Study for CS477 Examination April 10, 2024

In any problem involving grasphs or directed graphs, we use n to denote the number of vertices and m the number of edges, or arcs.

1. Fill in the blanks.

   (i) Name two greedy algorithms introduced in class this semester. _____ _____

   (ii) In closed hashing, collisions are resolved by the use of _____sequences.

   (iii) The items in a priority queue represent _____ _____.

   (iv) Name three kinds of search structures.

   _____ _____ _____

   (v) _____ hashing, which can be used by a compiler to identify reserved words, does not have collisions.

   (vi) In closed hashing, if a collision occurs, a _____ _____ can be used to locate an unused position in the hash table.

   (vii) In a _____ hash table, each item has two or more possible locations, and must be stored in one of those.

   (viii) _____ Which of the following three statements is closest to the truth?
   (1) In SHA256 hashing, collisions are impossible.
   (2) In SHA256 hashing, collisions occur no more than once a year in practice.
   (3) In SHA256 hashing, collisions are so unlikely that industry experts claim they never occur.

   (ix) The worst case time complexity of quicksort on a list of length n.

   _____

   (x) The average case time complexity of quicksort on a list of length n, if pivots are chosen at random.

   _____

   (xi) The worst case time complexity of building a treap with n items.

   _____

   (xii) The average case time complexity of building a treap with n items.

   _____

   (xiii) In an open hash table of size m holding n data items, the items at each index of the table are typically shown as linked list. However, that structure is only efficient if m/n is fairly small. In general, we should use a _____ at each table index.
   Pick one of these answers:
   heap
   stack
   search structure

(xiv) A directed graph is defined to be _____ if, given any two vertices x and y, the graph contains a path from x to y.

2. For each of these recursive subprograms, write a recurrence for the time complexity, then solve that recurrence.

(i)
```
void george(int n)
{
 if(n > 0)
  {
   for(int i = 0; i < n; i++) cout << "hello" << endl;
   george(n/2); george(n/3);
  }
}
```

(ii)
```
void martha(int n)
{
 if(n > 0)
  for(int i = 1; i < n; i++)
   for(int j = 1; j < i; j++)
    cout << "hello world";
  martha(n/2);
}
```

3. A 3-dimensional $10 \times 20 \times 12$ rectangular array A is stored in main memory in column major order, and its base address is 1024. Each item of A takes two words of main memory, that is, two addressed location. Find the address, in main memory, of A[5][13][7].

4. You are trying to construct a cuckoo hash table of size 8, where each of the 8 names listed below has the two possible hash values indicated in the array. Put the items into the table, if possible. Instead of erasing ejected items, simply cross them out, so that I can tell that you worked it properly.

	h1	h2
Ann	1	0
Bob	4	2
Cal	0	7
Dan	1	6
Eve	1	0
Fay	6	2
Gus	5	3
Hal	3	7

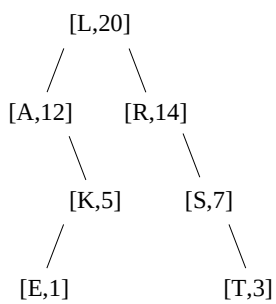
5. Let $\sigma = x_1, x_2, \ldots x_n$ be a sequence of numbers with both positive and negative terms. Write an $O(n)$ time dynamic program which finds the maximum sum of any contiguous subsequence of σ. For example, if the sequence is $-1, 4, -3, 2, 7, -5, 3, 4, -8, +6$ then the answer is $4 - 3 + 2 + 7 - 5 + 3 + 4 = 12$.

   $X[i]$ = maximum sum of any legal subsequence of $x_1, \ldots x_i$
   $Y[i]$ = maximum sum of any legal subsequence which ends at x_i.

   The program starts out as:
   $X[0] = 0$
   $Y[1] = x_1$
   etc.

   I have not gone over this algorithm in class. But you can do it anyway.

6. The figure below shows a treap, where the data are letters and the nodes of the tree are memos, where the first component is the <u>key</u>, a letter, and the second component is a the <u>priority</u>, a random integer. Insertion of the letter G, where the priority is chosen (at random) to be 17. Show the steps.



7. Explain how to implement a sparse array using a search structure.

8. Consider the following two C++ subprograms.

```
int f(int n)
  {
   if(n > 0)
    return f(n/2)+f(n/3)+f(n/6)+n*n;
   else
    return 0;
  }

void computef(int n)
 {
  f[0] = 0;
  for(int i = 0; i <= n; i++)
   f[i] = f[i/2]+f[i/3]+f[1/6]+n*n;
 }
```

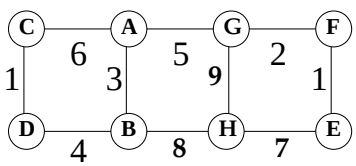
(i) The first of those subprograms is a recursive function. What is the asymptotic value of f(n)?

