Design and Analysis of Algorithms

Part I: Divide and Conquer

Lecture 4: Maximum Contiguous Subarray Problem


Ke Xu and Yongxin Tong

（许可 与 童咏昕）

School of CSE, Beihang University

Outline

- Introduction to Part I

- Maximum Contiguous Subarray Problem
  - Problem definition
  - A brute force algorithm
  - A data-reuse algorithm
  - A divide-and-conquer algorithm
  - Analysis of the divide-and-conquer algorithm

Introduction to Part I

- Divide-and-conquer (D&C) is an important algorithm design paradigm.
  - **Divide**
    Dividing a given problem into two or more subproblems (ideally of approximately equal size)
  - **Conquer**
    Solving each subproblem (directly if small enough or recursively)
  - **Combine**
    Combining the solutions of the subproblems into a global solution

Introduction to Part I

- In Part I, we will illustrate Divide-and-Conquer using several examples:
  - Maximum Contiguous Subarray (最大子数组)
  - Counting Inversions (逆序计数)
  - Polynomial Multiplication (多项式乘法)
  - QuickSort and Partition (快速排序与划分)
  - Randomized Selection (随机化选择)
  - Lower Bound for Sorting (基于比较的排序下界)

Outline

- Introduction to Part I

- Maximum Contiguous Subarray Problem
  - Problem definition
  - A brute force algorithm
  - A data-reuse algorithm
  - A divide-and-conquer algorithm
  - Analysis of the divide-and-conquer algorithm

Maximum Contiguous Subarray (MCS) Problem

ACME Corp[1] – PROFIT HISTORY

Year	1	2	3	4	5	6	7	8	9
Profit M$	-3	2	1	-4	5	2	-1	3	-1

Between years 1 and 9:

- ACME earned $-3+2+1-4+5+2-1+3-1=4$ M$

Between years 2 and 6:

- ACME earned $2+1-4+5+2=6$ M$

Between years 5 and 8:

- ACME earned $5+2-1+3=9$ M$

如果所有数组元素都是非负数，整个数组和肯定是最大

Problem: Find the span of years in which ACME earned the most

Answer: Year 5-8, 9 M$

[1]A Company that Makes Everything

Formal Definition

- Input: An array of reals $A[1...n]$
- The value of subarray $A[i...j]$ is

$$V(i,j) = \sum_{x=i}^{j} A(x)$$

Definition (Maximum Contiguous Subarray Problem)

Find $i \leq j$ such that $V(i,j)$ is maximized.

Outline

- Introduction to Part I

- Maximum Contiguous Subarray Problem
  - Problem definition
  - A brute force algorithm
  - A data-reuse algorithm
  - A divide-and-conquer algorithm
  - Analysis of the divide-and-conquer algorithm

A Brute Force Algorithm

Calculate the value of $V(i, j)$ for each pair $i \leq j$ and return the maximum value

```
VMAX ← A[1];
for i ← 1 to n do
    for j ← i to n do
        // calculate V(i,j)
        V ← 0;
        for x ← i to j do
            V ← V + A[x];
        end
        if V > VMAX then
            VMAX ← V;
        end
    end
end
return VMAX
```

$O(n^3)$ arithmetic additions

Outline

- Introduction to Part I

- Maximum Contiguous Subarray Problem
  - Problem definition
  - A brute force algorithm
  - A data-reuse algorithm
  - A divide-and-conquer algorithm
  - Analysis of the divide-and-conquer algorithm

A Data-Reuse Algorithm

Idea:

- don't need to calculate each V(*i, j*) from scratch
- exploit the fact: $V(i,j) = \sum_{x=i}^{j} A[x] = V(i,j-1) + A[j]$

```
VMAX ← A[1];
for i ← 1 to n do
    V ← 0;
    for j ← i to n do
        // calculate V(i,j)
        V ← V + A[j];
        if V > VMAX then
            VMAX ← V;
        end
    end
end
return VMAX
```

