

ALGORITMOS ADAPTATIVOS PARA O MÉTODO GMRES(m)

Tífani Teixeira Gonçalez

Pontificia Universidade Católica do Rio de Janeiro
Departamento de Matemática
Rio de Janeiro - RJ
E-mail: tifani@mat.puc-rio.br

Rudnei Dias da Cunha

Univerdade Federal do Rio Grande do Sul
Instituto de Matemática
Porto Alegre - RS
E-mail: rudnei.cunha@ufrgs.br.

1. Introdução

Nesse trabalho são estudados aprimoramentos ao método iterativo “Generalized Minimum Residual” (GMRES), proposto por Saad e Schultz em 1986 [8], e algumas de suas variantes, utilizados na solução de sistemas de equações lineares da forma

$$Ax = b$$

onde $A \in \mathfrak{R}^{n \times n}$ e $x, b \in \mathfrak{R}^n$.

No GMRES original, conforme proposto por Saad e Schultz, o resíduo inicial $r_0 = b - Ax_0$ é utilizado para gerar uma base ortonormal para um subespaço de Krylov,

$$\kappa_k(A, r_0) = \text{subespaço}(r_0, Ar_0, \dots, A^{k-1}r_0)$$

onde k é o número de iterações do método GMRES, através do processo de Arnoldi, o qual é uma modificação do processo de ortonormalização modificado de Gram-Schmidt. A solução do sistema acima é tomada como solução de um problema de mínimos quadrados que envolvem uma matriz e um vetor gerados durante o processo de Arnoldi.

Dentre algumas características do GMRES, destaca-se a convergência em, no máximo, n iterações. Obviamente, para n muito grande isso não é uma vantagem, pois o processo de Arnoldi requererá o armazenamento de n vetores de dimensão n , além de uma matriz de dimensão $(n+1) \times n$ – o que levará a uma complexidade espacial de $O(n^2)$ – e exigirá n produtos matriz-vetor.

Dessa forma em 1986, Saad e Schultz propuseram uma versão chamada *reinicializada*, GMRES(m), onde m é a dimensão da base do

subespaço de Krylov, na qual se produz uma base ortonormal para o subespaço de Krylov de dimensão m ; isto é, a cada iteração k , calcula-se o resíduo $r_k = b - Ax_k$, calcula-se os m vetores ortonormais através do processo de Arnoldi e, a partir deles resolve-se um problema de mínimos quadrados, obtendo-se uma aproximação x_{k+1} para a solução do problema. Infelizmente, ao se optar pelo método GMRES(m), perde-se a garantia de convergência apresentada pelo método GMRES, a menos que m seja suficientemente grande. Em [8] existe um teorema que foi proposta para a escolha de m ; no entanto, tal teorema é aplicável apenas a matrizes que sejam diagonalizáveis e cujos autovalores estejam distribuídos de uma maneira específica, o que não ocorre na maioria das matrizes. Em vista disso, tal teorema torna-se muito limitado, tornando-se necessário um critério mais abrangente e que não apresente tantas limitações quanto às características da matriz. Para fins de referência, o algoritmo GMRES(m) é apresentado a seguir:

1. **Início:** Escolha x_0 e calcule $r_0 = b - Ax_0$,
 $v_1 = r_0 / \|r_0\|_2$
2. **Iterações:** Para $k = 1, 2, \dots$, maxit
3. Para $j = 1, 2, \dots, m$
4. $h_{ij} = (Av_j, v_i); i = 1, 2, \dots, j$,
5. $w_{j+1} = Av_j - \sum_{i=1}^j h_{ij} v_i$
6. $h_{j+1,j} = \|w_{j+1}\|_2$
7. se $|h_{j+1,j}| < \epsilon$, vá para 10.
8. $v_{j+1} = w_{j+1} / h_{j+1,j}$
9. fim para
10. $x_k = x_0 + V_k y_k$, onde y_k minimiza

$$\|e_1 - H_k y_k\|_2$$

11. Se $\|r_k = b - Ax_k\|_2 < \varepsilon$, páre.

12. $x_0 = x_k$, $v_1 = r_k / \|r_k\|_2$

13. fim para

Nesse trabalho apresentamos maneiras de determinar dinamicamente, ao longo do processo de solução do sistema $Ax=b$, um valor adequado para m , de forma a garantir a convergência do método GMRES(m) e de algumas de suas variantes, bem como obter a solução com o menor gasto computacional possível, em termos de número de operações de ponto-flutuante.

