

Un estimado simple para los autovalores más pequeños de matrices simétricas

Augusto González

Instituto de Cibernética, Matemática y Física, Calle E 309, Vedado, Ciudad de La Habana; agonzale@icmf.inf.cu

Recibido el 2/05/10. Aprobado en versión final el 09/09/10

Sumario. Combinando la Teoría de Perturbaciones con el método de Kirzhnits y el denominado Principio de Mínima Sensitividad, se muestra que es posible, a través de un procedimiento muy simple, estimar los autovalores mas pequeños (o mas grandes) de matrices simétricas de elevada dimensión. Este resultado podría ser de gran interés para regularizar series perturbativas que aparecen en la Física como, por ejemplo, la serie de Moller-Plesset para calcular las energías de correlación en átomos o moléculas.

Abstract. A simple method for the estimation of the lowest (highest) eigenvalues of very large symmetric matrices is shown, which is based on the combination of Perturbation Theory, the Kirzhnits method, and the so called Principle of Minimal Sensitivity. The procedure could be employed in order to regularize perturbation series arising in Physics such as, for example, the Moller-Plesset series for the correlation energy of atoms or molecules.

Palabras clave. Teoría de matrices, 02.10.Yn, Mecánica Cuántica, 03.65.-w

1 Introducción

El estimado de algunos autovalores de matrices de dimensión relativamente grande es un problema que aparece con mucha frecuencia en diversos contextos. En la Mecánica Cuántica, por ejemplo, el operador de energía (el Hamiltoniano) se representa como una matriz en el espacio de Hilbert donde está definido. Aún cuando el espacio sea truncado y la matriz Hamiltoniana tenga dimensión finita, la dimensión toma usualmente valores extra grandes. En un problema de N partículas que pueden ocupar M estados, por ejemplo, la dimensión es del orden de $M!/N!(M-N)!$. En un átomo pequeño como el Oxígeno de $N=8$ electrones, con sólo $M=25$ estados la dimensión es ya 10^6 . Sin embargo, a veces no es necesario conocer todos los autovalores, pues los mas bajos determinan muchas de las propiedades electrónicas, ópticas o de otra índole.

Para calcular algunos de los autovalores de una matriz extra grande simétrica se han desarrollado varias estrate-

gias, la mas conocida de las cuales es el algoritmo de Lanczos ¹, que requiere de una función que multiplique la matriz dada por vectores arbitrarios. En el presente trabajo mostramos un método que representa una economía notable con respecto al de Lanczos. A grosso modo podemos decir que basta la diagonal de la matriz y algunas filas para estimar los autovalores mas bajos. El procedimiento que se presenta se basa en tres pilares: la conocida Teoría de Perturbaciones ², la menos conocida formulación de Kirzhnits del problema de autovalores como ecuaciones de evolución ³ y el denominado Principio de Mínima Sensitividad ⁴, cuyo contenido es intuitivamente simple. En el trabajo ilustramos, con matrices de dimensión 1000, que la metodología propuesta arroja estimados razonables y simples.

2 Enunciado del problema

Sea $H=T+V$, donde T es una matriz diagonal cuyos elementos son todos distintos o, al menos, los mas peque-

ños son distintos. Sin pérdida de generalidad podemos asumir que están ordenados. La matriz V es simétrica. Asumiremos también que H no tiene una estructura de bloques, pues en ese caso podríamos trabajar con los bloques en vez de con toda la matriz.

