NOTICIERO

ISSN 1514-9595 (web)

Miradas Matemáticas

Subespacios admisibles y aproximaciones de matrices

Pedro Massey
Centro de Matemática de La Plata-FCEx-UNLP
& IAM-CONICET

Resumen

En esta nota breve comentamos algunos aspectos relacionados con el cálculo de aproximaciones de matrices rectangulares $A \in \mathbb{C}^{m\times n}$ a través de matrices $B \in \mathbb{C}^{m\times n}$ que verifican la restricción $\mathrm{rk}(B) = \dim B(\mathbb{C}^n) \leq h$. El análisis clásico de varios métodos iterativos para la construcción de aproximaciones de la forma $B = P_{\mathcal{T}}A$ para subespacios $\mathcal{T} \subset \mathbb{C}^m$ con $\dim \mathcal{T} = h$, se basa en estimaciones de la distancia entre el subespacio $\mathcal{T}$ y el llamado subespacio dominante a izquierda de A de dimensión h. Sin embargo, es sabido que el cálculo de aproximaciones de la forma $B = P_{\mathcal{T}}A$ puede estar bien condicionado en casos en que el cálculo de aproximaciones del subespacio dominante de dimensión h de A está mal condicionado, lo que afecta al análisis clásico de los métodos iterativos. En lo que sigue desarrollamos algunas cuestiones relacionadas con esta aparente paradoja y compartimos algunas de las consideraciones que, en trabajo conjunto con Francisco Arrieta Zuccalli, nos han llevado a proponer una visión alternativa para el análisis de los métodos iterativos en términos de una nueva clase de subespacios.

Introducción

En la actualidad, la información digitalizada se presenta frecuentemente en términos de una matriz $A \in \mathbb{C}^{m\times n}$. Estas matrices de información suelen ser de gran tamaño, lo que corresponde a que modelan grandes volúmenes de datos. En ese contexto, las matrices no pueden ser cargadas de forma completa en la memoria rápida de las computadoras y típicamente se accede solo a partes de estas matrices (por ejemplo a filas y columnas). Este último hecho hace que sea posible calcular la acción de la matriz en vectores Ax, lo que a su vez permite definir expresiones del tipo A^*A ó AA^* y sus potencias (que sólo siguen siendo parcialmente accesibles en la memoria rápida de una computadora). Sin embargo, en general no es posible calcular de forma explícita algunas formas canónicas o representaciones conocidas del álgebra lineal: por ejemplo, no es posible calcular de forma directa la descomposición polar, ó la descomposición en valores singulares de A. En general, tampoco es posible calcular los llamados vectores singulares e incluso los valores singulares de forma directa y completa.

Por otro lado, el hecho de que la matriz A provenga de datos reales muchas veces garantiza que su información más relevante puede quedar representada por matrices con cierta estructura especial. Por ejemplo, la información representada por A contiene unas pocas tendencias bien marcadas; ó la información representada por A tiene una precisión demasiado alta para nuestro interés, y podemos pensar en comprimir esta información de forma que la versión comprimida sea aún de interés. En este contexto, la idea general es tratar de aproximar esta información relevante de A a través de matrices B, que admitan una manipulación algebraica directa en la computadora. Por supuesto, el planteo anterior es muy general; en lo que sigue vamos a considerar un modelo bien definido para describir el problema de aproximar la información más relevante de A y veremos algunos de los problemas asociados a su cálculo. También describiremos algunos de los problemas teóricos asociados a este tipo de modelo y plantearemos una posible perspectiva (nueva) para su análisis.

1. Aproximaciones ideales de A con restricciones

A partir de la matriz $A \in \mathbb{C}^{m\times n}$ proponemos representar su información más relevante (con cierto nivel de precisión) a través de matrices $B \in \mathbb{C}^{m\times n}$, que tengan una estructura más sencilla. En nuestro contexto, vamos a elegir de forma conveniente matrices $B \in \mathbb{C}^{m\times n}$, que satisfagan $\mathrm{rk}(B) = \dim B(\mathbb{C}^n) \leq h$, donde $1 \leq h \ll \text{mín}\{m, n\}$ es un parámetro que podemos establecer en función de distintos criterios (fijo de aquí en adelante). En principio, estamos interesados en elegir las matrices B_{op} que sean soluciones del problema de mejor aproximación con restricciones:

$$\|A - B_{\text{op}}\| = \text{mín}\{\|A - B\| \, : \, B \in \mathbb{C}^{m\times n} \, , \, \mathrm{rk}(B) \leq h\} \, .$$

En este contexto, dada una matriz C, $\|C\|$ denotará la norma espectral $\|C\|_{\text{sp}} = \text{máx}\{\|Cx\| : \|x\| \leq 1\}$ o la norma Frobenius $\|A\|_F = \text{tr}(A^*A)^{1/2}$. Las matrices B_{op} extraen la información más relevante de A, en el sentido de que el residuo de aproximación $\|A - B_{\text{op}}\|$ es lo más pequeño posible.

Aquí conviene notar que la restricción sobre el rank de B_{op} permite deducir representaciones útiles (computables) de esta matriz. De hecho, para cualquier $B \in \mathbb{C}^{m\times n}$ con $\mathrm{rk}(B) \leq h$ podemos calcular de forma efectiva representaciones de $B = F_1 F_2^*$, donde $F_1 \in \mathbb{C}^{m\times h}$ y $F_2 \in \mathbb{C}^{n\times h}$ son matrices que tienen solo h columnas (recordemos que $h \ll \text{mín}\{m, n\}$), lo que permite calcular representaciones canónicas de B con la computadora. Esto indica que el modelo propuesto más arriba da una respuesta al problema de hallar la manera de manipular de forma computable aproximaciones convenientes de A.

Por otro lado, notamos que las soluciones B_{op} del problema de mejor aproximación pueden no ser únicas y pueden depender de la norma de matrices que usemos. Sin embargo, existen matrices $A_h \in \mathbb{C}^{m\times n}$ que son mejores aproximaciones de A bajo las restricciones anteriores, con respecto a las normas Frobenius y espectral (y muchas otras más) simultáneamente. Para poder describir estas matrices A_h, necesitamos detallar una descomposición muy importante del análisis matricial, denominada descomposición en valores singulares.

Para comenzar, recordemos que los valores singulares de $A \in \mathbb{C}^{m\times n}$, notados por $\sigma_1(A) \geq \cdots \geq \sigma_p(A)$, donde $p = \text{mín}\{m, n\}$, están dados por $\sigma_i^2 = \lambda_i(A^*A)$, $1 \leq i \leq p$, donde $\lambda_1(A^*A) \geq \cdots \geq \lambda_n(A^*A) \geq 0$ denotan los autovalores de la matriz semidefinida positiva $A^*A \in \mathbb{C}^{n\times n}$, contando multiplicidades y ordenados en forma decreciente. A partir de los valores singulares podemos describir una descomposición en valores singulares (DVS) de A como una factorización $A = U\Sigma V^*$, donde $U \in \mathbb{C}^{m\times m}$ y $V \in \mathbb{C}^{n\times n}$ son matrices unitarias, y $\Sigma \in \mathbb{R}^{m\times n}$ es una matriz *diagonal* que satisface $\Sigma_{ii} = \sigma_i(A)$ $(= \sigma_i)$, para $1 \leq i \leq p$. Si denotamos por $u_1, \dots, u_m$ y $v_1, \dots, v_n$ las columnas de las matrices U y V en la DVS de A anterior, entonces se puede verificar que $Au_i = \sigma_i v_i$, para $1 \leq i \leq p$. Los vectores u_i y v_i son llamados los i-ésimos vectores singulares a izquierda y derecha de A, respectivamente. El hecho de que $u_1, \dots, u_m$ sean las columnas de una matriz unitaria implica que forman una base ortonormal de $\mathbb{C}^m$; de forma similar, $v_1, \dots, v_n$ es una base ortonormal de $\mathbb{C}^n$. Un resultado básico, pero muy importante, del análisis matricial garantiza la existencia de descomposiciones en valores singulares para toda matriz $A \in \mathbb{C}^{m\times n}$ (ver [1]).

