

THE TORTOISE AND THE HARE RESTART GMRES

MARK EMBREE*

Abstract. When solving large nonsymmetric systems of linear equations with the restarted GMRES algorithm, one is inclined to select a relatively large restart parameter in the hope of mimicking the full GMRES process. Surprisingly, cases exist where small values of the restart parameter yield convergence in fewer iterations than larger values. Here, two simple examples are presented where GMRES(1) converges exactly in three iterations, while GMRES(2) stagnates. One of these examples reveals that GMRES(1) convergence can be extremely sensitive to small changes in the initial residual.

Key words. Restarted GMRES, Krylov subspace methods.

AMS subject classifications. 65F10, 37N30

1. Introduction. GMRES is an iterative method for solving large nonsymmetric systems of linear equations, $\mathbf{Ax} = \mathbf{b}$ [8]. Throughout science and engineering, this algorithm and its variants routinely solve problems with millions of degrees of freedom. Its popularity is rooted in an optimality condition: At the k th iteration, GMRES computes the solution estimate \mathbf{x}_k that minimizes the Euclidean norm of the residual $\mathbf{r}_k = \mathbf{Ax}_k - \mathbf{b}$ over a subspace of dimension k ,

$$(1.1) \quad \|\mathbf{r}_k\| = \min_{\substack{p \in \mathcal{P}_k \\ p(0)=1}} \|p(\mathbf{A})\mathbf{r}_0\|,$$

where \mathcal{P}_k denotes those polynomials with degree not exceeding k , and $\mathbf{r}_0 = \mathbf{b} - \mathbf{Ax}_0$ is the initial residual. As each iteration enlarges the minimizing subspace, the residual norm decreases monotonically.

GMRES optimality comes at a cost, however, since each iteration demands both more arithmetic and memory than the one before it. A standard work-around is to restart the process after some fixed number of iterations, m . The resulting algorithm, GMRES(m), uses the approximate solution \mathbf{x}_m as the initial guess for a new run of GMRES, continuing this process until convergence. The global optimality of the original algorithm is lost, so although the residual norms remain monotonic, the restarted process can stagnate with a non-zero residual, failing to ever converge [8]. Since GMRES(m) enforces local optimality on m -dimensional spaces, one anticipates that increasing m will yield convergence in fewer iterations. Many practical examples confirm this intuition.

We denote the k th residual of GMRES(m) by $\mathbf{r}_k^{(m)}$. To be precise, one cycle between restarts of GMRES(m) is counted as m individual iterations. Conventionally, then, one expects $\|\mathbf{r}_k^{(m)}\| \leq \|\mathbf{r}_k^{(\ell)}\|$ for $\ell < m$. Indeed, this must be true when $k \leq m$.

Surprisingly, increasing the restart parameter sometimes leads to *slower* convergence: $\|\mathbf{r}_k^{(m)}\| > \|\mathbf{r}_k^{(\ell)}\|$ for $\ell < m < k$. The author encountered this phenomenon while solving a discretized convection-diffusion equation described in [4]. In unpublished experiments, de Sturler [1] and Walker and Watson [11] observed similar behavior arising in practical applications. One wonders, how much smaller than $\|\mathbf{r}_k^{(m)}\|$ might $\|\mathbf{r}_k^{(\ell)}\|$ be? The smallest possible cases compare GMRES(1) to GMRES(2) for 3-by-3 matrices. Eiermann, Ernst, and Schneider present such an example for which

*Oxford University Computing Laboratory, Wolfson Building, Parks Road, Oxford OX1 3QD, United Kingdom (mark.embree@comlab.ox.ac.uk). Supported by UK Engineering and Physical Sciences Research Council Grant GR/M12414.

$\|\mathbf{r}_4^{(1)}\|/\|\mathbf{r}_4^{(2)}\| = 0.2154\dots$ [2, pp. 284–285]. Otherwise, the phenomenon we describe has apparently received little attention in the literature.

The purpose of this article is twofold. First, we describe a pair of extreme examples where GMRES(1) converges exactly at the third iteration, while GMRES(2) seems to never converge. The second example leads to our second point: Small perturbations to the initial residual can dramatically alter the convergence behavior of GMRES(1).

