Gregory Knapp

Dr. Grimsman

9/30/2021

CS 312

Project 2 Convex Hull Solver Report

# Time and Space Complexity Analysis of Project 2 – Breakdown by Subsection

The time complexity for each function found in convex\_hull.py of the Project 2 code is as follows and pulled from the comments for each function in the code.

DNCHull(self, arr): Lines: 82 - 106

```
# This is the core divide and conquer algorithm of the project. It takes in the full x-value sorted array from the # main compute_hull function of size N, divides it by 2 over and over until sub-arrays of size 5 or smaller are # made. These 5 item or smaller arrays make up the base case for the function and each sub-array's convex hull # is found using the brute force convex hull algorithm below. When the array does not meet the base case, it is # split in half and the recursive DNCHull() call is made on the left and right sub-arrays.

# # Once the base cases are handled and the recursive calls are returned, the function then passes the arrays of # points for the convex hulls of the left and right sub-arrays to mergeHulls(). mergeHull then processes these # arrays using the upper and lower tangents to return an array of points that make up the combined convex hulls. This result is then either returned up the recursive stack or back to # compute_hull() if it is the final call. The final result is the array of points making up the convex hull of the # entire set of points.

# # The time complexity of this function comes down to the divide and conquer nature. A more complete theoretical # analysis of the Divide and Conquer function is found in the full report. Since the function is breaking down # the larger array of points into two sub-problems of N/2 size. This makes the time complexity of the Divide and Conquer aspect of the algorithm O(log n).
```

```
# The time complexity is directly affected by the merging function since it
is considered post-work to the recursive
# part of the function. Time complexity for mergeHull() is O(N), which is
detailed below.
#
# This makes our final time complexity for the Divide and Conquer Convex Hull
Algorithm = O(N log N).
#
# Space complexity for this algorithm is ultimately O(N) as it has the one
array containing all of the points that
# is being split up and worked on by the various helper functions
```

#### mergeHulls(self, left, right) - Lines 129 - 151

```
(though it is more likely to be 2 or 3
# in the smallest cases). Because the difference between 1 and 5 is
insignificant, the space complexity can just be
# called O(s).
```

### clockwisesort(self, arr) - Lines 286 - 319

```
# Since this is a type of sorting algorithm, the result array will have the exact size of the input array, just in
# a different order. We do create a second array of duples that is used to store angle and index values, which would
# be of the same length, just with a second column of values per row. In total, we essentially have 3 arrays of size
# s created and used during this function plus the input array of O(s0,
# so our space complexity could be labeled O(4s) or just O(s).
# three times: Once to get the sum of x and y values to find the center, a second time to get the angles of each
# line from the center, and a third to add the points in the correct order to the result array. This would normally
# be a time complexity of O(N^3), but using the guaranteed small array size I have designated as s, this time
# complexity would be O(s^3)
#
# Same as the time complexity, the space complexity only deals with a largest size of s for the result array.
# Since this is a type of sorting algorithm, the result array will have the exact size of the input array, just in
# a different order. We do create a second array of duples that is used to store angle and index values, which would
# be of the same length, just with a second column of values per row. In total, we essentially have 3 arrays of size
# s created and used during this function plus the input array of O(s0,
# so our space complexity could be labeled O(4s) or just O(s).
```

#### getRightMostX(self, arr) - Lines 369 - 390

```
# A helper function used in getuppertangent() and getlowertangent() in order to find the starting value of the # left side array for merging. Finding the left most x-value in the right side array is trivial, since it was # purposefully added first to the clockwise sorted arrays, it can be extracted directly. However, the rightmost # x-value which is used to make the starting line for testing for the convex hull tangent lines is no longer found # at the final value of the left points array after the clockwise sort of the base case hulls. # # This helper function is just a simple for loop through the left array to find the max x-value and return the index # it is found at. The code is separated into the helper function to reduce duplicate code since it is used by # both the upper and lower tangent functions. # # This is a simple looping function meant to identify the max x value from an array. The arrays used in this # can be as small as s to N/2, since this function is only called on the left sub-hull of the hulls to be merged. # In the context of this function, the time complexity could be said to be O(N) since it is a for loop over the # entire array of points. However, in the larger picture of the overall function, it acts as more of a O(1) constant # (more accurately O(N/2) time function, since it never iterates over the entire array of N points, even on the
```

```
# final step of recursion. Because of this, it does not impact the overall time complexity of the divide and # conquer # # The space complexity of this function is similarly dependant on the input size, but has a max size of O(N/2) # where N/2 is the size of the largest array that will be passed into it. Similar to time complexity, in the # context of the function, the space complexity if O(N), however, in the bigger picture it is more accurately # O(N/2) which is more similar to O(1).
```

