Numerical Optimisation: Assignment 2

Student Number:13064947

January 2018

1 Exercise 1 - Rosenberg function Descent b),c),d)

1.1 Chart explanations

Given the severe restrictions on space for the assignment. I have given four pages of plots, all with the same format. I will give a prose description. For the various descents, I have used lines, but also placed X's or *'s for each iteration in a different colour. This shows periods where solid progress is made compared to where the descent is a tiny step. This can also be seen on the bottom charts showing alpha's versus iterations.
For the steepest descent trajectory I have shown two charts. The left being the full descent trajectory. On the left chart I have also placed a box where the optimisation is close to the optimum (sometimes very small). The right hand plot shows this box zoomed in, the title shows the final number of iterations performed within this box - sometimes several thousand. Finally, when comparing alphas - due to the vastly different scale of alpha steps for SD v Newton, I use a semi-log scale for the former.

- Top Row: Steepest Descent contour plots (Left: Full Trajectory, Right: Zoom in of final N iterations, N shown in title)
- Middle Row: Newton Direction descent contour plot - Full Trajectory
- Bottom Row: Iterations v Size of Alpha Step plots (Left: Steepest Descent - semi-log y-axis, Right: Newton - standard axis)

1.2 Page Order

The order of the pages is as follows:

- Page 2: Start point (1.2,1.2) - Backtracking (rho0.1), figs 1, 2.
- Page 3: Start point (-1.2,1.0) - Backtracking (rho0.1), figs 3, 4.
- Page 4: Start point (1.2,1.2) - Strong Wolfe, figs 5, 6.
- Page 5: Start point (-1.2,1.0) - Strong Wolfe, figs 7, 8.
- Page 6: Bullet Point summary.

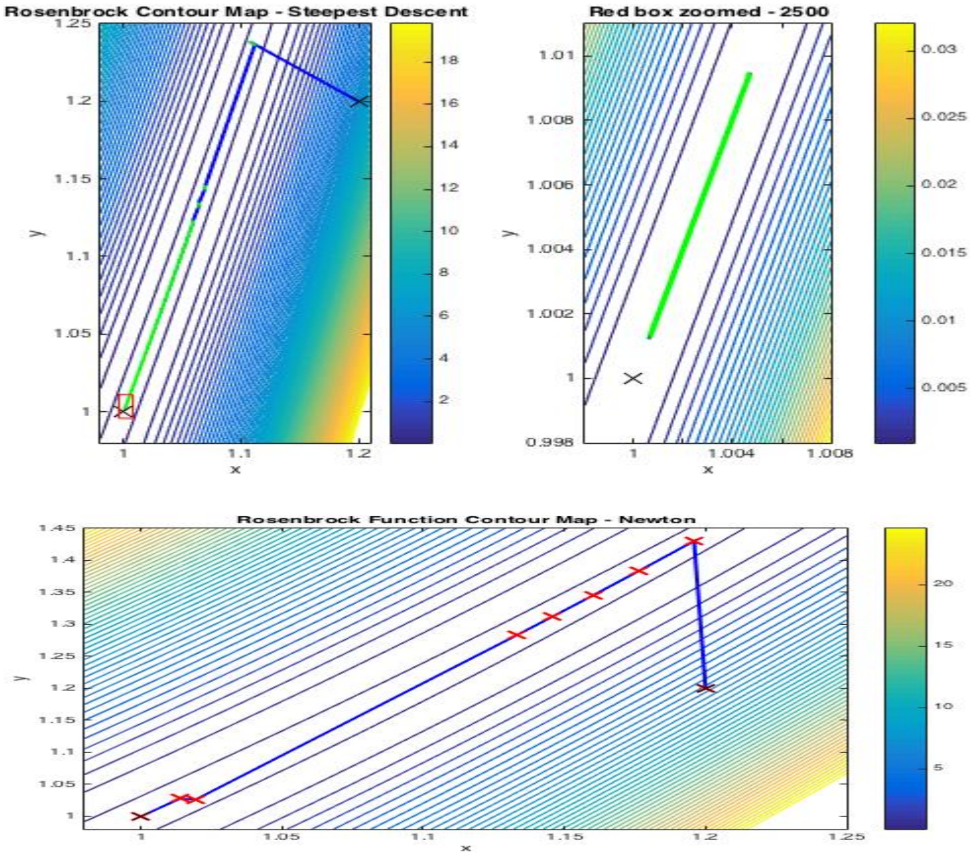
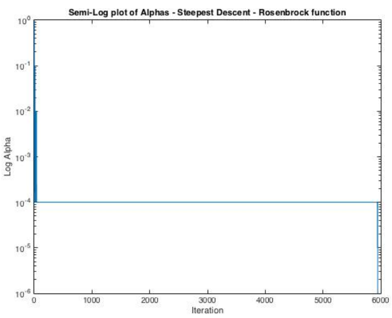
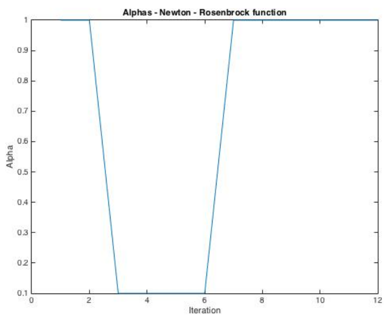


