Study Guide for the Final Exam

OR/ISE 501: Operations Research — Fall 2021

The Final Exam will be an individual in-class exam. The exam will be open notes and closed computer. The questions will be a mix of short answer and questions that are computationally easy enough to solve "by hand," but you can use a non-programmable calculator. The exam will be comprehensive but based primarily on the material covered in the final Review for the Final Exam section of the class.

Keep in mind that you will not have access to Julia during the exam, so you should be able to solve small, by-hand versions of the problems covered in class using a calculator. The exam will consist of (a) questions related to material not covered in Exams 1 and 2, mostly material not suitable for the take-home format of Exams 1 and 2; and (b) questions related to the problems in the HWs and on Exams 1 and 2 that can include (i) solving a small instance of the problem that can be solved with a calculator, (ii) justifying the posted solution approach, and (iii) possible extensions of the problems.

To prepare for the exam, study the following: (a) HWs 2–7; (b) the posted solutions for the HWs and Exams 1 and 2; (c) be able to apply by hand the key algorithmic procedures implemented in Julia to small problem instances; and (d) the following additional study problems:

1. Determine the local maxima, local minima, and inflection points of $f(x) = x(1 - x^2)$.

2. Determine the minimum of $f(x) = 5|x - 10| + 3|x - 30|$.

3. There are four stations located along a corridor, and a single testing device is to be located anywhere along the corridor so that it can receive samples from the stations. A technician must carry each sample from each station to the device. The distance in meters of the switches from the beginning of the corridor is 22, 5, 82, and 30, respectively, and the average number of daily samples per station is expected to be 3, 5, 4, and 2, respectively. Determine where the device should be located to minimize the total distance that the technician must travel.

4. Determine the minimum value of the function $x - x^3$ between –1.5 and $\sqrt{1/3}$ with a tolerance of 0.25 using Golden Section Search.

5. A testing center has a device that is able to use ultrasonic waves transmitted within a special solution to detect defects in parts like aircraft engines. The part is placed in a tank that, in order to operate correctly, is filled with just enough solution to submerge the part; the solution is then drained from the tank and disposed of since it can become contaminated during testing. Currently, test results are being delayed because it is time-consuming to fill the tank with the solution. In order to speed up testing, a suggestion has been made to consider pre-filling the tank and then either adding or draining the solution as needed to

submerge a part. The following demand data is available regarding the cubic volume of solution that was needed to test recent parts: 237, 214, 161, 146, and 331. Determine the cubic volume of solution that should be pre-filled in the tank to minimize the time required to submerge the part, given that the tank can be filled or drained at the same rate.

6. A new vaccine fulfillment center has been established that has the capacity to serve two customers. The dispensing and sterile packaging equipment can only fill one common-size container per customer because the set-up time to change the container size is very time-consuming. A single 235-unit container will be used for each customer's demand; any demand beyond the container capacity will be lost. The revenue received per unit of demand is \$5, and the cost to provide each unit is \$1. Once opened, each container must be used over a short period of time, and since the customers are dispersed geographically, it is not possible for them to share the vaccine. Determine the total profit given demands of 247 and 214 units per customer.

