Pares de Ritz e Pares Harmônicos de Ritz

Luiz Mariano Carvalho

16 de maio de 2024

Discutiremos aproximações para autovalores das matrizes que aparecem durante a aplicação dos métodos iterativos para a solução dos sistemas lineares. Há uma vasta literatura sobre o tema, por exemplo [2, 4, 10, 12, 16].

Vamos apresentar, inicialmente, uma abordagem feita em [15], onde se considera que o **método de Rayleigh-Ritz** é o melhor para uma boa aproximação em um dado subespaço de autovalores de uma matriz real e simétrica. Baseando-se em [11, seção 11.4], os autores dizem que esse método comporta-se bem para o cálculo de autovalores exteriores e seus autovetores associados, mas que, no entanto, o mesmo não ocorre para autovalores interiores ao espectro da matriz [7, 9, 14]. Há estudos para se tentar ultrapassar os problemas com o cálculo dos autopares interiores e os autores citam os esforços feitos em [14] e, em particular, em [9], aonde a inversão do operador (no nosso caso da matriz) pode ser tratada implicitamente (veremos como, no Teorema 4). A proposta recebeu o nome de **método harmônico de Rayleigh-Ritz** em [12]. As aproximações, valores harmônicos de Ritz, correspondentes a esse método têm recebido considerável atenção dada a sua ligação com polinômios de métodos iterativos para sistemas lineares baseados no resíduo minimal.

1 Três resultados clássicos

Vamos introduzir três resultados que discutem autovalores pelo viés da otimização. Segundo C. Meyer em [8, pág. 651], se V é uma matriz $m \times k$, com $k < m$, com colunas ortonormais (por exemplo, no processo de Arnoldi surge uma matriz como essa), então $V^H AV = \mathrm{H}$ não é uma transformação de similaridade[1], logo seria errado concluir que os autovalores de A são iguais aos autovalores de H. Apesar disso, é comum que os autovalores de H sejam uma boa aproximação para os autovalores extremos de A, em particular quando A é hermitiana. Veremos uma motivação dessa possibilidade, nos próximos teoremas. Lembremo-nos que, por serem reais, os autovalores de uma matriz hermitiana são ordenáveis, $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_m$.

[1] Duas matrizes A e B de ordem m são similares se existe S, de ordem m e não singular tal que $A = S^{-1}BS$.

Teorema 1 (Teorema de Rayleigh-Ritz). *Sejam λ_1 o maior autovalor de $A \in \mathbb{C}^{m\times m}$, matriz hermitiana, e λ_m o menor. Então*

$$\lambda_{\max} = \lambda_1 = \max_{\|x\|_2=1} x^H Ax = \max_{x\neq 0} \frac{x^H Ax}{x^H x} \tag{1}$$

$$\lambda_{\min} = \lambda_m = \min_{\|x\|_2=1} x^H Ax = \min_{x\neq 0} \frac{x^H Ax}{x^H x}. \tag{2}$$

Observação 1. *Essa forma de definir autovalores é denominada **formulação variacional** e os quocientes que aparecem no Teorema 1 recebem o nome de **quocientes de Rayleigh-Ritz** .*

Vamos apresentar um resultado que estende essa caracterização para todos os demais autovalores de uma matriz hermitiana.

Teorema 2 (Teorema de Courant-Fischer). *Os autovalores $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_m$ de uma matriz hermitiana $A \in \mathbb{C}^{m\times m}$ são*

$$\lambda_i = \max_{\dim \mathcal{V}=i} \min_{\substack{x\in\mathcal{V}\\ \|x\|_2=1}} x^H Ax \quad e \quad \lambda_i = \min_{\dim \mathcal{V}=m-i+1} \max_{\substack{x\in\mathcal{V}\\ \|x\|_2=1}} x^H Ax.$$

A demonstração desse teorema clássico, que é baseada, assim como a do teorema de Rayleigh-Ritz, na decomposição espectral de uma matriz hermitiana, pode ser vista em [6, pág. 179] ou [8, pág. 550].

