STABILITY OF THE APPROXIMATION OF A REGULAR SOLUTION BRANCH

Gabriel Caloz

Abstract. In this paper we extend the so called inf-sup conditions to non linear problems having regular solution branches. The inf-sup conditions for non linear problems are introduced for instance in Caloz and Rappaz [1994]. Here we present a more general approach to include turning points on a solution branch. Our abstract results are applied to model examples.

1. Introduction

Let X, Y be two reflexive Banach spaces and $F : X \to Y'$ be of class C^p, $p \geq 1$. Our goal is to construct a framework to study approximations of the problem: find $x \in X$ such that

$$F(x) = 0 \quad \text{or equivalently} \quad \forall y \in Y \quad \langle F(x), y\rangle_{Y'Y} = 0. \tag{1.1}$$

We shall focus on the following type of approximations. Let $\{X_h\}_{0<h\leq 1}$ be a family of finite dimensional subspaces of X and $\{Y_h\}_{0<h\leq 1}$ be a family of finite dimensional subspaces of Y. Then an approximation of problem (1.1) reads: find $x_h \in X_h$ such that

$$\forall y_h \in Y_h \quad \langle F(x_h), y_h\rangle_{Y'Y} = 0. \tag{1.2}$$

This approximation is called a Petrov-Galerkin approximation of (1.1).

Let $u_0 \in X$ be a solution to problem (1.1). If $u_0 \in X$ satisfies

$$\begin{cases} F(u_0) = 0, \\ DF(u_0) \in \mathscr{L}(X; Y') \quad \text{is an isomorphism,} \end{cases} \tag{1.3}$$

then we will call u_0 an isolated regular solution. If $u_0 \in X$ satisfies

$$\begin{cases} F(u_0) = 0, \\ DF(u_0) \in \mathscr{L}(X; Y') \quad \text{is surjective,} \\ \dim\ker DF(u_0) = m, \end{cases} \tag{1.4}$$

then u_0 is called a regular solution. Notice that the assumption (1.4) is equivalent to say that $DF(u_0)$ is a Fredholm operator of index m, surjective.

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

To study Petrov-Galerkin approximations (1.2) of problem (1.1) in a neighborhood of a solution u_0, we need to impose further assumptions on u_0 such as for instance u_0 is an isolated regular solution or u_0 is a regular solution.

There is a quite extensive literature on the approximation of nonlinear problems in Banach spaces. The most commonly used tools are the inverse and the implicit function theorems, see KELLER [1977], RHEINBOLDT [1986] or CROUZEIX and RAPPAZ [1989] and the references therein. There are also approaches based on monotonicity and maximum principle, see for instance CONRAD and CORTEY [1987], ŽENÍŠEK [1990], and BARRETT [1992].

In POUSIN and RAPPAZ [1992] and in CALOZ and RAPPAZ [1994], Petrov-Galerkin approximations of problems (1.1) have been considered in the case when u_0 is an isolated regular solution of (1.1). When u_0 is a regular solution, $X = \mathbb{R}^n \times Z$, Z being a Banach space, and $D_z F(u_0) \in \mathscr{L}(Z; Y')$ is an isomorphism, then u_0 belongs to a solution set which can be parametrized by $\lambda \in \mathbb{R}^n$. Such a study has been considered by different authors, see CALOZ and RAPPAZ [1994].

Our goal is to present a framework which is as simple as the one developed in CALOZ and RAPPAZ [1994] to study the approximations (1.2) under the more general assumption (1.4) on u_0. Only the case of $m = 1$ is studied in this paper.

The outline of the paper goes as follows. First we present well-known results useful later in our approach. Then in Section 3, we will develop an abstract framework, and finally we will apply it to examples in Section 4.

2. PRELIMINARIES

Let m be a positive integer, X and Z be two reflexive real Banach spaces. Given a C^p mapping $G : \mathbb{R}^m \times X \to Z$, $p \geq 1$, we consider the problem to find $(\lambda, x) \in \mathbb{R}^m \times X$ such that

$$G(\lambda, x) = 0. \tag{2.1}$$

We assume that $(\lambda_0, x_0) \in \mathbb{R}^m \times X$ satisfies

$$\begin{cases} G(\lambda_0, x_0) = 0, \\ D_x G(\lambda_0, x_0) \in \mathscr{L}(X; Z) \text{ is an isomorphism.} \end{cases} \tag{2.2}$$

So $(\lambda_0, x_0) \in \mathbb{R}^m \times X$ is a regular solution to (2.1). There exists a neighborhood $I \subset \mathbb{R}^m$ of λ_0, a neighborhood $U \subset X$ of x_0 and a C^p mapping $x : I \to U$ such that $\{(\lambda, x(\lambda)); \lambda \in I\}$ is a regular solution set of (2.1). Moreover if $\{(\lambda, x(\lambda)); \lambda \in I\}$ and $\{(\lambda, \tilde{x}(\lambda)); \lambda \in \tilde{I}\}$ are two regular solution sets containing (λ_0, x_0) then $x(\lambda)$ and $\tilde{x}(\lambda)$ coincide on the connected component of $I \cap \tilde{I}$ containing λ_0.

We introduce now a family $\{G_h\}_{0<h\leq 1}$ of C^p mappings

$$G_h : (\lambda, x) \in \mathbb{R}^m \times X \to G_h(\lambda, x) \in Z$$

which are approximations of G. Our goal is to study the existence and the convergence of solutions of the problem: find $(\lambda, x) \in \mathbb{R}^m \times X$ such that

$$G_h(\lambda, x) = 0, \tag{2.3}$$

in a neighborhood of the regular solution (λ_0, x_0).

To the solution (λ_0, x_0) and for $h \in (0,1]$, we associate an element $\tilde{x}_h \in X$. Notice that most often the choice $\tilde{x}_h = x_0$ will be adequate.

We assume there exist two positive constants η and L such that for all $(\lambda, v) \in \overline{B}\big((\lambda_0, \tilde{x}_h), \eta\big)$, for all $h \in (0,1]$

$$\|DG_h(\lambda_0, \tilde{x}_h) - DG_h(\lambda, v)\|_{\mathbb{R}^m \times X; Z} \leq L\big(|\lambda - \lambda_0| + \|v - \tilde{x}_h\|_X\big) \tag{2.4}$$

We assume the consistency

$$\lim_{h \to 0} \|G_h(\lambda_0, \tilde{x}_h)\|_Z = 0 \tag{2.5}$$

and the stability : there exists a constant C such that for all $h \in (0,1]$, $D_x G_h(\lambda_0, \tilde{x}_h) \in \mathscr{L}(X; Z)$ is an isomorphism and

$$\|D_x G_h(\lambda_0, \tilde{x}_h)^{-1}\|_{Z;X} \leq C. \tag{2.6}$$

Theorem 2.1. *In the framework introduced above, we assume that the assumptions (2.2), (2.4), (2.5), and (2.6) hold. Then there exist positive constants h_0, ϵ_0, δ_0, and for all $h \in (0, h_0]$ a C^p mapping $x_h : \lambda \in B(\lambda_0, \epsilon_0) \to x_h(\lambda) \in X$ such that*

$$\big(G_h(\lambda, x_h) = 0 \text{ and } x_h \in B(\tilde{x}_h, \delta_0)\big) \iff x_h = x_h(\lambda).$$

Moreover $\{(\lambda, x_h(\lambda)); \lambda \in B(\lambda_0, \epsilon_0)\}$ is a regular solution set and the following estimates hold

$$\|x_h(\lambda) - \tilde{x}_h\|_X \leq 2\gamma_{1h}\big(|\lambda - \lambda_0| + \|G_h(\lambda_0, \tilde{x}_h)\|_Z\big) \tag{2.7}$$

and

$$\|x_h(\lambda_0) - x(\lambda)\|_X \leq 2\gamma_{2h}\big(|\lambda - \lambda_0| + \|G(\lambda_0, x_h(\lambda_0))\|_Z\big) \tag{2.8}$$

with

$$\begin{aligned}
\gamma_{1h} &= \max\big(\|D_x G_h(\lambda_0, \tilde{x}_h)^{-1}\|_{Z;X}, 1 + \|D_\lambda G_h(\lambda_0, \tilde{x}_h)\|_{\mathbb{R}^m;Z}\|D_x G_h(\lambda_0, \tilde{x}_h)^{-1}\|_{Z;X}\big) \\
\gamma_{2h} &= \max\big(\|D_x G(\lambda_0, x_h(\lambda_0))^{-1}\|_{Z;X}, \\
&\qquad 1 + \|D_\lambda G(\lambda_0, x_h(\lambda_0))\|_{\mathbb{R}^m;Z}\|D_x G(\lambda_0, x_h(\lambda_0))^{-1}\|_{Z;X}\big) \quad \square
\end{aligned}$$

The technique used to prove Theorem 2.1 is based on a variant of the implicit function theorem, see for instance RHEINBOLDT [1986], CROUZEIX and RAPPAZ [1989] for an account. Here we also can apply Theorem 2.3 in CALOZ and RAPPAZ [1994].

