FGMRES DYNAMICS AND THE GEOMETRIC MEAN

*

Abstract. Flexible GMRES refers to a particular modification of preconditioned GMRES in which the preconditioner expressions vary. The algorithm has met with varying success and we show theoretical results detailing particular equivalences of FGMRES to GMRES systems using a construction similar to one in [11], and use this in order to detail new convergence results for FGMRES which include a relation between FGMRES and a geometric mean of properties of certain sequences of GMRES systems. These convergence results show that FGMRES is not as appealing as certain aspects of simply preconditioned GMRES, and suggest a conjecture on the connection between FGMRES and GMRES[14]. Further, we build on [14] with an analysis of the dynamical properties of FGMRES using the convergence conditions instantiated by [28, 29, 30].

Key words. Flexible or inner-outer Krylov methods, variable preconditioning, nonsymmetric linear system, iterative solver, dynamical systems

AMS subject classifications. 65F08 65F10 65B99 65N12 65Y20 37P05

1. Introduction.

Since the advent of GMRES[19], many adaptive methods have followed suit[1, 2]. This has incentivized the creation of an adaptive GMRES algorithm, or FGMRES[20]. Since then, many other similar methods have also been developed[10, 18, 24, 25].

Here, we consider the iteration of a linear system of equations $Ax = b$ using FGMRES, and aim to build on the convergence results first shown in [20], the backward analysis of FGMRES provided in [9] (which followed a similar result for GMRES[6]), and problems posed in [17]. However, unlike [9], we will not look at backward-error analysis, but rather a different way in which we can compare the convergence analysis to the original convergence results of GMRES[19].

In the course of extending the GMRES results, we will establish a relation between GMRES and FGMRES by introducing a matrix Y (which we will refer to in this paper as 'the Y matrix') for which GMRES using Y is equivalent in performing our original FGMRES iteration on $Ax = b$ and is constructed in a manner similar to the key construction found in [11]. Using this, we can extend results of GMRES on systems with symmetric part positive definite, and this result will have as a consequence a relation between GMRES and FGMRES using the geometric mean.

From there we will detail the consequences of this theoretical property. This includes a spectral relation between Y and our original constituting matrices, and examples that indicate a strong dependence of FGMRES on the initial vector chosen in the iteration or right-hand side. Finally, we leave off with a conjecture for the relation of convergence between GMRES and FGMRES.

Because of the dependence of the Y matrix on the initial vector, this necessitates a discussion on the influence this has on possibly stagnating. To this end, we further the discussion from [14] and use the results from [28, 29, 30] in order to apply a dynamical systems approach to studying FGMRES.

The remainder of the paper is organized as follows. In section 2 we recall some basic facts of convergence of GMRES that we wish to extend to FGMRES. Section 3 is dedicated to convergence results. In section 4 we illustrate with some numerical

*(FGMRES@OPTOUTOFDATABASES.33MAIL.COM). QUESTIONS, COMMENTS, OR CORRECTIONS TO THIS DOCUMENT MAY BE DIRECTED TO THAT EMAIL ADDRESS.

examples which includes our dynamical analysis as well, and we conclude in section 5.

2. GMRES. Let $K_k := span\{r_0, Ar_0, A^2r_0, \cdots, A^{k-1}r_0\}$ be a Krylov subspace corresponding to a given matrix A and its residual r_0. It is known that GMRES on A and r_0 generates an orthogonal basis corresponding to this Krylov subspace. This allows us to write the iterates of GMRES at step m as $x_m = x_0 + V_m y_m$, where V_m forms an orthonormal basis of the Krylov subspace produced by the Arnoldi method.[13, 15, 22]. We can alternatively express this observation as the Arnoldi relation $AV = VH$, with H being upper-Hessenberg.

In particular, and to provide a comparison to FGMRES later, the standard GMRES algorithm may be expressed as:

ALGORITHM 2.1 (GMRES).

$r_0 = b - Ax_0, \beta := ||r_0||_2, v_1 = \frac{r_0}{\beta}$

For $j = 1, 2, ..., m$

  $w_j := Av_j$

  For $i = 1, ..., j$

    $h_{ij} := (w_j, v_i); w_j := w_j - h_{ij} v_i$

  End

  $h_{j+1,j} = ||w_j||_2$ *Arnoldi*

  If $h_{j+1,j} = 0$

    $m := j$

    break

  $v_{j+1} = \frac{w_j}{h_{j+1,j}}$

End

$H_m := [h_{ij}]$

Find $y_m = min||\beta e_1 - H_m y||_2$ *via a Givens rotation QR process, keeping in mind that* H *is Hessenberg.*

$x_m := x_0 + V_m y_m$ [15, 19, 22]

With this description of GMRES, we can now outline the theoretical results with which will be compared with FGMRES.

The following theorem will be similar to theorem 3.2 by relating the solution of GMRES at each step to a polynomial of powers of the given matrix A. This is a concept which will be important later in the proof of lemma 3.4.

THEOREM 2.2. *[15, 22] Let* x_m *be the mth step approximate solution obtained by GMRES, and* $r_m := b - Ax_m$*. Then*

$$x_m = x_0 + q_m(A) r_0 \tag{2.1}$$

and

$$||r_m||_2 = ||(I - Aq_m(A))r_0||_2 = min_{q \in \mathbf{P}_{m-1}} ||(I - Aq(A))r_0||_2 \tag{2.2}$$

Where q *is a polynomial of degree not exceeding* $m - 1$.

With this theorem we have only to recall one more important convergence result that can be compared with that of FGMRES. In particular, this illustrates the convergence of GMRES on matrices with symmetric part positive definite.

