A Posteriori Analysis of a Non-Linear Gross-Pitaevskii type Eigenvalue Problem

Geneviève Dusson, Yvon Maday

A Posteriori Analysis of a Non-Linear Gross-Pitaevskii type Eigenvalue Problem

GENEVIÈVE DUSSON†
Laboratoire Jacques Louis Lions, Université Pierre et Marie Curie,
4 Place Jussieu, 75252, Paris Cedex 05 France and
Institut du Calcul et de la Simulation, Université Pierre et Marie Curie

AND

YVON MADAY‡
Laboratoire Jacques Louis Lions, Université Pierre et Marie Curie,
4 Place Jussieu, 75252, Paris Cedex 05 France and
Institut Universitaire de France and
Divison of Applied Mathematics, Brown University

[Received on 27 August 2013; revised on xx]

In this paper, we provide a first full *a posteriori* error analysis for variational approximations of the ground state eigenvector of a non-linear elliptic problems of the Gross-Pitaevskii type, more precisely of the form $-\Delta u+Vu+u^3=\lambda u$, $\|u\|_{L^2}=1$, with periodic boundary conditions in one dimension. Denoting by (u_N,λ_N) the variational approximation of the ground state eigenpair (u,λ) based on a Fourier spectral approximation and (u_N^k,λ_N^k) the approximate solution at the k^{th} iteration of an algorithm used to solve the non-linear problem, we first provide a precised *a priori* analysis of the convergence rates of $\|u-u_N\|_{H^1}$, $\|u-u_N\|_{L^2}$, $|\lambda-\lambda_N|$ and then present original *a posteriori* estimates in the convergence rates of $\|u_-u_N^k\|_{H^1}$ when N and k go to infinity. We introduce a residue standing for the global error $R_N^k=-\Delta u_N^k+Vu_N^k+(u_N^k)^3-\lambda_N^k u_N^k$ and we divide it into two residues characterizing respectively the error due to the discretization of the space and the finite number of iterations when solving the problem numerically. We show that the numerical results are coherent with this *a posteriori* analysis.

Keywords: a posteriori analysis, non-linear eigenvalue problem, iterative solution algorithm, plane wave approximation, stopping criteria.

1. Introduction

Non-linear eigenvalue problems are involved in many application fields such as non-linear mechanics, theoretical physics and electronic structure calculations. The numerical simulation of these problems demands a lot of computational resources both due to the accuracy that is generally required in the applications which implies the use of a large number of degrees of freedom and also due to the non-linear nature of the models that leads to iterative solution techniques that necessitate a large number of steps. The tuning of the two above ingredients involved in the approximation methods (number of

†Email: genevieve.dusson@student.ecp.fr
‡Corresponding author. Email: maday@ann.jussieu.fr

degrees of freedom and number of iterations) is, most of the times, guided in the best cases both by empirical reasons and by the available volume of computing resources. From the mathematical point of view, these questions are related to the numerical analysis of the discretization approaches that allow to state in a rigorous way the link from the discretization and solution parameters to the error between the approximate solution(s) and the exact one(s). In the range of the numerical analysis, and related to error bounds, we can distinguish between two types of contributions: the *a priori* analysis and the *a posteriori* analysis. The *a priori* version allows to qualify the tendency of the approximation properties as a function of the number of degrees of freedom and/or the amount of work necessary for the computation of the discrete solution. This is generally done by upper bounding the error by a constant times the best approximation given by the projection of the exact solution onto the discrete space. The above constant that appears in the *a priori* analysis is generally not fully known, nor actually the distance between the solution and its projection. Most of the times, the latter is evaluated from the regularity property of the solution that is at best only roughly estimated. On the contrary, the *a posteriori* analysis provides a (more or less) precise upper bound of the actual error after a computation has been performed. This bound involves only quantities that are or can easily be evaluated at the same cost as the computation of the discrete solution. Note that a posteriori analysis, thanks to the notion of indicators, may tell you what to do to improve the accuracy, but will not tell you what to do to diminish the current error by e.g. a factor 2. On the contrary, a posteriori analysis provides a stopping criteria when the desired accuracy is reached, an a priori estimator fails to do so.

The (*a priori*) numerical analysis of such non-linear eigenvalue problems is quite recent and relies in the papers (2), (3), (7), (8), (16), (12) and the references therein. These papers only consider the discretization error due to the use of a given number of degrees of freedom in order to approximate the problem of interest. For an analysis of the convergence of the iterative algorithms to solve the non-linear eigenvalue problem (or the associated non-linear minimization problem), the papers issued from (5), (1), (11) provide *a priori* convergence results and allow to understand the basics for the failure of some classical approaches and how to remedy.

As is standard, all these *a priori* approaches allow to state that, provided that you put enough computing resources, the approximation will be good. Such results are classically insufficient because the amount of required computing resources for large problems is very often out of the possibility that you can afford. This is the reason why *a posteriori* approaches (estimators and indicators) have been designed. As far as we know, the first paper in the direction of *a posteriori* estimates is (13), where the analysis of the Hartree Fock problem was performed and error bounds (i.e. upper and lower bounds) for the ground state energy was proposed. We refer also to the more recent contributions (10), (9).

The present paper is the first of a series that aims at providing precise information on the accuracy of the approximation as a function of the number of degrees of freedom that are used and the number of iterations at which we stop the numerical process. For the sake of clarity in the tools that we use, the analysis is explained on a non-linear equation that enters in the class of Gross-Pitaevskii equations ((14)) and we focus on a one dimensional example to present both the theory and the numerical simulation that illustrate it. This allows us to propose *a posteriori* estimates and indicators based on residual techniques that discriminate the effect of the discretization parameter (the number of degrees of freedom) from the parameters attached to the solution procedure (i.e. the number of iterations). The generalization of these tools for the more difficult problem of the Kohn Sham problem involves a series of technical difficulties and is on its way. We refer to (4) for a general presentation of the mathematical models and approaches for their simulations in computational quantum chemistry.

In this paper, we focus on the following non-linear eigenvalue problems arising in the study of variational models of the form

$$I = \inf\left\{E(v),\ v \in X,\ \int_\Omega v^2 = 1\right\} \tag{1.1}$$

where Ω is here simply the unit cell $(0,1)$ of a periodic lattice $\mathscr{R}$ of $\mathbb{R}$, the energy functional E is of the form

$$E(v) = \frac{1}{2}\int_\Omega |\nabla v|^2 + \frac{1}{2}\int_\Omega V v^2 + \frac{1}{4}\int_\Omega v^4,$$

and $X = H^1_\#(\Omega)$ is the Sobolev space, defined in the more general settings for any $s \in \mathbb{R}$,

$$H^s_\#(\Omega) = \{v|_\Omega,\ v \in H^s_{\rm loc}(\mathbb{R}) \mid v \text{ is 1-periodic}\},$$

provided with the norm denoted as $\|.\|_{H^s}$ and for any $k \in \mathbb{N}$,

$$C^k_\#(\Omega) = \left\{v|_\Omega,\ v \in C^k(\mathbb{R}) \mid v \text{ is 1-periodic}\right\}.$$

For $p \in [1,\infty]$, let us denote by C_p the Sobolev constant such that for any $v \in X$, $\|v\|_{L^p} \leqslant C_p \|v\|_{H^1}$. In addition, we remind the following Gagliardo Nirenberg inequality[1]

$$\forall v \in H^1_\#(\Omega), \quad \|v\|^2_{L^\infty(\Omega)} \leqslant \sqrt{5}\|v\|_{L^2(\Omega)}\|v\|_{H^1(\Omega)}. \tag{1.2}$$

In what follows, we shall assume that $V \in L^p(\Omega)$ for some $p > 1$. It was shown in (2) that (1.1) has exactly two solutions: u and $-u$ in X with $u > 0$ in Ω. From the embedding of $H^1(\Omega)$ into $C^0(\overline{\Omega})$ (valid because we are in one dimension) $u \in C^0(\overline{\Omega})$. Moreover, E is Gâteaux differentiable on X and for any $v \in X$, $E'(v) = A_v v$ where

$$A_v = -\Delta + V + v^2. \tag{1.3}$$

Under the previous assumptions E is twice differentiable at any $v \in X$ and, by denoting $E''(v)$ the second derivative of E at v, we have for any $v, w, z \in X$

$$\langle E''(v)w, z\rangle_{X',X} = \langle A_v w, z\rangle_{X',X} + 2\int_\Omega v^2 wz = \int_\Omega \nabla w.\nabla z + \int_\Omega V wz + 3\int_\Omega v^2 wz. \tag{1.4}$$

Note that A_v defines a self-adjoint operator on $L^2(\Omega)$, with form domain X (see *e.g.* (15)). The function u therefore is solution to the Euler equation

$$\forall v \in X, \quad \langle A_u u - \lambda u, v\rangle_{X',X} = 0 \tag{1.5}$$

[1]For any $v \in H^1(\Omega)$ we can indeed write

$$\forall x, y \in \Omega, \quad v^2(x) \leqslant v^2(y) + 2\sqrt{\int_\Omega v^2}\sqrt{\int_\Omega v'^2}.$$

from which we deduce, after integration in the y variable that

$$\forall x \in \Omega, \quad v^2(x) \leqslant \int_\Omega v^2 + 2\sqrt{\int_\Omega v^2}\sqrt{\int_\Omega v'^2} \leqslant \sqrt{5}\sqrt{\int_\Omega v^2}\sqrt{\int_\Omega v^2 + \int_\Omega v'^2}$$

for some $\lambda \in \mathbb{R}$ (the Lagrange multiplier associated with the constraint $\|u\|_{L^2}^2 = 1$) and equation (1.5), complemented with the constraint $\|u\|_{L^2} = 1$, takes the form of the following non-linear eigenvalue problem

$$\left\{ \begin{array}{l} A_u u = \lambda u \\ \|u\|_{L^2} = 1, \end{array} \right. \quad \text{or again} \quad \left\{ \begin{array}{l} -\Delta u + Vu + u^3 = \lambda u \\ \|u\|_{L^2} = 1, \end{array} \right. \tag{1.6}$$

which can be rewritten in a weak form as

$$\left\{ \begin{array}{l} \forall v \in X, \displaystyle\int_\Omega \nabla u \cdot \nabla v + \int_\Omega Vuv + \int_\Omega u^3 v = \lambda \int_\Omega uv \\ \displaystyle\int_\Omega u^2 = 1. \end{array} \right. \tag{1.7}$$

Let us remark that for any $v \in X$,

$$\langle E''(u)u - \lambda u, v \rangle_{X',X} = 2 \int_\Omega u^3 v. \tag{1.8}$$

In should be noted in addition,that λ is the ground state eigenvalue of the linear operator A_u. An important result is that λ is a *simple* eigenvalue of A_u (see *e.g.* the Appendix of (2)).