In fact, working with $\lambda = \lambda_0$ fixed, and restricting h_0 if necessary, we can prove that

$$\|x_h(\lambda_0) - \tilde{x}_h\|_X \leq 2\|D_x G_h(\lambda_0, \tilde{x}_h)^{-1}\|_{Z;X}\|G_h(\lambda_0, \tilde{x}_h)\|_Z \tag{2.9}$$

and

$$\|x_h(\lambda_0) - x_0\|_X \leq 2\|D_x G(\lambda_0, x_h(\lambda_0))^{-1}\|_{Z;X}\|G(\lambda_0, x_h(\lambda_0))\|_Z. \tag{2.10}$$

Let us check (2.9) for instance. By the Taylor expansion we get

$$\begin{aligned}\tilde{x}_h - x_h(\lambda_0) = D_x G_h(\lambda_0, \tilde{x}_h)^{-1}\big[G_h(\lambda_0, \tilde{x}_h) + \int_0^1 \big\{D_x G_h(\lambda_0, \tilde{x}_h)\\ - D_x G_h(\lambda_0, x_h(\lambda_0) + t(\tilde{x}_h - x_h(\lambda_0)))\big\}(\tilde{x}_h - x_h(\lambda_0))\, dt\big].\end{aligned}$$

From the estimate (2.7) and the consistency (2.5), we deduce

$$\lim_{h \to 0} \|\tilde{x}_h - x_h(\lambda_0)\|_X = 0.$$

Then with the equicontinuity assumption (2.4), the stability (2.6), we deduce the existence of a $\tilde{h} \leq h_0$ such that for $h \in (0, \tilde{h}]$

$$\begin{aligned}\|D_x G_h(\lambda_0, \tilde{x}_h)^{-1}\|_{Z;X}\| \int_0^1 \big\{D_x G_h(\lambda_0, \tilde{x}_h)\\ - D_x G_h(\lambda_0, x_h(\lambda_0) + t(\tilde{x}_h - x_h(\lambda_0)))\big\} dt\|_{X;Z} \leq \frac{1}{2}.\end{aligned}$$

So

$$\|\tilde{x}_h - x_h(\lambda_0)\|_X \leq \frac{1}{2}\|\tilde{x}_h - x_h(\lambda_0)\|_X + \|D_x G_h(\lambda_0, \tilde{x}_h)^{-1}\|_{Z;X}\|G_h(\lambda_0, \tilde{x}_h)\|_Z,$$

from which we deduce the estimate (2.9).

The estimates (2.7) and (2.9) will lead to a priori error estimates, while (2.8) and (2.10) will give a posteriori estimates. A technique used to improve computational results consists in optimizing the residual $\|G(\lambda_0, x_h(\lambda_0))\|_Z$. It is simple to check that the error $\|x_h(\lambda_0) - x_0\|_X$ can be bounded from below by the residual. Indeed we have by the Taylor expansion

$$G(\lambda_0, x_h(\lambda_0)) = \int_0^1 D_x G(\lambda_0, x_0 + t(x_h(\lambda_0) - x_0))(x_h(\lambda_0) - x_0)\, dt$$

and so for h sufficiently small we get

$$\|G(\lambda_0, x_h(\lambda_0))\|_Z \leq 2\|D_x G(\lambda_0, x_0)\|_{X;Z}\|x_h(\lambda_0) - x_0\|_X. \tag{2.11}$$

3. Abstract framework

Let X, Y be two reflexive real Banach spaces and $F : X \to Y'$ be a C^p mapping, $p \geq 2$. Let $u_0 \in X$ be a solution to

$$F(x) = 0 \tag{3.1}$$

such that

$$\begin{cases} DF(u_0) \in \mathscr{L}(X; Y') \text{ is surjective,} \\ \dim \ker DF(u_0) = 1. \end{cases} \tag{3.2}$$

Notice that if $p = 1$, then DF needs to be Lipschitzian at u_0.

Under the assumption (3.2), we can prove that there exists a mapping $u : [-\epsilon, \epsilon] \to X$ of class C^p, a constant $\delta_0 > 0$ such that

$$\big(F(x) = 0 \text{ and } x \in B(u_0, \delta_0)\big) \Longleftrightarrow x = u(s) \text{ for some } s \in [-\epsilon, \epsilon].$$

Moreover $u(0) = u_0$. The proof of the result is based on the construction of a problem equivalent to (3.1). Since the method is useful to handle the approximated problem, we recall it here.

Under the assumption (3.2), there exists an element $\varphi \in X$, $\|\varphi\|_X = 1$, such that

$$\ker DF(u_0) = \operatorname{span}\{\varphi\}.$$

Let a functional $B \in X'$ be such that

$$B(\varphi) = 1.$$

Then we define the mapping $\Phi : \mathbb{R} \times X \to \mathbb{R} \times Y'$ by

$$\Phi(t, x) = \big(B(x - u_0) - t, F(x)\big),$$

and notice that

$$\begin{aligned} \Phi(0, u_0) &= 0, \\ D_x\Phi(0, u_0)v &= \big(B(v), DF(u_0)v\big). \end{aligned}$$

Clearly $D_x\Phi(0, u_0) \in \mathscr{L}(X; \mathbb{R} \times Y')$ is an isomorphism. The existence of $u : [-\epsilon, \epsilon] \to X$ solution to (3.1) is a consequence of the implicit function theorem applied to Φ.

Remark that in a neighborhood of $u_0 \in X$, the problem (3.1) and the following one

$$\Phi(t, x) \equiv \big(B(x - u_0) - t, F(x)\big) = 0 \tag{3.3}$$

are equivalent. Indeed the uniqueness of u means: there exists a neighborhood $\mathscr{V}$ of u_0 in X such that

$$\big(t \in [-\epsilon, \epsilon], u \in \mathscr{V}, \text{ and } \Phi(t,u) = 0\big) \Longleftrightarrow u = u(t).$$

Clearly if $\Phi(t,u) = 0$, then $F(u) = 0$. On the other hand, if u is such that $F(u) = 0$ and $(t \equiv B(u - u_0), u) \in [-\epsilon, \epsilon] \times \mathscr{V}$, then $\Phi(t,u) = 0$ and $u = u(t)$.

Since for $t \neq 0$, $B(u(t) - u_0) \neq 0$, the solution set $\{u(t); t \in [-\epsilon, \epsilon]\}$ to $F(x) = 0$ is not reduced to the isolated solution u_0. In fact $u(t_1) \neq u(t_2)$, for all t_1, t_2 in $[-\epsilon, \epsilon]$, $t_1 \neq t_2$.

