Numerical solution of nonlinear matrix equations arising from Green's function calculations in nano research

Chun-Hua Guo[a,1], Yueh-Cheng Kuo[b,2], Wen-Wei Lin[c,3]

[a] *Department of Mathematics and Statistics, University of Regina, Regina, SK S4S 0A2, Canada*
[b] *Department of Applied Mathematics, National University of Kaohsiung, Kaohsiung 811, Taiwan*
[c] *Department of Applied Mathematics, National Chiao Tung University, Hsinchu 300, Taiwan*

Abstract

The Green's function approach for treating quantum transport in nano devices requires the solution of nonlinear matrix equations of the form $X + (C^* + \mathrm{i}\eta D^*)X^{-1}(C + \mathrm{i}\eta D) = R + \mathrm{i}\eta P$, where R and P are Hermitian, $P + \lambda D^* + \lambda^{-1}D$ is positive definite for all λ on the unit circle, and $\eta \to 0^+$. For each fixed $\eta > 0$, we show that the required solution is the unique stabilizing solution X_η. Then $X_* = \lim_{\eta\to 0^+} X_\eta$ is a particular weakly stabilizing solution of the matrix equation $X + C^*X^{-1}C = R$. In nano applications, the matrices R and C are dependent on a parameter, which is the system energy $\mathcal{E}$. In practice one is mainly interested in those values of $\mathcal{E}$ for which the equation $X + C^*X^{-1}C = R$ has no stabilizing solutions or, equivalently, the quadratic matrix polynomial $P(\lambda) = \lambda^2 C^* - \lambda R + C$ has eigenvalues on the unit circle. We point out that a doubling algorithm can be used to compute X_η efficiently even for very small values of η, thus providing good approximations to X_*. We also explain how the solution X_* can be computed directly using subspace methods such as the QZ algorithm by determining which unimodular eigenvalues of $P(\lambda)$ should be included in the computation. In some applications the matrices C, D, R, P have very special sparsity structures. We show how these special structures can be expoited to drastically reduce the complexity of the doubling algorithm for computing X_η.

Keywords: nonlinear matrix equation, weakly stabilizing solution, structure-preserving algorithm, Green's function
2000 MSC: 15A24, 65F30

1. Introduction

In this paper we study nonlinear matrix equations of the form

$$X + (C^* + \mathrm{i}\eta D^*)X^{-1}(C + \mathrm{i}\eta D) = R + \mathrm{i}\eta P, \tag{1}$$

where R and P are Hermitian, $P + \lambda D^* + \lambda^{-1}D$ is positive definite for all λ on the unit circle $\mathbb{T}$, and $\eta \geq 0$. The special case where $P = I, D = 0$ and C, R are real arises in nano research [1, 3, 12, 13] and has been studied in [7, 9]. We now briefly explain how the general equation (1) also arises in nano applications.

A main goal of basic research in molecular electronics is to advance the understanding of electron transport through molecules. In [17], a method for calculating the current is described for a system that consists

Email addresses: chguo@math.uregina.ca (Chun-Hua Guo), yckuo@nuk.edu.tw (Yueh-Cheng Kuo), wwlin@math.nctu.edu.tw (Wen-Wei Lin)

[1]The work of this author was supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada.
[2]The work of this author was partially supported by the National Science Council in Taiwan.
[3]The work of this author was partially supported by the National Center for Theoretical Sciences in Taiwan, and by ST Yau Center and the Center of Mathematical Modelling and Scientific Computing at National Chiao Tung University.

of a molecule connected between two semi-infinite metallic electrodes, and is implemented in a program that assumes a local-orbital picture and requires as input the Hamiltonian and overlap matrix elements between orbitals.

The system Hamiltonian is a bi-infinite Hermitian matrix of the form

$$H = \begin{bmatrix} H_L & H_{LM} & 0 \\ H_{LM}^* & H_M & H_{MR} \\ 0 & H_{MR}^* & H_R \end{bmatrix}, \tag{2}$$

where H_M, H_L, H_R are the Hamiltonians for the molecule, the left electrode, and the right electrode, respectively, and the overlap matrix is a Hermitian positive definite matrix partitioned in the same way and is given by

$$S = \begin{bmatrix} S_L & S_{LM} & 0 \\ S_{LM}^* & S_M & S_{MR} \\ 0 & S_{MR}^* & S_R \end{bmatrix}.$$

In [17] the blocks H_L and H_{LM} in (2) are shifted by $s_L S_L$ and $s_L S_{LM}$, respectively, where s_L is a proper energy shift, and the blocks H_R and H_{MR} are shifted similarly. These shifts do not change the structure of the matrix H in (2). So we can simply assume that the matrix H in (2) has already gone through the shifting procedure.

The Green's function (of the full interacting system) is defined by

$$G = \left((\mathcal{E} + \mathrm{i}0^+)S - H\right)^{-1} = \lim_{\eta \to 0^+} \left((\mathcal{E} + \mathrm{i}\eta)S - H\right)^{-1},$$

where $\mathcal{E}$ is energy. We note that for each $\eta > 0$ the infinite matrix $(\mathcal{E} + \mathrm{i}\eta)S - H = \mathcal{E}S - H + \mathrm{i}\eta S$ is known to be invertible by Bendixson's theorem (see [10, Lemma 3.3]), but the existence of the above one-sided limit is something assumed. The molecule Green's function G_M is that part of G corresponding to the block for the molecule and is obtained from

$$G_M = \left((\mathcal{E} + \mathrm{i}0^+)S_M - H_M - \Sigma_L - \Sigma_R\right)^{-1},$$

where

$$\Sigma_L = (\mathcal{E}S_{LM} - H_{LM})^* G_L (\mathcal{E}S_{LM} - H_{LM}), \quad \Sigma_R = (\mathcal{E}S_{MR} - H_{MR}) G_R (\mathcal{E}S_{MR} - H_{MR})^*,$$

with

$$G_L = \left((\mathcal{E} + \mathrm{i}0^+)S_L - H_L\right)^{-1}, \quad G_R = \left((\mathcal{E} + \mathrm{i}0^+)S_R - H_R\right)^{-1}.$$

Then [17] the net current is detemined through a definite integral of the transmission function given by

$$T(\mathcal{E}) = \mathrm{tr}(\Gamma_L G_M \Gamma_R G_M^*),$$

where

$$\Gamma_L = \mathrm{i}(\Sigma_L - \Sigma_L^*), \quad \Gamma_R = \mathrm{i}(\Sigma_R - \Sigma_R^*),$$

and $\mathrm{tr}(\cdot)$ denotes the trace of a matrix. Note that $T(\mathcal{E})$ is a real function of $\mathcal{E}$ since Γ_L and Γ_R are Hermitian.

We now explain how the matrix Σ_R is computed. The computation of Σ_L is similar. The matrices H_R and S_R can be written as [17]

$$H_R = \begin{bmatrix} H_s & H_{sb} & & & \\ H_{sb}^* & H_b & H_{bb} & & \\ & H_{bb}^* & H_b & H_{bb} & \\ & & H_{bb}^* & H_b & \ddots \\ & & & \ddots & \ddots \end{bmatrix}, \quad S_R = \begin{bmatrix} S_s & S_{sb} & & & \\ S_{sb}^* & S_b & S_{bb} & & \\ & S_{bb}^* & S_b & S_{bb} & \\ & & S_{bb}^* & S_b & \ddots \\ & & & \ddots & \ddots \end{bmatrix},$$

where $H_s, S_s \in \mathbb{C}^{q\times q}$ and $H_b, S_b \in \mathbb{C}^{n\times n}$, and we suppose that $H_M, S_M \in \mathbb{C}^{p\times p}$. The size of H_s and S_s has been taken sufficiently large so that all nonzero elements of the matrices H_{MR} and S_{MR} are in the $p \times q$ block on the left. This means that we only need G_s, the $q \times q$ block in the upper-left corner of G_R, for the computation of Σ_R. It is easy to see [17] that G_s is determined through

$$G_s = (U_s - U_{sb}G_bU'_{sb})^{-1},$$

where $U_s = zS_s - H_s, U_{sb} = zS_{sb} - H_{sb}, U'_{sb} = zS^*_{sb} - H^*_{sb}$ with $z = \mathcal{E} + \mathrm{i}\eta$ and $\eta \to 0^+$, and G_b is the $n \times n$ block in the upper-left corner of the inverse of

$$\begin{bmatrix} zS_b - H_b & zS_{bb} - H_{bb} & & \\ zS^*_{bb} - H^*_{bb} & zS_b - H_b & zS_{bb} - H_{bb} & \\ & zS^*_{bb} - H^*_{bb} & zS_b - H_b & \ddots \\ & & \ddots & \ddots \end{bmatrix}. \tag{3}$$

Note that the above matrix is invertible by Bendixson's theorem since the matrix

$$\mathcal{T}_R = \begin{bmatrix} S_b & S_{bb} & & \\ S^*_{bb} & S_b & S_{bb} & \\ & S^*_{bb} & S_b & \ddots \\ & & \ddots & \ddots \end{bmatrix} \tag{4}$$

is positive definite when S is positive definite. Note also that $\mathcal{T}_R$ is positive definite if and only if $S_b + \lambda S_{bb} + \lambda^{-1}S^*_{bb}$ is positive definite for all λ on $\mathbb{T}$.