$O(n^2)$ arithmetic additions

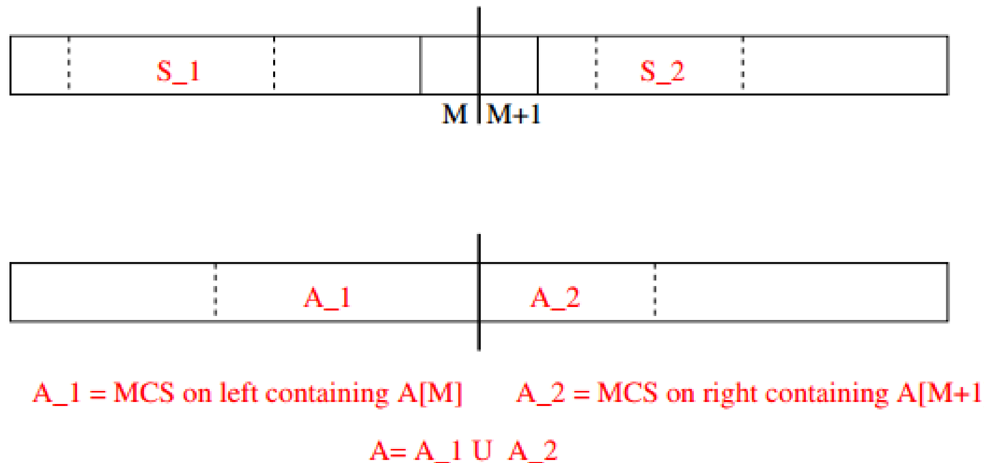
Outline

- Introduction to Part I

- Maximum Contiguous Subarray Problem
  - Problem definition
  - A brute force algorithm
  - A data-reuse algorithm
  - A divide-and-conquer algorithm
  - Analysis of the divide-and-conquer algorithm

A Divide-and-Conquer Algorithm

Set $m = \lfloor (n+1)/2 \rfloor$



The MCS S must be one of

1. S_1: the MCS in $A[1 \ldots m]$
2. S_2: the MCS in $A[m+1 \ldots n]$
3. A: the MCS across the cut.

So,

$$S = \text{the best among } \{S_1, S_2, A\}$$

最终，在S_1, S_2和A(跨越中点的最大子数组)这三种情况中选取和最大者

An Example of Divide-and-Conquer Algorithm

1	-5	4	2	-7	3	6	-1	2	-4	7	-10	2	6	1	-3

- $S_1 = [3, 6]$ and $S_2 = [2, 6, 1]$

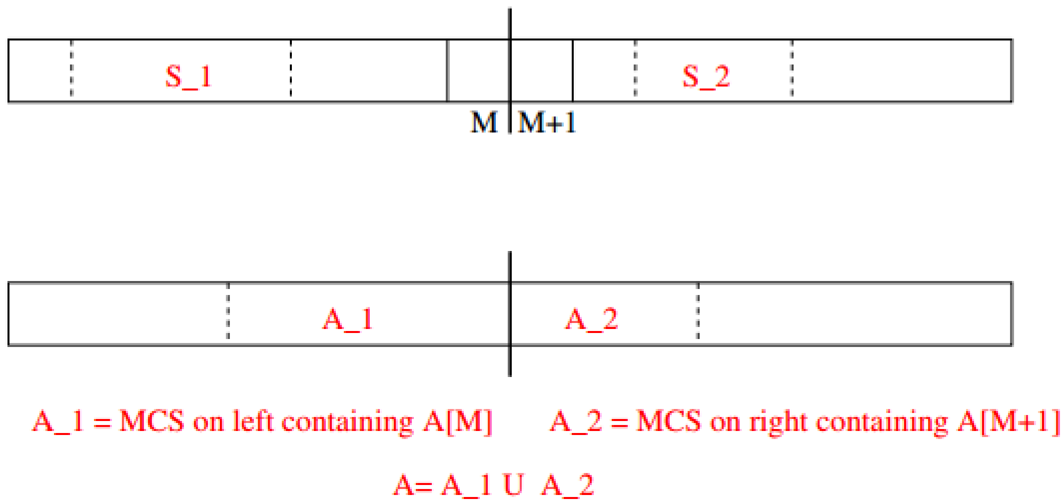
1	-5	4	2	-7	3	6	-1	2	-4	7	-10	2	6	1	-3

- $A_1 = [3, 6, -1]$ and $A_2 = [2, -4, 7]$
- $A = A_1 \cup A_2 = [3, 6, -1, 2, -4, 7]$

