行政院國家科學委員會專題研究計畫 成果報告

沉浸有限元素法於彈性界面與流構耦合問題的計算與誤差估計
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 99-2115-M-009-001-
執 行 期 間 ：99 年 08 月 01 日至 100 年 07 月 31 日
執 行 單 位 ：國立交通大學應用數學系（所）

計 畫 主 持 人：吳金典

計畫參與人員：碩士班研究生-兼任助理人員：蔣明虔
碩士班研究生-兼任助理人員：楊濟俊

報 告 附 件 ：出席國際會議研究心得報告及發表論文

處 理 方 式 ：本計畫涉及專利或其他智慧財產權，2 年後可公開查詢

中 華 民 國 100 年 10 月 29 日

Elsevier Editorial System(tm) for Journal of Computational and Applied Mathematics
Manuscript Draft

Manuscript Number: CAM-D-11-00770

Title: Numerical studies on structure-preserving algorithms for surface acoustic wave simulations

Article Type: Research Paper

Section/Category: 65Fxx

Keywords: Palindromic, structure-preserving, surface acoustic wave, finite element

Corresponding Author: Dr. Chin-Tien Wu,

Corresponding Author's Institution: Mathematical Modelling and Scientific Computation

First Author: Tsung-Ming Huang

Order of Authors: Tsung-Ming Huang; Chin-Tien Wu; Tiexiang Li

Numerical Studies on Structure-Preserving Algorithms for Surface Acoustic Wave Simulations[1]

Tsung-Ming Huang[a], Chin-Tien Wu[b,*], Tiexiang Li[c]

[a] *Department of Mathematics, National Taiwan Normal University, Taipei 116, Taiwan. E-mail:* min@math.ntnu.edu.tw.
[b] *Department of Applied Mathematics, National Chiao Tung University, Hsinchu 300, Taiwan. E-mail:* ctw@math.nctu.edu.tw.
[c] *Department of Mathematics, Southeast University, Nanjing 211189, Peoples Republic of China. E-mail:* txli@seu.edu.cn.

Abstract

We study the generalized eigenvalue problems (GEPs) derived from modeling the surface acoustic wave in piezoelectric materials with periodic inhomogenuity. The eigenvalues appear in the reciprocal pairs due to periodic boundary conditions in the modeling. By transforming the GEP into a T-palindromic quadratic eigenvalue problem (TPQEP), the reciprocal relationship of the eigenvalues can be maintained. In this paper, we outline four recently developed structure-preserving algorithms, SA, SDA, TSHIRA and GTSHIRA, for solving the TPQEP. Numerical comparisons on the accuracy and the computational costs of these algorithm are presented. The eigenvalues close to unit circle on the complex plane are of interests in the area of filter and sensor designs. Our numerical results show that the Arnoldi-type structure-preserving algorithms TSHIRA and GTSHIRA with "re-sympletic" and "re-bi-isotropic", respectively, are as accurate as the SA and SDA algorithm, and more efficient in finding these eigenvalues.

1. Introduction

In this paper we consider the generalized eigenvalue problem (GEP) of the form

$$\begin{bmatrix} M_1 & G \\ F^\top & 0 \end{bmatrix} \begin{bmatrix} \psi_i \\ \psi_\ell \end{bmatrix} + \lambda \begin{bmatrix} 0 & F \\ G^\top & M_2 \end{bmatrix} \begin{bmatrix} \psi_i \\ \psi_\ell \end{bmatrix} = 0, \tag{1}$$

where $M_1^\top = M_1 \in \mathbb{C}^{n\times n}$, $M_2^\top = M_2 \in \mathbb{C}^{m\times m}$, F and $G \in \mathbb{C}^{n\times m}$ with $m \ll n$, and the supscript "$\top$" denotes the complex transpose. If M_1 and M_2 are nonsingular, then (1) can be reduced as the T-palindromic quadratic eigenvalue problem (TPQEP) of the form

$$\mathcal{P}(\lambda)x \equiv (\lambda^2 A_1^\top + \lambda A_0 + A_1)x = 0, \tag{2}$$

*Corresponding author: Chin-Tien, Wu, Tel: +886-3-5712121-ext-56424.

where

$$\begin{aligned} x &= \psi_\ell, & \psi_i &= -M_1^{-1}(\lambda F + G)\psi_\ell, \\ A_1 &= F^\top M_1^{-1} G, & A_0 &= F^\top M_1^{-1} F + G^\top M_1^{-1} G - M_2; \end{aligned} \tag{3}$$

or

$$\begin{aligned} x &= \psi_i, & \psi_\ell &= -\lambda^{-1} M_2^{-1}(F^\top + \lambda G^\top)\psi_i, \\ A_1 &= G M_2^{-1} F^\top, & A_0 &= F M_2^{-1} F^\top + G M_2^{-1} G^\top - M_1. \end{aligned} \tag{4}$$

By taking the transpose of $\mathcal{P}(\lambda)$ in (2) and multiplying it by $1/\lambda^2$ it is easily seen that the eigenvalues of $\mathcal{P}(\lambda)$ appear in the reciprocal pairs $(\lambda, 1/\lambda)$ (including 0 and ∞). Since the nullity of $A_1 = GM_2^{-1}F^\top$ in (4) is larger or equal to $n-m$, $\mathcal{P}(\lambda)$ in (2) with A_0 and A_1 defined in (4) has $n-m$ trivial zero and infinite eigenvalues which are not interested. We are only interested in finding $2m(\ll 2n)$ nontrivial eigenpairs of $\mathcal{P}(\lambda)$.

The GEP (1) can be solved by traditional methods such as QZ and Arnoldi method. But it does not guarantee that half of the computed eigenvalues lie inside of the unit circle and the others are outside [9]. For solving TPQEP (2) with small and dense matrices A_0 and A_1, some pioneering works [7, 13, 14] have been done for preserving the reciprocity of the eigenvalues basing on a good linearization of (2) which transforms (2) into the form $\lambda Z^\top + Z$. Some structure-preserving methods [7, 18, 19] were proposed for solving $(\lambda Z^\top + Z)u = 0$. A structure-preserving doubling algorithm for solving (2) was developed in [5] via the computation of a solvent of a nonlinear matrix equation associated with (2). Another structure-preserving algorithm based on $(\mathcal{S}+\mathcal{S}^{-1})$-transform [12] and Patel's approach [17] was developed in [8]. For problems with large and sparse matrices A_0 and A_1, a structure-preserving algorithm using $(\mathcal{S}+\mathcal{S}^{-1})$-transform and implicity-restarted shift-and-invert Arnoldi method was also developed for searching eigenvalues in a specified region of interests [8].

The GEP (1) typically arises in many application areas including rail vibrations of fast train, surface acoustic wave (SAW) in filter design and crack modeling, etc [6]. In these areas, an accurate and efficient eigensolver which preserves the reciprocal relationship of the associated eigenpairs is needed. In this paper, we would like to compare the accuracy and computational costs of the above mentioned algorithms for computing reciprocal eigenpairs in a SAW device [22]. The SAW filter plays an important role in telecommunication filters [4, 16] and sensor technologies [2] etc. These filters are built on the physical property of piezoelectric materials, that electrical charges induce mechanical deformations and vice versa. The main component (or cell) of a SAW filter composes of a piezoelectric substrate and the input and output interdigital transducers (IDT). An input electrical signal from the input IDT produces a surface acoustic wave, traveling through periodically arranged electrodes and the output IDT picks up the output electrical signal. Depending on the material properties of the piezoelectric substrate (PZT) and the metallic electrode, and the gap length between the electrodes, frequencies in a desired range can be stopped or filtered out. In the filter design, it is important to know the stop band width and the center frequency f_c of the filter where $f_c = \frac{v_s}{\lambda_s}$ here v_s and λ_s are the wave velocity and wave length of the incident wave. The center frequency and

stop band width can be determined by experiments or computation. In computational approach, the dispersion diagram needs to be generated in which a GEP of the form (1) associated with each frequency in the search range has to be solved [9].

This paper is organized as follows. We shall first introduce finite element modelling for a simple SAW resonator in Section 2. For more finite element simulations of piezoelectric devices in two dimension (2D) and three dimension (3D), one can refer to the works done by Allik, Koshiba, Lerch, Buchner and Mohamed etc.,[1, 3, 10, 11]. In Section 3, we introduce four structure-preserving algorithms developed in [5, 8] to solve the TPQEP (2) and the GEP (1) resulted from our FEM model. Our numerical experiments in Section 4 compare the efficiency and accuracy of the structure-preserving algorithms for solving the GEP (1). Finally, we conclude the paper in Section 5.

2. Surface wave propagation

To model the wave propagation in a SAW device, we assume that a large number of electrodes are placed equally-spaced along a straight line on the PZT substrate. According to the Floquet-Bloch theory, one can reduce the problem to a single cell domain with one electrode by assuming the wave ψ is quasi-periodic of the form

$$\psi(x_1, x_2) = \psi_p(x_1, x_2)e^{(\alpha+\imath\beta)x_1}, \quad \psi_p(x_1 + p, x_2) = \psi_p(x_1, x_2),$$

where x_1 is the wave propagation direction, p is the length of the unit cell (i.e. the periodic interval), α and β are the attenuation and phase shift along the wave propagation direction, respectively.

Let Ω denote the piezoelectric substrate with a single IDT as shown in Figure 1, and Γ_ℓ and Γ_r denote the left and right boundary segments of Ω, respectively. For the general anisotropic PZT substrates, under the assumption of linear piezoelectric coupling, the elastic and electric fields interact following the general material constitution law below

$$\begin{aligned} T &= c^E S - e^\top E, \\ D &= eS + \varepsilon^S E, \end{aligned} \tag{5}$$

where vectors T, S, D and E are the mechanical stress, strain, dielectric displacement and the electric field, respectively, and the matrices c^E, ε^S and e are the elasticity constant, dielectric constant and piezoelectric constant matrices measured at constant electric and constant strain fields at constant temperature. By applying the virtual work principle to the equation (5), the equilibrium state satisfies the following equation:

$$\begin{aligned} &\int_\Omega (\delta S)^\top \left[c^E S + e^\top(\nabla\phi)\right] \, dV + \int_\Omega (\nabla\delta\phi)^\top \left[eS - \varepsilon^S(\nabla\phi)\right] \, dV + \int_\Omega (\delta u)^\top \rho\ddot{u} \, dV \\ = \; &\int_{\Gamma_l \cup \Gamma_r} \left[(\delta u)^\top (T \cdot \vec{n}) + (\delta\phi)^\top (D \cdot \vec{n})\right] \, dA, \end{aligned} \tag{6}$$

where ρ is the mass density, $u = [u_1, u_2, u_3]^\top$ is the displacement vector, ϕ is the electric potential that satisfies $\nabla\phi = E$, $S = [\frac{\partial u_1}{\partial x}, \frac{\partial u_2}{\partial y}, \frac{\partial u_3}{\partial z}, \frac{\partial u_2}{\partial z} + \frac{\partial u_3}{\partial y}, \frac{\partial u_3}{\partial x} + \frac{\partial u_1}{\partial z}, \frac{\partial u_1}{\partial y} + \frac{\partial u_2}{\partial x}]^\top$, and δu, $\delta\phi$, δS are virtual displacement, potential and strain vectors, respectively. Let

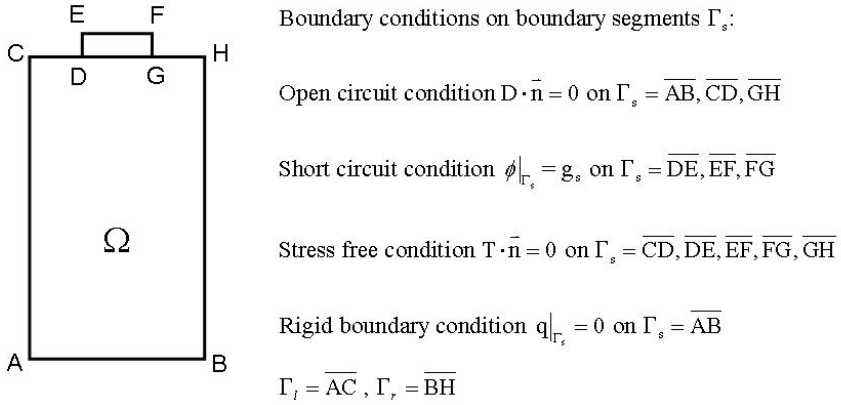


Figure 1: A 2D single cell domain of a LSAW resonator and boundary conditions

the notation $\psi=[u^\top,\phi]^\top$ and the subscript i,ℓ and r refer to nodal point index in the interior, the left boundary and the right boundary of the domain Ω, respectively. Using the periodic boundary conditions, proposed by Buchner [3],

$$T_r\cdot n_r=-\gamma T_\ell\cdot n_\ell,\quad D_r\cdot n_r=-\gamma D_\ell\cdot n_\ell\ \text{ with }\ \gamma=e^{-(\alpha+\imath\beta)},$$

the finite element discretization to (6) on the domain Ω [9] can be written in the following matrix form

$$C(\omega)\psi\equiv[K-\omega^2M+\imath\omega(\kappa_1K+\kappa_2M)]\psi=0, \tag{7}$$

where $\kappa_1,\ \kappa_2>0$ are the viscous damping and mass damping respectively. By ordering the nodal unknown ψ according the order of subscripts ℓ, i and r, the matrices K and M, and the vector ψ can be partitioned as following:

$$K=\begin{bmatrix}K_{\ell\ell}&K_{i\ell}^\top&0\\K_{i\ell}&K_{ii}&K_{ir}\\0&K_{ir}^\top&K_{rr}\end{bmatrix},\quad M=\begin{bmatrix}M_{\ell\ell}&M_{i\ell}^\top&0\\M_{i\ell}&M_{ii}&M_{ir}\\0&M_{ir}^\top&M_{rr}\end{bmatrix},$$

where $K_{ii},M_{ii}\in\mathbb{R}^{n\times n}$, $K_{\ell\ell},K_{rr},M_{\ell\ell},M_{rr}\in\mathbb{R}^{m\times m}$, $K_{i\ell},K_{ir},M_{i\ell},M_{ir}\in\mathbb{R}^{n\times m}$, and $\psi=[\psi_\ell^\top,\psi_i^\top,\psi_r^\top]^\top$ with $\psi_i\in\mathbb{C}^n$, $\psi_\ell,\psi_r\in\mathbb{C}^m$ $(m\ll n)$. Obviously the matrix $C(\omega)$ in (7) can also be partitioned into

$$C(\omega)\equiv C\equiv\begin{bmatrix}C_{\ell\ell}&C_{i\ell}^\top&0\\C_{i\ell}&C_{ii}&C_{ir}\\0&C_{ir}^\top&C_{rr}\end{bmatrix}$$

By setting $\psi_r = \lambda\psi_\ell$, the equation (7) leads to the generalized eigenvalue problem

$$\left(\begin{bmatrix} C_{ii} & C_{i\ell} \\ C_{ir}^\top & 0 \end{bmatrix} - \lambda \begin{bmatrix} 0 & C_{ir} \\ C_{i\ell}^\top & C_{bb} \end{bmatrix}\right) \begin{bmatrix} \psi_i \\ \psi_\ell \end{bmatrix} = 0, \tag{8}$$

where $C_{bb} := C_{\ell\ell} + C_{rr}$.

Since the viscosity is small for PZT substrates and metals in SAW devices, the attenuation factor α of surface waves is close to zero. As a result, the propagation factor λ are generally near the unit circle thereafter denoted by $\mathbb{U}$. Furthermore, for frequency ω in the stopping band, the frequency shift parameter β shall be close to π when the periodic interval p (i.e. the domain width here) equals to half of the incident wave length λ_s. Therefore, we are interesting in finding λ close to $\mathbb{U}$, especially for those are near -1 on the complex plane. Notice that eigenvalues of (2) appear in the reciprocal pairs $(\lambda, 1/\lambda)$. In the following sections, we aim to discuss the efficiency and accuracy of the structure-preserving algorithms [5, 8] for solving the eigen-curves $\lambda(\omega)$ and the associated eigenvectors of (8).

3. Structure-preserving Algorithms

In this section, we shall introduce four structure-preserving algorithms developed in [5, 8] to solve the TPQEP (2) and discuss the computation costs of these algorithms in solving the GEP (1). In the following, we suppose $\mathfrak{m}$ reciprocal pairs of eigenvalues near $\mathbb{U}$ are desired.

3.1. structure-preserving doubling algorithm

For solving the TPQEP (2) with $A_0, A_1 \in \mathbb{C}^{m\times m}$ defined in (3), a structure-preserving doubling algorithm (SDA) was developed in [5] via the computation of a solvent of a nonlinear matrix equation associated with (2). That is $\mathcal{P}(\lambda)$ can be factorized as

$$\mathcal{P}(\lambda) = (\lambda A_1^\top - X)X^{-1}(\lambda X - A_1) \tag{9}$$

for some nonsingular X with $X^\top = X$ if and only if X satisfies the following nonlinear matrix equation (NME):

$$A_1^\top X^{-1} A_1 + X + A_0 = 0.$$

Combining SDA in [5], the GEP (1) can be solved by Algorithm 1. The advantages of Algorithm 1 are as following: (i) the computed eigenvalues are guaranteed to appear in reciprocal pair since the eigenvalues of the matrix pencils $\lambda A_1^\top - X$ and $\lambda X - A_1$, which are reciprocal pairs, are the eigenvalues of $\mathcal{P}(\lambda)$ in (9) and (ii) the convergence rate of the SDA is proved to be quadratic [5] if there are no eigenvalues of $\mathcal{P}(\lambda)$ located on unit circle.

Next, let's discuss the computational costs of Algorithm 1. To mimic the computation cost in the LU factorization of the matrix M_1 obtained from finite element discretization, we reorder the nodal indices so that the matrix M_1 has narrower band structure. Let $M_1 = LU$ be the LU factorization of M_1. Then, computing A_0 and A_1 in Step 3.1 of

Algorithm 1 GE_SDA

Input: matrices F, G, M_2 and M_1, tolerance η and the number $\mathfrak{m}$ of desired eigenvalues.

Output: eigenpairs $\{(\gamma_j, [(\psi_{i,j}^{(1)})^\top, (\psi_{\ell,j}^{(1)})^\top]^\top), (\gamma_j^{-1}, [(\psi_{i,j}^{(2)})^\top, (\psi_{\ell,j}^{(2)})^\top]^\top)\}_{j=1}^{\mathfrak{m}}$ of (1).

1: Compute $A_0 = F^\top M_1^{-1} F + G^\top M_1^{-1} G - M_2$ and $A_1 = F^\top M_1^{-1} G$.
2: Set $k = 0$, $Y_k = A_1$, $X_k = -A_0$ and $Z_k = 0$.
3: **repeat**
4: Compute $Y_{k+1} = Y_k (X_k - Z_k)^{-1} Y_k$, $X_{k+1} = X_k - Y_k^\top (X_k - Z_k)^{-1} Y_k$, and $Z_{k+1} = Z_k + Y_k (X_k - Z_k)^{-1} Y_k^\top$;
5: Set $k = k + 1$;
6: **until** $\|X_k - X_{k-1}\| \le \eta \|X_k\|$
7: Compute the left and right eigenpairs $\{(\lambda_j, \psi_{\ell,j}^{(1)}), (\lambda_j, \psi_{\ell,j}^{(r)})\}_{j=1}^{m}$ of $X_k \psi_\ell = \lambda A_1^\top \psi_\ell$;
8: Choose the eigenpairs which associated eigenvalues are near the unit circle, said $\{(\lambda_j, \psi_{\ell,j}^{(1)}), (\lambda_j, \psi_{\ell,j}^{(r)})\}_{j=1}^{\mathfrak{m}}$;
9: Solve $(\lambda_j X_k - A_1)\psi_{\ell,j}^{(2)} = X_k \psi_{\ell,j}^{(r)}$ and set $\gamma_j = \lambda_j^{-1}$ for $j = 1, \ldots, \mathfrak{m}$;
10: Compute

$$\psi_{i,j}^{(1)} = -M_1^{-1} \left(\gamma_j F \psi_{\ell,j}^{(1)} + G \psi_{\ell,j}^{(1)} \right), \ \psi_{i,j}^{(2)} = -M_1^{-1} \left(\gamma_j^{-1} F \psi_{\ell,j}^{(2)} + G \psi_{\ell,j}^{(2)} \right)$$

for $j = 1, \ldots, \mathfrak{m}$.

Algorithm 1 requires solving $\widetilde{F} \equiv U^{-1} L^{-1} F$ and $\widetilde{G} \equiv U^{-1} L^{-1} G$, and matrix multiplications of $F^\top \widetilde{F}$, $G^\top \widetilde{G}$ and $F^\top \widetilde{G}$. In Steps 3.1-3.1, one LU factorization ($2m^3/3$ flops), two forward and back substitutions ($4m^3$ flops) and three matrices multiplications ($6m^3$ flops) are required for each iterate k. Next, computing the left and right eigenpairs in Step 3.1 and solving $\psi_{\ell,j}^{(2)}$ in Step 3.1 take $100m^3$ flops and $2\mathfrak{m}m^3/3$ flops, respectively. Finally, it also requires $2\mathfrak{m}$ forward and back substitutions to compute $\{\psi_{i,j}^{(1)}, \psi_{i,j}^{(2)}\}_{j=1}^{\mathfrak{m}}$ in Step 3.1. The total cost of Algorithm is summarized in Table 1.

3.2. structure-preserving algorithm

Another structure-preserving algorithm (SA) developed in [8] is based on the $(\mathcal{S} + \mathcal{S}^{-1})$-transform [12] and Patel's approach [17] for solving the TPQEP (2) with A_0, $A_1 \in \mathbb{C}^{m \times m}$ defined in (3). The idea is, first, to linearize the TPQEP as the following special GEP:

$$(\mathcal{M} - \lambda \mathcal{L}) \begin{bmatrix} x \\ y \end{bmatrix} = 0, \tag{10}$$

where $\lambda y = A_1 x$, and

$$\mathcal{M} = \begin{bmatrix} A_1 & 0 \\ -A_0 & -I \end{bmatrix}, \quad \mathcal{L} = \begin{bmatrix} 0 & I \\ A_1^\top & 0 \end{bmatrix}. \tag{11}$$

Obviously, the matrix pencil $\mathcal{M} - \lambda \mathcal{L}$ is $\top$-symplectic, i.e., it satisfies $\mathcal{M} \mathcal{J} \mathcal{M}^\top = \mathcal{L} \mathcal{J} \mathcal{L}^\top$ where $\mathcal{J} = \begin{bmatrix} 0 & I_m \\ -I_m & 0 \end{bmatrix}$. As a result, the eigenvalues of $(\mathcal{M}, \mathcal{L})$ appear in the reciprocal

pairs $(\lambda, 1/\lambda)$. Secondly, the $(\mathcal{S}+\mathcal{S}^{-1})$-transform is applied on $\mathcal{M}-\lambda\mathcal{L}$ and the pencil is now transformed into a $\top$-skew-Hamiltonian pencil $\mathcal{K}-\mu\mathcal{N}$, i.e., $(\mathcal{K}\mathcal{J})^\top = -\mathcal{K}\mathcal{J}$, $(\mathcal{N}\mathcal{J})^\top = -\mathcal{N}\mathcal{J}$:

$$\begin{aligned} \mathcal{K}-\mu\mathcal{N} &\equiv \left[(\mathcal{L}\mathcal{J}\mathcal{M}^\top + \mathcal{M}\mathcal{J}\mathcal{L}^\top) - \mu\mathcal{L}\mathcal{J}\mathcal{L}^\top\right]\mathcal{J}^\top \\ &= \begin{bmatrix} A_0 & A_1^\top - A_1 \\ A_1 - A_1^\top & A_0 \end{bmatrix} - \mu \begin{bmatrix} -A_1 & 0 \\ 0 & -A_1^\top \end{bmatrix}. \end{aligned} \tag{12}$$

The two eigenvalues λ and μ are then related by the relationship $\mu = \lambda + 1/\lambda$. The relationship between eigenpairs of the TPQEP in (2) and the $\top$-skew-Hamiltonian pair $(\mathcal{K},\mathcal{N})$ in (12) is stated in the following theorem.

Theorem 3.1. [8] *Let $(\mathcal{K},\mathcal{N})$ be defined in (12). If $z_s = [z_1^\top, z_2^\top]^\top$ with $z_1, z_2 \in \mathbb{C}^m$ is an eigenvector of $(\mathcal{K},\mathcal{N})$ corresponding to eigenvalue μ and ν satisfies $\nu + \frac{1}{\nu} = \mu$, then $\frac{1}{\nu}z_1 - z_2$ and $\nu z_1 - z_2$ are eigenvectors of the TPQEP in (2) corresponding to eigenvalues ν and $\frac{1}{\nu}$, respectively.*

Finally, based on Patel's approach [17], the matrix pair $(\mathcal{K},\mathcal{N})$ can further be reduced to a block triangular structure as following

$$\mathcal{K} := Q^\top \mathcal{K} Z = \begin{bmatrix} K_{11} & K_{12} \\ 0 & K_{11}^\top \end{bmatrix}, \qquad \mathcal{N} := Q^\top \mathcal{N} Z = \begin{bmatrix} N_{11} & N_{12} \\ 0 & N_{11}^\top \end{bmatrix}, \tag{13}$$

where $K_{11} \in \mathbb{C}^{m\times m}$ is upper Hessenberg, $N_{11} \in \mathbb{C}^{m\times m}$ is upper triangular, and Q, Z are unitary satisfying

$$Q = \mathcal{J}^\top Z \mathcal{J}.$$

We then apply the QZ algorithm to (K_{11}, N_{11}) for computing the m eigenpairs $\{(\mu_k, y_k)\}_{k=1}^m$. Consequently, $\{(\mu_k, Z\begin{bmatrix} y_k \\ 0 \end{bmatrix})\}_{k=1}^m$ are the m eigenpairs of $(\mathcal{K},\mathcal{N})$. Combining the above procedures and the structure-preserving algorithm in [8], the GEP (1) can be solved by Algorithm 2.

The computational costs in Steps 3.2 and 3.2 of Algorithm 2 are the same that in Steps 3.1 and 3.1 of Algorithm 1. The SA processes in Steps 3.2-3.2 of Algorithm 2 require approximately $50m^3$ flops [8] to compute the eigenpairs of the TPQEP (2) with small size matrices A_0 and A_1 in (3). The comparison of the computation costs for GE_SDA and GE_SA is listed in Table 1.

3.3. $\top$-skew-Hamiltonian implicit-restarted Arnoldi algorithm

In the above mentioned GE_SDA and GE_SA algorithms, the GEP (1) is transformed into the TPQEP (2) through equations in (3) where $M_1^{-1}F$ and $M_1^{-1}G$ are solved by LU factorization on the matrix M_1. The computation costs in this step increase in the amount of $2m$ times n^2. Since the GE_SDA and GE_SA algorithms are then working on the TPQEP where the size of matrices is $m \times m$, $m \ll n$, the computation cost in solving

Algorithm 2 GE_SA

Input: matrices F, G, M_2 and M_1, and the number $\mathfrak{m}$ of desired eigenvalues.

Output: eigenpairs $\{(\gamma_j, [(\psi_{i,j}^{(1)})^\top, (\psi_{\ell,j}^{(1)})^\top]^\top), (\gamma_j^{-1}, [(\psi_{i,j}^{(2)})^\top, (\psi_{\ell,j}^{(2)})^\top]^\top)\}_{j=1}^{\mathfrak{m}}$ of (1).

1: Compute $A_0 = F^\top M_1^{-1} F + G^\top M_1^{-1} G - M_2$ and $A_1 = F^\top M_1^{-1} G$.
2: Form the pair $(\mathcal{K}, \mathcal{N})$ as in (12);
3: Reduce $(\mathcal{K}, \mathcal{N})$ to block upper triangular forms in (13) using unitary transformations;
4: Compute eigenpairs $\{(\mu_k, y_k)\}_{k=1}^{m}$ of (K_{11}, N_{11}) defined in (13) by using the QZ algorithm;
5: Compute eigenvalues ν_k and ν_k^{-1} of $\mathcal{P}(\lambda)$ by solving $\nu^2 - \mu_k \nu + 1 = 0$;
6: Choose the eigenvalues which are near the unit circle, said $\{\nu_{k_j}, \nu_{k_j}^{-1}\}_{j=1}^{\mathfrak{m}}$;
7: Compute $z_j = Z \begin{bmatrix} y_{k_j} \\ 0 \end{bmatrix} \equiv \begin{bmatrix} z_{j1} \\ z_{j2} \end{bmatrix}, j = 1, 2, \ldots, \mathfrak{m}$;
8: Compute eigenvectors $\psi_{\ell,j}^{(1)} \equiv \gamma_j^{-1} z_{j1} - z_{j2}$ and $\psi_{\ell,j}^{(2)} \equiv \gamma_j z_{j1} - z_{j2}$ corresponding to eigenvalues $\gamma_j \equiv \nu_{k_j}$ and $\nu_{k_j}^{-1}$, respectively, for $j = 1, 2, \ldots, \mathfrak{m}$;
9: Compute

$$\psi_{i,j}^{(1)} = -M_1^{-1} \left(\gamma_j F \psi_{\ell,j}^{(1)} + G \psi_{\ell,j}^{(1)} \right), \ \psi_{i,j}^{(2)} = -M_1^{-1} \left(\gamma_j^{-1} F \psi_{\ell,j}^{(2)} + G \psi_{\ell,j}^{(2)} \right)$$

for $j = 1, \ldots, \mathfrak{m}$.

		GE_SDA	GE_SA
Compute $M_1 = LU$		1	1
Compute A_0, A_1	Solve $Lx = b_1$	$2m$	$2m$
	Solve $Ux = b_2$	$2m$	$2m$
	Compute $F^\top d_1$	$2m$	$2m$
	Compute $G^\top d_2$	m	m
Compute $\psi_i^{(1)}$, $\psi_i^{(2)}$	Solve $Lx = b_1$	$2\mathfrak{m}$	$2\mathfrak{m}$
	Solve $Ux = b_2$	$2\mathfrak{m}$	$2\mathfrak{m}$
	Compute Fe_1	$2\mathfrak{m}$	$2\mathfrak{m}$
	Compute Ge_2	$2\mathfrak{m}$	$2\mathfrak{m}$
Solve dense TPQEP		$(100 + \frac{32}{3}k + \frac{2}{3}\mathfrak{m})m^3$ flops	$50m^3$ flops

Table 1: The computational costs of GE_SDA and GE_SA where k denotes the total iterations to obtain convergent X_k in Lines 3.1-3.1 of GE_SDA.

the TPQEP is relatively small.

In the following, we introduce two Arnoldi-type algorithms in which the GEP (1) is transformed into the TPQEP (2) through equations in (4). Since the matrix size m of M_2 is much smaller, the cost in solving $M_2^{-1} F^\top$ and $M_2^{-1} G^\top$ by LU factorization of M_2 can now be ignored. Following the same idea in Section 3.2, the TPQEP (2) with $A_0, A_1 \in \mathbb{C}^{n \times n}$ is also transformed into the $\top$-skew-Hamiltonian pencil $\mathcal{K} - \mu \mathcal{N}$ through

the equations (10) and (12) with $\mathcal{J} = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}$. Instead of taking Patel's approach, we seek the eigenvalues of the matrix pair $(\mathcal{K}, \mathcal{N})$ by some implicit-restart Arnoldi algorithms. Although the Arnoldi algorithm is working on the matrices with size $2n \times 2n$ now, saving on computation costs is expected when fast convergence of the Arnoldi iterations can be achieved. In the following, we sketch the key steps and theorems that are employed in developing Arnoldi algorithm.