The block Toeplitz structure of the matrix (3) implies that G_b satisfies the matrix equation

$$G_b = (U_b - U_{bb}G_bU'_{bb})^{-1}, \tag{5}$$

where $U_b = zS_b - H_b, U_{bb} = zS_{bb} - H_{bb}, U'_{bb} = zS^*_{bb} - H^*_{bb}$.

For any $W \in \mathbb{C}^{n\times n}$, we can write $W = W_R + \mathrm{i}W_I$, where the Hermitian matrices

$$W_R = \frac{1}{2}(W + W^*), \quad W_I = \frac{1}{2\mathrm{i}}(W - W^*)$$

are called the real part and the imaginary part of W, respectively. We are only interested in $\mathcal{E}$ values for which the required solution G_b of (5) has a nonzero imaginary part (in the limit $\eta \to 0^+$) since otherwise G_b and then G_s would be Hermitian, which would imply that Σ_R is Hermitian and then $T(\mathcal{E}) = 0$ for the transmission function.

Now we let

$$X = G_b^{-1}, \quad C = \mathcal{E}S^*_{bb} - H^*_{bb}, \quad D = S^*_{bb}, \quad R = \mathcal{E}S_b - H_b, \quad P = S_b.$$

Then the equation (5) becomes (1).

2. Characterization of the solution G_b

The matrix equation (1) may have many different solutions. So what solution X do we need so that $X^{-1} = G_b$ is the required solution of (5)?

Let $A = C + \mathrm{i}\eta D,\ B = C^* + \mathrm{i}\eta D^*,\ Q = R + \mathrm{i}\eta P$. Then (1) becomes

$$X + BX^{-1}A = Q. \tag{6}$$

As before, R and P are Hermitian and $P + \lambda D^* + \lambda^{-1}D$ is positive definite for all λ on $\mathbb{T}$. Let

$$M = \begin{bmatrix} A & 0 \\ Q & -I \end{bmatrix}, \quad L = \begin{bmatrix} 0 & I \\ B & 0 \end{bmatrix}. \tag{7}$$

Then X is a solution of (6) if and only if

$$M \begin{bmatrix} I \\ X \end{bmatrix} = L \begin{bmatrix} I \\ X \end{bmatrix} X^{-1}A. \tag{8}$$

Therefore, every solution of (6) can be obtained from a suitable invariant subspace for the pencil $M - \lambda L$.

Lemma 1. *For any $\eta \neq 0$, the matrix pencil $M - \lambda L$ has no eigenvalues on $\mathbb{T}$.*

Proof. Suppose that $\lambda \in \mathbb{T}$ and $(M - \lambda L)x = 0$ for a vector $x = (x_1^\top, x_2^\top)^\top$ with $x_1, x_2 \in \mathbb{C}^n$. Then

$$Ax_1 = \lambda x_2, \quad Qx_1 - x_2 = \lambda B x_1. \tag{9}$$

By eliminating x_2 in (9) we have

$$Wx_1 \equiv \left(\lambda B - Q + \lambda^{-1}A\right) x_1 = 0. \tag{10}$$

The imaginary part of $x_1^* W x_1$ is $-\eta x_1^*(P - \lambda D^* - \lambda^{-1}D)x_1$. Since $P - \lambda D^* - \lambda^{-1}D$ is positive definite, it follows from (10) that $x_1 = 0$. By (9) we have $x_2 = 0$. Thus, $M - \lambda L$ has no eigenvalues on $\mathbb{T}$. □

Theorem 2. *For any $\eta \neq 0$, the matrix pencil $M - \lambda L \in \mathbb{C}^{2n\times 2n}$ has n eigenvalues inside $\mathbb{T}$ and n eigenvalues outside $\mathbb{T}$.*

Proof. We consider the matrix pencils

$$H(t,\lambda) = \begin{bmatrix} tA & 0 \\ tR + \mathrm{i}\eta P & -I \end{bmatrix} - \lambda \begin{bmatrix} 0 & I \\ tB & 0 \end{bmatrix}$$

obtained from the pencil $M - \lambda L$ by replacing C, D, R with tC, tD, tR. For each $t \in [0,1]$ and $\lambda \in \mathbb{T}$, $P + \lambda(tD^*) + \lambda^{-1}(tD) = (1-t)P + t(P + \lambda D^* + \lambda^{-1}D)$ is positive definite. From Lemma 1 we know that $H(t,\lambda)$ has no eigenvalues on $\mathbb{T}$ for all $t \in [0,1]$. Hence, $H(1,\lambda) = M - \lambda L$ and $H(0,\lambda)$ have the same numbers of eigenvalues inside $\mathbb{T}$. But it is clear that $H(0,\lambda)$ has n eigenvalues at 0 and n eigenvalues at ∞. □

Note that the pencil $M - \lambda L$ is a linearization of the quadratic polynomial $P(\lambda) = \lambda^2 B - \lambda Q + A$.

The basic fixed-point iteration for finding a solution of (6) is $X_{k+1} = \mathcal{F}(X_k)$, where $\mathcal{F}(X) = Q - BX^{-1}A$. The Fréchet derivative of $\mathcal{F}$ at X is the linear map $\mathcal{F}'_X : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$ given by $\mathcal{F}'_X(Z) = (BX^{-1})Z(X^{-1}A)$. A solution X of (6) is said to be stabilizing if $\rho(\mathcal{F}'_X) < 1$ or, equivalently, $\rho(BX^{-1})\rho(X^{-1}A) < 1$, where $\rho(\cdot)$ denotes the spectral radius. Note that the basic fixed-point iteration is locally convergent at a stabilizing solution.

Let X be any solution of (6). Then we have

$$P(\lambda) = (\lambda BX^{-1} - I)X(\lambda I - X^{-1}A).$$

So the eigenvalues of $X^{-1}A$ are n eigenvalues of $P(\lambda)$, and the eigenvalues of BX^{-1} are the reciprocals of the remaining n eigenvalues of $P(\lambda)$. It then follows from Theorem 2 that a solution X of (6) is stabilizing if and only if $\rho(X^{-1}A) < 1$ and that there is at most one stabilizing solution.

Let the invariant subspace of $M - \lambda L$ corresponding to its n eigenvalues inside $\mathbb{T}$ be spanned by the columns of the matrix $\begin{bmatrix} U \\ V \end{bmatrix}$, where $U, V \in \mathbb{C}^{n\times n}$. Then the existence of a stabilizing solution can be established by showing that U and V are both invertible. The stabilizing solution is then $X = VU^{-1}$. For the case where the matrices C, D, R, P in (6) are all real, an elementary proof for the invertibility of U has already been given in [8] and we note that the invertibility of V can be proved in the same way. It is also shown in [8] that the imaginary part of the stabilizing solution is positive definite for $\eta > 0$. The proofs can be carried over to the complex case here with only very minor changes.

Here, however, we are going to use an advanced result on linear operators to show the existence of a stabilizing solution since this approach will also explain that the stabilizing solution is precisely the solution we need for the nano application. The treatment is very similar to the one in [9] for a special case of the equation (6). So our presentation here will be very brief.

Recall that $G_b = \lim_{\eta\to 0^+} G_b(\eta)$, where $G_b(\eta)$ is the $n \times n$ matrix in the upper-left corner of T^{-1} with T given by (3) for each $\eta > 0$. Using the current notation, we have

$$T = \begin{bmatrix} Q & B & & \\ A & Q & B & \\ & A & Q & \ddots \\ & & \ddots & \ddots \end{bmatrix}. \tag{11}$$

Associated with T is the rational matrix function $\phi(\lambda) = \lambda A + Q + \lambda^{-1}B$. We already know from Bendixson's theorem (see [10, Lemma 3.3]) that T is invertible for each $\eta > 0$. Thus, by a result on linear operators (see [6, Chapter XXIV, Theorem 4.1] and [15]) we know that $\phi(\lambda)$ has a factorization

$$\phi(\lambda) = (I - \lambda^{-1}L)X(I - \lambda U) \tag{12}$$

with X invertible, $\rho(L) < 1$ and $\rho(U) < 1$. From (12) we see that

$$A = -XU, \quad B = -LX, \quad Q = X + LXU.$$

Thus $X + BX^{-1}A = Q$ and $\rho(X^{-1}A) < 1$. In other words, X is the unique stabilizing solution of (6).

By [6, Chapter XXIV, Theorem 4.1] the $n \times n$ matrix in the upper-left corner of T^{-1} is precisely X^{-1}. We thus have the following characterization of $G_b(\eta)$.

Theorem 3. *For any $\eta > 0$, the matrix $G_b(\eta)$ is the inverse of the unique stabilizing solution of* (6).

3. Computation of the stabilizing solution

Let M and L be as in (7). Then the stabilizing solution X of (6) satisfies (8) with $\rho(X^{-1}A) < 1$.

We remark that the equation (6) with *real* matrices C, D, R, P also arises in the study of a quadratic eigenvalue problem from the vibration analysis of fast trains, where the required solution is also the stabilizing solution and a doubling algorithm is used to find the solution [10]. We can get similar results for our more general equation (6). The situation here is slightly more complicated since we no longer have $B = A^\top$ and the stabilizing solution is no longer complex symmetric.

