Homework II
Restart GMRES(m)

Moritz Wolter

9, December, 2015

1 Abstract

When solving $\|\mathbf{Ax}-\mathbf{b}\|$ for very large $\mathbf{A}$ the memory requirements do not allow it to let the `GMRES` algorithm iterate until convergence, as memory and computational cost increases with each step. Therefore a modified version of `GMRES` called restarted `GMRES(m)` has been proposed by Saad and Schultz in their original paper.[1] After a fixed amount of iterations `m` the algorithm is restarted using $\mathbf{x}_m$ as initial condition for the next run. This changed version avoids memory problems at the cost of convergence. One would expect that the larger `m` is chosen the better the convergence properties of the restarted algorithm would be, since without restart when the largest possible `m` is chosen convergence is guaranteed. However this is not always the case. In this report curious behavior of `GMRES(m)` will be investigated following the paper "The Tortoise and the Hare, Restart GMRES" by Marc Embree.[2]

2 Results

2.1 `Gmres(m)` convergence on a larger random matrix

Before the results from the paper are considered `gmres(m)` will be run on a random matrix of dimension 60, the restart parameter will be varied and the effects on convergence observed. Results are shown in figure 1. In the plot a large spice is observed. In fact `gmres(34)` seems to have better convergence properties, then `gmres(35)`.

2.2 Experiments done in the paper

Embree considers two examples in his paper, the first one using

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} 2 \\ -4 \\ 1 \end{pmatrix}. \tag{1}$$

[1] A Generalized minimal residual algorithm for solving nonsymmetric linear systems, SIAM J. Sci. Stat. Comput. Vol 7, No.3 1986

[2] SIAM Rev., 45(2), 259266. (8 pages) The Tortoise and the Hare Restart GMRES, Marc Embree, `http://epubs.siam.org/doi/abs/10.1137/S003614450139961`

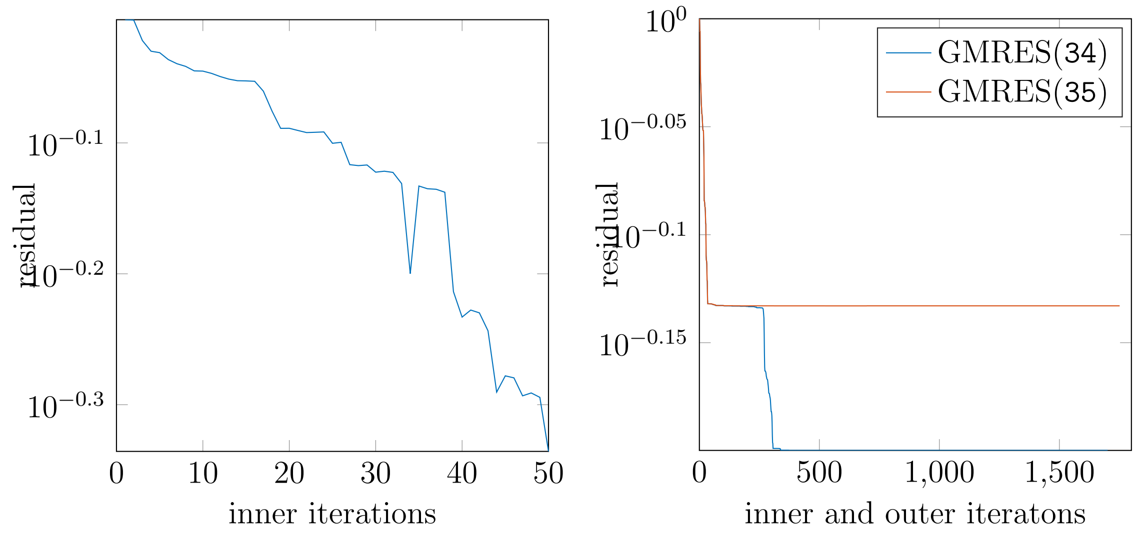

Figure 1: Convergence plot of `gmres(m)` for various different `m`. On the right the residuals of `m = 34` and `m = 35` are shown for comparison.