Figure 1: Steepest Descent and Newton - Backtrack- pt1



(a) Semi-Log Alphas versus Iterations SD



(b) Alphas versus Iterations Newton

Figure 2: Backtrack alphas SteepDescent versus Newton pt1

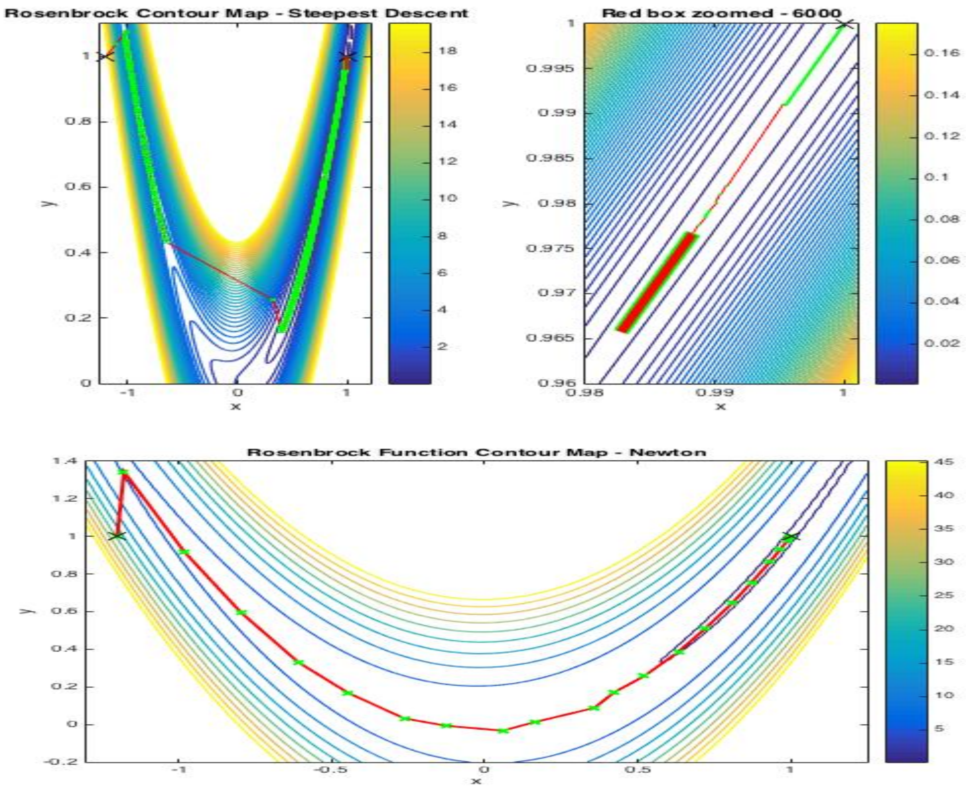
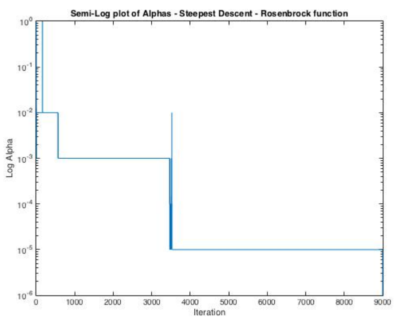
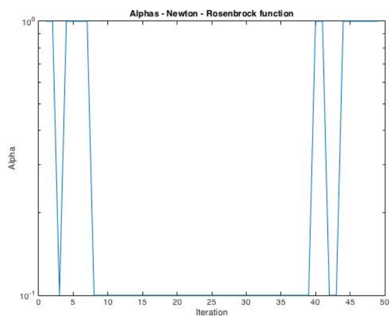


