A METHOD TO COMPUTE THE INVERSE OF A COMPLEX N-BLOCK TRIDIAGONAL QUASI-HERMITIAN MATRIX

Elena M. Godfrin
Departamento Fuentes Renovables y Uso Racional de la Energía
Comisión Nacional de Energía Atómica
1429 Buenos Aires - Argentina

Abstract

This paper presents a method to compute the inverse of a complex n-block tridiagonal quasi-hermitian matrix using adequate partitions of the complete matrix. This type of matrix is very usual in quantum mechanics and, more specifically, in solid state physics (e.g., interfaces and super-lattices), when the tight-binding approximation is used. The efficiency of the method is analysed comparing the required CPU time and work-area for different usual techniques.

1 Introduction

The linear combination of atomic orbitals method [1,2], specially within the tight-binding approximation, is one of the most used technics in quantum mechanics for the analysis of properties related with elementary excitations (electrons, phonons, etc.). Its application to the study of molecules and solids gives Hamiltonian operators which, written in an appropriate base, usually have many null elements. In periodical systems such as crystal solids and super-lattices the Hamiltonian may be represented as a quasi-hermitian[1]. block-tridiagonal matrix, in many cases of high dimension (e.g. [3]). The density of states as a function of the energy is obtained from the trace of the Green function matrix G [1,2] defined as:

$$G = (EI - H)^{-1} ,$$

[1] $M_{ij} \in C^{p \times q}$, $M_{ij} = M_{ji}^{\dagger}$ if $i \neq j$ and $det\ M_{ii} \neq 0$, where M^{+} indicates the transposed and conjugate matrix of M

where E is a scalar and I, the identity matrix. Usual numerical techniques are not appropriate to solve this equation since they do not take into account the particular structure of H (essentially, the large ammount of null elements), being inefficient both from CPU time and memory requirements.

This paper presents a method to compute the inverse of a complex n-block tridiagonal quasi-hermitian matrix which considers the structure of H using adequate partitions of the complete matrix. Its efficiency is analysed comparing the required CPU time for different usual techniques. It should be mentioned that a first version of this method (which only compute the diagonal blocks of the inverse matrix) has been used in several cases to study electronic properties of semiconductor interfaces (e.g., see Refs. [4,5]).

2 Mathematical problem

Let M be the n-block tridiagonal quasi-hermitian matrix,

$$\begin{pmatrix} M_{11} & M_{21}^{+} & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ M_{21} & M_{22} & M_{32}^{+} & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & M_{32} & M_{33} & M_{43}^{+} & \cdots & \cdots & \cdots & \\ \vdots & & & \ddots & & & \vdots & \\ 0 & \cdots & \cdots & \cdots & 0 & M_{n-1,n-2} & M_{n-1,n-1} & M_{n,n-1}^{+} \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & M_{n,n-1} & M_{n,n} \end{pmatrix},$$

the problem consist of finding the matrix G such that $M \times G = I$, where I is the identity matrix.

Partitioning the matrix G in the same way as M,

$$G = \begin{pmatrix} G_{11} & G_{12} & \cdots & G_{1n} \\ G_{21} & G_{22} & \cdots & G_{2n} \\ \vdots & \vdots & & \vdots \\ G_{n1} & G_{n2} & \cdots & G_{nn} \end{pmatrix};$$

the system $M \times G = I$ can be written as:

$$\begin{cases} M_{11}G_{11} + M_{21}^{+}G_{21} & & & & & = I \\ M_{21}G_{11} + M_{22}G_{21} + M_{32}^{+}G_{31} & & & & & = 0 \\ M_{32}G_{21} + M_{33}G_{31} + M_{43}^{+}G_{41} & & & & & = 0 \\ \vdots & & & & & \\ M_{n,n-1}G_{n-1,1} + M_{nn}G_{n1} & & & & & = 0 \end{cases}$$

$$\vdots$$

$$\begin{cases} M_{11}G_{1n} & +M_{21}^{+}G_{2n} & & & & = 0 \\ M_{21}G_{1n} & +M_{22}G_{2n} & +M_{32}^{+}G_{3n} & & & = 0 \\ & M_{32}G_{2n} & +M_{33}G_{3n} & +M_{43}^{+}G_{4n} & & = 0 \\ & & & \vdots & & \\ & & & & M_{n,n-1}G_{n-1,n} \; +M_{nn}G_{nn} & = I \end{cases},$$

where I is the identity matrix with adecuate dimension.

