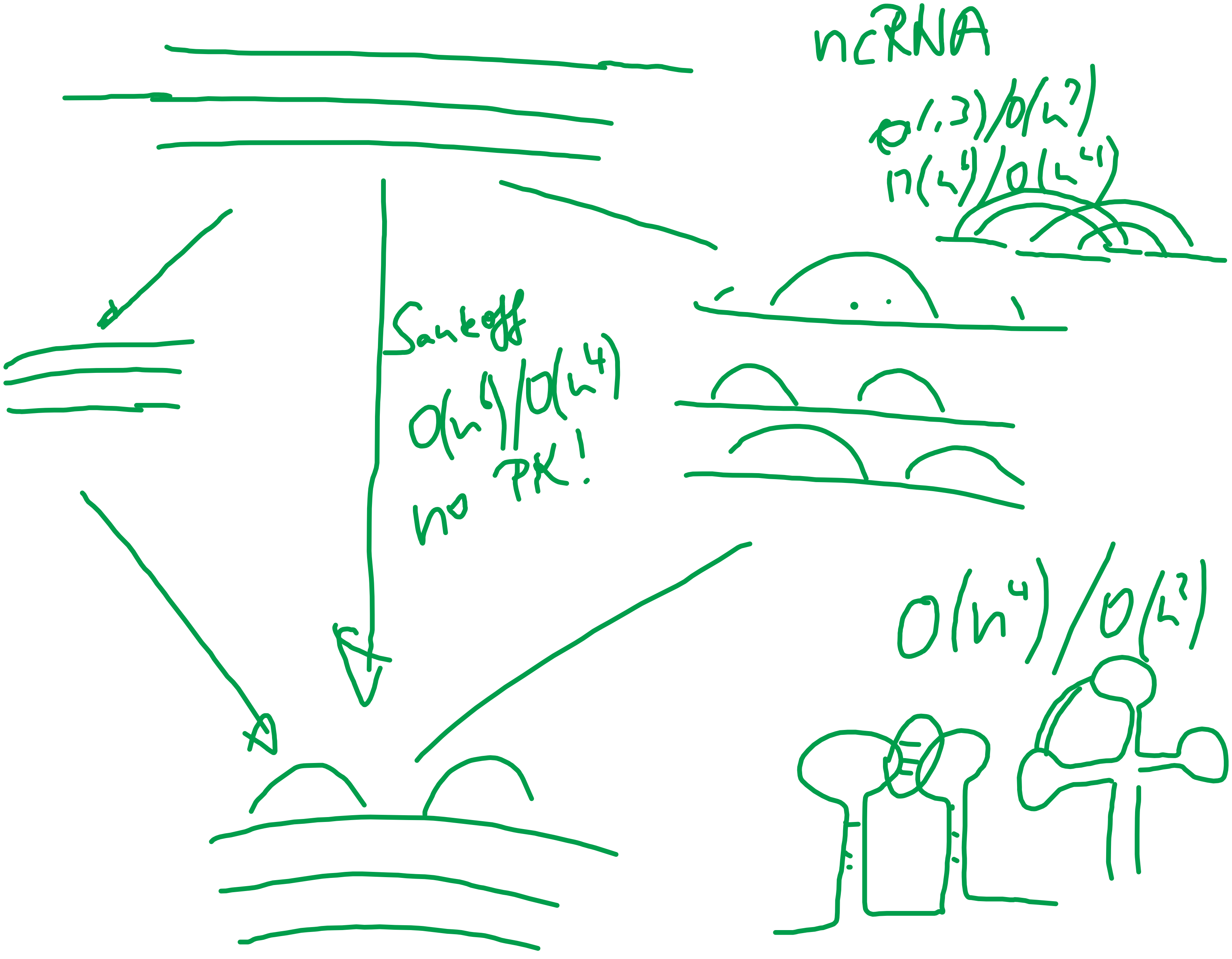
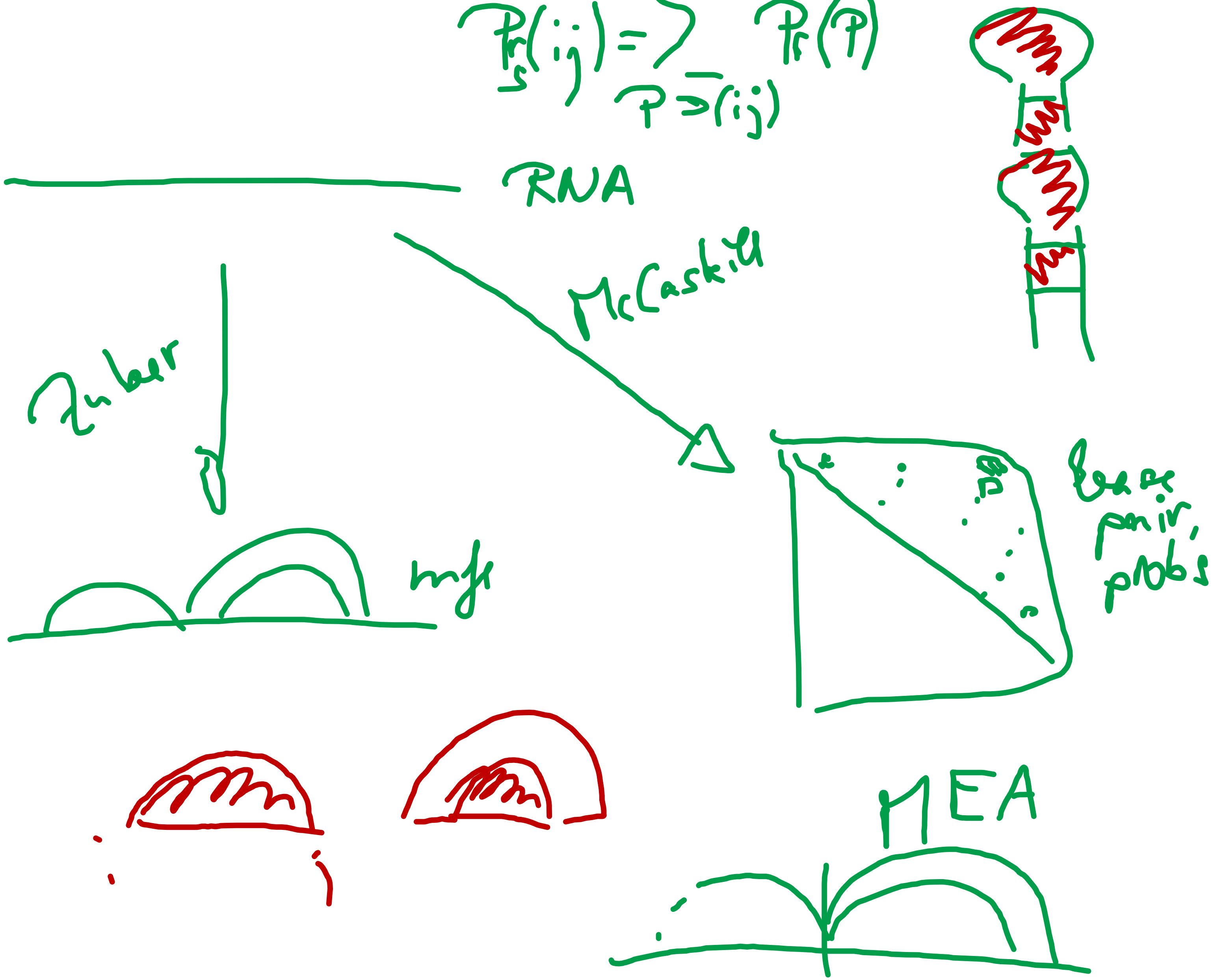