The second being

$$\mathbf{B} = \begin{pmatrix} 1 & 2 & -2 \\ 0 & 2 & 4 \\ 0 & 0 & 3 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} \tag{2}$$

Using series expressions as provided in the paper the convergence plots for `GMRES(1)` and `GMRES(2)` have been computed, the results shown in figure 2 resemble plots one and three in the paper. However to truly understand the gravity of the phenomenon $\mathbf{b}$ has been replaced with

$$\mathbf{r_0} = \begin{pmatrix} \xi \\ \eta \\ 1 \end{pmatrix} . \tag{3}$$

If ξ and $\eta \in [-10, 10]$ the plots shown in figures 3 and 4. Can be computed, which are the same as figure 2 and 4, if the different coloring is neglected. These plots show the residual of the two schemes after 30 iterations on a logarithmic scale. Blue areas indicate convergence.

2.3 Additional experiments

If $A - 0.35 * \mathbf{I}$ is considered the condition number changes from $\kappa(\mathbf{A}) = 14.2950$ to $\kappa(\mathbf{A} - 0.35 * \mathbf{I}) = 39.1873$. At the same time the convergent area of `GMRES(1)` shown in figure 5, shrinks considerably in comparison to the plot 3. Additionally as the conditioning worsened the white foggy areas without convergence grew significantly.

In a second series of experiments, designed to further investigate the effect of changed conditioning, $\mathbf{A}$ will be filled with entries drawn from the standart normal distribution ($\mu = 0$, $\sigma^2 = 1$). The values turned out to be:

$$\mathbf{R} = \begin{pmatrix} -0.3034 & 0.8884 & -0.8095 \\ 0.2939 & -1.1471 & -2.9443 \\ -0.7873 & -1.0689 & 1.4384 \end{pmatrix} \tag{4}$$

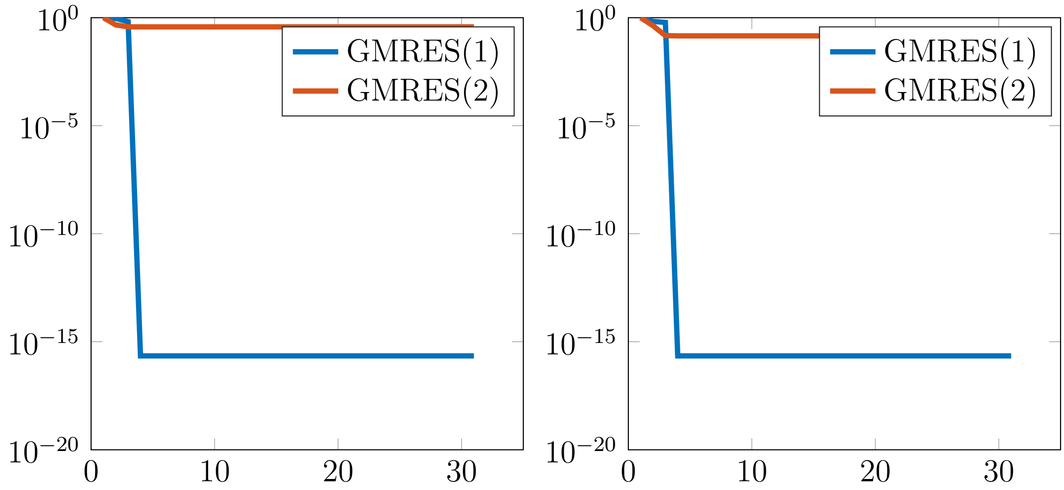

Figure 2: Convergence plot of GMRES(1) and GMRES(2). For matrix **A** (left), as seen in figure 1 of Embree's paper and for matrix **B** (right), as shown in figure three of Embree's paper.

