

Computing the Full Spectrum of Large Sparse Palindromic Quadratic Eigenvalue Problems arising from Surface Green's Function calculations

Tsung-Ming Huang^{a,*}, Wen-Wei Lin^b, Heng Tian^{c,**}, Guan-Hua Chen^c

^a*Department of Mathematics, National Taiwan Normal University, Taipei, 116, Taiwan*

^b*Department of Applied Mathematics, National Chiao Tung University, Hsinchu 300, Taiwan*

^c*Department of Chemistry, The University of Hong Kong, Hong Kong*

Abstract

Full spectrum of a large sparse \mathbb{T} -palindromic quadratic eigenvalue problem (\mathbb{T} -PQEP) is considered arguably for the first time in this article. Such a problem is posed by calculation of surface Green's functions (SGFs) of mesoscopic transistors with a tremendous non-periodic cross-section. For this problem, general purpose eigensolvers are not efficient, nor is advisable to resort to the decimation method *etc.* to obtain the Wiener-Hopf factorization. After reviewing some rigorous understanding of SGF calculation from the perspective of \mathbb{T} -PQEP and nonlinear matrix equation, we present our new approach to this problem. In a nutshell, the unit disk where the spectrum of interest lies is broken down adaptively into pieces small enough that they each can be locally tackled by the generalized \mathbb{T} -skew-Hamiltonian implicitly restarted shift-and-invert Arnoldi (GTSHIRA) algorithm with suitable shifts and other parameters, and the eigenvalues missed by this divide-and-conquer strategy can be recovered thanks to the accurate estimation provided by our newly developed scheme. Notably the novel non-equivalence deflation is proposed to avoid as much as possible duplication of nearby known eigenvalues when a new shift of GTSHIRA is determined. We demonstrate our new approach by calculating the SGF of a realistic nanowire whose unit cell is described by a matrix of size 4000×4000 at the density functional tight binding level, corresponding to a $8 \times 8 \text{ nm}^2$ cross-section. We believe that quantum transport simulation of realistic nano-devices in the mesoscopic regime will greatly benefit from this work.

Keywords: palindromic quadratic eigenvalue problem, GTSHIRA, non-equivalence deflation, surface Green's function, quantum transport

*Corresponding author

**Corresponding author

Email addresses: min@ntnu.edu.tw (Tsung-Ming Huang), tianheng@yangtze.hku.hk (Heng Tian)

1. Introduction to the \top -PQEP and the SGF calculation

In recent two decades, nonlinear eigenvalue problem gains more and more attention [27] in linear algebra community, of which probably the simplest one is the quadratic eigenvalue problem [33] (QEP),

$$(\lambda^2 A_2 + \lambda A_1 + A_0) x = 0. \quad (1)$$

Of particular interest is the QEP with some structures in A_2, A_1, A_0 , which results in certain symmetry in the spectrum. For example, a QEP is called a gyroscopic QEP (GQEP) [9, 28], if $A_2^\top = A_2, A_1^\top = -A_1, A_0^\top = A_0$, where A^\top denotes transpose of A . This is associated with a gyroscopic system described by a second-order differential equation [9, 28]. A more important QEP we have been working on during the past years is the palindromic QEP (PQEP) [6], with $A_2^\top = A_0, A_1^\top = A_1$ or $A_2^H = A_0, A_1^H = A_1$ as two particular types. Here, A^H denotes Hermitian transpose of A , and ‘palindromic’ refers to an interesting property that the QEP is invariant if the order of the transposed coefficient matrices is reversed, in plain language.

In this article, we only consider PQEP of the kind

$$\mathcal{Q}_p(\lambda)x = (\lambda^2 A^\top - \lambda Q + A) x = 0, \quad Q = Q^\top, \quad (2)$$

where the coefficients matrices $A, Q \in \mathbb{R}^{n \times n}$. We simply call it \top -PQEP. Immediately, we see the important *symplectic* property of the spectrum of \top -PQEP that

$$\lambda \in \sigma(\mathcal{Q}_p(\lambda)) \Leftrightarrow 1/\lambda \in \sigma(\mathcal{Q}_p(\lambda)) \quad (3)$$

where $\sigma(\cdot)$ denotes the spectrum, including 0 and $1/0 := \infty$.

The \top -PQEP is actually not new to quantum physicists [1, 3, 20, 21, 22, 23, 29, 31, 32, 34, 35], but only until recently has its solution as well as its relation to SGF calculation been systematically and rigorously studied [10, 11, 12, 13, 18]. Here we give a brief introduction to the \top -PQEP in the context of SGF calculations in quantum transport simulation only, though it also appears in other fields [7, 15, 16, 36].

In quantum transport setup, a nano-device, say, a nano-transistor illustrated in Fig. 1, made of a molecule or a segment of nanotube/nanowire is contacted with two semi-infinite electrodes on both left and right side. These electrodes are comprised of atoms regularly placed in a unit cell which is repeated to either left or right unidirectionally, periodically and freely. With the help of some atomic basis set, quantum Hamiltonian operator of right electrode is discretized into a semi-infinite block tridiagonal Toeplitz matrix \mathbf{H}_R , while the overlap between basis set is represented by \mathbf{S}_R with the same structure,

$$\mathbf{H}_R = \begin{pmatrix} H_0 & H_1^\top & & & \\ H_1 & H_0 & H_1^\top & & \\ & H_1 & \ddots & \ddots & \\ & & \ddots & \ddots & \ddots \end{pmatrix}, \mathbf{S}_R = \begin{pmatrix} S_0 & S_1^\top & & & \\ S_1 & S_0 & S_1^\top & & \\ & S_1 & \ddots & \ddots & \\ & & \ddots & \ddots & \ddots \end{pmatrix}. \quad (4)$$

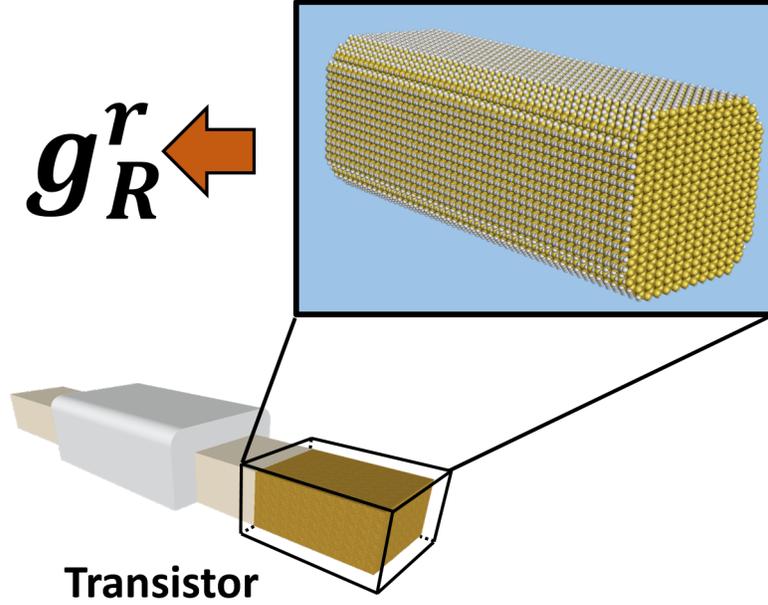


Figure 1: Illustration of a nano-transistor and its electrodes

Let the size of H_0, S_0, H_1, S_1 be $n \times n$. In addition, \mathbf{H}_R is usually real symmetric while \mathbf{S}_R is real symmetric positive-definite. The left electrode is similar. Due to short-ranged interaction between different unit cell, the so-called retarded SGF $\mathbf{g}_R^r(\omega)$ defined as

$$\mathbf{g}_R^r(\omega) = \lim_{\eta \rightarrow 0^+} \left[((\omega + i\eta)\mathbf{S}_R - \mathbf{H}_R)^{-1} \right]_{11}, \Im\omega \geq 0, \quad (5)$$

i.e. the top left block of size $n \times n$ of the inverse of the semi-infinite matrix, is sufficient to capture atomistic details of electrodes necessary for the simulation of quantum transport.

In this work, we only consider $\mathbf{g}_R^r(\omega)$ with $\omega \in \mathbb{R}$. Here we always assume that the condition number of $\mathbf{g}_R^r(\omega + i\eta)$ is always bounded whenever $\eta > 0$, thus, $\mathbf{g}_R^r(\omega)$ is invertible, despite some physically sensible exceptions. After applying Schur complement, we readily have

$$\mathbf{g}_R^r(\omega) = [\omega S_0 - H_0 - (\omega S_1 - H_1)^\top \mathbf{g}_R^r(\omega) (\omega S_1 - H_1)]^{-1}.$$

From here on, if it is necessary to emphasize the presence of the parameter ω in our problem, we will introduce subscription ω into some related quantities. Let $A_\omega = \omega S_1 - H_1$, $Q_\omega = \omega S_0 - H_0$, then $X_\omega = [\mathbf{g}_R^r(\omega)]^{-1}$ satisfies the nonlinear matrix equation (NME)

$$X + A_\omega^\top X^{-1} A_\omega = Q_\omega. \quad (6)$$

X_ω is also the unique solution [10, 11, 12, 13] to NME (6). Note that $X_\omega^{-1}A_\omega$ is called transfer matrix T_ω by quantum physicists. Requirement that T_ω^m be bounded for any $m \in \mathbb{N}$ leads to $\rho(X_\omega^{-1}A_\omega) \leq 1$, where $\rho(\cdot)$ denotes the spectral radius. We call such solution X_ω weakly stabilizing.

Although the solution X is unknown so far, the eigenvalue problem (EP) $X^{-1}A_\omega x = \lambda x$, which can be transformed to the \top -PQEP,

$$0 = (A_\omega - \lambda X)x = A_\omega x - \lambda(Q_\omega - A_\omega^\top X^{-1}A_\omega)x = \mathcal{Q}_{p,\omega}(\lambda)x \quad (7)$$

with $\mathcal{Q}_{p,\omega}(\lambda) := \lambda^2 A_\omega^\top - \lambda Q_\omega + A_\omega$, is well-defined. If this \top -PQEP can be solved, then the eigendecomposition of T_ω is known and we have $X^{-1} = (Q_\omega - A_\omega^\top T_\omega)^{-1}$. This is essentially the physicists' approach to SGF in Refs. [1, 3, 20, 21, 22, 23, 29, 31, 32, 34, 35]. Conversely, \top -PQEPs can be solved by the following solvent approach. Since any solution X to NME

$$X + A^\top X^{-1}A = -Q \quad (8)$$

leads to the Wiener-Hopf factorization of \top -PQEP (2),

$$\mathcal{Q}_p(\lambda) = (\lambda A^\top + X)X^{-1}(\lambda X + A),$$

the \top -PQEP (2) is reduced into two generalized eigenvalue problems (GEPs) for matrix pencils $X - \lambda(-A^\top)$ and $A - \lambda(-X)$, which are well studied.