Let τ be a shift value and $\tau \notin \sigma(\mathcal{M}, \mathcal{L})$ where $\sigma(\mathcal{A}, \mathcal{B})$ denotes the set of all eigenvalues of any matrix pair $(\mathcal{A}, \mathcal{B})$. Then, we have $\mu_0 \equiv \tau + \tau^{-1} \notin \sigma(\mathcal{K}, \mathcal{N})$. Define the shift-invert transformation $\widehat{\mathcal{K}} - \widehat{\mu}\widehat{\mathcal{N}}$ for $\mathcal{K} - \mu\mathcal{N}$ with $\widehat{\mu} = \frac{1}{\mu - \mu_0}$ and

$$\widehat{\mathcal{K}} \equiv -\tau\mathcal{N} = \tau \begin{bmatrix} A_1^\top & 0 \\ 0 & A_1 \end{bmatrix}, \tag{14a}$$

$$\widehat{\mathcal{N}} \equiv -\tau(\mathcal{K} - \mu_0\mathcal{N}) = (\mathcal{M} - \tau\mathcal{L})\,\mathcal{J}\left(\mathcal{M}^\top - \tau\mathcal{L}^\top\right)\mathcal{J}^\top, \tag{14b}$$

where $\widehat{\mathcal{K}}$ and $\widehat{\mathcal{N}}$ are $\top$-skew-Hamiltonian. Furthermore, from the definition of $\widehat{\mathcal{N}}$ in (14b), $\widehat{\mathcal{N}}$ can be factorized as $\widehat{\mathcal{N}} = \mathcal{N}_1\mathcal{N}_2$, where

$$\mathcal{N}_1 = \mathcal{M} - \tau\mathcal{L}, \quad \mathcal{N}_2 = \mathcal{J}(\mathcal{M}^\top - \tau\mathcal{L}^\top)\mathcal{J}^\top \tag{15}$$

are nonsingular and satisfy $\mathcal{N}_2^\top\mathcal{J} = \mathcal{J}\mathcal{N}_1$. The GEP $\widehat{\mathcal{K}}z = \widehat{\mu}\widehat{\mathcal{N}}z$ is then equivalent to the eigenvalue problem $\mathcal{B}\tilde{z} = \widehat{\mu}\tilde{z}$, where

$$\mathcal{B} \equiv \mathcal{N}_1^{-1}\widehat{\mathcal{K}}\mathcal{N}_2^{-1} \tag{16}$$

is $\top$-skew-Hamiltonian, i.e., $\mathcal{J}\mathcal{B}^\top = \mathcal{B}\mathcal{J}$, and $\tilde{z} = \mathcal{N}_2\hat{z}$. Now, according to the following two main theorems proved in [8, 15], the $\top$-skew-Hamiltonian implicity-restarted Arnoldi (TSHIRA) algorithm as shown in Algorithm 4 can be employed to solve this eigenvalue problem.

Let's define the Krylov matrix with respect to u_1 by

$$K_j \equiv K_j[\mathcal{B}, u_1] = [u_1,\, \mathcal{B}u_1,\, \ldots,\, \mathcal{B}^{j-1}u_1],\ 1 \le j \le n.$$

The two main theorems in [8, 15] are as follows:

Theorem 3.2. [15] *Let $\mathcal{B} \in \mathbb{C}^{2n\times 2n}$ be $\top$-skew-Hamiltonian and K_j be a Krylov matrix with rank$(K_j) = j$. Then span(K_j) is $\top$-isotropic and if $K_j = U_jR_j$ is a QR-factorization, then*

$$\mathcal{B}U_j = U_jH_j + \tilde{u}_{j+1}e_j^\top,$$

where $H_j \in \mathbb{C}^{j\times j}$ is unreduced upper Hessenberg, $U_j \in \mathbb{C}^{2n\times j}$ is orthonormal and $\top$-isotropic, and $\tilde{u}_{j+1} \in \mathbb{C}^{2n}$ is a suitable vector such that

$$U_j^H\tilde{u}_{j+1} = 0 \quad \textit{and} \quad U_j^\top\mathcal{J}\tilde{u}_{j+1} = 0.$$

Theorem 3.3. [8, 15] *Let $\mathcal{B} \in \mathbb{C}^{2n\times 2n}$ be $\top$-skew-Hamiltonian. If rank$(K_n) = n$, then there is a unitary $\top$-symplectic matrix $\mathcal{U}$ with $\mathcal{U}e_1 = u_1$ such that*

$$\mathcal{U}^H \mathcal{B}\mathcal{U} = \begin{bmatrix} H_n & S_n \\ 0 & H_n^\top \end{bmatrix},$$

where $H_n \equiv [h_{ij}]$ is unreduced upper Hessenberg and S_n is $\top$-skew-symmetric.

Based on Theorem 3.2, the jth step of TSHIRA is given by

$$h_{j+1,j}u_{j+1} = \mathcal{B}u_j - \sum_{i=1}^{j} h_{ij}u_i, \tag{17}$$

where $h_{ij} = u_i^H \mathcal{B}u_j$, $i = 1, \ldots, j$ and $h_{j+1,j} > 0$ is chosen so that $\|u_{j+1}\|_2 = 1$. In order to guarantee the $\top$-isotropic property of the space span$\{u_1, \ldots, u_{j+1}\}$ is preserved within machine precision, reorthogonalizing u_{j+1} against $\mathcal{J}U_j$ is necessary. As a result, the equation (17) is modified into

$$h_{j+1,j}u_{j+1} = \mathcal{B}u_j - \sum_{i=1}^{j} h_{ij}u_i - \sum_{i=1}^{j} t_{ij}\mathcal{J}\bar{u}_i,$$

where $t_{ij} = -u_i^\top \mathcal{J}\mathcal{B}u_j$, $i = 1, \ldots, j$. The above procedure is stated in Algorithm 3.

Finally, we present TSHIRA with Krylov-Schur restart to solve the eigenvalue problem $\mathcal{B}\tilde{z} = \widehat{\mu}\tilde{z}$ in Algorithm 4. Once the eigenpair $(\hat{\mu}, \tilde{z})$ is obtained, one can recover the eigenpair (μ, z) of $(\mathcal{K}, \mathcal{N})$ from the relationship $\hat{u} = \frac{1}{\mu-\mu_0}$ and the solution of the linear system $\mathcal{N}_2 z = \tilde{z}$. The reciprocal eigenpair $(\lambda, \frac{1}{\lambda})$ and the associated eigenvectors of the TPQEP (2) are then followed from Theorem 3.1.

Algorithm 3 The jth $\top$-isotropic Arnoldi step

Input: $\top$-skew-Hamiltonian $\mathcal{B}$ and $U_j = [u_1, \cdots, u_j]$ with $U_j^H U_j = I_j$ and $U_j^\top \mathcal{J}U_j = 0$.
Output: $[h_{1,j}, \cdots, h_{j+1,j}]$ and u_{j+1}.

1: Compute $u_{j+1} = \mathcal{B}u_j$;
2: **for** $i = 1, \ldots, j$ **do**
3: $\quad h_{ij} = u_i^H u_{j+1}$, $u_{j+1} = u_{j+1} - h_{ij}u_i$
4: **end for**
5: **for** $i = 1, \ldots, j$ **do**
6: $\quad t_{ij} = u_i^\top \mathcal{J}^\top u_{j+1}$, $u_{j+1} = u_{j+1} - t_{ij}\mathcal{J}\bar{u}_i$
7: **end for**
8: Set $h_{j+1,j} := \|u_{j+1}\|_2$ and $u_{j+1} := u_{j+1}/h_{j+1,j}$.

3.4. Generalized $\top$-skew-Hamiltonian implicity-restarted Arnoldi algorithm

Recall that an additional linear system $\mathcal{N}_2 z = \tilde{z}$ has to be solved for recovering the eigenpair (μ, z) of $(\mathcal{K}, \mathcal{N})$ when TSHIRA is employed to solve the GEP $\widehat{\mathcal{K}}z = \widehat{\mu}\widehat{\mathcal{N}}z$ in (14).

Algorithm 4 [8] TSHIRA for solving $\mathcal{B}\tilde{z} = \widehat{\mu}\tilde{z}$

Input: $\top$-skew-Hamiltonian matrix $\mathcal{B}$ with starting vector u_1.

Output: eigenpairs $(\widehat{\mu}_i, \tilde{z}_i)$, $i = 1, \ldots, \mathfrak{m}$ of $\mathcal{B}$.

1: Use Algorithm 3 with starting vector u_1 to generate the $\mathfrak{m}$th step of $\top$-isotropic Arnoldi decomposition:

$$\mathcal{B}U_{\mathfrak{m}} = U_{\mathfrak{m}}H_{\mathfrak{m}} + h_{\mathfrak{m}+1,\mathfrak{m}}u_{\mathfrak{m}+1}e_{\mathfrak{m}}^{\top};$$

2: **repeat**

3: Use Algorithm 3 to extend the $\mathfrak{m}$th step of $\top$-isotropic Arnoldi decomposition to the $(\mathfrak{m}+p)$th step of $\top$-isotropic Arnoldi factorization:

$$\mathcal{B}U_{\mathfrak{m}+p} = U_{\mathfrak{m}+p}H_{\mathfrak{m}+p} + h_{\mathfrak{m}+p+1,\mathfrak{m}+p}u_{\mathfrak{m}+p+1}e_{\mathfrak{m}+p}^{\top}.$$

4: Use Krylov-Schur restarting scheme [20, 21] to reform a new $\top$-isotropic Arnoldi decomposition with order $\mathfrak{m}$.

5: **until** wanted $\mathfrak{m}$ eigenpairs of $\mathcal{B}$ are convergent

This may result in losing some accuracy of the eigevector z. In order to eliminate this extra computational cost and to prevent the inaccuracy, a generalized $\top$-skew-Hamiltonian implicity-restarted Arnoldi (GTSHIRA) algorithm is proposed in [8]. The idea is to solve the GEP $\widehat{\mathcal{K}}z = \widehat{\mu}\widehat{\mathcal{N}}z$ in (14) directly through bi-reorthogonalization and bi-$\top$-isotropic processes. The GTSHIRA algorithm is based on following two theorems.

Theorem 3.4. [8] *Let $\mathcal{B} \equiv \mathcal{N}_1^{-1}\widehat{\mathcal{K}}\mathcal{N}_2^{-1}$ with $\widehat{\mathcal{N}} = \mathcal{N}_1\mathcal{N}_2$ be $\top$-skew-Hamiltonian. Let $K_j \equiv K_j[\mathcal{B}, u_1]$ be the Krylov matrix with rank$(K_j) = j$. If*

$$\mathcal{N}_2^{-1}K_j = Z_jR_{2,j} \quad \text{and} \quad \mathcal{N}_1K_j = Y_jR_{1,j}$$

are QR-factorizations, where $Z_j, Y_j \in \mathbb{C}^{2n\times j}$ are orthonormal and $R_{2,j}, R_{1,j}$ are nonsingular upper triangular, then we have

$$\widehat{\mathcal{K}}Z_j = Y_j\widehat{H}_j + \widehat{y}_{j+1}e_j^{\top} \tag{18}$$

and

$$\widehat{\mathcal{N}}Z_j = Y_j\widehat{R}_j, \tag{19}$$

where $\widehat{H}_j \in \mathbb{C}^{j\times j}$ is unreduced upper Hessenberg, $\widehat{R}_j \in \mathbb{C}^{j\times j}$ is nonsingular upper triangular, and Y_j and Z_j are $\top$-bi-isotropic such that

$$Y_j^H\widehat{y}_{j+1} = 0 \quad \text{and} \quad Z_j^{\top}\mathcal{J}\widehat{y}_{j+1} = 0,$$

for a suitable $\widehat{y}_{j+1} \in \mathbb{C}^{2n}$.

Theorem 3.5. [8] *Let $\mathcal{B} = \mathcal{N}_1^{-1}\widehat{\mathcal{K}}\mathcal{N}_2^{-1}$ with $\widehat{\mathcal{N}} = \mathcal{N}_1\mathcal{N}_2$ be $\top$-skew-Hamiltonian and $K_n \equiv K_n[\mathcal{B}, u_1]$ be the Krylov matrix with rank$(K_n) = n$. Then there are unitary*

matrices $\mathcal{U}$ *and* $\mathcal{V}$ *satisfying* $\mathcal{V} = \mathcal{J}^\top \mathcal{U} \mathcal{J}$, $\mathcal{U}e_1 = u_1$ *and* $\mathcal{V}e_1 = \mathcal{N}_1 u_1 / \|\mathcal{N}_1 u_1\|_2$ *such that*

$$\mathcal{V}^\top \widehat{\mathcal{K}} \mathcal{U} = \begin{bmatrix} \widehat{H}_n & \widehat{S}_n \\ 0 & \widehat{H}_n^\top \end{bmatrix}, \quad \mathcal{V}^\top \widehat{\mathcal{N}} \mathcal{U} = \begin{bmatrix} \widehat{R}_n & \widehat{T}_n \\ 0 & \widehat{R}_n^\top \end{bmatrix},$$

where $\widehat{H}_n$ *is unreduced upper Hessenberg,* $\widehat{R}_n$ *is nonsingular upper triangular and* $\widehat{S}_n$, $\widehat{T}_n$ *are* $\top$*-skew-symmetric.*

Based on Theorems 3.4 and assuming that the first $(j-1)$th step in GTSHIRA follows the generalized $\top$-isotropic Arnoldi process, i.e.,

$$\widehat{\mathcal{N}} Z_{j-1} = Y_{j-1} \widehat{R}_{j-1}, \tag{20}$$

by comparing the jth columns of both sides in (19) at the jth step, one has

$$\widehat{\mathcal{N}} z_j = \sum_{i=1}^{j-1} \widehat{r}_{ij} y_i + \widehat{r}_{jj} y_j. \tag{21}$$

With (20), (21) can be rewritten as

$$\widehat{r}_{jj}^{-1} z_j = \widehat{\mathcal{N}}^{-1} y_j - \sum_{i=1}^{j-1} \widetilde{r}_{ij} z_i, \tag{22}$$

where

$$[\widetilde{r}_{1j}, \ldots, \widetilde{r}_{j-1,j}]^\top := -\widehat{r}_{jj}^{-1} \widehat{R}_{j-1}^{-1} [\widehat{r}_{1j}, \ldots, \widehat{r}_{j-1,j}]^\top,$$

and $\widehat{r}_{jj}$ in (22) is chosen so that $\|z_j\|_2 = 1$. Since $Z_j^H Z_j = I_j$, the coefficient $\widetilde{r}_{ij}$ in (22) can be evaluated by

$$\widetilde{r}_{ij} = z_j^H \widehat{\mathcal{N}}^{-1} y_j, \quad i = 1, \ldots, j-1.$$

Finally, from (18), the vector y_{j+1} in the jth step of the generalized $\top$-isotropic Arnoldi process is given by

$$\widehat{h}_{j+1,j} y_{j+1} = \widehat{\mathcal{K}} z_j - \sum_{i=1}^{j} \widehat{h}_{ij} y_i,$$

where

$$\widehat{h}_{ij} = y_i^H \widehat{\mathcal{K}} z_j,$$

and $\widehat{h}_{j+1,j} > 0$ is chosen so that $\|y_{j+1}\|_2 = 1$.

Notice that, in theory, z_j and y_{j+1} are orthogonal to $\mathcal{J}\bar{Y}_j$ and $\mathcal{J}\bar{Z}_j$, respectively, in exact arithmetic. However, in practice, roundoff errors may cause $y_i^\top \mathcal{J}^\top z_j$ and $z_i^\top \mathcal{J}^\top y_{j+1}$, $i = 1, \ldots, j$, to be some nonzero tiny values. Therefore, in order to preserve the $\top$-bi-isotropic property of Y_j and Z_j, reorthogonalization of z_j against $\mathcal{J}\bar{Y}_j$ or y_{j+1} against

Algorithm 5 [8] The jth generalized $\top$-isotropic Arnoldi step

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

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

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

Input: $\top$-skew-Hamiltonian matrices $\widehat{\mathcal{K}}$, $\widehat{\mathcal{N}}$, starting vector y_1 and shift value τ.

Output: $\mathfrak{m}$ eigenpairs of $(\widehat{\mathcal{K}},\widehat{\mathcal{N}})$.

1: Use Algorithm 5 with starting vector y_1 to generate a generalized $\top$-isotropic Arnoldi decomposition with order $\mathfrak{m}$:

$$\begin{aligned}\widehat{\mathcal{K}}Z_{\mathfrak{m}}&=Y_{\mathfrak{m}}H_{\mathfrak{m}}+h_{\mathfrak{m}+1,\mathfrak{m}}y_{\mathfrak{m}+1}e_{\mathfrak{m}}^\top,\\ \widehat{\mathcal{N}}Z_{\mathfrak{m}}&=Y_{\mathfrak{m}}R_{\mathfrak{m}}.\end{aligned}$$

2: **repeat**
3: $\quad$Use Algorithm 5 to extend the generalized $\top$-isotropic Arnoldi decomposition with order $\mathfrak{m}$ to order $(\mathfrak{m}+p)$:

$$\begin{aligned}\widehat{\mathcal{K}}Z_{\mathfrak{m}+p}&=Y_{\mathfrak{m}+p}H_{\mathfrak{m}+p}+h_{\mathfrak{m}+p+1,\mathfrak{m}+p}y_{\mathfrak{m}+p+1}e_{\mathfrak{m}+p}^\top,\\ \widehat{\mathcal{N}}Z_{\mathfrak{m}+p}&=Y_{\mathfrak{m}+p}R_{\mathfrak{m}+p}.\end{aligned}$$

4: $\quad$Use Krylov-Schur restarting scheme [20, 21] to reform a new generalized $\top$-isotropic Arnoldi decomposition with order $\mathfrak{m}$.
5: **until** wanted $\mathfrak{m}$ eigenpairs of $(\widehat{\mathcal{K}},\widehat{\mathcal{N}})$ are convergent

Algorithm 7 GE_GTSHIRA/GE_TSHIRA

Input: matrices F, G, M_2 and M_1, shift value τ and the number $\mathfrak{m}$ of desired eigenvalues.

Output: eigenpairs $\{(\gamma_j, [(\psi_{i,j}^{(1)})^\top, (\psi_{\ell,j}^{(1)})^\top]^\top), \ (\gamma_j^{-1}, [(\psi_{i,j}^{(2)})^\top, (\psi_{\ell,j}^{(2)})^\top]^\top)\}_{j=1}^{\mathfrak{m}}$ of (1) where $\gamma_j + \gamma_j^{-1}$ for $j = 1, \ldots, \mathfrak{m}$ are the closest to shift value $\tau + \tau^{-1}$.

1: Compute eigenpairs $\{(\widehat{\mu}_j, z_j \equiv [z_{j1}^\top, z_{j2}^\top]^\top)\}_{j=1}^{\mathfrak{m}}$ of $(\widehat{\mathcal{K}}, \widehat{\mathcal{N}})$ by using GTSHIRA or Compute eigenpairs $\{(\widehat{\mu}_j, \tilde{z}_j)\}_{j=1}^{\mathfrak{m}}$ of $\mathcal{B}$ by using TSHIRA and solve $\mathcal{N}_2[z_{j1}^\top, z_{j2}^\top]^\top = \tilde{z}_j$, for $j = 1, \ldots, \mathfrak{m}$.

2: Compute eigenvalues γ_j and γ_j^{-1} of TPQEP in (2) by solving

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

Compute eigenvectors

$$\psi_{i,j}^{(1)} \equiv \gamma_j^{-1} z_{j1} - z_{j2}, \quad \psi_{i,j}^{(2)} \equiv \gamma_j z_{j1} - z_{j2}$$

corresponding to γ_j, γ_j^{-1}, respectively, for $j = 1, 2, \ldots, \mathfrak{m}$.

3: Compute

$$\psi_{\ell,j}^{(1)} = -M_2^{-1}\left(\gamma_j^{-1} F^\top \psi_{i,j}^{(1)} + G^\top \psi_{i,j}^{(1)}\right), \ \psi_{\ell,j}^{(2)} = -M_2^{-1}\left(\gamma_j F^\top \psi_{i,j}^{(2)} + G^\top \psi_{i,j}^{(2)}\right)$$

for $j = 1, \ldots, \mathfrak{m}$.

$\mathcal{J}\bar{Z}_j$ is needed. Summarizing above processes, we state the jth step of the generalized $\top$-isotropic Arnoldi process in Algorithm 5. The reorthogonalization steps just mentioned are Step 5 and Step 14, respectively, in Algorithm 5. Moreover, the GTSHIRA algorithm based on the generalized $\top$-isotropic Arnoldi process is presented in Algorithm 6 for finding eigenpairs of the matrix pair $(\widehat{\mathcal{K}}, \widehat{\mathcal{N}})$.

In the above TSHIRA and GTSHIRA algorithms, the main costs arise in computing $u_{j+1} = \mathcal{B}u_j$ and solving linear system $\widehat{\mathcal{N}}z_j = y_j$ at the jth $\top$-isotropic and generalized $\top$-isotropic Arnoldi steps, respectively. From (14b), (15) and (16), computing these vectors u_{j+1} and z_j require to solve the following linear systems

$$\mathcal{N}_1 v_1 = b_1, \quad \mathcal{N}_2 v_2 = b_2. \tag{23}$$

By the definitions of $\mathcal{M}$ and $\mathcal{L}$ in (11), we see that solving (23) is equivalent to solve

$$\begin{aligned} (\tau^2 A_1^\top + \tau A_0 + A_1) v_{11} &= b_{11} - \tau b_{12}, \\ v_{12} &= -b_{12} - (A_0 + \tau A_1^\top) v_{11}, \end{aligned} \tag{24}$$

and

$$\begin{aligned} (\tau^2 A_1 + \tau A_0 + A_1^\top) v_{22} &= b_{22} + (A_0 + \tau A_1) b_{21}, \\ v_{21} &= \tau v_{22} - b_{21}, \end{aligned} \tag{25}$$

where $v_1 = [v_{11}^\top, v_{12}^\top]^\top$, $v_2 = [v_{21}^\top, v_{22}^\top]^\top$, $b_1 = [b_{11}^\top, b_{12}^\top]^\top$ and $b_2 = [b_{21}^\top, b_{22}^\top]^\top$. By the definitions of A_0 and A_1, it holds that

$$\tau^2 A_1^\top + \tau A_0 + A_1 = (G + \tau F) M_2^{-1} (F^\top + \tau G^\top) - \tau M_1 \tag{26}$$

and

$$\tau^2 A_1 + \tau A_0 + A_1^\top = (F + \tau G) M_2^{-1} (G^\top + \tau F^\top) - \tau M_1. \tag{27}$$

Let $M_1 = LU$ be the LU factorization of M_1 and set

$$E_1 = L^{-1} \left(\frac{1}{\tau} G + F \right), \quad E_2 = U^{-\top} (F + \tau G). \tag{28}$$

By the Sherman-Morrison-Woodbury formula, (26) and (27) imply that

$$\left(\tau^2 A_1^\top + \tau A_0 + A_1 \right)^{-1} = -\frac{1}{\tau} U^{-1} \left[I + E_1 \left(M_2 - E_2^\top E_1 \right)^{-1} E_2^\top \right] L^{-1}$$

and

$$\left(\tau^2 A_1 + \tau A_0 + A_1^\top \right)^{-1} = -\frac{1}{\tau} L^{-\top} \left[I + E_2 \left(M_2 - E_1^\top E_2 \right)^{-1} E_1^\top \right] U^{-\top},$$

respectively.

Obviously, from (28), we need m forward substitutions and m backward substitutions to obtain E_1 and E_2, respectively. Furthermore, in addition to the cost in solving small linear systems $\left(M_2 - E_2^\top E_1 \right)^{-1}$ and $\left(M_2 - E_1^\top E_2 \right)^{-1}$, only two forward substitutions (L^{-1}, $U^{-\top}$) and two backward substitutions ($L^{-\top}$, U^{-1}) are required to obtain the solutions of (24) and (25) for generating Krylov subspace at each iterative step. Recall that, for GE_SDA and GE_SA, in order to form the matrices A_0 and A_1 in (3), one needs to compute $M_1^{-1}F$ and $M_1^{-1}G$ which require $2m$ forward and backward substitutions. Since the shift-and-invert Arnoldi method is known to converge very fast when a proper shift is known, the overall computational costs of GE_GTSHIRA and GE_TSHIRA, including computing E_1, E_2 and solving linear systems in each iterative steps, can be only about half amount of the computation cost needed in GE_SDA and GE_SA. Our numerical results in Table 5 confirm this observation. Finally, we summarize the process of applying TSHIRA/GTSHIRA to solve the GEP in (1) in Algorithm 7 and show the comparison of the computational costs for TSHIRA and GTSHIRA in Table 2.

4. Numerical results

In this section, we tests the above mentioned four types of structure preserving algorithms on computing the dispersion diagram of the frequency that are close to the stopping frequency of the SAW filter. The piezoelectric substrate of the filter is made of 15^o rotated quartz. The configuration of our computational domain is shown in Figure 1 where the domain width $\overline{AB}$ and height $\overline{CD}$ are set to be 10^{-6} and 3×10^{-6}, respectively, the ratio of the electrode width $\overline{EF}$ versus the domain width is set to be $\frac{1}{2}$ and the ratio of the electrode thickness $\overline{DE}$ versus the domain height is $\frac{1}{15}$. In our numerical studies, the viscous damping coefficient κ_1 is set to be 10^{-14} and the mass damping coefficient κ_2 is taken as $1 - \kappa_1$ to account for the effect from the electrode weight. All computations are carried out in MATLAB 2010b on a HP workstation with an Intel Quad-Core Xeon X5570 2.93GHz and 60 GB main memory, using IEEE double-precision floating-point

		TSHIRA	GTSHIRA
Compute E_1, E_2	$M_1 = LU$	1	1
	$F + \xi G$	2	2
	Solve $Lx = b_1$	m	m
	Solve $U^\top y = c_2$	m	m
	$E_2^\top E_1$ (flops)	$2m^2n$	$2m^2n$
jth step Arnoldi	Solve $Lx = b_1$, $L^\top y = c_1$	1	1
	Solve $Ux = b_2$, $U^\top y = c_2$	1	1
	Compute Fd_1, $F^\top c_1$, Gd_2, $G^\top c_2$	3	3
	Compute $M_1 b$	2	2
	Compute $E_1 d_1$, $E_1^\top c_1$, $E_2 d_2$, $E_2^\top c_2$	1	1
	Saxpy and inner products (flops)	$8nj + 15n$	$16nj + 18n$
Schur restarting	Matrix product (flops)	$2(\mathfrak{m}+p)^2n$	$4(\mathfrak{m}+p)^2n$

Table 2: Computational costs for TSHIRA and GTSHIRA.

arithmetic.

Suppose $\mathfrak{m}$ reciprocal pairs of eigenvalues near $\mathbb{U}$ are desired. For TSHIRA and GTSHIRA, the restart procedure will be activated when the desired eigenpairs don't converge before the dimension of the Krylov subspace reaches $5\mathfrak{m}$. This is done by setting the value of p in Step 4 of Algorithms 4 and 6 to $4\mathfrak{m}$. In the following discussion, we take $\mathfrak{m} = 5$ and the matrix dimensions of C_i and C_b are $n = 63960$ and $m = 723$, respectively. An example of computed reciprocal eigenpairs near $\mathbb{U}$ at frequency $\omega = 1.2757/(2\pi)\times 10^{10}$ is shown in Figure 2. The dispersion diagrams of the attenuation constant α and the propagation constant β associated with the eigenvalue $\lambda(\omega)$ are shown in Figure 3, for frequency ω around the stopping band, where the eigenpair most close to -1 on the complex plane is plotted.

4.1. Accuracy of structure-preserving eigensolvers

In this subsection, we compare the accuracy of the eigenpairs, computed by structure-preserving Algorithms 1, 2 and 7, respectively, for the GEP (8). Recall that the Krylov subspace $\mathcal{U}_j$ generated by the $\top$-Hamiltonian matrix $\mathcal{B}$ is automatically $\top$-isotropic in Theorem 3.2, and the subspaces $\mathcal{Z}_j$ and $\mathcal{Y}_{j+1}$ generated in Theorem 3.4 are automatically $\top$-bi-isotropic. As mentioned in Subsections 3.3 and 3.4, isotropic re-orthogonalization in Step 3 of Algorithm 3 and Steps 5 and 5 of Algorithm 5 is important in maintaining the $\top$-isotropic property. Moreover, Theorem 3.3 and 3.5 both show that the multiplicities of eigenvalues of $(\mathcal{K}, \mathcal{N})$ are all even. In other words, no duplicate eigenpairs need to be computed theoretically when the $\top$-isotropic property is kept during the computation. On the other hand, without the isotropic re-orthogonalization process, extra computation cost can arise in computing the duplicate eigenpairs. We would like to address this issue by numerical studies shown in the following. We also like to point out that the accuracy of the computed eigenpairs can be affected by different approaches in isotropic re-orthogonalization.

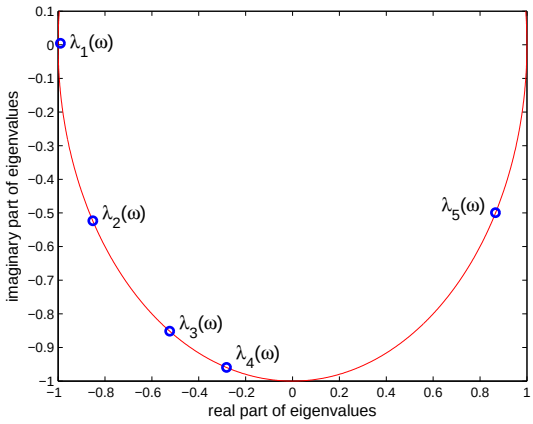


Figure 2: The distribution of the eigenvalues which are close to and inside of $\mathbb{U}$.

First, let us denote the algorithm that applying TSHIRA without the re-symplectic process as T_NoSymp and the algorithm that applying GTSHIRA without these re-bi-isotropic processes as GT_NoBiIso. In Table 3, the convergent eigenvalues obtained by T_NoSymp and GT_NoBiIso at frequency $\omega = 1.2757/(2\pi) \times 10^{10}$ are listed. Obviously, one can see that, in case only two eigenpairs $\{(\lambda_1, \lambda_1^{-1}), (\lambda_2, \lambda_2^{-1})\}$ are needed here, the algorithms T_NoSymp and GT_NoBiIso return four convergent eigenpairs in which two of them are indeed the duplicated pairs.

	T_NoSymp	GT_NoBiIso
$(\lambda, \frac{1}{\lambda})$	$-0.85175542558 - 0.52335156640\imath$	$-0.85175542559 - 0.52335156640\imath$
	$-0.85228028786 + 0.52367406214\imath$	$-0.85228028785 + 0.52367406213\imath$
	$-0.85175542557 - 0.52335156639\imath$	$-0.85175542556 - 0.52335156641\imath$
	$-0.85228028787 + 0.52367406214\imath$	$-0.85228028786 + 0.52367406216\imath$
	$-0.98999503056 + 0.00448884999\imath$	$-0.98999503056 + 0.00448884999\imath$
	$-1.01008531402 - 0.00457994365\imath$	$-1.01008531402 - 0.00457994365\imath$
	$-0.98999503056 + 0.00448884999\imath$	$-0.98999503056 + 0.00448884999\imath$
	$-1.01008531402 - 0.00457994365\imath$	$-1.01008531402 - 0.00457994365\imath$

Table 3: Convergent eigenvalues computed by T_NoSymp and GT_NoBiIso at $\omega = 1.2757/(2\pi) \times 10^{10}$.

Next, let's compare the accuracy of the computed eigenpairs obtained from three different isotropic re-orthogonalization approaches in GTSHIRA. One or two steps of re-bi-isotropic process can be performed by the for-loops in Steps 5-5 and 5-5.

To distinguish among various re-bi-isotropic processes, we use notations "FullIso", "zIsoY" and "yIsoZ" defined as follows:

- FullIso: Algorithm 5 with two for-loops in Steps 5-5 and 5-5.
- zIsoY: Algorithm 5 with one for-loop in Steps 5-5 and omitting for-loop in Steps 5-5.

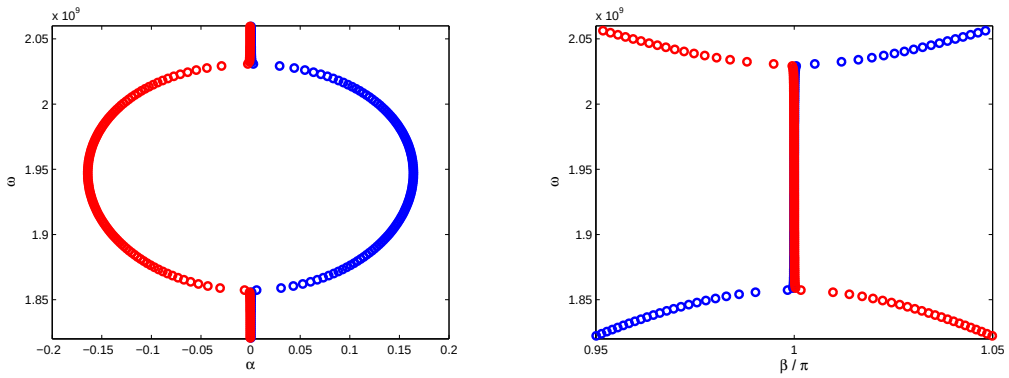


Figure 3: Dispersion diagrams of α and β near the stopping band.

