An Iterative Projection Method for Solving Large-Scale Nonlinear Eigenproblems with Application to Next-Generation Accelerator Design

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

January 10, 2006

1 Introduction

Recognizing the emerging needs to solve large-scale nonlinear eigenvalue problems arising from electronic structure calculations, simulations of MEMS devices and the electromagnetics field of a particle accelerator, the organizers of the 2005 SIAM Annual Meeting included an invited minisymposium on nonlinear eigenvalue problems in the program. The two-part (eight-talk) minisymposium brought together numerical analysts and application scientists. The sessions featured cutting-edge results from both communities, along with discussions of ongoing challenges [1].

In a recent survey paper [10], Mehrmann and Voss also called for more research on efficient algorithm development and computational theory for nonlinear eigenvalue problems by stating "*More research into new numerical methods and appropriate software for nonlinear eigenvalue problems is urgently needed.*" and "*For nonlinear eigenvalue problems, there are essentially no analogous packages that reach the standard of those for linear problems.*

In this paper, we consider large-scale nonlinear eigenvalue problems of the form

$$T(\lambda)x = (A - \lambda B + E(\lambda))x = 0, \tag{1}$$

where the term $A - \lambda B$ is the linear component and $E(\lambda)$ is the nonlinear component, namely, the elements of $E(\lambda)$ are nonlinear functions of λ. A number of applications give rise of nonlinear eigenvalue problems of such form. In the vibration study of fluid solid structures, see [14] and references therein,

$$T(\lambda) = A + \lambda B - \sum_{j=1}^{p} \frac{\lambda}{\sigma_j - \lambda} C_j,$$

where both A and B are symmetric positive definite, and C_j are symmetric. In an eigencomputation problem from fiber optic design, see [6],

$$T(\lambda) = A - \lambda I + s(\lambda) e_n e_n^T,$$

*Department of Mathematics, University of California, Davis. {liao@math.ucdavis.edu}

†Advanced Computations Department, Stanford Linear Accelerator Center. {liequan@slac.stanford.edu}

‡Department of Computer Science and Department of Mathematics, University of California, Davis. {bai@cs.ucdavis.edu}

§Advanced Computations Department, Stanford Linear Accelerator Center. {kwok@slac.stanford.edu}

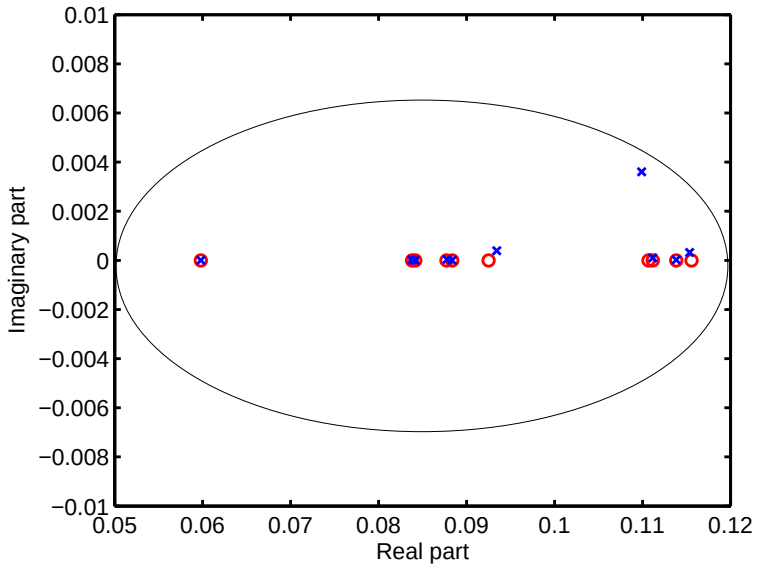

Figure 1: 10 smallest eigenvalues of the linear component (A, B) (in circle) and 10 closet eigenvalues of $T(\lambda)$ (in cross).

where $s(\lambda)$ involves the appropriate Bessel functions, and e_n is the last column of the $n \times n$ identity matrix.

