Solution of a Fredholm integral equation of physical interest

1995 J. Phys. A: Math. Gen. 28 391

(http://iopscience.iop.org/0305-4470/28/2/015)

J. Phys. A: Math. Gen. 28 (1995) 391–405. Printed in the UK

Solution of a Fredholm integral equation of physical interest

F J Fernández-Velicia†, F García-Molinerǂ and V R Velascoǂ

† Departamento de Física de los Materiales, Facultad de Ciencias, UNED, Senda del Rey s/n, 28040 Madrid, Spain

ǂ Instituto de Ciencia de Materiales, CSIC, Serrano 123, 28006 Madrid, Spain

Received 8 September 1994, in final form 15 November 1994

Abstract. An integral equation of the Fredholm type which appears in problems of physical interest—like the dielectric response of an inhomogeneous confined quasi-tridimensional electron gas—is studied when the kernel contains an infinite number of dyadic terms. In terms of a representation, this amounts to inverting an infinite matrix. The conditions for the existence of a bound inverse are established and an explicit non-recursive algorithm is developed which can be used in practical calculations to generate successive approximations which converge to the exact answer.

1. Introduction and statement of the problem

Several problems of physical interest take the form of an integral equation of the Fredholm type:

$$f(r) = g(r) + \int \mathrm{d}r'\, k(r, r') f(r') \tag{1}$$

where g is given and f is the unknown. For instance, in the standard theories of screening of an external potential V_{ext} by an electron gas [1], g is the known V_{ext}, f is the unknown total potential $V_{\mathrm{tot}} = V_{\mathrm{ext}} + V_{\mathrm{ind}}$ and k is essentially the polarizability of the electron gas.

Physical analysis usually leads to an expression for k as a sum of an infinite number of dyadic terms. This is often truncated so the number of dyadic terms is finite by virtue of some approximation which, in practice, is made without formal justification. This paper is concerned with the case in which the number of dyadic terms is infinite.

The problem can be one-, two- or three-dimensional and we ignore the trivial situation of translational invariance, when k is a function of $r - r'$ and the problem can be simply solved by Fourier transform. Lack of translational invariance is encountered, for instance, in the one-dimensional problems posed by the layered epitaxial heterostructures which are the subject of intense current research [2]. An interesting example is the screening of an external potential by the electron gas confined in a quantum well. These systems are obtained by some form of controlled 'modulation doping' and detailed calculations show that the resulting electron gas can be very strongly inhomogeneous. It was the study of this problem which motivated the present work. However, the mathematical analysis to be presented here holds more generally, with r a one-, two- or three-dimensional position variable.

With this proviso we study the integral equation (1) for the case of a one-dimensional position variable z. Defining

$$K(z, z') = \delta(z - z') - k(z, z') \tag{2}$$

0305-4470/95/020391+15$19.50

the problem is to solve

$$g(z) = \int dz' \, K(z, z') f(z') \tag{3}$$

that is, to find $M(z, z')$ so that

$$f(z) = \int dz' \, M(z, z') g(z') \tag{4}$$

for which we must have

$$\begin{aligned} \int dz'' \, K(z, z'') M(z'', z') &= \delta(z - z') \\ \int dz'' \, M(z, z'') K(z'', z') &= \delta(z - z'). \end{aligned} \tag{5}$$

Now, the kernels we consider here are of the form

$$k(z, z') = \sum_{\mu, \nu} L_\mu^*(z) k_{\mu\nu} S_\nu(z'). \tag{6}$$

The relevance of this for the above mentioned screening problem in particular and the relationship of the L_μ and S_ν to the one-electron eigenfunctions of the quantum well has been established elsewhere [4] in the random phase approximation [1].

A full discussion of this problem is of sufficient complexity to constitute a separate publication by itself [4] but a brief indication of the basic setup wil clarify the motivation for the mathematical analysis presented here. The electronic states are the product of free-electron plane waves in the (x, y) plane and localized one-dimensional wavefunctions $\phi_n(z)$. The focus is on the latter, which are labelled by the discrete quantum number n. The spectrum also contains, in the higher-energy region, delocalized states, more or less distorted by the potential of the quantum well. To avoid formal complications with a continuous spectrum—which could also be formally handled but would simply add an inessential formal complication—we can choose confining infinite barriers at sufficiently large distances from the well edges and then all ϕ_n form a discrete spectrum. On performing the RPA analysis, one finds [4] a kernel $k(z, z')$ of the form

$$k(z, z') = \sum_{m,n'} \left[\int dz'' \, G(z, z'') \phi_n(z'') \phi_{n'}^*(z'') \right] P_{nn'} \left[\phi_n^*(z') \phi_{n'}(z') \right]$$

where the $P_{nn'}$ are polarizability terms the detailed structure of which is irrelevant at this stage, and $G(z, z'')$ is, after 2D Fourier transform, the Green function of the Poisson equation relating the induced charge density to the induced potential. The single products $\phi_n(z')\phi_{n'}(z')$ are short-range functions of z' while the integral containing $G(z, z'')$ is a long-range function of z reflecting the long-range nature of the Coulomb field. We denote these functions, respectively, as $S_{n,n'}(z')$ and $L_{nn'}^*(z)$. Now, the indices n, n' run from 0 to ∞, but they form denumerable sets. Therefore, we can redefine a different labelling $n, n' \to \mu$, so $k(z, z')$ is a sum of products $L_\mu^*(z) P_\mu S_\mu(z')$. Thus we have a kernel of the form (6) where, in this particular case, $k_{\mu\nu} = P_\mu \delta_{\mu\nu}$. However, the problem can be solved for arbitrary $k_{\mu\nu}$ and it is interesting to do so because there are other problems of physical interest where the mathematical problem is the same with $k_{\mu\nu}$ non-diagonal.

The heart of the problem is that, in principle, the summation in (6) involves an infinite number of terms and even though truncation to a finite number could constitute a good physical approximation to a direct expression of k, the ensuing matrix inversion process, presently to be discussed, requires a proper mathematical analysis and a formal justification based on a proof of the existence of an inverse for the initial problem before truncation.