Observação 2. *No caso em que $i = m$ a formulação* max min *reduz-se à apresentada em* (2)*, quando $i = 1$ a formulação* min max *torna-se igual a* (1)*.*

O próximo teorema é uma aplicação do teorema de Courant-Fischer e fornece informações sobre autovalores de matrizes relacionadas por transformações unitárias ou ortogonais.

Teorema 3 (Teorema de Entrelaçamento [17, pág. 42]). *Sejam $A \in \mathbb{C}^{m\times m}$, hermitiana, com autovalores $\lambda_{\max} = \lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_m = \lambda_{\min}$ e $U \in \mathbb{C}^{m\times n}$ com colunas ortonormais. Temos $U^H AU$ e seus autovalores $\mu_{\max} = \mu_1 \geqslant \mu_2 \geqslant \ldots \geqslant \mu_n = \mu_{\min}$. Então*

$$\lambda_i \geqslant \mu_i \geqslant \lambda_{m-n+1}, \quad i = 1 : n.$$

Esse resultado recebe o nome de teorema de entrelaçamento porque caso $n = m - 1$ então

$$\lambda_1 \geqslant \mu_1 \geqslant \lambda_2 \geqslant \mu_2 \geqslant \ldots \lambda_{m-1} \geqslant \mu_{m-1} \geqslant \lambda_m,$$

ou seja, os autovalores da matrizes são entrelaçados pelos autovalores aproximados. Caso U seja uma matriz identidade de ordem menor do que m então $U^H AU$ é uma submatriz principal de A, e esse resultado é válido também para submatrizes principais.

2 Pares de Ritz

A seguir vamos caracterizar aproximações de autovalores e autovetores da matriz A, associada a um sistema linear. As caracterizações seguintes servem para matrizes quaisquer e não apenas para matrizes hermitianas, como os teoremas anteriores.

Definição 1 (Par de Ritz)**.** *Para qualquer subespaço $\mathcal{S} \subset \mathbb{C}^m$, um vetor $x \in \mathcal{S}$, com $x \neq 0$, é um **vetor de Ritz** da matriz $A \in \mathbb{C}^{m\times m}$ associado ao **valor de Ritz** $\theta \in \mathbb{C}$, se*

$$w^H(Ax - \theta x) = 0, \forall w \in \mathcal{S} \quad ou \quad Ax - \theta x \perp \mathcal{S} \tag{3}$$

*$(x, \theta) \in \mathcal{S} \times \mathbb{C}$ é chamado **par de Ritz**.*

Uma representação gráfica de um par de Ritz pode ser vista na Figura 1.

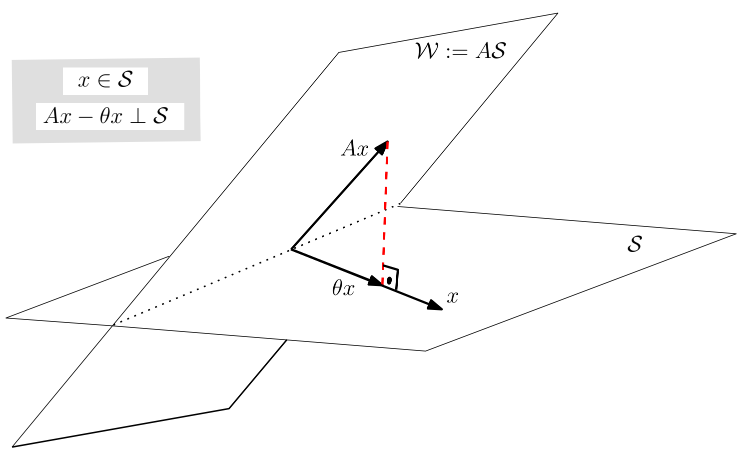

Figura 1: Representação esquemática de um par de Ritz.

