Análisis de Convergencia de una Iteración Inexacta RQ Truncada

Cristina Flores Navarro, William Carlos Echegaray Castillo
Facultad de Ciencias, Universidad Nacional de Ingeniería
E-mail: cnavarrof@uni.edu.pe, williamechegaray@yahoo.com.br

Recibido el 01 diciembre del 2005; aceptado el 15 de diciembre del 2005

Este trabajo presenta una iteración inexacta RQ truncada y su análisis de convergencia. El presente algoritmo sirve para encontrar los autovalores (k autovalores, $k \leq n$) de una matriz $A \in \mathbb{C}(k,k)$ que puede ser esparza o densa, está dirigido principalmente para matrices de gran tamaño por ejemplo de orden 200. Parte de la iteración RQ por Givens que es semejante a la QR. Con el fin de evitar el número de cálculos se reduce la matriz a la forma Hessenberg (H) y a esta matriz reducida se le aplica la iteración RQ para producir una sucesión de transformaciones ortogonales hasta llevar a H a una triangular superior, donde los elementos expuestos en su diagonal son los autovalores. Para acelerar la convergencia se elige determinados desplazamientos $\{u_j\}$ y se procede como el anterior. Para matrices de gran tamaño es casi imposible hacer tantas iteraciones y factorizaciones RQ. Para ello después de un número considerable de iteraciones se procede a truncar en un k paso, de tal forma que se siga actualizando la porción principal, aquí surge unas ecuaciones lineales que deben solucionarse, el análisis se centra en encontrar estas soluciones, buscando un método directo apropiado (iteración TRQ), luego se soluciona estas ecuaciones con un método iterativo precondicionado (iteración ITRQ). Finalmente se hace un análisis de su convergencia, llegando a mostrar que la TRQ es cuadrática y es cúbica si la matriz A es hermitiana. Y la iteración ITRQ es lineal.

Palabras claves: Cálculos de Valores Propios, Iteración RQ.

In this work, present an inexact truncated RQ iteration and its convergence analysis. This iteration can be used to find eigenvalues of a matrix $A \in \mathbb{C}(k,k)$ (k eigenvalues, $k < n$) where the matrix A is large sparse or dense, for example the dimension of the matrix is 200×200. It begins with RQ Givens algorithm that is similar to the familiar QR algorithm. In order, reduction the calculates the algorithm begins with a complete reduction of A to upper Hessenberg form (H), the RQ iteration in the applied to H to produce a sequence of orthogonal transformations which eventually drives H into an upper triangular form with eigenvalues exposed on the diagonal. To acceleration convergence a set of shifts selected $\{u_j\}$ to acceleration the convergence and again proceeds as above. For large scale matrix is not possible do many iterations RQ. It would be desirable to truncate this update procedure after k steps to maintain and update only the leading portion of the factorizations occurring in this sequence. Here emerge a set of linear equations that requieres solving. When these equations can be solved accurately by a direct solver, its called Truncated RQ iteration (TRQ) and whether the equations are solved iteratively with some error its called inexact TRQ iteration. We analyze the convergence of an inexact TRQ iteration. We will view to TRQ, the convergence of each eigenvalue is quadratic in general and cubic is A is hermitiana and the convergence rate of the inexact TRQ is at least linear with a small convergence factor.

Keywords: Computing the eigenvalues, RQ Iteration.

1. Introducción

El origen del problema del autovalor

$$Ax = \lambda x, \quad \text{Donde} : A \in \mathbb{C}(n,n), \quad x \neq 0$$

surge de dos clases de aplicaciones. La primera consiste de los problemas relacionados al Análisis de Vibraciones. Este típicamente genera el problema del autovalor simétrico generalizado.

La segunda es la clase de problemas relacionados al Análisis de Estabilidad, tal como por ejemplo el análisis de estabilidad de un circuito eléctrico. En esta segunda clase de problemas se genera matrices no simétricas.

Se desea dar una solución numérica para problemas de autovalor algebraico de gran escala. El enfoque está en una clase de método llamado, método de proyección del subespacio de Krylov, este método tiene por objetivo reducir una matriz densa en una matriz de Hessenberg de orden menor del orden original. El método de Arnoldi se basa en este método de proyección, y es una generalización para el caso no simétrico.

Un desarrollo casi reciente y prometedor es el Método Implícito Reinicializado de Arnoldi (IRA), este método puede verse como un truncamiento del algoritmo implícito desplazado QR, (donde Q es una matriz ortogonal y R es una matriz triangular superior). Con el mismo espíritu para el método IRA, se introduce una nueva iteración llamada Iteración RQ Truncada (TRQ). Este método es muy tratable para la aceleración de la convergencia.

2. Conceptos Preliminares

El problema del autovalor y definiciones

Se define como autovalores a las n raíces del polinomio característico representado por la ecuación:

$$P_A(\lambda) = det(A - \lambda I) = \prod_{j=1}^{n} (\lambda - \lambda_j)^{d_j}, \qquad (1)$$

donde A, I son matrices de orden n, λ es el autovalor y x es un autovector. P_A: Polinomio característico de orden n, donde el exponente d_j es la multiplicidad algebraica del autovalor λ_j ($\in \mathbb{R}$ ó $\mathbb{C}$).

El conjunto discreto formado por las n raíces de la ecuación (1) es llamado el espectro de A, denotado por

$$\sigma(A) = \{\lambda_1, \lambda_2, \ldots, \lambda_n\} \tag{2}$$

Teorema 1. *Sea el espacio vectorial X. Entonces, sean V y W matrices cuyos vectores columna son dos bases ortonormales distintos de X si y sólo si $V = WP$ y $W = VQ$ donde P y Q son matrices de cambio de base respectivamente. Además: $PQ = I$ (con P, Q: matrices unitarias).*

Demostración:

Sean:

$$\{v_1, v_2, \ldots, v_k\}, \quad \{w_1, w_2, \ldots, w_k\}$$

dichas bases ortonormales de X.

Definamos:

$$V = [v_1, v_2, \cdots, v_k]$$

$$W = [w_1, w_2, \cdots, w_k]$$

Entonces, escribiendo una de ellas como combinación lineal de la otra:

$$v_j = \sum_{i=1}^{k} \lambda_{ij} w_i, \quad j = 1, \ldots, k$$

expresando matricialmente:

$$[v_1, v_2, \cdots, v_k] = [w_1, w_2, \cdots, w_k] \underbrace{\begin{pmatrix} \lambda_{11} & \cdots & \lambda_{1k} \\ \vdots & \ldots & \vdots \\ \lambda_{k1} & \cdots & \lambda_{kk} \end{pmatrix}}_{P} \tag{3}$$

análogamente la otra se escribe:

$$w_j = \sum_{i=1}^{k} \beta_{ij} w_i, \quad j = 1, \ldots, k$$

expresando matricialmente:

$$[w_1, w_2, \cdots, w_k] = [v_1, v_2, \cdots, v_k] \underbrace{\begin{pmatrix} \beta_{11} & \cdots & \beta_{1k} \\ \vdots & \ldots & \vdots \\ \beta_{k1} & \cdots & \beta_{kk} \end{pmatrix}}_{Q} \tag{4}$$

luego reemplazando, (4) en (3):

$$[v_1, v_2, \cdots, v_k] = [v_1, v_2, \cdots, v_k] \underbrace{QP}_{I}$$

análogamente (3) en (4):

$$[w_1, w_2, \cdots, w_k] = [w_1, w_2, \cdots, w_k] \underbrace{PQ}_{I}$$

Por lo tanto,

$$P = Q^{-1} \qquad \blacksquare$$

Subespacios invariantes y transformaciones semejantes

Definición 1. *Un subespacio invariante es definido por:*

$$x \in S \Rightarrow Ax \in S, \quad S \subset \mathbb{C}^n \tag{5}$$

donde: S - subespacio de dimensión k, $(k \leq n)$.

Si v es un valor propio de A, entonces $S = span\{v\}$ es invariante en relación a la matriz A.

Sea X una matriz cuyas columnas contienen los vectores que forman una base que genera S luego X es de orden $n \times k$ y X satisface:

$$AX = XG \text{ donde } G \in \mathbf{M}(k, k) \tag{6}$$

y se define el Problema del Autovalor de G, como:

$$Gy = \lambda y \text{ donde } \lambda \text{ es un autovalor de } G. \tag{7}$$

Multiplicando por y a la ecuación (6) se tiene:

$$A(Xy) = \lambda(Xy) \tag{8}$$

definimos $x = Xy$ donde λ es autovalor de A, luego:

$$\sigma(G) \subset \sigma(A) \tag{9}$$

Si A y G son del mismo orden: $\sigma(A) = \sigma(G)$, luego A es semejante a G, bajo la transformación de X.