- yIsoZ: Algorithm 5 with one for-loop in Steps 5-5 and omitting for-loop in Steps 5-5.

To measure the accuracy of computed eigenpairs of (8), we consider the relative residual of an eigenpair (λ, ψ) where $\psi = [\psi_i^\top, \psi_\ell^\top]^\top$ which is defined as following:

$$\frac{\left\| \begin{bmatrix} C_i & C_{i\ell} \\ C_{ir}^\top & 0 \end{bmatrix} \psi - \lambda \begin{bmatrix} 0 & C_{ir} \\ C_{i\ell}^\top & C_b \end{bmatrix} \psi \right\|_F}{\left\| \begin{bmatrix} C_i & C_{i\ell} \\ C_{ir}^\top & 0 \end{bmatrix} \right\|_F \|\psi\|_F + |\lambda| \left\| \begin{bmatrix} 0 & C_{ir} \\ C_{i\ell}^\top & C_b \end{bmatrix} \right\|_F \|\psi\|_F},$$

here $\| * \|_F$ is the Frobenius matrix norm. The relative residuals of the convergent eigenpairs computed by "FullIso", "zIsoY" and "yIsoZ" are shown in Figure 4. From the numerical results in Figure 4, we see that the accuracy of the convergent eigenpairs computed by "yIsoZ" is higher than those by "FullIso" and "zIsoY". This result can be explained from the accumulation of the errors in the equalities (18) and (19). Let $\xi_{j,K} \equiv \|\widehat{\mathcal{K}} Z_j - Y_j \widehat{H}_j - \widehat{h}_{j+1,j} y_{j+1} e_j^\top\|_2$ and $\xi_{j,N} \equiv \|\widehat{\mathcal{N}} Z_j - Y_j \widehat{R}_j\|_2$, denote these errors in the jth iteration. The error $\xi_{j,N}$ depends on the accuracy of the solution of the linear systems in (23). If z_j is reorthogonalized to $\mathcal{J}\bar{Y}_j$, then the error produced by this reorthogonization will reduce the accuracy of $\xi_{j,N}$. Therefore, $\xi_{j,N}$ produced by "FullIso" and "zIsoY" are greater than that by "yIsoZ" as shown in Figure 5.(a). On the other hand, the error $\xi_{j,K}$ only depends on the accuracy of matrix product vector and vector inner product. Obviously, the amount of $\xi_{j,K}$ is much less than the amount of $\xi_{j,N}$. Consequently, even though the accuracy of $\xi_{j,K}$ can reduced by the errors from reorthogonalization y_{j+1} to $\mathcal{J}\bar{Z}_j$ as shown Figure 5.(b), the reorthogonalization process "yIsoZ" is much accurate than the "FullIso" and "zIsoY" reorthogonalization processes.

Finally, we compare the accuracy of the eigenpairs $(\lambda(\omega), u(\omega))$ obtained from GE_SDA, GE_SA, GE_TSHIRA and GE_GTSHIRA with "yIsoZ" re-bi-isotropic process. The relative residuals resulted from these algorithms in computing four reciprocal eigenpairs $(\lambda_i(\omega), u_i(\omega))$, for $i = 1, \ldots, 4$, that are closest to -1 on the complex plane are plotted

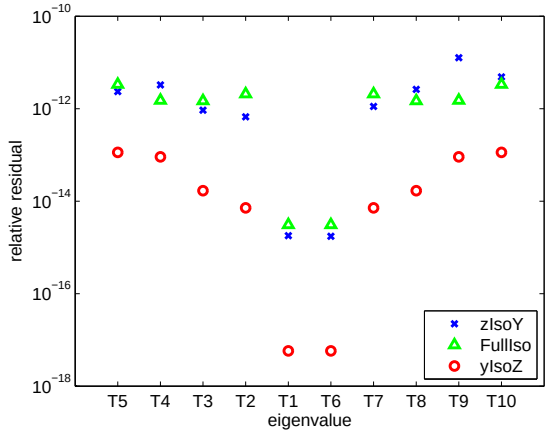


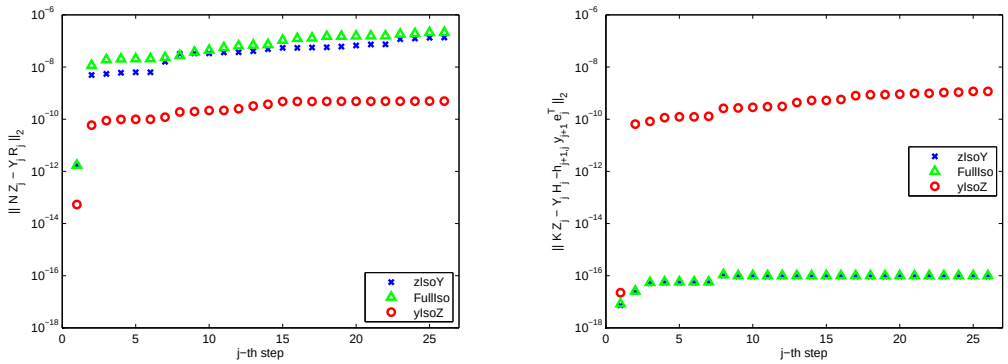
Figure 4: The relative residual of the computed eigenpairs produced by different re-bi-isotropic processes in Algorithm 5 with shift value $\tau = -0.99$.

in Figure 6 for each frequency ω near the stopping band. Obviously, one can see that the accuracy of the eigenpairs obtained from GE_SDA and GE_SA are higher than those obtained by GE_TSHIRA and GE_GTSHIRA.

4.2. Comparison with computational costs

In this subsection, we discuss the computational costs of structure-preserving Algorithms 1, 2 and 7 in computing $\mathfrak{m} = 5$ desired eigenpairs. Our numerical results show that the desired eigenpairs are convergent within $5\mathfrak{m}$ $\top$-isotropic Arnoldi steps without restart for GE_TSHIRA and GE_GTSHIRA. On the other hand, it requires total 18 iterations to obtain a convergent X_k in Steps 3.1-3.1 for the SDA algorithm. As we mentioned in Subsection 3.3, the number of forward and backward substitutions needed for GE_TSHIRA and GE_GTSHIRA is only about half the amount of these substitutions that needed to transform the GEP into TPQEP in GE_SDA and GE_SA. Since only additional 25 forward substitutions and backward substitutions are needed in GE_TSHIRA and GE_GTSHIRA for solving linear systems $Lx = b$ and $Uy = c$, we expect GE_TSHIRA and GE_GTSHIRA to be more robust than GE_SA and GE_SDA. The following numerical results support this observation.

To give an overall comparison for GE_SDA, GE_SP, GE_TSHIRA and GE_GTSHIRA, in Table 4, computational intensive items in these algorithms are listed in the first column and the sums of the CPU times for each associated item are listed in the other four columns. From the results in Table 4, the dominant computational costs in GE_TSHIRA and GE_GTSHIRA are the costs for computing E_1, E_2, $E_2^\top E_1$ and LU factorization of C_i. For GE_SDA and GE_SA, the cost in computing the matrices A_0 and A_1 of the TPQEP is the main cost comparing to the other costs. Obviously, the numbers shown in Table 4 indicate that GE_TSHIRA and GE_GTSHIRA are more efficient than GE_SDA and GE_SA. We also plot the overall CPU times for GE_SDA and GE_GTSHIRA with



(a) $\|\widehat{\mathcal{N}}Z_j - Y_j\widehat{R}_j\|_2$ (b) $\|\widehat{\mathcal{K}}Z_j - Y_j\widehat{H}_j - \widehat{h}_{j+1,j}y_{j+1}e_j^\top\|_2$

Figure 5: The errors of the equalities in (18) and (19) for "FullIso", "zIsoY" and "yIsoZ".

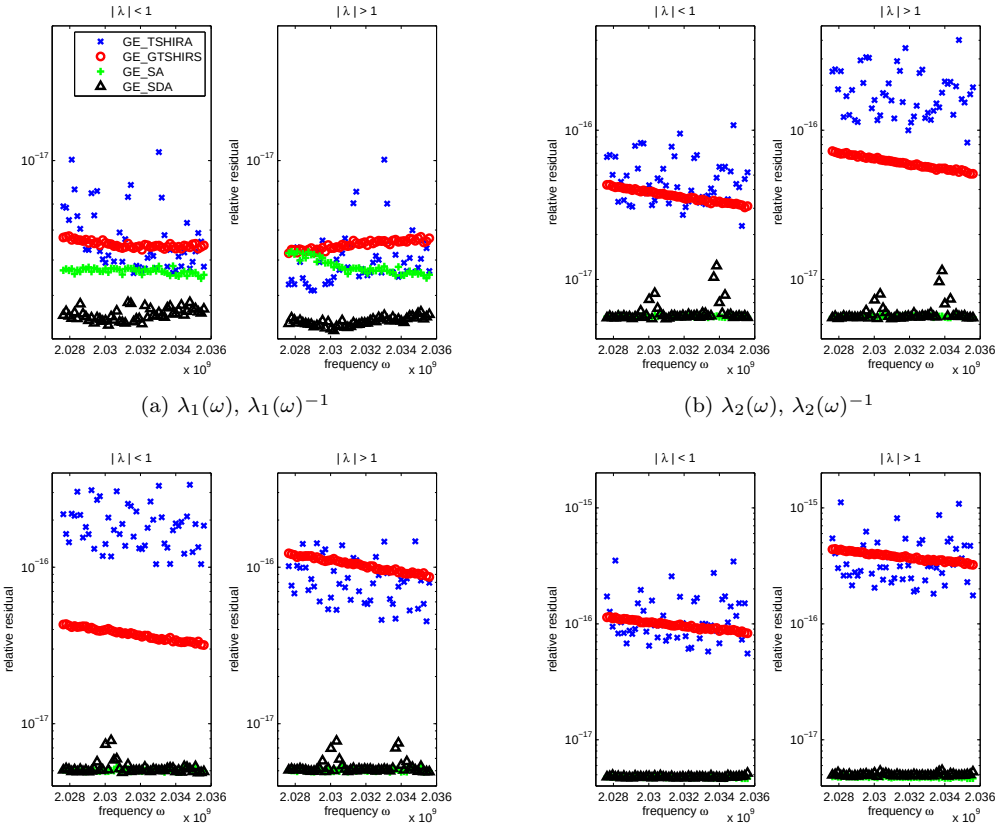
frequency from $1.274/(2\pi) \times 10^{10}$ to $1.279/(2\pi) \times 10^{10}$ in Figure 7. From Figure 7, one can see that the total CPU times needed in GE_SDA and GE_SA are 40% more than the CPU time needed in GE_TSHIRA and GE_GTSHIRA for computing 5 desired eigepairs.

	TSHIRA	GTSHIRA	SDA	SA
Compute $C_i = LU$	191.31	191.31	191.31	191.31
Compute E_1, E_2, $E_2^\top E_1$	243.75	243.75		
Compute A_0, A_1			533.94	533.94
Solve dense TPQEP			34.145	
Solve $Lx = b_1$	4.9850	4.9850	1.9940	1.9940
Solve $Ux = b_2$	3.9775	3.9775	1.5910	1.5910
Solve $U^\top y = c_2$	33.597	27.998		
Solve $L^\top y = c_1$	36.300	30.250		
Compute $E_1 d_1$, $E_2^\top c_2$	4.9150	4.9150		
Compute $E_1^\top c_1$, $E_2 d_2$	5.7930	4.8275		

Table 4: CPU times (sec.) for GE_TSHIRA, GE_GTSHIRA, GE_SDA and GE_SA.

5. Conclusion

In this paper, we have discussed the structure-preserving methods for solving the generalized eigenvalue problem arising in the surface acoustic wave propagation on a simple resonator with an interdigital transducer (IDT) where electrodes are arranged periodically on piezoelectric substrates (PZT) such as 15^o rotated Quartz. With given periodic boundary conditions, the eigenvalues of the GEP appear in the reciprocal pairs (λ, λ^{-1}). In order to preserve the reciprocal relationship of the eigenvalues, the GEP is transformed to two types of T-palindromic quadratic eigenvalue problems, one with



(a) $\lambda_1(\omega)$, $\lambda_1(\omega)^{-1}$ (b) $\lambda_2(\omega)$, $\lambda_2(\omega)^{-1}$

(c) $\lambda_3(\omega)$, $\lambda_3(\omega)^{-1}$ (d) $\lambda_4(\omega)$, $\lambda_4(\omega)^{-1}$

Figure 6: Relative residuals for GE_SDA, GE_SA, GE_TSHIRA and GE_GTSHIRA with shift value $\tau = -0.89$.

large coefficient matrices and the other with small coefficient matrices. The structure-preserving algorithms GE_SDA and GE_SA in Algorithms 1, 2 are employed to solve the TPQEP (3) with small-size coefficient matrices and GE_TSHIRA and GE_GTSHIRA in Algorithms 7 are employed to solve the TPQEP (4) with large-size coefficient matrices.

In finding the five eigenpairs that are near $\mathbb{U}$ and close to -1, we observed duplicate eigenpairs appear when applying GE_TSHIRA and GE_GTSHIRA without re-symplectic and re-bi-isotropic processes, respectively. On the other hand, no duplicate eigenpairs are observed when re-sympletic and re-bi-isotropic processes are integrated in GE_TSHIRA and GE_GTSHIRA. Three different re-bi-isotropic processes in GE_GTSHIRA has been tested. We have found that using the re-bi-isotropic processes in Steps 5-5 of Algorithm 5 achieves the best accuracy. Moreover, our numerical results show that the relative residuals of the eigenpairs produced by GE_SDA/GE_SA and GE_TSHIRA/GE_GTSHIRA can be less than 10^{-17} and 10^{-15}, respectively. Although the accuracy of GE_SDA

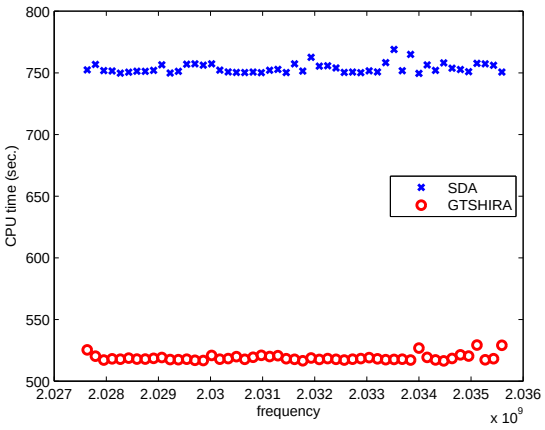


Figure 7: CPU times for GE_SDA and GE_GTSHIRA.

and GE_SA is marginally higher than that of GE_TSHIRA and GE_GTSHIRA, we further found that the total CPU times required for computing the five desired eigepairs by GE_SDA and GE_SA are about 40% more than that are required by GE_TSHIRA and GE_GTSHIRA. Therefore, by transforming the GEP into the TPQEP (4), the structure-preserving Arnoldi type algorithm GE_TSHIRA or GE_GTSHIRA with one "re-sympletic" or " re-bi-isotropic" processes provide an accurate and efficient way in finding the reciprocal eigenpairs of the GEP (1).

Acknowledgments

This work is partially supported by the National Science Council and the National Center for Theoretical Sciences in Taiwan. The author Chin-Tien Wu like to thank the support from National Science Council under the grant number 99-2115-M-009-001.

References

[1] H. Allik and T. Hughes. Finite element method for piezoelectric vibration. *Int. J. Numer. Methods Eng.*, 2:151–157, 1970.

[2] M. B. Angel, M. I. Rocha-Gaso, M. I. Carmen, and A.V. Antonio. Surface generated acoustic wave biosensors for detection of pathogens: A review. *Sensors*, 9:5740–5769, 2009.

[3] M. Buchner, W. Ruile, A. Dietz, and R. Dill. FEM analysis of the reflection coefficient of SAWS in an infinite periodic array. In *Proc. IEEE Ultrason. Symp.*, 371-375, 1991.

[4] C. K. Campbell. *Surface Acoustic Wave Devices for Mobile and Wireless Communications*. Academic Press, INC., 1998.

[5] 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.

[6] W. W. Lin E. K. Chu, T. M. Huang and C. T. Wu. Palindromic eigenvalue problems: a brief survey. *Taiwan J. Math.*, 14, 3A:743–779, 2010.

[7] A Hilliges, C. Mehl, and V. Mehrmann. On the solution of palindramic eigenvalue problems. In *Proceedings 4th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS)*, Jyväskylä, Finland, 2004.

[8] 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, 2009.

[9] T.-M. Huang, W.-W. Lin, and C.-T. Wu. Structure-preserving Arnoldi-type algorithms for solving palindromic quadratic eigenvalue problems in leaky surface wave propagation. Technical report, National Center for Theoretical Sciences, National Tsing Hua University, Taiwan, Preprints in Mathematics 2011-2-001, 2011.

[10] M. Koshiba, S. Mitobe, and M. Suzuki. Finite-element solution of periodic waveguides for acoustic waves. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, 34, No. 4:472–477, 1987.

[11] R. Lerch. Simulation of piezoelectric devices by two- and three-dimensional finite elements. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, 37, No. 2:1990, 1990.

[12] W.-W. Lin. A new method for computing the closed-loop eigenvalues of a discrete-time algebraic Riccatic equation. *Linear Alg. Appl.*, 96:157–180, 1987.

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

[14] D.S. Mackey, N. Mackey, C. Mehl, and V. Mehrmann. Vector spaces of linearizations for matrix polynomials. *SIAM J. Matrix Anal. Appl.*, 28:971–1004, 2006.

[15] V. Mehrmann and D. Watkins. Structure-preserving methods for computing eigenpairs of large sparse skew-Hamiltonian/Hamiltonian pencils. *SIAM J. Sci. Comput.*, 22:1905–1925, 2001.

[16] M. Mohamed, EL Gowini, and W. A. Moussa. A finite element model of a mems-based surface acoustic wave hydrongen sensor. *Sensors*, 10:1232–1250, 2010.

[17] R. V. Patel. On computing the eigenvalues of a symplectic pencil. *Linear Alg. Appl.*, 188:591–611, 1993.

[18] C. Schröder. A QR-like algorithm for the palindromic eigenvalue problem. Technical report, Preprint 388, TU Berlin, Matheon, Germany, 2007.

[19] C. Schröder. URV decomposition based structured methods for palindromic and even eigenvalue problems. Technical report, Preprint 375, TU Berlin, MATHEON, Germany, 2007.

[20] G.W. Stewart. A Krylov–Schur algorithm for large eigenproblems. *SIAM J. Matrix Anal. Appl.*, 23:601–614, 2001.

[21] G.W. Stewart. Addendum to "A Krylov–Schur algorithm for large eigenproblems". *SIAM J. Matrix Anal. Appl.*, 24:599–601, 2002.

[22] S. Zaglmayr. Eigenvalue problems in saw-filter simulations. Diplomarbeit, Institute of Computational Mathematics, Johannes Kepler University Linz, Linz, Austria, 2002.

Structure-Preserving Arnoldi-Type Algorithm for Solving Eigenvalue Problems in Leaky Surface Wave Propagation

Tsung-Ming Huang[a], Wen-Wei Lin[b], Chin-Tien Wu[c,*]

[a]*Department of Mathematics, National Taiwan Normal University, Taipei 116, Taiwan. E-mail:* min@ntnu.edu.tw.
[b]*Department of Mathematics, National Taiwan University, Taipei 106, Taiwan. E-mail:* wwlin@math.ntu.edu.tw.
[c]*Department of Applied Mathematics, National Chiao Tung University, Hsinchu 300, Taiwan. E-mail:* ctw@math.nctu.edu.tw.

Abstract

We study the generalized eigenvalue problems (GEP) arising from modeling leaky surface waves propagation in a acoustic resonator with infinitely many periodically arranged interdigital transducers. The constitution equations are discretized by finite element method with mesh refinement along the electrode interface and corners. The associated GEP is then transformed to a T-palindromic quadratic eigenvalue problem so that the eigenpairs can be accurately and efficiently computed by using a structure-preserving algorithm incorporating a generalized T-skew-Hamiltonian implicity-restarted Arnoldi method. Our numerical results show that the eigenpairs produced by the proposed structure-preserving method not only preserve the reciprocal property but also possess high efficiency and accuracy.

Key words: Leaky SAW, structure-preserving, palindromic quadratic eigenvalue problem, GTSHIRA, mesh refinement

1. Introduction

Waveguide devices have been widely used in controlling and interconnecting guided electromagnetic waves. Advances in the thin film technology and efficient transducers further encourage investigations on more sophisticated waveguide concepts in acoustic system. Acoustic wave guide devices are widely employed in applications including telecommunication filters [8, 25] and sensor technologies [2]. One of the basic element in most acoustic wave filters is a resonator which generally consists of reflectors externally coupled through one or two interdigital transducers (IDT). The IDT is primarily made by depositing periodical metallic grating electrodes on a piezoelectric film substrate as shown

*Corresponding author. Tel: +886-3-5712121-ext-56424

in Figure 1(a). Extensive theoretical and experimental works have been done specially on the Rayleigh surface acoustic wave [4, 9, 24, 25]. Finite element simulations of piezoelectric devices in two dimension (2D) and three dimension (3D) have been studied by Allik, Koshiba, Lerch, Buchner, Mohamed and others etc., [1, 6, 17, 19]. In the filter design, it is important to know the stop band width and the center frequency f_c of the filter where $f_c = v_s/\lambda_s$ here v_s and λ_s are the wave velocity and wave length of the incident wave. The center frequency and stop band width can visually be determined by plotting the dispersion diagram in which an eigenvalue problem associated with each frequency in the search range has to be solved.

Due to slower propagation velocity of the Rayleigh SAW, filters based on Rayleigh SAW design are usually limited to an operational frequency range less than 1GHZ. For frequency higher than 1GHZ, more recent attention has been paid to the so-called leaky surface acoustic wave (LSAW) because of its faster propagation speed in crystal cuts such as 64^o YX-$LiNbO_3$ and 36^o YX-$LiTaO_3$, and higher electromechanical coupling and minimal propagation loss in crystal cuts such as $40^o \sim 42^o$ YX-$LiTaO_3$ [8]. Searching a better crystal cut among various piezoelectric substrates (PZT) to increase LSAW velocity becomes one of the major issues in high frequency filter design. For each crystal cut, one has to solve many eigenvalue problems to plot the dispersion diagram. An efficient and accurate algorithm for solving eigenvalue problem resulted from mathematical model of a LSAW resonator is desired.

The eigenvalue problem obtained from the finite element modeling of the SAW or LSAW resonance can be represented as

$$\begin{bmatrix} M_1 & G \\ F^\top & 0 \end{bmatrix} \begin{bmatrix} \psi_i \\ \psi_\ell \end{bmatrix} + \lambda \begin{bmatrix} 0 & F \\ G^\top & M_2 \end{bmatrix} \begin{bmatrix} \psi_i \\ \psi_\ell \end{bmatrix} = 0, \tag{1}$$

where $M_1^\top = M_1 \in \mathbb{C}^{n\times n}$, $M_2^\top = M_2 \in \mathbb{C}^{m\times m}$, F and $G \in \mathbb{C}^{n\times m}$ with $m \ll n$, and the supscript "$\top$" denotes the complex transpose. The scalar $\lambda \in \mathbb{C}$ is called the eigenvalue of (1) and the nonzero vector $\begin{bmatrix} \psi_i^\top & \psi_\ell^\top \end{bmatrix}^\top$ is the associated eigenvector. The generalized eigenvalue problem (GEP) (1) can be reformulated into a T-palindromic quadratic eigenvalue problem (TPQEP) of the form

$$\mathcal{P}(\lambda)\psi_i \equiv (\lambda^2 A_1^\top + \lambda A_0 + A_1)\psi_i = 0 \tag{2a}$$

with

$$A_1^\top = FM_2^{-1}G^\top, \quad A_0 = FM_2^{-1}F^\top + GM_2^{-1}G^\top - M_1. \tag{2b}$$

By taking the transpose of $\mathcal{P}(\lambda)$ in (2a) and multiplying it by $1/\lambda^2$ it is easily seen that the eigenvalues of $\mathcal{P}(\lambda)$ appear in reciprocal pairs $(\lambda, 1/\lambda)$ (including 0 and ∞).

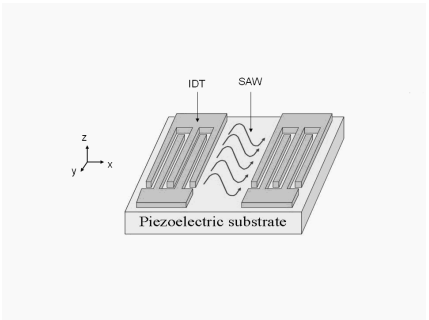
The GEP (1) can be solved by traditional methods such as the QZ and Arnoldi method. However, the reciprocal property of the eigenvalues of (1) can be destroyed easily and large numerical errors can be generated in computation [16]. To remedy the drawback, we transform the GEP (1) into the TPQEP (2a) so that the desired eigenpairs can be computed by structure-preserving methods [11, 13, 15, 21, 22, 29, 30]. For solving the TPQEP with small and dense coefficient matrices A_0 and A_1, a structure-preserving doubling algorithm for solving (2a) was developed in [11] via the computation of a solvent of a nonlinear matrix equation associated with (2a). Another structure-preserving algorithm based on $(\mathcal{S} + \mathcal{S}^{-1})$-transform [20] and Patel's approach [27] was developed in [15]. For problems with large and sparse matrices A_0 and A_1, a structure-preserving algorithm using the $(\mathcal{S} + \mathcal{S}^{-1})$-transform and the implicity-restarted shift-and-invert Arnoldi method was also developed for searching eigenvalues in a specified region of interests [15].

In this paper, we apply the generalized T-skew-Hamiltonian implicity-restarted Arnoldi method developed in [15] to solve the TPQEP (2a). Based on the shift-and-invert technique, the desired eigenpairs can be easily extracted. For solving the linear systems, although the coefficient matrices A_0 and A_1 in (2b) is large but not sparse, we derive a new formula by using the Sherman-Morrison-Woodbury formula so that the corresponding linear system can be efficiently solved. Comparing with the traditional Arnoldi method, our proposed structure-preserving method not only preserve the reciprocal property but also possess high efficiency and accuracy.

This paper is organized as follows. We shall first introduce finite element modeling for a simple resonator in Section 2. In Section 3, we introduce the efficient structure-preserving algorithm to solve the large and sparse generalized eigenvalue problems resulted from our FEM model. Our numerical experiments in Section 4 show that the proposed structure-preserving algorithm for solving the GEP in (1) is efficient and accurate. Finally, we conclude the paper in Section 5.

2. Finite Element Model for SAW

In contrast to the well known Rayleigh waves which consists of partial longitudinal waves and shear waves, the LSAW mainly propagates in the shear direction on the sagittal plane and is trapped at substrate surface and satisfies the stress free boundary condition on the surface. These properties allow one to reduce the general mode analysis in 3D to a 2D problem as shown in Figure 1(b) [28]. Furthermore, the boundary conditions for displacement can naturally be set to be rigid on the bottom boundary and stress-free on the top surface, and the boundary conditions for the electric potential can be set to be short-circuited for the electrodes on the top boundary and open-circuited elsewhere [10]. As proved in Auld's book [4], these boundary conditions guarantee the mode orthogonality and further ensure the mode excitation is determined



(a) A standard configuration of SAW resonators

(b) A 2D model of a LSAW resonator on the sagittal plane

Figure 1:

by the applied traction force and potential on the free surface. Therefore, on the sagittal plane, the usual 2D mode analysis can be applied to analyze the LSAW on the resonators with IDTs. In the following, we only consider the LSAW resonator on a 2D plane (the sagittal plane associated with crystal cuts).

To model the wave propagation in an infinite domain with periodically arranged electrodes, thanks to the Floquet-Bloch Theorem, one can reduce the problem to a single cell domain with one IDT by assuming the wave ψ is quasi-periodic of the form $\psi(x_1, x_2) = \psi_p(x_1, x_2)e^{(\alpha+i\beta)x_1}$ where x_1 is the wave propagation direction, p is the length of the unit cell (i.e. the periodic interval), α and β are the attenuation and phase shifts along the wave propagation direction, respectively, and ψ_p satisfies $\psi_p(x_1 + p, x_2) = \psi_p(x_1, x_2)$. Let Ω denote the PZT with a single IDT as shown in Figure 2, and Γ_l and Γ_r denote the left and right boundary segments of Ω. For the general anisotropic PZT substrates, under the assumption of linear piezoelectric coupling, the elastic and electric fields interact following the general material constitutions below

$$\begin{aligned} T &= c^E S - e^\top E, \\ D &= eS + \varepsilon^S E, \end{aligned} \tag{3}$$

where vectors T, S, D and E are the mechanical stress, strain, dielectric displacement and the electric field, respectively, and the matrices c^E, ε^S and e are the elasticity constant, dielectric constant and piezoelectric constant matrices measured at constant electric and constant strain fields at constant temperature. For various crystal cut of the PZT, the material constant matrices c^E, ε^S and e depend on the Euler angle θ of the cut. By applying the Bond strain transformation matrix N_θ [5] and the usual coordinates transformation matrix M_θ to the strain field and electric field, respectively, the material constant matrices for the cut angle θ can be obtained by

$$c^E := [N_\theta]^\top c_0^E [N_\theta], \quad e := [M_\theta]^\top e_0 [N_\theta], \quad \text{and,} \quad \varepsilon^S := [M_\theta]^\top \varepsilon_0^S [M_\theta],$$

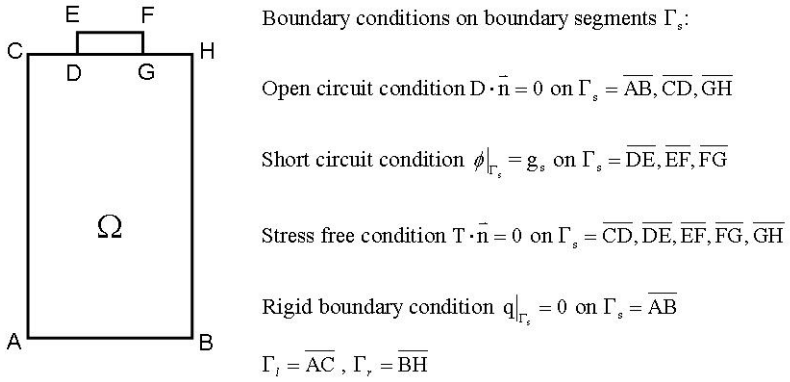


Figure 2: A 2D single cell domain of a LSAW resonator and boundary conditions

here c_0^E, e_0, and ε_0^S denote the material constant matrices of the crystal cut at Euler angle $\theta = [0^o, 0^o, 0^o]$.

