CME 323: Distributed Algorithms and Optimization

Instructor: Reza Zadeh (rezab@stanford.edu)
TA: Anuj Nagpal (anujnag@stanford.edu)
HW#1 – Due Thursday April 14 (on Gradescope)

1. **(5 points)** The Karatsuba algorithm multiplies two integers x and y. Assuming each has n bits where n is a power of 2, it does this by splitting the bits of each integer into two halves, each of size $n/2$. For any integer x we will refer to the low order bits as x_l and the high order as x_h. The algorithm computes the result as follows:[1]

   function $\mathbf{km}(x, y, n)$:
   if $n = 1$ **then**
     return $x \times y$
   else
     $a \leftarrow \mathbf{km}(x_l, y_l)$
     $b \leftarrow \mathbf{km}(x_h, y_h)$
     $c \leftarrow \mathbf{km}(x_l + x_h, y_l + y_h)$
     $d \leftarrow c - a - b$
     return $(b2^n + d2^{n/2} + a)$
   end if

   Note that multiplying by 2^k can be done just by shifting the bits over k positions.

   (a.) Assuming addition, subtraction, and shifting take $O(n)$ work and $O(n)$ depth what is the work and depth of **km**?

   (b.) Assuming addition, subtraction, and shifting take $O(n)$ work and $O(\log n)$ depth what is the work and depth of **km**?

2. **(5 points)** Suppose a square matrix is divided into blocks:

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

   where all the blocks are the same size. The *Schur complement* of block D of M is $S = A - BD^{-1}C$. The inverse of the matrix M can then be expressed as:

$$M^{-1} = \begin{bmatrix} S^{-1} & S^{-1}BD^{-1} \\ -D^{-1}CS^{-1} & D^{-1} + D^{-1}CS^{-1}BD^{-1} \end{bmatrix}$$

   This basically defines a recursive algorithm for inverting a matrix which makes two recursive calls (to calculate D^{-1} and S^{-1}), several calls to matrix multiply, and one each to elementwise add and subtract two matrices. Assuming that matrix multiply has work $O(n^3)$ and depth $O(\log n)$ what is the work and depth of this inversion algorithm?

[1] If you have seen this before, you might have thought of it as a sequential algorithm, but actually it is a parallel algorithm, since in particular, the three recursive calls to **km** can be made in parallel.

3. **(8 points) Curly Brace Matching:** We want to make an algorithm that solves curly brace matching problem i.e. given a code snippet as a string with whitespaces removed, we want to identify whether each opening curly brace '{' in the string has a corresponding closing curly brace '}'. For example, "{{a = 2}{{b += 4}{c = a - b}}}" has valid curly brace matching but "{{{a = 4}{b = 2}}}}" or "{a*=2}{{b=3}}{c=a+b{}" does not.

   (a) Give a sequential algorithm to solve the curly brace matching problem using a stack. Your algorithm must take $O(n)$ work and $O(n)$ extra space.

   (b) Give a sequential algorithm to solve the problem without a stack such that your algorithm takes $O(n)$ work but $O(1)$ extra space.

   (c) Give a parallel algorithm to compute the total number of left curly braces "{" and the right parentheses "}", respectively (if these two numbers do not match, we directly know that the string does not have a valid curly brace matching). Your algorithm must run in $O(n)$ work and $O(\log n)$ depth. You can use $O(n)$ extra space.

   (d) Give a parallel algorithm to solve the curly brace matching problem. Your algorithm must take $O(n)$ work and $O(\log n)$ depth. You can use $O(n)$ extra space. Prove the cost of your algorithm.

4. **(7 points)** Describe a divide-and-conquer algorithm for merging two sorted arrays of lengths n into a sorted array of length $2n$. It needs to run in $O(n)$ work and $O(\log^2 n)$ depth. You can write the pseudocode for your algorithm so that it looks like your favorite sequential language (C, Java, Matlab, ...), but with an indication of which loops or function calls happen in parallel. For example, use `parallel for` for a parallel for loop, and something like:

```
parallel {
    foo(x,y)
    bar(x,y)
}
```

   to indicate that `foo` and `bar` are called in parallel. You should prove correctness at the level expected in an algorithms class (e.g. CME305 or CS161).

5. **(6 points)** Given the price of a stock at each day for n days, we want to determine the biggest profit we can make by buying one day and selling on a later day. For example, the following stock prices have a best profit of 5:

   $[12, 11, 10, 8, 5, 8, 9, 6, 7, 7, 10, 7, 4, 2]$

   since we can buy at 5 on day 5 and sell at 10 on day 11. This has a simple linear time serial solution. Give an algorithm to solve this problem that runs in $O(n)$ work and $O(\log n)$ depth. Give pseudocode as in the previous problem.

6. **(10 points)** In this problem, we'll look at how fast the maximum of a set of n elements can be computed when allowing for concurrent writes. In particular we allow the arbitrary write rule for "combining" (i.e. if there are a set of parallel writes to a

location, one of them wins). Show that this can be done in $O(\log\log n)$ depth and $O(n)$ work.

(a.) Describe an algorithm for maximum that takes $O(n^2)$ work and $O(1)$ depth (using concurrent writes).

(b.) Use this to develop an algorithm with $O(n)$ work and $O(\log\log n)$ depth. Hint: use divide and conquer, but with a branching factor greater than 2.

7. **(8 points) Sorting Student IDs:** Consider the task of sorting a list of 8-digit Stanford Student IDs of the form $d_1d_2...d_8$ where each d_i is a positive integer with $0 \leq d_i \leq 9$. This task can be done using standard sorting algorithms in $O(n \log n)$ time complexity but we can utilize the fact that all the numbers we want to sort are positive integers, whose each digit can be bucketed into 10 categories.