Observação 3. *Usando a notação da Definição 1, sejam $V \in \mathbb{C}^{m\times n}$ uma matriz cujas colunas são ortonormais ($V^H V = I \in \mathbb{C}^{n\times n}$) e $\mathcal{S} := \mathrm{Im}(V)$. Sejam $\upsilon \in \mathbb{C}^n$, $z \in \mathbb{C}^n$, $x = V\upsilon$ e $w = Vz$. A equação (3) pode ser escrita:*

$$z^H(V^H AV\upsilon - \theta\upsilon) = 0, \forall z \in \mathbb{C}^n, \tag{4}$$

que se torna um problema padrão de cálculo de autovalores

$$V^H AV\upsilon = \theta\upsilon, \tag{5}$$

onde $x = V\upsilon$, com $\upsilon \in \mathbb{C}^n$ e $x \in \mathbb{C}^m$, ou seja, υ é a representação do vetor de Ritz x na base $\{v_1, v_2, \ldots, v_n\}$.

A Observação 3 nos permite fazer uma conexão entre a formulação variacional de Rayleigh-Ritz e o cálculo de um par de Ritz. Da equação (5) temos que

$$\upsilon^H V^H A V \upsilon = \theta \upsilon^H \upsilon \Rightarrow \theta = \frac{x^H A x}{x^H x}.$$

Vale insistir que o Teorema 1 tem como hipótese a matriz ser hermitiana e essa hipótese não é necessária à definição dos valores de Ritz.

3 Pares de Ritz harmônicos

Vamos a uma nova definição, trata-se de uma adaptação do conceito de pares de Ritz para favorecer a análise de certos métodos de Krylov.

Definição 2 (Valor Harmônico de Ritz [12])**.** *Seja $\mathcal{S} \subset \mathbb{C}^m$. O escalar $\theta \in \mathbb{C}$ é um **valor harmônico de Ritz** de A em relação um dado espaço $\mathcal{W} \subset \mathbb{C}^m$, caso θ^{-1} seja um valor de Ritz de A^{-1} com relação $\mathcal{W} := A\mathcal{S}$.*

Uma representação gráfica para essa definição pode ser vista na Figura 2. No entanto, usaremos uma outra formulação equivalente, proposta em [16], para

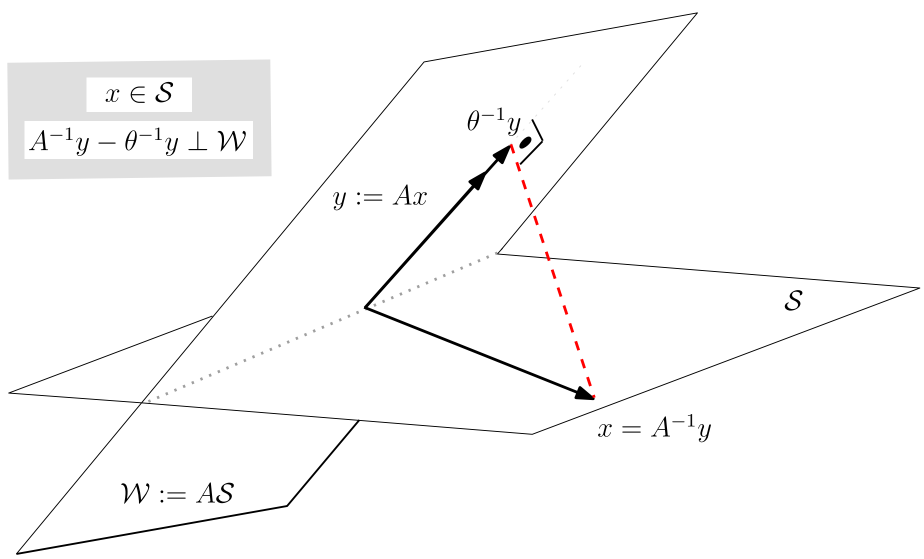

Figura 2: Representação esquemática de um par harmônico de Ritz usando a representação proposta na Definição 2.

os valores harmônicos de Ritz.