By applying the virtual work principle to the equation (3), the equilibrium state under the external body force f, the electrical field g and the above mentioned boundary conditions of the LSAW resonator, we have

$$\begin{aligned}
&\int_\Omega [\delta S]^\top C^E [S]\, dV + \int_\Omega [\delta S]^\top e^\top [\nabla\phi]\, dV + \int_\Omega [\nabla\delta\phi]^\top e[S]\, dV \\
&- \int_\Omega [\nabla\delta\phi]^\top \varepsilon^S [\nabla\phi]\, dV + \int_\Omega \delta q^\top \rho\ddot{q}\, dV \\
&= \int_\Omega \delta q^\top f\, dV + \int_\Omega [\nabla\delta\phi]^\top g\, dV + \int_{\Gamma_l\cup\Gamma_r} \delta q^\top (T\cdot\vec{n})\, dA + \int_{\Gamma_l\cup\Gamma_r} \delta\phi^\top (D\cdot\vec{n})\, dA.
\end{aligned} \tag{4}$$

Here, ρ is the mass density, $\vec{n}$ is the boundary normal, $q = [u, v, w]^\top$ is the displacement vector, ϕ is the electric potential that satisfies $\nabla\phi = E$, $S = [\frac{\partial u}{\partial x}, \frac{\partial v}{\partial y}, \frac{\partial w}{\partial z}, \frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}, \frac{\partial w}{\partial x} + \frac{\partial u}{\partial z}, \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}]^\top$, and δq, $\delta\phi$ and δS are the corresponding virtual displacement, potential and strain vectors, respectively. The equation can then be discretized by finite element method [1, 10]. Following the usual free mode analysis, we consider $f = 0$, $g = 0$ and a time harmonic quasi periodic solution vector $\psi_\omega(x, t) = \psi(x)e^{i\omega t}$. The spatial function $\psi(x) = [q(x), \phi(x)]$ satisfies the boundary conditions shown in Figure 2 in which the periodic boundary

conditions, proposed by Buchner [6],

$$\begin{aligned} \psi_r &= \psi(x_1+p,x_2) = \gamma\psi(x_1,x_2) = \gamma\psi_l, && (5)\\ T_r\cdot n_r &= T(\psi_r)\cdot n_r = -\gamma T(\psi_l)\cdot n_l = -\gamma T_l\cdot n_l, && (6)\\ D_r\cdot n_r &= D(\psi_r)\cdot n_r = -\gamma D(\psi_l)\cdot n_l = -\gamma D_l\cdot n_l, && (7) \end{aligned}$$

are enforced on the left and right boundaries, Γ_l and Γ_r, here n_l and n_r are the normal vector of Γ_l and Γ_r respectively and $\gamma = e^{-(\alpha+i\beta)}$. By plugging ψ_w into (4), the equation can be rewritten in the following matrix form:

$$\begin{bmatrix} K^{qq}-\omega^2 M^{qq} & K^{q\phi} \\ K^{\phi q} & -K^{\phi\phi} \end{bmatrix} \begin{bmatrix} q \\ \phi \end{bmatrix} = \begin{bmatrix} F_l+F_r \\ Q_l+Q_r \end{bmatrix}, \tag{8}$$

where

$$K^{qq} = \int_\Omega \delta q^\top B_S^\top C^E B_S q\, dV, \quad M^{qq} = \int_\Omega \delta q^\top \rho q\, dV, \quad \text{and} \quad K^{\phi\phi} = \int_\Omega \delta\phi^\top B_\phi^\top e B_S q\, dV,$$

$$K^{q\phi} = \int_\Omega \delta q^\top B_S^\top e^\top B_\phi \phi\, dV \quad \text{and} \quad K^{\phi q} = \int_\Omega \delta\phi^\top B_\phi^\top e B_S q\, dV,$$

$$F_l = \int_{\Gamma_l} \delta q^\top T_l \vec{n_l}\, dA \quad \text{and} \quad F_r = \int_{\Gamma_r} \delta q^\top T_r \vec{n_r}\, dA,$$

$$Q_l = \int_{\Gamma_l} \delta\phi^\top D_l \vec{n_l}\, dA \quad \text{and} \quad Q_r = \int_{\Gamma_r} \delta\phi^\top D_r \vec{n_r}\, dA,$$

and the matrices $B_S = \begin{bmatrix} \frac{\partial}{\partial x} & 0 & 0 & 0 & \frac{\partial}{\partial z} & \frac{\partial}{\partial y} \\ 0 & \frac{\partial}{\partial y} & 0 & \frac{\partial}{\partial z} & 0 & \frac{\partial}{\partial x} \\ 0 & 0 & \frac{\partial}{\partial z} & \frac{\partial}{\partial y} & \frac{\partial}{\partial x} & 0 \end{bmatrix}^\top$ and $B_\phi = \begin{bmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \end{bmatrix}^\top$.
Mechanical damping effects can also be considered by using the Rayleigh damping assumption in which the matrix $K^{qq}-\omega^2M^{qq}$ in (8) are modified into $K^{qq}+i\omega(\kappa_1 K^{qq}+\kappa_2 M^{qq})-\omega^2 M^{qq}$. Here κ_1 and κ_2 are coefficients associated with the viscous damping and mass damping, respectively.

To obtain the palindromic quadratic eigenvalue problem associated with the propagation parameter γ, following Hofer's approach [14], the nodal unknowns are splitted into the inner nodes $\psi_i = [q_i, \phi_i]$, the left boundary nodes $\psi_l = [q_l, \phi_l]$ and the right boundary nodes $\psi_r = [q_r, \phi_r]$. The matrix equation (8) can be recasted into the following form:

$$\begin{bmatrix} K_{ii} & K_{il} & K_{ir} \\ K_{li} & K_{ll} & K_{lr} \\ K_{ri} & K_{rl} & K_{rr} \end{bmatrix} \begin{bmatrix} \psi_i \\ \psi_l \\ \psi_r \end{bmatrix} = \begin{bmatrix} 0 \\ R_l \\ R_r \end{bmatrix}, \tag{9}$$

here R_l and R_r are vectors obtained from the discretization of the terms F_l+Q_l and F_r+Q_r, respectively. From the periodic boundary conditions (5), (6) and (7), (9) becomes

$$\begin{bmatrix} K_{ii} & K_{il} & K_{ir} \\ K_{li} & K_{ll} & K_{lr} \\ K_{ri} & K_{rl} & K_{rr} \end{bmatrix} \begin{bmatrix} I_i & 0 \\ 0 & I_l \\ 0 & \gamma I_l \end{bmatrix} \begin{bmatrix} \psi_i \\ \psi_l \end{bmatrix} = \begin{bmatrix} 0 \\ I_l \\ -\gamma I_l \end{bmatrix} R_l. \tag{10}$$

Furthermore, by multiplying the matrix

$$\begin{bmatrix} I_i & 0 & 0 \\ 0 & \gamma I_l & I_l \end{bmatrix}$$

to (10), the GEP associated with the propagation parameter γ is obtained:

$$\left(\begin{bmatrix} K_{ii} & K_{il} \\ K_{ir}^\top & 0 \end{bmatrix} + \gamma \begin{bmatrix} 0 & K_{ir} \\ K_{il}^\top & K_{ll} + K_{rr} \end{bmatrix} \right) \begin{bmatrix} \psi_i \\ \psi_l \end{bmatrix} = 0. \tag{11}$$

Since the viscosity is small for most crystalline solids, the attenuation factor α is close to zero. As a result, the propagation factors λ near the unit circle, denoted by $\mathbb{U}$, is desired. Moreover, for the frequency ω in the stopping band, the frequency shift parameter β shall be close to π when the periodic interval p (i.e. the domain width here) equals to half of the incident wave length λ_0. Therefore, for the eigenvalue problem (11), we are interesting in finding the eigenvalues λ close to $\mathbb{U}$, especially for those are near -1 on the complex plane.

Notice that the nonzero eigenvalues of (11) appear in the reciprocal pairs $(\lambda, 1/\lambda)$. The reciprocal relation is very sensitive to numerical errors when they are close to $\mathbb{U}$. On the other hand, it is well known that the solution of a general elliptic problem have singularities around corners [12] and, in addition, the solution may become less regular near the interface between the electrode and PZT substrate. It is inevitable that the error from discretization may be amplified in computing the reciprocal pairs. Therefore, it is important to minimize the accuracy deterioration due to singularities and lower regularity in finite element solutions. One can resolve the singularity by constructing the singular elements in which the mesh points are clustered to the singular source according to the order of the singularity [3]. In our calculation, we simply employee the locally refined meshes. An additional benefit from using the locally refined meshes is that we can discretize equation (8) using linear elements instead of using high order finite element discretization [6]. However, drawbacks include (i) the matrices from the discretization of (8) is large and sparse and (ii) the sparse pattern of the matrices is unstructured. These make the efficient computation of eigenvalues for large GEP in (11) a challenge. Moreover, for pizoelectric crystals, the elastic constant matrix c^E is 10^{20} greater than the electric constant matrix ε^S. To compute the eigenvalues and eigenvectors, proper scaling between the mechanic field and the electrical field is required. Hence, the eigen solutions obtained from the scaled problem must be accurate enough in order to disregard the round off error in the re-scaling process. Therefore, for solving the large sparse eigenvalue problem (11), an efficient algorithm, not only preserves the reciprocal eigen-structure but is also accurate enough to prevent error amplification from rounding and discretization, is desired. In the next section, we shall introduce an efficient structure-preserving algorithm that ensure the accuracy of the eigen-curves $\lambda(\omega)$ and the associated eigenvectors of (11).

3. Structure-preserving Arnoldi-type Algorithm

To compute the eigenvalues of the GEP in (11) derived from SAW filter, one can apply traditional eigensolvers such as LAPACK [14]. However, as the GEP is large and its sparse pattern is unstructured, the traditional eigensolvers are not efficient and the eigenvalues obtained are not guaranteed to satisfy the reciprocal relationship. In this paper, since only the eigenvalues closed to $\mathbb{U}$ are of interest, we choose iterative methods to find the eigenparis. Our numerical results in Section 4 show that the proposed algorithm converges efficiently in only a few steps and is very accurate in computing the reciprocal eigenpairs.

Observing that the imaginary part of $K_{ll} + K_{rr}$ in (11) is symmetric and positive definite, by Bendixson Theorem, $K_{ll} + K_{rr}$ is invertible. The second equation of (11) gives

$$\psi_l = -\frac{1}{\gamma}(K_{ll} + K_{rr})^{-1} K_{ir}^{\top} \psi_i - (K_{ll} + K_{rr})^{-1} K_{il}^{\top} \psi_i. \tag{12}$$

Letting

$$M_1 = K_{ii},\ M_2 = K_{ll} + K_{rr},\ F = K_{ir},\ G = K_{il},\ x = \psi_i,\ \lambda = \gamma, \tag{13}$$

and substituting (12) into the first equation of (11), we obtain the TPQEP in (2).

To solve (2) in a structure-preserving way, we first transform the TPQEP in (2a) into a $\top$-skew-Hamiltonian pencil $(\widehat{\mathcal{K}}, \widehat{\mathcal{N}})$ through the following procedure:

(i) The TPQEP is linearized into a special GEP [15],

$$(\mathcal{M} - \lambda \mathcal{L}) \begin{bmatrix} x \\ y \end{bmatrix} = 0, \tag{14}$$

where $\lambda y = A_1 x$, and

$$\mathcal{M} = \begin{bmatrix} A_1 & 0 \\ -A_0 & -I \end{bmatrix}, \quad \mathcal{L} = \begin{bmatrix} 0 & I \\ A_1^{\top} & 0 \end{bmatrix}. \tag{15}$$

The reciprocal eigenpairs $(\lambda, 1/\lambda)$ are kept in the matrix pencil $(\mathcal{M}, \mathcal{L})$ because the matrix pencil $\mathcal{M} - \lambda\mathcal{L}$ is $\top$-symplectic, i.e., it satisfies $\mathcal{M}\mathcal{J}\mathcal{M}^{\top} = \mathcal{L}\mathcal{J}\mathcal{L}^{\top}$ where $\mathcal{J} = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}$.

(ii) Using the $(\mathcal{S} + \mathcal{S}^{-1})$-transform, the matrix pencil $\mathcal{M} - \lambda\mathcal{L}$ is further transformed to a $\top$-skew-Hamiltonian pencil $\mathcal{K} - \mu\mathcal{N}$, i.e., $(\mathcal{K}\mathcal{J})^{\top} = -\mathcal{K}\mathcal{J}$, $(\mathcal{N}\mathcal{J})^{\top} = -\mathcal{N}\mathcal{J}$ and μ is the eigenvalue of the pencil:

$$\mathcal{K} - \mu\mathcal{N} \equiv \left[(\mathcal{L}\mathcal{J}\mathcal{M}^{\top} + \mathcal{M}\mathcal{J}\mathcal{L}^{\top}) - \mu\mathcal{L}\mathcal{J}\mathcal{L}^{\top} \right] \mathcal{J}^{\top}. \tag{16}$$

From the relationship $\mu = \lambda + 1/\lambda$, one can relate the two eigenvalues λ and μ and further implies that the multiplicity of the eigenvalue μ is even.

(iii) Let τ be a shift value and $\tau \notin \sigma(\mathcal{M}, \mathcal{L})$ where $\sigma(\mathcal{A}, \mathcal{B})$ denotes the set of all eigenvalues of any matrix pair $(\mathcal{A}, \mathcal{B})$. Since $\mu_0 \equiv \tau + 1/\tau \notin \sigma(\mathcal{K}, \mathcal{N})$, one can define the shift-invert transformation $\widehat{\mathcal{K}} - \widehat{\mu}\widehat{\mathcal{N}}$ for $\mathcal{K} - \mu\mathcal{N}$ with $\widehat{\mu} = (\mu - \mu_0)^{-1}$ where

$$\widehat{\mathcal{K}} \equiv -\tau\mathcal{N} = \tau \begin{bmatrix} A_1^\top & 0 \\ 0 & A_1 \end{bmatrix}, \tag{17a}$$

$$\widehat{\mathcal{N}} \equiv -\tau(\mathcal{K} - \mu_0\mathcal{N}) = (\mathcal{M} - \tau\mathcal{L})\,\mathcal{J}\left(\mathcal{M}^\top - \tau\mathcal{L}^\top\right)\mathcal{J}^\top, \tag{17b}$$

and $\widehat{\mathcal{K}}$ and $\widehat{\mathcal{N}}$ are $\top$-skew-Hamiltonian.

The relationship between eigenpairs of the TPQEP in (2) and the $\top$-skew-Hamiltonian pencil $(\widehat{\mathcal{K}}, \widehat{\mathcal{N}})$ in (17) is stated in the following theorem.

Theorem 3.1. [15] *Let $(\widehat{\mathcal{K}}, \widehat{\mathcal{N}})$ be defined in* (17) *and τ be a shift value with $\tau \notin \sigma(\mathcal{M}, \mathcal{L})$. If $z_s = [z_1^\top, z_2^\top]^\top$ with $z_1, z_2 \in \mathbb{C}^n$ is an eigenvector of $(\widehat{\mathcal{K}}, \widehat{\mathcal{N}})$ corresponding to eigenvalue $\widehat{\mu}$ and ν satisfies $\tau + \frac{1}{\tau} + \frac{1}{\widehat{\mu}} = \nu + \frac{1}{\nu}$, then $z_1 + \frac{1}{\nu}z_2$ and $z_1 + \nu z_2$ are eigenvectors of the TPQEP in* (2) *corresponding to eigenvalues ν and $\frac{1}{\nu}$, respectively.*

Next, from the definition of $\widehat{\mathcal{N}}$ in (17b), $\widehat{\mathcal{N}}$ can be factorized as $\widehat{\mathcal{N}} = \mathcal{N}_1\mathcal{N}_2$, where

$$\mathcal{N}_1 = \mathcal{M} - \tau\mathcal{L}, \quad \mathcal{N}_2 = \mathcal{J}(\mathcal{M}^\top - \tau\mathcal{L}^\top)\mathcal{J}^\top \tag{18}$$

are nonsingular and satisfy $\mathcal{N}_2^\top\mathcal{J} = \mathcal{J}\mathcal{N}_1$. Let $\mathcal{B} \equiv \mathcal{N}_1^{-1}\widehat{\mathcal{K}}\mathcal{N}_2^{-1}$ and u_1 be an initial vector. Define the Krylov matrix with respect to u_1 by

$$K_n \equiv K_n[\mathcal{B}, u_1] = [u_1,\, \mathcal{B}u_1,\, \cdots,\, \mathcal{B}^{n-1}u_1].$$

The following Theorem 3.2 guarantees that the Arnoldi process can be executed in a way that the $\top$-skew-Hamiltonian structure of the matrix pencil is preserved. As a result, a generalized $\top$-skew-Hamiltonian implicity-restarted Arnoldi (GTSHIRA) algorithm proposed in [15] can be employed to solve the eigenvalue problem $\widehat{\mathcal{K}}z = \widehat{\mu}\widehat{\mathcal{N}}z$.

Theorem 3.2. [15] *Let $\mathcal{B} = \mathcal{N}_1^{-1}\widehat{\mathcal{K}}\mathcal{N}_2^{-1}$ with $\widehat{\mathcal{N}} = \mathcal{N}_1\mathcal{N}_2$ be $\top$-skew-Hamiltonian and $K_n \equiv K_n[\mathcal{B}, u_1]$ be the Krylov matrix with rank$(K_n) = n$. Then there are unitary matrices $\mathcal{U}$ and $\mathcal{V}$ satisfying $\mathcal{V} = \mathcal{J}^\top\mathcal{U}\mathcal{J}$, $\mathcal{U}e_1 = u_1$ and $\mathcal{V}e_1 = \mathcal{N}_1u_1/\|\mathcal{N}_1u_1\|_2$ such that*

$$\mathcal{V}^\top\widehat{\mathcal{K}}\mathcal{U} = \begin{bmatrix} \widehat{H}_n & \widehat{S}_n \\ 0 & \widehat{H}_n^\top \end{bmatrix}, \quad \mathcal{V}^\top\widehat{\mathcal{N}}\mathcal{U} = \begin{bmatrix} \widehat{R}_n & \widehat{T}_n \\ 0 & \widehat{R}_n^\top \end{bmatrix}, \tag{19}$$

where $\widehat{H}_n$ is unreduced upper Hessenberg, $\widehat{R}_n$ is nonsingular upper triangular and $\widehat{S}_n$, $\widehat{T}_n$ are $\top$-skew-symmetric.

Finally, the unitary matrices $\mathcal{U}$ and $\mathcal{V}$ in Theorem 3.2 can also be generated in a structure-preserving way (GTSHIRA) as follows. Recall that the $\top$-bi-isotropic orthonormal matrices $\widehat{Z}_j, \widehat{Y}_j \in \mathbb{C}^{2n\times j}$ are computed iteratively according to the following structure-preserving Arnoldi process:

$$\widehat{\mathcal{K}}\widehat{Z}_j = \widehat{Y}_j\widehat{H}_j + \widehat{h}_{j+1,j}\widehat{y}_{j+1}e_j^\top \tag{20}$$

and

$$\widehat{\mathcal{N}}\widehat{Z}_j = \widehat{Y}_j\widehat{R}_j \tag{21}$$

with

$$\widehat{Y}_j^H\widehat{y}_{j+1} = 0 \quad \text{and} \quad \widehat{Z}_j^\top\mathcal{J}\widehat{y}_{j+1} = 0, \tag{22}$$

where $\widehat{H}_j, \widehat{R}_j \in \mathbb{C}^{j\times j}$ are unreduced upper Hessenberg and nonsingular upper triangular, respectively. By defining $\mathcal{U}_j \equiv \left[\widehat{Z}_j,\ -\mathcal{J}\bar{\widehat{Y}}_j\right]$ and $\mathcal{V}_j \equiv \left[\widehat{Y}_j,\ -\mathcal{J}\bar{\widehat{Z}}_j\right]$ where $\bar{\widehat{Y}}_j$ and $\bar{\widehat{Z}}_j$ denote the conjugate matrices of $\widehat{Y}_j$ and $\widehat{Z}_j$, respectively, it is easily seen that

$$\mathcal{V}_j^H\widehat{\mathcal{K}}\mathcal{U}_j = \begin{bmatrix} \widehat{H}_j & -\widehat{Y}_j^H\widehat{\mathcal{K}}\mathcal{J}\bar{\widehat{Y}}_j \\ 0 & \widehat{H}_j^\top \end{bmatrix}, \quad \mathcal{V}_j^H\widehat{\mathcal{N}}\mathcal{U}_j = \begin{bmatrix} \widehat{R}_j & -\widehat{Y}_j^H\widehat{\mathcal{N}}\mathcal{J}\bar{\widehat{Y}}_j \\ 0 & \widehat{R}_j^\top \end{bmatrix}$$

which implies that the $\top$-skew-Hamiltonian property is preserved in each iteration step.

Notice that Theorem 3.1 indicates that although the number of the eigenvectors associated with the eigenvalue μ is even, only half of the eigenvectors are needed to extract all the eigenvectors corresponding to the eigenvalues ν and $\frac{1}{\nu}$. Furthermore, through the above mentioned structure-preserving Arnoldi process, the $\top$-skew-Hamiltonian structure of the matrix pencil in Theorem 3.2 is preserved and the even multiplicity of the eigenvalue μ is automatically obtained. Therefore, the required halves of the eigenvectors associated with the eigenvalue μ can be easily computed when the desired eigenpairs are convergent. In fact, the desired eigenpairs $(\widehat{\mu}_i, z_i)$ of $(\widehat{\mathcal{K}}, \widehat{\mathcal{N}})$ can be computed from the matrix pair $(\widehat{H}_j, \widehat{R}_j)$ with $\widehat{H}_j\hat{s}_i = \widehat{\mu}_i\widehat{R}_j\hat{s}_i$ and $z_i = \widehat{Z}_j\hat{s}_i$. From Theorem 3.1, one can compute the desired eigenpairs of $(\mathcal{M}, \mathcal{L})$ from $(\widehat{\mu}_i, z_i)$ and preserve the reciprocal relationship of the eigenvalues of the GEP algebraically.

We summarize the above procedures for computing the reciprocal eigenpairs of the GEP (11) in Algorithm 1. Notice that, in step 1 of the Algorithm 1, the linear systems

$$\mathcal{N}_1 v_1 = b_1, \quad \mathcal{N}_2 v_2 = b_2, \tag{23}$$

have to be solved in order to obtain $\widehat{Z}_j$ from (21). This is indeed the most time-consuming step in the proposed structure-preserving algorithm. In the following, we discuss how to solve (23) efficiently.

Algorithm 1 Structure-preserving algorithm for solving GEP (11)

Input: matrices F, G, M_2 and M_1, shift value τ and the number $\mathfrak{m}$ of desired eigenvalues.

Output: eigenpairs $\{(\gamma_j, [(\psi_{i,j}^{(1)})^\top, (\psi_{l,j}^{(1)})^\top]^\top), (\gamma_j^{-1}, [(\psi_{i,j}^{(2)})^\top, (\psi_{l,j}^{(2)})^\top]^\top)\}_{j=1}^{\mathfrak{m}}$ of the GEP in (11) where $\gamma_j + \gamma_j^{-1}$ for $j = 1, \ldots, \mathfrak{m}$ are the closest to shift value $\tau + \tau^{-1}$.

1: Compute eigenpairs $\{(\widehat{\mu}_j, z_j \equiv [z_{j1}^\top, z_{j2}^\top]^\top)\}_{j=1}^{\mathfrak{m}}$ of $(\widehat{\mathcal{K}}, \widehat{\mathcal{N}})$ by using GT-SHIRA.

2: Compute eigenvalues γ_j and $\frac{1}{\gamma_j}$ of TPQEP in (2) by solving

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

Compute eigenvectors

$$\psi_{i,j}^{(1)} \equiv \frac{1}{\gamma_j} z_{j1} - z_{j2}, \quad \psi_{i,j}^{(2)} \equiv \gamma_j z_{j1} - z_{j2}$$

corresponding to $\gamma_j, \frac{1}{\gamma_j}$, respectively, for $j = 1, 2, \ldots, \mathfrak{m}$.

3: Compute

$$\psi_{l,j}^{(1)} = -M_2^{-1}\left(\gamma_j^{-1} F^\top \psi_{i,j}^{(1)} + G^\top \psi_{i,j}^{(1)}\right), \; \psi_{l,j}^{(2)} = -M_2^{-1}\left(\gamma_j F^\top \psi_{i,j}^{(2)} + G^\top \psi_{i,j}^{(2)}\right)$$

for $j = 1, \ldots, \mathfrak{m}$.

By the definitions of $\mathcal{M}$ and $\mathcal{L}$ in (15), we have

$$\begin{bmatrix} I & -\tau I \\ 0 & I \end{bmatrix} (\mathcal{M} - \tau\mathcal{L}) = \begin{bmatrix} \tau^2 A_1^\top + \tau A_0 + A_1 & 0 \\ -A_0 - \tau A_1^\top & -I \end{bmatrix} \tag{24a}$$

and

$$\begin{bmatrix} I & -A_0 - \tau A_1 \\ 0 & I \end{bmatrix} (\mathcal{M}^\top - \tau\mathcal{L}^\top) = \begin{bmatrix} \tau^2 A_1 + \tau A_0 + A_1^\top & 0 \\ -\tau I & -I \end{bmatrix}. \tag{24b}$$

From (18) and (24), we see that solving (23) is equivalent to solve

$$\begin{aligned} (\tau^2 A_1^\top + \tau A_0 + A_1) v_{11} &= b_{11} - \tau b_{12}, \\ v_{12} &= -b_{12} - (A_0 + \tau A_1^\top) v_{11}, \end{aligned} \tag{25}$$

and

$$\begin{aligned} (\tau^2 A_1 + \tau A_0 + A_1^\top) v_{22} &= b_{22} + (A_0 + \tau A_1) b_{21}, \\ v_{21} &= \tau v_{22} - b_{21}, \end{aligned} \tag{26}$$

where $v_1 = [v_{11}^\top, v_{12}^\top]^\top$, $v_2 = [v_{21}^\top, v_{22}^\top]^\top$, $b_1 = [b_{11}^\top, b_{12}^\top]^\top$ and $b_2 = [b_{21}^\top, b_{22}^\top]^\top$. By the definitions of A_0 and A_1, it holds that

$$\tau^2 A_1^\top + \tau A_0 + A_1 = (G + \tau F) M_2^{-1} (F^\top + \tau G^\top) - \tau M_1 \tag{27}$$

and

$$\tau^2 A_1 + \tau A_0 + A_1^\top = (F + \tau G) M_2^{-1} (G^\top + \tau F^\top) - \tau M_1. \tag{28}$$

Since the bandwidth of the matrix M_1 can be greatly reduced by reordering the unknowns as shown in Figure 3, the LU factorization of $M_1 = LU$ can be computed efficiently. Set

$$E_1 = L^{-1} \left(\frac{1}{\tau} G + F \right), \quad E_2 = U^{-\top} (F + \tau G).$$

By the Sherman-Morrison-Woodbury formula, (27) and (28) can be further factorized as following,

$$\begin{aligned} \left(\tau^2 A_1^\top + \tau A_0 + A_1\right)^{-1} &= -\frac{1}{\tau} U^{-1} \left[I - E_1 M_2^{-1} E_2^\top \right]^{-1} L^{-1} \\ &= -\frac{1}{\tau} U^{-1} \left[I + E_1 \left(M_2 - E_2^\top E_1 \right)^{-1} E_2^\top \right] L^{-1}, \end{aligned} \tag{29}$$

and

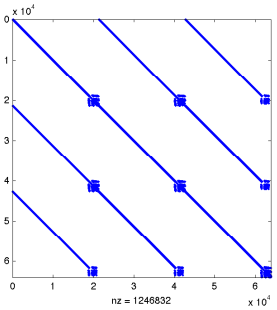
$$\left(\tau^2 A_1 + \tau A_0 + A_1^\top\right)^{-1} = -\frac{1}{\tau} L^{-\top} \left[I + E_2 \left(M_2 - E_1^\top E_2 \right)^{-1} E_1^\top \right] U^{-\top}. \tag{30}$$

Now, obviously, the solutions of (25) and (26) can be obtained by two forward substitutions (L^{-1}), two backward substitutions (U^{-1}) and solving small linear systems $\left(M_2 - E_2^\top E_1\right)^{-1}$ and $\left(M_2 - E_1^\top E_2\right)^{-1}$. As a result, Algorithm 1 is very efficient.

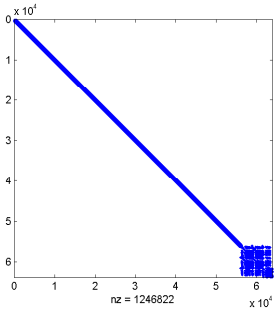
4. Numerical results

In this section, we firstly conduct numerical results to validate the convergence for our finite element model. Secondly, we report the numerical comparisons with our structure-preserving method and the traditional Arnoldi method for solving the GEP (11) to demonstrate the accuracy and efficiency of the proposed eigenvalue solver. All computations are carried out in MATLAB 2010b on a HP workstation with an Intel Quad-Core Xeon X5570 2.93GHz and 60 GB main memory, using IEEE double-precision floating-point arithmetic.

To make our numerical computation reliable, first, we scale the displacement field q and potential field ϕ by 10^{-5} and 10^5, respectively, and scale the mass density ρ accordingly. The entries of the stiffness matrices K^{qq}, $K^{q\phi}$, $K^{\phi q}$ and $K^{\phi\phi}$, and the mass matrix M^{qq} are about the same order (close to $O(1)$) after the scaling. Secondly, we reorder the interior nodes so that the matrix M_1 in



(a) Sparse pattern of matrix M_1 without permutation from locally refined meshes with mesh length $p/80$.



(b) Sparse pattern of the permuted matrix M_1 from locally refined meshes with mesh length $p/80$.

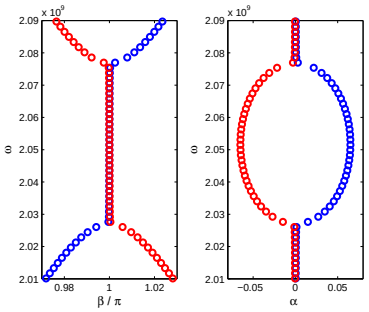
Figure 3: Sparsity of matrix M_1.

(13) has narrower band structure. As a result, the LU factorization of M_1 can be computed easily and solutions of the linear systems in (25) and (26) can be obtained efficiently from (29) and (30). The sparsity patterns of the matrices M_1 obtained from FEM descritization on a locally refined mesh are shown in Figure 3.(a) without nodal reordering, and shown in Figure 3.(b) with nodal reordering. A mesh that are locally refined twice near the interface over an uniform mesh is shown in Figure 5.

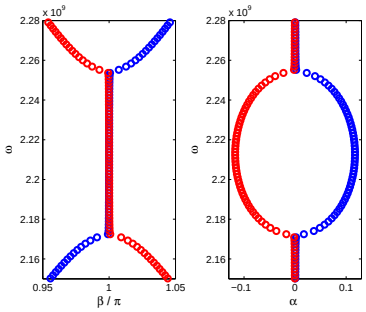
The configuration of our computational domain shown in Figure 2 is as follows. The domain width $\overline{AB}$ and height $\overline{CD}$ are set to be 10^{-6} and 3×10^{-6}, respectively. The ratio of the electrode width $\overline{EF}$ versus the domain width is set to be $\frac{1}{2}$ and the ratio of the electrode thickness $\overline{DE}$ versus the domain height is $\frac{1}{15}$. The material constants of $LiTaO_3$ and $LiNbO_3$ are taken from measurements obtained by Kushibiki, Takanaga and Sannomiya [18]. Also, it has been shown that the viscous damping coefficient $\kappa_1 \approx O(10^{-8})$ in the 10 KHZ operation range and $\kappa_1 \approx O(10^{-10})$ in the MHZ operation range for a family of PZT materials [23, 26]. In general, κ_1 depends on the operation frequency ω. The viscous damping coefficient is extrapolated to GHZ operation range according to the reciprocal rule $\kappa_1 \propto \frac{1}{\omega}$ [7]. In our numerical studies, the viscous damping coefficient κ_1 is set to be 10^{-14} and the mass damping coefficient κ_2 is taken as $1 - \kappa_1$ to account for the effect from the electrode weight.

4.1. Accuracy and convergence of finite element approximation

Firstly, we show that our finite element model gives accurate results in predicting the center of stopping band of LSAW on the filters with aluminum electrodes on top of piezoelectric substrates 36^o YX-$LiTaO_3$ and 64^o YX-$LiNbO_3$. The dispersion diagrams of the attenuation constant α and the propagation constant β associated with the eigenvalue $\lambda(\omega)$, that is most close to -1 on



(a) Dispersion diagram of $LiTaO_3$ near the stopping band



(b) Dispersion diagram of $LiNbO_3$ near the stopping band

Figure 4: Dispersion diagrams

the complex plane for frequency ω around the stopping bands, are shown in Figure 4(a) and Figure 4(b) for crystals 36^o YX-$LiTaO_3$ and 64^o YX-$LiNbO_3$, respectively. A typical shear wave displacement associated with the eigenvector of the computed eigenvalue is shown in Figure 5.

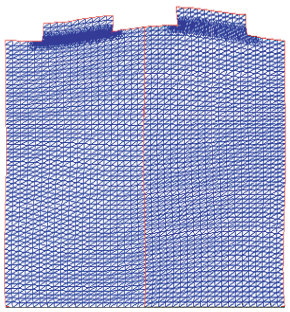


Figure 5: A shear wave displacement.