Nos proponemos hallar estimados sencillos para los autovalores mas bajos de H , que denominaremos E_n , con $n=1,2,\dots$

3 La formulación de Kirzhnits

Escribamos $H(g)=T+g V$, donde g es real. Queremos estimar los autovalores de $H=H(1)$. Notar que $H(0)=T$. En un contexto de Mecánica Cuántica, Kirzhnits mostró³ que el problema de autovalores para $H(g)$:

$$H(g)u(g) = E_n(g)u(g), \quad (1)$$

donde hemos escrito explícitamente la dependencia con g de los autovalores y autovectores, es completamente equivalente al siguiente sistema de ecuaciones no lineales de evolución respecto al parámetro g :

$$\frac{dE_n(g)}{dg} = V_{nn}(g), \quad (2)$$

$$\begin{aligned} \frac{dV_{nm}(g)}{dg} = & \sum_{k \neq n} \frac{V_{nk}(g)V_{km}(g)}{E_n(g) - E_k(g)} \\ & + \sum_{k \neq m} \frac{V_{nk}(g)V_{km}(g)}{E_m(g) - E_k(g)} \end{aligned} \quad (3)$$

siendo $V_{nm}(g) = u_n^t(g) V u_m(g)$. La demostración es muy simple. Tenemos la Ec. (1) y las relaciones de ortogonalidad, $u_n^t(g) \cdot u_m(g) = \delta_{nm}$, las cuales podemos derivar respecto del parámetro g , así como derivar $V_{nm}(g)$, para obtener (2) y (3).

Las ecuaciones de Kirzhnits permiten enunciar dos Lemas que nos serán útiles mas abajo:

LEMA 1 (no cruzamiento): Los autovalores se conectan continuamente al evolucionar g .

Es decir, el primer elemento de T evoluciona continuamente hasta el autovalor mas bajo de H , etc. Rigurosamente, este Lema es consecuencia de que H no tiene estructura de bloques. En el lenguaje de la Mecánica Cuántica diríamos que trabajamos con una representación irreducible. En términos de las ecuaciones de Kirzhnits, los denominadores, $E_n(g)-E_k(g)$, hacen que cuando dos autovalores se acercan, los correspondientes V_{nm} y V_{kk} crecen con signos opuestos y dan lugar a la “repulsión” entre autovalores (Fig. 1).

La consecuencia de este Lema es que si nos interesan los autovalores mas bajos de $H(1)$ debemos seguir la evolución de los primeros elementos de T .

Por otro lado, la matriz $H(g)$ en la base formada por los autovectores de $H(g_0)$ toma la forma $H_{g_0}(g)=T(g_0)+(g-g_0)V(g_0)$, donde $T(g_0)$ es una matriz diagonal cuyos elementos son los autovalores $E_n(g_0)$ y la matriz $V(g_0)$ tiene elementos $u_n(g_0)^t V u_m(g_0)$. Entonces:

LEMA 2: La familia de matrices $\{H_{g_0}(g)\}$ es isoespectral con $H(g)$ para todo g .

En nuestro estimado, partiremos el intervalo (0,1) en dos: (0, g_0) y (g_0 ,1). $T(g_0)$ y $V(g_0)$ los hallaremos por medio de la Teoría de Perturbaciones, mientras que a partir de g_0 seguiremos con $H_{g_0}(g)$.

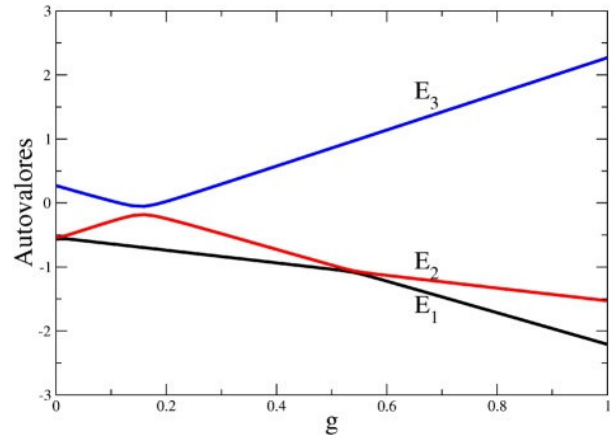


Figura 1. Ejemplo de repulsión entre los autovalores de una matriz de dimensión 3.