2. First Example. Consider using restarted GMRES to solve $\mathbf{Ax} = \mathbf{b}$ for

$$(2.1) \quad \mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 2 \\ -4 \\ 1 \end{pmatrix}.$$

Taking $\mathbf{x}_0 = \mathbf{0}$ yields the initial residual $\mathbf{r}_0 = \mathbf{b}$. Using the fact that \mathbf{A} and \mathbf{r}_0 are real, we can derive explicit formulas for GMRES(1) and GMRES(2) directly from the GMRES optimality condition (1.1). The recurrence for GMRES(1),

$$(2.2) \quad \mathbf{r}_{k+1}^{(1)} = \mathbf{r}_k^{(1)} - \frac{\mathbf{r}_k^{(1)\top} \mathbf{A} \mathbf{r}_k^{(1)}}{\mathbf{r}_k^{(1)\top} \mathbf{A}^\top \mathbf{A} \mathbf{r}_k^{(1)}} \mathbf{A} \mathbf{r}_k^{(1)},$$

was studied as early as the 1950s [3, §71],[7]. For the \mathbf{A} and $\mathbf{r}_0 = \mathbf{b}$ defined in (2.1), this iteration converges *exactly* at the third step:

$$\mathbf{r}_1^{(1)} = \begin{pmatrix} 3 \\ -3 \\ 0 \end{pmatrix}, \quad \mathbf{r}_2^{(1)} = \begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{r}_3^{(1)} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Expressions for one GMRES(2) cycle can likewise be derived using elementary calculus. The updated residual takes the form $\mathbf{r}_{k+2}^{(2)} = p(\mathbf{A})\mathbf{r}_k^{(2)}$, where $p(z) = 1 + \alpha z + \beta z^2$ is a quadratic whose coefficients $\alpha = \alpha(\mathbf{A}, \mathbf{r}_k^{(2)})$ and $\beta = \beta(\mathbf{A}, \mathbf{r}_k^{(2)})$ are given by

$$\alpha = \frac{(\mathbf{r}_k^{(2)\top} \mathbf{A} \mathbf{A} \mathbf{r}_k^{(2)})(\mathbf{r}_k^{(2)\top} \mathbf{A}^\top \mathbf{A} \mathbf{A} \mathbf{r}_k^{(2)}) - (\mathbf{r}_k^{(2)\top} \mathbf{A} \mathbf{r}_k^{(2)})(\mathbf{r}_k^{(2)\top} \mathbf{A}^\top \mathbf{A}^\top \mathbf{A} \mathbf{A} \mathbf{r}_k^{(2)})}{(\mathbf{r}_k^{(2)\top} \mathbf{A}^\top \mathbf{A} \mathbf{r}_k^{(2)})(\mathbf{r}_k^{(2)\top} \mathbf{A}^\top \mathbf{A}^\top \mathbf{A} \mathbf{A} \mathbf{r}_k^{(2)}) - (\mathbf{r}_k^{(2)\top} \mathbf{A}^\top \mathbf{A} \mathbf{A} \mathbf{r}_k^{(2)})(\mathbf{r}_k^{(2)\top} \mathbf{A}^\top \mathbf{A} \mathbf{A} \mathbf{r}_k^{(2)})},$$

$$\beta = \frac{(\mathbf{r}_k^{(2)\top} \mathbf{A} \mathbf{r}_k^{(2)})(\mathbf{r}_k^{(2)\top} \mathbf{A}^\top \mathbf{A} \mathbf{A} \mathbf{r}_k^{(2)}) - (\mathbf{r}_k^{(2)\top} \mathbf{A} \mathbf{A} \mathbf{r}_k^{(2)})(\mathbf{r}_k^{(2)\top} \mathbf{A}^\top \mathbf{A} \mathbf{r}_k^{(2)})}{(\mathbf{r}_k^{(2)\top} \mathbf{A}^\top \mathbf{A} \mathbf{r}_k^{(2)})(\mathbf{r}_k^{(2)\top} \mathbf{A}^\top \mathbf{A}^\top \mathbf{A} \mathbf{A} \mathbf{r}_k^{(2)}) - (\mathbf{r}_k^{(2)\top} \mathbf{A}^\top \mathbf{A} \mathbf{A} \mathbf{r}_k^{(2)})(\mathbf{r}_k^{(2)\top} \mathbf{A}^\top \mathbf{A} \mathbf{A} \mathbf{r}_k^{(2)})}.$$