2. Descrição dos métodos

Os experimentos foram realizados utilizando-se o método GMRES(m); uma variante do GMRES(m) utilizando transformações de Householder, ao invés do processo de Arnoldi, proposta por Walker [10], a qual chamaremos de GMRESH(m); o “Simpler GMRES” e sua variante usando transformações de Householder [11], os quais chamaremos de SGMRES(m) e SGMRESH(m); o “LooseGMRES”, proposto Baker et al. [1], chamados de LGMRES(m) e LGMRESH(m); o GMRES adaptativo proposto em [3], chamado de A-GMRES(m); e o BC-GMRES(m), proposto em [7].

Apresentamos agora uma breve explicação das variantes do GMRES, com suas principais características e vantagens.

O GMRESH(m) apresenta melhores propriedades numéricas que o GMRES(m), especialmente nas iterações finais, quando os resíduos tornam-se cada vez mais próximos, e utiliza menos armazenamento, apesar de requerer duas vezes mais operações aritméticas do que o GMRES(m). Segundo [10], no GMRESH(m) os vetores de Householder são de tamanho menor, e eles, junto com a matriz triangular superior gerada pelo algoritmo, podem ser armazenados em uma matriz de ordem $n \times (\text{maxit}+1)$, onde maxit é o número máximo de iterações. Isto é um pouco menos armazenamento do que o requerido pelo GMRES(m), o qual armazena toda a base de

vetores ortonormal, bem como as matrizes triangulares superiores. Assim, como em geral temos que $\text{maxit} < n$, então essa seria a maior vantagem entre esses dois algoritmos.

O Simpler GMRES surgiu em 1994 com Homer F. Walker e Lu Zhou [11], sendo apresentado com uma modificação no processo de Arnoldi original do GMRES. Segundo [11], na implementação usual do GMRES, o processo de Arnoldi é aplicado com $v_1 = Ar_0 / \|Ar_0\|_2$. A ortogonalização em cada passo do algoritmo pode ser feita usando a modificação de Gram Schmidt ou Transformação de Householder. Note que o SGMRES é resultante apenas da modificação feita no v_1 , tornando assim o problema de mínimos quadrados em um problema de mínimos quadrados triangular superior, eliminando a matriz de triangularização da matriz de Hessemberg.

O Loose GMRES surgiu em 2003 com Baker, Jessup e Manteuffel; veja [1]; sendo apresentado com uma técnica para acelerar a convergência do GMRES. O LGMRES(m, k) é muito semelhante ao método do gradiente conjugado completo (CG) com condicionador polinomial, e sua implementação não exige muitas modificações no GMRES(m). Segundo [1], uma motivação para propor o LGMRES se deu pelo fato de que o GMRES(m) não mantém a ortogonalidade entre espaços aproximados generalizados em sucessivas reinicializações, tornando assim sua convergência vagarosa ou até mesmo levando a divergência. Também é sabido, conforme [2], que métodos aumentados formam uma classe de técnicas de aceleração; esses métodos buscam evitar a divergência melhorando assim as informações obtidas com o GMRES em cada reinicialização. Tipicamente, um subespaço A-invariante é agrupado ao subespaço aproximado de Krylov, resultando assim em um subespaço de Krylov aumentado. O subespaço A-invariante associado com os menores autovalores é comumente usado. A idéia então consiste de adicionar vetores ao subespaço aproximado de Krylov que representaria o espaço aproximado para cada ciclo de inicialização. O LGMRES(m, k) tipicamente

não requer mais iterações do que o GMRES(m), e o fato do mesmo ser um acelerador de uma possível convergência do GMRES(m) não significa que o mesmo não possa ser aplicado sem o uso de pré-condicionadores.