4 La teoría de perturbaciones

La teoría de perturbaciones busca la solución del problema de autovalores para $H(g)$ como una serie alrededor de $g=0$. Se conoce que estas son series asintóticas⁵, es decir, su radio de convergencia puede ser incluso cero, pero los primeros términos dan una buena aproximación en cierto intervalo de g . Escribiendo en la Ec. (1):

$$E_n(g) = E_n(0) + E_n^{(1)}g + E_n^{(2)}g^2 + \dots \quad (4)$$

$$u_n(g) = u_n(0) + u_n^{(1)}g + u_n^{(2)}g^2 + \dots, \quad (5)$$

e igualando potencias de g , resulta, hasta primer orden (TP1):

$$E_n^{(TP1)}(g) = E_n(0) + V_{nn}(0)g, \quad (6)$$

$$\begin{aligned} V_{nm}^{(TP1)}(g) = & V_{nm}(0) + g \sum_{k \neq n} \frac{V_{nk}(0)V_{km}(0)}{E_n(0) - E_k(0)} \\ & + g \sum_{k \neq m} \frac{V_{nk}(0)V_{km}(0)}{E_m(0) - E_k(0)}. \end{aligned} \quad (7)$$

Mientras que, a segundo orden (TP2):

$$E_n^{(TP2)}(g) = E_n^{(TP1)}(g) + g^2 \sum_{k \neq n} \frac{V_{nk}(0)V_{kn}(0)}{E_n(0) - E_k(0)} \quad (8)$$

$$\begin{aligned} V_{nm}^{(TP2)}(g) = & V_{nm}^{(TP1)}(g) \\ & + \frac{g^2}{2} \sum_{k \neq n} \frac{V_{nk}(0)V_{km}(0)(V_{nn}(0) - V_{kk}(0))}{(E_n(0) - E_k(0))^2} \\ & + \frac{g^2}{2} \frac{V_{nk}(0)a_{km} + a_{nk}V_{km}(0)}{E_n(0) - E_k(0)} \end{aligned}$$

$$\begin{aligned}
& + \frac{g^2}{2} \sum_{k \neq m} \frac{V_{nk}(0)V_{km}(0)(V_{mm}(0) - V_{kk}(0))}{(E_m(0) - E_k(0))^2} \\
& + \frac{g^2}{2} \frac{V_{nk}(0)a_{km} + a_{nk}V_{km}(0)}{E_m(0) - E_k(0)}, \quad (9)
\end{aligned}$$

donde a_{nm} es, por definición, la expresión que multiplica a g en la Ec. (7). En la Fig. 2 se muestra el comportamiento típico de TP1 y TP2 para el autovalor mas bajo de una matriz de dimensión 1000.

5 La metodología propuesta

Consideremos el menor de $H(g)$ que contiene los α primeros elementos de T . Si diagonalizamos este menor y tomamos su autovalor mas bajo obtendremos una cota superior para $E_l(g)$. Este es un teorema del Algebra de matrices, que en la Mecánica Cuántica se conoce como Principio Variacional. Este estimado puede, sin embargo, ser muy malo, ya que sólo depende de un menor de dimensión α y no de toda la matriz.

Por otro lado, supóngase que evolucionamos inicialmente hasta g_0 siguiendo la Teoría de Perturbaciones. Y a partir de ahí tomamos el menor de $H_{g_0}(g)$, al que denominaremos $h_{g_0}(g)$, que contiene los α primeros elementos de $T(g_0)$. El caracter variacional se pierde ya que la evolución inicial es aproximada, pero $h_{g_0}(g)$ contiene mas información de la matriz inicial. En TP1, por ejemplo, $h_{g_0}(g)$ depende de T y de las primeras α filas de V . En TP2 muchos mas términos de V entran en el cálculo, pero el costo computacional es muy inferior al del método de Lanczos ya que el menor $h_{g_0}(g)$ tiene dimensión reducida y se calcula una sola vez.