Definición 2. *La descomposición de la matriz A a la forma Hessenberg está dada por:*

$$H = V^T A V \tag{10}$$

donde $V \in \mathbf{M}(n, n)$, V es ortogonal, H es Hessenberg.

Subespacio de Krylov

El subespacio de Krylov, es el subespacio de orden k definido por:

$$K_k(A, v_1) = span\{v_1, Av_1, A^2 v_1, \ldots, A^{k-1} v_1\} \tag{11}$$

donde $A \in \mathbf{M}(n, n)$, $v_1 \neq \overline{0}$.

El subespacio de Krylov es el subespacio de todos los vectores pertenecientes a $\mathbb{C}^n$ que pueden ser escritos como:

$$w = q(A) v_1 \tag{12}$$

donde q es un polinomio de grado a lo más $k - 1$.

Definición 3. *Un vector $x \in K_k$ es un vector de Ritz con un correspondiente valor de Ritz θ si la condición de Galerkin se satisface:*

$$\langle w, Ax - x\theta \rangle = 0, \ \forall w \in K_k(A, v_1) \tag{13}$$

donde: $\langle \cdot, \cdot \rangle$ es el símbolo de producto interno y (x, θ) es el par de Ritz.

La aplicación de esta condición está en la aproximación de autovalores y autovectores. Luego tenemos algunas consecuencias inmediatas:

Sea W una matriz cuyas columnas forman una base ortonormal de $K_k(A, v_1)$, digamos

$$W = [w_1, w_2, \cdots, w_k] \in \mathbb{R}(n, k)$$

Sea P una proyección ($P^2 = P$) además autoadjunta ($P^H = P$), definida:

$$P = WW^H \tag{14}$$

Definamos:

$$\widehat{A} = PA \quad G = W^H AW \tag{15}$$

donde $P \in \mathbf{M}(n, n)$, $\widehat{A} \in \mathbf{M}(n, n)$ y $G \in \mathbf{M}(k, k)$.

Lema 1. *(x, θ), es un par de Ritz si y solamente si $x = Wy$ con $Gy = y\theta$. Además es independiente de la base ortonormal $\{w_1, \ldots, w_k\}$ (vectores columna de W) que se elija.*

Demostración:

Este lema es una consecuencia de la condición de Galerkin, como veremos:

($\Longrightarrow$)

Como (x, θ) es un par de Ritz, entonces

$$\langle w, Ax - x\theta \rangle = 0, \ \forall w \in K_k(A, v_1)$$

como $\langle w, Ax - x\theta \rangle = w^H(Ax - x\theta), \ \forall w \in K_k(A, v_1)$. En particular para todas las columnas de W que son base ortonormal de $K_k(A, v_1)$ se tiene

$$W^H(Ax - x\theta) = \vec{0}$$

si $x = Wy$, donde $y \in \mathbb{R}^k$

$$W^H(AWy - Wy\theta) = \vec{0}$$

entonces,

$$\underbrace{W^H AW}_{G} y = \underbrace{W^H W}_{I_{(k\times k)}} y\theta$$

de la ecuación (15) aqui

$$Gy = y\theta$$

Por lo tanto:

$$x = Wy \text{ con } Gy = y\theta$$

($\Longleftarrow$)

Por hipótesis

$$Gy = y\theta \quad si \quad x = Wy \tag{16}$$

sustituyendo (15) en (16):

$$\begin{aligned} W^H AWy &= y\theta \\ W^H AWy - y\theta &= 0 \\ W^H AWy - W^H Wy\theta &= 0 \end{aligned}$$

$$W^H(AWy - Wy\theta) = 0 \tag{17}$$

De la ecuación (16), y de la ecuación (17) se tiene:

$$W^H(Ax - x\theta) = 0,$$

sea w^H, una fila de W^H entonces w es columna de W. Y como las columnas de W forman una base ortonormal de $K_k(A, v_1)$, en la ultima relación se tiene

$$\langle w_i, Ax - x\theta \rangle = 0, \quad w_i \in \{w_1, \ldots, w_k\}, \quad i = 1, \ldots, k$$

Sea $w \in K_k(A, v_1)$

$$w = \sum_{i=1}^{k} \alpha_i w_i \in K_k, \quad \alpha_i \text{: escalar}, \quad (w_i)_{i=1}^k \text{ base de } K_k$$

$$\begin{aligned} \langle w, Ax - x\theta \rangle &= \langle \sum_{i=1}^{k} \alpha_i w_i, Ax - x\theta \rangle \\ &= \sum_{i=1}^{k} \alpha_i \langle w_i, Ax - x\theta \rangle \\ &= \sum_{i=1}^{k} \alpha_i 0 \\ &= 0, \quad \forall w \in K_k \end{aligned}$$

Por lo tanto (x, θ) es un par de Ritz.

Veamos que independiente es la elección de $\{w_1, \ldots, w_k\}$:

Asumiendo que $\{w_1, w_2, \ldots, w_k\}$ es la base ortonormal de $K_k(A, v_1)$ y estos son los k vectores columna de W.

Sea $\{v_1, v_2, \ldots, v_k\}$ otra base ortonormal de $K_k(A, v_1)$, y por el teorema 1 tenemos

$$V = WQ \tag{18}$$

donde:

Q - matriz unitaria de orden k.

$V = [v_1, v_2, \ldots, v_k]$.

Sea

$$H = V^H AV \tag{19}$$

se hace el siguiente cambio de base, sustituyendo (18) en (19):

$$H = Q^H \underbrace{W^H AW}_{G} Q \tag{20}$$

por la ecuación (15),

$$H = Q^H GQ \tag{21}$$

Finalmente la ecuación (21), demuestra que H es semejante a G, y se justifica por el uso de transformación unitaria Q.

Luego H es semejante a G sobre una transformación unitaria Q

Por lo tanto, matrices unitarias del mismo orden, poseen los mismos autovalores y con ello la conservación de los valores de Ritz.

En cuanto a los vectores de Ritz:

sea x un vector de Ritz

$$\begin{aligned} x &= Wy \\ &= VQ^H y \end{aligned}$$

colocando $\widehat{y} = Q^H y$.

Tenemos, que

$$x = V\widehat{y} \tag{22}$$

Además $\widehat{G}\widehat{y} = \widehat{y}\theta$ donde $\widehat{G} = \widehat{W}^H A\widehat{W}$, por la ecuación (15) ■

Las propiedades del subespacio de Krylov pueden ser mejor analizadas usando polinomios [5].

Se establecerá una base ortonormal. Utilizando transformaciones Givens, para generar una matriz unitaria Q tal que:

$$V = WQ \tag{23}$$

Luego V es una base. Se formará una matriz H (Hessenberg superior). Teniéndose la reducción completa de Hessenberg superior

$$H = V^H AV \quad \text{donde } A, V, H \in \mathbf{M}(n,n) \tag{24}$$

En las k primeras columnas de la ecuación (24) se va cumplir una relación muy importante, esto se puede ver igualando de (24) las k primeras columnas de ambos miembros.

$$AV_k = V_{k+1}\overline{H}_k \quad \text{ó} \quad AV_k = V_k H_k + h_{k+1,k} v_{k+1} e_k^T \tag{25}$$

Las ecuaciones en (25) se conocen como la Factorización de Arnoldi, o también como los k pasos de Arnoldi. El propósito es investigar el uso de estas factorizaciones para obtener autovalores y autovectores aproximados. Los pares de Ritz, que cumplen la condición de Galerkin, expuesta en (13) se obtienen inmediatamente de los autovalores y autovectores de la matriz H_k.

Si $H_k y = y\theta$ entonces el vector $x = V_k y$ satisface:

$$\|Ax - x\theta\| \quad = \quad |\beta_k e_k^H y| \quad \text{donde} \quad \beta_k = \|f_k\| \tag{26}$$

Siendo (θ, x) un par de Ritz. El número $|\beta_k e_k^H y|$ es llamado la estimación de Ritz. Además:

$$\theta = x^H Ax \tag{27}$$

es un Cociente de Rayleigh (asumiendo $\|x\| = 1$) y se asocia el Cociente Residual de Rayleigh:

$$r(x) = Ax - x\theta \tag{28}$$

que satisface, por ecuación (26):

$$\|r(x)\| = |\beta_k e_k^H y|. \tag{29}$$

3. Iteración truncada RQ (TRQ)

La iteración truncada RQ, está basado en una recursión que se desarrolla en las k primeras columnas de la iteración implícita desplazada RQ. La iteración implícita desplazada RQ, es análogo a la conocida QR. La factorización RQ, se obtendrá por medio de las rotaciones de Givens.