   (a) Based on the bucketization idea, design a sequential algorithm that starts scanning the numbers from digit d_8 to d_1 and sorts them in $O(8n)$ time and $O(n)$ space complexity. Also provide pseudocode for same. (note that the idea behind the factor of 8 in time complexity is to hint that the algorithm can be extended to k-digit integers with $O(kn)$ time complexity).

   (b) Using the same bucketization idea, design a parallel algorithm that starts scanning the numbers from digit d_1 instead and sorts them with $O(n \log_{10} n)$ work, $O(n)$ depth and $O(n)$ memory requirement (assume that this bucketization can't be done in parallel). Also provide pseudocode for same.

   (c) Explain how can we improve the depth to $O(\log_{10} n)$ if we can also do the bucketization of elements in parallel. Assume you have sufficient number of processors.

8. **(8 points) Scheduling to Minimize Lateness** Consider a problem where we have a single resource and a set of n requests to use the resource for an interval of time. Assume each request has a deadline d_i, and a length t_i required to complete the request. For each job we need to assign a start time $s(i)$ (its finish time is $f(i) = s(i) + t_i$). We say request i is late if $f(i) > d_i$, and the lateness of task i is given by $l_i = max\{0, f(i) - d_i\}$. Design a greedy algorithm that minimizes the maximum lateness across all jobs, and prove that the resulting schedule is optimal.

9. **(15 points) Solving Linear Systems**

   Lower Triangular Systems Consider the task of solving the linear system $Ax = b$ where we assume A is lower triangular. A popular method for solving $Ax = b$ is *forward substitution.* The forward substitution algorithm can be represented as the following series of serial updates:

   $x_1 \leftarrow b_1/a_{11}$
   for $i = 2, \ldots, n$ **do**
   $\quad x_i \leftarrow \left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j\right)/a_{ii}$
   end for

   (a) What is the computation complexity of the forward substitution algorithm?

(b) The parallel forward substitution algorithm operates by parallelizing the serial forward substitution algorithm. Note that the y_j updates can all be executed in parallel.

$x_1 \leftarrow b_1/a_{11}$
for $j = 1, \ldots, n$ **do**
  $y_j \leftarrow a_{j1}x_1$
end for
for $i = 2, \ldots, n$ **do**
  $x_i \leftarrow (b_i - y_i)/a_{ii}$
  for $j = i+1, \ldots, n$ **do**
    $y_j \leftarrow y_j + a_{ji}x_i$
  end for
end for

What is the depth of the DAG representing the parallel forward substitution algorithm?

Tridiagonal Systems We now consider solving the system $Ax = b$ where A is tridiagonal. Explicitly, $a_{ij} = 0$ if $|i - j| \geq 2$. Note that this is equivalent to solving the following system of linear equations:

$$\begin{aligned} g_1x_1 + h_1x_2 &= b_1 \\ f_ix_{i-1} + g_ix_i + h_ix_{i+1} &= b_i, \quad i = 2, \ldots, n-1 \\ f_nx_{n-1} + g_nx_n &= b_n \end{aligned}$$

where g_i are the diagonal elements of A, f_i the entries below the diagonal, and h_i the entries above the diagonal. The idea behind *even-odd reductions* is to recursively reduce the above system to one of half the size. Explicitly, if none of the diagonal entries are zero, we can solve for each x_i in terms of x_{i-1} and x_{i+1}. If we do this for all odd i, and substitute the expression back in, we obtain a system on just the even indexed variables.

(a) Using the above system of equations, derive a tridiagonal system of equations on just the even indexed variables.

(b) What is the computational complexity of computing the coefficients of the reduced system?

The above procedure can be recursively applied until the problem is reduced to a single equation. Then we work backwards to solve for the value of the eliminated variables.

(c) What is the computational complexity of solving for the eliminated variables?

(d) Construct the DAG representing this algorithm.

(e) What is the runtime of the even odd reduction algorithm on $O(n)$ processors?

10. **(8 points) Givens Rotations:** Givens Rotations are used to zero out the subdiagonal entries of the matrix A one at a time. Crucially, a Givens rotation only affects two rows of the matrix. We will use this fact to derive a parallel implementation of the Givens rotation algorithm. Specifically, if two successive Givens rotations affect disjoint sets of rows, then they can be computed in parallel.

    (a) When n rows are available, what is the maximum number of Givens rotations we can apply simultaneously?

    (b) Implementing the Givens rotations in parallel ultimately comes down to deriving a schedule of the entries to eliminate at a particular step. We consider two functions $T(j,k)$ and $S(j,k)$ where $T(j,k)$ represents the iteration in which the jk-th entry is eliminated, and j and $S(j,k)$ are the rows the Givens rotation operates on. To simulateneously implement the Givens rotations, we require that $T(j,k)$ and $S(j,k)$ satisfy:

    - If $T(j,k) = T(j',k')$ and $(j,k) \neq (j',k')$ then $\{j, S(j,k)\} \cap \{j', S(j',k')\} = \emptyset$.
    - If $T(j,k) = t$ and $S(j,k) = i$, then $T(j,l) < t$ and $T(i,l) < t$ for all $l < k$.

    Prove that the schedule given by

    $$T(j,k) = n - j + 2k - 1$$
    $$S(j,k) = j - 1$$

    satisfies the above conditions.

    (c) What is the maximum number of stages required by this schedule?