Our goal is to develop a framework well-suited to study Petrov-Galerkin approximations of (3.1). Let $\{X_h\}_{0<h\leq 1}$ be a family of finite dimensional subspaces of X and $\{Y_h\}_{0<h\leq 1}$ be a family of finite dimensional subspaces of Y. We shall study the problem: find $x_h \in X_h$ such that

$$\text{for all } y_h \in Y_h \quad \langle F(x_h), y_h \rangle_{Y'Y} = 0. \tag{3.4}$$

To introduce the stability conditions, we define the bilinear form $b : X \times Y \to \mathbb{R}$ by

$$b(x,y) = \langle DF(u_0)x, y \rangle_{Y'Y}. \tag{3.5}$$

We assume the stability conditions

$$\inf_{\substack{y_h \in Y_h \\ \|y_h\|_Y = 1}} \sup_{\substack{x_h \in X_h \\ \|x_h\|_X = 1}} b(x_h, y_h) \geq \beta > 0, \tag{3.6}$$

$$\dim X_h = \dim Y_h + 1, \tag{3.7}$$

and the consistency

$$\lim_{h \to 0} \inf_{x_h \in X_h} \|u_0 - x_h\|_X = 0. \tag{3.8}$$

Let us study the problem (3.4) in a neighborhood of the solution u_0 satisfying the assumption (3.2).

An other way to write the assumptions (3.6), (3.7) is the following. For any $y_h \in Y_h$, $\|y_h\|_Y = 1$, we have

$$\sup_{\substack{x_h \in X_h \\ \|x_h\|_X = 1}} \langle DF(u_0)x_h, y_h \rangle_{Y'Y} \geq \beta > 0,$$

which implies that the mapping $DF(u_0) : X_h \to Y_h'$ is surjective. Moreover $\dim X_h = \dim Y_h + 1$. So there exists $\varphi_h \in X_h$, $\|\varphi_h\|_X = 1$ such that

$$\text{for all } y_h \in Y_h \quad \langle DF(u_0)\varphi_h, y_h \rangle_{Y'Y} = 0.$$

Now let $B_h \in X'$ be such that

$$B_h(\varphi) = 1, \qquad B_h(\varphi_h) = 1, \qquad \text{and } \|B_h\|_{X'} = 1.$$

We define then the mapping $\Phi_h : \mathbb{R} \times X \to \mathbb{R} \times Y'$ by

$$\Phi_h(t, x) = (B_h(x - u_0) - t, F(x)). \tag{3.9}$$

Remark that

$$\Phi_h(0, u_0) = 0.$$

Let us consider then the problem: find $t \in \mathbb{R}$, $x_h \in X_h$ such that

$$\begin{cases} B_h(x_h - u_0) - t = 0, \\ \text{for all } y_h \in Y_h \quad \langle F(x_h), y_h\rangle_{Y'Y} = 0. \end{cases} \tag{3.10}$$

Clearly if $(t, x_h) \in \mathbb{R} \times X_h$ is a solution of (3.10), then $x_h \in X_h$ is also a solution to (3.4).

To study the approximated problem (3.10), we shall apply the results stated in Section 2. First of all, we need to construct a family of mappings $\{G_h\}_{0<h\leq 1}$, with $G_h : \mathbb{R}\times X \to \mathbb{R} \times Y'$, such that problem (3.10) is equivalent to solving

$$G_h(t, x) = 0.$$

Then for that family $\{G_h\}_{0<h\leq 1}$ we need to check the equicontinuity, consistency, and stability conditions.

Let $\tilde{b}_h : X \times (\mathbb{R} \times Y) \to \mathbb{R}$ be the bilinear form defined by

$$\tilde{b}_h(x, (\delta, y)) = \langle D_x\Phi_h(0, u_0)x, (\delta, y)\rangle_{\mathbb{R}\times Y'\,\mathbb{R}\times Y}.$$

To construct the mappings G_h we need to define projectors onto X_h and Y_h. It will be done with the bilinear form $\tilde{b}_h$, which satisfies the well known inf-sup conditions.

Lemma 3.1. *The following relations hold*

$$\inf_{\substack{(\delta,y_h)\in\mathbb{R}\times Y_h \\ |\delta|+\|y_h\|_Y=1}} \sup_{\substack{x_h\in X_h \\ \|x_h\|_X=1}} \tilde{b}_h(x_h, (\delta, y_h)) \geq \alpha > 0, \tag{3.11}$$

$$\dim X_h = \dim(\mathbb{R} \times Y_h). \tag{3.12}$$

Proof. The relation (3.12) is an immediate consequence of (3.7). Let us now check (3.11). Let $(\delta, y_h) \in \mathbb{R} \times Y_h$ be such that

$$|\delta| + \|y_h\|_Y = 1.$$

From the assumption (3.6) and the choice of B_h, we deduce that there exists an element $\overline{x}_h \in X_h$ such that

$$\begin{cases} \langle DF(u_0)\overline{x}_h, y_h\rangle_{Y'Y} \geq \frac{\beta}{4}\|y_h\|_Y, \\ \|\overline{x}_h\|_X = 1, \\ B_h(\overline{x}_h) = 0. \end{cases}$$

We choose x_h of the form

$$x_h = \frac{\overline{x}_h + a\varphi_h}{\|\overline{x}_h + a\varphi_h\|_X},$$

with $a \in \mathbb{R}$ to be determined. We have

$$\frac{1}{\|\overline{x}_h + a\varphi_h\|_X} \geq \frac{1}{1+|a|}.$$

Then

$$\begin{aligned} \delta B_h(x_h) &+ \langle DF(u_0)x_h, y_h\rangle_{Y'Y} \\ &= \frac{1}{\|\overline{x}_h + a\varphi_h\|_X}\big[\delta a + \langle DF(u_0)\overline{x}_h, y_h\rangle_{Y'Y}\big] \\ &\geq \frac{\delta a + (1-|\delta|)\beta/4}{1+|a|}. \end{aligned}$$

We take

$$a = \frac{\operatorname{sgn}(\delta)}{2},$$

then

$$\begin{aligned} \delta B_h(x_h) + \langle DF(u_0)x_h, y_h\rangle_{Y'Y} &\geq \frac{2}{3}\min(1,\beta/4)\big[\frac{|\delta|}{2} + (1-|\delta|)\big] \\ &\geq \frac{1}{3}\min(1,\beta/4), \end{aligned}$$

and we can conclude. □

We define now the two projectors $\Pi_{X_h} \in \mathscr{L}(X; X_h)$ and $\Pi_{\mathbb{R}\times Y_h} \in \mathscr{L}(\mathbb{R}\times Y; \mathbb{R}\times Y_h)$ by

$$\text{for all } (\delta, y_h) \in \mathbb{R}\times Y_h \quad \tilde{b}_h(x - \Pi_{X_h}x, (\delta, y_h)) = 0, \tag{3.13}$$

$$\text{for all } x_h \in X_h \quad \tilde{b}_h(x_h, (\delta, y) - \Pi_{\mathbb{R}\times Y_h}(\delta, y)) = 0. \tag{3.14}$$

Then we construct a family $\{G_h\}_{0<h\leq 1}$ of mappings $G_h : \mathbb{R}\times X \to \mathbb{R}\times Y'$ by setting for all $(t,x) \in \mathbb{R}\times X$ and $(\delta, y) \in \mathbb{R}\times Y$

$$\begin{aligned} \langle G_h(t,x), (\delta,y)\rangle_{\mathbb{R}\times Y'\,\mathbb{R}\times Y} &= \\ &\langle \Phi_h(t,x), \Pi_{\mathbb{R}\times Y_h}(\delta,y)\rangle_{\mathbb{R}\times Y'\,\mathbb{R}\times Y} + \tilde{b}_h(x, (\delta,y) - \Pi_{\mathbb{R}\times Y_h}(\delta,y)). \end{aligned} \tag{3.15}$$

Then we will analyze the problem: find $(t,x) \in \mathbb{R}\times X$ such that

$$G_h(t,x) = 0. \tag{3.16}$$

Lemma 3.2. *Problems (3.10) and (3.16) are equivalent, that is if* $(t, x_h) \in \mathbb{R} \times X_h$ *satisfies (3.10) then* (t, x_h) *satisfies (3.16), conversely if* $(t, x) \in \mathbb{R} \times X$ *satisfies (3.16), then* $x \in X_h$ *and* (t, x) *satisfies (3.10).*

Proof. Let $(t, x_h) \in \mathbb{R} \times X_h$ satisfy (3.10). Since we have $x_h \in X_h$, then for all $(\delta, y) \in \mathbb{R} \times Y$

$$\tilde{b}_h(x_h, (\delta, y) - \Pi_{\mathbb{R}\times Y_h}(\delta, y)) = 0,$$

which implies that (t, x_h) is a solution of (3.16).

