Mathematisches Forschungsinstitut Oberwolfach

Report No. 56/2013

DOI: 10.4171/OWR/2013/56

Numerical Solution of PDE Eigenvalue Problems

Organised by
Andrew Knyazev, Cambridge MA
Volker Mehrmann, Berlin
Jinchao Xu, University Park

17 November – 23 November 2013

Abstract. This workshop brought together researchers from many different areas of numerical analysis, scientific computing and application areas, ranging from quantum mechanics, acoustic field computation to material science, working on eigenvalue problems for partial differential equations. Major challenges and new research directions were identified and the interdisciplinary cooperation was strengthened through a very lively workshop with many discussions.

Mathematics Subject Classification (2000): 35P30, 45C05, 65L15.

Introduction by the Organisers

The numerical solution of eigenvalue problems for partial differential equations (PDEs) is an important task in many application areas such as:

- dynamics of electromagnetic fields;
- electronic structure calculations;
- band structure calculations in photonic crystals;
- vibration analysis of heterogeneous material structures;
- particle accelerator simulations;
- vibrations and buckling in mechanics, structural dynamics;
- neutron flow simulations in nuclear reactors; and many more.

The topic involves theoretical research in several different areas of mathematics ranging from operator theory and matrix computation to modern numerical treatment of partial differential equations. It is also related to computer science, since the novel mathematical ideas, related to efficient computation of eigenvalues

and invariant subspaces, need to be efficiently implemented in modern high performance software. This must be highly parallel, taking advantage of availability of thousands of multi-core computer processors, which adds significant constraints on possible algorithms and brings new practical and theoretical challenges.

In recent years major research developments in the area of PDE eigenvalue problems have taken place including the following:

- meshless and generalized finite element method approximation methods;
- adaptive finite element methods;
- methods for polynomial and other nonlinear eigenvalue problems
- a priori and a posteriori eigenvalue and eigenvector error estimation;
- convergence theory for preconditioned and inexact eigensolvers;
- multigrid, domain decomposition and incomplete factorization based preconditioning for eigenproblems;
- public software implementing efficient eigensolvers for parallel computers.

Novel research directions have appeared for non-linear, non-selfadjoint, and parameter-dependent problems. New homotopy approaches are combined with PDE eigensolvers in order to deal with optimization problems, where the PDE eigenvalue problem appears in the inner loop. Very recently, a new perturbation/error analysis has evolved that applies directly to nonlinear eigenvalue problems.

Nevertheless, many difficult questions remain open even for linear eigenvalue problems including the design of good error estimators, the solution effective recycling of computed information in homotopy or optimization methods, and the treatment of multiple eigenvalues and other ill-conditioned problems. As computers continue getting more powerful, the size of matrices involved in eigenvalue and singular value computations keeps growing. Numerical solution of billion-size problems is now typical in quantum mechanics as well as in many engineering applications. The issues of numerical stability and round-off error analysis thus attract renewed attention.

These topics were addressed during the workshop, successfully taking advantage of the interdisciplinary interaction between researchers representing many different scientific fields related to eigenvalue problems and PDEs. Major challenges and further research directions were discussed and the road for further research cooperation was paved.

Workshop: Numerical Solution of PDE Eigenvalue Problems

Table of Contents

Abstracts

The Life Cycle of an Eigenvalue Problem

Mark Embree
(joint work with Jeffrey Hokanson and Charles Puelz)

This talk set the stage for our workshop on the *Numerical Solution of PDE Eigenvalue Problems* by describing the numerous mathematical moves connecting a motivating physical problem to numerically computed eigenvalues. We break this process into five steps:

(1) physical problem → mathematical model;
(2) mathematical model → linear operator eigenvalue problem;
(3) linear operator eigenvalue problem → large discretization matrix;
(4) large discretization matrix → small projected matrix;
(5) small projected matrix → eigenvalues (ideally with high relative accuracy).

We argue that much insight can be gained by working across multiple stages of this process, while by focusing too narrowly on one stage one might end up answering the wrong question. Here we briefly discuss a few examples that illustrate this point. (For others who take a similar perspective, see, e.g., [3, 9, 11].)

(1) Historically and pedagogically, mechanical vibrations give rise to the first physical eigenvalue problems [13]. Despite this pedigree, reconciling mathematical models to the true vibrations of a damped string can prove quite tricky. The vacuum chamber experiments and improved exponential fitting algorithms of Hokanson [8] show how difficult it can be to measure high frequency eigenvalues, though these are precisely the values that differentiate between distinct damping models and play a crucial role in inverse spectral theory for strings [4].
(2) Nonlinear eigenvalue problems can be linearized in a variety of ways. We show spectral approximations obtained by a simple linearization of an exponential eigenvalue problem from a delay differential equation [10], as well as a quadratic eigenvalue problem modeling vibrations of a damped, hinged beam. The latter example, discretized with piecewise cubic Hermite finite elements, can lead to highly inaccurate computed eigenvalues. At the discretization matrix level, Higham et al. [7] illustrate how careful scaling can deliver more accurate eigenvalues for this problem. We obtain similar results simply by using (a discretization of) the correct physical norm, then argue that, ideally, the conditioning of the eigenvalues of the discretization matrix should match the norms of the spectral projectors of the operator: if the eigenvalues are sensitive to perturbations at the operator level, the discretization should capture that feature of the model.
(3) After outlining some classical results describing how eigenvalues of a discretization matrix converge to eigenvalues of the underlying operator [1],

we explore a few pathologies. For example, truncation of an infinite domain can introduce spurious eigenvalues [2], while the presence of essential spectrum can lead to the phenomenon of *spectral pollution* [6]. With a multiplication operator proposed by Boulton, we show how a shift-invert eigensolver applied to the discretization will be drawn to the polluting eigenvalues, while applying the shift-invert transformation before discretization avoids this problem. Finally, we address the question of how many eigenvalues one seeks to compute. While typical PDE problems require only a small number of eigenvalues, in some circumstances one needs the entire spectrum. As an example, we discuss a Schrödinger operator modeling a quasicrystal, whose spectrum is a Cantor set; see, e.g., [5] for details.

(4) PDE eigenvalue problems are often solved by applying a Krylov subspace method to the shift-invert transformation of the discretization matrix, since convergence of such methods applied to the discretization matrix itself converge very slowly. At the operator level, the shift-invert transformation is the only natural mode of operation: domain considerations prevent one from building a Krylov subspace with the operator itself, but one can readily do so with the inverted operator. Functional analysis suggests the right matrix approach.

(5) Finally, the computation of discretization matrix eigenvalues is complicated by the limitations on the relative accuracy of the computed eigenvalues. Generally the smallest eigenvalues of the discretization matrix are associated with lowest frequency modes, and thus converge most rapidly to the operator eigenvalues. Standard eigenvalue algorithms applied to the $n \times n$ discretization matrix A_n deliver the exact eigenvalues of $A_n + E$, where $\|E\| = O(n\|A_n\|\varepsilon_{\text{mach}})$ and $\varepsilon_{\text{mach}}$ reflects the precision of the floating point arithmetic. Since A_n discretizes an unbounded operator, $\|A_n\| \to \infty$ as $n \to \infty$, so the relative accuracy of the smallest eigenvalue degrades with large n (a fact explored in greater detail in the talk by Qiang Ye). This problem becomes more acute for higher order differential operators, where $\|A_n\|$ grows more rapidly. At the linear algebra level one might apply more robust algorithms that preserve high relative accuracy. Alternatively, one could use higher order discretizations (e.g., spectral methods) for which the lowest frequency eigenvalues converge more rapidly, before accuracy is overwhelmed by $n\|A_n\|\varepsilon_{\text{mach}}$.

Spanning across several of these all levels are issues related to non-self-adjointness (or, more precisely, nonnormality). With a simple convection-diffusion problem we illustrate three distinct effects of this depature from normality [12]: transient growth in evolution problems; delayed convergence of matrix eigenvalues to operator eigenvalues; inaccurate calculation of eigenvalues of the discretization matrix.

References

[1] I. Babuška and J. Osborn, *Eigenvalue problems*, in Handbook of Numerical Analysis, P. G. Ciarlet and J. L. Lions, eds., vol. 2, Elsevier, Amsterdam, 1991.

[2] B. M. Brown and M. Marletta, *Spectral inclusion and spectral exactness for singular non-self-adjoint Hamiltonian systems*, Proc. Roy. Soc. London A, 459 (2003), pp. 1987–2009.
[3] F. Chatelin, *Spectral Approximation of Linear Operators*, Academic Press, New York, 1983.
[4] S. J. Cox and M. Embree, *Reconstructing an even damping from a single spectrum*, Inverse Problems, 27 (2011), p. 035012 (18pp).
[5] D. Damanik, M. Embree, and A. Gorodetski, *Spectral properties of Schrödinger operators arising in the study of quasicrystals*, Tech. Rep. TR 12-21, Rice University, Department of Computational and Applied Mathematics, October 2012. Submitted for inclusion in the monograph, *Directions in Aperiodic Order*.
[6] E. B. Davies and M. Plum, *Spectral pollution*, IMA J. Numer. Anal., 24 (2004), pp. 417–438.
[7] N. J. Higham, D. S. Mackey, F. Tisseur, and S. D. Garvey, *Scaling, sensitivity and stability in the numerical solution of quadratic eigenvalue problems*, Int. J. Num. Methods Engrg., 73 (2008), pp. 344–360.
[8] J. M. Hokanson, *Numerically stable and statistically efficient algorithms for large scale exponential fitting*, PhD thesis, Rice University, 2013.
[9] J. Liesen and Z. Strakoš, *Krylov Subspace Methods: Principles and Analysis*, Oxford University Press, Oxford, 2013.
[10] W. Michiels and S.-I. Niculescu, *Stability and Stabilization of Time-Delay Systems: An Eigenvalue-Based Approach*, SIAM, Philadelphia, 2007.
[11] M. Plum, *Eigenvalue problems for differential equations*, in Wavelets, Multilevel Methods and Elliptic PDEs, M. Ainsworth, J. Levesley, M. Marletta, and W. A. Light, eds., Oxford University Press, Oxford, 1997, pp. 39–83.
[12] L. N. Trefethen and M. Embree, *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*, Princeton University Press, Princeton, NJ, 2005.
[13] C. Truesdell, *The Rational Mechanics of Flexible or Elastic Bodies, 1638–1788*, Leonhardi Euleri Opera Omnia, Introduction to Volumes X and XI, Second Series, Orell Füssli, Zürich, 1960.

AGMG: from academic research to industrial software

Yvan Notay

The AGMG software [1] is nowadays routinely used in industry to solve large sparse linear systems arising from scalar elliptic PDEs. For instance, the Rolls-Royce company integrated it in its combustion code. This is not so frequent for a project that started as a purely academic research, with first results reported six years ago at the 2007 Preconditioning Conference in Toulouse, and three subsequent publications [2, 3, 4] mainly focusing on theoretical aspects. Of course, this requires some luck: it is only after significant investigations that one may assess the industrial potential of an academic research, and quite often the answer is rather negative.

However, considering how the AGMG project has been developed, it turns out that the relatively rapid emergence of an industrial code has been possible thanks to several design choices that are far from standard in the numerical mathematics community. Indeed, it is nowadays well admitted that a paper cannot be good if the expectations of the potential readership are not taken into account. However, one often ignores the consequences of the similar statement: "to develop a good

method, first consider the needs of those who could benefit from it". In the talk, we review the implications of this viewpoint, and explain how they contributed to the success of the AGMG software.

In fact, not all academic developments has industrial potential.

- If one focuses too much on industrial applications, one likely misses real innovation.
- Academic research is a long term collaborative effort: explicit collaboration, implicit collaboration (one elaborates on others' results and what one does aims at being useful to further others).

  Industrial applications are only at the end of the chain

  AGMG is at cross-point of multigrid and numerical algebra, and is much indebted to the many ones who contributed these fields

  (Industrial applicability of CG in '56?)

But industrial potential may be missed even when it is present because communication (papers, talks, algorithm description) is organized to be appreciated by authors' scientific community, which often does not contain any *real* user. This induces some practices which do not help to identify the real practical scope of the research:

- numerical results focus on iteration counts or other statistics, disregarding timings and comparison;
- focus is on the detailed analysis of a few examples leaving asides robustness;
- the methods contain various parameters defined in a way that is obscure for non experts;
- the method is too complex to be recoded but the code is not made available.

A key point is the willingness to publish software codes. This requires additional work: clean up, comment crucial parts, etc. But this is often worthwhile, even for the author of the code; in fact, a code that has not been published most often perishes.

Once one decides to publish his/her code, this induces a Copernican revolution: one has to take into account potential users. Otherwise, it is a bit like if one would consider his/her own notes as manuscript ready to be submitted. And, all in all **the constraints induced by the Copernican revolution are all what is needed to reveal the industrial potential**.

References

[1] Y. Notay, *AGMG software and documentation.*
See *http://homepages.ulb.ac.be/ ynotay/AGMG.*

[2] Y. Notay, *An aggregation-based algebraic multigrid method*, Electronic Trans. Numer. Anal., 37 (2010), pp. 123–146.

[3] A. Napov and Y. Notay, *An algebraic multigrid method with guaranteed convergence rate*, SIAM J. Sci. Comput., 34 (2012), pp. A1079–A1109.

[4] Y. Notay, *Aggregation-based algebraic multigrid for convection-diffusion equations*, SIAM J. Sci. Comput., 34 (2012), pp. A2288–A2316.

The H^1-stability of the L_2-projection onto finite element spaces and its meaning for the Rayleigh-Ritz method

Harry Yserentant
(joint work with Randolph E. Bank)

The H^1-stability of the L_2-projection onto a finite element space is a valuable tool in many areas of finite element analysis and is easy to prove for uniform or quasiuniform grids. With the widespread use of adaptive and more general classes of nonuniform meshes, there is interest in generalizing this result to the nonuniform mesh case. At first glance, it seems that this should not be a difficult problem. The mass matrix, while no longer comparable to the identity matrix independent of the (now local) mesh size, does remain comparable to its own diagonal. One expects that the exponential decay of matrix elements away from the diagonal in the inverse of the mass matrix should also remain valid even in the nonuniform mesh case. However, the central difficulty is that this exponential decay might potentially be offset by exponential growth due to grading of the finite element mesh. The work of many authors addressed this issue by imposing certain local or global growth constraints on the mesh. Recently we have proved a very general, more or less concluding result of this type [1] that allows the inclusion of high order elements and meshes generated by many commonly used adaptive meshing strategies. This result can be used to derive some new error estimates for the eigenvalues and eigenfunctions obtained by the Rayleigh-Ritz method [2]. The errors are bounded in terms of the error of the best approximation of the eigenfunction under consideration by functions in the finite element space. In contrast to most of the classical theory, the approximation error of eigenfunctions other than the given one does not enter into these estimates.

Although our technique easily transfers to more general situations and can be applied to a large variety of different finite element spaces, we restrict ourselves in this note for the ease of presentation to the classical case of piecewise polynomial conforming elements. Starting point is a conforming triangulation $\mathcal{T}$ of a polygonal domain Ω in two or three space dimensions, built up from triangles in two space dimensions and tetrahedrons in the three-dimensional case. Associated with $\mathcal{T}$ is a conforming finite element space $\mathcal{S}$ of the usual kind, consisting of continuous, piecewise polynomial functions of at first fixed degree, determined by their nodal values. Our object of study is the L_2-orthogonal projection

$$Q : L_2(\Omega) \to \mathcal{S}$$

from $L_2(\Omega)$ onto the finite element space $\mathcal{S}$. We want to estimate the H^1-seminorm of the projection Qu, the L_2-norm of its first order derivatives, of a function u in the Sobolev space H^1 by the H^1-seminorm of u itself.

We subdivide the elements in the triangulation into elements of different levels. This level structure is associated with a constant $\mu \geq 1$ that measures the local grading of the mesh. To each finite element $T \in \mathcal{T}$ we assign a nonnegative integer $k(T)$, the level of the element, such that $\mu^{-k(T)}$ is roughly proportional to

the diameter $h(T)$ of T, in the sense that there are constants $\alpha > 0$ and $\beta > 0$ with

$$\alpha\mu^{-k(T)} \leq h(T) \leq \beta\mu^{-k(T)}. \tag{1}$$

The actual size of these two constants is of no significance; only their ratio β/α enters into our estimates. The triangulation $\mathcal{T}$ can be highly nonuniform and can contain finite elements from a very wide range of levels. We require, however, that the level of two neighboring elements differs at most by one. By neighbor, we refer to all elements that share a vertex with a given element.

Such a level structure can be imposed to any shape regular triangulation. Let the diameter of the elements surrounding a vertex differ at most by the factor $\mu > 1$. Let h_0 be the maximum diameter of an element in the triangulation and set $k(T) = k$ for an element T of diameter $h_0\mu^{-k} \leq h(T) < h_0\mu^{-k+1}$. The estimate (1) then holds with the constants $\alpha = h_0$ and $\beta = \mu h_0$ and the level of two neighboring elements differs by the choice of μ at most by one. Other choices of α and β offer a greater flexibility in the choice of μ. Consider, for example, the red green-refinement in two or three space dimensions and let the level of an element T count the number of refinement steps that are needed to generate T from its ancestor in the initial triangulation. Under the additional constraint to the refinement process that the level of two elements sharing a common vertex must not differ by more than one, (1) then holds with the generic constant $\mu = 2$. The ratio of the constants α and β reflects in such cases the degree of non-uniformity of the initial triangulation but remains independent of the degree of refinement.

The main ingredient of our proof for the H^1-stability of the L_2-orthogonal projection Q, as well as of other proofs in the literature, is its strong localization properties. We study them with help of an iterative procedure. Let $\mathcal{S}_1$ be the subspace of $\mathcal{S}$ of the functions vanishing on the boundaries of the single elements and $\mathcal{S}_0$ its L_2-orthogonal complement. The projection Q splits then into the sum $Q = Q_0 + Q_1$ of the L_2-orthogonal projection Q_0 onto the subspace $\mathcal{S}_0$ and the L_2-orthogonal projection Q_1 onto $\mathcal{S}_1 = \mathcal{S}_0^{\perp}$. As the contributions from the single elements to Q_1 do not interact, the localization properties of the projection Q are completely determined by those of of the projection Q_0 onto $\mathcal{S}_0$.

We label the vertices of the finite elements by the integers $i = 1, 2, \ldots n$. The vertex i is surrounded by the patch U_i, the union of the finite elements that share this vertex. Let $\mathcal{V}_i$ be the space that consists of the functions in the space $\mathcal{S}_0$ that vanish outside U_i. Let P_i be the L_2-orthogonal projection onto $\mathcal{V}_i$ and let

$$C = P_1 + P_2 + \ldots + P_n.$$

We construct with help of this operator approximations of the projection Q_0u of a given square integrable function u onto $\mathcal{S}_0$. For that purpose, we first define finite element functions $u^{(\nu)} \in \mathcal{S}_0$ recursively by $u^{(0)} = 0$ and

$$u^{(\nu+1)} = u^{(\nu)} + C\,(u - u^{(\nu)})$$

and recombine them in cg-like manner to weighted averages $w^{(\ell)}$. Then

$$\|Q_0u - w^{(\ell)}\|_0 \leq \frac{2\,q^{\ell}}{1 + q^{2\ell}}\,\|Q_0u\|_0,$$

where the convergence rate

$$q = \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}$$

with respect to the L_2-norm is determined by the spectral condition number κ of the operator C. It turns out that this spectral condition number is completely determined by the reference element and does not depend on the triangulation under consideration. For linear elements, when the given iterative procedure reduces to the polynomially accelerated Jacobi method, the condition number is

$$\kappa = 4, \quad \kappa = 5$$

in the two and the three-dimensional case, respectively. For elements up to order twelve in the two-dimensional and up to order seven in the three-dimensional case,

$$q < \frac{1}{2}.$$

Our main theorem links the convergence rate q of this iterative method, which depends only on the type of the finite elements, with the constant μ from assumption (1), which reflects the local grading of the mesh. If the product of these constants is less than one, one gets the weighted L_2-norm estimate

$$\|h^{-1}Qv\|_0 \leq c\,\|h^{-1}v\|_0 \tag{2}$$

for the functions v in the Sobolev space $H^1(\Omega)$, where the function h takes the value $h(T)$ on the triangle T. The constant depends only on the product of the constants μ and q and on the ratio β/α of the constants from assumption (1). From this estimate, the H^1-stability

$$|Qv|_1 \leq c\,|v|_1. \tag{3}$$

of the L_2-orthogonal projection Q follows by a standard argument similar to the quasiuniform case. The constant in (3) differs from that in (2) and depends additionally, via the inverse inequality, on the shape regularity of the elements.

This means in particular that the L_2-orthogonal projection is H^1-stable for finite elements up to order twelve in two and up to order seven in three space dimensions as long as $\mu \leq 2$. This is, for example, the case for the meshes generated by the red-green refinement process in two and three space dimensions. Similar results holds for simple *hp*-methods. For linear elements, the L_2-projection is H^1-stable if

$$\mu < 3, \quad \mu < \frac{\sqrt{5}+1}{\sqrt{5}-1} = 2.6180\ldots$$

in two and three space dimensions, respectively, a rather mild condition. The L_2-orthogonal projection remains, however, in general no longer H^1-stable if the grading becomes too extreme, as simple examples show [1].

Next we discuss what stability estimates like (3) mean for the Rayleigh-Ritz method. Our results are of general nature and do not only apply to the finite element case. We start from the usual abstract framework with two real Hilbert spaces $\mathcal{H}_0$ and $\mathcal{H}_1 \subseteq \mathcal{H}_0$ and a symmetric, coercive, and bounded bilinear form $a : \mathcal{H}_1 \times \mathcal{H}_1 \to \mathbb{R}$. The inner product on $\mathcal{H}_0$ is denoted by (u, v) and the induced

norm by $\|u\|_0$. We equip $\mathcal{H}_1$ for simplicity with the energy norm $\|u\|$ induced by the bilinear form $a(u,v)$. For convenience we assume that $\mathcal{H}_1$ is compactly embedded into $\mathcal{H}_0$ and that both spaces are infinite dimensional. Then there exists an infinite sequence $0 < \lambda_1 \le \lambda_2 \le \dots$ of eigenvalues of finite multiplicity tending to infinity and an assigned sequence of eigenvectors $u_1, u_2, \dots$ in $\mathcal{H}_1$ for which

$$(u_k, u_\ell) = \delta_{k\ell}, \quad a(u_k, u_\ell) = \lambda_k \delta_{k\ell}.$$

For second order elliptic eigenvalue problems over bounded domains Ω, $\mathcal{H}_0 = L_2(\Omega)$, and $\mathcal{H}_1$ is a subspace of $H^1(\Omega)$, depending on the boundary conditions.

The aim is to approximate the eigenvalues λ_k and the vectors in the assigned eigenspaces. For this, one chooses an n-dimensional subspace $\mathcal{S}$ of $\mathcal{H}_1$, say the finite element spaces considered above. Then there exist discrete eigenvectors $u'_1, u'_2, \dots, u'_n$ in $\mathcal{S}$ for eigenvalues $0 < \lambda'_1 \le \lambda'_2 \le \dots \le \lambda'_n$, satisfying the relations

$$(u'_k, u'_\ell) = \delta_{k\ell}, \quad a(u'_k, u'_\ell) = \lambda'_k \delta_{k\ell}.$$

The method thus replicates the weak form of the original eigenvalue problem and is determined by the choice of the subspace $\mathcal{S}$ replacing its solution space $\mathcal{H}_1$.

We will measure the approximation properties of the chosen subspace $\mathcal{S}$ in terms of the a-orthogonal projection operator $P : \mathcal{H}_1 \to \mathcal{S}$ defined by

$$a(Pu, v) = a(u, v), \quad v \in \mathcal{S}.$$

With respect to the energy norm the projection Pu is the best approximation of $u \in \mathcal{H}_1$ by an element of $\mathcal{S}$. Our main assumption is that the correspondingly defined $\mathcal{H}_0$-orthogonal projection Q from $\mathcal{H}_0$ onto $\mathcal{S}$ is stable in the energy norm, that is, that there exists another constant κ with

$$\|Qv\| \le \kappa \|v\|, \quad v \in \mathcal{H}_1. \tag{4}$$

This constant κ must be independent of hidden discretization parameters. This trivially holds for spectral methods in which the approximation spaces $\mathcal{S}$ are built up from eigenfunctions of a nearby eigenvalue problem, say, in the case of a second order problem, from eigenfunctions of the Laplace operator. The finite element case is more complicated. For second order problems, (4) is the H^1-stability (3) of the L_2-orthogonal projection onto the finite element space.

Our analysis starts from an error representation for the eigenvectors. The error between an eigenvector $u \in \mathcal{H}_1$ of the original problem for the eigenvalue λ and its $\mathcal{H}_0$-orthogonal projection onto the space spanned by the discrete eigenvectors u'_k in $\mathcal{S}$ for the eigenvalues λ'_k in a given neighborhood Λ of λ possesses the representation

$$u - \sum_{\lambda'_k \in \Lambda} (u, u'_k)\, u'_k = R\,(u - Pu) + (I - Q)(u - Pu),$$

where the mapping $R : \mathcal{H}_0 \to \mathcal{S}$ is defined by the expression

$$Rf = \sum_{\lambda'_k \notin \Lambda} \frac{\lambda'_k}{\lambda'_k - \lambda}\, (f, u'_k)\, u'_k.$$

It leads rather immediately to an error estimate in the $\mathcal{H}_0$-norm, to

$$\Big\| u \; - \sum_{\lambda'_k \in \Lambda} (u, u'_k)\, u'_k \Big\|_0 \leq \max(1, \gamma) \|u - Pu\|_0,$$

where the constant γ asymptotically measures the separation of the eigenvalue λ under consideration from the continuous eigenvalues outside Λ and is given by

$$\gamma \; = \; \max_{\lambda'_k \notin \Lambda} \Big| \, \frac{\lambda'_k}{\lambda'_k - \lambda} \, \Big|.$$

With help of the assumption (4) one obtains correspondingly the error estimate

$$\Big\| u \; - \sum_{\lambda'_k \in \Lambda} (u, u'_k)\, u'_k \Big\| \; \leq \; (\gamma + 1) \kappa \, \|u - Pu\|$$

in the energy norm, and finally, by the same type of arguments, the error estimate

$$\min_{\lambda'_k \geq \lambda} (\lambda'_k - \lambda) \; \leq \; (\alpha \kappa)^2 \, \|u - Pu\|^2$$

for the eigenvalues, provided that already a discrete eigenvalue $\lambda'_k \geq \lambda$ exists for which $\lambda'_k - \lambda \leq \lambda$. The constant α takes the value $\alpha = 1$ if there is no discrete eigenvalue $\lambda'_k < \lambda$ and otherwise the value

$$\alpha \; = \; \max_{\lambda'_k < \lambda} \; \frac{\lambda}{\lambda - \lambda'_k}.$$

For eigenvalues greater than the minimum eigenvalue, the size of the prefactors depends asymptotically on the separation of the eigenvalue under consideration from the smaller eigenvalues. The speed with which the discrete eigenvalues converge to their continuous counterparts is asymptotically determined by the speed with which the square of the best energy norm approximation error of the assigned eigenfunctions tends to zero. As with the error estimates for the eigenfunctions, pollution effects arising from the approximation error for other eigenfunctions than the one under consideration do not occur.

References

[1] R. E. Bank, H. Yserentant: On the H^1-stability of the L_2-projection onto finite element spaces. *Numer. Math.*, in print, DOI 10.1007/s00211-013-0562-4.

[2] H. Yserentant: A short theory of the Rayleigh-Ritz method. *CMAM* 13:495–502, 2013.

Parallel short-range O(N) complexity algorithm for approximate invariant subspace calculation of dimension N in electronic structure

Jean-Luc Fattebert

(joint work with Daniel Osei-Kuffuor)

Unlike classical physics problem where the number of variables (such as temperature, pressure, etc.) is fixed and does not grow with the system size, quantum mechanics models have a number of fields — electronic wave functions — proportional to the system size. This leads to $O(N^2)$ degrees of freedom to represent $O(N)$ electronic wave functions for a problem composed of N atoms. From a

mathematical point of view, the solution to that problem requires to calculate an invariant subspace of dimension N and thus typically leads to $O(N^3)$ operations for standard eigensolvers.

To make an efficient use of tomorrow's largest exascale computers, algorithms with $O(N)$ complexity and short-range communications are needed, so that one can simulate a number of atoms directly proportional to the number of processors available, for hundreds of thousands of atoms using hundreds of thousands of processors.

A lot of research has been carried out in the last 20 year in the physics and chemistry communities in an effort to develop $O(N)$ algorithms for electronic structure calculations [1]. Most $O(N)$ algorithms introduce some approximations or truncations of terms to reduce computational complexity. It thus becomes important to evaluate and control the accuracy of the resulting algorithms. A sufficient level accuracy often means that these $O(N)$ algorithms become competitive only at large scale (more than 500 atoms). But $O(N)$ complexity is not enough if one hopes to make an efficient use of exascale computers. Optimal algorithms also need to avoid global communications.

We present the first truly scalable First-Principles Molecular Dynamics algorithm with O(N) complexity and fully controllable accuracy, capable of simulating systems of sizes that were previously impossible with this degree of accuracy. By avoiding global communication, we have extended W. Kohn's condensed matter "nearsightedness" principle [2] to a practical computational scheme capable of extreme scalability. Accuracy is controlled by the mesh spacing of the finite difference discretization, the size of the localization regions in which the electronic wavefunctions are confined, and a cutoff beyond which the components of the overlap matrix can be omitted when computing selected elements of its inverse. We demonstrate the algorithm's excellent parallel scaling for up to 101,952 atoms on 23,328 processors, with a wall-clock time of the order of one minute per molecular dynamics time step.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

References

[1] D.R. Bowler and T. Miyazaki, *O(N) methods in electronic structure calculations*, Reports on Progress in Physics, **75** (2012), 036503.

[2] V. Kohn, *Density functional and density matrix method scaling linearly with the number of atoms*, Phys. Rev. Lett **76** (1996), 3168–317.

Solving Kohn-Sham algebraic nonlinear eigenvalue problem via rapid iterative diagonalization

Zhaojun Bai
(joint work with Yunfeng Cai, John Pask and N. Sukumkar)

The importance of electronic structure calculations stems from their underlying quantum-mechanical nature, yielding insights inaccessible to experiment and robust, predictive power unattainable by more approximate, empirical schemes. However, because such quantum-mechanical (QM) calculations are computationally intensive, a vast range of real materials problems of utmost importance to the Laboratory and wider scientific community remain inaccessible by such rigorous, QM approaches. The bottleneck in all such calculations is the solution of the large, sparse, numerical eigenproblems produced. This is due to the nonorthogonal basis sets employed in modern electronic-structure methods, such as the partition-of-unity finite element (PUFE) method and APW+lo. The resulting numerical eigenvalue problems are ill-conditioned. It is especially pronounced when the basis is saturated with orbital basis functions with long tails. Specifically, at each SCF-cycle, the linear generalized eigenvalue problem has ill-conditioned coefficient matrices with a large common near-null subspace. There are highly clustered eigenvalues with no obvious gap between the eigenvalues that are sought and the rest, which is particularly severe for magnetic and metallic systems. It is well documented that existing widely used eigensolvers, such as those based on the Davidson method, have proven to be no longer satisfactory. Furthermore, large off-diagonal entries in the coefficient matrices, such as from local orbital components of the basis, render standard diagonal-based preconditioners ineffective.

To address this deficiency, in the past two and half years, under the support of the UC-Lab Fees Research Program, we have focused on the development of new eigensolution algorithms and implementations in the context of a new real-space PUFE QM method. By virtue of its highly efficient orbital-polynomial basis, PUFE with our new eigensolver has shown order-of-magnitude reductions in basis size relative to state-of-the-art planewave based methods to attain the same accuracies for a variety of physical systems.

In this talk, we present our contributions in the following two major aspects:

(1) To address the issues of ill-conditioned coefficient matrices, non-diagonally dominant, and clustered eigenvalues, we present an asymptotic convergence analysis of the widely used preconditioned steepest descent method and its variants. We establish the notion of theoretically optimal preconditioners, and propose highly effective locally accelerated preconditioners for individual eigenpairs of interest. We have called the resulting new method the Locally Accelerated Block Steepest Descent (LABPSD) method [2].

(2) To efficiently implement the LABPSD method, we develop a two-stage scheme to apply highly effective locally accelerated preconditioners. At the first stage, we exploit the underlying "sparse + low rank" structure of the coefficient matrices to perform pre-processing to obtain good starting vectors for an inner

linear solver. At this stage, it involves a direct sparse indefinite complex matrix factorization at the first SCF iteration. At the second stage, we use an iterative linear solver with the pre-processed starting vectors produced by the first stage to apply targeted preconditioners of desired eigenpairs [1].

We have conducted extensive proof-of-principle tests for a variety of materials systems. In this talk, we highlight a simulation result for a triclinic metallic system, CeAl, which is a hard test case due to the following properties: (a) The potentials of the atoms are deep, producing strongly localized solutions that require larger basis sets to resolve. (b) The atoms are heavy, with many electrons in valence, requiring many eigenfunctions to be computed. (c) Because this system contains Ce, it requires 17 enrichment functions per atom (as opposed to e.g., 2 for Li), which increases basis size substantially for PUFE: yet, total degrees of freedom (DOF) are still a factor of 5 fewer than for planewaves. (d) The lattice is distorted, and atoms are displaced from ideal positions. This provides a completely general problem, with no special symmetries to exploit. Finally, for the simulation, we do not assume a band gap, but rather solve the completely general metallic problem.

We present numerical simulation results to show that in the convergence of the total energy computed by the PUFE method with the new LABPSD eigensolver, the average number of outer LABPSD iterations per SCF iteration and the average number of inner MINRES iterations per outer LAPBSD iteration for each eigenpair are all between 2 and 4. This is a significant achievement since it is comparable with the typical number of iterations for the standard LOBPCG method used in ABINIT on the well-conditioned standard eigenvalue problems produced by that method [3]. This indicates that for generalized eigenproblems with ill-conditioned coefficient matrices (as occur in electronic structure methods using nonorthogonal bases), the performance of the LABPSD is clearly superior to current state-of-the-art methods such as LOBPCG and more recent algorithm of Blaha and co-workers [4].

In summary, our studies with PUFE reveal for the first time that a systematically improvable real-space approach can attain the required accuracies in quantum-mechanical materials calculations with not only fewer but substantially fewer degrees of freedom than current state-of-the-art planewave based methods, as implemented in VASP, ABINIT, Qbox, and a host of other codes in current use. The LABPSD eigensolver has proven to efficiently solve the ill-conditioned generalized complex eigenproblem produced by PUFE.

References

[1] Y. Cai, Z. Bai, J. Pask and N. Sukumar, *Hybrid preconditioning for iterative diagonalization of ill-conditioned generalized eigenvalue problems in electronic structure calculations*, Journal of Computational Physics, **255** (2013), 16–33.

[2] Y. Cai, Z. Bai, J. Pask and N. Sukumar, *A locally accelerated block preconditioned steepest descent method for generalized Hermitian eigenvalue problems*, in preparation, 2013

[3] F. Bottin, S. Leroux, A. Knyazev, and G. Zerah. *Large-scale ab initio calculations based on three levels of parallelization*, Comput. Mater. Sci., **42** (2008), 329 – 336.

[4] P. Blaha, H. Hofstätter, O. Koch, R. Laskowski, and K. Schwarz. *Iterative diagonalization in augmented plane wave based methods in electronic structure calculations*, J. Comput. Phys., *229* (2010), 453–460.

Finite Dimensional Approximations of Nonlinear Eigenvalue Problems in Density Functional Models

Aihui Zhou

(joint work with Huajie Chen, Xiaoying Dai, Xingao Gong, and Lianhua He)

Density functional theory (DFT) has been widely and successfully used in computational materials science, quantum chemistry, and quantum biology [9, 11, 12, 13, 14, 18]. We see that the Thomas-Fermi-von Weizsäcker (TFvW) type equations and Kohn-Sham equations, which are nonlinear eigenvalue problems, play a crucial role in DFT [8, 11, 12, 15]. Hence it is significant to mathematically understand why DFT and its numerics work so well and to design new efficient numerical methods for such nonlinear eigenvalue equations.

To our knowledge, there are only a few works on numerical analysis of nonlinear eigenvalue problems in literature. We refer to [5, 6, 10, 16, 17] for convergence of finite dimensional approximations and [1, 2, 4, 7] for a priori convergence rates. We note that numerical analysis given in [1, 16, 17] are for problems with convex energy functionals only while [6] gives a priori error upper bound for a general case of TFvW type equations and [2] provides an a priori error estimate for planewave discretizations for the Kohn-Sham LDA equations under a coercivity assumption. We refer to [3, 4, 7] for a systematic study on mathematical justification for finite dimensional approximations of both TFvW type equations and Kohn-Sham equations as well as the associated directly numerical minimizing energy functional methods, including some understanding of several existing approximate methods in electronic structure calculations based on DFT, from a priori error analysis to a posteriori error estimations, and from designing adaptive finite element algorithms to analyzing their convergence and complexity.

Let us informally describe several recent results on finite dimensional approximations of nonlinear eigenvalue problems resulting from DFT in our group. Under some reasonable assumptions, we prove in [4, 6, 7, 17] that all the limit points of finite dimensional approximations are ground states of the system, and every eigenpair can be well approximated by the finite dimensional approximations when the associated local isomorphism condition is satisfied. We obtain convergence of ground state energy approximations [6, 17] and convergence rates of both eigenvalue and eigenfunction approximations [4, 7]. We also propose and analyze two adaptive finite element algorithms, which are based on the residual type a posteriori error estimators. We derive the a posteriori error estimates and show in [3, 5, 7] that under some reasonable assumptions, all limit points of the adaptive finite element approximations are ground states, some ground states can be approximated by adaptive finite element approximations with some convergence

rate. In addition, we obtain the quasi-optimal complexity of adaptive finite element approximations [3, 7], too.

References

[1] E. Cancès, R. Chakir, and Y. Maday, *Numerical analysis of nonlinear eigenvalue problems*, J. Sci. Comput., **45** (2010), 90-117.

[2] E. Cancès, R. Chakir, and Y. Maday, *Numerical analysis of the planewave discretization of some orbital-free and Kohn-Sham models*, M2AN, **46** (2012), 341-388.

[3] H. Chen, X. Dai, X. Gong, L. He, and A. Zhou, *Adaptive finite element approximations for Kohn-Sham models*, arXiv:1302.6896.

[4] H. Chen, X. Gong, L. He, Z. Yang, and A. Zhou, *Numerical analysis of finite dimensional approximations of Kohn-Sham equations*, Adv. Comput. Math., **38** (2013), 225-256.

[5] H. Chen, X. Gong, L. He, and A. Zhou, *Adaptive finite element approximations for a class of nonlinear eigenvalue problems in quantum physics*, Adv. Appl. Math. Mech., **3** (2011), 493-518.

[6] H. Chen, X. Gong, and A. Zhou, *Numerical approximations of a nonlinear eigenvalue problem and applications to a density functional model*, Math. Meth. Appl. Sci., **33** (2010), 1723-1742.

[7] H. Chen, L. He, and A. Zhou, *Finite element approximations of nonlinear eigenvalue problems in quantum physics*, Comput. Methods Appl. Mech. Engrg., **200** (2011), 1846-1865.

[8] H. Chen and A. Zhou, *Orbital-free density functional theory for molecular structure calculations*, Numer. Math. Theor. Meth. Appl., **1** (2008), 1-28.

[9] W. Kohn, *Density functional and density matrix method scaling linearly with the number of atoms*, Phys. Rev. Lett. **76** (1996), 3168-3171.

[10] B. Langwallner, C. Ortner, and E. Süli, *Existence and convergence results for the Galerkin approximation of an electroniic density functional*, M^3AS, **12** (2010), 2237-2265.

[11] C. Le Bris (ed.), *Handbook of Numerical Analysis, Vol. X. Special issue: Computational Chemistry*, North-Holland, Amsterdam, 2003.

[12] R. M. Martin, *Electronic Structure: Basic Theory and Practical Method*, Cambridge University Press, Cambridge, 2004.

[13] R. G. Parr and W. T. Yang, *Density-Functional Theory of Atoms and Molecules*, Oxford University Press, New York, Clarendon Press, Oxford, 1994.

[14] Y. Saad, J. R. Chelikowsky, and S. M. Shontz, *Numerical methods for electronic structure calculations of materials*, SIAM Review, **52** (2010), 3-54.

[15] Y. A. Wang and E. A. Carter, *Orbital-free kinetic-energy density functional theory*, in: Theoretical Methods in Condensed Phase Chemistry (S. D. Schwartz, ed.), Kluwer, Dordrecht, 2000, 117-184.

[16] A. Zhou, *An analysis of finite-dimensional approximations for the ground state solution of Bose-Einstein condensates*, Nonlinearity, **17** (2004), 541-550.

[17] A. Zhou, *Finite dimensional approximations for the electronic ground state solution of a molecular system*, Math. Meth. Appl. Sci., **30** (2007), 429-447.

[18] A. Zhou, *Hohenberg-Kohn theorem for Coulomb type systems and its generalization*, J. Math. Chem., **50** (2012), 2746–2754.

Fast algorithms for Kohn-Sham density functional theory

Lin Lin

Kohn-Sham density functional theory (KSDFT) is by far the most widely used electronic structure theory for condensed matter systems. However, the computational cost of the standard method for solving KSDFT increases cubically with

respect to the number of electrons in the system (N). The cubic scaling hinders the application of KSDFT to systems of large size such as nano-scale systems.

Our aim is to design efficient algorithms for solving KSDFT for both insulating and metallic systems [3, 2, 5, 1]. Our method focuses on the property that the electron density ρ depends only on the diagonal of the Fermi-Dirac operator (β: inverse temperature; μ: chemical potential)

$$\rho = \text{diag } f(H) \equiv \text{diag } \frac{2}{1 + e^{\beta(H-\mu)}}. \tag{1}$$

Our strategy is to expand the Fermi-Dirac operator into resolvents (Green's functions)

$$\rho \approx \text{diag } \sum_{i=1}^{P} \frac{\omega_i}{H - z_i}, \tag{2}$$

where $\omega_i, z_i \in \mathbb{C}$. By choosing the proper weight ω_i and position of poles z_i, the pole expansion achieves by far the most efficient representation cost that scales as $P \sim \mathcal{O}(\log \beta \Delta E)$ with small pre-constant. Numerical example indicates that $P = 80$ is more than sufficient even in the extreme case where $\beta \Delta E \approx 10^6$ [3].

Another key component in the PEXSI method is the selected inversion method, which allows the accurate and efficient computation of selected elements of a Green's function for a Kohn-Sham system [5, 4], taking the form $(H - zS)^{-1}$, where H is the Hamiltonian operator and S is the overlap matrix, and z is a complex shift. The selected inversion method reveals the connection between the inverse of a sparse matrix and its associated Cholesky factor: any element in the inverse matrix corresponding to a nonzero element of its associated Cholesky factor can be evaluated without using any element outside the nonzero pattern of the Cholesky factor in the inverse matrix. In particular, if exact arithmetic can be used (without round-off error), then the selected inversion method is an exact method for computing these selected elements needed for the electronic structure calculation. Since the Cholesky factor is sparse compared to the full inverse $(H - zS)^{-1}$, selected inversion significantly reduces the computational complexity from $\mathcal{O}(N^3)$ to at most $\mathcal{O}(N^2)$ *without loss of accuracy*, where N is the number of atoms in the system.

Using a 2D tight binding system on a structured grid, the selected inversion method can solve a system with 4.3 billion degrees of freedom under 25 minutes using 4096 processors in parallel [4]. When PEXSI is applied to Hamiltonian matrices discretized by atomic orbitals, one can perform first principle KSDFT calculation for a large carbon nanotube with more than 10,000 atoms on a single processor [1] using single-zeta basis function. To fully realize the capability of the PEXSI method and accelerate the electronic structure calculation for large scale systems in practice, we developed a general purpose massively parallel code with the same name PEXSI. The parallel PEXSI code is able to use Department of Energy (DOE) high performance machines with more than 100,000 cores. It can be used to solve problems that contain 10,000 to 100,000 atoms. The PEXSI method is now being integrated into SIESTA [6], one of the most popular electronic

structure software packages based on atomic orbitals. The resulting SIESTA-PEXSI method will be described in a forthcoming publication.

References

[1] L. Lin, M. Chen, C. Yang, and L. He. Accelerating atomic orbital-based electronic structure calculation via pole expansion and selected inversion. *J. Phys. Condens. Matter*, 25:295501, 2013.

[2] L. Lin, J. Lu, R. Car, and W. E. Multipole representation of the Fermi operator with application to the electronic structure analysis of metallic systems. *Phys. Rev. B*, 79:115133, 2009.

[3] L. Lin, J. Lu, L. Ying, and W. E. Pole-based approximation of the Fermi-Dirac function. *Chin. Ann. Math.*, 30B:729, 2009.

[4] L. Lin, C. Yang, J. Lu, L. Ying, and W. E. A fast parallel algorithm for selected inversion of structured sparse matrices with application to 2D electronic structure calculations. *SIAM J. Sci. Comput.*, 33:1329, 2011.

[5] L. Lin, C. Yang, J. Meza, J. Lu, L. Ying, and W. E. SelInv – An algorithm for selected inversion of a sparse symmetric matrix. *ACM. Trans. Math. Software*, 37:40, 2011.

[6] J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal. The SIESTA method for ab initio order-N materials simulation. *J. Phys.: Condens. Matter*, 14:2745–2779, 2002.

Adaptive Wavelet Methods for Calculating Excitonic Eigenstates in Disordered Quantum Wires

Christian Mollet

A novel adaptive approach to compute the eigenenergies and eigenfunctions of the two-particle (electron-hole) Schrödinger equation including Coulomb attraction is presented. We are looking for the energetically lowest exciton state of a thin one-dimensional semiconductor quantum wire in the presence of disorder which arises from the non-smooth interface between the wire and surrounding material. The problem of two interacting particles, a hole and an electron, is described by a *time-dependent two-particle Schrödinger equation* of the form

$$i\hbar \frac{\partial}{\partial t} p\left(x_e, x_h, t\right) = \left(E_g + \hat{H}_{\text{kin}} + \hat{H}_{\text{attr}} + \hat{H}_{\text{dis}}\right) p\left(x_e, x_h, t\right) - \hat{\mu} E\left(x_h, t\right) \delta\left(x_e - x_h\right), \tag{1}$$

which describes the (complex-valued) state function $p\left(x_e, x_h, t\right)$ of the electron-hole pair. Here

$$\hat{H}_{\text{kin}} := -\frac{\hbar^2}{2m_e^*}\frac{\partial^2}{\partial x_e^2} - \frac{\hbar^2}{2m_h^*}\frac{\partial^2}{\partial x_h^2}$$

denotes the Hamiltonian of two free particles,

$$\hat{H}_{\text{attr}} := \frac{-e^2}{4\pi\hat{\varepsilon}_0\hat{\varepsilon}_r\left(\min\{|x_e - x_h|, |x_e - x_h \pm L|\} + \hat{\gamma}\hat{R}\right)}$$

describes the electron-hole attraction and

$$\hat{H}_{\text{dis}} = V_{\text{dis},e}\left(x_e\right) + V_{\text{dis},h}\left(x_h\right) \tag{2}$$

models the disorder of the interface of the wire which will be specified below. The term

$$-\mu E\,(x_h,t)\,\delta\,(x_e-x_h)$$

describes the optical excitation where the Dirac delta $\delta(x)$ models the excitation and $E\,(x_h,t)$ denotes the function of the electric field of the optical excitation, i.e., the electric field of the incident light. Here m_e^* is the effective mass of an electron and m_h^* the effective mass of a hole, $e = 1e = 1{,}602 \times 10^{-19}\,C$ is the elementary charge, $\hat{\varepsilon}_0 = 8{,}854 \times 10^{-12}\,\frac{C}{Vm}$ is the electric constant, $\hat{\varepsilon}_r$ the relative permittivity, $\hat{\gamma}\hat{R}$ a regularization parameter, $\hat{\mu}$ denotes the optical dipole-matrix-element and $L > 0$ the length of the quantum wire.

The solution, i.e., the wavefunction, provide information on the optical properties of the wire, whereas the energies of the excitons are determined by the eigenvalues of the Hamiltonian. To this end, we are interested in solving the eigenvalue problem

$$\hat{E}\,X(x) = \hat{H}\,X(x), \qquad x \in (0,L)^2,$$

with Hamiltonian

$$\hat{H} := E_{\mathrm{g}} + \hat{H}_{\mathrm{kin}} + \hat{H}_{\mathrm{attr}} + \hat{H}_{\mathrm{dis}}. \tag{3}$$

The eigenvalue problem (3) is equipped with periodic boundary conditions and zero initial conditions.

Due to production processes involving a random disorder of the interface with the surrounding material, we have to deal with a non-ideal wire. The potential functions $V_{\mathrm{dis},h}$ and $V_{\mathrm{dis},e}$ appearing in (2) may therefore be assumed to be a periodic potential function on $(0,L)$ or, more realistically, may be modeled as a stochastic perturbation on $(0,L)$. This we describe using a piecewise constant function with randomly chosen step heights,

$$V_{\mathrm{dis},e}(x_e) := \textstyle\sum_{i=1}^{M} \mathrm{Ran}_{\mathrm{dis},e}(i)\,\mathrm{Char}_{\left[(i-1)\frac{L}{M},\,i\frac{L}{M}\right)}(x_e),$$

where $\mathrm{Char}_{\hat{I}}(x_e) := 1$ for $x_e \in \hat{I}$ and zero otherwise denotes the characteristic function on an interval $\hat{I}$ and M is the number of steps. Furthermore, $\mathrm{Ran}_{\mathrm{dis},e}(i) \sim \mathcal{N}\left(0,\sigma^2\right)$ for all $i \in \mathbb{N}$ are the corresponding randomly chosen step heights, that is, $\mathrm{Ran}_{\mathrm{dis},e}(i)$ is for each $i \in \mathbb{N}$ a normally distributed random number with expectation zero and variance σ^2.

We reformulate the eigenvalue problem (3) in an appropriate weak form whose bilinear form, after introducing a shift, can be arranged to be symmetric, continuous, and coercive. We obtain our final problem formulation:
Find $u \in H^1_{\mathrm{per}}((0,L)^2)$ and $E \in \mathbb{C}$ such that

$$a(u,v) = E\,(u,v)_{L_2((0,L)^2)} \quad \text{for all } v \in H^1_{\mathrm{per}}((0,L)^2), \tag{4}$$

where $a(\cdot,\cdot)$ defines the derived (shifted) bilinear form and E the shifted eigenvalue of (3).

In order to calculate the smallest eigenpair of (4), we will apply an adaptive wavelet method. To this end, we consider a suitable Riesz basis $\mathbf{\Psi} := \{\psi_i : i \in \mathcal{I}\}$ of $H^1_{\mathrm{per}}((0,L)^2)$ with

$$\|\sum_{i\in\mathcal{I}} v_i\psi_i\|_{H^1((0,L)^2)} \sim \|\mathbf{v}\|_{\ell_2}, \qquad \text{for all } \mathbf{v} \in \ell_2(\mathcal{I}).$$

This leads to an *equivalent* generalized eigenvalue problem over $\ell_2(\mathcal{I})$

$$A\mathbf{u} = EB\mathbf{u}, \qquad \mathbf{u} \in \ell_2(\mathcal{I}), \tag{5}$$

with bi-infinite matrices $A := (a(\psi_j, \psi_\ell))_{\ell,j\in\mathcal{I}}$ and $B := ((\psi_j, \psi_\ell)_{L_2((0,L)^2)})_{\ell,j\in\mathcal{I}}$. A detailed analysis of adaptive wavelet computations of eigenvalues for the present problem is described in [4]. Considering (5), a preconditioned inverse iteration scheme (PINVIT) yields an ideal solution algorithm to calculate the smallest eigenpair. In order to obtain an algorithm which is numerically feasible, we introduce a perturbation. This leads to the following *perturbed preconditioned inverse iteration* (PPINVIT) introduced in [6], essentially based on [2], of the form

$$\mathbf{v} \leftarrow \mathbf{v} - P^{-1}(A_\epsilon(\mathbf{v}) - \mu_\epsilon(\mathbf{v})B_\epsilon(\mathbf{v})), \tag{6}$$

with appropriate properly scaled (diagonal) preconditioner P^{-1} and Rayleigh quotient $\mu(\mathbf{v}) := \frac{\langle A\mathbf{v},\mathbf{v}\rangle}{\langle B\mathbf{v},\mathbf{v}\rangle}$ with $\ell_2(\mathcal{I})$-inner product $\langle\cdot,\cdot\rangle$. The perturbation in (6) is introduced by an inexact operator application indicated by the index ϵ. Since we deal with wavelet bases, we are able to control the error. The inexact operator application can essentially be separated into two main parts. First, the *prediction* of a suitable index set ensuring exactness of the operator application and second, the *efficient evaluation* of the corresponding inner products. The first task was elaborated in [1], where a PREDICTION scheme was introduced which yields a desired index set with an asymptotically optimal size and optimal computational effort by using tree structured index sets. The evaluation of the inner products can be done efficiently with the evaluation scheme based on local polynomial representations introduced in [5]. It can be shown, that the overall scheme can be applied to the present situation and allows for a convergence proof together with asymptotically optimal complexity estimates, see [3].

The numerical results demonstrate the benefit of the adaptive scheme. Figure 1 shows the numerical solution without disorder (left), the corresponding adaptive grid after some iteration steps (middle) and the error related to the used degrees of freedom on uniform grids and adaptive grids (right).

References

[1] A. Cohen, W. Dahmen, R. DeVore, *Sparse Evaluation of Compositions of Functions using Multiscale Expansions*, SIAM J. Math. Anal. **35**(2003), 279–303.

[2] A. Knyazev and K. Neymeyr, *Gradient flow approach to geometric convergence analysis of preconditioned eigensolvers,* SIAM J. Matrix Anal. **31** (2009), 621–628.

[3] C. Mollet, *Excitonic Eigenstates in Disordered Semiconductor Quantum Wires: Adaptive Computation of Eigenvalues for the Electronic Schrödinger Equation based on Wavelets*, Shaker-Verlag, DOI: 10.2370/OND000000000098, (2011).

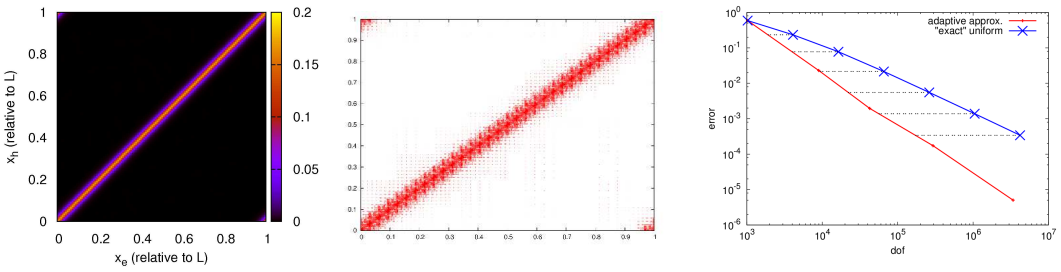


FIGURE 1. Eigenfunction w.r.t. smallest eigenvalue (left), adaptive grid (middle) and performance (right).

[4] C. Mollet, A. Kunoth, T. Meier, *Excitonic Eigenstates of Disordered Semiconductor Quantum Wires: Adaptive Wavelet Computation of Eigenvalues for the Electron-Hole Schrödinger Equation*, Commun. Comput. Physics **14** (2013), 21–47.

[5] C. Mollet, and R. Pabel, *Efficient Application of Nonlinear Stationary Operators in Adaptive Wavelet Methods – The Isotropic case*, Numerical Algorithms **63** (2013), 615–643.

[6] T. Rohwedder, R. Schneider, A. Zeiser, *Peturbed preconditioned inverse iteration for operator eigenvalue problems with applications to adaptive wavelet discretization*, Adv. Comp. Math. **34** (2010), 43–66.

Backward errors in the inexact Arnoldi process

CHRISTIAN SCHRÖDER

(joint work with Ute Kandler, Leo Taslaman)

Arnoldi's method is a standard tool in numerical linear algebra. Given a matrix $A \in \mathbb{C}^{n,n}$ and a normed starting vector $v_1 \in \mathbb{C}^n$ it generates the matrices V_{k+1} and $\underline{\mathrm{H}}_k$ satisfying the well known Arnoldi relation $AV_k = V_{k+1}\underline{\mathrm{H}}_k$.

In numerous applications including tensor computations and mixed precision arithmetic vector operations like matrix-vector multiplications, but also summation are subject to inaccuracies. This talk considered the question whether these inaccuracies in Arnoldi's method can be interpreted as backward error, that is, whether the resulting matrices $V_{k+1}, \underline{\mathrm{H}}_k$ are exact for a perturbed matrix $A + E$ for some matrix E (hopefully of small norm).

The talk consisted of two parts. In the first part we considered the shift-invert Arnoldi method which consists of computing

$$w_{i+1} = (A - \tau I)^{-1} v_i, \quad [v_{i+1}, h_i] = \text{orthonormalize}(w_{i+1}, V_i) \tag{1}$$

for $i = 1, 2, \ldots k$. After k steps the method returns an isometric matrix $V_{k+1} = [v_1, \ldots, v_{k+1}] \in \mathbb{C}^{n,k+1}$ and a Hessenberg matrix $\underline{\mathrm{H}}_k \in \mathbb{C}^{k+1,k}$ defined recursively by

$$\underline{\mathrm{H}}_k = \begin{bmatrix} \underline{\mathrm{H}}_{k-1} & h_k \\ 0 & \end{bmatrix}.$$

Considering the case that the linear systems in (1) are solved only approximately, we now assume

$$\|v_i - (A - \tau I)w_{i+1}\|_2 \leq \varepsilon_1 \tag{2}$$

for some tolerance level ε_1. This means that the linear systems are solved up to a residual of small norm. Then we can show that there is a backward error E_k (depending on the iteration count k) such that a Arnoldi relation for $A + E_k$,

$$(A + E_k - \tau I)^{-1} V_k = V_{k+1} \underline{\mathrm{H}}_k \tag{3}$$

holds with

$$\|E_k\| / \|A - \tau I\| = \mathcal{O}(\varepsilon_1).$$

In other words, *the shift-invert-Arnoldi method is backward stable with respect to small residuals in the linear systems.*

Unfortunately, a small residual is not always possible, e.g., when ε_1 is intended to be on the order of machine precision. So, we relaxed the assumption (2) to

$$\|v_i - (A - \tau I)w_{i+1}\|_2 \leq \varepsilon_1 \|v_i\|_2 + \varepsilon_2 \|(A - \tau I)w_{i+1}\|.$$

Additionally, we now allow inaccuracies in the orthonormalization phase of (1). More precisely, we assume that the obtained values of v_{i+1}, h_i fulfill

$$\|w_{i+1} - V_{i+1} h_i\|_2 \leq \varepsilon_3 \|w_{i+1}\|_2, \quad \kappa(V_k) \leq 1 + \varepsilon_4$$

for some $\varepsilon_3, \varepsilon_4$. It turns out that then (3) holds for some matrix E_k with

$$\|E_k\| / \|A - \tau I\| = \mathcal{O}(\varepsilon_1) + \mathcal{O}(\varepsilon_2 + \varepsilon_3 + \varepsilon_4)\kappa(\underline{\mathrm{H}}_k).$$

So, *with inaccurate orthogonalization the shift-invert Arnoldi method could be backwards-unstable when $\kappa(\underline{H}_k)$ is large*, e.g., if τ is close to an eigenvalue of A. This is work in progress, more details will be provided in [2].

The second part of the talk considered the standard Arnoldi method (i.e., without shift-invert transformation), but with a non-standard orthogonalization scheme called compensated Gram-Schmidt method (ComGS). In addition to V_k and H_k it constructs the matrix $D_k := V_k^H V_k$.

Algorithm 1. *Inexact Arnoldi method*

Input: $A \in \mathbb{C}^{n,n}, v_1 \in \mathbb{C}^n$ normed, $k \in \mathbb{N}$
Output: $V_{k+1} \in \mathbb{C}^{n,k+1}$, $H_k \in \mathbb{C}^{k,k}$, $h_{k+1,k} \in \mathbb{C}$, $D_{k+1} \in \mathbb{C}^{k+1,k+1}$

1: $V_1 = v_1$, $D_1 = 1$, $H_0 = [\,] \in \mathbb{C}^{0,0}$
2: **for** $i = 1, 2, 3, \ldots, k$ **do**
3: $\quad w_{i+1} = Av_i - f_{i+1}^{(M)}$
4: $\quad [v_{i+1}, h_{1:i,i}, h_{i+1,i}, D_{i+1}] =$ ComGS(w_{i+1}, V_i, D_i)
5: $\quad H_i = \begin{bmatrix} H_{i-1} & h_{1:i-1,i} \\ h_{i,i-1} e_{i-1}^T & h_{i,i} \end{bmatrix}$
6: $\quad V_{i+1} = [V_i, v_{i+1}]$
7: **end for**

Algorithm 2. *ComGS*

Input: $w_{i+1} \in \mathbb{C}^n$, $V_i \in \mathbb{C}^{n,i}$, $D_i \in \mathbb{C}^{i,i}$
Output: $v_{i+1} \in \mathbb{C}^n$, $h_{1:i,i} \in \mathbb{C}^i$, $h_{i+1,i} \in \mathbb{C}$, $D_{i+1} \in \mathbb{C}^{i+1,i+1}$

1: $s^{(0)} = D_i^{-1} V_i^H w_{i+1}$
2: $l_{i+1}^{(0)} = w_{i+1} - V_i s^{(0)} - f_{i+1}^{(0)}$
3: $s^{(1)} = D_i^{-1} V_i^H l_{i+1}^{(0)}$
4: $l_{i+1}^{(1)} = l_{i+1}^{(0)} - V_i s^{(1)} - f_{i+1}^{(1)}$
5: $h_{i+1,i} = \|l_{i+1}^{(1)}\|_2$, $h_{1:i,i} = s^{(0)} + s^{(1)}$
6: $v_{i+1} = (l_{i+1}^{(1)} - f_{i+1}^{(S)}) / h_{i+1,i}$
7: $D_{i+1} = \begin{bmatrix} D_i & V_i^H v_{i+1} \\ v_{i+1}^H V_i & v_{i+1}^H v_{i+1} \end{bmatrix}$

Here $f_i^{(M)}$, $f_i^{(0)}$, $f_i^{(1)}$,and $f_i^{(S)}$ model the perturbations in matrix-vector multiplication, in orthogonalization and in vector scaling, respectively. We assume that they

are bounded by $\|f_{i+1}^{(M)}\|_2 \le i\varepsilon\|A\|_2$, $\|f_{i+1}^{(0)}\|_2 \le i\varepsilon\|w_{i+1}\|_2$, $\|f_{i+1}^{(1)}\|_2 \le i\varepsilon\|l_{i+1}^{(0)}\|_2$, and $\|f_i^{(S)}\|_2 \le \varepsilon\|l_{i+1}^{(1)}\|_2$, respectively, where ε describes the level of accuracy of the vector operations.

To measure the distance to orthogonality of the basis V_k produced by the inexact Arnoldi method, Algorithm 1, the quantity $\|C_k - I_k\|_F$ is used where C_k is the Cholesky factor of D_k. We formulate the bounds for the cases with and without reorthogonalization (steps 3 and 4) within ComGS.

Theorem 1. *Let $\ell \in \{0,1\}$ be the number of reorthogonalization steps used in ComGS. Let $\kappa_k := \max_{i=1,\dots,k} \|h_{1:i,i}\|_2/h_{i+1,i}$ and $\kappa_0 = 0$. Then for sufficiently small $\varepsilon > 0$ the Cholesky factor C_k of D_k satisfies*

$$\|C_k - I_k\|_F \le \frac{\sqrt{k^5/2}\left(2 + (k\varepsilon)^\ell \kappa_{k-1}\right)\cdot\varepsilon}{1 - (\sqrt{k^5}(2 + (k\varepsilon)^\ell\kappa_{k-1}) + k(\ell+1) + 2)\varepsilon}.$$

The bound for $\|C_k - I_k\|_F$ is of the form $\alpha_1\varepsilon/(1-\alpha_2\varepsilon)$. Hence the bound is useful if $\max\{\alpha_1,\alpha_2\} \ll \varepsilon^{-1}$. This is the case whenever $\varepsilon^\ell\kappa_k$ is not large. Then, C_k differs from the identity by order ε, which means that Algorithm 1 generates an almost orthonormal basis V_k. Moreover, the algorithm provides implicitly an orthonormal basis of the search space by $\hat{V}_k := V_k C_k^{-1}$. It is implicit since building $\hat{V}_k$ would involve inaccurate vector additions.

We now com back to considering backward errors and restrict the scope to Hermitian matrices A. The output of Algorithm 1 satisfies the perturbed Arnoldi relation, $A\hat{V}_k = \hat{V}_k\hat{H}_k + \hat{v}_{k+1}\hat{h}_{k+1,k}e_k^T + \hat{F}_k$ where

$$\hat{V}_k = V_k C_k^{-1}, \begin{bmatrix}\hat{H}_k \\ \hat{h}_{k+1,k}e_k^T\end{bmatrix} := C_{k+1}\begin{bmatrix}H_k \\ h_{k+1,k}e_k^T\end{bmatrix} C_k^{-1}, \hat{F}_k := F_k C_k^{-1}.$$

In order to reformulate this as an unperturbed Krylov relation of a nearby matrix $A + E_k$, it turns out we have to replace H_k.

Theorem 2. *Let $A \in \mathbb{C}^{n,n}$ be Hermitian and $B_k \in \mathbb{C}^{k,k}$. Then there exists a Hermitian $E_k \in \mathbb{C}^{n,n}$ such that*

$$(A + E_k)\hat{V}_k = \hat{V}_k B_k + \hat{v}_{k+1}\hat{h}_{k+1,k}e_{k+1}^T.$$

holds if and only if B_k is Hermitian. In particular for B_k from the table

B_k	α	β
$\frac{1}{2}(H_k + H_k^H)$	$1+\sqrt{2}$	1
$\frac{1}{2}(\hat{H}_k + \hat{H}_k^H)$	$1+\sqrt{2}$	0
$\frac{1}{2}\mathrm{tridiag}(H_k + H_k^H)$	$2+\sqrt{2}$	2
$\frac{1}{2}\mathrm{tridiag}(\hat{H}_k + \hat{H}_k^H)$	$2+\sqrt{2}$	0

a corresponding E_k is bounded by $\|E_k\|_F \le \alpha\|\hat{F}_k\|_F + \beta\|\underline{\hat{H}}_k\|_F\frac{\zeta_k+\zeta_{k+1}}{1-\zeta_{k+1}}$ provided $\zeta_k := \|C_k - I_k\|_2 < 1$.

In conclusion we have shown that for Hermitian A the inexact Arnoldi method yields an exact Krylov relation of a nearby matrix $A + E_k$. Several choices are

possible for the small matrix B_k and we have proven bounds for the corresponding E_k.

Acknowledgment For proofs, details, numerical experiments etc. see [1]. This work is supported by deutsche Forschungsgemeinschaft, DFG under project ME 790/28-1 "Scalable Numerical Methods for Adiabatic Quantum Preparation".

References

[1] U. Kandler and C. Schröder, *Backward error analysis of an inexact Arnoldi method using a certain Gram Schmidt variant*, Preprint 10-2013, Institut f. Mathematik, TU Berlin, Germany, 2013, submitted.

[2] C. Schröder and L. Taslaman, *Backward error analysis of shift-and-invert Krylov methods*, in preparation.

A priori convergence analysis for inexact Hermitian Krylov methods

Ute Kandler

(joint work with Christian Schröder)

I real life applications typically only a small subset of eigenvalues and the corresponding invariant subspace of a large, sparse matrix is desired. Under these conditions the most prominent methods search for approximations of eigenvectors/invariant subspaces within Krylov subspaces

$$\mathcal{K}_k := \mathcal{K}_k(A, v_1) := \operatorname{span}(v_1, Av_1, A^2v_1, \ldots, A^{k-1}v_1)$$

of increasing dimension k.

Usually, an iterative numerical method for the Hermitian eigenvalue problem will employ the Lanczos process [2], resulting in an orthonormal basis $V_k = [v_1, \ldots, v_k]$ of $\mathcal{K}_k$, a $k \times k$ tridiagonal matrix T_k, and a scalar $t_{k+1,k}$ such that the Lanczos relation

$$AV_k = V_kT_k + v_{k+1}t_{k+1,k}e_k^H. \tag{1}$$

is satisfied. The Lanczos process consists mainly of matrix vector multiplications and orthogonalizations. Both may be inaccurate in multiple scenarios of practical interest like tensor computations (when $v_1, \ldots, v_k$, and A are high-dimensional tensors allowing only approximate operations), mixed precision arithmetic (when the computations are carried out in double precision, but the vectors are stored in single precision) or inexact solves in a shift-invert setting (when $A = (\mathcal{A} - \sigma I)^{-1}$). Often [1, 4] in these cases the perturbation can be interpreted as a backward error with respect to A so that a relation similar to (1) of the form

$$(A + E_k)\tilde{V}_k = \tilde{V}_kB_k + \tilde{v}_{k+1}b_{k+1}^H. \tag{2}$$

holds. Here, $\tilde{V}_k$ is still orthogonal, $b_{k+1} \in \mathbb{C}^k$, and $B = B^H \in \mathbb{C}^{k\times k}$ are known whereas $E_k = E_k^H \in \mathbb{C}^{n\times n}$ is unknown, but small in norm. Relation (2) implies that $\tilde{V}_k$ is a basis of a Krylov subspace

$$\tilde{\mathcal{K}}_k := \mathcal{K}_k(A + E_k, \tilde{v}_1)$$

of a Hermitian matrix $A + E_k$ close to A.

We use the relation (2) as a starting point for our a priori convergence analysis, i.e. we investigate how well an invariant subspace of A is contained in $\tilde{\mathcal{K}}_k$ which is a Krylov subspace of a perturbed matrix $A + E_k$. More precisely we consider the question: If l iterations have been performed without converging, how many more iteration steps of a Krylov subspace method are necessary to ensure convergence? Therefore we generalize a classic result of Saad [3, Theorem 6.3] that bounds of the angle between an eigenvector and the k-th Krylov subspace. As a suitable measure of the quality of an approximate invariant subspaces we use the angle between the exact and the approximated subspace since this measure does not depend on the choice of bases of the subspace. Our a priori bound constitutes a generalization of Saad's theorem in several respects:

i) The search space is chosen as Krylov subspace of a perturbed matrix $A + E_k$ (instead of A itself). Thus the setting of inexact Krylov methods is covered.

ii) Instead of just eigenvectors we consider invariant subspaces. This allows to treat clusters of eigenvalues as a whole.

iii) The eigenvalues corresponding to the subspace $\mathcal{X}$ need not be well separated from the remaining spectrum of A for the bound to be meaningful.

iv) The dimension l of the Krylov subspace on the right hand side is allowed to be larger than one. This is first of all necessary for the theorem to be meaningful, but also useful if information about the angle l-th Krylov subspace and the exact subspace are available.

To sum up, we provide a bound on how well the exact subspace $\mathcal{X}$ is contained in the search space $\tilde{\mathcal{K}}_k$. This bound is in terms of eigenvalues of A, their gaps and uses Chebychev polynomials. It is suitable for perturbations and a small gap between the desired and the remaining eigenvalues and the bound depends on how well the subspace $\mathcal{X}$ is contained in the search space $\tilde{\mathcal{K}}_l$ after l iterations.

Theorem 1. *Let $A \in \mathbb{C}^{n\times n}$ be Hermitian with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. Let $J_1 \subseteq J_2 \subseteq J_3 \subseteq J_4$ be nested nonempty subsets of $\{1, 2, \ldots, n\}$ such that $\max(J_3) < n$ and J_2, J_3 are intervals. Denote by $J_L := \{1, 2, \ldots, \min(J_3) - 1\}$ the leading and by $J_T := \{\max(J_3) + 1, \ldots, n\}$ the trailing indices. Let $\Lambda_i := \{\lambda_j : j \in J_i\}$ and $\Lambda_{-i} = \{\lambda_j : j \in \{1, \ldots n\} \setminus J_i\}$ for $i \in \{1, \ldots, 4, L, T\}$. Let $\mathcal{X}_1$ and $\mathcal{X}_4$ be invariant subspaces of A corresponding to Λ_1 and Λ_4, respectively. Let $k > \max(J_3)$. For $j = 1, \ldots, k$ let $\tilde{\mathcal{K}}_j := \mathcal{K}_j(A + E_k, \tilde{v}_1)$ for some Hermitian $E_k \in \mathbb{C}^{n\times n}$ and some $\tilde{v}_1 \in \mathbb{C}^n$. For $i \in \{2, 3\}$ let $\tilde{\mathcal{X}}_i$ be an invariant subspace of $A + E_k$ corresponding to $\{\lambda_j(A + E_k) : j \in J_i\}$. If $2\|E_k\|_2 < \min\{\mathrm{gap}(\Lambda_1, \Lambda_{-2}), \mathrm{gap}(\Lambda_3, \Lambda_{-4}), \mathrm{gap}(\Lambda_2, \Lambda_L \cup \Lambda_T)\}$ then for every $l = 1, \ldots, k - |J_L|$, we have that*

$$\begin{aligned}
\angle^{\frown}_{\max}\left(\mathcal{X}_1, \tilde{\mathcal{K}}_k\right) &\leq \angle^{\frown}_{\max}\left(\tilde{\mathcal{X}}_2, \tilde{\mathcal{K}}_k\right) + \delta_{12} \\
&\leq \arctan\left(\varrho_{kl} \cdot \tan \angle^{\frown}_{\max}(\tilde{\mathcal{X}}_3, \tilde{\mathcal{K}}_l)\right) + \delta_{12} \\
&\leq \arctan\left(\varrho_{kl} \cdot \tan_{\leq \frac{\pi}{2}}(\angle^{\frown}_{\max}(\mathcal{X}_4, \tilde{\mathcal{K}}_l) + \delta_{34})\right) + \delta_{12}
\end{aligned}$$

where $\tan_{\leq \frac{\pi}{2}}(\alpha) := \tan(\min\{\alpha, \frac{\pi}{2}\})$ *and*

$$\varrho_{kl} := \left(\sum_{i\in J_2} \frac{\tilde{\theta}_i^2}{\psi_{k-l-|\Lambda_L|}(1+2\tilde{\eta}_i)^2}\right)^{\frac{1}{2}}, \quad \delta_{ij} := \arctan\left(\frac{\|E_k\|_2}{\mathrm{gap}(\Lambda_i,\Lambda_{-j})-2\|E_k\|_2}\right),$$
$$\tilde{\theta}_i := \prod_{j\in J_L} \frac{|\lambda_j-\lambda_n|+2\|E_k\|_2}{|\lambda_j-\lambda_i|-2\|E_k\|_2} > 0, \qquad \tilde{\eta}_i := \frac{\mathrm{gap}(\lambda_i,\Lambda_T)-2\|E_k\|_2}{\mathrm{spread}(\Lambda_T)+2\|E_k\|_2} > 0,$$

where ψ_j *denotes the Chebychev polynomial of degree* j*.*

References

[1] U. Kandler and C. Schröder. Backward error analysis of an inexact Arnoldi method using a certain Gram Schmidt variant. Preprint 10-2013, TU Berlin, 2013.

[2] C. Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *Journal of Research of the National Bureau of Standards*, 45:255–282, 1950.

[3] Y. Saad. *Numerical methods for large eigenvalue problems.* Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, rev. ed. edition, 2011.

[4] V. Simoncini and D. B. Szyld. Theory of inexact Krylov subspace methods and applications to scientific computing SIAM J. Sci. Comput., 25 (2003), pp. 454–477 (electronic).

A posteriori error estimate for nonconforming approximation of multiple eigenvalues

Daniele Boffi

(joint work with Ricardo G. Durán, Francesca Gardini, Lucia Gastaldi)

We discuss a posteriori error estimates for the nonconforming finite element approximation of the eigenvalue problem associated with Laplace equation. This topic has been the object of the results presented in [4]. Here, we are particularly interested in the case when multiple eigenvalues are present.

To fix the notation, let us consider a Hilbert space V which is compactly embedded in a Hilbert space H (typically equal to $L^2(\Omega)$); we are given bilinear, continuous, and symmetric forms $a : V \times V \to \mathbb{R}$ and $b : H \times H \to \mathbb{R}$ so that our problem reads: find $\lambda \in \mathbb{R}$ such that for $u \in V$ it holds

$$a(u, v) = \lambda b(u, v) \quad \forall v \in V$$

Given $V_h \subset H$, a nonconforming approximation usually needs a discrete form $a_h : (V + V_h) \times (V + V_h) \to \mathbb{R}$ (typically constructed element by element) so that the discrete problem reads: find $\lambda_h \in \mathbb{R}$ such that for $u_h \in V_h$ it holds

$$a_h(u_h, v) = \lambda b(u_h, v) \quad \forall v \in V_h$$

In our case, $V = H_0^1(\Omega)$, the bilinear forms are $a(u, v) = (\mathbf{grad}\, u, \mathbf{grad}\, v)$, $b(u, v) = (u, v)$, $a_h = \sum_T (\mathbf{grad}\, u, \mathbf{grad}\, v)_T$, and V_h is the standard Crouzeix–Raviart nonconforming space.

We introduce the following estimators:

$$\eta_T = h_T \|\lambda_h u_h\|_{L^2(T)} \qquad \eta^2 = \sum_T \eta_T^2$$

$$\mu_T = \|\,\mathbf{grad}\,\tilde{u}_h - \mathbf{grad}_h\, u_h\|_{L^2(T)} \quad \mu^2 = \sum_T \mu_T^2$$

Here $\tilde{u}_h$ is the *conforming* p.w. $\mathcal{P}_1$ function obtained by averaging the values of u_h at the vertices of the triangulation.

In [4] it is shown that the following estimates hold:

$$\|\,\mathbf{grad}_h(u - u_h)\|_0 \le \mu + C\eta + \text{h.o.t.}$$
$$\eta_T \le C\|\,\mathbf{grad}_h(u - u_h)\|_{L^2(\mathcal{T}^*)} + \text{h.o.t.}$$
$$\mu_T \le C\|\,\mathbf{grad}_h(u - u_h)\|_{L^2(\mathcal{T})}$$

As usual, $\mathcal{T}$ refers to the mesh and $\mathcal{T}^*$ to the elements of the mesh in a small neighborhood of T.

It is clear that the way these estimates are written makes them only useful in the case of simple eigenvalues. In particular, expressions like $u - u_h$ make little sense in the case of multiple eigenvalues. In this case, the gap between subspaces should be used, by adopting expressions like $\mathrm{dist}_a(u_h, E_\lambda)$ or $\mathrm{dist}_b(u_h, E_\lambda)$. This is what has been done, for instance, in [6] where results by [5] and [7] have been generalized. A more detailed discussion on the approximation of multiple eigenvalues has been performed in [8]; see also [1], [2], and [3].

Concerning the links between the introduced a posteriori estimator and the approximation of the eigenvalues, in [4] the analysis is performed only for the first *simple* eigenvalue. Recently, by using techniques borrowed by [8], we extended the analysis to general (possibly multiple) eigenvalues. In particular, we obtained the following result.

Theorem. Let λ_i be an eigenvalue of multiplicity q and suppose that we have

$$\lambda_{i-1} < \lambda_i = \cdots = \lambda_{i+q-1} < \lambda_{i+q}$$

Let $\lambda_{i,h} \le \cdots \le \lambda_{i+q-1,h}$ be approximations of λ_i so that

$$\lambda_{j,h} \le \lambda_i \quad \text{for } j = i, \ldots, i+q-1$$

Then we have

$$\frac{\lambda_i - \lambda_{j,h}}{\lambda_i} \le \|(I - P_h^C + P^C_{1,\ldots,i-1,h})P_{i,\ldots,j,h}\|^2$$

Here $P_{(\cdot)}$ and $P^C_{(\cdot)}$ refer to suitable elliptic projections onto the space of non-conforming, respectively conforming, finite element subspaces, in the spirit of [8]. It turns out that the right hand side of the final estimate can be directly related to the introduced estimator.

References

[1] D. Boffi, *Finite element approximation of eigenvalue problems*, Acta Numerica **19** (2010), 1–120.

[2] D. Boffi, F. Gardini, L. Gastaldi, *Some remarks on eigenvalue approximation by finite elements*, in Frontiers in Numerical Analysis - Durham 2010, Springer Lecture Notes in Computational Science and Engineering **85** (2012), 1–77.

[3] D. Boffi, L. Gastaldi, *Some remarks on finite element approximation of multiple eigenvalues*, Appl. Numer. Math., to appear.

[4] E.A. Dari, R.G. Durán, C. Padra, *A posteriori error estimates for non-conforming approximation of eigenvalue problems*, Appl. Numer. Math. **62** (2012), 580–591.

[5] R.G. Durán, C. Padra, R. Rodríguez, *A posteriori error estimates for the finite element approximation of eigenvalue problems*, Math. Models Methods Appl. Sci. **13** (2003), 1219–1229.

[6] E.M. Garau, P. Morin, *Convergence and quasi-optimality of adaptive FEM for Steklov eigenvalue problems*, IMA J. Numer. Anal. **31** (2011), 914–946.

[7] S. Giani, I.G. Graham, *A convergent adaptive method for elliptic eigenvalue problems*, SIAM J. Numer. Anal. **47** (2009), 1067–1091.

[8] A. Knyazev, J.E. Osborn, *New a priori FEM error estimates for eigenvalues*, SIAM J. Numer. Anal. **43** (2006), 2647–2667.

[9] P. Solin, S. Giani, *An iterative adaptive finite element method for elliptic eigenvalue problems*, J. Comput. Appl. Math. **236** (2012), 4582–4599.

Two-Grid Methods for Maxwell Eigenvalue Problems

Long Chen

(joint work with Xiaozhe Hu, Shi Shu, Liuqiang Zhong, Jie Zhou)

We develop an efficient algorithm for computing the Maxwell eigenvalue problem, which is a basic and important computational model in computational electromagnetism, in regard to electromagnetic waveguides and resonances in cavities. The governing equations are

$$\operatorname{curl}(\mu_r^{-1}\operatorname{curl}\boldsymbol{u}) = \omega^2\varepsilon_r\boldsymbol{u} \quad \text{in } \Omega, \tag{1}$$

$$\operatorname{div}(\varepsilon_r\boldsymbol{u}) = 0 \quad \text{in } \Omega, \tag{2}$$

$$\gamma_{\boldsymbol{t}}\boldsymbol{u} = 0 \quad \text{on } \partial\Omega, \tag{3}$$

where $\Omega \subset \mathbb{R}^n (n = 2, 3)$ is a bounded Lipschitz polyhedron domain and $\gamma_{\boldsymbol{t}}\boldsymbol{u}$ is the tangential trace of $\boldsymbol{u}$. The coefficients μ_r and ε_r are the real relative magnetic permeability and electric permittivity, respectively, that satisfy the Lipschitz continuous condition, whereas ω is the resonant angular frequency of the electromagnetic wave for cavity Ω. In the sequel, we will use the conventional notation λ to replace ω^2.

We focus on speeding up the inverse or Rayleigh quotient iterations by using a two-grid approach. The two-grid method for elliptic eigenvalue problems [3] is developed by Xu and Zhou. The main idea is to reduce the solution of an eigenvalue problem on a given fine grid with mesh size h to the solution of the same eigenvalue problem on a much coarser grid with mesh size $H \gg h$, which can be easily solved as the size of the discrete eigenvalue problem is significantly

smaller than the original eigenvalue problem on the fine grid, and the solution of a linear problem on the same fine grid, which can be solved by mature and efficient numerical algorithms.

It is important to note that the standard two-grid method (Xu and Zhou [3]) for elliptic eigenvalue problems works when the order of error in the L^2 norm is one order higher than the error in the energy norm. In terms of approximation of Maxwell equations, it is known that establishing an L^2 norm error estimate is a very challenging task. For example, for the first family edge element, we cannot expect the error in the L^2 norm has higher convergence rate than the error in the energy norm. As a result, in order to make the two-grid algorithm work, on the fine grid, we must solve a linear Maxwell equation derived from the shifted inverse iteration (or Newton's method). This idea is proposed in [2, 4] as an acceleration scheme for the standard two-grid method of elliptic eigenvalue problems.

We adopt this idea and develop efficient two-grid methods for solving the Maxwell eigenvalue problem. Specifically, we first solve a Maxwell eigenvalue problem on a coarse grid, and then solve a linear Maxwell equation on a fine grid. Essentially, the procedure is similar to performing only one step of a Rayleigh quotient iteration. We present our algorithm below.

(1) Solve a Maxwell eigenvalue problem on the coarse grid $\mathcal{T}_H$:
Find $(\lambda_H, \boldsymbol{u}_H) \in \mathbb{R} \times \boldsymbol{V}_H$ and $\boldsymbol{u}_H \not\equiv 0$ satisfying

$$a(\boldsymbol{u}_H, \boldsymbol{v}_H) = \lambda_H b(\boldsymbol{u}_H, \boldsymbol{v}_H), \quad \text{for all } \boldsymbol{v}_H \in \boldsymbol{V}_H.$$

(2) Solve an indefinite Maxwell equation on the fine grid $\mathcal{T}_h$:
Find $\boldsymbol{u}^h \in \boldsymbol{V}_h$ such that

$$a(\boldsymbol{u}^h, \boldsymbol{v}_h) - \lambda_H b(\boldsymbol{u}^h, \boldsymbol{v}_h) = b(\boldsymbol{u}_H, \boldsymbol{v}_h), \quad \text{for all } \boldsymbol{v}_h \in \boldsymbol{V}_h.$$

(3) Use the Rayleigh quotient to compute the approximate eigenvalue on the fine grid:

$$\lambda^h = \frac{a(\boldsymbol{u}^h, \boldsymbol{u}^h)}{b(\boldsymbol{u}^h, \boldsymbol{u}^h)}.$$

Several non-trivial theoretical and practical issues must be addressed when generalizing the two-grid approach to the Maxwell eigenvalue problems. First for the shifted inverse iteration, we need to solve an indefinite and nearly singular Maxwell equation on the fine grid. It is difficult to solve this equation such that very efficient solvers are required. We will use the preconditioned GMRES and the HX preconditioner [1] for the corresponding definite linear equation. Because we are interested in small eigenvalues, the wave number of this indefinite Maxwell problem is relatively small. Our numerical computation shows that the solver converges in a few steps and that the solver is almost uniform to the size of the problem.

Another problem introduced by the shifted inverse iteration is the divergence-free constraint which only holds weakly on the coarse grid. It is possible to explicitly impose this constraint in the fine grid by projecting the obtained approximated eigenfunction on the coarse grid to the discrete divergence-free space on the fine grid by solving an extra Poisson equation. However, our analysis, which is based on the Helmholtz decomposition and an estimate of the differences between the

weakly divergence-free functions on coarse and fine grids, show that even without the projection step, our two-grid method produce an approximation λ^h to λ, and remain asymptotically convergence rate for $H^3 = h$ when the domain is smooth or convex.

Theroem. *Let $(\lambda^h, \boldsymbol{u}^h)$ be computed by our two grid method, and under the assumptions the coarse grid size H is small enough, then there exists an eigenfunction $\boldsymbol{u} \in M(\lambda)$ such that*

$$\min_{\alpha\in\mathbb{R}} \|\boldsymbol{u} - \alpha\boldsymbol{u}^h\|_{L^2} \le C(h^{1/2+\delta} + H^{3(1/2+\delta)}), \tag{4}$$

$$\min_{\alpha\in\mathbb{R}} \|\boldsymbol{u} - \alpha\boldsymbol{u}^h\|_{\mathrm{curl}} \le C(h^{1/2+\delta} + H^{3(1/2+\delta)}). \tag{5}$$

And for the eigenvalue, we have

$$|\lambda - \lambda^h| \le C(h^{1+2\delta} + H^{3(1+2\delta)}), \tag{6}$$

where the constant C and $0 < \delta \le 1/2$ depend only on μ_r, ε_r, ρ, λ, and $\boldsymbol{u}$.

When Ω is smooth or convex, we have $\delta = 1/2$ and

$$\min_{\alpha\in\mathbb{R}} \|\boldsymbol{u} - \alpha\boldsymbol{u}^h\|_{L^2} \le C(h + H^3),$$

$$\min_{\alpha\in\mathbb{R}} \|\boldsymbol{u} - \alpha\boldsymbol{u}^h\|_{\mathrm{curl}} \le C(h + H^3),$$

$$|\lambda - \lambda^h| \le C(h^2 + H^6).$$

Note that $H^3 = h$ implies that a very coarse mesh can be used–which saves considerable computational cost and time, especially in three dimensions. For example, for a three-dimensional unit cube, $h = 1/64$, the number of unknowns is $1,872,064$ whereas for $H = h^{1/3} = 1/4$ there are only 604 unknowns.

References

[1] R. Hiptmair, J. Xu, Nodal auxiliary space preconditioning in H(curl) and H(div) spaces. *SIAM J. Numer. Anal*, 45(6): 2483-2509, 2007.

[2] X. Hu and X. Cheng. Acceleration of a two-grid method for eigenvalue problems. *Math. Comp*, 80(275):1287–1301, 2011.

[3] J. Xu and A. Zhou. A two-grid discretization scheme for eigenvalue problems. *Math. Comp.*, 70(233):17–25, 2001.

[4] Y. Yang and H. Bi, Two-grid finite element discretization schemes based on shifted-inverse power method for elliptic eigenvalue problems. *SIAM J. Numer. Anal*, 49(4):1602–1624, 2011.

Self-adjoint Curl-Operators

Ralf Hiptmair

(joint work with P.R. Kotiuga, S. Tordeux)

Force free magnetic fields. According to Ampere's law a stationary current $\mathbf{j} : \Omega \to \mathbb{R}^3$, $\Omega \subset \mathbb{R}^3$ bounded, spawns a magnetic field $\mathbf{H}$ that satisfies $\mathbf{curl}\,\mathbf{H} = \mathbf{j}$. Since the Lorenz force is proportional to $\mathbf{j} \times \mathbf{H}$, magnetic fields that do not exert a force onto the moving charges causing $\mathbf{j}$ must fulfill $\mathbf{curl}\,\mathbf{H} = \alpha(\boldsymbol{x})\mathbf{H}$ with some

scalar function $\alpha : \Omega \to \mathbb{R}$. If α is constant, we face the linear eigenvalue equation $\mathbf{curl}\,\mathbf{H} = \alpha\mathbf{H}$ for the **curl**-operator and its solutions are called (linear) *Beltrami fields*, first studied by E. Beltrami in a fluid dynamics context. Beltrami fields also play a role in the study of stable plasmas [6, 15, 1, 7, 17]. They possess fascinating topological properties [5].

Eigenvalue problem for curl. Understanding the eigenvalue problem $\mathbf{curl}\,\mathbf{H} = \alpha\mathbf{H}$ entails understanding the spectral properties of the *unbounded operator* **curl** on $(L^2(\Omega))^3$. This approach was pioneered by R. Picard [18, 19, 20], see also [22], and a key step is to specify the domain $\mathcal{D}(\mathbf{curl})$ by imposing suitable *boundary conditions*. Of particular interest are boundary conditions that render the associated **curl**-operator self-adjoint with compact resolvent. Then abstract spectral theory of unbounded operators tells us that the eigenfunctions of this **curl** provide an orthonormal basis of $(L^2(\Omega))^3$ and its eigenvalue form discrete sequence accumulating at $\pm\infty$.

Self-adjoint extensions. The Green's formula for **curl**

$$\int_\Omega \mathbf{curl}\,\mathbf{u}\cdot\mathbf{v} - \mathbf{curl}\,\mathbf{v}\cdot\mathbf{u}\,\mathrm{d}\mathbf{x} = \int_{\partial\Omega} (\mathbf{u}\times\mathbf{v})\cdot\mathbf{n}\,\mathrm{d}S\,, \tag{1}$$

shows that **curl** is a symmetric operator on the sense subset $(C_0^\infty(\Omega))^3 \subset (L^2(\Omega))^3$. In an abstract fashion all possible self-adjoint extensions are accessible through the Glazman-Krein-Naimark Theorem [10, 11]. To apply it we equip the tangential trace space of $\boldsymbol{H}(\mathbf{curl},\Omega)$, usually denoted by $\boldsymbol{H}^{-\frac{1}{2}}(\mathbf{curl}_\Gamma,\partial\Omega)$ [4], with the symplectic form $[\boldsymbol{\nu},\boldsymbol{\mu}] := \int_{\partial\Omega}(\boldsymbol{\nu}(\boldsymbol{x})\times\boldsymbol{\mu}(\boldsymbol{x}))\cdot\mathbf{n}\,\mathrm{d}S(\boldsymbol{x})$, where $\boldsymbol{n}$ is the exterior unit normal of $\partial\Omega$. Then there is a *one-to-one* mapping between *complete Lagrangian subspaces* L of the symplectic space $(\boldsymbol{H}^{-\frac{1}{2}}(\mathbf{curl}_\Gamma,\partial\Omega),[\cdot,\cdot])$ and *self-adjoint extensions* of **curl**, characterized through their domains. This mapping is given by $L \mapsto \mathcal{D}_L(\mathbf{curl})$, $\mathcal{D}_L(\mathbf{curl}) := \{\mathbf{u}\in\boldsymbol{H}(\mathbf{curl},\Omega) : \mathbf{u}_{\boldsymbol{t}}\in L\}$, where $\mathbf{u}_{\boldsymbol{t}}$ denotes the tangential component of $\mathbf{u}$.

Remark. A fascinating property of the symplectic form $[\cdot,\cdot]$ is its invariance under continuous deformations of Ω. It has its roots in the fact that the *curl*-operator incarnates the exterior derivative for 1-forms in three dimensions. This makes the set of all self-adjoint extensions of **curl** a topological invariant of Ω.

Special Lagrangian subspaces of $(\boldsymbol{H}^{-\frac{1}{2}}(\mathbf{curl}_\Gamma,\partial\Omega),[\cdot,\cdot])$. The starting point is the "$L^2(\partial\Omega)$-orthogonal" *Hodge decomposition*

$$\boldsymbol{H}^{-\frac{1}{2}}(\mathbf{curl}_\Gamma,\partial\Omega) = \mathbf{grad}_\Gamma H^{\frac{1}{2}}(\partial\Omega)\oplus\mathbf{curl}_\Gamma H^{\frac{3}{2}}(\partial\Omega)\oplus\mathcal{H}^1(\partial\Omega)\,, \tag{2}$$

where $\mathcal{H}^1(\partial\Omega)$ is the space of harmonic tangential vector fields on $\partial\Omega$, whose dimension is 2β, $\beta\in\mathbb{N}_0$ standing for the first Betti number of Ω. It turns out that $(\mathcal{H}^1(\partial\Omega),[\cdot,\cdot])$ is a symplectic space as well, of which a canonical basis can be built as follows: denote by $\{\gamma_1,\ldots,\gamma_\beta,\gamma_1',\ldots,\gamma_\beta'\}$ a set of fundamental cycles of $\partial\Omega$, for which the γ_j are bounding w.r.t. $\mathbb{R}^3\setminus\Omega$, whereas the γ_j' are bounding

w.r.t. Ω. Moreover, γ_j and γ_j' are dual to each other [12, Ch. 5]. Then

$$\mathcal{L}_{\mathcal{H}} := \left\{ \{ \boldsymbol{\eta} \in \mathcal{H}^1(\partial\Omega) : \int_{c_j} \boldsymbol{\eta} \cdot \mathrm{d}s = 0 \, , c_j \in \{\gamma_j, \gamma_j'\}, \, j = 1, \ldots, \beta \} \right\} \tag{3}$$

provides all complete Lagrangian subspaces of $\mathcal{H}^1(\partial\Omega)$ [14, Sect. 6.3].

Theorem 1 ([14, Sect. 6])**.** *If $L \in \mathcal{L}_{\mathcal{H}}$, then*

(i) $\mathbf{grad}_\Gamma H^{\frac{1}{2}}(\partial\Omega) + L$ *(closed traces), and*
(ii) $\mathbf{curl}_\Gamma H^{\frac{3}{2}}(\partial\Omega) + L$ *(co-closed traces)*

are complete Lagrangian subspaces of $(\boldsymbol{H}^{-\frac{1}{2}}(\mathbf{curl}_\Gamma, \partial\Omega), [\cdot,\cdot])$.

Consequently, for $\beta = 0$, extension of **curl** to the following domains

$$\text{(i)} \quad \mathcal{D}_0 := \{ \mathbf{v} \in \boldsymbol{H}(\mathbf{curl}, \Omega) : \operatorname{curl}_\Gamma \mathbf{v}_{\boldsymbol{t}} = 0 \text{ on } \partial\Omega \} \, ,$$

$$\text{(ii)} \quad \mathcal{D}_\perp := \{ \mathbf{v} \in \boldsymbol{H}(\mathbf{curl}, \Omega) : \int_{\partial\Omega} \mathbf{v}_{\boldsymbol{t}} \cdot \mathbf{grad}_\Gamma \varphi \, \mathrm{d}S = 0 \; \forall \varphi \in H^{\frac{1}{2}}(\partial\Omega) \} \, .$$

will be self-adjoint.

Spectral properties. The self-adjoint **curl**-operators are invertible on the $L^2(\Omega)$-orthogonal complements of their kernels, all of which contain $\mathbf{grad}\, H_0^1(\Omega)$. Thanks to well-known compact embedding results like [13, Sect. 4.1]

$$\{ \mathbf{v} \in \boldsymbol{H}(\mathbf{curl}, \Omega) : \operatorname{div} \mathbf{v} = 0, \, \mathbf{v} \cdot \mathbf{n} = 0 \} \hookrightarrow (L^2(\Omega))^3 \, ,$$

$$\{ \mathbf{v} \in \boldsymbol{H}(\mathbf{curl}, \Omega) : \operatorname{div} \mathbf{v} = 0, \, \operatorname{div}_\Gamma \mathbf{v}_{\boldsymbol{t}} = 0 \} \hookrightarrow (L^2(\Omega))^3 \, ,$$

their inverses are *compact*. Therefore from classical spectral theory we conclude that self-adjoint **curl**-operators have a pure point spectrum with $\pm\infty$ as sole accumulation points and that they possess a complete $L^2(\Omega)$-orthonormal system of eigenfunctions.

Numerical approximation. Approaches to the numerical computation of Beltrami fields [21, 2, 3] have mainly considered the self-adjoint **curl**-operator $\mathbf{curl}_{|\mathcal{D}_0}$ based on closed traces. Throughout, they relied on the squaring approach and converted the eigenvalue problem for **curl** into (mixed) variational eigenvalue problems for $\mathbf{curl}\,\mathbf{curl}$ [21, Sect. 3], e.g., seek $\mathbf{H} \in \mathcal{D}_0$

$$\int_\Omega \mathbf{curl}\, \mathbf{H} \cdot \mathbf{curl}\, \mathbf{v} \, \mathrm{d}\boldsymbol{x} = \alpha^2 \int_\Omega \mathbf{H} \cdot \mathbf{v} \, \mathrm{d}\boldsymbol{x} \quad \forall \mathbf{v} \in \mathcal{D}_0 \, . \tag{4}$$

After Galerkin discretization the eigenfunctions have to be assigned to eigenvalues $\pm|\alpha|$ in a post-processing step. Numerical analysis of the discretized variational problem (4) can be based on the theory from [8, 9].

A direct Galerkin discretization of the eigenvalue problem for the self-adjoint **curl**-operator based on closed traces reads (for Ω topologically equivalent to a ball, $\beta = 0$): seek $\mathbf{H}_h \in V_h := \mathcal{W}_0^1(\Omega_h) + \mathbf{grad}\, \mathcal{W}_\partial^0(\Omega_h)$ such that

$$\int_\Omega \mathbf{curl}\, \mathbf{H}_h \cdot \mathbf{v}_h \, \mathrm{d}\boldsymbol{x} = \alpha \int_\Omega \mathbf{H}_h \cdot \mathbf{v}_h \, \mathrm{d}\boldsymbol{x} \quad \forall \mathbf{v}_h \in V_h \, . \tag{5}$$

Here $\mathcal{W}_0^1(\Omega_h) \subset \boldsymbol{H}_0(\mathbf{curl}, \Omega)$ is a space of lowest order edge finite elements on a (curvilinear) simplicial triangulation of Ω, and $\mathcal{W}_\partial^0(\Omega_h)$ is spanned by piecewise linear nodal basis functions associated with nodes on $\partial\Omega$. The numerical analysis of (5) is still open. Even more so the analysis of Galerkin approximations of eigenvalue problems for $\mathbf{curl}_{|\mathcal{D}_\perp}$ involving co-closed traces, where the boundary condition has to imposed as a linear constraint, *cf.* the definition of $\mathcal{D}_\perp$.

Remark. The Galerkin matrix arising from the left-hand-side of (5), also called the helicity matrix [16], does depend only on the topology of the triangulation. This property is closely related to the homeomorphic invariance of the symplectic form $[\cdot, \cdot]$ mentioned above.

References

[1] V. Arnold and B. Khesin, *Topological Methods in Hydrodynamics*, vol. 125 of Applied Mathematical Sciences, Springer, New York, 1998.

[2] T. Boulmezaoud and T. Amari, *Approximation of linear force-free fields in bounded 3-d domains*, Math. Comput. Modelling, 31 (2000), pp. 109–129.

[3] T.-Z. Boulmezaoud, Y. Maday, and T. Amari, *On the linear force-free fields in bounded and unbounded three-dimensional domains*, Math. Model. Numer. Anal., M2AN, 33 (1999), pp. 359–393.

[4] A. Buffa, M. Costabel, and D. Sheen, *On traces for* $\mathbf{H}(\mathbf{curl}, \Omega)$ *in Lipschitz domains*, J. Math. Anal. Appl., 276 (2002), pp. 845–867.

[5] J. Cantarella, *Topological Structure of Stable Plasma Flows*, PhD thesis, University of Pennsylvania, Philadelphia, PA, 1999.

[6] S. Chandrasekhar and P. Kendall, *On force-free magnetic fields*, Astrophysical Journal, 126 (1957), pp. 457–460.

[7] J. Crager and P. Kotiuga, *Cuts for the magnetic scalar potential in knotted geometries and force-free magnetic fields*, IEEE Trans. Magnetics, 38 (2002), pp. 1309–1312.

[8] J. Descloux, N. Nassif, and J. Rappaz, *On spectral approximation. Part I. The problem of convergence*, R.A.I.R.O. Numerical Analysis, 12 (1978), pp. 97–112.

[9] ———, *On spectral approximation. Part II. Error estimates for the Galerkin method*, R.A.I.R.O. Numerical Analysis, 12 (1978), pp. 113–119.

[10] W. Everitt and L. Markus, *Complex symplectic geometry with applications to ordinary differential equations*, Trans. American Mathematical Society, 351 (1999), pp. 4905–4945.

[11] ———, *Complex symplectic spaces and boundary value problems*, Bull. Amer. Math. Soc., 42 (2005), pp. 461–500.

[12] P. Gross and P. Kotiuga, *Electromagnetic Theory and Computation: A Topological Approach*, vol. 48 of Mathematical Sciences Research Institute Publications, Cambridge University Press, Cambridge, UK, 2004.

[13] R. Hiptmair, *Finite elements in computational electromagnetism*, Acta Numerica, 11 (2002), pp. 237–339.

[14] R. Hiptmair, P. Kotiuga, and S. Tordeux, ***Self-adjoint curl operators*, Annali di Matematica Pura ed Applicata, 191 (2012), pp. 431–457.**

[15] A. Jette, *Force-free magnetic fields in resistive magnetohydrostatics*, J. Math. Anal. Appl., 29 (1970), pp. 109–122.

[16] P. Kotiuga, *Analysis of finite element matrices arising from discretizations of helicity functionals*, J. Appl. Phys., 67 (1990), p. 5815.

[17] ———, *Topology-based inequalities and inverse problems for near force-free magnetic fields*, IEEE. Trans. Magnetics, 40 (2004), pp. 1108–1111.

[18] R. Picard, *Ein Randwertproblem in der Theorie kraftfreier Magnetfelder*, Z. Angew. Math. Phys., 27 (1976), pp. 169–180.

[19] ———, *On a selfadjoint realization of curl and some of its applications*, Riceche di Matematica, XLVII (1998), pp. 153–180.
[20] ———, *On a selfadjoint realization of curl in exterior domains*, Mathematische Zeitschrift, 229 (1998), pp. 319–338.
[21] R. Rodriguez and P. Venegas, *Numerical approximation of the spectrum of the curl operator*, Math. Comp., 83 (2014), pp. 553–577.
[22] Z. Yoshida and Y. Giga, *Remarks on spectra of operator rot*, Math. Z., 204 (1990), pp. 235–245.

Constructing both lower and upper bounds of eigenvalues by nonconforming finite element methods

Jun Hu
(joint work with Yunqing Huang, Rui Ma, Qun Lin, Quan Shen)

The first aim of this talk is to introduce a new systematic method that can produce lower bounds for eigenvalues. The main idea is to use nonconforming finite element methods. The conclusion is that if local approximation properties of nonconforming finite element spaces are better than total errors (sums of global approximation errors and consistency errors) of nonconforming finite element methods, corresponding methods will produce lower bounds for eigenvalues. More precisely, under three conditions on continuity and approximation properties of nonconforming finite element spaces we analyze abstract error estimates of approximate eigenvalues and eigenfunctions. Subsequently, we propose one more condition and prove that it is sufficient to guarantee nonconforming finite element methods to produce lower bounds for eigenvalues of symmetric elliptic operators. We show that this condition hold for most low–order nonconforming finite elements in literature. In addition, this condition provides a guidance to modify known nonconforming elements in literature and to propose new nonconforming elements. In fact, we enrich locally the Crouzeix-Raviart element such that the new element satisfies the condition; we also propose a new nonconforming element for second order elliptic operators and prove that it will yield lower bounds for eigenvalues. Finally, we prove the saturation condition for most nonconforming elements. We also present a guidance for how to design the nonconforming finite element methods which can produce the lower bounds for the eigenvalues of the elliptic operators.

The second aim of the talk is to, based on such nonconforming discrete eigenfunctions, propose a simple method to produce the upper bounds of the eigenvalues. More precisely, we construct a conforming approximation of the exact eigenfunction by the projection average interpolation of the nonconforming discrete eigenfunction. After showing the approximation property of the projection average interpolation, we prove that the Rayleigh–quotient of the aforementioned conforming approximation is convergent to the exact eigenvalues from above. Finally, we combine the lower and upper bounds of the eigenvalues to obtain a high accuracy approximation of the eigenvalues. Numerical examples verify our theoretical results.

References

[1] J. Hu, Y. Q. Huang, H. M. Shen. The lower approximation of eigenvalues by lumped mass finite element method, J.Comput. Math.2004(22), pp 545-556.

[2] J. Hu, Y. Q. Huang and Q. Lin. The lower bounds for eigenvalues of elliptic operators-by nonconforming finite element methods. arXiv:1112.1145v1[math.NA], 2011.

[3] J. Hu and Y. Q. Huang. The correction operator for the canonical interpolation operator of the Adini element and the lower bounds of eigenvalues, Science in China:Mathematics 55(2012),187-196.

[4] J. Hu and Y. Q. Huang. Lower Bounds for Eigenvalues of the Stokes Operator, Adv. Appl. Math. Mech., 5(2013), pp. 1–18.

[5] J. Hu, Y. Q. Huang and Q. Shen. A high accuracy post-processing algorithm for the eigenvalues of elliptic operators. Journal of Scientific Computing, 2011, (DOI)10.1007/s10915-011-9552-9.

[6] J. Hu, Y. Q. Huang and Q. Shen. Constructing both Lower and Upper Bounds for the Eigenvalues of the Elliptic Operators by the nonconforming finite element methods, preprint 2012.

Cluster robust estimates for eigenvalues and eigenfunctions of convection–diffusion–reaction operators

Luka Grubišić

(joint work with Stefano Giani, Agnieszka Miedlar and Jeffrey S. Ovall)

We present a collection of direct residual error estimates for finite element approximations of eigenvalues and eigenfunctions of linear convection–diffusion–reaction operators in bounded polygonal domains $\Omega \subset \mathbb{R}^2$, as given by the formal differential expression

$$\mathcal{A}\psi := -\nabla \cdot A\nabla\psi + b \cdot \nabla\psi + c\psi = \lambda\psi \ . \tag{1}$$

In general we assume that $A \in \left[L^\infty(\Omega)\right]^{2\times 2}$, $b \in \left[L^\infty(\Omega)\right]^2$ with $\nabla \cdot b \in L^\infty(\Omega)$, and $c \in L^\infty(\Omega)$. Our model problems will be mostly form the class of eigenvalue problems which include Fokker-Planck operators of the spectral type, see [4]. This is a representable class of analytically well understood benchmark model problems which will be used to test numerical procedures. We note that their numerical analysis is also a challenging problem in its own right. We concentrate on features relevant for testing numerical procedures.

Example 1. *Let $\mathcal{A}u := -\nabla \cdot (A\nabla u) + b \cdot \nabla u + cu$, where the coefficients A, b and c satisfy the conditions above. Define the multiplication operator $\mathcal{X}u := e^\beta u$ for some function $\beta \in W^{1,\infty}(\Omega)$. If $A^{-1}b$ is a conservative vector field, then we choose β such that $\nabla\beta = \frac{1}{2}A^{-1}b$, and determine that*

$$\mathcal{H}u = \mathcal{X}^{-1}\mathcal{A}\mathcal{X}u = -\nabla \cdot (A\nabla u) + \left(c - \frac{1}{2}\nabla \cdot b + \frac{1}{4}b \cdot (A^{-1}b)\right) u \ , \tag{2}$$

is self adjoint. From this argument we also deduce that (λ, ϕ) is an eigenpair of $\mathcal{H}$ if and only if $(\lambda, e^\beta\phi, e^{-\beta}\phi)$ is an eigentriple of $\mathcal{A}$.

In what follows we call the operator $\mathcal{A}$, which is similar in the sense of (2) to a (in general) normal operator, a *diagonalizable* operator.

Residual. For a scalar μ and $\phi \in H_0^1(\Omega) \setminus \{0\}$ we define the residual functional $\mathfrak{r}(\mu)[\psi,\cdot] = B(\psi,\cdot) - \mu(\psi,\cdot)$ and use $\|\mathfrak{r}(\mu)[\psi,\cdot]\|_{-1}$ to denote its negative order Sobolev norm. Let now $\Psi := \{\psi_1, \cdots, \psi_n\} \subset H_0^1(\Omega)$ be a linearly independent set and let μ_i, $i = 1, \cdots, n$ be given. By P we denote the L^2 orthogonal projection onto Ψ and Q denotes the orthogonal projection onto an isolated component of the spectrum which consists of semisimple eigenvalues of joint multiplicity n. Then we show

$$\min_{\xi \in \mathrm{Spec}(\mathcal{A})} \frac{|\mu_i - \xi|}{\sqrt{|\mu_i|\xi}} \leq O\Big(\frac{\|\mathfrak{r}(\mu_i)[\psi_i,\cdot]\|_{-1}}{\sqrt{\widehat{\mu}}}\Big)\,, \tag{3}$$

$$\min_{\phi \in \mathrm{Ran}(Q)} \|\phi - \psi_i\|_1 \leq O\big(\|\mathfrak{r}(\mu_i)[\psi_i,\cdot]\|_{-1}\big)\,, \tag{4}$$

$$\|Q - P\|_{HS} \leq O\Big(\sqrt{\sum_{i=1}^{n} \|\mathfrak{r}(\mu_i)[\psi_i,\cdot]\|_{-1}^2}\Big). \tag{5}$$

Here $\|\cdot\|_{HS}$ denotes the Hilbert-Schmidt operator norm. Note that we have placed no Galerkin orthogonality constraints on either the scalars μ_i or vectors ψ_i. This is particularly useful if one wants to incorporate the effects of inexact numerical linear algebra when numerically solving the discrete eigenvalue problems. To this end we followed the approach of [2] which we combined with spectral calculus from the theory of diagonalizable operators—equivalently terminology is scalar operators as presented in the monograph of Dunford-Schwartz—with the analysis of the associated Sobolev scale.

A Sobolev scale for non-selfadjoint operators. Define the form $B(w,v) = \int_\Omega A\nabla w \cdot \nabla \overline{v} + (b \cdot \nabla w + cw)\overline{v}\, dx$, then there exists the sectorial operator $\mathcal{A}$ which represents the form B in the sense of $B(\psi,\phi) = (\mathcal{A}\psi,\phi)_{L^2(\Omega)}$, $\phi \in H_0^1(\Omega)$ and $\psi \in \mathrm{Dom}(\mathcal{A}) \subset H_0^1(\Omega)$ and $\mathcal{A}^{1/2} = \frac{2}{\pi}\int_0^\infty (I + t^2\mathcal{A})^{-1}\mathcal{A}\, dt$ defines the operator square root. Kato's square root theorem [1, 3] yields $\mathrm{Dom}(\mathcal{A}^{1/2}) = \mathrm{Dom}(\mathcal{A}^{*1/2}) = H_0^1(\Omega)$, where $\mathcal{A}^{*1/2}$ denotes the dual operator. This allows us to define the norms $\|\cdot\|_{1/2} = \|\mathcal{A}^{1/2}\cdot\|_{L^2(\Omega)}$ and $\|\cdot\|_{*1/2} = \|\mathcal{A}^{*1/2}\cdot\|_{L^2(\Omega)}$ which are equivalent to the norm $\|\cdot\|_1$ on $H_0^1(\Omega)$. Subsequently, we have that the dual norms $\|\cdot\|_{-1/2}$ and $\|\cdot\|_{-*1/2}$ are equivalent to the $H^{-1}(\Omega)$ norm. Since residual functionals for finite element approximations are (typically) elements of $H^{-1}(\Omega)$, this allows us to build our error estimation theory on the analysis in the Sobolev scales associated to operators $\mathcal{A}$ and $\mathcal{A}^*$. The constants in (3)–(5) depend on the norms of $\mathcal{X}$ and $\mathcal{X}^{-1}$—which measure the departure from normality of $\mathcal{A}$—the separation of spectral components, and the equivalence constants between the norms $\|\cdot\|_{1/2}$, $\|\cdot\|_{*1/2}$ and $\|\cdot\|_1$. For details see the preprint which is published as `Matheon Preprint #1008`.

Numerical experiments. Let us now discretize our model problem (1) using *hp*-finite element spaces. Let $\mathcal{T} = \mathcal{T}_h$ be a triangulation of Ω with the piecewise constant mesh function $h : \mathcal{T}_h \to (0,1)$, $h(T) = \mathrm{diam}(T)$ for $T \in \mathcal{T}_h$ and let

$p : \mathcal{T}_h \to \mathbb{N}$ denote the polynomial degree distribution function. We define the space

$$V = V_h^p = \{v \in H_0^1(\Omega) \cap C(\overline{\Omega}) : \ v\big|_T \in \mathbb{P}_{p(T)} \text{ for each } T \in \mathcal{T}_h\} \ ,$$

where $\mathbb{P}_{p(T)}$ is the collection of polynomials of total degree not greater than p. For further technical details please consult the preprint mentioned above.

For $\psi_i \in H_0^1(\Omega)$ set $R_T(\psi_i) := \psi_i - \mu_i^{-1}(-\nabla \cdot A\nabla\psi_i + b \cdot \nabla\psi_i + c\psi_i)$ and $R_\varepsilon(\psi_i) := \mu_i^{-1}(-(A\nabla\psi_i)_{|_T} \cdot \mathbf{n}_T - (A\psi_i)_{|_{T'}} \cdot \mathbf{n}_{T'})$, and

$$\eta(\psi_i)^2 := \sum_{T\in\mathcal{T}} \left(\frac{h(T)}{p(T)}\right)^2 \|R_T(\psi_i)\|_{0,T}^2 + \sum_{\varepsilon\in\mathcal{E}} \frac{h(\varepsilon)}{p(\varepsilon)} \|R_\varepsilon(\psi_i)\|_{0,\varepsilon}^2$$

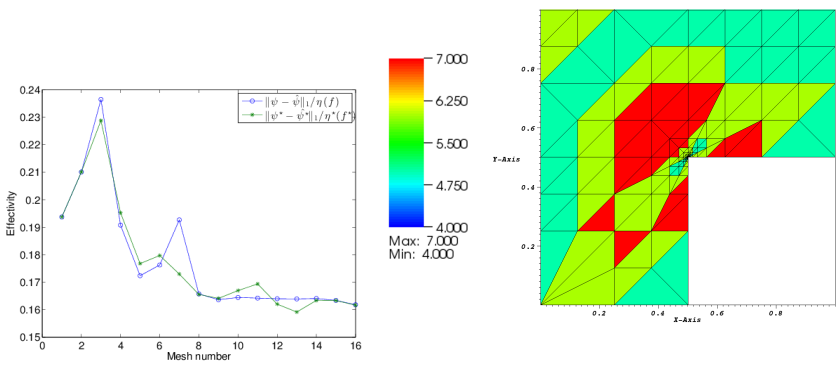
Then $\|\mathfrak{r}(\mu_i)[\psi_i, \cdot]\|_{-1} \leq C\mu_i\eta(\psi_i)$, and the constants depend only on the hp shape regularity constant as given in [5]. Analogous statement hods for the dual operator. Here $\mathcal{E}$ denotes the set of interior edges.

Note that starting from this formulation there are several upwinding finite element schemes to discretize the drift-diffusion operator, eg. [6]. Our focus is on providing error estimates regardless of the origin of the approximation functions ψ_i. Given the assumptions—no restrictions on the coefficients—estimates (3)–(5) are sharp. However, if we assume that the residuals are so small that the estimates guarantee $\|P - Q\|_{HS} < 1$ then we can identify a matching between a basis $\{\psi_i\}$ of $\mathrm{Ran}(P)$ and a basis $\{\phi_i\}$ of $\mathrm{Ran}(Q)$ such that

$$\frac{|\mu_i - \lambda|}{\mu_i} \leq C\mu_i\eta(\psi_i)\eta^d(\psi_i^d), \qquad \|\phi_i - \psi\|_1 \leq C\mu_i\,\eta(\psi_i) \ .$$

Note that there is also an efficiency estimate $\mu_i\eta(\psi_i) \leq C(\|\phi_i - \psi_i\|_1 + |\mu_i - \lambda|)$. The constant—as in the boundary value problem case from [5]—depends on maximal $p(T)$. We use these estimates for experiments. Note, equivalent statements hold for left eigenvectors and the eigenvalue λ is assumed semisimple of multiplicity n.

Obligatory L-Shape benchmark. We consider the operator $\mathcal{A} = -\Delta + b \cdot \nabla$, where $b = (2, 2)$ and Ω is the L-shaped domain.



First effectivity indices of the left and right eigenvectors (eigenfunctions) corresponding to the first eigenvalue on the L-shaped domain. The singularities of

the left and right eigenvectors at the origin (the reentrant corner) have been recognized by our adaptive scheme, which does heavy h-refinement near the origin. The exponential convergence rates, estimated with least-squares fitting, were $\alpha = 0.2697, 0.2671$.

References

[1] P. Auscher and P. Tchamitchian. Square roots of elliptic second order divergence operators on strongly lipschitz domains: L^2 theory. *J. Anal. Math.*, 90 (1):1–12, 2003.

[2] C. Beattie and I. C. F. Ipsen. Inclusion regions for matrix eigenvalues. *Linear Algebra Appl.*, 358:281–291, 2003. Special issue on accurate solution of eigenvalue problems (Hagen, 2000).

[3] T. Kato. Fractional powers of dissipative operators. *J. Math. Soc. Japan*, 13:246–274, 1961.

[4] D. Liberzon and R. W. Brockett. Spectral analysis of Fokker-Planck and related operators arising from linear stochastic differential equations. *SIAM J. Control Optim.*, 38(5):1453–1467, 2000.

[5] J. M. Melenk and B. I. Wohlmuth. *hp*-FEM. *Adv. Comput. Math.*, 15(1-4):311–331 (2002), 2001. A posteriori error estimation and adaptive computational methods.

[6] J. Xu and L. Zikatanov. A monotone finite element scheme for convection-diffusion equations. *Math. Comp.*, 68(228):1429–1446, 1999.

Hierarchically enhanced adaptive finite element method for PDE eigenvalue/eigenvector approximations

Agnieszka Międlar

(joint work with Luka Grubišić and Jeffrey S. Ovall)

1. Introduction and Preliminaries

In this work we are interested in solving generalized eigenvalue problems associated with finite element disretization of PDE eigenvalue problems up to the accuracy guaranteed by the higher order finite elements while keeping the computational cost of the lower finite elements approximation, i.e., obtaining approximations of the $\mathbb{P}_2$ finite elements accuracy within the cost of $\mathbb{P}_1$ finite elements computations. We are generally interested is solving the PDE problems of the form

Find $(\lambda, \psi, \psi^\star) \in \mathbb{R} \times H_0^1(\Omega) \times H_0^1(\Omega)$ *such that*

$$B(\psi, v) = \lambda(\psi, v), \quad B(v, \psi^\star) = \overline{\lambda}(\psi^\star, v) \quad \textit{for all } v \in H_0^1(\Omega).$$

where $\psi^\star, \psi \neq 0$ for all $v \in H_0^1(\Omega)$. We assume that $B(\cdot, \cdot)$ is bounded and coercive, and it defines the compact solution operator which maps the function $f \in L^2(\Omega)$, to $u(f) \in H_0^1(\Omega)$, i.e., $u(\psi) = \frac{1}{\lambda}\psi$. Therefore, the eigenvalue problems can be easily transformed to the boundary value problem of the form

Find $u(f), u^\star(f) \in v \in H_0^1(\Omega)$ *such that*

$$B(u(f), v) = (f, v), \quad B(v, u^\star(f) = (f, v) \quad \textit{for all } v \in H_0^1(\Omega).$$

Standard finite element discretization is obtained by solving the problem in the finite dimensional space

$$V_p = \Big\{ v \in C(\overline{\Omega}) \cap H_0^1(\Omega) : v_{|T} \in \mathbb{P}_p \text{ for each } T \in \mathcal{T} \Big\},$$

where $\mathcal{T}$ defines a conforming, shape-regular triangulation of domain $\Omega \subset \mathbb{R}^2$, with internal nodes and edges $\mathcal{V}$, $\mathcal{E}$, respectively. The most common basis for V_p is the so-called Lagrange (nodal) basis, globally continuous, piecewise polynomials of degree at most p ($\mathbb{P}_p$), i.e., $V_p = span\{\ell_z\}$, for node $z \in \mathcal{V}$. In contrast, the *p-hierarchical* basis for V_p contain functions of various degrees suggested by the corresponding hierarchical splitting [Ban96]

$$Q = V \oplus W,$$

with $V, W \subset H_0^1(\Omega)$. Here we consider, the following hierarchical splitting

$$Q = V_2, \quad V = V_1 = span\{\ell_z\}_{z\in\mathcal{V}}, \quad W = V_2 \backslash V_1 = span\{b_e\}_{e\in\mathcal{E}},$$

where $b_e = 4\ell_z\ell_{z'}$, where z, z' are the two vertices of the edge $e \in \mathcal{E}$. The system matrices, stiffness matrix K^{LB}, K^{HB} and mass matrix M^{LB}, M^{HB}, corresponding to the choice of Lagrange and p-hierarchical basis, respectively, possess the similar block structure

$$K^{LB} = \begin{bmatrix} K_{11}^{LB} & A_{12}^{LB} \\ K_{21}^{LB} & A_{22}^{LB} \end{bmatrix}, \quad K^{HB} = \begin{bmatrix} K_{11}^{HB} & K_{12}^{HB} \\ K_{21}^{HB} & K_{22}^{HB} \end{bmatrix}.$$

However, the corresponding blocks of both matrices have severely different properties, see [BO13]. The diagonal blocks K_{11}^{LB}, K_{22}^{LB} are both well-conditioned, whereas the off-diagonal blocks are strongly coupled and therefore highly ill-conditioned, which causes problems in the numerical computations. In contrast, the ill-conditioning of K^{HB} is concentrated in the diagonal block K_{11}^{HB} which can be treated numerically very well and the off-diagonal coupling is very mild due to the strengthened Cauchy-Schwarz inequality between spaces V_1 and $V_2\backslash V_1$ [Ban96]. Let us now restrict our investigation to the aforementioned choice of the hierarchical splitting of the finite element space $V_2 = V_1 \oplus (V_2\backslash V_1)$ which results in the system matrices of the following block structure

$$K^{\mathbb{P}_2} = \begin{bmatrix} K^{\mathbb{P}_1} & R \\ R^T & D \end{bmatrix}, \quad M^{\mathbb{P}_2} = \begin{bmatrix} M^{\mathbb{P}_1} & LB \\ LB^T & BB \end{bmatrix}.$$

2. Hierarchically enriched AFEM algorithm

The properties of the system matrix $K^{\mathbb{P}_2}$ and $M^{\mathbb{P}_2}$ accordingly, allow us to introduce a very effective adaptive finite element eigensolver. The main idea of the method is to exploit the hierarchical splitting of the finite element space and the hierarchical residual representation presented in [HOS11] to design a cheap a posteriori error estimator which will be used not only to conduct the proper mesh refinement, but in particular to improve the quality of the approximate eigentriples. The hierarchical residual representation combined with the hierarchical

splitting $Q = V \oplus W$ of the finite element space allow us to state our original problem in the following form

$$B(\widehat{u}(f), v) = (f, v) \quad \text{for all } v \in V, \tag{1}$$

$$B(\varepsilon(f), v) = (f, v) - B(\widehat{u}(f), v) \quad \text{for all } v \in W, \tag{2}$$

where $u(f) \in H_0^1(\Omega)$, $\widehat{u}(f) \in V$ and $\varepsilon(f) \in W$. Equation (1) determine the solution of the original problem in the $\mathbb{P}_1$ finite element space. This approximation have to be very accurate because of the ill-conditioning of the $K^{\mathbb{P}_1}$ block. The second equation (2) allows to determine the bubble residual term $\varepsilon(f)$, which will be used to steer the mesh refinement and improve the $\mathbb{P}_1$ finite element solution $\widetilde{\psi}_h$, i.e., the new hierarchically enhanced approximation will be given as $\widetilde{\psi}_h + \varepsilon$. Therefore, the approximate solution can be found using the following Algorithm.

Hierarchically enriched AFEM algorithm

Input: $\mathcal{T}_h, tol$

$(\widehat{\lambda}_h, \widehat{\psi}_h) = \mathrm{eig}(K_h^{P_1}, M_h^{P_1}, \mathcal{T}_h)$ ▷ Solve (1).

$\widehat{u}(\widehat{\psi}_h) = \frac{1}{\widehat{\lambda}_h} \widehat{\psi}_h$

$rhs = \widehat{\psi}_{bubbles} - R^T \widehat{u}(\widehat{\psi}_h)$ ▷ Determine the right-hand side of (2).

$\widetilde{\varepsilon}_h = bubble\_solve(D, rhs)$ ▷ Solve (2).

while $\widehat{\eta}_h := |||\widetilde{\varepsilon}_h||| \geq tol$ **do**

  $Refine\ \mathcal{T}_h$ ▷ Iterate further with inexact solve

  $(\widetilde{\lambda}_h, \widetilde{\psi}_h) = eig(K_h^{P_1}, M_h^{P_1}, \mathcal{T}_h)$ ▷ Solve EVP with refined eigenvector $\widehat{\psi}_h$

  $\widetilde{u}(\widetilde{\psi}_h) = \frac{1}{\widetilde{\lambda}_h} \widetilde{\psi}_h$

  $rhs = \widetilde{\psi}_{bubbles} - R^T \widetilde{u}(\widetilde{\psi}_h)$

  $\widetilde{\varepsilon}_h = inexact\_bubble\_solve(D, rhs)$

  $\widehat{\psi}_h = refined\_vect(\widetilde{\psi}_h, \widetilde{\varepsilon}_h) = \underset{v^* = \widetilde{\psi}_h + \alpha \widetilde{\varepsilon}_h}{\text{arg-min}} B(v^*, v^*)$

end while

Output: $(\widehat{\lambda}_h, \widehat{\psi}_h)$

We can show that for $\widehat{\psi}, \widehat{\psi}^\star \in V$, $(\widehat{\psi}, \widehat{\psi}^\star) \neq 0$ being the Galerkin approximations of $\psi, \psi^\star$, the following estimates hold

$$\|\psi - \widehat{\psi}\|_1 \leq C_{\widehat{\lambda}} \max\{K_1, K_2\}(\|\varepsilon(\widehat{\psi})\|_1 + \mathrm{osc}),$$

$$\|\psi^\star - \widehat{\psi}^\star\|_1 \leq C_{\widehat{\lambda}} \max\{K_1^\star, K_2^\star\}(\|\varepsilon(\widehat{\psi}^\star)\|_1 + \mathrm{osc}^\star).$$

This allow us to use the bubble residual vector ε to obtain a more accurate eigentriples with the minor additional cost of view iterative steps of some simple linear system solver, e.g., Gauss-Seidel.

3. Numerical experiments

Some preliminary numerical examples for the Laplace eigenvalue problem on the uniformly refined L-shape domain confirm the efficiency of the presented algorithm.

#DOFs	$\mathbb{P}_1$ error	$\mathbb{P}_1 - \mathbb{P}_2$ error	$\mathbb{P}_2$ error
9	1.3674×10^{-1}	3.6258×10^{-3}	$- - -$
49	3.7372×10^{-2}	3.1688×10^{-4}	2.2472×10^{-4}
225	9.5626×10^{-3}	2.1867×10^{-5}	1.4345×10^{-5}
961	2.4048×10^{-3}	1.4040×10^{-6}	9.0148×10^{-7}
3969	6.0209×10^{-4}	8.8361×10^{-8}	$- - -$

References

[Ban96] R. E. Bank, *Hierarchical bases and the finite element method*, Acta Numer., 1996, 1–43.

[BO13] S. Le Borne and J. S. Ovall, *Rapid error reduction for block Gauss-Seidel based on p-hierarchical bases*, Numer. Linear Algebra Appl., **20**, 2013, pp. 743–760.

[HOS11] M. Holst, J. S. Ovall, and R. Szypowski, *An efficient, reliable and robust error estimator for elliptic problems in* $\mathbb{R}^3$, Appl. Numer. Math. **61**(5), 2011, pp. 675–695.

[Kol05] K. Kolman, *A two-level method for nonsymmetric eigenvalue problems*, Acta Math. Appl. Sin. Engl. Ser. **21**(1), 2005, pp. 1–12.

Adaptive C^0 interior penalty method for biharmonic eigenvalue problems

Joscha Gedicke

(joint work with Susanne C. Brenner, Li-Yeng Sung)

This talk presents a residual based *a posteriori* error estimator for biharmonic eigenvalue problems and the C^0 interior penalty method. Biharmonic eigenvalue problems occur in the analysis of vibrations and buckling of plates. The *a posteriori* error estimator is proven to be reliable and efficient for sufficiently large penalty parameter $\sigma \geq 1$ and sufficiently small global mesh size H. The theoretical results are verified in numerical experiments.

The weak formulation of the eigenvalue problem for the vibration of plates seeks an eigenfunction $u \in H_0^2(\Omega)$ in the polygonal Lipschitz domain $\Omega \subset \mathbb{R}^2$ with $b(u,u) = 1$ and nonzero eigenvalue $\lambda \in \mathbb{R}$ such that

$$a(u,v) = \lambda b(u,v) \quad \text{for all } v \in H_0^2(\Omega),$$

where the bilinear forms $a(\cdot,\cdot)$ and $b(\cdot,\cdot)$ read

$$a(u,v) = \int_\Omega D^2u : D^2v \, dx = \int_\Omega \sum_{i,j=1}^{2} \frac{\partial^2 u}{\partial x_i \partial x_j} \frac{\partial^2 v}{\partial x_i \partial x_j} \, dx,$$

$$b(u,v) = \int_\Omega uv \, dx.$$

For simplicity, this talk is restricted to simple eigenvalues. Let $\mathcal{T}_\ell$ be a (shape regular) triangulation with set of edges $\mathcal{E}_\ell$ and set of interior edges $\mathcal{E}_\ell^i$. For any $v \in H^2(\Omega, \mathcal{T}_\ell) := \{v \in H_0^1(\Omega) : v|_T \in H^2(T) \text{ for all } T \in \mathcal{T}_\ell\}$ and any $w \in H^3(\Omega, \mathcal{T}_\ell)$, the jump of the derivative in normal direction and the average of the second

derivative in normal-normal direction along the edge $E = T_+ \cap T_-$, $T_\pm \in \mathcal{T}_\ell$, with normal n_E pointing from T_- to T_+, are defined by

$$\left[\!\left[\frac{\partial v}{\partial n} \right]\!\right] = \left. \frac{\partial v_+}{\partial n_E} \right|_E - \left. \frac{\partial v_-}{\partial n_E} \right|_E \qquad \text{and} \qquad \left\{\!\!\left\{ \frac{\partial^2 w}{\partial n^2} \right\}\!\!\right\} = \frac{1}{2} \left(\left. \frac{\partial^2 w_+}{\partial n_E^2} \right|_E + \left. \frac{\partial^2 w_-}{\partial n_E^2} \right|_E \right).$$

The C^0 interior penalty method [1, 3] avoids the use of complicated C^1 finite elements but uses standard Lagrange finite elements of total degree $k \geq 2$. This method is nonconforming in the sense that $P_k(\mathcal{T}_\ell) \cap H_0^1(\Omega) := V_\ell \not\subset H_0^2(\Omega)$, and the associated nonconforming bilinear form reads

$$\begin{aligned} a_{NC}(u_\ell, v_\ell) = &\sum_{T \in \mathcal{T}_\ell} \int_T D^2 u_\ell : D^2 v_\ell \, dx \\ &+ \sum_{E \in \mathcal{E}_\ell} \int_E \left(\left\{\!\!\left\{ \frac{\partial^2 u_\ell}{\partial n^2} \right\}\!\!\right\} \left[\!\left[\frac{\partial v_\ell}{\partial n} \right]\!\right] + \left[\!\left[\frac{\partial u_\ell}{\partial n} \right]\!\right] \left\{\!\!\left\{ \frac{\partial^2 v_\ell}{\partial n^2} \right\}\!\!\right\} \right) ds \\ &+ \sum_{E \in \mathcal{E}_\ell} \frac{\sigma}{h_E} \int_E \left[\!\left[\frac{\partial u_\ell}{\partial n} \right]\!\right] \left[\!\left[\frac{\partial v_\ell}{\partial n} \right]\!\right] ds \end{aligned}$$

for all $u_\ell, v_\ell \in V_\ell$. Note that the second term results from consistency, the third term realises symmetry and the last term is a penalty term with the penalty parameter $\sigma \geq 1$. The bilinear form $a_{NC}(\cdot, \cdot)$ is symmetric, continuous and coercive for sufficiently large penalty parameter $\sigma \geq 1$ [1, 3]. Hence, the discrete eigenvalue problem, to seek $u_\ell \in V_\ell$ with $b(u_\ell, u_\ell) = 1$ and $\lambda_\ell \in \mathbb{R}$ such that

$$a_{NC}(u_\ell, v_\ell) = \lambda_\ell b(u_\ell, v_\ell) \quad \text{for all } v_\ell \in V_\ell,$$

leads to a sequence of positive real eigenvalues $0 < \lambda_{\ell,1} \leq \lambda_{\ell,2} \leq \ldots \leq \lambda_{\ell,\dim(V_\ell)}$. Similar to the *a posteriori* error estimator for the source problem [2], the *a posteriori* error estimator for the eigenvalue problem and polynomial degree $k = 2$ reads

$$\eta_\ell^2 := \sum_{T \in \mathcal{T}_\ell} h_T^4 \lambda_\ell \|u_\ell\|_{L_2(T)}^2 + \sum_{E \in \mathcal{E}_\ell} \frac{\sigma^2}{h_E} \|[\![\partial u_\ell / \partial n]\!]\|_{L_2(E)}^2 + \sum_{E \in \mathcal{E}_\ell^i} h_E \|[\![\partial^2 u_\ell / \partial n^2]\!]\|_{L_2(E)}^2.$$

The *a posteriori* error estimator η_ℓ is proven to be reliable and efficient for the mesh dependent norm

$$\|v\|_{H^2(\Omega,\mathcal{T}_\ell)}^2 = \sum_{T \in \mathcal{T}_\ell} |v|_{H^2(T)}^2 + \sigma \sum_{E \in \mathcal{E}_\ell} h_E^{-1} \|[\![\partial v / \partial n]\!]\|_{L_2(E)}^2 \quad \text{for all } v \in H^2(\Omega, \mathcal{T}_\ell),$$

in the sense that up to generic constants

$$\begin{aligned} \|u - u_\ell\|_{H^2(\Omega,\mathcal{T}_\ell)} &\lesssim \eta_\ell + \|\lambda u - \lambda_\ell u_\ell\|_{L^2(\Omega)}, \\ \eta_\ell &\lesssim \|u - u_\ell\|_{H^2(\Omega,\mathcal{T}_\ell)} + H^2 \|\lambda u - \lambda_\ell u_\ell\|_{L_2(\Omega)}, \end{aligned}$$

where $\|\lambda u - \lambda_\ell u_\ell\|_{L^2(\Omega)}$ is of higher order compared to η_ℓ. The reliability and efficiency of η_ℓ^2 for the eigenvalue error $|\lambda - \lambda_\ell|$ is verified in numerical experiments for varying penalty parameters, sizes of eigenvalues, and for convex and non-convex domains.

References

[1] S.C. Brenner, C^0 *Interior Penalty Methods*, Frontiers in numerical analysis–Durham 2010, 79–147, Lect. Notes Comput. Sci. Eng., 85, Springer, Heidelberg, 2012.

[2] S.C. Brenner, T. Gudi, L.-Y. Sung, *An a posteriori error estimator for a quadratic C^0-interior penalty method for the biharmonic problem*, IMA J. Numer. Anal. **30** (2010), no. 3, 777–798.

[3] S.C. Brenner, L.-Y. Sung, C^0 *interior penalty methods for fourth order elliptic boundary value problems on polygonal domains*, J. Sci. Comput. **22/23** (2005), 83–118.

An Optimal Adaptive FEM for Eigenvalue Clusters

Dietmar Gallistl

Let $\Omega \subseteq \mathbb{R}^d$, $d \geq 2$, be a bounded Lipschitz domain with polyhedral boundary. The adaptive finite element approximation of multiple eigenvalues of the model problem $-\triangle u = \lambda u$ leads to the situation of eigenvalue clusters because the eigenvalues of interest and their multiplicities may not be resolved by the initial mesh. The optimality analysis of adaptive finite element methods in the literature is based on the comparison of the finite element solutions on different meshes. In the case of multiple eigenvalues, this leads to the difficulty that the discrete orthonormal systems of eigenfunctions produced by the adaptive algorithm may change in each step of the adaptive loop. The work [1] for multiple eigenvalues introduces the innovative methodology to use one bulk criterion for all discrete eigenfunctions in the algorithm for automatic mesh refinement and proves equivalence to the simultaneous error of the discrete eigenvalue approximation to the fixed orthonormal basis of the exact eigenspace. In practice, little perturbations in coefficients or in the geometry immediately lead to an eigenvalue cluster of finite length. This talk is based on the work [3] and extends the approach of [1] to the more practical case of eigenvalue clusters.

Suppose that the eigenvalues and the discrete eigenvalues are enumerated

$$0 < \lambda_1 \leq \lambda_2 \leq \dots \quad \text{and} \quad 0 < \lambda_{\ell,1} \leq \dots \leq \lambda_{\ell,\dim(V_\ell)}.$$

Let $(u_1, u_2, u_3, \dots)$ and $(u_{\ell,1}, u_{\ell,2}, \dots, u_{\ell,\dim(V_\ell)})$ denote some L^2-orthonormal systems of corresponding eigenfunctions. For a cluster of eigenvalues $\lambda_{n+1}, \dots, \lambda_{n+N}$ of length $N \in \mathbb{N}$ define the index set $J := \{n+1, \dots, n+N\}$ and the spaces $W := \operatorname{span}\{u_j\}_{j\in J}$ and $W_\ell := \operatorname{span}\{u_{\ell,j}\}_{j\in J}$. The eigenspaces $E(\lambda_j)$ may differ for different $j \in J$. Let the cluster be contained in a bounded interval $[A, B]$.

The adaptive algorithm is driven by the element-wise sum of the residual-based error estimator contributions [2] of all $u_{\ell,j}$ $(j \in J)$ and runs the following loop

$$\textsc{Solve} \to \textsc{Estimate} \to \textsc{Mark} \to \textsc{Refine}$$

based on an initial triangulation $\mathcal{T}_0$, and the bulk parameter $0 < \theta \leq 1$ for the Dörfler marking.

Let $\|\cdot\|$ denote the L^2 norm and $|||\cdot|||$ denote the H^1 seminorm and set

$$\mathcal{A}_\sigma := \big\{ v \in V \;\big|\; |v|_{\mathcal{A}_\sigma} := \sup_{m\in\mathbb{N}} m^\sigma \inf_{\mathcal{T}\in\mathbb{T}(m)} |||(1-\Pi^0_{\mathcal{T}})Dv|| < \infty \big\}.$$

(H1)	$M_J := \sup_{\mathcal{T}_\ell \in \mathbb{T}} \max_{j \in \{1,\dots,\dim(V_\ell)\} \setminus J} \max_{k \in J} \frac{\lambda_k}{|\lambda_{\ell,j} - \lambda_k|} < \infty$
(H2)	$\|h_0\|_\infty^{2s} B^2 C_{\mathrm{drel}}^2 (1 + M_J)^2 \leq 1$
(H3)	$\varepsilon := \max_{j \in J} \|u_j - \Lambda_\ell u_j\| \leq \sqrt{1 + (2N)^{-1}} - 1$
(H4)	$(1 + M_J)^2 (B C_{\mathrm{qo}} \|h_0\|_\infty^{2s} + B^2 C_{\mathrm{reg}}^2 \|h_0\|_\infty^{2+2s}) < \min\left\{1, \frac{1-\rho_1}{K C_{\mathrm{drel}}^2}\right\}/4$

TABLE 1. The constants C_{reg}, C_{qo}, C_{drel} depend only on Ω and its geometry and on the set $\mathbb{T}$ of admissible triangulations.

Here, $\mathbb{T}(m)$ is the set of admissible triangulations whose cardinality differs from that of the initial triangulation $\mathcal{T}_0$ by at most m and $\Pi_{\mathcal{T}}^0$ is the L^2 projection onto piecewise constants. Let $(\lambda_j \mid j \in J)$ denote the cluster under consideration with (possibly different) eigenspaces $E(\lambda_j)$. Provided that all eigenfunctions in the cluster belong to $\mathcal{A}_\sigma$, the error quantities

$$\left(\frac{|\lambda_k - \lambda_{\ell,k}|}{\lambda_{\ell,k}}\right)^{1/2} \quad \text{and} \quad \sup_{j \in J} \sup_{\substack{w \in E(\lambda_j) \\ \|w\|=1}} \inf_{v_\ell \in W_\ell} |||w - v_\ell|||$$

decay as $(\mathrm{card}(\mathcal{T}_\ell) - \mathrm{card}(\mathcal{T}_0))^{-\sigma}$. One subtle aspect is the dependence of the parameters on the smallness of the initial mesh and the initial resolution of the cluster and its length. An overview of sufficient conditions for optimal convergence is given in Table 1. The precise statement of the main result is as follows.

Theorem 1. *Provided the bulk parameter $\theta \ll 1$ is sufficiently small and the initial mesh size $\|h_0\|_\infty := \max\{\mathrm{diam}(T) \mid T \in \mathcal{T}_0\}$ satisfies the conditions (H1)–(H4) of Table 1, the adaptive algorithm computes discrete eigenpairs $((\lambda_{\ell,j}, u_{\ell,j})_{j \in J})_\ell$ with optimal rate of convergence in the sense that, for some constants C, C_{opt} and all $k \in J$,*

$$\begin{aligned}(1 + M_J^2 B^2 C)^{-1/2} \left(\frac{|\lambda_k - \lambda_{\ell,k}|}{\lambda_{\ell,k}}\right)^{1/2} + \sup_{j \in J} \sup_{\substack{w \in E(\lambda_j) \\ \|w\|=1}} \inf_{v_\ell \in W_\ell} |||w - v_\ell||| \\ \leq 2 C_{\mathrm{ba}} (1 + (1 + M_J) B \|h_0\|_\infty^s)(1 + B\|h_0\|_\infty^2) C_{\mathrm{opt}} \\ (\mathrm{card}(\mathcal{T}_\ell) - \mathrm{card}(\mathcal{T}_0))^{-\sigma} \left(\sum_{j \in J} |u_j|_{\mathcal{A}_\sigma}^2\right)^{1/2}.\end{aligned}$$

The proof is concerned with the analysis of the eigenfunction approximation. The estimate for the eigenvalues then follows from the results of [4].

A theoretical non-computable error estimator is employed which allows a proof of equivalence to the refinement indicator of the adaptive algorithm. In contrast to the case of one multiple eigenvalue, care has to be taken that the reliability and equivalence constants of the error estimator do not depend on the cluster or its length. The non-computable error estimator allows reliable and efficient error estimates and is locally equivalent to the computable explicit residual-based

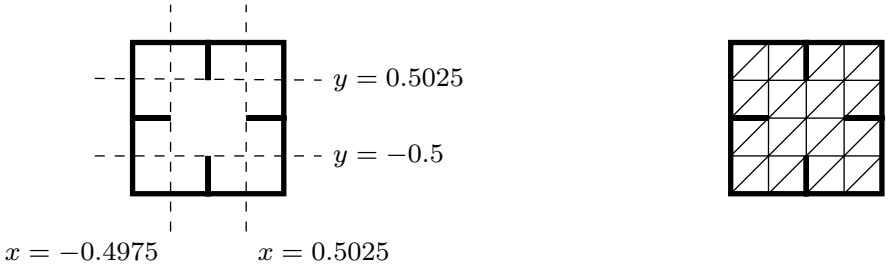


FIGURE 1. Square domain $(0,1)^2$ with perturbed symmetric slits and coarse initial triangulation $\mathcal{T}_0$ with 5 interior vertices.

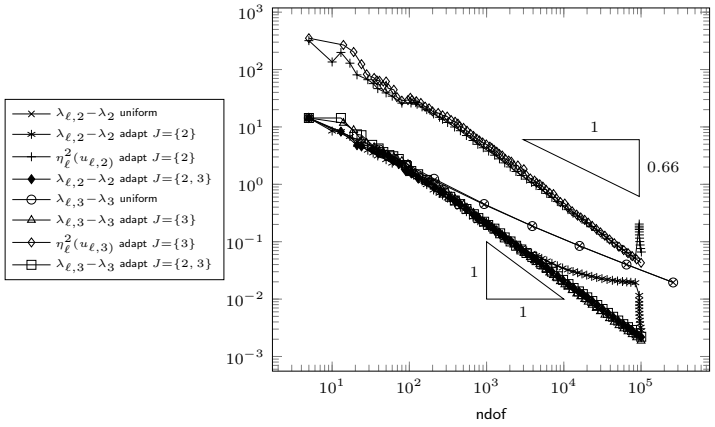


FIGURE 2. Convergence history for $J \subseteq \{2,3\}$ based $\theta = 0.1$.

error estimator. The proof of this property requires a careful analysis and the condition (H3) on the initial mesh-size. The equivalence of error estimators allows to consider the theoretical error estimator in the analysis with some modified bulk parameter. This leads to estimator reduction and contraction properties.

The following numerical test suggests that the adaptive cluster approximation seems to be superior compared to the use of an adaptive scheme for each eigenvalue separately, even if all eigenvalues on the continuous level are simple.

The exact second and third eigenvalues on the domain of Figure 1 read as $\lambda_2 = 17.6455, \quad \lambda_3 = 17.6626$. This cluster $J = \{2,3\}$ is not resolved on coarse or moderately fine meshes. The adaptive algorithm with bulk parameter $\theta = 0.1$ and an initial triangulation with 5 degrees of freedom yields the results displayed in Figure 2. In the case of $J = \{2\}$ one can observe some pre-asymptotic effect up to 10^5 degrees of freedom for the second eigenvalue $\lambda_{\ell,2}$. Indeed, the discrete eigenfunctions are not resolved on moderately refined meshes. The choice $J = \{2,3\}$ seems to resolve the cluster in a better way, in the sense that the discrete eigenfunctions are correctly separated on coarse meshes and the simultaneous approximation produces optimal rates.

The author is supported by the DFG Research Center Matheon (Berlin).

References

[1] X. Dai, L. He, and A. Zhou, *Convergence rate and quasi-optimal complexity of adaptive finite element computations for multiple eigenvalues*, arXiv Preprint 1210.1846v1 (2012).

[2] R.G. Durán, C. Padra and R. Rodríguez, *A posteriori error estimates for the finite element approximation of eigenvalue problems*, Math. Models Methods Appl. Sci., **13**(2003), 1219–1229.

[3] D. Gallistl, *An optimal adaptive FEM for eigenvalue clusters*, submitted (2013).

[4] A. Knyazev and J. Osborn, *New a priori FEM error estimates for eigenvalues*, SIAM J. Numer. Anal., **43**(2006), 2647–2667.

Adaptive Nonconforming Crouzeix-Raviart FEM for Eigenvalue Problems

Mira Schedensack

(joint work with C. Carstensen, D. Gallistl)

The nonconforming approximation of eigenvalues is of high practical interest because it allows for guaranteed upper and lower eigenvalue bounds [4] and for a convenient computation via a consistent diagonal mass matrix in 2D as well as a low-order discretization of elasticity and fluid problems. The underlying paper [1] of this presentation proves quasi-optimal convergence of an adaptive algorithm of the form

$$(\text{inexact Solve \& Estimate}) \to \text{Mark} \to \text{Refine}$$

with respect to the number of degrees of freedom.

Given a bounded polygonal Lipschitz domain $\Omega \subset \mathbb{R}^2$, the Laplace eigenvalue problem seeks eigenpairs $(\lambda, u) \in \mathbb{R} \times H^1_0(\Omega)$ with $\|u\|_{L^2(\Omega)} = 1$ and

$$\int_\Omega \nabla u \cdot \nabla v \, dx = \lambda \int_\Omega uv \, dx \quad \text{for all } v \in H^1_0(\Omega). \tag{1}$$

For simplicity, the analysis is carried out for the smallest eigenvalue, but the results remain true for higher simple eigenvalues. The nonconforming P_1 finite element space reads

$$\mathrm{CR}^1_0(\mathcal{T}_\ell) := \left\{ v \in L^2(\Omega) \;\middle|\; \begin{array}{l} v|_T \in P_1(T) \text{ for all } T \in \mathcal{T}_\ell \text{ and } v \text{ is continuous} \\ \text{in midpoints of interior edges and vanishes} \\ \text{in midpoints of boundary edges} \end{array} \right\}$$

for a regular triangulation $\mathcal{T}_\ell$ of Ω. The discretization of (1) seeks $(\lambda_\ell, u_\ell) \in \mathbb{R} \times \mathrm{CR}^1_0(\mathcal{T}_\ell)$ with

$$\int_\Omega \nabla_{\mathrm{NC}} u_\ell \cdot \nabla_{\mathrm{NC}} v_\ell \, dx = \lambda_\ell \int_\Omega u_\ell v_\ell \, dx \qquad \text{for all } v_\ell \in \mathrm{CR}^1_0(\mathcal{T}_\ell)$$

with the piecewise gradient ∇_{NC}.

Given $\kappa > 0$, the inexact Solve of the adaptive algorithm allows the computation of an approximation $(\tilde\lambda_\ell, \tilde u_\ell) \in \mathbb{R} \times \mathrm{CR}^1_0(\mathcal{T}_\ell)$ of the discrete eigenpair (λ_ℓ, u_ℓ) with

$$|||u_\ell - \tilde u_\ell|||^2_{\mathrm{NC}} + |\lambda_\ell - \tilde\lambda_\ell|^2 \le \kappa \min\{\eta^2_\ell, \eta^2_{\ell-1}\}$$

with the error estimator η_ℓ from ESTIMATE with respect to $\tilde\lambda_\ell$ and $\tilde u_\ell$. Precisely, η_ℓ is defined by

$$\eta_\ell^2(T) := |T|\|\tilde\lambda_\ell \tilde u_\ell\|^2_{L^2(T)} + |T|^{1/2} \sum_{E\in\mathcal{E}_\ell(T)} \| [\nabla_{\mathrm{NC}} \tilde u_\ell]_E \cdot \tau_E\|^2_{L^2(E)}$$

for the edges $\mathcal{E}_\ell(T)$ of a triangle $T \in \mathcal{T}_\ell$, the tangential vector τ_E of E and the jump $[\bullet]_E$ along E and $\eta_\ell := \sqrt{\sum_{T\in\mathcal{T}_\ell} \eta_\ell^2(T)}$ (and $\eta_{-1} := \infty$). This coupling of INEXACT SOLVE and ESTIMATE was analysed in [3] for an adaptive algorithm with the lowest-order conforming FEM for eigenvalue problems. The steps MARK and REFINE consist of Dörfler marking with bulk parameter θ and newest-vertex bisection.

The quasi-optimal convergence of the adaptive algorithm is stated in terms of an approximation semi-norm, which is defined by

$$|u|_{\mathcal{A}_\sigma} := \sup_{N\in\mathbb{N}} N^\sigma \inf_{\mathcal{T}\in\mathbb{T}(N)} \|\nabla u - \Pi_{\mathcal{T}} \nabla u\|_{L^2(\Omega)}$$

for some $\sigma > 0$ and the L^2 projection to piecewise constants $\Pi_{\mathcal{T}}$. Here, $\mathbb{T}(N)$ denotes the set of all regular triangulations which are created from the initial triangulation $\mathcal{T}_0$ using newest-vertex bisection and which consist of less than N newly created triangles.

The following theorem states optimal convergence rates for the adaptive algorithm.

Theorem 1. *For any $\sigma > 0$ such that $|u|_{\mathcal{A}_\sigma} < \infty$ and sufficiently small parameters $\theta \ll 1$, $\kappa \ll 1$ and initial mesh-size $\|h_0\|_{\infty,\Omega} \ll 1$, the adaptive algorithm computes sequences of triangulations $(\mathcal{T}_\ell)_\ell$ and discrete approximations $(\tilde\lambda_\ell, \tilde u_\ell)_\ell$ of optimal rate of convergence in the sense that there exists $C > 0$ with*

$$(\mathrm{card}(\mathcal{T}_\ell) - \mathrm{card}(\mathcal{T}_0))^\sigma \|\nabla_{\mathrm{NC}}(u - \tilde u_\ell)\|_{L^2(\Omega)} \le C|u|_{\mathcal{A}_\sigma} \quad \ell = 0, 1, 2, \dots$$

The proof consists of five main ingredients. The first tool is a best-approximation result for the nonconforming FEM, which proves the equivalence

$$\|\nabla_{\mathrm{NC}}(u - u_\ell)\|_{L^2(\Omega)} \approx \|\nabla u - \Pi_{\mathcal{T}_\ell} \nabla u\|_{L^2(\Omega)}\,.$$

The second ingredient is the L^2 and eigenvalue control

$$|\lambda - \lambda_\ell| + \|u - u_\ell\|_{L^2(\Omega)} \le C\|h_0\|^s_{\infty,\Omega} \|\nabla_{\mathrm{NC}}(u - u_\ell)\|_{L^2(\Omega)}$$

with $0 < s \le 1$ depending on the regularity of the domain. The third to fifth ingredients are the quasi-orthogonality, the contraction property and the discrete reliability for the approximation of the discrete eigenpair.

Numerical experiments suggest that the optimality is obtained for $\theta = 0.1$ and $\kappa = 0.01$ even for very coarse initial meshes.

The n-dimensional discrete reliability [2] allows the generalisation of the quasi-optimal convergence to three dimensions.

The author is supported by the Berlin Mathematical School.

References

[1] C. Carstensen, D. Gallistl, M. Schedensack *Adaptive nonconforming Crouzeix-Raviart FEM for eigenvalue problems*, Math. Comp., accepted for publication.

[2] C. Carstensen, D. Gallistl, M. Schedensack, *Discrete reliability for Crouzeix-Raviart FEMs*, SIAM J. Numer. Anal. **51** (2013), 2935–2955.

[3] C. Carstensen, J. Gedicke, *An adaptive finite element eigenvalue solver of quasi-optimal computational complexity*, SIAM J. Numer. Anal. **50** (2012), 1029–1057.

[4] C. Carstensen, J. Gedicke, *Guaranteed lower bounds for eigenvalues*, Math. Comp., accepted for publication.

Variational Approximation for Self-adjoint Eigenvalue Problems

Christopher Beattie

(joint work with Friedrich Goerisch)

This work presents inequalities for the inertia of a quadratic form restricted to arbitrary subspaces contained within its domain of definition. These inequalities are the basis of a new approach for computing rigorous lower bounds to eigenvalues of self-adjoint, semi-bounded operators, such as are commonly associated with boundary value problems in engineering and mathematical physics. The bounds obtained are complementary to those obtainable by the Rayleigh-Ritz procedure and together provide eigenvalue estimates with absolute error bounds.

Let A be a self-adjoint operator in a Hilbert space $\mathcal{H}$, densely defined on a domain $Dom(A) \subset \mathcal{H}$. Let $a(u)$ denote the closure of the associated quadratic form $\langle u, Au\rangle$ in $\mathcal{H}$. The Rayleigh-Ritz procedure proceeds as follows: Pick $\mathcal{R}_N = \text{span}\{r_1, r_2, \cdots, r_N\} \subset Dom(a)$; Assemble and solve the matrix eigenvalue problem:

$$\mathbf{Ax} = \Lambda\mathbf{Bx} \quad \text{with} \quad \begin{cases} \mathbf{A} = [a(r_i, r_j)]_{i,j=1}^N \\ \mathbf{B} = [\langle r_i, r_j\rangle]_{i,j=1}^N \end{cases} \in \mathbb{C}^{N\times N}$$

for $\Lambda_1^{(N)} \le \Lambda_2^{(N)} \le \cdots \le \Lambda_N^{(N)}$.

The bounding properties of $\Lambda_1^{(N)} \le \Lambda_2^{(N)} \le \cdots \le \Lambda_N^{(N)}$ are founded on the min-max characterization of the (lower) eigenvalues of self-adjoint operators, an important tool in spectral analysis:

Theorem *(Courant, Fischer, Weyl): Let $\lambda_\infty(A)$ denote the lowest point of the essential spectrum for A. Define $\lambda(p)$ for any finite index p as*

$$\lambda(p) = \inf_{\substack{\mathcal{U}\subset Dom(A) \\ \dim\mathcal{U}\le p}} \sup_{\substack{u\in\mathcal{U} \\ u\neq 0}} \frac{\langle u,\, Au\rangle}{\langle u,\, u\rangle}.$$

Then if $\lambda(p) < \lambda_\infty(A)$,

- *there exist at least p eigenvalues of A below $\lambda_\infty(A)$ and*
- *the p^{th} algebraically smallest eigenvalue of A (counting multiplicity) is given by $\lambda_p(A) = \lambda(p)$.*

The bounding properties of the Rayleigh-Ritz procedure follows directly from this, since

$$\lambda_\ell(A) = \min_{\dim \mathcal{U}=\ell} \max_{v\in\mathcal{U}} \frac{\langle v,\, Av\rangle}{\langle v,\, v\rangle} \leq \min_{\substack{\dim\mathcal{U}=\ell \\ \mathcal{U}\subset\mathcal{R}_N}} \max_{v\in\mathcal{U}} \frac{\langle v,\, Av\rangle}{\langle v,\, v\rangle} = \Lambda_\ell^{(N)}$$

The computation of complementary eigenvalue lower bounds is intrinsically more difficult than the computation of upper bounds, and it appears necessary to incorporate *a priori* spectral information that the Rayleigh-Ritz procedure does not require. Lower bounds can be obtained by several methods, each with different needs for such *a priori* information.

The method of intermediate problems, originated by Alexander Weinstein [5], is founded on the observation that oftentimes there can be found an operator, *a base operator*, whose eigenvalues give lower bounds (possibly quite crude) to the eigenvalues of the original problem and whose eigenfunctions are known and simple enough to be used in numerical computations. Starting with such a base operator, a sequence of intermediate operators is formed in such a way so that the eigenvalues of each operator are never less than the corresponding eigenvalues of the preceding operator and yet never larger than the corresponding eigenvalues of the original problem. Most importantly, the intermediate operators are formed so as to allow the resolution of the corresponding eigenvalue problems explicitly using the eigenvalues and eigenfunctions of the base operator.

Intermediate problem techniques often are able to provide accurate lower bounds for eigenvalues in many applications. However, the necessity of using base operator eigenfunctions in numerical computations can produce serious computational difficulties as well severely restricting the choice of base operators.

The Lehmann-Maehly [1, 2] method for computation of eigenvalue lower bounds makes use of a parameter which separates two consecutive eigenvalues with known index. This is tantamount to requiring a sufficiently accurate lower bound to an eigenvalue of a chosen index. Such information is at times either unavailable or must be obtained independently by other methods – but when such a parameter is known, the Lehmann-Maehly method can be a very effective tool.

The method of Weinberger [4] for computation of eigenvalue lower bounds is based on theoretical foundations that are somewhat distinct from that of the method of intermediate problems and of Lehmann-Maehly methods while at the same time generalizing them both. While being a substantial theoretical advancement over earlier methods, in its practical application it generally suffers from the same difficulties as the method of intermediate problems.

For the methods presented in this work, we make use of a base operator as well, though the only *a priori* spectral data used are lower bounds to the base problem eigenvalues. Neither eigenfunction information nor exact eigenvalue information for the base operator is necessary for the method to produce rigorous results. Unlike the Lehmann-Maehly method, our approach does not require a separation parameter for the problem under study. However, if such a parameter is known, we are able to recover Lehmann-Maehly bounds with our method.

The first set of results that are offered are applicable to self-adjoint operators projected and restricted to subspaces within their domain of definition. In particular, suppose A is a restriction of a self-adjoint operator A_0 in the sense that for some closed subspace, $\mathcal{P}$, of $\mathcal{H}$ (and associated orthogonal projection, $\mathbb{P}$), suppose that $A = \mathbb{P}A_0|_{\mathcal{P}}$ on $Dom(A) = \mathcal{P} \cap Dom(A_0)$. Notice that $a(u) = a_0(\mathbb{P}u)$ and $Dom(A_0) \supset Dom(A)$. Thus A_0 constitutes a base operator relative to A.

Theorem 1. *Suppose a value τ_0 and an index $m_0 > 1$ are known so that*

$$\lambda_{m_0-1}(A_0) < \tau_0 \leq \lambda_{m_0}(A_0).$$

Define the "Temple quotient relative to A_0" as

$$\widehat{T}_{\tau_0}(u) = \tau_0 + \frac{\|\mathbb{P}(A_0 - \tau_0)u\|^2}{\langle u, (A_0 - \tau_0)u\rangle}$$

and the cone in $\mathcal{H}$: $\mathcal{Y}_0 = \{u \in Dom(A_0) \mid \langle u, (A_0 - \tau_0)u\rangle \leq 0\}$.

Then:

$$\lambda_{m_0-\ell}(A) = \max_{\substack{\dim \mathcal{S}=\ell \\ \mathcal{S}\subset\mathcal{Y}^0}} \min_{u\in\mathcal{S}} \widehat{T}_{\tau_0}(u)$$

Restricting the max to a finite dimensional subspace leads to a computationally feasible problem yielding lower bounds: Choose trial functions,

$$\mathcal{Q}_N = span\{q_1, q_2, \cdots, q_N\} \subset Dom(A_0).$$

Then, analogously to the argument above that shows the Rayleigh-Ritz method produces eigenvalue upper bounds, we have

$$\lambda_{m_0-\ell}(A) = \max_{\substack{\dim \mathcal{S}=\ell \\ \mathcal{S}\subset\mathcal{Y}^0}} \min_{u\in\mathcal{S}} \widehat{T}_{\tau_0}(u) \geq \max_{\substack{\dim \mathcal{S}=\ell \\ \mathcal{S}\subset\mathcal{Y}^0\cap\mathcal{Q}_N}} \min_{u\in\mathcal{S}} \widehat{T}_{\tau_0}(u) = \tau_0 + \frac{1}{\Theta_\ell}$$

where $\Theta_1\, \Theta_1\, \ldots, \Theta_\ell$ are the (negative) eigenvalues of

$$\widehat{\mathbf{F}}\,\mathbf{x} = \Theta\,\widehat{\mathbf{G}}\,\mathbf{x} \quad \text{with} \quad \begin{cases} \widehat{\mathbf{F}} = [\langle (A_0 - \tau_0)q_i, q_j\rangle], \quad \text{and} \\ \widehat{\mathbf{G}} = [\langle (A_0 - \tau_0)q_i, \mathbb{P}(A_0 - \tau_0)q_j\rangle] \end{cases}$$

When $\mathbb{P} = I$ (so that $A = A_0$), this approach reduces to the Lehmann-Maehly method.

The second set of results that are offered follow the same pattern seen above for restrictions of quadratic forms, but instead for operators that are defined as a sum of operators (or more generally via a sum of closed quadratic forms). Specifically, suppose $A = A_1 + A_2$ is self-adjoint and densely defined on $Dom(A) = Dom(A_1) \cap Dom(A_2)$ with A_1, A_2 self-adjoint, semibounded, and densely defined on $Dom(A_1)$ and $Dom(A_2)$, respectively. An analogous variational characterization of eigenvalues of operator sums then appears as follows:

Theorem 2. *Suppose separating parameters for A_1 and A_2 are known: $\lambda_{m_1-1}(A_1) < \tau_1 \le \lambda_{m_1}(A_1)$ and $\lambda_{m_2-1}(A_2) < \tau_2 \le \lambda_{m_2}(A_2)$.*

For $\mathbf{u} = \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} \in Dom(A_1) \oplus Dom(A_2)$, *define*

$$\widetilde{T}(\mathbf{u}) = \tau_1 + \tau_2 + \frac{\|(A_1 - \tau_1)u_1 + (A_2 - \tau_2)u_2\|^2}{\langle u_1, (A_1 - \tau_1 I)u_1\rangle + \langle u_2, (A_2 - \tau_2)u_2\rangle}$$

and the cone in $\mathcal{H} \oplus \mathcal{H}$*:*

$$\mathcal{Y}^{\oplus} = \left\{ \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \in Dom(A_1) \oplus Dom(A_2)) \,|\langle u_1, (A_1 - \tau_1)u_1\rangle + \langle u_2, (A_2 - \tau_2)u_2\rangle \le 0 \right\}$$

Then with $k = m_1 + m_2 - 1$

$$\lambda_{k-\ell}(A) = \max_{\substack{\dim \mathcal{S} = \ell \\ \mathcal{S} \subset \mathcal{Y}^{\oplus}}} \min_{\mathbf{u} \in \mathcal{S}} \widetilde{T}(\mathbf{u})$$

As before, restricting the max to a finite dimensional subspace will lead to a computationally feasible problem that yields lower bounds: Choose trial functions:

$$\mathcal{P}_N = span\{p_1, p_2, \cdots, p_N\} \subset Dom(A_1), \text{ and}$$

$$\mathcal{Q}_M = span\{q_1, q_2, \cdots, q_M\} \subset Dom(A_2).$$

Define $\mathcal{Z}_{NM} = \mathcal{P}_N \oplus \mathcal{Q}_M$. Then,

$$\lambda_{k-\ell}(A) \ge \max_{\substack{\dim \mathcal{S} = \ell \\ \mathcal{S} \subset \mathcal{Y}^{\oplus} \cap \mathcal{Z}_{NM}}} \min_{\mathbf{u} \in \mathcal{S}} \widetilde{T}(\mathbf{u}) = \tau_1 + \tau_2 + \frac{1}{\Theta_\ell}$$

where $\Theta_1\, \Theta_1\, \ldots, \Theta_\ell$ are the (negative) eigenvalues of

$$\begin{bmatrix} \widetilde{\mathbf{F}}_1 & 0 \\ 0 & \widetilde{\mathbf{F}}_2 \end{bmatrix} \mathbf{x} = \Theta \begin{bmatrix} \widetilde{\mathbf{G}}_{11} & \widetilde{\mathbf{G}}_{12} \\ \widetilde{\mathbf{G}}_{21} & \widetilde{\mathbf{G}}_{22} \end{bmatrix} \mathbf{x} \tag{1}$$

$$\text{with} \begin{cases} \widetilde{\mathbf{F}}_1 = [\langle (A_1 - \tau_1)p_i, p_j\rangle], & \widetilde{\mathbf{G}}_{11} = [\langle (A_1 - \tau_1)p_i, (A_1 - \tau_1)p_j\rangle] \\ \widetilde{\mathbf{F}}_2 = [\langle (A_2 - \tau_2)q_i, q_j\rangle], & \widetilde{\mathbf{G}}_{22} = [\langle (A_2 - \tau_2)q_i, (A_2 - \tau_2)q_j\rangle] \\ \widetilde{\mathbf{G}}_{12} = [\langle (A_1 - \tau_1)p_i, (A_2 - \tau_2)q_j\rangle] & \widetilde{\mathbf{G}}_{21} = [\langle (A_2 - \tau_2)q_i, (A_1 - \tau_1)p_j\rangle] \end{cases}$$

As a simple illustration, consider a model of a rotating uniform beam

$$\frac{d^4u}{dx^4} - \frac{\alpha^2}{2}\frac{d}{dx}\left(1 - x^2\right)\frac{du}{dx} = \lambda u \quad \text{with} \quad u(0) = u'(0) = u''(1) = u'''(1) = 0$$

Express the operator as $A = A_1 + A_2$ where

- $A_1 = \dfrac{d^4u}{dx^4}$ with $Dom(A_1) = \left\{ u \in H^4(0,1) \,\middle|\, \begin{matrix} u(0) = u'(0) = 0 \\ u''(1) = u'''(1) = 0 \end{matrix} \right\}$

and

- $A_2 = -\frac{\alpha^2}{2}\frac{d}{dx}\left(1-x^2\right)\frac{du}{dx}$ with

$$Dom(A_2) = \left\{ u \in H^2(0,1) \;\middle|\; \begin{array}{c} u(0) = 0 \\ \lim_{x\to 1}(1-x)u'(x) = 0 \end{array} \right\}$$

Observe that eigenvalues of A_1 are λ^4 for roots, λ, of $\cosh(\lambda)\cos(\lambda)+1=0$. Eigenvalues of A_2 are explicitly computable as $\lambda_k(A_2) = \alpha^2\, k(2k-1)$ for $k = 1, 2, \ldots$ We take trial functions $p_k(x) = \cos(k\pi x) + a_k \cos\left((k+1)\pi x\right) - (1+a_k)$ for $k = 1, \ldots, N$ with $a_1, a_2, \ldots$ chosen so that p_k match boundary conditions for $Dom(A_1)$ and $q_j(x) = P_{2j-1}$ for $j = 1, 2, \ldots, M$ where $P_i(x)$ are Legendre polynomials of order i (which are also eigenfunctions of A_2). For $\alpha = 10$, we compute upper bounds to $\lambda_2(A)$ with a Rayleigh-Ritz problem of order $N = 3$ using trial functions p_1, p_2, , p_3, providing the upper bound

$$\lambda_2(A) \le \Lambda_2^{(3)} \le 2304.$$

An *a priori* lower bound for $\lambda_2(A)$ may be computed directly from the eigenvalues of A_1 and A_2 as

$$\max\{\lambda_1(A_1) + \lambda_2(A_2),\ \lambda_2(A_1) + \lambda_1(A_2)\} = 612.3 \le \lambda_2(A)$$

An improved lower bound may be obtained by taking first

$$\lambda_1(A_1) < \tau_1 = 485. < \lambda_2(A_1) \quad \text{and} \quad \lambda_2(A_1) < \tau_2 = 1400. < \lambda_3(A_2).$$

Then next forming and evaluating the generalized eigenvalue problem (1) for $N = 1$ and $M = 2$ and using Theorem 2, which yields

$$\tau_1 + \tau_2 + \frac{1}{\Theta_2} = 1080.7 \le \lambda_2(A).$$

Note that the main computation involved two 3×3 matrix eigenvalue problem and yields a rigorous[1] conclusion: $\lambda_2(A) \in (1080,\ 2304)$.

References

[1] N. J. Lehmann: Beiträge zur numerischem Lösung linearer Eigenwertprobleme I, II. Z. Angew. Math. Mech. **29**, 341–365 (1949); **30**, 1–16 (1950)

[2] H. J. Maehly: Ein neues Variationsverfahren zur genäherten Berechnung der Eigenwerte hermitesher operatoren, Helv. Phys. Acta **25**, 547–568 (1952)

[3] Siegfried M. Rump: *INTLAB Interval Laboratory*, Springer (1999)

[4] H. F. Weinberger: Variational Methods for Eigenvalue Problems, Philadelphia: SIAM, 1974

[5] A. Weinstein: Etude des spectres des equations aux derivees partielles de la theorie des plaques elastiques, Mémorial des sciences mathématiques, vol. **88**, Gauthier-Villars, Paris (1937)

[1] extra care using rigorous interval methods is needed for this last step, in fact. We omit discussion of this issue but direct the reader to [3].

Lyapunov Inverse Iteration for Rightmost Eigenvalues of Generalized Eigenvalue Problems

Howard Elman
(joint work with Minghao Wu)

This project concerns an efficient algorithm for computing a few rightmost eigenvalues of generalized eigenvalue problems. We are concerned with problems of the form

$$\mathcal{J}(\alpha)x = \mu \mathbf{M}x \tag{1}$$

arising from linear stability analysis (see [2]) of the dynamical system

$$\mathbf{M}u_t = f(u, \alpha). \tag{2}$$

$\mathbf{M} \in \mathbb{R}^{n\times n}$ is called the mass matrix, and the parameter-dependent matrix $\mathcal{J}(\alpha) \in \mathbb{R}^{n\times n}$ is the Jacobian matrix $\frac{\partial f}{\partial u}(\overline{u}(\alpha), \alpha) = \frac{\partial f}{\partial u}(\alpha)$, where $\overline{u}(\alpha)$ is the steady-state solution to (2) at α, *i.e.*, $f(\overline{u}, \alpha) = 0$. Let the solution path be the following set: $\mathcal{S} = \{(\overline{u}, \alpha) | f(\overline{u}, \alpha) = 0\}$. We seek the critical point $(\overline{u}_c, \alpha_c)$ associated with transition to instability on $\mathcal{S}$. While the method developed in this study works for any dynamical system of the form (2), our primary interest is the ones arising from spatial discretization of 2- or 3-dimensional time-dependent partial differential equations (PDEs). Therefore, we assume n to be large and $\mathcal{J}(\alpha), \mathbf{M}$ to be sparse throughout this paper.

The conventional method of locating the critical parameter α_c is to monitor the rightmost eigenvalue(s) of (1) while marching along $\mathcal{S}$ using numerical continuation (see [2]). In the stable regime of $\mathcal{S}$, the eigenvalues μ of (1) all lie to the left of the imaginary axis. As $(\overline{u}, \alpha)$ approaches the critical point, the rightmost eigenvalue of (1) moves towards the imaginary axis; at $(\overline{u}_c, \alpha_c)$, the rightmost eigenvalue of (1) has real part zero, and finally, in the unstable regime, some eigenvalues of (1) have positive real parts. The continuation usually starts from a point $(\overline{u}_0, \alpha_0)$ in the stable regime of $\mathcal{S}$ and the critical point is detected when the real part of the rightmost eigenvalue of (1) becomes nonnegative. Consequently, robustness and efficiency of the eigenvalue solver for the rightmost eigenvalue(s) of (1) are crucial for the performance of this method. Direct eigenvalue solvers such as the QR and QZ algorithms (see [4]) compute all the eigenvalues of (1), but they are too expensive for large n. Existing iterative eigenvalue solvers [4] are able to compute a small set ($k \ll n$) of eigenvalues of (1) near a given shift (or target) $\sigma \in \mathbb{C}$ efficiently. For example, they work well when k eigenvalues of (1) with smallest modulus are sought, in which case $\sigma = 0$. One issue with such methods is that there is no robust way to determine a good choice of σ when we have no idea where the target eigenvalues may be. In the computation of the rightmost eigenvalue(s), the most commonly used heuristic choice for σ is zero, *i.e.*, we compute k eigenvalues of (1) with smallest modulus and hope that the rightmost one is one of them. When the rightmost eigenvalue is real, zero is a good choice. However, such an approach is not robust when the rightmost eigenvalues consist of a complex conjugate pair: the rightmost pair can be far away from zero and it is not clear how big k should

be to ensure that they are found. Such examples can be found in the numerical experiments of this study.

Meerbergen and Spence [3] proposed the *Lyapunov inverse iteration* method, which estimates the critical parameter α_c without computing the rightmost eigenvalues of (1). Assume $(\overline{u}_0, \alpha_0)$ is in the stable regime of $\mathcal{S}$ and is also in the neighborhood of the critical point $(\overline{u}_c, \alpha_c)$. Let $\lambda_c = \alpha_c - \alpha_0$ and $\mathbf{A} = \mathcal{J}(\alpha_0)$. Then the Jacobian matrix $\mathcal{J}(\alpha_c)$ at the critical point can be approximated by $\mathbf{A} + \lambda_c \mathbf{B}$ where $\mathbf{B} = \frac{d\mathcal{J}}{d\alpha}(\alpha_0)$. It is shown in [3] that λ_c is the eigenvalue with smallest modulus of the eigenvalue problem

$$\mathbf{A}Z\mathbf{M}^T + \mathbf{M}Z\mathbf{A}^T + \lambda(\mathbf{B}Z\mathbf{M}^T + \mathbf{M}Z\mathbf{B}^T) = 0 \tag{3}$$

of Lyapunov structure and that λ_c can be computed by a matrix version of inverse iteration. Estimates of the rightmost eigenvalue(s) of (1) at α_c can be obtained as by-products. Elman *et al.* [1] refined the Lyapunov inverse iteration proposed in [3] to make it more robust and efficient and examined its performance on challenging test problems arising from fluid dynamics. Various implementation issues were discussed, including the use of inexact inner iterations, the impact of the choice of iterative method used to solve the Lyapunov equations, and the effect of eigenvalue distribution on performance. Numerical experiments demonstrated the robustness of their algorithm.

The method proposed in [1, 3], although it allows us to estimate the critical value of the parameter without computing the rightmost eigenvalue(s) of (1), only works in the neighborhood of the critical point $(\overline{u}_c, \alpha_c)$. In [1], for instance, the critical parameter value α_c of all numerical examples is known *a priori*, so that we can pick a point $(\overline{u}_0, \alpha_0)$ close to $(\overline{u}_c, \alpha_c)$ and apply Lyapunov inverse iteration with confidence. In reality, α_c is unknown and we start from a point $(\overline{u}_0, \alpha_0)$ in the stable regime of $\mathcal{S}$ that may be distant from the critical point. In this scenario, the method of [1, 3] cannot be used to estimate α_c, since $\mathcal{J}(\alpha_c)$ cannot be approximated by $\mathbf{A} + \lambda_c \mathbf{B}$. However, quantitative information about how far away $(\overline{u}_0, \alpha_0)$ is from $(\overline{u}_c, \alpha_c)$ can still be obtained by estimating the distance between the rightmost eigenvalue of (1) at α_0 and the imaginary axis: if the rightmost eigenvalue is far away from the imaginary axis, then it is reasonable to assume that $(\overline{u}_0, \alpha_0)$ is far away from the critical point as well, and therefore we should march along $\mathcal{S}$ using numerical continuation until we are close enough to $(\overline{u}_c, \alpha_c)$; otherwise, we can assume that $(\overline{u}, \alpha_0)$ is already in the neighborhood of the critical point and the method of [1, 3] can be applied to estimate α_c.

In this study, we develop a robust method to compute a few rightmost eigenvalues of (1) in the stable regime of $\mathcal{S}$. We show that the distance between the imaginary axis and the rightmost eigenvalue of (1) is the eigenvalue with smallest modulus of an eigenvalue problem similar in structure to (3). As a result, this eigenvalue can be computed efficiently by Lyapunov inverse iteration. We present numerical results for several examples arising from fluid dynamics, which demonstrate the fast convergence of this method, and we give an analysis that provides insight into the fast convergence. In addition, based on this analysis, we propose

a more efficient version of Lyapunov inverse iteration and a way of validating its results.

References

[1] H. C. Elman, K. Meerbergen, A. Spence, and M. Wu, *Lyapunov inverse iteration for identifying Hopf bifurcations in models of incompressible flow*, SIAM J. Sci. Comput., 34 (2012), pp. A1584–A1606.

[2] W. Govaerts, *Numerical Methods for Bifurcations of Dynamical Equilibria*, SIAM, Philadelphia, 2000.

[3] K. Meerbergen and A. Spence, *Inverse iteration for purely imaginary eigenvalues with application to the detection of Hopf bifurcation in large scale problems*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 1982–1999.

[4] G. W. Stewart, *Matrix Algorithms Volume II: Eigensystems*, SIAM, Philadelphia, 2001.

Field of values type eigenvalue inclusion regions

Michiel Hochstenbach

(joint work with David A. Singer, Paul F. Zachlin, Ian N. Zwaan)

In the talk we have given a brief overview of recent developments in the generation of fast and tight spectral inclusion regions for large sparse matrices, based on the field of values

$$W(A) = \{\, \mathbf{x}^* A \mathbf{x} \;:\; \|\mathbf{x}\|_2 = 1 \,\}$$

and generalizations thereof. We give a brief overview of some key aspects.

Matrix scaling

Denote the spectral radius by $\rho(A) = \max |\lambda|$, and the numerical radius by $r(A) = \max_{z \in W(A)} |z|$. For a tight spectral inclusion region, we hope that

$$\frac{r(A)}{\rho(A)} \approx 1.$$

However, for some matrices, such as `tols4000` [9], we have $r(A) \gg \rho(A)$. We may try to improve this situation using scaling, which leaves the eigenvalues invariant but may change the field of values. We get

$$\frac{r(D^{-1}AD)}{\rho(D^{-1}AD)} = \frac{r(D^{-1}AD)}{\rho(A)} \ll \frac{r(A)}{\rho(A)}$$

if we manage to find a scaling matrix D such that $r(D^{-1}AD) \ll r(A)$. In view of the "squeeze theorem" [6, p. 331]

$$\tfrac{1}{2}\,\|A\| \le r(A) \le \|A\|,$$

we try to find D such that $\|D^{-1}AD\| \ll \|A\|$. Various (Krylov) scaling techniques for this situation were investigated and developed in [5] (see also [1]), showing that scaling of a matrix may be a very helpful technique for generating tight spectral inclusion regions based on a field of values. In fact, we believe that the combination

of matrix scaling and a field of values based on an Arnoldi decomposition gives an eigenvalue inclusion region that is very hard to beat both in quality and efficiency.

Computing interior eigenvalues of large matrices may be very challenging. However, sometimes computing *exterior* eigenvalues also may take many matrix-vector products. It is quite surprising that with (say) 20 matrix-vector products we can often get a good idea of the location of *all* eigenvalues. In Figure 1, we plot $W(H_{20})$, an eigenvalue inclusion region, and several exclusion regions for the matrix `grcar` of dimension $n = 1000$; see also the rest of this extended abstract for more information.

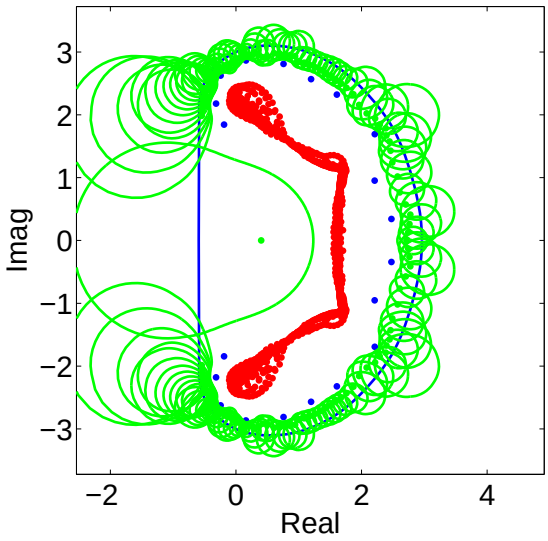


FIGURE 1. Eigenvalue inclusion and exclusion region after 20 Krylov steps for the `grcar` matrix of dimension $n = 1000$. Indicated are the eigenvalues (red), $W(H_{20})$ (blue), the Ritz value (blue), and exclusion regions based on several τ (green).

Projection

As for the solution of large-scale linear systems and eigenvalue problems, projection onto a Krylov subspace is a very useful technique to approximate the field of values. Let

$$AV_k = V_k\, H_k + h_{k+1,k}\, \mathbf{v}_{k+1}\, \mathbf{e}_k^*$$

be the Arnoldi decomposition after k steps, starting with $\mathbf{v}_1$ with unit norm. The approximation

$$W(H_k) \subset W(H_{k+1}) \subset W(A)$$

has already been described by Manteuffel [7] and also mentioned in [8, 10]. The convergence of the sets $W(H_k)$ to the set $W(A)$ is often quite rapid. In Figure 2, we plot the relative change of the area of $W(H_k)$:

$$\frac{\text{area}(W(H_{k+1})) - \text{area}(W(H_k))}{\text{area}(W(H_k))}$$

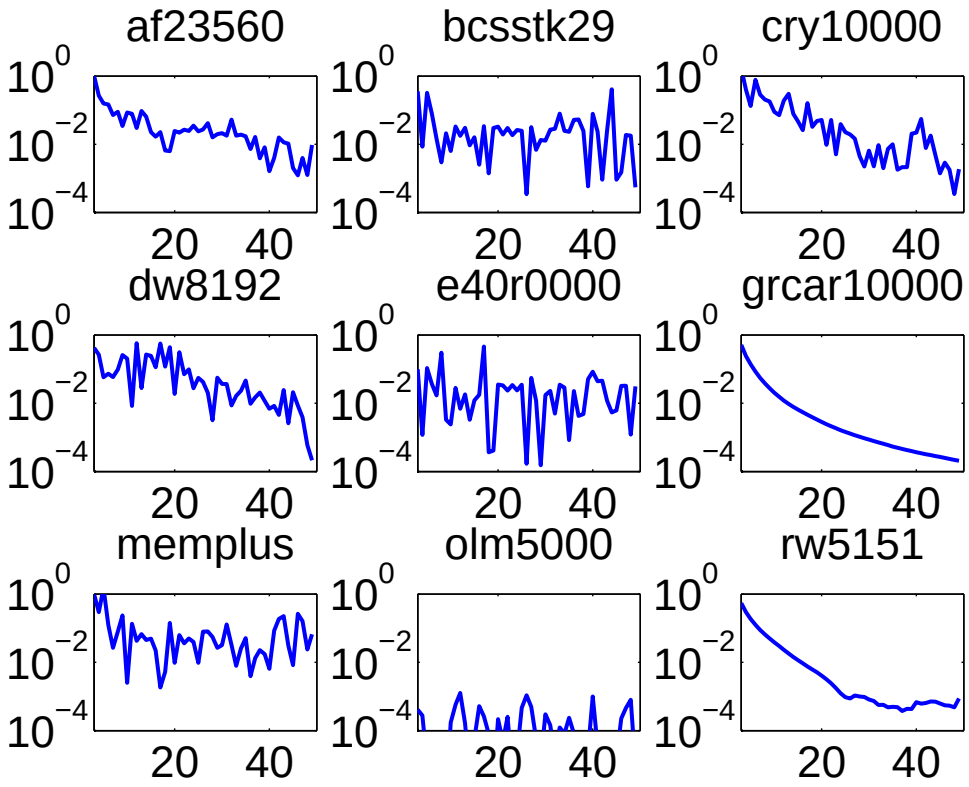


FIGURE 2. The relative increase of the area of $W(H_k)$ as a function of k for several large matrices.

as a function of k for various large matrices from [9].

Exclusion regions

The set $W(H_k)$ is always a convex and compact inclusion region. In [3] (see also [4]), the following equality was proved:

$$\Lambda(A) = \bigcap_{\tau \in \mathbb{C} \setminus \Lambda(A)} \frac{1}{W((A-\tau I)^{-1})} + \tau,$$

and used in a practical way by automatically selecting a finite number of targets τ, and using Krylov projections instead of the large matrices.

Generalized eigenvalue problem

Field of values based inclusion regions for the generalized eigenvalue problem $A\mathbf{x} = \lambda B\mathbf{x}$ (that is, matrix pencils) are studied in [2]. Key aspects are approximations to the sets $W(B^{-1}A)$, $W(AB^{-1})$, $\frac{1}{W((B^{-1}A-\tau I)^{-1})} + \tau$, and $\frac{1}{W((AB^{-1}-\tau I)^{-1})} + \tau$, obtained by projection onto a Krylov space.

REFERENCES

[1] T.-Y. CHEN AND J. W. DEMMEL, *Balancing sparse matrices for computing eigenvalues*, Linear Algebra and Its Applications, 309 (2000), pp. 261–287.

[2] M. E. Hochstenbach, *Fields of values and inclusion regions for matrix pencils*, Electron. Trans. Numer. Anal., 38 (2011), pp. 98–112.
[3] M. E. Hochstenbach, D. A. Singer, and P. F. Zachlin, *Eigenvalue inclusion regions from inverses of shifted matrices*, Linear Algebra Appl., 429 (2008), pp. 2481–2496.
[4] ———, *Numerical approximation of the field of values of the inverse of a large matrix*, Textos de Matematica, 44 (2013), pp. 59–71.
[5] M. E. Hochstenbach and I. N. Zwaan, *Matrix balancing for field of values type inclusion regions*, preprint, TU Eindhoven, October 2013. Submitted.
[6] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.
[7] T. A. Manteuffel, *Adaptive procedure for estimating parameters for the nonsymmetric tchebychev iteration*, Numer. Math., 31 (1978), pp. 183–208.
[8] T. A. Manteuffel and G. Starke, *On hybrid iterative methods for nonsymmetric systems of linear equations*, Numer. Math., 73 (1996), pp. 489–506.
[9] *The Matrix Market.* `http://math.nist.gov/MatrixMarket`, a repository for test matrices.
[10] K.-C. Toh and L. N. Trefethen, *Calculation of pseudospectra by the Arnoldi iteration*, SIAM J. Sci. Comput., 17 (1996), pp. 1–15.

Something about Numerical Approximation of PDE Eigenvalue Problems

Zhimin Zhang

1. How many numerical eigenvalues can we trust? [9]

When approximating PDE eigenvalue problems by numerical methods such as finite difference and finite element methods, it is a common knowledge that only a small portion of numerical eigenvalues are reliable. However, this knowledge is only qualitative rather than quantitative in the literature. Here we investigate the number of "trusted" eigenvalues by the finite element approximation of $2m$-th order elliptic PDE eigenvalue problems. Our two model problems are the Laplace and bi-harmonic operators. We show that the number of reliable numerical eigenvalues can be estimated in terms of the total degree of freedom N of resulting discrete systems. The result is worse than what we used to believe in that the percentage of reliable eigenvalues decreases with an increased N. As an example, we consider eigenvalues of the Laplace operator calculated by linear, bilinear finite element methods, or by 5-point, 9-point finite difference schemes. If we want numerical eigenvalues converge at least linearly, the number of reliable eigenvalues are roughly $O(\sqrt{N})$ and the portion of the trusted eigenvalues is then $O(\frac{1}{\sqrt{N}})$.

2. Spectral and spectral collocation methods for eigenvalues of differential and integral operators [1, 5].

On the other hand, spectral methods have advantage for eigenvalue approximation in the sense that when eigenfunctions are sufficiently smooth, the portion of reliable eigenvalues is $\left(\frac{2}{\pi}\right)^d$, where d is the dimension. We see that although the advantage diminishes with increased d, the percentage of reliable eigenvalues is fixed for any fixed dimension.

As an example, we propose and analyze a C^0 spectral element method for a model eigenvalue problem with discontinuous coefficients in the one dimensional setting. A super-geometric rate of convergence is proved for piecewise constant coefficients case. For the model problem with one jump in the middle of the solution interval, the machine epsilon is reached at polynomial degree $N = 6$ on two elements.

We also propose and analyze a spectral collocation method to solve eigenvalue problems of compact integral operators, particularly, piecewise smooth operator kernels and weakly singular operator kernels of the form $|t - s|^{-\mu}$, $0 < \mu < 1$. The convergence rate of eigenvalue approximation depends upon the smoothness of the corresponding eigenfunctions, especially, for an eigenfunction with r^α type singularity, the typical "double" convergent rate $N^{-4\alpha}$ (comparing with $h^{2\alpha}$ of the h-version) of the p-version method is observed.

3. Enhance eigenvalue approximation by gradient recovery techniques [3, 4, 6].

The polynomial preserving recovery (PPR) is used to enhance the finite element eigenvalue approximation. Remarkable fourth order convergence is observed for linear elements under structured meshes as well as unstructured initial meshes (produced by the Delaunay triangulation) with the conventional bisection refinement. As for singular eigenfunctions under adaptive meshes, superconvergence is also achieved by using PPR.

Furthermore, function value recovery techniques for linear finite elements are discussed. Using the recovered function and its gradient, we are able to enhance the eigenvalue approximation and increase its convergence rate to $h^{2\alpha}$, where $\alpha > 1$ is the superconvergence rate of the recovered gradient. This is true in both symmetric and nonsymmetric eigenvalue problems.

4. Eigenvalue approximation from below by some non-conforming finite elements [7, 8, 10].

Consider general nonconforming finite element methods for eigenvalue problems in the abstract form:

$$a_h(u_h, v) = \lambda b(u_h, v).$$

Let u_I be any interpolation in the nonconforming finite element space, we have the following identity,

$$\lambda - \lambda_h = \|u - u_h\|_h^2 - \lambda_h \|u_I - u_h\|_b^2 + \lambda_h(\|u_I\|_b^2 - \|u_h\|_b^2) + 2a_h(u - u_I, u_h).$$

By estimating each term on the right, it is possible to show that the first term is the dominate one for some non-conforming elements. As a consequence, λ_h is smaller than λ (in contrary to conforming elements) and hence approximates λ from below. Here is a partial list of nonconforming finite elements that yield lower bounds for particular eigenvalue problems under certain meshes.

1) Wilson element, 2nd-order elliptic operator, rectangle and cube
2) Q_1^{rot} and EQ_1^{rot}, 2nd-order elliptic operator, Steklov eigen-problem, rectangle
3) Crouzeix-Raviart element, 2nd-order elliptic operator, n-simplex mesh;

Steklov eigen-problem, triangular mesh
4) Extended Crouzeix-Raviart element, 2nd-order elliptic operator, Stokes, and Steklov eigen-problems, triangular mesh
5) Morley element, 4th-order elliptic operators, n-simplex mesh
6) Adini element, 4th-order elliptic operator, rectangle and cube

References

[1] Can Huang, Hailong Guo, and Zhimin Zhang, A spectral collocation method for eigenvalue problems of compact integral operators, *Journal of Integral Equations and Applications* 25-1 (2013), 79-101.
[2] Ronald H.W. Hoppe, Haijun Wu, and Zhimin Zhang, Adaptive finite element methods for the Laplace eigenvalue problem, *Journal of Numerical Mathematics* 18-4 (2010), 281-302.
[3] Ahmed A. Naga and Zhimin Zhang, Function value recovery and its application in eigenvalue problems, SIAM Journal on Numerical Analysis 50-1 (2012), 272-286.
[4] Ahmed A. Naga, Zhimin Zhang, and Aihui Zhou, Enhancing eigenvalue approximation by gradient recovery, *SIAM Journal on Scientific Computing* 28-4 (2006), 1289-1300.
[5] Lin Wang, Ziqing Xie, and Zhimin Zhang, Super-geometric convergence of spectral element method for eigenvalue problems with jump coefficients, *Journal of Computational Mathematics* 28-3 (2010), 418-428.
[6] Haijun Wu and Zhimin Zhang, Enhancing eigenvalue approximation by gradient recovery II: adaptive meshes, *IMA Journal of Numerical Analysis* 29-4 (2009), 1008-1022.
[7] Yidu Yang, Fubiao Lin, and Zhimin Zhang, N-simplex Crouzeix-Raviart element for the second-order elliptic/eigenvalue problems, *International Journal for Numerical Analysis and Modeling* 6-4 (2009), 615-626.
[8] Yidu Yang, Zhimin Zhang, and Fubiao Lin, Eigenvalue approximation from below using non-conforming elements, *Science in China – Series A* 53-1 (2010), 137-150.
[9] Z. Zhang, How many numerical eigenvalues can we trust? `http://arxiv.org/abs/1312.6773`, 2013.
[10] Zhimin Zhang, Yidu Yang, and Zhen Chen, Eigenvalue approximation from below by Wilson's element, *Chinese Journal of Numerical Mathematics and Applications* 29-4 (2007), 81-84.

On the Convergence of the Residual Inverse Iteration for Nonlinear Eigenvalue Problems

Daniel Kressner

(joint work with Cedric Effenberger)

We consider nonlinear eigenvalue problems of the form

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

where $T : D \to \mathbb{C}^{n\times n}$ is a continuously differentiable matrix-valued function on some open interval $D \subset \mathbb{R}$.

In the following, $T(\lambda)$ is supposed to be Hermitian for every $\lambda \in D$. Moreover, we assume that the scalar nonlinear equation

$$x^* T(\lambda)x = 0 \tag{2}$$

admits a unique solution $\lambda \in D$ for every vector x in an open set $D_\rho \subset \mathbb{C}^n$. The resulting function $\rho : D_\rho \to D$, which maps x to the solution λ of (2), is called *Rayleigh functional*, for which we additionally assume that

$$x^* T'(\rho(x))x > 0 \qquad \forall x \in D_\rho.$$

The existence of such a Rayleigh functional entails a number of important properties for the eigenvalue problem (1), see [4, Sec. 115.2] for an overview. In particular, the eigenvalues in D are characterized by a min-max principle and thus admit a natural ordering. Specifically, if

$$\lambda_1 := \inf_{x \in D_\rho} \rho(x) \in D$$

then λ_1 is the first eigenvalue of T.

It is of interest to study the convergence of Neumaier's residual inverse iteration [3] for computing the eigenvalue λ_1 of T and an associated eigenvector x_1. In the Hermitian case, this iteration takes the form

$$v_{k+1} = \gamma_k(v_k + P^{-1}T(\rho(v_k))v_k), \quad k = 0, 1, \ldots, \tag{3}$$

with normalization coefficients $\gamma_k \in \mathbb{C}$, an initial guess $v_0 \in \mathbb{C}^n$, and a Hermitian preconditioner $P \in \mathbb{C}^{n \times n}$. Usually, $P = -T(\sigma)$ for some shift σ not too far away from λ_1 but the general formulation (3) allows for more flexibility, such as the use of multigrid preconditioners.

In [3, Sec. 3], it was shown that (3) with $P = T(\sigma)$ converges linearly to an eigenvector belonging to a simple eigenvalue, provided that σ is sufficiently close to that eigenvalue. Jarlebring and Michiels [2] derived explicit expressions for the convergence rate by viewing (3) as a fixed point iteration and considering the spectral radius of the fixed point iteration matrix.

Our new convergence analysis is tailored to the particular situation of having a Rayleigh functional, and differs significantly from [2, 3]. Our major motivation for reconsidering this question was to establish mesh-independent convergence rates when applying (3) with a multigrid preconditioner to the finite element discretization of a nonlinear PDE eigenvalue problem. The analyses in [2, 3] do not seem to admit such a conclusion, at least it is not obvious. On the other hand, such results are well known for the linear case $T(\lambda) = \lambda I - A$, for which (3) comes down to the preconditioned inverse iteration (PINVIT). In particular, the seminal work by Neymeyr establishes tight expressions for the convergence of the eigenvalue and eigenvector approximations produced by PINVIT. Unfortunately, the elegance of Neymeyr's mini-dimensional analysis of the Rayleigh-quotient is strongly tied to linear eigenvalue problems; there seems little hope to carry it over to the general nonlinear case. Our approach proceeds by directly analysing the convergence of the eigenvector. Although leading to weaker bounds than Neymeyr's analysis in the linear case, the obtained results still allow to establish mesh-independent convergence rates.

In the first step, we show that

$$\tan \phi_P(v_{k+1}, x_1) \leq \gamma \cdot \tan \phi_P(v_k, x_1) + O(\varepsilon^2),$$

where x_1 is an eigenvector belonging to λ_1 and ϕ_P denotes the angle in the geometry induced by P. In the second step, we show that $\gamma < 1$ (independent of h) for a multigrid preconditioner of $T(\sigma)$ with σ sufficiently close to λ_1.

References

[1] Effenberger, E., Kressner, D.: On the convergence of the residual inverse iteration for nonlinear eigenvalue problems admitting a Rayleigh functionals. In preparation, 2013.

[2] Jarlebring, E., Michiels, W.: Analyzing the convergence factor of residual inverse iteration. BIT **51**(4), 937–957 (2011).

[3] Neumaier, A.: Residual inverse iteration for the nonlinear eigenvalue problem. SIAM J. Numer. Anal. **22**(5), 914–923 (1985)

[4] Voss, H.: Nonlinear eigenvalue problems. In: L. Hogben (ed.) Handbook of Linear Algebra. Chapman & Hall/CRC, FL (2013).

Rational Krylov for nonlinear eigenvalue problems arising from PDEs

Karl Meerbergen

(joint work with Roel Van Beeumen and Wim Michiels)

The solution of the nonlinear eigenvalue problem, in its most general form, written as

$$A(\lambda)x = 0 \quad , \quad x \neq 0$$

where λ is the eigenvalue, is appearing more and more often in applications arising from PDEs. We present three examples.

The first example is the delay eigenvalue problem, which is related to the delay differential equation. Indeed, the delayed term leads to a term involving exponentials. For example, the standard delay equation

$$u'(t) + Au(t) + Bu(t-\tau) = f(t)$$

leads to the eigenvalue problem

$$\lambda x + Ax + e^{-\lambda\tau}Bx = 0.$$

The second example is the 1D Schroedinger equation on an infinite domain whose potential is flat everywhere except in an interval. The eigenvalue problem is formulated as the classical equation with boundary conditions that reflect the exponential decay of the eigenfunction for a flat potential. These boundary conditions lead to an eigenvalue problem of the form

$$(-D + \mathrm{diag}(U) + \Sigma(\lambda))x = \lambda x$$

where D is the Laplacian, U is the discrete potential, and Σ is a matrix that is zero everywhere except on both ends of the main diagonal, where terms $\exp(i\sqrt{\lambda - s_i})$, $i = 1, 2$, appear. As a result, Σ is a nonlinear term of the eigenvalue problem. The extension to the 2D equation leads to Σ with more nonzero elements, also including exponentials of square roots [4].

The third and last example arises from the study of nonlinear damping in mechanical engineering. New damping models are more and more often used for

more accurate numerical simulations. The example that we report on here is a model of a clamped sandwhich beam [3] that takes the form

$$K_e x + \frac{G_0 + G_\infty (i\lambda\tau)^\alpha}{1 + (i\lambda\tau)^\alpha} K_v - \lambda^2 M x = 0$$

with $\alpha = 0.675$ and $\tau = 8.230$.

We now discuss numerical methods for solving the nonlinear eigenvalue problem. All problem mentioned higher can be written as

$$(A_0 + \lambda A_1 + \sum_{i=1}^{m} f_i(\lambda) B_i) x = 0$$

where f_i are scalar functions in the complex plane. For the nonlinear eigenvalue problem, we have local search methods and global search methods. The prototype examples of local search methods are Newton's method and residual inverse iteration [2]. Global method first build a rational or polynomial approximation to f_j, $j = 1, \ldots, m$ and then solve the related polynomial or rational eigenvalue problem [1].

In this talk, we discuss methods that lie in between these two classes of methods. The Newton method and the residual inverse iteration method can be seen as methods that approximate $A(\lambda)$ by a polynomial of degree one. We build an interpolating polynomial of degree k for $A(\lambda)$ in the points $\sigma_0, \ldots, \sigma_k \in \mathbb{C}$ and then perform k iterations of the rational Krylov method on the linearization of the resulting polynomial eigenvalue problem. When we use Newton polynomials and choose the poles of the rational Krylov method equal to the nodes of the Newton interpolation, then the algorithm can be organized in such a way that the nodes need not be determined in advance. This allows for tuning these parameters during the execution of the algorithm. In this way, we obtain a method that converges in less iterations than, e.g, the Newton method. The method thus behaves like a local search method, but can be used for building a global approximation, but in a dynamic way [3].

An appealing extension of the method lies in the use of rational Newton polynomials, that also lead to a similar recurrence relation as the classical polynomials and also allow for an efficient implementation of the rational Krylov method. In this case, we choose the poles and nodes as Leja-Bagpy points for optimal approximation of the nonlinear functions. The method can then be used as a global search method by selecting poles on the boundary of the search domain, in a Leja fashion.

The efficient and reliable implementation uses matrix functions for computing the divided differences required for Newton interpolation, proper scaling of the base functions and takes into account low rank of the nonlinear terms.

References

[1] C. Effenberger and D. Kressner. Chebyshev interpolation for nonlinear eigenvalue problems. *BIT*, pages 1–19, 2012.

[2] A. Neumaier. Residual inverse iteration for the nonlinear eigenvalue problem. *SIAM J. Numer. Anal.*, 22:914–923, 1998.
[3] R. Van Beeumen, K. Meerbergen, and W. Michiels. A rational Krylov method based on Hermite interpolation for nonlinear eigenvalue problems. *SIAM J. Sci. Comput.*, 35(1):A327–A350, 2013.
[4] W. Vandenberghe, M. V. Fischetti, R. Van Beeumen, K. Meerbergen, W. Michiels, and C. Effenberger. Determining bound states in a semiconductor device with contacts using a non-linear eigenvalue solver, 2013. In preparation.

Solving symmetric quadratic eigenvalue problems with SLEPc

Jose E. Roman
(joint work with Carmen Campos)

This work [1] is framed in the context of SLEPc, the Scalable Library for Eigenvalue Problem Computations [2]. SLEPc contains parallel implementations of various eigensolvers for different types of eigenproblems. The linear eigenvalue problem solver (`EPS`) contains basic methods as well as more advanced algorithms, including Krylov-Schur, Generalized Davidson, Jacobi-Davidson, Rayleigh-quotient conjugate gradient or the contour integral spectral slicing technique. Eigensolvers can be combined with spectral transformations (such as shift-and-invert) or preconditioners in the case of preconditioned eigensolvers. Linear solvers as well as data structures for matrices and vectors are provided by PETSc [3], on which SLEPc is based. Further functionality of SLEPc includes solvers for the partial singular value decomposition, as well as quadratic and general nonlinear eigenproblems.

For quadratic eigenvalue problems (`QEP`) we provide a solver based on linearization, as well as various memory-efficient solvers. The former explicitly builds the matrices of a companion linearization of the quadratic problem and then invokes a linear eigensolver from `EPS` to obtain the solution. The latter include the Q-Arnoldi and TOAR methods, that aim at exploiting the structure of the linearization in such a way that memory requirements for storing the Krylov basis are roughly divided by two.

In this work we investigate how to adapt the Q-Arnoldi method [4] for the case of symmetric quadratic eigenvalue problems, that is, we are interested in computing a few eigenpairs (λ, x) of $(\lambda^2 M + \lambda C + K)x = 0$ with M, C, K symmetric $n \times n$ matrices. This problem has no particular structure, in the sense that eigenvalues can be complex or even defective. Still, symmetry of the matrices can be exploited to some extent. For this, we perform a symmetric linearization $Ay = \lambda By$, where A, B are symmetric $2n \times 2n$ matrices but the pair (A, B) is indefinite and hence standard Lanczos methods are not applicable. We implement a symmetric-indefinite Lanczos method [5] and enrich it with a thick-restart technique [6]. This method uses pseudo inner products induced by matrix B for the orthogonalization of vectors (indefinite Gram-Schmidt). Restarting the pseudo-Lanczos recurrence requires special ways of solving the projected problems, using techniques such as those described in [7], that try to minimize the use of non-orthogonal transformation to try to elude instability.

The next step is to write a specialized, memory-efficient version that exploits the block structure of A and B, referring only to the original problem matrices M, C, K as in the Q-Arnoldi method. This results in what we have called the Q-Lanczos method. The Q-Arnoldi method may suffer from instability when the Hessenberg matrix of the Arnoldi relation has large norm, and so may Q-Lanczos. Therefore, we need to define a stabilized variant analog of the TOAR method, which represents the basis vectors of the pseudo-Lanczos recurrence as the product of two matrices that are orthogonal with respect to some non-standard inner product (STOAR). Restarting in this case is more complicated and involves computing the SVD of a small matrix.

We show results obtained with parallel implementations of all methods in SLEPc, when solving several problems from the NLEVP collection [8]. Although the methods relying on an indefinite inner product are not guaranteed to be stable, we observe reasonably good behaviour of the pseudo-Lanczos method operating on the explicit linearization as well as the STOAR variant.

References

[1] C. Campos and J. E. Roman, *Restarted Q-Arnoldi-type methods exploiting symmetry in quadratic eigenvalue problems*, submitted (2013).

[2] V. Hernandez, J. E. Roman, and V. Vidal, *SLEPc: A scalable and flexible toolkit for the solution of eigenvalue problems*, ACM Trans. Math. Software **31** (2005), 351–362.

[3] S. Balay, J. Brown, K. Buschelman, V. Eijkhout, W. Gropp, D. Kaushik, M. Knepley, L. C. McInnes, B. Smith, and H. Zhang, PETSc users manual, Tech. Report ANL-95/11 - Revision 3.4, Argonne National Laboratory, (2013).

[4] K. Meerbergen, The Quadratic Arnoldi method for the solution of the quadratic eigenvalue problem, SIAM J. Matrix Anal. Appl. **30** (2008), 1463–1482.

[5] B. N. Parlett and H. C. Chen, *Use of indefinite pencils for computing damped natural modes*, Linear Algebra Appl. **140** (1990), 53–88.

[6] K. Wu and H. Simon, *Thick-restart Lanczos method for large symmetric eigenvalue problems*, SIAM J. Matrix Anal. Appl. **22** (2000), 602–616.

[7] F. Tisseur, *Tridiagonal-diagonal reduction of symmetric indefinite pairs*, SIAM J. Matrix Anal. Appl. **26** (2004), 215–232.

[8] T. Betcke, N. J. Higham, V. Mehrmann, C. Schröder, and F. Tisseur, *NLEVP: a collection of nonlinear eigenvalue problems*, ACM Trans. Math. Softw. **39** (2013), 7:1–7:28.

Computation of the $\mathcal{H}_\infty$-Norm for Large-Scale Systems

Matthias Voigt

(joint work with Peter Benner and Ryan Lowe)

In this short report we consider linear time-invariant descriptor systems

$$E\dot{x}(t) = Ax(t) + Bu(t), \quad y(t) = Cx(t),$$

where E, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$, $x(t) \in \mathbb{R}^n$ is the descriptor vector, $u(t) \in \mathbb{R}^m$ is the input vector, and $y(t) \in \mathbb{R}^p$ is the output vector. Assuming that

the pencil $\lambda E - A$ is regular, the relationship between inputs and outputs in the frequency domain is given by the transfer function

$$G(s) := C(sE - A)^{-1}B.$$

By $\mathcal{RH}_\infty^{p\times m}$ we denote the Banach space of all rational $p \times m$ matrix-valued functions that are analytic and bounded in the open right half-plane $\mathbb{C}^+ := \{s \in \mathbb{C} : \mathrm{Re}(s) > 0\}$. For $G \in \mathcal{RH}_\infty^{p\times m}$, the $\mathcal{H}_\infty$-norm is defined by

$$\|G\|_{\mathcal{H}_\infty} := \sup_{s\in\mathbb{C}^+} \|G(s)\|_2 = \sup_{\omega\in\mathbb{R}} \|G(\mathrm{i}\omega)\|_2 .$$

The aim is to compute this norm under the assumption that all the matrices E, A, B, C are large and sparse and that m, $p \ll n$. We propose two approaches to achieve this.

The first approach is presented in [1] and considers perturbed transfer functions

$$G_\Delta(s) = C(sE - (A + B\Delta C))^{-1}B.$$

We define the structured complex stability radius by

$$r_\mathbb{C}(E, A, B, C) := \inf \left\{ \|\Delta\|_2 : G_\Delta \notin \mathcal{RH}_\infty^{p\times m} \text{ for some } \Delta \in \mathbb{C}^{m\times p} \right\}.$$

It can be shown that

$$r_\mathbb{C}(E, A, B, C) = \begin{cases} 1/\|G\|_{\mathcal{H}_\infty} & \text{if } G \not\equiv 0, \\ \infty & \text{if } G \equiv 0. \end{cases}$$

The condition that $G_\Delta \notin \mathcal{RH}_\infty^{p\times m}$ can be achieved in three ways. First, it can happen that $G_\Delta(\cdot)$ is not well-defined which is the case when the pencil $\lambda E - (A + B\Delta C)$ is singular. Second, $G_\Delta(\cdot)$ might be improper, i.e., unbounded at infinity. These two cases are treated separately. The algorithm concentrates on the third case, namely $G_\Delta(\cdot)$ has poles on the imaginary axis. This computation is based on structured ε-pseudospectra

$$\Pi_\varepsilon(E, A, B, C) = \left\{ s \in \mathbb{C} : s \text{ is a pole of } G_\Delta(\cdot) \text{ for a } \Delta \in \mathbb{C}^{m\times p} \text{ with } \|\Delta\|_2 < \varepsilon \right\}.$$

To compute $r_\mathbb{C}(E, A, B, C)$ we have to find the value of ε for which $\Pi_\varepsilon(E, A, B, C)$ touches the imaginary axis. This is done in a nested iteration, similarly as in [2]. In the inner iteration we compute the rightmost point of $\Pi_\varepsilon(E, A, B, C)$ for a *fixed* value of ε. This is done by computing an appropriate perturbation Δ that moves one of the poles of $G(\cdot)$ to the boundary of $\Pi_\varepsilon(E, A, B, C)$. We exploit the fact that an optimizing perturbation is of rank one, i.e., $\Delta = \varepsilon uv^H$ with $u \in \mathbb{R}^m$, $v \in \mathbb{R}^p$ and $\|u\|_2 = \|v\|_2 = 1$. In the outer iteration, ε is varied by applying Newton's method. To determine the pole that should be perturbed in the inner iteration we compute some dominant poles of $G(\cdot)$ [3]. This is particularly important to find *global* instead of local optimizers.

The second method goes back to [4] and has been generalized to descriptor systems in [5]. There we consider even matrix pencils of the form

$$\mathcal{H}_\gamma(\lambda) := \left[\begin{array}{cc|cc} 0 & -\lambda E^T - A^T & -C^T & 0 \\ \lambda E - A & 0 & 0 & -B \\ \hline -C & 0 & \gamma I_p & 0 \\ 0 & -B^T & 0 & \gamma I_m \end{array}\right].$$

If $\lambda E - A$ has no finite, purely imaginary eigenvalues and $\gamma > \min_{\omega\in\mathbb{R}} \|G(\mathrm{i}\omega)\|_2$, then $\|G\|_{\mathcal{H}_\infty} \geq \gamma$ if and only if $\mathcal{H}_\gamma(\lambda)$ has finite, purely imaginary eigenvalues. This can be used to implement an algorithm that iterates over γ and checks in every step whether $\mathcal{H}_\gamma(\lambda)$ has finite, purely imaginary eigenvalues. These eigenvalues also determine the boundary points of the components of the level-set

$$\Omega_\gamma := \{\omega \in \mathbb{R} : \|G(\mathrm{i}\omega)\|_2 > \gamma\}.$$

As discussed in [5], is important to find *all* finite, purely imaginary eigenvalues to obtain the entire level-set and to ensure global convergence to the $\mathcal{H}_\infty$-norm. However, in the large-scale setting, we cannot compute all eigenvalues of $\mathcal{H}_\gamma(\lambda)$, but we can only use iterative methods to determine *some* eigenvalues close to a number of prespecified shifts. However, heuristically the $\mathcal{H}_\infty$-norm is attained close to a dominant pole. Therefore, we use the dominant poles to determine shifts for the even eigensolver presented in [6]. In this way, we cannot ensure to find the whole level-set Ω_γ, but we can still find one of its components that contains the optimizing frequency ω. The results of this approach and a comparison to the pseudospectral method are discussed in [7]. Numerical examples show that both methods work well, even for rather difficult examples.

References

[1] P. Benner and M. Voigt, *A structured pseudospectral method for $\mathcal{H}_\infty$-norm computation of large-scale descriptor systems*, Math. Control Signals Systems (2013).

[2] N. Guglielmi, M. Gürbüzbalaban, and M. L. Overton, *Fast approximation of the H_∞ norm via optimization over spectral value sets*, SIAM J. Matrix Anal. Appl. **34** (2013), 709–737.

[3] J. Rommes and N. Martins, *Efficient computation of multivariate transfer function dominant poles using subspace acceleration*, IEEE Trans. Power Syst. **21** (2006), 1471–1483.

[4] N. A. Bruinsma and M. Steinbuch, *A fast algorithm to compute the H_∞-norm of a transfer function matrix*, Systems Control Lett. **14** (1990), 287–293.

[5] P. Benner, V. Sima, and M. Voigt, *$\mathcal{L}_\infty$-norm computation for continuous-time descriptor systems using structured matrix pencils*, IEEE Trans. Automat. Control **57** (2012), 233–238.

[6] V. Mehrmann, C. Schröder, and V. Simoncini. *An implicitly-restarted Krylov subspace method for real symmetric/skew-symmetric eigenproblems*, Linear Algebra Appl. **436** (2012), 4070–4087.

[7] R. Lowe and M. Voigt, *$\mathcal{L}_\infty$-norm computation for large-scale descriptor systems using structured iterative eigensolvers*, Preprint MPIMD/13-20, Max Planck Institute Magdeburg (2013).

Inner-outer methods for large-scale two-sided eigenvalue problems

Patrick Kürschner
(joint work with Melina Freitag)

In this work we investigate the numerical solution of large-scale, two-sided, non-Hermitian eigenvalue problems

$$Ax = \lambda Mx \quad \text{and} \quad y^H A = \lambda y^H M$$

with $A,\ M \in \mathbb{C}^{n\times n}$, eigenvalues $\lambda \in \mathbb{C}$, and right, left eigenvectors $x,\ y \in \mathbb{C}^n \backslash \{0\}$. We focus on basic iterative methods such as two-sided inverse and Rayleigh quotient iteration [5] for computing a single eigentriple $(\lambda,\ x,\ y)$. On the one hand, the left eigenvectors y are useful in important applications, e.g., modal truncation of large dynamical systems. On the other hand, their incorporation can also be beneficial for the performance of eigenvalue methods for non-normal problems.
Two-sided inverse and Rayleigh quotient iteration can be summarized as follows:

1. Choose a shift μ_k.
2. Solve the linear systems

$$(A - \mu_k M)\hat{u} = Mu_k \quad \text{and} \quad (A - \mu_k M)^H \hat{v} = M^H v_k \tag{1}$$

   for $\hat{u}$, $\hat{v}$.
3. Normalize the vectors via $u_{k+1} = \hat{u}/\|\hat{u}\|$, $v_{k+1} = \hat{v}/\|\hat{v}\|$.
4. Compute a new eigenvalue approximation by the two-sided Rayleigh quotient $\theta_{k+1} = \dfrac{v_{k+1}^H A u_{k+1}}{v_{k+1}^H M u_{k+1}}$.
5. Test for convergence, e.g., using the eigenvalue residuals $r_{u_{k+1}} = (A - \theta_{k+1}M)u_{k+1}$ and $r_{v_{k+1}} = (A - \theta_{k+1}M)^H v_{k+1}$. Repeat steps 1-5 if the approximations $\theta_{k+1},\ u_{k+1},\ v_{k+1}$ are not accurate enough.

The above iteration is initialized by vectors $u_0 \approx x,\ v_0 \approx y$. Two-sided inverse iteration (TII) uses a fixed shift $\mu_k = \mu \approx \lambda$ in step 1 which has to be specified beforehand, whereas two-sided Rayleigh quotient iteration (TRQI) uses $\mu_k = \theta_k$. If the iteration converges, i.e., $(\theta_k,\ u_k,\ v_k) \to (\lambda,\ x,\ y)$, the local rate of convergence of TII and TRQI is linear and cubic, respectively, if the linear systems in (1) are solved exactly [1, 5, 4]. Since solving these linear systems is the most expensive part of the above iteration, this is often done inexactly, e.g., by employing Krylov subspace methods for linear systems. The iterations of this linear solver are commonly referred to as *inner iterations* in contrast to the *outer iterations* of the method for the eigenvalue computation. It can be shown that the convergence rates of the exact methods can be reestablished under certain conditions [1]. In general, it can be said that with inexact solves we obtain the same convergence rate as for exact solves if the solve tolerance is chosen proportional to the eigenvalue residuals [3, 2].

In this report we focus on the usage of preconditioners for the Krylov subspace methods used within the inner iterations which is typically necessary for the large-scale case. For the application in the considered specific eigenvalue iteration,

specially tailored preconditioning strategies [3] can be employed which further improve the performance of the preconditioned Krylov subspace solver. In [3] it is shown that, when a preconditioner $\mathbb{P}$ satisfies, e.g., $\mathbb{P}u_k = Mu_k$ for the first linear system in (1), the number of necessary inner iterations to obtain an approximation to $\hat{u}$ of a certain precision is notably reduced. As result one obtains the so called *tuned preconditioners* which are constructed as rank-one updates of standard preconditioners $P \approx A - \mu_k M$:

$$\mathbb{P} = P + (Mu_k - Pu_k)u_k^H.$$

For the adjoint linear system in (1) one can choose

$$\mathbb{Q} = P^H + (M^H v_k - P^H v_k)v_k^H,$$

which satisfies $\mathbb{Q}v_k = M^H v_k$. Another possibility is to replace M by A which leads to tuned preconditioners satisfying $\mathbb{P}u_k = Au_k$ and $\mathbb{Q}v_k = A^H v_k$. The application of $\mathbb{P}$ and $\mathbb{Q}$ introduces only minor additional computations [3].
The coefficient matrices (1) are adjoint to each other such that both linear systems can be dealt with simultaneously by certain Krylov subspace methods, e.g., BiCG and QMR which are short-recurrence methods based on the two-sided Lanczos process. For the use in such two-sided Krylov subspace methods, the tuned preconditioners have to be adapted accordingly since a tuned preconditioner $\mathbb{S}$ has to satisfy both

$$\mathbb{S}u_k = Mu_k \quad \text{and} \quad \mathbb{S}^H v_k = M^H v_k$$

at the same time. It turns out that for these conditions to hold it is not sufficient to modify a standard preconditioner P by a rank-one update. In [1] a rank-two update of P is proposed as

$$\mathbb{S} = P + [Mu_k, Pu_k] \begin{bmatrix} v_k^H Pu_k + 1 & -1 \\ -1 & 0 \end{bmatrix} [M^H v_k, P^H v_k]^H$$

with the normalization $v_k^H Mu_k = 1$. Note that M can also be replaced by A. It is worth mentioning that the M-variant of $\mathbb{S}$ allows to draw novel connections [1] between TRQI and a simplified version of the two-sided Jacobi-Davidson algorithm [4]. The costs for applying $\mathbb{S}$ are about the same as for applying both $\mathbb{P}$ and $\mathbb{Q}$ together.

As short numerical experiment we consider the matrices given by the `anemo` system of the Oberwolfach model reduction benchmark collection. The dimension is $n = 29008$ and we look for $\lambda = -305.35$ using TII with initial guess $\mu = -300$. The outer iteration is terminated when the norm of the eigenvalue residuals r_{u_k} and r_{v_k} are smaller than 10^{-10}. As inner solver we apply preconditioned BiCG which is terminated when the residuals of the linear systems are smaller than $0.1 \cdot \min\left(1, \|r_{u_k/v_k}\|\right)$. The results in Table 2 show that using $\mathbb{S}$ reduces the number of inner iterations in BiCG compared to the usage of the standard preconditioners. Further experiments can be found in [1], where in several cases using BiCG with $\mathbb{S}$ even outperforms the separate solution of (1) with GMRES using $\mathbb{P}$ and $\mathbb{Q}$.

TABLE 2. Results of inexact TII using standard and tuned preconditioners for the `anemo` example.

Preconditioner	outer	inner	avg.	# precs	time
standard	6	1530	306	1530	15.8
tuned, $\mathbb{S}x = Mx$	6	1116	223	1126	9.9
tuned, $\mathbb{S}x = Ax$	7	1120	187	1132	10.4

One further, currently investigated research perspective is the generalization of the concept of tuned preconditioners to inexact, one- and two-sided Newton-type eigenvalue algorithms for nonlinear eigenvalue problems, e.g., nonlinear inverse iteration and Rayleigh functional iteration [6, 7]. Preliminary experiments confirm that tuned preconditioners can also be applied there and lead again to a reduction of inner iterations.

REFERENCES

[1] M. FREITAG AND P. KÜRSCHNER, *Tuned preconditioners for inexact two-sided inverse and Rayleigh quotient iteration*, MPI Magdeburg Preprint MPIMD Preprint/13-04, 2013. Available at `http://www.mpi-magdeburg.mpg.de/preprints/2013/04/`.

[2] M. FREITAG AND A. SPENCE, *Convergence theory for inexact inverse iteration applied to the generalised nonsymmetric eigenproblem*, Electron. Trans. Numer. Anal., 28 (2007), pp. 40–64.

[3] M. FREITAG, A. SPENCE, AND E. VAINIKKO, *Rayleigh quotient iteration and simplified Jacobi-Davidson with preconditioned iterative solves for generalised eigenvalue problems*, technical report, Dept. of Mathematical Sciences, University of Bath, 2008.

[4] M. E. HOCHSTENBACH AND G. L. G. SLEIJPEN, *Two-sided and alternating Jacobi–Davidson*, Linear Algebra and its Applications, 358(1-3) (2003), pp. 145–172.

[5] B. N. PARLETT, *The Rayleigh quotient iteration and some generalizations for nonnormal matrices*, Mathematics of Computation, 28 (1974), pp. 679–693.

[6] K. SCHREIBER, *Nonlinear Eigenvalue Problems: Newton-type Methods and Nonlinear Rayleigh Func- tionals*, Ph.D thesis, Department of Mathematics, TU Berlin, 2008.

[7] D. SZYLD, F. XUE, *Local convergence analysis of several inexact Newton-type algorithms for general nonlinear eigenvalue problems*, Numerische Mathematik, 123 (2013), pp. 333–362.

Computer-assisted existence and multiplicity proofs for semilinear elliptic boundary value problems via numerical eigenvalue bounds

MICHAEL PLUM

Many boundary value problems for semilinear elliptic partial differential equations allow very stable numerical computations of approximate solutions, but are still lacking analytical existence proofs. We propose a method which exploits the knowledge of a "good" numerical solution, in order to provide a rigorous proof of existence of an exact solution close to the approximate one. This goal is achieved by a fixed-point argument (similar to the Newton-Cantorovich Theorem) which takes all numerical errors into account, and thus gives a mathematical proof which is not "worse" than any purely analytical one. The main tool in the proof are bounds

for eigenvalues of the elliptic operator generated by linearization of the nonlinear problem at the computed approximate solution. These eigenvalue bounds are again obtained by computer-assisted means, by first computing "good" approximate eigenpairs and then using variational arguments to get error bounds (and thus safe enclosures) for the crucial eigenvalues.

The method is used to prove existence and multiplicity statements for some examples, including cases where purely analytical methods had not been successful.

Accurate Computations of Eigenvalues of Differential Operators

Qiang Ye

We are concerned with accurate computations of a few eigenvalues/eigenvectors of a large symmetric positive definite matrix arising in discretization of differential operators. When a fine discretization is used, the matrix may be extremely ill-conditioned. In that case, the accuracy of the smaller eigenvalues that are computed for the matrix may be low.

Consider a general symmetric positive definite matrix A and let $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be its eigenvalues. Conventional dense matrix eigenvalue algorithms (such as the QR algorithm) are normwise backward stable, i.e., the computed eigenvalues $\widehat{\lambda}_i$ are the exact eigenvalues of $A+E$ with $\|E\|_2 = \mathcal{O}(\mathbf{u})\|A\|_2$, where $\mathbf{u}$ is the machine roundoff unit. Eigenvalues of large (sparse) matrices are typically computed by an iterative method (such as the Lanczos algorithm), which produces an approximate eigenvalue $\widehat{\lambda}_i$ and an approximate eigenvector $\widehat{x}_i$ whose residual $\|A\widehat{x}_i - \widehat{\lambda}_i\widehat{x}_i\|_2$ converges in a floating point arithmetic is at best (at convergence) $\mathcal{O}(\mathbf{u})\|A\|_2\|\widehat{x}_i\|_2$. In both cases, we have $|\widehat{\lambda}_i - \lambda_i| \leq \mathcal{O}(\mathbf{u})\|A\|_2$. This is sufficient to guarantee good relative accuracy for larger eigenvalues (i.e. for $\lambda_i \approx \lambda_n$), but for smaller eigenvalue (i.e. for $\lambda_i \approx \lambda_1$), we have

$$\frac{|\widehat{\lambda}_i - \lambda_i|}{\lambda_i} \leq \frac{\|E\|_2}{\lambda_i} \leq \mathcal{O}(\mathbf{u})\frac{\lambda_n}{\lambda_i} \approx \mathcal{O}(\mathbf{u})\kappa_2(A) \tag{1}$$

where $\kappa_2(A) = \|A\|_2\|A^{-1}\|_2$. is the condition number in 2-norm. Therefore, the best relative accuracy of the smaller eigenvalues that one can compute depends on the condition number of A.

Consider discretizations of differential operators. The finite difference or the finite element methods lead to a large and sparse matrix eigenvalue problem. Here it is usually the smaller eigenvalues that are of interest and are well approximated by the discretization. However, as the discretization mesh size h decreases, the condition number increases and then the relative accuracy of smaller eigenvalues as computed by existing algorithms deteriorates. For a majority of problems, this deterioration in relative accuracy may be mild and harmless as it may be within the discretization error. Specifically, the condition numbers are typically of order $\mathcal{O}(h^{-2})$ for second order differential operators and of order $\mathcal{O}(h^{-4})$ for fourth order differential operators. However, for some important problems such as fourth order differential operators which arise in vibrational analysis of elastic

plates or beams, even for modestly small mesh size, the condition number of its discretization increases very rapidly in the order $\mathcal{O}(h^{-4})$. Hence, even for a modestly small value of h, smaller eigenvalues computed might have very low relative accuracy resulting an inaccurate final approximate eigenvalue.

In this work, we present algorithms that accurately compute the smaller eigenvalues of a differential operator. For second order differential operators, the finite difference discretization typically leads to a diagonally dominant matrix. We then use an accurate LDU factorization algorithm recently developed for diagonally dominant matrices to invert the matrix accurately in the shift-and-invert transformation, for which the larger eigenvalues and hence the corresponding smaller eigenvalues of the original problem are computed accurately. For fourth order differential operators, we consider some special finite difference discretizations that can be expressed as a product of two diagonally dominant matrices. This has been developed for 1 dimensional biharmonic operator with the clamped boundary condition. Then each diagonally dominant factor of the discretization can be factorized accurately and the discretization matrix can be accurately inverted. Again, in this way, the smaller eigenvalues are accurately computed. Numerical examples are presented to illustrate the effectiveness of the new algorithms.

Low-rank tensor methods with subspace correction for symmetric eigenvalue problems

André Uschmajew

(joint work with Daniel Kressner, Michael Steinlechner)

Low-rank tensor methods provide a possible way to solve, under some assumptions, even extremely high-dimensional PDEs or PDE eigenvalue problems, as they can occur, e.g., in electronic structure calculations, quantum spin systems, or stochastic PDEs, without having to face the curse of dimensionality. A recent overview over the expanding field of low-rank tensor methods is given in [3].

Consider the model problem

$$\begin{aligned} -\Delta u(\mathbf{x}) + V(\mathbf{x})u(\mathbf{x}) &= \lambda u(\mathbf{x}) \quad \text{for } \mathbf{x} \in \Omega = (a,b)^d, \\ u(\mathbf{x}) &= 0 \qquad\quad \text{for } \mathbf{x} \in \partial\Omega, \end{aligned} \tag{1}$$

on an appropriate function space with a given potential V. We seek for the p smallest eigenvalues. After discretization on a tensor product grid using n_μ grid points for variable x_μ we get a huge trace minimization problem,

$$\min\{\operatorname{trace}(\mathbf{U}^\mathsf{T}\mathbf{A}\mathbf{U})\colon \mathbf{U}^\mathsf{T}\mathbf{U} = I_p\}, \tag{2}$$

with the discretized operator $\mathbf{A}$ being of size $n_1 n_2 \cdots n_d \times n_1 n_2 \cdots n_d$. It is clear that this problem cannot be attacked using standard linear algebra algorithms when d is very large, say $d = 20$. In fact, one is then faced with the fundamental problem of storing and accessing the discrete eigenvectors $\mathbf{u}_\alpha$, $\alpha = 1, 2, \ldots, p$, not even speaking of matrix-vector operations.

The situation changes if we regard the discrete eigenvectors $\mathbf{u}_\mu$ as d-dimensional tensors of size $n_1 \times n_2 \times \cdots \times n_d$, and assume that they have a *low tensor rank*, or are at least very well approximable by tensors of low rank. As there are several notions of tensor rank for tensors of order higher than two (matrices), the precise meaning of this statement depends on the tensor decomposition (or *format*) one aims at. We assume the sought p eigenvectors being well approximable by tensors in the *block TT format* [1] with low rank. In this tensor format, which is related to Wilson's numerical renormalization group for one-dimensional quantum many-body systems [5], the eigenvectors $\mathbf{u}_\alpha$ (the columns of $\mathbf{U}$) are represented pointwise as matrix products

$$(\mathbf{u}_\alpha)_{i_1,i_2,\ldots,i_d} = U_1(i_1)\cdots U_{\mu-1}(i_{\mu-1})U_{\mu,\alpha}(i_\mu)U_{\mu+1}(i_{\mu+1})\cdots U_d(i_d), \quad 1 \le \alpha \le p,$$

with matrices $U_\nu(i_\nu) \in \mathbb{R}^{r_{\nu-1}\times r_\nu}$, $\nu \neq \mu$, and $U_{\mu,\alpha}(i_\mu) \in \mathbb{R}^{r_{\mu-1}\times r_\mu}$, $\alpha = 1, 2, \ldots, p$ (called *TT cores*). One fixes $r_0 = r_d = 1$ so that the product indeed results in a scalar for every multi-index $(i_1, i_2, \ldots, i_d)$. The other ranks r_ν one would like to keep as small as possible.

The position μ of the index α enumerating the eigenvectors is arbitrary, and can be changed. It can be moved to the right or left using singular value decomposition. Such a shift will change the involved TT cores, as well as the involved rank (e.g., r_μ when shifting from μ to $\mu + 1$). Having this tool at hand one can optimize the TT cores by *sweeping*: one fixes all cores except at position μ, minimizes (2) as a function of the $U_{\mu,\alpha}(i_\mu)$, shifts the index α to a neighboring core, and repeats the process [1, 5]. The minimization with respect to the $U_{\mu,\alpha}(i_\mu)$ is a projection of the huge trace minimization of (2) on an $(r_{\mu-1}n_\mu r_\mu)$-dimensional subspace only, and hence can be in principal addressed using a standard solver like LOPBCG as long as the ranks are kept moderate.

When $p \ge 2$, a surprising feature of the outlined alternating block optimization algorithm is *rank adaptivity*. Consequently, the final target ranks do not have to be fixed in advance, which would be a nontrivial task. However, the procedure is not rank adaptive when $p = 1$, i.e. when only an eigenvector for the minimal eigenvalue is sought, since there is no index to shift. In a recent work [2] on solving high-dimensional linear equations in TT format, Dolgov and Savostyanov presented the alternating minimal energy (AMEn) idea, which can be of use in this situation. It aims at accelerating the alternating core optimization by steepest descent steps. In such a step, a low-rank approximation of the *DMRG residual* (the residual with respect to *two* neighboring cores) is added to the current approximation. This addition increases the used ranks, and thus may also serve for rank adaption.

In our work [4] we transfer the AMEn idea to eigenvalue problems of the form (2). In this context it can be interpreted as a local subspace correction procedure which mimics the two-site DMRG algorithm [6] at least to first order. Additionally, the benefit and implementation of using preconditioned residuals is discussed. We test the algorithms for the Henon-Heiles potential and for the Newton potential.

For the named potentials, astonishingly high-dimensional eigenvalue problems can be solved by the outlined techniques. The theoretical proof of the approximability of eigenfunctions by low-rank tensors is usually difficult and an important research topic. Also the convergence properties of these successful algorithms are poorly understood so far, and deserve a closer look in the future.

References

[1] S.V. Dolgov, B.N. Khoromskij, I.V. Oseledets, D.V. Savostyanov, *Computation of extreme eigenvalues in higher dimensions using block tensor train format*, arXiv:1306.2269v1 (2013).

[2] S.V. Dolgov, D.V. Savostyanov, *Alternating minimal energy methods for linear systems in higher dimensions. Part I: SPD systems*, arXiv:1301.6068v1 (2013).

[3] L. Grasedyck, D. Kressner, C. Tobler, *A literature survey of low-rank tensor approximation techniques*, GAMM-Mitt. **36** (2013), 53–78.

[4] D. Kressner, M. Steinlechner, A. Uschmajew, *Low-rank tensor methods with subspace correction for symmetric eigenvalue problems*, MATHICSE Report **40.2013**, EPF Lausanne, (2013)

[5] I. Pižorn, F. Verstraete, *Variational numerical renormalization group: Bridging the gap between nrg and density matrix renormalization group*, Phys. Rev. Lett. **108** (2012), 067202.

[6] S.R. White, *Density matrix formulation for quantum renormalization groups*, Phys. Rev. Lett. **69** (1992), 2863–2866.

Hierarchical Multilevel Substructuring for PDE Eigenvalue Problems

Lars Grasedyck

(joint work with Peter Gerds)

In our paper we introduce a new method, called H-AMLS, which computes eigenpair approximations for an elliptic eigenvalue problem. The new method combines the automated multi-level substructuring (AMLS) method [2, 3, 4] with the concept of hierarchical matrices (H-matrices) [5, 6], and it allows us to compute a large amount of eigenvalue approximations at once in almost linear complexity.

AMLS is a substructuring method projecting the discretized eigenvalue problem onto a small subspace. A reduced eigenvalue problem is computed which delivers approximate solutions of the original problem. Whereas the AMLS method is very effective in the two-dimensional case, it is getting very expensive in the three-dimensional case, due to the fact that it computes the reduced eigenvalue problem via dense matrix operations. In the new method these dense matrices are approximated by data-sparse H-matrices and the corresponding matrix operations can be performed in almost linear complexity. Under certain assumptions based on the approximability of eigenfunctions in a finite element space of size N [1, 7] we provide computational bounds of the order $O(N \log N)$ for the new method. Beside the discretisation error two additional errors occur; the projection error of the AMLS method, and the error caused by the use of the H-matrix approximation which are controlled by several parameters. We investigate the influence of these parameters for the Laplace eigenvalue problem defined on a three-dimensional domain.

Whereas the computational cost of a classical shift-invert subspace solver depends at least linearly on the number M of sought eigenpairs, the H-AMLS method computes M eigenpairs at once in almost linear complexity $O((M+N)\log N)$. Furthermore, we note that the benchmarked large-scale three-dimensional problems could be easily solved with H-AMLS whereas the solution of these problems with classical AMLS is very expensive because of the $O(N^2)$ scaling of AMLS in the three-dimensional case.

References

[1] L. Banjai, S. Börm, and S. Sauter, *FEM for elliptic eigenvalue problems: how coarse can the coarsest mesh be chosen? An experimental study*, Comput. Vis. Sci. **11** (2008), 363–372.

[2] J.K. Bennighof and R.B. Lehoucq, *An automated multilevel substructuring method for eigenspace computation in linear elastodynamics*, SIAM J. Sci. Comput. **25** (2004), 2084–2106.

[3] R.R. Craig and M.C.C. Bampton, *Coupling of substructures for dynamic analysis*, AIAA Journal **6** (1968), 1313–1319.

[4] W. Gao, X.S. Li, C. Yang, and Z. Bai, *An implementation and evaluation of the AMLS method for sparse eigenvalue problems*, ACM Trans. Math. Software **34** (2008), 20, 28.

[5] L. Grasedyck, R. Kriemann, and S. LeBorne, *Parallel black box H-LU preconditioning for elliptic boundary value problems*, Comput. Vis. Sci. **11** (2008), 273–291.

[6] W. Hackbusch, *A sparse matrix arithmetic based on H-matrices. Part I: Introduction to H-matrices*, Computing **62** (1999), 89–108.

[7] S. Sauter, *hp-finite elements for elliptic eigenvalue problems: error estimates which are explicit with respect to lambda, h, and p*, SIAM J. Num. Anal. **48** (2010), 95–108.

Reporter: Ute Kandler

Participants

Prof. Dr. Zhaojun Bai
Department of Computer Science
University of California
One Shields Avenue
Davis, CA 95616
GERMANY

Prof. Dr. Randolph E. Bank
Department of Mathematics
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0112
UNITED STATES

Prof. Dr. Christopher Beattie
Department of Mathematics
Virginia Polytechnic Institute and
State University
Blacksburg, VA 24061-0123
UNITED STATES

Prof. Dr. Peter Benner
Max-Planck-Institut für
Dynamik komplexer techn. Systeme
Sandtorstr. 1
39106 Magdeburg
GERMANY

Prof. Dr. Michele Benzi
Department of Mathematics and
Computer Science
Emory University
400, Dowman Dr.
Atlanta, GA 30322
UNITED STATES

Prof. Dr. Daniele Boffi
Dipartimento di Matematica
Universita di Pavia
Via Ferrata, 1
27100 Pavia
ITALY

Prof. Dr. James Brannick
Department of Mathematics
Pennsylvania State University
University Park, PA 16802
UNITED STATES

Prof. Dr. Susanne C. Brenner
Department of Mathematics
Louisiana State University
Baton Rouge LA 70803-4918
UNITED STATES

Prof. Dr. Long Chen
Department of Mathematics
University of California, Irvine
Irvine, CA 92697-3875
UNITED STATES

Prof. Dr. Howard Elman
Department of Computer Science
University of Maryland
College Park, MD 20742
UNITED STATES

Prof. Dr. Mark Embree
Department of Mathematics
Rice University
6100 Main Street
Houston, TX 77005-1892
UNITED STATES

Dr. Jean-Luc Fattebert
Center for Applied Scientific Computing
Lawrence Livermore National
Laboratory
P.O.Box 808, L-561
Livermore CA 94550
UNITED STATES

Dietmar Gallistl
Fachbereich Mathematik
Humboldt Universität Berlin
Unter den Linden 6
10099 Berlin
GERMANY

Dr. Joscha Gedicke
Department of Mathematics
Louisiana State University
Baton Rouge LA 70803-4918
UNITED STATES

Prof. Dr. Ivan G. Graham
Dept. of Mathematical Sciences
University of Bath
Claverton Down
Bath BA2 7AY
UNITED KINGDOM

Prof. Dr. Lars Grasedyck
Institut für Geometrie und
Praktische Mathematik
RWTH Aachen
Templergraben 55
52056 Aachen
GERMANY

Dr. Luka Grubisic
Department of Mathematics
University of Zagreb
Bijenicka 30
10000 Zagreb
CROATIA

Prof. Dr. Dr.h.c. Wolfgang Hackbusch
Max-Planck-Institut für Mathematik
in den Naturwissenschaften
Inselstr. 22 - 26
04103 Leipzig
GERMANY

Prof. Dr. Ralf Hiptmair
Seminar für Angewandte Mathematik
ETH-Zentrum
Rämistr. 101
8092 Zürich
SWITZERLAND

Dr. Michiel Hochstenbach
Dept. of Mathematics & Computer
Science
Eindhoven University of Technology
5600 MB Eindhoven
NETHERLANDS

Dr. Jun Hu
School of Mathematical Sciences
Peking University
100 871 Beijing
CHINA

Prof. Dr. Ilse C.F. Ipsen
Department of Mathematics
North Carolina State University
Campus Box 8205
Raleigh, NC 27695-8205
UNITED STATES

Ute Kandler
Institut für Mathematik
Technische Universität Berlin
Sekr. MA 4-5
Strasse des 17. Juni 136
10623 Berlin
GERMANY

Prof. Dr. Andrew Knyazev
Mitsubishi Electric Research
Laboratories
201 Broadway
Cambridge MA 02139
UNITED STATES

Prof. Dr. Daniel Kressner
Section de Mathématiques
Station 8
École Polytechnique Fédérale de Lausanne
1015 Lausanne
SWITZERLAND

Patrick Kürschner
Max-Planck-Institut für Dynamik komplexer techn. Systeme
Sandtorstr. 1
39106 Magdeburg
GERMANY

Prof. Dr. Mats G. Larson
Department of Mathematics
University of Umea
901 87 Umea
SWEDEN

Dr. Lin Lin
Computational Research Division
Lawrence Berkeley National Laboratory
Mail Stop 50F-1650
1 Cyclotron Road
Berkeley, CA 94720
UNITED STATES

Dr. Karl Meerbergen
Departement Computerwetenschappen
Katholieke Universiteit Leuven
Celestijnenlaan 200A
3001 Heverlee
BELGIUM

Prof. Dr. Volker Mehrmann
Institut für Mathematik
Technische Universität Berlin
Sekr. MA 4-5
Strasse des 17. Juni 136
10623 Berlin
GERMANY

Agnieszka Miedlar
Institut für Mathematik
Sekr. MA 4-5
Technische Universität Berlin
Straße des 17. Juni 136
10623 Berlin
GERMANY

Christian Mollet
Mathematisches Institut
Universität zu Köln
Weyertal 86-90
50931 Köln
GERMANY

Prof. Dr. Yvan Notay
Université Libre de Bruxelles
Service de Metrologie Nucleaire
(CP 165-84)
50 av. F.D.Roosevelt
1050 Bruxelles
BELGIUM

Prof. Dr. Joseph E. Pasciak
Department of Mathematics
Texas A & M University
College Station, TX 77843-3368
UNITED STATES

Prof. Dr. Michael Plum
Institut für Analysis
Karlsruher Institut für Technologie
Kaiserstr. 89
76128 Karlsruhe
GERMANY

Prof. Dr. José E. Roman
Universitat Politècnica de València
D. Sistemes Informàtics i Computació
Camí de Vera, s/n
46022 Valencia
SPAIN

Mira Schedensack
Fachbereich Mathematik
Humboldt Universität Berlin
Unter den Linden 6
10099 Berlin
GERMANY

Prof. Dr. Joachim Schöberl
Institut für Analysis und
Scientific Computing
Technische Universität Wien
Wiedner Hauptstr. 8 - 10
1040 Wien
AUSTRIA

Dr. Christian Schröder
Institut für Mathematik
Technische Universität Berlin
Sekr. MA 4-5
Strasse des 17. Juni 136
10623 Berlin
GERMANY

Prof. Dr. Valeria Simoncini
Dipartimento di Matematica
Università degli Studi di Bologna
Piazza di Porta S. Donato, 5
40126 Bologna
ITALY

Andre Uschmajew
EPFL SB MATHICSE ANCHP
MA B2 525 (Batiment MA)
Station 8
1015 Lausanne
SWITZERLAND

Matthias Voigt
Max-Planck-Institut für
Dynamik komplexer techn. Systeme
Sandtorstr. 1
39106 Magdeburg
GERMANY

Prof. Dr. Jinchao Xu
Department of Mathematics
Pennsylvania State University
University Park, PA 16802
UNITED STATES

Prof. Dr. Qiang Ye
Department of Mathematics
University of Kentucky
715 Patterson Office Tower
Lexington, KY 40506-0027
UNITED STATES

Prof. Dr. Harry Yserentant
Institut für Mathematik
Technische Universität Berlin
Straße des 17. Juni 136
10623 Berlin
GERMANY

Prof. Dr. Zhimin Zhang
Beijing Computational Science Research
Center
No. 3, Heqing Road
Beijing 100084
CHINA

Prof. Dr. Aihui Zhou
Institute of Computational Mathematics
and Scientific/Engineering Computing
Academy of Mathematics and Systems
Sc.
Chinese Academy of Sciences
Beijing 100190
CHINA

Prof. Dr. Ludmil Zikatanov
Department of Mathematics
Pennsylvania State University
University Park, PA 16802
UNITED STATES

Dr. Ian Zwaan
Dept. of Mathematics & Computer Science
Eindhoven University of Technology
5600 MB Eindhoven
NETHERLANDS