For \top -PQEPs of small or moderate size, to compute the solution X , either the block cyclic reduction (BCR) method [2], which is called decimation or renormalization method [8, 30] in physics community, or the structured doubling algorithm (SDA) [26], is the first choice. Unfortunately, as the size of the \top -PQEP is multiplied and the sparsity of A, Q emerges, e.g. when the cross-section of the nano-device has become tremendous [4, 14], the performance of BCR *etc.* on a small computer cluster (e.g. no more than 100 cores) degrades dramatically. Even worse, it can happen that BCR *etc.* converges just linearly instead of quadratically or fails to compute the desired solution in some critical situations [5]. Therefore, the importance of directly solving the spectrum of the large sparse \top -PQEP stands out.

If only a very small subset of eigenvalues of the large sparse \top -PQEP is desired, the generalized \top -skew-Hamiltonian implicitly restarted shift-and-invert Arnoldi (G \top SHIRA) algorithm proposed in previous works [16, 18, 19] serves the purpose. But, as mentioned above, computation of X amounts to computation of the full spectrum of the \top -PQEP. Therefore, in this article we continue focusing on the \top -PQEP (2) whose coefficient matrices A, Q are large, sparse and real, and mainly aim at the *full* spectrum of such \top -PQEP using divide-and-conquer strategy, with the G \top SHIRA algorithm as our workhorse. To this end, there are many issues to be resolved. In the following we just enumerate several of them that are critical to the present work.

- How should the whole region be divided where the spectrum may be distributed, especially when the exact or approximate distribution can not be known in advance?

- What is the next shift after running the GTSHIRA algorithm with the present shift, if one does not want to miss too many target eigenvalues?
- How to reduce the influence of the converged eigenvalues close to the new shift? And how can we avoid duplicating eigenvalues just got?
- What to do with the eigenvalues left out by our spectrum slicing strategy, if they exist?

The solutions to all these questions constitute our main contributions in the present work. Namely,

- Novel non-equivalence deflation method for \mathbb{T} -PQEP is proposed (cf. Sec. 4) to push some converged eigenvalues to a remote place while keeping the rest intact.
- The region of concern of the eigenvalues is the unit disk and is partitioned by a series of concentric circles with shrinking radius. We propose a rule of thumb to determine the number of target eigenvalues, the radius, the shift for each step of calculation and which of the converged eigenvalues to be deflated (cf. Sec. 5).
- After one sweep of all chosen circles, location of missing eigenvalues can be well approximated by our newly developed scheme (cf. Sec. 5.5). On the basis of these approximations, the missing eigenvalues and the associated eigenvectors will be easily recovered.

These techniques are valuable in their own right. For audience who are already familiar with \mathbb{T} -PQEP, they can just jump to relevant sections of this article to acquaint themselves with these innovations. For a layperson, it is better to start with the following definitions which frequently appear in this work.

- (i) $\mathcal{J}_{2n} = \begin{bmatrix} 0_n & I_n \\ -I_n & 0_n \end{bmatrix}$, where I_n is the identity matrix of size $n \times n$ and 0_n is the zero matrix of size $n \times n$.
- (ii) $\mathcal{U} \in \mathbb{C}^{2n \times 2n}$ is called \mathbb{T} -symplectic if $\mathcal{U}^\top \mathcal{J}_{2n} \mathcal{U} = \mathcal{J}_{2n}$; $\mathcal{M} - \lambda \mathcal{L} \in \mathbb{C}^{2n \times 2n}$ is called \mathbb{T} -symplectic if $\mathcal{M} \mathcal{J}_{2n} \mathcal{M}^\top = \mathcal{L} \mathcal{J}_{2n} \mathcal{L}^\top$.
- (iii) $\mathcal{H} \in \mathbb{C}^{2n \times 2n}$ is called \mathbb{T} -Hamiltonian or \mathbb{T} -skew-Hamiltonian if $(\mathcal{H} \mathcal{J}_{2n})^\top = \mathcal{H} \mathcal{J}_{2n}$ or $(\mathcal{H} \mathcal{J}_{2n})^\top = -\mathcal{H} \mathcal{J}_{2n}$, respectively. (Please do not confuse this ‘Hamiltonian’ with the Hamiltonian in quantum or classical mechanics.)
- (iv) $\mathcal{K} - \lambda \mathcal{N} \in \mathbb{C}^{2n \times 2n}$ is called \mathbb{T} -skew-Hamiltonian if both \mathcal{K} and \mathcal{N} are \mathbb{T} -skew-Hamiltonian.
- (v) $X \in \mathbb{C}^{2n \times m}$, $1 \leq m \leq n$, is called \mathbb{T} -isotropic if $X^\top \mathcal{J}_{2n} X = 0_m$; $X, Y \in \mathbb{C}^{2n \times m}$, $1 \leq m \leq n$, are called \mathbb{T} -bi-isotropic if $X^\top \mathcal{J}_{2n} Y = 0_m$.

This article is organized as follows. In Sec. 2, \top -PQEP (2) will be linearized in a structure-preserving manner, and some facts and theorems that are important to the SGF calculation will be recapitulated. Next in Sec. 3, in order to actually solve \top -PQEP (2), $\mathcal{S} + \mathcal{S}^{-1}$ transformation of the \top -symplectic pair will be introduced and the GTSHIRA algorithm for the resulting \top -skew-Hamiltonian pair will be reviewed. After that the non-equivalence deflation method for \top -PQEP (2) will be developed in Sec. 4. Then in Sec. 5, practical implementation of our \top -PQEP solver will be elaborated on. And then in Sec. 6, some numerical results will be provided to demonstrate that our \top -PQEP solver indeed achieves the goal stated in the title. Finally, in Sec. 7, we summarize this work and put forward some related research questions.

2. some distinct features of the solution to NME(6)

It can be verified that X is a solution of (6) if and only if

$$\mathcal{M} \begin{bmatrix} I \\ X \end{bmatrix} = \mathcal{L} \begin{bmatrix} I \\ X \end{bmatrix} (X^{-1}A_\omega), \quad (9a)$$

where

$$\mathcal{M} = \begin{bmatrix} A_\omega & 0_n \\ Q_\omega & -I_n \end{bmatrix} \quad \text{and} \quad \mathcal{L} = \begin{bmatrix} 0_n & I_n \\ A_\omega^\top & 0_n \end{bmatrix}. \quad (9b)$$

It is easy to see that the matrix pair $(\mathcal{M}, \mathcal{L})$ is a \top -symplectic pair, i.e. we have $\mathcal{M}\mathcal{J}_{2n}\mathcal{M}^\top = \mathcal{L}\mathcal{J}_{2n}\mathcal{L}^\top$. On seeing this, we know the eigenvalues of $(\mathcal{M}, \mathcal{L})$ form reciprocal pairs $(\lambda, 1/\lambda)$, where possibly $\lambda = 0, \infty$. From (9), eigenvalues of $X^{-1}A_\omega$ are inevitably those of $(\mathcal{M}, \mathcal{L})$. This means that the spectrum of the \top -PQEP (2) is completely equivalent to that of the \top -symplectic pencil $\mathcal{M} - \lambda\mathcal{L}$. So far as we know, this pencil $\mathcal{M} - \lambda\mathcal{L}$ is one of the few structure-preserving linearizations of the \top -PQEP (7), and arguably the most numerically viable one [18, 19]. In passing, other linearization of the \top -PQEP (7), e.g.

$$\begin{bmatrix} 0_n & I_n \\ -A_\omega & Q_\omega \end{bmatrix} \begin{bmatrix} x \\ \lambda x \end{bmatrix} = \lambda \begin{bmatrix} I_n & 0_n \\ 0_n & A_\omega^\top \end{bmatrix} \begin{bmatrix} x \\ \lambda x \end{bmatrix},$$

which is popular among quantum physicists, breaks the intrinsic symmetry of the spectrum, and therefore is not always reliable.

Because X is the (weakly) stabilizing solution of the NME (6), it follows that the desired solution of NME (6) is determined by $X = X_2X_1^{-1}$, where $[X_1^\top, X_2^\top]^\top$ forms a suitable deflating subspace of the \top -symplectic pencil $\mathcal{M} - \lambda\mathcal{L}$. In Eq. (9), the condition $\rho(X^{-1}A_\omega) \leq 1$ requires that the (right) deflating subspace associated with eigenvalues $\{\lambda : |\lambda| < 1\}$ of $(\mathcal{M}, \mathcal{L})$ be kept and that subspace associated with $\{\lambda : |\lambda| > 1\}$ be discarded. This is sufficient if $\omega \notin \sigma(\mathbf{H}_R, \mathbf{S}_R)$. Otherwise, special care must be taken of the eigenvectors associated with unimodular eigenvalues of $(\mathcal{M}, \mathcal{L})$.

In light of two basic facts that for $\omega \in \mathbb{R}$, $X_\omega = \lim_{\eta \rightarrow 0^+} X_{\omega+i\eta}$, and that the spectrum of $\mathcal{Q}_{p,\omega+i\eta}(\lambda)$ does not intersect the unit circle $\mathbb{T} = \{\lambda \in \mathbb{C}; |\lambda| = 1\}$ for

$\forall \eta > 0$, it is certain that by introducing a small perturbation $i\eta$ to ω with $\eta > 0$, all unimodular eigenvalues of $\mathcal{Q}_{p,\omega}(\lambda)$ will deviate from \mathbb{T} . It is the eigenspaces associated with unimodular eigenvalues which will move inwards due to this $i\eta$ that are needed to determine X . In Ref. [11, 12], we have formulated this observation in the rigorous linear algebra language as follows.

Theorem 1. *Let λ_0 be a semi-simple unimodular eigenvalue of $\mathcal{Q}_{p,\omega}(\lambda)$ with multiplicity m_0 and let an orthonormal basis $Y \in \mathbb{C}^{n \times m_0}$ form the right eigenspace of λ_0 . Then $C = iY^H(2\lambda_0 A_\omega^\top - Q_\omega)Y$ is a nonsingular Hermitian matrix. Denote $B = Y^H(S_0 - \lambda_0 S_1^\top - \lambda_0^{-1} S_1)Y$. Let $d_j, j = 1, 2, \dots, \ell$, be distinct eigenvalues of the Hermitian definite pair (C, B) with multiplicities m_{0j} , and let $\xi_j \in \mathbb{C}^{m_0 \times m_{0j}}$ form an B -orthonormal basis of the eigenspace corresponding to d_j . Then for $\eta > 0$ sufficiently small*

$$\lambda_{j,k,\eta} = \lambda_0 \left(1 - \frac{\eta}{d_j} \right) + \mathcal{O}(\eta^2), \quad k = 1, 2, \dots, m_{0j}, \quad \text{and} \quad y_{j,\eta} = Y\xi_j + \mathcal{O}(\eta) \quad (10)$$

are perturbed eigenvalues and a basis of the corresponding invariant subspaces of $\mathcal{Q}_{p,\omega+i\eta}(\lambda)$, respectively.