Let now $(t, x_h) \in \mathbb{R} \times X$ satisfy (3.16). Choosing the test function $(\delta_0, y_0) - \Pi_{\mathbb{R}\times Y_h}(\delta_0, y_0)$ with $(\delta_0, y_0) \in \mathbb{R} \times Y$, we get

$$\tilde{b}_h(x_h, (\delta_0, y_0) - \Pi_{\mathbb{R}\times Y_h}(\delta_0, y_0)) = 0$$

and

$$\tilde{b}_h(x_h - \Pi_{X_h} x_h, (\delta_0, y_0) - \Pi_{\mathbb{R}\times Y_h}(\delta_0, y_0)) = 0.$$

For all $(\delta_0, y_0) \in \mathbb{R} \times Y$, the following holds

$$\tilde{b}_h(x_h - \Pi_{X_h} x_h, (\delta_0, y_0)) = 0$$

that is

$$\langle D_x\Phi_h(0, u_0)(x_h - \Pi_{X_h} x_h), (\delta_0, y_0)\rangle_{\mathbb{R}\times Y'\,\mathbb{R}\times Y} = 0$$

or

$$\delta_0 B_h(x_h - \Pi_{X_h} x_h) + \langle DF(u_0)(x_h - \Pi_{X_h} x_h), y_0\rangle_{Y'Y} = 0.$$

If we take $y_0 = 0$ and $\delta_0 = 1$, then

$$B_h(x_h - \Pi_{X_h} x_h) = 0.$$

If we take $\delta_0 = 0$ and $y_0 \in Y$, then

$$\langle DF(u_0)(x_h - \Pi_{X_h} x_h), y_0\rangle_{Y'Y} = 0.$$

So we get

$$x_h - \Pi_{X_h} x_h = \alpha\varphi, \quad \text{for some } \alpha \in \mathbb{R}.$$

With our choice of B_h, we deduce that $\alpha = 0$. We can conclude that x_h is in X_h. □

To study the approximated problem (3.16), we apply the results stated in Section 2. The family $\{G_h\}_{0<h\le 1}$ of C^p mappings $G_h : (t, x) \in \mathbb{R} \times X \to G_h(t, x) \in \mathbb{R} \times Y'$ is defined in (3.15). We will study (3.16) in a neighborhood of $(0, u_0)$. Notice that $\Phi_h(0, u_0) = 0$. For simplicity we take $\tilde{x}_h = u_0$. To apply Theorem 2.1, we need to check the assumptions (2.4), (2.5), and (2.6).

Let us check (2.4), that is there exist two positive constants η and L such that for all $(t, v) \in \overline{B}((0, u_0), \eta)$, for all $h \in (0, 1]$

$$\|DG_h(0, u_0) - DG_h(t, v)\|_{\mathbb{R}\times X;\mathbb{R}\times Y'} \le L(|t| + \|v - u_0\|_X).$$

Going back to the definitions of G_h in (3.15) and of Φ_h in (3.9), we have for $(\tau, x) \in \mathbb{R}\times X$ and $(\delta, y) \in \mathbb{R} \times Y$

$$\begin{aligned}\langle (DG_h(0,u_0) - DG_h(t,v))(\tau,x),(\delta,y)\rangle_{\mathbb{R}\times Y'\,\mathbb{R}\times Y}\\ = \langle (D\Phi_h(0,u_0) - D\Phi_h(t,v))(\tau,x),\Pi_{\mathbb{R}\times Y_h}(\delta,y)\rangle_{\mathbb{R}\times Y'\,\mathbb{R}\times Y}.\end{aligned}$$

So

$$\|DG_h(0,u_0) - DG_h(t,v)\|_{\mathbb{R}\times X;\mathbb{R}\times Y'} \le \|DF(u_0) - DF(v)\|_{X;Y'}\|\Pi_{\mathbb{R}\times Y_h}\|_{\mathbb{R}\times Y;\mathbb{R}\times Y}.$$

Now we check that $\|\Pi_{\mathbb{R}\times Y_h}\|_{\mathbb{R}\times Y;\mathbb{R}\times Y}$ is uniformly bounded in h. We have for $(\delta, y) \in \mathbb{R} \times Y$ with $|\delta| + \|y\|_Y = 1$

$$\begin{aligned}\alpha\|\Pi_{\mathbb{R}\times Y_h}(\delta,y)\|_{\mathbb{R}\times Y} &\le \sup_{\substack{x_h\in X_h\\ \|x_h\|_X=1}} \tilde{b}_h(x_h,\Pi_{\mathbb{R}\times Y_h}(\delta,y))\\ &= \sup_{\substack{x_h\in X_h\\ \|x_h\|_X=1}} \tilde{b}_h(x_h,(\delta,y)) \le \|\tilde{b}_h\|.\end{aligned} \tag{3.17}$$

We have proved that $\|\Pi_{\mathbb{R}\times Y_h}\|_{\mathbb{R}\times Y;\mathbb{R}\times Y}$ is uniformly bounded in h. With the assumption $p \ge 2$, we can conclude.

Let us check now the stability assumption (2.6), that is $D_xG_h(0, u_0) \in \mathscr{L}(X; \mathbb{R} \times Y')$ is an isomorphism with a uniformly bounded inverse. First of all for $(\delta, y) \in \mathbb{R} \times Y$ and $x \in X$

$$\begin{aligned}\langle D_xG_h(0,u_0)x,(\delta,y)\rangle_{\mathbb{R}\times Y'\,\mathbb{R}\times Y} &= \langle D_x\Phi_h(0,u_0)x,\Pi_{\mathbb{R}\times Y_h}(\delta,y)\rangle_{\mathbb{R}\times Y'\,\mathbb{R}\times Y}\\ &\quad + \langle D_x\Phi_h(0,u_0)x,(\delta,y) - \Pi_{\mathbb{R}\times Y_h}(\delta,y)\rangle_{\mathbb{R}\times Y'\,\mathbb{R}\times Y}\\ &= \langle D_x\Phi_h(0,u_0)x,(\delta,y)\rangle_{\mathbb{R}\times Y'\,\mathbb{R}\times Y}.\end{aligned}$$

We will check that $D_x\Phi_h(0, u_0) \in \mathscr{L}(X; \mathbb{R} \times Y')$ is an isomorphism with a uniformly bounded inverse. Let $\mathrm{span}\{\varphi\} \oplus X_1 = X$ be a decomposition of X. Given any $\delta \in \mathbb{R}$, $y' \in Y'$, let $x = \alpha\varphi + x_1$ be such that

$$D_x\Phi_h(0,u_0)x = (\delta, y')$$

that is

$$\begin{cases} B_h(\alpha\varphi + x_1) = \delta,\\ DF(u_0)(\alpha\varphi + x_1) = y'.\end{cases}$$

Notice that $DF(u_0)_{|X_1} : X_1 \to Y'$ is an isomorphism. Then for x_1 we have the following bound not depending on h,

$$\|x_1\|_X \le C\|y'\|_{Y'}.$$

The number α is determined by

$$\alpha = \delta - B_h(x_1).$$

So

$$\begin{aligned}|\alpha| &\le |\delta| + |B_h(x_1)| \\ &\le |\delta| + C\|y'\|_{Y'}.\end{aligned}$$

Finally

$$|\alpha| + \|x_1\|_X \le \max(1, 2C)(|\delta| + \|y'\|_{Y'}),$$

which means that $D_x\Phi_h(0, u_0) \in \mathscr{L}(X; \mathbb{R} \times Y')$ has a uniformly bounded inverse.

