Homework 2

Bastian Haase

1 Exercise 1

We will compare the performance of the steepest decent method to the quasi-Newton method in this section. For the quasi-Newton method, we will use three different techniques to modify the hessian if it is not positive definite.

The first technique that we will abbreviate with *fcol* is the one described in Algorithm 3.3 in our textbook of Nocedal and Wright. The main idea is to add a small multiple of the identity to the hessian until choleski factorization works.

The second technique, *spec*, analyzes the spectrum of the matrix. If the smallest eigenvalue λ is negative, we multiply the identity with a scalar $|\lambda| + \epsilon$ for small ϵ and add this to our hessian. We have chosen $\epsilon = 10^{-3}$ for this comparison.

Lastly, instead of modifying the hessian, we just try to solve the resulting system of linear equations with the conjugate gradient method, *cg*. In case the hessian is not positive definite, we still get an approximation of the solution of this system.

In all cases, we will determine the step length by use of a backtracking method starting with a length of 1 and dividing it by 2 if the Wolfe condition is not satisfied.

1.1 Setup and Parameters

We are considering the Rosenbrock function

$$f(x, y) = (1 - x)^2 + 100(y - x^2)^2$$

which attains its minimum at $(1, 1)$. The following table lists the parameters we have chosen.

Parameter	Tolerance	Max Iterations	Max Iterations Step Size	c_1
Value	10^{-6}	10^5	10^2	10^{-4}

Table 1: Parameters for Comparison

Here, *Max Iterations Step Size* denotes the maximum number of modifications of our step length at each step of our iteration. We will compare all four methods with a starting point where the hessian is spd and at a point where it is not. Additionally, we will determine the region of convergence by applying all methods to a 32×32 mesh of the set $[-2, 2] \times [-1, 3]$.

1.2 Hessian spd

We can see that the hessian of the Rosenbrock function at the point $(-1.9375, -0.9375)$ is positive definite, as the eigenvalues are both positive. The table illustrates the performance of the four methods with this starting point and the setup described above.

Method	**Iterations**	**Function Evaluations**	**Jacobian Evaluations**
steep	17914	122537	10001
fcol	25	83	26
spec	25	83	26
cg	25	83	26
Method	**Hessian Evaluations**	**Matrix Factorizations**	**Multiplications with Hessian**
steep	0	0	0
fcol	25	25	25
spec	25	50	25
cg	25	0	50

Table 2: Performance SPD

As we can see from the values in the table, the quasi-Newton methods all perform similarly when starting at a point with positive definite hessian that is close to the minimum. Contrary, the steepest descent method converges much slower, which can not be compensated by the avoidance of any computations with the hessian.

The figures illustrate how the norm of the gradient changes over the iterations.