A natural discretization in the periodic settings consists in using a Fourier basis. We denote by $(X_N)_{N>0}$ the family of finite-dimensional subspaces of X defined by

$$X_N = \mathrm{Span}\left\{ e_k : x \mapsto e^{2ik\pi x}, |k| \leqslant N, k \in \mathbb{Z} \right\}.$$

Remind now that, for any $v \in L^2(\Omega)$,

$$v(x) = \sum_{\boldsymbol{k} \in \mathbb{Z}} \widehat{v}_k e_k(x),$$

where $\widehat{v}_k$ is the k^{th} Fourier coefficient of v:

$$\widehat{v}_k := \int_\Omega v(x) \overline{e_k(x)}\, dx = \int_\Omega v(x)\, e^{-2ik\pi x}\, dx.$$

For any real number s, we now endow the Sobolev space $H^s_\#(\Omega)$ with the equivalent norm expressed in Fourier modes as follows

$$\|v\|_{H^s} = \left(\sum_{k \in \mathbb{Z}} \left(1 + |k|^2\right)^s |\widehat{v}_k|^2 \right)^{1/2}.$$

We obtain that for any $r \in \mathbb{R}$, and all $v \in H^r_\#(\Omega)$, the best approximation of v in $H^s_\#(\Omega)$ for any $s \leqslant r$ is

$$\Pi_N v = \sum_{\boldsymbol{k} \in \mathbb{Z}, |\boldsymbol{k}| \leqslant N} \widehat{v}_k e_k.$$

The more regular v (the regularity being measured in terms of the Sobolev norms H^r), the faster the convergence of this truncated series to v: for any real numbers r and s with $s \leqslant r$, we have (see e.g. (6))

$$\forall v \in H^r_\#(\Omega), \quad \|v - \Pi_N v\|_{H^s} \leqslant \frac{1}{N^{s-r}} \|v\|_{H^r}. \tag{1.9}$$

and in particular for the solution u :

$$\min\{\|u-v_N\|_{H^1},\ v_N \in X_N\} \underset{N\to+\infty}{\longrightarrow} 0. \tag{1.10}$$

Let us now consider the variational approximation of (1.1) consisting in solving

$$I_N = \inf\left\{E(v_N),\ v_N \in X_N,\ \int_\Omega v_N^2 = 1\right\}. \tag{1.11}$$

Problem (1.11) has at least one minimizer u_N, which satisfies for some $\lambda_N \in \mathbb{R}$

$$\forall v_N \in X_N, \quad \langle A_{u_N} u_N - \lambda_N u_N, v_N\rangle_{X',X} = 0 \tag{1.12}$$

that is

$$\forall v_N \in X_N, \int_\Omega \nabla u_N \cdot \nabla v_N + \int_\Omega V u_N v_N + \int_\Omega u_N^3 v_N - \lambda_N \int_\Omega u_N v_N = 0. \tag{1.13}$$

A possible algorithm used to solve the equation numerically in the space X_N is the following : starting from a given pair (u_N^0, λ_N^0), we solve at each step the linear equation

$$\Pi_N\left(-\Delta u_N^{k*} + V u_N^{k*} + (u_N^{k-1})^2 u_N^{k*}\right) = \lambda_N^{k-1} u_N^{k-1}. \tag{1.14}$$

The discrete solution u_N^{k*} is completely determined by the knowledge of $(\lambda_N^{k-1}, u_N^{k-1})$. Since u_N^{k*} is a priori a non-normalized vector, we normalize it and define u_N^k by

$$u_N^k = \frac{u_N^{k*}}{\|u_N^{k*}\|_{L^2}}. \tag{1.15}$$

Finally, we define the approximation of the eigenvalue λ_N^k as a Rayleigh quotient being

$$\lambda_N^k = \frac{\int_\Omega (\nabla u_N^{k*})^2 + \int_\Omega V(u_N^{k*})^2 + \int_\Omega (u_N^{k*})^4}{\int_\Omega (u_N^{k*})^2} = \int_\Omega (\nabla u_N^k)^2 + \int_\Omega V(u_N^k)^2 + \int_\Omega (u_N^k)^4. \tag{1.16}$$

It should be noticed that the above algorithm corresponds to an extension of the inverse power method to this non-linear eigenvalue problem. We can check numerically that such an algorithm converges (at least in all the simulations we have performed, eventually with a relaxation parameter — see the numerical results below). Moreover we can derive that the limit (λ_N, u_N) is a good approximation of the solution to problem (1.5). More precisely we can prove the following lemma.

Lemma 1 Let us assume that there exists $u_N^* \in H^1_\#$ with $\int_\Omega [u_N^*]^2 = 1$, such that the sequence $(u_N^{k*})_{k\geqslant 1}$ converges to u_N^* in H^1-norm when k goes to infinity, then

- the sequence $(\lambda_N^k)_{k\geqslant 1}$ converges to $\lambda_N^* = \int_\Omega (\nabla u_N^*)^2 + \int_\Omega V(u_N^*)^2 + \int_\Omega (u_N^*)^4$
- the sequence $(u_N^k)_{k\geqslant 1}$ converges to u_N^* in H^1-norm
- the limit (u_N^*, λ_N^*) verifies the non-linear eigenvalue equation (1.13).

Proof. The strong convergence of u_N^{k*} to u_N^* implies that the limit of the L^2-norm of u_N^{k*} is 1 and thus the sequence (u_N^k, λ_N^k) converges to (u_N^*, λ_N^*) in H^1-norm. Then, for any $\varepsilon > 0$, there exists $k_0 \in \mathbb{N}$ such that for any $k \geqslant k_0$, $k \in \mathbb{N}$, the following assertions hold:

$$\begin{aligned} |\lambda_N^k - \lambda_N^*| &\leqslant \varepsilon, \\ \|u_N^k - u_N^*\|_{H^1} &\leqslant \varepsilon, \\ \|u_N^k - u_N^{k-1}\|_{H^1} &\leqslant \varepsilon. \end{aligned} \tag{1.17}$$

For any $k > k_0$, using (1.2), we deduce that $\|u_N^k\|_{L^\infty} \leqslant \sqrt[4]{5}\sqrt{\varepsilon + \|u_N^*\|_{H^1}}$. Let us now take $v_N \in X_N$, we have:

$$\begin{aligned} &\int_\Omega \nabla u_N^* \nabla v_N + \int_\Omega V u_N^* v_N + \int_\Omega (u_N^*)^3 v_N - \lambda_N^* \int_\Omega u_N^* v_N \\ &\qquad = \int_\Omega \nabla u_N^k \nabla v_N + \int_\Omega V u_N^k v_N + \int_\Omega (u_N^{k-1})^2 u_N^k v_N - \lambda_N^{k-1} \int_\Omega u_N^{k-1} v_N \\ &\qquad + \int_\Omega \nabla (u_N^* - u_N^k) \nabla v_N + \int_\Omega V (u_N^* - u_N^k) v_N + \int_\Omega ((u_N^*)^3 - (u_N^k)^3) v_N \\ &\qquad + \int_\Omega ((u_N^k)^3 - (u_N^{k-1})^2 u_N^k) v_N + \lambda_N^{k-1} \int_\Omega u_N^{k-1} v_N - \lambda_N^* \int_\Omega u_N^* v_N \end{aligned}$$

From (1.14) the second line is zero so we are left with

$$\begin{aligned} &\int_\Omega \nabla u_N^* \nabla v_N + \int_\Omega V u_N^* v_N + \int_\Omega (u_N^*)^3 v_N - \lambda_N^* \int_\Omega u_N^* v_N \\ &\qquad = \int_\Omega \nabla (u_N^* - u_N^k) \nabla v_N + \int_\Omega V (u_N^* - u_N^k) v_N + \int_\Omega ((u_N^*)^2 + u_N^* u_N^k + (u_N^k)^2)(u_N^* - u_N^k) v_N \\ &\qquad\quad + \int_\Omega u_N^k (u_N^k + u_N^{k-1})(u_N^k - u_N^{k-1}) v_N + (\lambda_N^{k-1} - \lambda_N^*) \int_\Omega u_N^{k-1} v_N + \lambda_N^* \int_\Omega (u_N^{k-1} - u_N^*) v_N. \end{aligned}$$

Hence

$$\begin{aligned} &\left| \int_\Omega \nabla u_N^* \nabla v_N + \int_\Omega V u_N^* v_N + \int_\Omega (u_N^*)^3 v_N - \lambda_N^* \int_\Omega u_N^* v_N \right| \\ &\qquad \leqslant \|\nabla (u_N^* - u_N^k)\|_{L^2} \|\nabla v_N\|_{L^2} + \|V\|_{L^p} \|u_N^* - u_N^k\|_{L^\infty} \|v_N\|_{L^2} \\ &\qquad + (\|u_N^*\|_{L^\infty}^2 + \|u_N^*\|_{L^\infty} \|u_N^k\|_{L^\infty} + \|u_N^k\|_{L^\infty}^2) \|u_N^* - u_N^k\|_{L^2} \|v_N\|_{L^2} \\ &\qquad + \|u_N^k\|_{L^\infty} (\|u_N^k\|_{L^\infty} + \|u_N^{k-1}\|_{L^\infty}) \|u_N^k - u_N^{k-1}\|_{L^2} \|v_N\|_{L^2} \\ &\qquad + |\lambda_N^{k-1} - \lambda_N^*| \|u_N^{k-1}\|_{L^2} \|v_N\|_{L^2} + |\lambda_N^*| \|u_N^{k-1} - u_N^*\|_{L^2} \|v_N\|_{L^2}. \end{aligned}$$

Then from (1.17) and (1.2) we derive

$$\begin{aligned} &\left| \int_\Omega \nabla u_N^* \nabla v_N + \int_\Omega V u_N^* v_N + \int_\Omega (u_N^*)^3 v_N - \lambda_N^* \int_\Omega u_N^* v_N \right| \\ &\qquad \leqslant \varepsilon \|\nabla v_N\|_{L^2} + \varepsilon \Bigg(\sqrt[4]{5}\, \|V\|_{L^p} + \|u_N^*\|_{L^\infty}^2 + \sqrt[4]{5}\, \|u_N^*\|_{L^\infty} \sqrt{\varepsilon + \|u_N^*\|_{H^1}} \\ &\qquad\qquad + \sqrt{5}(\varepsilon + \|u_N^*\|_{H^1}) + 2\sqrt{5}(\varepsilon + \|u_N^*\|_{H^1}) + 1 + |\lambda_N^*| \Bigg) \|v_N\|_{L^2}. \end{aligned}$$

We then easily deduce that the limit (u_N^*, λ_N^*) verifies the non-linear eigenvalue equation (1.13). This concludes the proof of the lemma 1. □

The following property will be used several times in the analysis:

Lemma 2 For any $v, w \in X$ such that $\int_\Omega v^2 = \int_\Omega w^2 = 1$,

$$\int_\Omega v(v-w) = 1 - \int_\Omega vw = \frac{1}{2}\|v-w\|_{L^2}^2. \tag{1.18}$$

Proof. From the L^2-normalization of v and w we write

$$\begin{aligned} \int_\Omega v(v-w) &= 1 - \int_\Omega vw = \frac{1}{2}\int_\Omega v^2 - \int_\Omega vw + \frac{1}{2}\int_\Omega w^2 \\ &= \frac{1}{2}\|v-w\|_{L^2}^2 \end{aligned}$$

which concludes the proof of the lemma. □

In the remainder of this paper, we denote by u the unique positive solution of (1.1) and by u_N a minimizer of the discretized problem (1.11) such that $(u_N, u)_{L^2} \geqslant 0$.

2. A priori analysis

The purpose of this section is to provide a precised *a priori* analysis for the approximation of problem (1.1) by (1.13), more precisely establish error bounds on $\|u_N - u\|_{H^1}$, $\|u_N - u\|_{L^2}$, $|\lambda_N - \lambda|$ and $E(u_N) - E(u)$. Actually, we follow step by step the paper (2) where the *a priori* analysis was done in a more general framework.

We provide this *a priori* analysis of (1.1) for two reasons. Firstly the particular form of the energy functional and the fact that the problem in one-dimensional allows to simplify the proofs and understand better the basic ingredients that will be used in the next section. Secondly, and more importantly, we need to be as precise as possible in order to provide an accurate evaluation of the various constants that are involved in the error bounds of the *a posteriori* analysis.

Lemma 3 **(precise version of Lemma 1 of (2))** There exist $\beta > 0$, $M_1, M_3 \in \mathbb{R}_+$ and $\gamma > 0$ such that for any $v \in X$ and any $N \in \mathbb{N}$,

$$0 \leqslant \langle (A_u - \lambda)v, v\rangle_{X',X} \leqslant M_1\|v\|_{H^1}^2 \tag{2.1}$$

$$\beta\|v\|_{H^1}^2 \leqslant \langle (E''(u) - \lambda)v, v\rangle_{X',X} \leqslant M_3\|v\|_{H^1}^2. \tag{2.2}$$

$$\gamma\|u_N - u\|_{H^1}^2 \leqslant \langle (A_u - \lambda)(u_N - u), (u_N - u)\rangle_{X',X}. \tag{2.3}$$

Moreover the constants are

$$M_m = 1 + \|V\|_{L^p} C_{2p'}^2 + m\|u\|_{L^\infty}^2 + |\lambda| \tag{2.4}$$

$$\beta = \frac{1}{2}\frac{\eta}{\eta+\chi}, \qquad \eta = \min(\lambda_2 - \lambda, 2), \qquad \chi = |\lambda| + 1 + \frac{5\|V\|_{L^p}^2}{2} \tag{2.5}$$

$$\gamma = \frac{1}{2}\frac{\eta}{\eta + 2\chi} \tag{2.6}$$

where $\lambda_2 > \lambda$ is the the second smallest eigenvalue of the equation (1.5) and $p' = (1-p^{-1})^{-1}$.