Executing GMRES(2) on the matrix and right hand side (2.1) reveals

$$\mathbf{r}_1^{(2)} = \begin{pmatrix} 3 \\ -3 \\ 0 \end{pmatrix}, \quad \mathbf{r}_2^{(2)} = \frac{1}{2} \begin{pmatrix} 3 \\ 0 \\ 3 \end{pmatrix}, \quad \mathbf{r}_3^{(2)} = \frac{1}{28} \begin{pmatrix} 24 \\ -27 \\ 33 \end{pmatrix}, \quad \mathbf{r}_4^{(2)} = \frac{1}{122} \begin{pmatrix} 81 \\ -108 \\ 162 \end{pmatrix}.$$

The inferiority of GMRES(2) continues well beyond the fourth iteration. For example:

k	$\ \mathbf{r}_k^{(2)}\ /\ \mathbf{r}_0\ $
5	0.376888290025532...
10	0.376502488858910...
15	0.376496927936533...
20	0.376496055944867...
25	0.376495995285626...
30	0.376495984909087...

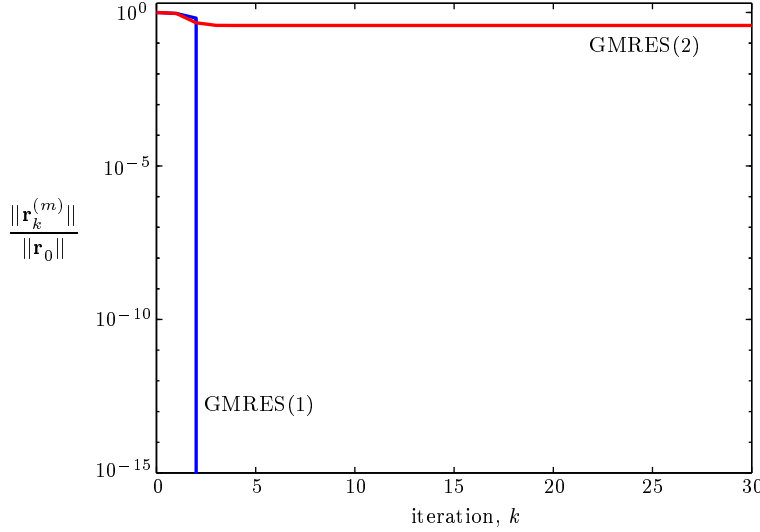


FIG. 1. Convergence curves for GMRES(1) and GMRES(2) applied to (2.1) with $\mathbf{x}_0 = \mathbf{0}$.

The entire convergence curve for the first thirty iterations is shown in Figure 1, based on performing GMRES(2) in exact arithmetic using Mathematica.

The particular value of \mathbf{b} (and thus \mathbf{r}_0) studied above is exceptional, as it is unusual for GMRES(1) to converge exactly in three iterations. Remarkably, though, GMRES(1) maintains superiority over GMRES(2) for a wide range of initial residuals. For this matrix \mathbf{A} , GMRES(2) converges exactly in one cycle for any initial residual with zero in the third component, so we restrict attention to residuals normalized to the form $\mathbf{r}_0 = (\xi, \eta, 1)^T$. Figure 2 indicates that GMRES(2) makes little progress for most such residuals, while GMRES(1) converges to high accuracy for the vast majority of these \mathbf{r}_0 values. The color in each plot reflects the magnitude of $\|\mathbf{r}_{100}^{(m)}\|/\|\mathbf{r}_0\|$: Blue indicates satisfactory convergence, while red signals little progress in one hundred iterations. (To ensure this data's fidelity, we performed these computations in both double and quadruple precision arithmetic; differences between the two were negligible.)

To gain an appreciation for the dynamics behind Figure 2, we first examine the action of a single GMRES(1) step. From (2.2) it is clear that GMRES(1) will completely stagnate only when $\mathbf{r}_0^T \mathbf{A} \mathbf{r}_0 = 0$. For the matrix \mathbf{A} specified in (2.1) and $\mathbf{r}_0 = (\xi, \eta, 1)^T$, this condition reduces to

$$(2.3) \quad \xi^2 + \xi\eta + \eta^2 + \xi + 3\eta + 1 = 0,$$

the equation for an oblique ellipse in the (ξ, η) plane.