For this we proceed in two steps. Firstly the problem is solved for a finite number of terms so that, in the summation of (6), $\mu, \nu = 1, \ldots, N$. We do this with an algorithm which lends itself to the study of the limit $N \to \infty$. Then this limit is studied so one can establish the conditions for the existence of the inverse. We thus obtain both the formal justification and an algorithm for performing practical calculations for any desired finite N.

2. Solution for finite N

Le us define $\boldsymbol{L}/\boldsymbol{S}$ as the vector with components L_μ/S_ν and $\mathbf{k}$ as the matrix of elements $k_{\mu\nu}$. We then write in compact form

$$M(z, z') = \delta(z - z') + \boldsymbol{L}^*(z) \cdot \mathbf{m} \cdot \boldsymbol{S}(z'). \tag{7}$$

We assume all the functions $L_\mu(z)$ and $S_\nu(z)$ are linearly independent, a property which holds in many situations of physical interest and, in particular, in the case of the dielectric response of the inhomogeneous electron gas [4]. Then these functions form a basis in their N-dimensional space. Otherwise we can always form a basis by making a suitable selection of the functions $L_\mu(z)$ and $S_\nu(z)$. Now, our unknown is the matrix $\mathbf{m}$ which represents a kernel $m(z, z')$ in the dual basis $\{L^*_\mu(z), S_\nu(z')\}$, just as the matrix $\mathbf{k}$ of (6) represents the kernel $k(z, z')$ in the same basis. We define the matrix β of elements

$$\beta_{\mu\nu} = \int \mathrm{d}z''\, S_\mu(z'') L^*_\nu(z'') \tag{8}$$

and then the first of equations (5), by (2), (6) and (7), reads

$$-\boldsymbol{L}^*(z) \cdot \mathbf{k} \cdot \boldsymbol{S}(z') + \boldsymbol{L}^*(z) \cdot \mathbf{m} \cdot \boldsymbol{S}(z') - \boldsymbol{L}^*(z) \cdot \mathbf{k} \cdot \beta \cdot \mathbf{m} \cdot \boldsymbol{S}(z') = 0. \tag{9}$$

Since the $\{L_\mu, S_\nu\}$ form a basis, this yields a matrix equation from which we obtain the formal solution

$$\mathbf{m} = (\mathbf{I}_N - \mathbf{k} \cdot \beta)^{-1} \cdot \mathbf{k} \tag{10}$$

where $\mathbf{I}_n$ is the $N \times N$ unit matrix. The same result is obtained by starting from the second of equations (5).

There are well known standard algorithms for matrix inversion. However, these do not lend themselves to the study of the limit $N \to \infty$, which is our ultimate goal, so we develop a different algorithm which suits this purpose.

Consider any given matrix $\mathbf{a}$—which in this case would be $\mathbf{I}_N - \mathbf{k} \cdot \beta$. Following a standard practice we can factorize it as the product

$$\mathbf{a} = \mathbf{P} \cdot \mathbf{T} \tag{11}$$

of a unitary matrix $\mathbf{P}$ and an upper triangular matrix $\mathbf{T}$ satisfying

$$T_{ij} = 0 \qquad (i > j). \tag{12}$$

Starting from

$$\sum_{r=1}^{j} P_{ir} T_{rj} = a_{ij} \tag{13}$$

multiplying by P^*_{im}, summing over i and using the unitary character of $\mathbf{P}$, we obtain

$$T_{mj} = \sum_{i=1}^{N} P^*_{im} a_{ij}. \tag{14}$$

Furthermore, multiplying (13) by its complex conjugate and summing over i we also obtain

$$\sum_{i=1}^{N} |a_{ij}|^2 = \sum_{r=1}^{i} |T_{rj}|^2. \tag{15}$$

All these equalities are invariant under a transformation of the type

$$P_{ij} = e^{-i\theta_i} p_{ij} \qquad T_{ij} = e^{-i\theta_i} t_{ij}. \tag{16}$$

Since every T_{rr} is $t_{rr} \exp(i\phi_r)$, we can choose the θ_i so all T_{jj} are real and equal to their moduli:

$$t_{jj} \geqslant 0. \tag{17}$$

We denote by $\mathbf{P}$ and $\mathbf{T}$ the corresponding $\mathbf{P}$ and $\mathbf{T}$ matrices. Now, in (13), we separate out the term $r = j$ and write

$$p_{ij} t_{jj} = a_{ij} - \sum_{r=1}^{(j-1)}{}' p_{ir} t_{rj} \equiv q_{ij}. \tag{18}$$

The prime on the summation indicates that the sum is nil by definition when $j = 1$ and the second equality defines q_{ij}. Obviously, q_{i1} is just a_{i1}. Now multiply (18) by a^*_{ij} and sum over i. This yields

$$\sum_{i=1}^{N} a^*_{ij} q_{ij} = t^2_{jj}. \tag{19}$$

Also, from the definition of q_{ij} we obtain

$$\sum_{i=1}^{N} |q_{ij}|^2 = t^2_{jj}. \tag{20}$$

whence

$$\sum_{i=1}^{N} a^*_{ij} q_{ij} = \sum_{i=1}^{N} |q_{ij}|^2. \tag{21}$$

Then, by using (14) in (21) and expressing every $p_{\ell m}$ as $q_{\ell m}/t_{mm}$, we obtain the recurrence relation

$$\begin{aligned} q_{i1} &= a_{i1} \\ q_{ij} &= a_{ij} - \sum_{s=1}^{N} \left[\sum_{k=1}^{(j-1)\prime} \left(\frac{q_{ik} q_{sk}^*}{\Sigma^{(k)}} \right) \right] a_{sj} \\ \Sigma^{(k)} &\equiv \sum_{m=1}^{N} a_{mk}^* q_{mk} \end{aligned} \tag{22}$$

from which follow

$$\begin{aligned} p_i &= \frac{q_{ij}}{\left[\sum_{m=1}^{N} |q_{mj}|^2 \right]^{1/2}} \\ t_{ij} &= \sum_{s=1}^{N} \frac{q_{si}^* a_{sj}}{\left[\sum_{m=1}^{N} |q_{mi}|^2 \right]^{1/2}}. \end{aligned} \tag{23}$$

This is not yet the final form in which we shall obtain **P** and **T** but these results will be used presently.