THEOREM 2.3 (Residual Minimization).

$$
\begin{array}{lcc}
||r_{k+1}||_2^2 & \leq & [1-\frac{\mu^2}{\sigma^2}]||r_k||_2^2 \\
where & & \\
\sigma & := & ||A||_2 = \rho^{\frac{1}{2}}(A^TA) \\
\mu & := & \lambda_{min}(A_S) = \lambda_{min}(\frac{1}{2}(A+A^T))
\end{array} \tag{2.3}
$$

[15, 22]

With this background, we can now discuss the connections between GMRES and FGMRES, and use this connection to extend the convergence results in FGMRES.[14]

3. FGMRES. As outlined above for GMRES, the Arnoldi loop constructs the following orthogonal basis of a preconditioned Krylov subspace:

$$Span(r_0, AM^{-1}r_0, \cdots, (AM^{-1})^{m-1}r_0) \tag{3.1}$$

In which the new vector is obtained from the previous vector in the process. The last step is a linear combination of the previous vectors $z_i = M^{-1}v_i, i = 1, \cdots, m$. Here, we need only apply M^{-1} to V_my_m. However, if we allow the preconditioner to change at each step, we would have

$$z_j = M_j^{-1}v_j \tag{3.2}$$

If we perform this modification, we may modify the above algorithm to create GMRES with flexible preconditioning, or FGMRES:

Algorithm 3.1 (FGMRES).

Let x_0 be an initial vector, m a preset dimension of the Krylov subspace, and define $H_m \in \Re_{(m+1)\times m}$.

Perform Arnoldi

Compute $r_0 = b - Ax_0, \beta = ||r_0||_2, v_1 = \frac{r_0}{\beta}$

For $j = 1, \cdots, m$ do

  $z_j := M_j^{-1}v_j$

  $w := Az_j$

  For $i = 1, \cdots, j$ do $h_{i,j} :=< w, v_i >, w := w - h_{i,j}v_i$

  $h_{j+1,j} = ||w||_2$, *if $h_{j+1,j} = 0$ break, $v_{j+1} = \frac{w}{h_{j+1,j}}$*

$Z_m := (z_1, \cdots, z_m)$

The approximate solution is then $x_m = x_0 + Z_my_m$, where y_m is the solution to the linear least squares problem $H_my = \beta e_1$.

It is clear that the above algorithm is mathematically equivalent to preconditioned GMRES when $M_j = M$ for $j = 1, \cdots, m$. [3, 22, 26]

In order to compare FGMRES with GMRES (before extending FGMRES results past the current literature and provide comparisons with the results listed in the previous section), we recall a basic and established properties of FGMRES which mimics theorem 2.2 above.

Theorem 3.2. $min_{x\in x_0+span(Z_m)}||b - Ax||_2 = ||b - Ax_m||$[22]

Finally, in order to extend these GMRES results, and add to the results on FGMRES, we note that FGMRES is equivalent to GMRES on a particular matrix.

Theorem 3.3. *FGMRES applied to a linear system $Ax = b$ is equivalent to applying GMRES to a linear system $Yx = b$ (with the same initial vector x_0) for some $n \times n$ matrix Y.*

Proof.

Notice that if FGMRES uses a sequence of preconditioners $M_i^{-1}A$, then FGMRES minimizes the residual over a linear combination of the vectors $z_0, z_1, \ldots$, and thus similar to what was done in [11] $Yz_i = z_{i+1}, 0 \le i \le n-1$ defines the matrix for which performing GMRES on Y is equivalent to applying FGMRES on A with the sequence of preconditioners M_i. Specifically, Y can be found algebraically as:

$$\begin{aligned} Y &= (z_1|z_2|z_3|\ldots)(z_0|z_1|z_2|\cdots)^{-1} \\ &= Z_1 Z_0^{-1} \end{aligned} \tag{3.3}$$

In this way, FGMRES is equivalent to GMRES on Y, and Y describes the convergence behavior of FGMRES. □

From this point, the Y matrix defined in the previous theorem forms a lynchpin of our analysis.[14]

In order to use this expression for Y carefully, to combine it with the result of [28], and establish very limited convergence results to compare with the previously exhibited GMRES convergence results, we will need the following lemma. This is essentially a stability result of GMRES that is applied to FGMRES and places a restriction on the variation of the preconditioners from one iteration to another (similar to, yet looser than, some results in [9, 10, 21]). A much stricter bound can be found using [6], but for the purposes of this study we neither need such strict results, and we will use the following result to build a connection between the behavior of the residual norm of FGMRES and the geometric mean of the behavior of the residual norm of the individually preconditioned GMRES iterations.

LEMMA 3.4. *Assume that* $||M_i^{-1} - M_j^{-1}|| \le \epsilon$.

Let the initial vector be given as x_0 *and* $r_0 := b - Ax_0$.

Let x_k *be the solution after* k *steps of FGMRES* ($x_k \ne x$) *and the Hessenberg matrix* H_k *be nonsingular.*

Let $a_1 = M_1^{-1}r_0$, *and define inductively* $a_k = \Sigma_{j=1}^{k} M_j^{-1}(\Sigma_{i=1}^{k-1}\alpha_{i,j,k}M_i a_i + \gamma_{k-1,j}Aa_{k-1})$ *where* $\alpha_{i,j,k}, \gamma_{k-1,j}$ *are arbitrary.*

Define the Y*-matrix so that* $Ya_i = a_{i+1}$ *as in theorem 3.3.*

Let y_k *be the solution after* k *steps of GMRES on* Y *with* $M_1^{-1}r_0$ *in place of* r_0.

Then $||x_k - y_k|| \le C_k\epsilon$ *for some constant* C_k *or* $||b - Ax_k|| \le ||b - Ay_k||$.

Proof.

We leave the proof of this in Appendix A.

□

[14]

With this, we may now establish some basic results for FGMRES, showing not only in a certain case that FGMRES converges if GMRES does (a basic theoretical result for many adaptive methods that build on top of a previous, non-adaptive version), but that FGMRES shares an amiability with matrices that have symmetric part positive definite like GMRES, paralleling theorem 2.3.

THEOREM 3.5 (Y is Positive Definite). *If each of the matrices* $M_i^{-1}A$ *has symmetric part positive definite parts and* $||M_i^{-1} - M_j^{-1}|| \le \epsilon$, *then FGMRES converges.*

Proof.