Now writing $\mathbf{r}_k^{(1)} = (\xi, \eta, 1)^T$, consider the map $\mathbf{r}_k^{(1)} \mapsto \mathbf{s}_{k+1}^{(1)}$ that projects $\mathbf{r}_{k+1}^{(1)}$ into the (ξ, η) plane,

$$\mathbf{s}_{k+1}^{(1)} = (\mathbf{r}_{k+1}^{(1)})_3^{-1} \begin{pmatrix} (\mathbf{r}_{k+1}^{(1)})_1 \\ (\mathbf{r}_{k+1}^{(1)})_2 \end{pmatrix},$$

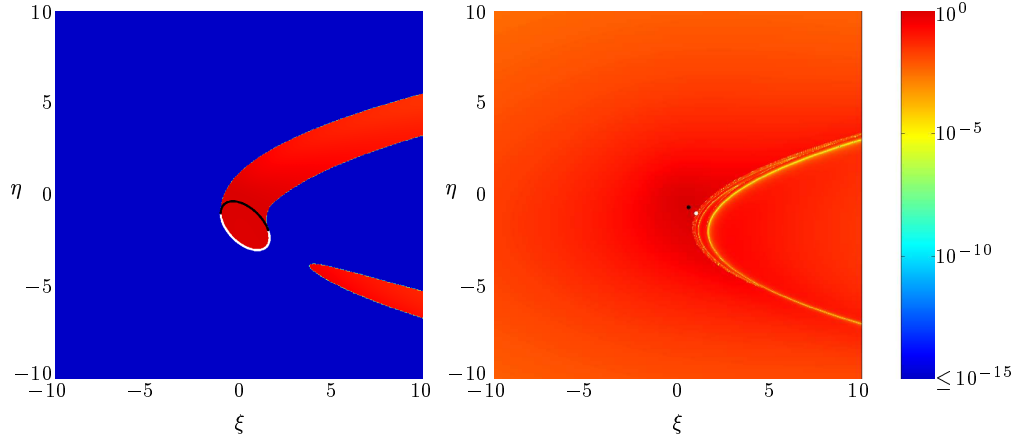


FIG. 2. Convergence of GMRES(1) (left) and GMRES(2) (right) for the matrix in (2.1) over a range of initial residuals of the form $\mathbf{r}_0 = (\xi, \eta, 1)^T$. The color indicates $\|\mathbf{r}_{100}^{(m)}\|/\|\mathbf{r}_0\|$ on a logarithmic scale: blue regions correspond to initial residuals that converge satisfactorily, while the red regions show residuals that stagnate or converge very slowly.

where $(\mathbf{r}_{k+1}^{(1)})_j$ denotes the j th entry of $\mathbf{r}_{k+1}^{(1)}$, which itself is derived from $\mathbf{r}_k^{(1)}$ via (2.2). For the present example, we have

$$(2.4) \quad \mathbf{s}_{k+1}^{(1)} = \begin{pmatrix} \frac{-\eta^3 - 4\eta^2 + 3\xi\eta + 9\xi - 4\eta - 1}{\eta^2 + \xi\eta + \xi + 5\eta + 10} \\ \frac{\eta^3 + \xi\eta^2 - 3\xi^2 + 2\eta^2 - 2\xi\eta - 3\xi + \eta - 3}{\eta^2 + \xi\eta + \xi + 5\eta + 10} \end{pmatrix}.$$

We can classify the fixed points (ξ, η) satisfying (2.3) by investigating the Jacobian of (2.4). One of its eigenvalues is always one, while the other eigenvalue varies above and below one in magnitude. In the left plot of Figure 2, we show the stable portion of the ellipse (2.3) in black and the unstable part in white.

We can similarly analyze GMRES(2). This iteration will never progress when, in addition to the stagnation condition for GMRES(1), \mathbf{r}_0 also satisfies $\mathbf{r}_0^T \mathbf{A} \mathbf{A} \mathbf{r}_0 = 0$. For the present example, this requirement implies

$$\xi^2 + 2\xi\eta + \eta^2 + 5\xi + 6\eta + 1 = 0,$$

the equation for an oblique parabola. This curve intersects the ellipse (2.3) at two points, drawn as dots in the right plot of Figure 2, the only stagnating residuals $(\xi, \eta, 1)^T$ for GMRES(2). We can analyze their stability as done above for GMRES(1). The projected map for this iteration, $\mathbf{r}_k^{(2)} \mapsto \mathbf{s}_{k+2}^{(2)}$, takes the form

$$(2.5) \quad \mathbf{s}_{k+2}^{(2)} = \begin{pmatrix} \frac{3}{\eta^2 - 3\xi + 4\eta + 9} \\ \frac{-\eta - 4}{\eta^2 - 3\xi + 4\eta + 9} \end{pmatrix}.$$

Analyzing the Jacobian for this GMRES(2) map at the pair of fixed points, we find one to be unstable (shown in black in the right plot of Figure 2) while the other is stable (shown in white). This stable fixed point is an attractor for stagnating residuals.