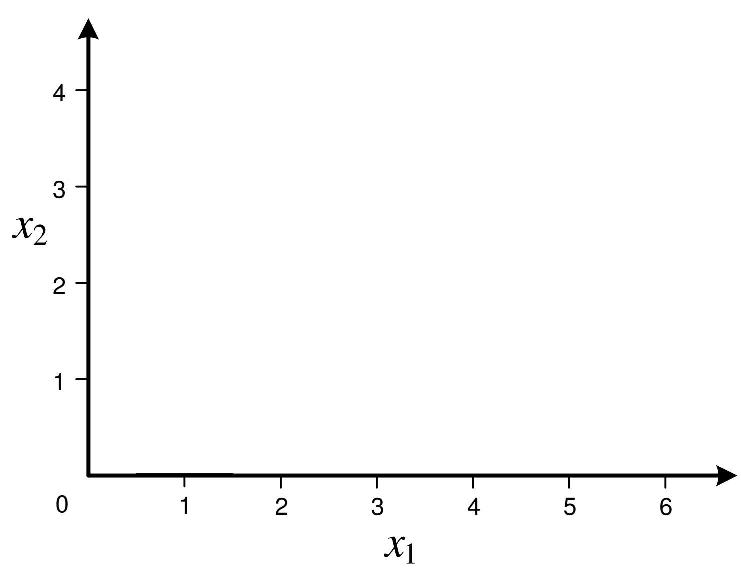
7. Explain how Brent's method is able to converge to the optimum solution faster than interval search.

8. Under what circumstances is using a loss function for regression that does not allow full information recovery a good feature?

9. In HW2-Q2, why was it necessary to use several different starting points to find your solution?

10. Draw the feasible region of the LP below along with the line of the objective passing through the optimal point.

$$\begin{aligned} \max \quad & 6x_1 + 8x_2 \\ \text{s.t.} \quad & 2x_1 + 3x_2 \le 11 \\ & 2x_1 \qquad\ \ \le 7 \\ & x_1, x_2 \ge 0 \end{aligned}$$

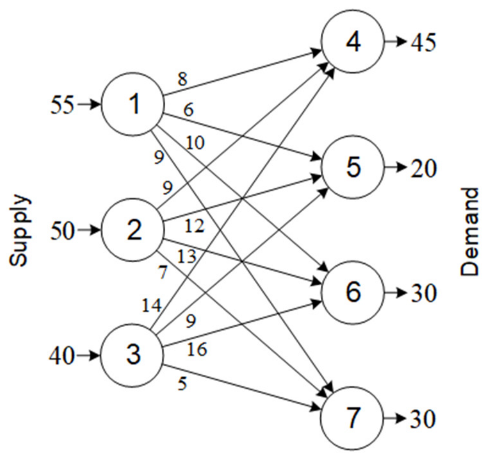


11. In Ex 4 of LP-3, why were continuous decision variables used to represent the number of employees even though each employee is a discrete person?

12. Explain why the simplex method always terminates.

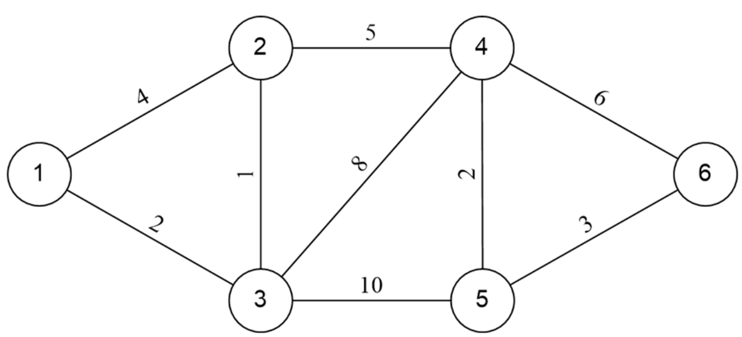
13. Given an LP with the decision variables x_1, and x_2 representing the production in tons of products 1 and 2, respectively, determine a constraint that can represent the requirement that product 1 must represent at least a quarter of total production.

14. For the network below,
   (a) Determine the solution returned by the following greedy procedure.
   (b) Explain why the supply at node 3 can increase without affecting the solution.
   (c) Explain why copies of the supply and demand arrays, `s,d = copy(s),copy(d)`, are made inside of the procedure, but a copy of `C` is not made.




```
function greedytrans(C,s,d)
    s, d = copy(s), copy(d)
    F = zeros(size(C))
    while !all(d .== 0)
        idxs = findall(s .!= 0)   # Available supply
        idxd = findall(d .!= 0)   # Unmet demand
        idx = argmin(C[idxs,idxd])
        # Min of avail supply and unmet demand
        f = min(s[idxs[idx[1]]],d[idxd[idx[2]]])
        F[idxs[idx[1]],idxd[idx[2]]] = f
        s[idxs[idx[1]]] -= f
        d[idxd[idx[2]]] -= f
    end
    TC = sum(C.*F)
    return (F=F, TC=TC)
end
```

15. The cost of each arc is shown in the network below. Use Dijkstra's procedure to determine the least-cost path from node 1 to node 6.



