Research Assistant Work

JuMP:

Introduction - Read

Read lightly over "Getting started with sets and indexing"

- Sets are often used to represent some variable in a sample with multiple observations

- Unordered sets can be used for representing non-numeric vectors, such as Strings. They are most commonly initialized through a standard vector. Other methods include using a dictionary (each key is associated with some value), and using Julia's Set functionality, which allows you to iterate through elements of a vector/dictionary and assign the result to a new set. This method will also remove duplicate elements

- Numbered sets are used for numeric variables. These are often initialized using ranges to automatically create a vector that includes each element within the range. You can use the syntax [start:step:stop] to initialize a vector that goes from start to stop in steps the size of "step".

- You can also use Tuples as sets through [(tuple\_name)] or iterating through tuple combinations. For multi-dimensional sets, you can construct containers using x[tuple\_1,tuple\_2], where tuple\_1 is the first dimension and tuple\_2 is the second dimension.

- Set operations include setdiff(x,y) which returns all elements that are not common between x and y. Another operation is union(x,y) which combines the elements of x and y vectors. Length(x) returns the number of elements in a vector.

- You can test for set membership using the "in" keyword, and combine with "for" and "if" keywords in expressions to filter and create new sets.

- For compound sets, where each element includes a function(s) of other elements, such as distance between cities, we can use a matrix to represent the set. We can set model objectives using some function of the variables. We can also use containers as indices by converting tuples to index values.

- Data conversion to a more intuitive form is useful.

Tutorial - Linear Programming

- Some (linear) optimization function according to linear constraints.

- Mixed-integer Linear Programs (MILPs) enforce some level of integrality where variables take discrete values.

- Knapsack problem: Maximize some value (c) associated with each state of i (objective function can be considered nutritional value, e.g.) according to some constraint (cost) on the number of i's that can be used through the variable w (cost per item). x is an indicator function for whether each i is activated (purchased). Can be written as a function of the cost and value vectors multiplied by a transposed "x" vector, with the summed elements of the resulting vectors corresponding to the objective and constraint functions. x is to be determined by the model here as the optimal set of decisions x for each i.

- Syntax: @variable to determine decision variables, @constraint to determine constraint function (cost), @objective to determine objective function. Use function optimize!(model) to solve program and solution\_summary(model) to summarize the results. @assert sets important variables for the model to run. Bin means a binary variable.

- Once you have created your model, it is a good idea to build a function around it that takes vectors/variables as inputs and computes the results, returning the decision variables outputted by the solver.

Time: 4 Hours

Week 2:

Diet Problem

* Objective function: Minimize cost based on some quantity of food >=0
* Constraint: Total amount of each macronutrient across all foods must be within the acceptable range. This is done for each macronutrient and summed across all foods
* Data: Create dataframe with each food, including cost and macronutrient profile (per quantity)
  + Then create dataframe with minimum and maximum bounds for each macronutrient
  + Set decision variable >=0, with the number of variables equal to the number of foods
  + We then set the objective function (min cost -> min sum(food\_cost \* x) over all food items f)
  + We then define our constraints, (min macro < sum(x\*macronutrient per unit) over all foods f < max macro, with separate constraints for each type of macronutrient)
  + Can code easily using row in eachrow(table), and then use each row for individual constraints, rather than manually creating each constraint
  + You can add or delete constraints, and use set\_integer to force a variable to take an integer

Cannery Problem

* Moving products from production plants “p” to endmarkets “m”. Demand for each endmarket is denoted “d(m)”, while the capacity for each production facility is denoted “c(p)”. The shipping cost of transferring a unit from a production facility to a market (can also be thought of as distance) is denoted “d(p,m)”. Objective is to minimize total cost (d(m,p)\*x(m,p) summed over M, summed over P).
* Constraints include production in each plant (or shipments from each plant) being less than or equal to capacity c(p), and shipments to each endmarket being at least equal to the demand in that market d(m)
* Data: In this example, data is loaded from JSON using JSON.parse function. Data provided includes markets with demand figures, production facilities with capacity figures, and distances for all combinations of production facilities and endmarkets
* Constructs a dictionary\*\* Would like to review this syntax, nested dictionary?\*\*
* Function to compute distance from plant to market \*\*want to review this function syntax\*\*
* Decision variables are initialized for all combinations of plants and markets (matrix)
* Constraint that each plant cannot ship more than capacity \*\*syntax\*\*
* Constraint that each market must receive at least its demand
* Objective is set as minimizing transportation distance, or minimize the product of units delivered and distance
  + Would likely (?) be best to minimize based on batches (or units/flight), instead of individual units, although this doesn’t seem to be an issue with the final answer

Factory Schedule Example

* Assumes we are optimizing production of some good from a set of factories “f” over the course of 12 months “m”.
* Each unit of production carries a variable cost “c(f)” per unit (depends on factory), and fixed cost “a(f)” per month.
* Z(m,f) is an indicator variable for whether the factory runs or shuts down. Where z = 0 (factory shut down), no fixed or variable costs are incurred
* Minimize the total cost, or
* Where monthly demand “d(m)” is exactly satisfied each month. In other words, production across all factories is equal to demand in each month.
* However, this model may be flawed, as a sufficiently high demand against maximum production capacities may lead to infeasibility. We can counter this by adding a new variable to account for unmet demand (purchase from other suppliers or running overtime), with an arbitrarily large cost factor
* Data: File as .txt, read into Julia as CSV and convert to dataframe type using CSV.read()
* Validating data can be a useful practice, as it helps avoid typos that lead to flawed figures. In this example, data is screened to ensure that minimum production does not exceed maximum production in any instances, and that variables are not less than 0.
* For sensitivity analysis, we analyze how the optimal objective value changes as variable costs change
* In this example, variable costs are scaled at factors between 0 and 1.5

Transportation Problem

* Set of factories and retail stores, aim is to minimize cost. Factory set is defined as a vector of instances “i” (origins), and retail stores are represented by a vector of instances “j” (destinations)
* Maximum supply (capacity) at each factory is “s(i)”, and demand from each store is “d(j)”
* Decision variable is x(i,j), representing the number of units from “i” to “j”, with per-unit shipping cost c(i,j)
* Minimize total cost, or cost\*units for each factory-endmarket pair