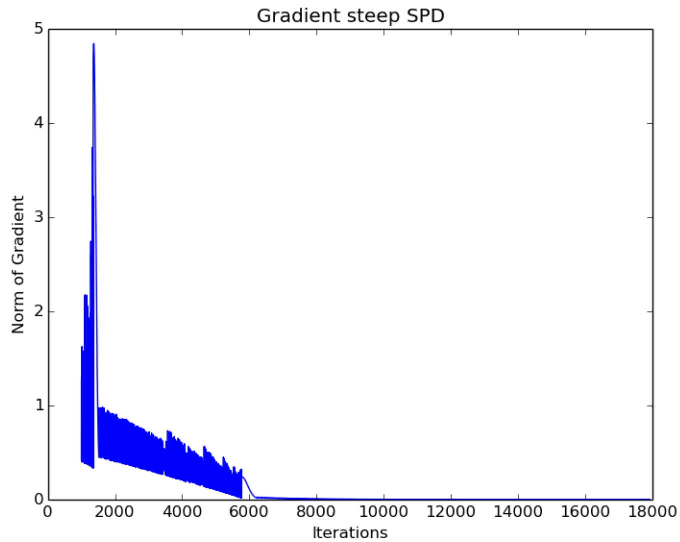

Figure 1: Gradient steep SPD

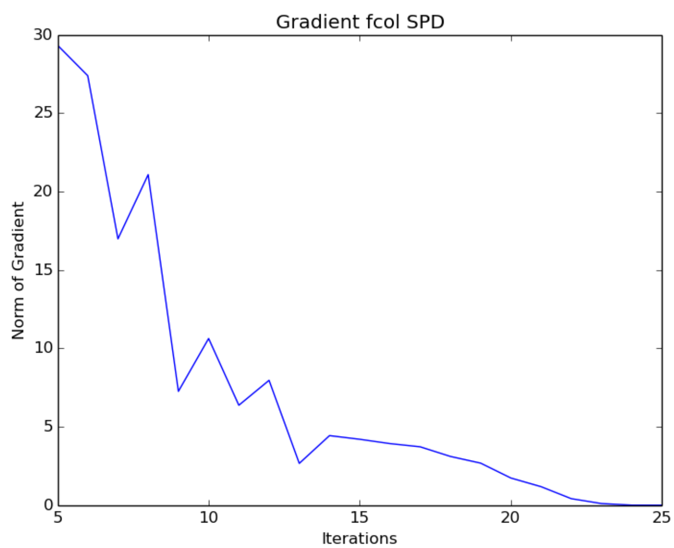

Figure 2: Gradient fcol SPD

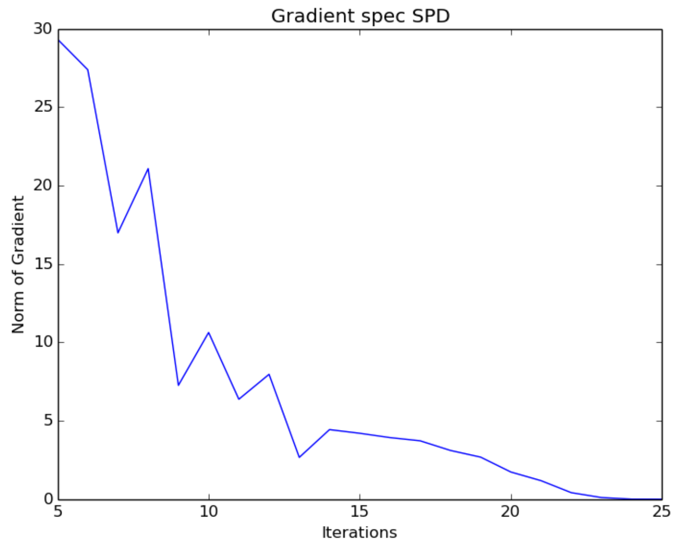

Figure 3: Gradient spec SPD

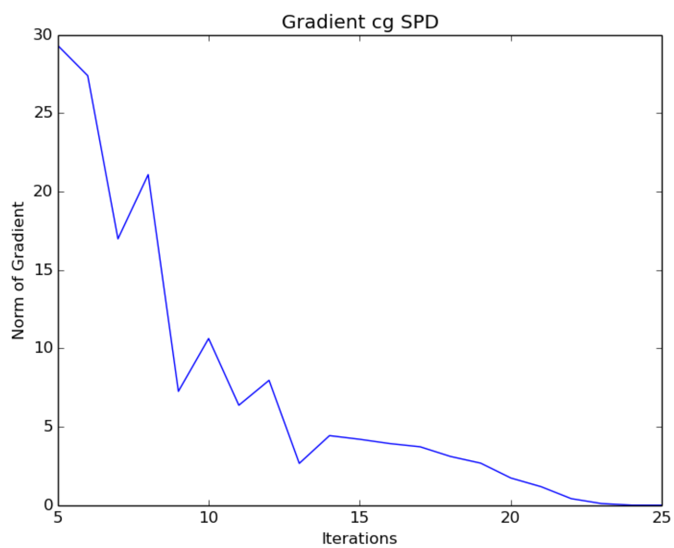

Figure 4: Gradient cg SPD

In this case, the figures support the conclusion of the last table, they show that the quasi-Newton methods perform similarly and they clearly outperform steepest descent.