16. A single product is produced in two stages. Determine the total production and inventory cost over the three-month planning horizon. The forecasted demand is 20, 10, and 15 tons per month. Inventory costs are \$5 and \$25 per ton per month, respectively, and production costs are \$200 and \$800 per ton, respectively. Each month, 45, 0, 35 and 45, 0, 0 tons, respectively, are produced at each stage. Initially, no inventory is in storage at both stages, and 35 and 0 tons are in storage for stages 1 and 2, respectively, at the end of three months.

17. In a multi-period production-inventory flow network, why is it incorrect to include the cost of carrying both the initial and final inventory in the total inventory cost calculation?

18. Explain how it is possible to force an arc to be used in a shortest path determination.

19. Determine the relative and absolute gap associated with an ILP objective value of 26 and a best objective bound of 31.00.

20. For the ILP shown below, the following table represents the LP relaxation solutions found by adding branch constraints. Each row represents the solution found after adding the less-than-or-equal and greater-than-or-equal branch constraints on the two decision variables, with the right-hand-side value of each constraint listed.

(a) Determine a sequence of nodes for a branch-and-bound tree that solves the ILP, listing, for each node, its predecessor node in the tree and the objective of the LP relaxation (XO), ILP objective value (OV), and the best objective bound (OB). (Note: not all rows need to be included in the tree.)

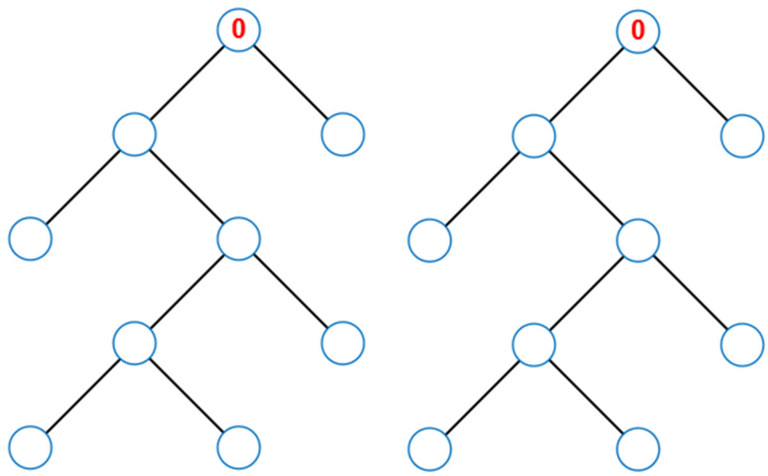
(b) Determine what the solution would be (i.e., the ILP objective and decision variable values) if the brand-and-bound procedure stopped after its relative gap (Gap) was within 10%.

$$\begin{aligned} \text{Maximize} \quad & 6x_1 + 8x_2 \\ \text{subject to} \quad & 2x_1 + 3x_2 \le 11 \\ & 2x_1 \le 7 \\ & x_1, x_2 \ge 0, \text{integer} \end{aligned}$$

