# THE MATRIX EQUATION $X + A^T X^{-1} A = Q$ AND ITS APPLICATION IN NANO RESEARCH*

CHUN-HUA GUO† AND WEN-WEI LIN‡

**Abstract.** The matrix equation $X + A^T X^{-1} A = Q$ has been studied extensively when $A$ and $Q$ are real square matrices and $Q$ is symmetric positive definite. The equation has positive definite solutions under suitable conditions, and in that case the solution of interest is the maximal positive definite solution. The same matrix equation plays an important role in Green's function calculations in nano research, but the matrix $Q$ there is usually indefinite (so the matrix equation has no positive definite solutions) and one is interested in the case where the matrix equation has no positive definite solutions even when $Q$ is positive definite. The solution of interest in this nano application is a special weakly stabilizing complex symmetric solution. In this paper we show how a doubling algorithm can be used to find good approximations to the desired solution efficiently and reliably.

**Key words.** nonlinear matrix equation, complex symmetric solution, stable solution, fixed-point iteration, doubling algorithm, Newton's method, Green's function

**AMS subject classifications.** 15A24, 65F30, 65H10

**1. Introduction.** The non-equilibrium Green's function formalism provides a powerful conceptual and computational framework for treating quantum transport in nanodevices [6]. Typically the system Hamiltonian $H$ is a bi-infinite or semi-infinite block tridiagonal real symmetric matrix [1, 6, 15, 16, 21]. The semi-infinite case is slightly easier to handle. So we assume $H$ is bi-infinite, and in that case $H$ is usually of the form

$$H = \begin{bmatrix} H_L & H_{L,S} & \\ H_{L,S}^T & H_S & H_{S,R} \\ & H_{S,R}^T & H_R \end{bmatrix},$$

where $H_S$ is the Hamiltonian for the scattering region, which is an $n_s \times n_s$ symmetric matrix, $H_L$ is the Hamiltonian for the left lead, given by

$$H_L = \begin{bmatrix} \ddots & \ddots & & \\ \ddots & B_L & A_L & \\ & A_L^T & B_L & A_L \\ & & A_L^T & B_L \end{bmatrix},$$

*Version of May 25, 2010.

†Department of Mathematics and Statistics, University of Regina, Regina, SK S4S 0A2, Canada (chguo@math.uregina.ca). The work of this author was supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada and by the National Center for Theoretical Sciences in Taiwan.

‡Department of Applied Mathematics, National Chiao Tung University, Hsinchu 300, Taiwan (wwlin@math.nctu.edu.tw). The work of this author was partially supported by the National Science Council and the National Center for Theoretical Sciences in Taiwan.

where all matrix blocks are $n_l \times n_l$, $A_L \neq 0$ and $B_L$ is symmetric, and $H_R$ is the Hamiltonian for the right lead, given by

$$H_R = \begin{bmatrix} B_R & A_R & & \\ A_R^T & B_R & A_R & \\ & A_R^T & B_R & \ddots \\ & & \ddots & \ddots \end{bmatrix},$$

where all matrix blocks are $n_r \times n_r$, $A_R \neq 0$ and $B_R$ is symmetric. The matrix $H_{L,S}$ represents the coupling between the scattering region and the left lead and is given by

$$H_{L,S} = \begin{bmatrix} \vdots \\ 0 \\ 0 \\ C_{L,S} \end{bmatrix},$$

where $C_{L,S}$ is $n_l \times n_s$, and the matrix $H_{S,R}$ represents the coupling between the scattering region and the right lead and is given by

$$H_{S,R} = [D_{S,R}\ 0\ 0\ \ldots],$$

where $D_{S,R}$ is $n_s \times n_r$.

The Green's function $G$ is defined [5, 7] by

$$G = ((\mathcal{E} + i0^+)I - H)^{-1} = \lim_{\eta \to 0^+} ((\mathcal{E} + i\eta)I - H)^{-1},$$

where $\mathcal{E}$ is energy, a real number that may be negative, and $I$ is the identity operator. We will also use $I_r$ (or simply $I$) to denote the identity matrix of dimension $r$. Since $H$ is real symmetric, it is easily seen [5, 7] that $G = G_1 + iG_2$ with $G_1$ and $G_2$ real symmetric (so $G$ is complex symmetric). In practice, one is only interested in $G_S$, the Green's function corresponding to the scattering region. It is easily found [5, 15] that

$$\begin{aligned} G_S &= \left((\mathcal{E} + i0^+)I - H_S - H_{L,S}^T G_L H_{L,S} - H_{S,R} G_R H_{S,R}^T\right)^{-1} \\ &= \left((\mathcal{E} + i0^+)I - H_S - C_{L,S}^T G_{L,S} C_{L,S} - D_{S,R} G_{S,R} D_{S,R}^T\right)^{-1}, \end{aligned}$$

where $G_L = ((\mathcal{E} + i0^+)I - H_L)^{-1}$, $G_R = ((\mathcal{E} + i0^+)I - H_R)^{-1}$, $G_{L,S}$ is the $n_l \times n_l$ matrix in the lower-right corner of $G_L$, and $G_{S,R}$ is the $n_r \times n_r$ matrix in the upper-left corner of $G_R$. Moreover, one is only interested in the values of $\mathcal{E}$ for which $G_{L,S}$ and $G_{S,R}$, respectively, have nonzero imaginary parts [16]. Computing $G_{L,S}$ and $G_{S,R}$ has been a challenging problem for nano-scientists.

It is easily seen [15] that $G_{L,S}$ satisfies the matrix equation

$$G_{L,S} = ((\mathcal{E} + i0^+)I - B_L - A_L^T G_{L,S} A_L)^{-1},$$

and $G_{S,R}$ satisfies the matrix equation

$$G_{S,R} = ((\mathcal{E} + i0^+)I - B_R - A_R G_{S,R} A_R^T)^{-1}.$$

For each fixed $\mathcal{E}$, we take $\eta > 0$ sufficiently small and compute $G_{L,S}(\eta)$, the $n_l \times n_l$ matrix in the lower-right corner of $G_L(\eta) = ((\mathcal{E} + i\eta)I - H_L)^{-1}$. In [1, 15, 21], the

values of $\mathcal{E}$ are between $-5$ and 5 and the smallest $\eta$ used is $10^{-6}$. Now $G_{L,S}(\eta)$ satisfies the matrix equation

$$X = ((\mathcal{E} + i\eta)I - B_L - A_L^T X A_L)^{-1}, \tag{1.1}$$

and $G_{L,S}(\eta)$ is taken to be an approximation to $G_{L,S}$. Similarly, $G_{S,R}$ is approximated by $G_{S,R}(\eta)$, which is the $n_r \times n_r$ matrix in the upper-left corner of $G_R(\eta) = ((\mathcal{E} + i\eta)I - H_R)^{-1}$ and is a particular solution of

$$X = ((\mathcal{E} + i\eta)I - B_R - A_R X A_R^T)^{-1}. \tag{1.2}$$

Since (1.1) and (1.2) have the same type, we only need to study (1.1).

A basic numerical method for finding $G_{L,S}(\eta)$ is the fixed-point iteration [21]

$$X^{(k+1)} = ((\mathcal{E} + i\eta)I - B_L - A_L^T X^{(k)} A_L)^{-1} \tag{1.3}$$

with $X^{(0)} = ((\mathcal{E} + i\eta)I - B_L)^{-1}$. It has been observed that the iteration converges in practice and the limit is the required matrix $G_{L,S}(\eta)$ (rather than a different solution of (1.1)). The convergence of the iteration has been observed to be very slow for $\eta$ close to 0. Note that we cannot take $\eta = 0$ for this iteration. Since otherwise the sequence $X^{(k)}$ (even if well defined) would be real, and would not approximate $G_{L,S}$, which is to have a nonzero imaginary part. To speed up the convergence of (1.3) the following strategy is suggested in [21]: once $X^{(k+1)}$ has been computed from $X^{(k)}$, replace $X^{(k+1)}$ by $(X^{(k+1)} + X^{(k)})/2$ and proceed with the iteration. Since $G_{L,S}(\eta)$ needs to be computed for many different values of $\mathcal{E}$, it is also suggested in [1, 21] that the $G_{L,S}(\eta)$ computed at a given energy be used as the initial guess for $G_{L,S}(\eta)$ at the next nearby energy point.

Other methods have also been considered. In [15] the equation (1.1) is rewritten as

$$A_L^T X A_L X - ((\mathcal{E} + i\eta)I - B_L)X + I = 0 \tag{1.4}$$

and Newton's method with exact line searches (as in [14]) is used, possibly preceded by the fixed-point iteration (1.3). In [15] $\eta = 0^+$ in theory, but $\eta$ is taken to be $2\times10^{-6}$ in actual computations. Whether this Newton method will always converge to the solution $G_{L,S}(\eta)$ is left as an open problem in [15]. Another approach is suggested in [16]. In that approach one would obtain the equation

$$A_L^T Y^2 - ((\mathcal{E} + i\eta)I - B_L)Y + A_L = 0 \tag{1.5}$$

by post-multiplying (1.4) by $A_L$, where $Y = XA_L$. The equations (1.4) and (1.5) are equivalent when $A_L$ is nonsingular. In [16] $\eta = 0$ is assumed. If the *right* solution $Y$ of (1.5) is found, then $G_{L,S} = YA_L^{-1}$ is obtained. To find $Y$ the auxiliary matrix

$$\begin{bmatrix} 0 & I \\ -(A_L^T)^{-1}A_L & (A_L^T)^{-1}(\mathcal{E}I - B_L) \end{bmatrix} \tag{1.6}$$

is used. Assuming this matrix is diagonalizable, the required solution $Y$ can be found (see [13]) by selecting $n_l$ linearly independent eigenvectors of (1.6). However, there is no explanation in [16] how these $n_l$ eigenvectors can be selected from the $2n_l$ eigenvectors. Moreover, this approach does not work when $A_L$ is singular. Indeed, for a simple example given in [16], $A_L$ is singular but the solution $G_{L,S}$ can be found analytically.