(ii) What is the asymptotic time complexity of the computation of f(n) using the recursive function?

(iii) The second subprogram uses dynamic programming, and stores values in an array. What is the asymptotic time complexity of that computation?

(iv) What is the asymtotic time complexity of a computation of f(n) using memoization? (Hint: it's a power of log n.) ---

9. Walk through Kruskal's algorithm to find the minimum spanning tree of the weighted graph shown below. Show the evolution of the union/find structure. Whenever there is choice between two edges of equal weight, choose the edge which has the alphabetically largest vertex. Whenever there is a union of two trees of equal weight, choose the alphabetically larger root to be the root of the combined tree. Indicate path compression when it occurs.



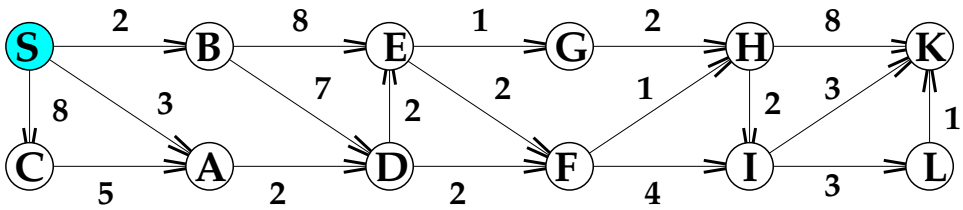
10. Write the prefix expression equivalent to the infix epression $-a * b - (-c - d) \wedge e$
(Don't forget that $\wedge$ means exponentiation.)

11. Walk through the stack algorithm to change the infix expression $-a + b \wedge c \wedge -f$ to postfix. Show the stack at each step.

12. Up to now, no one has written a polynomial time algorithm for the subset sum problem, given below. However, there is a pseudopolynomial time algorithm. The problem is to find a subsequence of a sequence of positive integers $x[1], \ldots x[n]$ whose sum is K. The time complexity of the algorithm is is $O(nK)$, so you might think the algorithm is polynomial time, but it isn't. Write code or pseudocode for the pseudopolynomial time algorithm for deciding whether there is a solution to a given instance of the subset sum problem.

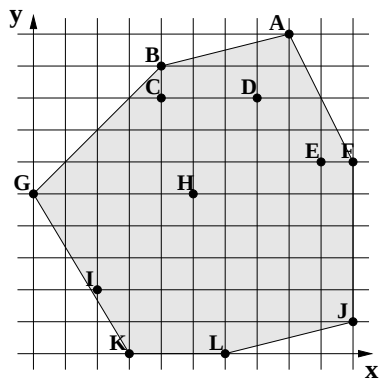
13. For the algorithm to be polynomial time, its time complexity must be a polynomially bounded funtion of the input size of the instance. For the instance given in the previous problem, the input size is not actually $O(nK)$. What is it?

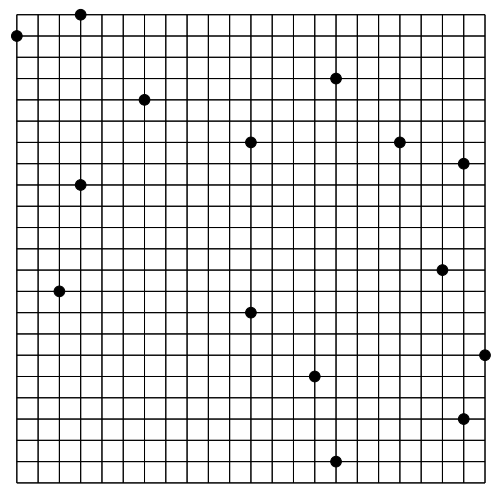
14. Walk through Dijkstra's algorithm for the following graph.



15. The convex hull of a set of a finite set of points in a plane is the smallest convex polygon which encloses the points, together with its interior. Walk through Graham Scan to find the convex hull of the points in the plane given in this figure. (I have not gone over Graham Scan yet.)

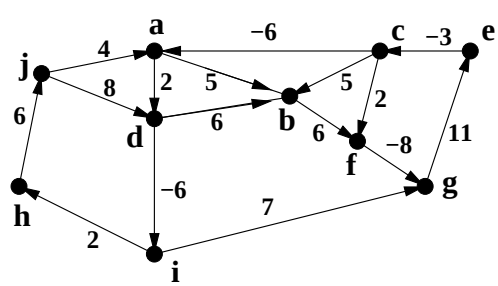
Here is an example, showint the convex hull of the set {A,B,C,D,E,F,G,H,I,J,K}.