O método GMRES-Adaptativo foi proposto por Maria Sosonkina Driver em 1997 [3], e o denotamos por A-GMRES. A diferença básica em relação ao GMRES(m) consiste no fato de que o A-GMRES apresenta, se necessário, alguns testes para incrementar o m e então reinicializar o método. De acordo com [3] a essência do A-GMRES é obter um parâmetro m adequado para a reinicialização do problema.

O BC-GMRES foi proposto por Takashi Nodera e Kentaro Moriya em 2003 [7], e é um método que escolhe o ciclo de reinicialização do m baseado no teste da norma residual e na distribuição dos zeros do polinômio residual do GMRES(m). Essa estratégia foi baseada em algumas conclusões apresentadas em [9]. O BC-GMRES executa a reinicialização quando a distribuição dos zeros do polinômio residual do GMRES(m) tornam-se ideais, ou seja, pode-se dizer que quanto mais espalhados estão os zeros do polinômio residual mais provável que a distribuição dos zeros aproximados seja similar aos zeros ideais. No BC-GMRES a decisão da reinicialização é tomada através da análise de duas condições, a saber: a distribuição dos zeros aproximados e o teste de convergência da norma residual.

3. Critério de escolha do valor de m

Apresentamos agora o critério desenvolvido para a escolha do m adequado, sendo que os objetivos para a elaboração do tal critério foram a redução da norma residual e a variação do m , ou seja, buscamos um critério o qual poderia crescer ou decrescer o m .

Inicialmente tentamos desenvolver um critério que apresentasse resultados satisfatórios quanto ao número de convergência dos métodos adaptativos e ao número de não convergências. Foram elaborados cinco critérios ao longo de todo o trabalho, sendo que os três primeiros

utilizavam o número de dígitos significativos exatos (DIGSE) como ferramenta para a comparação da variação do m . A idéia aqui era determinar a quase estagnação da convergência através das normas de dois resíduos sucessivos. Conforme pode-se observar em [4], os critérios elaborados com o DIGSE não apresentaram resultados satisfatórios, visto que o m praticamente não apresentou variação.

Optou-se então por alterar o valor de m de acordo com o logaritmo (base 10) das normas dos resíduos, e estabelecendo faixas de variação dentro das quais alteraríamos o valor de m de forma adequada. Por exemplo, se $\log\|r\|_2 > 0$, então o valor de m é duplicado (até um certo limite pré-especificado m_{max} , o qual pode ser escolhido em função de fatores como a memória disponível, por exemplo).

Se $\log\|r_i\|_2 < 0$ e $\log\|r_i\|_2 > 2/3 \log \epsilon$, então o método já está convergindo, e cabe agora verificar se ele está convergindo lentamente, caso em que adiciona-se a m o valor de m_{orig} (o valor inicialmente estipulado para m); caso contrário, podemos reduzir o valor de m , subtraindo dele $m_{orig}/3$ (desde que nunca seja menor do que m_{orig}).

Se $\log\|r_i\|_2 < 0$ e $\log\|r_i\|_2 < 2/3 \log \epsilon$, então o método já está quase que alcançando a tolerância especificada, e pode-se ainda alterar o valor de m , porém com quantidades menores, já que talvez seja dispendioso, por exemplo, aumentar muito m , já que a convergência quase foi alcançada. Por isso, quando a norma do resíduo encontra-se nessa faixa, porém a taxa de convergência é lenta, então adiciona-se a m o valor de $m_{orig}/2$; caso contrário, reduzimos o valor de m , subtraindo dele $m_{orig}/4$.

Em nosso critério, utilizamos como fator de determinação da taxa de convergência do método a relação $\|r_{i-5}\|_2/\|r_i\|_2$ e dizemos que ela é “rápida” se tal relação for superior a 2. Além disso, a fim de evitar muitas mudanças no valor de m , optamos por alterá-lo, se necessário, a cada cinco iterações. Tal idéia é baseada no algoritmo adaptativo para escolha do parâmetro de relaxação do método SOR conforme proposto por Hageman e Young em 1981 [5].

