An a posteriori error estimator based on shifts for positive hermitian eigenvalue problems

Athmane Bakhta, Damiano Lombardi

An *a posteriori* error estimator based on shifts for positive hermitian eigenvalue problems

Athmane Bakhta* Damiano Lombardi †

Abstract

This work deals with an a posteriori error estimator for hermitian positive eigenvalue problems. The proposed estimator is based on the residual and the definition of suitable shifts in the matrix spectrum. The mathematical properties (certification and sharpness) are investigated and some numerical experiments are proposed.

1 Introduction.

This work deals with an *a posteriori* error estimator for hermitian positive eigenvalue problems. Having a sharp error estimator for eigenvalues is, in general, a difficult task. Several works were proposed in the literature: in [2, 3] error estimators are proposed for the finite element discretization of eigenvalues of elliptic operators. In the reduced basis context for Stokes equations, an a posteriori error estimator based on Babuška stability theory [1] is proposed in [8]. A general formulation of a posteriori error bounds that can be applied to several situations is given in [5]. We also refer the reader to thesis manuscript [7] where error analysis is carried on for numerous problems: coercive, noncoercive, parabolic, Stokes and eigenvalue. In this work, that was originally motivated by problems involving classical periodic Schrödinger operators, an error estimator for the spectrum of positive self-adjoint operators is proposed, based on the problem residual and the definition of shifts. In particular, the aim is to estimate the error done when the problem finely discretized is projected on a low dimensional basis, giving rise to a coarsely discretized problem.

The document is structured as follows: after having introduced the notation, the expression of the estimator is derived in section 2.1. The main results on the analysis of the estimator are presented in section 3. The implementation and some numerical experiments are presented in section 4.

2 Notation and problem setting.

Let U be a Hilbert space and $\mathcal{A}$ be a linear positive self-adjoint operator $\mathcal{A} \in \mathcal{L}(U, U)$. Consider the associated eigenvalue problem:

$$\mathcal{A} u_{\text{exact}} = \lambda_{\text{exact}} u_{\text{exact}}, \tag{2.1}$$

where $\lambda_{exact} \in \mathbb{R}_+^*$ is the real positive eigenvalue, and $u_{exact} \in U$ is the associated eigenfunction.

A discrete fine approximation of Problem (2.1) is introduced as follows: let $\mathcal{N} \in \mathbb{N}^*$, and $A \in \mathbb{C}^{\mathcal{N} \times \mathcal{N}}$ be a matrix, obtained by discretizing Eq.(2.1). The finely discretized problem reads:

$$A u = \lambda u, \tag{2.2}$$

where, $u \in \mathbb{C}^{\mathcal{N}}$ denotes the finely discretized eigenvector and λ the corresponding eigenvalue.

A coarse approximation is solved by projecting the finely discretized problem on a low dimensional basis. More precisely, let $N \in \mathbb{N}^*$ such that $N \ll \mathcal{N}$ and denote by $W \in \mathbb{C}^{\mathcal{N} \times N}$, $W = [w_1, \ldots, w_N]$ the matrix whose columns are the finely discretized basis functions. Let A_N denote

*Ecole des Ponts ParisTech, athmane.bakhta@cermics.enpc.fr
†INRIA, damiano.lombardi@inria.fr

the coarse matrix defined as $A_N := W^H A W$. Thus, the coarse approximation of the eigenpairs is obtained by solving:

$$A_N \phi = \lambda_N \phi, \tag{2.3}$$

that provides $\lambda_N > \lambda$, an approximation from above of λ and where $u_N \in \mathbb{C}^{\mathcal{N}}$, $u_N := W\phi$, an approximation of the associated eigenvector reconstructed at the fine level of discretization. Throughout the whole document, the scalar product in $\mathbb{C}^{\mathcal{N}}$ will be denoted by: $\langle u, v \rangle = u^H v$, for $u, v \in \mathbb{C}^{\mathcal{N}}$.

In the sequel, we forget about the exact eigenvalues λ_{exact} and our aim is to estimate the error between the finely approximated value λ and the coarsely approximated value λ_N.

2.1 Derivation of the estimator.

As for most of the methods of *a posteriori* estimation, a relationship between the residual of the problem and the error on the solution is sought. The residual reads: $r_N \in \mathbb{C}^{\mathcal{N}}$, $r_N := Au_N - \lambda_N u_N$. Since the coarse problem is obtained by Galerkin projection, it follows that $\langle r_N, u_N \rangle = 0$. Let $e := u_N - u$ denote the error in the eigenvector approximation. The following equation for the residual is obtained:

$$r_N = (A - \lambda)e - (\lambda_N - \lambda)u_N. \tag{2.4}$$

Projecting Eq.(2.4) on u_N leads to:

$$0 = \langle u_N, (A - \lambda)u_N - u \rangle - (\lambda_N - \lambda) \;\Rightarrow\; \varepsilon := \lambda_N - \lambda = \langle u_N, (A - \lambda)u_N \rangle \geq 0. \tag{2.5}$$