In order to measure the convergence of the eigenvalues, tests over three successively refined meshes with initial mesh size $h = \frac{p}{20}$ are performed. The dimensions of matrices M_1 and M_2 associated with uniform meshes and locally refined meshes are list in second and third columns of Table 1, respectively. We set $\lambda_{[i,\xi]}$ to be the eigenvalue obtained from meshes with mesh length $p/(10 \times i)$. Here the index $\xi =$ "u" and $\xi =$ "ℓ" denote the mesh is uniform without and with local refinement, respectively. Using $\lambda_{[16,u]}$ as exact value, the convergence of eigenvalues can be verified from $|\lambda_{[16,u]} - \lambda_{[i,u]}|$ and $|\lambda_{[16,u]} - \lambda_{[i,\ell]}|$ for $i = 2, 4, 8$.

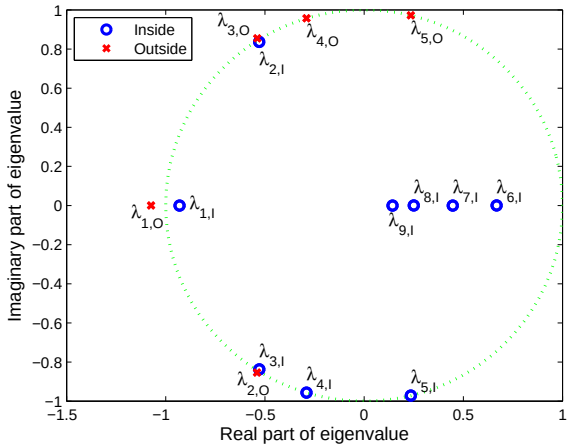


Figure 6: Distribution of eigenvalues for 64^o YX-$LiNbO_3$ at frequency 2.180 GHZ.

The values of $|\lambda_{[16,u]} - \lambda_{[i,u]}|$ and $|\lambda_{[16,u]} - \lambda_{[i,\ell]}|$ $(i = 2, 4, 8)$ for 36^o YX-$LiTaO_3$ and 64^o YX-$LiNbO_3$ at ω_s and ω_e are shown in Table 2 where ω_s and ω_e are the frequencies for which the stopping band starts and ends, respectively. From Table 2, it can be seen that the accuracy of eigenvalue is increased as the mesh length being reduced and is improved by using locally refined meshes. Moreover, it is known that the wave propagation velocity is about $4112m/s$ for 36^o YX-$LiTaO_3$ and about $4478m/s$ for 64^o YX-$LiNbO_3$. Since the domain width $p = 1 \times 10^{-6}$, clearly, the center of the stopping band is about 2.056 GHZ and 2.239 GHZ, respectively. We compute the center of the stoping band by averaging ω_s and ω_e on different mesh lengths and show it in Table 3. Obviously, one can see that the central frequency is monotonically converged to a constant when the mesh length is reduced. The numerical error from our finite element simulations is less than 0.2% and 1.2% for 36^o YX-$LiTaO_3$ and 64^o YX-$LiNbO_3$, respectively.

4.2. Comparison of Algorithm 1 and traditional Arnoldi method

From Tables 2 and 3, we already show that the accuracy of the computed eigenvalues and central frequency of stopping band obtained from the locally refined mesh with mesh length $p/80$ is almost the same as those obtained from uniform mesh with mesh length $p/160$. Therefore, in the following numerical computations, we only consider the coefficient matrices in the GEP (11) that are generated by the finite element discretization on the mesh that is locally refined twice over the uniform mesh with mesh length $p/80$.

Let the pair $(\lambda_{k,I}, \lambda_{k,O})$, $k = 1, \ldots, N$, denote the reciprocal pairs of eigenvalues of (11) where $\lambda_{k,I}$ and $\lambda_{k,O}$ lie inside and outside $\mathbb{U}$, respectively. Fig-

	Uniform		Local refine	
mesh length	M_1	M_2	M_1	M_2
$p/20$	3554	183	4968	183
$p/40$	14548	363	17192	363
$p/80$	58856	723	63960	723
$p/160$	236752	1443		

Table 1: Dimension information of matrices M_1 and M_2 obtained from FEM discretization

	36^o YX-$LiTaO_3$		64^o YX-$LiNbO_3$	
ω (GHZ)	$\omega_s = 2.028$	$\omega_e = 2.075$	$\omega_s = 2.177$	$\omega_e = 2.257$
$\lvert\lambda_{[16,u]} - \lambda_{[2,u]}\rvert$	0.0222	0.0161	0.0955	0.0797
$\lvert\lambda_{[16,u]} - \lambda_{[2,\ell]}\rvert$	0.0178	0.0141	0.0857	0.0751
$\lvert\lambda_{[16,u]} - \lambda_{[4,u]}\rvert$	0.0076	0.0056	0.0299	0.0458
$\lvert\lambda_{[16,u]} - \lambda_{[4,\ell]}\rvert$	0.0036	0.0042	0.0166	0.0329
$\lvert\lambda_{[16,u]} - \lambda_{[8,u]}\rvert$	0.0016	0.0015	0.0060	0.0087
$\lvert\lambda_{[16,u]} - \lambda_{[8,\ell]}\rvert$	0.0001	0.0006	0.0007	0.0016

Table 2: The values of $|\lambda_{[16,u]} - \lambda_{[i,\xi]}|$ for different mesh lengths at frequencies ω_s and ω_e.

mesh length		$h_u = \frac{p}{40}$	$h_\ell = \frac{p}{40}$	$h_u = \frac{p}{80}$	$h_\ell = \frac{p}{80}$	$h_u = \frac{p}{160}$
f_c	$LiTaO_3$	2.05246	2.05222	2.05206	2.05191	2.05183
	$LiNbO_3$	2.21623	2.21464	2.21385	2.21305	2.21305

Table 3: Computed center frequency f_c (GHZ) of stopping bands of LSAW on various meshes. Here, h_u and h_ℓ denote the mesh length of meshes without and with local refinement, respectively.

ure 6 displays the eigenvalues $\{\lambda_{1,I}, \ldots, \lambda_{9,I}, \lambda_{1,O}, \ldots, \lambda_{5,O}\}$ of the $LiNbO_3$ at frequency $\omega = 2.180$ GHZ in which reciprocal pairs $(\lambda_{k,I}, \lambda_{k,O})$ for $k = 1, \ldots, 5$ close to $\mathbb{U}$ may be of interests. The notation $\mathbb{O}(\lambda)$ represents the set of all the eigenvalues that cluster at the origin of the complex plane. Suppose $2N$ eigenvalues near $\mathbb{U}$ are desired. The Arnoldi process in (20)–(21) for GTSHIRA is set to restart if the desired eigenpairs are not convergent when the dimension j of the subspace span$\{\widehat{Y}_j\}$ grows more than $5N$. The number of restarting Arnoldi process is denoted by "#Iter" in the following.

A standard iterative approach for solving the GEP (11) is to apply Arnoldi method on the equation directly. However, the reciprocal property of the eigenvalues is not guaranteed to be preserved in the computation. For Algorithm 1, based on the $(\mathcal{S} + \mathcal{S}^{-1})$-transform, if λ and μ are the eigenvalues of (11) and (16), respectively, then λ and μ satisfy the relation $\mu = \lambda + \lambda^{-1}$. As a result, we can obtain the k-th reciprocal pair $(\lambda_{k,I}, \lambda_{k,O} \equiv 1/\lambda_{k,I})$ by solving the algebraic equation $\mu_k = \lambda_{k,I} + \lambda_{k,I}^{-1}$ after the k-th eigenvalue μ_k of $\mathcal{K}z = \mu\mathcal{N}z$ is computed. Hence, the reciprocity is automatically preserved. Two numerical comparisons on preserving reciprocal property, between Algorithm 1 and traditional Arnoldi method, are listed in the following where eigenvalues of 64^o YX-$LiNbO_3$ at frequency $\omega = 2.180$ GHZ are computed.

- Traditional Arnoldi method does not guarantee that half of the computed eigenvalues lie inside of the unit circle and the others are outside. For example, when we use Arnoldi method to compute four eigenvalues (i.e., $2N = 4$) of (11) which are near -1, the four convergent eigenvalues are $\lambda_{1,I}, \lambda_{1,O}, \lambda_{2,I}$ and $\lambda_{3,I}$. Clearly, the reciprocal property of eigenvalues is lost.

- Suppose one wants to compute the five reciprocal pairs $(\lambda_{k,I}, \lambda_{k,O})$ for $k = 1, \ldots, 5$. As shown in Figure 6, no matter what shift value τ is chosen, there exists some $k^\star \in \{1, \ldots, 5\}$ such that the eigenvalues in $\mathbb{O}(\lambda)$ are closer to τ than the eigenvalue $\lambda_{k^\star,I}$ or $\lambda_{k^\star,O}$. As a result, the eigenvalue $\lambda_{k^\star,I}$ or $\lambda_{k^\star,O}$ would not be discovered by the Arnoldi method. For example, if we take a shift value $\tau = -2.89$, then the desired reciprocal pair $(\lambda_{5,I}, \lambda_{5,O})$ near $\mathbb{U}$ can not be discovered by the Arnoldi method. On the contrary, in Algorithm 1, according to the relationship $\mu = \lambda + \lambda^{-1}$, the eigenvalue μ of $\mathcal{K}z = \mu\mathcal{N}z$ is far away from the shift value $\tau + 1/\tau$ when λ is closed to the origin. Naturally, Algorithm 1 will not converge to those unwanted eigenvalues in $\mathbb{O}(\lambda)$. As a result, all the desired eigenvalues can be discovered more easily by Algorithm 1 than the traditional Arnoldi method. Our numerical results in Table 4 show that not only all the desired eigenvalues are found by Algorithm 1, even when the number of desired eigenvalues is set to $2N = 18$, it also converges much faster than the traditional Arnoldi method. In fact, it only takes two restarting steps for Algorithm 1 to converge for all cases shown in Table 4. In addition,

from the rightmost column of Table 4, one can see that all the computed eigenvalues indeed preserve the reciprocal property. On the contrary, the reciprocity of the convergent eigenvalues obtained by Arnoldi method are diminished about 3 significant digits.

From the above comparison, Algorithm 1 preserves the reciprocal property of the eigenvalues of the GEP (11) effectively. For measuring the accuracy of Algorithm 1, let us define the relative residual of an eigenpair (λ, u) of (11), where $u = [\psi_i, \psi_l]^\top$, as following:

$$\frac{\left\| \begin{bmatrix} M_1 & G \\ F^\top & 0 \end{bmatrix} u - \lambda \begin{bmatrix} 0 & F \\ G^\top & M_2 \end{bmatrix} u \right\|_F}{\left\| \begin{bmatrix} M_1 & G \\ F^\top & 0 \end{bmatrix} \right\|_F \|u\|_F + |\lambda| \left\| \begin{bmatrix} 0 & F \\ G^\top & M_2 \end{bmatrix} \right\|_F \|u\|_F},$$

where $\| * \|_F$ is the Frobenius matrix norm. The maximal relative residuals of the ten desired eigepairs for 64^o YX-$LiNbO_3$ with various frequency are shown in Figure 7. From these numerical results, one can see that the eigenpairs produced by Algorithm 1 possess high accuracy in terms of relative residual error.

Next, let's compare the efficiency of Algorithm 1 and traditional Arnoldi method. The CPU times in computing ten desired eigenpairs (i.e., $2N = 10$) for 64^o YX-$LiNbO_3$ with various frequencies by using Algorithm 1 and traditional Arnoldi method are shown in Figure 8. On average, Algorithm 1 only takes 476 seconds of CPU time to compute the desired eigen pairs for all frequency ω in the search range. Obviously, the proposed Algorithm 1 is more efficient compared to the traditional Arnoldi method which takes 527 seconds of CPU time to get all the desired eigenpairs.

Method	$2N$	Computed eigenvalues	#Iter	$\max\{\lvert\lambda_{k,I}\lambda_{k,O} - 1\rvert\}$
Arnoldi	8	$\{(\lambda_{k,I}, \lambda_{k,O})\}_{k=1}^{4}$	2	1.7×10^{-13}
	10	$\{(\lambda_{k,I}, \lambda_{k,O})\}_{k=1}^{4}$, $\{\tilde{\lambda}_k\}_{k=1}^{2}$	5	1.7×10^{-13}
	12	$\{(\lambda_{k,I}, \lambda_{k,O})\}_{k=1}^{4}$, $\{\tilde{\lambda}_k\}_{k=1}^{4}$	4	1.7×10^{-13}
Algorithm 1	8	$\{(\lambda_{k,I}, \lambda_{k,O})\}_{k=1}^{4}$	2	1.1×10^{-16}
	10	$\{(\lambda_{k,I}, \lambda_{k,O})\}_{k=1}^{5}$	2	1.1×10^{-16}
	12	$\{(\lambda_{k,I}, \lambda_{k,O})\}_{k=1}^{6}$	2	1.1×10^{-16}
	18	$\{(\lambda_{k,I}, \lambda_{k,O})\}_{k=1}^{9}$	2	2.2×10^{-16}

Table 4: Convergent reciprocal pairs and the associated errors of reciprocity for 64^o YX-$LiNbO_3$ at frequency 2.180 GHZ versus different eigensolvers with various "2N" which denotes the number of interested eigenvalues. Here, $\tilde{\lambda}_k \in \mathbb{O}(\lambda)$ for $k = 1, \ldots, 4$.

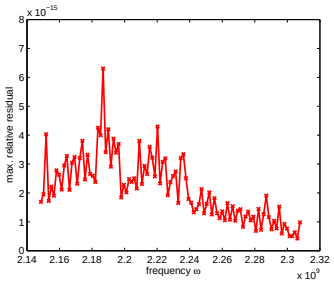


Figure 7: Maximal relative residuals of the ten desired eigenvalues produced by Algorithm 1.

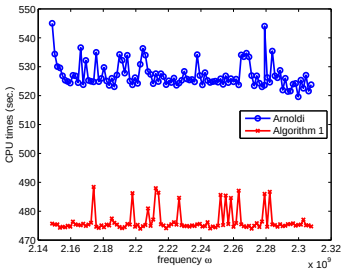


Figure 8: CPU times for computing ten desired eigenpairs by the traditional Arnoldi method and Algorithm 1.

5. Conclusions

In this paper, we have modeled the leaky surface acoustic wave propagation on a simple resonator with an interdigital transducer (IDT) where electrodes are arranged periodically on piezoelectric substrates (PZT) such as 64^o YX-$LiNbO_3$ and 36^o YX-$LiTaO_3$. The energy conservation equation (4) is discretized by finite element method (FEM) on a single cell domain with proper periodic boundary conditions as shown in Figure 2. Equation (4) is discretized on locally refined meshes in order to increase the accuracy of our numerical solutions. Our FEM simulation for predicting the center frequency of the stopping bands of the resonator is convergent and accurate within an error about 1% compared to experimental data as shown in Tables 2 and 3. For computing the dispersion diagram near the center of stopping band of the resonator, we transform the GEP (11) into the TPQEP (2) to reveal the important reciprocal relationship of the eigenvalues in which the eigenvalues appear in reciprocal pairs $(\lambda, 1/\lambda)$. The TPQEP (2) is then solved by GTSHIRA so that the reciprocal relationship of the eigenvalues can be automatically preserved. Our numerical results show that the traditional Arnoldi method converges slowly and fails to preserve the reciprocal property of the eigenvalues near the unit circle. On the other hand, the proposed structure-preserving method in Algorithm 1 not only converges to those eigenpairs faster than that of the traditional Arnoldi method but also possesses high accuracy in terms of relative residual error. Furthermore, the reciprocal property of the eigenpairs are kept nicely under machine precision.

Searching a good crystal cut of various PZTs for high frequency filter design based on LSAW is important. Our numerical studies here show that the dispersion diagram of a resonator with a prescribed crystal cut on its PZT substrate can be computed accurately and efficiently by discretizing the model equation on locally refined meshes and solving the resulted GEP by the proposed structure-preserving method in Algorithm 1. As a result, the computation time

in searching effective crystal cuts can be shorten and the computed dispersion diagrams can be more accurate and reliable.

Acknowledgments

This work is partially supported by the National Science Council, the Taida Institute of Mathematical Sciences, the Center for Advanced Study in Theoretical Sciences and the National Center for Theoretical Sciences in Taiwan. The corresponding author C. T. Wu thanks the support from National Science Council in Taiwan under the grant number 99-2115-M-009-001.

References

[1] H. Allik and T. Hughes. Finite element method for piezoelectric vibration. *Int. J. Numer. Methods Eng.*, 2:151–157, 1970.

[2] M. B. Angel, M. I. Rocha-Gaso, M. I. Carmen, and A.V. Antonio. Surface generated acoustic wave biosensors for detection of pathogens: A review. *Sensors*, 9:5740–5769, 2009.

[3] S. N. Atluri and M. Nakagaki. Computational methods for plane problems of fracture. *Computational Method in Mechanics of Fracture*, North-Holland Publishing Co, Amsterdam:170–227, 1986.

[4] B. A. Auld. *Acoustic Fields and Waves in Solids.* Krieger Publishing Company, Malabar, Florida, 1990.

[5] W. Bond. The mathematics of the physical properties of crystals. *BSTJ*, 22:1–72, 1943.

[6] M. Buchner, W. Ruile, A. Dietz, and R. Dill. FEM analysis of the reflection coefficient of SAWS in an infinite periodic array. In *Proc. IEEE Ultrason. Symp.*, 371-375, 1991.

[7] C. Cai, H. Zheng, M. S. Khan, and K. C. Hung. Modeling of material damping properties in ansys. *Inernational ANSYS conference Proceedings*, 2002.

[8] C. K. Campbell. *Surface Acoustic Wave Devices for Mobile and Wireless Communications.* Academic Press, INC., 1998.

[9] D. P. Chen and H. A. Haus. Analysis of metal-strip SAW gratings and transducers. *IEEE Trans. on Sonics and Ultrasonics*, 32, No. 3:395–408, 1985.

[10] L. C. Chin, V. V. Vardan, and V. K. Vardan. Hybrid finite element formulation for periodic piezoelectric arrays subjected to fluid loading. *Int. J. Numer. Meth. Engrg.*, 37:2987–3003, 1994.

[11] 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.

[12] P. Grisvard. *Elliptic Problems in nonsmooth domains*. Pitman Publishing INC., 1985.

[13] A Hilliges, C. Mehl, and V. Mehrmann. On the solution of palindramic eigenvalue problems. In *Proceedings 4th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS)*, Jyväskylä, Finland, 2004.

[14] M. Hofer, N. Finger, G. Kovacs, J. Schöberl, S. Zaglmayr, U. Langer, and R. Lerch. Finite-element simulation of wave propagation in periodic piezoelectric SAW structure. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, 53 No. 6:1192–1201, 2006.

[15] 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.

[16] C.F. Ipsen. Accurate eigenvalues for fast trains. *SIAM News*, 37, 2004.

[17] M. Koshiba, S. Mitobe, and M. Suzuki. Finite-element solution of periodic waveguides for acoustic waves. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, 34, No. 4:472–477, 1987.

[18] J. Kushibiki, I. Takanaga, and T. Sannomiya. Accurate measurements of the acoustical physical constants of $LiNbO_3$ and $LiTaO_3$ single crystals. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, 46:1315–1323, 1999.

[19] R. Lerch. Simulation of piezoelectric devices by two- and three-dimensional finite elements. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, 37, No. 2:1990, 1990.

[20] W.-W. Lin. A new method for computing the closed-loop eigenvalues of a discrete-time algebraic Riccatic equation. *Linear Alg. Appl.*, 96:157–180, 1987.

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

[22] D.S. Mackey, N. Mackey, C. Mehl, and V. Mehrmann. Vector spaces of linearizations for matrix polynomials. *SIAM J. Matrix Anal. Appl.*, 28:971–1004, 2006.

[23] J. E. San Miguel Medina, F. Buiochi, and J. C. Adamowski. Numerical modeling of a circular piezoelectric ultrasonic transduer radiating in water. *ABCM Symposium Series in Mechatronics*, 2:458–464, 2006.

[24] S. Mitobe, M. Koshiba, and M. Suzuki. Theoretical determination of equivalent circuit parameters for reflective SAW metallic gratings. *Electr. Commun. Jpn*, 70, No. 6:37–45, 1987.

[25] M. Mohamed, EL Gowini, and W. A. Moussa. A finite element model of a mems-based surface acoustic wave hydrongen sensor. *Sensors*, 10:1232–1250, 2010.

[26] G. Nadar, E. C. Nelli Silva, and J. C. Adamowski. Effective damping value of piezoelectric transduer determined by experimental techniques and numerical analysis. *ABCM Symposium Series in Mechatronics*, 1:271–279, 2004.

[27] R. V. Patel. On computing the eigenvalues of a symplectic pencil. *Linear Alg. Appl.*, 188:591–611, 1993.

[28] U. Rösler, D. Cohrs, A. Dietz, G. Fisherauer, W. Ruile, P. Russer, and R. Weigel. Determination of leaky saw propagation, reflection and coupling on $LiTaO_3$. *IEEE Ultrasonics Symposium*, pages 247–250, 1995.

[29] C. Schröder. A QR-like algorithm for the palindromic eigenvalue problem. Technical report, Preprint 388, TU Berlin, Matheon, Germany, 2007.

[30] C. Schröder. URV decomposition based structured methods for palindromic and even eigenvalue problems. Technical report, Preprint 375, TU Berlin, MATHEON, Germany, 2007.

Numerical Studies on the Monge-Ampère Equation and the Reconstruction of 1-D Reflector using Hermite Finite Element

Yu-Lin Tsai[1], Ming-Chiang Jiang[2] and Chin-Tien Wu[2]

1. Department of Photonics, National Chiao-Tung University, Hsinchu, Taiwan
2. Department of Applied Mathematics, National Chiao-Tung University, Hsinchu, Taiwan

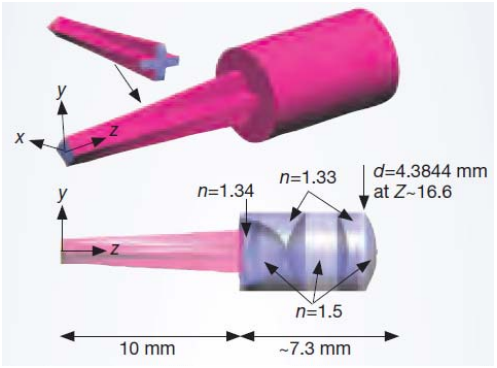
The 7th EASIAM 2011, Waseda University, Kitakyushu Campus, Japan

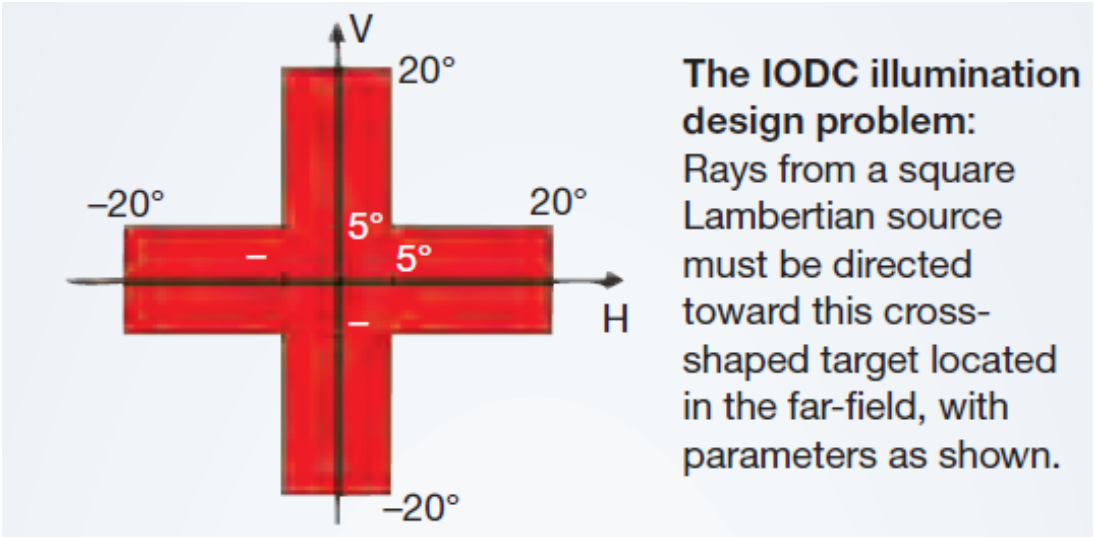
Outline

- Introduction and Motivations
- Modeling of Free-Form Surface Design
- Monge-Apère Equation
- Numerical Methods for Solving Monge-Apère Equation
- Numerical Results
- Reconstruction Algorithm for 1-D Surface
- Conclusions

Free-form surface design problem

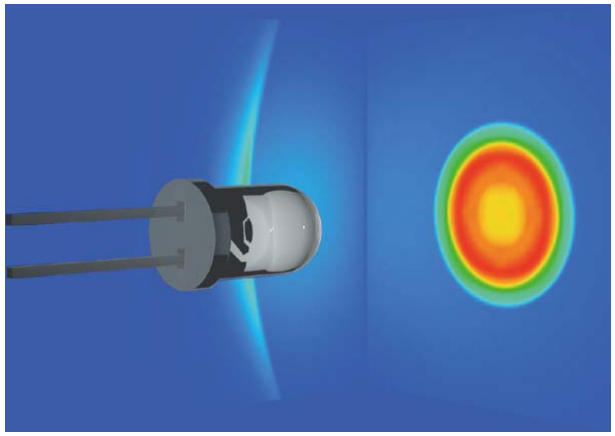
International Optical Design Conference 2006



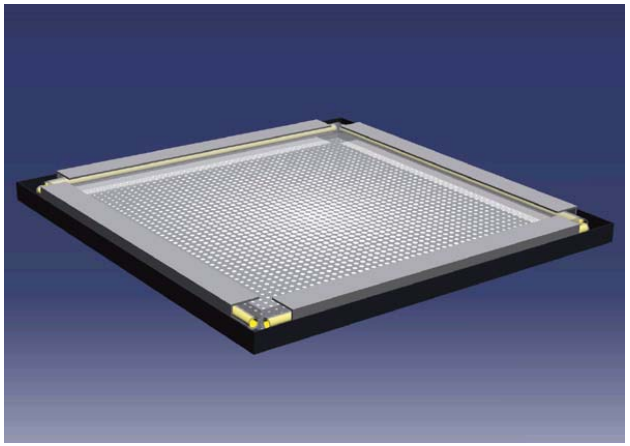


Question: How do we construct an optical system to do this such that the energy loss is minimized?

Some applications in industry



Illuminance distribution created by bullet-type LED



Sidelight-type backlight unit

interior lighting of passenger car

Radar Antenna Reflector Design

The equation appears in various fields

Differential geometry

- L. A. Caffarelli and X. Cabre, "Fully nonlinear elliptic equations," Vol. 43, AMS, 1995.
- S. Y. Chen and S. T. Yau, On the regularity of the Monge-Ampere equation, Comm. Pure Appl. Math., 30, pp.41-68, 1977.

Optimal control and mass transportation

- L. A. Caffarelli and M. Milman, Monge Ampere Equation: Applications to Geometry and Optimization. Contemporary Mathematics, AMS, 1999.

Meteorology and Geostrophic fluid

- R. J. McCann and A. M. Oberman. Exact semi-geostrophic flows in an elliptical ocean basin. Nonlinearity, 17(5), pp.1891-1922, 2004

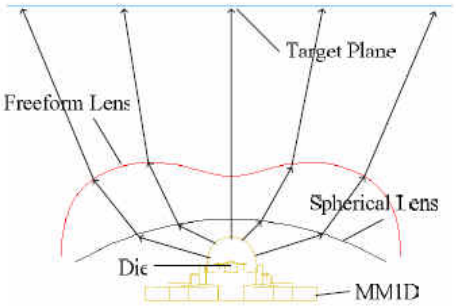
Optical Design

- P. Benítez and J. C. Miñano, The Future of ILLUMINATION DESIGN. OPN 2007
- Pengfei Guan and Xu-Jia Wang, On a Monge-Ampere equation arising in geometric optics, J. Diff. Geometry, Vol. 48, pp 205-223, 1998.
- V. Oliker and E. Newman, The energy conservation equation in the reflector mapping problem. Appl. Math. Lett. Vol. 6, pp. 91-95, 1993.

Freeform LED lens for uniform illumination

Yi Ding, Xu Liu*, Zhen-rong Zheng, and Pei-fu Gu, Optics express, 2008.

State Key Laboratory of Modern Optical Instrumentation, Zhejiang University,Hang Zhou 310027, China



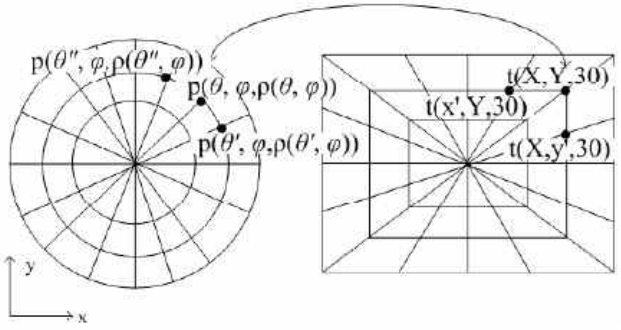


Fig. 3. The topological mapping from source to target plane.

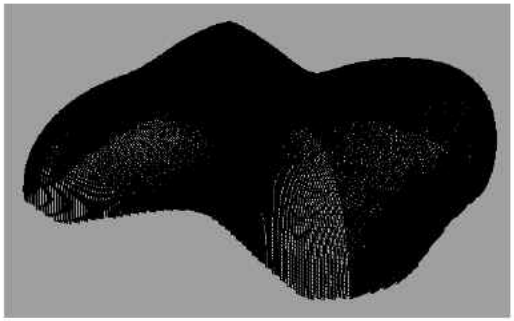


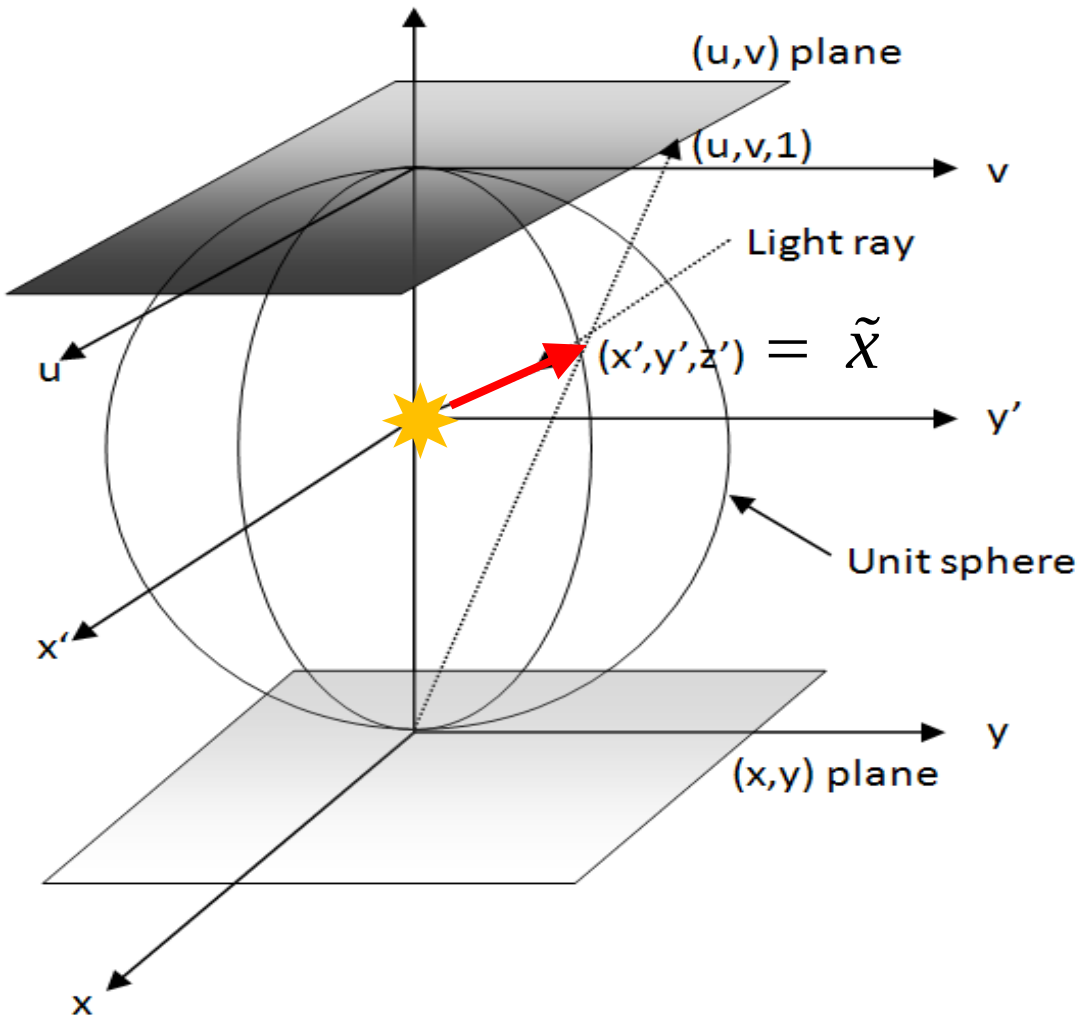
Fig. 4. The freeform lens.