Proof. From (1.3), we have for every $v \in X$,

$$\begin{aligned} |\langle (A_u - \lambda)v, v\rangle_{X',X}| &\leqslant \|\nabla v\|_{L^2}^2 + \|V\|_{L^p}\|v\|_{L^{2p'}}^2 + \|u^2\|_{L^\infty}\|v\|_{L^2}^2 + |\lambda|\|v\|_{L^2}^2 \\ &\leqslant \left(1 + \|V\|_{L^p} C_{2p'}^2 + \|u^2\|_{L^\infty} + |\lambda|\right)\|v\|_{H^1}^2, \end{aligned}$$

where $p' = (1-p^{-1})^{-1}$. Moreover, from (1.4)

$$\begin{aligned} |\langle (E''(u) - \lambda)v, v\rangle_{X',X}| &\leqslant |\langle (A_u - \lambda)v, v\rangle_{X',X}| + 2\|u^2\|_{L^\infty}\|v\|_{L^2}^2 \\ &\leqslant \left(1 + \|V\|_{L^p} C_{2p'}^2 + 3\|u^2\|_{L^\infty} + |\lambda|\right)\|v\|_{H^1}^2, \end{aligned}$$

hence the upper bounds in (2.1) and (2.2) with constants M_1 and M_3 defined in (2.4).

The fact that λ, the lowest eigenvalue of A_u, is simple (see the Appendix of (2)) provides the lower bound in (2.1). Indeed, the operator $A_u - \lambda$ is positive over the set $u^\perp$ defined as

$$u^\perp = \left\{ v \in X \mid \int_\Omega uv = 0 \right\}, \tag{2.7}$$

more precisely we have, for any $v \in X$,

$$\langle (A_u - \lambda)v, v\rangle_{X',X} \geqslant (\lambda_2 - \lambda)(\|v\|_{L^2}^2 - |(u,v)_{L^2}|^2) \geqslant \eta(\|v\|_{L^2}^2 - |(u,v)_{L^2}|^2) \geqslant 0, \tag{2.8}$$

with $\eta = \min(\lambda_2 - \lambda, 2)$. On the one hand for any $v \in X$,

$$\begin{aligned} \langle (E''(u) - \lambda)v, v\rangle_{X',X} &= \langle (A_u - \lambda)v, v\rangle_{X',X} + 2\int_\Omega u^2 v^2 \\ &\geqslant \eta(\|v\|_{L^2}^2 - |(u,v)_{L^2}|^2) + 2\int_\Omega u^2 v^2 \\ &\geqslant \eta\|v\|_{L^2}^2 + 2\int_\Omega u^2 v^2 - \eta\left(\int_\Omega uv\right)^2 \\ &\geqslant \eta\|v\|_{L^2}^2 + (2-\eta)\left(\int_\Omega uv\right)^2 \\ &\geqslant \eta\|v\|_{L^2}^2 \end{aligned} \tag{2.9}$$

On the other hand for any $v \in X$,

$$\begin{aligned} \langle (A_u - \lambda)v, v\rangle_{X',X} &\geqslant \|\nabla v\|_{L^2}^2 - \|V\|_{L^p}\|v\|_{L^\infty}^2 - |\lambda|\|v\|_{L^2}^2 \\ &\geqslant \|v\|_{H^1}^2 - \sqrt{5}\|V\|_{L^p}\|v\|_{L^2}\|v\|_{H^1} - (|\lambda|+1)\|v\|_{L^2}^2, \end{aligned}$$

by using the Gagliardo-Nirenberg inequality (1.2). Thanks to the inequality between arithmetic and geometric means applied to $\|v\|_{H^1}$ and $\|V\|_{L^p}\|v\|_{L^2}$, we deduce that

$$\langle (A_u - \lambda)v, v\rangle_{X',X} \geqslant \frac{1}{2}\|v\|_{H^1}^2 - \left(|\lambda| + 1 + \frac{5\|V\|_{L^p}^2}{2}\right)\|v\|_{L^2}^2. \tag{2.10}$$

Combining (2.9) with (2.10) we get the lower bound in (2.2) with the constant β defined in (2.5).

To prove (2.3) we notice from (1.18) and the positivity of $(u,u_N)_{L^2}$ that

$$\|u_N\|_{L^2}^2 - |(u,u_N)_{L^2}|^2 \geqslant 1-(u,u_N)_{L^2} = \frac{1}{2}\|u_N-u\|_{L^2}^2.$$

It therefore readily follows from (1.5) and (2.8) that

$$\langle (A_u-\lambda)(u_N-u),(u_N-u)\rangle_{X',X} = \langle (A_u-\lambda)(u_N),(u_N)\rangle_{X',X} \geqslant \frac{\eta}{2}\|u_N-u\|_{L^2}^2. \tag{2.11}$$

We also have from (2.10) that

$$\langle (A_u-\lambda)(u_N-u),(u_N-u)\rangle_{X',X} \geqslant \frac{1}{2}\|u_N-u\|_{H^1}^2 - \chi\|u_N-u\|_{L^2}^2 \tag{2.12}$$

with χ defined in (2.5). From (2.11) and (2.12) we can write

$$\langle (A_u-\lambda)(u_N-u),(u_N-u)\rangle_{X',X} \geqslant \frac{1}{2}\frac{\eta/2}{\eta/2+\chi}\|u_N-u\|_{H^1}^2$$

Hence (2.3) with γ defined in (2.6). □

For $w \in X'$, we denote by ψ_w in $u^\perp$ defined in (2.7) the unique solution to the adjoint problem

$$\begin{cases} \text{find } \psi_w \in u^\perp \text{ such that} \\ \forall v \in u^\perp, \quad \langle (E''(u)-\lambda)\psi_w, v\rangle_{X',X} = \langle w,v\rangle_{X',X}, \end{cases} \tag{2.13}$$

The existence and uniqueness of the solution to (2.13) is a straightforward consequence of (2.2) and the Lax-Milgram lemma that also provides the estimate,

$$\forall w \in L^2(\Omega), \quad \|\psi_w\|_{H^1} \leqslant \beta^{-1}\|w\|_{X'} \leqslant \beta^{-1}\|w\|_{L^2}. \tag{2.14}$$

Besides this existence and stability result, the (very) simple elliptic regularity result follows

Lemma 4 Assume $V \in L^p(\Omega)$ then, there exists a constant $\tilde{C} = \frac{3\sqrt[4]{5}}{\beta}(\|V\|_{L^p}+1) + \frac{\lambda}{\beta} + 1$ such that

$$\text{If } p \geqslant 2, \quad \|\psi_w\|_{H^2} \leqslant \tilde{C}\|w\|_{L^2}. \tag{2.15}$$

$$\text{If } p < 2, \quad \|\psi_w\|_{W^{2,p}} \leqslant \tilde{C}\|w\|_{L^2}. \tag{2.16}$$

Let us now state the first *a priori* result of this section.

Theorem 1 Under the previous assumptions,

$$u_N \text{ converges strongly to } u \text{ in } H^1(\Omega) \text{ for } N \to +\infty \tag{2.17}$$

In addition, there exists $C^E \in \mathbb{R}_+$ such that for any $N \in \mathbb{N}$,

$$\frac{\gamma}{2}\|u_N-u\|_{H^1}^2 \leqslant E(u_N)-E(u) \leqslant C^E\|u_N-u\|_{H^1}^2, \tag{2.18}$$

there exists $C^\lambda \in \mathbb{R}_+$ such that for any $N \in \mathbb{N}$,

$$|\lambda_N - \lambda| \leqslant C^\lambda \left(\|u_N - u\|_{H^1}^2 + \|u_N - u\|_{L^2} \right), \tag{2.19}$$

there exists $N_0 \in \mathbb{N}$ and $C^{H1} \in \mathbb{R}_+$ such that for any $N \geqslant N_0, N \in \mathbb{N}$,

$$\|u_N - u\|_{H^1} \leqslant C^{H1} \min_{v_N \in X_N} \|v_N - u\|_{H^1}, \tag{2.20}$$

and there exists $N_1 \in \mathbb{N}$ and $C^{L2} \in \mathbb{R}_+$ such that for any $N \geqslant N_1, N \in \mathbb{N}$,

$$\|u_N - u\|_{L^2}^2 \quad \leqslant \quad C^{L2} \|u_N - u\|_{H^1} \min_{\psi_N \in X_N} \|\psi_{u_N - u} - \psi_N\|_{H^1}. \tag{2.21}$$

REMARK 2.1 It should be noticed at this level that, even if the constants above C^λ, C^{H1} and C^{L2} can be estimated quite accurately, they involve u so as does $\min_{v_N \in X_N} \|v_N - u\|_{H^1}$: it results that these estimates are not constructive.

Proof. We have

$$\begin{aligned} E(u_N) - E(u) &= \frac{1}{2}\langle A_u u_N, u_N\rangle_{X',X} - \frac{1}{2}\langle A_u u, u\rangle_{X',X} + \frac{1}{2}\int_\Omega \left(\frac{u_N^4}{2} - \frac{u^4}{2} - u^2(u_N^2 - u^2) \right) \\ &= \frac{1}{2}\langle (A_u - \lambda)(u_N - u), (u_N - u)\rangle_{X',X} + \frac{1}{4}\int_\Omega (u_N^2 - u^2)^2. \end{aligned} \tag{2.22}$$

Using (2.3) and the fact that the second term on the right hand side is positive we get

$$E(u_N) - E(u) \geqslant \frac{\gamma}{2}\|u_N - u\|_{H^1}^2.$$

From the definition of $\Pi_N u$ the H^1-projector of u on X_N, $(\Pi_N u)_{N>0}$ converges to u in X when N goes to infinity. Denoting by $\widetilde{u}_N = \|\Pi_N u\|_{L^2}^{-1} \Pi_N u$ (which is well defined, since $u > 0$ means that it is not with zero average, hence $\Pi_N u$ is never null), we also have

$$\lim_{N \to +\infty} \|\widetilde{u}_N - u\|_{H^1} = 0.$$

The functional E being strongly continuous on X, we obtain from (2.3)

$$\|u_N - u\|_{H^1}^2 \leqslant \frac{2}{\gamma}(E(u_N) - E(u)) \leqslant \frac{2}{\gamma}(E(\widetilde{u}_N) - E(u)) \underset{N \to +\infty}{\longrightarrow} 0$$

that is (2.17). It follows that there exists $N_1 \in \mathbb{N}$ such that

$$\forall N > N_1, N \in \mathbb{N}, \quad \|u_N - u\|_{H^1} \leqslant \frac{1}{2} \quad \text{and} \quad \|u_N\|_{H^1} \leqslant 2\|u\|_{H^1}. \tag{2.23}$$

Moreover

$$\begin{aligned} E(u_N) - E(u) &= \frac{1}{2}\langle (A_u - \lambda)(u_N - u), (u_N - u)\rangle_{X',X} + \frac{1}{4}\int_\Omega (u_N^2 - u^2)^2 \\ &\leqslant \frac{M_1}{2}\|u_N - u\|_{H^1}^2 + \frac{1}{4}\|u_N - u\|_{L^\infty}^2 \|u_N + u\|_{L^2}^2 \text{ (from (2.1))} \\ &\leqslant \left(\frac{M_1}{2} + \sqrt{5} \right) \|u_N - u\|_{H^1}^2 \text{(from (1.2) using } \|u\|_{L^2(\Omega)} = \|u_N\|_{L^2(\Omega)} = 1). \end{aligned}$$

Hence the upper bound in (2.18) with $C^E = \frac{M_1}{2} + \sqrt{5}$.

From (1.13) with $v_N = u_N$ and (1.7) with $v = u_N - u$, we remark that

$$\begin{aligned}
\lambda_N - \lambda &= \lambda_N \int_\Omega u_N^2 - \lambda \int_\Omega u^2 \\
&= \int_\Omega (\nabla u_N)^2 + \int_\Omega V u_N^2 + \int_\Omega u_N^4 - \left(\int_\Omega (\nabla u)^2 + \int_\Omega V u^2 + \int_\Omega u^4 \right) \\
&= \int_\Omega \nabla (u_N - u)^2 + \int_\Omega V (u_N - u)^2 + 2 \int_\Omega \nabla u \cdot \nabla (u_N - u) + 2 \int_\Omega V u (u_N - u) + \int_\Omega u_N^4 - u^4 \\
&= \langle A_u (u_N - u), (u_N - u) \rangle_{X',X} - \int_\Omega u^2 (u_N - u)^2 + 2\lambda \int_\Omega u(u_N - u) - 2 \int_\Omega u^3 (u_N - u) + \int_\Omega u_N^4 - u^4 \\
&= \langle (A_u - \lambda)(u_N - u), (u_N - u) \rangle_{X',X} + \int_\Omega u_N^2 (u_N + u)(u_N - u) \text{ (from (1.18))} \qquad (2.24)
\end{aligned}$$

we also obtain

$$\begin{aligned}
\left| \int_\Omega u_N^2 (u_N + u)(u_N - u) \right| &\leqslant \| u_N^2 (u_N + u) \|_{L^2} \| u_N - u \|_{L^2} \\
&\leqslant \| u_N \|_{L^\infty}^2 \| u_N + u \|_{L^2} \| u_N - u \|_{L^2} \\
&\leqslant \sqrt{5} \| u_N \|_{L^2} \| u_N \|_{H^1} \| u_N + u \|_{L^2} \| u_N - u \|_{L^2} \text{(from (1.2))} \\
&\leqslant 4\sqrt{5} \| u \|_{H^1} \| u_N - u \|_{L^2} \text{(from (2.23)),}
\end{aligned}$$

and from (2.1)

$$\begin{aligned}
|\lambda_N - \lambda| &\leqslant M_1 \| u_N - u \|_{H^1}^2 + 4\sqrt{5} \| u \|_{H^1} \| u_N - u \|_{L^2} \\
&\leqslant C^\lambda \left(\| u_N - u \|_{H^1}^2 + \| u_N - u \|_{L^2} \right) \qquad (2.25)
\end{aligned}$$

with $C^\lambda = \max \left(M_1, 4\sqrt{5} \| u \|_{H^1} \right)$.