Multi-commodity Flow

* The multi-commodity problem is an extension to the transportation problem with multiple types of products. We add the extra parameter, “k” (instance of vector of product types) to existing constraints, and create a new constraint that limits the total number of units (across products) that can be transported between i and j pairs

Network Multi-commodity Flow Problem

* Similar to the multi-commodity flow problem, except that instead of having factories and endmarkets as distinct nodes, each node has both a supply capacity and demand level (although either can also be zero). Each node is represented by some instance “i” in vector N. Instances (i,j) in E represents the connections between all nodes, with shipment cost c(i,j,p), where “p” represents different commodity (or product) types
* Supply capacity for each node “u(I,p) and shipping capacity “u(I,j) between nodes cannot be exceeded
* Deicision variables are “x(i,j,p)”, or the quantity of product “p” being transported from “i” to “j”, and s(i,p), or the production quantity of each factory s(i,p)
* Total demand must be met by the sum of supply and net inflow

Tips and Tricks

* Absolute Value function, where t>=|x|
  + Option 1: Create two linear inequality constraints using both x and negative x. Therefore, regardless of x’s sign, the absolute value will be tested
  + Option 2: Use two non-negative variables and create expressions t+x and t-x. Can be used to test t>=-x (negative x) and t>=x (positive x)
  + Option 3: Using MOI.NormOneCone function
* L1-norm (Manhattan distance/vector magnitude sum)
  + To calculate the minimum Manhattan distance, use MOI.NormOneCone function
* Infinity-norm (Largest magnitude among each element of a vector, regardless of direction)
  + To calculate the smallest infinity-norm in a set of vectors \*\*not sure of this\*\*, use MOI.NormInfinityCone function
* Max
  + To model t>=max{x,y}, create constraints for both t>=x and t>=y and condition on both of them being true
* Min
  + To model t<=min{y,x}, use the same approach as above except with t<=y,x conditions and require both to be true
* Modulo (common divisor and remainder -> 5%2 = 2 (remainder 1)
  + Set constraint where y is between 0 and n-1 (where y is the remainder)
  + And set a constraint where z is any number greater than or equal to 0 (common divisor)
  + Set x greater than or equal to 0 (dividend)
  + Set x == n\*z + y, or ensure that x is equal to the remainder plus the divisor\*modulo quotient
* Boolean Operators
  + Or (x3=x1 OR x2, where x is binary)
    - Create 3 element binary vector (0 or 1), and set a constraint where x[3] must be <= x[1], x[2] <= x[3], and x[3] <= x[1] + x[2]
  + And (x3 = x1 AND x2), where x is binary
    - Create 3 element binary vector where x[3] <= x[1], x[3] <= x[2], and x[3] >= x[1] + x[2] -1
  + Not (x1 not x2)
    - Set constraint where x[1] + x[2] = 1, where x is a binary vector
  + Implies (x1, therefore x2)
    - Set x[1] <= x[2], such that if x[1] is 0, then x[2] can be 1 or 0. If x[1] is 1, then x[2] must be 1
  + Disjunctions (“or”)
    - Use “big-M” multiplied by a binary variable to relax one of the constraints
    - Introduce binary variable y that takes values 1 or 0, activating “big M” (large constant) to make one of the expressions true, and only testing the other expression
  + Indicator constraints, where an indicator activates a constraint
    - To model x1+x2 <= 1 if z=1, use “-->” operator, where z is a binary variable and x is a 2 element vector in the function z-->{sum(x) <=1})
    - For modelling the same constraint when z=0, use !z --> {sum(x)<=1})
    - If the --> operator does not work, use big M
  + Semi-continuous Variables
    - Can be modelled using the semicontinuous() function, or using a binary reformulation
  + Semi-integer Variables
    - Can use the semiinteger() function, or using binary reformulation with x being initialized as an integer type
  + Special Ordered Sets of Type 1 (set of variable which at most 1 can take on a non-zero value)
    - Use MathOptInterface.SOS1
  + Special Ordered Sets of Type 2 (same as type 1, except two elements can be non-zero but they must be consecutive)
    - Use MathOptInterface.SOS2
  + Piecewise Linear Approximations
    - SOSII constraints are most often used
    - We create a function with a set of x-values and y-values
    - Set N equal to the number of points you have
    - Model in constraint using combination of lambda, y-hat and x-hat \*\*unsure how this calculation works\*\*

Time: 8.5 hours

Week 3:

Approximating Non-linear Functions

* Approximating non-linear functions with a mixed-integer linear program
* If a function that you are approximating is convex, you can minimize down onto it using an outer approximation
  + We can use the following inequality due to concave-up nature of function:
    - F(x) >= f(x1) + df(x1)(x-x1)
    - As we add more planes, the error between the true function and the piecewise linear approximation decreases
* If we want to approximate a concave (downward acceleration) function, we can maximize “up” to it using the following inequality:
  + f(x) <= f(x1) + df(x1)(x-x1)
  + Example is 1/x
  + Adding more points of x1 leads to more accurate approximation
* You can also use an inner approximation method for non-linear functions. \*\* Model formulation, not sure how lambda works\*\*
  + \*\*Essentially plots points across the function and connects them linearly, then sets minimum/maximum for the function\*\*
  + Can be done for both concave and convex
* Piecewise linear approximation
  + If a model is non-convex or non-concave (think: sin/cos), then traditional inner/outer approximation becomes infeasible. Outer breaks down where concave up switches to concave down, and inner does not stay near the line (?)
  + We can force inner approximation to stay on the linear lines by using SOS2, that ensures at most two elements of lambda can be non-zero, and if they are, they must be adjacent