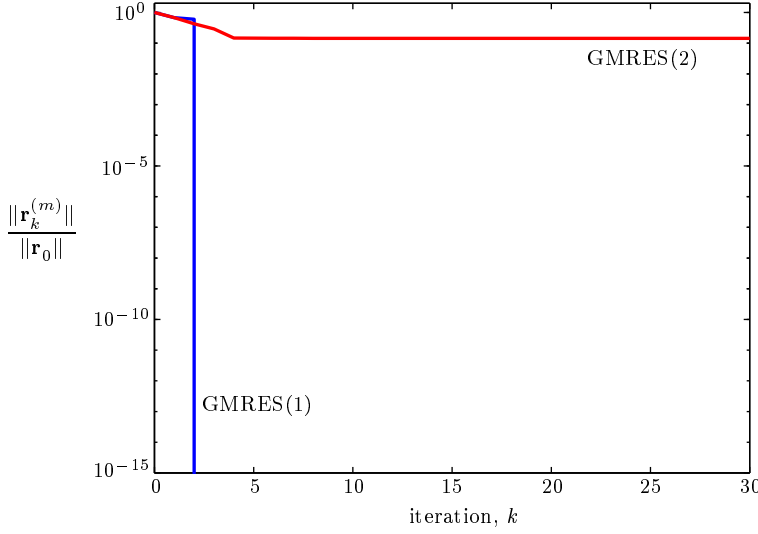


FIG. 3. Convergence curves for $\text{GMRES}(1)$ and $\text{GMRES}(2)$ applied to (3.1) with $\mathbf{x}_0 = \mathbf{0}$.

We return briefly to the initial residual $\mathbf{r}_0 = (2, -4, 1)^\top$. After the first few iterations, the angle between $\mathbf{r}_k^{(2)}$ and the fixed vector steadily converges to zero at the rate $0.6452\dots$ suggested by the Jacobian's dominant eigenvalue. We conclude with high confidence that $\text{GMRES}(2)$ never converges for this initial residual. (If one cycle of $\text{GMRES}(m)$ produces a residual parallel to \mathbf{r}_0 , then either $\mathbf{r}_m^{(m)} = \mathbf{r}_0$ or $\mathbf{r}_m^{(m)} = \mathbf{0}$. Thus a residual can't remain fixed in the finite (ξ, η) plane, but still converge to $\mathbf{0}$.)

3. Second Example. The matrix \mathbf{A} in (2.1) is nondiagonalizable, and one might be tempted to blame its surprising convergence behavior on this fact. To demonstrate that nondiagonalizability is not an essential requirement, we exhibit a diagonalizable matrix with eigenvalues $\{1, 2, 3\}$ for which restarted GMRES also produces extreme behavior. Take

$$(3.1) \quad \mathbf{A} = \begin{pmatrix} 1 & 2 & -2 \\ 0 & 2 & 4 \\ 0 & 0 & 3 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix},$$

with $\mathbf{x}_0 = \mathbf{0}$. Again, we construct the first few residuals. For $\text{GMRES}(1)$,

$$\mathbf{r}_1^{(1)} = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}, \quad \mathbf{r}_2^{(1)} = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{r}_3^{(1)} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

while $\text{GMRES}(2)$ yields

$$\mathbf{r}_1^{(2)} = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}, \quad \mathbf{r}_2^{(2)} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \quad \mathbf{r}_3^{(2)} = \frac{1}{17} \begin{pmatrix} 8 \\ 12 \\ -8 \end{pmatrix}, \quad \mathbf{r}_4^{(2)} = \frac{1}{67} \begin{pmatrix} -12 \\ 12 \\ -28 \end{pmatrix}.$$

Figure 3 illustrates the convergence curve for thirty iterations, again computed using exact arithmetic.