In order to evaluate the H^1-norm of the error $u_N - u$, we first notice that

$$\forall v_N \in X_N, \quad \| u_N - u \|_{H^1} \leqslant \| u_N - v_N \|_{H^1} + \| v_N - u \|_{H^1}, \qquad (2.26)$$

and from (2.2) that

$$\begin{aligned}
\| u_N - v_N \|_{H^1}^2 &\leqslant \beta^{-1} \langle (E''(u) - \lambda)(u_N - v_N), (u_N - v_N) \rangle_{X',X} \\
&= \beta^{-1} \Big(\langle (E''(u) - \lambda)(u_N - u), (u_N - v_N) \rangle_{X',X} \\
&\qquad + \langle (E''(u) - \lambda)(u - v_N), (u_N - v_N) \rangle_{X',X} \Big). \qquad (2.27)
\end{aligned}$$

For any $w_N \in X_N$, using (1.4), (1.5) and (1.13)

$$\begin{aligned}
\langle (E''(u)-\lambda)(u_N-u), w_N\rangle_{X',X} &= \langle (A_u-\lambda)(u_N-u), w_N\rangle_{X',X} + 2\int_\Omega u^2(u_N-u)w_N \\
&= \langle (A_u-\lambda)u_N, w_N\rangle_{X',X} + 2\int_\Omega u^2(u_N-u)w_N \\
&= \langle (A_u-A_{u_N})u_N, w_N\rangle_{X',X} + (\lambda_N-\lambda)\int_\Omega u_N w_N + 2\int_\Omega u^2(u_N-u)w_N \\
&= (\lambda_N-\lambda)\int_\Omega u_N w_N + \int_\Omega \left(u^2u_N - u_N^3 + 2u^2(u_N-u)\right) w_N \\
&= (\lambda_N-\lambda)\int_\Omega u_N w_N - \int_\Omega (u_N-u)^2(u_N+2u)w_N. \qquad (2.28)
\end{aligned}$$

By using (1.18) with $v = u_N$ and $w = v_N$, (2.19) and (2.23), we obtain that for any $N \geqslant N_1, N \in \mathbb{N}$ and all $v_N \in X_N$ such that $\|v_N\|_{L^2} = 1$,

$$\begin{aligned}
\left|\langle (E''(u)-\lambda)(u_N-u), (u_N-v_N)\rangle_{X',X}\right| &= \left|(\lambda_N-\lambda)\langle u_N, u_N-v_N\rangle_{X',X} - \int_\Omega (u_N-u)^2(u_N+2u)(u_N-v_N)\right| \\
&\leqslant \frac{1}{2}|\lambda_N-\lambda|\|u_N-v_N\|_{L^2}^2 + \|u_N-v_N\|_{L^\infty}\|u_N-u\|_{L^2}\|u_N-u\|_{L^\infty}\|u_N+2u\|_{L^2} \\
&\leqslant \frac{1}{2}C^\lambda\left(\|u_N-u\|_{H^1}^2 + \|u_N-u\|_{L^2}\right)\|u_N-v_N\|_{L^2}^2 \\
&\quad +3\sqrt{5}\|u_N-v_N\|_{H^1}\|u_N-u\|_{L^2}\|u_N-u\|_{H^1} \qquad (2.29)
\end{aligned}$$

It then follows from (2.2) that for any $N \geqslant N_1, N \in \mathbb{N}$ and all $v_N \in X_N$ such that $\|v_N\|_{L^2} = 1$,

$$\begin{aligned}
\|u_N-v_N\|_{H^1}^2 \quad \leqslant \quad & \beta^{-1}\Big(\frac{1}{2}C^\lambda(\|u_N-u\|_{H^1}^2 + \|u_N-u\|_{L^2})\|u_N-v_N\|_{L^2}^2 \\
& +3\sqrt{5}\|u_N-v_N\|_{H^1}\|u_N-u\|_{L^2}\|u_N-u\|_{H^1} + M_3\|u-v_N\|_{H^1}\|u_N-v_N\|_{H^1}\Big).
\end{aligned}$$

So

$$\begin{aligned}
\|u_N-v_N\|_{H^1} \quad \leqslant \quad & \beta^{-1}\left(\frac{C^\lambda}{2}\left(\|u_N-u\|_{H^1}^2 + \|u_N-u\|_{L^2}\right)\|u_N-v_N\|_{H^1}\right. \\
& \left. +3\sqrt{5}\|u_N-u\|_{L^2}\|u_N-u\|_{H^1} + M_3\|u-v_N\|_{H^1}\right),
\end{aligned}$$

that is

$$\begin{aligned}
&\|u_N-v_N\|_{H^1}\left(1-\beta^{-1}\frac{C^\lambda}{2}(\|u_N-u\|_{H^1}^2 + \|u_N-u\|_{L^2})\right) \\
&\qquad\qquad \leqslant 3\sqrt{5}\|u_N-u\|_{L^2}\|u_N-u\|_{H^1} + M_3\|u-v_N\|_{H^1}. \qquad (2.30)
\end{aligned}$$

Since $\|u_N-u\|_{H^1} \underset{N\to+\infty}{\longrightarrow} 0$, there exists $N_2 \in \mathbb{N}$ such that $\forall N \geqslant N_2$,

$$\beta^{-1}\frac{C^\lambda}{2}(\|u_N-u\|_{H^1}^2 + \|u_N-u\|_{L^2}) \leqslant \frac{1}{2},$$

i.e.
$$\|u_N - v_N\|_{H^1} \leqslant 6\sqrt{5}\|u_N - u\|_{L^2}\|u_N - u\|_{H^1} + 2M_3\|u - v_N\|_{H^1}.$$

Then

$$\begin{aligned}\|u_N - u\|_{H^1} &\leqslant \|u_N - v_N\|_{H^1} + \|v_N - u\|_{H^1}\\ &\leqslant 6\sqrt{5}\|u_N - u\|_{L^2}\|u_N - u\|_{H^1} + (2M_3+1)\|u - v_N\|_{H^1},\end{aligned}$$

hence
$$\|u_N - u\|_{H^1}\left(1 - 6\sqrt{5}\|u_N - u\|_{L^2}\right) \leqslant (2M_3+1)\|u - v_N\|_{H^1} \tag{2.31}$$

Thus, there exists $N_3 \in \mathbb{N}$ such that $\forall N \geqslant N_3$,

$$6\sqrt{5}\|u_N - u\|_{L^2} \leqslant \frac{1}{2}$$

Then $\forall N \geqslant N_3$, $\|u_N - u\|_{H^1} \leqslant C^M\|u - v_N\|_{H^1}$ where $C^M \leqslant 2(2M_3+1)$ (and $C^M \equiv C^M(N) \to 2M_3+1$ as $N \to \infty$). Hence for any $N \geqslant N_3$, $\|u_N - u\|_{H^1} \leqslant C^M J_N$ where $J_N = \min\limits_{v_N \in X_N, \|v_N\|_{L^2}=1} \|v_N - u\|_{H^1}$.

We now denote by
$$\widetilde{J_N} = \min_{v_N \in X_N} \|v_N - u\|_{H^1},$$

and by u_N^0 a minimizer of the above minimization problem. We know from (1.10) that u_N^0 converges to u in H^1 when N goes to infinity. Besides,

$$\begin{aligned} J_N &\leqslant \|u_N^0/\|u_N^0\|_{L^2} - u\|_{H^1}\\ &\leqslant \|u_N^0 - u\|_{H^1} + \frac{\|u_N^0\|_{H^1}}{\|u_N^0\|_{L^2}}\left|1 - \|u_N^0\|_{L^2}\right|\\ &\leqslant \|u_N^0 - u\|_{H^1} + \frac{\|u_N^0\|_{H^1}}{\|u_N^0\|_{L^2}}\|u - u_N^0\|_{L^2} \text{ (from the triangle inequality)}\\ &\leqslant \left(1 + \frac{\|u_N^0\|_{H^1}}{\|u_N^0\|_{L^2}}\right)\widetilde{J_N}.\end{aligned}$$

From the definition of u_N^0, we have $\|u_N^0 - u\|_{H^1} \leqslant \|u_N - u\|_{H^1}$, hence using (2.23) we have, for any $N \geqslant N_1$ $\|u_N^0 - u\|_{H^1} \leqslant 1/2$, and therefore $\|u_N^0\|_{H^1} \leqslant \|u\|_{H^1} + 1/2$ and $\|u_N^0\|_{L^2} \geqslant 1/2$, yielding $J_N \leqslant 2(\|u\|_{H^1}+1)\widetilde{J_N}$. Thus
$$\|u_N - u\|_{H^1} \leqslant 2C^M(\|u\|_{H^1}+1)\min_{v_N \in X_N}\|v_N - u\|_{H^1}$$

and (2.20) is proven with $C^{H1} \leqslant 4(2M_3+1)(\|u\|_{H^1}+1)$ (and $C^{H1} \equiv C^{H1}(N) \to (2M_3+1)\|u\|_{H^1}$ as $N \to \infty$).

Let $\widetilde{u_N}$ be the orthogonal projection, for the L^2 inner product, of u_N on the affine space $S = \left\{v \in L^2(\Omega) \mid \int_\Omega uv = 1\right\}$. One has
$$\widetilde{u_N} \in X, \qquad \int_\Omega u\widetilde{u_N} = 1, \qquad u_N - \widetilde{u_N} \in S^\perp \text{ i.e. } u_N - \widetilde{u_N} \text{ is colinear to } u.$$

As $\int_\Omega (\widetilde{u_N} - u)u = 0$, we have $\widetilde{u_N} - u \in u^\perp$. Moreover

$$\begin{aligned}\int_\Omega (\widetilde{u_N} - u_N)u &= 1 - \int_\Omega u_N u \\ &= \frac{1}{2}\int_\Omega u_N^2 - \int_\Omega u_N u + \frac{1}{2}\int_\Omega u^2 \\ &= \frac{1}{2}\int_\Omega (u_N - u)^2,\end{aligned}$$

hence

$$\widetilde{u_N} - u_N = \frac{1}{2}\|u_N - u\|_{L^2}^2 u \tag{2.32}$$

We can write the following

$$\begin{aligned}\|u_N - u\|_{L^2}^2 &= \int_\Omega (u_N - u)(\widetilde{u_N} - u) + \int_\Omega (u_N - u)(u_N - \widetilde{u_N}) \\ &= \int_\Omega (u_N - u)(\widetilde{u_N} - u) - \frac{1}{2}\|u_N - u\|_{L^2}^2 \int_\Omega (u_N - u)u \text{ (from (2.32))} \\ &= \int_\Omega (u_N - u)(\widetilde{u_N} - u) + \frac{1}{2}\|u_N - u\|_{L^2}^2 \left(1 - \int_\Omega u_N u\right) \\ &= \int_\Omega (u_N - u)(\widetilde{u_N} - u) + \frac{1}{4}\|u_N - u\|_{L^2}^4 \text{ (from (1.18))} \\ &= \langle (E''(u) - \lambda)\psi_{u_N - u}, \widetilde{u_N} - u\rangle_{X',X} + \frac{1}{4}\|u_N - u\|_{L^2}^4 \text{ (from (2.13))} \\ &= \langle (E''(u) - \lambda)\psi_{u_N - u}, u_N - u\rangle_{X',X} + \langle (E''(u) - \lambda)\psi_{u_N - u}, \widetilde{u_N} - u_N\rangle_{X',X} + \frac{1}{4}\|u_N - u\|_{L^2}^4 \\ &= \langle (E''(u) - \lambda)(u_N - u), \psi_{u_N - u}\rangle_{X',X} \\ &\quad + \frac{1}{2}\|u_N - u\|_{L^2}^2 \langle (E''(u) - \lambda)u, \psi_{u_N - u}\rangle_{X',X} + \frac{1}{4}\|u_N - u\|_{L^2}^4 \text{ (from (2.32))} \\ &= \langle (E''(u) - \lambda)(u_N - u), \psi_{u_N - u}\rangle_{X',X} \\ &\quad + \|u_N - u\|_{L^2}^2 \int_\Omega u^3 \psi_{u_N - u} + \frac{1}{4}\|u_N - u\|_{L^2}^4 \text{ (from (1.8)).}\end{aligned}$$