Factorización RQ

Análogo a la factorización QR tenemos la factorización RQ, donde la matriz A cuadrada es descompuesta con un procedimiento opuesto de QR, es decir posmultiplicando a la matriz A por matrices Givens, se generará ceros por columna.

Iteración RQ con desplazamiento

Sea A una matriz cuadrada de orden n y consideremos $A = RQ$ con $R = AG_1G_2 \ldots G_k$ y $Q = G_k^T G_{k-1}^T \ldots G_1^T$, donde R es una matriz triangular superior y Q una matriz ortogonal.

Similar a la iteración *implícita QR desplazada*, se implementará la iteración *implícita RQ desplazada*, como sigue:

Sea A una matriz cuadrada, la cual se quiere aplicar la factorización RQ. Para reducir el número de cálculos aritméticos comprendidos en la iteración RQ, se reducirá la matriz A, a la forma Hessenberg superior, por medio de una serie de transformaciones de semejanzas unitarias, con lo cual se tiene:

$AV = VH$ o sea $H = V^H AV$ ya que $V^H V = I$.

Así H es semejante a A (por Schur).

Iteración básica RQ aplicado a una matriz Hessenberg

Iniciando con

$H_1 \leftarrow Hessenberg(A)$ con

$$H_1 = {V_1}^H A V_1 \text{ tal que } {V_1}^H V_1 = I \tag{30}$$

Aplicando la factorización RQ a la matriz H_1

$$H_1 = R_1 Q_1. \tag{31}$$

Luego se construye H_2 como sigue:

$$H_2 = Q_1 H_1 {Q_1}^H \text{ resulta } V_2 = V_1 Q_1^H \tag{32}$$

Continuando este proceso se llega:

$$H_k = Q_{k-1} H_{k-1} Q_{k-1}^H, \quad \text{con} \quad V_k = (V_{k-1} Q_{k-1}^H) \tag{33}$$

Iteración RQ con desplazamiento

Se define un desplazamiento, como el reemplazo de una matriz A por la matriz $A - \mu I$, y a μ, se le conoce como el desplazamiento.

Se inicia con una reducción de la matriz A, a la forma Hessenberg.

$H_1 \leftarrow Hessenberg(A)$, luego

$$H_1 = V_1^H A V_1 \text{ tal que } {V_1}^H V_1 = I \tag{34}$$

Se inicia con la elección de un desplazamiento μ_1. Aplicando la factorización RQ a la matriz $H_1 - \mu_1 I$

$$H_1 - \mu_1 I = R_1 Q_1 \tag{35}$$

Si se toma H_2 como

$$H_2 = Q_1 H_1 Q_1^H \tag{36}$$

se sigue que $V_2 = V_1 Q_1^H$. Continuando con el proceso, se tiene: se elige un desplazamiento μ_{k-1} y se factoriza la matriz $H_{k-1} - \mu_{k-1} I$ como sigue

$$H_{k-1} = R_{k-1} Q_{k-1} \tag{37}$$

tomando la matriz H_k como

$$H_k = Q_{k-1}H_{k-1}Q_{k-1}^H, \quad \text{con} \quad V_k = (V_{k-1}Q_{k-1}^H) \tag{38}$$

Se observa que $A \sim H_1 \sim H_2 \sim \ldots \sim H_k$. Hasta que H_k sea triangular superior, para algún k. Luego aquellos elementos que queden en su diagonal principal serán sus autovalores de A.

4. Truncamiento de la Iteración RQ

Suponiendo que se tiene la matriz Hessenberg $H^{(j)}$, después de $(j-1)$ iteraciones, entonces

$$AV^{(j)} = V^{(j)}H^{(j)} \quad \text{tal que} \quad {V^{(j)}}^H V^{(j)} = I \tag{39}$$

Continuando con la iteración RQ con desplazamiento:

$$H^{(j)} - \mu_j I = R^{(j)}Q^{(j)} \tag{40}$$

donde $R^{(j)}$ y $Q^{(j)}$: Matriz triangular superior y ortogonal generada en la iteración j.

Luego premultiplicando por $V^{(j)}$ y usando (39) y (40) se tiene

$$(A - \mu_j I)V^{(j+1)} = V^{(j)}R^{(j)}, \quad j \geq 1 \tag{41}$$

De la ecuación (41), la primera columna:

$$(A - \mu_j I)v_1^{(j+1)} = v_1^{(j)}\rho_{1,1}^{(j)}, \text{ para todo } j \geq 1 \tag{42}$$

donde $\rho_{1,1} = e_1^T R e_1$.

Así, la sucesión $v_1^{(j)}$ de la primera columna, es una sucesión iterativa inversa, y se espera una rápida convergencia de las principales columnas de $V^{(j)}$ a vectores Schur de A.

Para matrices muy grandes va ser imposible obtener las iteraciones totales que involucran matrices ortogonales $n \times n$. Se podría truncar estos procedimientos, después de k-pasos y actualizar solo la porción principal de las factorizaciones que ocurren en estas sucesiones. Este truncamiento se obtiene de un conjunto de ecuaciones, que surgen cuando se completa parcialmente un paso RQ.

Para deducir estas relaciones, analicemos la partición

$$V = (V_k, \widehat{V}) \tag{43}$$

Además

$$H = \begin{array}{c} \begin{array}{cc} k & n-k \end{array} \\ \begin{pmatrix} H_k & M \\ \beta_k e_1 e_k^T & \widehat{H} \end{pmatrix} \end{array} \begin{array}{l} \}k \\ \}n-k \end{array} \tag{44}$$

y verifica

$$A(V_k, \widehat{V}) = (V_k, \widehat{V}) \begin{pmatrix} H_k & M \\ \beta_k e_1 e_k^T & \widehat{H} \end{pmatrix} \tag{45}$$

donde H_k es submatriz principal de H de orden k, $\widehat{H}$ submatriz de H de orden $(n-k)$, V_k matriz de orden $(n \times k)$, $\widehat{V}$ matriz de orden $(n \times n-k)$, e_1 y e_k vectores canónico columna de orden $n-k$ y k respectivamente, β_k elemento en la posición $(k+1, k)$ de H.

Usando la ecuación (44), para un desplazamiento μ de H:

$$H = \begin{pmatrix} H_k - \mu I_k & \widehat{M} \\ \beta_k e_1 e_k^T & \widehat{R} \end{pmatrix} \begin{pmatrix} I_k & 0 \\ 0 & \widehat{Q} \end{pmatrix} + \mu I$$

que reemplazando en la relación $AV = VH$ y factorizando se tiene:

$$(A - \mu I)(V_k, \widehat{V}) = (V_k, \widehat{V}) \begin{pmatrix} H_k - \mu I_k & \widehat{M} \\ \beta_k e_1 e_k^T & \widehat{R} \end{pmatrix} \begin{pmatrix} I_k & 0 \\ 0 & \widehat{Q} \end{pmatrix}$$

posmultiplicando por $\begin{pmatrix} I_k & 0 \\ 0 & \widehat{Q} \end{pmatrix}^H$, ambos lados, se tiene como resultado:

$$(A - \mu I)(V_k, \widehat{V}\widehat{Q}^H) = (V_k, \widehat{V}) \begin{pmatrix} H_k - \mu I_k & \widehat{M} \\ \beta_k e_1 e_k^T & \widehat{R} \end{pmatrix} \tag{46}$$

Hasta este punto se usó transformaciones Givens, en la formación de $\widehat{R}$ y $\widehat{Q}$. Se puede seguir aplicando rotaciones Givens para aniquilar los elementos subdiagonales de H. Sin embargo en este punto de las factorizaciones, hay un conjunto de ecuaciones que únicamente se determina por la primera columna de $\widehat{V}\widehat{Q}^H$. Si estas ecuaciones se pueden formular y solucionar, entonces las $(n-k-1)$ columnas que quedan de V y de H no necesitan ser formuladas ni factorizadas.

Para formular estas relaciones, se debe igualar las primeras $(k+1)$ columnas de ambos lados de la ecuación (46).