Teorema 4 (Caracterização dos Pares Harmônicos de Ritz [16, Teorema 5.1, pág. 279])**.** *Sejam $\mathcal{S} \subset \mathbb{C}^m$ e $\mathcal{W} = \{y \in \mathbb{C}^m ; \exists\, x \in S$ tal que $y = Ax\}$, ou seja,*

$\mathcal{W} := A\mathcal{S}$, então $\theta \in \mathbb{C}$ é um valor harmônico de Ritz de A em relação a $\mathcal{W}$, se e somente se,

$$w^H(Ax - \theta x) = 0, \forall w \in \mathcal{W}, \text{ para algum } x \in \mathcal{S}, \ x \neq 0. \tag{6}$$

*Denominaremos $x \in \mathcal{S}$ de **vetor harmônico de Ritz** associado a θ, e $(x, \theta) \in S \times \mathbb{C}$ de **par harmônico de Ritz**. Uma representação gráfica para essa caracterização pode ser vista na figura 3.*

Demonstração: Pelas Definições em 1 e 2, para θ ser um valor harmônico de Ritz de A em relação à $\mathcal{W}$, existem $y \neq 0 \in \mathcal{W}$ e $\theta \in \mathbb{C}$ tais que

$$w^H(A^{-1}y - \theta^{-1}y) = 0, \forall w \in \mathcal{W}, y \in \mathcal{W}, y \neq 0.$$

Basta apenas desenvolver para $y = Ax$, uma vez que $\mathcal{W} := A\mathcal{S}$

$$w^H(A^{-1}Ax - \theta^{-1}Ax) = 0 \Leftrightarrow \theta^{-1}w^H(\theta x - Ax) = 0 \Leftrightarrow w^H(\theta x - Ax) = 0.$$

■

É interessante observar que, no caso real, a equivalência entre as duas formulações de pares harmônicos de Ritz são simples relações de semelhança de triângulos retângulos, onde as hipotenusas são $x = A^{-1}y$, com $y \in \mathcal{W}, \ y \neq 0$, quando usamos a Definição 2, e θx, quando lançamos mão da caracterização proveniente do Teorema 4, ver Figura 4.

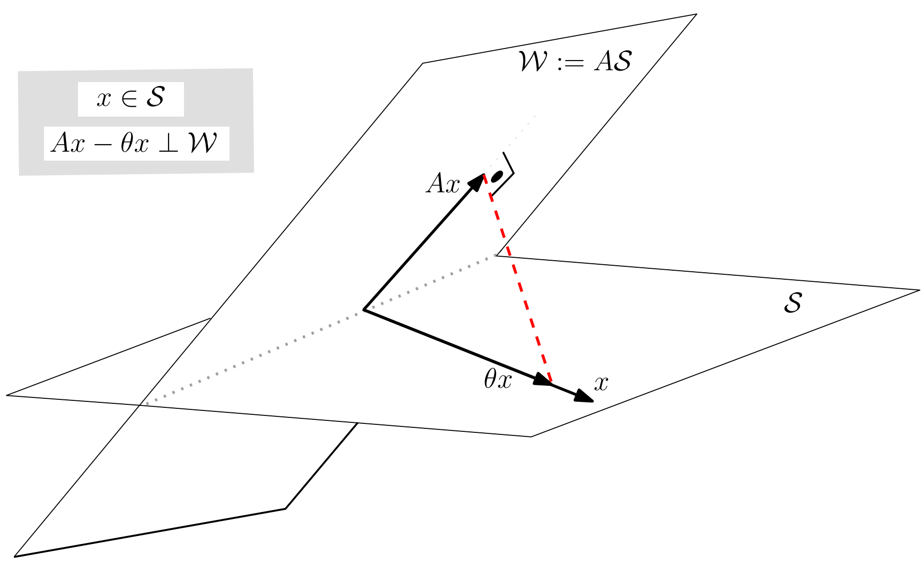

Figura 3: Representação esquemática de um par harmônico de Ritz usando a representação proposta no Teorema 4.

