COMP 2140 - Data Structures

Shahin Kamali

Topic 5 - Sorting
University of Manitoba

Based on notes by S. Durocher.
Overview

- Review: Insertion Sort
- Merge Sort
- Quicksort
- Heapsort
- Counting Sort

For further reading, refer to Open Data Structures Book (Chapter 11)
Sorting

- **Input:**
  - a sequence of \( n \) objects: \( A[0], \ldots, A[n - 1] \)
    (typically an array or a linked list)
  - a comparison predicate, \( \leq \), that defines a total order on \( A \)

- **Output:**
  - an ordered representation of the objects in \( A \)

Many sorting algorithms exist:
- bubble sort, insertion sort, merge sort, heapsort, radix sort, bucket sort, quicksort, etc.
Insertion Sort

- Go through the items in the array (list) one by one
- For each item \( x \) at index \( i \):
  - We know the sub-array \( A[0] \ldots A[i-1] \) is sorted
  - Insert \( x \) in its correct position in the sub-array \( A[i] \ldots A[i] \).

```
4  3  2  10  12  1  5  6
```
/* Function to sort an array using insertion sort */
void insertionSort(int arr[], int n)
{
    int i, key, j;
    for (i = 1; i < n; i++)
    {
        key = arr[i];
        j = i - 1;

        /* Move elements of arr[0..i-1], that are greater than key, to one position ahead of their current position */
        while (j >= 0 && arr[j] > key)
        {
            arr[j+1] = arr[j];
            j = j - 1;
        }
        arr[j+1] = key;
    }
}
Insertion Sort Summary

- **One Iteration of the Insertion Sort Algorithm:**
  - After the $i$th iteration, $A[0..i]$ is sorted.
  - Insert item $A[i + 1]$ in its proper place in $A[0..i]$.
  - In the worst case, $i$ items are moved in the $i + 1$th iteration!

![Diagram of Insertion Sort Algorithm]

- **Diagram Description:**
  - The diagram illustrates the process of insertion sort with an array $A$.
  - After the $i$th iteration, elements $0$ to $i$ are sorted.
  - The item at index $i + 1$ is being inserted into its proper place.
  - The process shows how items are moved from the right to the left to maintain the sorted order.

COMP 2140 - Data Structures
Insertion Sort Analysis

- In the worst case the array is sorted backwards.

\[
\begin{array}{ccccccc}
  n & n-1 & n-2 & \ldots & 3 & 2 & 1 \\
\end{array}
\]

\[
\begin{array}{ccccccc}
  \ldots & 3 & 2 & 1 \\
\end{array}
\]

\[
\begin{array}{ccccccc}
  \ldots & 3 & 2 & 1 \\
\end{array}
\]

\[
\vdots
\]

\[
\begin{array}{ccccccc}
  3 & 4 & 5 & \ldots & n & 2 & 1 \\
\end{array}
\]

\[
\begin{array}{ccccccc}
  2 & 3 & 4 & \ldots & n-1 & n & 1 \\
\end{array}
\]

\[
\begin{array}{ccccccc}
  1 & 2 & 3 & \ldots & n-2 & n-1 & n \\
\end{array}
\]

- The total number of moved items:

\[
1 + 2 + \ldots + n - 1 = n(n - 1)/2 \in (n^2)
\]
Insertion Sort Time Complexity

- The **worst-case** running time of insertion sort is \( O(n^2) \).
- As it turns out, the **average-case** running time is also \( O(n^2) \).
- Faster sorting algorithms exist. These include:

<table>
<thead>
<tr>
<th></th>
<th>worst case</th>
<th>average case</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quicksort</td>
<td>( O(n^2) )</td>
<td>( O(n \log n) )</td>
</tr>
<tr>
<td>Merge Sort</td>
<td>( O(n \log n) )</td>
<td>( O(n \log n) )</td>
</tr>
<tr>
<td>Heapsort</td>
<td>( O(n \log n) )</td>
<td>( O(n \log n) )</td>
</tr>
</tbody>
</table>

The lower bound on the worst-case time complexity of any comparison-based sorting algorithm is \( \Omega(n \log n) \).
Merge Sort