It follows from the properties of **P** and **T** that

$$\frac{\det |\mathbf{a}|}{\det |\mathbf{P}|} = \det |\mathbf{T}| = \prod_{j=1}^{N} t_{jj} \tag{24}$$

we note that $\det |\mathbf{P}|$ is a number of modulus equal to unity which never vanishes and thus the matrix **a** is singular if at least one of the t_{jj} vanishes. We shall prove that the necessary and sufficient condition for this to happen is that the jth column of **a** be a linear combination of the preceding $(j-1)$ columns. Indeed, let us assume that

$$a_{ij} = \sum_{s=1}^{(j-1)} \lambda_s a_{is} \qquad (i = 1, 2, \ldots, N). \tag{25}$$

Then (14) yields

$$t_{jj} = \sum_{s=1}^{(j-1)} \lambda_s \left(\sum_{i=1}^{N} p_{ij}^* a_{is} \right) = \sum_{s=1}^{(j-1)} \lambda_s t_{js} = 0. \tag{26}$$

Conversely, assume

$$t_{jj} = 0. \tag{27}$$

Then from the upper triangular nature of **T**:

$$\sum_{r=1}^{j} \mu_r t_{jr} = 0 \qquad (\mu_j \neq 0). \tag{28}$$

Using (14) again we have

$$\sum_{i=1}^{N} p_{ij}^{*}\left(\sum_{r=1}^{j}\mu_r a_{ir}\right)=0 \tag{29}$$

whence, multiplying by p_{mj}, summing over m and using the unitary character of **P**, we have

$$\sum_{i=1}^{N}\left(\sum_{j=1}^{N} p_{mj}p_{ij}^{*}\right)\left(\sum_{r=1}^{j}\mu_r a_{ir}\right)=\sum_{r=1}^{j}\mu_r a_{mr}=0 \tag{30}$$

where, by assumption, $\mu_j \neq 0$. Thus, since the argument holds for all $m = 1, 2, \ldots, N$, every a_{mj} is of the form (19) and the jth column of **a** is a linear combination of the preceding $(j-1)$ columns. We assume this is not the case, since we assume that the column vectors of **a** are all linearly independent.

The above argument is related to the identical vanishing, or not, of the t_{jj}. Now, in physics, all these matrix elements are functions of some parameters and it may happen that some t_{jj} vanish for some particular values of these parameters. This is typically associated with some normal mode of the system. For instance, in the case of the dielectric response of the quasi two-dimensional electron gas, i.e. confined in the z direction, the parameters are the frequency ω and the 2D wavevector Q. The equations of the form

$$t_{jj}(\omega, Q)=0 \tag{31}$$

then yield the dispersion relations $\omega(Q)$ for the plasma modes of the confined quasi 2D electron gas. However, when studying the problem of screening of an external potential with given independent ω and Q, these are not related by any normal mode relationship and (31) does not hold. After this clarification, we return to the inversion of **a**, given by

$$\mathbf{a}^{-1}=\mathbf{T}^{-1}\cdot\mathbf{p}^{\dagger} \tag{32}$$

which is reduced to the inversion of **T**.

Before studying the inversion of **a**, we re-examine the solution obtained in (22) and (23) for Q, **P** and **T**. From the **a** matrix, we define the vectors

$$a^{(s)}\equiv\{a_{1s}, a_{2s}, \ldots, a_{Ns}\} \tag{33}$$

and likewise for the vectors $P^{(s)}$ and $Q^{(s)}$ from the **P** and **T** matrices. In the N-dimensional space spanned by the basis formed by the vectors (33), assumed to be linearly independent, we define the scalar product and norm

$$(a^{(i)}\cdot a^{(j)})=\sum_{r=1}^{N} a_{ri}^{*}a_{rj} \qquad ||a^{(i)}||=(a^{(i)}\cdot a^{(i)})^{1/2} \tag{34}$$

and likewise for the $P^{(s)}$ and $Q^{(s)}$. The above recurrence formulae can then be concisely cast as

$$\begin{aligned} Q^{(1)}&=a^{(1)}\\ Q^{(j)}&=a^{(j)}-\sum_{k=1}^{(j-1)'} Q^{(k)}\frac{(Q^{(k)}\cdot a^{(j)})}{(Q^{(k)}\cdot Q^{(k)})}. \end{aligned} \tag{35}$$

Since $\mathbf{P}$ is unitary, the vectors $\boldsymbol{P}^{(i)}$ form an orthonormal basis while, by (18), the vectors $\boldsymbol{Q}^{(i)}$ form an orthogonal basis. Indeed it can be checked that (22) amounts to an orthogonalization procedure of the standard kind, and

$$(\boldsymbol{Q}^{(i)} \cdot \boldsymbol{Q}^{(j)}) = t_{ii}^2 \delta_{ij}. \tag{36}$$

However, we stress that the $\boldsymbol{a}^{(i)}$ are *not* assumed, in general, to be orthogonal. Now define the vectors

$$\boldsymbol{\sigma}^{(j)} \equiv \frac{1}{\Delta^{(j-1)}} \begin{vmatrix} (\boldsymbol{a}^{(1)} \cdot \boldsymbol{a}^{(1)}) & (\boldsymbol{a}^{(1)} \cdot \boldsymbol{a}^{(2)})^* & \dots & (\boldsymbol{a}^{(1)} \cdot \boldsymbol{a}^{(j-1)})^* & \boldsymbol{a}^{(1)} \\ (\boldsymbol{a}^{(2)} \cdot \boldsymbol{a}^{(1)})^* & (\boldsymbol{a}^{(2)} \cdot \boldsymbol{a}^{(2)}) & \dots & (\boldsymbol{a}^{(2)} \cdot \boldsymbol{a}^{(j-1)})^* & \boldsymbol{a}^{(2)} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ (\boldsymbol{a}^{(j)} \cdot \boldsymbol{a}^{(1)})^* & (\boldsymbol{a}^{(j)} \cdot \boldsymbol{a}^{(2)})^* & \dots & (\boldsymbol{a}^{(j)} \cdot \boldsymbol{a}^{(j-1)})^* & \boldsymbol{a}^{(j)} \end{vmatrix} \tag{37}$$

$$\boldsymbol{\sigma}^{(i)} \equiv \boldsymbol{a}^{(1)}$$

where $\Delta^{(j-1)}$ is the cofactor of $\boldsymbol{a}^{(j)}$ in (37) and this holds only for $j > 1$.