In lemma 3.4, let $\gamma_{2,j} = 1$ for $j = 1, \cdots, m$, else $\alpha, \gamma = 0$, let $z_0 = M_1^{-1}r_0$, then $a_1 = z_0$. Consequently, by theorem 3.3, the residual r_m resulting from using GMRES on Y defined by a_i cab be approximated by the actual residual term using FGMRES or bounds it from above.

Then by theorem 2.2 applied to GMRES on Y:

$$
\begin{aligned}
||r_m||^m & \\
&= (min_{\overline{\beta}}||z_0 - (\Sigma_{i=1}^m \beta_i \cdot M_i^{-1} A z_0)||)^m \\
&\leq \quad \Pi_{i=1}^m min_{\beta_i}||r_0 - \beta_i \cdot M_i^{-1} A z_0||
\end{aligned} \tag{3.4}
$$

With each item in the product is the minimum residual with respect to $M_i^{-1}A$, and thus by theorem 2.3:

$$
||r_m||^m \leq [\Pi_{i=1}^m (1 - \frac{\mu_i^2}{\sigma_i^2})]||M_1^{-1}||||r_0|| \tag{3.5}
$$

where $\mu_i = \lambda_{min}(M^{-1}A + (M^{-1}A)^T)/2, \sigma_i = ||M_i^{-1}A||_2$

□

[14]

Of particular interest in theorem 3.5 is the appearance of the geometric mean in equation (3.5). Following this, and recalling theorem 3.3, we wish to analyze this matrix Y to follow out this idea.

The result we obtain is very limited, but does point to the expanding viable connection with FGMRES and a specific geometric mean on its corresponding GMRES counterparts:

Theorem 3.6 (Unit Disk Convergence of Y). *If each of the matrices $I - M_i^{-1}A$ has norm < 1, $||M_i^{-1} - M_j^{-1}|| \leq \epsilon$, and the right-hand side vector $M_1^{-1}b$ has all nonzero entries under the Jordan basis of the matrix Y described above (and is nonsingular), then the residual norm of FGMRES at step k is identical to the residual norm of GMRES at step k on a matrix Y whose spectral radius is asymptotically bounded by the geometric mean of the norm of the matrices $I - M_i^{-1}A$.*

Proof.

Let b be replaced by $M_1^{-1}b$ and x_0 be replaced by 0, it will be useful to note a similar construction of the FGMRES matrix that if we consider minimizing the residual over the polynomial of the vectors $b, (I - AM_1^{-1})b, (I - AM_2^{-1})(I - AM_1^{-1})b, \ldots$, then by theorem 3.3 performing FMGRES is equivalent to performing GMRES on a Y with

$$
\begin{aligned}
Y \quad = \quad & [((I - AM_1^{-1})b|(I - AM_2^{-1})(I - AM_1^{-1})b|\cdots) \\
& (b|(I - AM_1^{-1})b|(I - AM_2^{-1})(I - AM_1^{-1})b|\cdots)^{-1}] \\
= \quad & Z_1 Z_0^{-1}
\end{aligned} \tag{3.6}
$$

We will perform the rest of the analysis with this Y. The rest of the result follows by applying the previous lemma 3.4 with coefficients chosen to give the matrix Y above, namely, $\alpha_{i,i,i} = 1$ and $\gamma_{i,i} = -1$ (else $\alpha, \gamma = 0$).

Assume that Y is nonsingular, and that λ_1 is an eigenvalue corresponding to the spectral radius, and if the Jordan canonical form of $Y = XJX^{-1}$ then the right hand side b is such that $e_m^T X^{-1} b \neq 0$ (by our hypothesis) where m is the geometric multiplicity of λ_1. Hence,

$$
\left(b|Yb|Y^2b|\cdots\right) c = Y^n b \tag{3.7}
$$

Let the Jordan form of $Y = XJX^{-1}$ with J ordered so that the first Jordan block contains λ_1 with geometric multiplicity m. Further, since X is nonsingular, there exists x_l such that $x_l^T X = e_m^T$. Finally, let $d = X^{-1}b$. With these simplifications, multiply the above through by x_l^T:

$$(0, 0, \ldots, 0, \Sigma_{i=1}^{n} c_i \lambda_1^{i-1}, 0, 0, \ldots, 0)d = x_l^T Y^n b \tag{3.8}$$

Since c_i satisfies the minimal polynomial of Y:

$$\lambda_1^n d_i = x_l^T Y^n b \tag{3.9}$$

By assumption $d_i \neq 0$, so then:

$$|\lambda_1|^n \leq \frac{\frac{|x_l^T Y^n b|}{||Y^n b||}}{\frac{d_i}{||b||}} \frac{||Y^n b||}{||b||} \tag{3.10}$$

Using Bunyakovsky-Cauchy-Schwartz inequality together with the fact that if $x_l^T X = e_m^T$ then $||x_l|| \leq ||X^{-1}||$:

$$|\lambda_1|^n \leq \frac{1}{\frac{d_i}{||X^{-1}||||b||}} \frac{||Y^n b||}{||b||} \tag{3.11}$$

Let $C := \frac{1}{\frac{d_i}{||X^{-1}||||b||}}$, then:

$$|\lambda_1| \leq C^{\frac{1}{n}} (\frac{||Y^n b||}{||b||})^{\frac{1}{n}} \tag{3.12}$$

Now using the fact that $Y^n b = \Pi(I - M_i^{-1}A)b$:

$$|\lambda_1| \leq C^{\frac{1}{n}} (||\Pi(I - M_i^{-1}A)||)^{\frac{1}{n}} \tag{3.13}$$

Thus, noting that $C^{\frac{1}{n}} \to 1$, if the geometric mean of the norm corresponding preconditioners ($||I - M_i^{-1}A||$) are < 1, the spectral radius is also < 1. □ [14]

4. Numerical Experiments and Examples. We now systematically illustrate the above two convergence theorems for FGMRES.

First, as an important remark, theorem 3.5 can be used to give some weak convergence bounds–*even* in the case where not all of $M_i^{-1}A$ have symmetric positive definite parts (since the minimum residual bound still holds).

