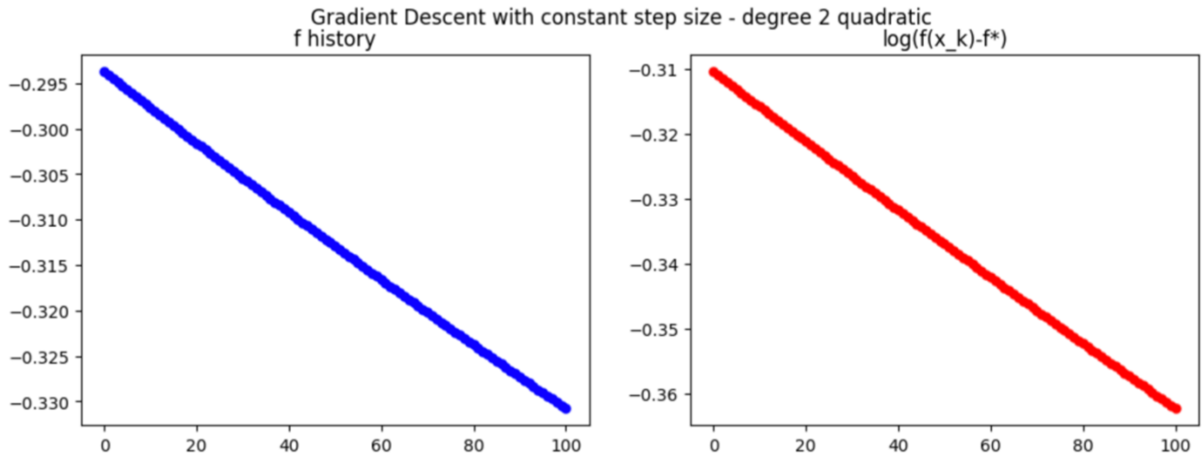
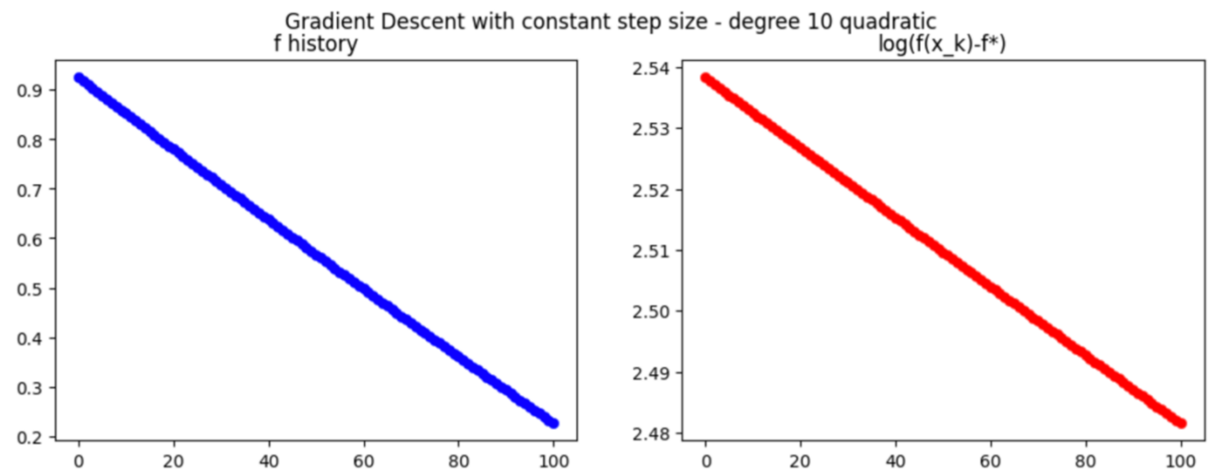
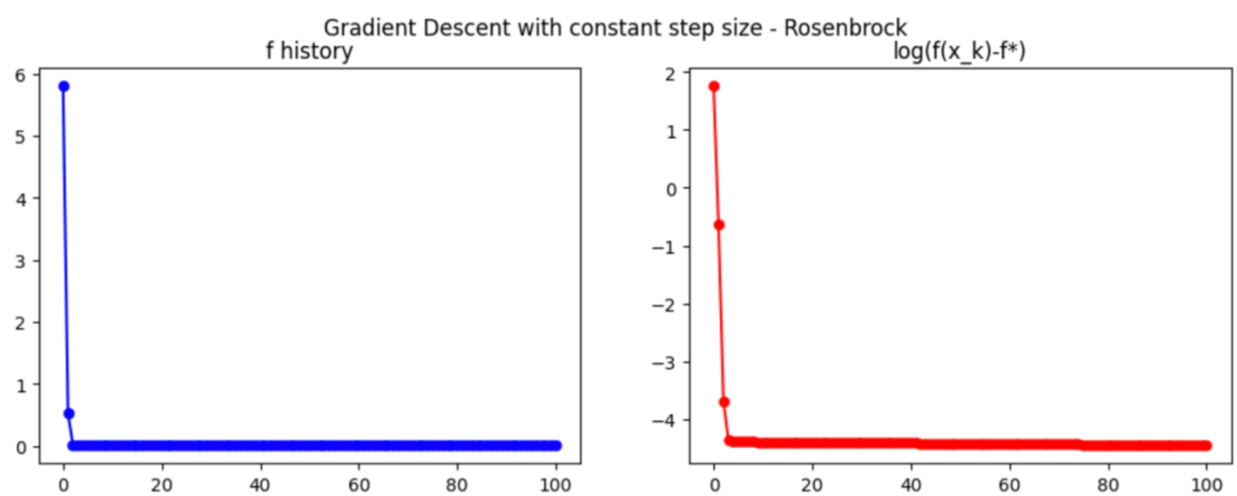
[4]