Starting with the matrices M and L in (7), we define the sequences $\{M_k\}$ and $\{L_k\}$, where

$$M_k = \begin{bmatrix} A_k & 0 \\ Q_k & -I \end{bmatrix}, \quad L_k = \begin{bmatrix} -P_k & I \\ B_k & 0 \end{bmatrix},$$

by the following structure-preserving doubling algorithm if no breakdown occurs.

Algorithm 1. *Let $A_0 = A, B_0 = B, Q_0 = Q, P_0 = 0$.*
For $k = 0, 1, \ldots$, compute

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

The above algorithm is the SDA-2 as presented in [2]. The next result shows that the doubling algorithm has some nice properties. In particular, it can compute the stabilizing solution X of (6) efficiently.

Theorem 4. *Let X be the stabilizing solution of (6) and $\widehat{X}$ be the stabilizing solution of the dual equation*

$$\widehat{X} + A\widehat{X}^{-1}B = Q.$$

Then

(a) *The sequences $\{A_k\}, \{B_k\}, \{Q_k\}, \{P_k\}$ in Algorithm 1 are well-defined.*

(b) *Q_k converges to X quadratically, A_k and B_k converge to 0 quadratically, $Q - P_k$ converges to $\widehat{X}$ quadratically, with*

$$\limsup_{k\to\infty} \sqrt[2^k]{\|Q_k - X\|} \le \rho(\widehat{X}^{-1}B)\rho(X^{-1}A),\ \limsup_{k\to\infty} \sqrt[2^k]{\|A_k\|} \le \rho(X^{-1}A),$$

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

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

Proof. The proof is very similar to that of [10, Theorem 4.1]. Although the statement of that theorem and the beginning of its proof refer to the specific problem under consideration in [10], the proof there is valid for this theorem after some minor changes. Here we only mention the following differences. In [10], $Q^\top = Q$ and $B = A^\top$ (this would be true here if the matrices C, D, R, P were all real), and in that case we can conclude that $B_k = A_k^\top, Q_k^\top = Q_k, P_k^\top = P_k$, and $\rho(\widehat{X}^{-1}B) = \rho(X^{-1}A)$. □

Remark 1. Although we no longer have $\rho(\widehat{X}^{-1}B) = \rho(X^{-1}A)$ in general, we always have $\rho(\widehat{X}^{-1}B) = \rho(X^{-1}B)$. In fact, the eigenvalues of $\widehat{X}^{-1}B$ are those of $\widehat{P}(\lambda) = \lambda^2 A - \lambda Q + B$ inside $\mathbb{T}$. We have mentioned earlier that the eigenvalues of BX^{-1} are the reciprocals of those eigenvalues of $P(\lambda) = \lambda^2 B - \lambda Q + A$ outside $\mathbb{T}$. However, the reciprocals of the eigenvalues of $P(\lambda)$ outside $\mathbb{T}$ are precisely the eigenvalues of $\widehat{P}(\lambda)$ inside $\mathbb{T}$. It is also well known that BX^{-1} and $X^{-1}B$ have the same eigenvalues.

Remark 2. As in [9], we can show that the basic fixed-point iteration (FPI)

$$X_{k+1} = Q - BX_k^{-1}A, \quad X_0 = Q$$

is also convergent and $X_{2^k-1} = Q_k$. So the convergence of the FPI is much slower than the doubling algorithm. However, we can use an averaging procedure for the FPI to speed up its convergence and for the nano application we can use the computed solution for one energy value as an initial guess for the solution for the next nearby energy value [17]. For the special case where $P = I, D = 0$ and C, R are real, convergence results for methods based on these ideas have been proved in [7] using the Earle–Hamilton theorem [4, 11]. The proofs there can be carried over to equation (6), with some minor changes, as long as $D = 0$ still holds (so C is any complex matrix, R is any Hermitian matrix, and P is any Hermitian positive definite matrix). However, when $D \neq 0$ we are unable to prove any non-local convergence results for those methods. The Earle–Hamilton theorem is not applicable since we no longer have $B = A^*$.

To emphasize its dependence on η, the stabilizing solution of (6) will be denoted by X_η. For the nano application, $G_b = \lim_{\eta\to 0^+} G_b(\eta)$ and $G_b(\eta) = X_\eta^{-1}$. So $G_b = X_*^{-1}$ with $X_* = \lim_{\eta\to 0^+} X_\eta$. It is easy to see that X_* is a particular weakly stabilizing solution of the matrix equation

$$X + C^* X^{-1} C = R, \tag{13}$$

with $\rho(X_*^{-1}C) \le 1$ and $\rho(C^* X_*^{-1}) \le 1$. The solution X_* can be approximated by computing X_η by the doubling algorithm for a sufficiently small η, but can also be computed directly by subspace methods, as we shall see in section 5. For now, we write $X_* = X_{*,R} + \mathrm{i}X_{*,I}$, where the Hermitian matrices $X_{*,R}$ and $X_{*,I}$ are the real part and the imaginary part of X_*, respectively, and we will examine the rank of $X_{*,I}$. Since the imaginary part of X_η is positive definite for $\eta > 0$, we know that $X_{*,I}$ is positive semi-definite.

4. Rank of $X_{*,I}$

We now denote the matrices A, B, Q in (6) by A_η, B_η, Q_η, respectively. So

$$A_\eta = C + \mathrm{i}\eta D, \quad B_\eta = C^* + \mathrm{i}\eta D^*, \quad Q_\eta = R + \mathrm{i}\eta P, \tag{14}$$

with C, D, R, P as before. Let

$$P_\eta(\lambda) = \lambda^2 B_\eta - \lambda Q_\eta + A_\eta.$$

We already know that P_η has no eigenvalues on $\mathbb{T}$ for $\eta \neq 0$. For $\eta = 0$ we get

$$P_0(\lambda) = \lambda^2 C^* - \lambda R + C.$$

It is quite possible that $P_0(\lambda)$ has some eigenvalues on $\mathbb{T}$. As we will see later, this is the case of primary interest for the nano application.

Theorem 5. *The number of eigenvalues (counting multiplicities) of $P_0(\lambda)$ on $\mathbb{T}$ must be even, say $2m$. Moreover, we have rank$(X_{*,I}) \leq m$.*

Proof. The matrix polynomial $P_0(\lambda)$ is $*$-palindromic. Thus μ is an eigenvalue of $P_0(\lambda)$ if and only if $1/\overline{\mu}$ is so, and they have the same algebraic, geometric, and partial multiplicities [14]. It follows that the total number of eigenvalues of $P_0(\lambda)$ on $\mathbb{T}$ must be even.

By $X_\eta + (C^* + \mathrm{i}\eta D^*)X_\eta^{-1}(C + \mathrm{i}\eta D) = R + \mathrm{i}\eta P$ we have

$$X_\eta + C^* X_\eta^{-1} C = R + \eta W_\eta, \tag{15}$$

where

$$W_\eta = \mathrm{i}P - \mathrm{i}D^* X_\eta^{-1} C - \mathrm{i}C^* X_\eta^{-1} D + \eta D^* X_\eta^{-1} D.$$

Taking imaginary parts on (15), we get

$$K_\eta - F_\eta^* K_\eta F_\eta = \eta T_\eta, \tag{16}$$

where $K_\eta = \mathrm{Im}(X_\eta), T_\eta = \mathrm{Im}(W_\eta), F_\eta = X_\eta^{-1} C$. Let $F = \lim_{\eta \to 0^+} F_\eta = X_*^{-1} C$. Then the eigenvalues of F consist of all $n - m$ eigenvalues of $P_0(\lambda)$ inside $\mathbb{T}$ plus m eigenvalues of $P_0(\lambda)$ on $\mathbb{T}$. Let

$$F = V_0 \begin{bmatrix} R_{0,1} & 0 \\ 0 & R_{0,2} \end{bmatrix} V_0^{-1} \tag{17}$$

be a spectral resolution of F, where $R_{0,1} \in \mathbb{C}^{m \times m}$ and $R_{0,2} \in \mathbb{C}^{(n-m)\times(n-m)}$ are upper triangular with $\sigma(R_{0,1}) \subseteq \mathbb{T}$ and $\sigma(R_{0,2}) \subseteq \mathbb{D} \equiv \{\lambda \in \mathbb{C} |\ |\lambda| < 1\}$. It follows from [16, Chapter V, Theorem 2.8] that there is a nonsingular matrix V_η such that

$$F_\eta = V_\eta \begin{bmatrix} R_{\eta,1} & 0 \\ 0 & R_{\eta,2} \end{bmatrix} V_\eta^{-1}, \tag{18}$$

and $R_{\eta,1} \to R_{0,1}$, $R_{\eta,2} \to R_{0,2}$, and $V_\eta \to V_0$, as $\eta \to 0^+$.