That is to say, $\mu_i \in \sigma(I) - \sigma(J_i)$, and when we vary the stiff subspace, a weak bound on when convergence still occurs can be thus given by the above result.

4.1. Numerical Test: Convergence rate of FGMRES with components that have symmetric positive definite parts. The following illustrates theorem 3.5. To calculate the bound we use the minimum residual bound, and then we use the bound formed by the geometric mean of the residual norm bound for each of the individually preconditioned GMRES iterations as in equation (3.5). Afterwards, we compare it with residual norm of FGMRES. Since the minimum residual bound is not tight, the overall bound is not tight.

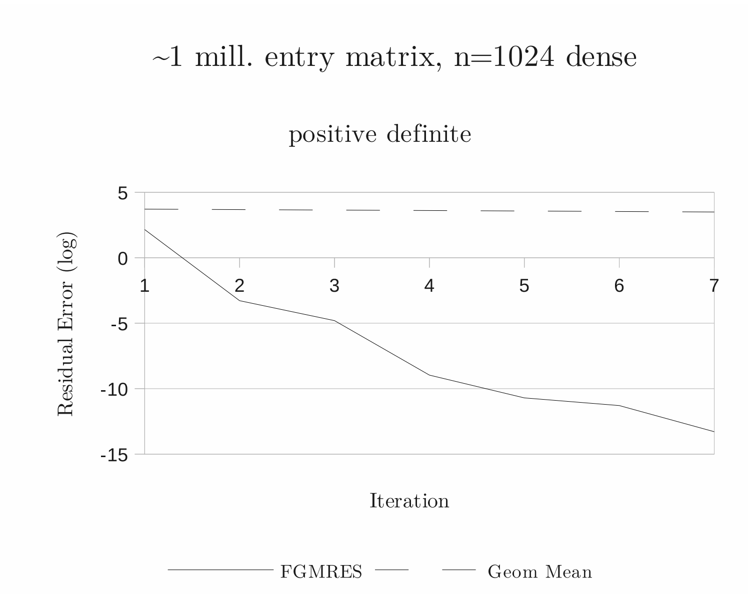

Figure 1. FGMRES bound for components with symmetric positive definite parts

A similar numerical generalization occurs with theorem 3. Theorem 3 does not show that in the unit norm case that the residual norm of FGMRES asymptotically approaches the geometric mean of the residual norm of each individual preconditioned GMRES (it only analyzes the spectra of the matrix Y), yet this claim is backed by numerical experiments as shown below.[14]

4.2. Numerical Test: FGMRES with a family of matrices where $||M_i^{-1}A|| < 1$ v. a bound which is the geometric mean of the residual norms of the individually preconditioned GMRES iterations. The following exhibits the tight bound the geometric mean of the residual norm of the individually preconditioned GMRES iterations gives. The places where FGMRES crosses the bound might be due to errors from the constant factor C in the proof of theorem 3.

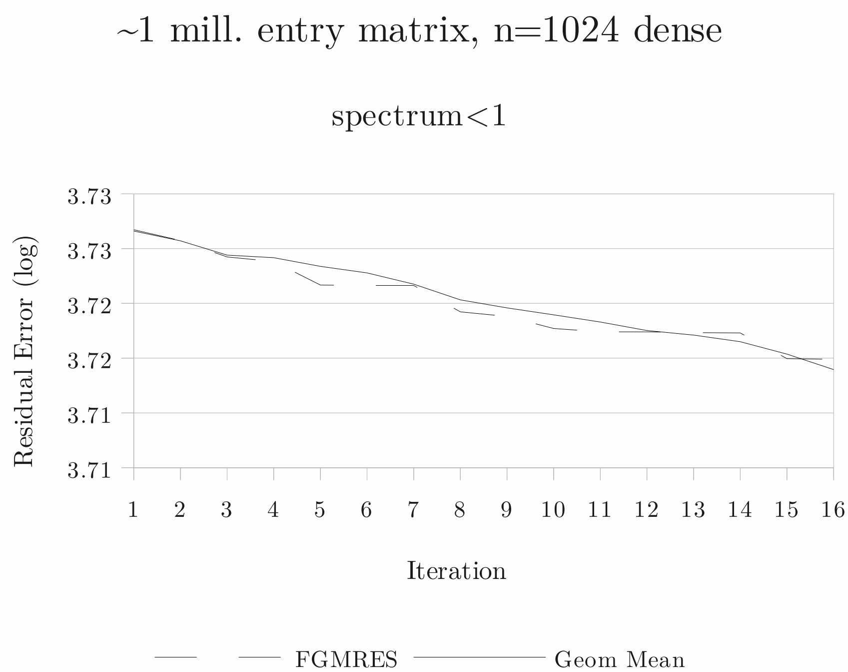

Figure 2. FGMRES residual norm v. geometric mean of GMRES residuals

Theorem 3 of the previous section included a restriction on the right-hand side of the linear system of the original system. This suggests that FGMRES might have a strong dependence in more general circumstances. This forms the bulwark of a conjecture that we propose in the conclusion of this paper. Before this, we first show an example that shows that we should indeed expect any sufficiently general theoretical property of FGMRES should include some dependence on the right-hand side of the original system or the initial vector iterate. This forms an important demonstration to introduce the dynamical properties in FGMRES.