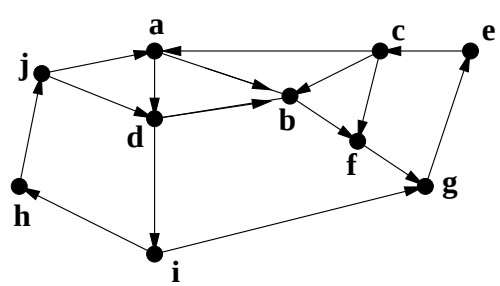


16. Figure (a) below shows an instance of the all-pairs minpath problem. Work the first part of Johnson's algorithm on that graph, showing the adjusted weights in Figure (b).

Do not complete the computation of Johnson's algorithm.



(a)



(b)

17. Walk through heapsort for the list BGHKRET.

18. True or False.

(i) _____ If there are 100 data items and 200 possible hash values, a collision is so unlikely that you can, in practice, assume that it won't happen.

(ii) _____ If there are 100 data items and 10000 possible hash values, collisions are so unlikely that you can, in practice, assume that they won't happen.

(iii) _____ Open hashing uses open addressing.

(iv) _____ Open hashing uses probe sequences.

(v) _____ You can avoid collisions in a hash table by making the table twice as large as the data set.

(vi) _____ False overflow for a queue can be avoided by implementing the queue as a circular list.

(vii) _____ If a stack is implemented as a linked list, the head of the linked list should hold the top item of the stack.

(i) _____ Kruskal's algorithm uses dynamic programming.

(ii) _____ There will be no collisions if the size of a hash table is at least ten times the number of data items.

(iii) _____ A hash function should appear to be random, but cannot actually be random.

19. Solve each recurrence, expressing each answer in terms of O, Ω, or Θ, whichever is most appropriate.

    (i) $F(n) = F(n/3) + F(2n/3) + 1$

    (ii) $G(n) = 2G(n/4) + \sqrt{n}$

    (iii) $H(n) = \log n + 1$

    (iv) Solve the recurrence: $H(n) < 4H(2n/5) + H(3n/5) + 2n^2$

    (v) Solve the recurrence: $G(n) = 4(G(n/2) + 5n^2$

    (vi) Solve the recurrence: $F(n) = F(n - \log n) + \log^2 n$

20. Find the time complexity of each of these code fragments in terms of n, using Θ notation.

    (i)
    ```
    for(int i = n; i > 1; i = i/2)
     cout < "Hello world!";
    ```

    (ii)
    ```
    for(int i = 2; i < n; i = i*i)
     cout < "Hello world!";
    ```

    (iii)
    ```
    for(int i = 1; i < n; i++)
     for(int j = i; j < n; j=2*j)
      cout < "Hello world!";
    ```

21. The asymptotic complexity of the Floyd/Warshall algorithm is _____.

22. The asymptotic complexity of Dijkstra's algorithm algorithm is _____.

23. Write pseudocode for the Floyd Warshall algorithm.

24. Write pseudocode for the Bellman Ford algorithm. Be sure to include the shortcut that stops execution when further computation is unnecessary.

25. Here is another coin-row problem. You have a row of coins of various values, where the value of the ith coin is $V[i] > 0$. Write pseudocode which finds the maximum value of a subset of coins, where the set may not contain coins which are either adjacent or just one apart in the row. That is, if the set contains the ith coin, it may not contain either the $(i+1)^{\text{st}}$ coin or the $(i+2)^{\text{nd}}$ coin. For example, if the coins are ⓐ ⓑ ⓒ ⓓ ⓔ ⓕ ⓖ ⓗ in that order, the subset may be {ⓐ, ⓓ, ⓗ}, but not {ⓑ, ⓓ, ⓖ}.

26. Execute heapsort for the list BXQVRST. Show the array at each step, and identify the step at which the array is a heap for the first time.

B	X	Q	V	R	S	T

27. A _____ hash function fills the hash table exactly with no collisions.

28. _____ algorithm finds a binary code so that the code for one symbol is never a prefix of the code for another symbol.

29. An acyclic directed graph with 9 vertices must have at least _____ strong components. (Must be exact answer.)

30. In _____ _____ there can be any number of items at a given index of the hash table.

31. The asymptotic expected time to find the median item in an unordered array of size n, using a randomized selection algorithm, is _____.

32. Give the asymptotic complexity, in terms of n, for each of these code fragments.

    (i)
    ```
    for(int i = 0; i < n; i++)
        for(int j = n; j > i; j = j/2)
    ```
    (ii)
    ```
    for(int i = 0; i < n; i++)
        for(int j = i; j > 0; j = j/2)
    ```

33. Solve each recurrence, giving asymptotic answers, using O, Ω, or Θ, whichever is most appropriate.

    (i) $F(n) \leq 4F(n/2) + n^2$

    (ii) $G(n) \geq G(4n/5) + G(3n/5) + n^2$

    (iii) $T(n) = T(3n/10) + T(n/5) + n$

34. You need to store the items A, B, and C, in that order, in a treap. The priority for A is 13, for B is 8, and for C is 14. Use maxheap order. Draw the resulting treap after each insertion, and show each rotation.