#### getuppertangent(self, left, right) - Lines 407 - 443

```
A helper function for mergehulls() using the left and right sub-hull arrays
```

```
# approach size of N between the two arrays. During the final recursive call,
the two hulls to combine will be
# of size N points, however, traversal of the upper tangent does not take
O(N) time to complete, since the furthest
# the tangent line could be would be about on the other side from the most
inside points, and often is much shorter
# than that. As a result, the time complexity is more accurately estimated in
a worst case scenario to be O(N/2),
# which is more similar to a constant O(1) time. In conjunction with
getlowertangent(), the time complexity for the
# two tangent functions together could be estimated at O(N) (which I have
done in the larger time complexity
# analysis of the entire divide and conquer) since it is possible to traverse
the almost the entire array in search
# of the tangent lines depending on the spread of points. Individually
however, it is more like O(N/2) time.
#
# Space complexity for this function is more straightforward. The final
recursive merges will have hulls containing
# essentially the full set of points to work with or O(N) points between the
left and right sub-hull arrays.
# Other than the input arrays, this function works in constant space
variables, so the O(N) is the total size
# complexity for this function.
```

### getlowertangent(self, left, right) - Lines 524 - 543

```
# A helper function for mergeHulls parallel to getuppertangent(). This function takes the same left and right # sub-hulls as getuppertangent and performs similar calculations on them, this time to get the lower tangent line of # the combined hull. # # The two functions are essentially identical with 4 differences. # 1) The first inner while loop (traversing the left sub-hull) now moves in a clockwise direction. # 2) The first inner while loop now checks for decreasing slope to know when to break the loop # 3) The second inner while loop (traversing the right sub-hull) now moves in a counter-clockwise direction. # 4) The second inner while loop now checks for increasing slope as its condition to break the loop # # In essence, getting the lower tangent is the same functionally, only reversing the movement and loop conditions # of the upper tangent function. Because of the similarities, the specific workings of the functions won't be # repeated again here. # # The justification for the time and space complexity of this function is the same as for getuppertangent, so it # will not be duplicated here. # The time complexity for getlowertangent in isolation is O(N/2), O(N) when considered together # with getuppertangent. #
```

### Theoretical Analysis of Project 2 Code

As depicted in the subsection time and space complexity breakdown of the code above, the overall time complexity of the Convex Hull Solver is dictated by two main parts: The divide and conquer function it self and the mergeHulls function called at the end of the divide and conquer.

The mergeHulls function is easy enough to evaluate complexity wise; the function iterates through N values across the two arrays in order to make the merge happen, giving us a time complexity of O(N) time. This value is our post-work time complexity in our divide and conquer algorithm.

Referring to the divide and conquer function, because we know our post-work time complexity, we can take a look the actual time complexity of the function itself. The DNCHull() function takes the array of size N and divides it into two sub problems, each one of size N/2. We now know that in the Master theorem, our a = 2, b = 2 and d = 1 because for  $O(N^d) = O(N)$ , d must equal 1.

Plugging this into for our recurrence relation, we get the following:

$$T(n) = 2T(N/2) + O(N)$$

If we then evaluate our Master theorem relationship for a, b, and d, we get the following:

$$a/(b^d) = 2/2^1 = 1 \rightarrow 1 == 1$$
 Therefore running time should be O(N^d log N).

Know that our d = 1, we get the expected running time of  $O(N \log N)$ , which matches up with the comments in the code found in the first section of this report.

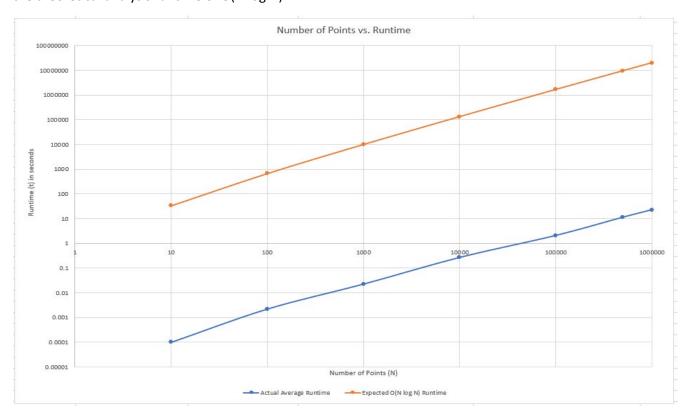
## **Empirical Analysis of Project 2 Code**