The objective is to express this quantity as $\langle r_N, B r_N \rangle$ where $B \in \mathbb{C}^{\mathcal{N} \times \mathcal{N}}$ is a hermitian matrix (which is the discretization, at fine level, of a self-adjoint operator that will be made precise later). This is done in two steps:

First, an expression for the matrix B is derived. To do so, let us assume that $\lambda_N \neq \lambda^{(i)}, \forall i,\ 1 \leq i \leq \mathcal{N}$, so that the matrix $(A - \lambda_N)$ is invertible. Here $\lambda^{(i)}$ denotes the i^{th} eigenvalue of A. Thus, reintroducing the residual r_N into Eq.(2.5) gives:

$$\lambda_N - \lambda = \langle u_N, (A - \lambda)u_N \rangle = \langle r_N, (A - \lambda_N)^{-1}(A - \lambda)(A - \lambda_N)^{-1} r_N \rangle. \tag{2.6}$$

By direct identification, the matrix B reads:

$$B = (A - \lambda_N)^{-1}(A - \lambda)(A - \lambda_N)^{-1}. \tag{2.7}$$

Note that, in this form, the matrix B cannot be directly used, since the eigenvalue λ is involved in its definition. Second, an approximation of the operator B is derived in order to obtain a computable and certified approximation of the error. Let $a, b \in \mathbb{R}$ be two scalars (whose optimal values will be precised later). Then, the matrix B is approximated as follows

$$\tilde{B} := (A - b)^{-1}(A - a)(A - b)^{-1}, \tag{2.8}$$

so that a, b are two *shifts* for the spectrum of $A - \lambda$ and $A - \lambda_N$ respectively.

2.2 A priori estimation for eigenvalues.

The proposed *a posteriori* error estimator is defined by exploiting an available *a priori* error estimator. Several of such estimators exist in the literature, and can be adapted to the problem of interest. For the present work, an *a priori* estimator based on traces is used, presented in [9]. Since in the present work, positive matrices are considered, the *a priori* lower value for λ is defined as:

$$\tilde{\lambda} = \max\{0, \tilde{\lambda}_{WS}\}, \tag{2.9}$$

where $\tilde{\lambda}_{WS}$ denotes the result of the estimation proposed in [9], which is not always guaranteed to be positive.

3 Analysis of the estimator: main results.

An error estimator is said to be *certified* if the estimated error is always larger than the actual error, and it is said to be *sharp* if the estimated error is as close as possible (in some sense depending on the problem) to the actual error. In this section, we give a precise definition of the proposed a posteriori error estimator and investigate under which conditions it is certified and sharp.

Definition 3.1. *The actual error $\varepsilon := \lambda - \lambda_N$ is estimated by the quantity*

$$\mu(a,b) := \langle r_N, \widetilde{B} r_N \rangle$$

where $\widetilde{B}$ is defined in (2.8) and the scalars $a, b \in \mathbb{R}$ depend on λ_N and on a priori lower and upper bounds $\tilde{\lambda} \leq \lambda \leq \tilde{\lambda}_+$.

We describe in the sequel, how to choose a, b in order to ensure that $\mu(a,b)$ is certified and sharp.

3.1 Certification

The goal of this section is to determine the values of a and b such that

$$\ell := \mu(a,b) - \varepsilon = \langle r_N, (\tilde{B} - B) r_N \rangle \geq 0. \tag{3.1}$$

Since A is hermitian and $B, \tilde{B}$ are obtained by shifts of A, they share the same eigenbasis and they commute. Hence, it follows that:

$$\langle r_N, (\tilde{B} - B) r_N \rangle = \sum_{i=1}^{\mathcal{N}} \langle r_N, u^{(i)} \rangle^2 \left(\frac{\lambda^{(i)} - a}{(b - \lambda^{(i)})^2} - \frac{\lambda^{(i)} - \lambda}{(\lambda_N - \lambda^{(i)})^2} \right), \tag{3.2}$$

where $\lambda^{(i)}$ and $u^{(i)}$ denote respectively the i–th eigenvalue and eigenvector of A. Thus, a sufficient condition for the estimator $\mu(a,b)$ to be certified is:

$$\frac{\lambda^{(i)} - a}{(b - \lambda^{(i)})^2} - \frac{\lambda^{(i)} - \lambda}{(\lambda_N - \lambda^{(i)})^2} \geq 0, \quad \forall 1 \leq i \leq \mathcal{N}. \tag{3.3}$$

We shall now determine the values of a and b so that (3.5) is satisfied. To this aim, let $\tilde{\lambda} \leq \lambda$ be the *a priori* lower bound defined in section 2.2 and let:

$$\begin{aligned} x_i &:= \lambda_N - \lambda^{(i)}, \\ \alpha &:= a - \lambda_N, \\ \beta &:= b - \lambda_N, \\ \varepsilon &:= \lambda_N - \lambda, \\ \tilde{\varepsilon} &:= \lambda_N - \tilde{\lambda}. \end{aligned} \tag{3.4}$$

After substitution into Eq.(3.3), the following is obtained for every $1 \leq i \leq \mathcal{N}$

$$Q(x_i; \alpha, \beta) = \frac{\alpha - x_i}{(\beta + x_i)^2} - \frac{\varepsilon - x_i}{x_i^2} \geq 0. \tag{3.5}$$