For any $\psi_N \in X_N$, it therefore holds

$$\begin{aligned}\|u_N - u\|_{L^2}^2 &= \langle (E''(u) - \lambda)(u_N - u), \psi_N\rangle_{X',X} + \langle (E''(u) - \lambda)(u_N - u), \psi_{u_N - u} - \psi_N\rangle_{X',X} \\ &\quad + \|u_N - u\|_{L^2}^2 \int_\Omega u^3 \psi_{u_N - u} + \frac{1}{4}\|u_N - u\|_{L^2}^4.\end{aligned} \tag{2.33}$$

From (2.28) with $w_N = \psi_N$, from (2.19) and (2.23), we obtain that for any $\psi_N \in X_N \cap u^\perp$,

$$\begin{aligned}
\left|\langle (E''(u)-\lambda)(u_N-u),\psi_N\rangle_{X',X}\right| &= (\lambda_N-\lambda)\int_\Omega (u_N-u)\psi_N - \int_\Omega (u_N-u)^2(u_N+2u)\psi_N \\
&\leqslant C^\lambda\left(\|u_N-u\|_{H^1}^2+\|u_N-u\|_{L^2}\right)\|u_N-u\|_{L^2}\|\psi_N\|_{L^2} \\
&\quad +\|u_N-u\|_{L^\infty}^2\|u_N+2u\|_{L^2}\|\psi_N\|_{L^2} \\
&\leqslant \left(3+C^\lambda\right)\Big(\|u_N-u\|_{L^\infty}^2 \\
&\quad +(\|u_N-u\|_{H^1}^2+\|u_N-u\|_{L^2})\|u_N-u\|_{L^2}\Big)\|\psi_N\|_{H^1} \qquad (2.34)
\end{aligned}$$

Let $\psi_N^0 \in X_N\cap u^\perp$ be such that

$$\|\psi_{u_N-u}-\psi_N^0\|_{H^1} = \min_{\psi_N\in X_N\cap u^\perp}\|\psi_{u_N-u}-\psi_N\|_{H^1}.$$

We deduce that $\|\psi_N^0\|_{H^1}\leqslant 2\|\psi_{u_N-u}\|_{H^1}\leqslant 2\beta^{-1}\|u_N-u\|_{L^2}$, then we obtain from (2.2), (2.23) (2.33) and (2.34) that for any $N\geqslant N_1$, $N\in\mathbb{N}$,

$$\begin{aligned}
\|u_N-u\|_{L^2}^2 &\leqslant 2\beta^{-1}\left(3+C^\lambda\right)\left(\sqrt{5}\|u_N-u\|_{H^1}+\|u_N-u\|_{H^1}^2+\|u_N-u\|_{L^2}\right)\|u_N-u\|_{L^2}^2 \\
&\quad +M_3\|u_N-u\|_{H^1}\|\psi_{u_N-u}-\psi_N^0\|_{H^1}+2\beta^{-1}\|u\|_{L^3}^3\|u_N-u\|_{L^2}^3+\frac{1}{4}\|u_N-u\|_{L^2}^4,
\end{aligned}$$

hence

$$\begin{aligned}
&\|u_N-u\|_{L^2}^2(1-2\beta^{-1}\left(3+C^\lambda\right)\left(\sqrt{5}\|u_N-u\|_{H^1}+\|u_N-u\|_{H^1}^2+\|u_N-u\|_{L^2}\right) \\
&\qquad -2\beta^{-1}\|u\|_{L^3}^3\|u_N-u\|_{L^2}-\frac{1}{4}\|u_N-u\|_{L^2}^2)\leqslant M_3\|u_N-u\|_{H^1}\|\psi_{u_N-u}-\psi_N^0\|_{H^1}
\end{aligned}$$

There exists $N_4\in\mathbb{N}$ such that for any $N\geqslant N_4$, $N\in\mathbb{N}$

$$\begin{aligned}
&2\beta^{-1}\left(3+C^\lambda\right)\left(\sqrt{5}\|u_N-u\|_{H^1}+\|u_N-u\|_{H^1}^2+\|u_N-u\|_{L^2}\right) \\
&\qquad +2\beta^{-1}\|u\|_{L^3}^3\|u_N-u\|_{L^2}+\frac{1}{4}\|u_N-u\|_{L^2}^2\leqslant\frac{1}{2}.
\end{aligned}$$

Then we have for any $N\geqslant N_4$, $N\in\mathbb{N}$,

$$\|u_N-u\|_{L^2}^2\leqslant 2M_3\|u_N-u\|_{H^1}\|\psi_{u_N-u}-\psi_N^0\|_{H^1}. \qquad (2.35)$$

Lastly, we denote by $\Pi_{X_N}^1$ the orthogonal projector on X_N for the H^1-inner product; the operator $\Pi_{X_N}^\perp v = \Pi_{X_N}^1 v - \dfrac{(u,\Pi_{X_N}^1 v)_{L^2}}{(u,\Pi_{X_N}^1 u)_{L^2}}\Pi_{X_N}^1 u$ is such that $\Pi_{X_N}^\perp v\in u^\perp$ and we have that

$$\begin{aligned}
\min_{v_N\in X_N\cap u^\perp}\|v_N-v\|_{H^1} &\leqslant \|\Pi_{X_N}^\perp v-v\|_{H^1} \\
&\leqslant \left(1+\frac{\|\Pi_{X_N}^1 u\|_{H^1}}{(u,\Pi_{X_N}^1 u)_{L^2}}\right)\min_{v_N\in X_N}\|v_N-v\|_{H^1}, \qquad (2.36)
\end{aligned}$$

hence (2.21) with $C^{L2} = 2M_3 \left(1 + \frac{\|\Pi^1_{X_N} u\|_{H^1}}{(u, \Pi^1_{X_N} u)_{L^2}}\right)$, which concludes the proof of theorem 1.

□

3. A posteriori analysis

In this section we derive *a posteriori* estimates for the approximation of the problem (1.6), in order to quantify the error done during the iterative resolution (1.14), (1.15), (1.16) of the non-linear eigenvalue problem; we introduce a residue measuring how close the approximate solution, obtained after — say — k iterations (u_N^k, λ_N^k) is to the exact one (u, λ).

We are in particular interested in an upper bound for the quantity

$$\left\|u - u_N^k\right\|_{H^1} = \max_{v \in H^1_\#} \frac{\int_\Omega \nabla(u - u_N^k) \cdot \nabla v + \int_\Omega (u - u_N^k) v}{\|v\|_{H^1}} \tag{3.1}$$

where — of course — the direct knowledge of u is not available : an alternative argument that is classically used then is the indirect knowledge that we have on (u, λ) which is problem (1.6) that (u, λ) satisfies. This naturally leads to the definition of the residue :

$$R_N^k = -\Delta u_N^k + V u_N^k + (u_N^k)^3 - \lambda_N^k u_N^k, \tag{3.2}$$

that evaluates in which sense the snapshot (u_N^k, λ_N^k) obtained after iteration k of the algorithm (1.14), (1.15), (1.16) in X_N fails to solve the problem (1.6) we look for.

As was said in the introduction to this paper, this global error between the exact and the approximated solutions stems from two main sources : (i) one is the finite dimension $2N+1$ of the Fourier space X_N, i.e. the discretization of the space X , (ii) the other one is the finite number of iterations k.

In order to identify each source of errors, between the discretization parameter N and the number of iterations k, we separate the global error into two components, each of them depending mainly on one parameter associated with the above sources of error. The discretization residue is based on the numerical scheme and hence can be naturally defined as

$$R_{disc} = -\Delta u_N^k + V u_N^k + (u_N^{k-1})^2 u_N^k - \|u_N^{k*}\|_{L^2}^{-1} \lambda_N^{k-1} u_N^{k-1}, \tag{3.3}$$

the quantity $\|R_{disc}\|_{H^{-1}}$ then measures the discretization error and depends on the finite dimension $(2N+1)$ of the Fourier space X_N on which we solve the problem.

The iteration residue is then defined such that $R_N^k = R_{disc} + R_{iter}$. Hence

$$R_{iter} = (u_N^k)^3 - (u_N^{k-1})^2 u_N^k - \lambda_N^k u_N^k + \|u_N^{k*}\|_{L^2}^{-1} \lambda_N^{k-1} u_N^{k-1} \tag{3.4}$$

the quantity $\|R_{iter}\|_{H^{-1}}$ is then the iteration error and depends mainly on the finite number of iterations k.

We now relate the error on the functional space X — which is here $\|u - u_N^k\|_{H^1}$ — to the error of this specific problem expressed by an upper bound of the global residue defined previously. Besides the bounds have to be *a posteriori* computable.

First we express the term $\int_\Omega \nabla(u-u_N^k)\cdot\nabla v$ that appears in the right-hand side of (3.1) with a maximum of *a posteriori* computable terms (i.e. contributions involving u_N^k and λ_N^k and not u nor λ). Then we deal with the remaining terms. Finally we gather everything and get the *a posteriori* estimates. From (1.7), we can write the following equalities, at least in the distributional sense if the functions are not smooth enough.

$$\begin{aligned}
-\Delta(u-u_N^k) &= \lambda u - Vu - u^3 + \Delta u_N^k \\
&= \Delta u_N^k - Vu_N^k - (u_N^k)^3 + \lambda_N^k u_N^k \\
&\quad +\lambda u - \lambda_N^k u_N^k \\
&\quad +V(u_N^k - u) \\
&\quad +(u_N^k)^3 - u^3 \\
&= -R_N^k \\
&\quad +(\lambda-\lambda_N^k)u \\
&\quad +\lambda_N^k(u-u_N^k) \\
&\quad +V(u_N^k-u) \\
&\quad +(u_N^k-u)((u_N^k)^2+uu_N^k+u^2).
\end{aligned}$$

Hence

$$\begin{aligned}
\left\|u-u_N^k\right\|_{H^1} &= \max_{v\in H^1_\#} \frac{\int_\Omega \nabla(u-u_N^k)\cdot\nabla v + \int_\Omega (u-u_N^k)v}{\|v\|_{H^1}} \\
&= \max_{v\in H^1_\#} \frac{1}{\|v\|_{H^1}}\Bigg[-\langle R_N^k, v\rangle_{X',X} \\
&\quad + \int_\Omega (\lambda-\lambda_N^k)uv \\
&\quad +(\lambda_N^k+1)\int_\Omega (u-u_N^k)v + \int_\Omega V(u_N^k-u)v \\
&\quad + \int_\Omega (u_N^k-u)((u_N^k)^2+uu_N^k+u^2)v\Bigg].
\end{aligned}$$

Let us now remark that the maximum in the first line above is achieved for $v=u-u_N^k$. This allows us to notice that $\int_\Omega (u_N^k-u)((u_N^k)^2+uu_N^k+u^2)v$ is negative so that

$$\begin{aligned}
\left\|u-u_N^k\right\|_{H^1} &\leqslant \max_{v\in H^1_\#} \frac{1}{\|v\|_{H^1}}\Bigg[-\langle R_N^k, v\rangle_{X',X} \\
&\quad + \int_\Omega (\lambda-\lambda_N^k)uv \\
&\quad +(\lambda_N^k+1)\int_\Omega (u-u_N^k)v + \int_\Omega V(u_N^k-u)v\Bigg]. \qquad (3.5)
\end{aligned}$$

The only other terms on the right-hand side above which are not *a posteriori* expressible are $\lambda - \lambda_N^k$ and $u - u_N^k$. Following the same lines as in the *a priori* analysis (see (2.24)) and using (1.16) instead of (1.13) we can write

$$\begin{aligned}\lambda_N^k - \lambda &= \langle (A_u - \lambda)(u_N^k - u), (u_N^k - u)\rangle_{X',X} + \int_\Omega (u_N^k)^2 (u_N^k + u)(u_N^k - u) \\ &= \langle (A_u - \lambda)(u_N^k - u), (u_N^k - u)\rangle_{X',X} + \int_\Omega 2(u_N^k)^3 (u_N^k - u) - \int_\Omega (u_N^k)^2 (u_N^k - u)^2.\end{aligned}$$

From (2.1) we get

$$|\lambda - \lambda_N^k| \leqslant M_1 \|u - u_N^k\|_{H^1}^2 + 2\|(u_N^k)^3\|_{L^2} \|u - u_N^k\|_{L^2} + \|u_N^k\|_{L^\infty}^2 \|u - u_N^k\|_{L^2}^2. \tag{3.6}$$

We now deal with $\|u - u_N^k\|_{L^2}$. Let us define $\widetilde{u}_N^k$ as the orthogonal projection, for the L^2 inner product, of u_N^k on the affine space $S = \left\{ v \in L^2(\Omega) \mid \int_\Omega uv = 1 \right\}$. Using the fact that $\|u_N^k\|_{L^2} = 1$ for any $k, N \in \mathbb{N}$ and following the *a priori* analysis, we show that (2.32) and (2.33) hold with u_N^k instead of u_N and $\widetilde{u_N^k}$ instead of $\widetilde{u_N}$.