This theorem provides us a practical rule to find the eigenspaces, i.e. to find those $y_{j,0}$ associated with positive d_j only.

Independently, physicists have also derived their criterion to pick out the appropriate eigenvectors for determining X from physics perspective [20, 21, 29, 32, 34, 35], though unimodular eigenvalues are usually assumed to be simple. Since their criterion is neither any simpler nor more general than the one stated in this theorem, we will not go into its details here.

What is more interesting is that for $\forall \omega \in \mathbb{R}$, the rank of $\Im(\mathbf{g}_R^r(\omega))$, which is also the rank of $\Im(-X_\omega) = \Im(A_\omega^\top X_\omega^{-1} A_\omega)$, is bounded by the halved number of unimodular eigenvalues of $\mathcal{Q}_{p,\omega}(\lambda)$, as the following theorem [11, 12] says.

Theorem 2. *For $\forall \omega \in \mathbb{R}$, the number of eigenvalues (counting multiplicities) of $\mathcal{Q}_{p,\omega}(\lambda)$ on \mathbb{T} must be even, say, $2m$. Let $X_I = \Im(X_\omega)$, then*

- (i) $\text{rank}(X_I) \leq m$;
- (ii) $\text{rank}(X_I) = m$ if all eigenvalues of $\mathcal{Q}_{p,\omega}(\lambda)$ on \mathbb{T} are semi-simple and $\|X_{\omega+i\eta}^{-1} A_\omega - X_\omega^{-1} A_\omega\|_2 = \mathcal{O}(\eta)$ for $\eta > 0$ sufficiently small;
- (iii) $\text{rank}(X_I) = m$ if all eigenvalues of $\mathcal{Q}_{p,\omega}(\lambda)$ on \mathbb{T} are semi-simple and each unimodular eigenvalue of multiplicity m_j is perturbed to m_j eigenvalues (of $\mathcal{Q}_{p,\omega+i\eta}(\lambda)$) inside \mathbb{T} or to m_j eigenvalues outside \mathbb{T} .

From physics perspective, this theorem is not hard to understand [29]. There is a one-to-one correspondence between the unimodular eigenvalues of $\mathcal{Q}_{p,\omega}(\lambda)$ and the left and right propagating states of the infinite periodic system. And only right propagating states can contribute to the density of states of the right electrode. And $\Im(\mathbf{g}_R^r(\omega))$ directly reflects the density of states for a given energy ω . Hence quantitatively we have the results stated in this theorem.

For the sake of simplicity, we will assume all unimodular eigenvalues of the pencil $\mathcal{M} - \lambda\mathcal{L}$ except ± 1 are semi-simple, while the eigenvalues ± 1 , if they exist, have partial multiplicities 2. In other words, so far we have not taken seriously into consideration the presence of Jordan form. In fact, numerical computation of Jordan form of the pencil $\mathcal{M} - \lambda\mathcal{L}$ is a very challenging issue, and will be discussed separately elsewhere.

3. the G \top SHIRA algorithm for the large sparse \top -PQEP

It is good that the matrix pair $(\mathcal{M}, \mathcal{L})$ in (9b) is a \top -symplectic pair, but to the best of our knowledge, there does not exist any Arnoldi-type algorithm for solving $(\mathcal{M}, \mathcal{L})$ directly that is both structure-preserving and stable. In [18, 19], an ingenious way has been invented to tackle a \top -symplectic pair indirectly, structure-preservingly and stably, which brings us much convenience. Specifically, the \top -symplectic pair $(\mathcal{M}, \mathcal{L})$ is transformed into a \top -skew-Hamiltonian pair $(\mathcal{K}, \mathcal{N})$ after applying the $(\mathcal{S} + \mathcal{S}^{-1})$ -transformation initiated in [24], then the G \top SHIRA algorithm is proposed to solve the resulting \top -skew-Hamiltonian pair. Compared with the usual unstructured Arnoldi algorithm, the G \top SHIRA algorithm can not only preserve the reciprocal property of the spectrum, but also accelerate the convergence when target eigenpairs are computed [19].

The $(\mathcal{S} + \mathcal{S}^{-1})$ -transformation of $(\mathcal{M}, \mathcal{L})$ in Eq. (9b)

$$\begin{aligned} (\mathcal{M}_s, \mathcal{L}_s) &\equiv (\mathcal{M}\mathcal{J}_{2n}\mathcal{L}^\top + \mathcal{L}\mathcal{J}_{2n}\mathcal{M}^\top, \mathcal{L}\mathcal{J}_{2n}\mathcal{L}^\top) \\ &= \left(\begin{bmatrix} A_\omega - A_\omega^\top & -Q_\omega \\ Q_\omega & A_\omega - A_\omega^\top \end{bmatrix}, \begin{bmatrix} 0_n & -A_\omega \\ A_\omega^\top & 0_n \end{bmatrix} \right) \end{aligned}$$

yields a \top -skew-Hamiltonian pair $(\mathcal{K}, \mathcal{N})$

$$(\mathcal{K}, \mathcal{N}) \equiv (\mathcal{M}_s, \mathcal{L}_s)\mathcal{J}_{2n}^\top = \left(\begin{bmatrix} Q_\omega & A_\omega - A_\omega^\top \\ A_\omega^\top - A_\omega & Q_\omega \end{bmatrix}, \begin{bmatrix} A_\omega & 0_n \\ 0_n & A_\omega^\top \end{bmatrix} \right). \quad (11)$$

The intimate connection between the GEP with \top -symplectic structure and that with \top -skew-Hamiltonian structure is unveiled as follows [18].

Theorem 3. *Let $(\mathcal{M}, \mathcal{L})$ be a \top -symplectic pair of size $2n \times 2n$ and $(\mathcal{K}, \mathcal{N})$ be its $(\mathcal{S} + \mathcal{S}^{-1})$ -transformation times \mathcal{J}_{2n}^\top . Then*

- (i) μ is a double eigenvalue of $(\mathcal{K}, \mathcal{N})$ if and only if $\nu, 1/\nu$ are eigenvalues of $(\mathcal{M}, \mathcal{L})$, where $\nu, 1/\nu$ are two roots of the equation $\mu = \lambda + 1/\lambda$.
- (ii) If $z = [z_1^\top, z_2^\top]^\top$ is an eigenvector of $(\mathcal{K}, \mathcal{N})$ corresponding to $\mu = \nu + 1/\nu \neq 2$, i.e. $(\mathcal{K} - \mu\mathcal{N})z = 0$, then

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \equiv \begin{bmatrix} \nu^{-1}z_1 - z_2 \\ -A_\omega z_1 + \nu^{-1}(Q_\omega z_1 - A_\omega^\top z_2) \end{bmatrix}, \quad \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \equiv \begin{bmatrix} \nu z_1 - z_2 \\ -A_\omega z_1 + \nu(Q_\omega z_1 - A_\omega^\top z_2) \end{bmatrix} \quad (12)$$

are the eigenvectors of $(\mathcal{M}, \mathcal{L})$ corresponding to ν and $1/\nu$, respectively.

Remark 1. From ((ii)) in Theorem 3, we know that the eigenvectors $[x_1^\top, x_2^\top]^\top$ of $(\mathcal{M}, \mathcal{L})$ corresponding to ν and $1/\nu$ can be computed from (12). However, it is more convenient to get $x_2 = Q_\omega x_1 - \nu A_\omega^\top x_1$ and $x_2 = Q_\omega x_1 - \nu^{-1} A_\omega^\top x_1$ from $\mathcal{M} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \nu \mathcal{L} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ with $x_1 = \nu^{-1} z_1 - z_2$ and $x_1 = \nu z_1 - z_2$, respectively.

The structure-preserving Arnoldi algorithm, namely, the GTSHIRA algorithm, has been proposed in Ref. [18] to compute a small portion of the spectrum of the large sparse pair $(\mathcal{K}, \mathcal{N})$ around a designated point. For a given $\lambda_0 \notin \sigma(\mathcal{M}, \mathcal{L})$, let $\mu_0 \equiv \lambda_0 + 1/\lambda_0 \notin \sigma(\mathcal{K}, \mathcal{N})$ be the designated point, then the shift-and-invert transformation $(\widehat{\mathcal{K}} - \hat{\mu}\widehat{\mathcal{N}})$ for $(\mathcal{K} - \mu\mathcal{N})$ with $\hat{\mu} = 1/(\mu - \mu_0)$ can be derived,

$$\begin{aligned}\widehat{\mathcal{K}} &\equiv -\lambda_0 \mathcal{N} = -\lambda_0 \mathcal{L} \mathcal{J}_{2n} \mathcal{L}^\top \mathcal{J}_{2n}^\top \\ \widehat{\mathcal{N}} &\equiv -\lambda_0 (\mathcal{K} - \mu_0 \mathcal{N}).\end{aligned}\quad (13)$$

Then GTSHIRA algorithm is then applied to generate a generalized \top -isotropic Arnoldi decomposition with order j :

$$\widehat{\mathcal{K}} Z_j = Y_j H_j + h_{j+1,j} y_{j+1} e_j^\top, \quad (14a)$$

$$\widehat{\mathcal{N}} Z_j = Y_j R_j, \quad (14b)$$

where $H_j \in \mathbb{C}^{j \times j}$ is unreduced upper Hessenberg, $R_j \in \mathbb{C}^{j \times j}$ is nonsingular upper triangular, and Y_j and Z_j are \top -bi-isotropic. With more detailed derivations and the implicitly restarted scheme found in Ref. [18], only the j -th generalized \top -isotropic Arnoldi iteration is stated in Algorithm 1 and using GTSHIRA algorithm to compute ℓ target eigenpairs of $\mathcal{M} - \lambda \mathcal{L}$ is stated in Algorithm 2.

Note that substituting μ_0 above into (13), $\widehat{\mathcal{N}}$ can be factorized into

$$\begin{aligned}\widehat{\mathcal{N}} &= -\lambda_0 \left(\mathcal{L} \mathcal{J}_{2n} \mathcal{M}^\top + \mathcal{M} \mathcal{J}_{2n} \mathcal{L}^\top - \left(\lambda_0 + \frac{1}{\lambda_0} \right) \mathcal{L} \mathcal{J}_{2n} \mathcal{L}^\top \right) \mathcal{J}_{2n}^\top \\ &= (\mathcal{M} - \lambda_0 \mathcal{L}) \mathcal{J}_{2n} (\mathcal{M}^\top - \lambda_0 \mathcal{L}^\top) \mathcal{J}_{2n}^\top.\end{aligned}\quad (15a)$$

From (9b), we have

$$(\mathcal{M} - \lambda_0 \mathcal{L})^{-1} = \begin{bmatrix} \mathcal{Q}_{p,\omega}(\lambda_0)^{-1} & 0_n \\ (\mathcal{Q}_\omega - \lambda_0 A_\omega^\top) \mathcal{Q}_{p,\omega}(\lambda_0)^{-1} & -I_n \end{bmatrix} \begin{bmatrix} I_n & -\lambda_0 I_n \\ 0_n & I_n \end{bmatrix}. \quad (15b)$$

Using factorizations in (15), the linear system $\widehat{\mathcal{N}} z_j = y_j$ in line 1 of Algorithm 1 can be efficiently solved.