1.3 Hessian not SPD

We can see that the hessian of the Rosenbrock function at the point $(-0.3125, 0.6875)$ is not positive definite, as one eigenvalue is negative. The table illustrates the performance of the four methods with this starting point and the setup described above.

Method	**Iterations**	**Function Evaluations**	**Jacobian Evaluations**
steep	13863	165561	13864
fcol	8	26	9
spec	9	46	10
cg	7	23	8
Method	**Hessian Evaluations**	**Matrix Factorizations**	**Multiplications with Hessian**
steep	0	0	0
fcol	8	13	8
spec	9	18	9
cg	7	0	12

Table 3: Performance SPD

As we can see from the values in the table, the quasi-Newton methods all perform similarly. But, you can still notice that the biggest advantage of the conjugate gradient method is the fact that no matrix factorizations are needed. On the downside, the number of matrix-vector multiplications is higher and we will later see that region of convergence is smaller than the one of the other methods. The fcol method has the smallest number of matrix vector multiplications but a higher number of matrix factorizations. The spec method has as its biggest downside the higher number of factorization. This is due to fact that the matrix was factored to determine the eigenvalues and to solve the linear system of equations. Note that this could be done (by not using cholfact to solve the system) with one factorization in the case when the hessian is SPD. Yet again, the performance of the steepest descent method is not competitive due to the slow convergence.

The figures illustrate how the norm of the gradient changes over the iterations.