4.3. Example: FGMRES Failure. Take as an example, an FGMRES algorithm which has $M_i^{-1}A$ as a matrix that permutes rows i and $i+1$ for $i < n-1$, $r_0 = e_i$, $M_{n-1}^{-1}A$ an arbitrary matrix, and $M_n^{-1}A = I$. Thus since $e_i^T e_j = 0$ $(i \neq j)$, the Arnoldi process will trivially produce:

$$\begin{aligned} Y &= \begin{pmatrix} e_1 & e_2 & \cdots & e_{n-1} & \overline{a} \end{pmatrix}^{-1} \begin{pmatrix} e_2 & e_3 & \cdots & \overline{a} & \overline{a} \end{pmatrix} \\ &= P_{1n} \begin{pmatrix} I_{n-1} & C \\ 0 & B \end{pmatrix} \end{aligned} \tag{4.1}$$

Where P_{1n} is the permutation matrix the permutes the first and last rows, and $\overline{a}$ can be made an arbitrary vector via appropriate choice of $M_{n-1}^{-1}A$ and r_0. Thus, in this simple example, the spectrum can depend wholely on r_0 (because P_{1n} is non-singular, then multiplication by it forms a homeomorphism; therefore, if B is made to vary its eigenvalue in modulus from 0 to ∞, then the same must occur under the homeomorphism [1]).[14]

4.4. Dynamical System Properties of FGMRES. What the previous example points out is that the Y matrix is dependent on the initial starting vectorof FGMRES. This gives an extra degree of variance that the original GMRES algorithm does not have. To this end we note the definition of the following functions of Zavorin [28, 29, 30]:

$$\begin{array}{rcl} F_V(y) & := & diag(y) \cdot V^H V y \\ G(\lambda_1, \lambda_2, \cdots, \lambda_n) & := & (-1)^{n+1} conj(\Pi_{k=1,k\neq j}^n \frac{\lambda_k}{\lambda_j - \lambda_k}) \end{array} \tag{4.2}$$

Since changing the initial vector can cause potentially large alterations in the matrix Y that GMRES is applied to, then if we are trying to determine the possibility of FGMRES to stagnate (and given that F_V is dependent on a relatively stable eigenproblem), we are left with studying the dynamics of G.

To this end, we consider G as a discrete dynamical system. We aim to show that the convergence properties of G are not of a straightforward kind that one would expect from an algorithm used to solve linear systems. In particular, even for the benign $n = 3$ case, there are heteroclinic connections and other possible manifolds that underlie the structure of GMRES. Thus, we consider $n = 3$, take G over $\Re^3$, and note that $G(a\overline{x}) = G(\overline{x})$. With these simplifications, then we can instead identify G as acting on S^2, and for $(\theta, \phi) \in S^2$, we define $\tilde{G}$ as G with domain and output restricted to the unit sphere. In particular, we define $\tilde{G}$ as:

$$\tilde{G}(\begin{bmatrix} \theta \\ \phi \end{bmatrix} = \begin{bmatrix} arccos(\sin\phi\cos\phi\sin^3(\theta)(\cos\phi - \sin\phi)) \\ arctan(\cot(\phi)\frac{\sin\theta\cos\phi - \cos\theta}{\sin\theta\sin\phi - \cos\theta}) \end{bmatrix} \tag{4.3}$$

We note that with such a definition that an analytic continuation can be made for when $x = y$, that therefore $(\theta, \phi) = (\frac{\pi}{2}, \frac{\pi}{4})$ is a fixed point, and that the eigenvalues of the Jacobian of $\tilde{G}$ at this point are $-2, 0$ respectively. Furthermore, there is a period two oscillation between $(\frac{\pi}{2}, 0)$ and $(\frac{\pi}{2}, \frac{\pi}{2})$ which both have Jacobians with eigenvalues of modulus < 1. Therefore, by [4][p. 126], we have one saddle hyperbolic point, and a convergent periodic hyperbolic point.

To explore this more, we look at the following picture showing a Newton-Raphson coloration of the domain of $\tilde{G}$, i.e., $[0, \pi] \times [-\frac{\pi}{2}, \frac{\pi}{2}]$. The lighter the shading the more stable the iteration of $\tilde{G}$. The curling line through the middle of the picture connects all of the hyperbolic points mentioned in the previous paragraph, and therefore forms a heteroclinic connection. This line through the middle of the picture is therefore the 'bad' non-convergent domain that shows where–unless the eigenbasis of the matrix happen to satisfy the non-stagnation condition–FGMRES is also likely to stagnate.

[1]The authors would like to thank Kyle Kloster and Jake Noparstak for this argument

Figure 3. Newton-Raphson like plot of iterations of $\tilde{G}$ function centered at $(\frac{\pi}{2}, \frac{\pi}{4})$

We strongly emphasize that this is not necessarily a bad property. The extra variance in Y upon each restart of FGMRES compared to the simple floating point variation of the entries of the original matrix A in GMRES can possibly force the system away from stagnation with great ease and may explain FGMRES' performance given the otherwise stultifying geometric mean conjecture. [2]

5. Conclusion. In summary, then, we have shown that FGMRES is equivalent to GMRES on a different matrix Y. Using this matrix, we were able to pull up basic convergence results from standard GMRES theory. Namely, that GMRES converges if the symmetric part of the preconditioned matrix is positive definite was extended into a similar yet appropriate form for FGMRES.

More importantly, however, was that this result introduced the concept of utilizing geometric mean to relate actions on FGMRES to individual GMRES iterations. We were able not only to show that FGMRES has a dependence on positive definiteness of the symmetric part of the corresponding preconditioned matrices, but that this this dependence is precisely that the geometric mean of the bounds for the convergence rate of GMRES formed a bound for the convergence rate of FGMRES.

Once this concept was established, we then were able to find similar properties hidden in FGMRES that followed a geometric mean relationship with similar key properties of FGMRES. Namely, a bound of the spectral radius of Y followed this relationship.

We then exhibited some numerical examples that both backed up the above described FGMRES convergence bound for which the symmetric part of the preconditioned matrices of A were all positive definite, as well as showed that this geometric mean property indeed works in more cases than we were able to definitively prove here. We also showed a specific example that shows that restrictions on the right-hand side–just as in our result regarding an asymptotic bound on the spectral radius

[2]The author would like to thank discussion with Kyle Kloster, Ahmed Sameh, Mestapha Oumouni, and Tim Kelley

of Y–should be expected in general for FGMRES.[14]

The final numerical experiment showing the dynamics of Zavorin's function is–in truth–a starting point. In actuality, and in higher dimensions, it is entirely possible that the dynamical properties will not be so simple as having a 1D heteroclinic connection. As can be evidenced in the first few iterations in the shading of Figure 3, the ranges of each successive iteration are incredibly complex. Therefore, it is very likely that higher-dimensional hyperbolic points will lead to much more interesting actions. Also, FGMRES still manages to hide this stagnating manifold into a rather small space. Furthermore, the introduction of a dynamical systems approach into FGMRES research can possibly lead into a rich area of study.