I.

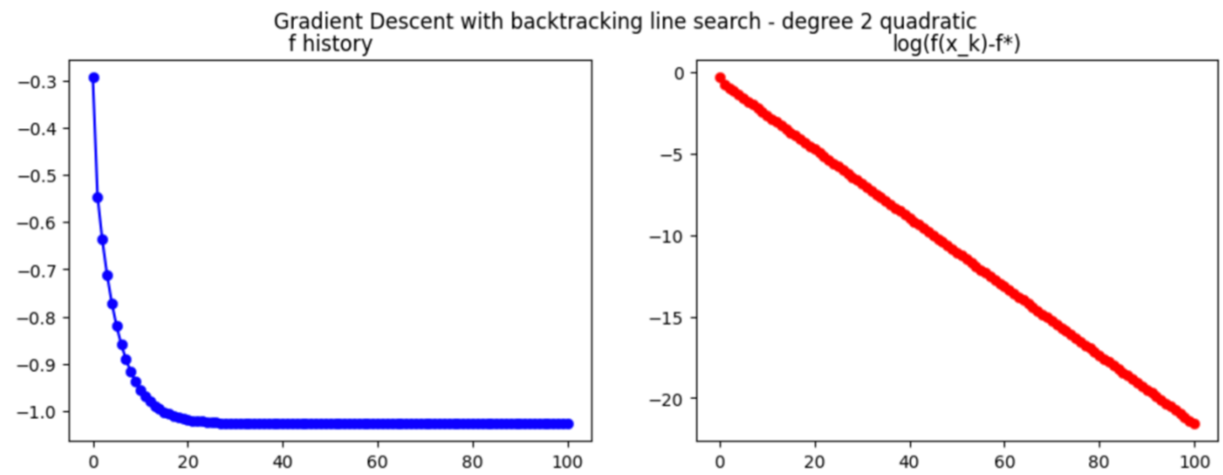
i) Gradient Descent with constant step size