In this paper we show how a doubling algorithm can find the desired solution $G_{L,S}(\eta)$ of (1.1) and the desired solution $G_{S,R}(\eta)$ of (1.2). Our numerical experiments indicate the efficiency and reliability of the doubling algorithm. The algorithm works very well even when $\eta$ is taken to be very close to 0.

Our starting point is to rewrite (1.1) as

$$X + A^T X^{-1} A = Q, \tag{1.7}$$

where

$$A = A_L, \quad Q = Q_L + i\eta I, \quad Q_L = \mathcal{E} I - B_L, \tag{1.8}$$

and the required solution is $X = (G_{L,S}(\eta))^{-1}$. Similarly, we rewrite (1.2) as

$$X + AX^{-1} A^T = Q, \tag{1.9}$$

where

$$A = A_R, \quad Q = Q_R + i\eta I, \quad Q_R = \mathcal{E} I - B_R, \tag{1.10}$$

and the required solution is $X = (G_{S,R}(\eta))^{-1}$.

We often have $A_L = A_R$ and $B_L = B_R$ (in [15] for example). In this case, the doubling algorithm is able to compute $G_{L,S}(\eta)$ and $G_{S,R}(\eta)$ simultaneously.

**2. Characterization of $G_{L,S}(\eta)$ and $G_{S,R}(\eta)$.** The matrices $G_{L,S}(\eta)$, $G_{S,R}(\eta)$ have been uniquely defined through the inverses of semi-infinite block Toeplitz matrices. They are also particular solutions of the matrix equations (1.1) and (1.2), respectively. Those matrix equations may have many other solutions. So we need to give a characterization for $G_{L,S}(\eta)$ and $G_{S,R}(\eta)$, in terms of the matrix equations (rather than the infinite matrices). We only need to derive results for $G_{L,S}(\eta)$, analogous results for $G_{S,R}(\eta)$ follow readily.

We have rewritten (1.1) as (1.7), by replacing $X$ in (1.1) with $X^{-1}$. When $Q$ is a real symmetric positive definite matrix (which is possible here only when $\eta = 0$), the matrix equation (1.7) has been studied extensively (see [3, 8, 10, 11, 12, 17, 20, 22]). In all those papers the desired solution is the maximal positive definite solution (when it has at least one positive definite solution). As we will explain later, in the nano application we are only interested in values of energy $\mathcal{E}$ for which the equation (1.7) with $Q = \mathcal{E} I - B_L$ has no positive definite solutions.

When $A$ is nonsingular, with $Y = X^{-1}A$ the equation (1.7) is equivalent to the quadratic matrix equation

$$A^T Y^2 - QY + A = 0,$$

for which we have the associated quadratic eigenvalue problem

$$P(\lambda)x = 0, \quad x \neq 0,$$

where $P(\lambda) = \lambda^2 A^T - \lambda Q + A$.

We will not use this connection to solve the equation (1.7) since the matrix $A$ may be singular in the application we have in mind. However, we will use the quadratic matrix polynomial $P(\lambda)$ in our discussions even when $A$ is singular.

THEOREM 2.1. *Let $A$ and $Q$ be as in* (1.8) *and $\mathbb{T}$ be the unit circle. Then the quadratic $P(\lambda) = \lambda^2 A^T - \lambda Q + A$ has no eigenvalues on $\mathbb{T}$ for any $\eta \neq 0$. In this*

*case it has exactly $n_l$ eigenvalues inside $\mathbb{T}$ (counting multiplicities) and exactly $n_l$ eigenvalues outside $\mathbb{T}$ (counting multiplicities and including eigenvalues at $\infty$).*

*Proof.* The quadratic $P(\lambda)$ has eigenvalues on $\mathbb{T}$ if and only if $\det P(\lambda) = 0$ for some $\lambda \in \mathbb{T}$, or equivalently $\det(\lambda A^T + \lambda^{-1} A - Q_L - i\eta I) = 0$ for some $\lambda \in \mathbb{T}$. However, for any $\lambda \in \mathbb{T}$ the matrix $\lambda A^T + \lambda^{-1} A - Q_L$ is Hermitian and thus has real eigenvalues $\nu_1, \nu_2, \ldots, \nu_{n_l}$. The eigenvalues of $\lambda A^T + \lambda^{-1} A - Q_L - i\eta I$ are then the nonzero numbers $\nu_1 - \eta i, \nu_2 - \eta i, \ldots, \nu_{n_l} - \eta i$. So $\det(\lambda A^T + \lambda^{-1} A - Q_L - i\eta I) \neq 0$ for any $\lambda \in \mathbb{T}$. Thus the quadratic $P(\lambda)$ has no eigenvalues on $\mathbb{T}$. The location of the eigenvalues must be as stated since $P(\lambda)$ is a $T$-palindromic matrix polynomial (see [18]). More precisely, the number $\xi$ is an eigenvalue of the quadratic $P(\lambda)$ if and only if $\xi^{-1}$ is so, and they have the same partial multiplicities (see [18, Theorem 2.2]). ∎

Let $X$ be any complex symmetric solution of (1.7) and $\rho(\cdot)$ denote the spectral radius. Then $\rho(X^{-1}A) = \rho((X^{-1}A)^T) = \rho(A^T X^{-1})$. We adopt the following definition.

DEFINITION 2.2. *A complex symmetric solution $X$ of* (1.7) *is said to be stabilizing if $\rho(X^{-1}A) < 1$, and weakly stabilizing if $\rho(X^{-1}A) = 1$.*

Let

$$M_0 = \begin{bmatrix} A & 0 \\ Q & -I \end{bmatrix}, \quad L_0 = \begin{bmatrix} 0 & I \\ A^T & 0 \end{bmatrix}. \tag{2.1}$$

Then the pencil $M_0 - \lambda L_0$ (also denoted by $(M_0, L_0)$) is a linearization of the $T$-palindromic polynomial $\lambda^2 A^T - \lambda Q + A$, and $X$ is a solution of (1.7) if and only if

$$M_0 \begin{bmatrix} I \\ X \end{bmatrix} = L_0 \begin{bmatrix} I \\ X \end{bmatrix} X^{-1} A. \tag{2.2}$$

It follows that each complex symmetric solution of (1.7) is determined by a suitable deflating subspace of the pencil $M_0 - \lambda L_0$. The required matrix $G_{L,S} = \lim_{\eta \to 0^+} G_{L,S}(\eta)$ is the inverse of a special complex symmetric solution of (1.7) with $\eta = 0$, and we are only interested in the values of $\mathcal{E}$ for which $G_{L,S}$ has a nonzero imaginary part. We will see later in this section that for those values of $\mathcal{E}$ the pencil $M_0 - \lambda L_0$ (or equivalently the quadratic $\lambda^2 A^T - \lambda Q + A$) must have eigenvalues on $\mathbb{T}$ and that $G_{L,S}$ is the inverse of one of the weakly stable solutions of (1.7) with $\eta = 0$. For $\eta > 0$, the pencil $M_0 - \lambda L_0$ has no eigenvalues on $\mathbb{T}$ by Theorem 2.1. This is a necessary condition for the existence of a (necessarily unique) stabilizing solution of (1.7). By using a deep result on linear operators, we will show that the equation (1.7) with $\eta > 0$ always has a stabilizing solution and its inverse is precisely the matrix $G_{L,S}(\eta)$ we are looking for.

Recall that $G_{L,S}(\eta)$ is the $n_l \times n_l$ matrix in the lower-right corner of $((\mathcal{E} + i\eta)I - H_L)^{-1}$. Note that

$$(\mathcal{E} + i\eta)I - H_L = \begin{bmatrix} \ddots & \ddots & & \\ \ddots & Q & -A & \\ & -A^T & Q & -A \\ & & -A^T & Q \end{bmatrix}.$$

It is easy to see that $G_{L,S}(\eta)$ is also the $n_l \times n_l$ matrix in the upper-left corner of $T^{-1}$

with

$$T = \begin{bmatrix} Q & -A^T & & \\ -A & Q & -A^T & \\ & -A & Q & \ddots \\ & & \ddots & \ddots \end{bmatrix}. \tag{2.3}$$

This change is largely notational and is not needed if we discuss $G_{S,R}(\eta)$. The symbol for the block Toeplitz matrix $T$ is $\phi(\lambda) = -\lambda A + Q - \lambda^{-1}A^T$.

Let $\ell_2$ be the usual Hilbert space of all square summable sequence of complex numbers, and let $\ell_2^m$ be the Cartesian product of $m$ copies of $\ell_2$. The infinite matrix (2.3) is then seen to be in $\mathcal{B}(\ell_2^{n_l})$, the set of all bounded linear operators on $\ell_2^{n_l}$. It is clear that $T = (\eta i)I - W$ with $W$ a self-adjoint operator in $\mathcal{B}(\ell_2^{n_l})$. It is well known that the spectrum of a self-adjoint operator is real. So, for each $\eta > 0$, $T$ has an inverse in $\mathcal{B}(\ell_2^{n_l})$ since $\eta i$ is not in the spectrum of $W$. Thus, by appealing to a deep result on linear operators (see [9, Chapter XXIV, Theorem 4.1] and [19]) we know that $\phi(\lambda)$ has a factorization

$$\phi(\lambda) = (I - \lambda^{-1}L)D(I - \lambda U) \tag{2.4}$$

with $D$ invertible, $\rho(L) < 1$ and $\rho(U) < 1$. From (2.4) we see that

$$A = DU, \quad A^T = LD, \quad Q = D + LDU.$$

Thus $D + A^T D^{-1} A = Q$ and $\rho(D^{-1}A) < 1$. In other words, $D$ is the unique stabilizing solution of the equation (1.7), which must be complex symmetric. By [9, Chapter XXIV, Theorem 4.1] the $n_l \times n_l$ matrix in the upper-left corner of $T^{-1}$ is precisely $D^{-1}$. We thus have the following characterization of $G_{L,S}(\eta)$ and $G_{S,R}(\eta)$.