The matrices $G_{i,j}$ are obtained solving the matricial equation system and thus, the inverse of M is computed.

3 Problem resolution

The general resolution for a general M n-block matrix is very tedious and does not contribute to qualify, thus only the 4-block problem will be solved.

Let then M and G be such that

$$\begin{pmatrix} M_{11} & M_{21}^{+} & 0 & 0 \\ M_{21} & M_{22} & M_{32}^{+} & 0 \\ 0 & M_{32} & M_{33} & M_{43}^{+} \\ 0 & 0 & M_{43} & M_{44}^{+} \end{pmatrix} \begin{pmatrix} G_{11} & G_{12} & G_{13} & G_{14} \\ G_{21} & G_{22} & G_{23} & G_{24} \\ G_{31} & G_{32} & G_{33} & G_{34} \\ G_{41} & G_{42} & G_{43} & G_{44} \end{pmatrix} = \begin{pmatrix} I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{pmatrix},$$

or, equivalently,

$$(1) \begin{cases} M_{11}G_{11} & +M_{21}^{+}G_{21} & & & = I \\ M_{21}G_{11} & +M_{22}G_{21} & +M_{32}^{+}G_{31} & & = 0 \\ & M_{32}G_{21} & +M_{33}G_{31} & +M_{43}^{+}G_{41} & = 0 \\ & & M_{43}G_{31} & +M_{44}G_{41} & = 0 \end{cases},$$

$$(2) \begin{cases} M_{11}G_{12} & +M_{21}^{+}G_{22} & & & = 0 \\ M_{21}G_{12} & +M_{22}G_{22} & +M_{32}^{+}G_{32} & & = I \\ & M_{32}G_{22} & +M_{33}G_{32} & +M_{43}^{+}G_{42} & = 0 \\ & & M_{43}G_{32} & +M_{44}G_{42} & = 0 \end{cases},$$

$$(3) \begin{cases} M_{11}G_{13} & +M_{21}^{+}G_{23} & & & = 0 \\ M_{21}G_{13} & +M_{22}G_{23} & +M_{32}^{+}G_{33} & & = 0 \\ & M_{32}G_{23} & +M_{33}G_{33} & +M_{43}^{+}G_{43} & = I \\ & & M_{43}G_{33} & +M_{44}G_{43} & = 0 \end{cases},$$

$$(4) \begin{cases} M_{11}G_{14} & +M_{21}^{+}G_{24} & & & = 0 \\ M_{21}G_{14} & +M_{22}G_{24} & +M_{32}^{+}G_{34} & & = 0 \\ & M_{32}G_{24} & +M_{33}G_{34} & +M_{43}^{+}G_{44} & = 0 \\ & & M_{43}G_{34} & +M_{44}G_{44} & = I \end{cases}.$$

Solving (1) and (2), and defining

$$\begin{aligned} X_4 &= 0,, \\ X_3 &= M_{43}^{+}[M_{44} - X_4]^{-1} M_{43}, \\ X_2 &= M_{32}^{+}[M_{33} - X_3]^{-1} M_{32}, \\ X_1 &= M_{21}^{+}[M_{22} - X_2]^{-1} M_{21}, \end{aligned}$$

$$\begin{aligned} Y_1 &= 0,, \\ Y_2 &= M_{21}[M_{11} - Y_1]^{-1} M_{21}^{+}, \\ Y_3 &= M_{32}[M_{22} - Y_2]^{-1} M_{32}^{+}, \\ Y_4 &= M_{43}[M_{33} - Y_3]^{-1} M_{43}^{+}, \end{aligned}$$

the following expressions are obtained:

$$\begin{aligned} G_{41} &= -[M_{44} - X_4]^{-1} M_{43} G_{31}, \\ G_{31} &= -[M_{33} - X_3]^{-1} M_{32} G_{21}, \\ G_{21} &= -[M_{22} - X_2]^{-1} M_{21} G_{11}, \\ G_{11} &= [M_{11} - X_1 - Y_1]^{-1}. \end{aligned}$$

and

$$\begin{aligned} G_{42} &= -[M_{44} - X_4]^{-1} M_{43} G_{32}, \\ G_{32} &= -[M_{33} - X_3]^{-1} M_{32} G_{22}, \\ G_{22} &= [M_{22} - X_2 - Y_2]^{-1}, \\ G_{12} &= -[M_{11} - Y_1]^{-1} M_{21}^{+} G_{22}. \end{aligned}$$