4. Non-equivalence deflation method (NEDM) for \top -PQEP

It is known that converged eigenvalues of \top -PQEP (2) will greatly affect the convergence of Ritz pairs of nearby eigenvalues to be calculated via GTSHIRA algorithm. Recently in Ref. [17], a novel NEDM has been proposed for some

Algorithm 1 [18] The j -th generalized \top -isotropic Arnoldi step

Input: \top -skew-Hamiltonian $\widehat{\mathcal{K}}$ and $\widehat{\mathcal{N}}$, upper triangular $R(1:j-1, 1:j-1)$, $Y_j = [y_1, \dots, y_j]$ and $Z_{j-1} = [z_1, \dots, z_{j-1}]$ with $Y_j^* Y_j = I_j$, $Z_{j-1}^* Z_{j-1} = I_{j-1}$ and $Y_j^\top \mathcal{J}_{2n} Z_{j-1} = 0$.

Output: $[h_{1,j}, \dots, h_{j+1,j}]$, $R(1:j, j)$, y_{j+1} and z_j .

- 1: Solve $\widehat{\mathcal{N}} z_j = y_j$;
 - 2: **for** $i = 1, \dots, j-1$ **do**
 - 3: $\widehat{r}_{ij} = z_i^* z_j$, $z_j = z_j - \widehat{r}_{ij} z_i$
 - 4: **end for**
 - 5: Reorthogonalize z_j to $\mathcal{J}_{2n} \bar{Y}_j$ as the for-loop in Steps 6-8 does:
 - 6: **for** $i = 1, \dots, j$ **do**
 - 7: $s_{ij} = y_i^\top \mathcal{J}_{2n}^\top z_j$, $z_j = z_j - s_{ij} \mathcal{J}_{2n} \bar{y}_i$
 - 8: **end for**
 - 9: Set $R(j, j) := \|z_j\|_2^{-1}$, $z_j := R(j, j) z_j$ and
 $R(1:j-1, j) := -R(j, j) R(1:j-1, 1:j-1) [\widehat{r}_{1j}, \dots, \widehat{r}_{j-1,j}]^\top$;
 - 10: Compute $y_{j+1} = \mathcal{K} z_j$;
 - 11: **for** $i = 1, \dots, j$ **do**
 - 12: $h_{ij} = y_i^* y_{j+1}$, $y_{j+1} = y_{j+1} - h_{ij} y_i$
 - 13: **end for**
 - 14: Reorthogonalize y_{j+1} to $\mathcal{J}_{2n} \bar{Z}_j$ as the for-loop in Steps 15-17 does:
 - 15: **for** $i = 1, \dots, j$ **do**
 - 16: $t_{ij} = z_i^\top \mathcal{J}_{2n}^\top y_{j+1}$, $y_{j+1} = y_{j+1} - t_{ij} \mathcal{J}_{2n} \bar{z}_i$
 - 17: **end for**
 - 18: Set $h_{j+1,j} := \|y_{j+1}\|_2$ and $y_{j+1} := y_{j+1} / h_{j+1,j}$.
-

Algorithm 2 G \top SHIRA for solving $\mathcal{M}x = \lambda \mathcal{L}x$

Input: matrices A_ω and Q_ω , nonzero shift λ_0 and the number ℓ of desired eigenvalues.

Output: eigenpairs $\{(\lambda_j, [(x_{1,j}^{(1)})^\top, (x_{2,j}^{(1)})^\top]^\top), (\lambda_j^{-1}, [(x_{1,j}^{(2)})^\top, (x_{2,j}^{(2)})^\top]^\top)\}_{j=1}^\ell$ of (9b).

- 1: Compute eigenpairs $\{(\widehat{\mu}_j, z_j \equiv [z_{1,j}^\top, z_{2,j}^\top]^\top)\}_{j=1}^\ell$ of $(\widehat{\mathcal{K}}, \widehat{\mathcal{N}})$ by G \top SHIRA.
- 2: Compute eigenvalues λ_j and λ_j^{-1} of symplectic pair $(\mathcal{M}, \mathcal{L})$ in (9b) by solving

$$\lambda^2 - (\lambda_0 + \lambda_0^{-1} + \widehat{\mu}_j^{-1})\lambda + 1 = 0;$$

Compute eigenvectors

$$\begin{aligned} x_{1,j}^{(1)} &= \lambda_j^{-1} z_{1,j} - z_{2,j}, & x_{2,j}^{(1)} &= Q_\omega x_{1,j}^{(1)} - \lambda_j A_\omega^\top x_{1,j}^{(1)}, \\ x_{1,j}^{(2)} &= \lambda_j z_{1,j} - z_{2,j}, & x_{2,j}^{(2)} &= Q_\omega x_{1,j}^{(2)} - \lambda_j^{-1} A_\omega^\top x_{1,j}^{(2)} \end{aligned}$$

corresponding to $\lambda_j, \lambda_j^{-1}$, respectively, for $j = 1, 2, \dots, \ell$.

other nonlinear EP than the \top -PQEP (2). Roughly speaking, in this method converged eigenvalues will be moved to infinity, hence pose no more threat, while the rest ones remain unchanged. This motivates us to develop a similar NEDM for \top -PQEP (2).

Assume $\lambda \neq -1$ and let

$$\nu = \frac{\lambda - 1}{\lambda + 1}, \text{ i.e. } \lambda = \frac{1 + \nu}{1 - \nu}, \quad (16a)$$

or assume $\lambda \neq 1$ and let

$$\nu = \frac{\lambda + 1}{\lambda - 1}, \text{ i.e. } \lambda = -\frac{1 + \nu}{1 - \nu}. \quad (16b)$$

For the convenience, (16a) and (16b) can be uniformly written as

$$\nu = \frac{\lambda \mp 1}{\lambda \pm 1}, \text{ i.e. } \lambda = \pm \frac{1 + \nu}{1 - \nu}. \quad (17)$$

Substitution of (17) into (2) leads to a GQEP

$$\mathcal{Q}_g(\nu)x \equiv (1 - \nu)^2 \mathcal{Q}_p \left(\pm \frac{1 + \nu}{1 - \nu} \right) x \equiv (\nu^2 M + \nu G + K) x = 0, \quad (18)$$

with

$$M \equiv A^\top \pm Q + A = M^\top, \quad (19a)$$

$$G \equiv 2A^\top - 2A = -G^\top, \quad (19b)$$

$$K \equiv A^\top \mp Q + A = K^\top. \quad (19c)$$

In passing, it is well known that the spectrum of a GQEP has a Hamiltonian structure, i.e. if $\nu \in \mathbb{C}$ with $\Re\nu \cdot \Im\nu \neq 0$ is an eigenvalue of $\mathcal{Q}_g(\nu)$, then so are $-\nu, \bar{\nu}, -\bar{\nu}$; while if $\nu \in \mathbb{R}$ or $\nu \in i\mathbb{R}$ is an eigenvalue of $\mathcal{Q}_g(\nu)$, then so is $-\nu$.

Let (Λ_m, X_m) with $\Lambda_m \in \mathbb{R}^{m \times m}$, $X_m \in \mathbb{R}^{n \times m}$ and $X_m^\top X_m = I_m$ be an eigenmatrix pair of $\mathcal{Q}_g(\nu)$, i.e.

$$MX_m \Lambda_m^2 + GX_m \Lambda_m + KX_m = 0, \quad (20)$$

where $\sigma(\Lambda_m)$ has the Hamiltonian structure mentioned above. Using (Λ_m, X_m) , we can define a new GQEP as

$$\tilde{\mathcal{Q}}_g(\nu)x \equiv (\nu^2 \tilde{M} + \nu \tilde{G} + \tilde{K}) x = 0, \quad (21)$$

where

$$\tilde{M} \equiv M - MX_m \Theta_m X_m^\top M = \tilde{M}^\top, \quad (22a)$$

$$\tilde{G} \equiv G + (KX_m \Lambda_m^{-1} \Theta_m X_m^\top M - MX_m \Theta_m \Lambda_m^{-\top} X_m^\top K) = -\tilde{G}^\top, \quad (22b)$$

$$\tilde{K} \equiv K + KX_m \Lambda_m^{-1} \Theta_m \Lambda_m^{-\top} X_m^\top K = \tilde{K}^\top, \quad (22c)$$

with

$$\Theta_m = (X_m^\top M X_m)^{-1}. \quad (23)$$

The following important theorem uncovers how the spectrum of $\tilde{\mathcal{Q}}_g(\nu)$ is related to that of $\mathcal{Q}_g(\nu)$ and Λ_m .

Theorem 4. *Let $\mathcal{Q}_g(\nu)$ and $\tilde{\mathcal{Q}}_g(\nu)$ be defined in (18) and (21), respectively. Assume that $\Lambda_m + \Theta_m \Lambda_m^{-\top} (X_m^\top K X_m)$ is nonsingular, then*

$$\sigma(\tilde{\mathcal{Q}}_g(\nu)) = \{\sigma(\mathcal{Q}_g(\nu)) \setminus \sigma(\Lambda_m)\} \cup \underbrace{\{\infty, \dots, \infty\}}_m.$$

Proof. From (20) and (22), we have

$$\begin{aligned} \tilde{\mathcal{Q}}_g(\nu) &= \mathcal{Q}_g(\nu) - (\nu M X_m \Theta_m - K X_m \Lambda_m^{-1} \Theta_m) (\nu X_m^\top M + \Lambda_m^{-\top} X_m^\top K) \\ &= \mathcal{Q}_g(\nu) - (M X_m (\nu I + \Lambda_m) + G X_m) \Theta_m (\nu X_m^\top M + \Lambda_m^{-\top} X_m^\top K). \end{aligned} \quad (24)$$

On the other hand, from (20), it follows that

$$\begin{aligned} \mathcal{Q}_g(\nu) X_m &= \nu^2 M X_m + \nu G X_m - M X_m \Lambda_m^2 - G X_m \Lambda_m \\ &= [M X_m (\nu I + \Lambda_m) + G X_m] (\nu I - \Lambda_m), \end{aligned}$$

which implies that

$$\mathcal{Q}_g(\nu)^{-1} [M X_m (\nu I + \Lambda_m) + G X_m] = X_m (\nu I - \Lambda_m)^{-1}. \quad (25)$$

Using the identity

$$\det(I_n + RS) = \det(I_m + SR),$$

where $R, S^\top \in \mathbb{R}^{n \times m}$, and (23)-(25), we have

$$\begin{aligned} &\det(\tilde{\mathcal{Q}}_g(\nu)) \\ &= \det(\mathcal{Q}_g(\nu)) \det(I - X_m (\nu I - \Lambda_m)^{-1} \Theta_m (\nu X_m^\top M + \Lambda_m^{-\top} X_m^\top K)) \\ &= \det(\mathcal{Q}_g(\nu)) \det(I - \Theta_m (\nu X_m^\top M + \Lambda_m^{-\top} X_m^\top K) X_m (\nu I - \Lambda_m)^{-1}) \\ &= \det(\mathcal{Q}_g(\nu)) \det(I - (\nu I + \Theta_m \Lambda_m^{-\top} (X_m^\top K X_m)) (\nu I - \Lambda_m)^{-1}) \\ &= \det(\mathcal{Q}_g(\nu)) \det((\nu I - \Lambda_m) - \nu I - \Theta_m \Lambda_m^{-\top} (X_m^\top K X_m)) \det(\nu I - \Lambda_m)^{-1} \\ &= \det(\mathcal{Q}_g(\nu)) \det(-\Lambda_m - \Theta_m \Lambda_m^{-\top} (X_m^\top K X_m)) \det(\nu I - \Lambda_m)^{-1}. \end{aligned}$$

By the nonsingular assumption, the proof is completed. \square

Furthermore, it is expected that the eigenvectors of $\mathcal{Q}_g(\nu)$ are also those of $\tilde{\mathcal{Q}}_g(\nu)$ except that those corresponding to Λ_m are deflated. This turns out true thanks to the following noteworthy theorem.

