

# Majority Algorithm Solution

Brian Rothstein

October 22, 1999

## 1 Problem

Given an array,  $A$ , consisting of  $N$  ( $N > 0$ ) positive integers, write a function which finds the integer that occurs a majority of the time in the array [if it exists] and return it. If it doesn't exist, return 0. The function must run in  $O(1)$  space and  $O(N)$  time.

## 2 Setup

It will be helpful to define the following terminology:

$$\delta(a, b) = \begin{cases} 1 & a = b \\ 0 & a \neq b \end{cases}$$

$$\#A_{i,j}(b) = \sum_{k=i}^j \delta(A[k], b)$$

$$M(A, i, j) = \begin{cases} 0 & \nexists b : \#A_{i,j}(b) > \frac{j-i+1}{2} \\ b & \exists b : \#A_{i,j}(b) > \frac{j-i+1}{2} \end{cases}$$

$$P(A, i, j) = b : \{\forall c : \#A_{i,j}(b) \geq \#A_{i,j}(c)\}$$

$\#A_{i,j}(b)$  is just the number of times that  $b$  occurs in  $A[i..j]$ .

$M(A, i, j)$  finds the majority element of  $A[i..j]$  if it exists. Otherwise it returns zero.

$P(A, i, j)$  finds the element that occurs at least as many times as any other element in  $A[i..j]$ .

### 3 Solution

#### 3.1 Simplification

For simplicity, we'll start off by considering

$$M'(A, i, j) = \begin{cases} \text{undefined} & \text{if } M(A, i, j) = 0 \\ b & \text{if } M(A, i, j) = b, b \neq 0 \end{cases}$$

Notice that  $M'(A, i, j)$  only needs to return a meaningful result if a majority exists in  $A[i..j]$ . For this reason, we can assume that there is a majority element, since if there's not, we can return anything.

So for now, assume  $M(A, 1, N) = m, m \neq 0$ . (Call this assumption  $[M*]$ )

#### 3.2 Two important theorems.

**Theorem 1.**  $[M*] \wedge \{M(A, 1, i) = 0\} \Rightarrow M(A, i + 1, N) = m$ .

What this says is that if there is a majority element in  $A[1..N]$  (as we just assumed) and there is no majority element in  $A[1..i]$  then there must be a majority element in  $A[i+1..N]$ . Furthermore, the majority element in  $A[i+1..N]$  must be the same as the majority element in  $A[1..N]$ .

**Proof.**

$$\text{Let } c = P(A, 1, i). \tag{1}$$

$$\Rightarrow \#A_{1,i}(m) \leq \#A_{1,i}(c). \tag{2}$$

$$M(A, 1, i) = 0 \Rightarrow \#A_{1,i}(c) \leq \frac{i}{2}. \tag{3}$$

$$\text{Now assume that } M(A, i + 1, N) \neq m. \tag{4}$$

$$\Rightarrow \#A_{i+1,N}(m) \leq \frac{N - i}{2}. \tag{5}$$

$$(2) \Rightarrow \#A_{1,i}(m) + \#A_{i+1,N}(m) \leq \#A_{1,i}(c) + \#A_{i+1,N}(m) \tag{6}$$

$$(3) \wedge (5) \Rightarrow \#A_{1,i}(c) + \#A_{i+1,N}(m) \leq \frac{i}{2} + \frac{N - i}{2} = \frac{N}{2} \tag{7}$$

$$(6) \wedge (7) \Rightarrow \#A_{1,i}(m) + \#A_{i+1,N}(m) \leq \frac{N}{2} \tag{8}$$

$$\Rightarrow \#A_{1,N}(m) \leq \frac{N}{2} \text{ which contradicts assumption } [M*].$$

So it must be that  $M(A, i + 1, N) = m$ .

**Theorem 2.**  $\{M(A, 1, i) = 0\} \wedge \{M(A, i + 1, j) = 0\} \Rightarrow M(A, 1, j) = 0.$

This says that if there is no majority element in  $A[1..i]$  and there is no majority element in  $A[i+1..j]$  then there is no majority element in  $A[1..j]$ .

**Proof.**

$$\text{Let } c = P(A, 1, i), \quad d = P(A, i + 1, j) \quad (9)$$

$$M(A, 1, i) = 0 \Rightarrow \#A_{1,i}(c) \leq \frac{i}{2} \quad (10)$$

$$M(A, i + 1, j) = 0 \Rightarrow \#A_{i+1,j}(d) \leq \frac{j-i}{2} \quad (11)$$

$$\text{Now suppose } M(A, 1, j) = e, \quad e \neq 0 \quad (12)$$

$$\Rightarrow \#A_{1,j}(e) > \frac{j}{2} \quad (13)$$

$$(9) \Rightarrow \#A_{1,i}(e) + \#A_{i+1,j}(e) \leq \#A_{1,i}(c) + \#A_{i+1,j}(d) \quad (14)$$

$$(10) \wedge (11) \Rightarrow \#A_{1,i}(c) + \#A_{i+1,j}(d) \leq \frac{i}{2} + \frac{j-i}{2} = \frac{j}{2} \quad (15)$$

$$\Rightarrow \#A_{1,j}(e) \leq \frac{j}{2} \text{ which contradicts (12), (13).}$$