La descomposición en valores singulares es una herramienta fundamental. Como consecuencia de esta descomposición, vemos que los valores singulares pueden ser usados para medir la matriz A. De hecho, usando esta descomposición y las propiedades de las matrices unitarias y la traza, es sencillo verificar que $\|A\|_{\text{sp}} = \sigma_1$ mientras que $\|A\|_F^2 = \sum_{i=1}^p \sigma_i^2$.

Notemos que si $A = U\Sigma V^*$ es una DVS entonces $A^*A = V\Sigma^*\Sigma V^*$, donde $D = \Sigma^*\Sigma \in \mathbb{C}^{n\times n}$ es la matriz diagonal tal que $D_{ii} = \sigma_i^2$, $1 \leq i \leq p$ y $D_{ii} = 0$, $p + 1 \leq i \leq n$. En este sentido, los vectores singulares a derecha $v_1, \dots, v_n$ corresponden a una diagonalización de A^*A a través de una base ortonormal. Más aún, es sencillo verificar que toda diagonalización $A^*A = \tilde{V}D\tilde{V}^*$ de A^*A corresponde a una DVS tal que $A = \tilde{U}\Sigma\tilde{V}^*$. Es bien sabido que si los autovalores de A^*A tienen multiplicidad mayor a uno, entonces existen diagonalizaciones de A^*A que son muy diferentes entre sí: por ejemplo, si $\sigma_1^2 = \lambda_1(A^*A)$ tiene multiplicidad dos, entonces podemos reemplazar a los vectores $v_1, v_2 \in \mathbb{C}^n$ de cualquier diagonalización $A^*A = VDV^*$ (con D como antes) por los vectores $\tilde{v}_1, \tilde{v}_2 \in \mathbb{C}^n$ tales que formen una base ortonormal del subespacio de $\mathbb{C}^n$ generado por $\{v_1, v_2\}$ y obtener una nueva diagonalización $A^*A = \tilde{V}D\tilde{V}^*$. A su vez, esto implicará que tendremos DVS's $A = U\Sigma V^*$ y $A = \tilde{U}\Sigma\tilde{V}^*$, con sus correspondientes vectores singulares muy diferentes entre sí. Por otro lado, el hecho de que exista un salto singular $\sigma_h = \lambda_h(A^*A)^{1/2} > \sigma_{h+1} = \lambda_{h+1}(A^*A)^{1/2}$ garantiza que el subespacio generado por $\{v_1, \dots, v_h\}$ está unívocamente asociado a A^*A (y por lo tanto a A). Notemos que podemos repetir todos los argumentos anteriores, pero ahora en términos de los vectores singulares a izquierda $\{u_1, \dots, u_m\}$ y la matriz semidefinida positiva $AA^* \in \mathbb{C}^{m\times m}$.

Finalmente, dada cualquier DVS de $A = U\Sigma V^*$ y $1 \leq h \leq p = \text{mín}\{m, n\}$ podemos definir la DVS truncada correspondiente, notada $A_h \in \mathbb{C}^{m\times n}$ y dada por $A_h = U_h \Sigma_h V_h^*$ donde $U_h \in \mathbb{C}^{m\times h}$ tiene columnas $u_1, \dots, u_h$, $V_h \in \mathbb{C}^{n\times h}$ tiene columnas $v_1, \dots, v_h$, y $\Sigma_h \in \mathbb{C}^{h\times h}$ es una matriz diagonal con $(\Sigma_h)_{ii} = \sigma_i$, $1 \leq i \leq h$. Notemos que $\text{rk}(A_h) \leq h$, por construcción. El llamado Teorema de Lidskii (ver [1]), que es un resultado central de la teoría de aproximación de matrices, garantiza las desigualdades

$$\|A - A_h\|_{\text{sp}} \leq \|A - B\|_{\text{sp}} \quad \text{y} \quad \|A - A_h\|_F \leq \|A - B\|_F \;\; \text{para toda } B \in \mathbb{C}^{m\times n}\,,\ \text{rk}(B) \leq h\,.$$

De hecho, tenemos que el residuo óptimo puede calcularse:

$$\|A - A_h\|_{\text{sp}} = \sigma_{h+1} \quad \text{y} \quad \|A - A_h\|_F = (\sum_{i=h+1}^{p} {\sigma_i}^2)^{1/2}\,. \tag{1.1}$$

Por otro lado, como consecuencia de la construcción de A_h se deduce de forma directa que

$$\sigma_i(A_h) = \sigma_i \ , \quad 1 \le i \le h \ \text{ y } \ \sigma_i(A_h) = 0 \ , \quad h+1 \le i \le p \, . \tag{1.2}$$

Notemos que la no unicidad de la DVS en presencia de valores singulares repetidos (múltiples), mencionada más arriba, implica la no unicidad de la matriz A_h en estos casos. Así, en general no hay una única A_h (y la notación puede resultar confusa). Por otro lado, la condición $\sigma_h > \sigma_{h+1}$ garantiza la unicidad de la mejor aproximación A_h, que resulta independiente de la DVS de A que hayamos considerado.

2. Aproximaciones computables de A con restricciones

Dada una matriz $A \in \mathbb{C}^{m\times n}$ (de tamaño grande), hemos visto en la sección anterior que existen matrices $A_h \in \mathbb{C}^{m\times n}$ que son las mejores aproximaciones de A entre las matrices $B \in \mathbb{C}^{m\times n}$ bajo la restricción $\mathrm{rk}(B) \le h$, para $1 \le h \ll p = \min\{m, n\}$ fijo. Cabe notar que estas matrices A_h se construyen a través de cualquier DVS de A. Sin embargo, como hemos mencionado en la introducción, en los casos de interés no es posible calcular (de forma explícita, usando una computadora) una DVS de A, debido a su tamaño. Para poder plantear una salida a esta encrucijada, conviene notar cierta estructura de las matrices A_h: dada una DVS $A = U\Sigma V^*$, si $U_h \in \mathbb{C}^{m\times h}$ denota la matriz con columnas $u_1, \dots , u_h$ entonces $A_h = U_h(U_h^* A)$: en efecto,

$$\begin{aligned} U_h(U_h^* A) &= U_h(U_h^* U)\Sigma V^* = U_h \begin{pmatrix} I_h & 0_{h\times m-h} \end{pmatrix} \Sigma V^* = U_h \, \Sigma_h \begin{pmatrix} I_h \\ 0_{n-h\times h} \end{pmatrix} V^* \\ &= U_h \, \Sigma_h (V \begin{pmatrix} I_h & 0_{h\times n-h} \end{pmatrix})^* = U_h \, \Sigma_h \, V_h^* = A_h \, . \end{aligned}$$

La representación anterior muestra el rol que cumple el subespacio $\mathcal{U}_h \subset \mathbb{C}^m$ generado por $\{u_1, \dots , u_h\} \subset \mathbb{C}^m$, denominado subespacio dominante a izquierda de A de dimensión h: de hecho, el producto $U_h U_h^* \in \mathbb{C}^{m\times m}$ coincide con el proyector ortogonal sobre este subespacio. Más aún, si fuera posible calcular U_h entonces la representación anterior $A_h = U_h(U_h^* A) = F_1 F_2^*$, con $F_1 = U_h \in \mathbb{C}^{m\times h}$ y $F_2 = A^* U_h \in \mathbb{C}^{n\times h}$, permitiría realizar manipulaciones algebraicas con A_h a través de manipulaciones de $F_2^* = U_h^* A$ que es una matriz con pocas filas.