35. The BFPRT selection algorithm has asymptotic time complexity $\Theta(n)$, which is proved using a recurrence. Give that recurrence.

36. Consider the function F computed by the recursive code given below.

    (i) What is the asymptotic complexity of $F(n)$?

    (ii) What is the asymptotic time complexity of the recursive code when it computes $F(n)$?

    (iii) What is the asymptotic time complexity of a memoization algorithm which computes $F(n)$?

```
int F(int n)
 {
  if(n < 3) return 1;
  else return F(n/3)+2*F((n+1)/3)+n*n;
 }
```

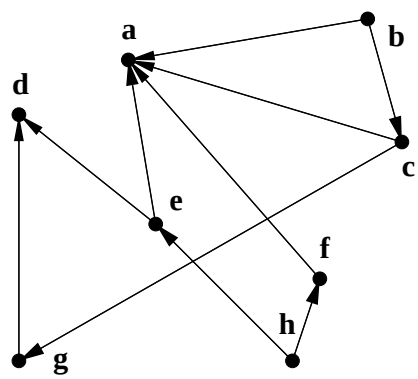
37. If the array $A[5][7]$ is stored in column-major order, how many predecessors does $A[3][4]$ have?

38. Explain how to implement a sparse array using a search structure.

39. You are implementing a 3D triangular array A where $A[i][j][k]$ is defined for $i \geq j \geq k \geq 0$, as a one-dimensional subarray of main memory, and you wish to store A in row-major order, with base address 1024. where $A[i][j][k]$ is defined for $i \geq j \geq k \geq 0$, Each term of A takes one place in main memory. What would be the address, in main memory, of $A[7][4][3]$?

40. You are given an acyclic directed graph $G = (V, E)$ where each arc has a positive weight. If (x, y) is an arc, we write $w(x, y)$ for the weight of that arc. Describe a dynamic programming algorithm which calculates the directed path through G of maximum weight. (The weight of a path is defined to be the sum of the weights of its constituent arcs.) (Hint: your algorithm should use words such as "in-neighbor" or "out-neighbor.")

41. Write the array of in-neighbor lists and the array of out-neighbor lists for the directed graph shown below.



a	
b	a,c
c	a,g
d	
e	a,d
f	a
g	d
h	e,f

a	b,c,e,f
b	
c	b
d	e,g
e	h
f	h
g	c
h	

42. Consider the following recursive C++ function.

```
int f(int n)
 {
  if(n > 0) return f(n/2)+f(n/4)+f(n/4 + 1)+n;
  else return 0;
 }
```

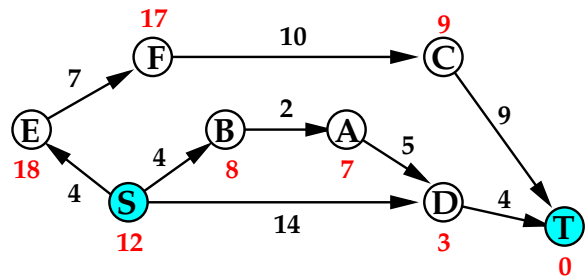
(i) What is the asymptotic complexity of f as a function of n, using Θ notation?

(ii) What is the asymptotic time complexity of this code as a function of n, using Θ notation?

(iii) Write pseudo-code for a dynamic programming algorithm to compute $f(n)$ for a given n. What is the asymptotic time complexity of your code as a function of n, using Θ notation?

(iv) Write pseudo-code for a memoization algorithm to compute $f(n)$ for a given n. What is the asymptotic complexity of the algorithm in terms of n, using Θ notation?

43. Walk through the A^* algorithm for the weighted directed graph shown below, where the pair is (S, T). The heuristic is shown as red numerals.



.

Show the arrays and the contents of the heap at each step. h is the heuristic, f is the current distance from the source, g is the sum of h and f, while back is the backpointer.

44. Find the Levenshtein edit distance from the word "mennoover" to the word "maneuver." Show the matrix.

    I will go over the algorithm on Monday.

45. (i) Fine the longest strictly monotone increasing subsequence of the sequence 1,5,2,2,4,8,7. The answer might not be unique. If there are choices, give just one answer.

    (ii) Write pseudocode for finding the length of the longest strictly monotone increasing subsequence of any given sequence of integers. (Hint: Use dynamic programming.)

    The algorithm is found in the handout titled, "Longest Monotone Subsequence," which you can find as lms.pdf, under Handouts.