Assignment 2

Callum Lau, 19102521

Numerical Optimisation

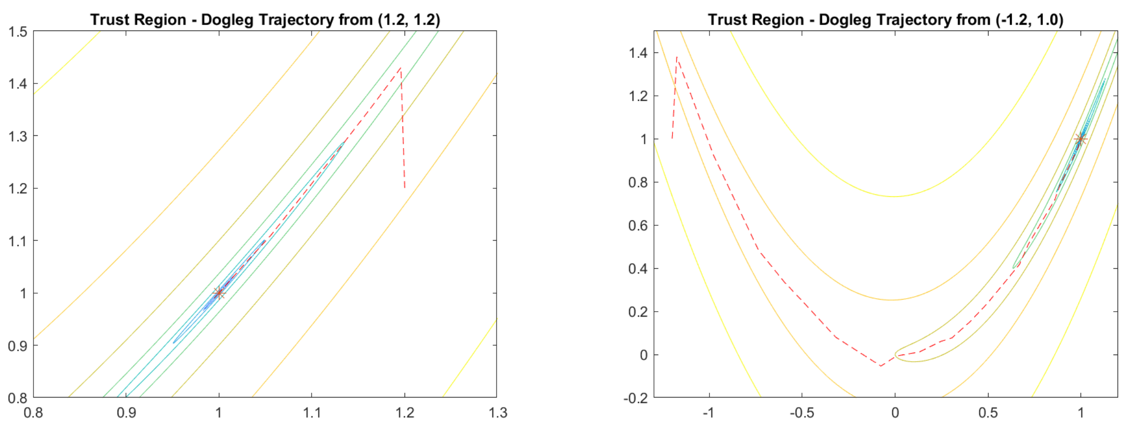
February 18, 2020

1 Exercise 5 - Trust Region Methods

The **Dogleg** quadratic solver was used in the following experiments. The initialisation for the parameters from both starting points $\boldsymbol{x}_0 = (1.2, 1.2)$ and $\boldsymbol{x}_0 = (-1.2, 1.0)$ is given below:

Parameter values		
$\boldsymbol{x}_0$	$(1.2, 1.2)$	$(-1.2, 1.0)$
eta	0.1	0.1
maxIter	100	100
tol	1e-6	1e-6
Delta	0.35	0.5

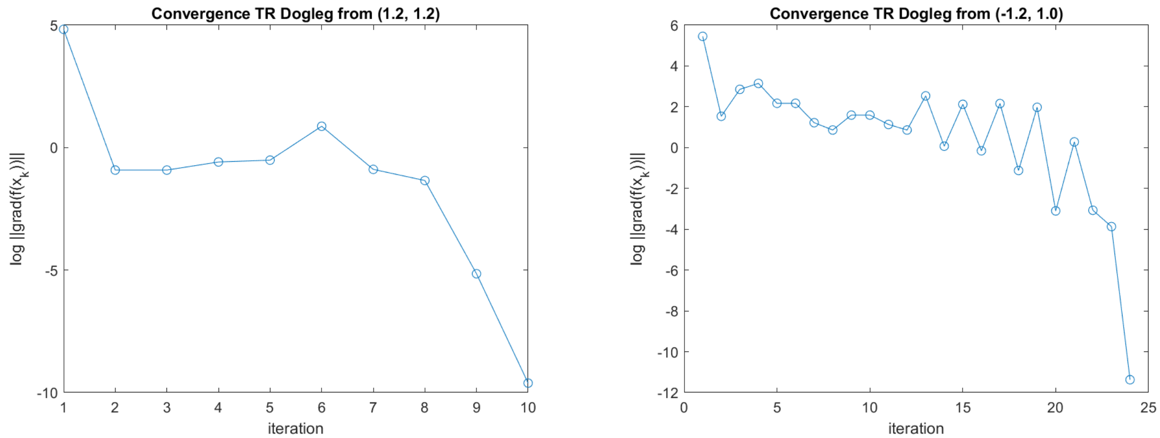
Furthermore the plots of the trajectories (with the Rosenbrock contours superimposed) are also provided:



(a) $\boldsymbol{x}_0 = (1.2, 1.2)$ (b) $\boldsymbol{x}_0 = (-1.2, 1)$

Figure 1: Dogleg trajectories

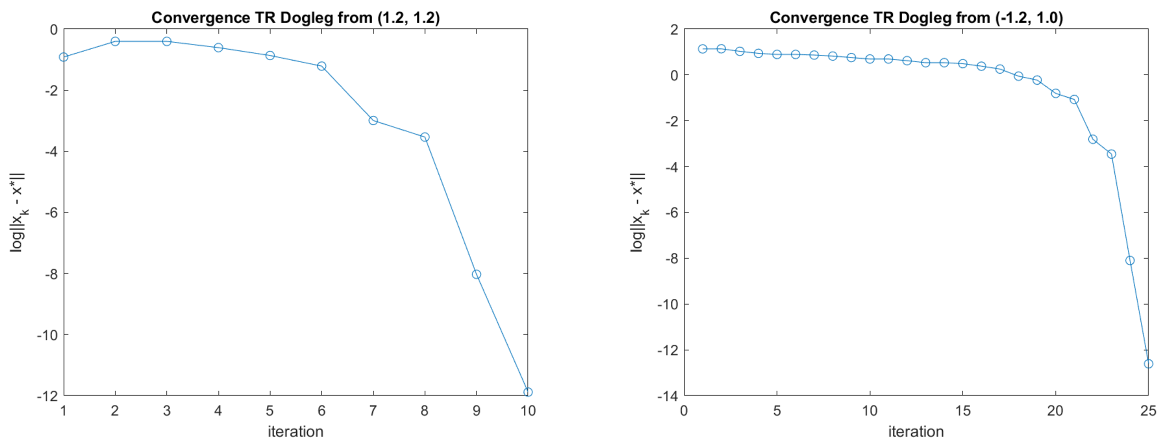
The convergence plots of $\log ||\nabla f(x_k)||$ from both starting points are shown below:



(a) $\boldsymbol{x}_0 = (1.2, 1.2)$ (b) $\boldsymbol{x}_0 = (-1.2, 1)$

Figure 2: Convergence of gradients

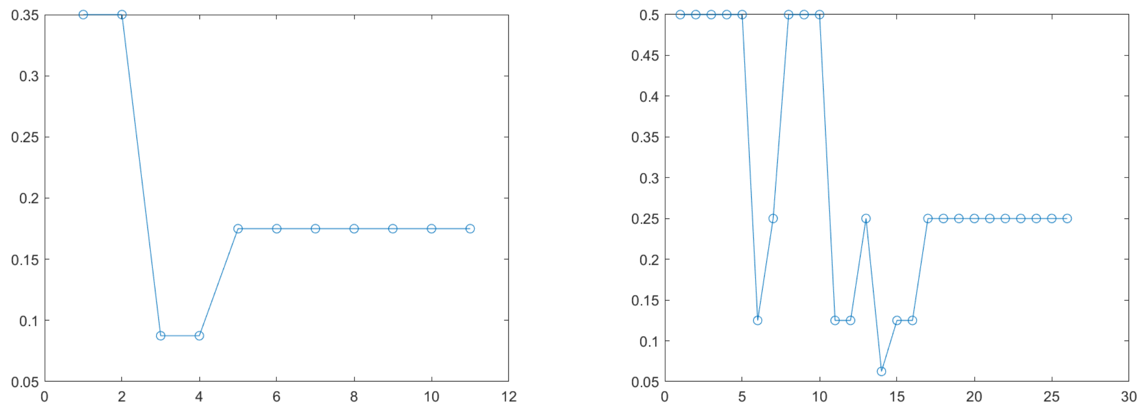
These plots conform to the theory presented in the lectures that for the Trust Region algorithm and approximate solutions p_k that satisfy the Cauchy Point sufficient reduction condition, then for $\eta \in (0, 1/4)$: $\lim_{k\to\infty} \nabla f(x_k) = 0$. This is because we observe that $\nabla f(x_k)$ does indeed decrease to zero, but also that the theory does not imply that $\nabla f(x_k)$ is monotonic decreasing - the plots show that this is the case - just that the values tend to zero. Furthermore, we plot below the convergence of the iterates $\log ||x_k - x^*||$ below:



(a) $\boldsymbol{x}_0 = (1.2, 1.2)$ (b) $\boldsymbol{x}_0 = (-1.2, 1)$

Figure 3: Convergence of iterates

We note that the convergence plots of the iterates show clearly the 'superlinear local convergence' property proved in the slides. Put loosely, for k sufficiently large and once the steps p_k become similar enough to Newton steps, the sequence x_k converges to x^* superlinearly. After iteration 6 from the starting point $x_0 = (1.2, 1.2)$ and after iteration 20 from the starting point $x_0 = (-1.2, 1.0)$ we note how the normed differences $||x_k - x^*||$ begin to decrease rapidly and hence exhibit this property.
Below we have plotted the trust region radii at each iteration:



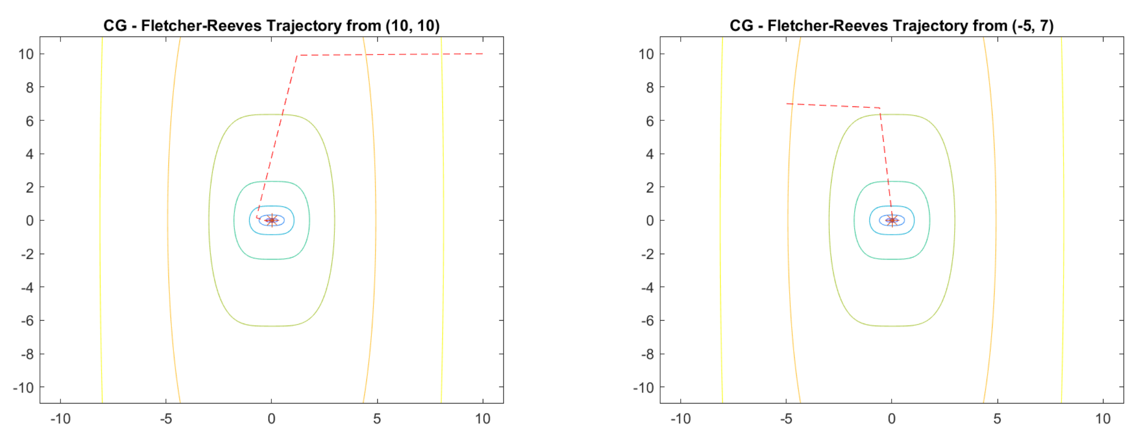
(a) $\boldsymbol{x}_0 = (1.2, 1.2)$ (b) $\boldsymbol{x}_0 = (-1.2, 1)$

Figure 4: Trust region radii

The trust region plots demonstrate the secondary point noted in the slide ‘superlinear local convergence’ - namely that for sufficiently large k and for p_k that satisfy the Cauchy point sufficient reduction criteria and are asymptotically similar to Newton steps, then the trust region bound Δ_k becomes inactive. This is because, after iteration 6 from the starting point $x_0 = (1.2, 1.2)$ and after iteration 20 from the starting point $x_0 = (-1.2, 1.0)$ (the same points with which we noted superlinear local convergence), the trust region radii remain constant until the iterates have converged to the minimum - i.e the bound becomes inactive for these sufficiently large k values.

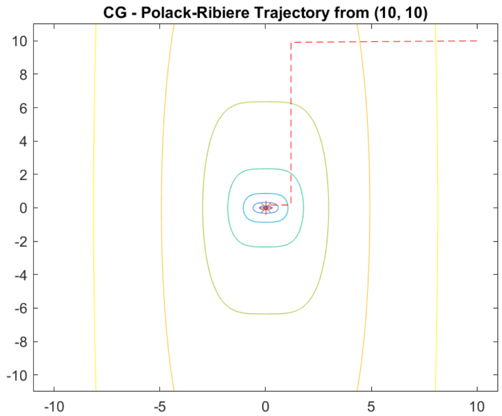
2 Exercise 7c.) - Conjugate Gradient Methods

The trajectories for the Fletcher-Reeves (FR) and Polack-Ribiere (PR) algorithms from both starting points are given below.