Por supuesto, el problema aún subsiste, dado que el cálculo de los vectores $\{u_1, \dots , u_h\}$ depende del cálculo de una DVS. Sin embargo, hay numerosos resultados del álgebra lineal numérica que garantizan aproximaciones del subespacio dominante a izquierda $\mathcal{U}_h$ a través de subespacios que pueden calcularse con el uso de la computadora. Esta última afirmación (que desarrollaremos más adelante) hace que esta perspectiva para el cálculo de matrices $B = P_{\mathcal{T}} A$ con $\dim \mathcal{T} = h$, y tales que sean buenas aproximaciones de A, sea computacionalmente viable.

Para llevar adelante el enfoque de aproximación anterior es necesario definir alguna noción de distancia (ó proximidad) entre subespacios. Vamos a considerar dos situaciones diferentes, según los subespacios tengan la misma dimensión o no, y las nociones en cada uno de estos casos se presentan de forma diferenciada. El primer paso es identificar un subespacio $\mathcal{S} \subset \mathbb{C}^n$ con su proyección ortogonal $P_{\mathcal{S}} \in \mathbb{C}^{n\times n}$, que satisface $P_{\mathcal{S}} = {P_{\mathcal{S}}}^* = {P_{\mathcal{S}}}^2$ y $P_{\mathcal{S}}(\mathbb{C}^n) = \mathcal{S}$. Es sencillo verificar que esta identificación es buena, en el sentido de que es una biyección entre subespacios y proyecciones ortogonales. Así, dados dos subespacios $\mathcal{S}, \mathcal{T} \subset \mathbb{C}^n$ de la misma dimensión, podemos definir la distancia entre ellos como $d(\mathcal{S}, \mathcal{T}) = \|P_{\mathcal{S}} - P_{\mathcal{T}}\|_{\mathrm{sp}}$: a partir de la definición (y del hecho de que $\| \cdot \|_{\mathrm{sp}}$ es una norma en $\mathbb{C}^{n\times n}$) se puede verificar que $d(\mathcal{S}, \mathcal{T})$ es una función distancia (o métrica) en

$$\text{Gr(n,k)} := \{\mathcal{S} \subset \mathbb{C}^n : \dim \mathcal{S} = k\},$$

para $1 \leq k \leq n$. Por otro lado, si $\dim \mathcal{S} < \dim \mathcal{T}$ entonces definimos la proximidad (o cercanía) de $\mathcal{S}$ a $\mathcal{T}$ dada por

$$d(\mathcal{S}, \mathcal{T}) = \|(I - P_{\mathcal{T}})P_{\mathcal{S}}\|_{\text{sp}} .$$

Notamos que cuando los subespacios tienen distinta dimensión, sus roles en la expresión $\|(I - P_{\mathcal{T}})P_{\mathcal{S}}\|_{\text{sp}}$ no son simétricos. Hemos usado la misma notación $d(\mathcal{S}, \mathcal{T})$ ya sea que $\dim \mathcal{S} = \dim \mathcal{T}$ o $\dim \mathcal{S} < \dim \mathcal{T}$. Una justificación de este hecho es la siguiente: dados dos subespacios cualesquiera $\mathcal{S}, \mathcal{T} \subset \mathbb{C}^n$ tales que $\dim \mathcal{S} \leq \dim \mathcal{T}$, podemos definir la apertura (o ángulo máximo) entre los subespacios como el valor

$$\theta_{\text{máx}}(\mathcal{S}, \mathcal{T}) = \text{máx}\{\arccos(|\langle s, t\rangle|) : s \in \mathcal{S}, t \in \mathcal{T}, \|s\| = \|t\| = 1\} \in [0, \pi/2] .$$

Es claro que esta cantidad describe la posición relativa entre los subespacios, es decir, es una noción geométrica. En este caso, sucede que $d(\mathcal{S}, \mathcal{T}) = \sin(\theta_{\text{máx}}(\mathcal{S}, \mathcal{T})) \in [0, 1]$, ya sea que las dimensiones de los subespacios sean iguales o no.

Ahora que hemos definido la noción de distancia o proximidad entre subespacios, podemos describir algunos de los resultados clásicos del álgebra lineal numérica relacionados con la aproximación de subespacios dominantes. Vamos a considerar un análisis determinista de un proceso iterativo que describimos a continuación; más aún, vamos a iniciar considerando un caso que llamaremos genérico (refiriéndonos a que se da con muchísima frecuencia), pero que no cubre todos los casos. Partimos de una DVS (arbitraria, pero fija) $A = U\Sigma V^* \in \mathbb{C}^{m\times n}$ y consideramos la dimensión objetivo $1 \leq h \ll p = \text{mín}\{m, n\}$. Vamos a asumir que, además, $\sigma_h > \sigma_{h+1}$, de forma que la mejor aproximación A_h está unívocamente determinada. Dado un subespacio inicial $\mathcal{X} \subset \mathbb{C}^n$ y tal que $h \leq r := \dim \mathcal{X} \ll p = \text{mín}\{m, n\}$ (genérico), entonces (ver [2, 9]) se verifica que para todo $q \geq 0$: si $\mathcal{U}_h$ y $\mathcal{V}_h$ son los subespacios dominantes a izquierda y derecha de A de dimensión h, generados por $\{u_1, \dots, u_h\}$ y $\{v_1, \dots, v_h\}$ respectivamente, entonces

$$d(\mathcal{U}_h, (AA^*)^q A(\mathcal{X})) \leq \left(\frac{\sigma_{h+1}}{\sigma_h}\right)^{2q+1} \tan(\theta_{\text{máx}}(\mathcal{X}, \mathcal{V}_h)) . \tag{2.1}$$