We show that there exist values of $\alpha, \beta, \tilde{\varepsilon}$ and consequently values of a and b such that the estimator is certified. We introduce the function

$$(\alpha, \beta) \in \mathbb{R}^2 \mapsto \mathcal{J}(\alpha, \beta) := \frac{\beta^2 [\beta^2 + 4\tilde{\varepsilon}(\alpha + \beta)]}{[2\beta + \alpha - \tilde{\varepsilon}]^2}$$

and the set

$$\mathcal{T} := \{(\alpha, \beta) \in \mathbb{R}^2 \mid \quad \mathcal{J}(\alpha, \beta) \leq x_i^2, \quad \forall 1 \leq i \leq \mathcal{N}\}.$$

Thus, it holds:

Proposition 3.2. *If $(\alpha, \beta) \in \mathcal{T}$ then the estimator $\mu(\alpha + \lambda_N, \beta + \lambda_N)$ is certified.*

The proof of the proposition is presented in Appendix A. Let us remark that, in the expression of $\mathcal{J}$ only quantities that can actually be computed appear. Moreover, it holds that $\mathcal{J} \to 0$ as $\beta \to 0$. This means that, for values of b which are not too far from λ_N, the estimator is certified.

3.2 Sharpness.

To study the sharpness of the estimator, an *upper apriori* bound $\tilde{\lambda}^+ \geq \lambda$ is introduced and assumed to satisfy $\tilde{\lambda}^+ > \lambda_N \geq \lambda$. Let us recall that the difference between the estimated error $\mu(a,b)$ and the actual error ε is:

$$\ell = \sum_{i=1}^{\mathcal{N}} \langle r_N, u^{(i)} \rangle^2 Q(x_i; \alpha, \beta). \tag{3.6}$$

where the form Q is defined in (3.5). Since we do not control the term $\langle r_N, u^{(i)} \rangle$, we consider the minimization of the form Q, in a worst case sense. Let $x_0, x_1 \in \mathbb{R}$ such that $x_0 < x_1$ and $x_0 < -\beta$ and consider the interval $I = \mathbb{R} \setminus [x_0, x_1]$. Thus, the following holds:

$$\ell \leq \left(\sup_{x \in I} Q(x; \alpha, \beta) \right) \|r_N\|_2^2. \tag{3.7}$$

The term $\sup_{x \in I} Q(x; \alpha, \beta)$ can be computed explicitly. It is reduced to find the zeros of a cubic polynomial (details in the proof). However, to get some insight in the estimation, some approximations are introduced. Let us consider the function

$$K : \mathbb{R} \ni y \mapsto K(y; \alpha, \beta) = \frac{2\beta + \alpha}{(y+\beta)^2} + \frac{\beta^2}{y(y+\beta)^2}$$

where the parameters α and β are defined in (3.4). The second main result reads:

Proposition 3.3. *Let $x_0 = \frac{\beta(\beta - \tilde{\varepsilon})}{(2\beta + \alpha)}$ and $x_1 > x_0$ and consider the interval $I = \mathbb{R}/[x_0, x_1]$. Thus,*

$$\beta + \left(\frac{2\beta + \alpha + 2\tilde{\varepsilon}}{K(x_0; \alpha, \beta)} \right)^{1/2} \leq \tilde{\lambda}^+ - \lambda_N \quad \Rightarrow \quad \sup_{x \in I} Q(x; \alpha, \beta) \leq K(x_0; \alpha, \beta).$$

Moreover, the optimal values α^* and β^* are the solution to the constrained minimization problem

$$(\alpha^*, \beta^*) \in \operatorname*{argmin}_{(\alpha,\beta) \in \mathbb{R}^2} K(x_0; \alpha, \beta) \tag{3.8}$$

$$\beta + \left(\frac{(2\beta + \alpha + 2\tilde{\varepsilon})}{K(x_0; \alpha, \beta)} \right)^{1/2} \leq \tilde{\lambda}^+ - \lambda_N. \tag{3.9}$$

The proof of the proposition is presented in Appendix B. We point out that all the quantities appearing can be actually computed.

3.3 A simplified version.

In this section, a simplified version is presented, for which some estimation can be performed analytically. Let $k \in \mathbb{R}^+$. Consider the following shifts:

$$\begin{aligned} b &= (1+k)\lambda_N, \\ a &= \tilde{\lambda} + k\lambda_N. \end{aligned} \tag{3.10}$$

This implies:

$$\begin{aligned} \beta &= k\lambda_N, \\ \alpha &= \beta - \tilde{\varepsilon}. \end{aligned} \tag{3.11}$$

Hence, Eq.(3.8) can be expressed as function of β only. It holds:

Proposition 3.4. *When α, β are defined as in Eq.(3.11),*

$$\ell \leq \frac{6\|r_N\|_2^2}{\lambda_N + \tilde{\lambda}}$$

The proof of the proposition is presented in Appendix C. Remark that for a sufficiently small value of k, the estimator is certified, since $k \to 0$ implies $\beta \to 0$.