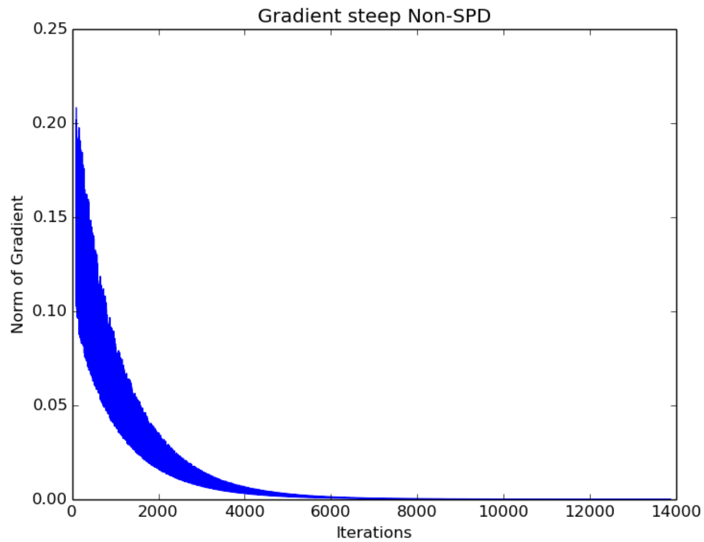

Figure 5: Gradient steep Non-SPD

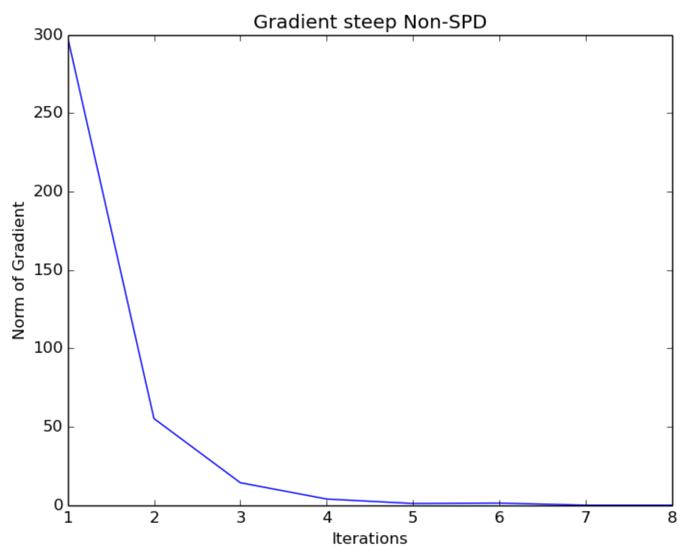

Figure 6: Gradient fcol Non-SPD

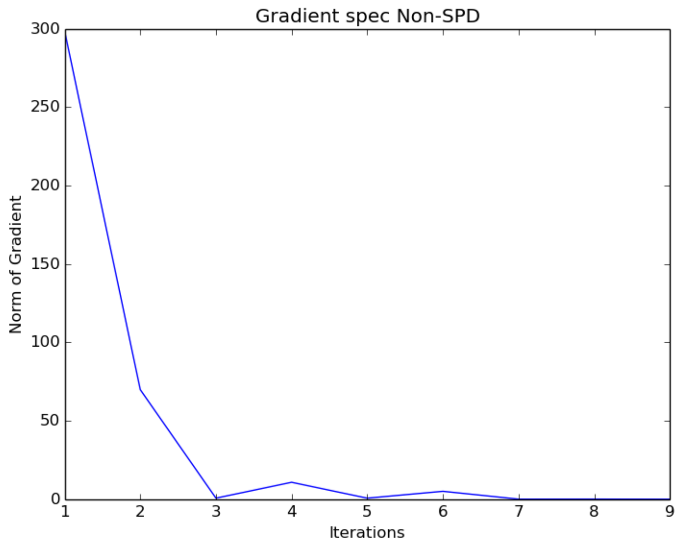

Figure 7: Gradient spec Non-SPD

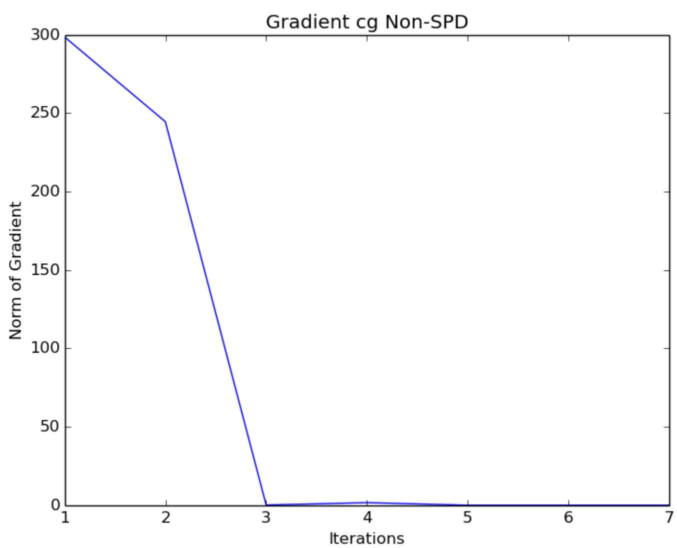

Figure 8: Gradient cg Non-SPD

Yet again, the figures support the conclusion of the last table, they show that the quasi-Newton methods perform similarly and they clearly outperform steepest descent.

1.4 Region of Convergence

The following plots illustrate the region of convergence. In these plots, you can see the convergence in the region $[-2, 2] \times [-1, 3]$ by creating a 32×32 equidistant mesh on it. The plot indicates the region and the numbering on the axis has to be understood relatively. The point $(3, 4)$ is not representing the point $(3, 4)$ in the cartesian plane but the corresponding point in the mesh.

All methods converge everywhere but cg. The region where it does not converge is white, while black represents convergence. This is a strong disadvantage of this method especially considering that these points are relatively close to the solution.