Let us check now the consistency assumption (2.5). We have for $(\delta, y) \in \mathbb{R} \times Y$

$$\begin{aligned}\langle G_h(0, u_0), (\delta, y)\rangle_{\mathbb{R}\times Y'\,\mathbb{R}\times Y} &= \langle \Phi_h(0, u_0), \Pi_{\mathbb{R}\times Y_h}(\delta, y)\rangle_{\mathbb{R}\times Y'\,\mathbb{R}\times Y} \\ &\quad + \tilde{b}_h(u_0, (\delta, y) - \Pi_{\mathbb{R}\times Y_h}(\delta, y)) \\ &= \tilde{b}_h(u_0, (\delta, y) - \Pi_{\mathbb{R}\times Y_h}(\delta, y)) \\ &= \tilde{b}_h(u_0 - \Pi_{X_h} u_0, (\delta, y)).\end{aligned}$$

Then

$$\begin{aligned}\|G_h(0, u_0)\|_{\mathbb{R}\times Y'} &\le \|\tilde{b}_h\| \|u_0 - \Pi_{X_h} u_0\|_X \\ &\le C\|u_0 - \Pi_{X_h} u_0\|_X\end{aligned}$$

where C is a constant independent of h. For any $x_h \in X_h$ we have

$$\begin{aligned}\|x_h - \Pi_{X_h} u_0\|_X &= \sup_{\substack{\varphi\in X' \\ \|\varphi\|_{X'}=1}} \langle \varphi, x_h - \Pi_{X_h} u_0\rangle_{X'X} \\ &= \sup_{\substack{\varphi\in X' \\ \|\varphi\|_{X'}=1}} \langle x_h - \Pi_{X_h} u_0, [D_x\Phi_h(0, u_0)^\star](D_x\Phi_h(0, u_0)^\star)^{-1}\varphi\rangle_{XX'} \\ &= \sup_{\substack{\varphi\in X' \\ \|\varphi\|_{X'}=1}} \tilde{b}_h(x_h - \Pi_{X_h} u_0, (D_x\Phi_h(0, u_0)^\star)^{-1}\varphi) \\ &= \sup_{\substack{\varphi\in X' \\ \|\varphi\|_{X'}=1}} \tilde{b}_h(x_h - u_0, \Pi_{\mathbb{R}\times Y_h}(D_x\Phi_h(0, u_0)^\star)^{-1}\varphi) \\ &\le C\|x_h - u_0\|_X \|\Pi_{\mathbb{R}\times Y_h}\|_{\mathbb{R}\times Y;\mathbb{R}\times Y} \|[D_x\Phi_h(0, u_0)^\star]^{-1}\|_{X';\mathbb{R}\times Y},\end{aligned}$$

since $D_x\Phi_h(0, u_0) \in \mathscr{L}(X; \mathbb{R} \times Y')$ is an isomorphism with a uniformly bounded inverse, so is the adjoint operator $D_x\Phi_h(0, u_0)^\star$. Finally with the estimate (3.17), we deduce that for $x_h \in X_h$

$$\begin{aligned}\|u_0 - \Pi_{X_h} u_0\|_X &\le \|u_0 - x_h\|_X + \|x_h - \Pi_{X_h} u_0\|_X \\ &\le (1 + C)\|u_0 - x_h\|_X.\end{aligned}$$

With the assumption (3.8), we conclude immediately that (2.5) holds.

We can apply Theorem 2.1 and get the following result.

Theorem 3.3. *We assume that the hypotheses (3.2), (3.6), (3.7), and (3.8) hold. Then there exist positive constants h_0, ϵ_0, δ_0, and for all $h \in (0, h_0]$ a C^p mapping $u_h : t \in [-\epsilon_0, \epsilon_0] \to u_h(t) \in X_h \subset X$ such that*

$$\big(G_h(t, u_h) = 0 \quad \text{and} \quad u_h \in B(u_0, \delta_0)\big) \iff u_h = u_h(t).$$

Moreover $\{(t, u_h(t)); t \in [-\epsilon_0, \epsilon_0]\}$ is a regular solution set (in the sense introduced in (1.4) with $m = 1$) and the following estimates hold

$$\|u_h(t) - u_0\|_X \le C\big(|t| + \|G_h(0, u_0)\|_{\mathbb{R}\times Y'}\big), \tag{3.18}$$

$$\|u_h(0) - u(t)\|_X \le C\big(|t| + \|\Phi_h(0, u_h(0))\|_{\mathbb{R}\times Y'}\big), \tag{3.19}$$

where $u(.)$ is the solution branch of (3.1) in a neighborhood of u_0. □

With Lemma 3.2, we can rewrite Theorem 3.3 in terms of problem (3.4).

Corollary 3.4. *We assume that the hypotheses (3.2), (3.6), (3.7), and (3.8) hold. Then there exist positive constants h_0, ϵ_0, δ_0, and for all $h \in (0, h_0]$ a C^p mapping $u_h : t \in [-\epsilon_0, \epsilon_0] \to u_h(t) \in X_h \subset X$ such that*

$$\big(\text{for all } y_h \in Y_h \quad \langle F(u_h), y_h\rangle_{Y'Y} = 0 \quad \text{and} \quad u_h \in B(u_0, \delta_0) \cap X_h\big)$$
$$\iff \quad t = B_h(u_h - u_0), \quad u_h = u_h(t).$$

Moreover $\{u_h(t); t \in [-\epsilon_0, \epsilon_0]\}$ is a regular solution set and the following estimates hold

$$\|u_h(t) - u_0\|_X \le C\big(|t| + \inf_{x_h \in X_h} \|u_0 - x_h\|_X\big), \tag{3.20}$$

$$\|u_h(0) - u(t)\|_X \le C\big(|t| + \|F(u_h(0))\|_{Y'}\big). \quad □ \tag{3.21}$$

Remark 3.5. It is possible to get error estimates for the derivatives $\|\frac{d^j}{dt^j}u_h(0) - \frac{d^j}{dt^j}u(0)\|$. We do not emphasize this point here, see Crouzeix and Rappaz [1989] or Caloz and Rappaz [1994] for details. □

4. Applications

4.1 The case of the natural parameter space.

Let $\mathscr{X}$, Y be two reflexive Banach spaces and $F : \mathbb{R} \times \mathscr{X} \to Y'$ be a C^p mapping, $p \ge 2$. We assume that $(\lambda_0, \kappa_0) \in \mathbb{R} \times \mathscr{X}$ satisfies

$$F(\lambda_0, \kappa_0) = 0 \quad \text{and} \quad D_\kappa F(\lambda_0, \kappa_0) \in \mathscr{L}(\mathscr{X}; Y') \quad \text{is an isomorphism.} \tag{4.1}$$

Under the assumption (4.1), we deduce the existence of a regular solution branch $\{(\lambda, \kappa(\lambda)); \lambda \in I\}$ containing (λ_0, κ_0). Here $\mathbb{R}$ is the parameter space which could be used conveniently to parametrize the solution set. This problem and approximations of it have been widely studied in the literature, see RHEINBOLDT [1986] or CROUZEIX and RAPPAZ [1989] for an account.

We apply the theory of Section 3, with F, $X = \mathbb{R} \times \mathscr{X}$, Y, $u_0 = (\lambda_0, \kappa_0)$. The mapping $DF(\lambda_0, \kappa_0) \in \mathscr{L}(\mathbb{R} \times \mathscr{X}; Y')$ is surjective,

$$\ker(DF(u_0)) = \operatorname{span}\{\varphi\}$$

with $\varphi = \tilde{\varphi}/\|\tilde{\varphi}\|_{\mathbb{R}\times\mathscr{X}}$ and

$$\tilde{\varphi} = \begin{pmatrix} 1 \\ -D_\kappa F(\lambda_0, \kappa_0)^{-1} D_\lambda F(\lambda_0, \kappa_0) \end{pmatrix}.$$

The functional $B \in (\mathbb{R} \times \mathscr{X})'$ could be chosen to be

$$B(\lambda, \kappa) = \lambda;$$

in fact to follow what has been done before, we should take B with $B(\varphi) = 1$. Our previous choice is also convenient. Then the mapping $\phi : \mathbb{R} \times (\mathbb{R} \times \mathscr{X}) \to \mathbb{R} \times Y'$ is

$$\phi(t, \lambda, \kappa) = \big((\lambda - \lambda_0) - t, F(\lambda, \kappa)\big),$$

which means that the auxiliary parameter t is in fact λ itself.