4 Implementation and numerical experiments.

In this section an efficient implementation of the proposed estimator is described and some numerical experiments are proposed to assess its properties.

4.1 Efficient implementation

The numerical implementation of the estimator is discussed in this section. One of the desirable properties of an a posteriori estimator, is to be cheap from a computational point of view. The expression of the present estimator is:

$$\mu = \langle r_N, (A-b)^{-1}(A-a)(A-b)^{-1} r_N \rangle. \tag{4.1}$$

This expression involves the inverse of the matrix $A-b$. Even if the matrix A is sparse (which is the case for most of the applications), the computation of the inverse would be prohibitive. Instead, the following computation is performed:

$$(A-b)y_N = r_N, \tag{4.2}$$
$$\mu = \langle y_N, (A-a)y_N \rangle, \tag{4.3}$$

hinting that a linear system for y_N has to be solved. This is cheaper than computing the inverse. In order for it to be even cheaper, an iterative method is used. Since A is hermitian, its eigenvalues are real, so that, since b induces a simple shift in the eigenvalues, their imaginary part remains zero. Thus, iterations of bi-conjugate gradient stabilized are effective [4]. As initial guess for the iteration, we take $y_N^{(0)} = u_N$. This is a particularly good guess, that would be the exact solution if $b = \lambda$. By doing so, only few matrix vector products are actually needed to have a good approximation of y_N. The cost of this operation is low, especially when A is sparse. Observe that, what it is important, is not the approximation of y_N *per se*, but the approximation of μ. A good stopping criterion for the iteration is thus $|\mu^{(k+1)} - \mu^{(k)}| < \delta$, where δ is a user defined tolerance, that can be chosen to fix a certain number of digits in the approximation of μ.

4.2 Random matrix smallest eigenvalue

For this first synthetic test, we create a random matrix $\tilde{A} \in \mathbb{R}^{\mathcal{N}\times\mathcal{N}}$ with $\mathcal{N} = 500$ and consider the positive symmetric matrix A defined as:

$$A = cI + \frac{1}{2}\left(\tilde{A} + \tilde{A}^T\right) \quad with \quad c = 1 - \min_{\lambda}\left[\frac{1}{2}\left(\tilde{A} + \tilde{A}^T\right) - \lambda I\right]. \tag{4.4}$$

As consequence, the matrix A is hermitian and positive definite, and its smallest eigenvalue, whose estimation error is investigated, is equal to 1. Then, the eigenvalue problem is projected in coarse basis consisting of the first N vectors of the canonical basis with $N < \mathcal{N}$, so that projecting A is equivalent to take the first N rows and columns of A.

Since the matrix is random, 32 samples were taken, and for each of them the test was performed. The values $N = [200, 250, 300, 350, 400]$ were used to compute an approximation of the spectrum. The estimator described in Section 3.3 was used for different values of $k = [0.01, 0.05, 0.1, 0.2]$. The mean of the true error and the mean of its a posteriori estimation are reported in Table 1. The estimator is rather sharp, and this is true in general for all the values of k and N. An interesting trend can be observed. For better discretizations (higher N), the parameter k that allows for better estimations is larger. This is in accordance with the theoretical estimation provided in Eq.(5.18). Indeed, for better discretization, λ_N is closer to λ, meaning that, the larger N, the lower λ_N. Consequently, the ratio $\frac{\tilde{\lambda}^+}{\lambda_N}$ becomes larger and so does the optimal value of k.

4.3 Lowest energy bands of a periodic Schrödinger operator

The second test case is related to the application that first motivated this work, the approximation of the lower energy states of the Schrödinger operator. For all $k \in \mathbb{Z}$, let $e_k(x) := \frac{1}{\sqrt{2\pi}} e^{ikx}$. For all $s \in \mathbb{N}^*$, let us define

$$X_s := \text{Span}\,\{e_k| \quad k \in \mathbb{Z},\ |k| \leq s\} \tag{4.5}$$

$k = 0.01$			$k = 0.05$			$k = 0.1$			$k = 0.2$		
N	ε	μ	N	ε	μ	N	ε	μ	N	ε	μ
200	6.72	7.71	200	6.81	7.89	200	6.84	8.26	200	6.74	9.51
250	5.36	6.31	250	5.38	6.18	250	5.44	6.26	250	5.28	6.28
300	4.14	5.08	300	4.12	4.81	300	4.16	4.68	300	4.05	4.47
350	2.98	3.91	350	2.99	3.64	350	2.99	3.40	350	2.97	3.09
400	1.93	2.85	400	1.93	2.58	400	1.94	2.33	400	1.95	2.08

Table 1: Error and error estimation for different values of the parameter k and basis size N.

and denote by $N_s := 2s+1$ the dimension of X_s and by $\Pi_{X_s} : L^2_{\text{per}} \to X_s$ the $L^2_{\text{per}}(0, 2\pi)$ orthogonal projector onto X_s. We consider the $2\pi-$ periodic real valued potential $V \in L^2_{\text{per}}(0, 2\pi)$ (plotted in Figure 1-(a)) and defined by the Fourier expansion

$$V(x) = \sum_{k=-3}^{3} \hat{V}_k e_k, \quad \hat{V}_0 = 2 \quad \text{and} \quad \hat{V}_k = 1 + 0.5\,\mathrm{i}, \quad \hat{V}_{-k} = \overline{\hat{V}_k}, \quad \forall k > 0.$$