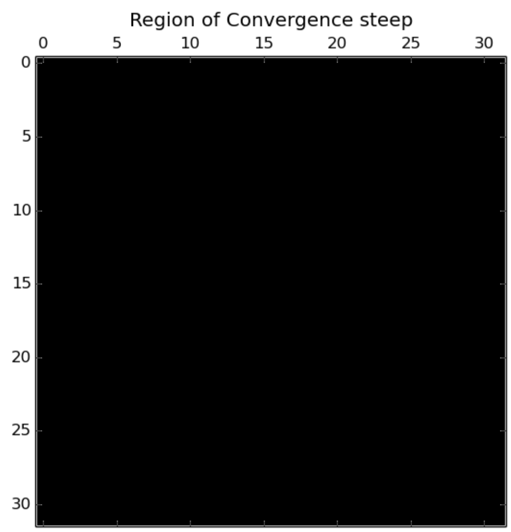

Figure 9: Gradient steep Non-SPD

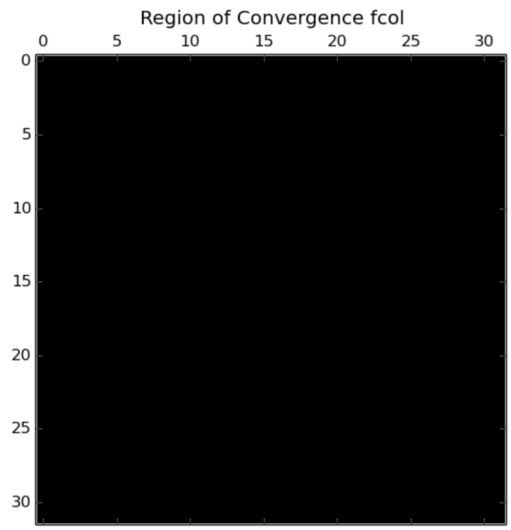

Figure 10: Gradient fcol Non-SPD

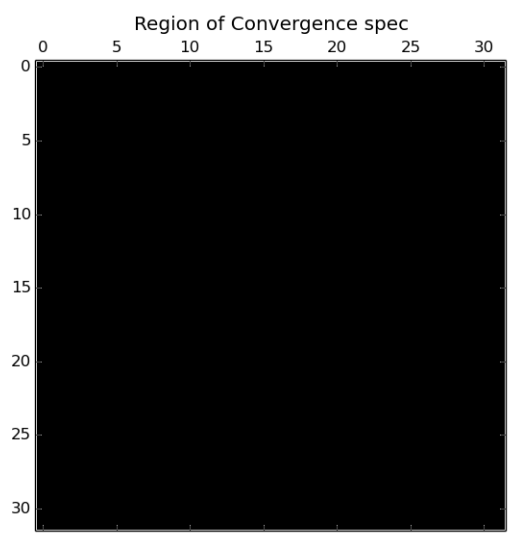

Figure 11: Gradient spec Non-SPD

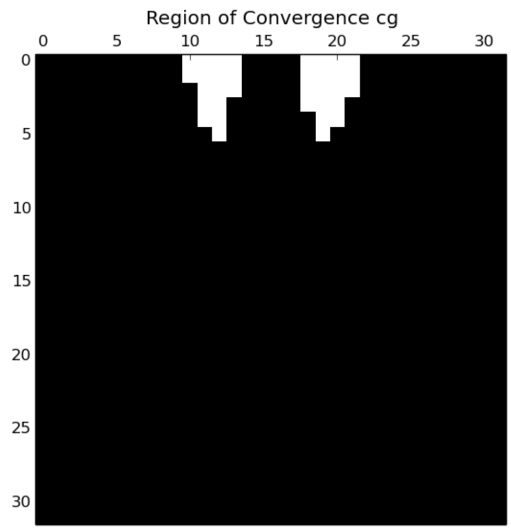

Figure 12: Gradient cg Non-SPD

2 Exercise 2

In this exercise we apply all four methods to the example from last homework. I will continue to use the notation from last homework and not recall the details given there for the sake of brevity. In order to apply all methods, we first have to determine the hessian.

The following plot indicates that the hessian that we determined is correct. Its formula is not relevant for further discussions so we do not include it here.