THEOREM 2.3. *The matrix $G_{L,S}(\eta)$ is the inverse of the unique stabilizing solution of* (1.7)*, and the matrix $G_{S,R}(\eta)$ is the inverse of the unique stabilizing solution of* (1.9)*.*

We will show in the next section that a doubling algorithm can compute $G_{L,S}(\eta)$ and $G_{S,R}(\eta)$ efficiently, even if $\eta$ is very close to 0. In the remainder of this section we answer the following question: for what values of $\mathcal{E}$ will $G_{L,S} = \lim_{\eta \to 0^+} G_{L,S}(\eta)$ have a nonzero imaginary part?

In the limiting case $\eta = 0$, the matrix $Q$ in (1.7) is real symmetric. For real symmetric (and more generally Hermitian) matrices $X$ and $Y$, we write $X \geq Y (X > Y)$ if $X - Y$ is positive semidefinite (definite). When $Q$ is real symmetric positive definite, necessary and sufficient conditions for the existence of a positive definite solution of (1.7) have been given in [8].

THEOREM 2.4. *Let $Q$ be positive definite. Then the equation* (1.7) *has a positive definite solution if and only if the rational matrix function $\psi(\lambda) = Q + \lambda A + \lambda^{-1}A^T$ is regular (i.e., the determinant of $\psi(\lambda)$ is not identically zero) and $\psi(\lambda) \geq 0$ for all $\lambda$ on $\mathbb{T}$. In this case the equation* (1.7) *has a maximal positive definite solution $X_+$ (i.e., $X_+ \geq X$ for any other positive definite solutions).*

The next result follows quite quickly from Theorem 2.4.

THEOREM 2.5. *Let $Q$ be real symmetric, and for each $\lambda$ on $\mathbb{T}$ the eigenvalues of the Hermitian matrix $Q + \lambda A + \lambda^{-1}A^T$ be $\mu_1(\lambda) \leq \mu_2(\lambda) \leq \ldots \leq \mu_n(\lambda)$. For $i = 1, 2, \ldots, n$, let $a_i = \min_{|\lambda|=1} \mu_i(\lambda)$ and $b_i = \max_{|\lambda|=1} \mu_i(\lambda)$, and let $\Delta_i = [a_i, b_i]$. Then the matrix equation*

$$X + A^T X^{-1} A = sI - Q \tag{2.5}$$

*has a positive definite solution for $s > b_n$, has a negative definite solution for $s < a_1$, and has no positive or negative definite solutions for $a_1 < s < b_n$. $\psi_s(\lambda) = Q + sI + \lambda A + \lambda^{-1} A^T$ The quadratic $P_s(\lambda) = \lambda^2 A^T - \lambda(sI - Q) + A$ has eigenvalues on $\mathbb{T}$ if and only if $s \in \cup_{1 \le i \le n} \Delta_i$.*

*Proof.* For each $i = 1, \ldots, n$, $\mu_i(\lambda)$ is a continuous function on $\mathbb{T}$. So $\mu_i(\mathbb{T}) = \Delta_i$. Since $\mu_1(\lambda) \le \mu_2(\lambda) \le \ldots \le \mu_n(\lambda)$, we have $a_1 \le a_2 \le \ldots \le a_n$ and $b_1 \le b_2 \le \ldots \le b_n$. Let $\psi_s(\lambda) = sI - Q - \lambda A - \lambda^{-1} A^T$. If $s > b_n$, then $\psi_s(\lambda) \ge sI - b_n I > 0$ for all $\lambda \in \mathbb{T}$. Note that in this case we must have $sI - Q > 0$. In fact, we have $sI - Q - (A + A^T) > 0$ for $\lambda = 1$ and $sI - Q + (A + A^T) > 0$ for $\lambda = -1$. Adding this two we get $2(sI - Q) > 0$. Now by Theorem 2.4 the equation (2.5) has a positive definite solution. If $s < a_1$, then $\widehat{\psi}_s(\lambda) = Q - sI + \lambda A + \lambda^{-1} A^T \ge a_1 I - sI > 0$. So $Q - sI > 0$ and by Theorem 2.4 the equation $X + A^T X^{-1} A = Q - sI$ has a positive definite solution $X_*$. So $-X_*$ is a negative definite solution of (2.5). $\widehat{\psi}_s(\lambda)$ (for $s = -b_n$). We now show that the equation (2.5) has no positive definite solutions for $s < b_n$. In fact, the existence would imply $\psi_s(\lambda) \ge 0$ for all $\lambda \in \mathbb{T}$ and then $\psi_{b_n}(\lambda) > 0$ for all $\lambda \in \mathbb{T}$. This is impossible since $\psi_{b_n}(\lambda)$ is singular for some $\lambda \in \mathbb{T}$. Similarly, the equation (2.5) has no negative definite solutions for $s > a_1$. The quadratic $P_s(\lambda) = \lambda^2 A^T - \lambda(sI - Q) + A$ has eigenvalues on $\mathbb{T}$ if and only if $\det P_s(\lambda) = 0$ for some $\lambda \in \mathbb{T}$, or equivalently $\det \psi_s(\lambda) = (s - \mu_1(\lambda))(s - \mu_2(\lambda)) \cdots (s - \mu_n(\lambda)) = 0$ for some $\lambda \in \mathbb{T}$, the latter is equivalent to $s \in \cup_{1 \le i \le n} \Delta_i$. □

We use one simple example to illustrate the results in Theorem 2.5.

*Example* 2.1. Let $n = 2$, and

$$A = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad Q = \begin{bmatrix} t+1 & t \\ t & t+1 \end{bmatrix}, \quad t \ge 0.$$

This example is a special case of an example in [16]. For $\lambda = e^{i\theta}$ we find

$$\mu_1(\lambda) = t + 1 - \sqrt{t^2 + 1 + 2t\cos\theta}, \quad \mu_2(\lambda) = t + 1 + \sqrt{t^2 + 1 + 2t\cos\theta}.$$

We then find

$$\Delta_1 = \begin{cases} [0, 2t], & 0 \le t \le 1, \\ [0, 2], & t \ge 1, \end{cases} \qquad \Delta_2 = \begin{cases} [2, 2(t+1)], & 0 \le t \le 1, \\ [2t, 2(t+1)], & t \ge 1. \end{cases}$$

By Theorem 2.5 or direct verification, the equation (2.5) has a positive definite solution for $s > 2(t + 1)$, has a negative definite solution for $s < 0$, and has no positive or negative definite solutions for $0 < s < 2(t+1)$. The quadratic $P_s(\lambda) = \lambda^2 A^T - \lambda(sI - Q) + A$ has eigenvalues on the unit circle if and only if $s \in \Delta_1 \cup \Delta_2$. In particular, when $t = 0$ the quadratic $P_s(\lambda)$ has eigenvalues on the unit circle only for $s = 0, 2$; when $t = 1$ the quadratic $P_s(\lambda)$ has eigenvalues on the unit circle for all $s \in [0, 4]$. For this example, the value $t = 1$ is the only one for which $\Delta_1 \cup \Delta_2$ is connected.

We can now give the following result about when $G_{L,S}$ has a nonzero imaginary part. A similar result can be given for $G_{S,R}$.

THEOREM 2.6. *For $\lambda \in \mathbb{T}$, let the eigenvalues of $\psi_L(\lambda) = B_L + \lambda A_L + \lambda^{-1} A_L^T$ be $\mu_{L,1}(\lambda) \le \ldots \le \mu_{L,n}(\lambda)$, where $n = n_l$. Let*

$$\Delta_{L,i} = \left[ \min_{|\lambda|=1} \mu_{L,i}(\lambda), \ \max_{|\lambda|=1} \mu_{L,i}(\lambda) \right],$$

*and $\Delta_L = \cup_{i=1}^n \Delta_{L,i}$. Then $G_{L,S}$ is a real symmetric matrix if $\mathcal{E} \notin \Delta_L$, and is a complex symmetric matrix if $\mathcal{E} \in \Delta_L$. When $\mathcal{E} \in \Delta_L$, the quadratic $P_L(\lambda) =$*

*$\lambda^2 A_L^T - \lambda(\mathcal{E}I - B_L) + A_L$ has eigenvalues on $\mathbb{T}$. In the generic case that all these eigenvalues on $\mathbb{T}$ are simple and nonreal, $G_{L,S}$ has a nonzero imaginary part.*