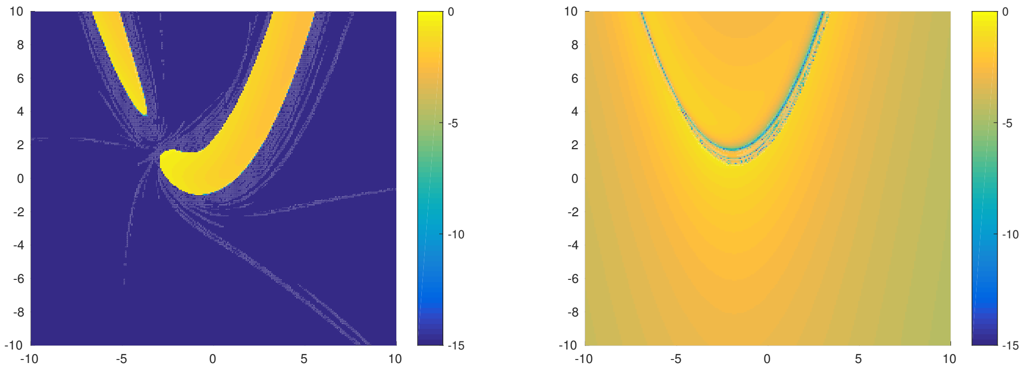

Figure 3: Residual magnitude plot for matrix **A**, as seen in the Embree paper GMRES(1)(left) and GMRES(2)(right) in figure two.

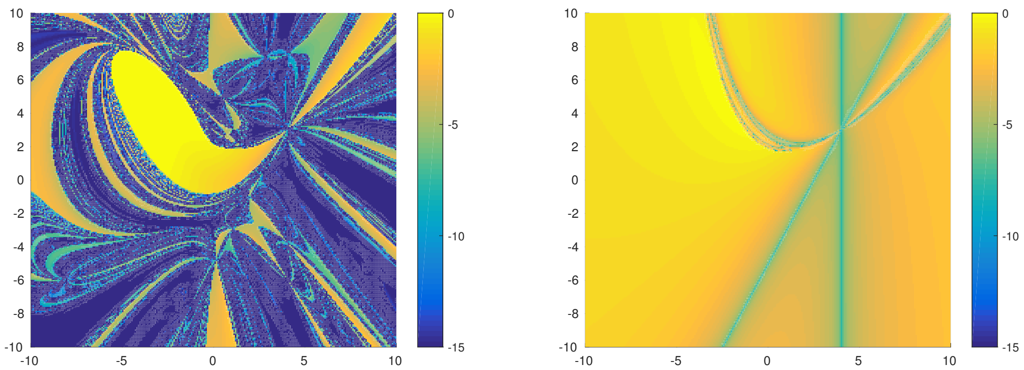

Figure 4: Residual magnitude Plot for matrix **B**, as seen in the Embree paper for GMRES(1)(left) and GMRES(2)(right) in figure 4.

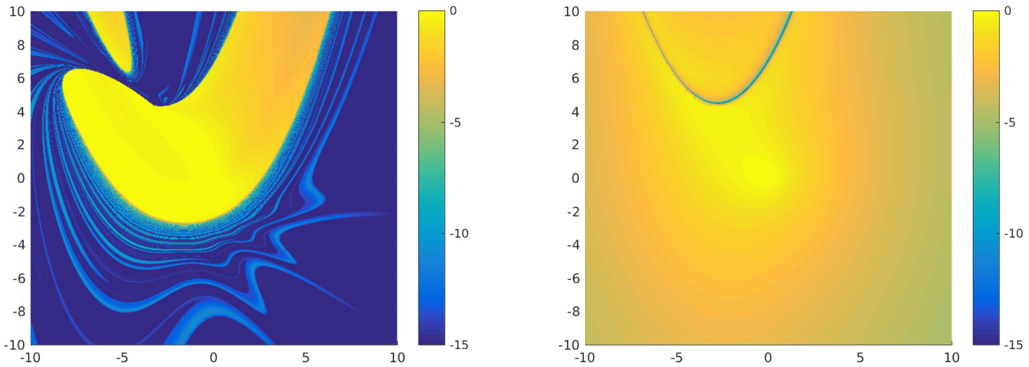

Figure 5: `GMRES(1)` and `GMRES(2)` convergence on $\mathbf{A} - 0.35 \cdot \mathbf{I}$.