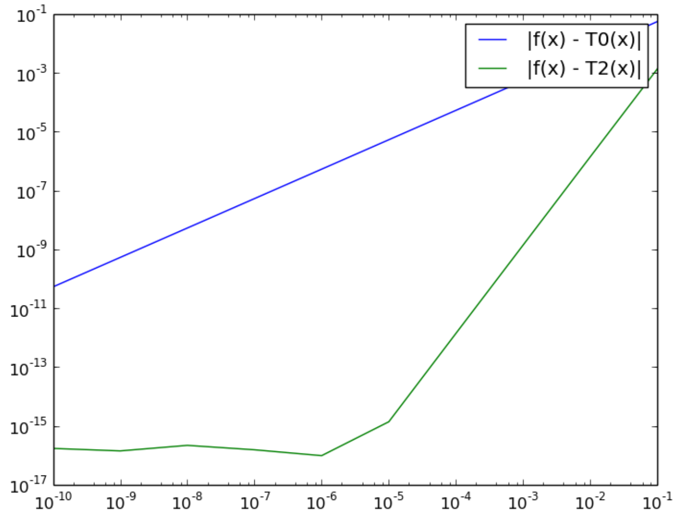

Figure 13: Check Hessian

We will now apply all four methods to the function where we have set $\alpha = 2$ and $\beta = 5$. The setup is identical to the one in exercise 1. The hessian at our starting point is not spd, in fact exactly one of the 18 eigenvalues is negative.

The following table illustrates ones more that the convergence of steepest descent is not competitive. Furthermore, fcol performs much weaker than the other two methods. Note that at our starting point, the matrix has only one negative eigenvalue and 17 positive ones. So, in some sense, it is close to being spd. It seems plausible that spec and cg can take advantage of this situation better as cg's approximation of a solution will be accurate and spec modifies the matrix more efficiently than fcol in this case. The high number of matrix factorizations that fcol has to perform supports this. The following plots show that all methods converge to the same solution.

Method	Iterations	Function Evaluations	Jacobian Evaluations
steep	331	994	332
fcol	17	55	18
spec	4	17	5
cg	5	17	6
Method	**Hessian Evaluations**	**Matrix Factorizations**	**Multiplications with Hessian**
steep	0	0	0
fcol	17	31	17
spec	4	8	4
cg	5	0	45

