Etan Soclof

"Median Finding" Using A Divide-And-Conquer Algorithm

A simple binary search will not be effective in determining the median of both arrays for many reasons. For one, a binary search requires that one know at the outset what is being searched for and in this problem we have not identified the median at the outset.
Instead, we will find the median of each array and note that the median of the combined arrays cannot be greater than the greater of the two medians and or less than the lesser of the two medians. Therefore, we can compose two subarrays by eliminating those values that meet the above criteria and recursively find the medians of those subarrays.

Algorithm:
Find median of array1, median1, and median of array2, median2.
While size of arrays > 2
  If median1 > median2
    If size of arrays is even
      eliminate numbers less than median2 from array2 and all numbers greater than median1 from array1 (except for the number in the index above median1)
    If size of arrays is odd
      eliminate numbers less than median2 from array2 and all numbers greater than median1 from array1
  If median2 > median1, eliminate ints less than median1 from array1 and ints greater than median2 from array2
    If size of arrays is even
      eliminate numbers less than median1 from array1 (except for the number in the index above median1) and all numbers greater than median2 from array2
    If size of arrays is odd
      eliminate numbers less than median1 from array1 and all numbers greater than median2 from array2
  Recursively run algorithm on new arrays
If size of arrays == 2
  Merge sort to a single array and return element at index 1.

Recurrence Relation:

$$Q(n) = \begin{cases} 4 & \text{if } n \text{ is 2} \\ Q(n/2) + O(1) & \text{otherwise} \end{cases}$$

By the Master theorem: a = 1, b = 2, d = 0. Since a = b^d, Q(n) = O(logn).

Proof:
Consider two arrays, array1 and array2, each of size n, in which the first index is referred to as index 1. When the arrays are combined into a single array, array3, of size 2n (with the first index referred to as index 1), the median of array3, median3, will be at index n.
Find the medians of each individual array - median1 and median2. After finding these two medians, we can split up the remaining numbers into four distinct sets:

A. Set A: Numbers larger than the larger of the two medians in the array in which the larger of the two medians is found are necessarily larger than both medians. These can be found from index $\lceil n/2 \rceil$ + 1 to index n of the array in which the larger of the two medians is found. There are a total of such $\lceil n/2 \rceil$ - 1 numbers in an array with an odd number of numbers and a total of n/2 such numbers in an array containing an even number of numbers.

B. Set B: Numbers lesser than the lesser of the two medians, in the array in which the lesser of the two medians is found are necessarily lesser than both medians. These can be found from index1 to index n of the array in which the larger of the two medians is found. There are a total of such $\lceil n/2 \rceil$ - 1 numbers in an array with an odd number of numbers and a total of n/2 - 1 such numbers in an array containing an even number of numbers.
C. Set C: Numbers larger than the lesser of the two medians in the array in which the lesser of the two medians is found might be lesser or greater than the larger of the two medians. There are a total of such $\lceil n/2 \rceil$ - 1 numbers in an array with an odd number of numbers and a total of n/2 - 1 such numbers in an array containing an even number of numbers.
D. Set D: Numbers lesser than the larger of the two medians in the array in which the larger of the two medians is found might be lesser or greater than the lesser of the two medians.There are a total of such $\lceil n/2 \rceil$ - 1 numbers in an array with an odd number of numbers and a total of n/2 such numbers in an array containing an even number of numbers.

When considering the total amount of numbers that are greater than the greater of the two medians, we must consider sets A and C. The cardinality of the union of sets A and C is 2($\lceil n/2 \rceil$ - 1) = n - 1 when dealing with an array with an odd number of elements and 2(n/2) = n when dealing with an array with an even number of elements. In the worst case, when all members of C are larger that the greater of the two medians, the members of A and C will necessarily comprise the last n - 1 or n indexes of array3, i.e. from index n + 1 or n + 2 to index 2n. Because we know that median3 is at index n of median3, we can conclude that numbers that are greater than the greater of the two medians cannot be median3. Because we know that all members of A are greater than the greater of the two medians, all members of A are greater than medians3.
When considering the total amount of numbers that are less than the lesser of the two medians, we must consider sets B and D. The cardinality of the union of sets B and D is 2($\lceil n/2 \rceil$ - 1) = n - 1. In the worst case, when all members of D are less that the lesser of the two medians, these numbers will necessarily comprise the last n - 1 indexes of array3, i.e. from index 1 to index n - 1. Because we know that median3 is at index n of median3, we can conclude that numbers that are lesser than the lesser of the two medians cannot be median3. Because we know that members of B are less than the lesser of the two medians, all members of B are less than medians3.

Lemma 1: Given an array with n unique values if n is even, the (n/2)th element is the median. In such a case. If an equal amount of elements greater than and less than the (n/2)th element are eliminated, the median will remain the same.

For all x in set A, x > median3. For all x in set B, x < median3. Therefore, we can eliminate all x in A U B. By lemma 1, if $|A| = |B|$, we can eliminate sets A and B from consideration and the median of the remaining numbers will remain the same. When both arrays are contain an odd number of elements, $|A| = |B|$. When both contain and even number of elements, $|A| > |B|$. To compensate for the latter case, only eliminate n/2 - 1 members of set A from consideration. Median3 can be found in the remaining subarrays.

Array1

C - numbers less than median1 in array1 (amounting to $\lceil n/2 \rceil$ - 1 ints)	median1	A - numbers greater than median1 in array1 (amounting to $\lfloor n/2 \rfloor$ ints)

Array2

B - numbers less than median2 in array2 (amounting to $\lceil n/2 \rceil$ - 1 ints)	median2	D - numbers greater than median2 in array2 (amounting to $\lfloor n/2 \rfloor$ ints)

When median1 > median2: Array3

B Some or all of D (amounting to at most 2($\lceil n/2 \rceil$ - 1) = n - 1 ints)	median2	Some or all of C and D	median1	A Some or all of C (amounting to at most n ints)