From (16) and (18) we have

$$V_\eta^* K_\eta V_\eta - \begin{bmatrix} R_{\eta,1}^* & 0 \\ 0 & R_{\eta,2}^* \end{bmatrix} V_\eta^* K_\eta V_\eta \begin{bmatrix} R_{\eta,1} & 0 \\ 0 & R_{\eta,2} \end{bmatrix} = \eta V_\eta^* T_\eta V_\eta. \tag{19}$$

Let

$$V_\eta^* K_\eta V_\eta = \begin{bmatrix} H_{\eta,1} & H_{\eta,3} \\ H_{\eta,3}^* & H_{\eta,2} \end{bmatrix}, \quad V_\eta^* T_\eta V_\eta = \begin{bmatrix} Z_{\eta,1} & Z_{\eta,3} \\ Z_{\eta,3}^* & Z_{\eta,2} \end{bmatrix}. \tag{20}$$

Then (19) becomes

$$H_{\eta,1} - R_{\eta,1}^* H_{\eta,1} R_{\eta,1} = \eta Z_{\eta,1}, \tag{21a}$$

$$H_{\eta,2} - R_{\eta,2}^* H_{\eta,2} R_{\eta,2} = \eta Z_{\eta,2}, \tag{21b}$$

$$H_{\eta,3} - R_{\eta,1}^* H_{\eta,3} R_{\eta,2} = \eta Z_{\eta,3}. \tag{21c}$$

As $\eta \to 0^+$, $R_{\eta,1} \to R_{0,1}$ with $\rho(R_{0,1}) = 1$, $R_{\eta,2} \to R_{0,2}$ with $\rho(R_{0,2}) < 1$, and $Z_{\eta,2}$ and $Z_{\eta,3}$ are bounded by the convergence of T_η. So we have $H_{\eta,2} \to 0$ from (21b) and $H_{\eta,3} \to 0$ from (21c). Since $X_{*,I} = \lim_{\eta\to 0^+} K_\eta$, it follows from (20) that $\operatorname{rank}(X_{*,I}) \le m$. □

We conjecture that equality holds in Theorem 5 when all eigenvalues of $P_0(\lambda)$ on $\mathbb{T}$ are simple.

For the nano application, the matrices C and R in $P_0(\lambda)$ are given by

$$C = \mathcal{E} S_{bb}^* - H_{bb}^*, \quad R = \mathcal{E} S_b - H_b.$$

If $P_0(\lambda)$ has no eigenvalues on $\mathbb{T}$ for an energy value $\mathcal{E}$, then X_* is Hermitian by Theorem 5 and $G_b = X_*^{-1}$ is also Hermitian. We then know that the transmission function $T(\mathcal{E})$ takes zero value, without solving any nonlinear matrix equations. So we are only interested in those $\mathcal{E}$ values for which $P_0(\lambda)$ has some eigenvalues on $\mathbb{T}$. The next simple result is thus relevant, where $S_b - \lambda S_{bb} - \lambda^{-1} S_{bb}^*$ is positive definite for all λ on $\mathbb{T}$.

Theorem 6. *For $\lambda \in \mathbb{T}$, let the eigenvalues of*

$$(S_b - \lambda S_{bb} - \lambda^{-1} S_{bb}^*)^{-1} (H_b - \lambda H_{bb} - \lambda^{-1} H_{bb}^*)$$

be $\mu_1(\lambda) \le \cdots \le \mu_n(\lambda)$. Let

$$\Delta_i = \left[\min_{|\lambda|=1} \mu_i(\lambda),\ \max_{|\lambda|=1} \mu_i(\lambda) \right],$$

and $\Delta = \bigcup_{i=1}^n \Delta_i$. Then the quadratic pencil $P_0(\lambda) = \lambda^2(\mathcal{E} S_{bb} - H_{bb}) - \lambda(\mathcal{E} S_b - H_b) + (\mathcal{E} S_{bb}^* - H_{bb}^*)$ has some eigenvalues on $\mathbb{T}$ if and only if $\mathcal{E} \in \Delta$.

Proof. The quadratic $P_0(\lambda)$ has some eigenvalues on $\mathbb{T}$ if and only if $\det(P_0(\lambda)) = 0$ for some $\lambda \in \mathbb{T}$ or, equivalently,

$$\det(-\lambda^{-1} P_0(\lambda)) = \det\left(\mathcal{E}(S_b - \lambda S_{bb} - \lambda^{-1} S_{bb}^*) - (H_b - \lambda H_{bb} - \lambda^{-1} H_{bb}^*) \right) = 0$$

for some $\lambda \in \mathbb{T}$, the latter is equivalent to $\mathcal{E} \in \Delta$. □

5. Direct computation of X_*

The solution X_* can be computed directly by subspace methods. We will need to include all eigenvalues of

$$P(\lambda) = \lambda^2 C^* - \lambda R + C \tag{22}$$

inside $\mathbb{T}$ and half of its eigenvalues on $\mathbb{T}$ — the half that would be perturbed to the inside of $\mathbb{T}$ when $P(\lambda)$ is perturbed to

$$P_\eta(\lambda) = \lambda^2 (C^* + \mathrm{i}\eta D^*) - \lambda(R + \mathrm{i}\eta P) + (C + \mathrm{i}\eta D). \tag{23}$$

Let

$$\mathcal{M} = \begin{bmatrix} C & 0 \\ R & -I \end{bmatrix}, \quad \mathcal{L} = \begin{bmatrix} 0 & I \\ C^* & 0 \end{bmatrix}. \tag{24}$$

Then the pencil $\mathcal{M}-\lambda\mathcal{L}$, also denoted by $(\mathcal{M},\mathcal{L})$, is a linearization of the quadratic matrix polynomial $P(\lambda)$. It is easy to check that y and z are the right and left eigenvectors, respectively, corresponding to an eigenvalue λ of $P(\lambda)$ if and only if

$$\begin{bmatrix} y \\ Ry-\lambda C^*y \end{bmatrix}, \quad \begin{bmatrix} z \\ -\overline{\lambda}z \end{bmatrix} \tag{25}$$

are the right and left eigenvectors of $(\mathcal{M},\mathcal{L})$, respectively.

The following result is a generalization of [7, Theorem 3.1] for the special case where $P=I, D=0$ and C,R are real. It shows which invariant subspace corresponding to unimodular eigenvalues of $P(\lambda)$ should be used in the computation of X_*, assuming they are all semi-simple.

Theorem 7. *Suppose that λ_0 is a semi-simple eigenvalue of $P(\lambda)$ on $\mathbb{T}$ with multiplicity m_0 and $Y\in\mathbb{C}^{n\times m_0}$ forms an orthonormal basis of right eigenvectors corresponding to λ_0. Then $\mathrm{i}Y^*(2\lambda_0C^*-R)Y$ is a nonsingular Hermitian matrix. Let d_j, $j=1,\ldots,\ell$, be the distinct eigenvalues of*

$$Y^*(P-\lambda_0D^*-\lambda_0^{-1}D)Y\,(\mathrm{i}Y^*(2\lambda_0C^*-R)Y)^{-1}$$

with multiplicities m_0^j, and let $\xi_j\in\mathbb{C}^{m_0\times m_0^j}$ form an orthonormal basis of the eigenspace corresponding to d_j. Then for $\eta>0$ sufficiently small and $j=1,\ldots\ell$

$$\lambda_{j,\eta}^{(k)}=\lambda_0-\lambda_0d_j\eta+O(\eta^2),\ k=1,\ldots,m_0^j,$$

and

$$y_{j,\eta}=Y\left((Y^*(P-\lambda_0D^*-\lambda_0^{-1}D)Y\right)^{-1}\xi_j+O(\eta) \tag{26}$$

are perturbed eigenvalues and a basis of the corresponding invariant subspace of $P_\eta(\lambda)$, respectively.

Proof. Since $P(\lambda_0)Y=0$ with $Y^*Y=I_{m_0}$ and $|\lambda_0|=1$, we have

$$0^*=(P(\lambda_0)Y)^*=\overline{\lambda}_0^2Y^*(\lambda_0^2C^*-\lambda_0R+C).$$

It follows that Y forms an orthonormal basis for left eigenvectors of $P(\lambda)$ corresponding to λ_0. From (25), we obtain that the column vectors of

$$\mathcal{Y}_R=\begin{bmatrix} Y \\ RY-\lambda_0C^*Y \end{bmatrix} \text{ and } \mathcal{Y}_L=\begin{bmatrix} Y \\ -\overline{\lambda}_0Y \end{bmatrix}$$

form a basis of left and right eigenspaces of $\mathcal{M}-\lambda\mathcal{L}$ corresponding to λ_0, respectively. Since λ_0 is semi-simple, the matrix

$$[Y^*,-\lambda_0Y^*]\mathcal{L}\begin{bmatrix} Y \\ RY-\lambda_0C^*Y \end{bmatrix}=-Y^*(2\lambda_0C^*-R)Y=-Y^*P'(\lambda_0)Y$$

is nonsingular. Let

$$\widetilde{\mathcal{Y}}_R=-\mathcal{Y}_R\left(Y^*P'(\lambda_0)Y\right)^{-1},\quad \widetilde{\mathcal{Y}}_L=\mathcal{Y}_L.$$

Then we have

$$\widetilde{\mathcal{Y}}_L^*\mathcal{L}\widetilde{\mathcal{Y}}_R=I_{m_0} \text{ and } \widetilde{\mathcal{Y}}_L^*\mathcal{M}\widetilde{\mathcal{Y}}_R=\lambda_0I_{m_0}. \tag{27}$$

For $\eta>0$ sufficiently small, we let

$$\mathcal{M}_\eta=\begin{bmatrix} C+\mathrm{i}\eta D & 0 \\ R+\mathrm{i}\eta P & -I \end{bmatrix},\quad \mathcal{L}_\eta=\begin{bmatrix} 0 & I \\ C^*+\mathrm{i}\eta D^* & 0 \end{bmatrix}.$$