Table 4: Performance SPD

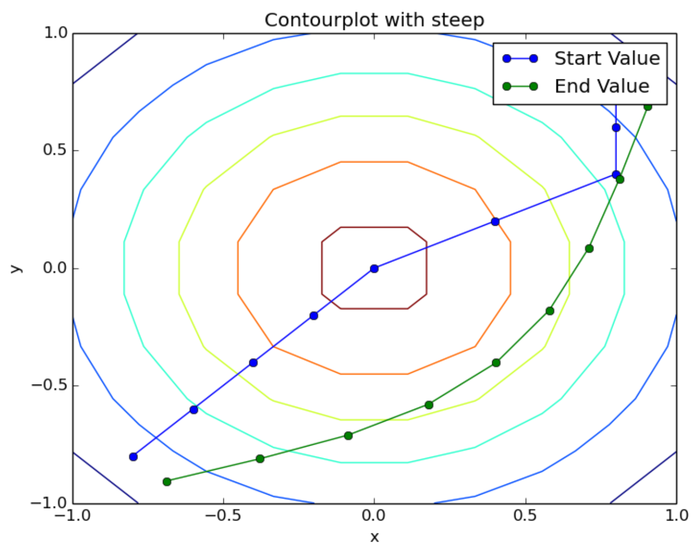

Figure 14: Convergence steep

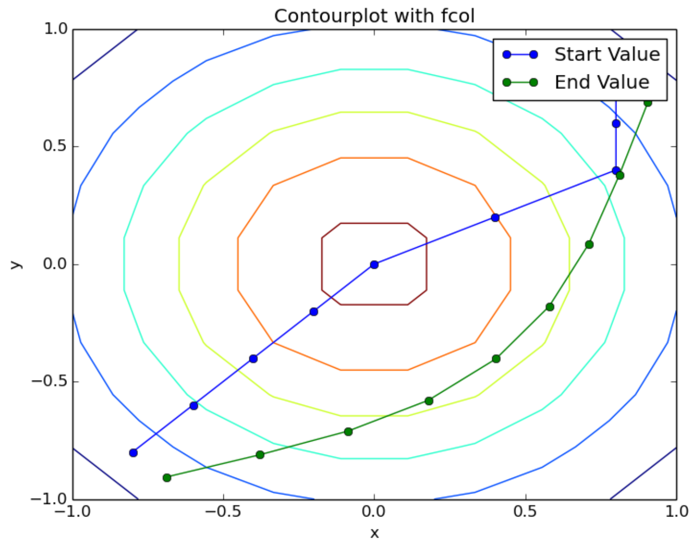

Figure 15: Convergence fcol

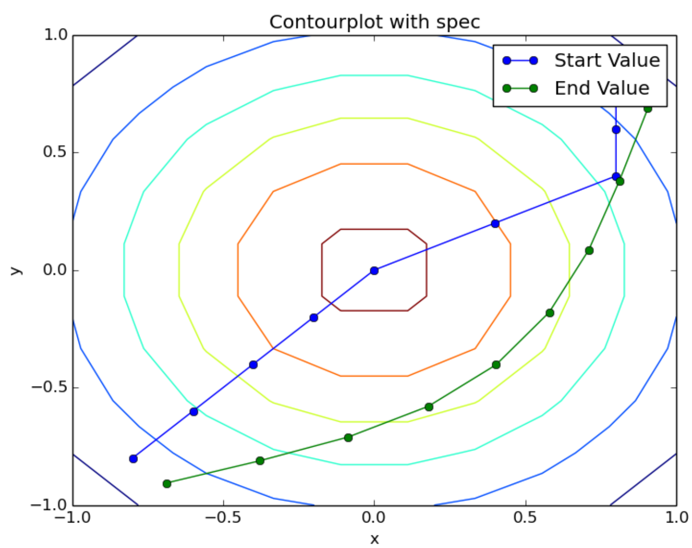

Figure 16: Convergence spec

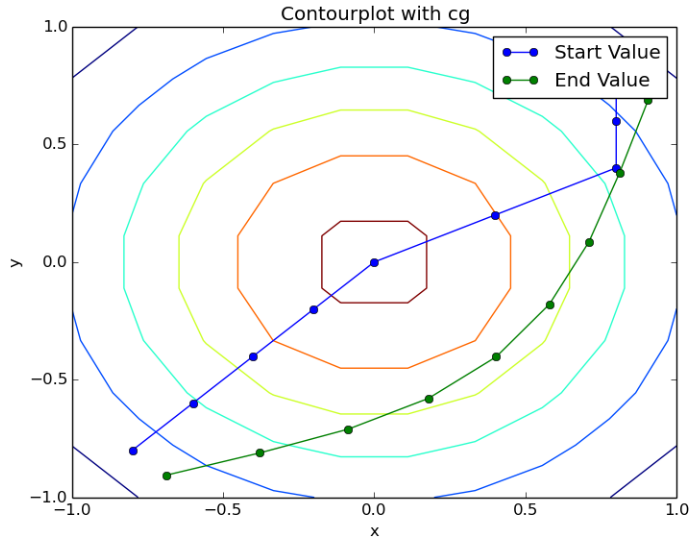

Figure 17: Convergence cg

The following figures now show how the norm of the gradient changes over each iteration.