Theorem 5. Let $(\Lambda_m, X_m) \in \mathbb{R}^{m \times m} \times \mathbb{R}^{n \times m}$ and $(\Lambda_r, X_r) \in \mathbb{R}^{r \times r} \times \mathbb{R}^{n \times r}$ be two eigenmatrix pairs of $\mathcal{Q}_g(\nu)$, with $\sigma(\Lambda_m), \sigma(\Lambda_r)$ having the Hamiltonian structure. Suppose $\sigma(\Lambda_m) \cap \sigma(\Lambda_r) = \emptyset$, then the orthogonality relation holds

$$X_m^\top K X_r + \Lambda_m^\top (X_m^\top M X_r) \Lambda_r = 0, \quad (26)$$

and (Λ_r, X_r) is also an eigenmatrix pair of $\tilde{\mathcal{Q}}_g(\nu)$.

Proof. Assumption gives the equations

$$\Lambda_m^\top X_m^\top M X_r - X_m^\top G X_r + \Lambda_m^{-\top} X_m^\top K X_r = 0, \quad (27)$$

$$X_m^\top M X_r \Lambda_r + X_m^\top G X_r + X_m^\top K X_r \Lambda_r^{-1} = 0, \quad (28)$$

which implies that

$$(\Lambda_m^\top X_m^\top M X_r \Lambda_r + X_m^\top K X_r) \Lambda_r^{-1} + \Lambda_m^{-\top} (\Lambda_m^\top X_m^\top M X_r \Lambda_r + X_m^\top K X_r) = 0.$$

The uniqueness of the solution to the Sylvester equation leads to the result (26).

Since (Λ_r, X_r) is an eigenmatrix pair of $\mathcal{Q}_g(\nu)$, we have

$$M X_r \Lambda_r^2 + G X_r \Lambda_r + K X_r = 0. \quad (29)$$

From (22), (26) and (29) it follows that

$$\begin{aligned} & \tilde{M} X_r \Lambda_r^2 + \tilde{G} X_r \Lambda_r + \tilde{K} X_r \\ &= -M X_m \Theta_m X_m^\top M X_r \Lambda_r^2 + K X_m \Lambda_m^{-1} \Theta_m X_m^\top M X_r \Lambda_r \\ & \quad - M X_m \Theta_m \Lambda_m^{-\top} X_m^\top K X_r \Lambda_r + K X_m \Lambda_m^{-1} \Theta_m \Lambda_m^{-\top} X_m^\top K X_r \\ &= K X_m \Lambda_m^{-1} \Theta_m \Lambda_m^{-\top} (\Lambda_m^\top X_m^\top M X_r \Lambda_r + X_m^\top K X_r) \\ & \quad - M X_m \Theta_m \Lambda_m^{-\top} (\Lambda_m^\top X_m^\top M X_r \Lambda_r + X_m^\top K X_r) \Lambda_r = 0. \end{aligned}$$

□

Corollary 1. All eigenvectors associated with finite eigenvalues of $\tilde{\mathcal{Q}}_g(\lambda)$ are also eigenvectors of $\mathcal{Q}_g(\lambda)$.

Now, substituting (17) back to (21), we obtain a deflated \top -PQEP

$$\begin{aligned} \tilde{\mathcal{Q}}_p(\lambda) &= (\lambda \pm 1)^2 \tilde{\mathcal{Q}}_g \left(\frac{\lambda \mp 1}{\lambda \pm 1} \right) = (\lambda \mp 1)^2 \tilde{M} + (\lambda^2 - 1) \tilde{G} + (\lambda \pm 1)^2 \tilde{K} \\ &= \lambda^2 (\tilde{M} + \tilde{G} + \tilde{K}) - \lambda \left[\pm (2\tilde{M} - 2\tilde{K}) \right] + (\tilde{M} - \tilde{G} + \tilde{K}) \\ &\equiv \lambda^2 \tilde{A}^\top - \lambda \tilde{Q} + \tilde{A}, \end{aligned} \quad (30)$$

where

$$\begin{aligned}
\tilde{A} &\equiv \tilde{M} + \tilde{K} - \tilde{G} \\
&= M + K - G + KX_m\Lambda_m^{-1}\Theta_m(\Lambda_m^{-\top}X_m^\top K - X_m^\top M) \\
&\quad + MX_m\Theta_m(\Lambda_m^{-\top}X_m^\top K - X_m^\top M) \\
&= M + K - G + (KX_m\Lambda_m^{-1} + MX_m)\Theta_m(\Lambda_m^{-\top}X_m^\top K - X_m^\top M), \quad (31a)
\end{aligned}$$

$$\begin{aligned}
\tilde{Q} &\equiv 2(\tilde{M} - \tilde{K}) = \pm 2[(M - K) - KX_m\Lambda_m^{-1}\Theta_m\Lambda_m^{-\top}X_m^\top K - MX_m\Theta_mX_m^\top M] \\
&= \pm 2 \left[(M - K) - [KX_m \quad MX_m] \begin{bmatrix} \Lambda_m^{-1}\Theta_m\Lambda_m^{-\top} & 0 \\ 0 & \Theta_m \end{bmatrix} \begin{bmatrix} X_m^\top K \\ X_m^\top M \end{bmatrix} \right]. \quad (31b)
\end{aligned}$$

This is the NEDM for \top -PQEP. Obviously, the G \top SHIRA algorithm can be applied to the deflated \top -PQEP $\tilde{\mathcal{Q}}_p(\lambda)$ again.

According to (15), now in each iteration of G \top SHIRA algorithm we need to solve linear systems $\tilde{\mathcal{Q}}(\lambda_0)z_1 = d_1$ and $\tilde{\mathcal{Q}}(\lambda_0)^\top z_2 = d_2$, for which we develop the following tricks.

Let

$$U_m = KX_m, \quad V_m = MX_m\Lambda_m, \quad \Phi_m = \Lambda_m^\top\Theta_m^{-1}\Lambda_m.$$

Substituting (19) into (31), \tilde{A} and \tilde{Q} can be rewritten as

$$\tilde{A} = 4A + (U_m + V_m)\Phi_m^{-1}(U_m^\top - V_m^\top), \quad (32a)$$

$$\tilde{Q} = 4Q \mp 2 [U_m \quad V_m] \begin{bmatrix} \Phi_m^{-1} & 0 \\ 0 & \Phi_m^{-1} \end{bmatrix} \begin{bmatrix} U_m^\top \\ V_m^\top \end{bmatrix}. \quad (32b)$$

From (32), it follows that

$$\begin{aligned}
&\tilde{\mathcal{Q}}_p(\lambda_0) \\
&= 4\mathcal{Q}_p(\lambda_0) + [(\lambda_0 \pm 1)^2 U_m + (1 - \lambda_0^2)V_m \quad (\lambda_0^2 - 1)U_m - (\lambda_0 \mp 1)^2 V_m] \begin{bmatrix} \Phi_m^{-1} U_m^\top \\ \Phi_m^{-1} V_m^\top \end{bmatrix}
\end{aligned}$$

and that

$$\tilde{\mathcal{Q}}_p(\lambda_0)^\top = 4\mathcal{Q}_p(\lambda_0)^\top + [U_m\Phi_m^{-1} \quad V_m\Phi_m^{-1}] \begin{bmatrix} (\lambda_0 \pm 1)^2 U_m^\top + (1 - \lambda_0^2)V_m^\top \\ (\lambda_0^2 - 1)U_m^\top - (\lambda_0 \mp 1)^2 V_m^\top \end{bmatrix}.$$

Define

$$\begin{aligned}
W_m &= \mathcal{Q}_p(\lambda_0)^{-1} [(\lambda_0 \pm 1)^2 U_m + (1 - \lambda_0^2)V_m \quad (\lambda_0^2 - 1)U_m - (\lambda_0 \mp 1)^2 V_m], \\
Z_m &= \mathcal{Q}_p(\lambda_0)^{-\top} [U_m \quad V_m].
\end{aligned}$$

Then applying the Sherman-Morrison-Woodbury formula we have

$$\begin{aligned}
&\tilde{\mathcal{Q}}_p(\lambda_0)^{-1} \\
&= \frac{1}{4}\mathcal{Q}_p(\lambda_0)^{-1} - \frac{1}{4}W_m \left(4 \begin{bmatrix} \Phi_m & 0 \\ 0 & \Phi_m \end{bmatrix} + \begin{bmatrix} U_m^\top \\ V_m^\top \end{bmatrix} W_m \right)^{-1} \begin{bmatrix} U_m^\top \\ V_m^\top \end{bmatrix} \mathcal{Q}_p(\lambda_0)^{-1}
\end{aligned}$$

and

$$\begin{aligned}
& \tilde{\mathcal{Q}}_p(\lambda_0)^{-\top} \\
&= \frac{1}{4} \mathcal{Q}_p(\lambda_0)^{-\top} - \frac{1}{4} Z_m \left(4 \begin{bmatrix} \Phi_m & 0 \\ 0 & \Phi_m \end{bmatrix} + \begin{bmatrix} (\lambda_0 \pm 1)^2 U_m^\top + (1 - \lambda_0^2) V_m^\top \\ (\lambda_0^2 - 1) U_m^\top - (\lambda_0 \mp 1)^2 V_m^\top \end{bmatrix} Z_m \right)^{-1} \\
&\quad \times \begin{bmatrix} (\lambda_0 \pm 1)^2 U_m^\top + (1 - \lambda_0^2) V_m^\top \\ (\lambda_0^2 - 1) U_m^\top - (\lambda_0 \mp 1)^2 V_m^\top \end{bmatrix} \mathcal{Q}_p(\lambda_0)^{-\top}.
\end{aligned}$$

5. Practical implementation

5.1. Estimating how many eigenvalues near the shift

Keep in mind that the number ℓ of desired eigenvalues is one necessary input for Algorithm 2. In this subsection, we propose an empirical rule to determine ℓ when a specific shift λ_0 in Algorithm 2 is given.