A nonlinear eigenvalue problem we are particularly interested is from the finite element analysis of the resonant frequencies and external Q of a waveguide loaded cavity, as currently being studied by researchers at Stanford Linear Accelerator Center (SLAC) for the next-generation accelerator design [3, 8, 9]. It is a part of DOE SciDAC projects [7, 16]. The nonlinear eigenvalue problem is of the form

$$T(\lambda) = A - \lambda^2 B + \mathbf{i} \sum_j \sqrt{\lambda^2 - \sigma_j^2} W_j, \tag{2}$$

where A is symmetric semi-positive definite and B is symmetric positive definite. $W_j s$ are real symmetric. $\sigma_j s$ are given parameters and $\mathbf{i} = \sqrt{-1}$.

We assume that the linear component of $T(\lambda)$ is dominant or leading in the sense that the eigenvalues of $T(\lambda)$ are "small" perturbation of the eigenvalues of the linear component $A - \lambda B$. The plot of Figure 1 illustrates such a notion for the eigenvalue problem (2).

Because of problem size, the classical nonlinear eigensolvers such as inverse iteration [13], residual inverse iteration [12], and method of successive linear problems [13] are no longer suitable. *Self Consistent Field (SCF)* iteration algorithm is a widely used technique to solve large scale nonlinear eigenvalue problems. The method starts with linear solutions without the nonlinear terms and iterates to the norm of residual of the solution reaches to an acceptable order. It is basically a fixed point iteration algorithm. The convergence property of SCF is heavily rely on the initial approximations. Although this algorithm often converges in practice, it has to work on the original problem size in each iteration and convergence rate is unpredictable.

In this paper, we study an iterative subspace projection method, first proposed in [15]. We focus on the critical stages of algorithms, such as the choice of initial projection subspace, and the expansion and refinement of projection subspace. Using the new ideas on the choice of initial projection subspace, and the expansion and the refinement of projection subspace, we present a notable improvement over the early work [15] to the nonlinear eigenvalue problem (2).

2 A Framework of Subspace Projection Algorithms

We begin with the discussion of a general framework of the well-known Rayleigh-Ritz subspace projection approximation techniques, and the highlight the critical computational steps and the gist of structure-preserving. A high-level description of the Rayleigh-Ritz orthogonal subspace projection approximation technique consists of the following steps.

1. Select a proper projection subspace $\mathcal{Q}$.
2. Construct an orthonormal basis Q_n of $\mathcal{Q}$.
3. Seek an approximate eigenpair (θ, z) to satisfy the Galerkin condition:

$$z \in \mathcal{Q} \text{ such that } T(\theta)z \perp \mathcal{Q}. \tag{3}$$

4. Refine and/or expand the projection subspace $\mathcal{Q}$ if necessary.

We call θ a Ritz value and z a Ritz vector. The accuracy of a Ritz pair (θ, z) as approximate eigenpair is typically assessed through the residual vector $r = T(\theta)z$.

Note that at step 3, since $z \in \mathcal{Q}, z = Q_n g$ for some n-vector g, where n is the dimension of the subspace $\mathcal{Q}$. Hence, step 3 is equivalent to solve the reduced eigenvalue problem of determining θ and g satisfying

$$T_n(\theta)g = 0 \tag{4}$$

where $T_n(\theta) = Q_n^H T(\theta) Q_n$ is an $n \times n$ matrix. With proper chosen and refined projection subspace $\mathcal{Q}$, it is typical that satisfactory approximations to few desired eigenpairs of $T(\lambda)$ occur for a small dimension of projection subspace $\mathcal{Q}$, i.e., $n \ll N$. The dimension of the reduced eigenvalue problem (4) is significantly smaller than the original one (1). The reduce small eigenvalue problems can then be solved using established dense eigencomputation techniques.

Sometimes, it is also beneficial to introduce another subspace $\mathcal{P}$ and to consider an oblique subspace projection technique. In this case, the Galerkin condition (3) becomes the Petrov-Galerkin condition:

$$z \in \mathcal{Q} \text{ such that } T(\theta)z \perp \mathcal{P}. \tag{5}$$

It is clear that for a practical implementation of the Rayleigh-Ritz technique, essential issues are:

1. what is an optimal projection subspace $\mathcal{Q}$ and how to stably and efficiently generate an orthonormal basis Q_n of $\mathcal{Q}$?