The following table is the results of my empirical testing on the Convex Hull Solver

N # of Points	Test 1	Test 2	Test 3
10	0.0001	0.0001	0.0001
100	0.002	0.002	0.003
1000	0.022	0.023	0.023
10000	0.266	0.266	0.266
100000	2.047	2.091	2.075
500000	11.201	11.208	11.305
1000000	22.636	22.378	22.46

Test 4		Test 5	Average Runtime	Expected O(N log N)	
	0.0001	0.0001	0.0001	33.21928095	
	0.002	0.002	0.0022	664.385619	
	0.022	0.022	0.0224	9965.784285	
	0.267	0.27	0.267	132877.1238	
	2.061	2.044	2.0636	1660964.047	
	11.207	11.323	11.2488	9465784.285	
	22.427	22.505	22.4812	19931568.57	

From this table, I created the following graph of points (N) against runtime (t) with logarithmic axis. The blue curve is the actual runtimes from my trial and the orange curve is the expected run times assuming the theoretical analysis runtime of O(N log N).



As is clear in the graph, the two curves, although similar in shape and curvature, are not even close in terms of time for each value of N. Because of the similarities in shape, I did not think that there was a mistake in the run time being  $g(n) = O(N \log N)$  for the convex hull algorithm, so I decided to look into constants of proportionality k that could make the CH(P) = k \* g(n).

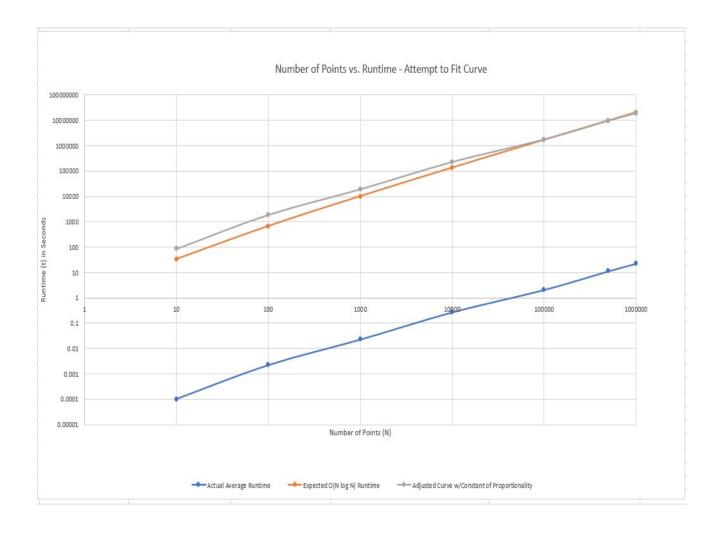
Based on the equation CH(P) = k \* g(n), I used my actual and expected values to add a column in my table for k as seen below.

N # of Points	Actual Average Runtime	Expected O(N log N) Runtime		
10	0.0001	33.21928095		
100	0.0022	664.395619		
1000	0.0224	9965.784285		
10000	0.267	132877.1238		
100000	2.0636	1660964.047		
500000	11.2488	9465784.258		
1000000	22.4812	19931568.57		

Adjusted Curve w/Constant of Proportionality	Constant of proportionality k = CH(Q) / g(N)
84.14928044	3.0103E-06
1851.28417	3.31128E-06
18849.43882	2.24769E-06
224678.5788	2.00938E-06
1736504.551	1.24241E-06
9465784.258	1.18836E-06
18917768.03	1.12792E-06

The final column "Constant of proportionality k = CH(Q) / g(N)" is the result for each row that I calculated k for. The second to final column is the values for each value of N when I applied the constant that I chose for my graph.

Thinking that Big O notation is concerned with the behavior of functions in relation to one another when input numbers get extremely large, I applied the second to last value of k that I got when I divided 11.2488 / 9465784.258 which was 1.18836e^-6, resulting in the values in the second to last column. When I plotted these new values along with the actual and expected curves, I got the following graph.



The gray line is the adjusted curve with the constant of proportionality while the orange and blue curves are the same as before. With  $1.18836e^{-6} = k$  applied to the curve, we see that the beginning of the curve does stretch above the expected O(N log N) values for the graph but comes down below it and settles in right below O(N log N) at around N = 1,000,000 and continues that way, so that g(n) does act like an upper bound for our adjusted curve.