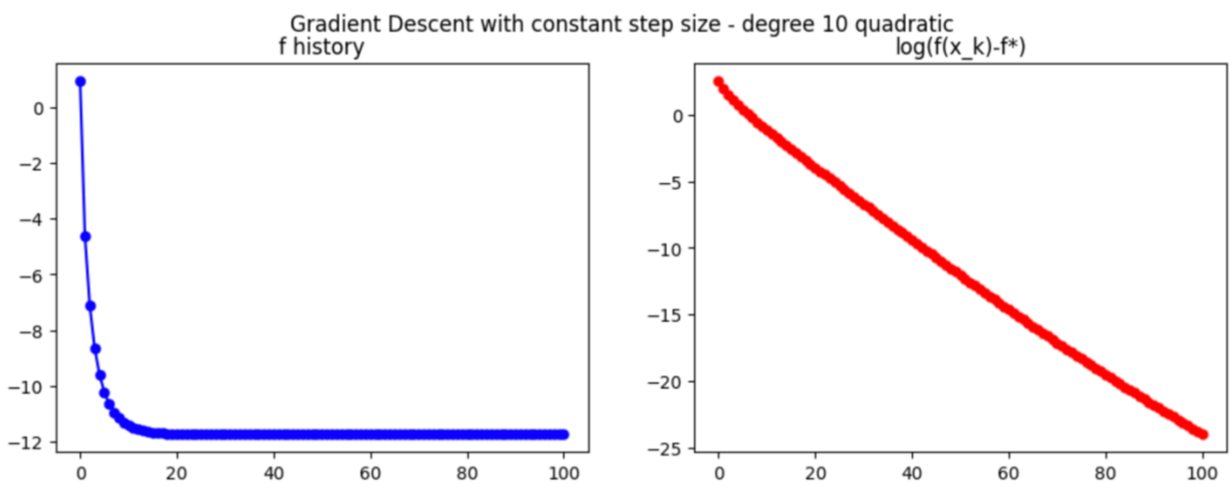


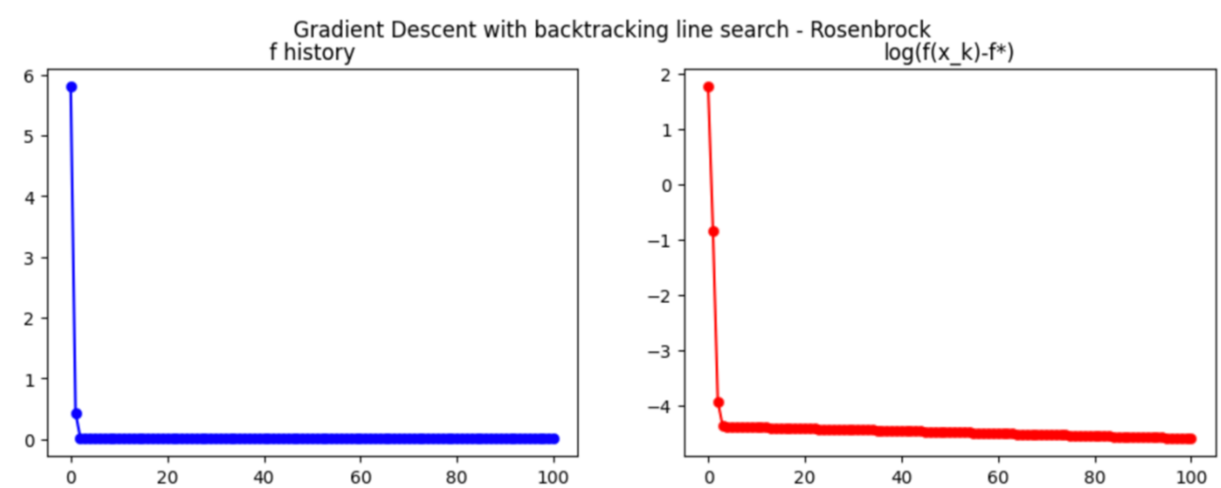




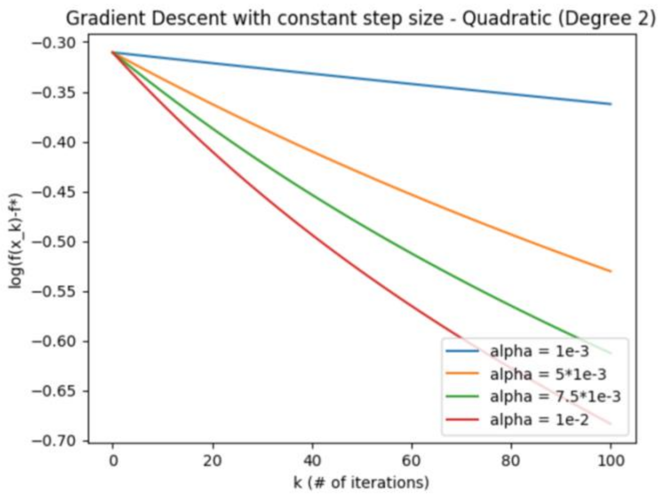