Con el fin de sólo ver las $(k+1)$ columnas de la ecuación (46), se restringe:

$$(A - \mu I) \underbrace{(V_k, \widehat{V}\widehat{Q}^H)}_{(k+1)\ columnas} = (V_k, \widehat{V}) \underbrace{\begin{pmatrix} H_k - \mu I_k & \widehat{M} \\ \beta_k e_1 {e_k}^T & \widehat{R} \end{pmatrix}}_{(k+1)\ columnas}$$

sea $v = \widehat{V}e_1$, $v_+ = \widehat{V}\widehat{Q}^H e_1$, $h = \widehat{M}e_1$, $\alpha = e_1^T \widehat{R} e_1$.

Igualamos las $(k+1)$ primeras columnas de ambos lados, de la ecuación (46) se tiene

$$(A - \mu I)(V_k, v_+) = (V_k, v) \begin{pmatrix} H_k - \mu I_k & h \\ \beta_k e_k^T & \alpha \end{pmatrix} \tag{47}$$

que es llamada la **ecuación de reducción TRQ**.

Esta ecuación matricial representa la forma reducida de la iteración RQ truncada.

Regresando a la ecuación (47), se sigue que v_+ satisface:

$$(A - uI)v_+ = V_k h + v\alpha \tag{48}$$

Como las columnas de (V_k, v_+) deben ser ortonormales: ${V_k}^H v_+ = 0$ y $\| v_+ \| = 1$.
Luego ecuación (48) se expresa como

$$\begin{pmatrix} A - uI & V_k \\ {V_k}^H & 0 \end{pmatrix} \begin{pmatrix} v_+ \\ -h \end{pmatrix} = \begin{pmatrix} v\alpha \\ 0 \end{pmatrix}, \tag{49}$$

que es llamada la **ecuación de reducción TRQ**, con $\|v_+\| = 1$ y esta ecuación es la ecuación matricial TRQ.

Nótese que son desconocidos v_+, h y α, que tendrán como soluciones a:

$$w = -(A-\mu I)^{-1}V_k z \quad v_+ = \frac{w}{\| w \|} \quad h = \frac{-z}{\| w \|} \quad \alpha = 0,$$

y satisfacen las ecuaciones TRQ.

Las ecuaciones TRQ se pueden usar para desarrollar una versión truncada en el paso k de la iteración implícita RQ desplazada. Si se tiene un paso k de la factorización de Arnoldi, entonces un paso k de la iteración TRQ se puede implementar en el siguiente algoritmo.

Algoritmo 1 Iteración Truncada RQ (TRQ)
Entrada: A, V_k, H_k, f_k con $AV_k = V_k H_k + f_k {e_k}^T$, ${V_k}^H V_k = I$

H_k: Hessenberg superior

Salida: V_k, H_k tal que $AV_k = V_k H_k$, $V_k^H V_k = I$, H_k triangular superior.

$\beta_k \leftarrow \|f_k\|$, $v \leftarrow f_k/\beta_k$

Para $j = 1, 2, 3, \ldots$ hasta que converja **hacer**

Selección de un desplazamiento μ_j, $\mu \leftarrow \mu_j$

Solución: $\begin{pmatrix} A-\mu I & V_k \\ {V_k}^H & 0 \end{pmatrix}\begin{pmatrix} v_+ \\ -h \end{pmatrix} = \begin{pmatrix} v\alpha \\ 0 \end{pmatrix}$,

con $\|v_+\| = 1$

Poniendo en factores RQ:

$$\begin{pmatrix} H_k - \mu I_k & h \\ \beta_k e_k^T & \alpha \end{pmatrix} = \begin{pmatrix} R_k & r \\ 0 & \rho \end{pmatrix}\begin{pmatrix} Q_k & q \\ \sigma e_k^T & \gamma \end{pmatrix}$$

$V_k \leftarrow V_k Q_k^H + v_+ q^H$

$\beta_k \leftarrow \sigma e_k^T R_k e_k$, $v \leftarrow v_k\bar{\sigma} + v_+\bar{\gamma}$

$H_k \leftarrow Q_k R_k + \mu I_k$

Fin de hacer

5. Implementación

En esta sección, se tratará algunas asociaciones prácticas con la implementación eficiente de la iteración TRQ.

Solución de las ecuaciones TRQ

La iteración truncada RQ será efectiva, si hay un intermedio eficiente para solucionar las ecuaciones TRQ.

Recordando que: A, H_k, V_k y $f_k = v\beta_k$ son los elementos que están en el paso k de la relación de Arnoldi. De las ecuaciones TRQ se tiene:

$$\begin{pmatrix} A-\mu I & V_k \\ V_k^H & 0 \end{pmatrix}\begin{pmatrix} w \\ z \end{pmatrix} = \begin{pmatrix} f_k \\ 0 \end{pmatrix} \qquad (50)$$

Se quiere evitar solucionar el sistema usando eliminación Gaussiana, para ello se busca una solución directa. Un algoritmo que da solución directa a las ecuaciones TRQ, se puede ver en [7].

Selección del desplazamiento

Otro tema importante a ser tratado en la iteración TRQ es la elección de desplazamientos. Se sugiere ver [7].

6. Iteración Inexacta TRQ

Si el costo de la factorización $A - \mu I$ es moderado, la Iteración TRQ provee un método claro y eficiente de obtener las aproximaciones para los autovalores. De otro modo debe recurrirse a otros métodos para solucionar la ecuación TRQ (49). Un candidato natural es una solución iterativa precondicionada. Se presentará un algoritmo basado en la idea de incorporar una solución iterativa en la Iteración TRQ, y se analizará la convergencia de este método.

Primero se examinará la consecuencia de reemplazar la solución exacta, v_+ de (49) con alguna aproximación $\widetilde{v}_+$.

Iteración Inexacta TRQ

De (50) se tiene que

$$w = (I - V_k V_k^H)(A-\mu I)^{-1}V_k s \qquad (51)$$

sea

$$w_k = (A-\mu I)^{-1}V_k s \qquad (52)$$

entonces

$$(A-\mu I)w_k = V_k s, \text{ para algún } s \neq 0 \qquad (53)$$

de esta ecuación se desea encontrar w_k. Sea $\widetilde{w}_k$ la solución a al ecuación (53) por aplicación de alguna solución iterativa y por la ecuación (51)

$$\widetilde{w} = (I - V_k V_k^H)\widetilde{w}_k \qquad \widetilde{w} \approx w \qquad (54)$$

Sea

$$\widetilde{v}_+ = \frac{\widetilde{w}}{\|\widetilde{w}\|} \qquad (55)$$

Se sabe que:

$$v_+ = \frac{w}{\|w\|} \qquad (56)$$

donde

$\widetilde{v}_+$ – es la solución aproximada para la ecuación (49).
v_+ – es la solución exacta para la ecuación (49).

La ortogonalización de la ecuación (54) y normalización de $\widetilde{w}$ garantizan que:

$$V_k^H \widetilde{v}_+ = 0 \quad \text{y} \quad \|\widetilde{v}_+\| = 1 \qquad (57)$$

Para continuar con la iteración TRQ, se deben calcular $\widetilde{h}$ y $\widetilde{\alpha}$ (ecuación (49)), de tal manera que verifiquen:

$$(A-\mu I)\widetilde{v}_+ = V_k\widetilde{h} + v\widetilde{\alpha} \quad \text{tal que} \quad (\widetilde{h}, \widetilde{\alpha}) \approx (h, \alpha) \qquad (58)$$

Sin embargo, el lema siguiente indica que generalmente es difícil encontrar un perfecto par $(\widetilde{h}, \widetilde{\alpha})$ para la ecuación (58).

Lema 2. *Supóngase que se soluciona la ecuación (53) por el método de Subespacio de Krylov para obtener una aproximación $\widetilde{w}$. Si $\widetilde{w} = (I - V_k V_k^H)\widetilde{w}_k \neq 0$, entonces $(A-\mu I)\widetilde{v}_+ \notin span\{v_1, v_2, \ldots, v_k, v\}$ donde $V_k = [v_1, \ldots, v_k]$*

Demostración:
Sean

$$V_k s \in span\{v_1, Av_1, A^2v_1, \ldots, A^{k-1}v_1\}$$
$$v \in span\{v_1, Av_1, A^2v_1, \ldots, A^k v_1\}.$$

Entonces

$$V_k s = \sum_{i=0}^{k-1} \alpha_i A^i v_1 \quad \text{osea} \quad V_k s = p(A)v_1 \tag{59}$$

para algún polinomio $p(\lambda)$ de grado a lo más $k-1$.