Then, for $j > 1$

$$\left(\boldsymbol{\sigma}^{(j)} \cdot \boldsymbol{a}^{(k)}\right) = 0 \qquad k = 1, 2, \dots, (j-1). \tag{38}$$

Now, for all $s = 1, \dots, j-1, j, \dots, N$, every $\boldsymbol{Q}^{(s)}$ is a linear combination of $\{\boldsymbol{a}^{(1)}, \boldsymbol{a}^{(2)}, \dots, \boldsymbol{a}^{(s)}\}$. Thus, the space spanned by $\{\boldsymbol{Q}^{(1)}, \boldsymbol{Q}^{(2)}, \dots, \boldsymbol{Q}^{(j-1)}\}$ is the same as that spanned by $\{\boldsymbol{a}^{(1)}, \boldsymbol{a}^{(2)}, \dots, \boldsymbol{a}^{(j-1)}\}$ and due to the orthogonality of the $\boldsymbol{Q}^{(r)}$ vectors (36) we have, in particular, also for $j > 1$,

$$(\boldsymbol{Q}^{(j)}, \boldsymbol{Q}^{(k)}) = 0 \qquad k = 1, 2, \dots, (j-1) \tag{39}$$

whence, for $j > 1$,

$$(\boldsymbol{Q}^{(j)}, \boldsymbol{a}^{(k)}) = 0 \qquad k = 1, 2, \dots, (j-1). \tag{40}$$

Thus, both $\boldsymbol{\sigma}^{(j)}$ and $\boldsymbol{Q}^{(j)}$ are orthogonal to all vectors of the space spanned by $\{\boldsymbol{a}^{(1)}, \boldsymbol{a}^{(2)}, \dots, \boldsymbol{a}^{(j-1)}\}$. From (38) and (40) we can now obtain an important result.

First we note that by expanding the numerator of (37), through the elements of the last column, the vector $\boldsymbol{\sigma}^{(j)}$ has the form

$$\boldsymbol{\sigma}^{(j)} = \boldsymbol{a}^{(j)} - \sum_{s=1}^{(j-1)'} \rho_{j,s} \boldsymbol{a}^s \tag{41}$$

and $\boldsymbol{Q}^{(j)}$, given in (23), is also of the form

$$\boldsymbol{Q}^{(j)} = \boldsymbol{a}^{(j)} - \sum_{s=1}^{(j-1)'} r_{j,s} \boldsymbol{a}^s. \tag{42}$$

Note that the form of (41) and (42) is not incompatible with (38) or (40) because the basis formed by the $\boldsymbol{a}^{(j)}$ is not orthogonal. Now consider the vector

$$\boldsymbol{Q}^{(j)} - \boldsymbol{\sigma}^{(j)} = \sum_{s=1}^{(j-1)'} (\rho_{j,s} - \gamma_{j,s}) \boldsymbol{a}^s. \tag{43}$$

This is a vector in the space spanned by $\{a^{(1)}, a^{(2)}, \ldots, a^{(j-1)}\}$ but at the same time, by (38) and (40), it is orthogonal to all the vectors of this basis. Therefore, this vector vanishes identically whence we have proved that

$$Q^{(j)} = \sigma^{(j)}. \tag{44}$$

Thus, the q_{ij}, first obtained in the form of a recurrent relation, is given by (37) which is a constructive, non-recurrent formula giving the q_{ij} directly from the a_{ij}.

Let us return to the determinant $\Delta^{(r)}$ defined in (37)—in this case $r = j - 1$. We shall denote this as $\Delta^{(r)}[a]$ to indicate that it is formed by scalar products of a vectors. Likewise, $\Delta^{(r)}[\alpha]$ will denote the same determinant formed by scalar vectors of another set $\{\alpha\}$. Let these be related to the vectors of the set $\{a\}$ by a linear transformation of the form

$$\alpha^{(r)} = \sum_{s=1}^{N} \omega_{i,s} a^{(s)} \qquad i = 1, 2, \ldots, r \tag{45}$$

such that the matrix of elements $\omega_{i,s}$ $(i, s = 1, 2, \ldots, r)$ has a non-vanishing determinant $\omega^{(r)}$. Then

$$\Delta^{(r)}[\alpha] = |\omega^{(r)}|^2 \Delta^{(r)}[a]. \tag{46}$$

Thus, if one of these determinants is non-vanishing then the other is also non-vanishing.

Now consider, in particular, the linear relationship between the $Q^{(i)}$ and the $a^{(s)}$ which results from using (44) and expanding the numerator of (37) through the elements of the last column. This is a linear transformation of the form (45) in which the matrix of ω coefficients is a lower triangular matrix with all the diagonal elements equal to unity. Thus,

$$\Delta^{(j)}[Q] = \Delta^{(j)}[a]. \tag{47}$$

However, we have seen that the $Q^{(i)}$ vectors form an orthogonal set. Therefore

$$\Delta^{(j)}[Q] = \prod_{r=1}^{j} ||Q^{(r)}||^2 \tag{48}$$

and this is always non-negative and real. Hence $\Delta^{(j)}$, defined in (30), is real and non-negative.