The absolutely continuous spectrum (see [6] for the details) of the periodic Shcrödinger operator $A = -\frac{d}{dx^2} + V$ is obtained as the union of the discrete spectra of the Bloch operators $A_q := (-\mathrm{i}\frac{d}{dx} + q)^2 + V$ where $q \in [-1/2, 1/2[$. Thus, for every $q \in [-1/2, 1/2[$ an eigenvalue problem

$$A_q u = \lambda u \tag{4.6}$$

is solved in the Hilbert space $U = H^1_{\text{per}}(0, 2\pi)$. The solutions of the eigenvalue problem (4.6) are numerically approximated using a Galerkin method in Fourier space X_s. For all $s \in \mathbb{N}^*$, we denote by $\lambda^s_{q,1} \leq \cdots \leq \lambda^s_{q,N_s}$ the eigenvalues (ranked in increasing order, counting multiplicity) of the operator $A^s_q := \Pi_{X_s} A_q \Pi^*_{X_s}$. For $Q \in \mathbb{N}^*$, we introduce a regular discretization grid of the Brillouin zone $[-1/2, 1/2]$ and denote by $\Gamma^*_Q := \{q_0 = -\frac{1}{2},\ q_1, q_2, \cdots, q_Q = \frac{1}{2}\}$.

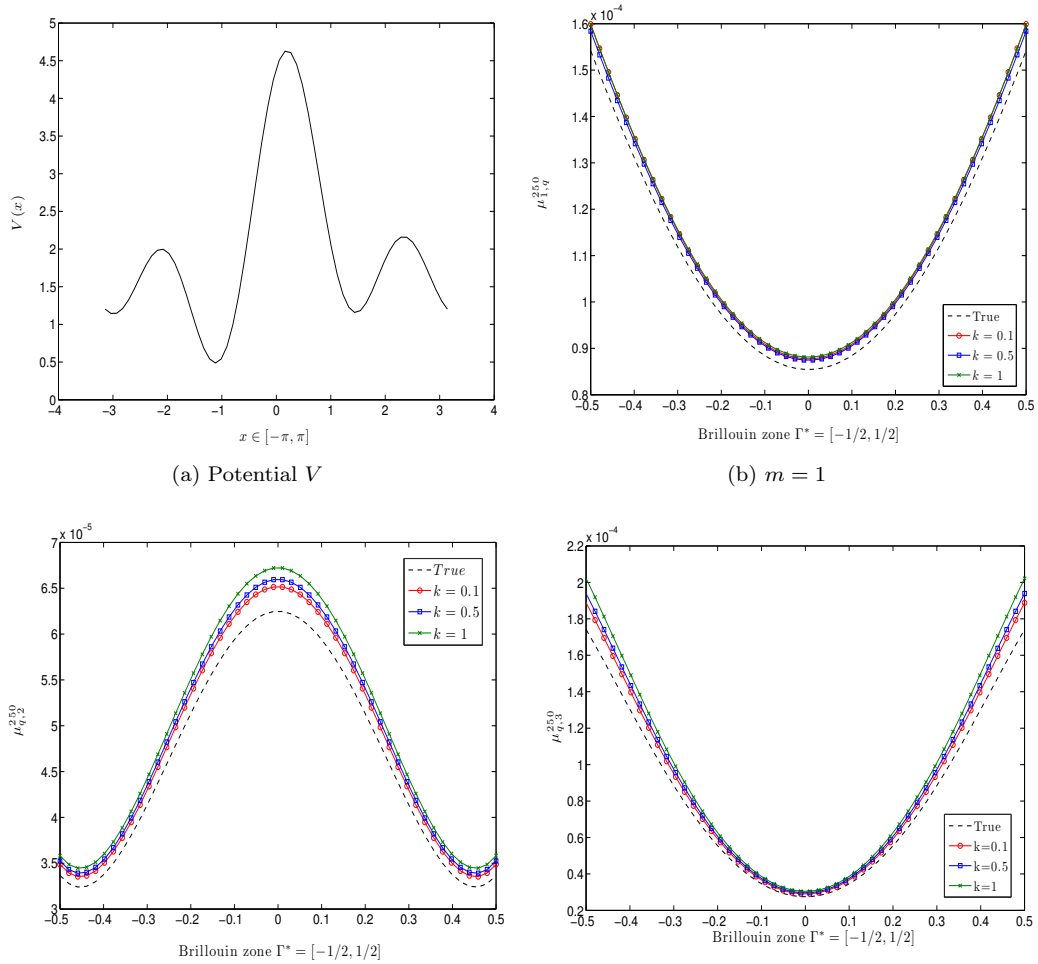
In the present test, we investigate the a posteriori error estimator presented in Section 3.3 on the three first energy bands $q \in \Gamma^*_Q \mapsto \lambda^s_{q,m}$ for $m = 1, 2, 3$ where the fine discretization of the operator is obtained with $\mathcal{N} = 501$ and the coarse discretization with $N = 13$ corresponding respectively to $s_{\text{ref}} = 250$ and $s = 6$. In Figure 1 are plotted: the true error $\varepsilon_{q,m} = \lambda^s_{q,m} - \lambda^{s_{\text{ref}}}_{q,m}$ and the a posteriori error estimator $\mu^s_{m,q}$ computed as explained previously for different values of the coefficient $k > 0$, namely $k = 0.1, 0.5, 1$.