In an analogous way G_{is} and G_{is} (for $1 \leq i \leq 4$) are obtained. The general rule for a general n-block matrix M is derived immediately:

$$\begin{aligned} G_{ii} &= [M_{ii} - X_i - Y_i]^{-1} \\ G_{ij} &= C_i G_{i-1,j} \quad for\ i > j \\ G_{ij} &= D_i G_{i+1,j} \quad for\ i < j \end{aligned}$$

for $1 \leq i, j \leq n$, where

$$\begin{aligned} X_n &= 0 \\ X_{n-i} &= M^{+}_{n-i+1,n-i}[M_{n-i+1,n-i+1} - X_{n-i+1}]^{-1} M_{n-i+1,n-i} \quad for\ 1 \leq i \leq n-1; \\ Y_1 &= 0 \\ Y_{i+1} &= M_{i+1,i}[M_{ii} - Y_i]^{-1} M^{+}_{i+1,i} \quad for\ 2 \leq i \leq n \\ C_i &= -[M_{ii} - X_i]^{-1} M_{i,i-1} \quad for\ 1 \leq i \leq n-1, \\ D_i &= -[M_{ii} - Y_i]^{-1} M^{+}_{i+1,i} \quad for\ 2 \leq i \leq n. \end{aligned}$$

4 Computational solution

The computer code of this algorithm is written in FORTRAN77, as a selfconsistent routine using a band storage mode for the matrix M (only the diagonal and infradiagonal blocks are stored); the required work-area[2] is less than or equal to $2 \times MDIM^2 \times (3 + NB)$, where $MDIM$ is the maximum block dimension and NB is the number of blocks of the matrix M.

The alternative methods selected to compare the results are LEQT2C and MA23A routines from IMSL and HARWELL librairies respectively (the only available in our computer center which accomplish the desired task), and a routine called INV, which is the simplest version of the inversion algorithm. All of them use full storage mode for the matrix M.

The LEQT2C routine cannot compute the inverse matrix for the different M proposed (the obtained return code indicates that the matrix is algorithmically singular).

[2] the storage area for the matrix M is not included

The numerical values obtained by the algorithm presented in this paper and by the MA23A and INV routines coincide, at least, until the forth significant digit. The CPU times required in a BASF-768 computer and the size of required work-area by the methods are resumed in the following tables:

Matrix dimension	30	50	100
MA23A	1050	2750	10500
this method	450(a)	650(b)-1600(c)	1150(d)-2600(e)

Table 1: Size of required work-area, in units of the length of a complex number (tipically, 8 bytes in single precision), for three matrix dimensions. Block dimensions used: (a) 5×5; (b) 5×5; (c) 10×10; (d) 5×5; (e) 10×10.

Matrix dimension	30	50	100
MA23A	0.54	1.88	7.70
INV	0.42	1.84	13.87
this method	0.26(a)	0.58(b)-1.25(c)	1.82(d)-3.74(e)

Table 2: CPU time involved, in seconds. Block dimensions used: (a) 5×5; (b) 5×5; (c) 10×10; (d) 5×5; (e) 10×10.

5 Conclusions

Considering the results described in the preceeding section, this method allows a more efficient use of memory, and computational times for the complex n-block tridiagonal quasi-hermitian matrices. This method, with the adecuate modifications, can also be applied to solve systems $GM = B$, where M is a matrix as described, specially if B is a diagonal matrix.

It is important to notice that the calculations of the matrices X and Y are independent, as it is for the infradiagonal and supradiagonal blocks; therefore, this method can be used in vectorial computers with the corresponding computer code.

Acknowledgments

This work has been partially supported by the Scientific Cooperation Contract Nr. ST2J-0254-7-E (EDB) between the European Communities and the Universidad Autó-

noma de Madrid (Spain). The author wishes to thank Prof. F. Flores for his support and stimulation during the development of this work. She also acknowledge A. Muñoz and J.C. Durán for drawing her attention to the problem discussed in this paper, and specially the second one, for reading the manuscript with fruitful criticism, which contributed to clarify its contents.

References

[1] J. M. Ziman, Principles of the Theory of Solids, Cambridge University (1963).

[2] N. W. Ashcroft y N. D. Mermin, Solid State Physics, Holt-Saunders International Editions (1981).

[3] F. Guinea, J. Sanchez-Dehesa y F. Flores, J. Physics C 16, 6499 (1983).

[4] A. Muñoz, J.C. Durán and F. Flores, Surface Science 181, L200 (1987).

[5] A. Muñoz, R. Pérez, J.C. Durán and F. Flores, Surface Science 211/212, 503 (1989)