Average Case Optimal Pancake Sorting

Team CANEPKA

Coleman Link
Nik Olivieri
Team Introduction

● Coleman Link
  ○ 4th year - Computer Science BS/MS

● Nik Olivieri
  ○ Bioinformatics MS
Pancake Sorting - Background

- Given a list: [3, 1, 2, 4, 5]
- A “prefix reversal” of that list is taking some sequential portion of the list, that must start with the first element, and reversing its order
- E.g. [3, 1, 2, 4, 5] → [2, 1, 3, 4, 5]
  - Prefix reversal of length 3
Pancake Sorting - Background

- Any list can be sorted using prefix reversals
- E.g. \([3, 1, 2, 4, 5] \rightarrow [2, 1, 3, 4, 5] \rightarrow [1, 2, 3, 4, 5]\)
- This is referred to as “Pancake Sorting”
  - The way you would sort a stack of pancakes by flipping some number of them with a spatula
Pancake Sorting - Background
Example

\[
\begin{array}{ccccccc}
5 & 4 & 3 & 2 & 1 & 1 & 2 \\
3 & 2 & 1 & 2 & 3 & 2 & 3 \\
1 & 1 & 2 & 3 & 4 & 4 & 4 \\
2 & 3 & 4 & 5 & 5 & 5 & 5 \\
4 & 5 & 5 & 5 & 5 & 5 & 5 \\
\end{array}
\]
Pancake Sorting - Background

- Each specific sequence has some minimum number of ‘flips’ (prefix reversals) that it can be sorted in.
- Determining the optimal (minimum) number of flips for some specific list is NP-Hard [Bulteau, Fertin, Rusu - 2011]
Pancake Sorting - Background

- “Pancake Number”
  - For a given list size, what is the maximum number of prefix reversals that may be necessary to sort a list of unique elements of that size?
  - Find the minimum number of steps necessary to sort each permutation of a given size, then take the maximum of all of these.
Pancake Sorting - Background

● “Pancake Number”
  ○ Determining this value is known to be computationally difficult as well.
  ○ Value are known for list sizes up to and including 17:

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>...</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>...</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td></td>
</tr>
</tbody>
</table>
Our Investigation

- “Pancake Number” measures the worst case number of reversals for a given list size.

- What about the average case?
Our Investigation

- Compute an estimate for the average case number of prefix reversals needed to sort a given list size.
- Use parallel programming to improve this computation.
Sequential algorithm

Average(list_length, num_iterations):
  Total = 0
  Do num_iterations times:
    Compute a random list of length <list_length>
    Compute the minimum number of reversals necessary to sort that list
    Add this minimum number of reversals to the Total

  Return <Total> divided by the number of iterations
Parallel algorithm

- Use an algorithm similar to sequential one, with loop iterations split up between threads.
- Perform a reduction between the total number of reversals from each thread.
Parallel algorithm

Average_parallel(list_length, num_iterations):
  Partition num_iterations between multiple threads
  Within each thread:
    Local_total = 0
    Do (number of iterations partitioned to this thread) times:
      Compute a random list of length <list_length>
      Compute the minimum # of reversals necessary to sort that list
      Add this minimum number of reversals to this thread’s local_total
  Reduce thread local totals together (using sum) into a global total
Return global total divided by the total number of iterations
Applications

- Pancake graphs
  - A graph for a given sequence length where each vertex corresponds to a permutation of that length
  - Edges between two vertices if the sequences can be made into one another with a single prefix reversal
  - Used in distributed computing network design
Applications

- Prefix reversal distance
  - Number of prefix reversals to turn one arbitrary sequence into another
  - Generalization of prefix reversal sorting
  - Used in comparative genomics

  - Prefix reversals represent a type of mutation that can occur in a genome
Sources