Recall that at the j -th iteration of G \top SHIRA algorithm, we obtain the \top -isotropic Arnoldi decomposition (14) with $H_j, R_j \in \mathbb{C}^{j \times j}$. Let (θ_i, v_i) be an eigenpair of (H_j, R_j) and let $z_i = Z_j v_i$ be a Ritz vector of the pencil $(\hat{\mathcal{K}} - \mu \hat{\mathcal{N}})$ corresponding to the Ritz value θ_i . Then from (14), we have the residual bound

$$\|\hat{\mathcal{K}} z_i - \theta_i \hat{\mathcal{N}} z_i\|_2 = \|h_{j+1,j}(e_j^\top v_i) y_{j+1}\|_2 = |h_{j+1,j}| |e_j^\top v_i|. \quad (33)$$

Intuitively, for a given j , it is likely that some residuals of the resulting j Ritz pairs are small enough that we can refine the corresponding Ritz values with only a little extra effort to obtain the desired eigenvalues. Accordingly, the number of the such Ritz pairs can be taken as good estimation of ℓ . Specifically, by trial and error we set $j = 70$, and set ℓ to the number of the resulting Ritz pairs whose residuals, using (33), satisfy

$$|h_{j+1,j}| |e_j^\top v_i| \leq 10^{-2} |\lambda_0|^{0.3}.$$

5.2. How to determine shifts

Given a circle with center origin and radius ρ , we choose the shift λ_0 along the circle and move λ_0 from ρ to $-\rho$ counterclockwise. Let $\lambda_1, \dots, \lambda_\ell$ be the computed eigenvalues of the pair $(\mathcal{M}, \mathcal{L})$ by Algorithm 2 with shift λ_0 . The associated theoretical convergence radius γ of G \top SHIRA is

$$\gamma = \max_{1 \leq i \leq \ell} \left\{ \left| \lambda_i + \frac{1}{\lambda_i} - \lambda_0 - \frac{1}{\lambda_0} \right| \right\}.$$

We can find on the circle a point $\rho e^{i\theta}$ which satisfies

$$\left| \rho e^{i\theta} + \frac{1}{\rho} e^{-i\theta} - \lambda_0 - \frac{1}{\lambda_0} \right| = \gamma,$$

with θ greater than the argument of λ_0 . Then the shift of the next step is set to

$$\lambda_{0,\text{new}} = \rho e^{i(\theta+\delta)},$$

where δ depends on the distribution of the eigenvalues $\lambda_1, \dots, \lambda_\ell$. If they are clustered in a small region, then so are the eigenvalues nearby, supposedly. Consequently, the parameter δ must be small, i.e. $\lambda_{0,\text{new}}$ is close to λ_0 , to hit as many eigenvalues as possible in the next step. Otherwise, δ can be large. Therefore, we define the density d of the distribution of $\lambda_1, \dots, \lambda_\ell$ as

$$d = \frac{\ell}{\theta_2 - \theta_1},$$

where θ_1 and θ_2 are the minimal and maximal argument of $\lambda_1, \dots, \lambda_\ell$, respectively. Using such d , we set δ to

$$\delta = \min \left\{ \frac{5\pi}{180}, \frac{5}{d} \right\}.$$

5.3. How to determine the radius of the circles

In the first step, we deal with the unit circle, i.e. $\rho = 1$. The initial shift λ_0 is set to $\lambda_0 = \rho$. Then, ℓ is set to, say, ℓ_1 , following the rule in Sec. 5.1, and after running Algorithm 2 we can obtain ℓ_1 target eigenvalues $\lambda_{1,1}, \dots, \lambda_{1,\ell_1}$ with $|\lambda_{1,1}| \leq \dots \leq |\lambda_{1,\ell_1}|$. Next, with $\lambda_{1,1}, \dots, \lambda_{1,\ell_1}$ known, we apply the rule in Sec. 5.2 to determine the new shift. Repeating these two steps, we can compute the eigenvalues with nonnegative imaginary part which are near the present circle. Suppose that for the present circle there are m shifts used and let $\lambda_{i,1}, \dots, \lambda_{i,\ell_i}$ be the computed eigenvalues using i -th shift with $|\lambda_{i,1}| \leq \dots \leq |\lambda_{i,\ell_i}|$, for $i = 1, \dots, m$. Then we set the radius of the next circle to $\rho = \min_{1 \leq i \leq m} \{|\lambda_{i,1}|\}$.

5.4. How to choose the converged eigenvalues to be deflated

Note that the eigenvalues to be deflated in the NEDM for \mathbb{T} -PQEP are not specified in Sec. 4. Let $\lambda_1, \dots, \lambda_k$ be the eigenvalues of $(\mathcal{M}, \mathcal{L})$ produced by Algorithm 2 with shift σ_0 . In this subsection, we propose an rule to determine which of these eigenvalues should be deflated when a new shift σ_1 is given. For this purpose, we reorder $\lambda_1, \dots, \lambda_k$ into a new sequence denoted by $\hat{\lambda}_1, \dots, \hat{\lambda}_k$ such that

$$|\mu_1 - \mu_\sigma| \leq |\mu_2 - \mu_\sigma| \leq \dots \leq |\mu_k - \mu_\sigma|,$$

where $\mu_\sigma = \sigma_1 + 1/\sigma_1$ and $\mu_i = \hat{\lambda}_i + 1/\hat{\lambda}_i$ for $i = 1, \dots, k$. Since the convergence rate of the GTSHIRA algorithm still obeys the pertinent theory for the usual unstructured Arnoldi algorithm, it is easy to know that the probability of recomputing $\hat{\lambda}_i$ by the GTSHIRA algorithm with shift σ_1 decreases monotonously from $i = 1$ to $i = k$. In other words, the smaller i is, the more likely $\hat{\lambda}_i$ will be deflated. However, it is unnecessary to deflate all convergent eigenvalues. Whichever of $\hat{\lambda}_1, \dots, \hat{\lambda}_k$ are far away from σ_1 can hardly be recomputed by the GTSHIRA algorithm, hence are of no concern. Specifically, by trial and error 20 convergent eigenvalues will be deflated, i.e. here $\hat{\lambda}_1, \dots, \hat{\lambda}_{20}$ will be deflated assuming $k \geq 20$.

5.5. Finding missing eigenpairs

Executing above processes, we can not guarantee that all the target eigenpairs are computed, though most of them are hit. In this subsection, we develop a scheme to find these missing eigenpairs.

Recall that the \top -symplectic pair $(\mathcal{M}, \mathcal{L})$ admits of the following real symplectic canonical form [25]

$$\mathcal{M}U = \mathcal{L}U\Lambda,$$

where Λ is block diagonal and the (right) eigenspace U satisfies

$$U^\top \mathcal{J}_{2n} U = \mathcal{J}_{2n},$$

or equivalently,

$$\mathcal{J}_{2n}^\top U^\top \mathcal{J}_{2n} = U^{-1}.$$

Assuming that \mathcal{L} is invertible, we have

$$\Lambda = U^{-1}(\mathcal{L}^{-1}\mathcal{M})U = (\mathcal{J}_{2n}^\top U^\top \mathcal{J}_{2n})(\mathcal{L}^{-1}\mathcal{M})U. \quad (34)$$

Let this matrix Λ be partitioned into $\Lambda = \text{diag}(\Lambda_{n-r}, \Lambda_r, \Lambda_{n-r}^{-1}, \Lambda_r^{-1})$, where the blocks $\Lambda_{n-r}, \Lambda_{n-r}^{-1}$ are known and the blocks $\Lambda_r, \Lambda_r^{-1}$ are to be known. Correspondingly, let the matrix $\text{span}(U_1) \in \mathbb{R}^{2n \times (2n-2r)}$ be the subspace of the known eigenvectors and $\text{span}(U_2) \in \mathbb{R}^{2n \times 2r}$ of the unknown eigenvectors. Then the eigenspace U mentioned above is $U = [U_1, U_2]P$, where the permutation matrix P is

$$P = \begin{bmatrix} I_{n-r} & 0 & 0 & 0 \\ 0 & 0 & I_{n-r} & 0 \\ 0 & I_r & 0 & 0 \\ 0 & 0 & 0 & I_r \end{bmatrix} \equiv \begin{bmatrix} P_1 & 0 \\ 0 & I_r \end{bmatrix},$$

Then, from (34), we have

$$P\Lambda P^\top = (P\mathcal{J}_{2n}^\top P^\top)PU^\top \mathcal{J}_{2n}(\mathcal{L}^{-1}\mathcal{M})UP^\top,$$

which implies

$$\text{diag}(\Lambda_r, \Lambda_r^{-1}) = \mathcal{J}_{2r}^\top U_2^\top \mathcal{J}_{2n}(\mathcal{L}^{-1}\mathcal{M})U_2. \quad (35)$$

Strictly speaking, this result does not contain any new information, since neither Λ_r nor U_2 has been known so far. However, in practice, U_2 can be approximately known. Let $X \in \mathbb{R}^{2n \times 2r}$ be a randomly constructed matrix. Define

$$\widehat{X} = X - U_1 \mathcal{J}_{2n-2r}^\top U_1^\top \mathcal{J}_{2n} X,$$

then

$$U_1^\top \mathcal{J}_{2n} \widehat{X} = U_1^\top \mathcal{J}_{2n} X - U_1^\top \mathcal{J}_{2n} U_1 \mathcal{J}_{2n-2r}^\top U_1^\top \mathcal{J}_{2n} X = 0,$$

which means that \widehat{X} is \mathcal{J} -orthogonal to U_1 and $\text{span}(\widehat{X})$ is an approximate subspace of the missing eigenvectors U_2 . From (35), we see that the eigenvalues of

$$\mathcal{J}_{2r}^\top \widehat{X}^\top \mathcal{J}_{2n}(\mathcal{L}^{-1}\mathcal{M})\widehat{X}y = \lambda y. \quad (36)$$

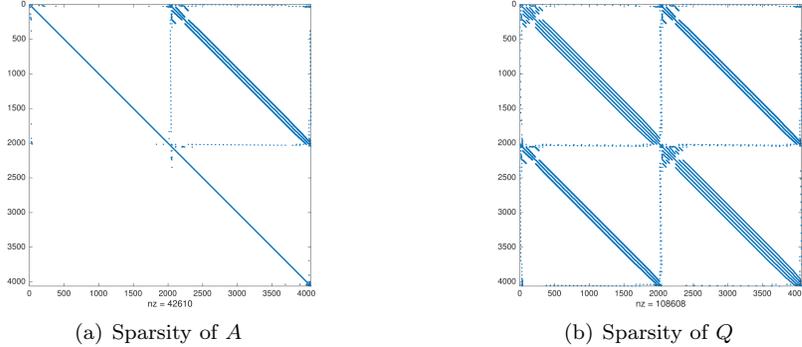


Figure 2: Sparsity patterns of A and Q

are the approximate missing eigenvalues. Then, we use these eigenvalues as shifts and apply the G^\top SHIRA algorithm to compute these missing eigenpairs.