*Proof.* By taking $Q = B_L, A = A_L$ and $s = \mathcal{E}$ in Theorem 2.5, we know that $P_L(\lambda)$ has eigenvalues on $\mathbb{T}$ if and only if $\mathcal{E} \in \Delta_L$. The matrix $G_{L,S}$ is known to be complex symmetric. If $\mathcal{E} \notin \Delta_L$, then we know by Theorem 2.3 that $G_{L,S}$ is determined by the deflating subspace of $M_0 - \lambda L_0$ corresponding to the $n$ eigenvalues inside $\mathbb{T}$, where $M_0$ and $L_0$ are given in (2.1) with $\eta = 0$. So $G_{L,S}$ is a real matrix since $A_L$ and $B_L$ are real. We now assume that $\mathcal{E} \in \Delta_L$ and that the eigenvalues on $\mathbb{T}$ are simple eigenvalues $\alpha_k \pm \beta_k i$ $(k = 1, \ldots, m)$, with $\beta_k > 0$. In this case, for each $\eta > 0$, $P_{L,\eta}(\lambda) = \lambda^2 A_L^T - \lambda((\mathcal{E} + i\eta)I - B_L) + A_L$ has no eigenvalues on $\mathbb{T}$ (see Theorem 2.1). After the introduction of $i\eta$, one of the pair $\alpha_k \pm \beta_k i$ is moved to the inside of $\mathbb{T}$ (the choice is independent of the size of $\eta > 0$ by the continuity of eigenvalues) and the other is moved to the outside of $\mathbb{T}$ (by the property of $T$-palindromic matrix polynomials). Therefore, $G_{L,S}$ is determined by the deflating subspace of $M_0 - \lambda L_0$ (with $\eta = 0$) corresponding to the $n - m$ eigenvalues inside $\mathbb{T}$ together with the $m$ eigenvalues on $\mathbb{T}$ that would be moved to the inside of $\mathbb{T}$ with the introduction of $i\eta$. These $n$ eigenvalues of $M_0 - \lambda L_0$ must be those of $G_{L,S}A$ by (2.2). In this case $G_{L,S}$ must have a nonzero imaginary part since otherwise the eigenvalues of $G_{L,S}A$ would appear in conjugate pairs. ∎

We remark that the numerical determination of $\Delta_L$ in the above theorem may require the computation of the eigenvalues of $\psi_L(\lambda)$ for many fixed values of $\lambda$ on $\mathbb{T}$. However, for each fixed $\lambda$ on $\mathbb{T}$, $\psi_L(\lambda)$ is just a Hermitian matrix. The computation of all its eigenvalues is a relatively easy task. We don't have to do the computation for too many points of $\lambda$ on $\mathbb{T}$. A rough numerical approximation of $\Delta_L$ would be sufficient. In this way, we can avoid the much more complicated computation of $G_{L,S}$ for many energy values $\mathcal{E}$ of no practical interest. Moreover, in many cases $\Delta_{L,i}$ and $\Delta_{L,i+1}$ overlap for $i = 1, \ldots, n-1$, so an accurate approximation of $\Delta_L$ only requires the computation of the extreme eigenvalues $\mu_{L,1}(\lambda)$ and $\mu_{L,n}(\lambda)$ for many $\lambda$ values on $\mathbb{T}$.

**3. Computation of $G_{L,S}(\eta)$ and $G_{S,R}(\eta)$.** We know that $G_{L,S}(\eta) = X_s^{-1}$, where $\eta > 0$ and $X_s$ is the unique stabilizing solution of (1.7). A doubling algorithm has been studied in [17] for the equation (1.7) with a real symmetric positive definite $Q$. In our case, $Q$ is complex symmetric. However, the more general presentation in [3, 4] can be used directly.

Let $M_0$ and $L_0$ be as given in (2.1). It is easy to verify that the pencil $M_0 - \lambda L_0$ is $T$-symplectic, i.e.,

$$M_0 J M_0^T = L_0 J L_0^T \quad \text{for} \quad J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}.$$

We can define the sequences $\{M_k\}$ and $\{L_k\}$, where

$$M_k = \begin{bmatrix} A_k & 0 \\ Q_k & -I \end{bmatrix}, \quad L_k = \begin{bmatrix} -P_k & I \\ A_k^T & 0 \end{bmatrix}, \tag{3.1}$$

by the following doubling algorithm [3, 4] if no breakdown occurs.

ALGORITHM 3.1. *Let $A_0 = A, Q_0 = Q, P_0 = 0$.*

*For $k = 0, 1, \ldots$, compute*

$$\begin{aligned} A_{k+1} &= A_k(Q_k - P_k)^{-1} A_k, \\ Q_{k+1} &= Q_k - A_k^T (Q_k - P_k)^{-1} A_k, \\ P_{k+1} &= P_k + A_k (Q_k - P_k)^{-1} A_k^T. \end{aligned}$$

We will show shortly that this algorithm will not break down, and $Q_k$ converges to $X_s$ much more quickly than the sequence $\{X_k\}$ generated by the following basic fixed-point iteration.

ALGORITHM 3.2.

$$\begin{aligned} X_0 &= Q, \\ X_{k+1} &= Q - A^T X_k^{-1} A, \quad k = 0, 1, \ldots. \end{aligned}$$

We will also show that the sequence $\{X_k\}$ is well-defined and converges to $X_s$. Note that the sequence $\{X^{(k)}\}$ from (1.3) is then well-defined and given by $X^{(k)} = X_k^{-1}$. So we indeed have $\lim X^{(k)} = G_{L,S}(\eta)$, as observed in numerical experiments.

THEOREM 3.1. *Let $A$ and $Q$ be as in* (1.8) *with $\eta > 0$. Let $X_s$ be the stabilizing solution of* (1.7) *and $\widehat{X}_s$ be the stabilizing solution of the dual equation*

$$X + AX^{-1}A^T = Q$$

(*The existence of $\widehat{X}_s$ is also guaranteed by the argument leading to Theorem* 2.3). *Then*

(a) *The sequences $\{A_k\}, \{Q_k\}, \{P_k\}$ in Algorithm* 3.1 *are well-defined, and $Q_k$ and $P_k$ are complex symmetric.*

(b) *The sequence $\{X_k\}$ in Algorithm* 3.2 *is well-defined and $X_k$ is complex symmetric.*

(c) *$Q_k = X_{2^k - 1}$ for each $k \geq 0$.*

(d) *$Q_k$ converges to $X_s$ quadratically, $A_k$ converges to 0 quadratically, $Q - P_k$ converges to $\widehat{X}_s$ quadratically, with*

$$\limsup_{k\to\infty} \sqrt[2^k]{\|Q_k - X_s\|} \leq (\rho(X_s^{-1}A))^2, \quad \limsup_{k\to\infty} \sqrt[2^k]{\|A_k\|} \leq \rho(X_s^{-1}A),$$

$$\limsup_{k\to\infty} \sqrt[2^k]{\|Q - P_k - \widehat{X}_s\|} \leq (\rho(X_s^{-1}A))^2,$$

*where $\|\cdot\|$ is any matrix norm.*

(e) *$X_k$ converges to $X_s$ linearly with*

$$\limsup_{k\to\infty} \sqrt[k]{\|X_k - X_s\|} \leq (\rho(X_s^{-1}A))^2.$$

*Proof.* Let $T_k$ be the block $k \times k$ matrix given by

$$T_k = \begin{bmatrix} Q & -A^T & & \\ -A & Q & \ddots & \\ & \ddots & \ddots & -A^T \\ & & -A & Q \end{bmatrix}.$$

For each $k \geq 1$ we can write $T_k = V_k + i\eta I$ with $V_k$ real symmetric. So $T_k$ has no zero eigenvalues and is thus invertible. The sequence $\{X_k\}$ is obtained by block Gaussian elimination performed on the matrix (2.3). In fact, $X_0 = Q$ is the $(1,1)$ block in (2.3); when the $(1,1)$ block is used to eliminate the $(2,1)$ block, the new $(2,2)$ block is $X_1$; when the new $(2,2)$ block is used to eliminate the $(3,2)$ block, the new $(3,3)$ block is $X_2$; and so on. Since $T_k$ is invertible for each $k \geq 1$, $X_k$ is well-defined and invertible for each $k \geq 0$. It is easily seen by induction that $X_k$ are complex symmetric since $Q$ is complex symmetric. So (b) is proved.

Let $W_k = Q_k - P_k$ in Algorithm 3.1. Then the sequence $\{W_k\}$ satisfies

$$W_{k+1} = W_k - A_k^T W_k^{-1} A_k - A_k W_k^{-1} A_k^T, \quad W_0 = Q.$$

It follows from [2, Theorem 13] that $W_k$ is nonsingular for each $k \geq 0$. The sequences $\{A_k\}, \{Q_k\}, \{P_k\}$ are then well-defined. Again, $Q_k$ and $P_k$ are complex symmetric since $Q$ is complex symmetric. This proves (a).

The proof of (c) is the same as in the proof of [11, Proposition 5], although $Q$ is complex symmetric here.