Therefore, for any $\psi_N \in X_N$,

$$\begin{aligned}\|u_N^k - u\|_{L^2}^2 &= \langle (E''(u) - \lambda)(u_N^k - u), \psi_N \rangle_{X',X} \\ &\quad + \langle (E''(u) - \lambda)(u_N^k - u), \psi_{u_N^k - u} - \psi_N \rangle_{X',X} \\ &\quad + \|u_N^k - u\|_{L^2}^2 \int_\Omega u^3 \psi_{u_N^k - u} + \frac{1}{4} \|u_N^k - u\|_{L^2}^4.\end{aligned}$$

At this step the *a posteriori* proof differs from the *a priori* one : For any $\psi_N \in X_N \cap u^\perp$,

$$\begin{aligned}\langle (E''(u) - \lambda)(u_N^k - u), \psi_N \rangle_{X',X} &= \langle (-\Delta + V + u^2 - \lambda)(u_N^k - u), \psi_N \rangle_{X',X} + 2\langle u^2 (u_N^k - u), \psi_N \rangle_{X',X} \\ &= \langle (-\Delta + V + u^2 - \lambda) u_N^k, \psi_N \rangle_{X',X} + 2\langle u^2 (u_N^k - u), \psi_N \rangle_{X',X} \\ &= \langle (-\Delta + V + (u_N^k)^2 - \lambda_N^k) u_N^k, \psi_N \rangle_{X',X} \\ &\quad + \langle (u^2 - (u_N^k)^2) u_N^k, \psi_N \rangle_{X',X} + (\lambda_N^k - \lambda)\langle u_N^k, \psi_N \rangle_{X',X} + 2\langle u^2 (u_N^k - u), \psi_N \rangle_{X',X} \\ &= \langle R_N^k, \psi_N \rangle_{X',X} + (\lambda_N^k - \lambda)\langle u_N^k, \psi_N \rangle_{X',X} - \langle (u - u_N^k)^2 (2u + u_N^k), \psi_N \rangle_{X',X} \\ &= \langle \Pi_N R_N^k, \psi_N \rangle_{X',X} + (\lambda_N^k - \lambda)\langle u_N^k - u, \psi_N \rangle_{X',X} - \langle (u - u_N^k)^2 (2u + u_N^k), \psi_N \rangle_{X',X}\end{aligned}$$

where, for this last line, we have used that $\psi_N \in u^\perp$. Hence

$$\begin{aligned}\left| \langle (E''(u) - \lambda)(u_N^k - u), \psi_N \rangle_{X',X} \right| &\leqslant \|\Pi_N R_N^k\|_{H^{-1}} \|\psi_N\|_{H^1} + |\lambda - \lambda_N^k| \|u_N^k - u\|_{L^2} \|\psi_N\|_{L^2} \\ &\quad + \|u_N^k - u\|_{L^\infty}^2 \|u_N^k + 2u\|_{L^2} \|\psi_N\|_{L^2}.\end{aligned} \tag{3.7}$$

Then from (2.2) for any $\psi_N \in X_N \cap u^\perp$,

$$\begin{aligned}\|u_N^k - u\|_{L^2}^2 &\leqslant \|\Pi_N R_N^k\|_{H^{-1}} \|\psi_N\|_{H^1} + |\lambda - \lambda_N^k| \|u_N^k - u\|_{L^2} \|\psi_N\|_{L^2} \\ &\quad + \|u_N^k - u\|_{L^\infty}^2 \|u_N^k + 2u\|_{L^2} \|\psi_N\|_{L^2} + M_3 \|u_N^k - u\|_{H^1} \|\psi_{u_N^k - u} - \psi_N\|_{H^1} \\ &\quad + \|u_N^k - u\|_{L^2}^2 \|u^3\|_{L^2} \|\psi_{u_N^k - u}\|_{L^2} + \frac{1}{4} \|u_N^k - u\|_{L^2}^4.\end{aligned} \tag{3.8}$$

Let $\psi_N^0 \in X_N \cap u^\perp$ be such that

$$\|\psi_{u_N^k-u} - \psi_N^0\|_{H^1} = \min_{\psi_N \in X_N \cap u^\perp} \|\psi_{u_N^k-u} - \psi_N\|_{H^1} \leqslant \|\psi_{u_N^k-u}\|_{H^1}.$$

We remark from the definition of ψ_N^0 and (2.14) that $\|\psi_N^0\|_{H^1} \leqslant 2\|\psi_{u_N^k-u}\|_{H^1} \leqslant 2\beta^{-1}\|u_N^k - u\|_{L^2}$.

From (3.8) applied to ψ_N^0 we get

$$\begin{aligned}
\|u_N^k - u\|_{L^2}^2 \quad \leqslant \quad & \frac{2}{\beta}\|\Pi_N R_N^k\|_{H^{-1}}\|u_N^k - u\|_{L^2} + \frac{2}{\beta}\|\lambda - \lambda_N^k\|\|u_N^k - u\|_{L^2}^2 \\
& + \frac{2}{\beta}\|u_N^k - u\|_{L^\infty}^2 \|u_N^k + 2u\|_{L^2}\|u_N^k - u\|_{L^2} \\
& + M_3\|u_N^k - u\|_{H^1}\|\psi_{u_N^k-u} - \psi_N^0\|_{H^1} + \frac{2}{\beta}\|u_N^k - u\|_{L^2}^3\|u^3\|_{L^2} + \frac{1}{4}\|u_N^k - u\|_{L^2}^4.
\end{aligned}$$

Moreover from (2.36), (at least if $p \geqslant 2$)

$$\begin{aligned}
\|\psi_{u_N^k-u} - \psi_N^0\|_{H^1} \quad \leqslant \quad & \left(1 + \frac{\|\Pi_{X_N}^1 u\|_{H^1}}{(u, \Pi_{X_N}^1 u)_{L^2}}\right) \min_{\psi_N \in X_N} \|\psi_{u_N^k-u} - \psi_N\|_{H^1} \\
\leqslant \quad & \left(1 + \frac{\|\Pi_{X_N}^1 u\|_{H^1}}{(u, \Pi_{X_N}^1 u)_{L^2}}\right) \frac{1}{N}\|\psi_{u_N^k-u}\|_{H^2} \\
\leqslant \quad & \frac{\tilde{C}}{N}\left(1 + \frac{\|\Pi_{X_N}^1 u\|_{H^1}}{(u, \Pi_{X_N}^1 u)_{L^2}}\right)\|u_N^k - u\|_{L^2} \text{ (from (2.15))}, \qquad (3.9)
\end{aligned}$$

(note that, if $p < 2$ we get $\|\psi_{u_N^k-u} - \psi_N^0\|_{H^1} \leqslant \frac{\tilde{C}}{\sqrt{N}}\left(1 + \frac{\|\Pi_{X_N}^1 u\|_{H^1}}{(u, \Pi_{X_N}^1 u)_{L^2}}\right)\|u_N^k - u\|_{L^2}$ instead of (3.9) and what follows can be adjusted accordingly)

From (3.6), (3.9) and (1.2) we have

$$\begin{aligned}
\|u_N^k - u\|_{L^2}^2 \quad \leqslant \quad & \frac{1}{\beta}\|\Pi_N R_N^k\|_{H^{-1}}\|u_N^k - u\|_{L^2} \\
& + \frac{1}{\beta}\left(M_1\|u - u_N^k\|_{H^1}^2 + 2\|(u_N^k)^3\|_{L^2}\|u - u_N^k\|_{L^2} + \|u_N^k\|_{L^\infty}^2\|u - u_N^k\|_{L^2}^2\right)\|u_N^k - u\|_{L^2}^2 \\
& + \frac{\sqrt{5}}{\beta}\|u_N^k - u\|_{H^1}\|u_N^k + 2u\|_{L^2}\|u_N^k - u\|_{L^2}^2 \\
& + \frac{\tilde{C}M_3}{N}\left(1 + \frac{\|\Pi_{X_N}^1 u\|_{H^1}}{(u, \Pi_{X_N}^1 u)_{L^2}}\right)\|u_N^k - u\|_{H^1}\|u_N^k - u\|_{L^2} \\
& + \frac{1}{\beta}\|u_N^k - u\|_{L^2}^3\|u^3\|_{L^2} + \frac{1}{4}\|u_N^k - u\|_{L^2}^4. \qquad (3.10)
\end{aligned}$$

Simplifying by $\|u_N^k - u\|_{L^2}$ we get

$$\|u_N^k - u\|_{L^2}\Bigg(1 - \frac{1}{\beta}\left(M_1\|u-u_N^k\|_{H^1}^2 + 2\|(u_N^k)^3\|_{L^2}\|u-u_N^k\|_{L^2} + \|u_N^k\|_{L^\infty}^2\|u-u_N^k\|_{L^2}^2\right)$$
$$- \frac{\sqrt{5}}{\beta}\|u_N^k - u\|_{H^1}\|u_N^k + 2u\|_{L^2} - \frac{1}{\beta}\|u_N^k - u\|_{L^2}\|u^3\|_{L^2} - \frac{1}{4}\|u_N^k - u\|_{L^2}^2\Bigg)$$
$$\leqslant \frac{1}{\beta}\|\Pi_N R_N^k\|_{H^{-1}} + \frac{\tilde{C}M_3}{N}\left(1 + \frac{\|\Pi_{X_N}^1 u\|_{H^1}}{(u, \Pi_{X_N}^1 u)_{L^2}}\right)\|u_N^k - u\|_{H^1}. \quad (3.11)$$

Since $\|u - u_N^k\|_{H^1} \underset{N,k\to\infty}{\longrightarrow} 0$, for any $\alpha_1 > 1$ there exists $N_{\alpha_1} \in \mathbb{N}$ and $k_{\alpha_1} \in \mathbb{N}$ such that for any $N \geqslant N_{\alpha_1}$, for any $k \geqslant k_{\alpha_1}$,

$$\frac{1}{\beta}\left(M_1\|u-u_N^k\|_{H^1}^2 + 2\|(u_N^k)^3\|_{L^2}\|u-u_N^k\|_{L^2} + \|(u_N^k)^2\|_{L^\infty}\|u-u_N^k\|_{L^2}^2\right)$$
$$+ \frac{\sqrt{5}}{\beta}\|u_N^k - u\|_{H^1}\|u_N^k + 2u\|_{L^2} + \frac{1}{\beta}\|u_N^k - u\|_{L^2}\|u^3\|_{L^2} + \frac{1}{4}\|u_N^k - u\|_{L^2}^2 \leqslant 1 - \frac{1}{\alpha_1}, \quad (3.12)$$

so that we are left with

$$\|u_N^k - u\|_{L^2} \leqslant \frac{\alpha_1}{\beta}\|\Pi_N R_N^k\|_{H^{-1}} + \frac{\alpha_1\tilde{C}M_3}{N}\left(1 + \frac{\|\Pi_{X_N}^1 u\|_{H^1}}{(u, \Pi_{X_N}^1 u)_{L^2}}\right)\|u_N^k - u\|_{H^1}. \quad (3.13)$$