Remark 2. Usually the missing eigenspace is well approximated by \hat{X} , then we solve the following GEP

$$\hat{X}^\top \mathcal{M} \hat{X} y = \lambda \hat{X}^\top \mathcal{L} \hat{X} y, \quad (37)$$

rather than (36).

6. Numerical results

In this section, our proof of principle is performed calculating the SGFs of a doped and hydrogen-saturated silicon nanowire with cross-section as large as $8 \times 8 \text{ nm}^2$ [4], illustrated in Fig. 1. Density functional tight binding modelling of such system results in semi-infinite symmetric Toeplitz matrices $\mathbf{H}_{L/R}$, $\mathbf{S}_{L/R}$, with sparse diagonal block H_0, S_0 and sparse sub-diagonal block H_1, S_1 of size around 4000×4000 . The sparsity pattern of diagonal and sub-diagonal block is shown in Fig. 2. It turns out such a large size already lies in the region of asymptotic scaling of almost all matrix algorithms on a small computer cluster. That means in this case the decimation method, naively calling LAPACK sub-routines, are unattractive. So here we dismiss them, but solely demonstrate the performance of our newly developed method.

In this work, all calculations are performed with MATLAB R2016b on an HP server equipped with the RedHat Linux operating system, two Intel Quad-Core Xeon E5-2643 3.33 GHz CPUs and 96 GB of main memory.

It is known that in quantum transport simulation SGF $g_R^r(\omega)$ is needed for a large number of points on real ω -axis the around Fermi level of the whole system, in order to calculate the electronic current through the nano-device [4]. Despite the fact that these real ω 's are usually confined in a relatively narrow interval,

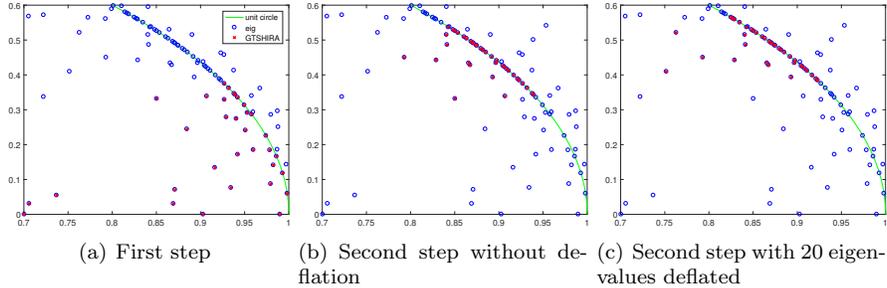


Figure 3: Effect of nonequivalent deflation for $\omega = 0.7$. Eigenvalues computed by GTSHIRA are marked by red ‘x’ and those computed by MATLAB function `eig` are marked by the blue ‘o’.

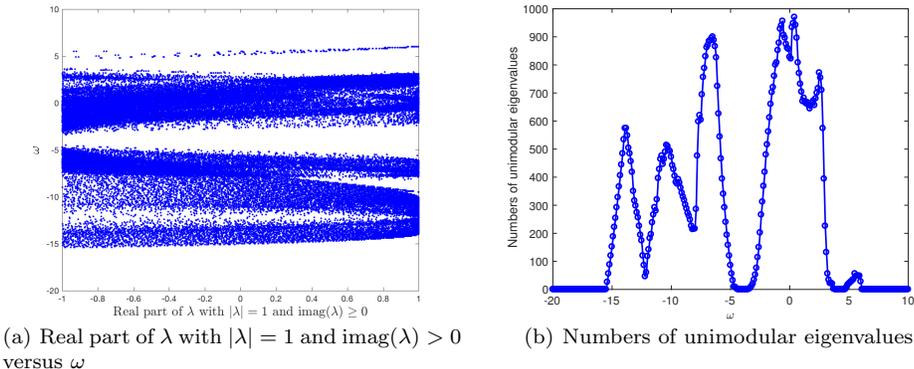


Figure 4: Distribution and numbers of all unimodular eigenvalues.

say, no longer than 1 electron Volt (eV), we just carry out our calculations at 300 equally distant points in a long interval $[-20, 10]$ without referring to the actual unit eV.

First, in Fig. 3, the superiority of NEDM is elucidated. Specifically, after 51 eigenvalues around 1, plotted in Fig. 3(a), are found by GTSHIRA algorithm, we choose the second shift on \mathbb{T} following the rule in Sec. 5.2 to find nearby $\ell = 46$ eigenvalues *without* and *with* NEDM for the \mathbb{T} -PQEP (2). Without NEDM, it is found that 7 of the 51 known eigenvalues are repeated and 39 new ones are got, as shown in Fig. 3(b); while with NEDM, after deflating 20 known eigenvalues following the rule in Sec. 5.4, there is no waste of computation and we indeed obtain 46 new eigenvalues, as shown in Fig. 3(c). The gain is $7/46 \approx 15\%$. Now that NEDM is very beneficial, we will use it for all later calculations without exception.

The rank of $\Im(\mathbf{g}_R^r(\omega))$ is very important to the model reduction in the elec-

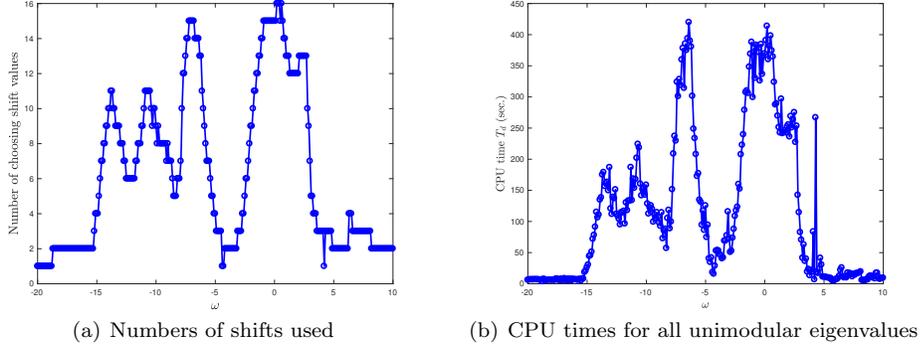


Figure 5: The performance of GTSHIRA with NEDM along \mathbb{T} .

tronic current calculation [4]. From Theorem 2, this rank is immediately known once our \mathbb{T} -PQEP solver finishes calculation along the unit circle \mathbb{T} . In Fig. 4, we show the distribution of the unimodular eigenvalues versus $\omega \in [-20, 10]$ and the total number of the unimodular eigenvalues.

Recall that in Sec. 5.2, we have proposed a rule to choose the shifts $\sigma_1, \dots, \sigma_m$ along $\mathbb{T}_\rho^+ \equiv \{x; |x| = \rho, \Im(x) \geq 0\}$ from ρ to $-\rho$. In Fig. 5(a) the number of shifts, i.e. m , used for computing unimodular eigenvalues at each ω are plotted. From Fig. 4(b), we can see that for some ω there are as many as 500 unimodular eigenvalues on \mathbb{T}_1^+ . Even so, m is no more than 17 throughout. Once the calculation for the last shift on \mathbb{T}_1^+ is finished, we sum up the wall clock time and plot the results versus ω in Fig. 5(b). The efficiency of our proposed method is clearly seen.

Following all the discussions in Sec. 5, we carefully carry out the calculation for all shifts within \mathbb{T}_1^+ , one circle by one circle. After all, we can not guarantee that all the target eigenvalues are found. If there are some missing eigenvalues, then we use the scheme proposed in Sec. 5.5 to find them. As shown in Fig. 6, the estimations of the missing eigenvalues using our scheme are very close to those produced by MATLAB function `eig`. That is the key to our success that all missing eigenpairs can be recovered. Even though there may be as many as 400 missing eigenvalues, shown in Fig. 7(a), our scheme in Sec. ?? still works well, which highlights the robustness of our scheme.

Besides, the accuracy of $\mathbf{g}_R^r(\omega)$ from our \mathbb{T} -PQEP solver is a matter of great concern. To measure the accuracy, we check the relative residual

$$\text{RRes}_\omega = \frac{\|X_\omega + A_\omega^\top X_\omega^{-1} A_\omega - Q_\omega\|}{\|X_\omega\| + \|A_\omega\|^2 \|X_\omega^{-1}\| + \|Q_\omega\|}, \quad (38)$$

after plugging the solution X_ω into the NME (6), where $\|\cdot\|$ is the spectral norm. We plot RRes_ω versus ω in Fig. 7(b). The accuracy is around 10^{-8} on

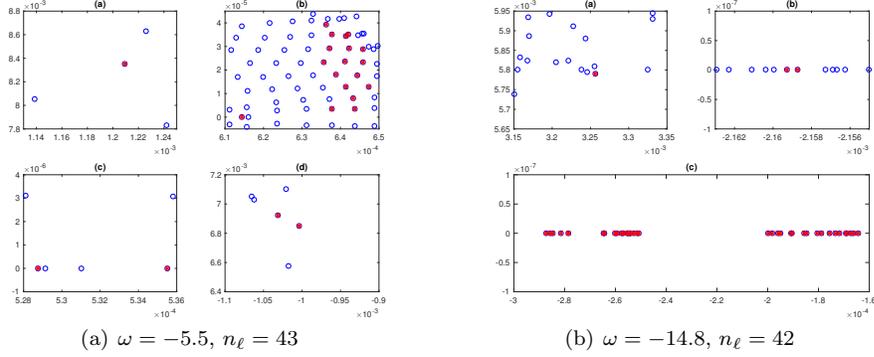


Figure 6: The missing eigenvalues estimated by the scheme in Sec. 5.5. The approximations are marked by red ‘x’ and eigenvalues from MATLAB function `eig` by blue ‘o’. n_ℓ is the number of missing eigenvalues.

average. Deterioration of accuracy is probably due to lack of awareness of the Jordan form, which should be carefully considered in future

7. Conclusion

In this work, we have strengthened the power of the GTSHIRA algorithm significantly so that the *full* spectrum and the associated eigenvectors of a large sparse Υ -PQEP can be efficiently, robustly and relatively accurately computed on a small computer cluster, putting aside the memory limit. This is accomplished owing to the sophisticated divide-and-conquer strategy, the scheme to find the missing eigenvalues, the trick to solve the sparse linear system in Algorithm 1 and the novel NEDM for the Υ -PQEP as well as GQEP, to name a few. In particular, it is necessary to reiterate the unique features of NEDM here. Over the usual equivalence transformation which always preserve eigenvalues, there are two major advantages of the non-equivalence deflation as follows.