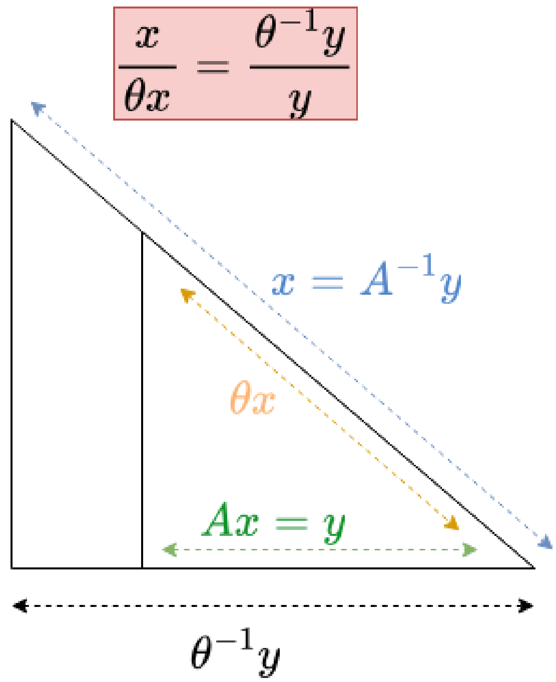

Figura 4: Relações em um triângulo retângulo das representações de um par harmônico de Ritz usando a Definição 2 e o Teorema 4. Ver Figuras 2 e 3.

Observação 4. *Usando a notação do Teorema 4, sejam* $V \in \mathbb{C}^{m\times n}$ *uma matriz cujas as colunas são ortonormais,* $V^H V = I \in \mathbb{C}^{n\times n}$ *e* $\mathcal{S} := \mathrm{Im}(V)$*. Sejam* $\chi \in \mathbb{C}^n$*,* $z \in \mathbb{C}^n$*,* $x = V\chi$ *e* $w = AVz$*. A equação* (6) *pode ser escrita como:*

$$z^H(V^H A^H AV\chi - \theta V^H A^H V\chi) = 0, \forall z \in \mathbb{C}^n, \tag{7}$$

levando a um problema generalizado de autovalores*:*

$$V^H A^H AV\chi = \theta V^H A^H V\chi. \tag{8}$$

Caso $V^H A^H V$ *seja uma matriz não singular, esse problema torna-se um problema de autovalores:*

$$\left(V^H A^H V\right)^{-1} V^H A^H AV\chi = \theta\chi. \tag{9}$$

A Observação 4 nos permite fazer a conexão entre uma variante da formulação variacional de Rayleigh-Ritz e o cálculo de um par harmônico de Ritz. Senão vejamos: da equação (8) temos que

$$\chi^H V^H A^H AV\chi = \theta\chi^H V^H A^H V\chi \Rightarrow \theta = \frac{(Ax)^H Ax}{(Ax)^H x}.$$

Aqui, novamente, vale o esclarecimento de que o Teorema 1 tem como hipótese a matriz ser hermitiana; hipótese desnecessária à definição dos valores harmônicos de Ritz.

Os pares Ritz e os pares harmônicos Ritz, e algumas de suas variações [2], são bastante utilizados nos métodos iterativos para cálculo de autovalores, ver [1]. Para matrizes hermitianas e para matrizes que não estejam muito longe de serem normais, o comportamento dos autovalores e de suas aproximações ajudam a compreender o histórico da convergência de alguns métodos de Krylov [3], [5]. Com isso, os métodos de Krylov, quando aplicados à solução de um sistema linear, podem fazer uso de aproximações de autovalores, que estão implícitas, como veremos a seguir. A princípio, o GMRES com recomeço desconsidera a maior parte da informação guardada durante a iteração anterior.

O teorema a seguir mostra a transformação do problema de cálculo de pares harmônicos de Ritz apresentado em (8) em um bem mais simples e demonstra uma propriedade relevante de ortogonalidade dos pares harmônicos de Ritz.