O critério de escolha do valor de m

pode ser apresentado de forma algorítmica como segue:

1. Se $\text{mod}(i, 5) = 0$
2. Se $\log\|r_i\|_2 > 0$, então
3. $m = \min(2m, m_{max})$
4. fim se
5. Se $\log\|r_i\|_2 < 0$ e $\log\|r_i\|_2 > 2/3 \log \varepsilon$, então
6. Se $i > 5$ e $\|r_{i-5}\|_2 / \|r_i\|_2 > 2$, então
7. $m = \max(m_{orig}, m - m_{orig}/3)$
8. senão
9. $m = \min(m + m_{orig}, m_{max})$
10. fim se
11. senão
12. Se $i > 5$ e $\|r_{i-5}\|_2 / \|r_i\|_2 > 2$, então
13. $m = \max(m_{orig}, m - m_{orig}/4)$
14. senão
15. $m = \min(m + m_{orig}/2, m_{max})$
16. fim se
17. fim se
18. fim se

onde i indica a i -ésima iteração.

Esse critério foi embutido em todos as variantes do GMRES apresentadas na seção 2, às quais foram acrescidas da letra A para que fosse possível notar a diferença em relação aos demais. Os métodos serão comparados como segue: SGMRES-A(m) com SGMRES(m); SGMRESH-A(m) com SGMRESH(m); BC-GMRES(m) com BC-GMRES-A(m); LGMRES-A(m, k) com LGMRES(m, k) e, por fim, o A-GMRES(m) com GMRESH(m) visto que estes dois utilizam a formulação do GMRES(m) com base nas transformações de Householder.

Os métodos utilizados para a realização dos testes foram empregados sem pré-condicionadores. e utilizou-se como critério de parada $\|r_i\|_2 < \varepsilon$, sendo $\varepsilon = 10^{-10}$. Os valores m_{orig} e m_{max} foram estipulados como 10 e 50, respectivamente (o que representa um intervalo de valores típicos para m encontrados na literatura). Os testes foram realizados utilizando o software MATLAB versões 5.3 e 6.2 em um computador PC com processador AMD Athlon XP 2.400 e 512MBytes de memória RAM.

Ao todo foram realizados 114 testes, com as seguintes matrizes: ADD32, ARC_130,

CAVITY05, CDDE1, E05R0000, E05R0300, FIDAPM05, FS_183_1, FS_760_1, GRE115, IMPCOL_A, JPWH991, NOS3, ORSIRR1, PDE900, SAYLR3, SHERMAN1 e STEAM1, presentes na coleção “Matrix Market” [6].

Para fins de explanação, apresentamos os gráficos comparativos da evolução da convergência do método adaptativo com o seu equivalente não-adaptativo, para as matrizes CDDE1 e E05R0300; além disso, num outro gráfico apresentamos a variação do valor de m juntamente com a curva de $\log \|r_i\|_2$. Salientamos que o eixo horizontal, o qual representa as iterações, apresenta escalas diferentes em cada método.

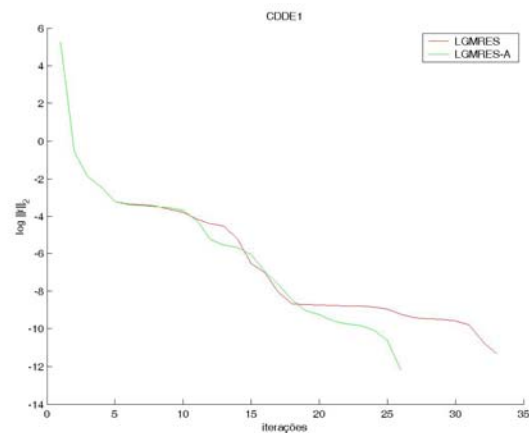


Figura 1 – Curvas residuais para o problema CDDE1, usando LGMRES($m, 2$).