- $Value(S_1) = 9$; $Value(S_2) = 9$; $Value(A) = 13$
- solution: A

Divide: MCS across The Cut

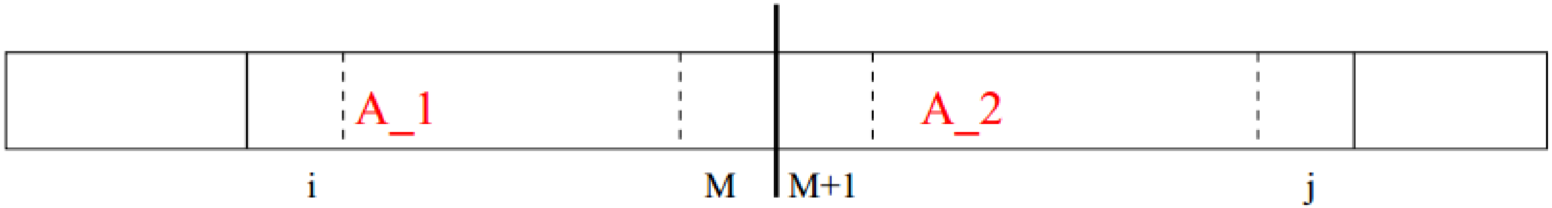
Set $m = \lfloor (n+1)/2 \rfloor$



$A = A_1 \cup A_2$

- A_1: MCS among contiguous subarrays ending at A[m]
- A_2: MCS among contiguous subarrays starting at A[m+1]

Conquer: Finding the "A_1" Subarrays



A_1 is in the form $A[i \dots m]$, $V(i, m) = V(i+1, m) + A[i]$

MAX $\leftarrow A[m]$;
SUM $\leftarrow A[m]$;
for $i \leftarrow m - 1$ **downto** *1* **do**
  SUM $\leftarrow$ SUM $+A[i]$;
  if *SUM* $>$ *MAX* **then**
    MAX $\leftarrow$ SUM;
  end
end
A_1=MAX;

Conquer: Finding "*A*" with A Linear Time

- There are only m sequences of the form
  - A_1 can be found in $O(m)$ time
- Similarly, A_2 is in the form $A[m+1 \ldots j]$
  - there are only $n-m$ such sequences
  - A_2 can be found in $O(n-m)$ time
- $A = A_1 \cup A_2$ can be found in $O(n)$ time
  - linear to the input size