Teorema 5. *Seja V uma matriz cujas as colunas formam uma base ortonormal para $\mathcal{K}^{k+1}(A, r_0)$. Suponhamos que o polinômio mínimo de r_0 em relação a A tem grau maior que $k+1$. Usando a notação do método de Arnoldi a equação para pares harmônicos de Ritz em* (8) *pode ser escrita como*

$$(H_k + h^2_{(k+1),k} H_k^{-H} e_k e_k^T)\chi = \theta\chi. \tag{10}$$

Seja (x, θ) um par harmônico de Ritz de A em relação a $A\mathcal{K}_k(A, r_0)$, tal que $x = V_k\chi$, com $\chi \in \mathbb{C}^k$. Então vale a seguinte relação de ortogonalidade:

$$\overline{H}_k^H(\overline{H}_k\chi - \theta\begin{pmatrix}\chi\\0\end{pmatrix}) = 0. \tag{11}$$

<u>Demonstração</u>:

$$V_k^H A^H A V_k \chi = \theta V_k^H A^H V_k \chi$$

usando uma das relações provenientes do método de Arnoldi, $AV_k = V_{k+1}\overline{H}_k$, temos

$$\overline{H}_k^H V_{k+1}^H V_{k+1} \overline{H}_k \chi = \theta \overline{H}_k^H V_{k+1}^H V_k \chi$$

como V_{k+1} é ortogonal, podemos simplificar para

$$\overline{H}_k^H \overline{H}_k \chi = \theta \overline{H}_k^H \begin{pmatrix} I_{k\times k} \\ 0 \end{pmatrix} \chi \Rightarrow \overline{H}_k^H \overline{H}_k \chi = \theta H_k^H \chi \tag{12}$$

escrevendo a matriz $\overline{H}_k$ em blocos, chegamos a

$$\begin{pmatrix} H_k^H & \begin{matrix} 0 \\ \vdots \\ 0 \\ h_{(k+1),k} \end{matrix} \end{pmatrix} \begin{pmatrix} & H_k & \\ 0 & \dots & 0 \;\; h_{(k+1),k} \end{pmatrix} = \theta H_k^H \chi$$

que pode ser simplificado para

$$(H_k^H H_k + h_{(k+1),k}^2 e_k e_k^T)\chi = \theta H_k^H \chi$$

como podemos assumir que H_k é não singular, graças à hipótese sobre o grau do polinômio mínimo de r_0 em relação à A, então

$$(H_k + h_{(k+1),k}^2 H_k^{-H} e_k e_k^T)\chi = \theta\chi.$$

Provando a relação (10).

Partindo da equação (12)

$$\overline{H}_k^H \overline{H}_k \chi = \theta \overline{H}_k^H \begin{pmatrix} I_{k\times k} \\ 0 \end{pmatrix} \chi$$

com uma simples reorganização, temos

$$\overline{H}_k^H (\overline{H}_k \chi - \theta \begin{pmatrix} \chi \\ 0 \end{pmatrix}) = 0.$$

$\blacksquare$

Cabe observar que o teorema anterior está fortemente ancorado no uso do método de Arnoldi para ortogonalização da matriz de Krylov.

O próximo teorema relaciona os resíduos dos cálculos dos valores harmônicos de Ritz com o resíduo de uma dada iteração do GMRES, ver [10].

Teorema 6. *Sejam $\overline{H}_k$ e $(\beta e_1 - \overline{H}_k y_k)$ provenientes do algoritmo do GMRES e sejam (x_i, θ_i) pares harmônicos de Ritz de A em relação a $A\mathcal{K}_k(A, r_0)$, tal que $x_i = V_k \chi_i$, com $\chi_i \in \mathbb{C}^k$. Suponhamos que o polinômio mínimo de r_0 em relação a A tem grau maior que $k+1$. Então*

$$\overline{H}_k \chi_i - \theta_i \begin{pmatrix} \chi_i \\ 0 \end{pmatrix} = \alpha_i (\beta e_1 - \overline{H}_k y_k), \tag{13}$$

com α_i escalares.