So it must be that  $M(A, 1, j) = 0.$

### 3.3 First attempt at a loop invariant

Theorems 1 and 2 suggest the following loop invariant for  $M'(A, 1, N)$ :

$$L(a, i, j, b) := \{M(A, 1, i - 1) = 0\} \wedge \{M(A, i, j) = b\}$$

With this invariant, we could have  $j$  iterate through  $A$  and set  $i$  to  $j$  every time that we find that there is no majority in  $A[i..j]$ . Theorem 2 would then allow us to merge  $\{M(A, 1, i - 1) = 0\} \wedge \{M(A, i, j) = 0\}$  into  $M(A, 1, j) = 0.$  Theorem 1 would then allow us to conclude that  $b$  is the majority winner (under  $[M*]$ ) when  $j = N.$

This loop invariant is close to the right one but it's not strong enough to allow us to find or prove an algorithm. This is because it doesn't allow us to figure out when  $M(A, i, j) = 0.$

### 3.4 Second attempt at a loop invariant

A better loop invariant is:

$$L(A, i, j, k, b) := \{M(A, 1, i - 1) = 0\} \wedge \{k > j\} \wedge \{\#A_{i,j}(b) = \frac{k - i + 1}{2}\}$$

Notice that  $L \Rightarrow M(A, i, j) = b$  since

$$\{\#A_{i,j}(b) = \frac{k - i + 1}{2}\} \wedge \{k > j\} \Rightarrow \#A_{i,j}(b) > \frac{j - i + 1}{2}$$

More importantly, we can determine from this invariant when  $M(A, i, j) = 0$ .

**Theorem 3.**  $L \wedge \{j + 1 = k\} \wedge \{A[j + 1] \neq b\} \Rightarrow M(A, i, j + 1) = 0$ .

**Proof.**

$$L \wedge \{A[j + 1] \neq b\} \Rightarrow \#A_{i,j}(b) = \#A_{i,j+1}(b) = \frac{k - i + 1}{2} \quad (16)$$

$$(16) \wedge \{j + 1 = k\} \Rightarrow \#A_{i,j+1}(b) = \frac{(j + 1) - i + 1}{2} \quad (17)$$

$$\Rightarrow M(A, i, j + 1) \neq b. \quad (18)$$

$$\text{Now suppose } M(A, i, j + 1) = c, c \neq 0, c \neq b. \quad (19)$$

$$\Rightarrow \#A_{i,j+1}(c) > \frac{(j + 1) - i + 1}{2} \quad (20)$$

$$\text{But } M(A, i, j) = b \Rightarrow \#A_{i,j}(c) < \frac{j - i + 1}{2} \quad (21)$$

$$\Rightarrow \#A_{i,j+1}(c) < \frac{j - i + 1}{2} + 1 = \frac{(j + 1) - i + 1 + 1}{2} \quad (22)$$

$$\Rightarrow 2 \cdot \#A_{i,j+1}(c) < (j + 1) - i + 1 + 1 \quad (23)$$

$$\Rightarrow 2 \cdot \#A_{i,j+1}(c) \leq (j + 1) - i + 1 \quad (24)$$

$$\Rightarrow \#A_{i,j+1}(c) \leq \frac{(j + 1) - i + 1}{2} \text{ which contradicts (20).}$$

So  $M(A, i, j + 1) = 0$ .

### 3.5 Algorithm for $M'(A, 1, N)$

```
M'(A, 1, N)
{
    i=j=1;
    k=2;
    b=A[1];
                                                    // (p1)

    while(j<N)
    {
                                                    // (p2)
        j++;
        if(A[j]==b) k+=2;
        else if(j==k) { i=j; b=A[j]; k=k+1; }
    }
                                                    // (p3)

    return b;
                                                    // (p4)
}
```

To prove that the algorithm is correct, we need to show 4 things.

- **Initialization:** The loop invariant,  $L$ , is true at (p1)
- **Maintenance:** If  $L$  is true at (p2), then it will be true at (p3).
- **Boundedness:** The loop will terminate in a finite number of steps.
- **Correctness:** After the loop (p4),  $L \wedge \{j \geq N\} \Rightarrow M'(A, 1, N) = b$ .

### 3.6 Proof of Algorithm $M'(A, 1, N)$

**Initialization.**

$$\begin{aligned} & \text{At (p1) we have } L(A, i, j, k, b) \\ & = L(A, 1, 1, 2, A[1]) \\ & = \{M(A, 1, 0) = 0\} \wedge \{2 > 1\} \wedge \{\#A_{1,1}(A[1]) = \frac{2-1+1}{2}\} \end{aligned}$$

which is clearly true.