On the other hand, $Q^{(j)}$ is of the form (35) and has the property (33). Hence, by (30),

$$(Q^{(j)} \cdot Q^{(j)}) = (Q^{(j)} \cdot a^{(j)}) = (a^{(j)} \cdot Q^{(j)}) = \frac{\Delta^{(j)}}{\Delta^{(j-1)}} \tag{49}$$

and

$$||Q^{(j)}|| = \left[\frac{\Delta^{(j)}}{\Delta^{(j-1)}}\right]^{\frac{1}{2}}. \tag{50}$$

With this we can write down the desired formulae for the p_{ij} and t_{ij} in final form. Since p_{ij} is given by (23), we have

$$p_{ij} = [\Delta^{(j-1)}\Delta^{j}]^{-1/2} \begin{vmatrix} (a^{(1)} \cdot a^{(1)}) & (a^{(1)} \cdot a^{(2)})^* & \ldots & (a^{(1)} \cdot a^{(j-1)})^* & a_{i1} \\ (a^{(2)} \cdot a^{(1)})^* & (a^{(2)} \cdot a^{(2)}) & \ldots & (a^{(2)} \cdot a^{(j-1)})^* & a_{i2} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ (a^{(j)} \cdot a^{(1)})^* & (a^{(j)} \cdot a^{(2)})^* & \ldots & (a^{(j)} \cdot a^{(j-1)})^* & a_{ij} \end{vmatrix} \tag{51}$$

for all $i, j = 1, \ldots, N$. It follows also from (23) that

$$t_{ij} = \frac{(Q^{(i)} \cdot a^{(j)})}{||Q^i||} \tag{52}$$

whence the explicit formula

$$t_{ij} = [\Delta^{(i-1)}\Delta^i]^{-1/2} \begin{vmatrix} (a^{(1)} \cdot a^{(1)}) & (a^{(1)} \cdot a^{(2)}) & \ldots & (a^{(1)} \cdot a^{(i-1)}) & (a^{(1)} \cdot a^{(j)}) \\ (a^{(2)} \cdot a^{(1)}) & (a^{(2)} \cdot a^{(2)}) & \ldots & (a^{(2)} \cdot a^{(i-1)}) & (a^{(2)} \cdot a^{(j)}) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ (a^{(i)} \cdot a^{(1)}) & (a^{(i)} \cdot a^{(2)}) & \ldots & (a^{(i)} \cdot a^{(i-1)}) & (a^{(i)} \cdot a^{(j)}) \end{vmatrix}. \tag{53}$$

We stress that the matrix elements p_{ij} and t_{ij} are obtained here not from the usual type of recurrence relations but by the explicit formulae (51) and (53). In particular

$$t_{ii} = \frac{||Q^{(i)}||^2}{||q^{(i)}||} = \left[\frac{\Delta^{(i)}}{\Delta^{(i-1)}}\right]^{\frac{1}{2}}. \tag{54}$$

These results allow us to formulate an algorithm for the inversion of $\mathbf{T}$ with which we can eventually study the limit $N \to \infty$.

3. The inversion of t and the limit $N \to \infty$

Let the subindex N denote a matrix of finite order N and consider, in particular, the upper triangular matrix $\mathbf{T}_{(N)}$. This can again be factorized as the product

$$\mathbf{T}_{(N)} = \tau_{(N)} \cdot \mathbf{d}_{(N)} \tag{55}$$

where $\mathbf{d}_{(N)}$ is the diagonal matrix of elements

$$(\mathbf{d}_{(N)})_{ij} = \delta_{ij} t_{jj} \tag{56}$$

and

$$(\tau_{(N)})_{ij} = \begin{cases} 0 & \text{if } i > j \\ 1 & \text{if } i = j \\ \dfrac{t_{ij}}{t_{jj}} & \text{if } i < j \end{cases} \qquad (i, j = 1, 2, \ldots, N). \tag{57}$$

For given $N > 2$, we introduce the $N - 1$ vectors $b^{(j)}$ with components

$$b_k^{(j)} = \frac{t_{k,j+1}}{t_{j+1,j+1}} \qquad (j = 1, 2, \ldots, N-1; k = 1, 2, \ldots, j) \tag{58}$$

and corresponding matrices

$$\mathbf{b}_{(N)}^{(j)} = \begin{Vmatrix} 0 \cdots 0 & b_1^{(j)} & 0 \cdots 0 \\ \ldots & \ldots & \ldots \\ 0 \cdots 0 & b_j^{(j)} & 0 \cdots 0 \\ 0 \cdots 0 & 0 & 0 \cdots 0 \\ \ldots & \ldots & \ldots \\ 0 \cdots 0 & 0 & 0 \cdots 0 \end{Vmatrix} \tag{59}$$

where the $(j+1)$th column consists of the j components of $b^{(j)}$ and $(N-j)$ zeros, for $j = 2, \ldots, N-1$.

Then

$$\tau_{(N)} = \mathbf{I}_{(N)} + \sum_{j=1}^{N-1} \mathbf{b}_{(N)}^{(j)}. \tag{60}$$

It is easily seen that

$$\mathbf{b}_{(N)}^{(i)} \cdot \mathbf{b}_{(N)}^{(j)} = 0 \qquad \text{if } i \leqslant j \tag{61}$$

whence

$$\tau_{(N)} = [\mathbf{I}_{(N)} + \mathbf{b}_{(N)}^{(N-1)}] \cdot [\mathbf{I}_{(N)} + \mathbf{b}_{(N)}^{(N-2)}] \cdots [\mathbf{I}_{(N)} + \mathbf{b}_{(N)}^{(1)}] \tag{62}$$

and

$$\tau_{(N)}^{-1} = [\mathbf{I}_{(N)} - \mathbf{b}_{(N)}^{(1)}] \cdot \ldots \cdot [\mathbf{I}_{(N)} - \mathbf{b}_{(N)}^{(N-2)}] \cdot [\mathbf{I}_{(N)} - \mathbf{b}_{(N)}^{(N-1)}]. \tag{63}$$

The matrices $\mathbf{b}_{(N)}^{(i)}$ and the products $\mathbf{b}_{(N)}^{(i)} \cdot \mathbf{b}_{(N)}^{(j)}$ entering (63) can be written in the form

$$\mathbf{b}_{(N)}^{(i)} = \begin{Vmatrix} \mathbf{b}_{(N-1)}^{(i)} & \mathbf{o} \\ \mathbf{o} & 0 \end{Vmatrix} \qquad \mathbf{b}_{(N)}^{(i)} \cdot \mathbf{b}_{(N)}^{(j)} = \begin{Vmatrix} \mathbf{b}_{(N-1)}^{(i)} \cdot \mathbf{b}_{(N-1)}^{(j)} & \mathbf{o} \\ \mathbf{o} & 0 \end{Vmatrix} \tag{64}$$

so that (63) reads

$$\tau_{(N)}^{-1} = \begin{Vmatrix} \tau_{(N-1)}^{-1} & \mathbf{o} \\ \mathbf{o} & 1 \end{Vmatrix} ; \cdot \left\| \mathbf{I}_{(N)} - \mathbf{b}_{(N)}^{(N-1)} \right\|. \tag{65}$$

