CAB301 Assignment 2: Empirical Comparison of Two Algorithms for Finding the Minimum Distance between Two Elements in an Array

Shane Havers N9477756

Andrew Grant N9167340

Date Submitted:

# Summary

The purpose of this report is to summarise and compare the results of experiments conducted on two algorithms that both find the smallest difference between two elements within a set of numbers. The tests were completed by implementing the algorithms and a test driver in C++ within the Code Blocks IDE, and outputting the results to Comma Separated Value files (.CSV). These final outputs were plotted in MATLAB, and the trends clearly reinforce the expectations that the second algorithm (MinDistance2) was indeed far more efficient than the first algorithm (MinDistance). As such, the end test results show a very high correlation with the estimated theoretical efficiency for both algorithms.

# Description of the Algorithms

The algorithms being compared in this report are algorithms that check the difference – or ‘distance’ – between each element of an input list of numbers, and seek to return the smallest difference. The first algorithm, MinDistance, was proposed by Levitin [1] and can be found in Appendix 1. MinDistance2, which can be found in Appendix 2, serves the same purpose as MinDistance but is proposed as a potentially more efficient version.

The MinDistance algorithm works by first setting the variable to an arbitrarily large value, ideally infinity. Then, it iterates over each element in the array with an index variable *,* and then again with an index variable *.* The algorithm then asks: first ensure the element is not being compared to itself (check ); and if not, is the difference between the two elements smaller than the current minimum (check ). If this difference is the smallest, set to this value. After checking every permutation of and , the algorithm returns the minimum value and exits.

The MinDistance2 starts the same way, by setting to a large value. Then, it iterates over every element of bar the final one (again with index ), and then loops over every value with an index larger than the index (i.e. the second loop goes from index to the end of the array). Then, it stores the difference between the values at indices and in a temporary variable, and if that variable is smaller than the current minimum, change to that value. After both loops have exited, the algorithm returns the minimum value and exits.

# Theoretical Analysis of the Algorithms

## Identifying the Algorithm’s Basic Operation

### MinDistance

### MinDistance2

## Choice of Problem Size

## Analysis of Theoretical Average Case Efficiency

# Methodology, Tools and Techniques

## Programming Environment

The algorithm and testing suite used to run them were implemented in C++ using the Code Blocks IDE. C++ is a well-established, fast and flexible programming language [2], making it an appropriate choice for comparing the algorithms. The code was then compiled and run on a Microsoft Surface Pro 4 running Microsoft’s flagship operating system, Windows 10.

## Implementation of Algorithms

Both algorithms were implemented as C++ methods that were called within the Main function. The integer primitive type was used to represent the list of numbers, as whole numbers were adequate to demonstrate both algorithms’ effectiveness while also minimising the amount of memory used to store the array. The setup of the C++ program can be found in Appendix 3, and the Main method can be found in Appendix 4.

The translation of both algorithms from pseudocode to C++ was straightforward, with the only significant difference being that the initial value of was set to the maximum allowable integer value of rather than the impossible value of infinity – the risk of the distance between two values being larger than is completely negligible, therefore is a suitable substitute for infinity in this case. The MinDistance and MinDistance2 methods can be found in Appendix 5 and Appendix 6 respectively.

## Generating Test Data and Running Experiments

To properly compare the algorithms, the test array was populated with pseudo-random numbers before being run through both algorithms. The pseudo-random numbers were generated using C++’s *‘random’* library by using a Uniform Integer Distribution, seeded by a Mersenne Twister 19937 generator, between after being adapted from cppreference.com’s implementation [3].

The size of the test array varied from 25 to 10,000, and increased by 25 at each step. The operations counting was only carried out once per array size, because the number of operations carried out by both algorithms is constant for a given input size. There is however variation in the execution time, so for each input size 50 different arrays were tested on both algorithms to obtain an average.

Finally, the test results were written to .CSV files. This was so that the results could easily be output to and read by MATLAB, and accurate plots could be obtained. Allain’s use of the *‘fstream’* and *‘iostream’* libraries were consulted to achieve this [4].