Then $\mathcal{M}_\eta - \lambda\mathcal{L}_\eta$ is a linearization of $P_\eta(\lambda)$. By (27) and [16, Chapter VI, Theorem 2.12] there are $\widehat{\mathcal{Y}}_R$ and $\widehat{\mathcal{Y}}_L$ such that $\left[\widetilde{\mathcal{Y}}_R\ \widehat{\mathcal{Y}}_R\right]$ and $\left[\widetilde{\mathcal{Y}}_L\ \widehat{\mathcal{Y}}_L\right]$ are nonsingular and

$$\begin{bmatrix} \widetilde{\mathcal{Y}}_L^* \\ \widehat{\mathcal{Y}}_L^* \end{bmatrix} \mathcal{M} \left[\widetilde{\mathcal{Y}}_R\ \widehat{\mathcal{Y}}_R\right] = \begin{bmatrix} \lambda_0 I_{m_0} & 0 \\ 0 & \widehat{\mathcal{M}} \end{bmatrix}, \quad \begin{bmatrix} \widetilde{\mathcal{Y}}_L^* \\ \widehat{\mathcal{Y}}_L^* \end{bmatrix} \mathcal{L} \left[\widetilde{\mathcal{Y}}_R\ \widehat{\mathcal{Y}}_R\right] = \begin{bmatrix} I_{m_0} & 0 \\ 0 & \widehat{\mathcal{L}} \end{bmatrix}.$$

Then, by [16, Chapter VI, Theorem 2.15] the column vectors of $\widetilde{\mathcal{Y}}_R + O(\eta)$ span the right eigenspace of $(\mathcal{M}_\eta, \mathcal{L}_\eta)$ corresponding to (Λ, I_{m_0}), where

$$\Lambda = \left((\lambda_0 I_{m_0} + \eta E_{11} + O(\eta^2)\right)\left(I_{m_0} + \eta F_{11} + O(\eta^2)\right)^{-1}$$

with

$$E_{11} = \widetilde{\mathcal{Y}}_L^* \begin{bmatrix} \mathrm{i}D & 0 \\ \mathrm{i}P & 0 \end{bmatrix} \widetilde{\mathcal{Y}}_R = Y^*(\lambda_0 \mathrm{i}P - \mathrm{i}D)Y\left(Y^*P'(\lambda_0)Y\right)^{-1},$$
$$F_{11} = \widetilde{\mathcal{Y}}_L^* \begin{bmatrix} 0 & 0 \\ \mathrm{i}D^* & 0 \end{bmatrix} \widetilde{\mathcal{Y}}_R = Y^*(\lambda_0 \mathrm{i}D^*)Y\left(Y^*P'(\lambda_0)Y\right)^{-1}.$$

Thus

$$\Lambda = \lambda_0 I_{m_0} + \eta(E_{11} - \lambda_0 F_{11}) + O(\eta^2) = \lambda_0 I_{m_0} - \eta\lambda_0 W + O(\eta^2),$$

where

$$W = Y^*(P - \lambda_0 D^* - \lambda_0^{-1} D)Y\left(\mathrm{i}Y^*P'(\lambda_0)Y\right)^{-1}. \tag{28}$$

The matrix $Z = \mathrm{i}Y^*P'(\lambda_0)Y = \mathrm{i}Y^*(2\lambda_0 C^* - R)Y$ in (28) is Hermitian since

$$Z - Z^* = \mathrm{i}Y^*(2\lambda_0 C^* + 2\overline{\lambda}_0 C - 2R)Y = 2\mathrm{i}\overline{\lambda}_0 Y^* P(\lambda_0)Y = 0.$$

Since $Y^*(P - \lambda_0 D^* - \lambda_0^{-1} D)Y$ is positive definite, all eigenvalues of W in (28) are real and there are m_0 linearly independent eigenvectors. Let d_j for $j = 1, \ldots, \ell$ be the distinct eigenvalues of W with multiplicities m_0^j, and let $\xi_j \in \mathbb{C}^{m_0 \times m_0^j}$ form an orthonormal basis of the eigenspace corresponding to d_j. Then we have

$$\Phi^{-1}\Lambda\Phi = \lambda_0 I_{m_0} - \eta\lambda_0 \mathrm{diag}\left(d_1 I_{m_0^1}, \ldots, d_\ell I_{m_0^\ell}\right) + O(\eta^2).$$

where $\Phi = [\xi_1, \ldots, \xi_\ell] \in \mathbb{C}^{m_0 \times m_0}$. Then for each $j \in \{1, 2, \ldots, \ell\}$, the perturbed eigenvalues $\lambda_{j,\eta}^{(k)}$, $k = 1, \ldots, m_0^j$, and a basis of the corresponding invariant subspace of $\mathcal{M}_\eta - \lambda\mathcal{L}_\eta$ with $\lambda_{j,\eta}^{(k)}|_{\eta=0} = \lambda_0$ can be expressed by

$$\lambda_{j,\eta}^{(k)} = \lambda_0 - \lambda_0 d_j \eta + O(\eta^2),\ k = 1, \ldots, m_0^j, \tag{29a}$$

and

$$\zeta_{j,\eta} = \mathcal{Y}_R\left(Y^*(P - \lambda_0 D^* - \lambda_0^{-1} D)Y\right)^{-1} \xi_j + O(\eta). \tag{29b}$$

The equation in (26) follows from (29b). □

For the pencil $(\mathcal{M}, \mathcal{L})$ given by (24), the relation $\mathcal{M}\begin{bmatrix} I \\ X \end{bmatrix} = \mathcal{L}\begin{bmatrix} I \\ X \end{bmatrix} X^{-1}A$ shows that the weakly stabilizing solution X_* of (13) is obtained by $X_* = X_2 X_1^{-1}$, where the colums of $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ form a basis for

the invariant subspace of $(\mathcal{M}, \mathcal{L})$ corresponding to its eigenvalues inside $\mathbb{T}$ and its eigenvalues on $\mathbb{T}$ that would be perturbed to the inside of $\mathbb{T}$ when $(\mathcal{M}, \mathcal{L})$ is perturbed to $(\mathcal{M}_\eta, \mathcal{L}_\eta)$ with $\eta > 0$.

We can use the QZ algorithm to determine this invariant subspace, with the aid of Theorem 7 when all unimodular eigenvalues of $(\mathcal{M}, \mathcal{L})$ are semi-simple. In practice, these unimodular eigenvalues are likely to be simple and the statements in our Theorem 7 can be simplified significantly. However, if 1 (or -1) happens to be an eigenvalue of $(\mathcal{M}, \mathcal{L})$, then it must have even multiplicity because (counting multiplicity) half of eigenvalues at 1 (or -1) will be perturbed to the inside of $\mathbb{T}$ and the other half to the outside. Typically 1 (or -1) will be a double eigenvalue of partial multiplicity 2, and the eigenvector corresponding to it should be used in the computation of X_*.

6. Exploiting sparsity

Subspace methods for finding X_* may be more efficient than the doubling algorithm that finds X_η for a sufficiently small η. However, it is possible for the doubling algorithm to exploit certain sparsity structures in the matrices A, B, Q in (6) while subspace methods couldn't.

For the nano application here and other applications, the matrices A, B, Q are from a semi-infinite block tridiagonal and block Toeplitz matrix, as given in the matrix T in (11). In some situations, the matrix T is block tridiagonal with the matrices on the three diagonals having some periodicity, but is not block Toeplitz when the submatrices in T are of the given sizes. To make T a block tridiagonal *and* block Toeplitz matrix, we would have to partition the matrix T into larger submatrices. To be more precise, the matrix T is given as in (11), and the matrices $A, B, Q \in \mathbb{C}^{n\times n}$ have the following structures.

$$Q = \begin{bmatrix} K_{1,1} & K_{1,2} & & 0 \\ K_{2,1} & K_{2,2} & \ddots & \\ & \ddots & \ddots & K_{p-1,p} \\ 0 & & K_{p,p-1} & K_{p,p} \end{bmatrix}, \quad A = \begin{bmatrix} 0 & K_{p+1,p} \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 \\ K_{p,p+1} & 0 \end{bmatrix}, \tag{30}$$

where $K_{j,j} \in \mathbb{C}^{n_j\times n_j}, K_{p+1,p} \in \mathbb{C}^{n_1\times n_p}, K_{p,p+1} \in \mathbb{C}^{n_p\times n_1}$.

We now use Algorithm 1 to compute the stabilizing solution X_s of (6). We remark that the equation (6) here is more general than the one studied in [10]. That equation arises in the vibration analysis of fast trains.

As in [10], the complexity of Algorithm 1 can be reduced drastically by using the special structures of the matrices Q, A, B given by (30). Write $Q_k = Q - \widehat{P}_k$. Then it is easily seen from Algorithm 1 that the matrices A_k, B_k, $\widehat{P}_k$ and P_k have the special forms

$$A_k = \begin{bmatrix} 0 & E_k \\ 0 & 0 \end{bmatrix}, \quad B_k = \begin{bmatrix} 0 & 0 \\ F_k & 0 \end{bmatrix}, \quad \widehat{P}_k = \begin{bmatrix} 0 & 0 \\ 0 & \widehat{G}_k \end{bmatrix}, \quad P_k = \begin{bmatrix} G_k & 0 \\ 0 & 0 \end{bmatrix},$$

where E_k, F_k, $\widehat{G}_k$ and G_k are $n_1 \times n_p$, $n_p \times n_1$, $n_p \times n_p$ and $n_1 \times n_1$ matrices, respectively. Algorithm 1 can be rewritten as the following simplified algorithm.