To prove (d), we start with

$$M_0 \begin{bmatrix} I \\ X_s \end{bmatrix} = L_0 \begin{bmatrix} I \\ X_s \end{bmatrix} X_s^{-1} A,$$

a special case of (2.2). It follows from [3] that for each $k \geq 0$

$$M_k \begin{bmatrix} I \\ X_s \end{bmatrix} = L_k \begin{bmatrix} I \\ X_s \end{bmatrix} (X_s^{-1} A)^{2^k}. \tag{3.2}$$

Substituting (3.1) into (3.2) yields

$$A_k = (X_s - P_k)(X_s^{-1} A)^{2^k}, \quad Q_k - X_s = A_k^T (X_s^{-1} A)^{2^k}. \tag{3.3}$$

Similarly we have

$$\widehat{M}_0 \begin{bmatrix} I \\ \widehat{X}_s \end{bmatrix} = \widehat{L}_0 \begin{bmatrix} I \\ \widehat{X}_s \end{bmatrix} \widehat{X}_s^{-1} A^T,$$

where

$$\widehat{M}_0 = \begin{bmatrix} A^T & 0 \\ Q & -I \end{bmatrix}, \quad \widehat{L}_0 = \begin{bmatrix} 0 & I \\ A & 0 \end{bmatrix}.$$

The pencil $\widehat{M}_0 - \lambda \widehat{L}_0$ is a linearization of $\lambda^2 A - \lambda Q + A^T$, which has the same eigenvalues as $\lambda^2 A^T - \lambda Q + A$. It follows that $\widehat{X}_s^{-1} A^T$ and $X_s^{-1} A$ have the same eigenvalues, and in particular $\rho(\widehat{X}_s^{-1} A^T) = \rho(X_s^{-1} A)$. For each $k \geq 0$ we now have

$$\widehat{M}_k \begin{bmatrix} I \\ \widehat{X}_s \end{bmatrix} = \widehat{L}_k \begin{bmatrix} I \\ \widehat{X}_s \end{bmatrix} (\widehat{X}_s^{-1} A^T)^{2^k},$$

where $\widehat{M}_k$ and $\widehat{L}_k$ are given by (3.1) and Algorithm 3.1 *when* $A_0 = A$ in Algorithm 3.1 is replaced by $A_0 = A^T$. It is easy to see that

$$\widehat{M}_k = \begin{bmatrix} A_k^T & 0 \\ \widehat{Q}_k & -I \end{bmatrix}, \quad \widehat{L}_k = \begin{bmatrix} -\widehat{P}_k & I \\ A_k & 0 \end{bmatrix}$$

with

$$\widehat{P}_k = Q - Q_k, \quad \widehat{Q}_k = Q - P_k. \tag{3.4}$$

So we now have

$$A_k^T = (\widehat{X}_s - \widehat{P}_k)(\widehat{X}_s^{-1} A^T)^{2^k}, \quad \widehat{Q}_k - \widehat{X}_s = A_k (\widehat{X}_s^{-1} A^T)^{2^k}. \tag{3.5}$$

By (3.3), (3.5) and (3.4), we have

$$\begin{aligned} Q_k - X_s &= A_k^T (X_s^{-1} A)^{2^k} \\ &= (\widehat{X}_s - \widehat{P}_k)(\widehat{X}_s^{-1} A^T)^{2^k} (X_s^{-1} A)^{2^k} \\ &= (Q_k - X_s + (X_s + \widehat{X}_s - Q))(\widehat{X}_s^{-1} A^T)^{2^k} (X_s^{-1} A)^{2^k}. \end{aligned}$$

Thus

$$(Q_k - X_s)(I - (\widehat{X}_s^{-1} A^T)^{2^k} (X_s^{-1} A)^{2^k}) = (X_s + \widehat{X}_s - Q)(\widehat{X}_s^{-1} A^T)^{2^k} (X_s^{-1} A)^{2^k}. \tag{3.6}$$

It follows that

$$\limsup_{k\to\infty} \sqrt[2^k]{\|Q_k - X_s\|} \le \rho(\widehat{X}_s^{-1} A^T)\rho(X_s^{-1} A) = (\rho(X_s^{-1} A))^2 < 1.$$

So $Q_k$ converges to $X_s$ quadratically. Then we know $\{\widehat{P}_k\}$ is bounded and have by the first equation in (3.5) that

$$\limsup_{k\to\infty} \sqrt[2^k]{\|A_k\|} \le \rho(X_s^{-1} A) < 1.$$

So $A_k$ converges to 0 quadratically. By the second equations in (3.5) and (3.4) we have

$$\limsup_{k\to\infty} \sqrt[2^k]{\|(Q - P_k) - \widehat{X}_s\|} \le (\rho(X_s^{-1} A))^2 < 1.$$

So $Q - P_k$ converges to $\widehat{X}_s$ quadratically. This completes the proof of (d).

Since $Q_k$ converges to $X_s$ and $X_{2^k-1} = Q_k$, we know that a subsequence of $\{X_k\}$ converges to $X_s$. Since $\rho(X_s^{-1} A) < 1$, $X_s$ is an attractive fixed point of the mapping $f$ defined by $f(X) = Q - A^T X^{-1} A$. It follows that the sequence $\{X_k\}$ also converges to $X_s$, and we have

$$\limsup_{k\to\infty} \sqrt[k]{\|X_k - X_s\|} \le (\rho(X_s^{-1} A))^2,$$

as in [12]. So (e) is proved. □

Algorithm 3.1 is said to be structure-preserving since for each $k \ge 0$ $M_k$ and $L_k$ have the structures given in (3.1) and the pencil $M_k - \lambda L_k$ is $T$-symplectic.

When $A_L = A_R$ and $B_L = B_R$, the dual equation of (1.7) with (1.8) is precisely the equation (1.9) with (1.10). In this case, Algorithm 3.1 can find $G_{L,S}(\eta)$ and $G_{S,R}(\eta)$ simultaneously.

To get good approximations to $G_{L,S}$, we need to take $\eta > 0$ to be sufficiently small. However, we cannot apply Algorithm 3.1 with $\eta = 0$ directly. If we take $\eta = 0$, then the sequence $\{Q_k\}$ is real and will not approximate $X_s$, which is to be complex symmetric.

We are only interested in the values of $\mathcal{E}$ for which $\lim_{\eta\to 0^+}\rho(X_s^{-1}A)=1$. Since we need to take $\eta$ sufficiently small, we always have $\rho(X_s^{-1}A)\approx 1$, and the convergence of Algorithm 3.2 will be very slow in general. The strategy proposed in [21] for improving the convergence of iteration (1.3) generates a new sequence $X^{(k)}$ as follows.

ALGORITHM 3.3. *Take* $X^{(0)}=Q^{-1}$. *For* $k=0,1,\ldots,$ *compute*

$$\begin{aligned} X^{(k+1)} &= (Q-A^TX^{(k)}A)^{-1},\\ X^{(k+1)} &\leftarrow (X^{(k)}+X^{(k+1)})/2. \end{aligned}$$

We now adapt this strategy for Algorithm 3.2 and get the following modified fixed-point method.

ALGORITHM 3.4. *Take* $X_0=Q$. *For* $k=0,1,\ldots,$ *compute*

$$\begin{aligned} X_{k+1} &= Q-A^TX_k^{-1}A,\\ X_{k+1} &\leftarrow (X_k+X_{k+1})/2. \end{aligned}$$

We remark that for $\{X^{(k)}\}$ from Algorithm 3.3 and $\{X_k\}$ from Algorithm 3.4 we no longer have $X^{(k)}=X_k^{-1}$ in general. To keep this relation, one would have to replace the last line of Algorithm 3.4 by

$$X_{k+1}\leftarrow \left((X_k^{-1}+X_{k+1}^{-1})/2\right)^{-1}.$$

This would make the algorithm slightly more expensive. Numerical experiments show that this change does not affect the performance of the algorithm in any significant way. So we prefer to use the update of $X_{k+1}$ as given in Algorithm 3.4. Numerical experiments also show that the convergence of Algorithm 3.4 is often much faster than that of Algorithm 3.2. But a rigorous convergence analysis remains an open problem.

In [16], $\eta=0$ is assumed and a diagonalization procedure is used on the auxiliary matrix (1.6) when $A=A_L$ is nonsingular. However, that approach will run into difficulties even when $A$ is perfectly conditioned and the auxiliary matrix is diagonalizable.

*Example* 3.1. Let

$$A_L=-I_3,\quad B_L=\begin{bmatrix} 4 & -1 & 0\\ -1 & 4 & -1\\ 0 & -1 & 4\end{bmatrix}.$$

It is easy to determine by Theorem 2.6 that

$$\Delta_{L,1}=[0.5858,4.5858],\quad \Delta_{L,2}=[2.0000,6.0000],\quad \Delta_{L,3}=[3.4142,7.4142].$$

So $\Delta_L=[0.5858,7.4142]$. For $\eta=0$ and $\mathcal{E}=4$, the matrix in (1.6) has all 6 eigenvalues on $\mathbb{T}$:

$$-\frac{\sqrt{2}}{2}\pm\frac{\sqrt{2}}{2}i,\quad \frac{\sqrt{2}}{2}\pm\frac{\sqrt{2}}{2}i,\quad \pm i.$$

The difficulty with the approach in [16] is that we do not know which 3 eigenvalues should be used to get $G_{L,S}$. There are 20 different choices. By the proof of Theorem 2.6, we now know that we can pick only one eigenvalue from each conjugate pair. This

reduces the number of choices to 6. Only by changing $\eta$ from 0 to a small positive number do we find that the 3 eigenvalues with negative imaginary parts are perturbed to the inside of $\mathbb{T}$. So these 3 eigenvalues are to be used to find $G_{L,S}$. We could have taken a small $\eta > 0$, say $\eta = 10^{-10}$, right in the beginning and use Algorithm 3.1 to compute $G_{L,S}(\eta)$ as a good approximation to $G_{L,S}$.

In general, the palindromic matrix polynomial $P(\lambda) = \lambda^2 A^T - \lambda Q + A$, where $A$ and $Q$ are given in (1.8) with $\eta = 0$, can have $2k$ eigenvalues on $\mathbb{T}$ with $k$ being any integer from 0 to $n_l$. For the above example, we find that $P(\lambda)$ has 0 eigenvalues on $\mathbb{T}$ for $\mathcal{E} \in (-\infty, 0.5857] \cup [7.4143, \infty)$, has 2 eigenvalues on $\mathbb{T}$ for $\mathcal{E} \in [0.5858, 2) \cup (6, 7.4142]$, has 4 eigenvalues on $\mathbb{T}$ for $\mathcal{E} \in [2, 3.4142] \cup [4.5858, 6]$, and has 6 eigenvalues on $\mathbb{T}$ for $\mathcal{E} \in [3.4143, 4.5857]$. For a large problem, the computed eigenvalues of $P(\lambda)$ (or the matrix (1.6)) may contain a number of conjugate pairs near $\mathbb{T}$, and the number of eigenvalues inside $\mathbb{T}$ may be different from the number of eigenvalues outside $\mathbb{T}$. In that case, it is hard to tell which eigenvalues should be used to compute $G_{L,S}$. For this reason, the eigenvalue approach will not be considered further in this paper.