Facility Location Problem

* Uncapacitated facility location
  + We are given a set of clients M = {1,…,m} and a set of sites N= {1,…,n} where a facility can be built
  + Decision variables are split into 2 categories:
    - Binary variable y(j) indicates whether factory “j” is built or not
    - Binary variable x(i,j) indicates whether client “i” is assigned to facility “j”
    - Objective: Minimize the total cost of serving all clients. This cost breaks down into 2 components:
      * Fixed cost of building a facility f(j)
      * Cost of serving clients from assigned facility (in this case, the Euclidean distance between the two) c(i,j)
    - Constraints:
      * Each customer must be served by exactly one facility
      * A facility cannot serve any client unless it is open
    - MILP Formation
      * Minimize total cost, or c(i,j)\*x(i,j) for all factory-client combinations + f(j)\*y(j) for all factory locations
      * Ensure that each client can only be served once (each client is served by exactly one facility)
  + Capacitated facility location
    - Introduces capacity constraints, where each client “i” has some demand a(i) >= 0, and each facility “j” has a finite capacity q(j) which cannot be exceeded
    - Additional constraint ensures that the sum of demand from each customer (0 unless x(i,j) = 1) for a facility must be less than or equal to the capacity of the built facility

Financial modelling problems

* Cashflow Requirements
  + Given a set of cash flow requirements (mandatory payments) for a company, and a set of capital sources, we can create a linear model to forecast how we can use each source of capital to meet our requirements and maximize our total money by some point in time
* Combinatorial Auctions
  + The set of items available in the auction is M = {1,2,…,m}, and a bid is defined as variable B(j) = (S(j), p(j)) where S(j) is a set of items from M that are being bid on, and p(j) is the price offer for the set
  + Suppose auctioneer receives “n” bids, the goal is to maximize revenue from all bids
  + Objective function is to maximize revenue, or the sum of offer price from all accepted bids. You can use an indicator function y(j) for whether you accept a bid
  + Subject to only being able to accept one bid per item

Geographical Clustering

* + Goal is to cluster a set of “n” cities into “k” groups that minimizes the total pairwise distance between cities and ensuring that the variance in the total population is relatively small between cities in the same group
  + Starts with a function to compute distance (haversine)
  + Creates lower triangular matrix by calling distance function on each city \*\*why a triangular matrix?\*\*
  + Building the model:
    - Add constraint: Each city must be part of exactly one group
    - Set a maximum difference between group population and target population. When this constraint is relaxed, your cities will be closer together but the variance in populations will be higher
    - We compute the total distance between cities in a group by iterating over all cities using an indicator function z(i,j) for cities within the same group
    - Objective is to minimize dot product of z with the distance matrix

Network Flow Problems

* A flow network is a directed graph with nodes and edges, where each edge has some capacity and receives a flow. Flow cannot exceed capacity
* Shortest Path Problem
  + Suppose that each arc (i,j) is assigned some scalar cost a(i,j). We want to find a path through all the nodes that minimizes cost
  + Minimize, for all pairs of (i,j) the cost of the pair times an indicator function for whether the pair is activated
  + \*\*Don’t understand constraint including purpose of “b” and what the variable “k” represents\*\*
* Assignment Problem
  + There are n people and n objects, where “i” represents each person and “j” represents each object. There is some benefit a(i,j) of matching person i with object j. We want to maximize this benefit
  + Additional constraint is that person i can only be assigned to object j if the pair (i,j) belongs to a given set of pairs in A
  + Want to find the set of person-object pairs that maximizes total benefit
  + Each object can only be assigned once, and \*\*each person (object) must be assigned to one object (person) \*\*
* Max-flow Problem
  + We have a graph with two special nodes, a sink (that only consumes) and a source (that only produces)
  + Objective is to move as much flow as possible from source to sink

Multi-objective Knapsack

* A multi-objective problem has multiple objective functions
* We can modify our original knapsack problem by adding an additional objective. Rather than just maximizing our total profit, we can also look to maximize another function: Total desirability rating
* \*\*How do these functions work? What is the behavior in a scenario with tradeoffs? Different forms?\*\*
* Multi-objective function models provide a range of solutions to illustrate trade-offs b between functions. Best way is to plot the decision variables for each solution. This allows the decision maker to choose a solution that fits their personal preferences

Sudoku

* A popular number puzzle. Rules:
  + The numbers 1-9 must appear in each 3x3 square
  + The numbers 1-9 must appear in each row
  + The numbers 1-9 must appear in each column
  + Objective is to satisfy these rules, in other words not a maximization or minimization problem
  + We can model this problem where all decision variables are binary, or 0-1 integer programming
  + We will initialize a binary variable as an indicator for each possible number in each possible cell occurring. This takes the form x[i,j,k], as an indicator for whether the value k is present at row i and column j.
  + We will then create a constraint that each column and row must contain one of each character
  + We can then create a constraint that each 3x3 square must include each of the 9 digits

N-Queens Problem

* Involves placing a number of queens on an N\*N chessboard such that none of the queens can attack each other
* We can create a variable that represents a matrix version of the chessboard, where each value takes 1 if a queen sits on the board and 0 otherwise
* We can initialize the following constraints to make the model work:
  + There must be exactly one queen in each row/column
  + There can only be one queen on any given diagonal

Constraint Programming

* JuMP supports a range of constraint-programming type constraints through sets in MathOptInterface (MOI)
* Due to the ability to convert models to MIPs, variables can only take on integer forms \*\*is this correct?\*\*
* MOI applications for constraints
  + MOI.AllDifferent set ensures that every element in a list takes a different integer value
  + MOI.BinPacking set can be useful for placing a set of items into different bins given some capacity for each bin
  + MOI.Circuit set is used for constructing a tour of a list of N variables. (From the textbook) “They will each be assigned an integer from 1 to N that describes the successor to each variable in the list” \*\*I don’t understand this. What is the successor?\*\*
  + MOI.CountAtLeast is used to test whether n elements of a set of variables belong to some set of values. For example, we want to ensure that at least one of x[1] and x[2] in our model takes on the value of 3
    - We can use a partition for testing multiple sets of variables simultaneously. The partition is a vector with the number of elements per variable set (or the number of variables per set)
    - Then we can define the value set that the variable elements must be a part of
    - And the number, n, of elements per variable set that must be in the value set
  + MOI.CountBelongs is used to count how many elements in a set of elements belong to some set of values
  + MOI.CountDistinct set is used to count the number of distinct elements in a set of variables
  + MOI.CountGreaterThan is used to limit the number of variables in a set that have a value equal to another variable
    - We compute some value n that is strictly greater than the number of times that variable set y appears in x
    - \*\*Don’t understand the application of this\*\*
  + MOI.Table is used to select a single row in a matrix / 2d list
  + \*\*How do you select the row #?\*\*