Lo anterior puede interpretarse de la siguiente manera: comenzando con el subespacio inicial $\mathcal{T}_1 = (AA^*)A(\mathcal{X})$ de dimensión baja (a lo sumo $r \ll p$), calculamos de forma iterativa los subespacios $\mathcal{T}_\ell = (AA^*)(\mathcal{T}_{\ell-1})$, para $2 \leq \ell \leq q$. De esta forma, la sucesión de subespacios satisface: $\mathcal{T}_q = (AA^*)^q A(\mathcal{X})$. En particular, dado que $\frac{\sigma_{h+1}}{\sigma_h} < 1$, los subespacios $\mathcal{T}_q$ aproximan arbitrariamente bien al subespacio $\mathcal{U}_h$, es decir, $d(\mathcal{U}_h, \mathcal{T}_q) \to 0$ cuando $q \to \infty$. Remarcamos que los subespacios dominantes no juegan ningún papel en la construcción de $\mathcal{T}_q$. El resultado anterior, basado en el llamado método de subespacios iterativos para la construcción de los subespacios $\mathcal{T}_q$, es el primer paso para la construcción computable de aproximaciones de A, dadas por matrices $B \in \mathbb{C}^{m\times n}$ que satisfacen que $\text{rk}(B) \leq h$. Si comenzamos con un subespacio genérico $\mathcal{X} \subset \mathbb{C}^n$ con $\dim \mathcal{X} = h$ entonces, de forma genérica, $\dim \mathcal{T}_q = h$; en tal caso $d(\mathcal{U}_h, \mathcal{T}_q) = \|P_{\mathcal{U}_h} - P_{\mathcal{T}_q}\|_{\text{sp}}$ y entonces, por ejemplo:

$$\|A - P_{\mathcal{T}_q}A\|_{\text{sp}} \leq \|A - A_h\|_{\text{sp}} + \|A_h - P_{\mathcal{T}_q}A\|_{\text{sp}} \leq \sigma_h + d(\mathcal{U}_h, \mathcal{T}_q)\,\|A\|_{\text{sp}} ,$$

donde usamos que $A_h = P_{\mathcal{U}_h}A$ y que

$$\|P_{\mathcal{U}_h}A - P_{\mathcal{T}_q}A\|_{\text{sp}} \leq \|P_{\mathcal{U}_h} - P_{\mathcal{T}_q}\|_{\text{sp}}\,\|A\|_{\text{sp}} = d(\mathcal{U}_h, \mathcal{T}_q)\,\|A\|_{\text{sp}} .$$

De esta forma, los residuos $\|A - P_{\mathcal{T}_q}A\|_{\rm sp}$ convergen al residuo óptimo $\|A - A_h\|_{\rm sp}$, cuando q se hace arbitrariamente grande. A partir de la Ec. (2.1) es posible concluir que $\sigma_i(P_{\mathcal{T}_q}A) \to \sigma_i$, cuando q se hace arbitrariamente grande, para $1 \le i \le h$.

3. El rol del salto singular $\sigma_h - \sigma_{h+1}$

Con las notaciones de la sección anterior, el análisis del comportamiento de la sucesión de subespacios $\mathcal{T}_q = (AA^*)^q A(\mathcal{X})$ que aparece en la Ec. (2.1) está fuertemente influenciado por el valor $\sigma_{h+1}/\sigma_h < 1$. Si este valor es muy cercano a 1, entonces la estimación (superior) de la distancia $d(\mathcal{U}_h, \mathcal{T}_q)$ no será realmente informativa.

Una primera pregunta es si se trata de una limitación del análisis realizado para llegar a la Ec. (2.1). Sucede que este no es el caso: desde un punto numérico, ejemplos realizados muestran que los valores (reales) de $d(\mathcal{U}_h, (AA^*)^q A(\mathcal{X}))$ dependen de cociente σ_{h+1}/σ_h y que cuando este cociente se acerca a 1, entonces $d(\mathcal{U}_h, (AA^*)^q A(\mathcal{X}))$ se mantiene cerca de 1, incluso para valores grandes de q. Por otro lado, desde un punto de vista teórico, se puede mostrar que el problema del cálculo de $\mathcal{U}_h$ se vuelve inestable, cuando el cociente σ_{h+1}/σ_h se acerca a 1 (i.e. cuando el salto singular $\sigma_h - \sigma_{h+1}$ se hace pequeño). Como medida de estabilidad del cálculo (en nuestro caso, de la aproximación) del subespacio $\mathcal{U}_h$, se puede considerar el llamado número de condición, que definimos a continuación. Dada $A \in \mathbb{C}^{m\times n}$ tal que $\sigma_h > \sigma_{h+1}$ como antes, consideramos $\mathcal{M}_h = \{C \in \mathbb{C}^{m\times n} :\ \sigma_h(C) > \sigma_{h+1}(C)\} \subset \mathbb{C}^{m\times n}$ (que resulta un conjunto abierto de $\mathbb{C}^{m\times n}$ con respecto a la topología inducida por la norma $\|\cdot\|_{\rm sp}$). Para cada $C \in \mathcal{M}_h$, notamos por $\mathcal{U}_h(C) \subset \mathbb{C}^m$ al subespacio dominante a izquierda de dimensión h, que está unívocamente definido. En este contexto, el número de condición del cálculo de $\mathcal{U}_h = \mathcal{U}_h(A)$ está dado por

$$\kappa(A, h) = \lim_{\varepsilon\to 0^+} \sup_{C\in\mathcal{M}_h,\, 0<\|A-C\|_{\rm sp}\le\varepsilon} \frac{d(\mathcal{U}_h, \mathcal{U}_h(C))}{\|A - C\|_{\rm sp}} \tag{3.1}$$

y satisface que $\kappa(A, h) = \alpha(\sigma_h(A) - \sigma_{h+1}(A))^{-1}$, para una constante $1 \le \alpha \le 2$ (ver [8]). Es decir, en la medida en que el salto singular $\sigma_h - \sigma_{h+1}$ sea muy pequeño (ó equivalentemente, en la medida en que σ_{h+1}/σ_h esté muy cercano a 1) el cálculo de $\mathcal{U}_h = \mathcal{U}_h(A)$ se vuelve muy inestable, en el sentido de que para matrices C cercanas a A, $\mathcal{U}_h(C)$ puede estar lejos de $\mathcal{U}_h$.

Cabe remarcar que hay otros métodos de aproximación de $\mathcal{U}_h$ a través de sucesiones de subespacios $\tilde{\mathcal{T}}_q \subset \mathbb{C}^m$ que son efectivamente computables: sin embargo, los análisis clásicos (del caso determinista) de estos algoritmos producen estimaciones de la distancia $d(\mathcal{U}_h, \tilde{\mathcal{T}}_q)$ que también dejan de ser informativas cuando $\sigma_h - \sigma_{h+1}$ se hace pequeño (ver [2]). De forma análoga, ejemplos numéricos de estos algoritmos muestran que los valores $d(\mathcal{U}_h, \tilde{\mathcal{T}}_q)$ efectivamente dependen del salto singular $\sigma_h - \sigma_{h+1}$ en el sentido anterior. Más aún, en numerosos ejemplos numéricos se verifica que $d(\mathcal{T}_{q-1}, \mathcal{T}_q)$ se mantiene por arriba de un umbral (inferior) estrictamente positivo como función de $q \ge 1$, lo que indica que la sucesión no se aproxima a ningún subespacio; más bien, parece vagar por $\mathrm{Gr}(n, h)$.

Es ahora cuando conviene recordar nuestro problema de aproximación inicial: dada la matriz A, estamos interesados en calcular aproximaciones computables B de la matriz A que satisfagan la restricción $\mathrm{rk}(B) \le h$ y tales que $\|A - B\|_{2,F}$ sea del orden de $\|A - A_h\|_{2,F}$ (error de aproximación óptimo). Hasta ahora, hemos considerado la estrategia clásica de resolver el problema anterior a través de aproximaciones de la forma $P_{\mathcal{T}} A$, donde $\mathcal{T} \subset \mathbb{C}^m$ denota un subespacio computable y tal que $\dim \mathcal{T} = h$ y $d(\mathcal{U}_h, \mathcal{T})$ sea pequeña. Pero si el salto singular $\sigma_h - \sigma_{h+1}$ es muy pequeño, entonces esta estrategia de resolver un problema auxiliar de calcular subespacios $\mathcal{T}_q$ que se acerquen a $\mathcal{U}_h$ (cuando q crece) no parece conveniente, porque el problema de aproximar el subespacio $\mathcal{U}_h$ se vuelve inestable. Un planteo natural es el de cambiar la estrategia y resolver el problema inicial de forma más directa, al menos en los casos en donde $\sigma_h - \sigma_{h+1}$ es muy pequeño.