Newton's method has been studied in [12] for equation (1.7) with a Hermitian positive definite $Q$. Since $Q$ is non-Hermitian in our case, there is no guarantee of convergence for Newton's method with $X_0 = Q$. With a given $X_0$, the Newton iteration for (1.7) is easily found to be

$$X_k - A^T X_{k-1}^{-1} X_k X_{k-1}^{-1} A = Q - 2A^T X_{k-1}^{-1} A. \tag{3.7}$$

If $X_{k-1}$ is a nonsingular complex symmetric matrix with $\rho(X_{k-1}^{-1} A) < 1$, then (3.7) has a unique solution $X_k$, which must be complex symmetric. Since $\rho(X_s^{-1} A) < 1$ for $\eta > 0$, the convergence of Newton's method is guaranteed if $X_0$ is complex symmetric and sufficiently close to $X_s$.

ALGORITHM 3.5 (Newton's method for (1.7)). *Take $X_0$ to be complex symmetric and sufficiently close to $X_s$. For $k = 1, 2, \ldots$, compute $L_k = X_{k-1}^{-1} A$, and solve*

$$X_k - L_k^T X_k L_k = Q - 2L_k^T A. \tag{3.8}$$

Note that the Stein equation (3.8) is uniquely solvable when $\rho(L_k) < 1$.

If we are to find $X_s$ (and then $G_{L,S}(\eta) = X_s^{-1}$) only for one fixed value of $\mathcal{E}$, there is little hope for Algorithm 3.5 to beat Algorithm 3.1. In practice, we need to determine $G_{L,S}$ (through $X_s$ and $G_{L,S}(\eta)$ for a small $\eta > 0$) for a range of energy values [21]. So for Algorithm 3.5, the $X_s$ computed at a given energy can be used as an initial guess for $X_s$ at the next nearby energy point. However, there is some danger associated with this practice. If $\eta$ is very small, then the convergence of Algorithm 3.5 is guaranteed only when $X_s$ for the previous energy is very close to the $X_s$ for the current energy. This means that the stepsize for the energy $\mathcal{E}$ must be very small. But it is hard to tell how small the stepsize should be for a given $\eta > 0$. On the other hand, Algorithm 3.1 is not a correction method, and thus we cannot use the $X_s$ computed at a given energy to compute $X_s$ at the next nearby energy point. But the convergence of $Q_k$ to $X_s$ is fast (see Theorem 3.1 and (3.6)) even though we use $Q_0 = Q$ for each energy $\mathcal{E}$.

**4. Numerical results.** In this section we present some numerical results to illustrate the convergence behavior of the algorithms for computing the stabilizing solution $X_s$ of the equation (1.7). We use DA, M_FPM and NM to denote the doubling algorithm (Algorithm 3.1), the modified fixed-point method (Algorithm 3.4) and Newton's

Method (Algorithm 3.5), respectively. All computations are performed in MATLAB R2008b using IEEE double-precision floating-point arithmetic (eps $\approx 2.2 \times 10^{-16}$) on the Linux system.

Suppose one complex flop is equivalent to four real flops and $n$ is the dimension of the matrices in (1.7). Then for each iteration DA requires about $(104/3)n^3$ real flops, M_FPM requires about $(38/3)n^3$ real flops, and NM requires about $(398/3)n^3$ real flops. We also recall that DA can compute the stabilizing solutions of (1.7) and (1.9) simultaneously if the matrices $A$ and $Q$ in these two equations are the same two matrices. This makes the comparison more favorable for DA.

To measure the accuracy of a computed stabilizing solution $X$ to (1.7) we use the relative residual (r_res)

$$\frac{\left\|X + A^T X^{-1} A - Q\right\|}{\|X\| + \|A\|^2 \|X^{-1}\| + \|Q\|},$$

where $\|\cdot\|$ is the spectral norm.

*Example* 4.1. We randomly generate a symmetric matrix $B_L$ and an arbitrary matrix $A_L$ of dimension 50. By Theorem 2.6 we find

$$\Delta_L = [-23.03, -19.94] \cup [-19.84, 15.83] \cup [16.12, 17.78] \cup [18.26, 21.28].$$

We divide the interval $[-23.03, 21.28]$ (the smallest interval containing $\Delta_L$) into $P$ subintervals using $P+1$ equally spaced nodes $\mathcal{E}_i = -23.03 + 44.31(i/P)$, $i = 0, \ldots, P$. We now choose $P = 1000$ and take $\eta = 10^{-10}$ in (1.8). Let $d_i$ be the distance between $\mathbb{T}$ and the set of eigenvalues of the pencil $(M_0, L_0)$ in (2.1), with $\mathcal{E} = \mathcal{E}_i$. We find that $d_i < 10^{-8}$ whenever $\mathcal{E}_i \in \Delta_L$.

We run DA for each $\mathcal{E}_i$ as well as for a few $\mathcal{E}$ values outside $[-23.03, 21.28]$. The algorithm is stopped when $\|Q_{k+1} - Q_k\| < 10^{-8}$ and $X = Q_{k+1}$ is taken to be a good approximation to the stabilizing solution of (1.7). We also compute the relative residual for $X$ as another check of accuracy. In Figure 4.1, we plot the relative residuals, the number of iterations and the distance between $\mathbb{T}$ and the set of eigenvalues of the pencil $(M_0, L_0)$. We see that DA converges to the stabilizing solution of (1.7) in about 40 iterations for $\mathcal{E} \in \Delta_L$, and that the convergence is very fast for other $\mathcal{E}$ values. We also see that r_res $< 10^{-10}$ for almost all $\mathcal{E}$ values. Generally speaking, there is no need to look for higher accuracy since we have taken $\eta = 10^{-10}$ rather than $\eta = 0^+$.

Note that DA computes the stabilizing solution for each $\mathcal{E}$ value without using the stabilizing solution already computed for a nearby $\mathcal{E}$ value. For this example, we have already found by DA the (approximate) stabilizing solution $X_{DA}^{(i)}$ for each $\mathcal{E} = \mathcal{E}_i$, $i = 0, \ldots, P$. We now examine whether it is possible to use NM to find all those stabilizing solutions more efficiently, with some help from DA. Let $q \geq 2$ be a factor of $P$ and $P = mq$. Suppose we only use DA to find the stabilizing solutions for $\mathcal{E} = \mathcal{E}_{rq}$, $r = 0, \ldots, m-1$. Then we may find the stabilizing solutions for $\mathcal{E} = \mathcal{E}_{rq+i}$ ($i = 1, \ldots, q_r$, with $q_r = q - 1$ for $r < m - 1$ and $q_r = q$ for $r = m - 1$) as follows. Take $X_0 = X_{DA}^{(rq)}$, apply NM until $\|X_{k+1} - X_k\| < 10^{-8}$ or $k + 1 = 30$, and take $X_{k+1}$ to be $X_{NM}^{(rq+1)}$. To get $X_{NM}^{(rq+i)}$ for $i = 2, \ldots, q_r$, we apply NM in the same way with $X_0 = X_{NM}^{(rq+i-1)}$. For $r = 0, \ldots, m-1$ and $i = 1, \ldots, q_r$, we say NM is successful if $\left\|X_{NM}^{(rq+i)} - X_{DA}^{(rq+i)}\right\| < 10^{-6}$. Suppose NM is successful $S$ times, we define the

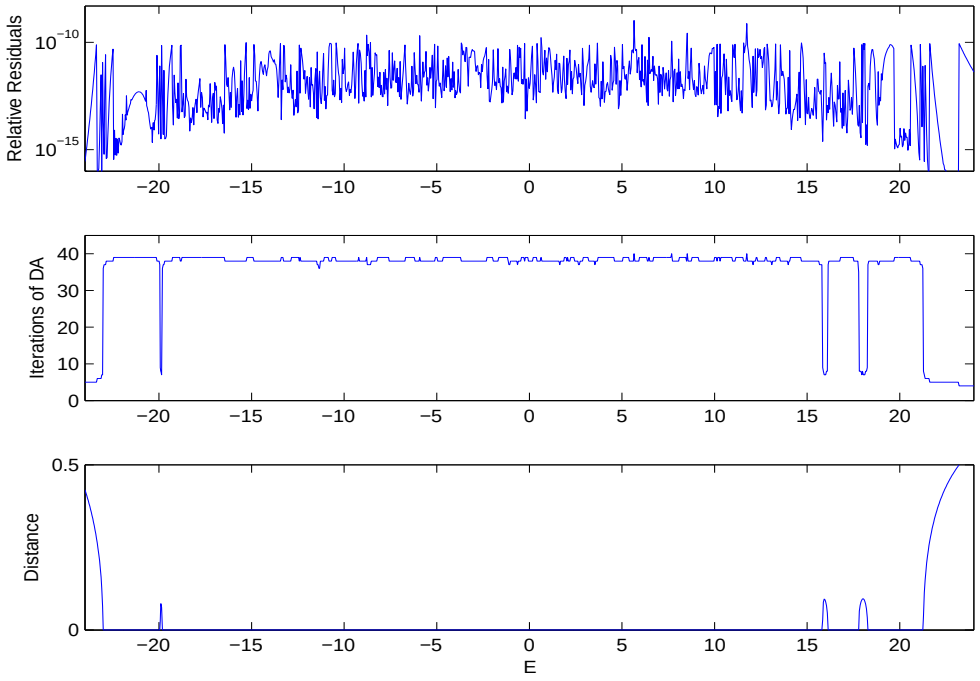


FIG. 4.1. *Relative residuals, the number of iterations for convergence of DA, and the distance between* $\mathbb{T}$ *and the spectrum of* $(M_0, L_0)$*,* $P = 1000$*.*

efficiency (or success rate) of NM by

$$\text{Eff} = \frac{S}{m(q-1)+1}. \tag{4.1}$$

For $P = 1000$ we find that NM converges (within 30 iterations) most of the time, with average number of iterations at about 10. However, it often converges to a wrong solution. In practice, $X_{DA}^{(rq+i)}$ is not computed if we choose to use NM to get $X_{NM}^{(rq+i)}$. But we could still check whether $X_{NM}^{(rq+i)}$ is approximating the stabilizing solution by computing the eigenvalues of $(X_{NM}^{(rq+i)})^{-1}A$. The efficiency of NM is shown in Table 4.1 for different values of $q$. Note that the success rate of NM is only 58%

| q | 40 | 20 | 10 | 5 | 2 |
|---|---|---|---|---|---|
| Eff | 15% | 27% | 33% | 48% | 58% |