ii) Gradient Descent with Backtracking line search

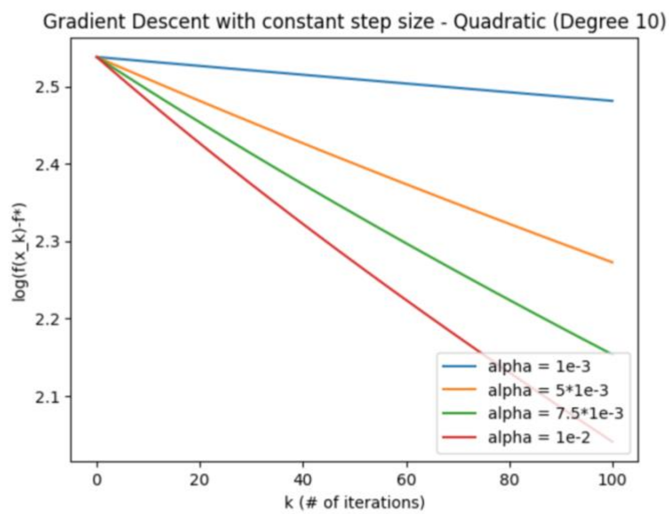


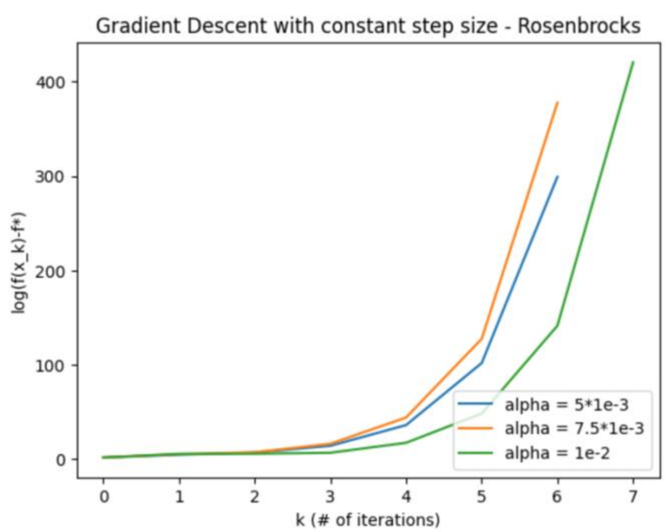




II.