Let $\{\mathscr{X}_h\}_{0<h\leq 1}$ be a family of finite dimensional subspaces of $\mathscr{X}$ and $\{Y_h\}_{0<h\leq 1}$ be a family of finite dimensional subspaces of Y. A Petrov-Galerkin approximation of

$$F(\lambda, \kappa) = 0 \tag{4.2}$$

reads: find $\lambda \in \mathbb{R}$ and $\kappa_h \in \mathscr{X}_h$ such that

$$\text{for all } y_h \in Y_h \quad \langle F(\lambda, \kappa_h), y_h \rangle_{Y'Y} = 0. \tag{4.3}$$

We assume that there exists $\alpha > 0$ such that for all $h \in (0, 1]$

$$\inf_{\substack{\kappa_h \in \mathscr{X}_h \\ \|\kappa_h\|_{\mathscr{X}}=1}} \sup_{\substack{y_h \in Y_h \\ \|y_h\|_Y=1}} \langle D_\kappa F(\lambda_0, \kappa_0)\kappa_h, y_h \rangle_{Y'Y} \geq \alpha > 0, \tag{4.4}$$

$$\dim \mathscr{X}_h = \dim Y_h, \tag{4.5}$$

and the consistency

$$\lim_{h\to 0} \inf_{\kappa_h \to \mathscr{X}_h} \|\kappa_0 - \kappa_h\|_{\mathscr{X}} = 0. \tag{4.6}$$

Problem (4.3) had been already studied in CALOZ and RAPPAZ [1994] under the assumptions (4.1), (4.4), (4.5), and (4.6). Here we check that our approach includes the previous result. Then we need to check that (3.6), (3.7), and (3.8) hold.

We define the bilinear form $b : (\mathbb{R} \times \mathscr{X}) \times Y \to \mathbb{R}$ by

$$b((\lambda, \kappa), y) = \langle DF(\lambda_0, \kappa_0)(\lambda, \kappa), y\rangle_{Y'Y}.$$

The approximation space X_h is $\mathbb{R} \times \mathscr{X}_h$. Clearly with (4.5) and (4.6), the assumptions (3.7) and (3.8) hold. We still need to check

$$\inf_{\substack{y_h \in Y_h \\ \|y_h\|_Y = 1}} \sup_{\substack{(\lambda, \kappa_h) \in \mathbb{R} \times \mathscr{X}_h \\ |\lambda| + \|\kappa_h\|_{\mathscr{X}} = 1}} b((\lambda, \kappa_h), y_h) \geq \beta > 0. \tag{4.7}$$

The assumptions (4.4) and (4.5) imply

$$\inf_{\substack{y_h \in Y_h \\ \|y_h\|_Y = 1}} \sup_{\substack{\kappa_h \in \mathscr{X}_h \\ \|\kappa_h\|_{\mathscr{X}} = 1}} \langle D_\kappa F(\lambda_0, \kappa_0)\kappa_h, y_h\rangle_{Y'Y} \geq \alpha > 0. \tag{4.8}$$

Let $y_h \in Y_h$ with $\|y_h\|_Y = 1$ be given. From (4.8), we know that there exists a $\overline{\kappa}_h \in \mathscr{X}_h$, $\|\overline{\kappa}_h\|_{\mathscr{X}} = 1$, with

$$\langle D_\kappa F(\lambda_0, \kappa_0)\overline{\kappa}_h, y_h\rangle_{Y'Y} \geq \alpha/2.$$

Then

$$\sup_{\substack{(\lambda, \kappa_h) \in \mathbb{R} \times \mathscr{X}_h \\ |\lambda| + \|\kappa_h\|_{\mathscr{X}} = 1}} b((\lambda, \kappa_h), y_h) \geq b((0, \overline{\kappa}_h), y_h) \geq \alpha/2$$

and (4.7) is checked.

Theorem 4.1. *We assume that the assumptions (4.1), (4.4), (4.5), and (4.6) hold. Then there exist positive constants h_0, ϵ_0, δ_0, and for all $h \in (0, h_0]$ a C^p mapping $\kappa_h : \lambda \in [\lambda_0 - \epsilon_0, \lambda_0 + \epsilon_0] \to \kappa_h(\lambda) \in \mathscr{X}_h \subset \mathscr{X}$ such that*

$$\Big(\text{for all } y_h \in Y_h \quad \langle F(\lambda, \kappa_h), y_h\rangle_{Y'Y} = 0 \text{ and } \kappa_h \in B(\kappa_0, \delta_0) \cap \mathscr{X}_h, \\ |\lambda - \lambda_0| \leq \epsilon_0\Big) \iff \kappa_h = \kappa_h(\lambda).$$

Moreover $\{(\lambda, \kappa_h(\lambda)); \lambda \in [\lambda_0 - \epsilon_0, \lambda_0 + \epsilon_0]\}$ is a regular solution set and

$$\|\kappa_h(\lambda) - \kappa_0\|_{\mathscr{X}} \leq C\{|\lambda - \lambda_0| + \inf_{\kappa_h \in \mathscr{X}_h} \|\kappa_0 - \kappa_h\|_{\mathscr{X}}\}, \tag{4.9}$$

$$\|\kappa_h(\lambda_0) - \kappa(\lambda)\|_{\mathscr{X}} \leq C\{|\lambda - \lambda_0| + \|F(\lambda_0, \kappa_h(\lambda_0))\|_{Y'}\}. \quad \square \tag{4.10}$$

Remark 4.1. We present Theorem 4.1 with λ as parameter. When applying Corollary 3.4, we should introduce the functional B_h. In fact in our particular case, it is simple to check that B_h can be replaced by B defined above and the result still holds. $\square$

4.2 The case of a simple limit point.

Let the functions $\alpha : \mathbb{R}_+ \to \mathbb{R}$, $\beta : \mathbb{R}_+ \to \mathbb{R}$, and $\theta : \mathbb{R} \to \mathbb{R}$ satisfy the following conditions

$$\begin{aligned} &\alpha \in L^\infty(\mathbb{R}_+) \cap L^1(\mathbb{R}_+), \\ &\beta \in L^\infty(\mathbb{R}_+) \cap L^2(\mathbb{R}_+), \\ &\theta \in C^p(\mathbb{R}),\ p \geq 3. \end{aligned} \tag{4.11}$$

We consider the problem to find $(\lambda, u) \in \mathbb{R} \times H_0^1(\mathbb{R}_+)$ such that

$$-u'' + \alpha u + \beta\theta(u)u^2 + \lambda u = 0 \quad \text{in} \quad \mathbb{R}_+. \tag{4.12}$$

Note that a finite element approximation of (4.12) has been studied in Descloux and Rappaz [1982]. The point in this example is the lack of compactness of the embeddings of the Sobolev spaces.

If $(\lambda, u) \in \mathbb{R} \times H_0^1(\mathbb{R}_+)$ is a solution to (4.12), with $u \not\equiv 0$, then

$$\lambda > 0. \tag{4.13}$$

Indeed (λ, u) satisfies

$$-u'' + pu + \lambda u = 0 \tag{4.14}$$

where

$$p = \alpha + \beta\theta(u)u \in L^1(\mathbb{R}_+) \cap L^\infty(\mathbb{R}_+).$$

The differential operator

$$Lu = -u'' + pu$$

has been widely studied; its essential spectrum $\sigma_e(L)$ is $[0, \infty)$, see Dunford and Schwartz [1958], Chapter XIII, Akhiezer and Glazman [1981], Naimark [1968], or Schechter [1971] for more details.

Let (λ_0, u_0) be a solution of (4.12) with $u_0 \not\equiv 0$ and $\lambda_0 > 0$. Our goal is to analyze the approximation of solutions of (4.12) in a neighborhood of (λ_0, u_0) making use of the abstract framework in Section 3.