Sin embargo, pasa algo insospechado: los ejemplos numéricos realizados sobre matrices $A \in \mathbb{C}^{m\times n}$ tales que $\sigma_h - \sigma_{h+1}$ es muy pequeño (incluso cuando $\sigma_h = \sigma_{h+1}$!), muestran que las aproximaciones de la forma $P_{\mathcal{T}_q}A$ obtenidas a partir del método del subespacio iterativo, i.e. $\mathcal{T}_q = (AA^*)^q A(\mathcal{X})$ para un subespacio genérico inicial $\mathcal{X} \subset \mathbb{C}^n$ con $\dim \mathcal{X} = h$, son buenas aproximaciones de A en el sentido de que los errores de aproximación $\|A - P_{\mathcal{T}_q}A\|_{2,F}$ decrecen hasta mantenerse muy cerca de $\|A - A_h\|_{2,F}$, cuando q crece. Más aún, lo anterior muestra que los valores singulares $\sigma_i(P_{\mathcal{T}_q}A)$ crecen hasta valores muy cercanos a $\sigma_i(A)$ cuando q se hace grande, para $1 \le i \le h$. Estos hechos parecen entrar en contradicción con el comportamiento de la distancia $d(\mathcal{U}_h, \mathcal{T}_q)$ mencionado más arriba (ver [6]).

Como veremos más adelante, los hechos anteriores pueden explicarse a través de la clase de subespacios admisibles a izquierda $\mathrm{Adm}_l(A,h)$ asociada a A y h, cuando $\sigma_h \approx \sigma_{h+1}$ (incluso si $\sigma_h = \sigma_{h+1}$): esta clase, que definiremos en breve, está formada por subespacios $\mathcal{S} \subset \mathbb{C}^m$, cada uno de los cuales induce una buena aproximación $P_{\mathcal{S}}A$ de A, y de forma que $\mathrm{Adm}_l(A,h)$ es lo suficientemente abundante como para garantizar que $\text{mín}\{d(\mathcal{T}_q, \mathcal{S}) : \ \mathcal{S} \in \mathrm{Adm}_l(A,h)\} \to 0$, cuando $q \to 0$. Estos hechos, junto con propiedades elementales de continuidad, validan teóricamente que el cálculo de aproximaciones de la forma $P_{\mathcal{T}_q}A$ es relevante en estos casos.

4. El caso extremo $\sigma_h = \sigma_{h+1}$.

Continuando con el contexto de las secciones anteriores, partamos ahora de una matriz $A \in \mathbb{C}^{m\times n}$, pero supongamos que $\sigma_h = \sigma_{h+1}$. En este caso definimos $j+1 = \text{mín}\{1 \le i \le h : \ \sigma_i = \sigma_h\}$ y $k = \text{máx}\{h \le i \le p : \ \sigma_i = \sigma_h\}$. En lo que sigue vamos a suponer que $1 \le j < h < k < \mathrm{rk}(A)$: de esta forma,

$$\sigma_j > \sigma_{j+1} = \cdots = \sigma_h = \cdots = \sigma_k > \sigma_{k+1} > 0\,. \tag{4.1}$$

Los saltos de los valores singulares que figuran arriba indican que los subespacios dominantes a izquierda $\mathcal{U}_j, \mathcal{U}_k \subset \mathbb{C}^m$ así como los subespacios dominantes a derecha $\mathcal{V}_j, \mathcal{V}_k \subset \mathbb{C}^n$ están unívocamente determinados. Por otro lado, hay una infinidad de subespacios dominantes a izquierda de A de dimensión h: más aún, en este caso los podemos describir a todos mediante el conjunto

$$\mathrm{Dom}_l(A,h) = \{\mathcal{U}_j \oplus \mathcal{D} : \ \mathcal{D} \subset \mathcal{U}_k \ominus \mathcal{U}_j\ ,\ \dim \mathcal{D} = h - j\} \subset \mathbb{C}^m\,, \tag{4.2}$$

donde $\mathcal{U}_k \ominus \mathcal{U}_j = \mathcal{U}_k \cap \mathcal{U}_j^{\perp}$. Una primera observación a tener en cuenta es que la clase $\mathrm{Dom}_l(A,h) \subset \mathrm{Gr}(m,h)$ tiene diámetro máximo: es decir,

$$\text{máx}\{d(\mathcal{S}, \tilde{\mathcal{S}}) : \ \mathcal{S}, \tilde{\mathcal{S}} \in \mathrm{Dom}_l(A,h)\} = 1\,.$$

Esto último indica que la clase de subespacios dominantes no se concentra en ningún lugar de $\mathrm{Gr}(m,h)$. Estas observaciones comienzan a explicar por qué la expresión en la Ec. (3.1) deja de tener sentido en el caso en que $\sigma_h = \sigma_{h+1}$: por un lado, la función que a la matriz A (y matrices cercanas) le asigna sus subespacios dominantes a izquierda de dimensión h resulta multivaluada, y en consecuencia no puede ser analizada por el cociente planteado en (3.1) (que sólo se aplica a expresiones funcionales univaluadas). Por otro lado, en el caso multivaluado, el conjunto que representa los posibles valores (es decir, los subespacios dominantes) a considerar está disperso (no concentrado) lo que no permite elegir algún elemento representativo de la clase.

Las observaciones anteriores ponen de manifiesto el cambio de naturaleza (para bien!) del problema de aproximación cuando $\sigma_h = \sigma_{h+1}$: hemos mencionado que un subespacio $\mathcal{T} \subset \mathbb{C}^m$ cercano a cualquier subespacio dominante $\mathcal{U}_h$ de A da lugar a una buena aproximación $P_{\mathcal{T}}A$ de A. En el contexto actual, la existencia de infinitos subespacios dominantes a izquierda de A de dimensión h induce el problema de determinar, dado $\mathcal{T} \subset \mathbb{C}^m$ con $\dim \mathcal{T} = h$,

$$d(\mathcal{T}, \mathrm{Dom}_l(A,h)) := \text{mín}\{d(\mathcal{T},\mathcal{U}) : \mathcal{U} \in \mathrm{Dom}_l(A,h)\}\,. \tag{4.3}$$

Dado que el diámetro del conjunto $\mathrm{Dom}_l(A,h)$ es grande, el problema anterior resulta no trivial. Por otro lado, notemos que bastaría con poder calcular una cota superior (informativa) del valor $d(\mathcal{T}, \mathrm{Dom}_l(A,h))$ para determinar la calidad del subespacio $\mathcal{T}$ con respecto a A, y posteriormente, la calidad de la aproximación $P_{\mathcal{T}}A$. El problema planteado en la Ec. (4.3) no parece haber sido considerado previamente en la literatura. De hecho, no parece haber resultados de convergencia correspondientes a un análisis determinístico como el descrito para los subespacios $\mathcal{T}_q$, cuando $\dim \mathcal{X} = h$ y $\sigma_h = \sigma_{h+1}$. Hay una buena razón para eso: como ya hemos mencionado, ejemplos numéricos muestran que los subespacios $\mathcal{T}_q$ no parecen estabilizarse alrededor de ningún subespacio. Sin embargo, en este contexto, resulta más adecuado estudiar el decaimiento de $d(\mathcal{T}_q, \mathrm{Dom}_l(A,h))$ como función de q. Notemos que en principio es posible que $d(\mathcal{T}_q, \mathrm{Dom}_l(A,h)) \to 0$ cuando $q \to \infty$, aún cuando la sucesión de subespacios $(\mathcal{T}_q)_{q\geq 1}$ no se estabilice alrededor de ningún subespacio fijo.

