

Notes on RNA Folding Recurrences

24th TBI Winterseminar in Bled
3rd Annual Meeting of the Bompfünowerer Consortium

Christian Höner zu Siederdisen

February 16, 2009

Recurrences (almost as) in ViennaRNA

$$F_{ij} = \min\{F_{i,j-1}, F_{i+1,j}, \min_{i < k \leq j} C_{ik} + F_{k+1,j}\}$$

$$C_{ij} = \min\{\mathcal{H}(i,j), \min_{i < k < l < j} C_{kl} + \mathcal{I}(i,j;k,l), \\ M_{i-1,j+1} + \alpha + \mathcal{M}(i,j)\}$$

$$M_{ij}^{1+} = \min\{M_{i,j-1}^{1+} + \beta, M_{i+1,j}^{1+} + \beta, \\ C_{ij} + \gamma, \min_{i < k < j} C_{ik} + M_{k+1,j}^{1+}\}$$

$$M_{ij} = \min\{M_{i,j-1} + \beta, M_{i+1,j} + \beta, \min_{i < k < j} C_{ik} + M_{k+1,j}^{1+}\}$$

Limitations and Runtime

- interior loops limited in size (typically 30)
- $O(n^2)$ space
- $O(n^3)$ time
- runtime can trivially be improved by minimizing only over $C_{ij} < \infty$
- let $P = \{c_{ij} \mid c_{ij} \in C, c_{ij} < \infty\}$

Faster Multiloop Calculations (Wexler)

Wexler et al. (2006) give the following ML recursion:

$$W_{ij} = \min\{C_{ij}, \min_{i < k < j} C_{ik} + W_{k+1,j} : C_{ik} < W_{ik}\}$$

- runtime: $O(n^2\phi(n))$
- $\phi(n) \approx 5.7$
- \Rightarrow multiloop calculations can be done in $O(n^2)$ time

Sparse Interior Loops with Linear Costs

- Kanehisi, Goad, 1982: $O(n^2)$ time
- Eppstein et al, 1989: $O(n + P \log \log \min(P, n^2/P))$
- where for each point in $P \ll n^2$ a lookup is performed
- the lookup data structure is a flat tree

Improved Sparse Interior Loops

- let $S \leq n$ a lookup structure
- let $s_k \in S, \text{col}(s_k) < j$ the lookup result for $p_{ij} \in P$
- for $p_{il}, l > j$:
 - s_k or
 - s_{k+1} is the result
 - depending on $\text{col}(s_{k+1}) < l$
- traverse S and $P_{i,k}$ in parallel for all k
- new runtime: $O(n + P + S^2)$

Concave Cost Functions

- reminder: $P \ll n^2$
- Eppstein et al, 1989: $O(n + P \log P \log \min(P, n^2/P))$
- Larmore, Schieber, 1991: $O(n^2)$
or $O(n + P \log \min(P, n^2/P))$

Interior Loops With Multiloop Optimization

- sparse solution interesting for $P \ll n^2$
- for each $p_{ij} \in P$: find minimal $s \in S$
- assume the following costs¹: $O(P \log S)$ where $S \leq n$
- remove s from S if $\text{energy}(s) > M_{ij} + \alpha$ with $\beta \leq 0$
- for $\beta > 0$ have S_{now} and S_{future}

¹or any optimization

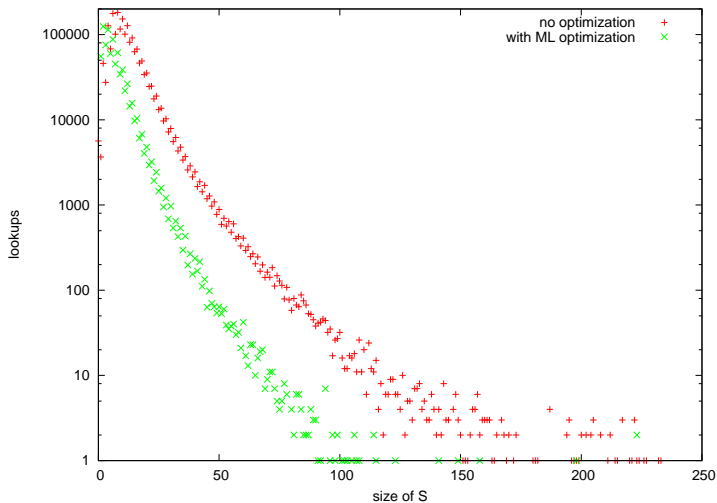
Measurements

Runtime for HIV1 (first 4000 nc)

- without optimization: 30.66 s
- with optimization: 23.64 s
- not using advanced Eppstein or Larmore optimizations

Size of S

(on a total of 1,912,283 C entries)



Comments

- folding without interior loop restrictions possible and fast
- Ninio correction for large interior loops?
- accuracy of the energy model for multiple loops and interior loops?
- how to verify? large interior loops require longer sequences than have been experimentally verified

Special Thanks to

- Ivo Hofacker
- Ilenia Boria
- everybody at the TBI