These results all converge around similar ideas, and we would like conclude by offering the following conjecture: that under some suitable restriction on A and b (given example 4.3 above) that if GMRES applied to each of the individual preconditioners of FGMRES converges, then not only will FGMRES converge; but its the residual norm at step k of FGMRES will asymptotically approach the geometric mean of the residual norm of each of the individually preconditioned GMRES iterations at step k.

It should be noted that if true, this exposes a flaw in the existing literature on FGMRES. Namely, if this conjecture is true, then FGMRES seems to no better than simply choosing the best preconditioner in the adaptive preconditioning in FGMRES. There are three explanations for this. First is that there might be computational advantages for not generating the preconditioner explicitly that mean that the algorithm must be theoretically treated as a FGMRES iteration[2]. Second, it may be theoretical impossible for the preconditioners to be generated *a priori*[1, 8]. Third, as mentioned previously, given the additional theoretical dependence on the right-hand side, this might allow us to dynamically push ourselves out of the domains where normal GMRES would have stagnated. However, given the strong restrictions on the preconditioners placed in lemma 3.4, or the unexpected conditions on the right-hand side exhibited in example 4.3 and theorem 3, the 'suitable restriction' on A and b might be stronger than anticipated. In either case, FGMRES remains to fill in an important gap in applications and still exhibits a strong relationship with the theoretical and practical advantages of GMRES.

Appendix A. FGMRES DECOMPOSITION LEMMA.

In this appendix, we restate and prove the lemma required to analyze the preconditioner dependence of FGMRES.

LEMMA A.1. *Assume that* $||M_i^{-1} - M_j^{-1}|| \leq \epsilon$.

Let the initial vector be given as x_0 *and* $r_0 := b - Ax_0$.

Let x_k *be the solution after* k *steps of FGMRES* ($x_k \neq x$) *and the Hessenberg matrix* H_k *be nonsingular.*

Let $a_1 = M_1^{-1} r_0$, *and define inductively* $a_k = \Sigma_{j=1}^{k} M_j^{-1}(\Sigma_{i=1}^{k-1} \alpha_{i,j,k} M_i a_i + \gamma_{k-1,j} A a_{k-1})$ *where* $\alpha_{i,j,k}, \gamma_{k-1,j}$ *are arbitrary.*

Define the Y*-matrix so that* $Y a_i = a_{i+1}$ *as in 3.3.*

Let y_k *be the solution after* k *steps of GMRES on* Y *with* $M_1^{-1} r_0$ *in place of* r_0.

Then $||x_k - y_k|| \leq C_k \epsilon$ *for some constant* C_k *or* $||b - Ax_k|| \leq ||b - Ay_k||$.

Proof.

We show this inductively.

$k = 1$:

Note that x_1 minimizes the residual over $x_0 + Span(M_1^{-1} r_0)$ by theorem 3.2 and y_1 minimizes the residual over $x_0 + Span(M_1^{-1} p(r_0)) = x_0 + Span(\alpha_1 M_1^{-1} r_0) = x_0 + Span(M_1^{-1} r_0)$ by theorem 2.2.

Thus $x_1 = y_1$.

$k = N$:

Assume that $||x_i - y_i|| \leq C_i\epsilon, \forall 1 < i < N$.

Note that $x_i - x_0$ is a linear combination of $z_1, z_2, \cdots$ up to z_i (where z_i are the same z_i vectors in the FGMRES algorithm).

Thus x_k minimizes the residual over $x_0 + Span(Z_k) = x_0 + Span(\{x_i - x_0 | 1 < i < N\} \cup z_k) = x_0 + Span(X_1)$ (by theorem 3.2 above).

Note that $y_i - x_0$ is a linear combination of $a_1, a_2, \cdots$ up to a_i.

Thus y_k minimizes the residual over $x_0 + Span(z_0, Yz_0, \cdots, Y^{i-1}z_0) = x_0 + Span(\{y_i - x_0 | 1 < i < N\} \cup a_k) = x_0 + span(X_2)$ (by theorem 2.2).

By algorithm 3.1 above, we have

$$z_k = M_k^{-1}(\Sigma_{i=1}^{k-1}\beta_i M_i z_i + \zeta_{k-1} A z_{k-1}). \tag{A.1}$$

And because the spans are the same as shown above, then under a different linear combination

$$z_k = M_k^{-1}(\Sigma_{i=1}^{k-1}\beta_i' M_i (x_i - x_0) + \zeta_{k-1}' A(x_{k-1} - x_0)) \tag{A.2}$$

Furthermore, by hypothesis of a_k as a polynomial expression of the prior a_i:

$$a_k = \Sigma_{j=1}^{k} M_j^{-1}(\Sigma_{i=1}^{k-1}\alpha_{i,j,k} M_i a_i + \gamma_{k-1,j} A a_{k-1}) \tag{A.3}$$

And because the spans are the same, then under a different linear combination

$$a_k = \Sigma_{j=1}^{k} M_j^{-1}(\Sigma_{i=1}^{k-1}\alpha_{i,j,k}' M_i (y_i - x_0) + \gamma_{k-1,j}' A(y_{k-1} - x_0)) \tag{A.4}$$

Note that $\gamma_{k-1,j}' \neq 0$ as when this $= 0$ this corresponds to the breakdown case of FGMRES, which can not happen since H_k is nonsingular. Now define z_k' as:

$$z_k' = \Sigma_{j=1}^{k}(\frac{\gamma_{k-1,j}'}{\zeta_{k-1}'} z_k - \Sigma_{i=1}^{k-1}\frac{\gamma_{k-1,j}'\beta_i'}{\zeta_{k-1}'}(x_i - x_0) + \Sigma_{i=1}^{k-1}\alpha_{i,j,k}'(x_i - x_0)) \tag{A.5}$$

Further, if $||M_i^{-1} - M_j^{-1}||_2 \leq \epsilon$ then for any vectors $\overline{p}, \overline{q}$

$$\begin{array}{rcl} ||\overline{p} + M_j^{-1} M_i \overline{q}|| & = & ||\overline{y} + (M_j^{-1} - M_i^{-1} + M_i^{-1})\overline{q}|| \\ & \leq & ||\overline{p} + \overline{q}|| + \epsilon||\overline{q}|| \\ & = & ||\overline{p} + \overline{q}|| + C||\epsilon|| \end{array} \tag{A.6}$$

We will use A.6 repeatedly in what follows.