Figure 3: Steepest Descent and Newton - Backtrack - pt2



(a) Point 2: Semi-Log Alphas versus Its



(b) Point 2: Alphas versus Its Newton

Figure 4: Backtrack alphas SteepDescent versus Newton- Pt2

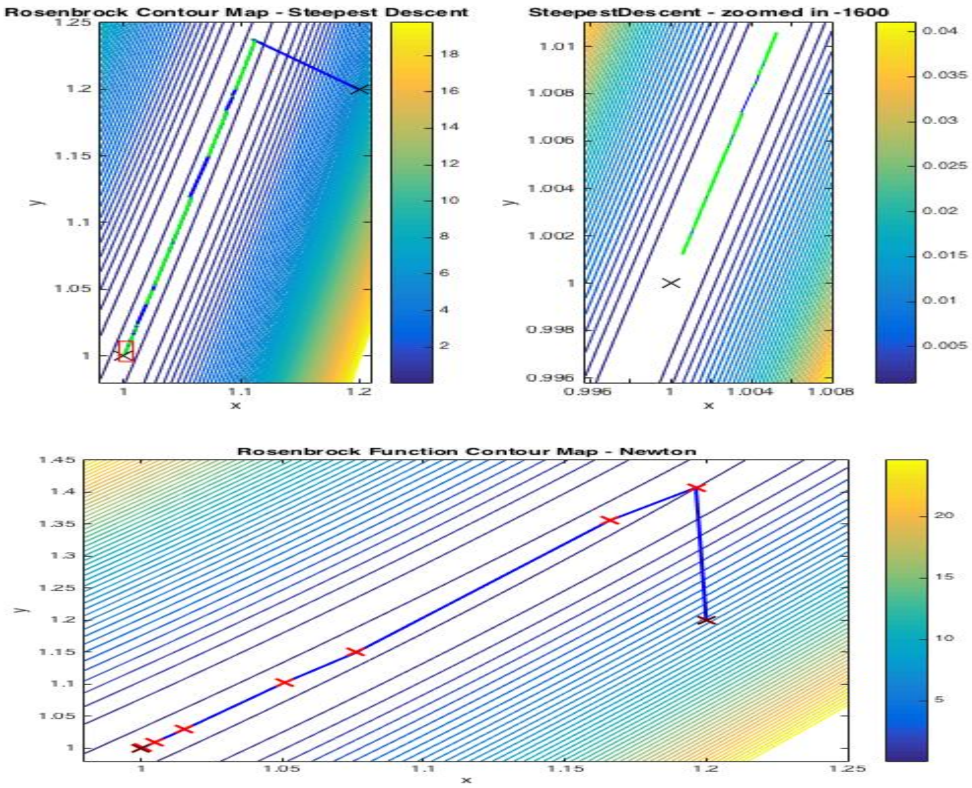
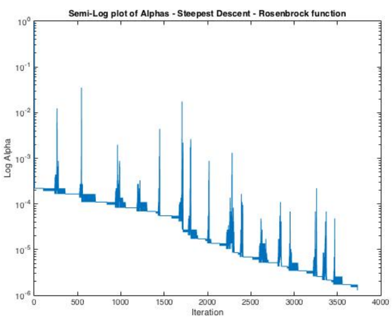
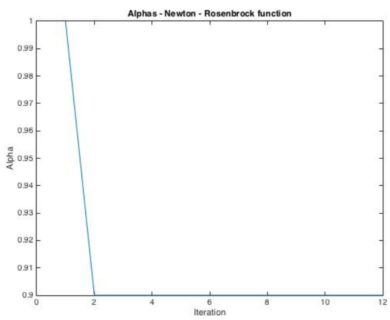