Callbacks

* \*\*Don’t understand callbacks\*\*

Sensitivity Analysis of an LP

* lp\_sensitivity\_report function can create a sensitivity report of an LP similar to excel solver
* Sensitivity analysis of an LP asks 2 questions
  + How much can the objective coefficients change before a different solution becomes optimal
  + How much can the right-hand side of a constraint change before the optimal solution changes
  + Sensitivity Report object
    - Provides a tuple for each variable in the objective function with lower and upper bounds that illustrate how much the objective coefficient can change before the optimal solution changes
    - A tuple for each constraint that explains how much the RHS can change before the optimal solution changes
    - We can use a function to create a dataframe for each constraint or variable with upper and lower bounds, reduced cost, and other columns

Week 3 Hours: 10

Non-Linear Programs (NLP)

* NLPs are a class of optimization problems where at least one constraint or objective function is nonlinear
* Mixed-integer nonlinear programs (MINLPs) enforce integrality for non-linear programs, similar to LPs
* For choosing a solver, use one that begins with NLP. Solvers that start with MI support integrality

Simple Examples

* For choosing a solver, use one that begins with NLP. Solvers that start with MI support integrality
* Rosenbrock Function
  + Define variables “x” and “y”
  + Set objective function to minimize a non-linear function of x and y (where random variables are multiplied by each other)
  + \*\*@Test keyword\*\*
* Clnlbeam
  + Initialize constants h, N, and alpha
  + initialize variables “t”, “x”, “u” that are vectors with N+1 elements (index i from 1 to 1001
  + Set constraints as functions of variables and constants
* Maximum Likelihood Estimation
  + “randn(x)” function produces a vector of length x with numbers sampled from a standard normal distribution
  + \*\*what does setting a start value do, particularly if a variable is not a list/only has a single element\*\*
* Quadratically Constrained Programs
  + Set variables x, y, and z (atomic)
  + Create sets of constraints that include variable addition and multiplication
  + Optimize
  + Print objective value and the values of “x” and “y”
  + Test that the model is feasible and locally solved \*\*what does locally solved mean in this context?\*\*

Rocket Control

* Objective is to maximize the final altitude of a rocket launched vertically, where “final” is the highest point of the rocket
* We must take into account the mass, fuel consumption, gravity (a function of altitude), and aerodynamic drag D(function of altitude and velocity) in each scenario
  + We use these variables to compute the measures of rate of ascent (velocity), acceleration (velocity derivative), and rate of mass loss (a function of thrust leading to fuel loss and lower mass)
  + These equations control the dynamics of the rocket
* We set variables required for dynamically computing the three measures over time and solving the model
  + We create initial constants that include constants for starting velocity, mass, gravity, and height, alongside final mass, the mass loss constant, the total amount of time measured in the model, and the number of steps (frames) where variables are computed. We can then calculate the time per step (total time/# of steps). We also set the thrust-to-fuel mass, drag scaling, and maximum thrust constants
  + Using these constants, we can initialize our variables velocity, height, mass, and thrust as vectors, such that each element must be greater than or equal to 0 and start at the initial value constants. The number of elements for each variable is the number of steps defined as a constant. Mass at any point must be greater than final mass (\*\*final mass assumes all fuel has been used\*\*?)
  + We can use the fix() function to force the first elements of variables to be equal to the starting (initial) values defined for each variable.
  + Objective is to maximize the height of the rocket at the final time period
  + Drag and gravity are defined as functions, with drag a function of height and velocity, and gravity a function of height
  + We set constraints to ensure that the variables are abide by the laws of physics \*\*what does ddt(x::Vector, t::Int) do?\*\*
  + We can create a function to plot our variables over time by defining x as seconds elapsed at each frame (deltaT(or seconds per frame)\* # of frames elapsed). We can then call our plotting function and set the y-variable to each of our variables.
  + \*\*This model only looks at the first 0.2 seconds of a rocket launch. But how would it change with different numbers of frames or a different duration (1 second)? Would like to test this model in Julia to understand it better\*\*

Portfolio Optimization

* Say that we will be investing x dollars in 3 stocks, so x is defined as x(i=1,2,3) where x is money invested in i. We want to earn at least $50 dollars of profit from our collective investments, or 5% (assuming we start with $1000). In earning our $50 of profit, we also want to minimize our risk.
* Risk is measured as the variance of portfolio returns, and can be simplified as the sum of the product of weights ($ invested in this case) for each security combination times the covariance of returns for that combination
* This simplifies to the transpose of the weight vector times the covariance matrix times the weight vector
* Understanding these rules, we can define our objective as minimizing the total variance of the portfolio returns. We set constraints, such as total return(or profit) being at least 50 and the total amount invested being less than or equal to 1000. You cannot invest negative amounts (x must be greater than 0 for each stock)

Week 5: 4 Hours

Optimal control for a Space Shuttle reentry trajectory:

* We aim to determine the trajectory of a space shuttle’s reentry using a set of equations to solve for variables such as height, velocity, angle, etc.
* We set constant values for some variables in our calculation, for example the force of gravity
* We set initial values for some of our dynamic variables, such as the height of the aircraft
* We set some values for the final point of the reentry, which indicate when the shuttle has completed its reentry
* The objective is to maximize latitude of the vehicle (or cross-range)
* We can use a set of discretized points, n, with time step size delta\_t, which we can choose and define as a variable (done in seconds, in this case)
* Begin by setting our constants, which include:
  + Global variables (remain the same regardless of dynamic variables)
  + Aerodynamic and atmospheric forces on the vehicle (
  + Initial conditions at time = 0 (or n=1)
  + Final conditions that mark the reentry complete (set for our final time point, n)
  + Our mesh points (\*\*) and integration scheme
* We can set our decision (or dynamic) variables using the @variable function
* We can then use expressions to define the mathematical relationships between variables \*\*why are these not defined as constraints but as expressions? Do they have the same effect in this case?\*\*
* We can create our own method of integration using brute-force rectangular or trapezoidal integration constraints
* We can then set our objective function to maximize our cross-range, or longitude
* We can then plot various decision variables over time