The corresponding notations are the following: $X = \mathbb{R} \times H_0^1(\mathbb{R}_+)$, $Y = H_0^1(\mathbb{R}_+)$, and $F : \mathbb{R} \times H_0^1(\mathbb{R}_+) \to H^{-1}(\mathbb{R}_+)$ given for $(\lambda, u) \in \mathbb{R} \times H_0^1(\mathbb{R}_+)$ by: for all $w \in H_0^1(\mathbb{R}_+)$

$$\begin{aligned} \langle F(\lambda, u), w\rangle_{H^{-1}(\mathbb{R}_+) H_0^1(\mathbb{R}_+)} &= \int_0^\infty u'(t)w'(t)\,dt + \int_0^\infty \alpha(t)u(t)w(t)\,dt \\ &+ \int_0^\infty \beta(t)u^2(t)\theta(u(t))w(t)\,dt + \lambda \int_0^\infty u(t)w(t)\,dt. \end{aligned} \tag{4.15}$$

The mapping F is of class C^p and its partial derivatives are at the point $(\lambda, u) \in \mathbb{R} \times H_0^1(\mathbb{R}_+)$: for all $v, w \in H_0^1(\mathbb{R}_+)$

$$\begin{aligned} \langle D_u F(\lambda, u)v, w\rangle_{H^{-1}(\mathbb{R}_+)H_0^1(\mathbb{R}_+)} &= \int_0^\infty v'(t)w'(t)\,dt + \int_0^\infty [\alpha(t) \\ +\lambda]v(t)w(t)\,dt &+ \int_0^\infty \beta(t)[2\theta(u(t))u(t) + \theta'(u(t))u^2(t)]v(t)w(t)\,dt, \end{aligned} \tag{4.16}$$

for all $w \in H_0^1(\mathbb{R}_+)$

$$\langle D_\lambda F(\lambda, u), w\rangle_{H^{-1}(\mathbb{R}_+)H_0^1(\mathbb{R}_+)} = \int_0^\infty u(t)w(t)\,dt. \tag{4.17}$$

The mapping $D_u F(\lambda_0, u_0) \in \mathscr{L}(H_0^1(\mathbb{R}_+); H^{-1}(\mathbb{R}_+))$ is a Fredholm operator of index 0 and $\dim\ker(D_u F(\lambda_0, u_0))$ is either 0 or 1, see Dunford and Schwartz [1958].

The assumption (3.2) for the solution (λ_0, u_0) is satisfied in the two cases:

$$\dim\ker(D_u F(\lambda_0, u_0)) = 0 \tag{4.18}$$

or

$$\dim\ker(D_u F(\lambda_0, u_0)) = 1 \quad \text{and} \quad D_\lambda F(\lambda_0, u_0) \notin \operatorname{range}(D_u F(\lambda_0, u_0)). \tag{4.19}$$

In the following, we assume that (λ_0, u_0) is a solution of (4.12) with $\lambda_0 > 0$, and satisfying to either (4.18) and (4.19). There exists a mapping $(\lambda, u) : [-\epsilon, \epsilon] \to \mathbb{R} \times H_0^1(\mathbb{R}_+)$, of class C^p, and a constant $\delta_0 > 0$ such that

$$\Big(F(\lambda, u) = 0 \text{ and } (\lambda, u) \in B((\lambda_0, u_0), \delta_0)\Big) \iff (\lambda, u) = (\lambda(\tau), u(\tau)) \text{ for some } \tau \in [-\epsilon, \epsilon];$$

moreover

$$\lambda(0) = \lambda_0, \quad u(0) = u_0.$$

We construct finite dimensional subspaces of $H_0^1(\mathbb{R}_+)$. Let $\{a_n\}_{n=1}^\infty \subset \mathbb{R}_+$ be the sequence of $a_n = n^\gamma$, $0 < \gamma < 1$. We set for $n \in \mathbb{N}$

$$h_n = \frac{a_n}{n}, t_i = ih_n \quad \text{for } i = 0, \ldots, n.$$

For notation simplicity we write simply h instead of h_n. The finite dimensional subspace V_h is then defined by

$$V_h = \{f : \mathbb{R}_+ \to \mathbb{R} \text{ continuous}; f(0) = 0, f(t) = 0 \text{ for } t \geq t_n, f|_{]t_i, t_{i+1}[} \in \mathscr{P}_1\}, \tag{4.20}$$

here $\mathscr{P}_1$ is the space of polynomials of degree 1. The Galerkin approximation to (4.12) reads: find $(\lambda_h, u_h) \in \mathbb{R} \times V_h$ such that

$$\text{for all } w_h \in V_h \quad \langle F(\lambda_h, u_h), w_h \rangle_{H^{-1}(\mathbb{R}_+) H_0^1(\mathbb{R}_+)} = 0. \tag{4.21}$$

To apply the theory of Section 3, we need to check the assumptions (3.6), (3.7), and (3.8), with $X_h = \mathbb{R} \times V_h$, $Y_h = V_h$. The assumption (3.7) is immediate and the (3.8) one is a consequence of standard interpolation results. Finally we need to check (3.6), that is

$$\inf_{\substack{w_h \in V_h \\ \|w_h\|_{1,\mathbb{R}_+}=1}} \sup_{\substack{(\delta, v_h) \in \mathbb{R} \times V_h \\ |\delta| + \|v_h\|_{1,\mathbb{R}_+}=1}} \langle DF(\lambda_0, u_0)(\delta, v_h), w_h \rangle_{H^{-1}(\mathbb{R}_+) H_0^1(\mathbb{R}_+)} \geq \beta > 0. \tag{4.22}$$

The partial derivative $D_u F(\lambda_0, u_0) \in \mathscr{L}(H_0^1(\mathbb{R}_+); H^{-1}(\mathbb{R}_+))$ is: for all $v, w \in H_0^1(\mathbb{R}_+)$

$$\langle D_u F(\lambda_0, u_0) v, w \rangle_{H^{-1}(\mathbb{R}_+) H_0^1(\mathbb{R}_+)} = \int_0^\infty v'(t) w'(t)\, dt + \int_0^\infty [\alpha(t) + \lambda_0 + K(t)] v(t) w(t)\, dt$$

with

$$K(t) = \beta(t)[2\theta(u_0(t)) u_0(t) + \theta'(u_0(t)) u_0^2(t)] \in L^1(\mathbb{R}_+) \cap L^\infty(\mathbb{R}_+).$$

Therefore we can choose $\eta \geq \lambda_0 > 0$ such that the bilinear form, for $v, w \in H_0^1(\mathbb{R}_+)$,

$$a(v, w) = \int_0^\infty v'(t) w'(t)\, dt + \int_0^\infty [\alpha(t) + \eta + K(t)] v(t) w(t)\, dt$$

is $H_0^1(\mathbb{R}_+)$-elliptic, that is there exists $\kappa > 0$ such that

$$\text{for all } v \in H_0^1(\mathbb{R}_+) \quad a(v, v) \geq \kappa \|v\|_{1,\mathbb{R}_+}^2.$$

Corresponding to $a(.,.)$, we can define the operator $T \in \mathscr{L}(H^{-1}(\mathbb{R}_+); H_0^1(\mathbb{R}_+))$ by: for $f \in H^{-1}(\mathbb{R}_+)$, $Tf \in H_0^1(\mathbb{R}_+)$ satisfies

$$\text{for all } v \in H_0^1(\mathbb{R}_+) \quad a(Tf, v) = \langle f, v \rangle_{H^{-1}(\mathbb{R}_+) H_0^1(\mathbb{R}_+)};$$

let $T_h \in \mathscr{L}(H^{-1}(\mathbb{R}_+); V_h)$ be the discrete analogue of T. Then we easily check that

$$T D_u F(\lambda_0, u_0) = I + (\lambda_0 - \eta) T$$

and

$$T_h D_u F(\lambda_0, u_0) = I + (\lambda_0 - \eta) T_h.$$

The operator $T D_u F(\lambda_0, u_0)$ is a Fredholm operator with index 0 and the dimension of $\ker(T D_u F(\lambda_0, u_0))$ is either 0 or 1. We have assumed in (4.18) and (4.19) that the operator $T DF(\lambda_0, u_0) \in \mathscr{L}(\mathbb{R} \times H_0^1(\mathbb{R}_+); H_0^1(\mathbb{R}_+))$ is surjective and has a kernel of dimension 1. To get a similar property for $T_h DF(\lambda_0, u_0)$, the difficulty lies in the fact that we cannot have T_h to tend to T, as operators from $L^2(\mathbb{R}_+)$ to $H_0^1(\mathbb{R}_+)$, in the operator norm since T is not compact. To prove (4.22), we need a careful analysis of the eigenvalue problem for the operator $T_h D_u F(\lambda_0, u_0)$.