Energy Conservation

$$\iint_{v(\Omega)} L(x,y)\,dxdy = \iint_{\Omega} I(u,v)\,d\Omega \qquad (1)$$

$L(x,y)$	illumination
I	intensity distribution
Ω	solid angles
$v(\Omega)$	the aperture region.

Stereographic Coordinate



Change variables

$$\tilde{x}(u,v)=\frac{1}{1+\frac{1}{4}w^2}\left(u,v,1-\frac{1}{4}w^2\right)$$

where $w^2=u^2+v^2$, we have

$$d\Omega=|\tilde{x}_u\times\tilde{x}_v|\,dudv=\left(1+\frac{1}{4}w^2\right)^{-2}dudv \qquad (2)$$

Substituting (2) into (1), then

$$\iint_{v(\Omega)} L(x,y)\,dxdy=\iint_{\Omega} I(u,v)\left(1+\frac{1}{4}w^2\right)^{-2}dudv \qquad (3)$$

Free-form surface of Reflector type

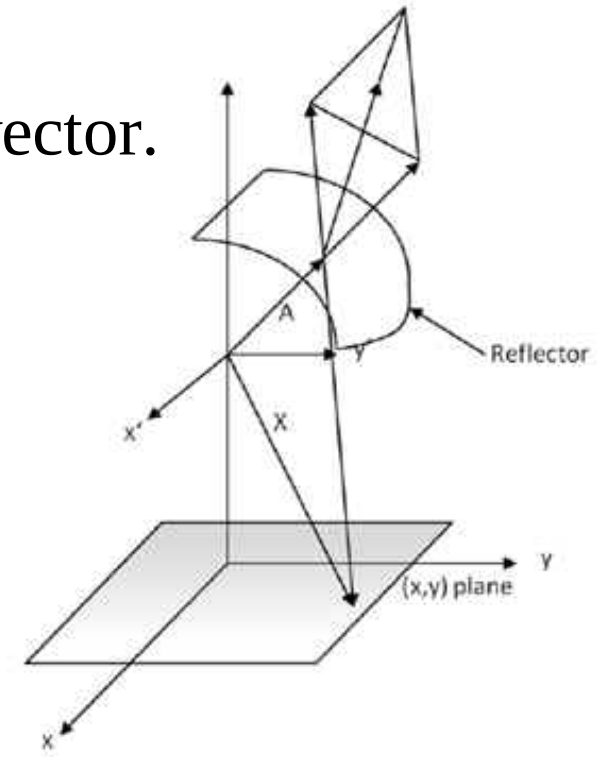
Consider a perfect reflecting surface

$A = \rho(u,v)\,\tilde{x}(u,v)$, where ρ is the distance from light source to the reflector A

Let $N = A_u \times A_v$ be the surface normal vector.

Law of reflection

$$\frac{\mathrm{A}}{\|\mathrm{A}\|} + \frac{A - X}{\|A - X\|} = \frac{2(A \cdot N)N}{\|A\|\|N\|^2}$$



$$\begin{bmatrix} x \\ y \\ -1 \end{bmatrix} = X = A + \|X - A\| \left[\tilde{x} - 2(\tilde{x} \cdot N) N / \|N\|^2 \right] \quad \cdots\cdots\cdots \begin{pmatrix} 4.1 \\ 4.2 \\ 4.3 \end{pmatrix}$$

where $N = A_u \times A_v = (\rho_u \tilde{x} + \rho \tilde{x}_u) \times (\rho_v \tilde{x} + \rho \tilde{x}_v)$

From the equation (4.3), one can obtain $\|X - A\|$.

Plug this result into (4.1) and (4.2), we have

$$x = uG + 2\rho\rho_u F$$
$$y = vG + 2\rho\rho_v F$$

where $G = G(u, v, \rho, \rho_u, \rho_v)$ and $F = F(u, v, \rho, \rho_u, \rho_v)$

$$G = \rho\left(1+\frac{1}{4}w^2\right)^{-1} + F\left[-\rho^2\left(1+\frac{1}{4}w^2\right)^{-2} + \rho_u^2 + \rho_v^2 - \rho\left(\rho_u u + \rho_v v\right)\left(1+\frac{1}{4}w^2\right)^{-1}\right]$$

$$F = \frac{1+\rho\left(1-\frac{1}{4}w^2\right)\left(1+\frac{1}{4}w^2\right)^{-1}}{\left(1-\frac{1}{4}w^2\right)\left[\rho^2\left(1+\frac{1}{4}w^2\right)^{-2} - \rho_u^2 - \rho_v^2\right] + 2\rho\left(\rho_u u + \rho_v v\right)\left(1+\frac{1}{4}w^2\right)^{-1}}$$

J. S. Schruben, Formulation of a reflector-design problem for a light fixture. J. Optical Society America, 62, 1972.

The equation (3) can now be rewritten as

$$\iint_{\Omega} L(u,v,\rho,\rho_u,\rho_v)\ D\ dudv = \iint_{\Omega} I(u,v)\ \left(1+\frac{1}{4}w^2\right)^{-2}\ dudv$$

where, considering x and y as functions of variables u,v,ρ,ρ_u,ρ_v, the Jacobian D can be represented as following:

$$D = \begin{vmatrix} x_u + \rho_u x_\rho + & x_v + \rho_v x_\rho + \\ \rho_{uu} x_{\rho_u} + \rho_{uv} x_{\rho_v} & \rho_{vu} x_{\rho_u} + \rho_{vv} x_{\rho_v} \\ y_u + \rho_u y_\rho + & y_v + \rho_v y_\rho + \\ \rho_{uu} y_{\rho_u} + \rho_{uv} y_{\rho_v} & \rho_{vu} y_{\rho_u} + \rho_{vv} y_{\rho_v} \end{vmatrix}$$

Monge-Ampère equation

$$\mathrm{L}\left(u,v,\rho,\rho_u,\rho_v\right)D = I\left(u,v\right)\left(1+\frac{1}{4}w^2\right)$$

where

$$D = J_{\rho_u\rho_v}\left(\rho_{uu}\rho_{vv}-\rho_{uv}^2\right)+\left(J_{\rho_u v}+\rho_v J_{\rho\rho_u}\right)\rho_{uu}+$$
$$\left(J_{\rho_v v}+J_{u\rho_u}+\rho_u J_{\rho\rho_u}+\rho_v J_{\rho_v\rho}\right)\rho_{uv}+$$
$$\left(J_{u\rho_v}+\rho_u J_{\rho\rho_v}\right)\rho_{vv}+J_{\rho v}\rho_u+J_{u\rho}\rho_v+J_{uv}$$

and

$$J_{\alpha\beta} = x_\alpha y_\beta - x_\beta y_\alpha \quad \text{for } \alpha,\beta\in\left\{u,v,\rho,\rho_u,\rho_v\right\}$$

A simplified expression by Oliker and Newman, 1993

$$D = \frac{1}{\left(\rho^2 + \rho_i \rho_j e^{ij}\right)^2} \frac{\det\left[2\rho\tilde{\nabla}_{ij}\rho - \left(\rho^2 - \rho_i\rho_j e^{ij}\right)e_{ij} - 4\rho_i\rho_j\right]}{\det\left[e_{ij}\right]}$$

where e_{ij} is jacobian associated with the first fundemental form of the unite sphere S, $e^{ij} = \left(e_{ij}\right)^{-1}$ and $\tilde{\nabla}_{ij}$ is the second covariant derivatives in the metric e on S.

In this formulation, the domain Ω and the aperture region $v(\Omega)$ are on the unit sphere S and the coordinate system is simply the spherical coordinate.

The Monge-Ampère equation

$$A\det\begin{bmatrix} u_{xx} & u_{xy} \\ u_{yx} & u_{yy} \end{bmatrix} + Bu_{xx} + Cu_{xy} + Du_{yy} + E = 0$$

where the coefficients A, B, C, D, E are functions depend on x, y, u, u_x, u_y

The equation appears in various fields

Differential geometry

- L. A. Caffarelli and X. Cabre, "Fully nonlinear elliptic equations," Vol. 43, AMS, 1995.
- S. Y. Chen and S. T. Yau, On the regularity of the Monge-Ampere equation, Comm. Pure Appl. Math., 30, pp.41-68, 1977.

Optimal control and mass transportation

- L. A. Caffarelli and M. Milman, Monge Ampere Equation: Applications to Geometry and Optimization. Contemporary Mathematics, AMS, 1999.

Meteorology and Geostrophic fluid

- R. J. McCann and A. M. Oberman. Exact semi-geostrophic flows in an elliptical ocean basin. Nonlinearity, 17(5), pp.1891-1922, 2004

Optical Design

- P. Benítez and J. C. Miñano, The Future of ILLUMINATION DESIGN. OPN 2007
- Pengfei Guan and Xu-Jia Wang, On a Monge-Ampere equation arising in geometric optics, J. Diff. Geometry, Vol. 48, pp 205-223, 1998.
- V. Oliker and E. Newman, The energy conservation equation in the reflector mapping problem. Appl. Math. Lett. Vol. 6, pp. 91-95, 1993.

Some results and review

- Boundary conditions: ρ, ρ_u, ρ_v should be given on $\partial\Omega$ to ensure the map $\partial\Omega \to \partial V(\Omega)$ is deffeomorphism.
- Existence and uniqueness are obtained by Mader (1981), Caffarelli and Oliker (1993, 2008).
- Regularity is estimated by Chen-Yau (1976), and recently by Pengfei Guan and Xu-Jia Wang (1998).

  In Guan and Wang's result, $u \in C^{1,1}(\Omega)$ if provided $L,\ I \in C^{1,1}(\Omega)$ and I is non-negative.
- Numerics are carried out recently by Oberman etc. (2008), Dean and Glowinski (2003) and Feng and Neilan (2007) for simple Monge-Ampere equations.

Flat Form Assumption

Assume the free-form is flat corresponding to a small aperture region, i.e. $w, \rho_u, \rho_v << 1$

The Jacobian can be further simplified as following:

$$D = \left(2 + \frac{2}{\rho}\right)^2 \det\begin{bmatrix} \rho_{uu} & \rho_{uv} \\ \rho_{vu} & \rho_{vv} \end{bmatrix} - \left(2 + \frac{2}{\rho}\right)\Delta u + 1$$

In our numerical studies, we only consider the simplest version: $\det\begin{bmatrix} \rho_{uu} & \rho_{uv} \\ \rho_{vu} & \rho_{vv} \end{bmatrix} = f(u,v).$

Benamou, Froese and Oberman Finite Difference Scheme

First method:

$$\left(D^2_{xx}u_{ij}\right)\left(D^2_{yy}u_{ij}\right)-\left(D^2_{xy}u_{ij}\right)^2=f_{ij} \tag{4}$$

where

$$\begin{aligned} D^2_{xx}u_{ij} &= \frac{1}{h^2}\left(u_{i+1,j}+u_{i-1,j}-2u_{i,j}\right) \\ D^2_{yy}u_{ij} &= \frac{1}{h^2}\left(u_{i,j+1}+u_{i,j-1}-2u_{i,j}\right) \\ D^2_{xy}u_{ij} &= \frac{1}{4h^2}\left(u_{i+1,j+1}+u_{i-1,j-1}-u_{i-1,j+1}-u_{i+1,j-1}\right) \end{aligned} \tag{5}$$

J. D. Benamou, B. D. Froese and A. D. Oberman, .Two Numerical Method for the Elliptic Monge-Ampere Equation.,Preprint, 2009

Rewrite the scheme of Monge-Ampere as a quadratic equation for $u_{i,j}$

$$u_{i,j} = \frac{1}{2}\left(a_1 + a_2\right) - \frac{1}{2}\sqrt{\left(a_1 - a_2\right)^2 + \frac{1}{4}\left(a_3 - a_4\right)^2 + h^4 f_{i,j}} \tag{6}$$

where

$$\begin{aligned} a_1 &= \left(u_{i+1,j} + u_{i-1,j}\right)/2 \\ a_2 &= \left(u_{i,j+1} + u_{i,j-1}\right)/2 \\ a_3 &= \left(u_{i+1,j+1} + u_{i-1,j-1}\right)/2 \\ a_4 &= \left(u_{i-1,j+1} + u_{i+1,j-1}\right)/2 \end{aligned} \tag{7}$$

Second method:
fixed point iteration for solving $u = T(u)$ where

$$T(u) = \triangle^{-1}\left(\sqrt{(\triangle u)^2 + 2\left(f - \det(D^2 u)\right)}\right) \tag{8}$$

the consists of iterating $u^{n+1} = T(u^n)$ by solving

$$\triangle u^{n+1} = \sqrt{\left(u_{xx}^n\right)^2 + \left(u_{yy}^n\right)^2 + 2\left(u_{xy}^n\right)^2 + 2f} \tag{9}$$

Feng and Neilan vanishing moment method

Feng and Neilan appling $-\epsilon\triangle^2 u^\epsilon$ to the Monge-Ampere equation and adding a boundary condition $\triangle u^\epsilon = \epsilon$ on $\partial\Omega$, find u^ϵ such that

$$\begin{aligned} -\epsilon\triangle^2 u^\epsilon + \det\left(D^2 u^\epsilon\right) &= f, \text{ in } \Omega \\ u^\epsilon &= g \text{ on } \partial\Omega \\ \triangle u^\epsilon &= \epsilon \text{ on } \partial\Omega \end{aligned} \tag{10}$$

the $h-\epsilon$ relation, ϵ can be small about h^2. h is mesh size.
Then Feng separate (10) into couple partial difference equations system

$$\begin{aligned} \sigma^\epsilon - D^2 u^\epsilon &= 0 \\ -\epsilon\triangle tr\left(\sigma^\epsilon\right) + \det\left(\sigma^\epsilon\right) &= f \end{aligned} \tag{11}$$

they use mixed finite element.

Xiaobing Feng and Michael Neilan. Mixed finite element methods for the fuiiy noniinear Monge-Ampere equation based on the vanishing moment method.. SIAM J. Numer. Anal.,47(2):1226{1250, 2009.

Our Approach

Following Feng and Neilan approach (10)

- Solve the regularized problem directly instead of decoupling the equation into low order system equations.

Using BCIZ element

- State variables are the nodal values and derivatives.
- The solution of BCIZ element is continuous
- Represent constant curvature

Nonlinear iteration

- Newton's iteration

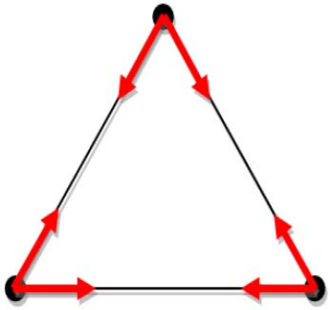
BCIZ Element

Merit

- Has been extensively tested for biharmonic problems.
- BCIZ is C^0 on uniform meshes that satisfies necessary condition of the regularity result obtained by John Urbas, 1988 (Math. Z. vol. 197).

Demerits

- The BCIZ element is sensitive to gird structure.



$$
\begin{aligned}
\varphi_1 &= \zeta_1^2(3-2\zeta_1) + 2\zeta_1\zeta_2\zeta_3 \\
\varphi_2 &= -\zeta_1^2(y_{12}\zeta_2 + y_{13}\zeta_3) - \frac{1}{2}(y_{12}+y_{13})\zeta_1\zeta_2\zeta_3 \\
\varphi_3 &= \zeta_1^2(x_{12}\zeta_2 + x_{13}\zeta_3) + \frac{1}{2}(x_{12}+x_{13})\zeta_1\zeta_2\zeta_3 \\
\varphi_4 &= \zeta_2^2(3-2\zeta_2) + 2\zeta_1\zeta_2\zeta_3 \\
\varphi_5 &= -\zeta_2^2(y_{23}\zeta_3 + y_{21}\zeta_1) - \frac{1}{2}(y_{23}+y_{21})\zeta_1\zeta_2\zeta_3 \\
\varphi_6 &= \zeta_2^2(x_{23}\zeta_3 + x_{21}\zeta_1) + \frac{1}{2}(x_{23}+x_{21})\zeta_1\zeta_2\zeta_3 \\
\varphi_7 &= \zeta_3^2(3-2\zeta_3) + 2\zeta_1\zeta_2\zeta_3 \\
\varphi_8 &= -\zeta_3^2(y_{31}\zeta_1 + y_{32}\zeta_2) - \frac{1}{2}(y_{31}+y_{32})\zeta_1\zeta_2\zeta_3 \\
\varphi_9 &= \zeta_1^2(x_{31}\zeta_1 + x_{32}\zeta_2) + \frac{1}{2}(x_{31}+x_{32})\zeta_1\zeta_2\zeta_3
\end{aligned}
$$

C. A. Felippa and B. Haugen, From the Individual Element Test to Finite Element Templates: Evolution the Patch test., Int. J. Numer. Methods Engr., Vol. 38, 199-229 (1995)

Linearization

Linearization $\det\left(D^2 u\right)$

$$\delta_u \det(D^2 u) = (D_{yy}\, u)\,(D_{xx}) + (D_{xx}\, u)\,(D_{yy}) - 2\,(D_{xy}\, u)\,(D_{xy}) \quad (12)$$

Newton step:

$$-\epsilon \triangle^2 (\delta u^n) + div\left(\begin{bmatrix} D_{yy}\, u^n & -D_{xy}\, u^n \\ -D_{xy}\, u^n & D_{xx}\, u^n \end{bmatrix} \nabla (\delta u^n)\right) = \quad (13)$$

$$f - \det(D^2 u^n) + \epsilon \triangle^2 u^n$$

Weak Formulation

The variational problem is:

Find $u^\epsilon \in H^2(\Omega)$, $u^\epsilon = g$, $\triangle u^\epsilon = \epsilon$ on $\partial\Omega$ such that

$$\epsilon \int_\Omega \triangle u^\epsilon \triangle v dx + \int_\Omega \left(cof\left(D^2 u^\epsilon\right) Du^\epsilon \right) \cdot Dv \; dx + \int_\Omega f v dx = 0$$

where $cof\left(D^2 u^\epsilon\right) = \begin{bmatrix} u^\epsilon_{yy} & -u^\epsilon_{xy} \\ -u^\epsilon_{xy} & u^\epsilon_{xx} \end{bmatrix}$

Error estimations from Feng and Neilan

$$\left\| u^\varepsilon - u_h^\varepsilon \right\|_{L^2} \le C(\varepsilon)\varepsilon^{-2}\left[\frac{h^l}{\sqrt{\varepsilon}} \left\| u^\varepsilon \right\|_{H^l} + \varepsilon^{-2} h^{2l-4} \left\| u^\varepsilon \right\|_{H^l}^2 \right] \quad \text{if } u^\varepsilon \text{ has } \mathrm{H}^4 \text{ regularity}$$

$$\left\| u^\varepsilon - u_h^\varepsilon \right\|_{H^1} \le C(\varepsilon)\varepsilon^{-2}\left[h^{l-1} + \varepsilon^{-5/2} h^{2(l-2)} \right]\left(\left\| \sigma^\varepsilon \right\|_{H^l} + \left\| u^\varepsilon \right\|_{H^l} \right)$$

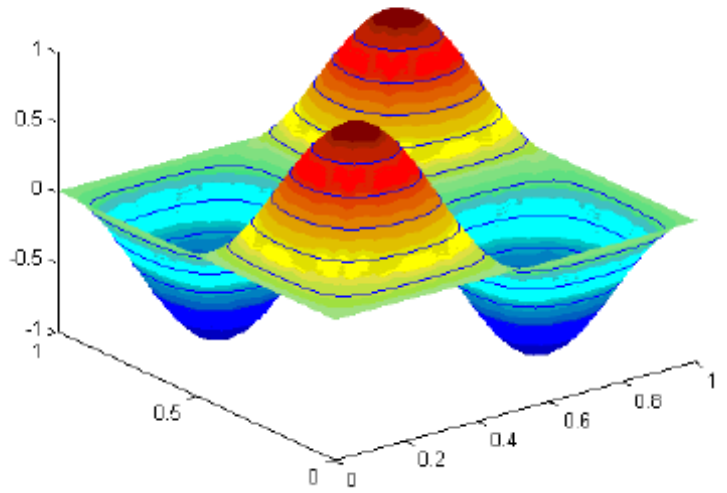
Error check for Laplace

The analytical solution

$u = \sin(2\pi x)\sin(2\pi y)$

$f = -8\pi^2 \sin(2\pi x)\sin(2\pi y)$

solution u_h :



Mash size	$\|u-u_h\|_\infty$	$\|u-u_h\|_{L^2}$	$\|u-u_h\|_{H^1}$
2^{-3}	0.001228	0.000803	0.043738
2^{-4}	0.000167	7.54E-05	0.004195
2^{-5}	1.32E-05	7.24E-06	0.000532
2^{-6}	8.83E-07	7.85E-07	0.000118

The convergence rate of L^2-norm is third order and H^1-norm is second order

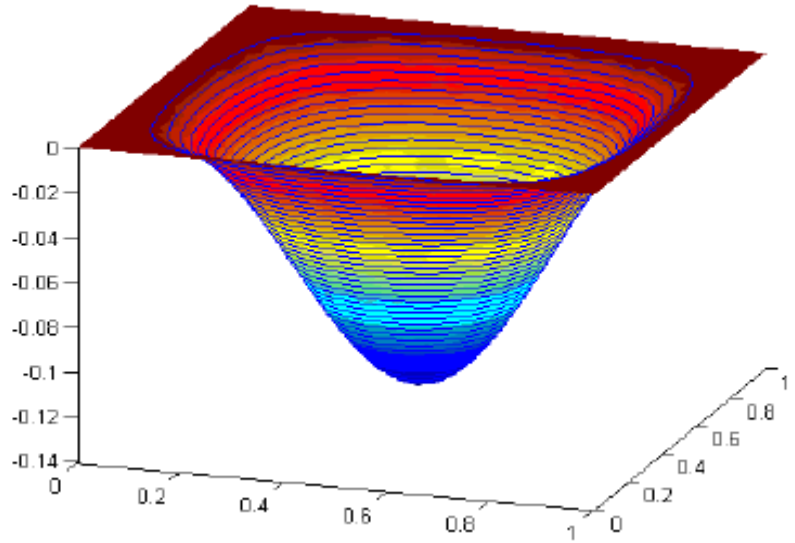
Error check for Biharnomic equation

The analytical solution

$u = (\cos(2\pi x) - 1)\,(y^2 - 2y^3 + y^4)$

$f = 0$

solution u_h :



Mash size	$\|u-u_h\|_\infty$	$\|u-u_h\|_{L^2}$	$\|u-u_h\|_{H^2}$
2^{-2}	0.01172	0.004154	0.645238
2^{-3}	0.004269	0.001568	0.307797
2^{-4}	0.001169	0.000445	0.150572
2^{-5}	0.000302	0.000117	0.074556
2^{-6}	7.67E-05	2.98E-05	0.037112
2^{-7}	1.93E-05	7.51E-06	0.018516

The convergence rate of L^2-norm is second order and H^2-norm is first order

Test case 1.

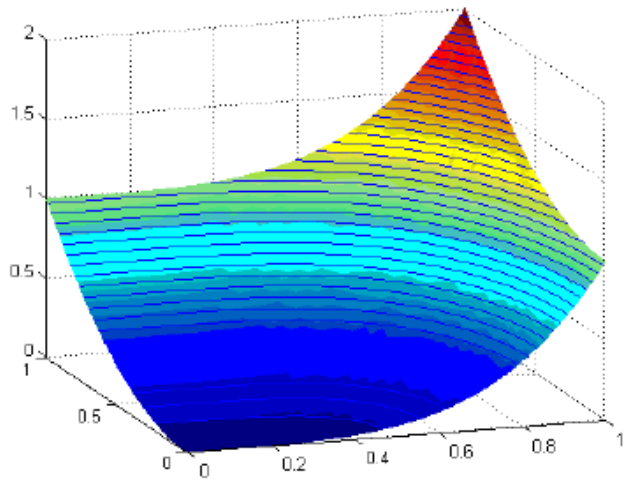
The analytical solution

$u = x^4 + y^2$

$f = 24x^2$

mesh size=$1/256$

solution u_h^ϵ :



ε	$\|u^\varepsilon - u_h\|_\infty$	$\|u^\varepsilon - u_h\|_{L^2}$	$\|u^\varepsilon - u_h\|_{H^2}$	Iter
1	0.305	0.161	5.319208	6
2^{-2}	0.230	0.121	4.673278	10
2^{-4}	0.113	0.0571	3.634048	10
2^{-6}	0.0428	0.0190	2.71177	8
2^{-8}	0.0145	5.67e-03	1.98896	8
2^{-10}	4.50e-03	1.60e-03	1.437771	8
2^{-12}	1.29e-03	4.33e-04	1.029425	9
2^{-14}	3.49e-04	0.000113	0.732763	10

Ours		Feng's	
ε	$\left\|u^{\varepsilon}-u_h\right\|_{L^2}$	ε	$\left\|u^{\varepsilon}-u_h\right\|_{L^2}$
2^{-1}	0.145332	0.5	0.082589
2^{-2}	0.120509	0.25	0.074746
2^{-3}	0.088363	0.1	0.051429
2^{-4}	0.057111	0.05	0.033436
2^{-6}	0.019037	0.0125	0.011590
2^{-9}	0.003027	0.0025	0.002939
2^{-11}	0.000836	0.0005	0.000679

ϵ	$\frac{\left\|u^0-u_h^\epsilon\right\|_{L^2}}{\epsilon}$	$\frac{\left\|u^0-u_h^\epsilon\right\|_{H^2}}{\sqrt[4]{\epsilon}}$
1	0.160842924	5.319207667
2^{-2}	0.482036757	6.609013597
2^{-4}	0.913779843	7.268095576
2^{-6}	1.218354047	7.670044224
2^{-8}	1.450567862	7.955838131
2^{-10}	1.637202829	8.133261803
2^{-12}	1.772707203	8.235401948
2^{-14}	1.85193153	8.290262971

Table 6. Change of $\|u^0 - u_h^\epsilon\|$ w.r.t. ϵ $(h = 2^{-8})$

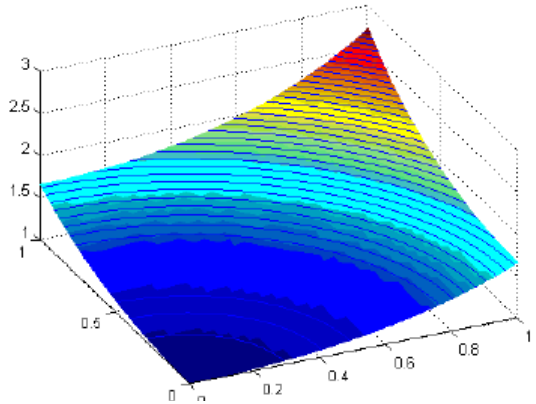
Test case 2.

The analytical solution

$$u = e^{\frac{x^2+y^2}{2}}$$

$$f = \left(1 + x^2 + y^2\right) e^{x^2+y^2}$$

solution u_h^ϵ :



mesh size=1/256

Ours		Oberman's		
N	iter	N	M1	M2
128	69	81	14621	59

ε	$\|u^\varepsilon - u_h\|_\infty$	$\|u^\varepsilon - u_h\|_{L^2}$	$\|u^\varepsilon - u_h\|_{H^2}$	Iter
1	0.1776024	0.100518	3.026106	29
2^{-2}	0.1413431	0.081722	2.715974	48
2^{-4}	0.0713641	0.044524	2.129656	38
2^{-6}	0.0226147	0.01562	1.567714	9
2^{-8}	0.0063037	0.004523	1.12653	8
2^{-10}	0.001811	0.001218	0.800411	8
2^{-12}	0.0004998	0.000316	0.566837	8
2^{-14}	0.0001327	8.00E-05	0.401068	9

Ours		Feng's	
ε	$\left\|u^{\varepsilon}-u_h\right\|_{L^2}$	ε	$\left\|u^{\varepsilon}-u_h\right\|_{L^2}$
2^{-1}	0.0936	0.5	0.0387
2^{-2}	0.0817	0.25	0.0410
2^{-3}	0.0644	0.1	0.0322
2^{-4}	0.0445	0.05	0.0223
2^{-6}	0.0156	0.0125	7.82e-03
2^{-9}	2.36e-03	0.0025	1.86e-03
2^{-11}	6.23e-04	0.0005	4.04e-04

ϵ	$\frac{\left\|u^0-u_h^\epsilon\right\|_{L^2}}{\epsilon}$	$\frac{\left\|u^0-u_h^\epsilon\right\|_{H^2}}{\sqrt[4]{\epsilon}}$
1	0.100518486	3.026105949
2^{-2}	0.32688765	3.840967403
2^{-4}	0.71238757	4.259311736
2^{-6}	0.99970787	4.434163657
2^{-8}	1.157804247	4.506118399
2^{-10}	1.24700059	4.527808719
2^{-12}	1.294427725	4.534697517
2^{-14}	1.309904805	4.53757113

Table 10. Change of $\|u - u_h^\epsilon\|$ w.r.t. ϵ $(h = 2^{-8})$

Test case 3.

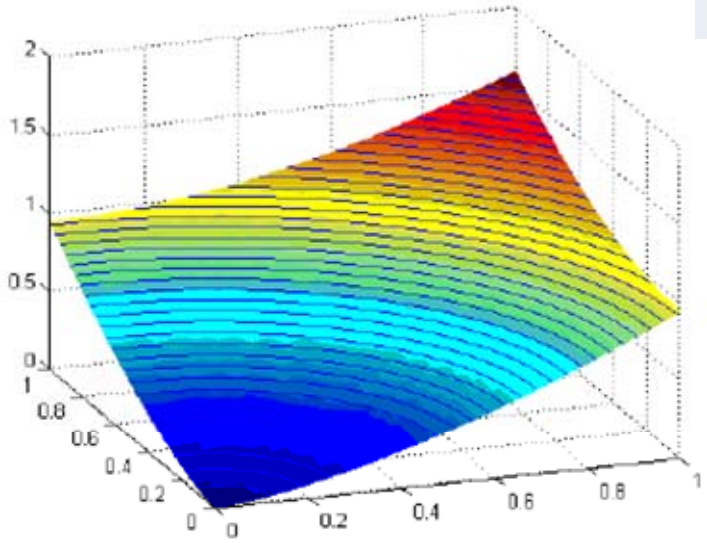
The analytical solution

$$u = \frac{2\sqrt{2}\left(x^2+y^2\right)^{\frac{3}{4}}}{3}$$

$$f = \frac{1}{\sqrt{x^2+y^2}}$$

mesh size=1/256

solution u_h^ϵ :



ε	$\left\|u^\varepsilon - u_h\right\|_\infty$	$\left\|u^\varepsilon - u_h\right\|_{L^2}$	$\left\|u^\varepsilon - u_h\right\|_{H^2}$	Iter
1	0.1413141	0.078907	2.164988	5
2^{-2}	0.122805	0.069333	2.010561	19
2^{-4}	0.075892	0.04483	1.63768	9
2^{-6}	0.0278187	0.018061	1.217238	10
2^{-8}	0.0080091	0.005569	0.894614	8
2^{-10}	0.0021189	0.001546	0.650581	8
2^{-12}	0.0005573	0.000408	0.468459	9
2^{-14}	0.0001439	0.000104	0.333575	11

Ours		Oberman's		
N	iter	N	M1	M2
128	79	101	23849	59

Test case 4.