Algorithm 2. *Let* $E_0 = K_{p+1,p},\ F_0 = K_{p,p+1},\ \widehat{G}_0 = 0,\ G_0 = 0.$

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

$$\begin{bmatrix} S_{k,1} \\ S_{k,2} \\ \vdots \\ S_{k,p} \end{bmatrix} = \left(Q - \begin{bmatrix} G_k & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \\ & & & & \widehat{G}_k \end{bmatrix} \right)^{-1} \begin{bmatrix} E_k \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{31}$$

$$\begin{bmatrix} T_{k,1} \\ T_{k,2} \\ \vdots \\ T_{k,p} \end{bmatrix} = \left(Q - \begin{bmatrix} G_k & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \\ & & & & \widehat{G}_k \end{bmatrix} \right)^{-1} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ F_k \end{bmatrix}, \tag{32}$$

where $S_{k,i} \in \mathbb{C}^{n_i \times n_p}$ and $T_{k,i} \in \mathbb{C}^{n_i \times n_1}$, and then compute

$$E_{k+1} = E_k S_{k,p},\ F_{k+1} = F_k T_{k,1},\ \widehat{G}_{k+1} = \widehat{G}_k + F_k S_{k,1},\ G_{k+1} = G_k + E_k T_{k,p}. \tag{33}$$

The main task of Algorithm 2 is to solve the large sparse linear systems in (31) and (32). This could be done by using the Sherman–Morrison–Woodbury formula, as in [10]. But here we present a new approach that is both simpler and less expensive.

Let

$$P = \begin{bmatrix} I_{n_1} & 0 & 0 \\ 0 & 0 & I_{n_p} \\ 0 & I_{n-n_1-n_p} & 0 \end{bmatrix}$$

be a permutation matrix and note that

$$P \left(Q - \begin{bmatrix} G_k & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \\ & & & & \widehat{G}_k \end{bmatrix} \right) P^\top = \left[\begin{array}{cc|c} K_{1,1} - G_k & 0 & \mathcal{V} \\ 0 & K_{p,p} - \widehat{G}_k & \\ \hline \multicolumn{2}{c|}{\mathcal{U}} & \mathcal{C} \end{array}\right],$$

where

$$\mathcal{V} = \begin{bmatrix} K_{1,2} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & K_{p,p-1} \end{bmatrix}, \tag{34}$$

$$\mathcal{U} = \begin{bmatrix} K_{2,1} & 0 \\ 0 & \vdots \\ \vdots & 0 \\ 0 & K_{p-1,p} \end{bmatrix}, \quad \mathcal{C} = \begin{bmatrix} K_{2,2} & K_{2,3} & & 0 \\ K_{3,2} & K_{3,3} & \ddots & \\ & \ddots & \ddots & K_{p-2,p-1} \\ 0 & & K_{p-1,p-2} & K_{p-1,p-1} \end{bmatrix}. \tag{35}$$

Then the matrices $S_{k,1}$, $S_{k,p}$, $T_{k,1}$ and $T_{k,p}$ of the solutions of (31) and (32) satisfy

$$\left(\begin{bmatrix} K_{1,1} - G_k & 0 \\ 0 & K_{p,p} - \widehat{G}_k \end{bmatrix} - \mathcal{V}\mathcal{C}^{-1}\mathcal{U} \right) \begin{bmatrix} S_{k,1} & T_{k,1} \\ S_{k,p} & T_{k,p} \end{bmatrix} = \begin{bmatrix} E_k & 0 \\ 0 & F_k \end{bmatrix}. \tag{36}$$

Note that the matrix $\mathcal{V}\mathcal{C}^{-1}\mathcal{U}$ is independent of k. Since $Q_k = Q - \widehat{P}_k$ and $\lim_{k\to\infty} Q_k = X_s$, we know that X_s is obtained from Q by replacing $K_{p,p}$ in the lower-right corner with $K_{p,p} - \widehat{G}_*$, where $\widehat{G}_* = \lim_{k\to\infty} \widehat{G}_k$.

The following algorithm gives a more detailed implementation of Algorithm 2 and computes the stabilizing solution X_s of (6).

Algorithm 3. *Computation of X_s.*
Input: $K_{j,j} \in \mathbb{C}^{n_j \times n_j}$, $K_{j,j+1} \in \mathbb{C}^{n_j \times n_{j+1}}$, $K_{j+1,j} \in \mathbb{C}^{n_{j+1} \times n_j}$, $j = 1, \ldots, n$, where $n_{p+1} = n_1$; tolerance τ.
Output: The stabilizing solution X_s of (6), where A, B, Q are given by (30).

Take $\mathcal{V}$,$\mathcal{U}$ and $\mathcal{C}$ in (34) and (35);

Compute $W = \begin{bmatrix} K_{1,1} & 0 \\ 0 & K_{p,p} \end{bmatrix} - \mathcal{V}\mathcal{C}^{-1}\mathcal{U}$;

$E_0 = K_{p+1,p}$, $F_0 = K_{p,p+1}$, $\widehat{G}_0 = 0$, $G_0 = 0$;

For $k = 0, 1, \ldots$

$$\begin{bmatrix} S_{k,1} & T_{k,1} \\ S_{k,p} & T_{k,p} \end{bmatrix} = \left(W - \begin{bmatrix} G_k & 0 \\ 0 & \widehat{G}_k \end{bmatrix} \right)^{-1} \begin{bmatrix} E_k & 0 \\ 0 & F_k \end{bmatrix};$$

$E_{k+1} = E_k S_{k,p}$, $F_{k+1} = F_k T_{k,1}$, $\widehat{G}_{k+1} = \widehat{G}_k + F_k S_{k,1}$,
$G_{k+1} = G_k + E_k T_{k,p}$;

If $\|F_k S_{k,1}\| \le \tau \|\widehat{G}_k\|$ *and* $\|E_k T_{k,p}\| \le \tau \|G_k\|$, *then*

$X_s \leftarrow Q$, $X_s(n - n_p + 1 : n, n - n_p + 1 : n) \leftarrow K_{p,p} - \widehat{G}_{k+1}$,
and stop.

In nano applications, we typically need the $n_1 \times n_1$ matrix in the upper-left corner of T^{-1}, where T is given by (11). We also know that X_s^{-1} is the $n \times n$ matrix in the upper-left corner of T^{-1}. So we are mainly interested in the $n_1 \times n_1$ matrix Y_1 in the upper-left corner of X_s^{-1}. Note that Y_1 is the same as the matrix $S_{k,1}$ in (31) *when* $G_k, \widehat{G}_k, E_k$ in (31) are replaced by $0, \widehat{G}_*, I_{n_1}$, respectively. Thus

$$Y_1 = [I_{n_1}, 0] \left(W - \begin{bmatrix} 0 & 0 \\ 0 & \widehat{G}_* \end{bmatrix} \right)^{-1} \begin{bmatrix} I_{n_1} \\ 0 \end{bmatrix},$$

where the matrix W has already been computed in Algorithm 3.

7. Numerical results

In this section we present some numerical results. We use the doubling algorithm to compute the stabilizing solution X_η of the equation

$$X + B_\eta X^{-1} A_\eta = Q_\eta, \tag{37}$$

where A_η, B_η, Q_η are given in (14). If these matrices have the special sparsity stuctures in (30), then Algorithm 3 is used. To measure the accuracy of a computed stabilizing solution X_η to (37), we use the relative residual

$$\mathrm{RRes}_\eta = \frac{\|X_\eta + B_\eta X_\eta^{-1} A_\eta - Q_\eta\|}{\|X_\eta\| + \|A_\eta\|\|B_\eta\|\|X_\eta^{-1}\| + \|Q_\eta\|},$$

where $\|\cdot\|$ is the spectral norm. To see whether X_η is a good approximation to the weakly stabilizing solution X_* of the equation $X + C^* X^{-1} C = R$, we compute

$$\mathrm{RRes} = \frac{\|X_\eta + C^* X_\eta^{-1} C - R\|}{\|X_\eta\| + \|C\|^2\|X_\eta^{-1}\| + \|R\|}. \tag{38}$$

We also use the QZ algorithm to compute X_* directly, and the relative residual RRes_0 is defined as in (38), with the computed X_η replaced by the computed X_*.