## Implementation of Basic Operations Counters

The basic operations were counted once within the main loop for each distinct array size. To achieve this, the algorithms had to be modified to include a counter that is incremented each time the basic operation was executed. The counter for MinDistance and MinDistance2 were the global long integers and respectively.

The implementation of the modified algorithms was the separate methods MinDistance\_OpsCount and MinDistance2\_OpsCount, which are included in Appendix 7 and Appendix 8 respectively. In MinDistance\_OpsCount the counter is incremented immediately before the first array element comparison, which was separated from the initial condition to properly implement the counter, and immediately after the second comparison. In MinDistance2\_OpsCount the counter is only incremented after the assignment of the variable, which performs the same array element comparison the first algorithm.

After the basic operations had been counted, the program moved on to the execution timing. Once this was complete, the results of this part of the experiment were written to the Ops CSV file.

## Implementation of Execution Timers

The execution time of each algorithm was measured within the main loop, and was measured 50 times for each distinct array size to find a reasonable average. Both algorithms were timed by using the *“chrono”* library by adapting Smistad’s method [5]. For the first algorithm, the timer was started, then the MinDistance method was run, and then the stop time was added to the global long integer – the same process was then performed for the second algorithm, except using the MinDistance2 method and the variable.

The process of timing the algorithm’s execution time was performed 50 times on each algorithm. UP TO HERE.

# Experimental Results

## Functional Testing

## Number of Basic Operations

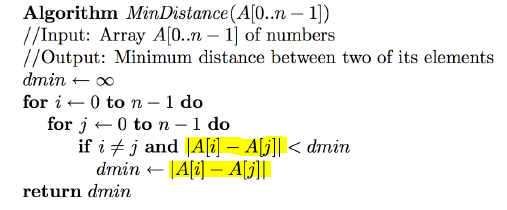
## Execution Time

# References

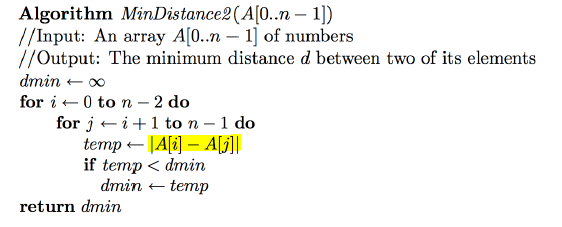
|  |  |
| --- | --- |
| [1] | A. Levitin, Introduction to the design and analysis of algorithms, 2nd ed., Pearson Addison-Wesley, 2007. |
| [2] | "Albatross", “C++: A Brief Description,” cplusplus.com, 2017. [Online]. Available: http://www.cplusplus.com/info/description/. [Accessed 17 May 2017]. |
| [3] | cppreference.com, “cppreference.com,” 2017. [Online]. Available: httpp://en.cppreference.com/w/cpp/numeric/random/uniform\_int\_distribution. [Accessed May 2017]. |
| [4] | A. Allain, “C++ File I/O,” Cprogramming.com, 2011. [Online]. Available: http://www.cprogramming.com/tutorial/lesson10.html. [Accessed May 2017]. |
| [5] | E. Smistad, “Measuring runtime in milliseconds using the C++ 11 chrono library,” Erik Smistad, 2012. [Online]. Available: https://www.eriksmistad.no/measuring-runtime-in-milliseconds-using-the-c-11-chrono-library/ . [Accessed May 2017]. |