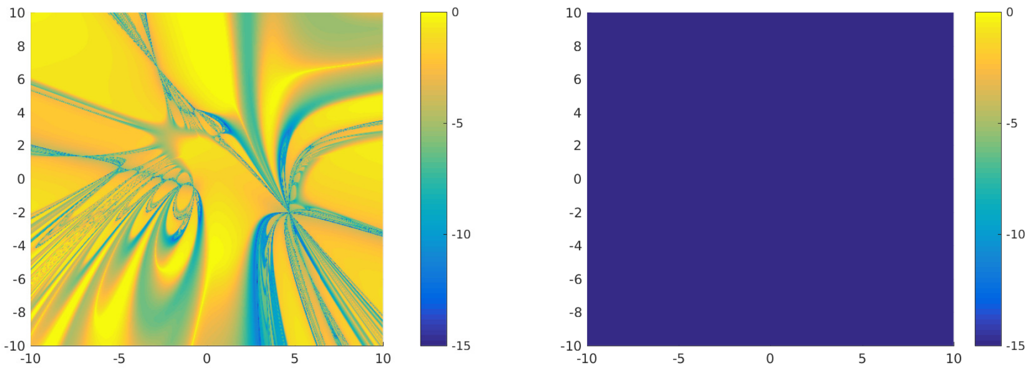

Figure 6: Convergence of `GMRES(1)`(left) and `GMRES(2)`(right) for the random matrix $\mathbf{R}$.

Results for `GMRES(1)` and `GMRES(2)`. Are shown in figure 6. Here it can be observed, that `GMRES(2)`, converges for all possible right hand sides $\mathbf{b}$, while `GMRES(1)` does not. This is probably the more common, but mathematically less interesting case. Often convergence improves significantly if the identity matrix is added to A. Unfortunately the condition number got worse by adding one multiple of the identity. It increased from $\kappa(\mathbf{R}) = 5.1025$ to $\kappa(\mathbf{R} + \mathbf{I}) = 32.4770$. Adding two times the identity matrix makes matters even worse as $\kappa(\mathbf{R} + 2\mathbf{I}) = 77.5446$, with results shown in figure 7. Finally adding three times the identity leads to $\kappa(\mathbf{R} + 3\mathbf{I}) = 6.4262$ and complete convergence for both algorithms figure 8. However the conditioning is still worse then it was originally, which indicates that conditioning alone cannot be used to predict `GMRES(m)` convergence.

2.4 Two-Dimensional case

At the end of the paper Embree proposes to take a closer look at:

$$\mathbf{C} = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} \quad \mathbf{r_0} = \begin{pmatrix} \xi \\ \eta \end{pmatrix} \tag{5}$$

The residuals show noisy convergence patterns in the bottom and top corners for `GMRES(1)` in figure 4. On the right side seemingly random patterns are shown at larger magnification. Similar ones appeared in figure 4 when zooming in. No plot exists for `GMRES(2)` is it always