**Maintenance.** Assume that  $L(A, i, j, k, b)$  holds at (p2). We need to show 3 things:

$$\{L(A, i, j, k, b)\} \wedge \{A[j+1] = b\} \Rightarrow L(A, i, j+1, k+2, b) \quad (25)$$

$$\begin{aligned} &\{L(A, i, j, k, b)\} \wedge \{A[j+1] \neq b\} \wedge \{j+1 = k\} \\ &\quad \Rightarrow L(A, j+1, j+1, k+1, A[j+1]) \end{aligned} \quad (26)$$

$$\begin{aligned} &\{L(A, i, j, k, b)\} \wedge \{A[j+1] \neq b\} \wedge \{j+1 \neq k\} \\ &\quad \Rightarrow L(A, i, j+1, k, b) \end{aligned} \quad (27)$$

These correspond to the three cases inside the loop.

**Proof of (25).**

$$\begin{aligned} &L(A, i, j, k, b) \wedge \{A[j+1] = b\} \Rightarrow \\ &\{M(A, 1, i-1) = 0\} \wedge \{k+2 > j+1\} \wedge \{\#A_{i,j+1}(b) = \frac{k-i+1}{2} + 1\} \\ &= L(A, i, j+1, k+2, b). \end{aligned}$$

**Proof of (26).** By theorem 3,

$$L(A, i, j, k, b) \wedge \{j+1 = k\} \wedge \{A[j+1] \neq b\} \Rightarrow M(A, i, j+1) = 0.$$

By theorem 2,

$$\{M(A, 1, i-1) = 0\} \wedge \{M(A, i, j+1) = 0\} \Rightarrow M(A, 1, j+1) = 0$$

Therefore,

$$\begin{aligned} &L(A, i, j, k, b) \wedge \{j+1 = k\} \wedge \{A[j+1] \neq b\} \Rightarrow \\ &\{M(A, 1, j+1) = 0\} \wedge \{k+1 > j+1\} \wedge \\ &\#A_{j+1,j+1}(A[j+1]) = \frac{k+1 - (j+1) + 1}{2} = \frac{k+1 - (k) + 1}{2} = 1 \\ &= L(A, j+1, j+1, k+1, A[j+1]). \end{aligned}$$

**Proof of (27).**

$$\begin{aligned} &L(A, i, j, k, b) \wedge \{A[j+1] \neq b\} \wedge \{j+1 \neq k\} \Rightarrow \\ &\{M(A, 1, i-1) = 0\} \wedge \{k > j+1\} \wedge \{\#A_{i,j+1}(b) = \frac{j-i+1}{2}\} \\ &= L(A, i, j+1, k, b). \end{aligned}$$

Thus the invariant is maintained through the loop.

**Boundedness.** The loop obviously terminates in  $N - 1$  steps. Since  $N$  is finite, the loop will terminate in a finite number of steps. More importantly, the number of steps is linear with respect to  $N$ .

**Correctness.** At (p4) we have,

$$\begin{aligned}j = N &\Rightarrow L(A, i, N, k, b) \\&= \{M(A, 1, i - 1) = 0\} \wedge \{k > N\} \wedge \{\#A_{i,N}(b) = \frac{k - i + 1}{2}\} \\&\Rightarrow \#A_{i,N}(b) > \frac{N - i + 1}{2} \\&\Rightarrow M(A, i, N) = b\end{aligned}$$

Therefore since  $\{M(A, 1, i - 1) = 0\} \wedge \{M(A, i, N) = b\}$ , we know by theorem 1 that  $M(A, 1, N) = b$ , which is what we wanted to show. So  $b$  is the majority element under  $[M^*]$  (i.e. assuming that there exists a majority element.) But now the implementation of  $M(A, 1, N)$  is obvious!

### 3.7 Algorithm for $M(A, 1, N)$

```
M(A, 1, N)
{
    b = M'(A, 1, N);
    count = 0;
    i = 1;
    while(i <= N)
    {
        if(A[i] == b) count++;
        i++;
    }

    if(count > N/2) return b;
    else return 0;
}
```

The proof is left as an exercise to the reader.

## 4 Notes

- In  $M'(A, 1, N)$ , the variable,  $i$ , is not used. Its only real use is to prove the loop invariant. Since  $b$  achieves the same value with or without  $i$ , you can get rid of it (or put it in comments so that the proof is still clear.)
- $M'(A, 1, N)$  can be optimized to return  $b$  if  $k > N$  since at that point we have

$$\#A_{i,j}(b) = \frac{k-i+1}{2} > \frac{N-i+1}{2} \Rightarrow M(i, N) = b.$$

- $M(A, 1, N)$  can be optimized by merging  $M'(A, 1, N)$  with it and returning  $b$  in the first loop (the  $M'$  loop) if  $\frac{k-i+1}{2} > \frac{N}{2}$  since then we have

$$\begin{aligned} \#A_{i,j}(b) &= \frac{k-i+1}{2} > \frac{N}{2} \\ \Rightarrow \#A_{1,N}(N) &\geq \#A_{i,j}(N) > \frac{N}{2} \\ \Rightarrow M(A, 1, N) &= b. \end{aligned}$$