The analytical solution

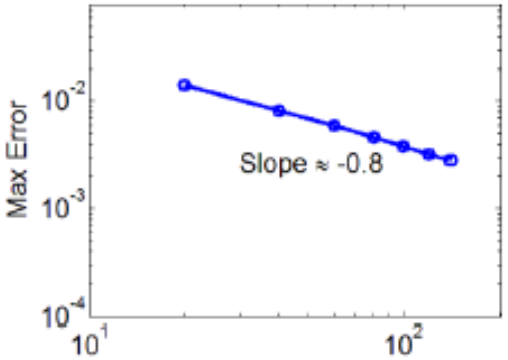
$u = \sqrt{x^2 + y^2}$

$f = 0$

mesh size=1/256

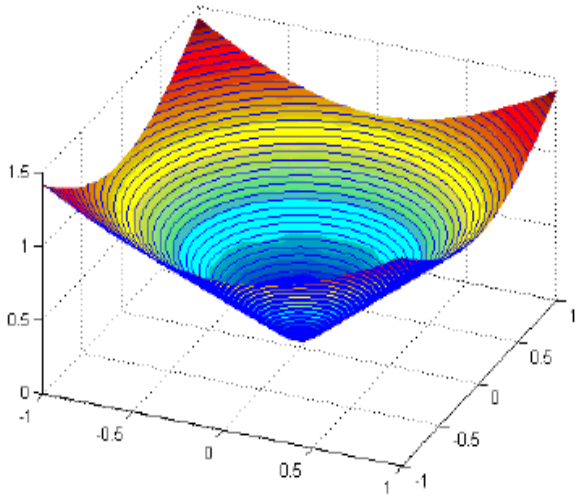
solution u_h^ϵ :

Oberman's result



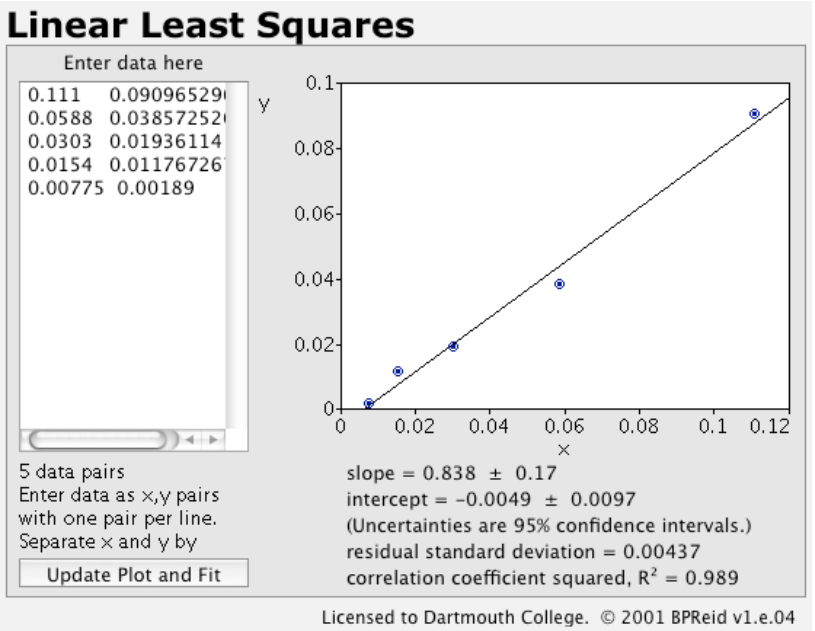
Number of points

Ours		Oberman's		
N	iter	N	M1	M2
127	62	121	39396	10486

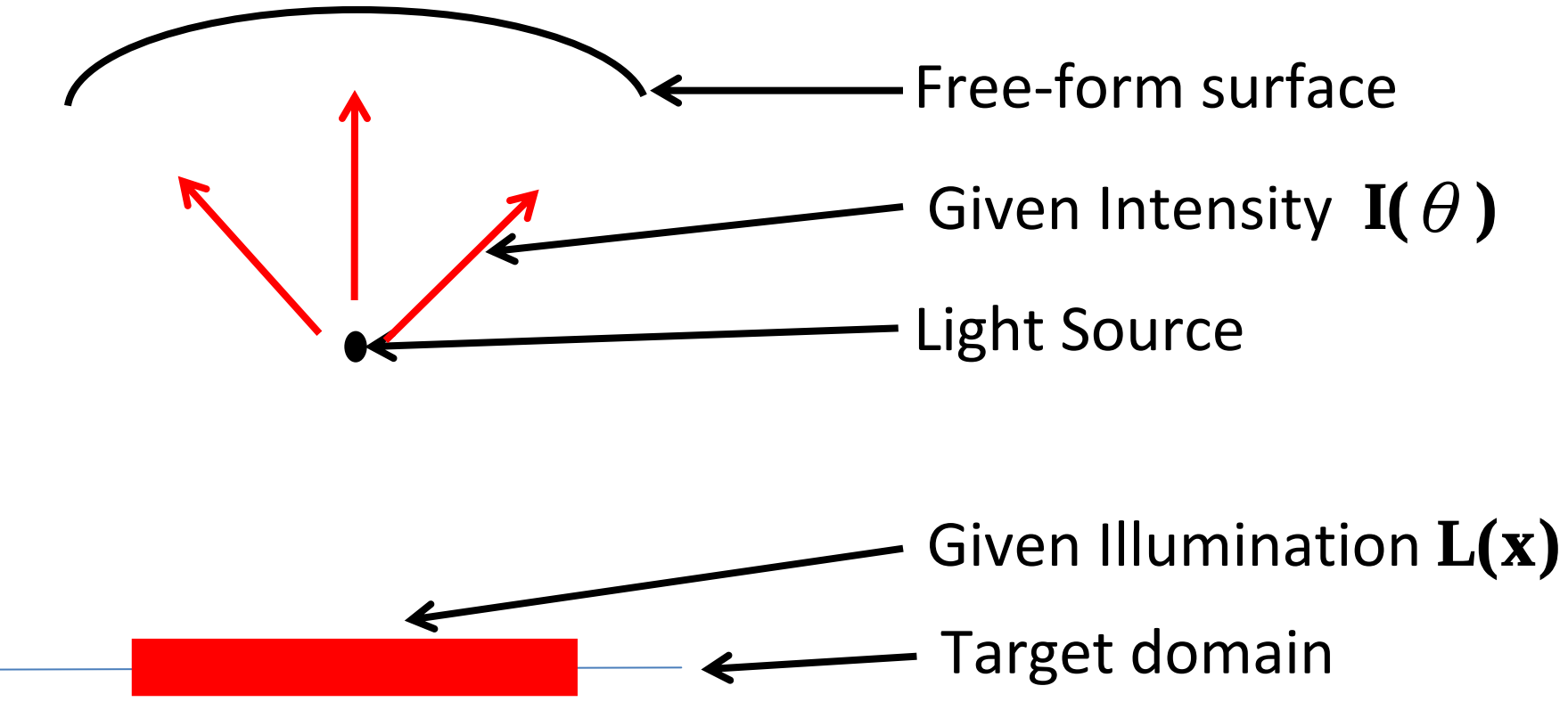


Error from our Simulation:

Slope=0.84



Reconstruction of 1-D Surface



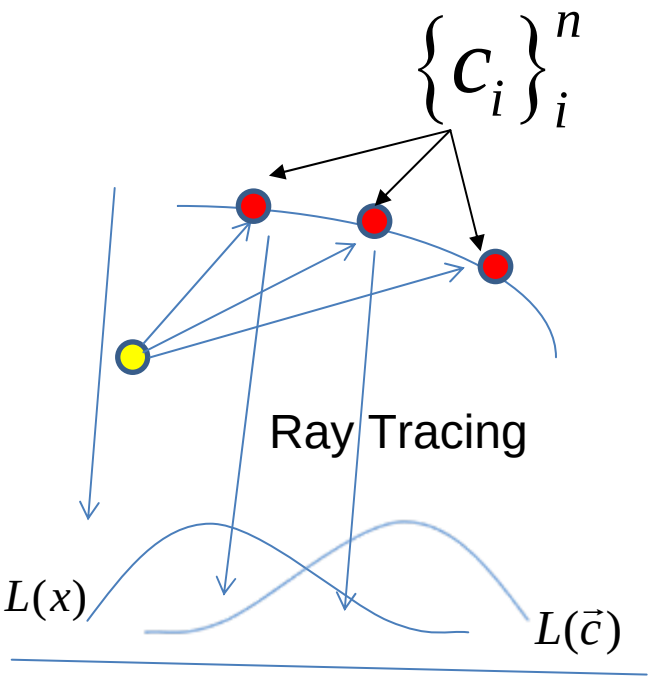
Previous Works

(1) Methods based on optimization and ray tracing. (Neubauer 94, Halstead 96, and Patow and Pueyo 03)

Given target illumination density, control points $\{c_i\}_i^n = \vec{c}$

Optimization: find $\vec{c}$ that minimize

$$\|L(x) - L(c)\|$$



Previous Works

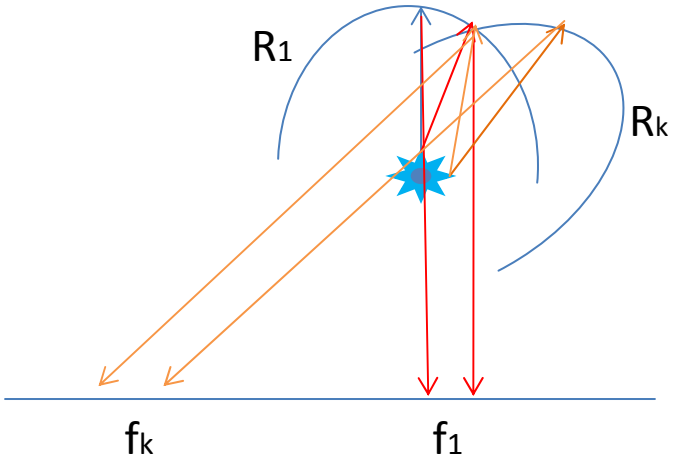
(2)Oliker's approximation

Kochengin and Oliker (2003):

Construct a parametric parabola such that

$$\sqrt{\sum_{k} (G_i(R_k) - f_i)^2} \le \varepsilon .$$

$$G_i(R) = \int_{V(y_i)} I(m) d\sigma(m) , \quad f_i = \int_{\Gamma_i} L(y)\, d\sigma(y)$$



(3)Methods based on solving Monge-Ampere equation

Benitez, etc. (2004) :

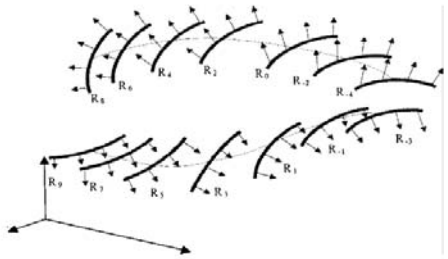


Fig. 3 SMS ribs generated from the seed rib R_0.

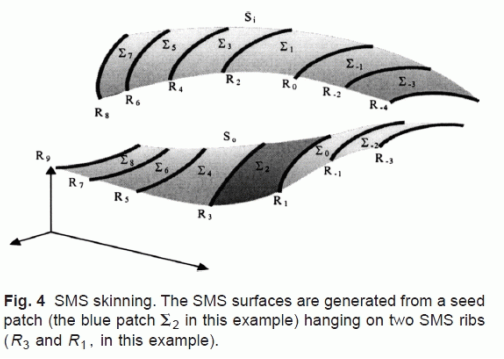
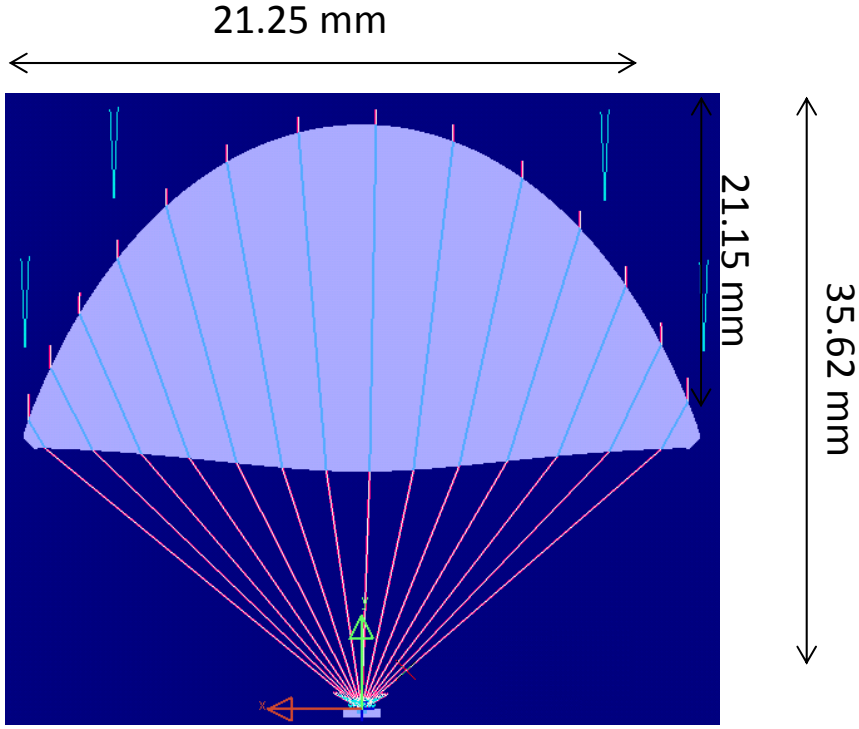


Fig. 4 SMS skinning. The SMS surfaces are generated from a seed patch (the blue patch Σ_2 in this example) hanging on two SMS ribs (R_3 and R_1, in this example).

(3)SMS method

From a point light source to a uniform line target.

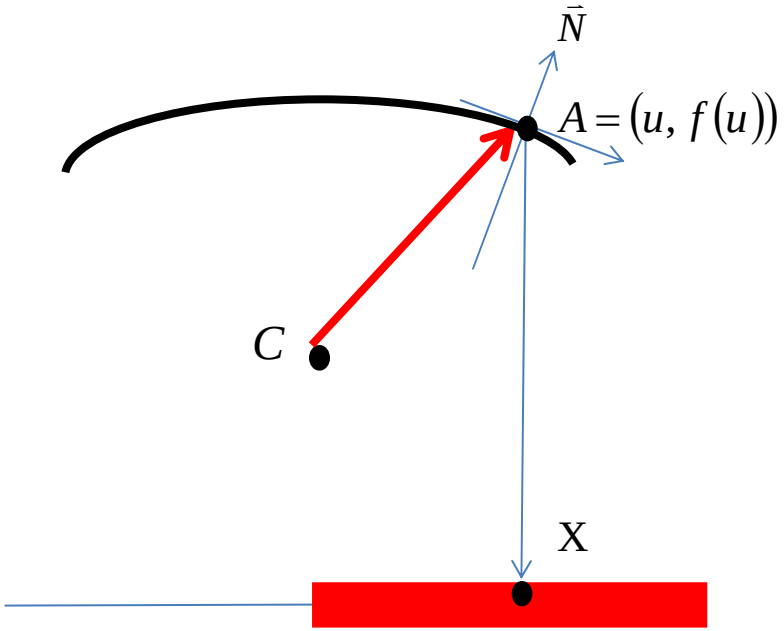
SMS2D lens:
LED length : 1.95 mm
divergence angle : ±2.5 deg



*P. Benítez , J. C. Miñano et. al., "Simultaneous multiple surface optical design method in three dimensions", *Opt. Eng*. **43** 1489-1502 (2004)

Modeling in global domain (I)

Reflection Law[1]:



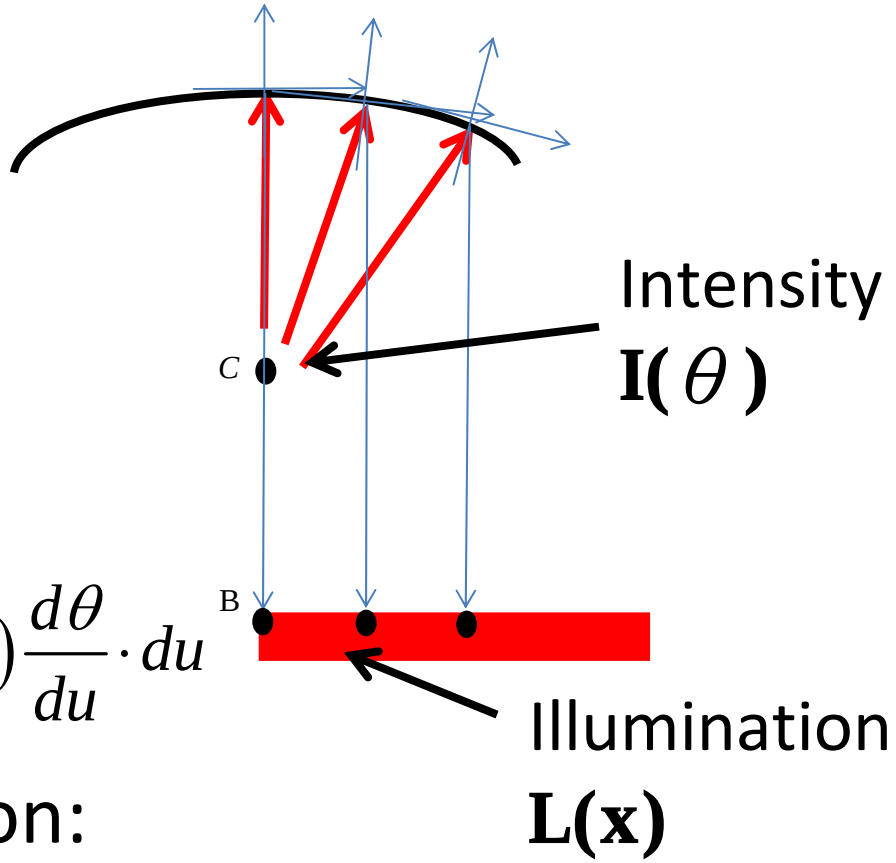
$$\frac{\overrightarrow{CA}}{|\overrightarrow{CA}|}+\frac{\overrightarrow{CA}-\overrightarrow{CX}}{|\overrightarrow{CA}-\overrightarrow{CX}|}=\frac{2\overrightarrow{N}(\overrightarrow{CA}\cdot\overrightarrow{N})}{|\overrightarrow{CA}||\overrightarrow{N}|^2}$$

[1]J. S. Scheruben, Formulation of a Reflector-Design Problem for a Lighting Fixture

Modeling in global domain (II)

- Global power conservation:

$$\int L(x)\,dx = \int I(\theta)\,d\theta$$



- Partition of Unity:

$$\int_{\Omega} \mathrm{L}(\mathrm{u})\frac{dx}{du}\cdot du = \int_{\Omega} I(u)\frac{d\theta}{du}\cdot du$$

$$\Rightarrow \sum_{i=1}^{n}\int_{\Omega_i} L(u)\frac{dx}{du}\cdot du = \sum_{i=1}^{n}\int_{\Omega_i} I(u)\frac{d\theta}{du}\cdot du$$

- Local power conservation:

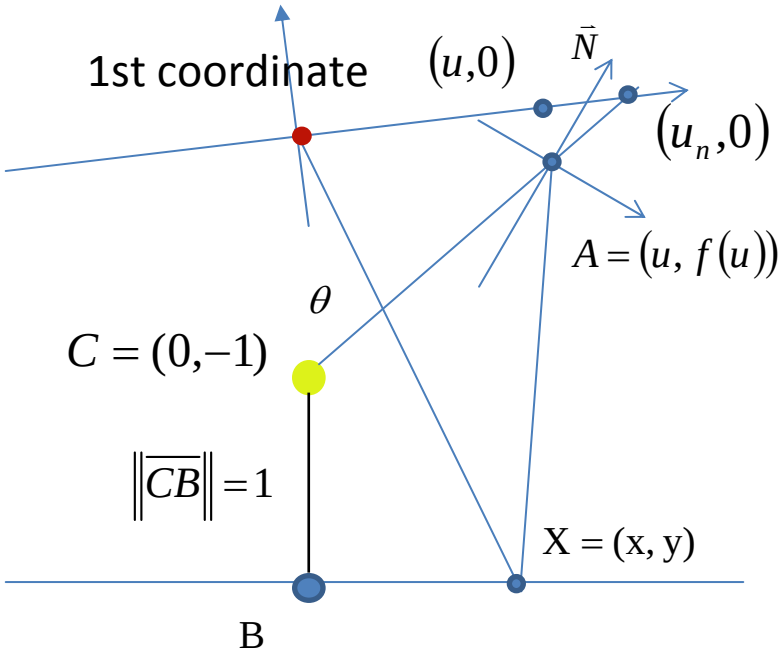
$$\int_{\Omega_i} L(u)\frac{dx}{du}\cdot du = \int_{\Omega_i} I(u)\frac{d\theta}{du}\cdot du$$

Modeling of local problem(I)

(i) Initialization: Set c=(0,-1) and define initial coordinate system according to the tangent plane of the reflector at starting point.

(ii) Normalization: set the distance between the light source C and the target plane equals to 1.

(iii) Represent the reflector surface in local coordinate and derive the Monge-Ampere equation from flat-form and local power conservation assumption.



Monge-Ampere equation in 1-D:

$$[L](f)=[A(u,f,f')]f''(u)+[B(u,f,f')]f'(u)+[C(u,f)]f(u)+[D(u)]$$

Modeling of Local Problem(II)

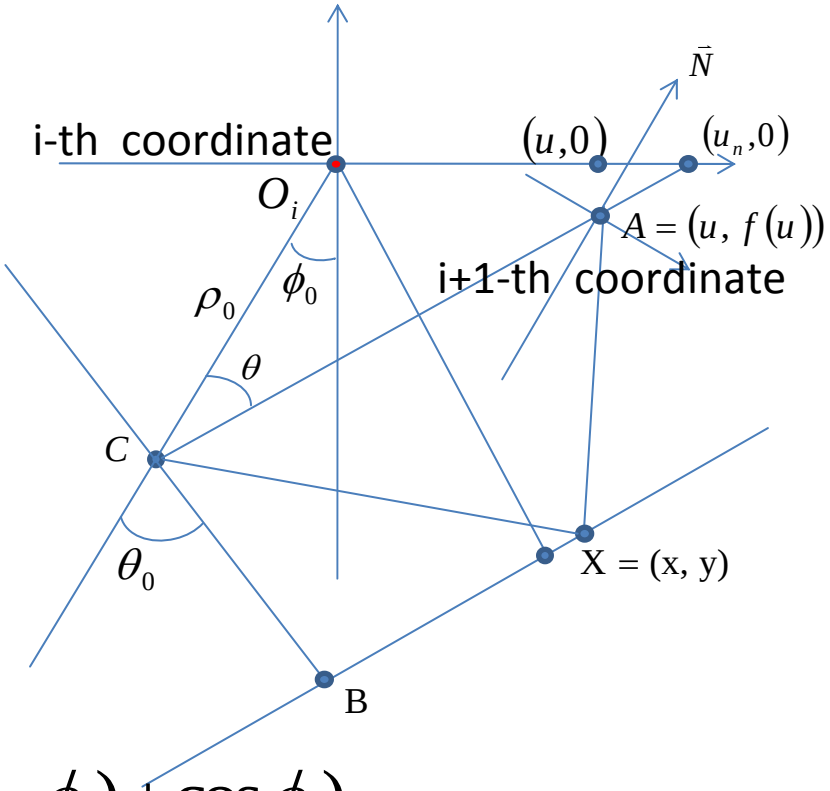
$$\left|\overrightarrow{CB}\right| = 1 \quad \left|\overrightarrow{OC}\right| = \rho_0$$

$$\vec{C} = \rho_0(-\sin\phi_0, -\cos\phi_0)_{i-th}$$

$$\overrightarrow{CB} = (\sin(\theta_0 - \phi_0), -\cos(\theta_0 - \phi_0))_{i-th}$$

$$\vec{OX} - \vec{CB} - \vec{OC}$$

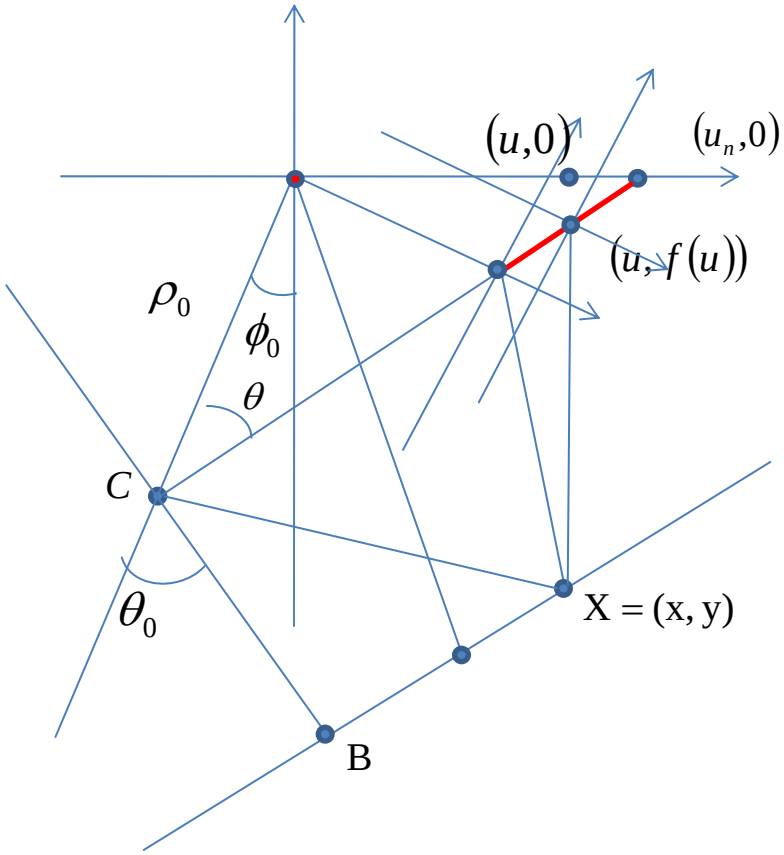
$$= (x - \sin(\theta_0 - \phi_0) + \sin\phi_0, y + \cos(\theta_0 - \phi_0) + \cos\phi_0)_{i-th}$$



Monge-Ampere equation in local coordinate: $[L]_{i+1} = [L]_i T_{i+1}^i,$

where T_{i+1}^i is the transformation matrix from (i+1)th coordinate to the ith coordinate.

Modeling of local problem (III)



(1) Boundary Condition: the red line is the possible region for f(u) based on the law of the reflection and convex assumption of the solution.

we choose the middle point to be the right boundary condition.

(2) We solve the local problem by finite element method (using Hermite element).

Numerical Tests

To verify our reconstruction algorithm:

1. We assume an ideal point light source and no energy loss on reflector surface.

2. The target illumination densities are obtained from ray tracing of reflectors of parabolic type and elliptic type.

3. We measure the errors between the given curves and the reconstructed curves.

Numerical tests

The curves to be reconstructed are as following:

- Parabola Equation:

$$(y - a + 1) = -\frac{1}{4a}x^2$$

  where a is the distance from top to focus

- Ellipse Equation:

$$\frac{x^2}{b^2} + \frac{(y+1.5)^2}{a^2} = 1,\ c^2 = a^2 - b^2$$

  where a is the major axis, b is the minor axis, and c is the distance from center to focus.

Error Measurement

Parameters for error measurement:

(1) J-th level of refinement: the number of sub-partition of each partition is 2^j

(2) Compress Ratio $|\Omega_\theta| / |\Omega_x|$:

the ratio between the range of the outgoing light angle and the range of the target illumination segment.

Error Measurement (Parabola)

(1) Surface error = e_s : The max-norm error between the given curves and reconstructed curves.

(2) Illumination error = e_{ill} : The relative error for each partition,

$Num\_Refl\_T(i) =$ number of reflected ray from given curves at ith partition

$Num\_Refl\_R(i) =$ number of reflected ray from reconstructed curves at ith partition

$$e_{ill}(i) = \frac{\left| Num\_Refl\_T(i) - Num\_Refl\_R(i) \right|}{Num\_Refl\_T(i)}$$

Error Measurement (Ellipse)

(1) Surface error = e_s : The max-norm error between the given curves and reconstructed curves.