- Merge sort is an example of a **divide-and-conquer** algorithm.
  - **Divide** the input into two or more disjoint subsets.
  - Recursively solve each sub-problem.
  - Combine solutions to the sub-problems to give the solution to the original problem.
Merge Sort Algorithm

- **Input:** an array $A[0..n-1]$ of comparable elements.
  - **Divide** $A$ into two subarrays $A[0..\lfloor n/2 \rfloor]$ and $A[\lfloor n/2 \rfloor + 1, n-1]$
  - **Recursively** sort each sub-array
  - **Combine** the two subarray via merging them

- The base of recursion is an array of size 1 which is sorted
  - In practice, when the length of sub-array is less than 100, selection sort is applied.
Merge Sort Scheme
Merge Sort Example

\[
\text{MS}(A[0..7]) = 13 \ -4 \ 7 \ 5 \ 6 \ 9 \ 2 \ 1
\]
Merging Sorted Sub-arrays

- Given two sorted arrays $A$ and $B$ of size $n$ and $m$, merge them into array $C$ of size $m + n$.
  - $i$, $j$, and $k$ are three indices moving on $A$, $B$, and $C$.
  - They are initially 0.
- At each step, copy the smaller of $A[i]$ and $B[j]$ to $C[k]$.
  - Increment $k$
  - If $A[i]$ is copied, increment $i$; otherwise increment $j$.
- If one array ends ($i = n$ or $j = m$), copy the remaining items of the other array to $C$. 

COMP 2140 - Data Structures
Merge Example

\[
\begin{align*}
\text{length } n & \quad A \quad i \\
\text{length } m & \quad B \quad j \\
\text{length } n + m & \quad C \quad k \\
\end{align*}
\]

<table>
<thead>
<tr>
<th>i</th>
<th>-5</th>
<th>-1</th>
<th>10</th>
<th>21</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>j</td>
<td>-3</td>
<td>-2</td>
<td>0</td>
<td>15</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>k</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>i</th>
<th>-5</th>
<th>-1</th>
<th>10</th>
<th>21</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>j</td>
<td>-3</td>
<td>-2</td>
<td>0</td>
<td>15</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>-5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>i</th>
<th>-5</th>
<th>-1</th>
<th>10</th>
<th>21</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>j</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>15</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>k</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
// Merge A[0..n] and B[0..m] into C[0..n+m-1]  
void mergeArrays(int arr1[], int arr2[], int n, int m, int arr3[])  
{
    int i = 0, j = 0, k = 0;  
    while (i < n && j < m)  
    {
        if (A[i] < B[j])  
        {
            C[k] = A[i]; i++; k++;  
        }  
        else  
        {
            C[k] = B[j]; j++; k++;  
        }  
    }  
    while (i < n)  
    {
        C[k] = A[i]; i++; k++;  
    }  
    while (j < m)  
    {
        C[k] = B[j]; j++; k++;  
    }  
}
Merge Sort Summary

- Recursively sort the left half of the input array $A$
- Recursively sort the left half of the input array $B$
- Merge the two sub-arrays into a new one
  - note that merging requires a new array, that is, it cannot be done in place.
What is the time complexity of merge-sort?

\[ T(n) = \begin{cases} 
1, & \text{if } n \leq 1 \\
2T(n/2) + O(n), & \text{if } n \geq 2
\end{cases} \]

We solve this using replacement method to get \( T(n) \in O(n \log n) \).
Analysis of Merge Sort

\[ T(n) = \begin{cases} 
1, & n \leq 1 \\
2T(n/2) + O(n) & n \geq 2
\end{cases} \]

Since merging two subarrays of size \( n/2 \) takes \( O(n) \), the merge time is at most \( cn \) for some constant value of \( c \) (think of \( c \) as \( M \) in the definition of big-Oh). So we have:

\[
T(n) \leq 2T(n/2) + c \cdot n \\
\leq 2(2T(n/4) + c \cdot n/2) + c \cdot n = 4T(n/4) + 2cn \\
\leq 4(2T(n/8) + c \cdot n/4) + 2c \cdot n = 8T(n/8) + 3cn \\
\leq \ldots \\
\leq 2^k T\left(\frac{n}{2^k}\right) + kcn \\
\ldots \\
\leq 2^{\log n} T\left(\frac{n}{2^{\log n}}\right) + \log ncn = nT(1) + cn \log n \in O(n \log n).
Should we use merge sort?