As with the first example, we investigate the performance of restarted GMRES for a range of $\mathbf{r}_0 = (\xi, \eta, 1)^T$, shown in Figure 4. GMRES(2) performs as before, making little progress for virtually all the residuals shown; there are two fixed points, one stable and the other not. The GMRES(1) phase plane, on the other hand, contains fascinating structure. Whether the iteration converges or stagnates appears sensitively dependent on the initial residual, highlighted in Figure 5. Red regions—indicating stagnation—are drawn towards the arc of stable fixed points (shown in black in Figure 4). The boundary between stagnating and converging residuals exhibits hallmark

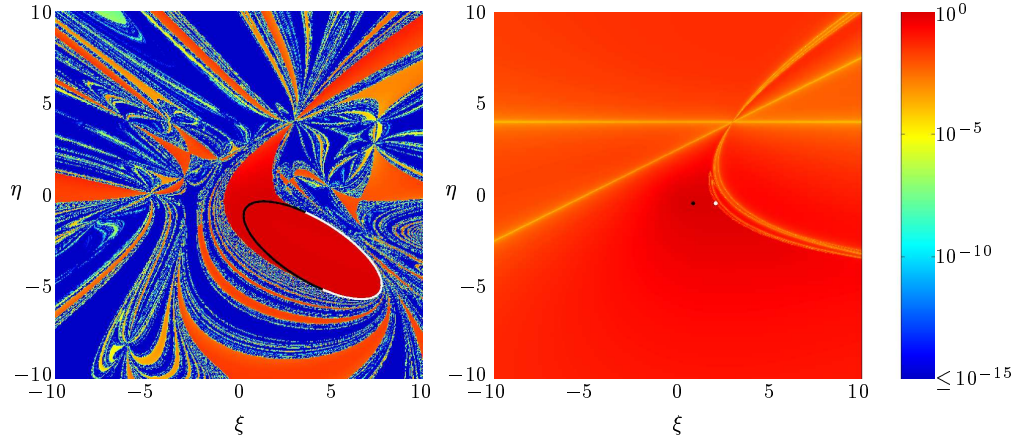


FIG. 4. Comparison of GMRES(1) (left) and GMRES(2) (right) as in Figure 2, but for the matrix in (3.1). (Double and quadruple precision computations differ notably only at the boundaries between convergence and stagnation.)

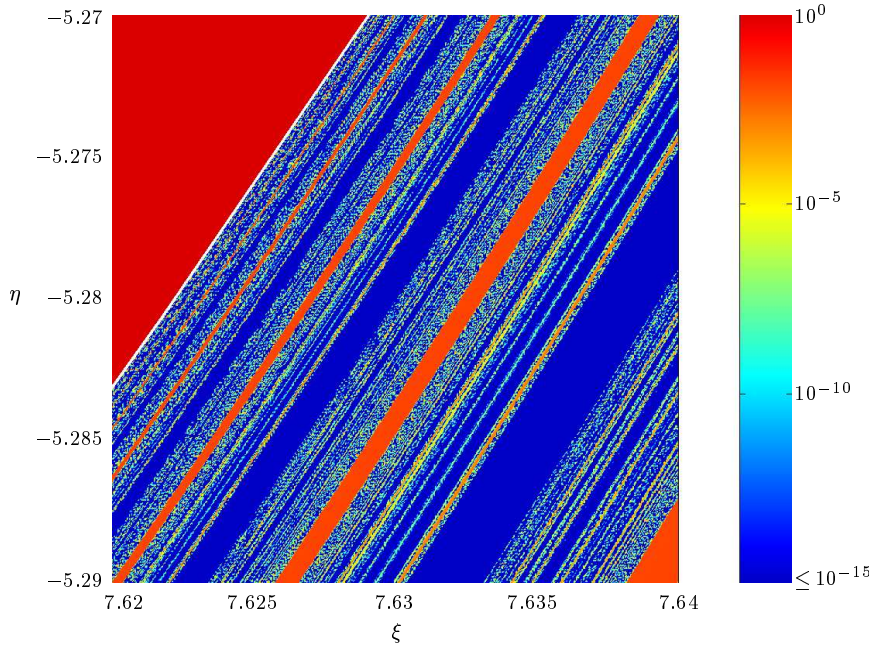


FIG. 5. Close-up of the left plot of Figure 4. The white curve denotes unstable fixed points of the map $\mathbf{s}_{k+1}^{(1)}$.