5 Conclusions and perspectives.

An a posteriori error estimator has been proposed, based on the residual and on the definition of shifts of the spectrum of the matrix. It is shown to be conditionally certified and it provides a sharp estimation. Future directions of investigation consist in extending this type of estimator to non-positive and potentially non-hermitian eigenvalue problems, and to apply this to realistic cases.

Acknowledgments

The present work would not have been possible without fruitful discussions with Virginie Ehrlacher and David Gontier.



(a) Potential V

(b) $m = 1$

(c) $m = 2$

(d) $m = 3$

Figure 1: A posteriori error estimator for the three lowest energy bands of a one-dimensional periodic Schrödinger operator.

Appendix A: proof of Proposition 3.2.

The aim is to prove that if (α, β) belong to the set $\mathcal{T}$ then the estimator $\mu(a, b)$ is certified where $a = \alpha + \lambda_N$ and $\beta = b + \lambda_N$. As we already mentioned in Section 3.1, a sufficient condition for the estimator to be certified is that the function $\mathbb{R} \ni x \mapsto Q(x; \alpha, \beta)$ defined in (3.5) in nonnegative.

Let assume that $x \neq 0$ and $\beta \neq -x$. The first condition corresponds to the fact that the estimated eigenvalue λ_N is exactly equal to one of the true eigenvalues of A. This is related to the approximation of the problem, not to the estimator. It can happen either if N is too small and the approximation is poor, or if $\lambda_N = \lambda$. This later case, that corresponds to a zero error, is solvable. Indeed, the denominator in Q vanishes but the residual is orthogonal to the eigenvector, thus making the whole term vanishing in the series in Eq.(3.2). The condition $\beta \neq -x$ is an actual condition for b that will be analyzed later on. This condition impacts the sharpness more than the fact that the method is certified.

The expression of Q is developed, leading to

$$Q(x; \alpha, \beta) = \frac{(2\beta + \alpha - \varepsilon)x^2 - \beta(2\varepsilon - \beta)x - \varepsilon\beta^2}{x^2(\beta + x)^2} \tag{5.1}$$

The sign of Q is determined by the sign of the numerator. The discriminant of the numerator is

$$\Delta(\alpha, \beta) = \beta^2(2\varepsilon - \beta)^2 + 4\varepsilon\beta^2(2\beta + \alpha - \varepsilon) = \beta^2[\beta^2 + 4\varepsilon(\alpha + \beta)]. \tag{5.2}$$

Two cases are possible for Q to be nonnegative. The first case is when $\Delta(\alpha, \beta) < 0$ and $(2\beta + \alpha - \varepsilon) \geq 0$. The values of α and β given by this situation are not useful in our context. The second case, which is more interesting, is when $\Delta(\alpha, \beta) \geq 0$ and $(2\beta + \alpha - \varepsilon) \geq 0$.

Assume that $(2\beta + \alpha - \varepsilon) \geq 0$ which translates to

$$2\beta + \alpha \geq \varepsilon \;\Leftarrow 2\beta + \alpha \geq \lambda_N - \tilde{\lambda}, \tag{5.3}$$

meaning that, if a, b are chosen such that $2\beta + \alpha$ is larger that the difference between the estimated value and the a priori lower bound, then the condition will be automatically satisfied. Let $x_{1,2}$, $x_2 \leq x_1$ denote the two real zeros of the numerator:

$$x_{1,2} = \frac{\beta(2\varepsilon - \beta) \pm \beta[\beta^2 + 4\varepsilon(\alpha + \beta)]^{1/2}}{2[2\beta + \alpha - \varepsilon]}. \tag{5.4}$$

Thus, $Q(x; \alpha, \beta)$ is nonnegative if $x \in \mathbb{R} \setminus (x_2, x_1)$. which automatically implies that $x \neq 0$, since the zeros are of opposite sign. Observe that the case $\beta = 0 \;\Rightarrow\; b = \lambda_N$ leads to a certified estimator $x_{1,2} = 0$, along with the condition $x \neq 0$. Nevertheless, this estimator is not sharpe.

Finally, we consider the gap between the two zeros:

$$(x_1 - x_2)^2 = \frac{\beta^2[\beta^2 + 4\varepsilon(\alpha + \beta)]}{[2\beta + \alpha - \varepsilon]^2}. \tag{5.5}$$

which can not be computed, since it depends upon ε. Nevertheless, the following holds

$$(x_1 - x_2)^2 \leq \frac{\beta^2[\beta^2 + 4\tilde{\varepsilon}(\alpha + \beta)]}{[2\beta + \alpha - \tilde{\varepsilon}]^2} = J(\alpha, \beta). \tag{5.6}$$

To conclude the proof, if $x^2 \geq (x_1 - x_2)^2$ then $x \in \mathbb{R} \setminus (x_2, x_1)$ which ensures that $Q(x; \alpha, \beta)$ is nonnegative implying that the estimator $\mu(\alpha + \lambda_N, \beta + \lambda_N)$ is certified.