Final step: From (3.5) (where we use the fact that the element that maximizes the expression is e.g. $v = u - u_N^k$) and (3.6), we can write that

$$\begin{aligned}\|u - u_N^k\|_{H^1} \quad \leqslant \quad & \|R_N^k\|_{H^{-1}} \\ & + M_1\|u-u_N^k\|_{H^1}^2 \\ & + 2\|(u_N^k)^3\|_{L^2}\|u-u_N^k\|_{L^2} \\ & + \|(u_N^k)^2\|_{L^\infty}\|u-u_N^k\|_{L^2}^2 \\ & + \left(|\lambda_N^k + 1| + \|V\|_{L^1}\right)\|u-u_N^k\|_{L^2}\end{aligned}$$

and from (3.13) we write

$$\begin{aligned}\|u - u_N^k\|_{H^1} \quad \leqslant \quad & \|R_N^k\|_{H^{-1}} \\ & + M_1\|u-u_N^k\|_{H^1}^2 + \|u_N^k\|_{L^\infty}^2\|u-u_N^k\|_{L^2}^2 \\ & + \left(2\|(u_N^k)^3\|_{L^2} + |\lambda_N^k + 1| + \|V\|_{L^1}\right) \\ & \Bigg(\frac{\alpha_1}{\beta}\|\Pi_N R_N^k\|_{H^{-1}} \\ & + \frac{\alpha_1\tilde{C}M_3}{N}\left(1 + \frac{\|\Pi_{X_N}^1 u\|_{H^1}}{(u, \Pi_{X_N}^1 u)_{L^2}}\right)\|u_N^k - u\|_{H^1}\Bigg).\end{aligned}$$

Hence

$$\|u-u_N^k\|_{H^1}\Bigg(1-M_1\|u-u_N^k\|_{H^1}-\|u_N^k\|_{L^\infty}^2\|u-u_N^k\|_{L^2} - \Big(2\|(u_N^k)^3\|_{L^2}+|\lambda_N^k+1|+\|V\|_{L^1}\Big)\frac{\alpha_1\tilde{C}M_3}{N}\left(1+\frac{\|\Pi^1_{X_N}u\|_{H^1}}{(u,\Pi^1_{X_N}u)_{L^2}}\right)\Bigg) \leqslant \|R_N^k\|_{H^{-1}}+\Big(2\|(u_N^k)^3\|_{L^2}+|\lambda_N^k+1|+\|V\|_{L^1}\Big)\frac{\alpha_1}{\beta}\|\Pi_N R_N^k\|_{H^{-1}}.$$

Since $\|u-u_N^k\|_{H^1}\underset{N,k\to\infty}{\longrightarrow}0$, for any $\alpha_2>1$ there exists $N_{\alpha_2}\in\mathbb{N}$ and $k_{\alpha_2}\in\mathbb{N}$ such that for any $N\geqslant N_{\alpha_2}$, for any $k\geqslant k_{\alpha_2}$,

$$M_1\|u-u_N^k\|_{H^1}+\|(u_N^k)^2\|_{L^\infty}\|u-u_N^k\|_{L^2} +\frac{1}{N}\Big(2\|(u_N^k)^3\|_{L^2}+|\lambda_N^k+1|+\|V\|_{L^1}\Big)\frac{\alpha_1\tilde{C}M_3}{\beta}\left(1+\frac{\|\Pi^1_{X_N}u\|_{H^1}}{(u,\Pi^1_{X_N}u)_{L^2}}\right)\leqslant 1-\frac{1}{\alpha_2}$$

Then we have for $N\geqslant N_{\alpha_2}$ and $k\geqslant k_{\alpha_2}$

$$\frac{1}{\alpha_2}\|u-u_N^k\|_{H^1} \leqslant \|R_N^k\|_{H^{-1}} +\Big(2\|(u_N^k)^3\|_{L^2}+|\lambda_N^k+1|+\|V\|_{L^1}\Big)\frac{\alpha_1}{\beta}\|\Pi_N R_N^k\|_{H^{-1}}$$

Let us now notice that, in the limit we have $\alpha_1\underset{N,k\to\infty}{\longrightarrow}1$ and $\alpha_2\underset{N,k\to\infty}{\longrightarrow}1$, so that we have

Theorem 2 We have the following a posteriori error bound

$$\|u-u_N^k\|_{H^1} \leqslant \theta\Bigg(\|R_N^k\|_{H^{-1}}+\Big(2\|(u_N^k)^3\|_{L^2}+|\lambda_N^k+1|+\|V\|_{L^1}\Big)\frac{1}{\beta}\|\Pi_N R_N^k\|_{H^{-1}}\Bigg) \quad (3.14)$$

with a constant $0<\theta\leqslant 1$ and as close to 1 as we wish as the error goes to zero.

This expression traduces the *a posteriori* relation between the error in H^1-norm and the global residue R_N^k. Let us now split the global residue into its two components in order to make the iteration and discretization errors explicit. From the definition of the discretization and iteration residues (3.3) and (3.4) we write

$$\|u-u_N^k\|_{H^1} \leqslant \theta\Bigg(\|R_{disc}+R_{iter}\|_{H^{-1}}+\Big(2\|u_N^k\|_{L^6}^3+|\lambda_N^k+1|+\|V\|_{L^1}\Big)\frac{1}{\beta}\|\Pi_N(R_{disc}+R_{iter})\|_{H^{-1}}\Bigg)$$

As the discretization residue R_{disc} has been defined from the numerical scheme described in the introduction, the projection of the residue on X_N is zero, that is $\Pi_N(R_{disc})=0$. Then, for any $\theta>0$, as close to 1 as we wish when the convergence is acknowledged, we have the following decoupled upper bound

$$\|u-u_N^k\|_{H^1} \leqslant \theta\Bigg(\|R_{disc}\|_{H^{-1}}+\bigg(1+\Big(2\|u_N^k\|_{L^6}^3+|\lambda_N^k+1|+\|V\|_{L^1}\Big)\frac{1}{\beta}\bigg)\|R_{iter}\|_{H^{-1}}\Bigg). \quad (3.15)$$

Thus the H^1-error can be *a posteriori* bounded by fully computable terms : indeed (i) in dimension 1 the Sobolev constants are known, (ii) for planewave discretization the norm H^{-1} is computable, (iii) moreover the constant β has been explicitly expressed in the *a priori* analysis and can be, if not computed directly, at least bounded by a computable term.

4. Numerical results

In this section we gather some numerical results that illustrate the results proven so far, in particular the *a posteriori* analysis of the non-linear eigenvalue problem (1.7). First we provide some results on the properties of the "inverse power" method (1.14), (1.15), (1.16). Then we show that the numerical results are coherent with the *a posteriori* analysis and that the separation of the error components is relevant. We also study the influence of the potential regularity on the H^1-convergence and the effect of the variation of the non-linearity on the convergence of the algorithm used to solve the problem.

In the next subsections 4.2, 4.3 and 4.4, we evaluate the *a posteriori* estimators found in the previous section and perform numerical simulations with a potential V given by its Fourier coefficients

$$\widehat{V}_k = -\frac{1}{\sqrt{2\pi}}\frac{1}{|k|^4-\frac{1}{4}}, \tag{4.1}$$

from which we deduce that $V \in L^p$ for any $p > 1$. The approximate potential we consider is calculated with 801 coefficients i.e. $k \leqslant 400$.

4.1 *General framework*

We first verify the convergence of the "inverse power" method (1.14), (1.15), (1.16) for different "strength" of the non-linear contribution. Let us define $\alpha \in \mathbb{R}^+$ the coefficient of non-linearity. Equation (1.7) becomes

$$\left\{\begin{array}{l} \forall v \in X, \displaystyle\int_\Omega \nabla u\cdot\nabla v + \int_\Omega Vuv + \alpha\int_\Omega u^3 v = \lambda \int_\Omega uv \\ \displaystyle\int_\Omega u^2 = 1. \end{array}\right. \tag{4.2}$$

Taking this coefficient into account, the algorithm we use to solve the equation in the space X_N is similar to (1.14, 1.15, 1.16)

This algorithm converges numerically for small values of α. However for large values of α the algorithm does not converge anymore. This non-convergence starts for α in the range of 10. In order to avoid this problem, two solutions have been considered. First the convergence is improved for larger dimension of the space X_N. Hence we can increase the dimension in the numerical "exact" space and in the approximate spaces X_N. Another solution is to introduce a relaxation coefficient η, $0 < \eta \leqslant 1$, such that for each k a relaxation step is added in the algorithm as we define $\widetilde{u_N^k} = \eta u_N^k + (1-\eta)u_N^{k-1}$. This improves the convergence of the algorithm. For example for $N = 80$ and $\eta = 0.3$ the algorithm converges in less than 100 iterations for α up to 15 and the number of iterations required to verify the condition $\|u_N^k - u_N^{k-1}\|_{H^1} < 10^{-12}$ increases from 27 to 80 when α increases from 4 to 15.

In all what follows, α is fixed equal to 1. The numerical "exact solution" is computed in the space $X_{300} = \mathrm{Span}\{e_k, |k| \leqslant 300\}$, and the number of iterations is pushed so that the global residue defined above for this numerical "exact solution" is:

$$\|R_N^{k_{max}}\|_{H^{-1}} = \|-\Delta u_N^{k_{max}} + V u_N^{k_{max}} + (u_N^{k_{max}})^3 - \lambda_N^{k_{max}} u_N^{k_{max}}\|_{H^{-1}} = 1.7283.10^{-12}$$

which is achieved with $k_{max} = 32$.

The total error is given by the *a posteriori* bound (3.14), that is

$$err_{total} = \|R_N^k\|_{H^{-1}} + \left(2\|u_N^k\|_{L^6}^3 + |\lambda_N^k + 1| + \|V\|_{L^1}\right)\frac{1}{\beta}\|\Pi_N R_N^k\|_{H^{-1}} \tag{4.3}$$

Two error components are defined from the bound (3.15): the k-error and the N-error which depend respectively mainly on k and N. More precisely we define the k-error by

$$err_k = \left(1 + \left(2\|u_N^k\|_{L^6}^3 + |\lambda_N^k + 1| + \|V\|_{L^1}\right)\frac{1}{\beta}\right)\|R_{iter}\|_{H^{-1}} \tag{4.4}$$

and the N-error by

$$err_N = \|R_{disc}\|_{H^{-1}} \tag{4.5}$$

Let us notice that the total error is not exactly the sum of the two error components. Indeed

$$err_{total} < err_k + err_N,$$

hence err_{total} is a sharper estimate than the sum of the two contributions $err_k + err_N$.

4.2 *With a large number of iterations*

In this subsection, we compute different approximate solutions using a given large number of iterations and varying the dimension of the Fourier space X_N (see table 1). The number of iterations is $k_{max} = 32$ in our case for N between 15 and 100. Recall that this value of k_{max} corresponds to the minimum of iterations required to complete the condition the residual is less than 10^{-12} for $N = 300$.

Let us first remark that the total error (4.3) (with θ chosen equal to 1) is larger than $\|u_{exact} - u_N^k\|_{H^1}$ which confirms the fact that the convergence of θ to 1 is very fast, since then the total error is an upper bound for $\|u_{exact} - u_N^k\|_{H^1}$). Secondly the k-error obtained is close to 10^{-12} and almost constant which depicts the fact that the k-error is independent of N and almost zero when the algorithm has converged. The N-error is then the main component of the total error and decreases from 10^{-6} to 10^{-10}.

The total error is very close to $\|u_{exact} - u_N^k\|_{H^1}$ which shows that the *a posteriori* bounds found in the previous analysis seem to be close to optimal when the algorithm has converged.

4.3 *In large dimension for the discretization space*

In this subsection, we compute the approximate solution using a large dimension for the discretization space ($N = 100$) and varying the number of iterations. The number of iterations varies from 1 to 32.

As in the previous subsection we remark that the total error is larger than $\|u_{exact} - u_N^k\|_{H^1}$.

N	k	N-error	k-error	Total error	$\|u_{exact} - u_N^k\|_{H^1}$
15	32	1.291758e-06	1.759282e-12	1.291759e-06	1.090845e-06
20	32	3.636445e-07	1.748157e-12	3.636463e-07	3.109467e-07
25	32	1.354259e-07	1.738445e-12	1.354277e-07	1.166171e-07
30	32	6.028068e-08	1.722262e-12	6.028240e-08	5.214103e-08
35	32	3.036443e-08	1.726690e-12	3.036616e-08	2.634541e-08
40	32	1.674949e-08	1.716547e-12	1.675121e-08	1.456524e-08
45	32	9.904824e-09	1.714087e-12	9.906538e-09	8.627907e-09
50	32	6.188268e-09	1.710524e-12	6.189979e-09	5.397729e-09
55	32	4.042411e-09	1.707464e-12	4.044119e-09	3.529819e-09
60	32	2.739745e-09	1.706193e-12	2.741451e-09	2.394466e-09
65	32	1.915264e-09	1.702199e-12	1.916966e-09	1.675139e-09
70	32	1.374711e-09	1.708393e-12	1.376420e-09	1.203118e-09
75	32	1.009446e-09	1.701410e-12	1.011148e-09	8.839268e-10
80	32	7.560881e-10	1.699141e-12	7.577872e-10	6.623852e-10
85	32	5.762858e-10	1.707072e-12	5.779929e-10	5.050752e-10
90	32	4.460840e-10	1.704587e-12	4.477886e-10	3.911054e-10
95	32	3.500942e-10	1.704633e-12	3.517989e-10	3.070460e-10
100	32	2.781826e-10	1.693670e-12	2.798763e-10	2.440477e-10

Table 1. Error components evolution with large number of iterations

We observe that k has an influence on the N-error so the N-error is not independent of k even though it depends mainly on N. and is much smaller than the k-error during the first iterations (up to a multiplicative factor equal to 10^{-4}) and becomes constant from $k = 20$: the N-error decreases up to 2.10^{-10} for $k \geqslant 20$.