Figure 5: Strong Wolfe Steepest Descent v Newton pt1



(a) Semi-Log Alphas v Its SteepDesc



(b) Alphas v Iterations Newton

Figure 6: Strong Wolfe - Alphas pt1

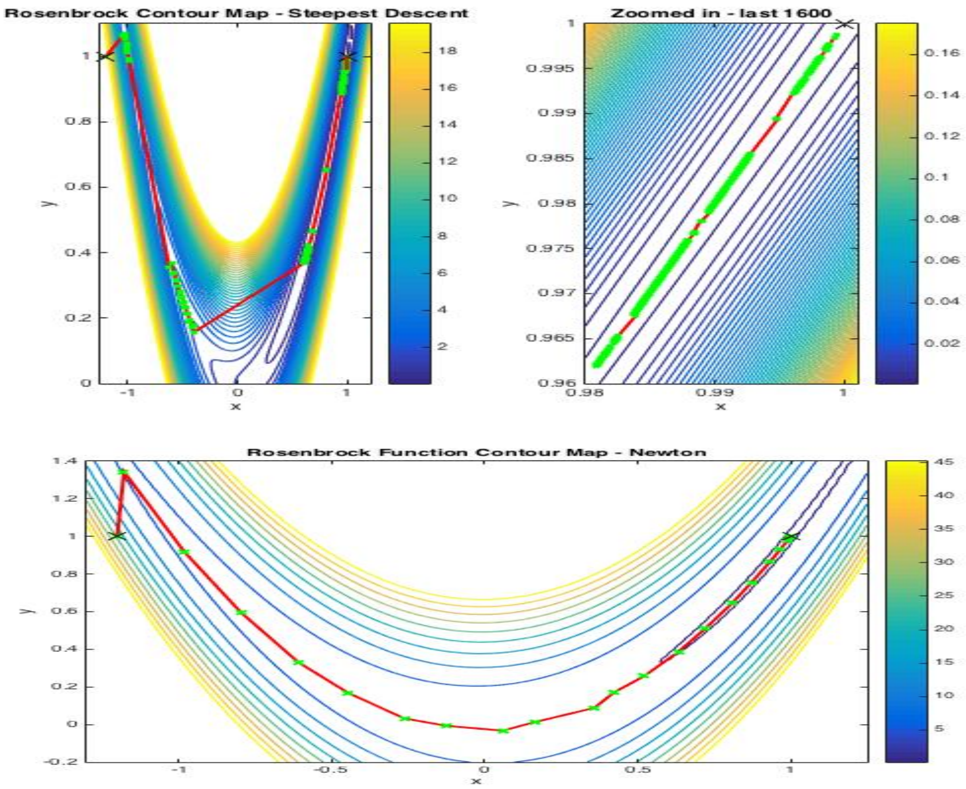
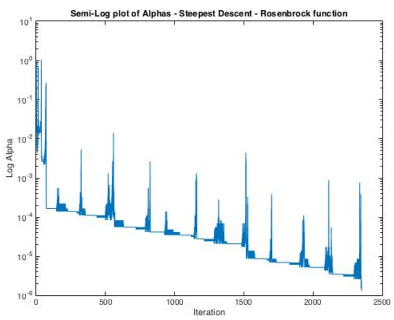
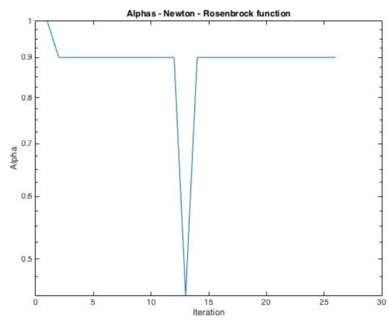


Figure 7: Strong Wolfe Steepest Descent v Newton pt2



(a) Semi-Log Alphas versus Its SD pt2



(b) Alphas versus Its Newton pt2

Figure 8: Strong Wolfe - Alphas point 2

2 Summary

Apologies for going over the space requirement for the assignment but I wish to give a brief summary of what is going on and so far its all images and no explanation.

2.1 Point one versus Point two

Starting point one is (1.2,1.2), point two is (-1.2,1.0). As can be seen on the contour plots point one is a far easier starting point than point two. Thinking geometrically one can imagine point one as being in the same valley up a hill from the optimal point (1.0,1.0), whereas point two is somewhere up another valley with a big mountain between itself and the optimal point. One can see this by comparing the number of **iterations** to convergence for each method starting from point one and two.

Point One

- Steepest Descent - Backtrack $\approx$ 6000.
- Newton - Backtrack $\approx$ 12.
- Steepest Descent - Strong Wolfe $\approx$ 4000.
- Newton - Strong Wolfe $\approx$ 12.

Point Two

- Steepest Descent - Backtrack $\approx$ 9000.
- Newton - Backtrack $\approx$ 50.
- Steepest Descent - Strong Wolfe $\approx$ 2500.
- Newton - Strong Wolfe $\approx$ 25.

2.2 Newton verses Steepest Descent

The biggest difference in this optimisation problem was between steepest descent and the newton method. This is regardless of the method used to choose step size and regardless of the starting point. This can be seen by the fact I had to use a semi-log scale for alpha versus iterations for steepest descent and that these alphas are sometimes orders of magnitude smaller than the Newton method.

For those who are visual it can also be seen by looking at fig 1, as an example. Using steepest descent the method starts with a blue line, but then quickly becomes a blur of green iterations. Each iteration being a tiny step.
Why is this? Geometrically what is happening is that at each iteration we are

effectively hopping from one side of the valley to another whilst making very little progress along the valley to the final optimum. In contrast the newton method shows long blue lines, with sporadic red X's at each iteration.
The key reason for this is that the newton method is a second order method and thus takes into account not just the steepest path but the curvature at that point. This would work extremely well in a quadratic bowl for example.

A final contrast is again the quick progress the newton method makes compared to steepest descent when close to an optimum, compare to the top right zoomed in plots of steepest descent, where it may still be several thousand iterations away from finishing.

2.3 Backtracking versus Strong Wolfe

I have really run out of space, but observe that the stronger conditions for strong Wolfe resulted in fewer iterations in this example to reach an optimum. Clearly there are dependencies here. The rho for backtracking used in this example was an arbitrary 0.1. This could of course be varied.