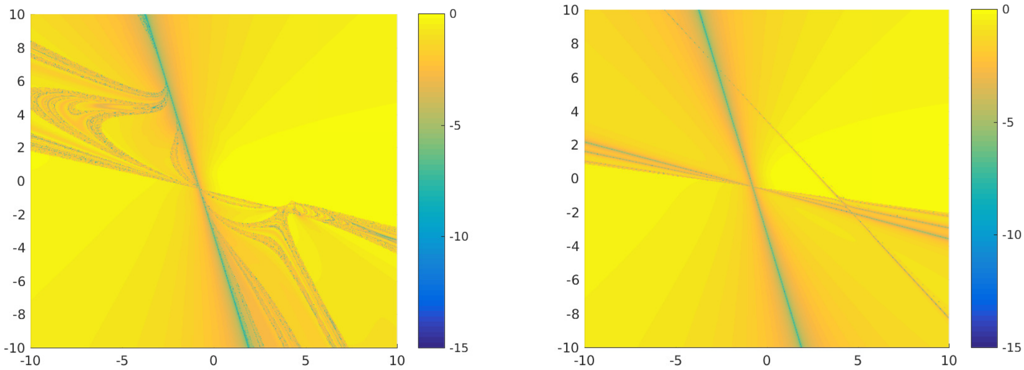

Figure 7: Convergence results for $\mathbf{R} + 2\mathbf{I}$

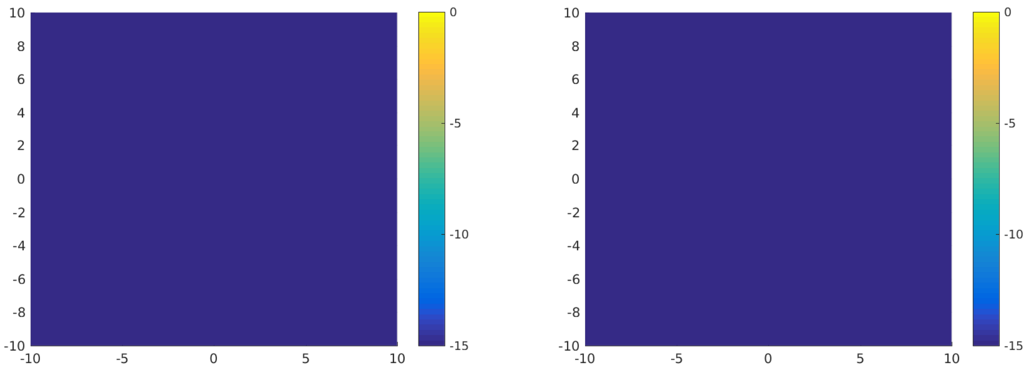

Figure 8: Convergence results for $\mathbf{R} + 3\mathbf{I}$

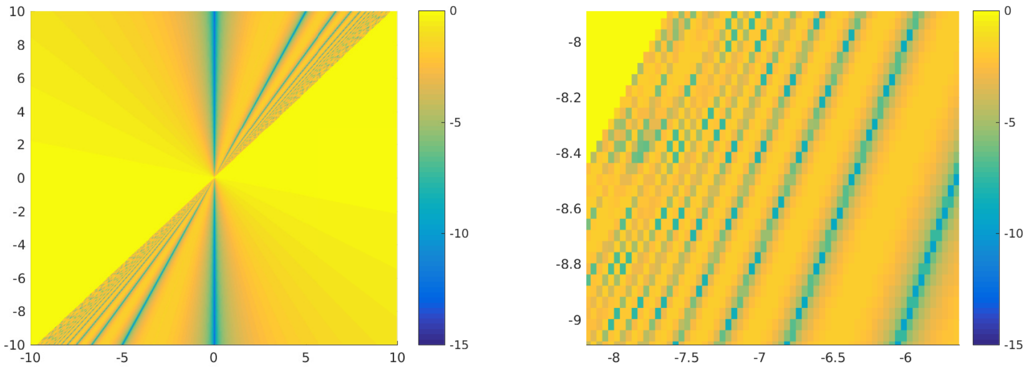

Figure 9: A plot of the convergence plain for the matrix **C** proposed at the end of Embree's paper. With zoom on an interesting region.