Appendix B: proof of the Proposition 3.3

The objective of this proof is to provide a bound for the infinity norm of $Q(x; \alpha, \beta)$ defined in Eq.(3.5). The proof is divided into two steps consisting in studying the function Q (see Figure 2) in two intervals, namely $x > x_1$ and $-\infty < x < x_*$, where x_* will be defined later on.

Let $x > x_1$.

Let us bound Q from above by means of a monotonically non-increasing function. It holds:

$$Q(x;\alpha,\beta) = \frac{(2\beta+\alpha-\varepsilon)x^2 - \beta(2\varepsilon-\beta)x - \varepsilon\beta^2}{x^2(\beta+x)^2} \leq \frac{2\beta+\alpha}{(x+\beta)^2} + \frac{\beta^2}{x(x+\beta)^2} := K(x). \tag{5.7}$$

The maximum of $Q(x)$, in $x > x_1$ is reached for a point x which is strictly larger than x_1, since $Q(x_1) = 0$. Furthermore, $\partial_x K \leq 0$, so that K is monotonically non-increasing. Thus, if $x_0 < x_1$, then $K(x_0) > K(x_1) > Q(x), \forall x > x_1$.

A point $x_0 < x_1$ is provided by $x_0 = \frac{\beta(\beta-\tilde{\varepsilon})}{(2\beta+\alpha)}$.

Thus, for $x > x_1$, the infinity norm is bounded by $K(x_0)$, which is a function of $\alpha, \beta, \tilde{\varepsilon}$.

The second part of the proof consists in studying Q for $x < -\beta$. This interval is relevant, especially when the lower eigenvalues are estimated. Indeed, $x = \lambda_N - \lambda^{(i)}$, so that most of the x are negative. Let λ be the n–th eigenvalue, the one we are currently estimating, and let λ^+ be the successive eigenvalue different from λ, *i.e.* $\lambda^+ > \lambda$.

First, it can be checked that:

$$Q \leq \frac{2\beta+\alpha+2\tilde{\varepsilon})}{(x+\beta)^2} := K_-(x), \tag{5.8}$$

by considering that $x < -\beta$ and $\tilde{\varepsilon} > \varepsilon$.

A point x_* is looked for, such that:

$$K_-(x^*) = K(x_0), \tag{5.9}$$

leading to:

$$x_* = -\beta \pm \frac{(2\beta+\alpha+2\tilde{\varepsilon})^{1/2}}{K^{1/2}(x_0)}. \tag{5.10}$$

To be sure that there are no values in proximity of the asymptotes, the sign minus can be picked.

To conclude, if $x < x_*$, then $Q(x) \leq K_-(x) \leq K_-(x^*) = K(x_0)$ and hence $\|Q\|_\infty \leq K(x_0)$ in the interval $(-\infty \leq x \leq x_*) \cup (x \geq x_1)$.

Lastly, the condition relating (α, β) to $x < x^*$ is detailed. Let us consider that, the x closer to $-\beta$ is the one for which $x = \lambda_N - \lambda^+$. Let $\tilde{\lambda}^+$ be a lower bound for the eigenvalue λ^+, that, we recall, it is the first successive eigenvalue different from λ. Then, it holds:

$$\beta + \frac{(2\beta+\alpha+2\tilde{\varepsilon})^{1/2}}{K^{1/2}(x_0)} \leq \tilde{\lambda}^+ - \lambda_N. \tag{5.11}$$

Appendix C: proof of the Proposition 3.4

The values of α, β for the simplified estimator are substituted into the System (3.8), leading to:

$$x_0 = \frac{\beta(\beta-\tilde{\varepsilon})}{(3\beta-\tilde{\varepsilon})}, \tag{5.12}$$

$$K(x_0;\alpha,\beta) = \frac{3\beta-\tilde{\varepsilon}}{(x_0+\beta)^2} + \frac{\beta^2}{x_0(x_0+\beta)^2}, \tag{5.13}$$

$$\beta + \frac{(3\beta+\tilde{\varepsilon})^{1/2}}{K^{1/2}(x_0)} \leq \tilde{\lambda}^+ - \lambda_N. \tag{5.14}$$

At the expense of a little bit of sharpness, this system can be further simplified:

$$x_0 = \frac{(\beta-\tilde{\varepsilon})}{3}, \tag{5.15}$$

$$K(x_0;\alpha,\beta) \leq \frac{6}{\beta-\tilde{\varepsilon}}, \tag{5.16}$$

$$\beta \leq \frac{1}{1+2/\sqrt{6}}\left(\tilde{\lambda}^+ - \lambda_N\right). \tag{5.17}$$