5. Subespacios admisibles

Veamos que podemos flexibilizar algunas de las hipótesis consideradas previamente, para poder extender la propuesta de análisis como en la Ec. (4.3) a casos más generales. Concretamente, vamos a asumir que $A \in \mathbb{C}^{m\times n}$ es una matriz tal que los valores singulares satisfacen

$$\sigma_{j+1} \geq \cdots \sigma_h \geq \cdots \geq \sigma_k \ \text{ con } \ 0 \leq \sigma_{j+1} - \sigma_k \approx 0,$$

con $1 \leq j < h \leq k < \mathrm{rk}(A)$, y que este grupo de valores singulares tiene una dispersión $\delta := \sigma_{j+1} - \sigma_k \geq 0$ muy pequeña. También vamos a asumir también que $\sigma_j > \sigma_{j+1}$ y $\sigma_k > \sigma_{k+1}$ son saltos singulares significativos. Como consecuencia de nuestras hipótesis, el salto singular $\sigma_h - \sigma_{h+1} \leq \delta$ resulta pequeño; este hecho indica que la desigualdad en la Ec. (2.1) no será informativa en este contexto. Este caso puede ser pensado como una perturbación del caso correspondiente a las condiciones en la Ec. (4.1); bajo estas hipótesis, podemos considerar una clase similar a la definida en la Ec. (4.2) para determinar la calidad de subespacios $\mathcal{T} \subset \mathbb{C}^m$ con $\dim \mathcal{T} = h$.

Motivados por las observaciones anteriores introducimos la clase de subespacios admisibles a izquierda de dimensión h, asociados a A y los índices $1 \leq j < h < k$, dada por

$$\mathrm{Adm}_l(A,h) = \{\mathcal{U}_j \oplus \mathcal{D} \ : \ \mathcal{D} \subset \mathcal{U}_k \ominus \mathcal{U}_j\,,\ \dim \mathcal{D} = h - j\} \subset \mathbb{C}^m\,. \tag{5.1}$$

Notemos que, como hemos asumido saltos singulares (significativos) $\sigma_j > \sigma_{j+1}$ y $\sigma_k > \sigma_{k+1}$, los subespacios dominantes $\mathcal{U}_j$ y $\mathcal{U}_k$ están unívocamente determinados. Así, la clase de subespacios admisibles está parametrizada por $\mathrm{Gr}(k-j, h-j)$, es decir, la variedad de todos los subespacios de dimensión $h-j$ dentro de $\mathbb{C}^{k-j}$. Se puede verificar de forma sencilla que cualquier $\mathcal{S} \in \mathrm{Adm}_l(A, h)$ es tal que $P_{\mathcal{S}} A$ es una buena aproximación de rango bajo de A, en el sentido de que los valores singulares $\sigma_i(P_{\mathcal{S}} A)$ son una buena aproximación de $\sigma_i(A)$, para $1 \leq i \leq h$, y el residuo $\|A - P_{\mathcal{S}} A\|$ es casi óptimo, en la medida en que la dispersión $\delta = \sigma_{j+1} - \sigma_k$ del grupo de valores singulares $\sigma_{j+1} \geq ... \geq \sigma_k$ sea pequeña. De forma explícita, se tienen las siguientes acotaciones para los valores singulares de $P_{\mathcal{S}} A$

$$\begin{aligned} \sigma_i(P_{\mathcal{S}} A) = \sigma_i \quad &\text{para} \quad 1 \leq i \leq j \quad \text{y} \\ \sigma_i - \delta \leq \sigma_i(P_{\mathcal{S}} A) \leq \sigma_i \quad &\text{para} \quad j+1 \leq i \leq h\,, \end{aligned} \tag{5.2}$$

y las siguientes acotaciones para las normas de los residuos (o errores de aproximación):

$$\|A - P_{\mathcal{S}}\|_{\mathrm{sp}} \leq \|A - A_h\|_{\mathrm{sp}} + \delta texty \|A - P_{\mathcal{S}}\|_F \leq \|A - A_h\|_F + \sqrt{k-h} \cdot \delta\,. \tag{5.3}$$

El concepto de subespacios admisibles a izquierda de dimensión h intenta formalizar un hecho que informalmente era bien conocido dentro del álgebra lineal numérica: en la medida que $\sigma_h \approx \sigma_{h+1}$, el subespacio dominante $\mathcal{U}_h$ es más difícil de calcular. En este caso, su papel es menos protagónico, en el sentido de que puede ser reemplazado de forma satisfactoria por otros subespacios (los subespacios admisibles) que cumplen un rol similar, en el sentido que valen las Ecs. (5.2) y (5.3), por ejemplo.

A partir de lo anterior, y haciendo uso de las propiedades de continuidad de valores singulares y residuos, concluimos que si $\mathcal{T} \subset \mathbb{C}^m$, $\dim \mathcal{T} = h$, es tal que

$$d(\mathcal{T}, \mathrm{Adm}_l(A, h)) = \mathrm{mín}\{d(\mathcal{T}, \mathcal{S}) \,:\, \mathcal{S} \in \mathrm{Adm}_l(A, h)\} \tag{5.4}$$

es pequeña, entonces $\mathcal{T}$ también dará lugar a una aproximación de rango bajo $P_{\mathcal{T}} A$ con buenas propiedades. Tal como en el caso de la Ec. (4.3), el problema de hallar cotas superiores informativas para $d(\mathcal{T}, \mathrm{Adm}_l(A, h))$ definida como en Ec. (5.4) no es trivial, dado que el conjunto $\mathrm{Adm}_l(A, h)$ tiene un diámetro grande. Por otro lado, el hecho de que $\mathrm{Adm}_l(A, h)$ tenga un diámetro grande puede ser considerado como una ventaja al momento de probar que las distancias $d(\mathcal{T}_q, \mathrm{Adm}_l(A, h))$ decrecen cuando el parámetro q crece, para la sucesión de subespacios $\mathcal{T}_q = (AA^*)^q A(\mathcal{X})$, generada a partir de un subespacio inicial $\mathcal{X} \subset \mathbb{C}^n$, $\dim \mathcal{X} = h$. Incluso más: en este contexto puede suceder que las distancias $d(\mathcal{T}_i, \mathrm{Adm}_l(A, h))$ decrezcan, aún cuando las variaciones $d(\mathcal{T}_i, \mathcal{T}_{i+1})$, $1 \leq i \leq q-1$, asuman valores grandes (es decir, aún cuando la sucesión finita $\{\mathcal{T}_i\}_{i=1}^q$ no se estabilice alrededor de algún subespacio (límite) $\mathcal{T}$). Estas observaciones indican que el análisis correspondiente al estudio de $d(\mathcal{T}_q, \mathrm{Adm}_l(A, h))$ definida como en la Ec. (5.4) corresponde a un análisis de cercanía a esta clase, y no a un análisis de convergencia a un subespacio particular.