The k-error decreases regularly from 10^2 to 10^{-12}. The main error component is at first the k-error and then the N-error for k between 29 and 32.

The factor between the total error and $\|u_{exact} - u_N^k\|_{H^1}$ decreases from 10^2 to 1.2. The large constant for the first iterations can be explained by some inequalities used in the *a posteriori* analysis verified uniquely for k and N large enough and simulated here starting from $k = 1$. Another possibility is that the negative term deleted in the *a posteriori* analysis which is of the form $-\int_\Omega (u_N^k - u)^2((u_N^k)^2 + uu_N^k + u^2)$ is not small for small k and therefore leads to a big difference between the value of the upper bound and the value of the real error. Taking this term into account in the *a posteriori* analysis could lead to better numerical results. This is one of the things we try to incorporate in the extension of this paper to the full Density Functional Theory framework.

This illustrates the fact that the k-error estimator can be used as a stopping criterion for the convergence of the “inverse power” iterative technique (1.14), (1.15), (1.16).

N	k	N-error	k-error	Total error	$\|u_{exact} - u_N^k\|_{H^1}$
100	1	4.493523e-02	1.355092e+02	1.355542e+02	1.740224e+00
100	3	1.025217e-02	7.378567e+00	7.388819e+00	5.747210e-01
100	5	2.054271e-03	7.888511e-01	7.909053e-01	9.440197e-02
100	7	2.137739e-04	9.629828e-02	9.651205e-02	1.423893e-02
100	9	2.696019e-05	1.369212e-02	1.371908e-02	2.059918e-03
100	11	3.684329e-06	1.943385e-03	1.947069e-03	2.936097e-04
100	13	5.170146e-07	2.756909e-04	2.762079e-04	4.170375e-05
100	15	7.312565e-08	3.911323e-05	3.918636e-05	5.918803e-06
100	17	1.036984e-08	5.549499e-06	5.559869e-06	8.398648e-07
100	19	1.496524e-09	7.873990e-07	7.888955e-07	1.191691e-07
100	21	3.477196e-10	1.117222e-07	1.120699e-07	1.691028e-08
100	23	2.797530e-10	1.585204e-08	1.613179e-08	2.411285e-09
100	25	2.782143e-10	2.249220e-09	2.527435e-09	4.186479e-10
100	27	2.781833e-10	3.191342e-10	5.973175e-10	2.487320e-10
100	29	2.781826e-10	4.527943e-11	3.234621e-10	2.441369e-10
100	31	2.781826e-10	6.421043e-12	2.846037e-10	2.440488e-10
100	32	2.781826e-10	2.420327e-12	2.806030e-10	2.440478e-10

Table 2. Error components evolution in large dimensional space

4.4 *Error balance*

We would now like to minimize the computational cost necessary to reach a given acceptable total error. So given a small number ε, we want to find N and k as small as possible such that the total error is not larger than ε. As the total error is less than the sum of two error components depending respectively mainly on k or N, we try to balance the two error components.

Let ε be the total acceptable error. The aim is to find (N,k) such that $err_{total} < \varepsilon$. The algorithm used is the following:

1. Set $N = 10$ and $k = 1$. Choose an initial pair (u_N^0, λ_N^0).
2. From $(u_N^{k-1}, \lambda_N^{k-1})$, find (u_N^k, λ_N^k) solution of (1.14) and (1.16).
3. Compute err_{total}
    - If $err_{total} < \varepsilon$, then return (N,k)
    - Else compute err_N and err_k
        - If $err_k > err_N$ then back to 2 with $k = k+1$.
        - Else back to 1 with $N = N+2$.

The table 3 shows the results of this study. Both the number of iterations and the dimension necessary to achieve a given accuracy increase when the accuracy increases.

N	k	k-error	N-error	Total error	$\|u_{exact} - u_N^k\|_{H^1}$
10	6	1.871314e-01	3.929046e-03	2.504778e-01	2.734769e-02
10	7	7.079140e-02	1.424871e-03	9.616493e-02	1.058219e-02
10	9	1.011118e-02	1.934565e-04	1.377042e-02	1.526087e-03
10	10	3.811613e-03	7.247742e-05	5.190456e-03	5.760686e-04
10	11	1.436084e-03	2.804418e-05	1.955458e-03	2.172673e-04
10	12	5.409683e-04	1.266597e-05	7.367374e-04	8.207722e-05
10	13	2.037707e-04	8.496406e-06	2.779414e-04	3.145792e-05
10	15	2.891183e-05	7.612298e-06	4.223632e-05	7.619566e-06
10	16	1.089043e-05	7.595828e-06	1.940165e-05	6.452118e-06
12	17	4.068677e-06	3.433588e-06	7.753818e-06	2.926068e-06
16	18	1.516084e-06	9.728359e-07	2.652929e-06	8.553814e-07
20	19	5.671734e-07	3.636831e-07	9.944701e-07	3.225775e-07
26	21	7.994209e-08	1.138046e-07	1.931176e-07	9.884332e-08
32	22	2.998194e-08	4.524367e-08	7.477359e-08	3.944923e-08
42	23	1.124065e-08	1.347480e-08	2.509457e-08	1.184888e-08
54	24	4.219241e-09	4.387882e-09	8.857335e-09	3.883516e-09
68	26	5.970657e-10	1.565134e-09	2.073886e-09	1.372402e-09
88	27	2.243446e-10	4.933429e-10	6.974298e-10	4.337893e-10

Table 3. Error balance

It is interesting to note that the way k and N increase as the error decreases in the following manner $k = -2.27\log(err_k) + 4.45$ while $log(N) = -0.224\log(err_N) - 0.144$, the figure 1 illustrates this point. Note that N is 10 at the beginning of the error balance which explains the left part of the right figure.

5. Conclusions

In this paper we have proposed a completely new a posteriori analysis for the solution technique applied to a simple nonlinear eigenvalue problem. The estimator is indeed cut into two pieces, each one dedicated to characterize the error due to the number of degrees of freedom used to discretize the problem and the number of iteration for the fixed point approach allowing to solve the non-linear problem. The numerical simulations that have been provided show that the very precise analysis that can be done on this one dimensional example is optimal since the ratio between the error estimate and the exact error appears close to 1. The extension of these ideas and techniques to a more complex framework of the Kohn-Sham problem is under consideration, following the techniques presented in (3).

Acknowledgements

The authors are grateful to Eric Cancès, Benjamin Stamm and Martin Vohralik for helpful discussions and comments on a preliminary version of the present article. Financial support from the French state funds managed by CALSIMLAB and the ANR within the Investissements d'Avenir programme under reference ANR-11-IDEX-0004-02 together with the ANR Manif and the are also acknowledged.

REFERENCES

E. CANCÈS (2001) *SCF algorithms for Kohn-Sham models with fractional occupation numbers*. J. Chem. Phys. 114 10616–10623.

E. CANCÈS, R. CHAKIR AND Y. MADAY (2010) *Numerical analysis of nonlinear eigenvalue problems*. Journal of Scientific Computing, 45(1-3), 90–117.

E. CANCÈS, R. CHAKIR AND Y. MADAY (2012) *Numerical analysis of the planewave discretization of some orbital-free and Kohn-Sham models*. ESAIM: ESAIM: Mathematical Modelling and Numerical Analysis, 46(02), 341–388.

E. CANCÈS, M. DEFRANCESCHI, W. KUTZELNIGG, C. LE BRIS AND Y. MADAY (2003) *Computational quantum chemistry: a primer*. in: Handbook of numerical analysis. Volume X: special volume: computational chemistry, Ph. Ciarlet and C. Le Bris eds (North-Holland, 2003) 3–270.

E. CANCÈS AND C. LE BRIS (2000) *Can we outperform the DIIS approach for electronic structure calculations?*. Int. J. Quantum Chem. 79 82–90.

C. CANUTO, M.Y. HUSSAINI, A. QUARTERONI AND T.A. ZANG (2007) *Spectral methods*. Springer, 2007.

H. CHEN, X. GONG, L. HE AND A. ZHOU (2012) *Convergence of adaptive finite element approximations for nonlinear eigenvalue problems*. arXiv preprint, http://arxiv.org/pdf/1001.2344

H. CHEN, X. GONG, AND A. ZHOU (2010) *Numerical approximations of a nonlinear eigenvalue problem and applications to a density functional model*, ?? M2AS 33 1723–1742.

X. DAI, L. HE, AND A. ZHOU (2012). *Convergence rate and quasi-optimal complexity of adaptive finite element computations for multiple eigenvalues*, arXiv preprint arXiv:1210.1846.

X. DAI, J. XU, AND A. ZHOU (2008). *Convergence and optimal complexity of adaptive finite element eigenvalue computations*, Numerische Mathematik, 110(3), 313–355.

K. -N. KUDIN, G.-E. SCUSERIA AND E. CANCÈS (2002) *A black-box self-consistent field convergence algorithm: One step closer*. The Journal of chemical physics, 116, 8255.

B. LANGWALLNER, C. ORTNER AND E. SÜLI (2010) *Existence and convergence results for the Galerkin approximation of an electronic density functional*. M3AS 20, 2237-2265.

Y. MADAY AND G. TURINICI (2003) *Error bars and quadratically convergent methods for the numerical simulation of the Hartree-Fock equations*. Numer. Math. 94 739–770.

L. P. PITAEVSKII AND S. STRINGARI (2003) *Bose-Einstein condensation* Clarendon Press.

M. REED AND B. SIMON (1975) *Methods of Modern Mathematical Physics*. volume I, Academic Press.

A. ZHOU (2007) *Finite dimensional approximations for the electronic ground state solution of a molecular system*, Math. Meth. Appl. Sci. 30 429–447.

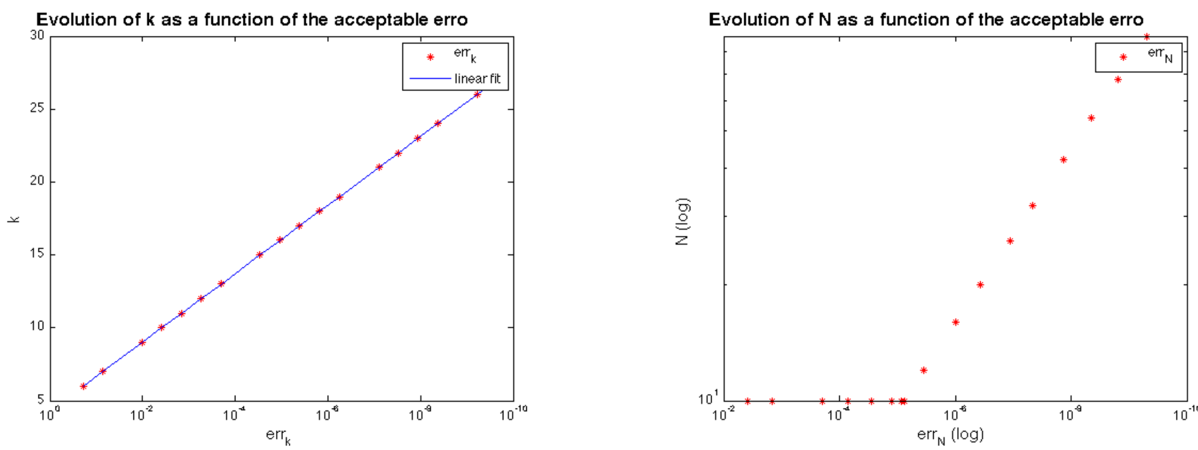

FIG. 1. Left: k evolution as err_k decreases (linear fit in log-scale for err_k). Right: N evolution as err_N decreases (linear fit in loglog-scale)