De la ecuación (53) se tiene $w_k = (A-\mu I)^{-1}V_k s$ que se soluciona aproximadamente por el método del subespacio de Krylov

$$w_k \approx q(A)V_k s$$

y sea

$$\widetilde{w}_k = q(A)V_k s \tag{60}$$

donde $q(\lambda)$ es otro polinomio asociado con la solución lineal krylov. Se sigue de las ecuaciones (60) y (59) que

$$\widetilde{w}_k = q(A)p(A)v_1$$

Haciendo

$$\psi(\lambda) = q(\lambda)p(\lambda)$$

Luego

- Si el $Grado(\psi) = k-1$ entonces
  $\widetilde{w}_k = \psi(A)v_1 \in span\{v_1, v_2, \ldots, v_k\}$.
  Se tiene que $(I - V_k V_k^H)V_k = 0$, es decir $(I - V_k V_k^H)$ es ortogonal a V_k y como $\widetilde{w}_k \in span\{v_1, v_2, \ldots, v_k\}$, entonces
  $$(I - V_k V_k^H)\widetilde{w}_k = 0 \tag{61}$$
  De las ecuaciones (54) y (61)
  $$\widetilde{w} = (I - V_k V_k^H)\widetilde{w}_k = 0 \tag{62}$$
  que es una contradicción con la hipótesis.

Por lo tanto: $Grado(\psi) \geq k$

Sea $z = V_k^H \widetilde{w}_k$. Ya que el espacio $\{v_1, v_2, \ldots, v_k\}$ es un subespacio de Krylov de dimension k, asociado con A y v_1, el vector $V_k z$ ($V_k z \in span\{v_1, v_2, \ldots, v_k\}$) puede expresarse como

$$V_k z = r(A)v_1 \tag{63}$$

para algún polinomio $r(\lambda)$ de grado a lo mas $k-1$.

Por lo tanto, si se hace

$$\beta = \|(I - V_k {V_k}^H)\widetilde{w}_k\| \tag{64}$$

entonces:

$$\begin{aligned}
(A-\mu I)\widetilde{v}_+ &= (A-\mu I)\frac{\widetilde{w}}{\|\widetilde{w}\|} \\
&= (A-\mu I)(I - V_k V_k^H)\widetilde{w}_k/\beta \text{ de (54) y (64)} \\
&= (A-\mu I)[\widetilde{w}_k - V_k z]/\beta \\
&= (A-\mu I)[\psi(A)v_1 - r(A)v_1]/\beta \text{ de (63)} \\
&= (A-\mu I)[\psi(A) - r(A)]v_1/\beta \\
&= \phi(A)v_1
\end{aligned}$$

donde:

$$\phi(\lambda) = (\lambda - \mu)[\psi(\lambda) - r(\lambda)]/\beta$$

ϕ : polinomio de grado al menos de $k+1$.

De donde se concluye que

$$(A-\mu I)\widetilde{v}_+ \notin span\{v_1, Av_1, \ldots, A^k v_1\} \qquad \blacksquare$$

De (58) se deduce:

$$\widetilde{h} = V_k^H A\widetilde{v}_+ \quad \text{y} \quad \widetilde{\alpha} = v^H(A-\mu I)\widetilde{v}_+$$

Ecuación de reducción Inexacta TRQ

Debido al error sobrante en la ecuación (58), la reducción RQ truncada, ecuación (47) ahora es inexacta. Se puede expresar esta reducción inexacta por

$$(A-\mu I)(V_k, \widetilde{v}_+) = (V_k, v)\begin{pmatrix} H_k - \mu I_k & \widetilde{h} \\ \beta_k e_k^T & \widetilde{\alpha} \end{pmatrix} + z e_{k+1}^T \tag{65}$$

igualando la ultima columna de (65):

$$z \equiv (A-\mu I)\widetilde{v}_+ - (V_k, v)\begin{pmatrix} \widetilde{h} \\ \widetilde{\alpha} \end{pmatrix} \tag{66}$$

definiéndose a z como el error residual, tenemos:

$$V_k^H z = 0 \quad \text{y} \quad v^H z = 0 \tag{67}$$

Ahora se procede a aplicar una sucesión de rotaciones Givens por la derecha a (65), estas afectarán solo a las columnas, con el fin de eliminar los elementos de la subdiagonal de la matriz

$$\begin{pmatrix} H_k - \mu I_k & \widetilde{h} \\ \beta_k e_k^T & \widetilde{\alpha} \end{pmatrix}$$

el error residual variará en todas las columnas de V_k. Por lo tanto, los vectores bases no son vectores de Arnoldi muy válidos. Sin embargo, como se mostrará posteriormente, la primera columna v_1^+ de esta base actualizada satisface $(A-\mu I)v_1^+ = \rho_{11}v_1 + z\sigma$ donde σ es un producto de senos asociado con las rotaciones Givens anteriormente mencionadas.

Esta observación nos muestra que una iteración inversa aproximada permanece en cada iteración inexacta TRQ.

El error asociado con esta iteración inversa probablemente se considere menor que $\|z\|$, debido al factor σ. De donde, una simple solución para corregir las bases de Arnoldi, es recalcular una factorización de Arnoldi de la primera columna de la actualizada V_k.

7. Análisis de Convergencia

Esta sección está enfocado en el análisis de convergencia del esquema inexacto TRQ. En particular es de interés comprender la precisión de la solución a la ecuación TRQ y la proporción de convergencia da cada autopar en la iteración TRQ. En toda la iteración TRQ usamos desplazamientos del cociente de Rayleigh, establecemos la convergencia lineal local del primer vector base de Arnoldi a un autovector de A. El factor de convergencia depende de $\|\delta z\|$, la magnitud del error residual de amortiguamiento en (66), y de la distancia entre dos autovalores consecutivos buscados.

Análisis de Convergencia

Para iniciar el análisis, se asume que una reducción Hessenberg inexacta (65) ya se obtuvo. La siguiente matriz corresponde a la ecuación inexacta de la ecuación (65)

$$\begin{pmatrix} H_k - \mu I_k & \widetilde{h} \\ \beta_k e_k^T & \widetilde{\alpha} \end{pmatrix} \in M(k+1, k+1) \tag{68}$$

al que se quiere eliminar sus elementos subdiagonales. Para el efecto se aplicaran rotaciones Givens. Para el análisis se aplicará $k-1$ rotaciones de orden $k+1$, $Q_1^H, Q_2^H, \ldots, Q_{k-1}^H$, donde cada uno tiene la forma:

$$Q_i^H = \begin{pmatrix} I_{k-1} & & & 0 \\ & \gamma_i & \sigma_i & \\ & -\sigma_i & \gamma_i & \\ 0 & & & I_{k-1} \end{pmatrix} \tag{69}$$

donde: $\sigma_i^2 + \gamma_i^2 = 1, i = 1, 2, \ldots, k-1$; σ – función seno, γ – función coseno.

Estas rotaciones se aplicarán por el lado derecho a la ecuación inexacta (65). Sea

$$P_{k-1} = Q_1^H Q_2^H \ldots Q_{k-1}^H \tag{70}$$

ver [7]

$$P_{k-1} = \begin{pmatrix} I_1 & 0 \\ 0 & \gamma_{k-1} \\ 0 & -\gamma_{k-2}\sigma_{k-1} \\ 0 & \gamma_{k-3}\sigma_{k-2}\sigma_{k-1} \\ \vdots & \vdots \\ 0 & (-1)^k \gamma_1 \sigma_2 \sigma_3 \ldots \sigma_{k-1} \\ 0 & (-1)^{k+1} \sigma_1 \sigma_2 \ldots \sigma_{k-1} \end{pmatrix}_{(k+1)\times(k+1)} \tag{71}$$

Multiplicando (71) por el lado derecho a la ecuación (65) se tiene:

$$(A - \mu I) \underbrace{(V_k, \widetilde{v}_+) P_{k-1}}_{parte1} = \underbrace{z e_{k+1}^T P_{k-1}}_{parte3} + \underbrace{(V_k, v) \begin{pmatrix} H_k - \mu I_k & \widetilde{h} \\ \beta_k e_k^T & \widetilde{\alpha} \end{pmatrix} P_{k-1}}_{parte2} \tag{72}$$

y sea $parte4 = (V_k, v)(parte2)$.

Se necesita encontrar los términos $parte1$, $parte2$, $parte3$ y $parte4$ mencionados en la ecuación (72).