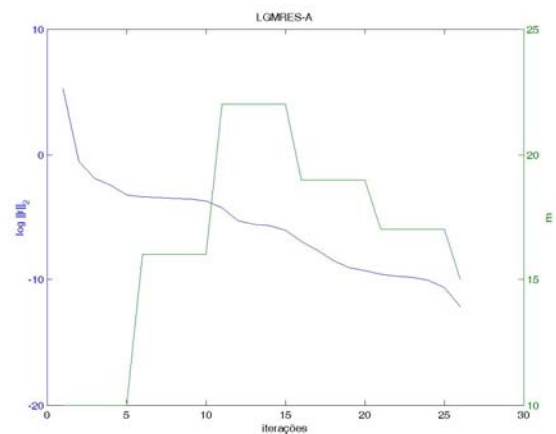


Figura 2 – Variação de m e a curva residual para o método adaptativo LGMRES-A($m, 2$).

As figuras 1 e 2 apresentam o comportamento típico do uso do critério adaptativo. Observe que o método adaptativo obteve a convergência em menos iterações e como, sempre que m

aumenta, há uma correspondente diminuição no valor do resíduo, como pode-se ver na figura 2. A Tabela 1 apresenta os resultados obtidos com todos os métodos estudados, para esse problema, onde pode-se observar que as versões adaptativas foram sempre melhores do que as não-adaptativas, garantindo a convergência e com menor custo computacional.

Método	N° de			FLOPS
	itera- ções	m mín	m máx	
A-GMRES	400	3	10	$5,08 \cdot 10^8$
GMRESH-A	30	2	20	$9,56 \cdot 10^7$
SGMRES	1000	7	10	$3,58 \cdot 10^8$
SGMRES-A	30	10	40	$4,27 \cdot 10^7$
SGMRESH	400	2	10	$3,26 \cdot 10^8$
SGMRESH-A	25	10	20	$4,49 \cdot 10^7$
BC-GMRES	1000	1	10	$1,87 \cdot 10^8$
BC-GMRES-A	70	1	60	$3,82 \cdot 10^7$
LGMRES(m,2)	35	10	10	$1,66 \cdot 10^7$
LGMRES-A(m,2)	14	10	20	$1,43 \cdot 10^7$
LGMRES(m,3)	45	10	10	$3,96 \cdot 10^7$
LGMRES-A(m,3)	40	10	20	$2,31 \cdot 10^7$

Tabela 1 – Resultados dos métodos para o problema CDDE.

As figuras 3 e 4, referentes ao problema E05R0300, apresentam um comportamento atípico, dentre todos os testes realizados, pois não se obteve convergência com qualquer dos métodos, nem tampouco foi o critério adaptativo capaz de garantir convergência.

De forma sucinta podemos apresentar os resultados da seguinte forma: 76 testes convergiram para os métodos adaptativos, ou seja, para os métodos acrescidos do critério desenvolvido nesse trabalho para a escolha do m adequado ao sistema estudado, apresentando menor número de flops e de iterações do que os demais; 24 métodos não adaptativos, ou seja, métodos que não apresentam nenhum critério para a escolha do m , convergiram apresentando menor número de flops do que as versões adaptativas e 12 testes não apresentaram convergência dentro da tolerância pré-especificada, em nenhum dos casos.

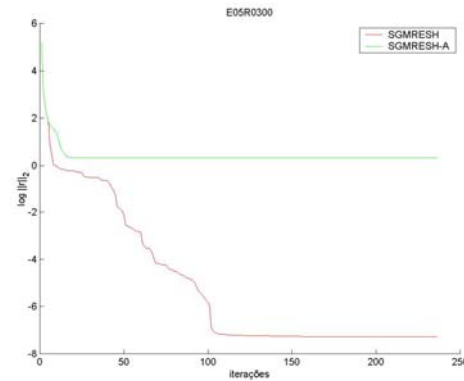


Figura 3 – Curvas residuais para o problema E05R0300, usando SGMRESH.

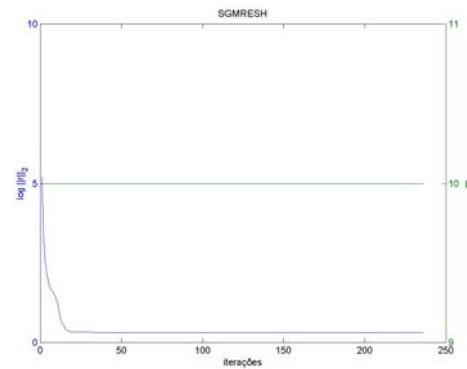


Figura 4 - Variação de m e a curva residual para o método adaptativo SGMRESH-A.