Then, also recalling that $||x_i - y_i|| \leq C_i\epsilon$:

$$||z_k' - a_k||_2 = ||\Sigma_{j=1}^{k}(\tfrac{\gamma_{k-1,j}'}{\zeta_{k-1}'} z_k - \Sigma_{i=1}^{k-1}\tfrac{\gamma_{k-1,j}'\beta_i'}{\zeta_{k-1}'}(x_i - x_0) + \Sigma_{i=1}^{k-1}\alpha_{i,j,k}'(x_i - x_0)) - a_k||_2 \tag{A.7}$$

By equation A.2:

$$
\begin{aligned}
(A.8)\quad &= ||\Sigma_{j=1}^{k}(\frac{\gamma'_{k-1,j}}{\zeta'_{k-1}} M_k^{-1}(\Sigma_{i=1}^{k-1}\beta'_i M_i(x_i - x_0) + \zeta'_{k-1}A(x_{k-1} - x_0)) \\
&\quad -\Sigma_{i=1}^{k-1}\frac{\gamma'_{k-1,j}\beta'_i}{\zeta'_{k-1}}(x_i - x_0) + \Sigma_{i=1}^{k-1}\alpha'_{i,j,k}(x_i - x_0)) - a_k||_2
\end{aligned}
$$

By equation A.4:

$$
\begin{aligned}
&= ||\Sigma_{j=1}^{k}((M_k^{-1}(\Sigma_{i=1}^{k-1}\frac{\gamma'_{k-1,j}\beta'_i}{\zeta'_{k-1}}M_i(x_i - x_0) \\
(A.9)\quad &\quad +\gamma'_{k-1,j}A(x_{k-1} - x_0)) - \Sigma_{i=1}^{k-1}\frac{\gamma'_{k-1,j}\beta'_i}{\zeta'_{k-1}}(x_i - x_0) + \Sigma_{i=1}^{k-1}\alpha'_{i,j,k}(x_i - x_0)) \\
&\quad -M_j^{-1}(\Sigma_{i=1}^{k-1}\alpha'_{i,j,k}M_i(y_i - x_0) + \gamma'_{k-1,j}A(y_{k-1} - x_0)))||
\end{aligned}
$$

Using A.6:

$$
\begin{aligned}
&\leq ||\Sigma_{j=1}^{k}(((\Sigma_{i=1}^{k-1}\frac{\gamma'_{k-1,j}\beta'_i}{\zeta'_{k-1}}(x_i - x_0) \\
&\quad +\gamma'_{k-1,j}M_k^{-1}A(x_{k-1} - x_0)) - \Sigma_{i=1}^{k-1}\frac{\gamma'_{k-1,j}\beta'_i}{\zeta'_{k-1}}(x_i - x_0) + \Sigma_{i=1}^{k-1}\alpha'_{i,j,k}(x_i - x_0)) \\
&\quad -(\Sigma_{i=1}^{k-1}\alpha'_{i,j,k}(y_i - x_0) + \gamma'_{k-1,j}M_j^{-1}A(y_{k-1} - x_0)))|| + C\epsilon \\
&= ||\Sigma_{j=1}^{k}(((\gamma'_{k-1,j}M_k^{-1}A(x_{k-1} - x_0)) + \Sigma_{i=1}^{k-1}\alpha'_{i,j,k}(x_i - x_0)) \\
&\quad -(\Sigma_{i=1}^{k-1}\alpha'_{i,j,k}(y_i - x_0) + \gamma'_{k-1,j}M_j^{-1}A(y_{k-1} - x_0)))|| + C\epsilon \\
&\leq \Sigma_{j=1}^{k}(||\gamma'_{k-1,j}M_k^{-1}A(x_{k-1} - x_0) - \gamma'_{k-1,j}M_j^{-1}A(y_{k-1} - x_0)|| \\
&\quad +||\Sigma_{i=1}^{k-1}\alpha'_{i,j,k}(x_i - x_0) - \Sigma_{i=1}^{k-1}\alpha'_{i,j,k}(y_i - x_0)||) + C\epsilon \\
&\leq \Sigma_{j=1}^{k}(||\gamma'_{k-1,j}M_k^{-1}A(x_{k-1} - x_0) - \gamma'_{k-1,j}M_j^{-1}A(y_{k-1} - x_0)|| \\
&\quad +\Sigma_{i=1}^{k-1}|\alpha'_{i,j,k}|||x_i - y_i||) + C\epsilon
\end{aligned}
\tag{A.10}
$$

Induction hypothesis:

$$
\begin{aligned}
&\leq \Sigma_{j=1}^{k}(||\gamma'_{k-1,j}M_k^{-1}A(x_{k-1} - x_0) - \gamma'_{k-1,j}M_j^{-1}A(y_{k-1} - x_0)||) + C\epsilon \\
&\leq \Sigma_{j=1}^{k}(||M_k^{-1}||||\gamma'_{k-1,j}A(x_{k-1} - x_0) - \alpha'_{k-1,j,k}M_kM_j^{-1}A(y_{k-1} - x_0)||) + C\epsilon
\end{aligned}
\tag{A.11}
$$

A.6 again:

$$
\leq \Sigma_{j=1}^{k}(||M_k^{-1}||||\gamma'_{k-1,j}A(x_{k-1} - y_{k-1})||) + C\epsilon \tag{A.12}
$$

Induction hypothesis:

$$
\leq C\epsilon \tag{A.13}
$$

Therefore, under the assumption that $\gamma'_{k-1,j} \neq 0$ for some j, then $span(X_1) = span(X'_1)$ where the last column of X'_1 is z'_k and $||z'_k - a_k||_2 \leq C\epsilon$.