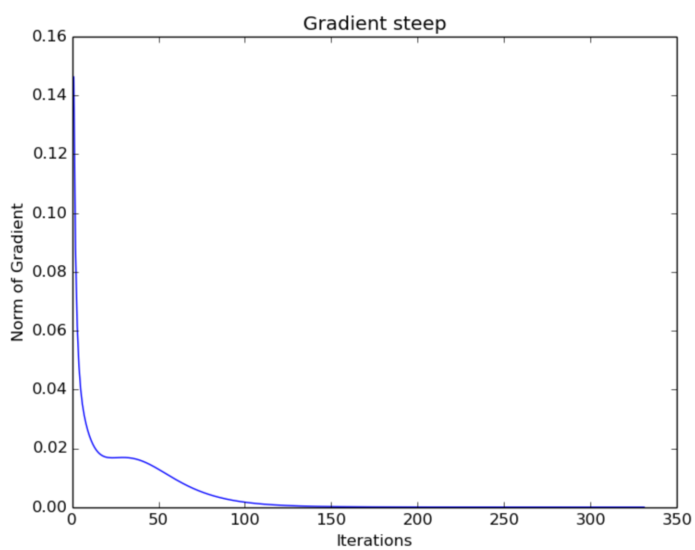

Figure 18: Gradient steep

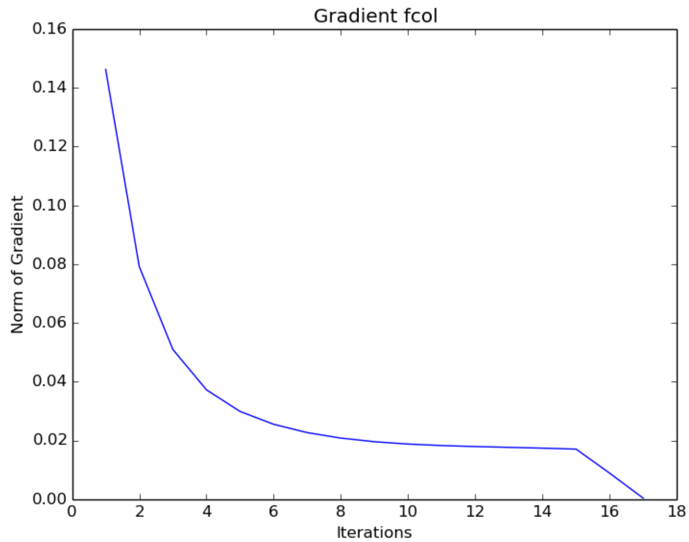

Figure 19: Gradient fcol

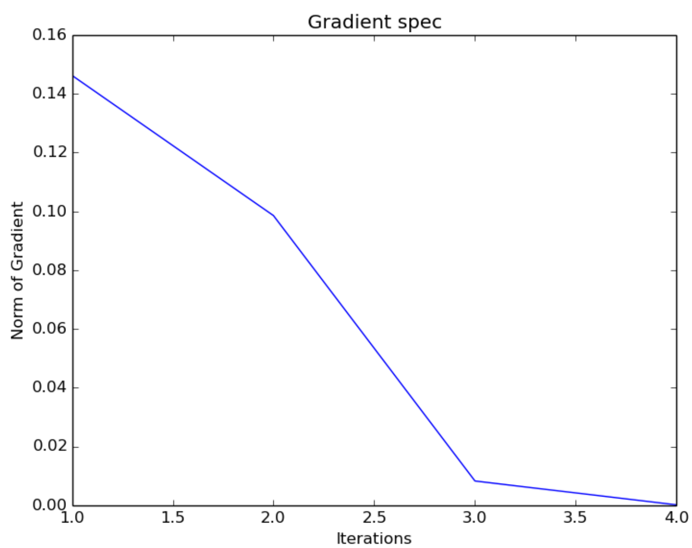

Figure 20: Gradient spec

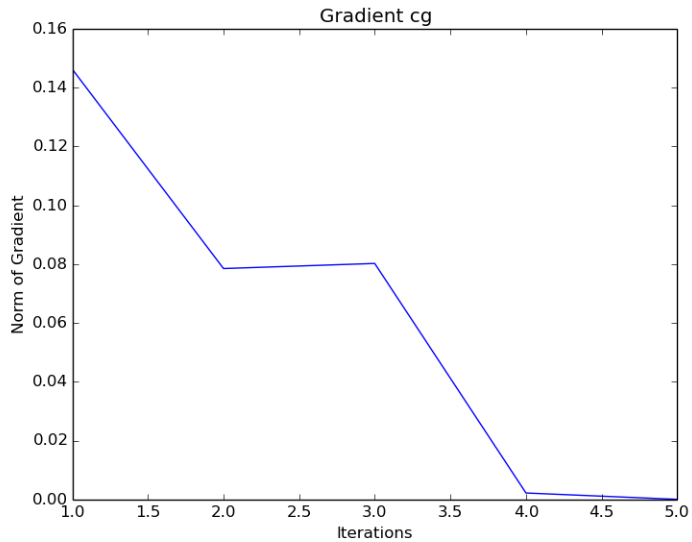

Figure 21: Gradient cg

As the table and these figures suggest, all quasi-Newton methods are superior to the steepest descent method. In this particular example, the spec method seems to perform the best.