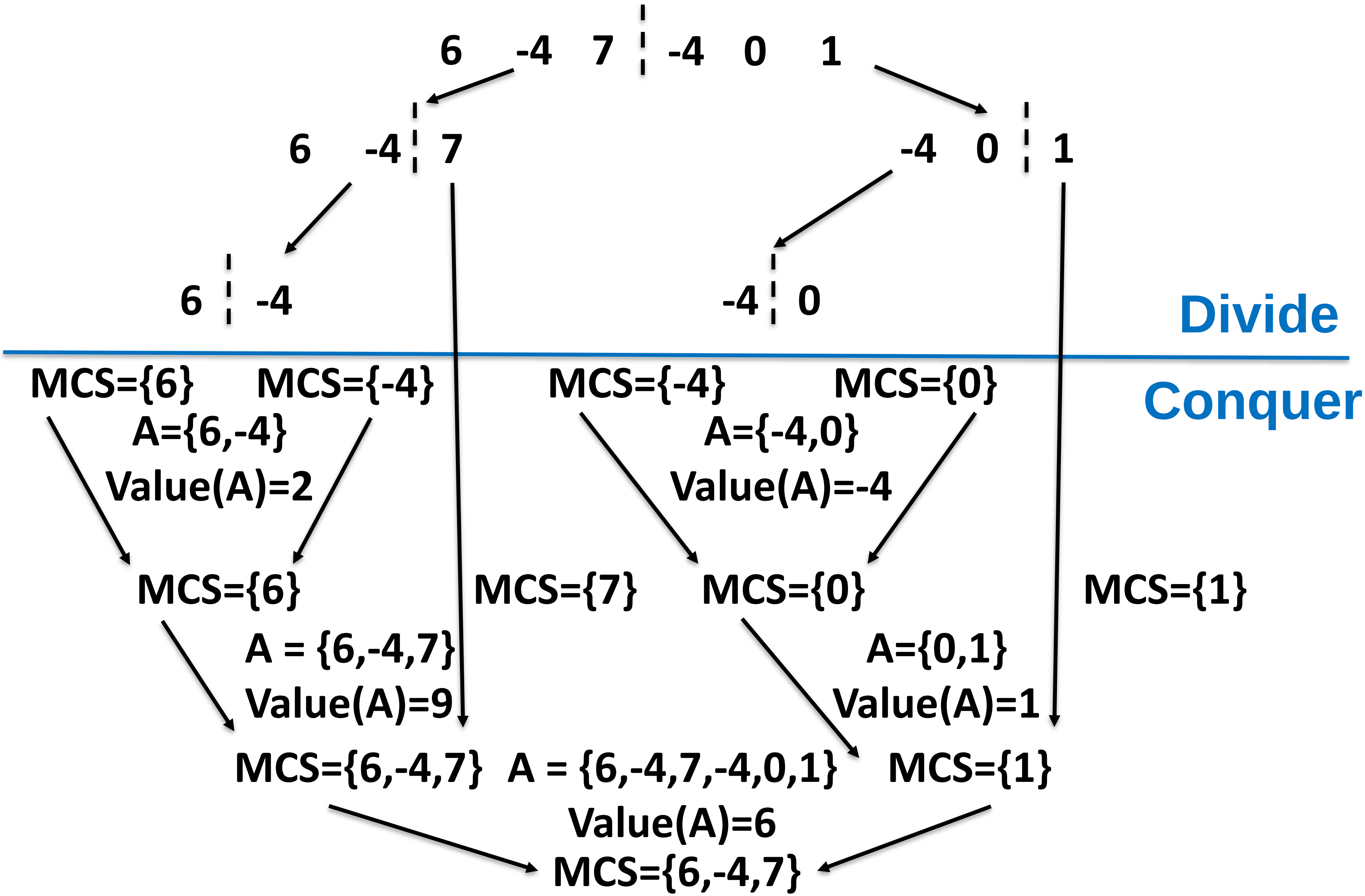
The Complete Divide-and-Conquer Algorithm

$MCS(A, s, t)$

Input: $A[s \dots t]$ with $s \leq t$
Output: MCS of $A[s \dots t]$
begin
    if $s = t$ **then return** $A[s]$;
    else
        $m \leftarrow \lfloor \frac{s+t}{2} \rfloor$;
        Find $MCS(A, s, m)$;
        Find $MCS(A, m+1, t)$;
        Find MCS that contains both $A[m]$ and $A[m+1]$;
        return maximum of the three sequences found
    end
end

First Call: *MCS*(*A, 1, n*)

A Full Illustration of the D&C Algorithm



Outline

- Introduction to Part I

- Maximum Contiguous Subarray Problem
  - Problem definition
  - A brute force algorithm
  - A data-reuse algorithm
  - A divide-and-conquer algorithm
  - Analysis of the divide-and-conquer algorithm

Analysis of the D&C Algorithm

- n: problem size ($n = t - s + 1$)
- $T(n)$: time needed to run $MCS(A, s, t)$

begin
  if $s = t$ **then return** $A[s]$ // O(1)
  else
    $m \leftarrow \lfloor \frac{s+t}{2} \rfloor$;
    Find $MCS(A, s, m)$; // $T(\lceil \frac{n}{2} \rceil)$
    Find $MCS(A, m+1, t)$; // $T(\lfloor \frac{n}{2} \rfloor)$
    Find MCS that contains *both* $A[m]$ and $A[m+1]$; // O(n)
    return maximum of the three sequences found // O(1)
  end
end