But since $||(y_i - x_0) - (x_i - x_0)|| \leq C_i\epsilon$, then $||X_1 - X_2||_2 \leq ||X_1 - X_2||_F \leq nC_X\epsilon$

Then by Wedin [5, 12, 27], we know that the forward error for a linear least square problem is $||x_k - y_k||_2 \leq (1 + 2\kappa_2(X_1))nC_X\epsilon = C_k\epsilon$.

Should $\gamma'_{k-1,j} = 0\forall j$, then this corresponds to $span(X_1) \supset span(X'_1)$ where the last column of X'_1 is z'_k and $||z'_k - a_k|| \leq C\epsilon$. Therefore, if we let x'_k correspond to the solution using X'_1, then similar to the previous line $||x'_k - y_k||_2 \leq C_k\epsilon$, and $||b - Ax_k|| \leq ||b - Ax'_k|| = ||b - Ay_k + A(y_k - x'_k)|| \leq ||b - Ay_k|| + C_k\epsilon$.

□

REFERENCES

[1] Baglama, J., Calvetti, D., Golub, G. H., & Reichel, L. (1998). Adaptively preconditioned GMRES algorithms. *SIAM Journal on Scientific Computing*, 20(1), 243-269.

[2] Burrage, K., & Erhel, J. (1998). On the performance of various adaptive preconditioned GMRES strategies. *Numerical linear algebra with applications*, 5(2), 101-121.

[3] Chapman, A., & Saad, Y. (1997). Deflated and augmented Krylov subspace techniques. *Numerical linear algebra with applications*, 4(1), 43-66.

[4] Dang-Vu, H., & Delcarte, C. (2000). *Bifurcations et chaos: une introduction la dynamique contemporaine avec des programmes en Pascal, Fortran et Mathematica*. Ellipses.

[5] Demmel, J. W. (1997). *Applied Numerical Linear Algebra*. SIAM.

[6] Drkoov, J., Greenbaum, A., Rozlonk, M., & Strako, Z. (1995). Numerical stability of GMRES. *BIT Numerical Mathematics*, 35(3), 309-330.

[7] Eaton, J. W., Bateman, D., & Hauberg, S. (1997). Gnu octave. Network theory.

[8] Erhel, J., Burrage, K., & Pohl, B. (1996). Restarted GMRES preconditioned by deflation. *Journal of computational and applied mathematics*, 69(2), 303-318.

[9] Giraud, L., Gratton, S., & Langou, J. (2007). Convergence in backward error of relaxed GMRES. *SIAM Journal on Scientific Computing*, 29(2), 710-728.

[10] Golub, G. H., & Ye, Q. (1999). Inexact preconditioned conjugate gradient method with inner-outer iteration. SIAM Journal on Scientific Computing, 21(4), 1305-1320.

[11] Greenbaum, A., Ptk, V., & Strako, Z. E. K. (1996). Any nonincreasing convergence curve is possible for GMRES. *SIAM Journal on Matrix Analysis and Applications*, 17(3), 465-469.

[12] Higham, Nicholas J. (2002). *Accuracy and Stability of Numerical Algorithms*. SIAM.

[13] Householder, A. S. (1964). *The Theory of Matrices in Numerical Analysis*. Dover Publications.

[14] Imberti, D. (2013). *Methods for Increasing Domains of Convergence in Iterative Linear System Solvers* (Doctoral Dissertation). Purdue University. Lafayette, Indiana, 2013.

[15] Kelley, C.T. (1995). *Iterative Methods for Linear and Nonlinear Equations*. SIAM.

[16] MATLAB and Statistics Toolbox Release 2012b, The MathWorks, Inc., Natick, Massachusetts, United States.

[17] Morikuni, K., Reichel, L., & Hayami, K. (2014). FGMRES for linear discrete ill-posed problems. *Applied Numerical Mathematics*, 75, 175-187.

[18] Notay, Y. (2000). Flexible conjugate gradients. *SIAM Journal on Scientific Computing*, 22(4), 1444-1460.

[19] Saad, Y., & Schultz, M. H. (1986). GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. SIAM Journal on scientific and statistical computing, 7(3), 856-869.

[20] Saad, Y. (1993). A flexible inner-outer preconditioned GMRES algorithm. *SIAM Journal on Scientific Computing*, 14(2), 461-469.

[21] Saad, Y. (1997). Analysis of augmented Krylov subspace methods. *SIAM Journal on Matrix Analysis and Applications*, 18(2), 435-449.

[22] Saad, Y. (2003). *Iterative Methods for Sparse Linear Systems*. SIAM.

[23] Sadok, H. (1981). *Methodes de Projection pour les Systemes Lineaires et Non Lineaires* (Doctoral Dissertation). University of Lille 1, Lille, France, 1981.

[24] Simoncini, V., & Szyld, D. B. (2002). Flexible inner-outer Krylov subspace methods. *SIAM Journal on Numerical Analysis*, 40(6), 2219-2239.

[25] Szyld, D. B., & Vogel, J. A. (2001). FQMR: A flexible quasi-minimal residual method with inexact preconditioning. *SIAM Journal on Scientific Computing*, 23(2), 363-380.

[26] Van der Vorst, H. A., & Vuik, C. (1994). GMRESR: A family of nested GMRES methods. Numerical linear algebra with applications, 1(4), 369-386.

[27] Wedin, P. . (1973). Perturbation theory for pseudo-inverses. *BIT Numerical Mathematics*, 13(2), 217-232.

[28] Zavorin, I. (2001). *Analysis of GMRES Convergence by Spectral Factorization of the Krylov Matrix* (Doctoral Dissertation). University of Maryland, College Park, August 2001.

[29] Zavorin, I. (2002). Spectral factorization of the Krylov matrix and convergence of GMRES.

[30] Zavorin, I., OLeary, D., & Elman, H. (2003). Complete stagnation of GMRES. *Linear Algebra and its Applications*, 367, 165-183.