(a) Con ayuda de la ecuación (71):

$$parte1 = (V_k, \widetilde{v}_+) P_{k-1} = (v_1, \widetilde{v}_2) \tag{73}$$

donde $\widetilde{v}_2$ representa la segunda columna de $(V_k, \widetilde{v}_+) P_{k-1}$, luego

$$\widetilde{v}_2 = (V_k, \widetilde{v}_+) Q_1^H \ldots Q_{k-1}^H e_2, \quad \text{por (70)} \tag{74}$$

Luego se demuestra que:

$$\widetilde{v}_2^H \widetilde{v}_2 = 1 \tag{75}$$

de la ecuación (73), $\widetilde{v}_2$ tiene la siguiente forma

$$\widetilde{v}_2 = (v_2, v_3, \cdots v_k, \widetilde{v}_+) \begin{pmatrix} \gamma_{k-1} \\ -\gamma_{k-2}\sigma_{k-1} \\ \vdots \\ (-1)^{k+1} \sigma_1 \ldots \sigma_{k-1} \end{pmatrix} \tag{76}$$

(b) Considerando que P_{k-1} se formó para eliminar los $k-1$ elementos subdiagonales de la matriz

$$\begin{pmatrix} H_k - \mu I & \widetilde{h} \\ \beta_k e_k^T & \widetilde{\alpha} \end{pmatrix}_{(k+1,k+1)}$$

entonces los elementos $\{h_{32}, \ldots, h_{(k,k-1)}, \beta_k\}$ se hacen ceros, luego sean

$$\eta = h_{12}\gamma_{k-1} + \ldots + \widetilde{h}_1((-1)^{k+1}\sigma_1 \ldots \sigma_{k-1})$$
$$\rho = (h_{22} - \mu)\gamma_{k-1} + \ldots + \widetilde{h}_2((-1)^{k+1}\sigma_1 \ldots \sigma_{k-1})$$

entonces

$$parte2 = \begin{pmatrix} h_{11} - \mu & \eta \\ h_{21} & \rho \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix} \tag{77}$$

de (72)

$$parte4 = v_1 \; v_2 \begin{pmatrix} h_{11} - \mu & \eta \\ h_{21} & \rho \end{pmatrix} \tag{78}$$

para simplificar mas la ecuación (78):

(i) Considerando el desplazamiento del cociente de Rayleigh:

$$\mu = h_{11} \tag{79}$$

(ii) De la ecuación (47), igualando solo la primera columna:

$$(A - \mu I) v_1 = v_2 h_{21} \tag{80}$$

(iii) Luego

$$\|(A-\mu I)v_1\|^2 \quad = \quad h_{21}^2, \quad \text{por (80)} \qquad (81)$$

Sea:

$$\epsilon = \|(A-\mu I)v_1\| \qquad (82)$$

Regresando a la ecuación (78):

$$parte4 \quad = \quad (v_1, v_2)\begin{pmatrix} 0 & \eta \\ \epsilon & \rho \end{pmatrix} \qquad (83)$$

(c) De la ecuación (72):

$$parte3 = ze_{k+1}^T P_{k-1} = (\overline{0}, z\widehat{\sigma}) \qquad (84)$$

donde: $\widehat{\sigma} = (-1)^{k+1}\sigma_1 \ldots \sigma_{k-1}$, $z = [z_1, \ldots, z_n]^T$

Regresando a la ecuación matricial (72) y reemplazando las ecuaciones obtenidas (73), (83) y (84), se tiene:

$$(A-\mu I)(v_1, \widetilde{v}_2) = (v_1, v_2)\begin{pmatrix} 0 & \eta \\ \epsilon & \rho \end{pmatrix} + (0, z\widehat{\sigma}) \qquad (85)$$

donde: $\widetilde{v}_2 = (V_k, \widetilde{v}_+)Q_1^H Q_2^H \ldots Q_{k-1}^H e_2$,
$\widehat{\sigma} = (-1)^{k+1}\sigma_1\sigma_2 \ldots \sigma_{k-1}$,
$\epsilon = \|(A-\mu I)v_1\|$.
Sean:

$$\tau = \sqrt{\epsilon^2+\rho^2}, \qquad \sigma_k = \frac{\epsilon}{\tau}, \qquad \gamma_k = \frac{\rho}{\tau} \qquad (86)$$

La ecuación (85) es el resultado de eliminar $(k-1)$
coeficientes subdiagonales, faltando aún un coeficiente por eliminar. Para ello se aplica una rotación Givens de orden dos, por el lado derecho a (85). Sea

$$Q_k^H = \begin{pmatrix} \gamma_k & \sigma_k \\ -\sigma_k & \gamma_k \end{pmatrix}$$

multiplicando Q_k^H por el lado derecho a (85)

$$(A-\mu I)\underbrace{(v_1, \widetilde{v}_2)\begin{pmatrix} \gamma_k & \sigma_k \\ -\sigma_k & \gamma_k \end{pmatrix}}_{(v_1^+, \widetilde{v}_2^+)} =$$

$$(v_1, v_2)\begin{pmatrix} 0 & \eta \\ \epsilon & \rho \end{pmatrix}\begin{pmatrix} \gamma_k & \sigma_k \\ -\sigma_k & \gamma_k \end{pmatrix} + (0, z\widehat{\sigma})\begin{pmatrix} \gamma_k & \sigma_k \\ -\sigma_k & \gamma_k \end{pmatrix}$$

donde

$$v_1^+ = \gamma_k v_1 - \sigma_k \widetilde{v}_2 \quad \text{y} \quad \widetilde{v}_2^+ = \sigma_k v_1 + \gamma_k \widetilde{v}_2 \qquad (87)$$

simplificando

$$(A-\mu I)(v_1^+, \widetilde{v}_2^+) = (v_1, v_2)\begin{pmatrix} -\sigma_k\eta & \gamma_k\eta \\ \frac{\rho}{\tau}\epsilon - \frac{\epsilon}{\tau}\rho & \frac{\epsilon}{\tau}\epsilon + \frac{\rho}{\tau}\rho \end{pmatrix} + (-\sigma_k z\widehat{\sigma}, z\widehat{\sigma}\gamma_k)$$

Ecuación que resulta de eliminar los elementos subdiagonales de la ecuación inexacta TRQ

La última ecuación produce:

$$(A-\mu I)(v_1^+, \widetilde{v}_2^+) = (v_1, v_2)\begin{pmatrix} -\sigma_k\eta & \gamma_k\eta \\ 0 & \tau \end{pmatrix} + (-\sigma_k z\widehat{\sigma}, z\widehat{\sigma}\gamma_k) \qquad (88)$$

donde

$$v_1^+ = \gamma_k v_1 - \sigma_k \widetilde{v}_2 \quad \text{y} \quad \widetilde{v}_2^+ = \sigma_k v_1 + \gamma_k \widetilde{v}_2 \qquad (89)$$

Esta ecuación es el resultado de eliminar los k elementos subdiagonales de la ecuación (65) (ecuación inexacta TRQ).

Análisis de convergencia

Se analizará la convergencia de la iteración inexacta TRQ examinando la norma de

$$r_+ = (A-\mu_+ I)v_1^+ \quad \text{donde} \quad \mu_+ = (v_1^+)^H A v_1^+ \qquad (90)$$

Se define el error residual amortiguado ν como:

$$\nu = \|\widehat{\sigma}z\| \qquad (91)$$

Notese que $|\widehat{\sigma}| = |\sigma_1\sigma_2 \ldots \sigma_{k-1}| < 1$, esto es porque cada σ_i es una función seno usado para construir las rotaciones de Givens en la ecuación (69).

Teorema 2. *Sea $r = (A-\mu I)v_1$ y $r_+ = (A-\mu_+ I)v_1^+$, donde v_1^+ es la definida en (89) y μ, μ_+ son cocientes de Rayleigh de A con respecto a v_1 y v_1^+ respectivamente, μ_+ es la definida en (90).*

(a) Si cada $A-\mu I$ es no singular y μ es convergente a un autovalor simple de A, entonces

$$\|r_+\| \leq \psi(\mu, \nu)\|r\| \qquad (92)$$

donde la magnitud de la función ψ depende de μ y del tamaño del error amortiguado ν definido en (91).

(b) Sea $V \equiv (V_k, \widehat{V}_{n-k})$ unitario, donde V_k consiste de vectores base de Arnoldi generados por el paso (2.7) en el algoritmo 4. Se particiona V como $V = (v_1, \widehat{V}_{n-1})$, y sean

$$C = \widehat{V}_{n-1}^H A \widehat{V}_{n-1} \quad y \quad \xi = \|(C-\mu I)^{-1}\|^{-1} \qquad (93)$$

si $\nu < \xi$, entonces

$$|\psi(\mu, \nu)| \leq \frac{|\epsilon\eta|}{\xi^2-\nu^2} + \frac{|\epsilon|\nu}{\xi^2-\nu^2} + \frac{\nu}{\sqrt{\xi^2-\nu^2}} \qquad (94)$$

donde

$$\epsilon = \|(A-\mu)v_1\| \quad y \quad \eta = v_1^H A v_2$$

(c) Más aún, si $\nu < \dfrac{\xi}{\sqrt{2}}$, entonces $|\psi(\mu, \nu)| < 1$.