fractal qualities, as one might establish by analyzing the projected GMRES(1) map

$$\mathbf{s}_{k+1}^{(1)} = \begin{pmatrix} \frac{2\eta^3 - \xi\eta^2 - 6\xi\eta + 2\eta^2 - 11\xi - \eta - 3}{\xi^2 - \eta^2 + \xi\eta - \xi + 2\eta - 10} \\ \frac{4\eta^3 - \xi^2\eta - 4\xi^2 - 8\eta^2 - 8\xi\eta + 8\xi + 7\eta - 12}{-2\xi^2 + 2\eta^2 - 2\xi\eta + 2\xi - 4\eta + 20} \end{pmatrix}.$$

4. Closing Remarks. Examples like (2.1) and (3.1) abound. We discovered these by varying the three upper right entries of \mathbf{A} and the first two components of \mathbf{b} among the integers from -5 to 5 , while fixing all other entries. Even among such a restricted set, numerous other examples exhibit similar behavior.

Our contrived examples are extreme models of a phenomenon experienced in practical computations. For the convection-diffusion discretization described in [4], GMRES(1) or GMRES(5) can outperform GMRES(20) on moderately refined grids. The optimal choice of restart parameter depends on the problem. Since, on average, one GMRES(ℓ) iteration is cheaper than one GMRES(m) iteration when $\ell < m$ [5, 6], the potential advantage of smaller restarts is acute.

There is much more to learn about this unusual restarting behavior. How common is sensitive dependence on \mathbf{r}_0 , especially for larger restarts? What characterizes susceptible matrices? One hopes an improved GMRES convergence theory will identify better practical guidelines for choosing the restart parameter. One also wonders if related algorithms, including GMRES restarted with an augmented subspace [2] and BiCGSTAB(ℓ) [9], exhibit similar unexpected behavior. Such effects might also arise from automatic shift-selection strategies in the restarted Arnoldi algorithm for calculating eigenvalues [10].

Acknowledgements. I am indebted to Nick Trefethen and Andy Wathen for advice that both influenced the course of this research and improved its presentation. I also thank Henk van der Vorst for his helpful comments.

REFERENCES

- [1] E. DE STURLER, *All good things come to an end, or the convergence of GMRES(m)*, July 1997. Talk at SIAM's 45th Anniversary Meeting, Stanford University.
- [2] M. EIERMANN, O. G. ERNST, AND O. SCHNEIDER, *Analysis of acceleration strategies for restarted minimal residual methods*, J. Comp. Appl. Math., 123 (2000), pp. 261–292.
- [3] D. K. FADDEEV AND V. N. FADDEEVA, *Computational Methods of Linear Algebra*, Freeman, San Francisco, 1963. Translated by Robert C. Williams.
- [4] B. FISCHER, A. RAMAGE, D. SILVESTER, AND A. J. WATHEN, *On parameter choice and iterative convergence of stabilised advection-diffusion problems*, Comput. Methods Appl. Mech. Engrg., 179 (1999), pp. 185–202.
- [5] Y. HUANG AND H. VAN DER VORST, *Some observations on the convergence behavior of GMRES*, Tech. Rep. 89–09, Faculty of Technical Mathematics and Informatics, Delft University of Technology, 1989.
- [6] W. A. JOUBERT, *On the convergence behavior of the restarted GMRES algorithm for solving nonsymmetric linear systems*, Numer. Lin. Alg. Applics., 1 (1994), pp. 427–447.
- [7] M. A. KRASNOSEL'SKIĬ AND S. G. KREĬN, *An iteration process with minimal residuals*, Mat. Sbornik N.S., 31(73) (1952), pp. 315–334. (In Russian).
- [8] Y. SAAD AND M. H. SCHULTZ, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput., 7 (1986), pp. 856–869.
- [9] G. L. G. SLEIJPEN AND D. R. FOKKEMA, *BiCGSTAB(ℓ) for linear equations involving unsymmetric matrices with complex spectrum*, Elect. Trans. Numer. Anal., 1 (1993), pp. 11–32.
- [10] D. C. SORENSSEN, *Implicit application of polynomial filters in a k -step Arnoldi method*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 357–385.
- [11] L. T. WATSON, *personal communication*, February 2001.