Tips and Tricks

* One useful functionality is to define functions that allow multiple outputs (a single vector output)
  + This may help with efficiency, particularly if the variables are related and share (computationally) expensive calculations to get there
  + You can also use memoization \*\*Can you explain how memoization works?\*\*
  + \*\*What does @operator do, seems like it integrates an existing function into a model. How is this different from @expression? Why is it useful?\*\*

User-defined Hessians

* + Hessian matrix contains all of the second partial derivatives of a multi-variate function. In other words, the second derivative of a function with respect to its variables in each possible combination (with respect to x, y, y then x, and x then y
  + We can define the hessian matrix of the rosenbrock function
  + Only need to fill in one lower-rectangular portion of a hessian matrix because it does not matter in what order you take partial derivatives – the matrix will ultimately be symmetrical
  + For operator functions, we can add additional functions to the operator equal to the first-order derivative (gradient vector for multi-variate) and second-order derivative (hessian matrix for multi-variate)

Nested Optimization Problems

* + We can look at nested optimization problems, where our ultimate optimization problem depends on the results of an initial optimization
  + Here, we can use an example of a minimization function subject to a constraint where some variable is the maximum of another function over a variable exclusively in the lower-level (sub) function
  + We can create a container function for our lower-level function that takes fixed variables as arguments (those that are not involved as sub-problem decision variables) and computes the value of the sub-problem using our decision variables y
  + However, we can also represent the subproblem optimization as a @operator function with the first order being our sub-problem optimization function, our second order being the gradient vector of the function, and our third order being the hessian matrix with respect to our objective variables (decision variables in the objective function)
    - The reason for only providing derivatives with respect to the x’s (obj function decision variables) is that we are maximizing our sub-problem optimization with respect to y (this will set our decision variables for y such that the objective is maximized absolutely), but our sub-problem value will depend on our x-values in the problem, so we want to maximize our overall objective function as a function of the x’s with respect to the sub-problem and with respect to the overall problem. For this reason, we look at how the sub-problem changes to with respect to x, such that we can determine the overall maximum of the objective function.
  + \*\*Why do we need to define our second and third-order functions with respect to the x’s (the objective function decision variables)?, What is the point of providing the first and second order derivatives, anyway if you’re going to brute force the calculation? In other words, do solvers work through fast brute force, or do they exercise calculus (derivatives) to determine maximum, minimum, etc. points for these models?\*\*
  + To improve performance, we can use a cache to avoid recalling the sub-problem’s order conditions

Week 7: 4 hours

Next meeting – Optimization Methods in Finance

<https://www.andrew.cmu.edu/user/gc0v/webpub/book.pdf>

Identify a question or so, and try to implement that model

Finding a question with a quadratic constraint

Optimization Methods in Finance Reading

An optimization can have a global and local solution, where the global solution is optimal for all feasible sets and the local solution is optimal for some subset of the feasible region that is some distance from the optimal point

\*\*Difference between global minimizer and strict global minimizer, same question for strict local vs. not strict local\*\*

Predictors of optimization model efficiency: Decision variables, number of constraints

Most successful ways of solving an LP are the interior-point and simplex methods

Quadratic programs:

Objective function has a degree of 2: 1/2 (x-transpose)\*Q(linear matrix)\*x + f(x)

Q is a linear symmetric and diagonal matrix that is defined as the linear scale of the x-variable in the quadratic term, (**0.2**x^2) and f(x) is the scale on the linear term.

The QP solver only understands a particular format (1/2 times quadratic term, <= and not >= for inequality constraints. \*\*Why is this?\*\*

Week 10: 6 hours

\*\*Dual of a ML program – can you explain what this is and what it represents?\*\*

Finance Applied Quadratic Optimization Models

Mean-Variance Optimization:

* + Takes into account the trade-off between expected return and risk of a portfolio
  + Consider assets (S1, S2… Sn) with randomly distributed returns at each period. We can define μ(i) as the expected return (conditional mean) of an asset’s returns and σ as the expected risk of the investment (conditional standard deviation of returns)
  + For i and j within S (Si, Sj), ρ is the correlation coefficient between the returns of the two assets. We can represent the covariance matrix of asset returns as Σ, which is a symmetric matrix
  + We can view μ as a vector of a securities’ expected returns (one-dimensional vector) and x as a vector of weight of the portfolio in each security. We can derive the expected return of the portfolio as the weighted average of asset expected return by weight. We can also represent this as μ-transpose \* x
  + The variance of a portfolio is denoted as the sum, over each (i,j) combination of assets, of the covariance multiplied by product of weight in each asset. This can also be represented by the formula x-transpose \* covariance matrix \* x

Week 11: 4 Hours

Mean Variance Optimization (MVO) issues:

* Often produce portfolios that lack diversification due to slight differences in mean returns and variance that are not realized in the market but lead the model to take advantage of arbitrage opportunities that do not exist
  + One way to counter this is to set a limit on the weight of each asset in the portfolio, or establish sector weights based on asset assignment
* Given the sensitivity of the model to the data, it is likely to produce drastically different weight outputs given small shifts in expected variance and returns. Therefore, if the model is being rerun to rebalance the portfolio, it is likely to result in large transaction costs

Week 12: 3 hours

Black-Litterman Model: Quadratic Optimization

The mean-variance optimization model takes our mean return expectations (usually an estimation of mean return per asset class historically) and expected variance/covariance of returns between asset classes to locate optimal portfolios. That said, this model has a few shortcomings, including the following:

1. The model does not incorporate the input of an investor’s own views relative to the market in a way that allows these views to adjust weights optimally.
2. The model is overly sensitive to small differences in expected return and risk, assigning large portions of the portfolio to take advantage of arbitrage opportunities that are the result of very small and likely trivial differences. It is typical for the model to construct under-diversified portfolios, as we saw with our own implementation.
3. Changes in asset returns and risk will likely lead to dramatic shifts in output weights. This results in overly-expensive rebalancing due to transaction costs of trading out of securities and into new ones.