6. Propiedades de los subespacios admisibles

En esta última sección consideramos algunas propiedades básicas de los subespacios admisibles, que permiten responder algunas de las cuestiones que hemos discutido previamente. Los resultados contenidos en esta sección han sido desarrollados en conjunto con Francisco Arrieta Zuccalli (ver [3]).

Vamos a recordar aquí las hipótesis generales que hemos considerado hasta ahora: por un lado, $A \in \mathbb{C}^{m \times n}$ es una matriz rectangular y $\sigma_1 \geq \cdots \geq \sigma_p \geq 0$ denotan sus valores singulares (aquí $p = \mathrm{mín}\{m, n\}$). Hemos fijado una dimensión objetivo $1 < h < \mathrm{rk}(A)$ (como restricción), hemos supuesto la existencia de índices $1 \leq j < h < k < \mathrm{rk}(A)$ tales que

$\sigma_j > \sigma_{j+1}$ y $\sigma_k > \sigma_{k+1}$ son saltos singulares significativos, y $0 \leq \delta := \sigma_{j+1} - \sigma_k \approx 0$; es decir, σ_h está dentro de grupo de valores singulares, muy parecidos entre sí (posiblemente el mismo, repetido $k-j$ veces). Notemos que entonces están unívocamente definidos los subespacios dominantes a izquierda y derecha $\mathcal{U}_j,\, \mathcal{U}_k \subset \mathbb{C}^m$ y $\mathcal{V}_j,\, \mathcal{V}_k \subset \mathbb{C}^n$.

Sobre la cercanía a la clase de subespacios admisibles: La distancia o proximidad entre subespacios que hemos considerado previamente satisface una importante propiedad de monotonía : si $\mathcal{S}' \subset \mathcal{S}$ y $\mathcal{T} \subset \tilde{\mathcal{T}}$ de forma que $\dim \mathcal{S} \leq \dim \mathcal{T}$ entonces:

$$d(\mathcal{S}', \mathcal{T}) \leq d(\mathcal{S}, \mathcal{T}) \ \text{ y } \ d(\mathcal{S}, \mathcal{T}) \leq d(\mathcal{S}, \tilde{\mathcal{T}})\,.$$

Recordemos que si $\mathcal{S} \in \mathrm{Adm}_l(A, h)$ entonces, en particular, $\mathcal{U}_j \subset \mathcal{S}$ y $\mathcal{S} \subset \mathcal{U}_k$. Así, si $\mathcal{T} \subset \mathbb{C}^m$ con $\dim \mathcal{T} = h$ entonces, por la propiedad de monotonía, $d(\mathcal{T}, \mathcal{U}_j) \leq d(\mathcal{T}, \mathcal{S})$ y $d(\mathcal{T}, \mathcal{U}_k) \leq d(\mathcal{T}, \mathcal{S})$. Así, si $\mathcal{T}$ es cercano a algún subespacio admisible $\mathcal{S} \in \mathrm{Adm}_l(A, h)$, entonces $\mathcal{U}_j$ es cercano a $\mathcal{T}$ y $\mathcal{T}$ es cercano a $\mathcal{U}_k$. Sucede que vale también una suerte de recíproca: si $h \leq \dim \mathcal{T} \leq k$ entonces (ver [3]):

$$d(\mathcal{T}, \mathrm{Adm}_l(A, h)) \leq d(\mathcal{T}, \mathcal{U}_j) + d(\mathcal{T}, \mathcal{U}_k)\,. \tag{6.1}$$

La desigualdad anterior es el punto de partida de una serie de consecuencias importantes para los subespacios admisibles. La prueba de esta desigualdad se obtiene eligiendo explícitamente un subespacio $\mathcal{S} \in \mathrm{Adm}_l(A, h)$ tal que $d(\mathcal{T}, \mathcal{S}) \leq d(\mathcal{T}, \mathcal{U}_j) + d(\mathcal{T}, \mathcal{U}_k)$.

Sobre el análisis de los algoritmos iterativos. Una de las principales motivaciones para considerar la clase de subespacios admisibles (bajo las hipótesis que estamos asumiendo) era la de proveer una clase bien comportada de subespacios, que sirviera para medir la calidad de un subespacio $\mathcal{T} \subset \mathbb{C}^m$ como base para la construcción de una aproximación con rango bajo $P_{\mathcal{T}} A$. De hecho, planteamos considerar $d(\mathcal{T}, \mathrm{Adm}_l(A, h))$ como medida de la calidad del subespacio $\mathcal{T}$. En este sentido, podemos indicar el siguiente comportamiento: dado un subespacio $\mathcal{X} \subset \mathbb{C}^n$ con $h \leq r := \dim \mathcal{X} \leq k$, en posición genérica, si consideramos la sucesión finita de subespacios computables, dada por $\mathcal{T}_q = (AA^*)^q A(\mathcal{X})$ para $q \geq 1$, entonces

$$d(\mathcal{T}_q, \mathrm{Adm}_l(A, h)) \leq \left(\frac{\sigma_{j+1}}{\sigma_j}\right)^{2q+1} \tan(\theta_{\text{máx}}(\mathcal{X},\, \mathcal{V}_j)) + \left(\frac{\sigma_{k+1}}{\sigma_k}\right)^{2q+1} \tan(\theta_{\text{máx}}(\mathcal{X},\, \mathcal{V}_k))\,. \tag{6.2}$$

Notemos que bajo las hipótesis que hemos asumido, los saltos singulares en los índices j y k son significativos (mientras que el salto singular en el índice h no lo es). En este caso, la estimación anterior resulta informativa y muestra que la sucesión de subespacios computables $\{\mathcal{T}_q\}_q$ se aproxima arbitrariamente a la clase de subespacios admisibles (aún cuando la sucesión no se estabilice alrededor de ningún subespacio). Este último hecho valida teóricamente que las matrices de la forma $P_{\mathcal{T}_q} A$ sean buenas aproximaciones con restricciones de rango de la matriz A, lo que era bien conocido en el plano de los experimentos numéricos con el método del subespacio iterativo.

Sobre la estabilidad de la clase de subespacios admisibles: recordemos que el número de condición, considerado previamente para subespacios dominantes en la Ec. (3.1), indica una medida de la estabilidad del cálculo de $\mathcal{U}_h$. En el presente contexto, veremos que es posible definir una cantidad similar para el cálculo de la clase de subespacios admisibles, y daremos una estimación para este nuevo número de condición. En efecto, bajo las hipótesis de esta sección, es posible elegir $\varepsilon_0(A) = \varepsilon_0 > 0$ suficientemente chico, tal que si $\|A - B\|_{\text{sp}} \leq \varepsilon_0$, entonces $\sigma_j(B) > \sigma_{j+1}(B)$ y $\sigma_k(B) > \sigma_{k+1}(B)$ son saltos singulares significativos y $0 \leq \delta(B) := \sigma_{j+1}(B) - \sigma_k(B) \approx 0$. Así, si $\|A - B\|_{\mathrm{sp}} \leq \varepsilon_0$, tiene sentido considerar la clase de subespacios admisibles a izquierda de dimensión h de B:

$$\mathrm{Adm}_l(B,h) = \{\mathcal{U}_j(B) \oplus \mathcal{D} \ : \ \mathcal{D} \subset \mathcal{U}_k(B) \ominus \mathcal{U}_j(B)\,,\ \dim \mathcal{D} = h - j\}\,.$$