Table 4.1. Efficiency of NM for $P = 1000$.

even for $q = 2$. Besides, the computational work for 10 NM iterations is not much less than that for 40 DA iterations. Therefore, we would prefer to use DA to compute the stabilizing solution for each $\mathcal{E}$ value. However, NM may be used to improve the accuracy of the solution obtained by DA if there is a need to do so. For example, when $\mathcal{E} = 5.64$ we find that r_res $= 2 \times 10^{-9}$ after 40 DA iterations, and starting with this approximate solution we get r_res $= 3 \times 10^{-13}$ after 1 NM iteration.

*Example* 4.2. We consider a semi-infinite Hamiltonian operator for a heterostruc-

tured semiconductor of the form

$$H(\psi,\vec{x}) = -\nabla\frac{\hbar}{2\varepsilon(\vec{x})}\nabla\psi + V(\vec{x})\psi, \quad \vec{x} \equiv \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \Omega, \tag{4.2}$$

where $\Omega \equiv \Omega_1 \cup \Omega_2$ with

$$\begin{cases} \Omega_1 = ([-9,-1]\times(-\infty,0]) \cup ([1,9]\times(-\infty,0])\,, \\ \Omega_2 = [-1,1]\times(-\infty,0], \end{cases}$$

$\hbar$ is the reduced Planck constant, $\psi$ is the associated wave function, $\varepsilon(\vec{x})$ is the electron effective mass with

$$\varepsilon(\vec{x}) = \begin{cases} \varepsilon_1, & \vec{x} \in \Omega_1, \\ \varepsilon_2, & \vec{x} \in \Omega_2, \end{cases}$$

and $V(\vec{x}) = \omega_1 x_1^2$ is the potential energy.

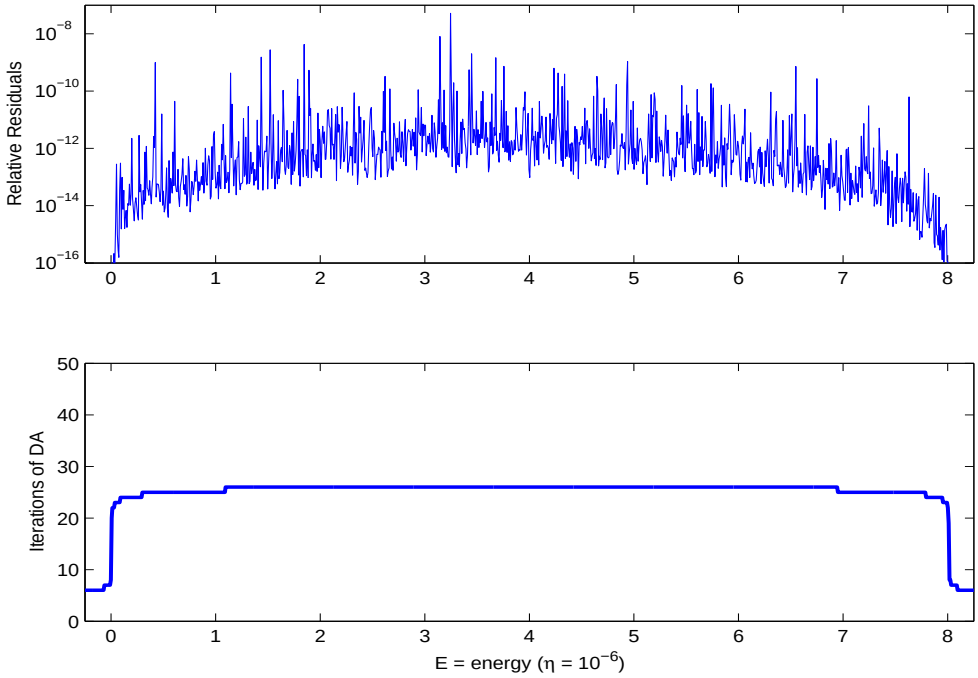


FIG. 4.2. *Relative residuals and the number of iterations for convergence of DA, $P = 1000$.*

Let $T_r$ be the tridiagonal matrix of dimension $r$ with 4 on the main diagonal and $-1$ on the two adjacent diagonals. We use the classical five-point central finite difference method to discretize the Hamiltonian operator (4.2) on uniform grid points in $\Omega$ with mesh size $h$. Then the corresponding matrices $B_L$ and $A_L$ in (1.8) are of the forms

$$\begin{aligned} B_L = {} & \delta_1 T_l \oplus (2(\delta_1+\delta_2)) \oplus \delta_2 T_m \oplus (2(\delta_1+\delta_2)) \oplus \delta_1 T_l \\ & - \delta_1\left(e_{l+1}e_l^T + e_l e_{l+1}^T\right) - \delta_2\left(e_{l+2}e_{l+1}^T + e_{l+1}e_{l+2}^T\right) \\ & - \delta_2\left(e_{n+1}e_n^T + e_n e_{n+1}^T\right) - \delta_1\left(e_{n+2}e_{n+1}^T + e_{n+1}e_{n+2}^T\right) \\ & + \omega_1 h^2 \mathrm{diag}\left((1-c)^2, (2-c)^2, \ldots, (N-c)^2\right) \end{aligned}$$

and

$$A_L = -\left[\delta_1 I_l \oplus \left(\frac{\delta_1+\delta_2}{2}\right) \oplus \delta_2 I_m \oplus \left(\frac{\delta_1+\delta_2}{2}\right) \oplus \delta_1 I_l\right],$$

where $\delta_i = \hbar/2h^2\varepsilon_i$ $(i = 1, 2)$, $e_j$ denotes the $j$th vector of the identity matrix, $l$ and $m$ are the numbers of grid points on the $x_1$ axis in $(-9, -1)$ and $(-1, 1)$, respectively, $n = l + m + 1$, $N = 2l + m + 2$, $c = (N + 1)/2$ and $\oplus$ denotes the direct sum of two matrices.

In our tests we take $l = 39$, $m = 9$, $\delta_1 = 1$, $\delta_2 = 0.1$, and $\omega_1 = 5 \times 10^{-4}$. By Theorem 2.6 we find $\Delta_L = [0.00386, 8.0103]$. We divide $\Delta_L$ into $P$ subintervals using $P + 1$ equally spaced nodes $\mathcal{E}_i$, $i = 0, \ldots, P$. We now choose $P = 1000$ and take $\eta = 10^{-6}$ in (1.8). Let $d_i$ be the distance between $\mathbb{T}$ and the set of eigenvalues of the pencil $(M_0, L_0)$ in (2.1), with $\mathcal{E} = \mathcal{E}_i$. We find that $d_i < 10^{-5}$ whenever $\mathcal{E}_i \in \Delta_L$.

We run DA for each $\mathcal{E}_i$ as well as for a few $\mathcal{E}$ values outside $\Delta_L$. The algorithm is stopped when $\|Q_{k+1} - Q_k\| < 10^{-8}$ and $X = Q_{k+1}$ is taken to be a good approximation to the stabilizing solution of (1.7). In Figure 4.2, we plot the relative residuals and the number of iterations. We see that DA converges to the stabilizing solution of (1.7) in about 26 iterations for $\mathcal{E} \in \Delta_L$. We also see that r_res $< 10^{-9}$ for almost all $\mathcal{E}$ values.

In this example, we have increased $\eta$ from $10^{-10}$ to $10^{-6}$, and as a result the usual number of DA iterations is reduced from 40 to 26. We have also tried $\eta = 10^{-10}$ for this example and find that DA still works very well with the number of iterations around 40, as in Example 4.1. In the nano literature, the smallest $\eta$ used is $10^{-6}$ since no powerful methods were available at that time. With DA studied in this paper, we can also take $\eta$ to be smaller than $10^{-10}$ if there is a need to do so.

Suppose we only use DA to find the stabilizing solution for $\mathcal{E} = \mathcal{E}_0$ and use NM (in the same way as in Example 4.1) to find the stabilizing solutions for $\mathcal{E} = \mathcal{E}_i$, $i = 1, \ldots, P$. In the notation of Example 4.1, we have $q = P$, $m = 1$, and Eff $= S/P$. We find that Eff $= 46.3\%$ and $46.5\%$ for $P = 1000$ and 2000, respectively. To have Eff $= 100\%$ one would have to take an extremely large $P$, which is not practical.

*Example* 4.3. We consider Example 2.1 with $t = 1$. The matrices $A$ and $Q$ in (1.8) are now given by $A = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ and $Q = (\mathcal{E} + i\eta)I_2 - \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. We take $\eta = 10^{-10}$. For $0 \le \mathcal{E} \le 4$, the distance between $\mathbb{T}$ and the set of eigenvalues of the pencil $(M_0, L_0)$ in (2.1) is less than $10^{-8}$.