	Node	Pred	$x_1 \le$	$x_1 \ge$	$x_2 \le$	$x_2 \ge$	x_1	x_2	*XO*	*OV*	*OB*	*Gap*
1			2			2	2.00	2.33	30.67			
2			3		1		3.00	1.00	26.00			
3			2			3	1.00	3.00	30.00			
4			1				1.00	3.00	30.00			
5				4					Infeas			
6						3	1.00	3.00	30.00			
7			3	3		2			Infeas			
8							3.50	1.33	31.67			
9			3			2	2.50	2.00	31.00			
10				1			3.50	1.33	31.67			
11			2		2	2	2.00	2.00	28.00			
12			2	3					Infeas			
13			3				3.00	1.67	31.33			
14			1			3	1.00	3.00	30.00			
15			2		1		2.00	1.00	20.00			
16					3		3.50	1.33	31.67			

21. Given the LP feasible solution to an ILP of $x_1 = 3.62$, $x_2 = 4.00$, and $x_3 = 7.12$, determine, as part of a branch and bound procedure, which variable would be selected using (a) the most infeasible branching rule and (b) the least infeasible branching rule.

22. In the branch and bound graphs below, the top node is labeled 0 and corresponds to the initial relaxation. (a) In the left graph, find a labeling of the remaining nodes assuming that node selection is done depth-first. (b) In the right graph, find a labeling of the remaining nodes assuming that node selection is done breadth-first.



23. The decision variables x_1, x_2, and x_3 represent whether activities 1–3, respectively, occur in a BIP, with 1 indicating occurrence.

(a) Add a constraint to the BIP to ensure that activity 1 occurs only when activity 2 or 3 occurs.

(b) Add a constraint to ensure the occurrence of at least one of the activities.

24. The decision variable x represents the amount of a bulk commodity in a MILP. Add any additional variables and constraints to the MILP to ensure that, if any of the commodity is used, the amount must be between five and twenty.

25. A single product is produced in a two-stage production process. Stage one has a capacity to produce 40 tons of the product per month, and there is a fixed cost of \$6000 per month and a variable cost of \$200 for each ton of product produced. For stage two, capacity is 80 tons per month, and there is a fixed cost of \$18,000 and a variable cost of \$800. The fixed costs are only incurred when any amount of production occurs at that stage for that month. Assuming that the annual inventory carrying rate is 0.4 \$/\$-yr, determine the inventory cost for each stage.

26. Explain how an implementation of the following mathematical programming formulation was used in the solution of one of the questions in HW 5.

$$
\begin{aligned}
\text{Minimize} \quad & \sum_{k \in N} y_k \\
\text{subject to} \quad & \sum_{k \in N} x_{ik} = 1, && i \in M \\
& x_{ik} + x_{jk} \le 1, && (i, j) \in W; k \in N \\
& x_{ik} \le y_k, && i \in M, k \in N \\
& y_i \in \{0, 1\}, && i \in M \\
& x_{ij} \in \{0, 1\}, && i \in M; j \in M
\end{aligned}
$$

27. Explain how the use of a Gomory cut can speed up the solution of an ILP.

28. Why does the per-unit cost of carrying inventory usually increase at each stage in a multi-stage production-inventory flow network while it remains constant across each period in a stage?

29. In the solution for Ex 1 of MIP-4,

    (a) Why was production cost separated into variable and fixed components in the model?