   The gist of s structure-preserving algorithm is that such a projection subspace should not only be rich in the direction of the desired eigenvectors, but also have the property that the structure of $T(\lambda)$ can be preserved after the projection onto such a subspace.

For example, if $T(\lambda)$ is a family of symmetric matrices $T(\lambda) = A - \lambda B + E(\lambda)$, then the reduced family of matrices $T_n(\theta) = A_n - \theta B_n + E_n(\theta)$ should be a family of symmetric matrices. The nonlinear eigenvalue problem from RF cavity design is one of the examples.

2. how to iteratively expand, refine and restart the projection subspace $\mathcal{Q}$, and how to stably and efficiently carry out the associated computation.

Because of the particular nonlinear eigenproblem (1), we will focus on discussing the choice of the initial projection subspace Q_n and expansion and refinement of Q_n in next section.

3 Iterative Projection Methods

There are several iterative projection methods [15, 2, 5]. The only difference is using different nonlinear inverse iteration technique to expand projection subspace. Regardless of this, they have the same procedure. Here, we will only discuss a so-called nonlinear Arnoldi method described in [15] by Voss. It is proposed to solve the nonlinear eigenvalue problems of the form (1). It suggests that to find approximate eigenpairs of the nonlinear eigenvalue problem (1) we execute the following steps.

1. Select or update a proper projection subspace $\mathcal{Q}$.
2. Construct an orthonormal basis Q_n of $\mathcal{Q}$.
3. Solve the projected NEP

$$T_n(\theta)y = (A_n - \theta B_n + E_n(\theta))y = 0$$

where $A_n = Q_n^H A Q_n$,$B_n = Q_n^H B Q_n$ and $E_n(\theta) = Q_n^H E(\theta) Q_n$.

4. Test whether to accept a Ritz pair $(\theta, Q_n y)$ as an approximate eigenpair.
5. Iterate if necessary.

Note that this algorithm is under the framework of the structure-persevering subspace algorithm as described in section 2.

It is clear that for the nonlinear eigenvalue problem as we described, the subspace defined by the eigenvectors of linear component $A - \lambda B$ corresponding to the eigenvalues in the region of interest is a natural choice for the initial projection subspace Q_n. In the RF cavity design, a strong dominant linear component $A - \lambda B$ means that the external Q values are very large and therefore the waveguide boundary conditions act like the perfect insulating boundary conditions (magnetic boundary conditions). In [15], it is advised to use Q_n based on the eigenvectors of the eigenvalue problem $T(\sigma)u = \theta u$, where σ is a shift close to the eigenvalue of interest.

As recommended in [15], the projection subspace Q_n can be expanded using the vector q from a nonlinear inverse iteration [13, 12] corresponding to the "next" eigenvalue and eigenvector to be found:

$$Q_n \longrightarrow [\, Q_n \;\; q \,]$$

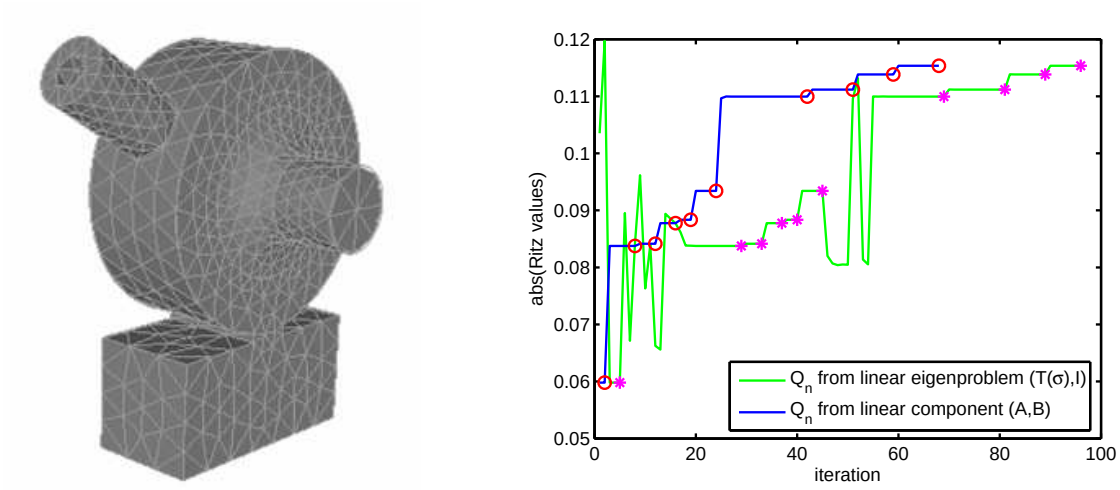

Figure 2: Left: test example of accelerator cavity design. Right: convergence history of Voss's nonlinear Arnoldi method and the modified one.

where q is orthonormalized with respect to Q_n. Specifically, the new search direction q is formed by following (*nonlinear Arnoldi*)

$$\begin{aligned} &v = T^{-1}(\sigma)T(\theta)z, \\ &v = v - Q_n Q_n^H v,\ q = v/\|v\|,\ Q_{n+1} = [\ Q_n\ q\ ] \end{aligned}$$

where (θ, z) is the Ritz pair of $T(\lambda)$ and σ is a fixed shift not too far away from the eigenvalue.

For refining Q_n, we first order the columns of initial Q_n corresponding to some ordering of eigenvalues of the linear component $A - \lambda B$. Once an eigenvector is convergent, it replaces the corresponding columns of Q_n. This is particular help for memory-saving when the restarting is needed because the projection matrix has already consisted of the convergent eigenvectors.

The improvement of the modifications is demonstrated in the next section.

4 Applications to Next-Generation Accelerator Design

We consider the cavity with the external waveguide coupling in the Accelerator RF cavity design. The frequency domain representation of the Maxwell equation (written in terms of the electric field E) has the form [4]

$$\begin{aligned} &\nabla \times (\nabla \times \mathbf{E}) = \lambda^2 \mathbf{E}, && \text{in } \Omega, \\ &n \times \mathbf{E} = 0, && \text{on electric boundary}, \\ &n \times (\nabla \times \mathbf{E}) + \mathrm{i} \textstyle\sum_j \sqrt{\lambda^2 - \sigma_j^2}\ n \times n \times \mathbf{E} = 0, && \text{on magnetic boundary}, \end{aligned} \tag{6}$$

where λ is an unknown cavity resonant frequency and n is the normal vector of the boundary at that point. The boundary conditions in second and third equations of (6) are perfectly conducting and waveguide boundary conditions, respectively. In this test example, we consider the cavity with waveguide coupling of two types (σ_0 and σ_1) as shown in the left plot of Figure 2. A hierarchical Nedelec-type [11] disrcetization yields a nonlinear eigenvalue problem of the form

$$T(\lambda)x := (A - \lambda^2 B + E(\lambda))x = 0 \tag{7}$$

Figure 3: ILC superstructure consisting of two accelerating cavities with power and HOM couplers.

where

$$E(\lambda) = \mathbf{i}\sqrt{\lambda^2 - \sigma_0^2}W_0 + \mathbf{i}\sqrt{\lambda^2 - \sigma_1^2}W_1,$$

and stiffness matrix $A \geq 0$ and mass matrix $B > 0$ are sparse real symmetric $N \times N$ matrices, $W_j s$ are sparse low rank real symmetric $N \times N$ matrices. The order N of the $T(\lambda)$ is 9956. The parameters $\sigma_0 = 0$ and $\sigma_1 = 0.043551$.

Using the new ideas on the choice of initial projection subspace Q_n, and the refinement of Q_n with convergent eigenvectors, we have seen a notable improvement over the early work of Voss [15] to the nonlinear eigenvalue problem (7). In this experiment, we want to find 10 eigenvalues of $T(\lambda)$ closest to the point about $\sigma = 0.05$, as shown in Figure 1. The right plot of Figure 2 shows the convergence history of the nonlinear Arnoldi method implemented according to [15] (green curve, or grey in black-and-white), and the modified one as we discussed above (blue curve, or dark curve in black-and-white). The horizontal axis is the dimension n of the subspace span$\{Q_n\}$. The vertical axis is the absolute Ritz values $|\theta|$ of the next one to be found. Once a Ritz pair $(\theta, Q_n y)$ is convergent, as measured by relative residual norm $\|T(\theta)Q_n y\|_2/\|Q_n y\|_2$, it is marked as magenta star for the nonlinear Arnoldi method and as red circle for the modified one. From the plot, we can easily see that the convergence rate of the first 6 eigenvalues of the modified method is much better. In addition, the modified algorithm took total 67 steps of iterations to find all 10 desired eigenvalues, 30% less than the number of steps required by the nonlinear Arnoldi method.

5 Concluding Remarks

We have pointed out that a variety of important nonlinear eigenvalue problems have the form (1) which consists of a dominant linear and positive pencil and a "small" nonlinear term. A rigorous study of the perturbation theory of a matrix pencil with nonlinear perturbation is a subject of future study. A case study for the RF cavity nonlinear eigenproblems have been conducted. This is particular important to the next-generation accelerator design when considering the International Linear Collider (ILC) superstructure (Figure 3) with two types of waveguide coupling. We have made progress by using linear eigenmodes as initial projection subspace and modifying the initial subspace. To improve the solver efficiency further, more research is needed to clarify the underlying structure of nonlinear eigenvalue problems of this particular type. The followings are some issues. Linear eigenmodes often provide good initial projection subspace, other alternatives to add the information of nonlinear eigenmodes need to be explored. This is particularly important when we consider the issue of restarting. A variant of nonlinear inverse iteration is currently used to subspace expansion [2, 5, 15]. More efficient techniques, including block version, also need to be studied.

References

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

[2] T. Betcke and H. Voss. A Jacobi-Davidson-type projection method for nonlinear eigenvalue problems. *Future Generation Computer Systems*, 20(3):363–372, 2004.

[3] H. Igarashi, Y. Sugawara, and T. Honma. A numerical computations of external Q of resonant cavities. *IEEE Trans.on Magnetics*, 31(3):1642–1645, 1995.

[4] J. D. Jackson. *Classical Electrodynamics*. Wiley, NY, 3rd edition, 1999.

[5] Elias Jarlebring and Heinrich Voss. Rational krylov for nonlinear eigenproblems, an iterative projection method, 2003. To Appear Appl. Math.

[6] L. Kaufman. Dispersion calculations in fiber optic design. Presentation at SIAM Annual Meeting, New Orleans, LA, July 2005.

[7] K. Ko, N. Folwell, L. Ge, A. Guetz, V. Ivanov, L.-Q. 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.

[8] L. Lee. Modeling RF cavity with external coupling. Presentation at SIAM Conference on Computational science and Engineering, Orlando, FL, Feb. 2005.

[9] L. Lee. Solving nonlinear eigenproblems in accelerator cavity design. Presentation at SIAM Annual Meeting, New Orleans, LA, July 2005.

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

[11] J. C. Nedelec. Mixed finite elements in R^3. *Numer. Math.*, 35:315–341, 1980.

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

[13] A. Ruhe. Algorithms for the nonlinear eigenvalue problem. *SIAM J. Numer. Anal.*, 10:674–689, 1973.

[14] H. Voss. A rational spectral problem in fluid-solid vibration. *Electr. Trans. Numer. Anal.*, 16:94–106, 2003.

[15] H. Voss. An Arnoldi method for nonlinear eigenvalue problems. *BIT Numerical Mathematics*, 44(2):387–401, 2004.

[16] Chao Yang. Solving large-scale eigenvalue problems in scidac applications. In *Proceedings of SciDAC*, Journal of Physics: Conference Series, 2005.