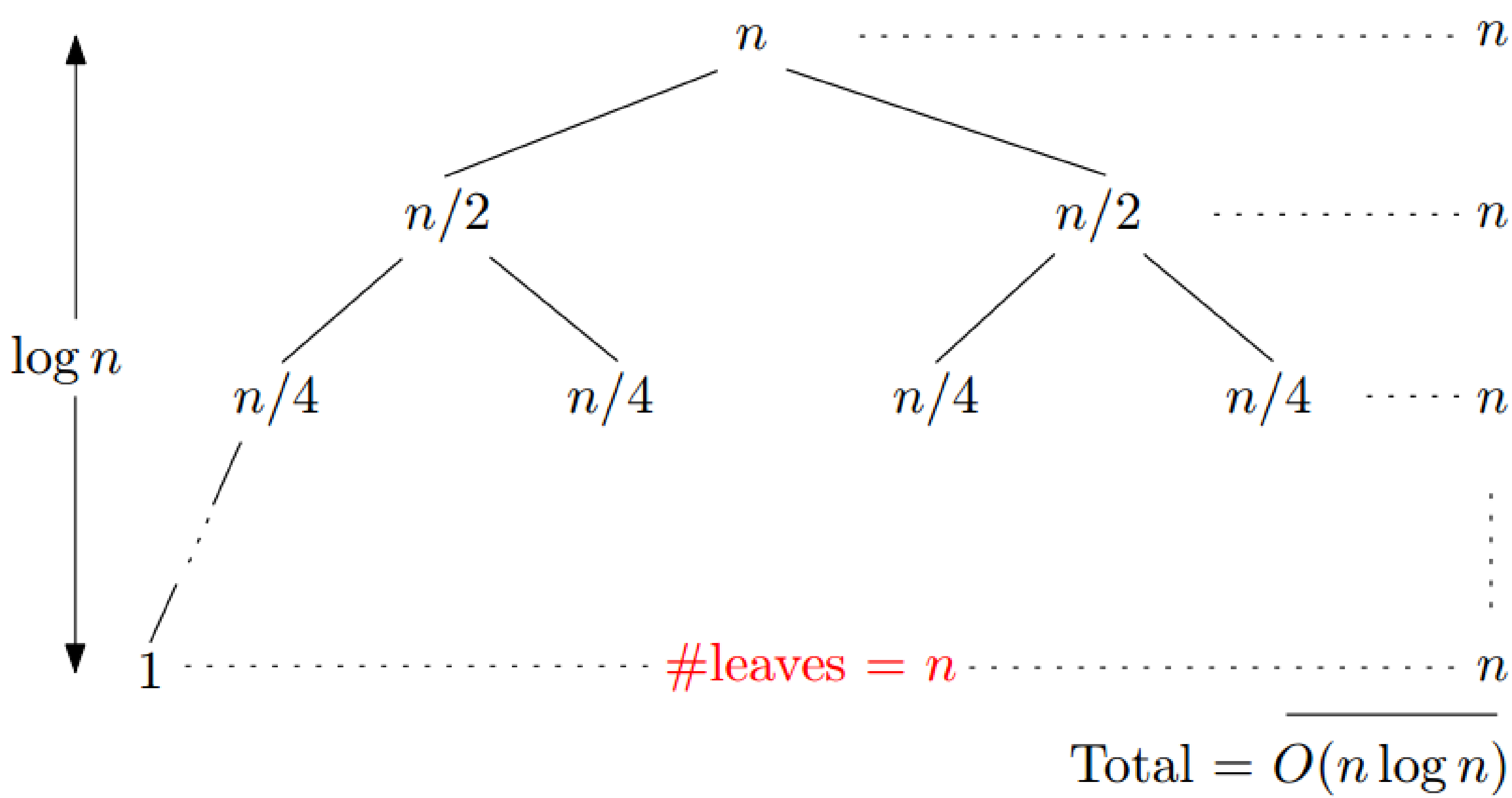
$$T(1) = O(1)$$

$$T(n) = T(\lceil n/2 \rceil) + T(\lfloor n/2 \rfloor) + O(n) \quad \text{for } n > 1$$

Analysis of the D&C Algorithm

To simplify the analysis, we assume that n is a power of 2

$$T(n) = \begin{cases} 2T(n/2) + n, & \text{if } n > 1, \\ 1, & \text{if } n = 1. \end{cases}$$



Summary

In the MCS problem, we saw 3 different algorithms for solving the maximum contiguous subarray problem

- A $O(n^3)$ brute force algorithm
- A $O(n^2)$ algorithm that reuses data
- A $O(n \log n)$ divide-and-conquer algorithm

Can you solve the problem in $O(n)$ time?

谢谢

BDA
Beihang University