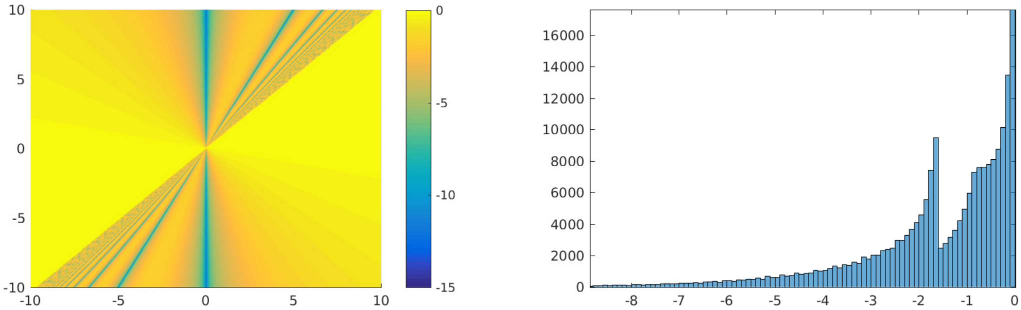

Figure 10: Convergence plain of matrix **C** along with a histogram of the logarithm of the residual norm.

converged completely within two iterations. In fact for all of the approximately ten two by two matrices that where tried did `GMRES(2)` converge to the exact result.

2.5 An attempt to learn more about the noise.

In the zoom on figure 9 reveals a noisy area where the outcome of running restart gmres cannot be predicted. This section aims at learning more about the noise. In a first attempt a histogram of the residual magnitude shown in figure 10, has been made. On the right a zoom on the bottom right part of the histogram is shown, such that parts of the line at zero are neglected. The bar at zero is therefore considerably higher then the one shown. However as most of the surface plot is colored in yellow which corresponds to an error size of 10^0, this does not come as a surprise that in a histogram of the same image the line at 10^0. It is rather of more interest do consider the distribution of errors smaller then $10^0 = 1$. And here another peak is observed in bin $[-1.7 - 1.6]$, which creates the impression that the noise consists of two superposed noise distribution. As shown in figure 11 noise with a similar distribution can be produced by superposing shifted absolute valued normal distributions. Finally 12 shows a histogram for the residuals found in the convergence plain of matrix **B**, interestingly it seems to consists of three superposed distribution. A link between matrix dimensions can not be confirmed by a histogram of **A** which only displays two very large bins like **C** does.

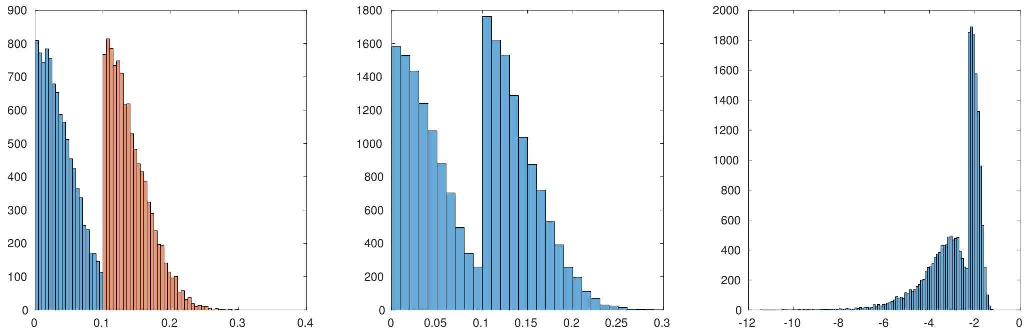

Figure 11: Plots of the distributions generated by `d1 = 0.00001 + 0.05*abs(randn(10000,1)); d2 = 0.1 + 0.05*abs(randn(10000,1));` . The last plot is on a semilogarithmic scale.