Dentre os valores apresentados devemos destacar que, para algumas matrizes, as versões adaptativas mantiveram-se absolutas, ou seja, todos os métodos adaptativos convergiram e apresentaram menor número de flops do que os demais, para as matrizes CDDE1, GRE115, NOS3, ORSIRR1, SAYLR3 e SHERMAN1. Em outros, pode-se notar que apenas as versões adaptativas obtiveram a convergência dentro da tolerância pré-especificada, como para FS_183_1; GRE115; LNS_131 e STEAM1, o que indica que o uso das versões adaptativas é indicado. Podemos citar ainda que entre 114 testes realizados, 5 apresentaram convergência apenas para as versões adaptativas, nenhuma apenas para as versões não-adaptativas, e dentre todos os

testes, somente 2 sistemas não obtiveram a convergência em nenhum dos casos. Salientamos que, para os sistemas utilizando as matrizes E05R0300 e a IMPCOL_A, não se obteve convergência em nenhuma das versões; e, nos testes realizados com a matriz ARC_130, alguns métodos apresentaram o mesmo número de flops para as versões adaptativas e não adaptativas, fazendo com que não pudéssemos concluir nada a respeito das diferenças apresentadas por essas versões.

Um fato interessante foi que o número de iterações necessárias para obter a convergência das versões adaptativas sempre foram menores do que as demais; isso, independentemente, do número de flops apresentados pelos métodos.

4. Conclusão

Em virtude dos dados apresentados podemos concluir que as versões adaptativas formam uma excelente alternativa para a resolução de sistemas não-simétricos esparsos, visto que tendem a diminuir o custo computacional e o tempo de execução dos mesmos. Note que os testes foram realizados sem o uso de pré-condicionadores, visto que o nosso objetivo nesse trabalho era de elaborar um critério adequado para a escolha da dimensão do subespaço de Krylov; devido a isso não nos detivemos nos mesmos.

Referências

[1] Baker, A., Jessup, E., e Manteuffel, T. A Technique for Accelerating the Convergence of Restarted GMRES. Technical Report CU-CS-945-03, Dept. of Computer Science, College of Engineering and Applied Science, University of Colorado at Boulder, 2003.

[2] Chapman, A. e Saad, Y. Deflated and augmented Krylov subspace techniques. *Numerical Linear Algebra with Applications*, 4:15–41, 1997.

[3] Driver, M. (1997). Parallel sparse linear algebra for homotopy methods. PhD thesis,

Faculty of the Virginia Polytechnic Institute and State University, 1997.

[4] Gonçalves, T.T. Algoritmos adaptativos para o método GMRES(m). Dissertação de mestrado, Programa de Pós-Graduação em Matemática Aplicada, UFRGS, Porto Alegre, janeiro de 2005.

[5] Hageman, L. e Young, D. *Applied Iterative Methods*. Academic Press, New York, 1981.

[6] Matrix Market: A visual repository of test data for use in comparative studies of algorithms for numerical linear algebra. <http://math.nist.gov/MatrixMarket/>. Visitado em outubro 2004.

[7] Nodera, T. e Moriya, K. New adaptive GMRES(m) method with choosing suitable restart cycle m . *PPAM*, 18(1):1105–1113, 2003.

[8] Saad, Y. e Schultz, M. GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal of Scientific and Statistical Computing*, 7:856–869, 1986.

[9] Tsuno, N. e Nodera, T. The speedup of the GMRES(m) method using the early restarting procedure. *IPSJ*, 40(4):1760–1773, 1999.

[10] Walker, H. Implementation of the GMRES method using Householder transformations. *SIAM Journal of Scientific and Statistical Computing*, 9(1):152–163, 1988.

[11] Walker, H. e Zhou, L. A simpler GMRES. *Numerical Linear Algebra with Applications*, 1:571–581, 1994.