One model that has been found to mitigate these concerns while providing an input for investor views that deviate from the market equilibrium is the Black-Litterman Model. I am not yet familiar with the logic that underlies this model or how exactly it alleviates concerns 1 and 2, but here is an outline of the model below:

1. Mean Returns: We assume that the return for any asset class can be divided into a market equilibrium return (what is expected by the market) and the investor’s own view of the return (where their views deviate from market consensus). We can calculate the mean equilibrium return using the Capital Asset Pricing Model (CAPM) and estimate the variance of the mean equilibrium return by taking the sampling distribution of the mean (in the textbook, they assume some “t”<1 that is multiplied by the variance of asset returns). Once we have determined our equilibrium risk and returns, we can begin building in our investor views:
   1. First, we predict out our excess returns against the equilibrium in a column vector. These represent our views and can be determined however we like.
   2. A black text on a white background

      Description automatically generatedA black and white symbol

      Description automatically generatedWe then create a diagonal matrix corresponding to our confidence in each view, where our confidence level is the variance of our noise vector with mean 0 and variance of the confidence. Given these elements, we incorporate them into an investor view matrix, “P” that fits into the model using the following equation:  
      Where “P” is our view matrix, “q” is our expected excess returns, and epsilon is our noise vector. We can then incorporate this to compute our mean return vector which corresponds to a multivariate normal distribution with mean mu-hat below:  
      Where tau\*Epsilon is our covariance matrix multiplied by a scalar <1, pi is our mean equilibrium return (vector?), and upsilon is our diagonal confidence matrix.
2. Once we have computed our new return vector, we can input it into the MVO model to compute the optimal portfolio.

Where “P” is our investor view matrix, mu is our mean return,

Week 13:

To do:

Clear up BL write-up

Access stock database for UBCTG & Plug to model

Explore efficiency

Week 14:

10 hours

Black-litterman research, understanding the structure of the model and its application for short-term trading

<https://web.stanford.edu/~boyd/papers/pdf/cvx_portfolio.pdf>

<https://community.wvu.edu/~krsubramani/courses/sp15/optfin/lecnotes/QPapps.pdf>

<https://www.daytrading.com/quadratic-programming>

<https://www.global.toshiba/ww/technology/corporate/rdc/rd/topics/23/2312-03.html>

<https://ieeexplore.ieee.org/document/10254556>

<https://ieeexplore.ieee.org/document/10292668>

<https://www.worldscientific.com/doi/abs/10.1142/9789812562586_0001>

<https://onlinelibrary.wiley.com/doi/epdf/10.1002/9781118593486.ch4?saml_referrer>

<https://www.reddit.com/r/algotrading/comments/14yg6c1/is_there_commercial_interest_for_superfast/>

What I am looking for is a high-frequency trading (HFT) model that relies on running efficiently to beat the market using a quadratic optimization method. In this sense, we can improve performance by making an existing strategy that relies on quadratic optimization run more efficiently through some approach and observe the results.

Paper:

<https://www.nature.com/articles/s41598-023-45392-w>

Title: “Best practices for portfolio optimization by quantum computing, experimented on real quantum devices”

This paper looks at the classic Markowitz MVO model, and aims to run the model with greater efficiency and accuracy than traditional methods using a quantum-native approach (termed Variational Quantum Eigensolver or VQE) on a real quantum computer. The research takes the MVO problem, converts it to a quadratic unconstrained binary optimization problem, (QUBO), and converts the QUBO to a set of Ising Hamiltonian quantum operators and uses VQE to approximate the solution. The paper aims to find the optimal hyperparameters (ansatz and optimizer) for the problem, and the results are compared among simulators, quantum computers, and a benchmark solution using the branch-and-bound method for a small dataset of four assets (Apple, IBM, Netflix, Tesla).

The research suggests that the traditional branch-and-bound method is capable of solving the optimization problem for up to 120 assets. In fact, for all solving approaches (approximate or exact) on classical computers, the solution tends to only be feasible up to a few hundred assets.

The classic (branch and bound) solution is the optimal solution. The research found that, having run the model on nine IBM quantum computers (known as Noisy Intermediate Scale Quantum devices or NISQ), three of them returned the optimal solution. It goes on to conclude that as quantum computers evolve to more appropriate characteristics, MVO across a wider range of assets will be possible.

Based on this research, one idea might be to look for an alternative approach to solve the MVO problem more efficiently for some number of assets with a classical approach or more accurately using a quantum-native approach (if we can access a quantum computer in some way). Based on what I’ve read from this research, it seems that this approach yields potential for improved solutions at a greater number of decision points (assets) as soon as quantum computers have evolved sufficiently.

<https://onlinelibrary.wiley.com/doi/epdf/10.1002/9781118593486.ch4?saml_referrer>

Title: “Portfolio Optimization: Handbook of High‐Frequency Trading and Modeling in Finance (Chapter 4: Applications in Quantum Computing)”

<https://ieeexplore.ieee.org/document/10292668>

Title: “Real-Time Trading System Based on Selections of Potentially Profitable, Uncorrelated, and Balanced Stocks by NP-Hard Combinatorial Optimization”

<https://ieeexplore.ieee.org/document/10254556>

Title: Pairs-Trading System Using Quantum-Inspired Combinatorial Optimization Accelerator for Optimal Path Search in Market Graphs

A pairs-trading strategy looks for divergence in performance among highly correlated stocks, going short one and going long another to take advantage of mispricing between both securities. This research aims to use a quadratic optimization to search for pair-trading opportunities in N securities. It does this by analyzing market graphs (like a network chart) where each node represents an asset and each edge (line connecting each node with each other node) represents the price difference and correlation factors between the assets (mathematically, correl(i,j)\*[bid(stock i)-ask(stock j)], for both orders of (i,j) and (j,i) for each edge, where correl is a measure of correlation or similarity) . The mechanism for the graph analysis is an embedded Ising machine for optimal path search, in the form of a simulated bifurcation algorithm

Given that this research involves creating the trading infrastructure behind the algorithm, the architecture is more involved than the above study. The research uses a hybrid FPGA/CPU system, where the FPGA portion creates order packets by running the model on market data and passes the result (order filled/lapsed) to the CPU portion for central control and position management.

The research uses a universe of 15 securities (with 210 pairs between different assets) and achieves a computational time of 30.6μs (microseconds). In live trading, the strategy achieved an order fill rate of 93.4% and a total latency of 33 microseconds, which is efficient enough to execute the strategy.

