1 Architecture Team: Discuss TSP interface.
2 Class: Discussion on TSP: Go over LP-relaxation?
3 Optical Mapping:
4 Quiz.
5 Optical Mapping Architecture Team.

Optical Mapping:

$\text{Restriction Map Model:}$

$\text{SMRM (Single Molecule Restriction Map)}$

A vector with ordered set of rational numbers on the open interval $(0, 1)$:

$$D_j = (s_{ij}, s_{2j}, \ldots, s_{Mjj}),$$

$$0 < s_{ij} < s_{2j} < \ldots < s_{Mjj} < 1, \quad s_{ij} \in \mathbb{Q}$$

$\text{Problem}$

Data: A collection of SMRM vectors:

$$D_1, D_2, \ldots, D_m$$

Desiderata: Compute a consensus vector

$$H = (h_1, h_2, \ldots, h_N)$$

such that $H$ is "consistent" with each $D_j$.

$$H^* = \arg \min_{H, j} \text{dist}(D_j, H).$$
Consensus:

\[ H^* = \underset{H, j}{\text{argmin}} \quad \text{dist} (D_j, H) \]

\[ D_j = (s_{1j}, s_{2j}, \ldots, s_{M_j j}) \]

\[ \Rightarrow D_j + c = (s_{1j} + c, s_{2j} + c, \ldots, s_{M_j j} + c) \quad c \in [0, 1) \]

\[ \text{dist} (D_j, H) = \text{dist} (D_j + c, H) \]

\[ D_j^R = (1 - s_{M_j j}, \ldots, 1 - s_{2j}, 1 - s_{1j}) \]

\[ \text{dist} (D_j^R, H) = \text{dist} (D_j, H) \]

Consensus:

\[ H^* = \underset{H, j}{\text{argmin}} \left\{ \text{dist} (D_j, H), \text{dist} (D_j^R, H) \right\} \]

or

\[ H^* = \underset{H, j}{\text{argmin}} \left\{ \text{dist} (D_j + c, H) \mid -s_{ij} < c < 1 - s_{ij} \right\} \]

or

\[ H^* = \underset{H, j}{\text{argmin}} \left\{ \text{dist} (D_j + c, H), \text{dist} (D_j^R + c, H) \mid -s_{ij} < c < 1 - s_{ij} \right\} \]

Assume some distribution generating \( D_j \)'s

\[ \langle H \rangle = \underset{H}{\text{argmin}} \sum_j \min \left\{ \text{dist} (D_j + c, H), \text{dist} (D_j^R + c, H) \mid 1 - s_{ij} < c < 1 - s_{M_j j} \right\} \]
**TOY EXAMPLE PROBLEM.**

(Unknown Orientation:)

Data: A set of ordered vectors with rational entries in the open interval $(0, 1)$:

\[ D_1, D_2, \ldots, D_k, D_{k+1}, \ldots, D_m \]

A rational number $p_c \in (0, 1)$ and an integer $N$.

An admissible alignment of the data can be represented as

\[ D'_1, D'_2, \ldots, D'_k, D'_{k+1}, \ldots, D'_m \]

where

\[ D'_j \in \{ D_j, D_j^R \} \quad (1 \leq j \leq k) \]

and

\[ D'_j = D_j \quad (j > k) \]

An Alignment \((\mathcal{A}_k)\)

For any rational number $h_i \in [0, 1]$, define an indicator variable

\[ m_{ijk} = \begin{cases} 1 & \text{if } h_i \in D'_i \\ 0 & \text{otherwise} \end{cases} \]

Define a characteristic function

\[ \chi_k : [0, 1] \to \{ 0, 1 \} \]

\[ h_i \mapsto \begin{cases} 1 & \text{if } \sum m_{ijk} > p_m \end{cases} \]
Desiderata: Find an admissible alignment $A_k$ such that
\[ |\{ h \in [0, 1] \mid X_k(h) = 1 \} | \geq N. \]

NP-Completeness
Consider an instance of a 3-SAT problem:
With $l$ variables:
\[ x_1, x_2, \ldots, x_l \]
And $n$ clauses:
\[ c_1, c_2, \ldots, c_n \quad (n \geq l) \]

- Assume that no clause contains a variable and its negation: $x_j$ and $\overline{x}_j$ (The clause is a tautology $\equiv T$)

- Restriction site associated with a clause $c_i$:
  \[ f_i = \frac{i}{2(n+1)} \quad f_i^R = 1 - f_i = \frac{2n-i+2}{2(n+1)} \]

\[ D_1 \quad f_i \quad \text{if } x_j \in c_i \]
\[ \vdots \]
\[ D_j \quad \text{OR} \quad \text{if } \overline{x}_j \in c_i \]
\[ \vdots \]
\[ D_k \]
\[ D_{k+1} \quad f_1, f_2, \ldots, f_n \]
\[ D_m \quad \vdots \]
\[ D_m \quad \vdots \]
Create a dataset \( D_1, D_2, \ldots, D_{k+1}, \ldots, D_m \)

with \( m = 2k - 1 \) as follows:

\( D_j \) has a cut at \( f_i \) or \( f_i^R \), only:

\[ f_i \in D_j \quad \text{iff} \quad x_j \in C_i \]

\( (f_i^R \in D_j \quad \text{iff} \quad \overline{x}_j \in C_i) \)

\( N = n, \quad p_c = \frac{1}{2} \)

CNF has a satisfying assignment

\[ \Rightarrow \text{Choose an admissible alignment in which} \]

\[ D'_j = \begin{cases} D_j & \text{if } x_j = \text{true} \\ D_j^R & \text{if } x_j = \text{false} \end{cases} \quad 1 \leq j \leq k, \quad 1 \leq j \leq m. \]

\[ \Leftrightarrow \text{For every } f_i, \ (1 \leq i \leq n) \text{ there are } (k-1) \text{ matches from } D_{k+1}, \ldots, D_m \]

\[ \& \text{at least one more from } D'_1, \ldots, D'_k \]

(by since each clause must be satisfied)

\[ \forall 1 \leq i \leq n \quad \sum_j m_{ijk} \geq k > \frac{2(k-1)}{2} = p_c m. \]

\[ \Rightarrow \{ h \in [0,1] \mid x_k(h) = 1 \} = \{ f_1, f_2, \ldots, f_n \} \]

\[ \Rightarrow \{ h \in [0,1] \mid x_k(h) = 1 \} = \{ f_1, f_2, \ldots, f_n \} \]
Conversely, if the CNF has no satisfying assignment, then for every admissible alignment there exists an $1 \leq i \leq n$

$$\forall k \exists i: \sum_j m_{ijk} = (l-1) < p_{em} \quad \text{and} \quad \left| \sum_{h \in [0,1]} \lambda_k(h) = 1 \right| < n. \quad \Box$$

Problem Generation:
Statistical Model:

- A model or hypothesis $H$
  $$= \{ h_1, h_2, \ldots, h_N \} \quad N \approx 40$$
  Distribution for $h_i$'s
  Exponential gaps
  or uniform gaps.

- $Pr[D_j \mid H]$
  $$D_j \sim H. \quad \left\{ \begin{array}{l}
  \text{Pairwise Conditional Indep.} \\
  Pr[D_j \mid D_j, \ldots, D_{jm}, H] = Pr[D_j, H]
  \end{array} \right.$$ 

- $Pr[\text{bad}], \quad Pr[\text{good}] = 1 - Pr[\text{bad}]$

$$Pr[D_j \mid H] = \frac{1}{2} \sum Pr[D_j^{(c,k)} \mid H, \text{good}] \cdot Pr[\text{good}]$$
  $$+ \frac{1}{2} \sum Pr[D_j^{(c,k)} \mid H, \text{bad}] \cdot Pr[\text{bad}]$$

($N$) $\Rightarrow$ Alignment.
\[ D_j^{(k)} = D_j \text{ or } D_j^* \text{ with equal probability:} \]

\[ D_j \text{ Good } \Rightarrow \]

Choose parameters \( p_c, \sigma, \nu \).

\[ h_i \in H \Rightarrow s_i \sim N(h_i, \sigma) \text{ with } pr = p_c. \]

\[ s_i = \text{absent with } pr = 1 - p_c. \]

Spurious cuts \( \Rightarrow \) Stochastic Poisson.

\[ e^{\lambda_f} \frac{\lambda_f^{F_{jk}}}{F_{jk}!} \]

\[ D_j \text{ Bad } \Rightarrow \]

Poisson:

\[ e^{-\lambda_n} \frac{\lambda_n^{M_j}}{M_j!} \]

\[ \mathbb{P}_r \left[ D_j^{(k)} \mid H_i \text{ good} \right] = \prod_{i=1}^{2} \left[ \mathbb{P}_c \left( p_c \frac{e^{-(s_{ij} - h_i)^2/2\sigma^2}}{\sqrt{2\pi}\sigma_i} \right)^{m_{ijk}} (1-p_c) (1-m_{ijk}) \right] \]

\[ \times e^{-\lambda_f} \frac{\lambda_f^{F_{jk}}}{F_{jk}!} \]

\[ \mathbb{P}_r \left[ D_j^{(k)} \mid H_i \text{ bad} \right] = e^{-\lambda_n} \frac{\lambda_n^{M_j}}{M_j!} \]