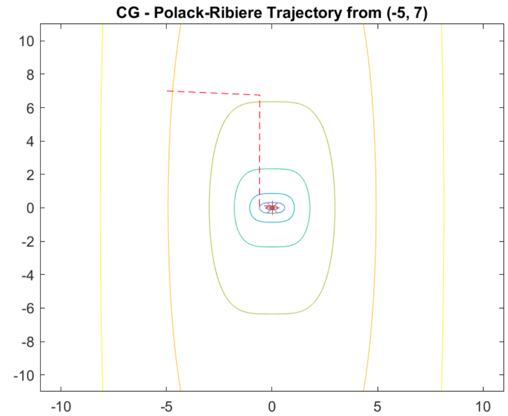


(a) $\boldsymbol{x}_0 = (10, 10)$ (b) $\boldsymbol{x}_0 = (-5, 7)$

Figure 5: Fletcher-Reeves trajectories



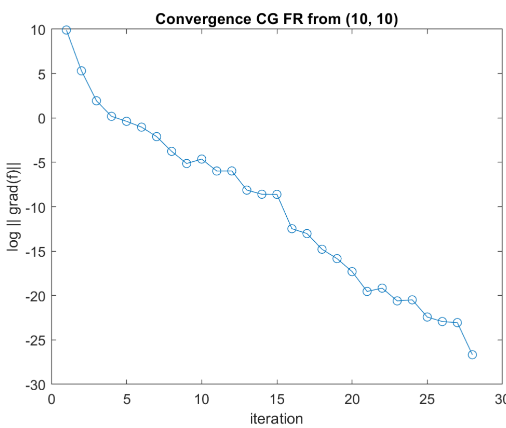
(a) $\boldsymbol{x}_0 = (10, 10)$



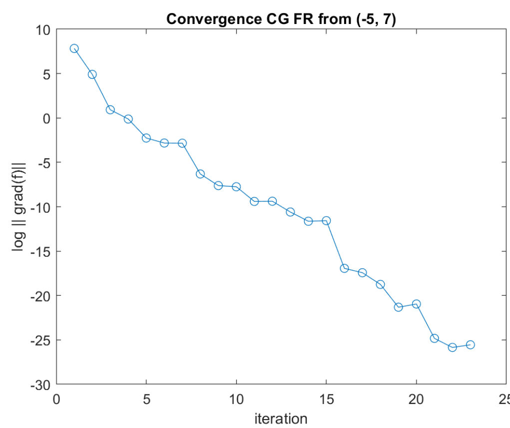
(b) $\boldsymbol{x}_0 = (-5, 7)$

Figure 6: Polack-Ribiere trajectories

Furthermore the convergence plots of the gradients for the FR algorithm and from both starting points are also given:



(a) $\boldsymbol{x}_0 = (10, 10)$



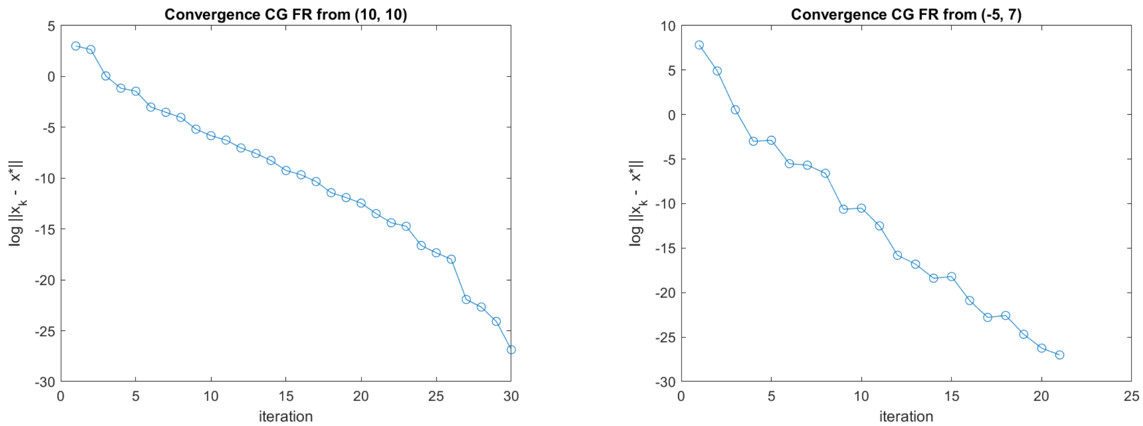
(b) $\boldsymbol{x}_0 = (-5, 7)$

Figure 7: Fletcher-Reeves convergence of gradients

One potential problem for the FR method is that if the line search for choosing step length α_k is not exact then p_k may fail to be a descent direction. This is suggested by the convergence graphs above since we observe that the norm of the gradients $||\nabla f_k||$ are not strictly decreasing. As suggested in the notes, this can be avoided by requiring that α_k satisfies the strong Wolfe conditions. Note that due to the theorem by Al-Baali, using the FR algorithm with line search satisfying the strong Wolfe conditions ensures global convergence:

$$\liminf_{k\to\infty} ||\nabla f_k|| = 0 \tag{1}$$

Below we also plot the convergence of the iterates for the FR algorithm and from both starting points:



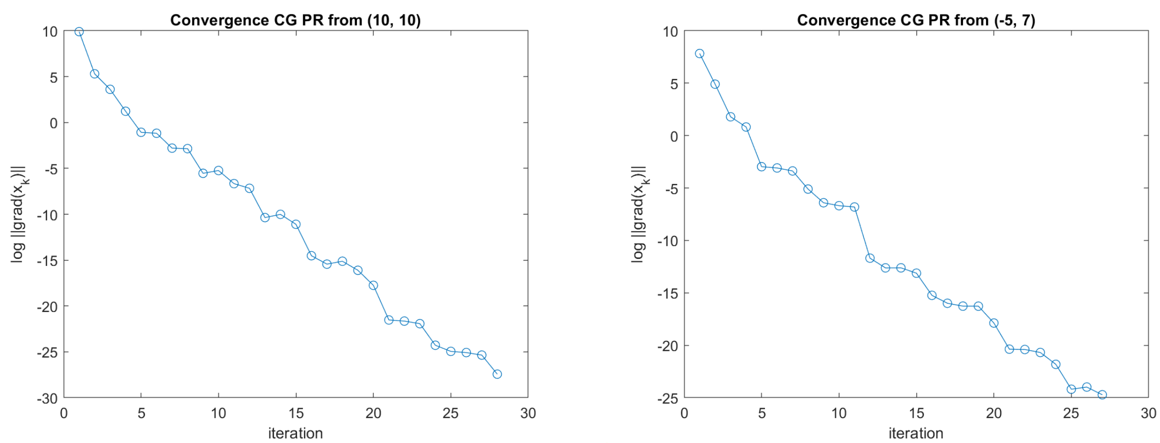
(a) $\boldsymbol{x}_0 = (10, 10)$ (b) $\boldsymbol{x}_0 = (-5, 7)$

Figure 8: Fletcher-Reeves convergence of iterates

Another potential problem for the FR algorithm is that if a bad direction and tiny step are generated then the next direction and step is likely poor. This is potentially seen be the very small distance between the some of the iterates in the graph of convergence above. Noting this in the case $f : \mathbb{R}^2 \to \mathbb{R}$ we see that this could become a major issue for larger dimensions. A solution to this problem follows the presentation in the notes, namely for a function $f : \mathbb{R}^n \to \mathbb{R}$ restarts of $\beta_k = 0$ every n steps serves to 'refresh the algorithm discarding old information that may not be beneficial'. In effect, when $\boldsymbol{x}_k$ is close enough to $\boldsymbol{x}^*$, by Taylors theorem f can be suitably approximated by a convex quadratic function and hence taking the steepest descent direction will enable the algorithm to behave like standard linear Conjugate Gradient, with finite termination within n steps from the restart. In fact, it is true that such restarting leads to n-step quadratic converge:

$$||\boldsymbol{x}_{k+n} - \boldsymbol{x}^*|| = O(||\boldsymbol{x}_k - \boldsymbol{x}^*||^2) \tag{2}$$

Finally we provide plots of the convergence of the PR algorithm from both starting points:



(a) $\boldsymbol{x}_0 = (10, 10)$ (b) $\boldsymbol{x}_0 = (-5, 7)$

Figure 9: Polack-Ribiere convergence

We note that the PR algorithm with step lengths satisfying the strong WC does not ensure that descent directions are generated. We could fix this by choosing $\beta_k = \max(0, \beta_k^{FR})$