- (1) The current converged eigenvalues will be moved to infinity or other remote places, as if some nontrivial projection is applied, while the rest eigenvalues are not changed. More importantly, the rest target eigenvalues can be computed by the structure-preserving Arnoldi algorithm with much faster convergence, also with shorter sequence of basis vectors and lower computational overhead.
- (2) The structure of the original nonlinear EP, say, QEP, under the nonequivalence deflation is preserved almost perfectly, except for the low rank updates in the coefficient matrices, which facilitates applying the structure-preserving Arnoldi algorithm.

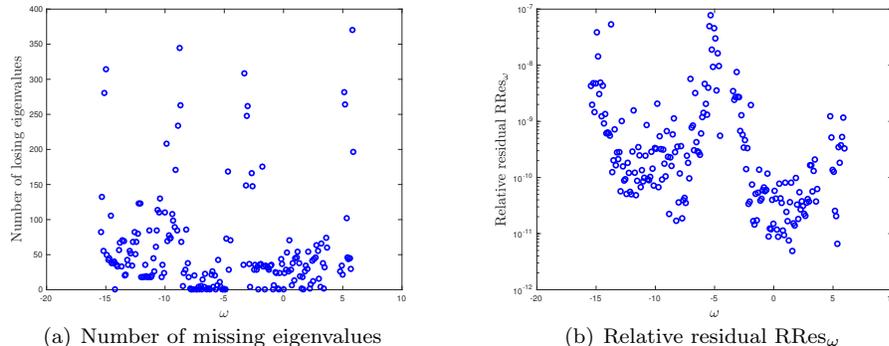


Figure 7: Number of missing eigenvalues and relative residual $RRes_\omega$ for $\omega \in \Omega$.

On the other hand, a deeper understanding of the SGF in quantum transport simulation is summarized from the perspective of \mathbb{T} -PQEP and NME. Especially, compared with physicists' approach to SGF mentioned in Sec. 1 and Sec. 2, our approach is more rigorous and more efficient owing to the linearization (9), transformation (11) and Algorithm 1 as well as Algorithm 2, which are all structure-preserving. Our proof-of-principle calculation is carried out on a Si nanowire of a $8 \times 8 \text{ nm}^2$ cross-section. Even though this system still lags behind the frontier of electronics industry, this work shows the great promise for filling this gap. We are now engaged in incorporating this new \mathbb{T} -PQEP solver into our software package for quantum transport simulation, and trying to simulate a more complicated system of a even larger cross-section with this new method.

Some prospective issues that are beyond the scope of the present work. In particular, in order to further strengthen our \mathbb{T} -PQEP solver, as mentioned in Sec. 2 and Sec. 6, we have to confront the numerical problem of Jordan form of the pencil $\mathcal{M} - \lambda\mathcal{L}$. Moreover, when the size of the coefficient matrices of \mathbb{T} -PQEP is so huge that solving the sparse linear system in line 1 of Algorithm 1 becomes the bottleneck, we have to find an appropriate sparse linear solver of $\mathcal{O}(n)$ complexity, which is definitely a highly nontrivial task. In addition, we have only discussed the serial implementation of our new \mathbb{T} -PQEP solver, however, the parallel implementation is expectedly even more challenging and needs further investigation.

8. Acknowledgements

Huang was partially supported by the Ministry of Science and Technology (MOST) 105-2115-M-003-009-MY3, National Center of Theoretical Sciences (NCTS) in Taiwan. Lin was partially supported by MOST, NCTS and ST Yau Center in Taiwan. Tian and Chen would like to acknowledge the support from Hong Kong University Grant Council (AoE/P-04/08). Tian is also grateful for

the chance of short-term visit to the ST Yau Center in NCTU, which breeds this joint project. Mr. Jun Li at University of Hong Kong (HKU) provided the original data of matrices H_0, S_0, H_1, S_1 and Dr. Shu-Guang Chen at HKU gave us valuable feedback and helped us to prepare Fig.1 in this article. Both of them deserve gratitude, too.

References

- [1] T. Ando. Quantum point contacts in magnetic fields. *Phys. Rev. B*, 44:8047, 1991.
- [2] D. A. Bini, L. Gemignani, and B. Meini. Computations with infinite Toeplitz matrices and polynomials. *Linear Algebra Appl.*, 343-344:21–61, 2002.
- [3] Y. C. Chang and J. N. Schulman. Complex band structures of crystalline solids: An eigenvalue method. *Phys. Rev. B*, 25:3975, 1982.
- [4] Q. Chen, J. Li, C.-Y. Yam, Y. Zhang, N. Wong, and G.-H. Chen. An approximate framework for quantum transport calculation with model order reduction. *J. Comput. Phys.*, 286:49–61, 2015.
- [5] 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.*, 31:227–247, 2009.
- [6] E. K.-W. Chu, T.-M. Huang, W.-W. Lin, and C.-T. Wu. Palindromic eigenvalue problems: a brief survey. *Taiwanese J. Math.*, 14:743–779, 2010.
- [7] 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.*, 219:237–252, 2008.
- [8] F. Guinea, C. Tejedor, F. Flores, and E. Louis. Effective two-dimensional Hamiltonian at surfaces. *Phys. Rev. B*, 28:4397, 1983.
- [9] C.-H. Guo. Numerical solution of a quadratic eigenvalue problem. *Linear Algebra Appl.*, 385:391–406, 2004.
- [10] C.-H. Guo, Y.-C. Kuo, and W.-W. Lin. Complex symmetric stabilizing solution of the matrix equation $X + A^\top X^{-1}A = Q$. *Linear Algebra Appl.*, 435:1187–1192, 2011.
- [11] C.-H. Guo, Y.-C. Kuo, and W.-W. Lin. Numerical solution of nonlinear matrix equations arising from Green’s function calculations in nano research. *J. Comput. Appl. Math.*, 236:4166–4180, 2012.
- [12] C.-H. Guo, Y.-C. Kuo, and W.-W. Lin. On a nonlinear matrix equation arising in nano research. *SIAM J. Matrix Anal. Appl.*, 33(1):235–262, 2012.

- [13] C.-H. Guo and W.-W. Lin. The matrix equation $X + A^T X^{-1} A = Q$ and its application in nano research. *SIAM J. Sci. Comput.*, 32:3020–3038, 2010.
- [14] M. Harb, V. Michaud-Rioux, Y. Zhu, L. Liu, L. Zhang, and H. Guo. Quantum transport modelling of silicon nanobeams using heterogeneous computing scheme. *J. Appl. Phys.*, 119:124304, 2016.
- [15] A Hilliges, C. Mehl, and V. Mehrmann. On the solution of palindromic eigenvalue problems. In *Proceedings 4th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS)*, Jyväskylä, Finland, 2004.
- [16] T.-M. Huang, T. Li, W.-W. Lin, and C.-T. Wu. Numerical studies on structure-preserving algorithms for surface acoustic wave simulations. *J. Comput. Appl. Math.*, 244:140–154, 2013.
- [17] T.-M. Huang, W.-W. Lin, and V. Mehrmann. A Newton-type method with nonequivalence deflation for nonlinear eigenvalue problems arising in photonic crystal modeling. *SIAM J. Sci. Comput.*, 38:B191–B218, 2016.
- [18] T.-M. Huang, W.-W. Lin, and J. Qian. Structure-preserving algorithms for palindromic quadratic eigenvalue problems arising from vibration on fast trains. *SIAM J. Matrix Anal. Appl.*, 30:1566–1592, 2008.
- [19] T.-M. Huang, W.-W. Lin, and C.-T. Wu. Structure-preserving Arnoldi-type algorithm for solving eigenvalue problems in Leaky surface wave propagation. *Appl. Math. Comput.*, 219:9947–9958, 2013.
- [20] P. A. Khomyakov and G. Brocks. Real-space finite-difference method for conductance calculations. *Phys. Rev. B*, 70:195402, 2004.
- [21] P. A. Khomyakov, G. Brocks, V. Karpan, M. Zwierzycki, and P. J. Kelly. Conductance calculations for quantum wires and interfaces: Mode matching and Green’s functions. *Phys. Rev. B*, 72:035450, 2005.
- [22] D. H. Lee and J. D. Joannopoulos. Simple scheme for surface-band calculations. I. *Phys. Rev. B*, 23:4988, 1981.
- [23] D. H. Lee and J. D. Joannopoulos. Simple scheme for surface-band calculations. II. The Green’s function. *Phys. Rev. B*, 23:4997, 1981.
- [24] W.-W. Lin. A new method for computing the closed-loop eigenvalues of a discrete-time algebraic Riccati equation. *Linear Algebra Appl.*, 96:157–180, 1987.
- [25] W.-W. Lin, V. Mehrmann, and H. Xu. Canonical forms for Hamiltonian and symplectic matrices and pencils. *Linear Algebra Appl.*, 302-303:469–533, 1999.

- [26] 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(1):26–39, 2006.
- [27] V. Mehrmann and H. Voss. Nonlinear eigenvalue problems: A challenge for modern eigenvalue methods. *GAMM-Mitteilungen*, 27(2):121–152, 2004.
- [28] J. Qian and W.-W. Lin. A numerical method for quadratic eigenvalue problems of gyroscopic systems. *J. Sound Vibr.*, 306:284–296, 2007.
- [29] I. Rungger and S. Sanvito. Algorithm for the construction of self-energies for electronic transport calculations based on singularity elimination and singular value decomposition. *Phys. Rev. B*, 78:035407, 2008.
- [30] M. P. L. Sancho, J. M. L. Sancho, J. M. L. Sancho, and J. Rubio. Highly convergent schemes for the calculation of bulk and surface Green functions. *J. Phys. F: Met. Phys.*, 15:851–858, 1985.
- [31] S. Sanvito, C. J. Lambert, J. H. Jefferson, and A. M. Bratkovsky. General Green’s-function formalism for transport calculations with *spd* Hamiltonians and giant magnetoresistance in Co- and Ni-based magnetic multilayers. *Phys. Rev. B*, 59:11936, 1999.
- [32] H. H. B. Sørensen, P. C. Hansen, D. E. Petersen, S. Skelboe, and K. Stokbro. Krylov subspace method for evaluating the self-energy matrices in electron transport calculations. *Phys. Rev. B*, 77:155301, 2008.
- [33] F. Tisseur and K. Meerbergen. The quadratic eigenvalue problem. *SIAM Rev.*, 43:235–286, 2001.
- [34] J. Velez and W. Butler. On the equivalence of different techniques for evaluating the Green function for a semi-infinite system using a localized basis. *J. Phys. Condens. Matter*, 21:R637–R657, 2004.
- [35] M. Wimmer. *Quantum transport in nano-structures: From computational concepts to spintronics in graphene and magnetic tunnel junctions*. PhD thesis, University of Regensburg, 2008.
- [36] S. Zaglmayr. *Eigenvalue Problems in SAW-Filter Simulations*. Diplomarbeit, Institute für Numerische Mathematik, Johannes Kepler Universität Linz, Linz, Austria, July 2002.