Demostración:

(a) De la ecuación (88), igualando las primeras columnas se tiene:

$$(A-\mu I)v_1^+ = v_1(-\sigma_k\eta) - \widehat{\sigma}\sigma_k z \qquad (95)$$

para claridad, se deja los subíndices de σ_k y γ_k en lo que sigue. Luego por (90) y por (95):

$$r_+ = (-\sigma\eta)v_1 + (\mu-\mu_+)v_1^+ + (-z\widehat{\sigma})\sigma \qquad (96)$$

deduciéndose relaciones necesarias:

$$V_k^H z = 0,\ v_1^H z = 0,\ v_1^H \widetilde{v}_2 = 0,\ {v_1^+}^H v_1^+ = 1.$$

La distancia entre μ_+ y μ se puede estimar como sigue:

$$\mu_+ - \mu = (v_1^+)^H A v_1^+ - \mu = -\gamma\sigma\eta + \sigma^2\widehat{\sigma}\widetilde{v}_2^H z \quad (97)$$

Se transformará r_+ a $V^H r_+$, antes cuidando y observando su norma.

Como V es unitario $V^H V = I = V V^H$. Luego:

$$\|r_+\|^2 = r_+^H V V^H r_+ = \|V^H r_+\|^2 \quad (98)$$

Se particiona $V = (v_1, \widehat{V}_{n-1})$.

Luego sean:

$$p = \widehat{V}_{n-1}^H \widetilde{v}_2 \quad \text{y} \quad \widehat{z} = \widehat{V}_{n-1}^H z \quad (99)$$

donde:

z – vector columna de orden n.

$\widehat{z}$ – vector columna de orden $(n-1)$.

Se necesitan los siguiente resultados para simplificar la expresión de $V^H r_+$:

$\widehat{V}_{n-1}^H v_1 = \overline{0}$, $V^H v_1^+ = \begin{pmatrix} \gamma \\ -\sigma p \end{pmatrix}$,

$V^H z = \begin{pmatrix} 0 \\ \widehat{z} \end{pmatrix} \begin{matrix} \}1 \\ \}n-1 \end{matrix}$,

$V^H z = \begin{pmatrix} \overline{0} \\ \widehat{V}^H z \end{pmatrix} \begin{matrix} \}k \\ \}n-k \end{matrix}$,

Particionando

$\widehat{V} = \left(v, \widehat{V}_{n-(k+1)}\right)$, $V^H z = \begin{pmatrix} 0 \\ \widehat{z} \end{pmatrix} \begin{matrix} \}1 \\ \}n-1 \end{matrix}$.

Y se deduce que las k primeras componentes de $\widehat{z}$ son ceros.

$\|\widehat{z}\|^2 = \|z\|^2$, $V^H v_1 = e_1$, $\widehat{V}_{n-1}\widehat{V}_{n-1}^H = I - v_1 v_1^H$,

$\|p\|^2 = 1$, $\left\| \begin{pmatrix} \sigma \\ \gamma p \end{pmatrix} \right\|^2 = \sigma^2 + \gamma^2 p^H p = 1$,

$\left\| \begin{pmatrix} \gamma \\ -\sigma p \end{pmatrix} \right\|^2 = \gamma^2 + \sigma^2 p^H p = 1$

Así:

$$V^H r_+ = (-\sigma\eta) \begin{pmatrix} \sigma^2 \\ \gamma\sigma p \end{pmatrix} + -\sigma^2(\widetilde{v}_2^H z)\widehat{\sigma} \begin{pmatrix} \gamma \\ -\sigma p \end{pmatrix} - \sigma \begin{pmatrix} 0 \\ \widehat{z}\widehat{\sigma} \end{pmatrix} \quad (100)$$

Luego $\|r_+\| \le \sigma^2|\eta| + \sigma^2\nu + |\sigma|\nu$, obteniéndose dos relaciones importantes:

$$\|r_+\| \le \sigma^2|\eta| + \sigma^2\|\widehat{\sigma}z\| + |\sigma|\|\widehat{\sigma}z\| \quad (101)$$

$$\|r_+\| \le \sigma^2|\eta| + \sigma^2\nu + |\sigma|\nu \quad (102)$$

Recordando que σ es generado para eliminar el elemento subdiagonal de la matriz

$$\begin{pmatrix} 0 & \eta \\ \epsilon & \rho \end{pmatrix},$$

que aparece en (85). Además por (86):

$$|\sigma| = \frac{|\epsilon|}{\sqrt{\epsilon^2 + \rho^2}} \le |\frac{\epsilon}{\rho}| \quad (103)$$

Igualando la definición de r dado en la hipótesis del teorema (2) con la ecuación (82), se tiene:

$$|\epsilon| = \|r\| = \|(A - \mu I)v_1\| \quad (104)$$

Luego:

$$\|r_+\| \le (\frac{|\epsilon\eta|}{\rho^2} + \frac{|\epsilon|\nu}{\rho^2} + \frac{\nu}{|\rho|})\|r\|$$

Se concluye que $\|r_+\| \le \psi(\mu,\nu)\|r\|$

donde $\psi(\mu,\nu) = \dfrac{|\epsilon\eta|}{\rho^2} + \dfrac{|\epsilon|\nu}{\rho^2} + \dfrac{\nu}{|\rho|}$

Claramente, el factor $\psi(\mu,\nu)$ puede ser acotado uniformemente si ν no es muy grande, y si $|\rho|$ puede ser acotado fuera del cero. Por supuesto, no podríamos saber el tamaño de ρ hasta la aplicación de $k-1$ rotaciones $Q_1, Q_2, \ldots, Q_{k-1}$.

(b) Los argumentos siguientes proveen una cota mínima a priori para $|\rho|$. Esto asegura que $|\rho|$ puede ser acotado inferiormente si ν es suficientemente pequeño.

Igualando la primera columna de la ecuación (88) se tiene:

$$(A - \mu I)v_1^+ = v_1(-\sigma\eta) + (-\sigma z\widehat{\sigma})$$

premultiplicando por V^H y usando el factor $VV^H = I$, por ser V unitario

$$\underbrace{V^H(A - \mu I)\underbrace{VV^H}_{I}v_1^+}_{factor1} = \underbrace{V^H v_1(-\sigma\eta) + V^H(-\sigma z\widehat{\sigma})}_{factor2} \quad (105)$$

Veamos algunas relaciones necesarias para simplificar: $v_1^H(A - \mu I)v_1 = 0$,

$\widehat{V}_{n-1}^H(A - \mu I)v_1 = \widehat{V}_{n-1}^H v_2 h_{21} = \epsilon e_1$, luego:

$$factor1 = \begin{pmatrix} 0 & h^H \\ \epsilon e_1 & C - \mu I \end{pmatrix} \begin{pmatrix} \gamma \\ -\sigma p \end{pmatrix} \quad (106)$$

donde

$$h^H = v_1^H A \widehat{V}_{n-1} \quad (107)$$

Luego se tiene:

$$factor2 = V^H v_1(-\sigma\eta) + V^H(-\sigma z\widehat{\sigma}) = \begin{pmatrix} -\sigma\eta \\ -\sigma\widehat{\sigma}\widehat{z} \end{pmatrix} \quad (108)$$

reemplazando en (105):

$$\begin{pmatrix} 0 & h^H \\ \epsilon e_1 & C - \mu I \end{pmatrix} \begin{pmatrix} \gamma \\ -\sigma p \end{pmatrix} = \begin{pmatrix} -\sigma\eta \\ -\sigma\widehat{\sigma}\widehat{z} \end{pmatrix} \quad (109)$$

Una primera ecuación de (109):

$$h^H p = \eta \tag{110}$$

Entonces:

$$\eta = h^H p = v_1^H A\widetilde{v}_2 \tag{111}$$

luego:

$$\rho e_1 - (C - \mu I)p = -\widehat{\sigma}\widehat{z} \tag{112}$$

Recordando que $\widehat{z}$ tiene sus k primeras componentes ceros, entonces $e_1^T\widehat{z} = 0$.