En resumen, lo que se propone es lo siguiente: Calculamos por TP el menor, $h_{g_0}(g)$, de la matriz $H_{g_0}(g)$ que contiene los α primeros elementos de $T(g_0)$. El estimado de los primeros autovalores, $E_n(g)$, resulta de diagonalizar $h_{g_0}(g)$. Los estimados para los autovalores, ε_n , quedan dependientes de dos parámetros: g_0 y α . g_0 se escoge suficientemente pequeño para que TP1 o TP2 sean válidas. Con respecto a α utilizaremos el

Principio de Mínima Sensitividad (PMS): α es un parámetro del cual los autovalores exactos no deben depender. El mejor estimado es el que evidencie poca sensibilidad a este parámetro, es decir: $d\varepsilon_n/d\alpha=0$.

El PMS ha sido utilizado otras veces para controlar series perturbativas ^{4,6}.

6 Ejemplos numéricos

Para probar la generalidad del método se diagonalizaron varias matrices de dimensión 1000. Los elementos fueron escogidos aleatoriamente, de forma que las matrices son completamente independientes unas de otras. T tiene elementos aleatorios con distribución uniforme entre 1 y 30. La matriz V se construyó simétrica, con elementos

aleatorios distribuidos uniformemente entre -4 y 4.

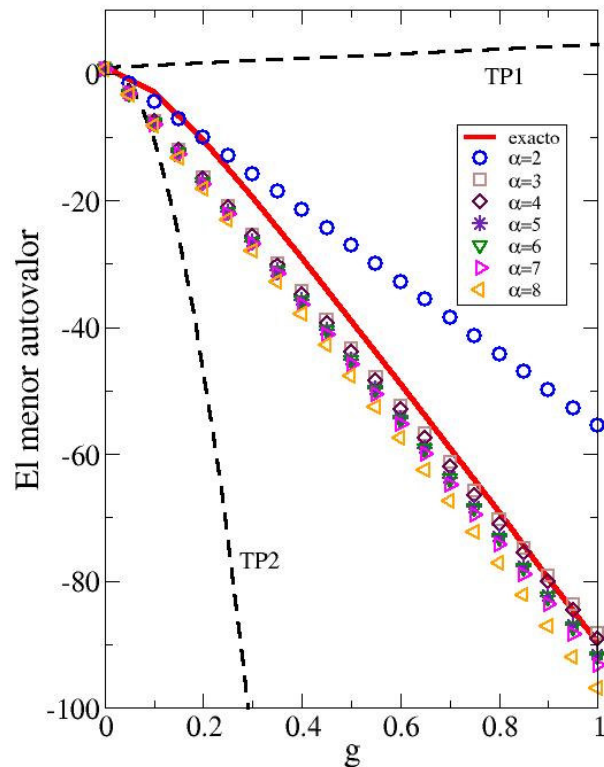


Figura 2. El menor autovalor de una matriz simétrica de dimensión 1000. En línea roja continua se grafica el autovalor exacto, los valores arrojados por TP1 y TP2 se muestran en línea negra discontinua, mientras que TP2 corregida, con $1 < \alpha < 9$, se representa por símbolos. Símbolos superpuestos corresponden a buenos estimados, de acuerdo con el PMS.

La Fig. 2 muestra los resultados para una de estas matrices. Tomamos $g_0=0.005$ y utilizamos las expresiones de TP2, Ecs. (8) y (9), para construir el menor $h_{g_0}(g)$ de dimensión α que debemos diagonalizar. Con $\alpha=2$ (círculos azules) el estimado es malo y exhibe una alta sensibilidad, como se puede ver si comparamos con $\alpha=3$ (cuadrados). Cuando $\alpha=3$ y 4, los estimados son casi indistinguibles en $g=1$, como lo son las curvas con $\alpha=5$ y 6. Los estimados que surgen son -89.07 y -91.63, que se comparan muy favorablemente con el resultado exacto -89.65.