The operator $T \in \mathscr{L}(H_0^1(\mathbb{R}_+); H_0^1(\mathbb{R}_+))$ is self-adjoint with respect to the scalar product $a(.,.)$. The spectrum of T, $\sigma(T)$, is in $[0, \infty)$. The same is true for $T_h \in \mathscr{L}(V_h; V_h)$, $\sigma(T_h) \subset [0, \infty)$.

From the analysis of Galerkin approximations of eigenvalue problem, see DESCLOUX [1981], we know that

(1) if $\mu \in (1/\eta, \infty)$, $\mu \notin \sigma(T)$, then there exist $h_0 > 0$ and $C > 0$ such that for $0 < h \leq h_0$ and $v_h \in V_h$

$$\|(T_h - \mu I)v_h\|_{1,\mathbb{R}_+} \geq C\|v_h\|_{1,\mathbb{R}_+};$$

(2) if $\mu \in (1/\eta, \infty) \cap \sigma(T)$, then there exist $h_0 > 0$, $\epsilon > 0$ such that for $0 < h \leq h_0$, $(\mu - \epsilon, \mu + \epsilon) \cap \sigma(T_h)$ contains only one eigenvalue μ_h of T_h; moreover μ_h is of multiplicity 1 and $\lim_{h\to 0} \mu_h = \mu$. If $\varphi \not\equiv 0$ denotes an eigenvector corresponding to μ, then there exists an eigenvector φ_h corresponding to μ_h and $\lim_{h\to 0} \|\varphi - \varphi_h\|_{1,\mathbb{R}_+} = 0$.

Since λ_0 is positive, $-1/(\lambda_0 - \eta)$ is in the interval $(1/\eta, \infty)$ and we can use the results (1) or (2). If we are in the case (1), we have no difficulty. If we are in the case (2), we use the selfadjointness of $T_h D_u F(\lambda_0, u_0) \in \mathscr{L}(V_h; V_h)$ with respect to $a(.,.)$, to show that the restriction $R_h \in \mathscr{L}(\{\varphi_h\}^\perp; \{\varphi_h\}^\perp)$ of $T_h D_u F(\lambda_0, u_0)$ to the orthogonal complement $\{\varphi_h\}^\perp$ has no eigenvalue in $(-\epsilon(\eta - \lambda_0), \epsilon(\eta - \lambda_0))$ for all $h \leq h_0$. Consequently R_h is an isomorphism with an inverse bounded by $1/\epsilon(\eta - \lambda_0)$, for all $0 < h \leq h_0$.

With this result, the assumption $TDF(\lambda_0, u_0) \in \mathscr{L}(\mathbb{R} \times H_0^1(\mathbb{R}_+); H_0^1(\mathbb{R}_+))$ surjective, and $\lim_{h\to 0} \|\varphi - \varphi_h\|_{1,\mathbb{R}_+} = 0$, we easily check (4.22).

Theorem 4.2. *Let $(\lambda_0, u_0) \in \mathbb{R} \times H_0^1(\mathbb{R}_+)$ be a solution of (4.12) satisfying to (4.18) or (4.19). Let V_h be given in (4.20). Then there exist positive constants h_0, ϵ_0, δ_0, and for all $0 < h \leq h_0$ a C^p mapping $(\lambda_h, u_h) : [-\epsilon_0, \epsilon_0] \to \mathbb{R} \times V_h$, such that*

$$\Big((\lambda_h, u_h) \text{ is a solution of (4.21) and } (\lambda_h, u_h) \in B((\lambda_0, u_0), \delta_0) \cap \mathbb{R} \times V_h\Big) \iff (\lambda_h, u_h) = (\lambda_h(\tau), u_h(\tau)) \text{ for some } \tau \in [-\epsilon_0, \epsilon_0].$$

Moreover there exists a constant C such that

$$|\lambda_h(\tau) - \lambda_0| + \|u_h(\tau) - u_0\|_{1,\mathbb{R}_+} \leq C(|\tau| + h). \quad \square \tag{4.23}$$

The estimate (4.23) is optimal for the norm of $u_h(\tau) - u_0$ and is immediately deduced from the (3.20) one with standard interpolation results. The estimate for the eigenvalue can be improved as mentioned below.

To derive a posteriori estimates, we can use the relation (3.19) to get

$$|\lambda_h(0) - \lambda_0| + \|u_h(0) - u_0\|_{1,\mathbb{R}_+} \le C\|F(\lambda_h(0), u_h(0))\|_{-1,\mathbb{R}_+}. \tag{4.24}$$

From standard techniques, see Crouzeix and Rappaz [1989], we can prove that if (λ_0, u_0) is a non degenerate simple limit point, that is

$$\lambda(0) = \lambda_0, \quad \frac{d}{d\tau}\lambda(0) = 0, \quad \frac{d^2}{d\tau^2}\lambda(0) \neq 0,$$

then there exists a unique $\tau_h \in [-\epsilon_0, \epsilon_0]$ such that $(\lambda_h(\tau_h), u_h(\tau_h))$ is a non degenerate simple limit point. The estimate for the difference $\lambda_0 - \lambda_h(\tau_h)$ can also be improved to

$$|\lambda_0 - \lambda_h(\tau_h)| \le Ch^2.$$

References

Akhiezer, N.I. and Glazman, I.M., *Theory of Linear Operators in Hilbert Spaces*, Vol. II, Pitman Advanced Publishing Program, Boston, 1981.

Barrett, J.W., *Finite element approximation of a non-lipschitz nonlinear eigenvalue problem*, M^2AN **26** (1992), 627–656.

Caloz, G. and Rappaz, J., *Numerical Analysis for Nonlinear and Bifurcation Problems*, Tech. Rpt. (1994), I.R.M.A.R., Université de Rennes I; to appear in the Handbook of Numerical Analysis.

Conrad, F. and Cortey-Dumont, P., *Nonlinear eigenvalue problems in elliptic variational inequalities: some results for the maximal branch*, Numer. Funct. Anal. Optim. **9 and 10** (1987), 1059–1090 and 1091–1114.

Crouzeix, M. and Rappaz, J., *On numerical approximation in bifurcation theory*, RMA 13, Masson, Paris, 1989.

Descloux, J., *Essential numerical range of an operator with respect to a coercive form and the approximation of its spectrum by the Galerkin method*, SIAM J. Numer. Anal. **18** (1981), 1128–1133.

Descloux, J. and Rappaz, J., *Numerical approximation of a nonlinear Sturm-Liouville problem on an infinite interval*, Proceedings of Equadiff. **V** (1982), 81–83, Teubner-Texte zur Mathematik.

Dunford, N. and Schwartz, J.T., *Linear operators*, Interscience, New York, 1958.

Keller, H.B., *Numerical solution of bifurcation and nonlinear eigenvalue problems*, Application of bifurcation theory (1977), 359–384 (P.H. Rabinowitz, eds.), Academic Press.

Naimark, M.A., *Linear differential operators*, Part 2, Harrap, London, 1968.

Pousin, J. and Rappaz, J., *Consistency, stability, a priori and a posteriori errors for Petrov-Galerkin methods applied to nonlinear problems*, Tech. Rpt. 04-92 (1992), D.M.A., École Polytechnique Fédérale de Lausanne; to appear in Numer. Math..

Rheinboldt, W.C., Numerical analysis of parametrized nonlinear equations, Wiley-Interscience, John Wiley and Sons, 1986.

Schechter, M., *Spectra of partial differential operators*, North-Holland Publishing Company, Amsterdam, 1971.

Ženíšek, A., *Nonlinear elliptic and evolution problems and their finite element approximations*, Computational mathematics and applications, Academic Press Limited, London, 1990.

IRMAR, Campus de Beaulieu, Université de Rennes I, 35042 Rennes Cedex

E-mail: caloz@lie.univ-rennes1.fr