Now, the case $N = 2$ is trivial and, for $N = 3$, (63) yields

$$\tau_{(3)}^{-1} = \mathbf{I}_{(3)} - \mathbf{b}_{(3)}^{(1)} - \mathbf{b}_{(3)}^{(2)} + \mathbf{b}_{(3)}^{(1)} \cdot \mathbf{b}_{(3)}^{(2)}. \tag{66}$$

From (63) and (65), it is easy to prove by induction that

$$\tau_{(N)}^{-1} = \sum_{r=0}^{N-1} (-1)^r \sum_{\{\alpha; r\}} [\mathbf{b}_{(N)}^{(1)}]^{\alpha_1} \cdot [\mathbf{b}_{(N)}^{(2)}]^{\alpha_2} \cdots [\mathbf{b}_{(N)}^{(N-1)}]^{\alpha_{N-1}} \tag{67}$$

where, by definition, all α_i take only the values 0, 1 and the symbol $\{\alpha; r\}$ in the second summation indicates that for every r the α_i satisfy the condition $\alpha_1 + \alpha_2 + \cdots + \alpha_{N-1} = r$.

Now, the inverse of $\tau_{(N)}$ is of the same form as $\tau_{(N)}$ and we can define $N-1$ vectors $\mathbf{S}^{(j)}$ and corresponding $N-1$ matrices $\mathbf{S}^{(j)}$ so that they have the property (61) and the formulae relating the $\mathbf{S}_{(N)}^{(j)}$ to the $\mathbf{S}^{(j)}$ and $\tau_{(N)}^{-1}$ to the $\mathbf{S}_{(N)}^{(j)}$ are isomorphic with (58) through (60). In particular,

$$\mathbf{S}_{(N)}^{(j)} = \begin{Vmatrix} 0 \cdots 0 & s_1^{(j)} & 0 \cdots 0 \\ \cdots & \cdots & \cdots \\ 0 \cdots 0 & s_j^{(j)} & 0 \cdots 0 \\ 0 \cdots 0 & 0 & 0 \cdots 0 \\ \cdots & \cdots & \cdots \\ 0 \cdots 0 & 0 & 0 \cdots 0 \end{Vmatrix} \tag{68}$$

$$\tau_{(N)}^{-1} = \mathbf{I}_{(N)} + \sum_{j=1}^{N-1} \mathbf{S}_{(N)}^{(j)}.$$

We are now ready to study the limit $N \to \infty$.

First, we note that if all the columns of the given infinite matrix $\boldsymbol{a}$ belong to the Hilbert space ℓ_2, then it follows from the unitary character of $\mathbf{P}$ that

$$\sum_{i=1}^{j} |b_i^{(j)}|^2 < \infty \tag{69}$$

for j finite or infinite. This is a basic property and we shall assume throughout that this condition is satisfied. We now consider in τ_∞ the succession of elements $\{b_{N-j}^{(N)}\}$, in general complex numbers. For each fixed value of $j = 0, 1, \ldots$ these are the elements of a given parallel to the principal diagonal. We keep each j fixed when $N \to \infty$ and consider three cases.

Case 1. $\forall j \in \{0, 1, 2 \ldots\}$

$$\lim_{N\to\infty} b_{N-j}^{(N)} \equiv \beta_j < \infty. \tag{70}$$

Case 2. $\exists j \in \{0, 1, 2 \ldots\}$ for which the limit of (70) does not exist but $\forall j \in \{0, 1, 2 \ldots\}$ either (i)

$$\lim_{N\to\infty} |b_{N-j}^{(N)}| < \infty \tag{71}$$

or (ii) the succession $\{|b_{N-j}^{(N)}|\}$ is oscillatory and bounded between finite values.

Case 3. $\exists j \in \{0, 1, 2 \ldots\}$ for which either (i)

$$\lim_{N\to\infty} |b_{N-j}^{(N)}| = \infty \tag{72}$$

or (ii) the succession $\{|b_{N-j}^{(N)}|\}$ is oscillatory and is not bounded between finite values.

Consider case 1. From (68), we can cast (64) as a recursion formula

$$\begin{aligned} s_{(N-j)}^{(N)} &= -b_{(N-j)}^{(N)} - {\sum_{r=0}^{j-1}}' s_{(N-j)}^{(N-1-r)} b_{(N-j)}^{(N)} \\ j &= 0, 1, 2, \ldots, N-1 \qquad N = 1, 2, \ldots, \infty \end{aligned} \tag{73}$$

from which it is clear that if the limit of (70) exists and if it is finite then

$$\lim_{N\to\infty} s_{N-j}^{(N)} \equiv \sigma_j < \infty \tag{74}$$

and defining

$$\hat{\sigma}_0 = 1 \qquad \hat{\sigma}_j = \sigma_{j-1} \qquad (j = 1, 2, \ldots) \tag{75}$$

we cast the limit $N \to \infty$ of (73) as

$$\hat{\sigma}_{j+1} = -{\sum_{r=0}^{j}}' \hat{\sigma}_{j-r} \beta_r. \tag{76}$$

Now consider a complex variable z which can have arbitrarily small modulus. From (76) we obtain

$$\sum_{j=0}^{\infty} \hat{\sigma}_{j+1} z^{j+1} = -z\left(\sum_{j=0}^{\infty} \hat{\sigma}_j z^j\right)\left(\sum_{j=0}^{\infty} \beta_j z^j\right) \tag{77}$$

which, recalling (75), is

$$\sum_{j=0}^{\infty} \sigma_j z^{j+1} = -1 + \frac{1}{1 + \sum_{j=0}^{\infty} \beta_j z^{j+1}}. \tag{78}$$

If $\beta_j = 0, \forall j \in \{0, 1, 2, \ldots\}$, then the same holds for all σ_j and then $\tau_{(\infty)}^{-1}$ exists and is bounded, as is $\tau_{(\infty)}$. Quite generally, we define the functions

$$\begin{aligned} f(z) &= 1 + \sum_{j=0}^{\infty} \beta_j z^{j+1} \\ g(z) &= \sum_{j=0}^{\infty} \sigma_j z^{j+1} \end{aligned} \tag{79}$$

and note that, on account of (69), we have

$$1 + \sum_{j=0}^{\infty} |\beta_j|^2 < \infty. \tag{80}$$

Hence, $f(z)$ is analytical, at least inside the circle $\bar{\mathcal{D}}(0, 1)$ of unit radius centred at the origin $z = 0$. Two situations arise, depending on whether or not $f(z)$ is analytical in the boundary $\mathcal{F}$ of $\bar{\mathcal{D}}(0, 1)$, that is the circumference $|z| = 1$. A different theorem can be proved for each of these situations.