# Appendix

1. MinDistance Algorithm



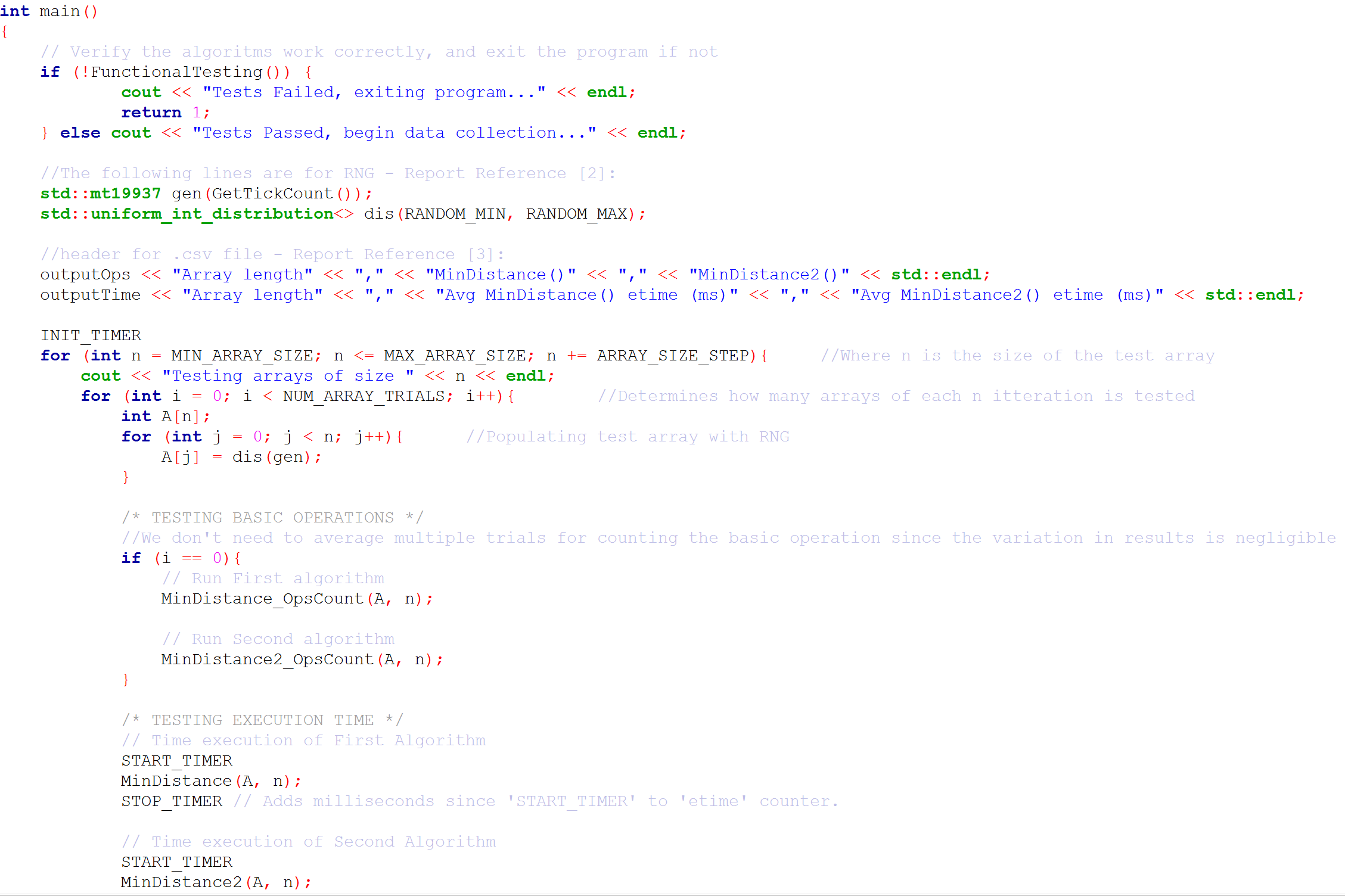
1. MinDistance2 Algorithm

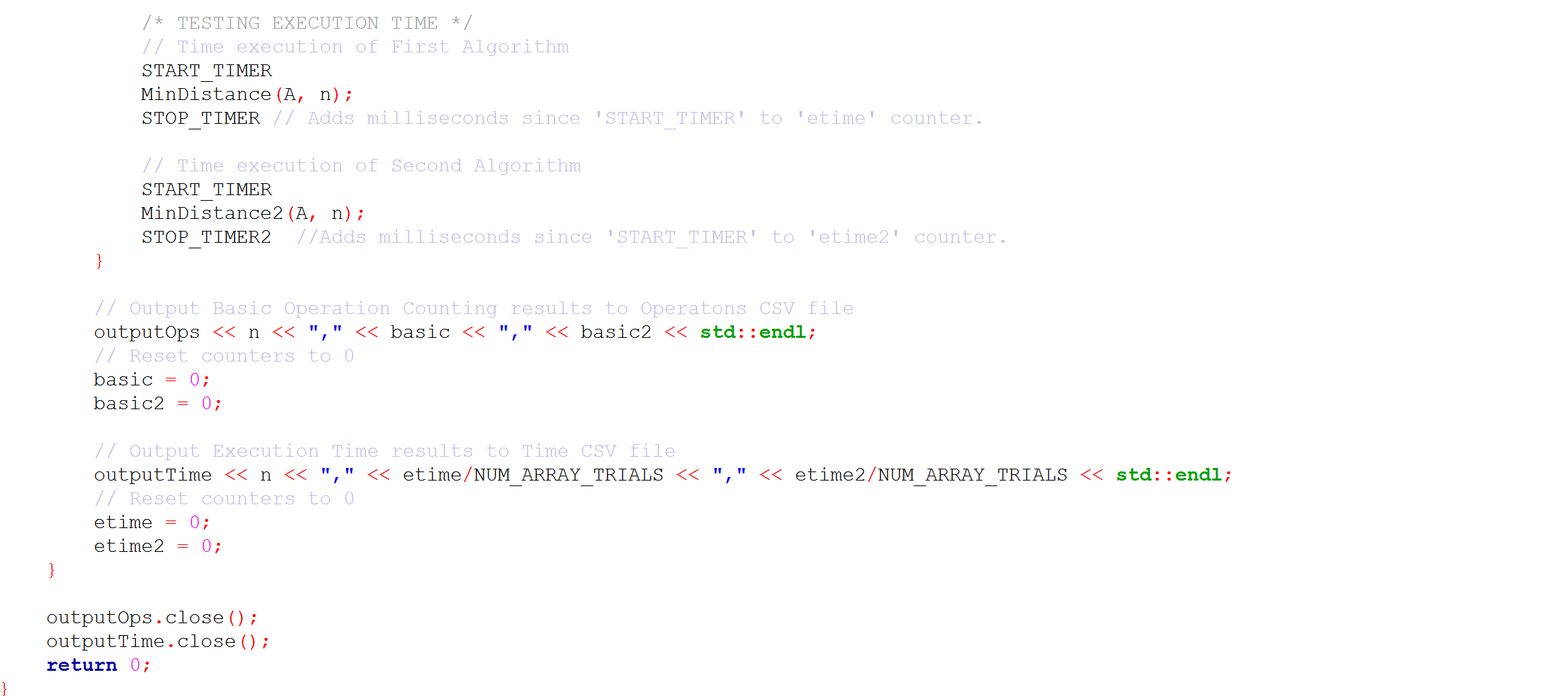


