SLAC- PUB- 12137
September 2006

Solving Large Scale Nonlinear Eigenvalue Problems in Next-Generation Accelerator Design*

Ben-Shan Liao† Zhaojun Bai‡ Lie-Quan Lee§ Kwok Ko§

Abstract

A number of numerical methods, including inverse iteration, method of successive linear problem and nonlinear Arnoldi algorithm, are studied in this paper to solve a large scale nonlinear eigenvalue problem arising from finite element analysis of resonant frequencies and external Q_e values of a waveguide loaded cavity in the next-generation accelerator design. We present a nonlinear Rayleigh-Ritz iterative projection algorithm, NRRIT in short and demonstrate that it is the most promising approach for a model scale cavity design. The NRRIT algorithm is an extension of the nonlinear Arnoldi algorithm due to Voss. Computational challenges of solving such a nonlinear eigenvalue problem for a full scale cavity design are outlined.

Key words. nonlinear eigenvalue problem, Rayleigh-Ritz technique, waveguide loaded cavity, external Q_e value.

AMS subject classifications. 65F15, 65F50

1 Introduction

We consider a nonlinear eigenvalue problem (NEP) of the form

$$T(\lambda)x = 0, \tag{1}$$

where

$$T(\lambda) = K - \lambda M + \mathbf{i}\sum_{j=1}^{p} (\lambda - \sigma_j^2)^{\frac{1}{2}}\ W_j, \tag{2}$$

and K, M and W_j are $N \times N$ symmetric matrices. Moreover, K is semi-positive definite and M is positive definite. σ_j are given nonnegative scalars and $\mathbf{i} = \sqrt{-1}$. λ and x are the so-called *eigenvalues* and *eigenvectors*, respectively.

The nonlinear eigenvalue problem of the form (1) is derived from a Nedelec-type finite element discretization of the frequency domain Maxwell's equation with waveguide boundary

*version of August 19, 2006. The research of Zhaojun Bai and Ben-Shan Liao were supported in part by the National Science Foundation under grant 0220104.....

†Department of Mathematics, University of California, Davis, CA 95616. (liao@math.ucdavis.edu)

‡Department of Computer Science and Department of Mathematics, University of California, Davis, CA95616. (bai@cs.ucdavis.edu)

§Advanced Computations Department, Stanford Linear Accelerator Center, Menlo Park, CA 94025. (liequan@slac.stanford.edu, kwok@slac.stanford.edu).

Work supported in part by Department of Energy contract DE- AC02- 76SF00515

conditions in waveguide loaded cavity design [7, 11, 14, 13]. The matrices K, M and W_j are referred to as stiffness, mass and damping, respectively, in electromagnetic simulations. σ_j are cutoff wave numbers of modes in the waveguides, and are simply referred as cutoff values.

Let $\kappa = \sqrt{\lambda}$. The resonant frequencies f and corresponding external Q_e values of the cavity are defined by

$$f(\kappa) = \frac{c}{2\pi} \cdot \mathrm{Re}(\kappa) \tag{3}$$

and

$$Q_e(\kappa) = \frac{1}{2} \cdot \frac{\mathrm{Re}(\kappa)}{\mathrm{Im}(\kappa)}, \tag{4}$$

where c denotes the velocity of light in vacuum. The quantity Q_e measures the electromagnetic coupling between the cavity and waveguide. It characterizes the energy loss through the waveguide. With a given resonant frequency $f_0 > 0$, the cavity designers would like to seek physically meaningful resonant frequencies f, which are close to and larger than f_0 and $Q_e > 1$. Correspondingly, the desired eigenvalues λ of the NEP (1) are the ones such that

$$\begin{cases} \kappa = \sqrt{\lambda} \text{ is close to the shift value } \sigma_0 = \frac{2\pi}{c} f_0 \text{ and} \\ \lambda \in \mathcal{D} = \left\{ \lambda \;\mid\; \lambda = \kappa^2,\ \mathrm{Re}(\kappa) > \sigma_0,\ \mathrm{Im}(\kappa) > 0 \text{ and } Q_e(\kappa) > \widehat{Q}_e \right\}, \end{cases} \tag{5}$$

where $\widehat{Q}_e > 1$ is a prescribed value. The shaded domain in Figure 1 is an illustration of such a region in term of the square root $\lambda^{1/2}$.

The NEP (1) is of great interest in the next-generation accelerator design [7, 14, 13, 10, 24]. For the waveguide loaded cavity with one cutoff value, i.e., $p = 1$, it can be recast as a standard quadratic eigenvalue problem (see section 2). In this case, by exploiting established numerical techniques for solving the quadratic eigenvalue problems, we have been successfully solved the NEP (1) with the degrees of freedom up to 3.2 millions [14, 13]. In this paper, we consider the case with more than one cutoff value. In this case, there is no known transformation to convert the NEP (1) to a standard linear or polynomial eigenvalue problems. In this paper, we study numerical methods to solve the NEP (1) directly. We consider model test problems of two cutoff values ($p = 2$) with the degrees of freedom about 10,000. Finally we examine the feasibility of numerical solvers for the NEP (1) with a large number of cutoff values, say $p = 8$, and of the degrees of freedom up to millions.

Among existing nonlinear eigensolvers, the inverse iteration [19] is the simplest. It is derived based on the Newton's method to solve the nonlinear system of equations associate with the NEP (1). With a proper initial approximation, the method converges asymptotically quadratic. However, it requires to solve linear systems of equations of the same dimension at each iteration. Therefore, it is generally too expensive to use for finding multiple number of eigenvalues of large scale NEPs. The method of successive linear problems (MSLP) [19] and related self consistent iteration (SCI) [12] are popular techniques to solve the NEPs. The MSLP is derived based on the linearization of $T(\lambda)$. In the inner loop of the MSLP, it solves linear eigenproblems of the same degree of freedom. Therefore, it is also a computational expensive method.

The nonlinear Arnoldi algorithm [22] uses a subspace projection to reduce to a smaller NEP. It is designed for solving large scale NEPs. It can be characterized as a generalization of the popular Rayleigh-Ritz subspace projection technique for solving large scale linear eigenvalue problems, such as the Arnoldi and Lanczos methods, see [1] and references therein. The

nonlinear Arnoldi algorithm has successfully solved rational Hermitian eigenvalue problems, namely, the elements of $T(\lambda)$ are rational functions of λ and $T(\lambda)$ is Hermitian. In this case, the related Hermitian linear eigenvalue problem provides initial approximate eigenpairs and subsequently, one can use the underlying min-max property of rational Hermitian eigenvalue problems to approximate a set of desired eigenvalues in order [20, 22, 4, 8].

In this paper, we study the inverse iteration, the method of successive linear problems and the nonlinear Arnoldi algorithm to solve the NEP (1). All three methods require initial approximate eigenpairs with the proper ordering corresponding to the number of desired eigenpairs in a prescribed domain. This is one of fundamental challenges to solve nonlinear eigenvalue problems. The first contribution of this paper is to propose such a set of initial approximate eigenvalues and their ordering. In addition, we propose a variant of the nonlinear Arnoldi algorithm. We present a real-valued projection subspace to preserve the real symmetric property of the original NEP. We characterize the nonlinear Arnoldi algorithm and its variant under a general framework of nonlinear Rayleigh-Ritz technique. Numerical results demonstrate that the nonlinear Rayleigh-Ritz technique is the most promising approach to solve the large scale NEP (1). We outline the feature work for solving the the NEPs in a full scale realistic cavity design.

The rest of the paper is organized as follows. In section 2 we consider the initial approximate eigenpairs and their ordering. In section 3 we first review the inverse iteration, MSLP and SCI for finding a single eigenpair. Then we extend these methods for finding a set of eigenvalues. In section 4, we present a framework of nonlinear Rayleigh-Ritz projection techniques and review the nonlinear Arnoldi algorithm under this framework. In section 5, we present an implementation of the nonlinear Rayleigh-Ritz technique, which is an extension of the nonlinear Arnoldi algorithm to solve the NEP (1). Numerical experiments of these study are reported in section 6. Concluding remarks and the outline of our future work are in section 7.

2 Initial approximate eigenpairs and ordering

All iterative NEP solvers require initial approximate eigenpairs with the proper ordering. It is a critical factor deciding whether an iterative solver converges and the rate of convergence. By the form of the coefficient matrix $T(\lambda)$ of the NEP (1), a natural choice of the initial approximate eigenpairs is to use a selected set of eigenpairs of the symmetric positive definite pencil $K - \lambda M$. However, note that these eigenvalues θ_i are real. The external Q_e value does not defined in this case. Therefore, these real initial approximate eigenpairs are generally not efficient to approximate complex eigenpairs and do not possess physical meaning.

Alternatively, let us consider the first order truncation of the Taylor series expansion of $T(\lambda)$ about $\lambda = \lambda_0$:

$$T(\lambda) \approx T(\lambda_0) + (\lambda - \lambda_0)T'(\lambda_0) \equiv \widehat{K}(\lambda_0) - \lambda\widehat{M}(\lambda_0), \tag{6}$$

where

$$\widehat{K}(\lambda_0) = T(\lambda_0) - \lambda_0 T'(\lambda_0) \quad \text{and} \quad \widehat{M}(\lambda_0) = -T'(\lambda_0)$$

and

$$T'(\lambda_0) = -M + \frac{1}{2}\mathbf{i}\sum_{j=1}^{p}(\lambda_0 - \sigma_j^2)^{-\frac{1}{2}}W_j.$$

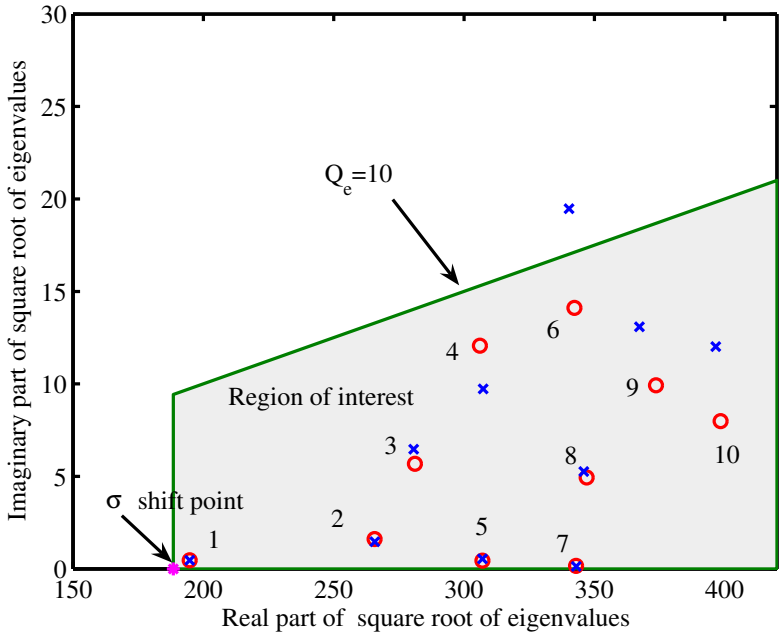

Figure 1: 10 eigenvalues of the linearized system $(\widehat{K}(\sigma_0^2), \widehat{M}(\sigma_0^2))$ (in circle) and 10 closet eigenvalues of $T(\lambda)$ (in cross) in the region of interest $\mathfrak{D}$ (shaded area).

It suggests that we can use selected eigenpairs (θ_ℓ, v_ℓ) of the matrix pair $(\widehat{K}(\sigma_0^2), \widehat{M}(\sigma_0^2))$ as the initial approximate eigenpairs. In practice, we can use Krylov subspace methods, such as the shift-and-invert Arnoldi method [21, 15], to find these desired eigenpairs (θ_ℓ, v_ℓ).

Assume that $\theta_\ell^{1/2}$ for $\ell = 1, 2, \ldots, n$ are the n eigenvalues of the matrix pair $(\widehat{K}(\sigma_0^2), \widehat{M}(\sigma_0^2))$ close to the shift value σ and θ_ℓ^2 are in the region $\mathcal{D}$ as described in (5). Then by the specification of the desired eigenvalues of the NEP (1) as described in section 1, we propose to order these initial approximate eigenvalues θ_ℓ according to the distance between $\theta_\ell^{1/2}$ and the shift point σ:

$$|\theta_1^{1/2} - \sigma_0| \le |\theta_2^{1/2} - \sigma_0| \le \cdots \le |\theta_n^{1/2} - \sigma_0|. \tag{7}$$

This ordering will be used in the following iterative methods to seek a set of desired eigenvalues of the NEP (1).

For example, consider the NEP (1) with a single waveguide cutoff value σ_1:

$$(K - \lambda M + \mathbf{i}(\lambda - \sigma_1^2)^{\frac{1}{2}}\, W_1)x = 0. \tag{8}$$

Let θ_i be the selected eigenvalues of the matrix pencil $\widehat{K}(\sigma_0^2) - \lambda\widehat{M}(\sigma_0^2)$, and be ordered according to (7). The numbered circles ("$\circ$") in Figure 1 show ten eigenvalues θ_ℓ of the matrix pencil $\widehat{K}(\sigma_0^2) - \lambda\widehat{M}(\sigma_0^2)$ close to the given shift σ_0 for a model NEP (8) of dimension 10124.

Note that the NEP (8) can be transformed to a quadratic eigenvalue problem:

$$\left[\nu^2 M - \nu\,(\mathbf{i}W_1) + (\sigma_1^2 M - K)\right] x = 0, \tag{9}$$

where $\nu = (\lambda - \sigma_1^2)^{1/2}$ [7, 14, 13]. We can extract few desired eigenvalues ν by a second-order Arnoldi algorithm [2]. The crosses ("$\times$") in Figure 1 show 10 eigenvalues λ_i of the model

NEP (8) computing by solving the quadratic eigenvalue problem (9). By Figure 1, we see that the eigenvalues θ_i of the matrix pencil $\widehat{K}(\sigma_0^2) - \lambda\widehat{M}(\sigma_0^2)$ are good approximations to the eigenvalues λ of the NEP (8). In section 6, we will study a model test NEP (1) with more than one waveguide cutoff value.

We note that the initial approximate eigenvectors discussed in this section can be used to define the initial projection subspace for a nonlinear Rayleigh-Ritz iterative projection method. We use such initial projection subspace in our implementation of the Rayleigh-Ritz technique discussed in section 5.

3 Classical nonlinear eigensolvers

In this section we review two classical nonlinear eigensolvers, namely the inverse iteration and the method of successive linear problems.

3.1 Inverse iteration

It is easy to see that the NEP (1) is equivalent to the nonlinear system of equations

$$\begin{bmatrix} T(\lambda)x \\ v^H x - 1 \end{bmatrix} = 0, \tag{10}$$

where v is a prescribed vector for the normalization of the eigenvector. an inverse iteration solver stems from applying Newton's method to solve the nonlinear system (10) [19, 23]. The following is an outline of the inverse iteration to find a single eigenpair (λ, x) of the NEP (1).

IIT: Inverse ITeration

1. Choose an initial approximate eigenpair $(\lambda^{(0)}, x^{(0)})$.
2. Determine a vector v such that $v^H x^{(0)} = 1$.
3. For $\ell = 0, 1, \cdots$, until convergence
   (a) solve the linear system $T(\lambda^{(\ell)})u = T'(\lambda^{(\ell)})x^{(\ell)}$ for u.
   (b) compute $w = v^H u$, $\lambda^{(\ell+1)} = \lambda^{(\ell)} - (v^H x^{(\ell)})/w$.
   (c) set $x^{(\ell+1)} = u/w$.

Note that at step 2, a common choice of the normalization vector v is $x^{(0)}/\|x^{(0)}\|_2^2$. The convergence of an approximate eigenpair $(\lambda^{(\ell)}, x^{(\ell)})$ is declared if the relative norm of the residual vector $r = T(\lambda^{(\ell)})x^{(\ell)}$ is smaller than a prescribed tolerance value ϵ, i.e., $\|r\|/\|x\| \leq \epsilon$. Under certain assumptions associated with the initial approximate eigenpair, the inverse iteration converges quadratically [19].

Note that the most expensive computational task of the inverse iteration is on solving the linear systems at step 3a. One can use sparse linear system solvers, such as UMFPACK [5] and SuperLU [6] for large-scale problems. Due to the cost of solving linear system of equations of the same dimension of the original NEP (1) in the inner loop, the inverse iteration is generally suitable for problems of moderate sizes. Variants of inverse iteration include residual inverse iteration and QR-type iteration [18, 23].

3.2 Method of successive linear problem and self consistent iteration

The method of successive linear problems (MSLP) is based on the first order truncation of the Taylor series expansion of the λ-matrix $T(\lambda)$ as given in (6). Assume that $\lambda^{(0)}$ is an initial approximate eigenvalue and $\lambda^{(\ell)}$ is the ℓ-th approximation, then the MSLP obtains the $(\ell+1)$-th approximation $\lambda^{(\ell+1)}$ by solving the linear eigenproblem

$$\widehat{K}(\lambda^{(\ell)})\, x^{(\ell+1)} = \lambda^{(\ell+1)}\, \widehat{M}(\lambda^{(\ell)})\, x^{(\ell+1)}, \tag{11}$$

where

$$\widehat{K}(\lambda) = K + E(\lambda) - \lambda E'(\lambda), \quad \widehat{M}(\lambda) = M - E'(\lambda), \quad E(\lambda) = \mathbf{i}\sum_{j=1}^{p} (\lambda - \sigma_j^2)^{\frac{1}{2}}\, W_j.$$

The $(\ell+1)$-th approximation $\lambda^{(\ell+1)}$ is the one closest to $\lambda^{(\ell)}$.

The following is an outline of the MSLP to find a single eigenpair of the NEP (1).

MSLP: Method of Successive Linear Problems

1. Choose an initial approximate eigenpair $(\lambda^{(0)}, x^{(0)})$.
2. For $\ell = 0, 1, 2, \cdots$, until convergence
   (a) solve the linear eigenproblem (11) for $(\lambda^{(\ell+1)}, x^{(\ell+1)})$.

At step 2a, a shift-and-invert Arnoldi algorithm [21, 15] can be applied to find the desired eigenpair $(\lambda^{(\ell+1)}, x^{(\ell+1)})$ of the linear eigenproblem (11) of the same dimension of the original NEP (1). As the inverse iteration, the MSLP is generally computational expensive and is suitable for problems of moderate sizes.

By ignoring the derivative term $E'(\lambda)$ in the linear eigenproblem (11), it leads to the eigenvalue problem

$$(K + E(\lambda^{(\ell)}))\, x^{(\ell+1)} = \lambda^{(\ell+1)} M\, x^{(\ell+1)}. \tag{12}$$

This is the so-called self consistent iteration (SCI) [14, 13]. One can expect the convergence behaviors of the MSLP and of the SCI are similar if the one matrix pencil is a small perturbation of the another one. In this paper, we will use MSLP in our numerical experiments since MSLP and SCI have almost the same computational complexity for the NEP (1), while the MSLP algorithm converges quadratically under certain conditions [19]. Note that SCI can be applied to more general cases such as eigenvalue problems with nonlinearity in eigenvector, i.e., the nonlinear eigenvalue problem of the form $H(x)x = \lambda x$ [12, 9].

3.3 Inverse iteration or MSLP to compute k eigenvalues

Using the initial approximate eigenpairs and the ordering as discussed in section 2, we are ready to use the inverse iteration or MSLP to compute a set of eigenpairs of the NEP (1). The following is a pseudo-code to find k approximate eigenpairs close to the given shift σ_0 with a prescribed accuracy ϵ measured by the relative residual norms.

IIT(k)/MSLP(k): Inverse iteration/MSLP to compute k eigenpairs

1. Compute n selected eigenpairs $\{(\theta_i, v_i)\}_{i=1}^{n}$ of the matrix pair $(\widehat{K}(\sigma_0^2), \widehat{M}(\sigma_0^2))$ as described in (5). Assume $n \geq k$.

2. Order $\{\theta_i\}_{i=1}^n$ according to (7)
3. Set $\ell = 1$ and the number of convergent eigenvalues $m = 0$.
4. Iterate while $m \leq k$ and $\ell \leq n$
   (a) start with the ℓ-th initial approximate pair (θ_ℓ, v_ℓ),
   (b) compute an approximate eigenpair (λ, x) of $T(\lambda)x = 0$ by IIT/MSLP,
      1) if $\|T(\lambda)x\|_2/\|x\|_2 < \epsilon$, then accept (λ, x) as m-th convergent eigenpair, set $m = m + 1$, and goto (4c),
      2) if the iteration number of IIT/MSLP exceeds the prescribed limit, declare failure to find the eigenpair associated with the initial approximate eigenpair (θ_ℓ, v_ℓ),
   (c) set $\ell = \ell + 1$,

Numerical results of the IIT(k) and MSLP(k) to find eigenpairs of the NEP (1) will be reported in section 6.

4 Nonlinear Rayleigh-Ritz technique

In the inner loop of the inverse iteration and the method of successive linear problems, it requires to solve linearized problems of the same dimension of the original NEP (1). It becomes very expensive for large scale problems. It is well-known that for large scale linear eigenproblems, Rayleigh-Ritz technique, which includes the well-known Arnoldi and Lanczos methods, is an effective approach through the reduction of linear eigenproblems to much smaller dimension, see [1] and reference therein.

Recently, the Rayleigh-Ritz technique has been generalized to solve large scale nonlinear eigenvalue problems under the names of nonlinear Arnoldi algorithm [22], nonlinear Jacobi-Davidson method [4], and rational Krylov method [20, 8]. All these methods can be summarized by the following general framework.

Nonlinear Rayleigh-Ritz technique

1. Select a proper projection subspace $\mathcal{V}$.
2. Compute a pair (θ, z) to satisfy the Galerkin condition:

$$z \in \mathcal{V} \text{ such that } T(\theta)z \perp \mathcal{V}. \tag{13}$$

3. Expand or restart the projection subspace $\mathcal{V}$ when necessary.

Note that at step 2, since $z \in \mathcal{V}$, $z = Vg$ for some n-vector g, where n is the dimension of the subspace $\mathcal{V}$, and V is an orthonormal basis. Hence, step 2 is equivalent to solve a reduced nonlinear eigenproblem of determining a pair (θ, g) satisfying

$$T_V(\theta)g = 0, \tag{14}$$

where $T_V(\theta) = V^H T(\theta)V$ is an $n \times n$ matrix. The value θ computed from this technique are referred to as *Ritz values* and the vectors z are the associated *Ritz vectors*. The accuracy of a Ritz pair (θ, z) as an approximate eigenpair is assessed through the residual vector $r = T(\theta)z$.

With a proper initial projection subspace $\mathcal{V}$ and their expansion, it is typical that satisfactory approximations of few eigenpairs of the $N \times N$ matrix $T(\lambda)$ occur for a small dimension of projection subspace $\mathcal{V}$, i.e., $n \ll N$. The reduced nonlinear eigenproblem (14) can be solved by using the inverse iteration or MSLP as described in section 3.

In a practical implementation of the Rayleigh-Ritz technique, two critical issues are

(a) what is a proper initial projection subspace $\mathcal{V}$ and how to order initial approximate eigenvalues?

(b) how to expand or restart the projection subspace $\mathcal{V}$ when necessary?

In the nonlinear Arnoldi algorithm [22], it is suggested to expand the subspace $\mathcal{V}$ by the residual inverse iteration. Such a subspace expansion strategy can also be explained by the rational Krylov method [20, 8]. Alternatively, the nonlinear Jacobi-Davidson method expands the projection subspace by solving the corresponding correction equations with proper preconditioning [4]. The performance tests of these algorithms for solving rational eigenvalue problems are reported in the recent survey papers [23, 17]. It concludes that the nonlinear Arnoldi method is more efficient than nonlinear Jacobi-Davidson and rational Krylov algorithms.

5 A nonlinear iterative projection algorithm

In this section, we present an implementation of the Rayleigh-Ritz technique to solve of the NEP (1). Our implementation starts with the nonlinear Arnoldi algorithm due to Voss [22].

As we have discussed in section 2, to find a set of desired eigenpairs of a NEP, the foremost issue is to decide a proper initial projection subspace $\mathcal{V}$ and an ordering of initial approximate eigenpairs. For the specific NEP (1), we will use the initial projection subspace $\mathcal{V}$ spanned by the eigenvectors $\{v_\ell\}$ of the matrix pencil $\widehat{K}(\sigma_0^2) - \lambda\widehat{M}(\sigma_0^2)$ corresponding to the eigenvalues $\{\theta_\ell\}$ as described in (5), and ordered as in (7).

Note that the eigenvectors $\{v_\ell\}$ are generally complex. Therefore, if we directly uses them in the Rayleigh-Ritz projection, the coefficient matrices of the resulting nonlinear eigenvalue problem will be complex. To preserve the real symmetric property of the coefficient matrices K, M and W_j in the original NEP (1), we propose to use a real orthonormal basis Q defined by

$$Q = \mathrm{orth}(\begin{bmatrix} \mathrm{Re}(V) & \mathrm{Im}(V) \end{bmatrix}),$$

where $\mathrm{orth}(X)$ stands for an orthonormal basis for the range of X. Subsequently, we solve a reduced nonlinear eigenvalue problem

$$T_Q(\theta)y = 0, \tag{15}$$

where

$$\begin{aligned} T_Q(\theta) &= Q^T T(\theta) Q \\ &= K_Q - \theta M_Q + \mathrm{i}\sum_{j=1}^{p} (\theta - \sigma_j^2)^{\frac{1}{2}} W_{Q,j} \end{aligned}$$

with real symmetric coefficient matrices $K_Q = Q^H K Q$, $M_Q = Q^T M Q$ and $W_{Q,j} = Q^T W_j Q$. The real symmetric property is preserved. Since the subspace spanned by the real basis vectors

Q contains the subspace spanned by the complex basis vectors V, the quality of the real reduced NEP will be at least as good as the case using complex projection.

Starting with ℓ-th initial approximate eigenpair (θ_ℓ, v_ℓ), we use the inverse iteration to find an approximate eigenpair (θ, y) of the reduced NEP (15). Subsequently, an approximate eigenpair, called Ritz pair, of the original NEP (1) is given by (θ, z), where $z = Qy$.

The accuracy of the Ritz pair (θ, z) is assessed through the norm of relative residual $\|T(\theta)z\|_2/\|z\|_2$. If the Ritz pair is acceptable, then the projection subspace is updated by replacing the initial eigenvector v_ℓ with the Ritz vector z. Otherwise, The projection subspace spanned by Q is expanded in the new direction

$$v = T^{-1}(\sigma_0^2)T(\theta)z.$$

This is essentially one step of residual inverse iteration [18, 22]. To maintain the orthogonality of Q, v is orthogonalized against the previous basis vectors Q

$$v := v - QQ^T v.$$

The projection subspace spanned by Q is expanded as the following:

$$Q := \begin{bmatrix} Q & \widetilde{Q} \end{bmatrix}$$

where $\widetilde{Q} = \text{orth}(\begin{bmatrix} \text{Re}(v) & \text{Im}(v) \end{bmatrix})$. Correspondingly, the coefficient matrices K_Q, M_Q and $W_{Q,j}$ are updated to define an updated reduced NEP $T_Q(\lambda)$, namely

$$K_Q := \begin{bmatrix} K_Q & Q^T(K\widetilde{Q}) \\ (\widetilde{Q}^T K)Q & \widetilde{Q}^T(K\widetilde{Q}) \end{bmatrix}, \quad M_Q := \begin{bmatrix} M_Q & Q^T(M\widetilde{Q}) \\ (\widetilde{Q}^T M)Q & \widetilde{Q}^T(M\widetilde{Q}) \end{bmatrix},$$

and

$$W_{Q,j} := \begin{bmatrix} W_{Q,j} & Q^T(W_j\widetilde{Q}) \\ (\widetilde{Q}^T W_j)Q & \widetilde{Q}^T(W_j\widetilde{Q}) \end{bmatrix},$$

for $j = 1, 2, \ldots, p$.

The procedure is then iterated with the initial approximate eigenpairs (θ_ℓ, v_ℓ) and updated the reduced NEP defined by $T_Q(\lambda)$.

Note that an iteration with an initial approximate eigenpair (θ_ℓ, v_ℓ) may fail to converge to an eigenpair of the original NEP. This is detected by the number of the iterations *iters* exceeding a prescribed maximal iteration number $i_{\max}$. If it occurs, we purge all those expanded vectors in Q associated the initial approximation, and furthermore, remove the associated starting vector v_ℓ from the starting matrix V.

To control the memory requirement requriement and computational cost, we can restart the projection subspace when the number of basis vectors Q larger than a prescribed maximal dimension. This is done after an eigenpair converged. The restarting projection subspace is spanned by $Q = \text{orth}(\begin{bmatrix} \text{Re}(V) & \text{Im}(V) \end{bmatrix})$, where V is updated with the converged Ritz vectors z and is reduced by purging those vectors v_ℓ, which lead to the failure of the convergence.

In summary, we have the following pseudocode for a nonlinear Rayleigh-Ritz iterative method to find eigenvalues of the NEP (1), where k is the number of desired eigenvalues, n is the number of eigenvectors used in initial projection subspace with $n \geq k$, $|Q|$ is the number of columns of the basis vectors of the projection subspace, $n_{\max}$ is the maximal size of $|Q|$ allowed. m is the number of the converged eigenpairs, *iters* is the number of iterations with respect to converge an eigenpair, and $i_{\max}$ is the maximal number of iterations allowed and ϵ is the prescribed accuracy for computed eigenpairs.

NRRIT: Nonlinear Rayleigh-Ritz ITerative method

1. Compute n selected eigenpairs $\{(\theta_j, v_j)\}_{j=1}^n$ of the matrix pair $(\widehat{K}(\sigma_0^2), \widehat{M}(\sigma_0^2))$ as described in (5).
2. Order $\{\theta_j\}_{j=1}^n$ according to (7).
3. Set $Q = \text{orth}(\begin{bmatrix} \text{Re}(V) & \text{Im}(V) \end{bmatrix})$ where $V = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$.
4. Set $\ell = 1$, $iters = 1$ and $m = 0$.
5. Iterate while $m \le k$ and $\ell \le n$
   (a) start with the ℓ-th initial approximate eigenpair (θ_ℓ, v_ℓ),
   (b) compute an eigenpair (θ, y) of the projected NEP $T_Q(\theta)y = 0$ by using the IIT with the starting pair $(\theta_\ell, Q^T v_\ell)$,
   (c) compute Ritz vector $z = Qy$ and residual $r = T(\theta)z$,
   (d) if $\|r\|_2/\|z\|_2 < \epsilon$, then
      i. accept the Ritz pair (θ, z) as an approximate eigenpair,
      ii. update V by setting $v_\ell = z$,
      iii. if $|Q| > n_{\max}$, then restart: $Q = \text{orth}(\begin{bmatrix} \text{Re}(V) & \text{Im}(V) \end{bmatrix})$, and update K_Q, M_Q and $W_{Q,j}$,
      iv. record the number of iteration $iters$; set $iters = 1$,
      v. set $\ell = \ell + 1$ and $m = m + 1$; goto step 5 and search for next eigenpair,
   (e) solve $T(\sigma_0^2)v = r$ for v,
   (f) orthogonalize v against Q: $v = v - QQ^T v$,
   (g) compute $\widetilde{Q} = \text{orth}(\begin{bmatrix} \text{Re}(v) & \text{Im}(v) \end{bmatrix})$,
   (h) expand $Q := \begin{bmatrix} Q & \widetilde{Q} \end{bmatrix}$, and update K_Q, M_Q and $W_{Q,j}$,
   (i) if $iters > i_{\max}$, then
      i. flag failure with the initial approximate eigenpair (θ_ℓ, v_ℓ),
      ii. purge all expanded vectors from Q generated in this loop,
      iii. update V by purging v_ℓ from V,
      iv. set $\ell = \ell + 1$, $iters = 0$ and search for next eigenpair,
   (j) set $iters = iters + 1$,

We note that at step 5e, it is required to solve a linear system $T(\sigma_0^2)v = r$ of the same dimension of the original NEP. This is computational most expansive step of the algorithm. However, note that the coefficient matrix $T(\sigma_0^2)$ of the linear system remains unchanged in the loop. Therefore, the algorithm is generally more efficient than other nonlinear RR iterative methods such as the one based on the Jacobi-Davidson correction equation [4], where the correction equation is modified at every iteration in the loop.

6 Numerical experiments

In this section, we present numerical results of solving the NEP (1) with the inverse iteration (IIT), the method of successive linear problem (MSLP) and the nonlinear Rayleigh-Ritz iterative (NRRIT) method. We highlight the issues associated the initial approximate eigenpairs

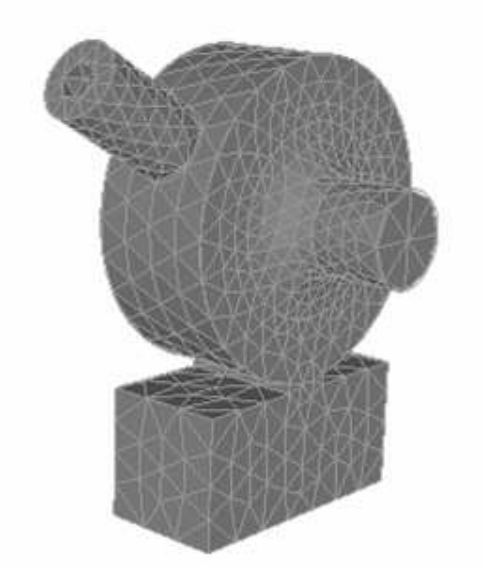

Figure 2: A simulation model of waveguide loaded cavity with two cutoff guides. The cavity part is the cylindrical part. The two waveguides are coaxial waveguide at the top left and the rectangular waveguide at the bottom.

and performance comparison. The main feature of the NRRIT is also highlighted. We consider the NEP (1) with two cutoff values, i.e.,

$$T(\lambda)x = \left[K - \lambda M + \mathbf{i}(\lambda - \sigma_1^2)^{\frac{1}{2}} W_1 + \mathbf{i}(\lambda - \sigma_2^2)^{\frac{1}{2}} W_2\right] x = 0, \tag{16}$$

where the coefficient matrices K, M, W_1 and W_2 are of order $N = 9956$. The cutoff values $\sigma_1 = 0$ and $\sigma_2 = 0.043551$. The computational task is to compute 10 resonant frequencies f, which are close to and larger than the prescribed frequency $f_0 = 2.7693$MHz and $Q_e > 10$. It means that we need to seek $k = 10$ eigenvalues λ of the NEP (16) in the region of interest as described in (5), with $\sigma_0 := 2\pi f_0/c = 0.058$ and $\widehat{Q}_e = 10$. The underlying physical model is a radio-frequency gun model as shown in Figure 2. The cavity itself is the big cylinder. The two waveguides are the coaxial waveguide (top left) and the rectangular waveguide (bottom).

EXPERIMENT 1. In this experiment, we examine the convergence behaviors of all three methods under the initial approximate eigenpairs obtained from the linearized matrix pencil $\widehat{K}(\sigma_0^2) - \lambda\widehat{M}(\sigma_0^2)$ and the ones from the linear component $K - \lambda M$ of $T(\lambda)$.

Using the shift σ_0^2, we compute 10 eigenpairs (θ_j, v_j) of the matrix pair (K, M) by the shift-and-invert Arnoldi method [16]. It took 11.8 seconds[1]. These 10 computed θ_i values are listed in the first column of Table 1, and ordered according to (7). Note that all those eigenvalues are real since (K, M) is a symmetric positive definite matrix pair.

We also use the shift-and-invert Arnoldi algorithm to find $n = 10$ eigenpairs (θ_j, v_j) of the matrix pair $(\widehat{K}(\sigma_0^2), \widehat{M}(\sigma_0^2))$ with the shift σ_0^2. It took 23.0 seconds. Those 10 complex initial approximate eigenvalues are listed in the first column of Table 2, and ordered according to (7).

[1]All numerical experiments reported in this paper were run in MATLAB 7.0.1 on a Pentium IV PC with 2.6GHz CPU and 1GB of core memory.

Table 1: The approximate eigenvalues (λ_j) and the corresponding resonant frequencies (f_j) computed by IIT(10), MSLP(10) and NRRIT with real initials (θ_j).

	Initial	Approximation	Resonant	IIT(10)	MSLP(10)	NRRIT	
j	$\theta_j^{1/2}$	$\lambda_j^{1/2}$	f_j (MHz)	iters	iters	iters	$\|Q\|$
1	(5.97e-2, 0)	(5.97e-2, 8.62e-7)	2.8549	2	1	3	24
2	(6.19e-2, 0)	(3.82e-2, 5.24e-2)	1.8261	6	7	20	64
3	(8.08e-2, 0)	(6.07e-2, 6.37e-2)	2.8986	8	6	16	96
4	(8.38e-2, 0)	(8.37e-2, 1.96e-5)	3.9997	2	1	6	108
5	(8.41e-2, 0)	(8.41e-2, 3.39e-6)	4.0179	2	1	5	118
6	(8.77e-2, 0)	(8.77e-2, 3.81e-5)	4.1905	2	1	5	128
7	(8.83e-2, 0)	(8.83e-2, 5.72e-6)	4.2185	2	1	4	136
8	(9.24e-2, 0)	(9.34e-2, 3.93e-4)	4.4607	3	2	6	148
9	(9.74e-2, 0)	(1.19e-1, 1.88e-2)	5.7046	5	3	13	174
10	(1.04e-1, 0)	(1.25e-1, 1.75e-2)	6.0057	6	3	15	204
Total no. of iters				38	26	93	
Total CPU time (s)				179.4	391.2	50.6	

With those starting pairs, all three methods converge to the same set of 10 approximate eigenpairs of the NEP (16), as shown in second columns of Tables 1 and 2. All relative residual norms are smaller than $\epsilon = 10^{-8}$.

We observe that for all three methods, the method with complex initials converges much faster than using real initials. For example, the last columns of Tables 1 and 2 list the number of iterations for NRRIT method with real initials and with complex initials, respectively. It took a total of 93 iterations by using the real initials. On the other hand, it only takes a total of 57 iterations by using the complex initials.

In term of quality of converged eigenpairs, we have the following observations:

1. There are 6 eigenpairs founded by both real and complex initials. They are number 1, and 4 to 8 in Table 1 and number 1 to 6 in Table 2.

2. The 2nd and 3rd approximated eigenpairs computed with the real initials in Table 1 are not physically meaningful since their corresponding Q_e values are less than 1.

3. The resonant frequencies f_9 and f_{10} corresponding to the 9th and 10th approximate eigenvalues in Table 1 are further away from the prescribed frequency $f_0 = 2.7693$. than the ones in Table 2. They not the desired 10 eigenvalues closest to the prescribed frequency f_0.

In summary, it clearly shows that for all three methods, the initial approximate eigenpairs obtained from the linearized matrix pencil $\widehat{K}(\sigma_0^2) - \lambda\widehat{M}(\sigma_0^2)$ as defined in (6) are better than the ones obtained directly from the linear component $K - \lambda M$ of $T(\lambda)$.

Experiment 2. In this experiment, we compare computational efficiency of IIT(10), MSLP(10) and NRRIT. We focus on the performance for the methods with complex initial approximate eigenpairs as shown in Table 2.

The most expensive computational costs of IIT(10) and MSLP(10) are solving linear systems and linear eigenvalue problems of original problem size, respectively. It is generally more

Table 2: The approximate eigenvalues (λ_j) and the corresponding resonant frequencies (f_j) computed by IIT(10), MSLP(10) and NRRIT with complex initials (θ_j).

	Initial	Approximation	Resonant	IIT(10)	MSLP(10)	NRRIT	
j	$\theta_j^{1/2}$	$\lambda_j^{1/2}$	f_j (MHz)	iters	iters	iters	$\lvert Q\rvert$
1	(5.97e-2, 8.62e-7)	(5.97e-2, 8.62e-7)	2.8549	1	1	1	20
2	(8.37e-2, 1.81e-5)	(8.37e-2, 1.96e-5)	3.9997	1	1	9	38
3	(8.41e-2, 3.40e-6)	(8.41e-2, 3.39e-6)	4.0179	1	1	5	48
4	(8.77e-2, 3.37e-5)	(8.77e-2, 3.81e-5)	4.1905	2	1	5	58
5	(8.83e-2, 5.86e-6)	(8.83e-2, 5.72e-6)	4.2185	1	1	4	66
6	(9.34e-2, 3.25e-4)	(9.34e-2, 3.93e-4)	4.4607	2	1	4	74
7	(1.08e-1, 4.58e-3)	(1.09e-1, 3.60e-3)	5.2472	3	2	11	96
8	(1.11e-1, 1.15e-4)	(1.11e-1, 1.03e-4)	5.3079	2	1	6	108
9	(1.13e-1, 2.19e-5)	(1.13e-1, 2.27e-5)	5.4353	1	1	6	120
10	(1.15e-1, 3.31e-4)	(1.15e-1, 3.18e-4)	5.5075	2	1	6	132
Total no. of iters				16	11	57	
Total CPU time (s)				95.5	99.9	46.3	

expensive to an eigenvalue problem than a linear system of equations. This is seen that in Table 2, although MSLP(10) took a total of 11 iterations and IIT(10) took 16 iterations, the total CPU time of MSLP(10) is greater than IIT(10). This phenomenon is also observed by Ruhe [19].

In the inner loop of NRRIT, it is required to solve a linear system of the same dimension of the original NEP. However, note that the coefficient matrix of the linear system remains unchanged in the loop. Therefore, we can repeatedly use the factorized form of the coefficient matrix to solve the linear systems with different right hand sides in the loop.

In summary, NRRIT is the most CUP-efficient. It is about a factor of 2 faster than IIT(10) and MSLP(10).

EXPERIMENT 3. In this experiment, we compare the performance of NRRIT with real projection basis Q and complex projection basis V.

The left plot of Figure 3 shows the convergence history of NRRIT with real projection Q (solid line) and with complex projection V (dash-dot line). The horizontal axis is the number of iteration steps. The vertical axis is the relative residual (in log of base 10) of the Ritz pair (θ, z). The relative residual norm of an initial approximate eigenpair is marked by cross ("$\times$") for real projection subspace and is marked by plus ("$+$") for complex projection subspace. Once a Ritz pair (θ, z) is convergent, as measured by relative residual norm less than $\epsilon = 10^{-8}$, it is marked by a circle ("$\circ$") for the real projection subspace, and by a star ("$*$") for the complex projection subspace.

The right plot of Figure 3 shows the number of iterations to find an approximate eigenpair of the original NEP with relative residual norm less than $\epsilon = 10^{-8}$ by NRRIT with real projection Q (dark bar) and with complex projection V (gray bar), where the horizontal axis is the index of the eigenvalue. The vertical axis is the number of iteration steps *iters*.

We have the following remarks:

1. The initial relative residual norms with real projection subspace Q are smaller the ones with complex projection subspace V, even though the initial approximate eigenpairs are

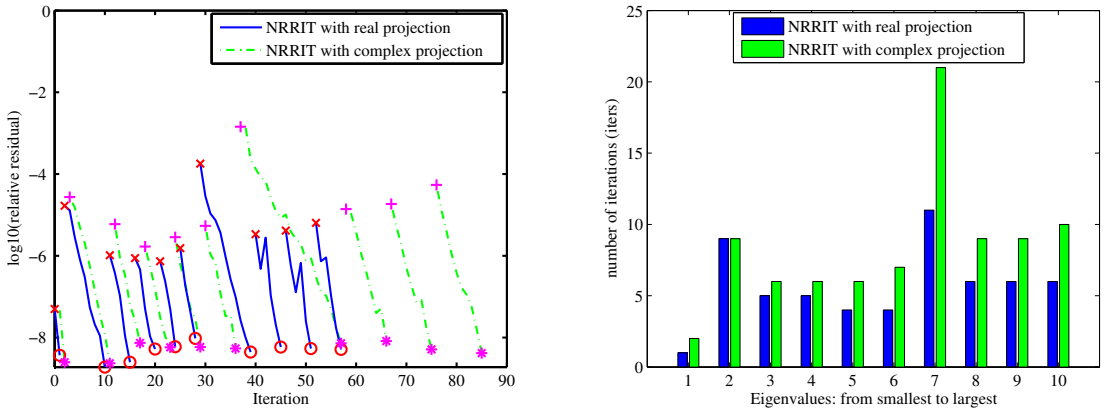

Figure 3: Left: convergence history of NRRIT with real projection subspace Q (solid line) and complex projection subspace V (dash-dot line). Right: the number of iterations required by NRRIT with real projection subspace Q (black bar) and complex projection subspace V (gray bar).

the same, see steps 5(a)-(c) in NRRIT algorithm.

2. NRRIT with real projection subspace Q requires fewer iterations (57 in total) than the complex projection subspace V (85 in total), 30% less in term of the number of iterations.

3. NRRIT with real projection Q took a total of 46.3 seconds to find 10 approximate eigenvalues, while NRRIT with complex projection V took a total of 60.5 seconds.

   It only makes a small portion of the total CPU-time to solve a projected NEP in NRRIT, although dimension of reduced NEP with real projection subspace Q doubles than the complex projection subspace V. The dimension of the projected NEP is $|Q|$, which is much smaller than the dimension of the original NEP.

In summary, NRRIT with real projection subspace outperforms the NRRIT with complex projection subspace. Preserving real symmetric property of the NEP (16) pays off.

7 Concluding Remarks

In this paper, we studied three numerical methods, namely IIT, MSLP and NRRIT, for computing a set of eigenpairs of the large scale NEP (1) arising from finite element analysis of resonant frequencies and external Q_e values of a waveguide loaded cavity in the next-generation accelerator design. We have shown how to determine the proper initial approximations and how to order them properly. IIT and MSLP are two classical methods, and are suitable for moderate size problems. NRRIT is a nonlinear Rayleigh-Ritz technique and is a generalization of nonlinear Arnoldi of Voss. It is developed for large scale NEPs. From our numerical experiments of model NEPs, we have seen that NRRIT outperforms IIT and MSLP.

Our current work includes incorporating an implementation of NRRIT into the simulation package Omega3P[2]. We also plan to study a 8-cavity cyromdule model in the near future.

[2]A package for electromagnetic modelings and simulations at Stanford Linear Accelerator Center.

This model involves at least 3 cutoff values and with DOFs up to millions. Our ultimate goal is to solve the NEPs arising from the International Linear Collider (ILC) superstructure, see discussion in [14, 13, 3]. The number of the cutoff values p can be up to 12.

References

[1] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst, eds., *Templates for the solution of Algebraic Eigenvalue Problems: A Practical Guide*, SIAM, Philadelphia, 2000.

[2] Z. Bai and Y. Su, *SOAR: A second-order arnoldi method for the solution of the quadratic eigenvalue problem*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 640–659.

[3] Z. Bai and C. Yang, *From self-consistency to SOAR: solving large-scale nonlinear eigenvalue problems.* SIAM NEWs, April 2006. Available at http://www.cs.ucdavis.edu/~bai/SIAMAN05.

[4] T. Betcke and H. Voss, *A Jacobi-Davidson-type projection method for nonlinear eigenvalue problems*, Future Generation Computer Systems, 20 (2004), pp. 363–372.

[5] T. Davis, *Algorihm 832: UMFPACK- an unsummetric-pattern multifrontal method*, Transcations on Mathematical Software, 30 (2004), pp. 196–199.

[6] J. W. Demmel, S. C. Eisenstat, J. R. Gilbert, X. S. Li, and J. W. H. Liu, *A supernodal approach to sparse partial pivoting*, SIAM J. Matrix Analysis and Applications, 20 (1999), pp. 720–755.

[7] H. Igarashi, Y. Sugawara, and T. Honma, *A numerical computations of external Q of resonant cavities*, IEEE Trans. on Magnetics, 31 (1995), pp. 1642–1645.

[8] E. Jarlebring and H. Voss, *Rational Krylov for nonlinear eigenproblems, an iterative projection method*, Applications of Mathematics, 50 (2005), pp. 543–554.

[9] T. Kerkhoven, A. Galick, U. Ravaioli, J. Arends, and Y. Saad, *Efficient numerical simulation of electron states in quantum wires*, J. Applied Physics, 68 (1990), pp. 3461–3469.

[10] K. Ko, N. Folwell, L. Ge, A. Guetz, V. Ivanov, L. Lee, Z. Li, I. Malik, W. Mi, C.-K. Ng, and M. Wolf, *Electromagnetic systems simulation - from simulation to fabrication.* SciDAC Report, Stanford Linear Accelerator Center, Menlo Park, CA, 2003.

[11] K. Ko, N. Folwell, L. Ge, A. Guetz, L. Lee, Z. Li, C. Ng, E. Prudencio, G. Schussman, R. Uplenchwar, and L. Xiao, *Advances in electromagnetic modelling through high performance computing*, Tech. Rep. SLAC-PUB-11789, Stanford Linear Accelerator Center, Menlo Park, CA, 2006.

[12] W. Kohn and L. J. Sham, *Self-consistent equations including exchange and correlation effects*, Phys. Rev., 140 (1965), pp. A1133–A1138.

[13] L. Lee, Z. Bai, W. Gao, L. Ge, P. Husbands, M. Kowalski, X. Li, Z. Li, C. Ng, C. Yang, E. G. Ng, and K. Ko, *Solving nonlinear eigenproblems in accelerator cavity design.* Presentation at SIAM Annual Meeting, New Orleans, LA, July 2005.

[14] L. Lee, L. Ge, Z. Li, C. Ng, K. Ko, B.-S. Liao, Z. Bai, W. Gao, C. Yang, P. Husbands, and E. G. Ng, *Modeling RF cavity with external coupling.* Presentation at SIAM Conference on Computational science and Engineering, Orlando, FL, Feb. 2005.

[15] R. Lehoucq, D. Sorensen, and C. Yang, *Arpack User's Guide: Solution of Large-Scale Eignevalue Problems with Implicitly Restarted Arnoldi Methods*, SIAM, Philadelphia, 1998.

[16] K. Meerbergen and A. Spence, *Implicitly restarted Arnoldi with purification for the shift-invert transformation*, Mathematics of Computation, 66 (1997), pp. 667–689.

[17] V. Mehrmann and H. Voss, *Nonlinear eigenvalue problems: A challenge for modern eigenvalue methods*, Tech. Rep. 83, Arbeitsbereich Mathematik, TU Hamburg-Harburg, 2004. Available at http://www.tu-harburg.de/mat/hp/voss/.

[18] A. Neumaier, *Residual inverse iteration for the nonlinear eigenvalue problem*, SIAM J. Numer. Anal., 22 (1985), pp. 914–923.

[19] A. Ruhe, *Algorithms for the nonlinear eigenvalue problem*, SIAM J. Numer. Anal., 10 (1973), pp. 674–689.

[20] ——, *A rational Krylov algorithm for nonlinear matrix eigenvalue problems*, Zapiski Nauchnyh Seminarov POMI, 268 (2000), pp. 176–180.

[21] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, Halstead Press, New York, 1992.

[22] H. Voss, *An Arnoldi method for nonlinear eigenvalue problems*, BIT Numerical Mathematics, 44 (2004), pp. 387–401.

[23] ——, *Numerical methods for sparse nonlinear eigenvalue problems*, in Proc. XVth Summer School on Software and Algorithms of Numerical Mathematic, Hejnice, Czech Republic, 2004.

[24] C. Yang, *Solving large-scale eigenvalue problems in SciDAC applications*, in Proceedings of SciDAC, no. 16 in Journal of Physics: Conference Series, 2005, pp. 425–434.