Gradient Descent with constant step size - Quadratic (Degree 2)
log(f(x_k)-f*)
k (# of iterations)
alpha = 1e-3
alpha = 5*1e-3
alpha = 7.5*1e-3
alpha = 1e-2

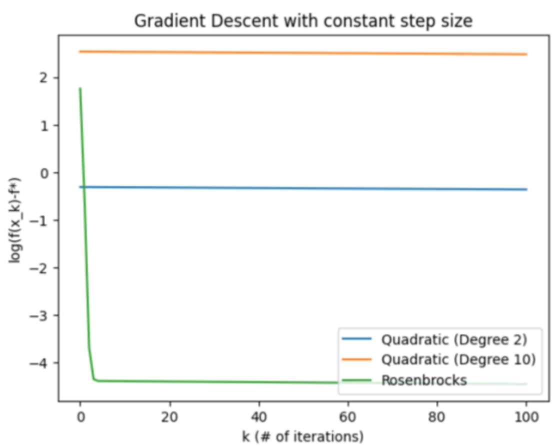

Gradient Descent with constant step size - Quadratic (Degree 10)
log(f(x_k)-f*)
k (# of iterations)
alpha = 1e-3
alpha = 5*1e-3
alpha = 7.5*1e-3
alpha = 1e-2

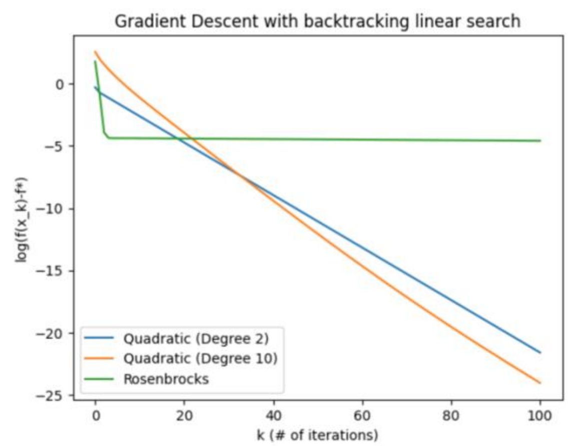

Gradient Descent with constant step size - Rosenbrocks
log(f(x_k)-f*)
k (# of iterations)
alpha = 5*1e-3
alpha = 7.5*1e-3
alpha = 1e-2

Overall, alpha of of 1e-2 converges the fastest for most times, and then 7.5*1e-3, 5.0*1e-3, so on. We can observe that if the constant is changing within a specific range, larger alpha usually performs better in converging. However in Gradient Descent with constant step size of using Rosenbrocks problem seems that it's loss (error) increases, which is obviously strange that is diverges from the same initial value of zero.

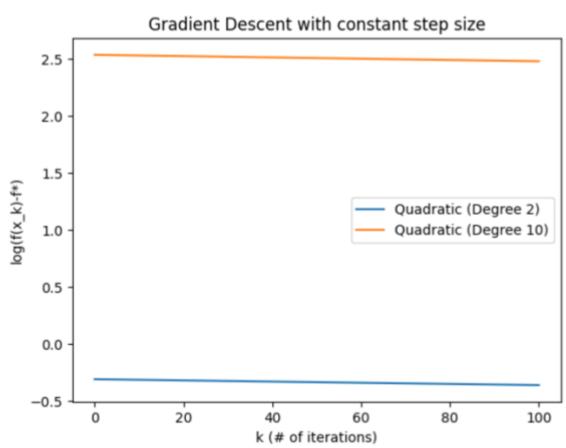
[5]

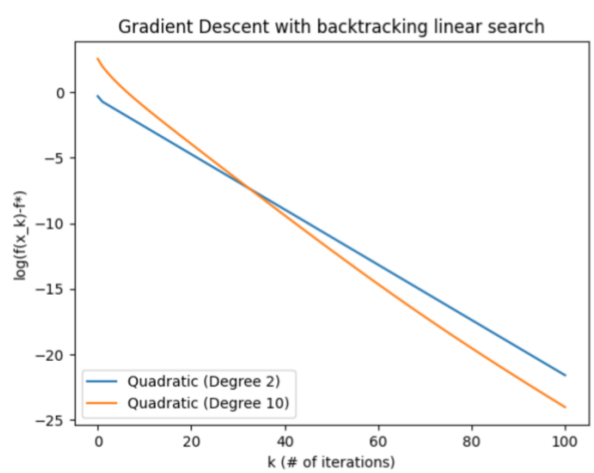
< Comparison of Quadratic problems (including Rosenbrocks) >