One idea for our research might be to use a similar quadratic optimization model and aim to improve its efficiency using our own methods. This way, we can improve our order fill rate. We could also backtest (assuming we can get data at this granularity) our model using different model designs, graph/edge values, and more to try and achieve the best possible results in live trading. One idea might be to use a quantum computer (as done in the study above) to accommodate a larger group of assets in our model. Because this problem has an application to high-frequency trading (HFT), every improvement in efficiency and breadth makes a meaningful difference in performance.

Other papers:

<https://web.stanford.edu/~boyd/papers/pdf/cvx_portfolio.pdf>

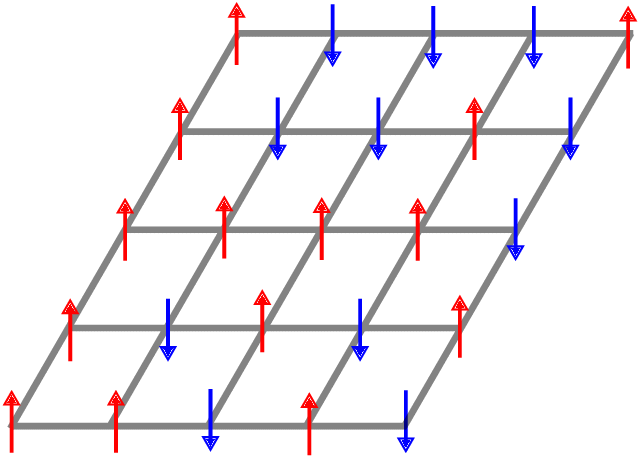
<https://www.hitachi.com/rd/news/topics/2021/2111_cmos.html>

<https://www.global.toshiba/ww/technology/corporate/rdc/rd/topics/23/2312-03.html>

Time: 5 hours

Notes: Ising Model (Source: Wikipedia)

The Ising model represents a square lattice split into a set of sites, indexed “i”. Each site has a binary state, 1 or -1. We can consider this model to represent a ferromagnet such as Iron. Iron can become a permanent magnet if placed under an external magnetic field that leads each “site” in the material to hold the same electron spin. In this case, the electron dipoles together form a negatively or positively charged magnetic force and can attract other magnetic materials.



In a traditional Ising model, each site not only has its binary state, or electron spin, but it also has some interaction, J(i,j) between neighboring sites that impacts the energy of the system. The traditional Ising model also features an external magnetic force h(j) for each point j. We can consider the total energy of some Ising model as the following formula, where <i,j> denotes neighboring pairs:

A close-up of a sign

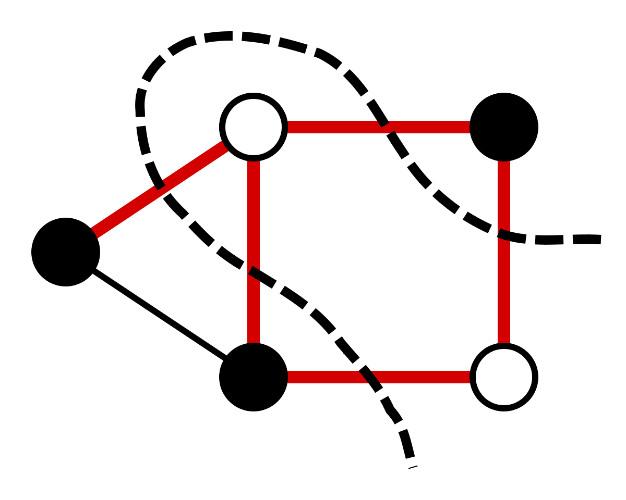
Description automatically generated

Here, the first term sums the product of neighboring sites’ spin with the interaction effect “J”, for each neighboring site. We can consider that if two sites have the same electron spin and a positive interaction effect, the energy of the lattice will decrease overall. Neighboring electrons with opposite spins lead to an increase in total lattice energy.

The second term of the total energy formula iterates through each site, taking the sum of the external field multiplied by the electron spin for each. It then multiplies this sum by mu, the dipole moment for the overall lattice.

However, we can consider a simplified version of the Ising model where there is no external magnetic field operating on the lattice sites. This would cause our second term in the total energy formula to evaluate to zero, and our total energy of the lattice becomes only a function of the first term. In this case, the problem becomes similar to a graph maximum cut (max-cut) problem. The max-cut problem will be described below, alongside its relation to the simplified Ising problem.

Consider a network graph, G, with a set of vertices (points) “V(G)” and a set of edges “E(G)”, where each edge E(G) has a corresponding weight “W(i,j)”. If we split the graph “G” into two groups, one with vertices denoted 1 and the other with vertices denoted -1, the maximum cut problem aims to select groups that maximize the total weight of edges between vertices from opposing groups. In another sense, one can consider this problem as cutting some line in the network graph that splits it into two groups and summing the weight of each connection that is crossed, or broken, by the line. As it relates to the Ising problem, the edges between vertices in opposing groups are equivalent to neighboring electrons with opposite electron spins. Below is a visualization of the maximum-cut problem, where the aim is to maximize the score of the edges crossed by a line. The main decision variable here is which line to draw (or which group to assign each vertex).



For the max-cut problem, we can define δ(V+) as the set of edges that connect vertices from separate groups (cut edges). The size of a cut (scenario) is equal to the following:

A mathematical equation with numbers and symbols

Description automatically generated

Where W(i,j) is the corresponding weight of the edge between the two vertices of different groups, similar to J(i,j) from the Ising model. We divide this sum by two to counteract the double-counting for (i,j) and (j,i) edges between the same two vertices. One popular problem applied to this model is, given some network graph, what is the cut that leads to the largest size.

We can relate this to the Ising problem as follows. First, we define E(V+) as edges between vertices (or neighboring electrons) with positive spins and E(V-) as edges between vertices with negative spins. In both cases, these are not “cut” edges because they have the same spin, or binary state. We can then define the total energy of our Ising lattice as follows:

A group of mathematical symbols

Description automatically generated

In the first transformation, the simplified Ising formula (recall, we assume no external magnetic field here) is split into its three parts. Edges between electrons with spins that are both positive, edges between electrons that are both spinning in the negative direction, and electrons spinning in opposite ways. In the second transformation, the first term takes the negative edge weight between all (neighboring) vertices in each group. However, edges between opposite groups should be additive to the total energy, so this must be added back to neutralize the negative effect and then again to reach a net-positive effect, which is why the second term is multiplied by two for cut edges.