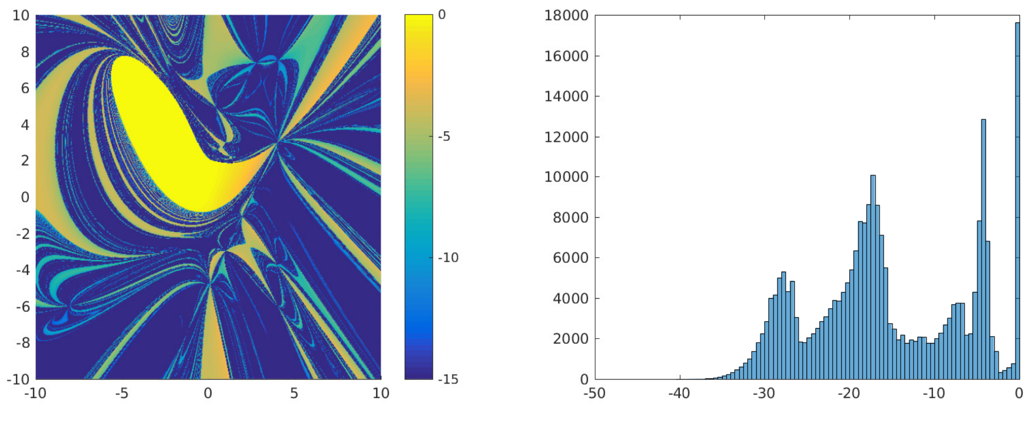

Figure 12: Convergence plain of matrix **C** along with a histogram of the logarithm of the residual norm.

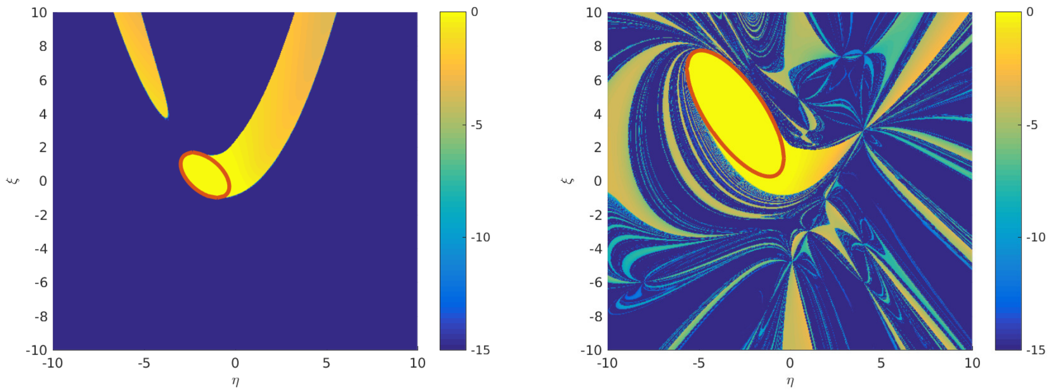

Figure 13

2.6 Stagnation condition

Using the recurrence relation for `gmres(1)`

$$\mathbf{r}_{k+1}^{(1)} = \mathbf{r}_k^{(1)} - \frac{\mathbf{r}_k^{(1)T}\mathbf{A}\mathbf{r}_k^{(1)}}{\mathbf{r}_k^{(1)T}\mathbf{A}^T\mathbf{A}\mathbf{r}_k^{(1)}}\mathbf{A}\mathbf{r}_k^{(1)}, \tag{6}$$

with $\mathbf{r}_0 = (\xi\ \eta\ 1)$ it follows that `gmres(1)` must stagnate if $\mathbf{r}_k^{(1)T}\mathbf{A}\mathbf{r}_k^{(1)} = 0$. Using the condition above the two elpsoids

$$\xi^2 + \xi\eta + \xi + \eta^2 + 3\eta + 1 = 0, \tag{7}$$
$$\xi^2 + 2\xi\eta - 2\xi + 2\eta^2 + 4\eta + 3 = 0 \tag{8}$$

can be found. Which is a set of points on which the algorithms stagnates.

3 Conclusion

In this report the results of Embree's paper "The Tortoise and the Hare, Restart GMRES "have been successfully reproduced. Just like in the tortoise the hare in the fairytale `GMRES(1)` is able to beat (`GMRES(2)`) if specific matrices are considered. As to why this happens remains unknown. Additional experiments confirmed the link between `GMRES(m)` convergence and the conditioning of the problem that was already observed in the first report. However this report showed clearly that the initial condition plays a key role as well.