Theorem 1.A. If the function $f(z)$, defined by (79), is also analytical in $\mathcal{F}$, then a necessary and sufficient condition for $\tau_{(\infty)}$ to have a bounded inverse $\tau_{(\infty)}^{-1}$ is that

$$f(z_0) \neq 0 \qquad \forall z_0 \in \bar{\mathcal{D}}(0, 1). \tag{81}$$

Proof. Assume that $f(z_0)$ vanishes for some $z_0 \in \bar{\mathcal{D}}(0, 1)$. Then, as a consequence of the analyticity of $f(z)$ in $\bar{\mathcal{D}}(0, 1)$, it follows from (78) and (79) that $g(z)$ has a pole at z_0. Therefore, the radius of convergence of the series defining $g(z)$ is < 1 and $\lim_{j\to\infty} \sigma_j = \infty$. This implies that the series

$$\sum_{j=0}^{\infty} |\sigma_j|, \sum_{j=0}^{\infty} |\sigma_j|^2 \tag{82}$$

diverge and then $\tau_{(\infty)}$ does not have a bounded inverse. Thus, (81) is a necessary condition.

Conversely, assume that (81) holds. Then, $g(z)$ is also analytical in $\bar{\mathcal{D}}(0, 1)$, including its boundary $\mathcal{F}$. Therefore, the radius of convergence of the series defining $g(z)$ is $\geqslant 1$, the two series of (82) converge and $\tau_{(\infty)}$ has a bounded inverse. Thus, (81) is also a sufficient condition. □

A different situation arises if $f(z)$ is analytical *inside* $\bar{\mathcal{D}}(0, 1)$ but *not* in its boundary $\mathcal{F}$. Let us write every $z \in \mathcal{F}$ in the form $z = \exp(i\theta)$, $\theta = [0, 2\pi]$. Then the series defining $f(z)$ in $\mathcal{F}$ is a particular case of a Fourier series of the form

$$\sum_{n=-\infty}^{\infty} c_n \exp(i\theta_n) \tag{83}$$

which must correspond to some function of θ in $[0, 2\pi]$, which, in principle, can only be $f(\exp(i\theta))$. The problem is that the c_n may or may not satisfy the conditions for the series of (83) to be of class C^1 with respect to θ and this is not always guaranteed. For instance, it is not when $c_n \sim n^{-\rho}$ with $1/2 < \rho < 1$. In this case, the Fourier series of (83) is not guaranteed to converge locally to $f(\exp(i\theta))$ for all $\theta \in [0, 2\pi]$ and there may even exist a set of points in this interval for which the series diverges. The question arising is the following: given that

$$\sum_{n=-\infty}^{\infty} |c_n|^2 < \infty \tag{84}$$

in what type of subsets of $\mathcal{F}$ is it possible that series (83) does not converge pointwise to $f(\exp(i\theta))$?

A first answer to this question can be stated as follows [5]: if $f(\exp(i\theta))$ is a square-summable function in $[0, 2\pi]$ then its corresponding Fourier series converges pointwise to f almost everywhere.

Therefore, in order to give a definitive answer to the question posed, we only need to show that if (84) holds then $f(\exp(i\theta))$ is square summable.

To this end, let $L^2([0, 2\pi])$ denote the set of all complex functions $\phi(\theta)$ having a Lebesgue measure in $[0, 2\pi]$ for which the norm

$$\|\phi\| = \left\{ \frac{1}{2\pi} \int_0^{2\pi} |\phi(\theta)|^2 \, d\theta \right\}^{1/2} < \infty. \tag{85}$$

With the standard inner product, the ϕ functions constitute a Hilbert space in which the functions $u_n(\theta) = \exp(in\theta)$ ($n \in \mathcal{Z}$) form an orthogonal basis which is maximal because the set of all trigonometric polynomials is dense in $L^2([0, 2\pi])$. Under these conditions, the Riesz–Fischer theorem [6] proves that if $\{c_n\}$ is a succession of complex numbers satisfying (84), then there exists a function $\phi(\theta) \in L^2([0, 2\pi])$ such that

$$c_n = \frac{1}{2\pi} \int_0^{2\pi} \phi(\theta) \exp(-in\theta) \, d\theta \tag{86}$$

which in our case can only be $f(\exp i\theta)$. Then it follows from Parseval's and Plancherel's theorems that

$$\|f(\exp(i\theta))\| = \sum_{n=-\infty}^{\infty} |c_n|^2 < \infty \tag{87}$$

i.e. f has finite norm. Then series (83), that is

$$1 + \sum_{n=0}^{\infty} \beta_n \exp[i(n+1)\theta] \tag{88}$$

converges pointwise to the function $f(\exp i\theta)$ everywhere in $[0, 2\pi]$ except possibly in a set of zero measure. We can now prove the following theorem.