Aquí debemos hacer un comentario sobre el parámetro g_0 . Si extendemos TP hasta un g_0 demasiado grande obtenemos resultados sin sentido, es decir, por debajo de TP2, aun con $\alpha=1$ o 2. Redujimos g_0 hasta que los estimados con $1 < \alpha < 10$ tuvieran sentido y ahí aplicamos el PMS. Una reducción posterior de g_0 hace necesario incrementar α y por tanto encarece computacionalmente el estimado.

7 Conclusiones

Podemos resumir el método propuesto en el presente

trabajo de la siguiente manera. Utilizando un menor de H podemos obtener una cota superior para el autovalor mas pequeño de una matriz simétrica. En particular, TP1 es una cota que surge de un menor de dimensión uno. Podemos mejorar notablemente este estimado si inicialmente evolucionamos por medio de TP1 o TP2 hasta g_0 y utilizamos un menor de H_{g_0} . El carácter variacional se pierde pues la evolución inicial es aproximada, pero el menor construido tiene mas información de la matriz inicial. La dimensión del menor, α , aparece como un parámetro adicional. El PMS nos dice que el mejor estimado para el autovalor es aquel que muestre poca sensibilidad a α . Demostramos, en ejemplos de juguete, es cierto, que basta $\alpha < 10$ para estimar con errores del orden del 2 % el autovalor mas bajo. Razonamientos análogos pueden hacerse para el mayor autovalor o para un conjunto de autovalores.

El método pudiera resultar muy interesante en problemas físicos, donde el operador Hamiltoniano en un subespacio finito del espacio de Hilbert tiene forma de matriz, cuya dimensión se hace, con mucha frecuencia, excesivamente grande. Un problema al que lo aplicaremos es a corregir o controlar la serie de Moller-Plesset⁷ para calcular la energía de correlación en átomos o moléculas. Igualmente interesante podría ser su uso en matrices extra grandes que aparecen en el tratamiento masivo de datos y para las cuales se están desarrollando métodos basados en muestreo aleatorio⁸.

Los resultados del presente trabajo pueden generalizarse a matrices hermíticas y también al caso en que T tiene espectro degenerado. En este último caso hay que resolver primero la degeneración utilizando teoría de perturbaciones.

Agradecimientos

El autor agradece a los participantes del Seminario de Física Teórica del ICIMAF por los comentarios críticos y sugerencias. El trabajo recibió apoyo del Programa Nacional de Ciencias Básicas (Cuba) y de la Red Caribeña de Mecánica Cuántica, Partículas y Campos (ICTP).

Referencias

1. J.K. Cullum and R.A. Willoughby, "Lanczos Algorithms for Large Symmetric Eigenvalue Computations" (Birkhauser, Stuttgart, 1985).
2. L.D. Landau y E.M. Lifshits, "Mecánica Cuántica, Teoría no relativista" (Mir, Moscu, 1967).
3. D.A. Kirzhnits and N.Zh. Takibaev, JETP Lett. 23, 323 (1976); Sov. J. Nucl. Phys. 25, 370 (1977); D.A. Kirzhnits, G.Yu. Kryuchkov and N.Zh. Takibaev, Fiz. Elem. Chast. Atom. Yadra 10, 741 (1979).
4. P.M. Stevenson, "Optimized perturbation theory", Phys. Rev. D 23, 2916-2944 (1981).
5. M.V. Fedoriuk, "Asintotica de integrales y series" (Nauka, Novosibirsk, 1987).
6. A. Gonzalez, J.D. Serna, R. Capote, G. Avendaño, "Higher Landau levels contribution to the energy of interacting electrons in a quantum dot" Physica E 30, 134-137 (2005).
7. M.L. Leininger, W.D. Allen, H.F. Schaefer, C.D. Sherrill, "Is Møller-Plesset Perturbation Theory a Convergent *Ab Initio* Method?", J. Chem. Phys. 112, 9213-9222 (2000).
8. N. Halko, P.G. Martinsson, and J.A. Tropp, "Finding structure with randomness: stochastic algorithms for constructing approximate matrix decompositions", arXiv.org : 0909.4061, ACM Report 2009-5, Caltech, Sep. (2009).