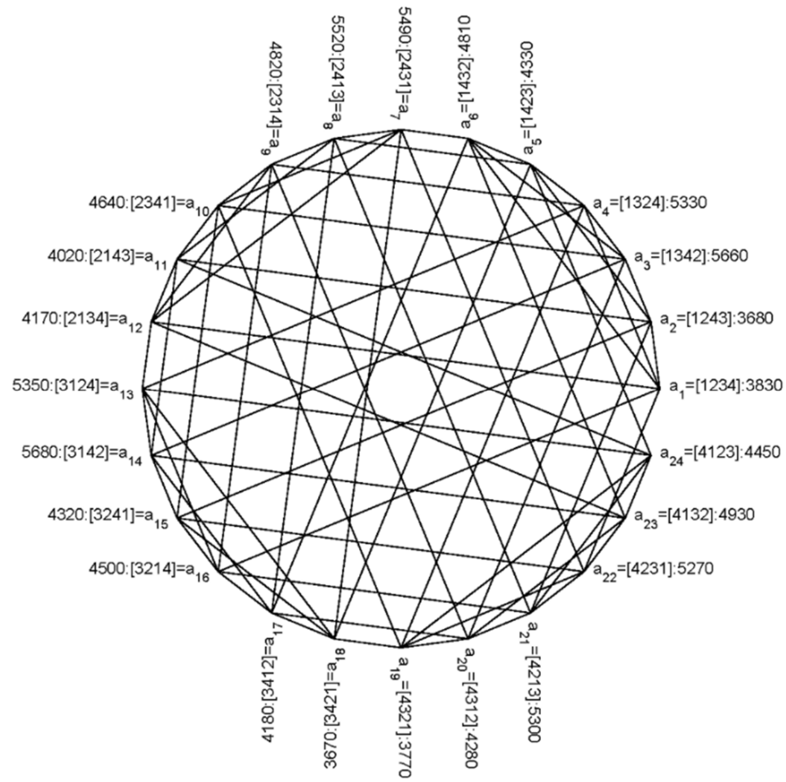
    (b) How was the fact that only 30 tons of product can be stored at each stage included in the model?

30. In the solution for Ex 2 of MIP-4, why was production cost not included in the objective function?

31. A couple is renovating three different rental properties that have just become vacant. They would like to finish the renovation as soon as possible and have received quotes for the work at each property to be completed next week from three different contractors. The first contractor has quoted 4, 2, and 5 thousand dollars for properties 1–3, respectively; the second contractor, 3, 4, and 4 thousand dollars; and the third contractor, 5, 5, and 8 thousand dollars. Determine which contractor should work on which property.

32. Using the *firstfit* procedure, assign samples of sizes 3, 6, 2, 5, 4, 4, 2, 4, 4, 2, 4, and 5 to containment devices in the given sequence. Each device has a size capacity of eight.

33. Starting with an initial assignment array of [2, 3, 1, 4], use the following pairwise interchange graph to determine the final assignment that would be found by the *SDPI* procedure:



34. Estimate the expected total cost associated with randomly assigning twenty resources to sites in a layout problem if the cost c_{ijkl} of assigning resource i to site k if resource j is assigned to site l is uniformly distributed between 0 and 1.

35. A hospital has four identical operating rooms available. Three different types of operations will be performed that each takes 50, 30, and 150 minutes, respectively, on average, and there are ten, fifteen, and six operations of each type, respectively, planned for tomorrow. Determine a lower bound on the maximum number of hours that it will take to complete all of the operations, assuming that it takes fifteen minutes to switch between each operation in the operating room no matter the type of operation or the type of the preceding operation.

36. Given two identical resources and four independent tasks with times of 1, 2, 3, and 4, respectively, that have initially been allocated to resources 1, 1, 2, and 2, respectively, explain why, in this instance, the new allocation that is found by using the *relocate* procedure minimizes the maximum time required to complete all of the tasks.

37. Determine the cube per order for the following item master and order dataset inline CSV files, where any missing SKU data should be dropped and any missing QTY data should be imputed with a representative value:

```
SKU,LENGTH,WIDTH,DEPTH,CUBE,WEIGHT,UOM
1,5,3,2,30,1.25,EA
2,3,2,4,24,4.75,EA
3,8,6,5,180,9.65,CA
4,4,4,3,32,6.35,EA
```

```
ORDER,SKU,QTY
1,1,
1,2,3
1,3,2
1,4,6
2,1,4
2,3,1
3,1,2
3,,
4,2,2
5,,1
```