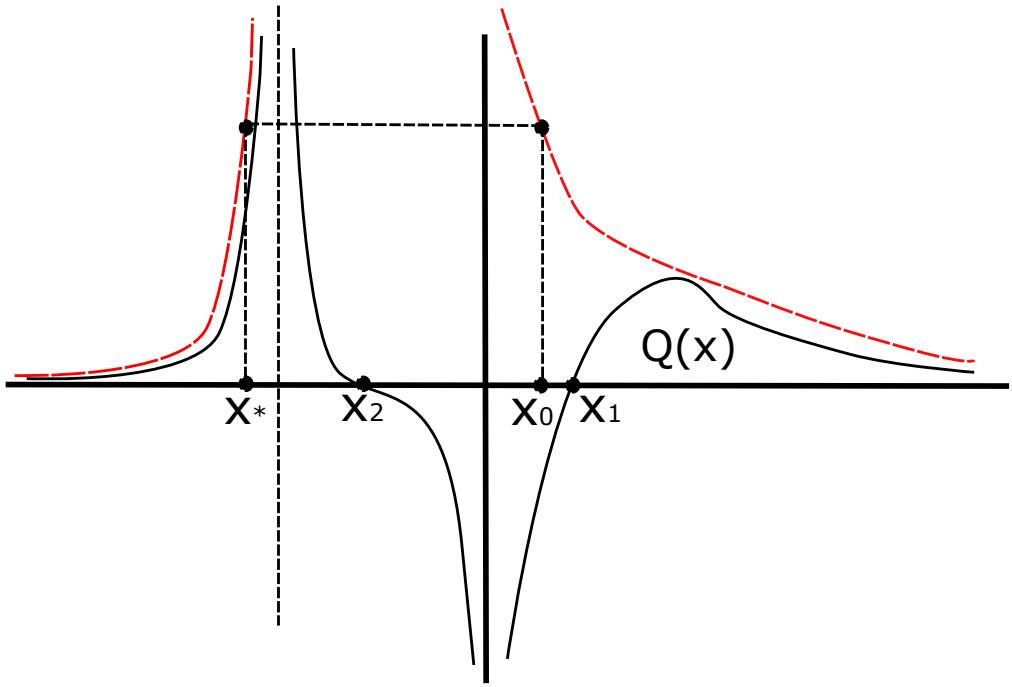


Figure 2: Picture of the functions involved: the function $Q(x)$ is in solid black line, the functions K, K_- are in dashed red line; the points $x_0, x_{1,2}$ and x_* are depicted as well.

Thus, if the estimated gap between λ and the successive eigenvalue different from λ is $\tilde{\lambda}^+ - \lambda_N$, then, the optimal β is:

$$\beta = \frac{1}{1+2/\sqrt{6}}\left(\tilde{\lambda}^+ - \lambda_N\right) \;\Rightarrow k = \frac{1}{1+2/\sqrt{6}}\left(\frac{\tilde{\lambda}^+}{\lambda_N} - 1\right). \tag{5.18}$$

For this:

$$\ell \leq \frac{6\|r_N\|_2^2}{(1+2/\sqrt{6})\tilde{\lambda}^+ - 2/\sqrt{6}\lambda_N + \tilde{\lambda}} \leq \frac{6\|r_N\|_2^2}{\lambda_N + \tilde{\lambda}}. \tag{5.19}$$

The last inequality is derived by considering that $\tilde{\lambda}^+ > \lambda_N$.

References

[1] Ivo Babuška and Werner C Rheinboldt. A-posteriori error estimates for the finite element method. *International Journal for Numerical Methods in Engineering*, 12(10):1597–1615, 1978.

[2] Ricardo G Durán, Claudio Padra, and Rodolfo Rodríguez. A posteriori error estimates for the finite element approximation of eigenvalue problems. *Mathematical Models and Methods in Applied Sciences*, 13(08):1219–1229, 2003.

[3] Vincent Heuveline and Rolf Rannacher. A posteriori error control for finite element approximations of elliptic eigenvalue problems. *Advances in Computational Mathematics*, 15(1):107–138, 2001.

[4] C. T. Kelley. *Iterative Methods for Linear and Nonlinear Equations*. Society for Industrial and Applied Mathematics, 1995.

[5] Yvon Maday, Anthony T Patera, and Jaume Peraire. A general formulation for a posteriori bounds for output functionals of partial differential equations; application to the eigenvalue problem. *Comptes Rendus de l'Académie des Sciences-Series I-Mathematics*, 328(9):823–828, 1999.

[6] M. Reed and B. Simon. *Methods of modern mathematical physics. IV: Analysis of operators.* Elsevier, 1978.

[7] Dimitrios Vasileios Rovas. *Reduced-basis output bound methods for parametrized partial differential equations.* PhD thesis, Massachusetts Institute of Technology, 2003.

[8] Gianluigi Rozza, DB Phuong Huynh, and Andrea Manzoni. Reduced basis approximation and a posteriori error estimation for stokes flows in parametrized geometries: roles of the inf-sup stability constants. *Numerische Mathematik*, 125(1):115–152, 2013.

[9] Henry Wolkowicz and George PH Styan. Bounds for eigenvalues using traces. *Linear algebra and its applications*, 29:471–506, 1980.