Además $\widehat{z} = (0, \breve{z})^T$, donde $\breve{z} \in M(n-2, 1)$, $0 \in \mathbb{R}$ y se observa que $\breve{z}$, tiene $k-1$ ceros.

Luego de (112):

$$(C - \mu I)p = \rho e_1 + \widehat{\sigma}\widehat{z} = \begin{pmatrix} \rho \\ \widehat{\sigma}\breve{z} \end{pmatrix} \tag{113}$$

Se demuestra que $(C - \mu I)$ es no singular, ver [7].

De esta manera de (113):

$$p = (C - \mu I)^{-1} \begin{pmatrix} \rho \\ \widehat{\sigma}\breve{z} \end{pmatrix} \tag{114}$$

De la ecuación (99), $p = \widehat{V}_{n-1}^H \widetilde{v}_2$ y p tiene longitud 1, de donde

$$1 = \|p\| \leq \|(C - \mu I)^{-1}\| \sqrt{\rho^2 + \widehat{\sigma}^2 \|z\|^2}, \quad \text{por (113)}$$

entonces

$$\left\| \frac{1}{(C - \mu I)^{-1}} \right\| \leq \sqrt{\rho^2 + \nu^2} \tag{115}$$

De (93), $\xi = \dfrac{1}{\|(C - \mu I)^{-1}\|}$.

Luego en (115) se tiene

$$\xi \leq \sqrt{\rho^2 + \nu^2} \tag{116}$$

Por la condición del teorema, si $\nu < \xi$ y elevando al cuadrado a (116), se tiene $0 < \xi^2 - \nu^2 \leq \rho^2$, tomando valor absoluto y sacando la raíz cuadrada, queda $\sqrt{\xi^2 - \nu^2} \leq |\rho|$ en consecuencia de $\dfrac{1}{\rho^2} \leq \dfrac{1}{\xi^2 - \nu^2}$ y como

$$\psi(\mu, \nu) = \frac{|\epsilon\eta|}{\rho^2} + \frac{|\epsilon|\nu}{\rho^2} + \frac{\nu}{|\rho|}$$

entonces

$$\psi(\mu, \nu) \leq \frac{|\epsilon\eta|}{\xi^2 - \nu^2} + \frac{|\epsilon|\nu}{\xi^2 - \nu^2} + \frac{\nu}{\sqrt{\xi^2 - \nu^2}} \tag{117}$$

donde por (104) y (111):

$$|\epsilon| = \|(A - \mu I)v_1\|$$
$$\eta = v_1^H A\widetilde{v}_2$$

(c) Por hipótesis del teorema, μ es el cociente de Rayleigh de A con respecto a v_1:

$$\mu = \frac{v_1^T A v_1}{v_1^T v_1} \Rightarrow v_1^T(Av_1 - \mu v_1) = 0 \tag{118}$$

y como μ es lo más cercano a un autovalor deseado, hipótesis (a):

$$Av_1 \approx v_1\mu \tag{119}$$

Por (104), (118) y (119)

$$|\epsilon| = \|(A - \mu I)v_1\| \approx 0 \tag{120}$$

Luego en la ecuación (117) se puede ignorar los dos primeros términos por (120) y enfocarnos en el tercer término:

$$\psi(\mu, \nu) \leq \frac{\nu}{\sqrt{\xi^2 - \nu^2}}$$

Por condición del teorema, si $\nu < \frac{\xi}{\sqrt{2}}$, tenemos:

$$\nu^2 < \frac{\xi^2}{2} \Rightarrow \xi^2 - \nu^2 > \frac{\xi^2}{2} \Rightarrow \frac{1}{\sqrt{\xi^2 - \nu^2}} < \frac{\sqrt{2}}{\xi} \tag{121}$$

Luego:

$$|\psi(\mu, \nu)| \leq |\nu| \frac{1}{\sqrt{\xi^2 - \nu^2}} < |(\frac{\xi}{\sqrt{2}})(\frac{\sqrt{2}}{\xi})| = 1$$

En este caso se puede espera la convergencia monótona.

□

8. Conclusiones

El análisis de convergencia realizado, es un análisis local.

La ecuación inexacta TRQ (88), es inexacta porque $z \neq 0$ e implica que el factor $(-\sigma_k z\widehat{\sigma}, z\widehat{\sigma}\gamma_k) \neq 0$ y este es lo que hace que la iteración TRQ sea inexacta.

Si suponemos que la ecuación inexacta TRQ (88) es exacta, es decir, $z = 0$ entonces $(-\sigma_k z\widehat{\sigma}, z\widehat{\sigma}\gamma_k) = 0$ tenemos de (88):

$$(A - \mu I)v_1^+ = v_1(-\sigma\eta) \tag{122}$$

por (90) y (122):

$$r_+ = v_1(-\sigma\eta) + (\mu - \mu_+)v_1^+ \tag{123}$$

luego $\mu_+ - \mu = -\gamma\sigma\eta$, $V^H r_+ = (-\sigma\eta)\begin{pmatrix} \sigma^2 \\ -\sigma\gamma p \end{pmatrix}$, $\|r_+\| = |\eta|\sigma^2|$ como $\widehat{z} = \widehat{V}_{n-1}^H z$ y se está considerando que $z = 0$ entonces de la ecuación (112):

$$\rho e_1 - (C - \mu I)p = 0 \text{ entonces } \|p\| = \|\rho e_1(C - \mu I)^{-1}\| \tag{124}$$

y como se sabe que $\|p\| = 1$. Luego:

$$\frac{1}{|\rho|} \leq \|(C - \mu I)^{-1}\| \tag{125}$$

Obteniéndose:

$$\frac{\|r_+\|}{\|r\|^2} \leq |\eta| \|(C - \mu I)^{-1}\|^2 < \infty \tag{126}$$

y se muestra que el método tiene convergencia cuadrática. Cuando la matriz A es hermitiana se tiene:

$$\widehat{V}_{n-1}^H(A-\mu I)v_1 = \epsilon e_1,$$

entonces:

$$\widehat{V}_{n-1}^H(A-\mu I)v_1 = \widehat{V}_{n-1}^H Av_1 - \mu \underbrace{\widehat{V}_{n-1}^H v_1}_{0} = \widehat{V}_{n-1}^H Av_1 = \epsilon e_1$$

entonces:

$$\widehat{V}_{n-1}^H Av_1 = \epsilon e_1 \tag{127}$$

luego por ser A hermitiana: $\|h\|^2 = \|r\|^2$ y $|\eta| = \|r\|$. Regresando a la ecuación (126)

$$\frac{\|r_+\|}{\|r\|^2} \leq |\eta|\|(C-\mu I)^{-1}\|^2 \leq \|r\|\|(C-\mu I)^{-1}\|^2 \tag{128}$$

entonces $\dfrac{\|r_+\|}{\|r\|^3} \leq \|(C-\mu I)^{-1}\|^2 < \infty$ consiguiéndose la proporción de convergencia cúbica.

Para problemas Hermitianos, ξ es aproximadamente, la distancia entre el autovalor para el cual la inexacta TRQ converge al autovalor más cercano a este. Esta cantidad puede estimarse examinando a $|\mu - \hat{\mu}|$, donde $\widehat{\mu}$ es el autovalor de H_k más cercano a μ.

La iteración Inexacta TRQ, bajo algunas condiciones que se asumen, converge linealmente con un factor de convergencia pequeño ver [7].

1. Chao Yang, *Convergence Analysis of an inexact truncated RQ-iteration.*
2. D.C. Sorensen And C. Yang, *A truncated RQ-iteration for large scale eigenvalue calculations*, SIAM J. Matrix Anal. Appl., 19(4):1045-1073, 1998.
3. Gene H. Golub And Charles F. Van Loan, *Matrix Computations Second Edition.*
4. *Carlos Chávez Vega,* Algebra Lineal *1993.*
5. *David Kincaid Y Ward Cheney,* Análisis Numérico, Las matamáticas del cálculo científico *1993.*
6. *George Oliveira Ainsworth Junior,* Desenvolvimento de um algoritmo baseado no método de Arnoldi para solução de problemas de autovalor generalizado *Rio de Janeiro, RJ-Brasil Abril de 2003.*
7. *Cristina Navarro Flores,* Análisis de Convergencia de una Iteración Inexacta TRQ. *PERU, UNI 2005.*
8. *Kyle A. Gallivan,* Set 10 Krylov Methods-Arnoldi-based. *School of Computational Science Florida State University - 2005.*
9. *Yousef Saad,* Iterative Methods for sparse. Linear Systems *Second Edition with corrections. January 3RD, 2000.*