ncRNA
O(n^3)/O(n^2)
O(n^5)/O(n^4)
Sankoff
O(n^6)/O(n^4)
no PK!
O(n^4)/O(n^3)

$$Pr_S(ij) = \sum_{P \ni (ij)} Pr(P)$$



$R = R[1] \ldots R[n]$; $P = (P_{ii'})_{ii'}$ RIBOSUM

$S = S[1] \ldots S[n]$; $Q = (Q_{jj'})_{jj'}$

$$\mathrm{Score}(A \mid R, S) = \sum_{(i,j)(i',j') \in \mathcal{A}} P(ii') \, Q(jj') \cdot \tau(R[i] R[i'] S[j] S[j'])$$

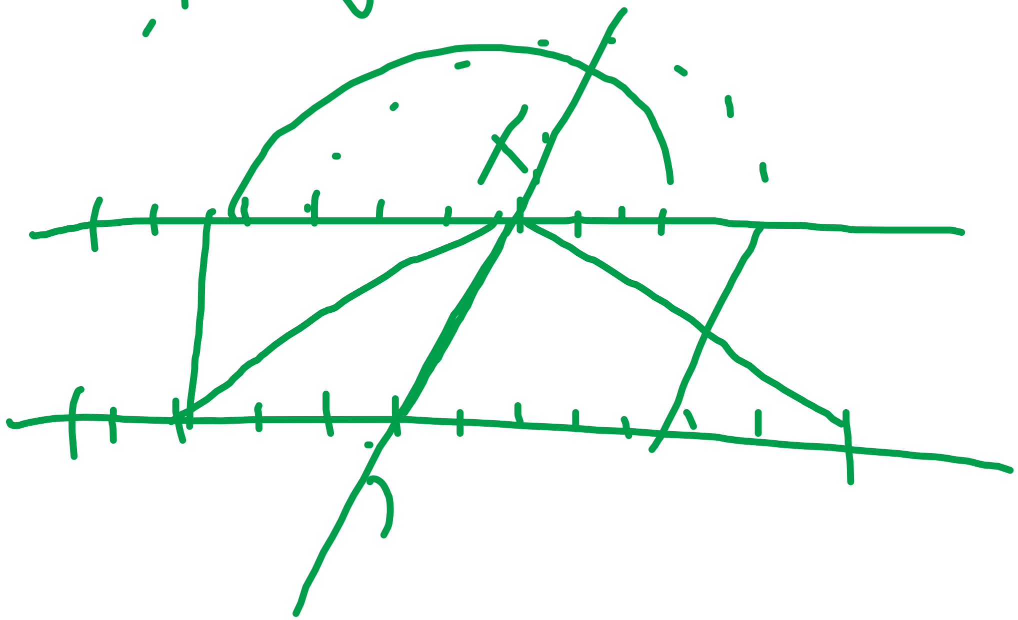
$$+ \alpha \sum_{(i,j) \in \mathcal{A}} P_u(i) \, Q_u(j) \cdot G(R[i] S[j])$$

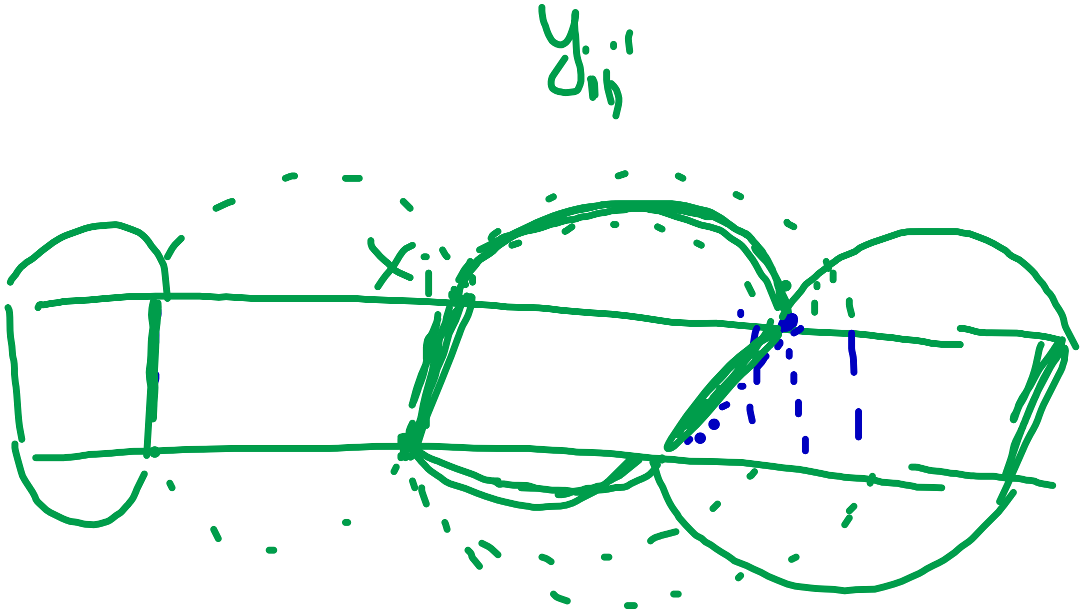
$$+ \text{gap cost}$$

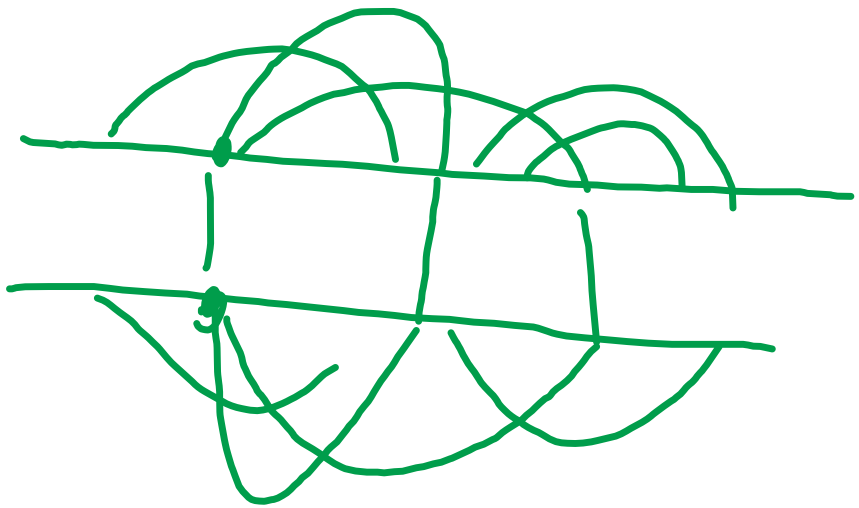
$X_1 \ldots\ldots X_n$

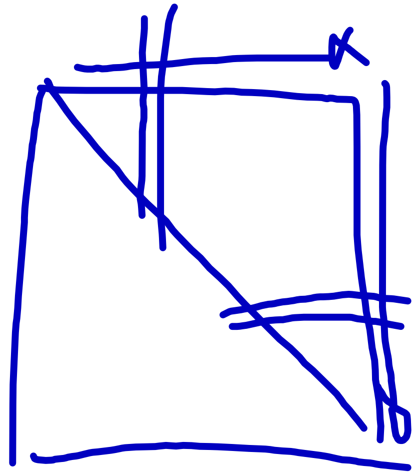
$X_i = j$

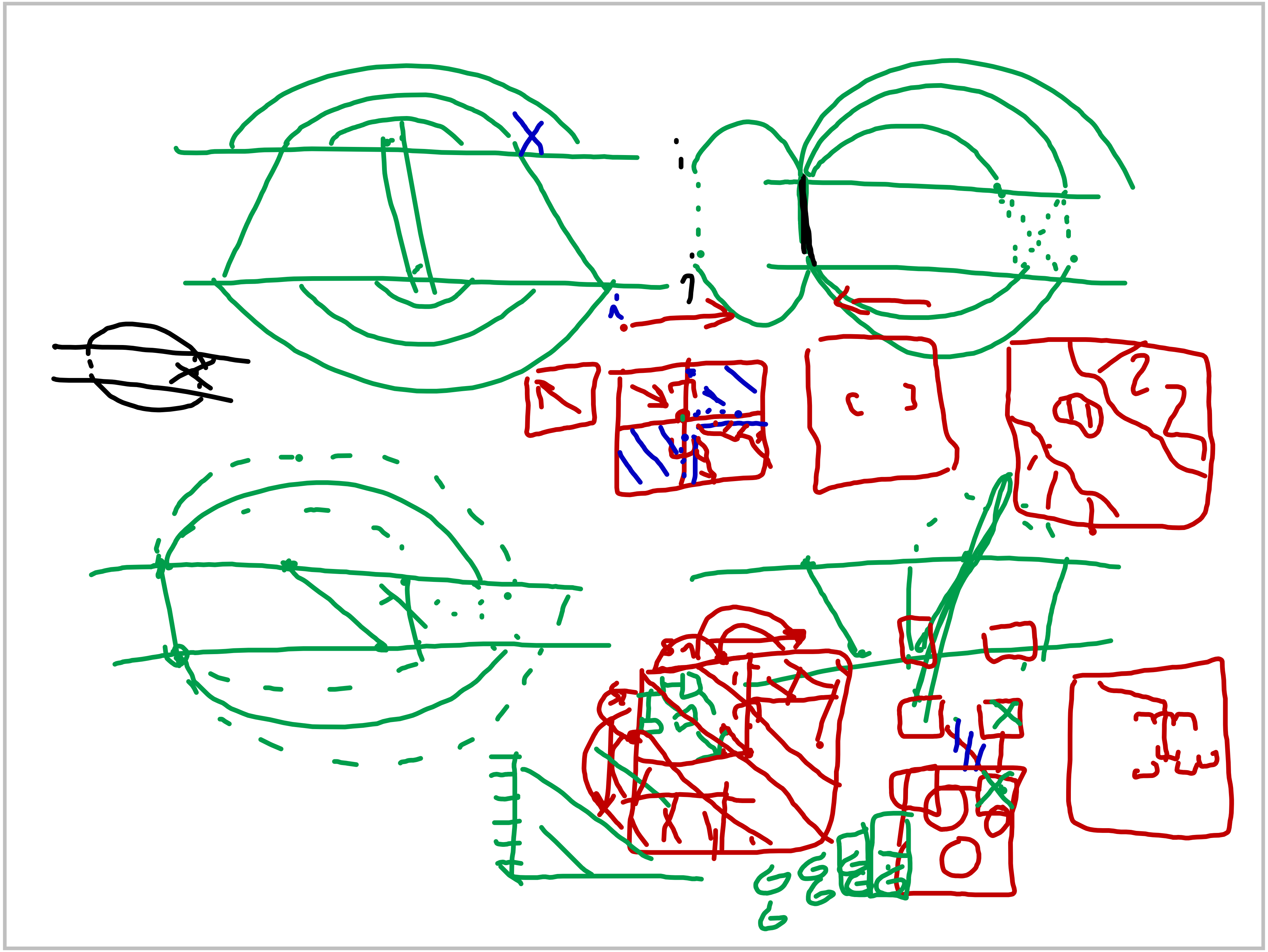
"edge $(i,j) \in \mathcal{A}$"




y_{ij}
x




x



$$X_1, \dots X_n \in \{-, 0, \dots, m\}, \; S \in \mathbb{N}$$

$X_i = j$:⟺ position i aligned to position j

$X_i = -$ " " " " to a gap

Sc ⟺ score

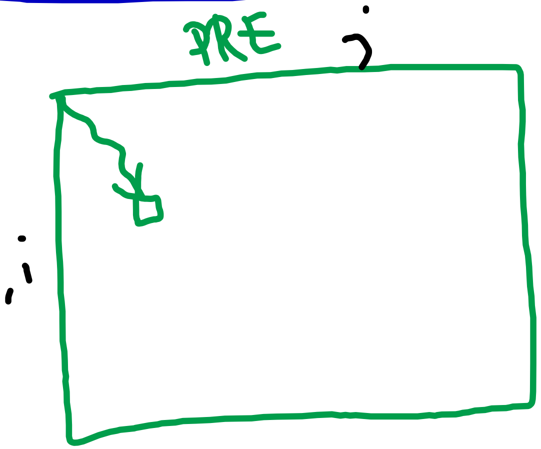
$\mathrm{sorted}(X_1, \dots, X_n)$

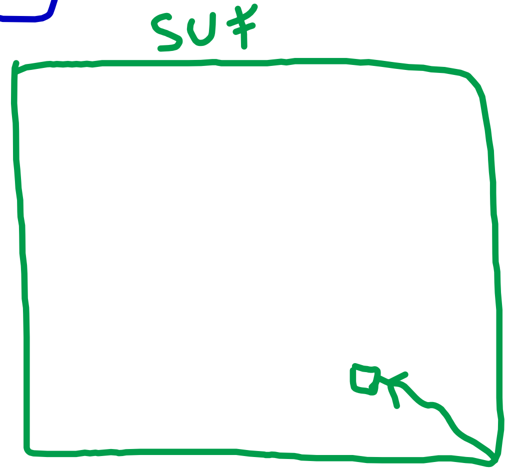
$\mathrm{alignmentScore}(X_1, \dots X_n, Sc)$ ≈ S is the score for the alignment $X_1, \dots, X_n$

R, S Sequences

P, Q Dot Plots

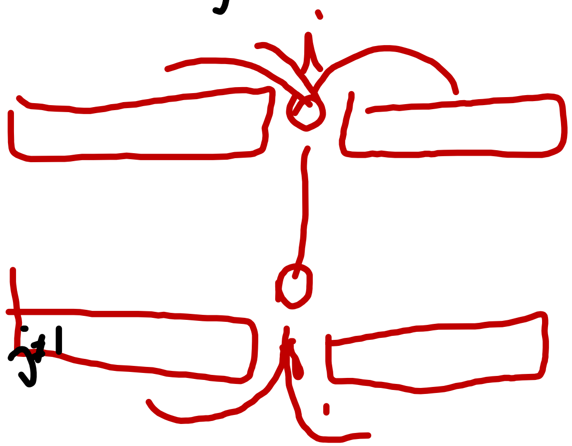
alignmentScan$(X_1, \ldots, X_m, Sc)$





$$PRE_{i,j} = \max \begin{cases} PRE_{i-1,j-1} + \beta(i,j) & \text{if } j \in D(X_i) \\ PRE_{i-1,j} + \gamma \\ PRE_{i,j-1} + \gamma \end{cases}$$

$SUF_{i,j}$



$$u(i \sim j) = PRE_{i-1,j-1} + \beta(i,j) + SUF_{i+1,j+1}$$

Simplest idea

$$\beta(i,j) = \sigma(R[i], S[j]) \cdot P_u(i) \cdot Q_u(j)$$

"match of unpaired bases"

$$+ \sum_{\substack{(i,i')\in P,\\ (j,j')\in Q,\\ j'\in D(X_{i'})}} \frac{1}{2} P(ii') \cdot Q(jj')\, \tau(R[i] R[i'] S[j] S[j'])$$

symmetrically

$$+ \sum_{\substack{(i',i)\in P\\ (j',j)\in Q\\ j'\in D(X_{i'})}} \frac{1}{2}$$

2. idea

$$\beta(i,j) = \sigma(R[i], S[j]) \cdot P_u(i) \cdot Q_u(j)$$

"match of unpaired bases"

$$+ \sum_{\substack{(i,i')\in P,\ j'\in D(X_{i'}) \\ (j,j')\in Q}} \max \tfrac{1}{2} P(i,i') \cdot Q(j,j')\, \tau(R[i]R[i']\, S[j]\, S[j'])$$

$$+ \sum_{\substack{(i',i)\in P \\ (j',j)\in Q \\ j'\in D(X_{i'})}} \max_{\substack{j'\in D(X_{i'}) \\ (i',i)\in Q}}{}' \ 2$$

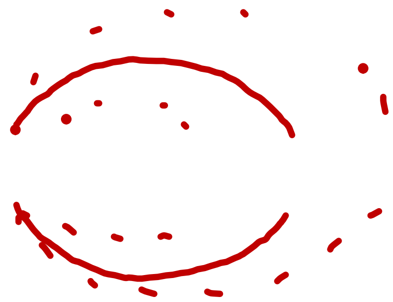
~~symmetrically~~

i	j	score

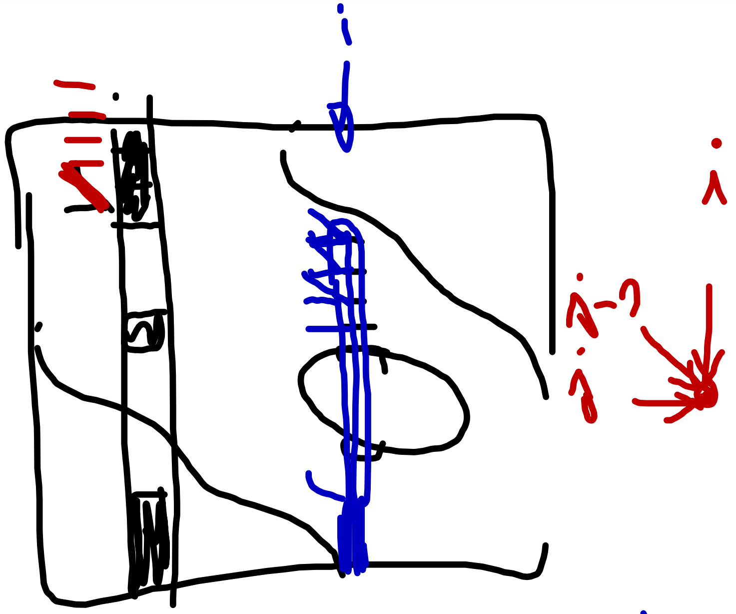
3. idea

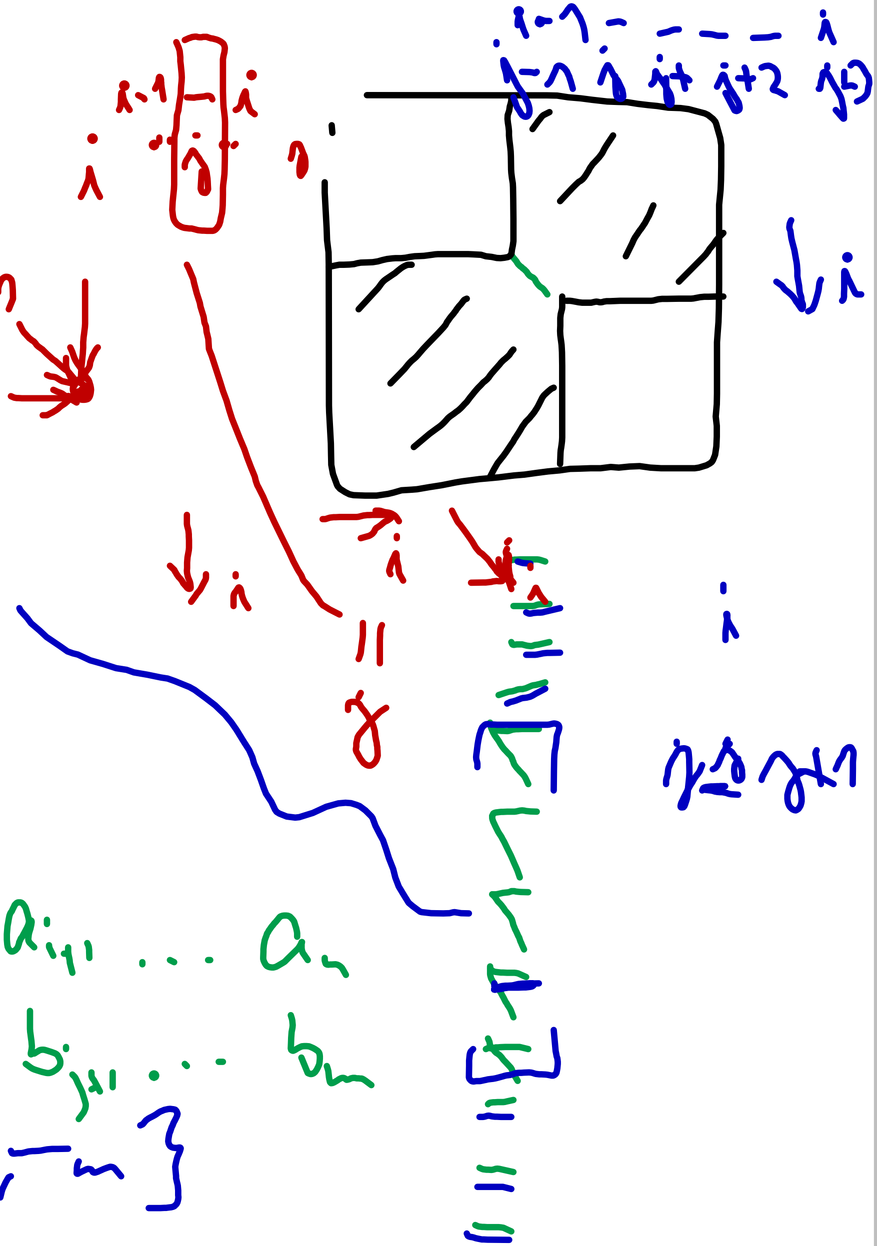
$$\beta(i,j) = \sigma(R[i], S[j]) \cdot P_u(i) \cdot Q_u(j)$$ "match of unpaired bases"