Example 1. We randomly generate two complex matrices C, D and two complex Hermitian matrices R, P of dimension 6. Let ϱ be the minimal eigenvalue of P and set

$$P := P + (2\|D\| - \varrho)I_6.$$

We verify that $P + \lambda D^* + \lambda^{-1}D$ is positive definite for all $\lambda \in \mathbb{T}$. We then compute the stabilizing solution X_η of (37) with $\eta = 10^{-4}$, 10^{-8}, 10^{-12}, respectively, by using Algorithm 1. In each case, Algorithm 1 is stopped when $\max\{\|A_{k+1}\|, \|B_{k+1}\|\} < 10^{-10}$ and Q_{k+1} is taken to be the computed X_η. When $\eta = 0$, $P_0(\lambda) = \lambda^2 C^* - \lambda R + C$ has $2m = 4$ eigenvalues on $\mathbb{T}$, given by

$$\Lambda = \{-0.9026 + 0.4304\mathrm{i},\ 0.5687 - 0.8226\mathrm{i},\ 0.9891 + 0.1472\mathrm{i},\ 0.1960 + 0.9806\mathrm{i}\}.$$

By Theorem 7 we determine that

$$\Lambda^s = \{0.5687 - 0.8226\mathrm{i},\ 0.9891 + 0.1472\mathrm{i}\}$$

is such that the perturbed eigenvalues of $P_\eta(\lambda)$ ($\eta > 0$) associated with each $\lambda^s \in \Lambda^s$ are inside $\mathbb{T}$. Then we compute the weakly stabilizing solution X_* of (37) by using the invariant subspace corresponding to stable eigenvalues and eigenvalues in Λ^s (the QZ algorithm). The numerical results are shown in Table 1.

Table 1: Relative residuals

η	10^{-4}	10^{-8}	10^{-12}	0
RRes_η	3.36×10^{-16}	4.03×10^{-15}	4.24×10^{-15}	3.09×10^{-16}

We know that $X_{\eta,I} = \frac{1}{2\mathrm{i}}(X_\eta - X_\eta^*)$ is positive definite for $\eta > 0$ and we know from Theorem 5 that $\mathrm{rank}(X_{*,I}) \le m = 2$. These are confirmed by the numerical results shown in Table 2, where $X_{0,I} = X_{*,I}$.

Table 2: The eigenvalues of $X_{\eta,I}$

η	The eigenvalues of $X_{\eta,I}$
10^{-4}	$2.3555,\ 1.2676,\ 5.87 \times 10^{-3},\ 1.89 \times 10^{-3},\ 1.21 \times 10^{-3},\ 1.10 \times 10^{-3}$
10^{-8}	$2.3510,\ 1.2639,\ 5.91 \times 10^{-7},\ 1.89 \times 10^{-7},\ 1.21 \times 10^{-7},\ 1.10 \times 10^{-7}$
10^{-12}	$2.3510,\ 1.2639,\ 5.91 \times 10^{-11},\ 1.89 \times 10^{-11},\ 1.21 \times 10^{-11},\ 1.10 \times 10^{-11}$
0	$2.3510,\ 1.2639,\ 5.41 \times 10^{-15},\ 1.61 \times 10^{-15},\ -5.12 \times 10^{-16},\ -4.18 \times 10^{-15}$

Example 2. We consider a semi-infinite Hamiltonian operator of the transverse magnetic (TM) mode for the two-dimensional photonic crystal of the form [5]

$$\begin{aligned} H(u, \vec{k}, \vec{x}) &= -\frac{1}{\varepsilon(\vec{x})}\left(\nabla + \mathrm{i}\vec{k}\right) \cdot \left(\nabla + \mathrm{i}\vec{k}\right) u(\vec{x}) \\ &= -\frac{1}{\varepsilon(\vec{x})}\left(\Delta + 2\mathrm{i}\vec{k} \cdot \nabla - \|\vec{k}\|^2\right) u(\vec{x}), \end{aligned} \tag{39}$$

where $\vec{k} = (k_1, k_2)$ is a wave number in the first Brillouin zone $\Omega^* = (-\pi, \pi]^2$, $\vec{x} \in \Omega = [-0.5, \infty) \times [-0.5, 0.5] = \Omega_1 \cup \Omega_2$ with

$$\begin{cases} \Omega_1 = \bigcup_{j=0}^{\infty} B_\rho(j), \\ \Omega_2 = \bigcup_{j=0}^{\infty} \left([-0.5 + j, 0.5 + j] \times [-0.5, 0.5]\right) \setminus B_\rho(j) \end{cases}$$

and $B_\rho(j) = \{(x_1, x_2)|(x_1 - j)^2 + x_2^2 \le \rho^2\}$, $0 < \rho < 0.5$, and $\varepsilon(\vec{x})$ is the dielectric function with

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

See Figure 1 for an illustration of the domain Ω. By Bloch's theorem, we assume that the boundary

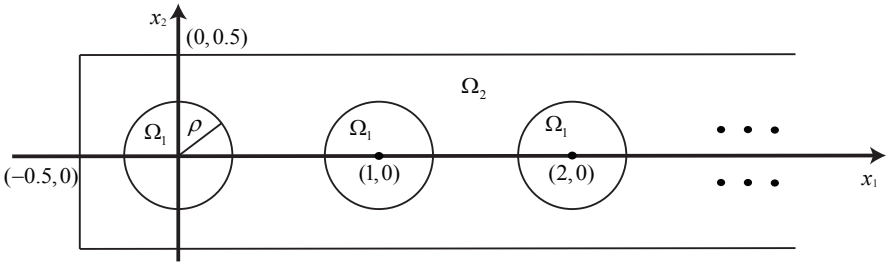

Figure 1: The domain $\Omega = \Omega_1 \cup \Omega_2$.

conditions are given by

$$\begin{cases} u(\vec{x}) = 0, \quad \vec{x} \in \{(-0.5, x_2)|x_2 \in [-0.5, 0.5]\}, \\ u(x_1, 0.5) = e^{\mathrm{i}k_2} u(x_1, -0.5), \ x_1 \in [-0.5, \infty]. \end{cases}$$

We use the classical five-point central finite difference method to discretize the operator (39) on the uniform grid points in Ω with mesh size $h = 1/n$. So n is the number of grid points on the x_2 axis in $[-0.5, 0.5)$. Let T_n be the tridiagonal matrix of dimension n with 4 on the main diagonal and -1 on the two adjacent diagonals and let D_n be the tridiagonal matrix of dimension n with 0 on the main diagonal and -1 and 1 on the supper-diagonal and the sub-diagonal, respectively. Let

$$\Phi = \frac{1}{h^2}(T_n - \delta e_1 e_n^\top - \bar{\delta} e_n e_1^\top) - \frac{\mathrm{i}k_2}{h}(D_n + \delta e_1 e_n^\top - \bar{\delta} e_n e_1^\top) + (k_1^2 + k_2^2) I_n$$

and

$$\Psi = \left(-\frac{1}{h^2} - \frac{\mathrm{i}k_1}{h}\right) I_n,$$

where $\delta = e^{\mathrm{i}k_2}$ and e_j denotes the jth column vector of the identity matrix. Then the system Hamiltonian H from the operator (39) is a semi-infinite block tridiagonal matrix with H_b on the main diagonal and H_{bb} and H_{bb}^* on the supper-diagonal and the sub-diagonal, respectively. The block matrices H_b and H_{bb} are of the forms

$$H_b = \begin{bmatrix} H_{1,1} & H_{1,2} & & \\ H_{1,2}^* & H_{2,2} & \ddots & \\ & \ddots & \ddots & H_{n-1,n} \\ & & H_{n-1,n}^* & H_{n,n} \end{bmatrix}, \quad H_{bb} = \begin{bmatrix} 0 & 0 \\ H_{n,n+1} & 0 \end{bmatrix} \in \mathbb{C}^{n^2 \times n^2},$$

where

$$H_{j,j} = \Gamma_j \Phi \Gamma_j, \ H_{j,j+1} = \Gamma_j \Psi \Gamma_{j+1}, \ j = 1, \ldots, n,$$

and $\Gamma_j = \mathrm{diag}(\Upsilon(:, j))$, $\Upsilon = [\Upsilon_{ij}] \in \mathbb{R}^{n \times n}$ with

$$\begin{cases} \Upsilon_{ij} = \sqrt{\frac{1}{\varepsilon_1}}, & (-0.5 + jh, 0.5 - ih) \in B_\rho(0), \\ \Upsilon_{ij} = \sqrt{\frac{1}{\varepsilon_2}}, & (-0.5 + jh, 0.5 - ih) \notin B_\rho(0). \end{cases}$$

We now apply the Green's function approach to the system Hamiltonian H with the overlap matrix being the identity. So we need to determine the $n^2 \times n^2$ block (particularly the $n \times n$ block) in the upper-left corner of the inverse of the matrix (3) (now with $S_b = I$ and $S_{bb} = 0$). This is done by solving the matrix equation (37). The matrices $A_\eta,\ B_\eta,\ Q_\eta \in \mathbb{C}^{n^2 \times n^2}$ in (37) now have the structures in (30), with $n_j = p = n$ and

$$K_{j,j} = zI_n - H_{j,j}, \quad K_{j,j+1} = -H_{j,j+1}, \quad K_{j+1,j} = -H^*_{j,j+1}, \quad j = 1, \ldots, n,$$

where $z = \mathcal{E} + \mathrm{i}\eta$ with $\mathcal{E} \in \mathbb{R}$ and $0 \leq \eta \ll 1$. We remark that the matrix equation here is a special case of (1) with $P = I$, $D = 0$, $C = A_\eta = B^*_\eta$ and $R = \mathcal{E}I - H_b$.