Apparently, only the second term of this final transformation is dependent on the total energy of the lattice, so maximizing (minimizing) this term is equivalent to maximizing (minimizing) the total energy of the lattice. In this sense, both the Ising problem and max-cut problem aim to maximize the weight of cut edges. A mathematical formula that relates the Ising problem to the Max-Cut problem is defined formally below:

A black text on a white background

Description automatically generated

Q: Any more context needed here, any areas I should investigate further before diving into the finance application of these problems?

Q: Why is it that the second term of the Ising reformulation is the only term related to the total energy of the lattice? This is the critical part that relates the simplified Ising problem to the max-cut.

Q: Why is it that we scale the max-cut size by ½ due to double-counting edges but we do not scale the Ising term by ½? Is it defined here that the Ising term only counts each edge once?

Q: Why is it that the Ising total energy, as a function of max-cut weight, multiplies the size by 4? What transformation is being done and what is the rationale?

Hours: 5

Paper Summary: “Pairs-Trading System Using Quantum-Inspired Combinatorial Optimization Accelerator for Optimal Path Search in Market Graphs”

Arbitrage involves selling the same or similar securities on two different markets, taking advantage of mispricing opportunities. The pairs-trading strategy is a form of this, where we look for two historically correlated securities and monitor performance for deviations from the historical correlation (for example, two positively correlated stocks move in opposite directions).

This research uses a network graph to depict the market, with nodes as securities and edges as representing the relationship between two securities.

Aim is to minimize the sum of weights activated by the network graph (cost function):

A black and white math symbol

Description automatically generated

Each solution will yield a net purchase and sale of two stocks, whether this is direct or bypass (a->b) vs. (c->a->b->c). This is because each pass through the graph must be cyclical. In the network graph, favorable paths may exist other than the optimal solution. The optimization is run multiple times using the tabu search method to yield different results.

The penalty function is defined as follows, which enforces constraints for the Tabu search and cyclicality:

A group of mathematical symbols

Description automatically generated

The overall optimization is a QUBO defined as a linear combination of the penalty and cost function: A black and yellow math equation

Description automatically generated

The Ising Machine finds the bit configuration (b(i,j)) that minimizes the QUBO cost function

Time: 2 Hours

To do:

Begin creating model in Julia using spreadsheet data for now

Solve for approximate solution, not using branch-bound

Notes on Pairs-Trade paper:

Resource for tick-level security data: <https://lobsterdata.com/>

Resource for tick-level crypto data: <https://www.binance.com/en/landing/data>

Resource for tick-level crypto data (reasonable pricing): <https://crypto-lake.com/subscribe/>

Resource for tick-level crypto data (less reasonable pricing but better coverage): <https://tardis.dev/>

Forums:

<https://quant.stackexchange.com/questions/71688/are-there-open-source-or-academic-only-limit-order-book-data-sets-available>

<https://quant.stackexchange.com/questions/38385/real-time-limit-order-book-data-of-desired-depth>

<https://quant.stackexchange.com/questions/76503/how-are-order-book-and-trade-data-consolidated-distilled-into-a-more-tractabl>

<https://quant.stackexchange.com/questions/78115/free-historical-limit-order-book-tick-data-with-decent-depth>

Link: <https://ieeexplore.ieee.org/document/10254556>

How are relative prices calculated?

In computing edge weights, the research paper uses bid and ask prices for different stocks. It explains them this way: “ask and bids are normalized by the base price on the day”. Does this suggest that we are looking at the daily movement from the base price?

How is the similarity factor calculated?

In computing the similarity factor, the research uses a dynamic time warping (DTW) distance between pairs of nodes (stocks). The DTW distance is computed on price sequences of the securities and

Dynamic Time Warping (DTW) Distance:

A close-up of a graph

Description automatically generated

DTW is a method of two time series that differ in speed or velocity. We can say that similarities in the trends between securities are happen at different times, and therefore want to look at the similarity between securities where trends occur at different times.

Restrictions:

Every index must be matched with at least one index from the other sequence

Each index from the first (last) sequence must be matched with the first (last) index from the second sequence.

Index mapping between sequences must be monotonically increasing. The lines between sequences can never cross.

Optimal match is one where the sum of differences between matched values is the lowest

Direct vs. Bypass Paths

As mentioned, the research uses a network model of markets and aims to iterate through nodes/edges to compute the best path based on achieving the maximum weight. Any pass through the graph will lead to a buy-sell pair between just two securities. These securities are on the edges of the path: (a -> b -> c) corresponds to a long position in security “a” and a short position in security “c”.

Order Sizing:

The size (# of shares) of each order placed depends on the minimum shares that can be purchased (a single lot; L), the price of a single lot (p), and the amount of money we want to spend per trade (“A”). We can compute trade size per security as: A / (L\*p).

Time: 6 hours

Written Notes:

The penalty function aims to enforce a set of conditions for the pairs-trade model. The overarching aim is to minimize the linear combination of the penalty and cost function, and the penalty function is written below:

A group of math symbols

Description automatically generated

The terms of the penalty serve a set of purposes, which we list below:

1. The first term forces the outflow of each node to be 1 or less
2. The second term forces the inflow of each node to be 1 or less
3. The third term forces the inflow and outflows of each node to be equal
4. The fourth term forces forbids traversing the same edge twice in different directions
5. The fifth term forbids choosing the pairs in the tabu list T(i,j)

Our total cost function is the combination of our cost and penalty functions, which we aim to minimize. This function takes the form of a quadratic unconstrained binary optimization (QUBO).

In generalizing the penalty function, we can consider 4 main cases that impact which terms of the penalty functions are activated, or not necessarily equal to 0:

1. i=i’, j **≠**j’
2. i=i’, j=j’
3. i **≠**i’, j=j’
4. i **≠** i’, j **≠** j’

Each of these cases will impact which of the first two terms of the penalty function are activated, if any. The third, fourth and fifth terms will be activated regardless.

1. Term 1: Yes, term 2: No
2. Term 1: No, term 2: No
3. Term 1: No, term 2: Yes
4. Term 1: Yes, term 2: Yes