1. Setup of C++ Program

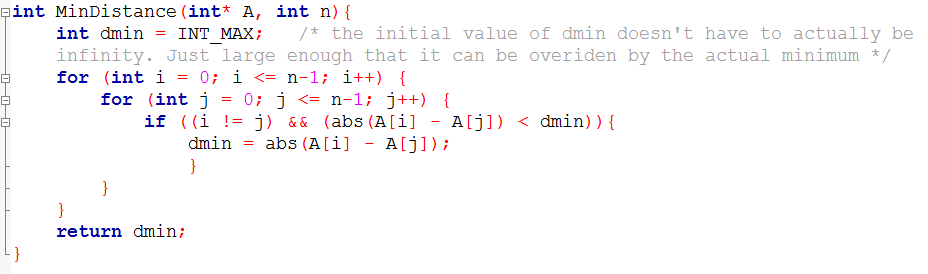


1. Main Method

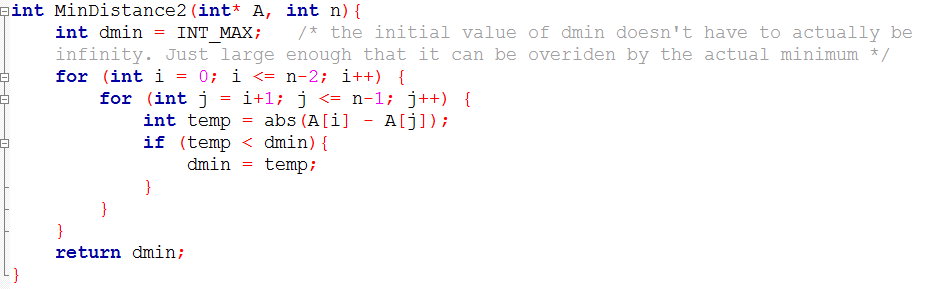




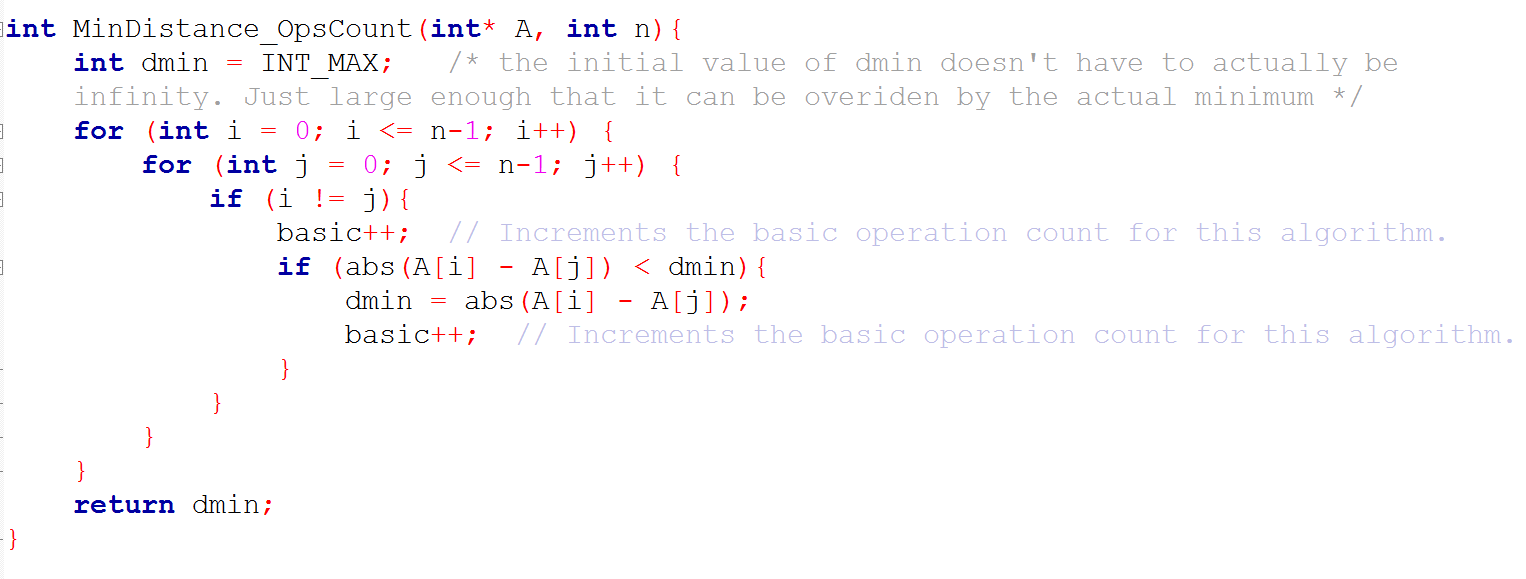
1. MinDistance C++ Implementation



1. MinDistance2 C++ Implementation



1. MinDistance\_OpsCount C++ Implementation



1. MinDistance2\_OpsCount C++ Implementation