We run DA as before. In Figure 4.3 we plot the relative residuals and the number of iterations for convergence of DA. In most cases, 37 iterations are needed. We know from Theorem 3.1(c) that in exact arithmetic the approximate solution obtained after 37 DA iterations is the same as the approximate solution obtained after $2^{37} - 1$ (which is over 137 billion) iterations of the basic fixed-point method (Algorithm 3.2). So we would not try to use Algorithm 3.2 even for this small problem. But M_FPM turns out to be quite useful. We run M_FPM for $\mathcal{E} = \mathcal{E}_i = 0.004i$, $i = 0, 1, \ldots, 1000$. The algorithm is stopped when r_res $< 10^{-10}$ or the number of iterations exceeds 10000. Figure 4.4 plots the relative residuals and the number of iterations. More specificly, we find that M_FPM converges in 37–100 iterations for 536 values of $\mathcal{E}_i$ with $i \in \{37$–$301, 303$–$305, 695$–$697, 699$–$963\}$, converge in 0 iterations for $\mathcal{E}_{500} = 2$, and does not converge in 100 iterations for the other 464 values of $\mathcal{E}_i$. Note that for $\mathcal{E}_{500} = 2$, the number of M_FPM iterations is 0 since $X_0 = Q$ happens to be the exact solution when $\eta = 0$ and since we take $\eta = 10^{-10}$, which is very close to 0.

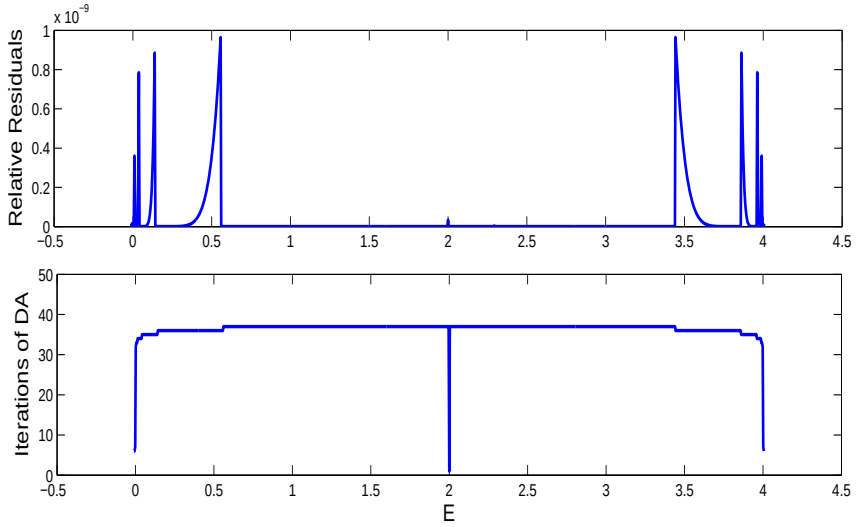


FIG. 4.3. *Relative residuals and the number of iterations for convergence of DA,* $\eta = 10^{-10}$.

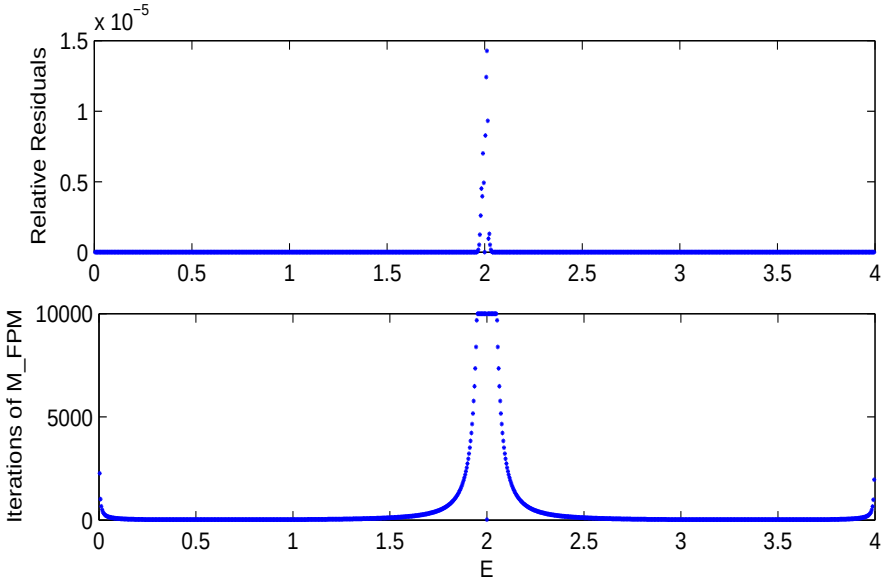


FIG. 4.4. *Relative residuals and the number of iterations for convergence of M_FPM,* $\eta = 10^{-10}$.

DA takes one iteration for $\mathcal{E} = 2$ only because a different stopping criterion is used. For $\mathcal{E} = \mathcal{E}_i, i = 302, 698$, the number of M_FPM iterations is 102, exceeding 100 only slightly. We may say that for this example, M_FPM is more efficient than DA most of the time. However, there is currently no theory about the convergence of M_FPM even for this simple example.

**5. Conclusion.** We have studied a doubling algorithm for Green's function calculations in nano research. Its convergence to the desired solution is guaranteed. The algorithm has been shown to be efficient and reliable. Newton's method cannot be used for this purpose by itself, but may be used as a correction method if very high accuracy is required. A modified fixed-point method may be more efficient than the doubling algorithm in some situations, but its convergence analysis remains an open problem.

REFERENCES


[1] I. Appelbaum, T. Wang, J. D. Joannopoulos, and V. Narayanamurti, *Ballistic hot-electron transport in nanoscale semiconductor heterostructures: Exact self-energy of a three-dimensional periodic tight-binding Hamiltonian*, Physical Review B, 69 (2004), Article Number 165301, 6 pp.

[2] D. A. Bini, L. Gemignani, and B. Meini, *Computations with infinite Toeplitz matrices and polynomials*, Linear Algebra Appl., 343–344 (2002), pp. 21–61.

[3] C.-Y. Chiang, E. K.-W. Chu, C.-H. Guo, T.-M. Huang, W.-W. Lin, and S.-F. Xu, *Convergence analysis of the doubling algorithm for several nonlinear matrix equations in the critical case*, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 227–247.

[4] E. K.-W. Chu, T.-M. Hwang, W.-W. Lin, and C.-T. Wu, *Vibration of fast trains, palindromic eigenvalue problems and structure-preserving doubling algorithms*, J. Comput. Appl. Math., 219 (2008), pp. 237–252.

[5] S. Datta, *Electronic Transport in Mesoscopic Systems*, Cambridge University Press, 1995.

[6] S. Datta, *Nanoscale device modeling: the Green's function method*, Superlattices and Microstructures, 28 (2000), pp. 253–278.

[7] E. N. Economou, *Green's Functions in Quantum Physics, 3rd edition*, Springer, 2006.

[8] J. C. Engwerda, A. C. M. Ran, and A. L. Rijkeboer, *Necessary and sufficient conditions for the existence of a positive definite solution of the matrix equation* $X + A^* X^{-1} A = Q$, Linear Algebra Appl., 186 (1993), pp. 255–275.

[9] I. Gohberg, S. Goldberg, and M. A. Kaashoek, *Classes of Linear Operators, Vol. II, Operator Theory: Advances and Applications, Vol. 63*, Birkhäuser, 1993.

[10] C.-H. Guo, *Convergence rate of an iterative method for a nonlinear matrix equation*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 295–302.

[11] C.-H. Guo, *Numerical solution of a quadratic eigenvalue problem*, Linear Algebra Appl., 385 (2004), pp. 391–406.

[12] C.-H. Guo and P. Lancaster, *Iterative solution of two matrix equations*, Math. Comp., 68 (1999), pp. 1589–1603.

[13] N. J. Higham and H.-M. Kim, *Numerical analysis of a quadratic matrix equation*, IMA J. Numer. Anal., 20 (2000), pp. 499–519.

[14] N. J. Higham and H.-M. Kim, *Solving a quadratic matrix equation by Newton's method with exact line searches*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 303–316.

[15] D. L. John and D. L. Pulfrey, *Green's function calculations for semi-infinite carbon nanotubes*, Physica Status Solidi B – Basic Solid State Physics, 243 (2006), pp. 442–448.

[16] A. Kletsov, Y. Dahnovsky, and J. V. Ortiz, *Surface Green's function calculations: A nonrecursive scheme with an infinite number of principal layers*, Journal of Chemical Physics, 126 (2007), Article Number 134105, 5 pp.

[17] W.-W. Lin and S.-F. Xu, *Convergence analysis of structure-preserving doubling algorithms for Riccati-type matrix equations*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 26–39.

[18] D. S. Mackey, N. Mackey, C. Mehl, and V. Mehrmann, *Structured polynomial eigenvalue problems: good vibrations from good linearizations*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 1029–1051.

[19] C. van der Mee, G. Rodriguez, and S. Seatzu, *LDU Factorization results for bi-infinite and semi-infinite scalar and block Toeplitz matrices*, Calcolo, 33 (1996), pp. 307–335.

[20] B. Meini, *Efficient computation of the extreme solutions of* $X + A^* X^{-1} A = Q$ *and* $X - A^* X^{-1} A = Q$, Math. Comp., 71 (2002), pp. 1189–1204.

[21] J. Tomfohr and O. F. Sankey, *Theoretical analysis of electron transport through organic molecules*, Journal of Chemical Physics, 120 (2004), pp. 1542–1554.

[22] X. Zhan, *Computing the extremal positive definite solutions of a matrix equation*, SIAM J. Sci. Comput., 17 (1996), pp. 1167–1174.