# Theoretical vs. Empirical Analysis - Conclusions

With the analysis above, I still feel comfortable concluding that although the actual and expected values from the theoretical and empirical analysis were very different, the runtime for the Convex Hull algorithm should be O(N log N).

The similarity in the shape of the two curves over the large data points initially made me think that it was not a different type of function, an O(N) curve would grow at an entirely different rate, and  $O(\log N)$  curve would be much less steep in terms of growth rate. This made me think that it is more likely to be a case of finding the right constant of proportionality such that  $g(n) = O(N \log N)$  would be an upper bound for CH(P).

I will admit, I am not familiar with curve fitting software or coding packages for python or any other language, so all of this work was done just with Excel and from observation and based on what was mentioned in the lab specs, so it is in no way the correct constant of proportion. However, the graph behavior with  $k = 1.18835e^{-6}$  added in does look much more like the relationship we would expect between the two functions, so I feel comfortable with it as base analysis for the results.

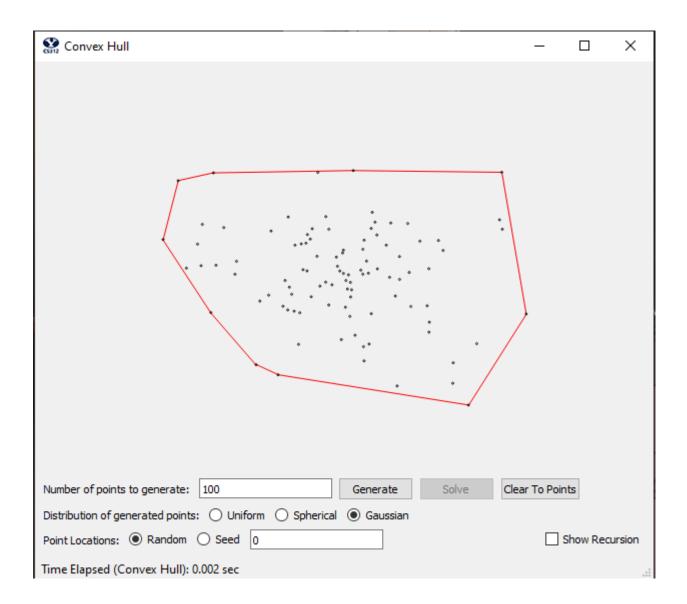
As for why the differences between the theoretical and empirical analysis came to be, it is hard to say. The first thing that comes to my mind is hardware power. It is hard to assume what speed processes are handled at when doing theoretical analysis, and this can only be evaluated by running empirical testing. Just from keeping up with the slack channel, I've seen students run the 1,000,000 test as low as nearly 10 seconds and over 200 seconds. I fall on the more average spectrum of about 22 seconds, which is still a huge difference from 200 seconds, so hardware certainly had an impact on the differences.

The biggest source for the difference in my mind comes from the N part of O(N log N). Assuming our time is in seconds, saying O(N) runtime for an algorithm would mean that for each one piece of input, we are taking 1 seconds to process it. In my algorithm, the O(N) part of O(N log N) stemmed from the mergeHulls() function that was called in the recursive divide and conquer function. Theoretically, I labeled it O(N) because it is iterating two different not nested times over the array of points in the hull. This should be the main time eating part of the algorithm, since O(N) is significantly slower than O(log N), however to the computer, O(N) processes really are trivial things. Until we reach exponential time, computers handle things in constant, linear, and lower polynomial time so easily that the O(N) functions really do seem more like constant function (even if they are not in reality and shouldn't be assumed to be so).

I think the discrepancy between expected time to run and actual time to run in the O(N) part of my algorithm would cause the large gap between expected runtime and actual runtime.

### **Working Screenshots**

Screenshot #1 – Convex Hull Solver ran with 100 Points



Convex Hull	_		×
Number of points to generate: 1000 Generate Solve C	ear To Poir	nts	
Distribution of generated points: O Uniform O Spherical O Gaussian			
Point Locations:   Random  Seed  O		Show Red	ursion
Time Elapsed (Convex Hull): 0.024 sec			:

### Source Code Files

### convex\_hull.py code file

```
from which pyqt import PYQT VER
if PYQT VER == 'PYQT5':
   from PyQt5.QtCore import QLineF, QPointF, QObject
    raise Exception('Unsupported Version of PyQt: {}'.format(PYQT VER))
RED = (255, 0, 0)
BLACK = (0, 0, 0)
PAUSE = 0.25
class ConvexHullSolver(QObject):
        self.view.clearLines(line)
```

```
def eraseHull(self, polygon):
   points.sort(key=QPointF.x) # Time Complexity for Python's sort is
                range(len(finalhullpoints))]
    t4 = time.time()
```

```
# compute hull() if it is the final call. The final result is the array
```

```
def mergeHulls(self, left, right): # Space complexity of input values:
    uppertangent = self.getuppertangent(left, right) # getuppertangent()
```

```
leftindex = 0 # Initialize leftindex - Time: O(1)
result.append(left[leftindex]) # Append value to result - Time: O(1)
while left[leftindex % len(left)] != uppertangent.p1(): # Time:
result.append(right[rightindex % len(right)]) # Append value to
    result.append(right[rightindex % len(right)]) # Append value to
    leftindex = leftindex + 1 # Iterate variable - Time: O(1)
```

```
if (a * testpoint.x()) + (b * testpoint.y()) > c: #
```

```
if pointone not in hullpoints: # Evaluate if statement -
   hullpoints.append(pointtwo) # Append value to
```

```
startline = QLineF(center, arr[minindex]) # Create line - Time: O(1)
```

```
angleandindex.append([i, angle]) # Append duple to array - Time:
```

```
def getuppertangent(self, left, right): # Space O(N) for two input
    rightsidestartindex = 0 # Initialize variable - Time: O(1)
```

```
tangentline = QLineF(p, q) # Make line - Time: O(1)
        r = left[(leftsidestartindex - leftiterator) % len(left)] #
       nextslope = (q.y() - r.y()) / (q.x() - r.x()) # Calculate
       leftiterator = leftiterator + 1  # Assign variable - Time:
```

```
currentslope = (q.y() - p.y()) / (q.x() - p.x()) # Calculate
```

```
p = left[leftsidestartindex] # Initialize variable - Time: O(1)
```

```
r = left[(leftsidestartindex + leftiterator) % len(left)] #
```

#### Proj2GUI.py

```
from which pyqt import PYQT VER
if PYQT VER == 'PYQT5':
  from PyQt5.QtWidgets import *
  from PyQt5.QtGui import *
  from PyQt5.QtCore import *
elif PYQT VER == 'PYQT4':
  from PyQt4.QtGui import *
  from PyQt4.QtCore import *
  raise Exception ('Unsupported Version of PyQt: {}'.format(PYQT VER))
  def displayStatusText(self, text):
  def clearPoints(self):
  def clearLines(self, lines=None):
```

```
self.lineList[color].remove(line)
   self.update()
   app.processEvents()
def paintEvent(self, event):
   painter = QPainter(self)
   painter.setRenderHint(QPainter.Antialiasing, True)
   tform = QTransform()
      c = QColor(color[0], color[1], color[2])
        pt = QPointF(w*point.x(), h*point.y())
```

```
super(Proj2GUI, self). init ()
def newPoints(self):
   if self.randBySeed.isChecked():
   ptlist = []
               ptlist.append( QPointF(xval, yval) )
               ptlist.append( QPointF(xval, yval) )
```

```
ptlist.append( QPointF(xval, yval) )
def clearClicked(self):
   self.view.clearLines()
   self.view.update()
   self.view.addPoints( self.points, (0,0,0) )
   self.view.update()
def solveClicked(self):
   app.processEvents()
   self.view.update()
   app.processEvents()
def randbytime(self):
def randbyseed(self):
def initUI( self ):
   self.statusBar = QStatusBar()
```

```
vbox = QVBoxLayout()
self.randByTime
self.showRecursion = QCheckBox('Show Recursion')
h = OHBoxLayout()
h.addStretch(1)
grp = QButtonGroup(self)
h.addStretch(1)
h.addWidget( self.randBySeed )
```

```
vbox.addLayout(h)

self.generateButton.clicked.connect(self.generateClicked)
self.solveButton.clicked.connect(self.solveClicked)
self.clearButton.clicked.connect(self.clearClicked)

self.randByTime.clicked.connect(self._randbytime)
self.randBySeed.clicked.connect(self._randbyseed)

self.randByTime.setChecked(True)
self.distribOval.setChecked(True)
self.generateClicked()

self.showRecursion.setChecked(False)
self.show()

if __name__ == '__main__':
    # This line allows CNTL-C in the terminal to kill the program
signal.signal(signal.SIGINT, signal.SIG_DFL)

app = QApplication(sys.argv)
w = Proj2GUI()
sys.exit(app.exec())
```

PYQT\_VER = 'PYQT5'