Theorem 1.B. Assume the function $f(z)$ defined by (79) is analytical only in $\mathcal{D}(0, 1)$, i.e. inside the domain $\bar{\mathcal{D}}(0, 1)$ but not in its boundary $\mathcal{F}$. Then the necessary and sufficient conditions of $\tau_{(\infty)}$ to have a bounded inverse $\tau_{(\infty)}^{-1}$ are the following:

(i) that $f(z) \neq 0\ \forall z \in \mathcal{D}(0, 1)$;

(ii) that if $z_0 = \exp(i\theta_0)$ is a point of $\mathcal{F}$ where $f(z_0) = 0$, then for a sufficiently small interval of values of θ about θ_0, we have

$$|f(\exp(i\theta))| \leqslant K|\theta - \theta_0|^{\rho} \tag{89}$$

with K and ρ real positive constants and $0 < \rho < \frac{1}{2}$. Moreover, the points of $\mathcal{F}$ satisfying these conditions must constitute at most a set of zero measure.

Proof. Assume $f(z)$ vanishes for some point of $\mathcal{D}(0, 1)$ where it is, by assumption, analytical. Using the same type of argument as employed in theorem 1.A, we then find that σ_N diverges for $N \to \infty$. Then the series of (82) diverge and $\tau_{(\infty)}$ has no bounded inverse. Thus, (i) is necessary. Furthermore, assume that in (89) $\rho > \frac{1}{2}$ and that the subset of $\mathcal{F}$ for which this holds does not have zero measure. Then the function

$$g(\exp(i\theta)) = -1 + \frac{1}{f(\exp(i\theta))} \tag{90}$$

does not have a finite norm, the series of (82) diverges and $\tau_{(\infty)}$ does not have a bounded inverse. Thus, (ii) is also necessary.

Now assume that these conditions hold. Then the function g of (90) has finite norm, the series of (82) converges and $\tau_{(\infty)}$ has a bounded inverse. Thus, (i) and (ii) are sufficient. □

We now consider case 2, so that for all j the succession $\{b_{N-j}^{(N)}\}$ has at least the property that all its elements are contained in a compact domain Ω of the complex plane. Then there are two possibilites:

(i) that beyond a certain $N_0 \in \mathcal{N}$ all elements $b_{N-j}^{(N)}$ with $N > N_0$ which are different form a finite set $\beta_j(k)$, that is to say, $k \in \mathcal{K}$ where $\mathcal{K} \in \mathcal{N}$ is finite; and

(ii) that this set is infinite.

In case (i), we must study the different functions

$$f_k(z) = 1 + \sum_{j=0}^{\infty} \beta_j(k) z^{j+1} \tag{91}$$

and verify whether or not the conditions of theorems 1.A and 1.B hold. This decides the conditions for the existence of a bounded $\tau_{(\infty)}^{-1}$ under these circumstances.

In case (ii), since the complex numbers form a metric space with a Euclidean distance and Ω is compact in this space, Ω is sequentially compact and also has the Bolzano–Weierstrass property. Therefore:

(i) out of an infinite succession of elements, all contained in Ω, we can always extract at least one infinite partial succession which converges to a limit in Ω; and

(ii) an infinite succession of different elements, all contained in Ω, contains at least one accumulation point.

Applied to the case under study, this implies that for each value of j for which the succession $\{b_{N-j}^{(N)}\}$ does not converge, there are at least two accumulation points and at least two convergent infinite successions. In general, denoting by $\beta_j(r)$, $r \in \mathcal{R}$ with $\mathcal{R}$ either

finite or denumerable infinite, for each different accumulation point there is at least one convergent subsuccession. Therefore, we must in this case study the functions

$$f_r(z) = 1 + \sum_{j=0}^{\infty} \beta_j(r) z^{j+1} \tag{92}$$

and investigate whether or not the conditions of theorems 1.A and 1.B hold, which will again establish the conditions for the existence of $\tau_{(\infty)}^{-1}$.

Finally, we consider case 3. Then there is no compact domain in the complex plane which contains all elements of the succession $\{b_{N-j}^{(N)}\}$ and the matrix $\tau_{(\infty)}$ has no bounded inverse.

4. Conclusion

We have established the general conditions for the existence of a bounded inverse of a given infinite matrix. Cases 2 and 3 have been studied for the sake of formal completeness. In practice, case 1 is the usual situation in problems of physical interest. In particular, in the problem which motivated this research—the dielectric response of a confined electron gas—all β_j vanish and the existence of $\tau_{(\infty)}^{-1}$ follows immediately.

In summary, the algorithm presented here serves: (a) to study the limit $N \to \infty$; and (b) to provide a practical method of calculation to generate successive approximations which converge to $\tau_{(\infty)}^{-1}$ as $N \to \infty$. In practice, the steps of the calculation are as follows.

(i) Given the initial matrix $\mathbf{a}$ to be inverted, this is factorized as in (11). For each finite N, the elements p_{ij} and t_{ij} are given by (51), (53) and (54).

(ii) The inversion of $\mathbf{a}$ is then reduced to the inversion of $\mathbf{T}$.

(iii) The matrix $\mathbf{T}$ is factorized as in (55). The inverse of $\mathbf{T}$ is then $\mathbf{d}^{-1} \cdot \tau^{-1}$.

(iv) Since $\mathbf{d}$ is diagonal, its inversion is immediate and the inverse of τ is given by (67).

(v) Thus in a direct non-recurrent formula:

$$\mathbf{a}^{-1} = \mathbf{d}^{-1} \cdot \tau^{-1} \cdot \mathbf{P}^{\dagger}. \tag{93}$$

(vi) Having proved the existence of $\mathbf{a}^{-1}$, this process can be carried out for increasing N up to the desired degree of accuracy.

Application of this formalism to a physical model of an inhomogeneous confined quasi-2D electron gas is currently in progress in our laboratory.

Acknowledgment

This work was partially supported by the Spanish CICYT under grant no MAT91-0738.

References

[1] Raines S 1972 *Many-Electron Theory* (Amsterdam: North-Holland)
Mahan G D 1990 *Many-Particle Physics* (New York: Plenum)
[2] Weisbuch C and Winter B 1991 *Quantum Semiconductor Structures. Fundamentals and Applications* (Boston, MA: Academic)
[3] Chico L, Jaskólski W and García-Moliner F 1992 *Phys. Scr.* 47 284
[4] Fernández-Velicia F J, Velasco V R and García-Moliner F work in progress (to be published)
[5] Carleson L 1969 *Acta Math.* **116** 135
[6] Rudin W 1987 *Real and Complex Analysis* 3rd edn (New York: McGraw-Hill)