- Unlike insertion sort, merge sort does not work in place.
- Merging two arrays requires temporary storage.
Quick Sort

- Like insertion sort, quicksort works in place.
- Like merge sort, quicksort employs a divide and conquer strategy.
- Quicksort is usually implemented as a randomized algorithm.
Quick Sort

- Select an arbitrary element in the array as a **pivot**.
- Partition the array such that elements less than or equal to the pivot appear to its left and elements greater than or equal to the pivot appear to its right.
- Recursively sort each partition.
- In the base case we have an array of size 1.
Quick Sort

- Any element in the array can be selected as the pivot.
- Typically, the pivot is selected randomly.
- Elements are partitioned into those less than or equal to the pivot and those greater than the pivot.
- In general, elements within a partition are not initially sorted.

**Partitioning can be performed in place.**
**Partition Algorithm**

- First, swap pivot with the first element.
- Store elements \( \leq \) the pivot at the front of the array and elements \( \geq \) the pivot at the back of the array.
  1. Scan the array starting from the front until an element is found that is \( > \) the pivot.
  2. Scan the array starting from the back until an element is found that is \( < \) the pivot.
  3. Swap these two items.
  4. Continue until the entire array has been partitioned.
/ Quicksort Partition Java Code
public static int partition(int[] A, int lo, int hi, int pivot) {
    int finalPivot;
    int left = lo + 1;
    int right = hi;
    swap(A, lo, pivot);
    while (left < right) {
        while (left < right && A[lo] >= A[left])
            left++;
        while (left < right && A[lo] <= A[right])
            right--;
        swap(A, left, right);
    }
    if (A[right] <= A[lo])
        finalPivot = right;
    else
        finalPivot = right - 1;
    swap(A, lo, finalPivot);
}
Quicksort Example

- Sort the following array using quicksort.

\[
\begin{array}{cccccccc}
6 & -3 & 5 & 1 & 2 & -4 & 3 & 7 \\
\end{array}
\]

- **Step 1.** Select a pivot element and swap it with the first element.

\[
\begin{array}{cccccccc}
6 & -3 & 5 & 1 & 2 & -4 & 3 & 7 \\
\end{array}
\]

\[
\begin{array}{cccccccc}
2 & -3 & 5 & 1 & 6 & -4 & 3 & 7 \\
\end{array}
\]

- Note that you could select any element as the pivot.
Quicksort Example

- We now partition the array such that elements to the left of the pivot are $\leq$ than the pivot and element to the right of the pivot are $\geq$ the pivot.
  - Initialize \texttt{left} to the leftmost element after the pivot and \texttt{right} to the rightmost element.

\begin{itemize}
  \item \textbf{step 2:}
  \begin{itemize}
    \item Increment \texttt{left} until we find an element that is greater than the pivot.
  \end{itemize}
\end{itemize}
Quicksort Example

- **Step 3:** Decrement **right** until we find an element that is less than the pivot.

```
2  -3  5  1  6  -4  3  7
```

```
left  right
```

- **Step 4:** Swap the elements at positions **left** and **right**.

```
2  -3  -4  1  6  5  3  7
```

```
left  right
```
Quicksort Example

- **Step 5:** Repeat until $left \geq right$.

![Diagram showing a list of numbers and an arrow indicating that left equals right.]

- **Step 6:** If the element at position $right$ is less than the pivot, then swap it with the pivot. Otherwise, swap the element at position $right - 1$ with the pivot.

![Diagram showing a list of numbers with a swap indicated between the first and second elements.]
Quicksort Example

- The pivot element is now in the correct position in the array.
- Elements to the left of the pivot are $\leq$ than it and elements to the right of the pivot are $\geq$ than it.

Once the left subarray is recursively sorted and the right subarray is recursively sorted, the entire array is sorted.
- The base case is reached when the array has size $\leq 1$. 