$$+ \min \begin{cases} \displaystyle\sum_{(i i') \in P} \max_{\substack{j' \in D(X_{i'}) \\ (j j') \in Q}} \tfrac{1}{2} P(i i') \cdot Q(j j')\, \tau(R[i] R[i'] S[j] S[j']) \\[2em] \displaystyle\sum_{(j j') \in Q} \max_{\substack{(i i') \in P \\ j' \in D(X_{i'})}} \tfrac{1}{2} P(i i') \cdot Q(j j')\, \tau(R[i] R[i'] S[j] S[j']) \end{cases}$$



and symm.





$X_1 \dots\dots X_m$

$$\begin{array}{ccc|ccc} a_1 & \dots & a_i & a_{i+1} & \dots & a_n \\ b_1 & \dots & b_j & b_{j+1} & \dots & b_m \end{array}$$

$$D(X_i) = \{1, \dots, m, -0, \dots, -m\}$$

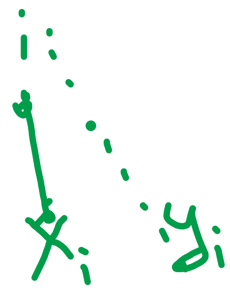
first try:

$\cancel{X_i = j}$ $\equiv$ "match $i \sim j$"

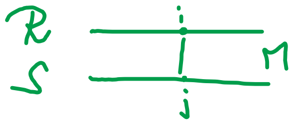
$y_i = j$ $\equiv$ "alignment can be decomposed into an alignment of $a_1 \dots a_i$, $b_1 \dots b_j$ and an alignment of $a_{i+1} \dots a_n$, $b_{j+1} \dots b_m$."

$X_i \leq y_i$ or $X_i = -$

$X_i \neq - \longrightarrow X_i = y_i$



Constr. model



for $i = 1 \ldots n$:

Var: $M_i = j$ "match $i \sim j$"

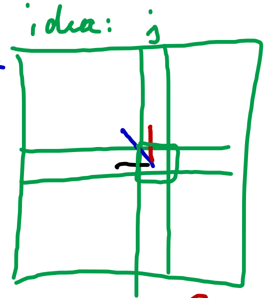
$G_i = j$ "gap $i \sim - \; j \; j{+}1$"

$H_i \ni j$ "gap $- \; i \; i{+}1 \sim j$"

($Sp_i = j$ "split $= \overset{}{\underset{j}{|}} =$")*

$Score \in \mathbb{N}$

idea:



Dom: $D(M_i) = \{1, \ldots, m, \perp\}$

$D(G_i) = \{0, \ldots, m, \perp\}$

$D(H_i) = \{\{\}, \ldots, \{1..m\}\}$

$(D(Sp_i) = \{0, \ldots, m\})^*$

$M_i = j$ $G_i = j$

$H_i \ni j$ $Sp_i = j$

Constraints:

at least one of M_i, G_i, H_i is not undef.

exactly one of M_i, G_i is $\perp$

$H_i \neq \{\} \rightarrow \min(H_i) - 1 = \begin{cases} M_i & M_i \neq \perp \\ G_i & G_i \neq \perp \end{cases}$

$\left.\begin{array}{l} \text{sorted}(M_1, \ldots, M_n) \\ \text{sorted}(G_1, \ldots, G_n) \\ \text{sortedSet}(H_1, \ldots, H_n) \end{array}\right\}$ likely to be repeated

$\text{alignmentScore}(\vec{M}, \vec{G}, \vec{H}, Score)$ * too complicated, replace by alignment score prop.

$(\text{sorted}(Sp_1 \ldots Sp_n))$

$\text{Sorted}(X_1 \ldots X_k)$

$\forall 1 \leq i < j \leq k$:

$X_i \neq \perp \wedge X_j \neq \perp$

$\longrightarrow X_i < X_j$

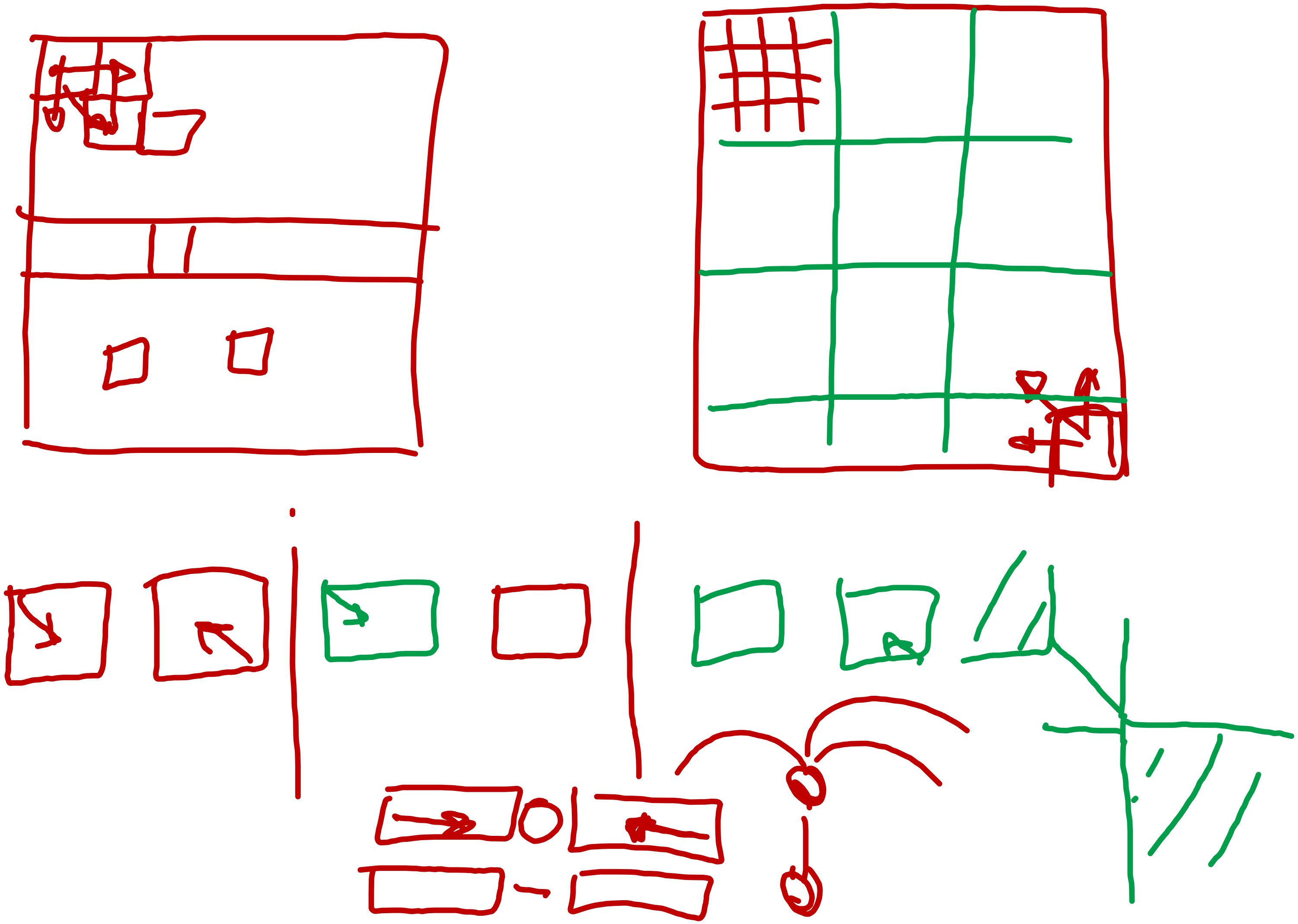
$\text{SortedSet}(X_1 \ldots X_k)$

$\forall 1 \leq i < j \leq k$:

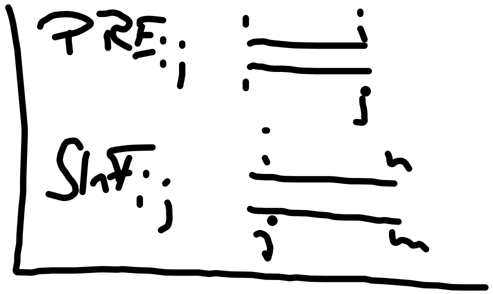
$X_i \neq \{\} \wedge X_j \neq \{\}$

$\rightarrow \max(X_i) < \min(X_j)$

alignmentScore$(\vec{M}, \vec{G}, \vec{H}, \text{Score})$



Pruning & Upper Bounds



$$U(i\sim j) = PRE_{i-1,j-1} + \beta(i,j) + SUF_{i+1,j+1}$$

$$U(i\sim j) < \min(\text{Score}) \rightarrow M_i \neq j$$

$$U\left(i\sim -_{j,j+1}\right) = PRE_{i-1,j} + \gamma + SUF_{i+1,j+1}$$

$$U\left(i\sim -_{j,j+1}\right) < \min(\text{Score}) \rightarrow G_i \neq j$$

$$U\left(-_{i,i+1}\sim j\right) = PRE_{i,j-1} + \gamma + SUF_{i+1,j+1}$$

$$U\left(-_{i,i+1}\sim j\right) < \min(\text{Score}) \rightarrow H_i \neq j$$