Para considerar una expresión formalmente similar a la dada en la Ec. (3.1), introducimos una función distancia d_H en $\mathcal{P}(\mathrm{Gr}(m,h))$, donde $\mathcal{P}(\mathrm{Gr}(m,h))$ denota el conjunto de partes de $\mathrm{Gr}(m,h)$. En efecto, consideramos la métrica de Hausdorff inducida por la función distancia en $\mathrm{Gr}(m,h)$: en este caso, dados $\mathcal{M},\ \mathcal{N} \in \mathcal{P}(\mathrm{Gr}(m,h))$ entonces

$$d_H(\mathcal{M},\mathcal{N}) = \max\{\sup_{M\in\mathcal{M}} d(M,\mathcal{N})\,,\ \sup_{N\in\mathcal{N}} d(N,\mathcal{M})\}\,,$$

donde, como antes, $d(N,\mathcal{M}) = \inf\{d(N,M) \ : \ M \in \mathcal{M}\}$ y similar para $d(M,\mathcal{N})$. Con el lenguaje anterior, consideramos la siguiente noción de número de condición para los subespacios admisibles en A:

$$\kappa(\mathrm{Adm}_l(A,h)) = \lim_{\varepsilon\to 0^+} \sup_{0<\|A-B\|_{\mathrm{sp}}<\varepsilon} \frac{d_H(\mathrm{Adm}_l(B,h),\mathrm{Adm}_l(A,h))}{\|A-B\|_{\mathrm{sp}}}\,. \tag{6.3}$$

Como consecuencia de la desigualdad en la Ec. (6.1), obtenemos la siguiente acotación de la variación de las clases de subespacios admisibles en A (ver [3]):

$$\kappa(\mathrm{Adm}_l(A,h)) \leq \alpha\left(\frac{1}{\sigma_j - \sigma_{j+1}} + \frac{1}{\sigma_k - \sigma_{k+1}}\right)\,, \tag{6.4}$$

para una constante $1 \leq \alpha \leq 2$. La estimación anterior indica el rol que cumplen los saltos singulares antes y después del grupo de valores singulares, en la estabilidad de la clase de subespacios admisibles a izquierda y de dimensión h, para matrices cercanas a A.

Si consideramos el caso particular en que $\sigma_{j+1} = \cdots = \sigma_h = \cdots = \sigma_k$ (de forma que $\sigma_h = \sigma_{h+1}$ y la expresión en la Ec. (3.1) no está definida), la Ec. (6.4) indica una cota superior para la variación de las clases de subespacios admisibles (con respecto a la métrica de Hausdorff) que está bien definida. Esto último muestra que los subespacios admisibles permiten tratar el caso multivaluado que corresponde a los subespacios dominantes, cuando hay repetición del valor singular σ_h.

7. Comentarios finales

Hemos considerado la estrategia clásica de evaluar la calidad de las aproximaciones de rango bajo de la forma $P_{\mathcal{T}}A$ de una matriz A, a través de la estimación de la distancia al subespacio dominante a izquierda $d(\mathcal{U}_h,\mathcal{T})$, para subespacios $\mathcal{T} \subset \mathbb{C}^m$, $\dim\mathcal{T} = h$. Hemos observado que la estabilidad del cálculo de aproximaciones de los subespacios $\mathcal{U}_h$ no es buena en el caso en que $\sigma_h \approx \sigma_{h+1}$. En el caso en que σ_h forma parte de un grupo de valores singulares cercanos (es decir, en donde $\sigma_h \approx \sigma_{h+1}$), hemos introducido la clase de subespacios admisibles y hemos mostrado que: por un lado, los subespacios admisibles $\mathcal{S} \subset \mathbb{C}^m$ dan lugar a buenas aproximaciones de la forma $P_{\mathcal{S}}\,A$. Por otro lado, los subespacios admisibles forman una clase abundante, en el sentido que $d(\mathcal{T}_q,\mathrm{Adm}_l(A,h))$ se hace arbitrariamente pequeña cuando q crece, para sucesiones $\{\mathcal{T}_q\}_q$ de subespacios construidos con métodos computables bien conocidos (ver Eq. (6.2)). Esto último permite explicar por qué las sucesiones $\{\mathcal{T}_q\}_q$ dan lugar a buenas aproximaciones de la forma $P_{\mathcal{T}_q}A$, aún cuando $d(\mathcal{T}_q,\mathcal{U}_h)$ sea grande.

La clase de subespacios admisibles tiene un rol de interés en otras áreas del álgebra lineal numérica. Por ejemplo, parece una clase natural para el estudio de los llamados subespacios activos (ver [4]). Por otro lado, en conjunto con Francisco Arrieta Zuccalli, estamos desarrollando aplicaciones en el contexto de las llamadas aproximaciones CUR de matrices (ver [5]).

Finalmente, cabe remarcar que el enfoque de los subespacios admisibles tiene sus limitaciones: de hecho, el éxito de este enfoque radica en disponer de matrices cuyos valores singulares estén agrupados en clases tales que cada clase tenga una pequeña dispersión, y que clases distintas se mantengan a una distancia significativa entre sí. Por ejemplo, en el caso de una matriz cuyos valores singulares tienen un decaimiento lineal de la forma $\sigma_i = b - a \cdot i$, $1 \leq i \leq p = \text{mín}\{m, n\}$, para $a,\ b > 0$ y tales que $a \approx 0$, entonces estos valores no pueden agruparse en el sentido anterior y la técnica de subespacios admisibles no tiene una aplicación evidente.

Agradecimientos. Quiero agradecer a Francisco Arrieta Zuccalli por permitirme hacer uso en este nota del material que hemos desarrollado en conjunto. Quiero agradecer también a los editores del “Noticiero de la UMA”, y en particular a Iván Angiono, por invitarme a participar de este espacio sobre Miradas Matemáticas.

Referencias

[1] Antezana, J., Stojanoff, D.: Análisis Matricial, Cursos y Seminarios de matemática Serie B (2008).

[2] Arrieta Zuccalli, F., Massey, P., Stojanoff, D.: Block subspace expansions for eigenvalue and eigenvector approximation (enviado). Accesible en: https://arxiv.org/pdf/2411.14578

[3] Arrieta Zuccalli, F., Massey, P.:Λ-admissible subspaces of self-adjoint matrices (enviado). Accesible en: https://arxiv.org/pdf/2602.11976

[4] Constantine, P. G.: Active subspaces. Emerging ideas for dimension reduction in parameter studies. SIAM Spotlights, 2. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2015.

[5] Drineas, P., Mahoney, M., Muthukrishnan, S.:Relative-error *CUR* matrix decompositions. SIAM J. Matrix Anal. Appl. 30 (2008), no. 2, 844-881

[6] Drineas, P., Ipsen, I.: Low-rank matrix approximations do not need a singular value gap. SIAM J. Matrix Anal. Appl. 40 (2019), no. 1, 299-319.

[7] Massey, P.: Admissible subspaces and the subspace iteration method. BIT 64 (2024), no. 1, Paper No. 12, 36 pp.

[8] Vannieuwenhoven, N.: The condition number of singular subspaces, revisited. Linear Algebra Appl. 687 (2024), 157–180

[9] Saibaba, A.: Randomized subspace iteration: analysis of canonical angles and unitarily invariant norms. SIAM J. Matrix Anal. Appl. 40 (2019), no. 1, 23-48.

Ir al índice general