(2) Illumination error = e_{ill} : The relative error for each partition,

$Num\_Refl\_T(i) =$ number of reflected ray from given curves at ith partition

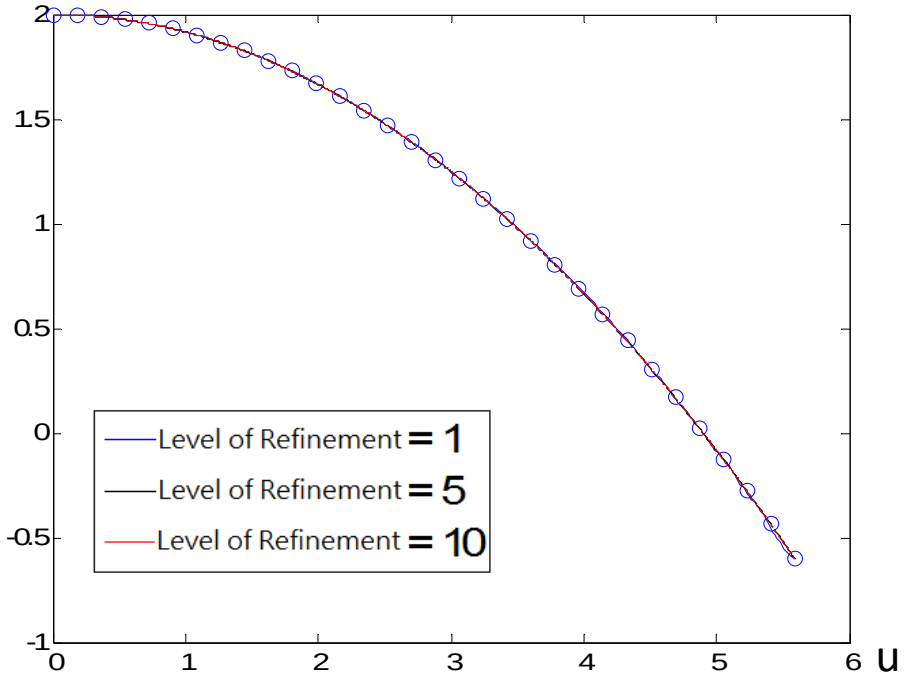
$Num\_Refl\_R(i) =$ number of reflected ray from reconstructed curves at ith partition

$$e_{ill} = \frac{\sum_i \left| Num\_Refl\_T(i) - Num\_Refl\_R(i) \right|}{\sum_i Num\_Refl\_T(i)}$$

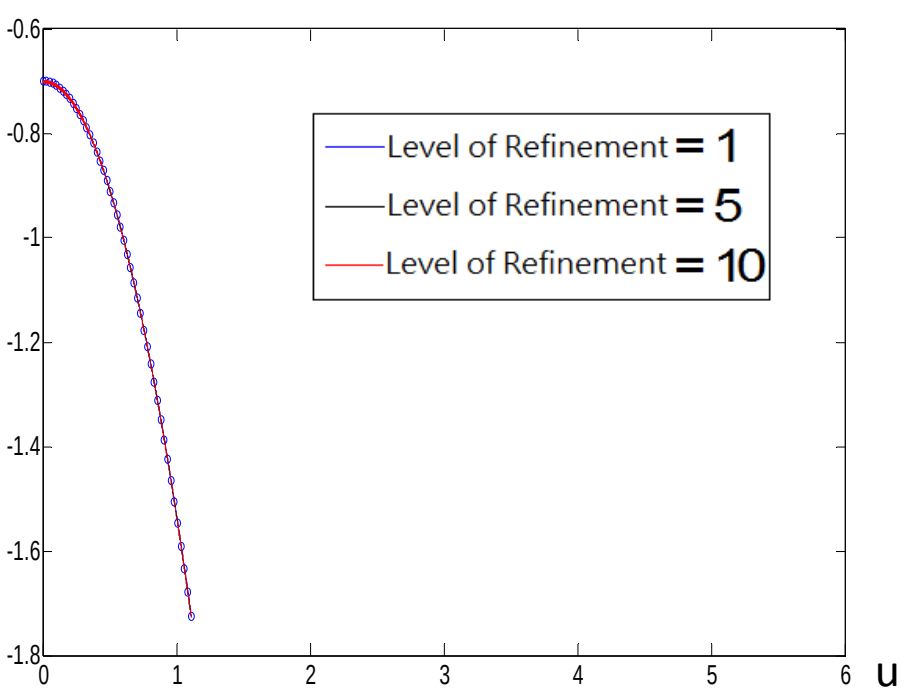
Numerical Results

Reconstructed parabolic curves:

case 1. a= 3:

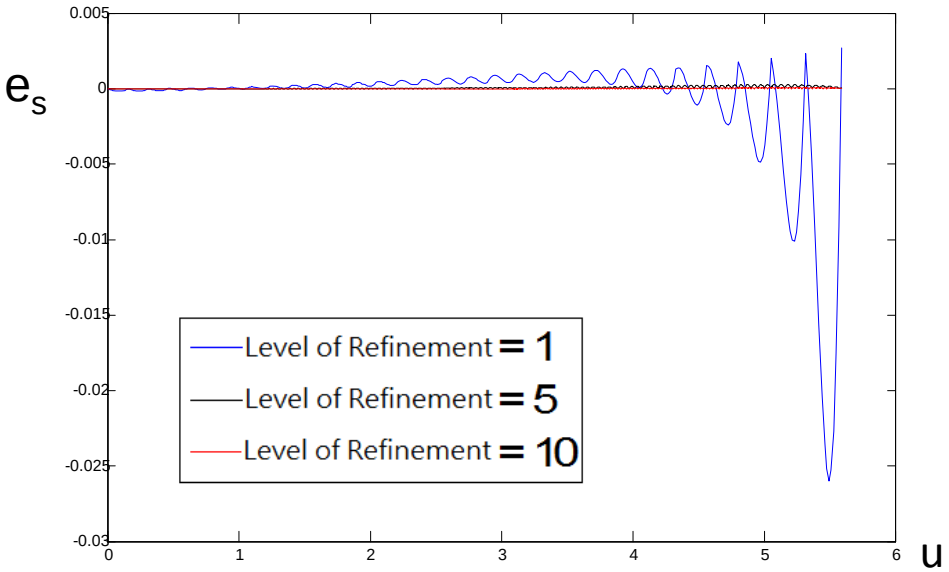


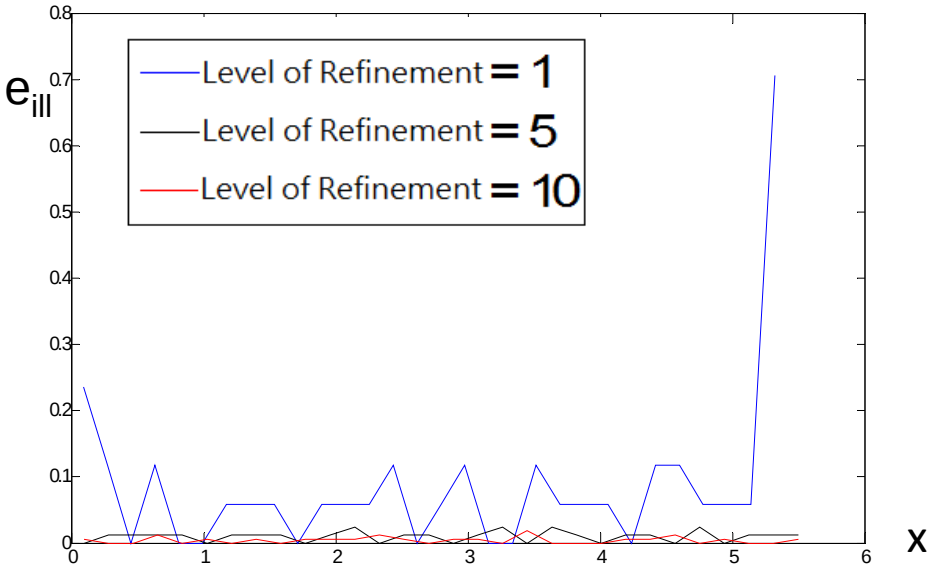
case 2. a=0.3:



Errors

Case (1)

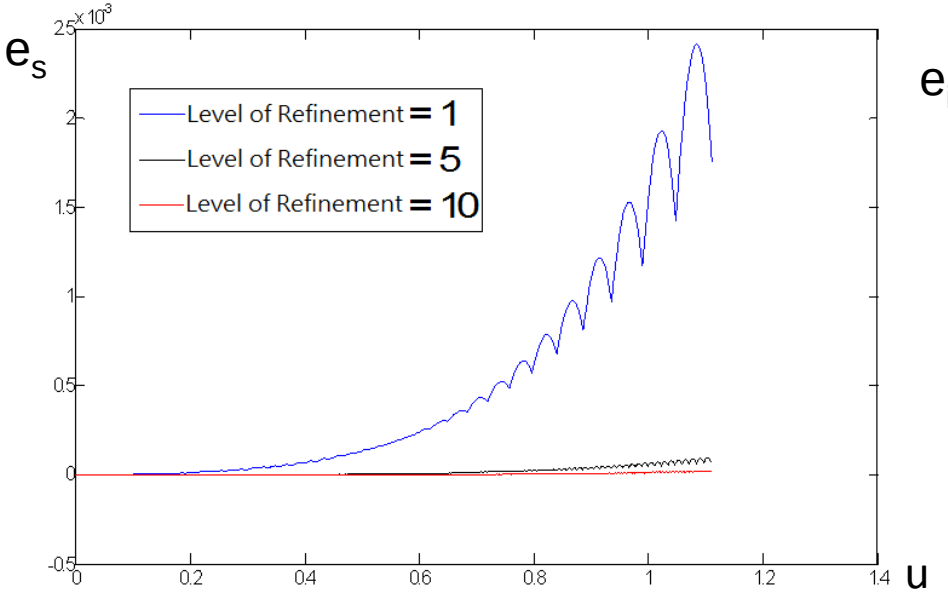


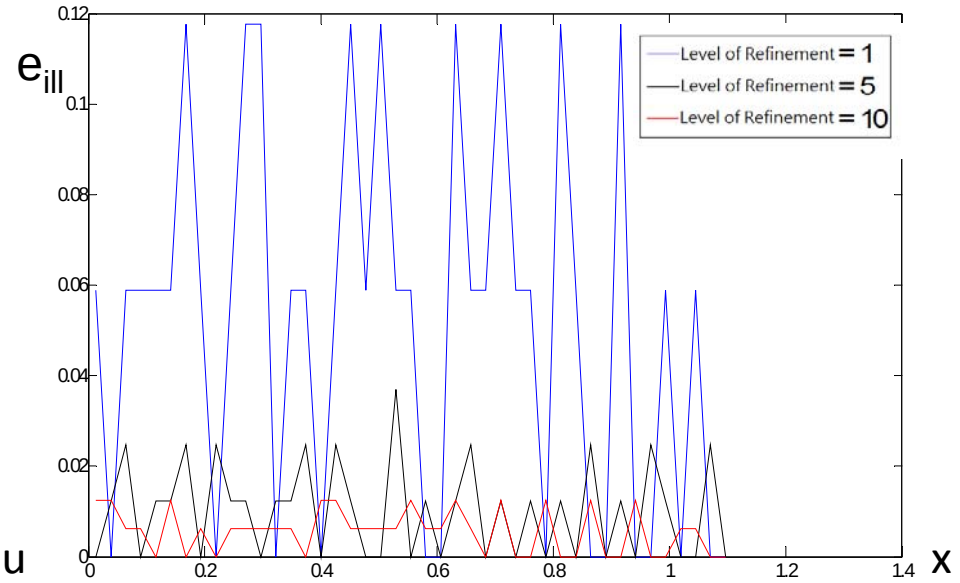


	Level of Refinement = 1	Level of Refinement = 5	Level of Refinement = 10	Level of Refinement = 50
Max Surface Error	0.026	2.7513*e-4	8.0356*e-5	6.8706*e-5
Avg. Illumination Error	0.0843	0.0103	0.0046	0.00149

Errors

Case (2)



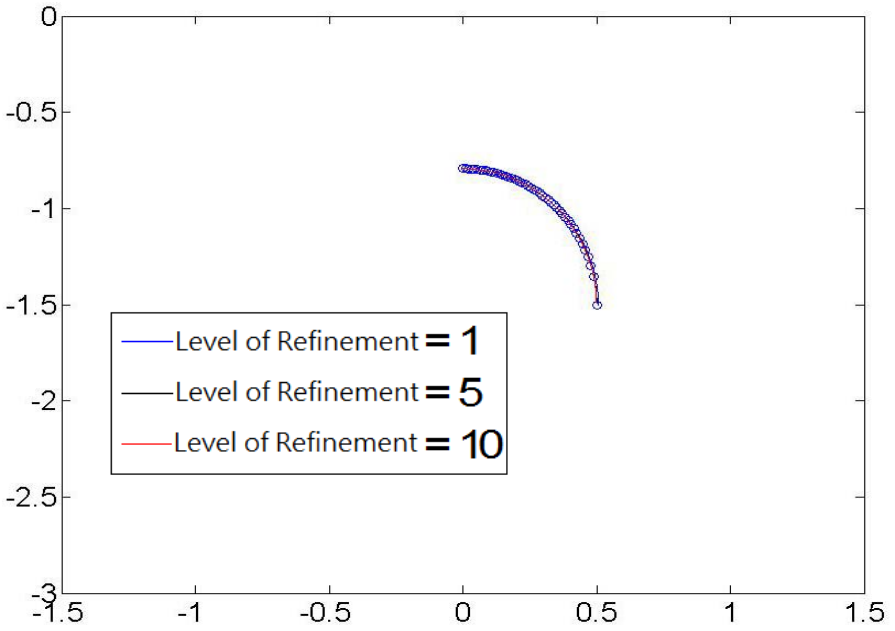


	Level of Refinement = 1	Level of Refinement = 5	Level of Refinement = 10	Level of Refinement = 50
Max Surface Error	0.0024	9.7213e-5	2.3461e-5	1.6704e-6
Avg. Illumination Error	0.0520	0.0103	0.0056	0.001538

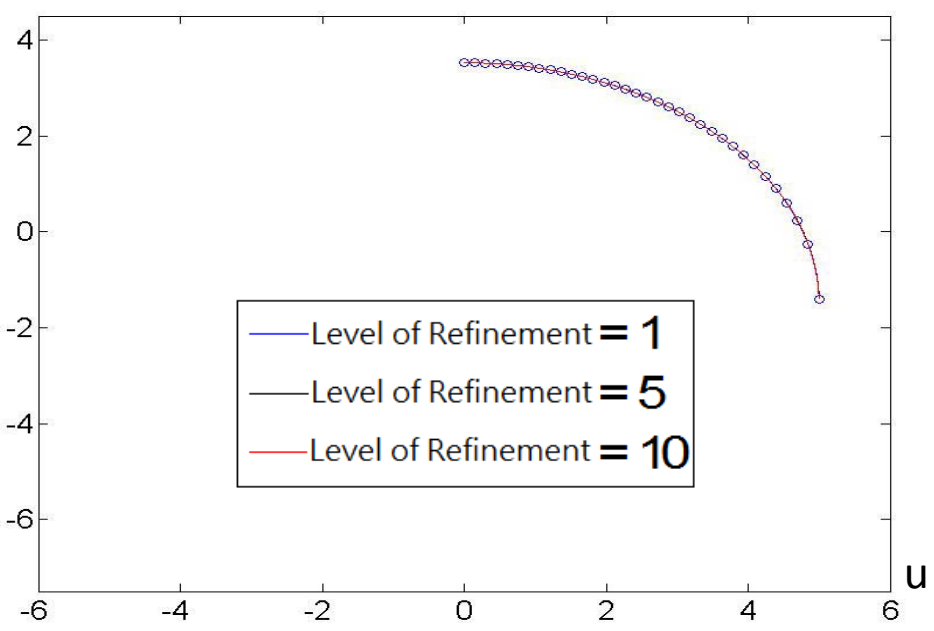
Numerical Results

Reconstructed elliptic curves:

case 1. c=0.5, b=0.5



case 2. c=0.5, b=5

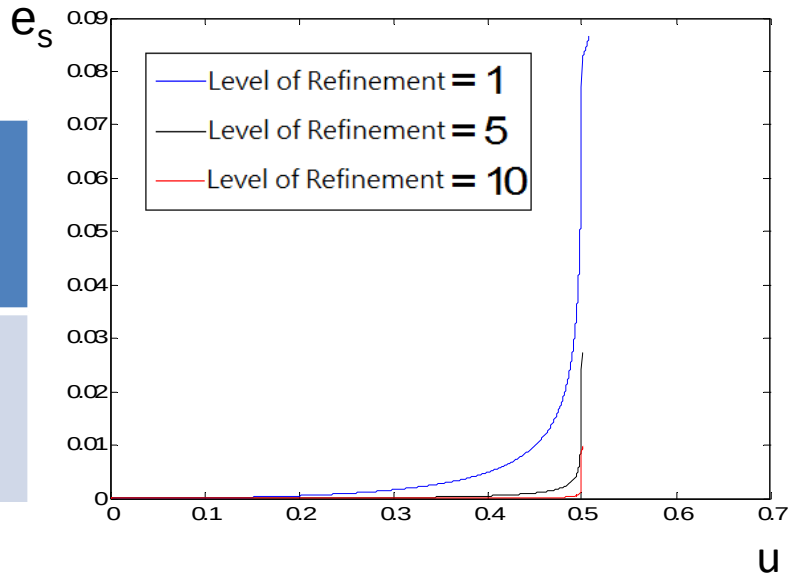


Errors

Case (1)

(i) Level of Refinement=5;

	Compress Ratio = 100	Compress Ratio = 1000	Compress Ratio = 10000
Max Surface Error	0.0868	0.0275	0.0097



(ii) Compress ratio = 1000;

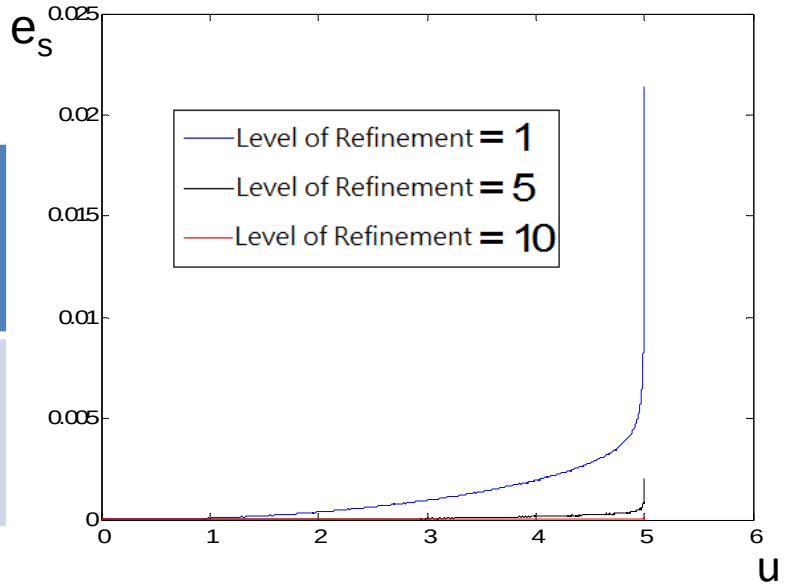
	Level of Refinement = 1	Level of Refinement = 5	Level of Refinement = 10	Level of Refinement = 30
e_s	0.8047	0.3155	0.1605	0.0729

Errors

Case (2)

(i) Level of Refinement=5;

	Compress Ratio = 100	Compress Ratio = 1000	Compress Ratio = 10000
Max Surface Error	0.0214	0.002	1.2685e-4



(ii) Compress ratio = 1000;

	Level of Refinement = 5	Level of Refinement = 10	Level of Refinement = 30	Level of Refinement =100	Level of Refinement =500	Level of Refinement = 1000
e_s	0.9368	0.9199	0.8171	0.5305	0.1707	0.1248

Conclusions

- Monge-Ampere can be solved accurately and efficiently by Fang-Neilan's approach using BCIZ element.
- Reconstruction algorithm based on flat form and local energy conservation assumptions can be employed to reconstruct 1D free-form effectively.
- Two dimensional free-form reconstruction using BCIZ element is under investigation.
- Convergence of the numerical solutions and error estimations are also under investigation.

Acknowledgement

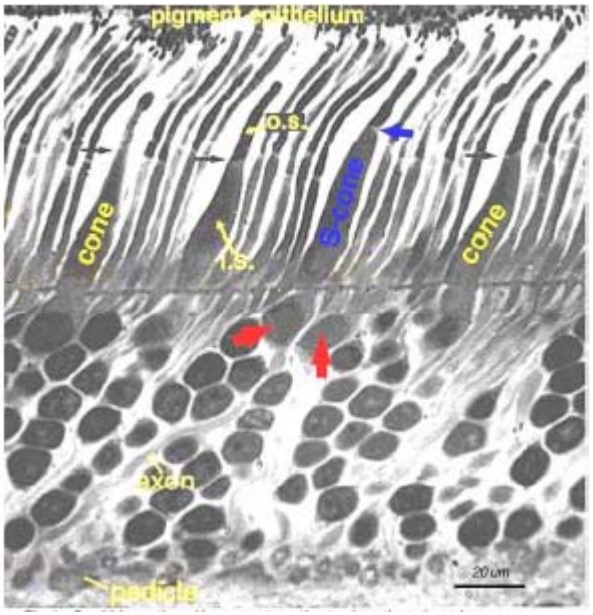
The authors would like to thank the support from NSC Taiwan under the grant 99-2115M-009-001.

Thanks for your attention.

How do the eyes see the color?

Cone cell: Cones are responsible for **color vision**. There are three types of cones, sensitive to red, green, and blue light respectively.

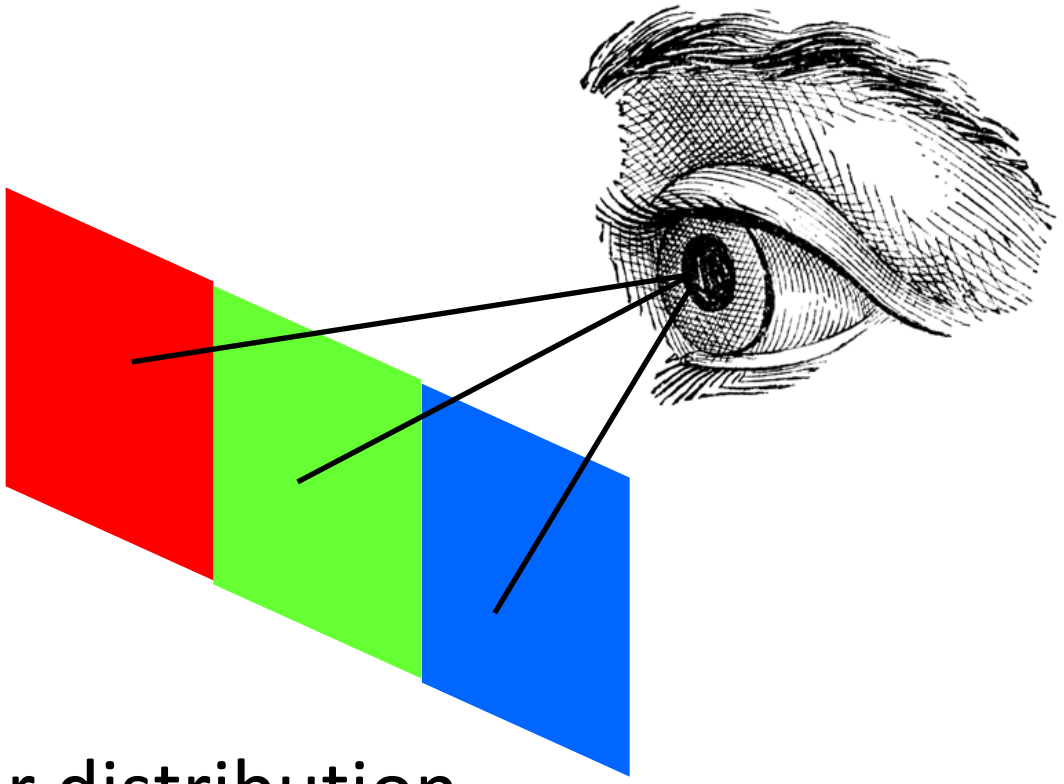
Rods: Rods cannot detect color but are responsible for black and white vision.



CIE 1931 color space

- Tristimulus Values:

$$X = \int_0^\infty I(\lambda)\overline{x}(\lambda)\,d\lambda$$
$$Y = \int_0^\infty I(\lambda)\overline{y}(\lambda)\,d\lambda$$
$$Z = \int_0^\infty I(\lambda)\overline{z}(\lambda)\,d\lambda$$



- $I(\lambda)$: spectral power distribution

CIE (1932). *Commission internationale de l'Eclairage proceedings, 1931*. Cambridge University Press, Cambridge.

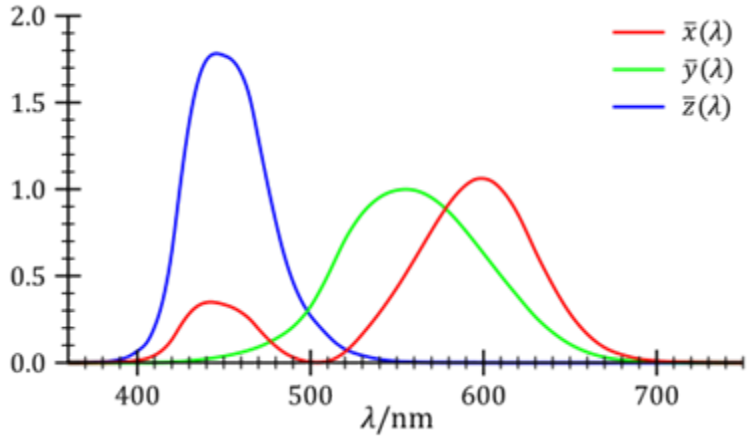
W. D. Wright, "A re-determination of the trichromatic coefficients of the spectral colours", *Transactions of the Optical Society* **30** (4) ,141–164 (1928)

J. Guild, "The colorimetric properties of the spectrum", Phil. Trans. R. Soc. Lond. A January 1, **230** 149-187 **30** (1932)

CIE 1931 color space

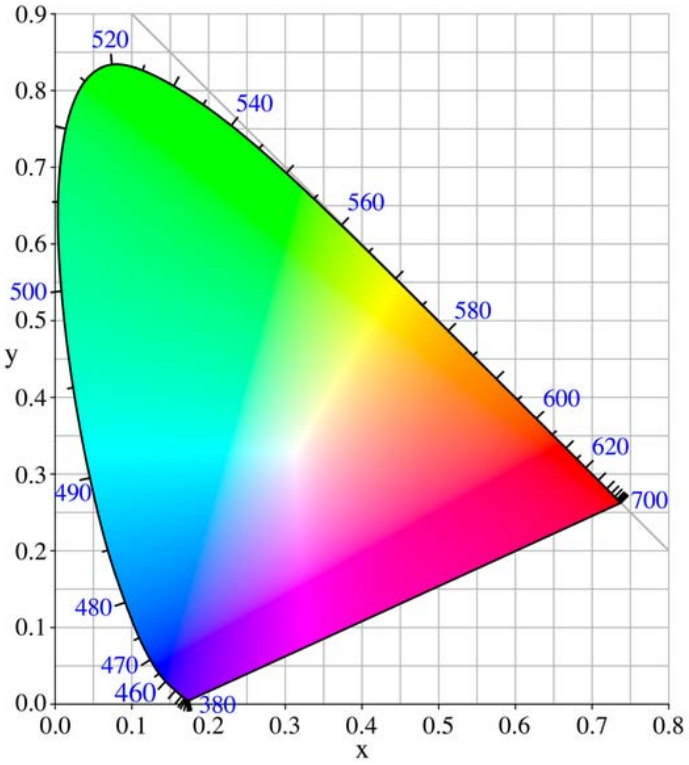
- Color Matching Functions:

$$\overline{x}(\lambda), \overline{y}(\lambda), \overline{z}(\lambda)$$



- Spectral sensitivity curves of three linear light detector

CIE 1931 color space



- CIE *xy* Chromaticity Diagram
  - Color space specified by x, y, and Y

$$x = \frac{X}{X + Y + Z}$$

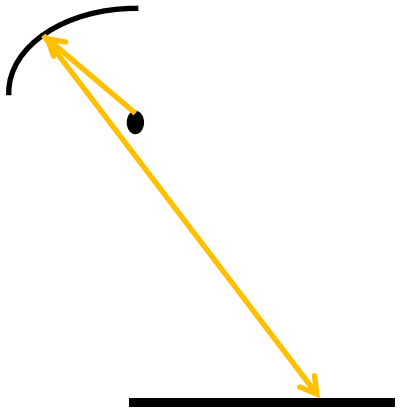
$$y = \frac{Y}{X + Y + Z}$$

  - Y is brightness or luminance of a color

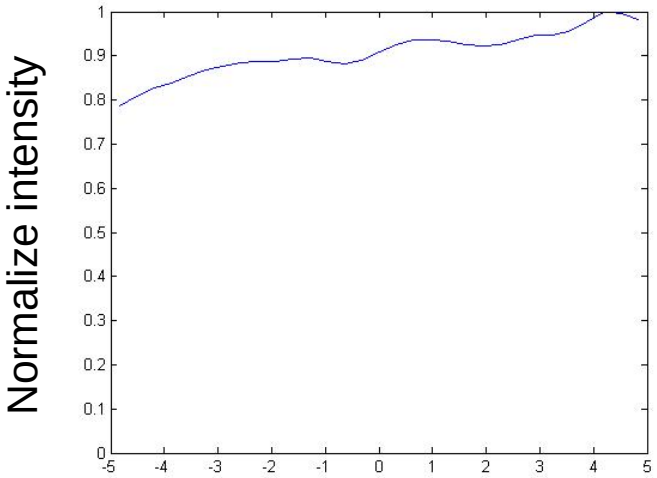
Intensity design

These are two per-defined distribution

Structure:

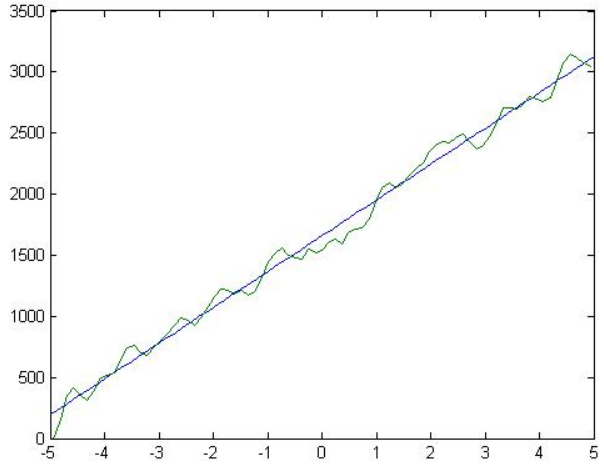


Uniform



Uniformity: 79%

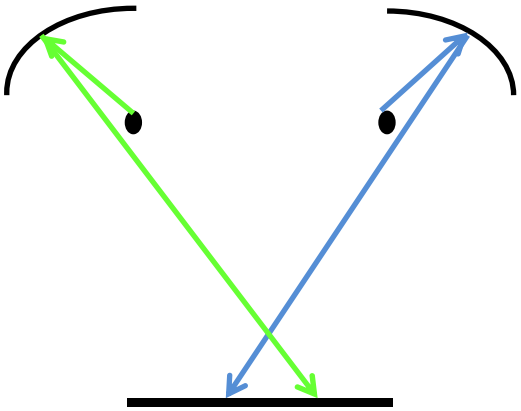
Linear



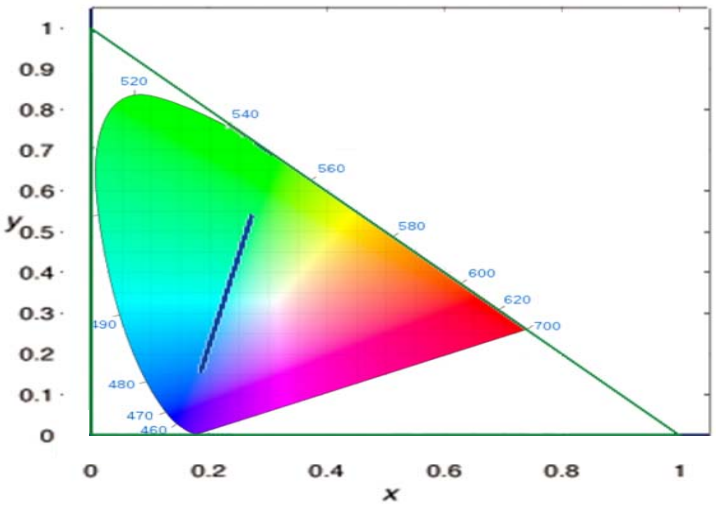
Color coded

The color of target plane is linear change from green to blue.

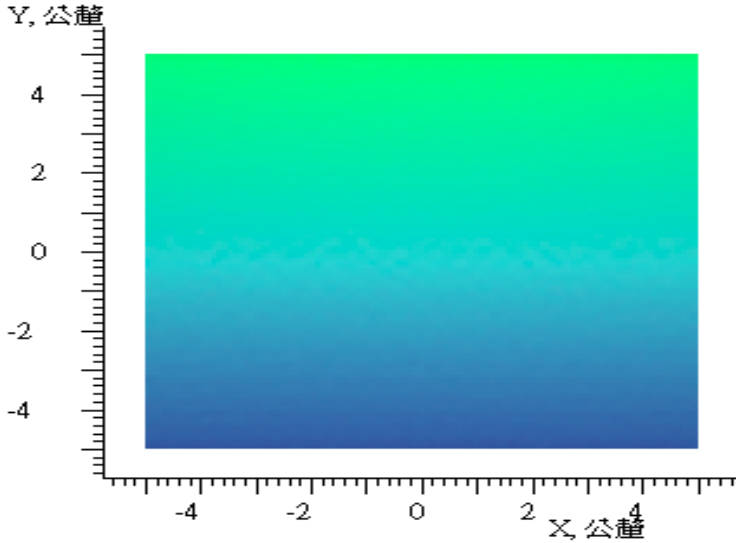
Structure:



Color chromaticity diagram :



Result:



國科會補助計畫衍生研發成果推廣資料表

日期:2011/10/24

國科會補助計畫	計畫名稱: 沉浸有限元素法於彈性界面與流構耦合問題的計算與誤差估計
	計畫主持人: 吳金典
	計畫編號: 99-2115-M-009-001- 學門領域: 數值分析與計算數學

無研發成果推廣資料

99 年度專題研究計畫研究成果彙整表

計畫主持人：吳金典	計畫編號：99-2115-M-009-001-
計畫名稱：沉浸有限元素法於彈性界面與流構耦合問題的計算與誤差估計	

成果項目			量化			單位	備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）
			實際已達成數（被接受或已發表）	預期總達成數(含實際已達成數)	本計畫實際貢獻百分比		
國內	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	0	0	100%		
		專書	0	0	100%		
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（本國籍）	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		
國外	論文著作	期刊論文	1	1	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	0	0	100%		
		專書	0	0	100%	章/本	
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（外國籍）	碩士生	2	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		

其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)	There are 2 more papers submitted to SCI journal. Please see the report.

	成果項目	量化	名稱或內容性質簡述
科教處計畫加填項目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與（閱聽）人數	0	

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

□達成目標

■未達成目標（請說明，以 100 字為限）

□實驗失敗

□因故實驗中斷

■其他原因

說明：

已完成二階橢圓方程沉浸有限元素法之誤差估計

已完成四階一維彈性桿的沉浸有限元素法程式

已完成非線性桿型變之程式測試

目前尚在研究流構界面之沉浸有限元之建造以確保介面條件能被滿足

2. 研究成果在學術期刊發表或申請專利等情形：

論文：■已發表 □未發表之文稿 □撰寫中 □無

專利：□已獲得 □申請中 ■無

技轉：□已技轉 □洽談中 ■無

其他：(以 100 字為限)

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）(以 500 字為限)

一成果所代表之意義； Immersed finite element method (IFEM) 對介面問題提供一個更一般的有限元素法的計算架構. In this proposal, we derive an effective a posteriori error indicator for the method. The accuracy of the numerical solutions obtained from IFEM can be improved through local mesh refinement by using our error indicator.