< Quadratic problems only >

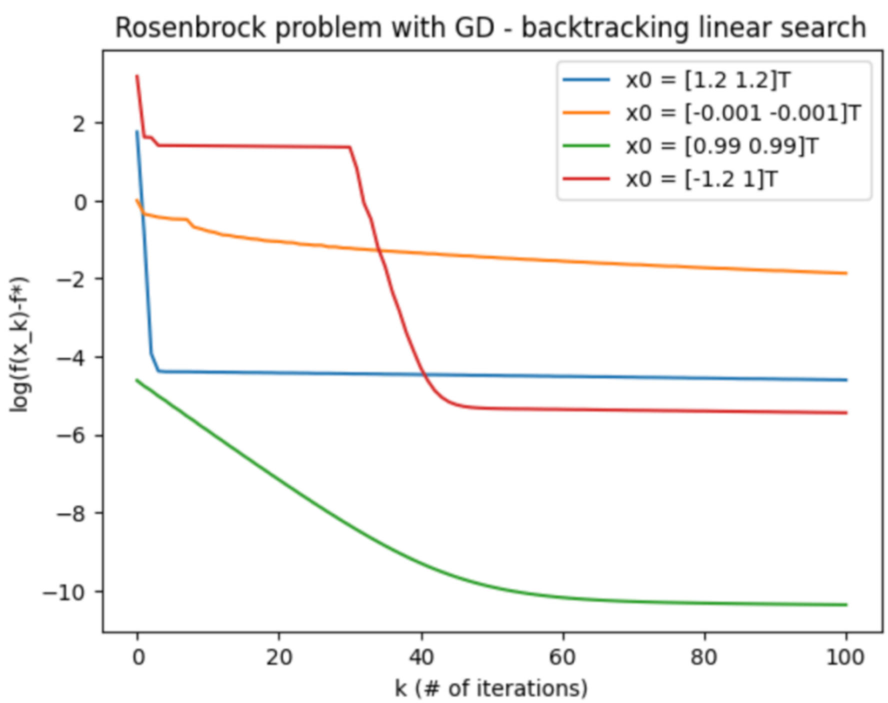




As shown in plots above, in constant step size, quadratic with degree 2 generally has smaller error than degree 10, and in backtracking line search, degree 10 converges slightly faster and both dimensions show a good performance rate.

With respect to Newton's method, it has converging action to search for local minimum in every step (every iteration). As A is positive definite so that the Hessian always is positive definite as well which makes the function to be strictly convex and have a strict local minimum.

[6]



As the starting points becomes larger, it has more abrupt decrease/change in its trend. It seems like the smaller number in positive performs better in terms of reducing its loss and convergence.

[7]

- In backtracking with Gradient Descent, the function iterates in alpha = 1.0 most of times, in quadratic problems.
- Gradient Descent methods are usually slower in its convergence and computationally not expensive than Newton's method which means it is quite simple update rules.

On contrary, Newton's method converges to the point much faster than GD which is an advantange. However, it is computationally expensive that it requires extra loop to do line search within each iteration, including a computation of Hessian and its inverse.

9.

As a student who does not have a good background in linear algebra and hard time studying materials to get familiar to them, it would be very great to have weekly quiz (which can be either easy or hard but creative, I personally prefer the first to check up if I am catching it up right) to check myself understand the material enough. Or, it will be also great to have review/discussion session so that students in the class can share their reasons, and more chances to get help from your peers/instructors. I would appreciate any of the changes or ideas. I'm sincerely thanks to instructors for being passionate and truly caring for students.