Demonstração: Pelo resultado encontrado em [13, corolário 1.39 do teorema 1.38, pág. 36], temos que $\beta e_1 - \overline{H}_k y_k$ é ortogonal a $\overline{H}_k y$, $\forall y \in \mathbb{C}^k$, onde

$$y_k = \arg\min_{y \in \mathbb{C}^k} \|\beta e_1 - \overline{H}_k y\|_2.$$

Então

$$(\beta e_1 - \overline{H}_k y_k) \perp \mathrm{Im}(\overline{H}_k).$$

Usando (11), podemos escrever que

$$(\overline{H}_k \chi_i - \theta_i \begin{pmatrix} \chi_i \\ 0 \end{pmatrix}) \perp \mathrm{Im}(\overline{H}_k).$$

Logo ambos os vetores pertencem ao $\text{Nuc}(\overline{H}_k)$, mas pela hipótese sobre o grau do polinômio mínimo de r_0 em relação a A, H_k tem posto completo, $\text{Nuc}(\overline{H}_k)$ tem dimensão 1, e

$$(\beta e_1 - \overline{H}_k y_k) \text{ é paralelo a } (\overline{H}_k \chi_i - \lambda_i \begin{pmatrix} \chi_i \\ 0 \end{pmatrix}).$$

■

Referências

[1] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst, editors. *Templates for the solution of Algebraic Eigenvalue Problems: A Practical Guide*. SIAM, Philadelphia, 2000.

[2] C. Beattie. Harmonic Ritz and Lehman bounds. *ETNA*, 7:18–39, 1998.

[3] J. K. Cullum. Iterative methods for solving $Ax = b$ GMRES-FOM versus QMR/BiCG. Technical Report TR-96-2, Institute for Advances Studies, University of Maryland, 1996.

[4] S. Goossens and D. Roose. Ritz and harmonic Ritz values and the convergence of FOM and GMRES. *Numerical linear algebra with applications*, 6(4):281–293, 1999.

[5] A. Greenbaum, V. Pták, and Z. Strakoš. Any nonincreasing convergence curve is possible for GMRES. *SIAM Journal on Matrix Analysis and Applications*, 17(3):465–469, July 1996.

[6] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1987.

[7] Z. Jia and G. W. Stewart. An analysis of the Rayleigh-Ritz method for approximating eigenspaces. *Mathematics of Computation*, 70(234):637–647, 2001.

[8] C. D. Meyer. *Matrix analysis and applied liner algebra*. SIAM, 2000.

[9] R. B. Morgan. Computing interior eigenvalues of large matrices. *Lin. Alg. and Its Applic.*, 154/156:289–309, 1991.

[10] R. B. Morgan. Implicitly restarted GMRES and Arnoldi methods for nonsymmetric systems of equations. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1112–1135, 2000.

[11] B. N. Parlett. *The symmetric eigenvalue problem*. SIAM, Philadelphia, 1998. Corrected reprint of the 1980 original.

[12] C. C. Paige, B. N. Parlett, and H. A. van der Vorst. Approximate solutions and eigenvalue bounds from Krylov subspaces. *Numerical Linear Algebra with Applications*, 2(2):115 – 133, 1995.

[13] Y. Saad. *Iterative Methods for Sparse Linear Systems*. SIAM, 2nd edition, 2003.

[14] D. S. Scott. The advantages of inverted operators in Rayleigh–Ritz approximations. *SIAM Journal on Scientific and Statistical Computing*, 3(1):68–75, 1982.

[15] G. L. G. Sleijpen and J. van den Eshof. On the use of harmonic Ritz pairs in approximating internal eigenpairs. *Linear Algebra and its Applications*, 358(1-3):115–137, January 2003.

[16] G. L. G. Sleijpen and H. A. van der Vorst. A Jacobi–Davidson iteration method for linear eigenvalue problems. *SIAM Review*, 42(2):267–293, 2000. This paper originally appeared in SIAM Journal on Matrix Analysis and Applications, Volume 17, Number 2, 1996, pages 401-425.

[17] G. W. Stewart. *Matrix Algorithms. Volume II: Eigensystems*. SIAM, 2001.