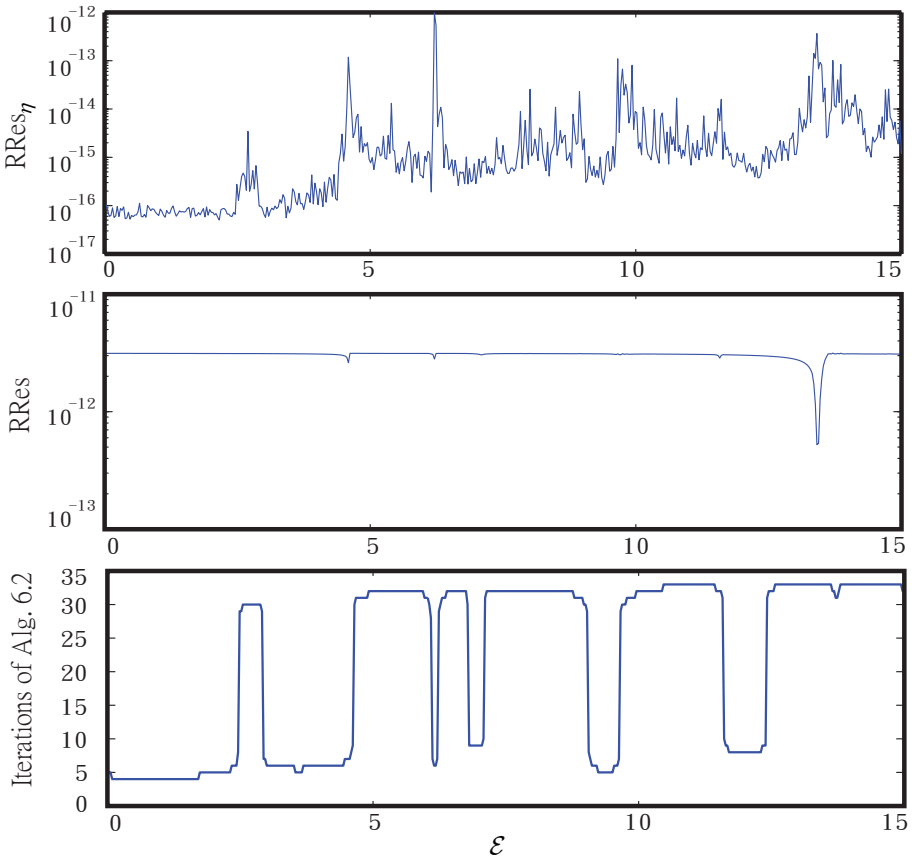

Figure 2: *RRes*$_\eta$, *RRes* and the number of iterations of Algorithm 3.

We now use Algorithm 3 to compute the solution X_η of (37). In our test we take $n = 50$, $\rho = 0.3$, $\varepsilon_1 = 1$, $\varepsilon_2 = 10$ and $(k_1, k_2) = (0.5, 0.7)$. We divide $[0, 15]$ into κ subintervals using $\kappa + 1$ equally spaced nodes $\mathcal{E}_i$, $i = 0, 1, \ldots, \kappa$. We now choose $\kappa = 500$ and run Algorithm 3 with $\eta = 10^{-8}$ and $\tau = 10^{-8}$ for each $\mathcal{E}_i$. In Figure 2, we plot the relative residuals $(\mathrm{RRes}_\eta, \mathrm{RRes})$ and the number of iterations of Algorithm 3. We see that very good approximations to X_η and X_* are obtained in no more than 33 iterations. We also determine the interested energy interval $\Delta = \bigcup_{i=1}^{n} \Delta_i$, where Δ_i are given in Theorem 6. The energy values in Δ are precisely those for which the pencil $(\mathcal{M}, \mathcal{L})$, where $\mathcal{M}$ and $\mathcal{L}$ are given in (24), has eigenvalues on $\mathbb{T}$. In Figure 3, we plot the number of eigenvalues of $(\mathcal{M}, \mathcal{L})$ on $\mathbb{T}$ and $\Delta \bigcap [0, 15]$. For this example, the number of such eigenvalues for $\mathcal{E} \in [0, 15]$ is $0, 2$, or 4. For some larger values of $\mathcal{E}$, we find the number of such eigenvalues to be 6, which turns out to be the maximal number for any energy value. As expected from our convergence results, Algorithm 3 requires more iterations when $(\mathcal{M}, \mathcal{L})$ has unimodular eigenvalues, but

it does not matter too much whether the actual number of unimodular eigenvalues is 2, 4, or any other positive even integer.

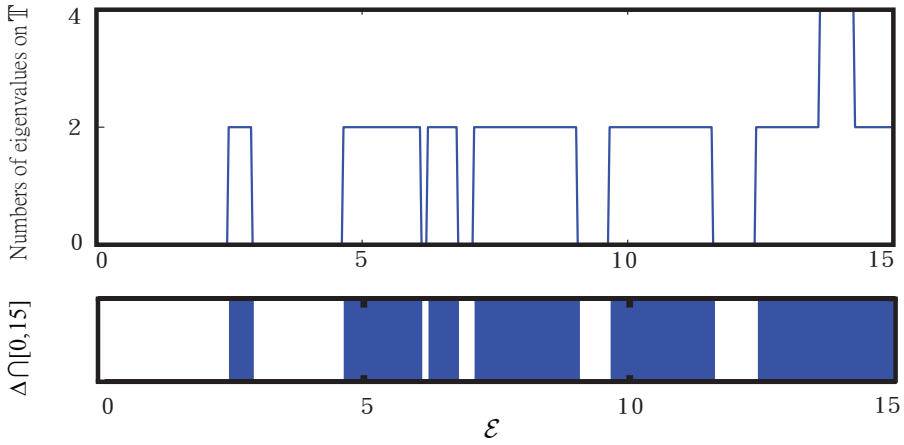

Figure 3: The number of eigenvalues on $\mathbb{T}$ and interested energy interval between 0 to 15.

8. Conclusion

We have introduced a class of nonlinear matrix equations that is wider than those studied earlier in the literature. The main motivation for studying this wider class is from the Green's function approach for treating quantum transport in nano devices. We have characterized the special solution of practical interest. We have shown how the doubling algorithm and subspace methods like the QZ algorithm can be used to find good approximations to the required solution. We have also shown how some special sparsity structures in the coefficent matrices of the equation can be expoited to drastically reduce the complexity of the doubling algorithm for computing the desired solution. The matrix equation from the nano application involves a parameter. At present it is not clear whether the solution computed for one value of the parameter can be used to reduce the computational work of some iterative methods in computing the solution for a nearby value of the parameter, with guaranteed convergence. This could be a topic for further research.

References

[1] I. Appelbaum, T. Wang, J.D. Joannopoulos, 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] C.-Y. Chiang, E.K.-W. Chu, C.-H. Guo, T.-M. Huang, W.-W. Lin, 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) 227–247.

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

[4] C.J. Earle, R.S. Hamilton, A fixed point theorem for holomorphic mappings, Global Analysis (Proc. Sympos. Pure Math., Vol. XVI, Berkeley, Calif., 1968), American Mathematical Society, Rhode Island, 1970, pp. 61–65.

[5] C. Engström, M. Wang, Complex dispersion relation calculations with the symmetric interior penalty method, Internat. J. Numer. Methods Engrg. 84 (2010) 849–863.

[6] I. Gohberg, S. Goldberg, M.A. Kaashoek, Classes of Linear Operators, Vol. II, Oper. Theory Adv. Appl., Vol. 63, Birkhäuser, 1993.

[7] C.-H. Guo, Y.-C. Kuo, W.-W. Lin, On a nonlinear matrix equation arising in nano research, SIAM J. Matrix Anal. Appl., to appear.

[8] C.-H. Guo, Y.-C. Kuo, W.-W. Lin, Complex symmetric stabilizing solution of the matrix equation $X + A^\top X^{-1} A = Q$, Linear Algebra Appl. 435 (2011) 1187–1192.

[9] C.-H. Guo, W.-W. Lin, The matrix equation $X + A^\top X^{-1} A = Q$ and its application in nano research, SIAM J. Sci. Comput. 32 (2010) 3020–3038.

[10] C.-H. Guo, W.-W. Lin, Solving a structured quadratic eigenvalue problem by a structure-preserving doubling algorithm, SIAM J. Matrix Anal. Appl. 31 (2010) 2784–2801.

[11] L.A. Harris, Fixed points of holomorphic mappings for domains in Banach spaces, Abstr. Appl. Anal., 2003, no. 5, pp. 261–274.
[12] D.L. John, D.L. Pulfrey, Green's function calculations for semi-infinite carbon nanotubes, Physica Status Solidi B – Basic Solid State Physics 243 (2006) 442–448.
[13] A. Kletsov, Y. Dahnovsky, 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.
[14] D.S. Mackey, N. Mackey, C. Mehl, V. Mehrmann, Structured polynomial eigenvalue problems: good vibrations from good linearizations, SIAM J. Matrix Anal. Appl. 28 (2006) 1029–1051.
[15] C. van der Mee, G. Rodriguez, S. Seatzu, *LDU* Factorization results for bi-infinite and semi-infinite scalar and block Toeplitz matrices, Calcolo 33 (1996) 307–335.
[16] G.W. Stewart, J.-G. Sun, Matrix Perturbation Theory, Academic Press, 1990.
[17] J. Tomfohr, O.F. Sankey, Theoretical analysis of electron transport through organic molecules, Journal of Chemical Physics 120 (2004) 1542–1554.