38. Demand is being estimated by a model that minimizes the *L1* loss function. Given the demand data in the array `d` given below along with the following partition of the data into training and test sets, determine the root mean squared error of the model:

```
using MLJ
d = [237, 214, 161, 146, 331, 159, 423, 332, 139, 327, 152, 98, 116]
train, test = partition(eachindex(d), 0.7, shuffle=true, rng=123)
```

```
([1, 11, 6, 12, 7, 4, 13, 8, 9], [10, 2, 3, 5])
```

39. Five email messages are available to train a spam filter. Two features have been identified as being the most useful for detecting spam. Two of the emails are spam and have values of 4 and 6 for feature one and 8 and 3 for feature two, respectively. The three other messages are not spam and have values of 1, 5, and 8 and 1, 6, and 4 for features one and two, respectively. Determine the information gain associated with using 7 as the first threshold for feature two using the *CART* procedure in creating a decision tree for spam detection.

40. With respect to Ex 2 of ML-3, the table contains the feature values for three passengers. Use the decision tree to determine the predicted probability that each passenger would survive.

Row	Pclass Int64	Sex Int64	Age Float64	Sibsp Int64	Parch Int64	Fare Float64
1	3	2	21.0	0	0	7.8292
2	3	1	8.0	8	2	69.55
3	3	1	18.0	0	1	14.4542

```
Feature 2, Threshold 1.5
L-> Feature 1, Threshold 2.5
    L-> Feature 6, Threshold 29.35625
        L-> Feature 3, Threshold 22.5
            L-> 2 : 12/12
            R-> 2 : 26/33
        R-> 2 : 61/61
    R-> Feature 6, Threshold 22.737499999999997
        L-> Feature 4, Threshold 0.5
            L-> 2 : 32/47
            R-> 1 : 11/20
        R-> Feature 6, Threshold 31.331249999999997
            L-> 1 : 9/9
            R-> 1 : 4/6
R-> Feature 6, Threshold 26.26875
    L-> Feature 3, Threshold 11.5
        L-> Feature 4, Threshold 2.0
            L-> 2 : 7/7
            R-> 1 : 1/1
        R-> Feature 3, Threshold 32.5
            L-> 1 : 130/150
            R-> 1 : 65/67
    R-> Feature 4, Threshold 2.5
        L-> Feature 3, Threshold 36.5
            L-> 2 : 20/33
            R-> 1 : 28/38
        R-> Feature 3, Threshold 3.5
            L-> 1 : 4/5
            R-> 1 : 8/8
```

41. The following are descriptions of a problem. Based on the models and solution procedures discussed in class, recommend both a model and solution procedure that would be best suited for solving the problem. Include a justification for both your model and solution procedure recommendations.

    a. *Problem*: A single generator will be used to power three different refrigeration units located on an industrial site. Where should the generator be placed so that the cost of running power lines to the units is minimized?

       Model:

       Solution procedure:

    b. *Problem*: A student would like to minimize their weekly expenditure for food while still maintaining a healthy diet. They have determined the recommended nutritional requirements along with the cost and nutritional information of several of their favorite foods. How should they determine what foods to purchase each week?

       Model:

       Solution procedure:

    c. *Problem*: A developer would like to create a software package to detect 4000 known computer viruses using 8000 substrings of 20 or more consecutive bytes from the viruses that are not found in normal code and can be used to detect viruses in code. In order to operate efficiently, only a small number of substrings, much less than 8000, can be used.

       Model:

       Solution procedure:

d. *Problem*: Boxes with different heights are stacked floor-to-ceiling in a room, and the area required for storage should be minimized.

   Model:

   Solution procedure:

e. *Problem*: NASA would like to determine the scaled sound pressure level, in decibels, of different size airfoils at various wind tunnel speeds and angles of attack.

   Model:

   Solution procedure:

f. *Problem*: Several different commodities are available for purchase. There is a limited capacity available for shipping the commodities, and there are limited funds available for the purchase of the commodities. Subject to a minimum order quantity, any amount of each commodity can be purchased.

   Model:

   Solution procedure: