Computer-Assisted Proofs
for Semilinear Elliptic Boundary Value Problems

Michael PLUM

Institut für Analysis, Universität Karlsruhe (TH)
Englerstraße 2, 76128 Karlsruhe, Germany
E-mail: michael.plum@math.uni-karlsruhe.de
URL: http://www.mathematik.uni-karlsruhe.de/mi1plum/~plum/

For second-order semilinear elliptic boundary value problems on bounded or unbounded domains, a general computer-assisted method for proving the existence of a solution in a "close" and explicit neighborhood of an approximate solution, computed by numerical means, is proposed. To achieve such an existence and enclosure result, we apply Banach's fixed-point theorem to an equivalent problem for the error, i.e., the difference between exact and approximate solution. The verification of the conditions posed for the fixed-point argument requires various analytical and numerical techniques, for example the computation of eigenvalue bounds for the linearization at the approximate solution. The method is used to prove existence and multiplicity results for some specific examples.

Key words: elliptic boundary value problem, semilinear, computer-assisted proof, existence, enclosures, error bounds, multiplicity

1. Introduction

Semilinear elliptic differential equations of the form

$$-\Delta u(x) + f(x, u(x)) = 0 \quad (x \in \Omega) \tag{1}$$

(with $\Omega \subset \mathbb{R}^n$ denoting some given domain, and $f \colon \Omega \times \mathbb{R} \to \mathbb{R}$ some given non-linearity), together with boundary conditions, e.g., of Dirichlet type

$$u(x) = 0 \quad (x \in \partial\Omega), \tag{2}$$

have been (and still are) extensively studied in the differential equations literature. Such *semilinear boundary value problems* have a lot of applications, e.g., in Mathematical Physics, and often serve as model problems for more complex mathematical situations, and last but not least, they form a very exciting and challenging object for purely mathematical investigations. Starting perhaps with Picard's successive iterations at the end of the 19th century, various analytical methods and techniques have been (and are being) developed to study *existence* and *multiplicity* of solutions to problem (1), (2), such as variational methods (including mountain pass methods), index and degree theory, monotonicity methods, fixed-point methods, and more.

In this article, we want to report on a supplement to these purely analytical methods by a *computer-assisted* approach, which in the recent years has turned out to be successful with various examples where purely analytical methods have failed.

Such an approach has the remarkable additional advantage (compared with a "theoretical" proof) of providing accurate *quantitative* information in form of tight and explicit *bounds* for the solution.

We include problems with an *unbounded* domain Ω, which have rarely been treated by computer-assisted methods before. For the easier case of a bounded domain, see e.g., [5, 26, 27, 28, 32], and also [22, 23, 24, 25]. Our approach has been applied successfully also to other types of problems; see, e.g., the remarks at the end of Section 2.

We start with an *approximate* solution ω to (1), (2), which can be obtained by any numerical method which gives approximations in the function space needed (to be specified later). In this first step, there is no need for any mathematical rigor, and the field is open for the whole rich variety of modern numerics.

Next, we use a Newton–Kantorovich-type argument to prove the existence of a solution to (1), (2) in some "close" and "explicit" neighborhood of ω. For this purpose, we consider the boundary value problem for the *error* $v = u - \omega$ and rewrite it as a fixed-point equation

$$v \in X, \quad v = T(v) \tag{3}$$

in a Banach space X, which we treat by Banach's fixed-point theorem. This gives the existence of a solution v^* of (3) in some suitable set $V \subset X$, provided that a contraction property holds on V (see (15) below), and that

$$T(V) \subset V. \tag{4}$$

Consequently, $u^* := \omega + v^*$ is a solution of (1), (2) (which gives the desired existence result), and the statement "$u^* \in \omega + V$" (implied by $v^* \in V$) gives the desired bounds, or enclosures, for u^*.

So the crucial condition to be verified, for some suitable set V, is (4). Restricting ourselves to *norm balls* V (centered at the origin), we find that (4) results in an inequality involving the radius of V, and various other terms generated by the "data" of our problem (1), and by the numerical approximation ω. All these terms are *computable*, either directly or via additional computer-assisted means (like the eigenvalue bounds discussed briefly in Section 3.3). In these computations (in contrast to the computation of ω mentioned above), all possible numerical errors have to be taken into account, in order to be able to check the aforementioned inequality (implying (4)) with mathematical rigor. For example, remainder term bounds need to be computed when quadrature formulas are applied, and interval arithmetic [17, 34] is needed to take rounding errors into account.

Computer-assisted means for obtaining enclosures for solutions to elliptic partial differential equations have been proposed by Collatz [8, 9] already more than 50 years ago. He used maximum-principle-type arguments to obtain two-sided bounds for the error function $u - \omega$, with ω denoting a numerical C^2-approximation. Schröder [35]–[37], Walter [38] and others generalized these ideas, which resulted

in the method of *differential inequalities.* It was successfully applied to many examples with first or second order ordinary differential equations, or with second order elliptic or parabolic differential equations. However, there are drawbacks of differential inequalities methods concerning the size of the class of problems (1), (2) to which they can be applied: At least for obtaining "tight" solution enclosures, all eigenvalues of the linearization L of (1), (2) at ω need to be *positive*, which excludes many interesting situations. Furthermore, differential inequalities techniques are essentially restricted to first- and second-order problems (with the exception of some fourth-order problems which can be handled as second-order systems). In contrast, the enclosure method proposed in this article requires the eigenvalues of the linearization L to be *non-zero* only (which is checked by *eigenvalue enclosures*), and at least in principle it can be used for elliptic problems of any (even) order; see also the remarks at the end of Section 2.

An existence and enclosure method similar to ours has been developed by Nakao and his group [22]–[25]. This approach avoids the computation of eigenvalue enclosures for L, which constitutes a significant advantage in some cases. Instead, a finite-dimensional *projection* of L is used, and treated by well-established means of verifying numerical linear algebra. However, also the (infinite-dimensional) *projection error* needs to be bounded in a suitable way, which is well possible for many bounded domains Ω, but problematic e.g., for PDE problems on unbounded domains.

Another more recent approach, which has at least the *potential* of being applicable to elliptic boundary value problems (although this has not been carried out so far), is based on the Conley index and the numerical verification of corresponding topological conditions; it is suited for proving the existence of stationary solutions for certain classes of problems, as well as for detecting global dynamics (see e.g., [10, 15]).

For *ordinary* differential equation problems (possibly originating from a partial differential equation after symmetry reductions), many existence and enclosure methods can be found in the literature, which we will not address in this article.

2.   Abstract formulation

It turns out to be useful to explain the basics of our computer-assisted approach first for the following abstract problem:

$$\text{Find } u \in X \text{ satisfying } \mathcal{F}(u) = 0, \tag{5}$$

with $(X, \langle \, \cdot \, , \, \cdot \, \rangle_X)$ and $(Y, \langle \, \cdot \, , \, \cdot \, \rangle_Y)$ denoting two real Hilbert spaces (also complex spaces can be admitted), and $\mathcal{F} \colon X \to Y$ some Fréchet differentiable mapping.

Let $\omega \in X$ denote some *approximate* solution to (5) (computed, e.g., by numerical means), and

$$L := \mathcal{F}'(\omega) \colon X \to Y \tag{6}$$

the Fréchet derivative of $\mathcal{F}$ at ω, i.e., $L \in \mathcal{B}(X, Y)$ (the Banach space of all bounded linear operators from X to Y), and $\|\mathcal{F}(\omega + h) - \mathcal{F}(\omega) - L[h]\|_Y = o(\|h\|_X)$ as $\|h\|_X \to 0$.

Suppose that we know constants δ and K, and a non-decreasing function $g \colon [0, \infty) \to [0, \infty)$ such that

$$\|\mathcal{F}(\omega)\|_Y \leq \delta, \tag{7}$$

i.e., δ bounds the *defect* (residual) of the approximate solution ω to (5),

$$\|u\|_X \leq K\|L[u]\|_Y \quad \text{for all } u \in X, \tag{8}$$

i.e., K bounds the *inverse* of the linearization L,

$$\|\mathcal{F}'(\omega + u) - \mathcal{F}'(\omega)\|_{\mathcal{B}(X,Y)} \leq g(\|u\|_X) \quad \text{for all } u \in X, \tag{9}$$

i.e., g majorizes the modulus of continuity of $\mathcal{F}'$ at ω, and

$$g(t) \to 0 \quad \text{as } t \to 0 \tag{10}$$

(which in particular requires $\mathcal{F}'$ to be continuous at ω).

The concrete computation of such δ, K, and g is the main challenge in our approach, with particular emphasis on K. We will however not address these questions in this section, i.e., on the abstract level, but postpone them to the more specific case of the boundary value problem (1), (2), to be treated in the following section. For now, we assume that (7)–(10) hold true.

In order to obtain a suitable fixed-point formulation (3) for our problem (5), we will need that the operator L is *onto*. (Note that L is one-to-one by (8).) For this purpose, we will make an additional *symmetry* assumption (see (13) below), and now assume that $Y = X'$, the (topological) *dual* of X, i.e., the space of all bounded linear functionals $l \colon X \to \mathbb{R}$. X' ($= \mathcal{B}(X, \mathbb{R})$) is a Banach space endowed with the usual operator sup-norm. Indeed, this norm is generated by an inner product (which therefore makes X' a Hilbert space) as explained in the following: Consider the linear mapping $\Phi \colon X \to X'$ given by

$$(\Phi[u])[v] := \langle u, v \rangle_X \quad (u, v \in X). \tag{11}$$

For all $u \in X$,

$$\|\Phi[u]\|_{X'} = \sup_{v \in X \setminus \{0\}} \frac{|(\Phi[u])[v]|}{\|v\|_X} = \sup_{v \in X \setminus \{0\}} \frac{|\langle u, v \rangle_X|}{\|v\|_X} = \|u\|_X,$$

i.e., Φ is an *isometry* (and hence one-to-one).

Furthermore, Φ is *onto* by Riesz' representation theorem for bounded linear functionals on a Hilbert space: Given any $r \in X'$, some (unique) $u \in X$ exists such that $r[v] = \langle u, v \rangle_X$ for all $v \in X$, i.e., $\Phi[u] = r$ by (11). Φ is therefore called the

canonical isometric isomorphism between X and X'. It immediately gives an inner product on X' by

$$\langle r, s \rangle_{X'} := \langle \Phi^{-1}[r], \Phi^{-1}[s] \rangle_X \quad (r, s \in X'), \tag{12}$$

and the norm generated by this inner product is the "old" norm $\| \cdot \|_{X'}$, because Φ is isometric.

In theoretical functional analysis, the Hilbert spaces X and X' are often identified via the isometric isomorphism Φ, i.e., they are not distinguished, which however we will *not* do because this might lead to confusion when X is a Sobolev function space, as it will be later.

To ensure that $L \colon X \to Y = X'$ is onto, we make the additional assumption that $\Phi^{-1}L \colon X \to X$ is *symmetric* with respect to $\langle \cdot, \cdot \rangle_X$, which by (11) amounts to the relation

$$(L[u])[v] = (L[v])[u] \quad \text{for all } u, v \in X. \tag{13}$$

This implies the denseness of the range $(\Phi^{-1}L)(X) \subset X$: Given any u in its orthogonal complement, we have, for all $v \in X$,

$$0 = \langle u, (\Phi^{-1}L)[v] \rangle_X = \langle (\Phi^{-1}L)[u], v \rangle_X,$$

and hence $(\Phi^{-1}L)[u] = 0$, which implies $L[u] = 0$ and thus $u = 0$ by (8).

Therefore, since Φ is isometric, the range $L(X) \subset X'$ is *dense*. For proving that L is onto, we are therefore left to show that $L(X) \subset X'$ is *closed*. For this purpose, let $(L[u_n])_{n \in \mathbb{N}}$ denote some sequence in $L(X)$ converging to some $r \in X'$. Then (8) shows that $(u_n)_{n \in \mathbb{N}}$ is a Cauchy sequence in X. With $u \in X$ denoting its limit, the boundedness of L implies $L[u_n] \to L[u]$ $(n \to \infty)$. Thus, $r = L[u] \in L(X)$, which proves closedness of $L(X)$.

We are now able to formulate and prove our main theorem, which is similar to the Newton–Kantorovich theorem:

THEOREM 1. *Let δ, K, g satisfy conditions* (7)–(10). *Suppose that some $\alpha > 0$ exists such that*

$$\delta \leq \frac{\alpha}{K} - G(\alpha), \tag{14}$$

where $G(t) := \int_0^t g(s)\, ds$, and

$$Kg(\alpha) < 1. \tag{15}$$

Then, there exists a solution $u \in X$ of the equation $\mathcal{F}(u) = 0$ satisfying

$$\|u - \omega\|_X \leq \alpha. \tag{16}$$

The solution is moreover unique under the side condition (16).

REMARK 1.

a) Due to (10), $G(t) = \int_0^t g(s)\,ds$ is *superlinearly* small as $t \to 0$. Therefore, the crucial condition (14) is indeed satisfied for some "small" α if K is "moderate" (i.e., not too large) and δ is sufficiently small, which means according to (7) that the approximate solution ω to problem (5) must be computed with *sufficient accuracy*, and (14) tells us *how* accurate the computation has to be. This meets the general philosophy of computer-assisted proofs: The "hard work" of the proof is left to the computer!

b) Due to (10), the additional contraction condition (15) is not too critical if α (computed according to (14)) is "small."

Proof of Theorem 1. We rewrite problem (5) as

$$L[u - \omega] = -\mathcal{F}(\omega) - \{\mathcal{F}(u) - \mathcal{F}(\omega) - L[u - \omega]\},$$

which due to the bijectivity of L amounts to the equivalent fixed-point equation

$$v \in X, \quad v = -L^{-1}[\mathcal{F}(\omega) + \{\mathcal{F}(\omega + v) - \mathcal{F}(\omega) - L[v]\}] =: T(v) \tag{17}$$

for the error $v = u - \omega$.

Now we are going to show the following properties of the fixed-point operator $T\colon X \to X$:

i) $T(V) \subset V$ for the closed norm ball

$$V := \{v \in X : \|v\|_X \leq \alpha\},$$

ii) T is contractive on V.

Then, Banach's fixed-point theorem gives a solution $v^* \in V$ of the fixed-point equation (17), whence by construction $u^* := \omega + v^*$ is a solution of $\mathcal{F}(u) = 0$ satisfying (16).

For proving i) and ii), we first note that for every differentiable function $f\colon [0,1] \to Y$, the real-valued function $\|f\|_Y$ is differentiable almost everywhere on $[0,1]$, and $(d/dt)\|f\|_Y \leq \|f'\|_Y$ a.e. on $[0,1]$. Hence, for every $v, \tilde{v} \in X$,

$$\|\mathcal{F}(\omega + v) - \mathcal{F}(\omega + \tilde{v}) - L[v - \tilde{v}]\|_Y$$
$$= \int_0^1 \frac{d}{dt}\|\mathcal{F}(\omega + (1-t)\tilde{v} + tv) - \mathcal{F}(\omega + \tilde{v}) - tL[v - \tilde{v}]\|_Y\,dt$$
$$\leq \int_0^1 \|\{\mathcal{F}'(\omega + (1-t)\tilde{v} + tv) - L\}[v - \tilde{v}]\|_Y\,dt$$
$$\leq \int_0^1 \|\mathcal{F}'(\omega + (1-t)\tilde{v} + tv) - L\|_{\mathcal{B}(X,Y)}\,dt \cdot \|v - \tilde{v}\|_X$$
$$\leq \int_0^1 g(\|(1-t)\tilde{v} + tv\|_X)\,dt \cdot \|v - \tilde{v}\|_X, \tag{18}$$

using (6) and (9) in the last step. Choosing $\tilde{v} = 0$ in (18) we obtain, for each $v \in X$,

$$\|\mathcal{F}(\omega + v) - \mathcal{F}(\omega) - L[v]\|_Y \leq \int_0^1 g(t\|v\|_X)\,dt \cdot \|v\|_X$$

$$= \int_0^{\|v\|_X} g(s)\,ds = G(\|v\|_X). \qquad (19)$$

Furthermore, (18) and the fact that g is non-decreasing imply, for all $v, \tilde{v} \in V$,

$$\|\mathcal{F}(\omega + v) - \mathcal{F}(\omega + \tilde{v}) - L[v - \tilde{v}]\|_Y \leq \int_0^1 g((1-t)\|\tilde{v}\|_X + t\|v\|_X)\,dt \cdot \|v - \tilde{v}\|_X$$

$$\leq g(\alpha)\|v - \tilde{v}\|_X. \qquad (20)$$

To prove i), let $v \in V$, i.e., $\|v\|_X \leq \alpha$. Now (17), (8), (7), (19), and (14) imply

$$\|T(v)\|_X \leq K\|\mathcal{F}(\omega) + \{\mathcal{F}(\omega + v) - \mathcal{F}(\omega) - L[v]\}\|_Y$$

$$\leq K(\delta + G(\|v\|_X)) \leq K(\delta + G(\alpha)) \leq \alpha,$$

which gives $T(v) \in V$. Thus, $T(V) \subset V$.

For proving ii), we note that (17), (8), and (20) imply, for $v, \tilde{v} \in V$,

$$\|T(v) - T(\tilde{v})\|_X = \|L^{-1}\{\mathcal{F}(\omega + v) - \mathcal{F}(\omega + \tilde{v}) - L[v - \tilde{v}]\}\|_X$$

$$\leq K\|\mathcal{F}(\omega + v) - \mathcal{F}(\omega + \tilde{v}) - L[v - \tilde{v}]\|_Y \leq Kg(\alpha)\|v - \tilde{v}\|_X,$$

whence (15) shows that T is contractive on V. This completes the proof of Theorem 1. $\square$

In the following section, we will apply the abstract approach developed in this section to the elliptic boundary value problem (1), (2). This can be done in different ways, i.e., by different choices of the Hilbert space X, resulting in different general assumptions (e.g., smoothness conditions) to be made for the "data" of the problem and the numerical approximation ω, and different conditions (7)–(9), (14), (15), as well as different "results," i.e., existence statements and error bounds (16).

In this article, we will concentrate on the choice $X = H_0^1(\Omega)$, i.e., we will aim at *weak* solutions of problem (1), (2). Our abstract setting can also be applied for obtaining *strong* solutions by choosing $X = H^2(\Omega) \cap H_0^1(\Omega)$ and $Y = L^2(\Omega)$ if Ω is H^2-regular (see e.g., [5, 26, 27, 28, 32]), and also to problems with Neumann boundary conditions by choosing $X = H^1(\Omega)$, and to *systems* of elliptic equations. Now we report briefly on some other applications of our abstract setting which we cannot discuss in more detail in this article.

For *parameter-dependent problems* (where $\mathcal{F}$ in (5), or f in (1), depends on an additional parameter λ), one is often interested in *branches* $(u_\lambda)_{\lambda \in I}$ of solutions. By additional perturbation techniques, our method can indeed be generalized to computer-assisted proofs for such solution branches, as long as the parameter-interval I defining the branch is compact [28]. Such branches may however contain

turning points (where a branch "returns" at some value λ^*) or *bifurcation points* (where several—usually two—branches cross each other). Near such points, the operator L defined in (6) is "almost" singular, i.e., (8) holds only with a very large K, or not all all, which makes our approach break down. However, there are means to overcome these problems:

In case of (simple) turning points, the well-known method of augmenting the given equation by a *bordering* equation can also be used here; the "new" operator $\mathcal{F}$ in (5) contains the "old" one *and* the bordering functional, and the "new" operator L is regular near the turning point if the bordering equation has been chosen appropriately [27].

In case of (simple) symmetry-breaking bifurcations, we can, in a first step, include the symmetry in the space X, which excludes the symmetry-breaking branch and regularizes the problem, whence an existence and enclosure result for the symmetric branch can be obtained. In a second step, we exclude the symmetric branch by some transformation (similar to the Lyapunov–Schmidt reduction), and defining a corresponding new operator $\mathcal{F}$ we can perform our method to obtain an existence and enclosure result also for the symmetry-breaking branch [29].

Non-selfadjoint eigenvalue problems have been treated in [19, 25], again using bordering equation techniques normalizing the unknown eigenfunction. So $\mathcal{F}$ now acts on pairs (u, λ), and is defined via the eigenvalue equation and the (scalar) normalizing equation. In this way, we were able to give the first known instability proof of the Orr–Sommerfeld equation with Blasius profile [19] and with Poiseuille profile [25], which is a fourth-order ODE eigenvalue problem on $[0, \infty)$ (for the Blasius profile) or on $[-1, 1]$ (for the Poiseuille profile), respectively.

Also (other) *higher order* problems are covered by our abstract setting. In [6], we could prove the existence of 36 travelling wave solutions of a fourth-order nonlinear beam equation on the real line. Biharmonic problems (with $\Delta\Delta u$ as leading term) are presently investigated by Fazekas; see also [11].

3. Weak solutions

Based on our abstract considerations, we will now investigate problem (1), (2). Suppose that $\Omega \subset \mathbb{R}^n$ is a (bounded or unbounded) domain with Lipschitz boundary. For uniform presentation, we assume that $n \geq 2$; the case $n = 1$ is actually much simpler. We choose the spaces

$$X := H_0^1(\Omega), \quad Y := H^{-1}(\Omega) \tag{21}$$

for our abstract setting, where $H^{-1}(\Omega) := (H_0^1(\Omega))'$ denotes the topological dual space of $H_0^1(\Omega)$, i.e., the space of all bounded linear functionals on $H_0^1(\Omega)$. We endow $H_0^1(\Omega)$ with the inner product

$$\langle u, v \rangle_{H_0^1} := \langle \nabla u, \nabla v \rangle_{L^2} + \sigma \langle u, v \rangle_{L^2} \tag{22}$$

(with some parameter $\sigma > 0$ to be chosen later), and $H^{-1}(\Omega)$ with the "dual" inner product given by (12), with Φ from (11).

For $\rho \in L^2(\Omega)^n$, we define the functional div $\rho \colon H_0^1(\Omega) \to \mathbb{R}$ as usual by

$$(\operatorname{div} \rho)[\varphi] := - \int_\Omega \rho \cdot \nabla \varphi \, dx \quad \text{for all } \varphi \in H_0^1(\Omega), \tag{23}$$

implying in particular that $|(\operatorname{div} \rho)[\varphi]| \leq \|\rho\|_{L^2} \|\nabla \varphi\|_{L^2} \leq \|\rho\|_{L^2} \|\varphi\|_{H_0^1}$, whence div ρ is indeed a *bounded* linear functional, i.e., div $\rho \in H^{-1}(\Omega)$, and

$$\|\operatorname{div} \rho\|_{H^{-1}} \leq \|\rho\|_{L^2}. \tag{24}$$

Using this definition of Δu ($= \operatorname{div}(\nabla u)$), it is easy to check that the canonical isometric isomorphism $\Phi \colon H_0^1(\Omega) \to H^{-1}(\Omega)$ defined in (11) is now given by (note (22))

$$\Phi[u] = -\Delta u + \sigma u \quad (u \in H_0^1(\Omega)). \tag{25}$$

We recall that Sobolev's embedding theorem [1, Theorem 5.4] gives $H_0^1(\Omega) \subset L^p(\Omega)$, with bounded embedding $H_0^1(\Omega) \hookrightarrow L^p(\Omega)$, i.e., there exists some constant $C_p > 0$ such that

$$\|u\|_{L^p} \leq C_p \|u\|_{H_0^1} \quad \text{for all } u \in H_0^1(\Omega), \tag{26}$$

for each

$$p \in [2, \infty) \text{ if } n = 2, \quad p \in \left[2, \frac{2n}{n-2}\right] \text{ if } n \geq 3. \tag{27}$$

The following lemma (see e.g., the appendix of [31] for a proof) provides explicit constants C_p satisfying (26); they are not always optimal but easy to compute.

LEMMA 1.   *Let $\rho^* \in [0, \infty)$ denote the minimal point of the spectrum of $-\Delta$ on $H_0^1(\Omega)$.*
a)   *Let $n = 2$ and $p \in [2, \infty)$. With ν denoting the largest integer $\leq p/2$, (26) holds for*

$$C_p = \left(\frac{1}{2}\right)^{\frac{1}{2} + \frac{2\nu - 3}{p}} \left[\frac{p}{2}\left(\frac{p}{2} - 1\right) \cdots \left(\frac{p}{2} - \nu + 2\right)\right]^{\frac{2}{p}} \frac{1}{\left(\rho^* + \frac{p}{2}\sigma\right)^{\frac{1}{p}}}$$

(where the bracket-term is put equal to 1 if $\nu = 1$).
b)   *Let $n \geq 3$ and $p \in \left[2, \frac{2n}{n-2}\right]$. With $s := n\left(\frac{1}{p} - \frac{1}{2} + \frac{1}{n}\right) \in [0, 1]$, (26) holds for*

$$C_p = \left(\frac{n-1}{\sqrt{n}(n-2)}\right)^{1-s} \left(\frac{s}{s\rho^* + \sigma}\right)^{\frac{s}{2}}$$

(where the second factor is put equal to 1 if $s = 0$).

With p in the range (27), and p' denoting its dual number (i.e., $p^{-1} + (p')^{-1} = 1$), we obtain by Hölder's inequality, combined with the above embedding, that for all $w \in L^{p'}(\Omega)$

$$\left|\int_\Omega w\varphi \, dx\right| \leq \|w\|_{L^{p'}} \|\varphi\|_{L^p} \leq C_p \|w\|_{L^{p'}} \|\varphi\|_{H_0^1},$$

implying $w \in H^{-1}(\Omega)$ (with the usual identification of functions with their canonically associated linear functionals), and $\|w\|_{H^{-1}} \leq C_p \|w\|_{L^{p'}}$. Consequently, the embedding $L^{p'}(\Omega) \hookrightarrow H^{-1}(\Omega)$ is bounded, with the same embedding constant C_p as in the "dual" embedding $H_0^1(\Omega) \hookrightarrow L^p(\Omega)$. Note that the range (27) for p amounts to the range

$$p' \in (1, 2] \text{ if } n = 2, \quad p' \in \left[\frac{2n}{n+2}, 2\right] \text{ if } n \geq 3 \tag{28}$$

for the dual number p'.

Our abstract setting requires that $f(\,\cdot\,, u) \in H^{-1}(\Omega)$ for each $u \in H_0^1(\Omega)$, and furthermore that

$$\mathcal{F}\colon \begin{cases} H_0^1(\Omega) \to H^{-1}(\Omega), \\ u \quad \mapsto -\Delta u + f(\,\cdot\,, u) \end{cases} \tag{29}$$

is Fréchet-*differentiable*. Since $\Delta\colon H_0^1(\Omega) \to H^{-1}(\Omega)$ is linear and bounded by (24), this amounts to the Fréchet-differentiability of

$$\mathcal{G}\colon \begin{cases} H_0^1(\Omega) \to H^{-1}(\Omega), \\ u \quad \mapsto f(\,\cdot\,, u). \end{cases} \tag{30}$$

For this purpose, we require that $\partial f/\partial y$ is continuous on $\bar{\Omega} \times \mathbb{R}$. But this is not sufficient; the main reason is that $H_0^1(\Omega)$ does not embed into $C(\bar{\Omega})$. We need additional *growth restrictions* on $f(x, y)$ or $(\partial f/\partial y)(x, y)$ as $|y| \to \infty$.

An important (but not the only) admissible class consists of those functions f which satisfy

$$f(\,\cdot\,, 0) \in H^{-1}(\Omega), \tag{31}$$

$$\frac{\partial f}{\partial y}(\,\cdot\,, 0) \text{ is a bounded function on } \Omega, \tag{32}$$

$$\left|\frac{\partial f}{\partial y}(x, y) - \frac{\partial f}{\partial y}(x, 0)\right| \leq c_1 |y|^{r_1} + c_2 |y|^{r_2} \quad (x \in \Omega, \ y \in \mathbb{R}), \tag{33}$$

with non-negative constants c_1, c_2, and with

$$0 < r_1 \leq r_2 < \infty \text{ if } n = 2, \quad 0 < r_1 \leq r_2 \leq \frac{4}{n-2} \text{ if } n \geq 3. \tag{34}$$

(A "small" r_1 will make condition (33) weak near $y = 0$, and a "large" r_2 will make it weak for $|y| \to \infty$.)

LEMMA 2. *Let f satisfy (31)–(33), besides the continuity of $\partial f/\partial y$. Then $\mathcal{G}$ given by (30) is well-defined and Fréchet-differentiable, with derivative $\mathcal{G}'(u) \in \mathcal{B}(H_0^1(\Omega), H^{-1}(\Omega))$ (for $u \in H_0^1(\Omega)$) given by*

$$(\mathcal{G}'(u)[v])[\varphi] = \int_\Omega \frac{\partial f}{\partial y}(\,\cdot\,, u) v\varphi \, dx \quad (v, \varphi \in H_0^1(\Omega)). \tag{35}$$

The proof of Lemma 2 is rather technical, and therefore omitted here.
According to (23) and (35), we have

$$
\begin{aligned}
(\mathcal{F}'(u)[\varphi])[\psi] &= \int_{\Omega}\left[\nabla\varphi\cdot\nabla\psi + \frac{\partial f}{\partial y}(\,\cdot\,,u)\varphi\psi\right]dx \\
&= (\mathcal{F}'(u)[\psi])[\varphi] \quad (u,\varphi,\psi\in H_0^1(\Omega))
\end{aligned}
\tag{36}
$$

for the operator $\mathcal{F}$ defined in (29), which in particular implies condition (13) (for
any $\omega\in H_0^1(\Omega)$; note (6)), in the setting (21), (29).

REMARK 2. If the domain Ω is *bounded*, several simplifications and exten-
sions are possible:

a)   The range $\sigma > 0$ for the parameter in (22) can be extended to $\sigma \geq 0$.

b)   Condition (33) can be simplified to

$$
\left|\frac{\partial f}{\partial y}(x,y)\right| \leq \tilde{c}_1 + \tilde{c}_2|y|^r \quad (x\in\Omega,\ y\in\mathbb{R})
\tag{37}
$$

for some r in the range (34). Condition (32) is satisfied automatically and
can therefore be omitted.

c)   In the case $n = 2$, the power-growth condition (33) (or (37)) is too restrictive
(for bounded domains). Instead, *exponential* growth can be allowed, based on
the Trudinger–Moser inequality [21, Theorem 1 and the first part of its proof]
which states that

$$
\frac{1}{\operatorname{meas}(\Omega)}\int_{\Omega}\exp\left[\left(\frac{u(x)}{c\|u\|_{H_0^1}}\right)^2\right]dx \leq 1 + \frac{1}{4\pi c^2 - 1} \quad (u\in H_0^1(\Omega))
\tag{38}
$$

for each $c > (4\pi)^{-\frac{1}{2}}$. In [32], we showed that, e.g., in the case $f(x,y) = -\lambda e^y$,
the Fréchet differentiability (and other properties) of the mapping $\mathcal{G}$ defined in
(30) can easily be derived from (38); see also the second example in Section 4.

We comment now on the computation of an approximate solution ω, and of
the terms δ, K, and g satisfying (7)–(10), needed for the application of Theorem 1,
in the setting (21), (29).

3.1. Computation of ω

By (21), ω needs to be in $X = H_0^1(\Omega)$. In the finite element context, this
allows, for example, the "usual" linear (or quadratic) triangular elements. We are
furthermore allowed to use approximations ω of the form

$$
\omega = \begin{cases} \omega_0 & \text{on } \Omega_0, \\ 0 & \text{on } \Omega\setminus\Omega_0, \end{cases}
\tag{39}
$$

with $\Omega_0 \subset \Omega$ denoting some bounded subdomain (the "computational" domain),
and $\omega_0 \in H_0^1(\Omega_0)$ some approximate solution of the differential equation (1) on Ω_0,

subject to Dirichlet boundary conditions on $\partial\Omega_0$. If Ω is bounded, one will usually choose $\Omega_0 := \Omega$.

For computing ω_0, one may use a Newton iteration, together with e.g., a Ritz–Galerkin or a collocation method, and some linear algebraic system solver, which possibly incorporates multigrid methods. To *start* the Newton iteration, a rough initial approximation is needed, which can e.g., be obtained by path-following methods, or by use of the numerical mountain pass algorithm proposed in [7].

An important remark is that, no matter how ω is put up or which numerical method is used, there is *no* need for any *rigorous* (i.e., error free) computation at this stage, i.e., the whole variety of numerical methods is at hand.

We pose the additional condition of ω being *bounded*, which on one hand is satisfied anyway for all practical numerical schemes, and on the other hand turns out to be very useful in the following.

3.2.   Defect bound δ

By (21) and (29), condition (7) for the defect bound δ now amounts to

$$\|-\Delta\omega + f(\,\cdot\,,\omega)\|_{H^{-1}} \le \delta. \tag{40}$$

The best general way for obtaining δ seems to be the following. First we compute an additional approximation $\rho \in H(\mathrm{div}, \Omega)$ to $\nabla\omega$. (Here, $H(\mathrm{div}, \Omega)$ denotes the space of all vector-valued functions $\tau \in L^2(\Omega)^n$ with weak derivative $\mathrm{div}\,\tau$ in $L^2(\Omega)$. Hence, obviously $H(\mathrm{div}, \Omega) \supset H^1(\Omega)^n$.) ρ can be computed e.g., by interpolation (or some more general projection) of $\nabla\omega$ in $H(\mathrm{div}, \Omega)$, or in $H^1(\Omega)^n$. It should be noted that ρ comes "for free" as a part of the approximation, if *mixed* finite elements are used to compute ω.

Furthermore, according to the arguments after Lemma 1, applied with $p = p' = 2$,

$$\|w\|_{H^{-1}} \le C_2\|w\|_{L^2} \quad \text{for all } w \in L^2(\Omega). \tag{41}$$

Due to Lemma 1, $C_2 = (\rho^* + \sigma)^{-\frac{1}{2}}$, with $\rho^* \in [0,\infty)$ denoting the smallest point of the spectrum of $-\Delta$ on $H_0^1(\Omega)$, is a valid embedding constant. By (24) and (41),

$$\|-\Delta\omega + f(\,\cdot\,,\omega)\|_{H^{-1}} \le \|\mathrm{div}(-\nabla\omega + \rho)\|_{H^{-1}} + \|-\mathrm{div}\,\rho + f(\,\cdot\,,\omega)\|_{H^{-1}}$$
$$\le \|\nabla\omega - \rho\|_{L^2} + C_2\|-\mathrm{div}\,\rho + f(\,\cdot\,,\omega)\|_{L^2}, \tag{42}$$

which reduces the computation of a defect bound δ (satisfying (40)) to computing bounds for two *integrals*. To accomplish this task, we now assume that ω is of the form (39), and that also ρ is supported in the bounded subdomain Ω_0.

In some cases the integrals can be calculated in closed form, by hand or by computer algebra routines, for example if f is polynomial and ω, ρ are piecewise polynomial (as they are if finite element methods have been used to compute them), or if $f(x,\,\cdot\,)$ is polynomial and $f(\,\cdot\,,y)$, ω, and ρ are trigonometric polynomial. The resulting formulas have to be evaluated *rigorously*, to obtain true upper bounds for

the integrals in (42). For this purpose, *interval arithmetic* [17, 34] must be used in this evaluation, in order to take rounding errors into account.

If closed form integration is impossible, a *quadrature formula* should be applied, possibly piecewise, to the integrals, again with evaluation in interval arithmetic. To obtain true upper bounds for the integrals, we need in addition a *remainder term bound* for the quadrature formula, which usually requires rough $\|\cdot\|_\infty$-bounds for some higher derivatives of the integrand. Such rough bounds can be obtained, e.g., by subdividing Ω_0 into (many) small boxes, and performing interval evaluations of the needed higher derivatives over each of these boxes (which gives true supersets of the function value ranges over each of the boxes, and thus, by union, over Ω_0).

There is an alternative way to compute δ if ω is of the form (39), with $\omega_0 \in H^2(\Omega_0) \cap H_0^1(\Omega_0)$, and with Ω_0 having a Lipschitz boundary. This situation can arise, e.g., if Ω is the whole of $\mathbb{R}^n$, and the "computational" domain $\Omega_0 = (-l_1, l_1) \times \cdots \times (-l_n, l_n)$ is chosen as a "large" rectangle, whence ω_0 can be put up, e.g., as a finite Fourier series on Ω_0.

Using partial integration on Ω_0, we obtain now

$$
\|-\Delta\omega + f(\,\cdot\,,\omega)\|_{H^{-1}}
$$
$$
\leq C_2\big[\|-\Delta\omega_0 + f(\,\cdot\,,\omega_0)\|^2_{L^2(\Omega_0)} + \|f(\,\cdot\,,0)\|^2_{L^2(\Omega\setminus\Omega_0)}\big]^{\frac{1}{2}} + C_{\mathrm{tr}}\left\|\frac{\partial\omega_0}{\partial\nu_0}\right\|_{L^2(\partial\Omega_0)},
$$
$$
\tag{43}
$$

with C_{tr} denoting a constant for the trace embedding $H^1(\Omega_0) \hookrightarrow L^2(\partial\Omega_0)$, and $\partial\omega_0/\partial\nu_0$ the normal derivative on $\partial\Omega_0$. It can be shown that

$$
C_{\mathrm{tr}} = \left[\frac{1}{\sigma}\left(\frac{1}{2}\sum_{i=1}^{n}\frac{1}{l_i} + \sqrt{\frac{1}{4}\left(\sum_{i=1}^{n}\frac{1}{l_i}\right)^2 + n\sigma}\right)\right]^{\frac{1}{2}} \tag{44}
$$

is a valid embedding constant; see the appendix of [31] for a proof.

3.3. Bound K for L^{-1}

According to (21), condition (8) now reads

$$
\|u\|_{H_0^1} \leq K\|L[u]\|_{H^{-1}} \quad \text{for all } u \in H_0^1(\Omega), \tag{45}
$$

with L, defined in (6), now given by (note (29), (30))

$$
L = -\Delta + \mathcal{G}'(\omega)\colon H_0^1(\Omega) \to H^{-1}(\Omega).
$$

Under the growth conditions (31)–(34), Lemma 2 (or (35)) shows that, more concretely,

$$
(L[\varphi])[\psi] = \int_{\Omega}\left[\nabla\varphi\cdot\nabla\psi + \frac{\partial f}{\partial y}(\,\cdot\,,\omega)\varphi\psi\right]dx \quad (\varphi,\psi \in H_0^1(\Omega)); \tag{46}
$$

the same formula holds true also in the exponential case mentioned in Remark 2 c).
So we will assume from now on that L is given by (46).

Making use of the isomorphism $\Phi\colon H_0^1(\Omega) \to H^{-1}(\Omega)$ given by (11) or (25), we obtain

$$\|L[u]\|_{H^{-1}} = \|\Phi^{-1}L[u]\|_{H_0^1} \quad (u \in H_0^1(\Omega)).$$

Since moreover $\Phi^{-1}L$ is $\langle\,\cdot\,,\,\cdot\,\rangle_{H_0^1}$-symmetric by (46) and (13), and defined on the whole Hilbert space $H_0^1(\Omega)$, and hence *selfadjoint*, we find that (45) holds for any

$$K \geq [\min\{|\lambda|\colon \lambda \text{ is in the spectrum of } \Phi^{-1}L\}]^{-1}, \tag{47}$$

provided that the min is positive (which is clearly an unavoidable condition for $\Phi^{-1}L$ being invertible with bounded inverse). Thus, in order to compute K, we need bounds for

i)   the essential spectrum of $\Phi^{-1}L$ (i.e., accumulation points of the spectrum, and eigenvalues of infinite multiplicity),

ii) isolated eigenvalues of $\Phi^{-1}L$ of finite multiplicity, more precisely those neighboring 0.

ad i)   If Ω is unbounded, we suppose again that ω is given in the form (39), with some bounded Lipschitz domain $\Omega_0 \subset \Omega$. If Ω is bounded, we may assume the same, simply choosing $\Omega_0 := \Omega$ (and $\omega_0 := \omega$).

Now define $L_0\colon H_0^1(\Omega) \to H^{-1}(\Omega)$ by (46), but with $(\partial f/\partial y)(x,\omega(x))$ replaced by $(\partial f/\partial y)(x,0)$. Using the Sobolev–Kondratchev–Rellich embedding theorem [1], implying the compactness of the embedding $H^1(\Omega_0) \hookrightarrow L^2(\Omega_0)$, we find that $\Phi^{-1}L - \Phi^{-1}L_0\colon H_0^1(\Omega) \to H_0^1(\Omega)$ is compact. Therefore, the perturbation result given in [16, IV, Theorem 5.35] shows that the essential spectra of $\Phi^{-1}L$ and $\Phi^{-1}L_0$ coincide. Thus, being left with the computation of bounds for the essential spectrum of $\Phi^{-1}L_0$, we can use e.g., Fourier transform methods if $\Omega = \mathbb{R}^n$ and $(\partial f/\partial y)(\,\cdot\,,0)$ is constant, or Floquet theory if $(\partial f/\partial y)(\,\cdot\,,0)$ is periodic. Alternatively, if

$$\frac{\partial f}{\partial y}(x,0) \geq c_0 > -\rho^* \quad (x \in \Omega), \tag{48}$$

with $\rho^* \in [0,\infty)$ denoting the minimal point of the spectrum of $-\Delta$ on $H_0^1(\Omega)$, we obtain by straightforward estimates of the Rayleigh quotient that the (full) spectrum of $\Phi^{-1}L_0$, and thus in particular the essential spectrum, is bounded from below by $\min\{1, (c_0 + \rho^*)/(\sigma + \rho^*)\}$.

ad ii)   For computing bounds to eigenvalues of $\Phi^{-1}L$, we choose the parameter σ in the H_0^1-product (22) such that

$$\sigma > \frac{\partial f}{\partial y}(x,\omega(x)) \quad (x \in \Omega); \tag{49}$$

thus, we have to assume that the right-hand side of (49) is bounded above. Furthermore, we assume that the infimum s_0 of the essential spectrum of

$\Phi^{-1}L$ is *positive*, which is true, e.g., if (48) holds. As a particular consequence of (49) (and (25)) we obtain that $s_0 \leq 1$ and all eigenvalues of $\Phi^{-1}L$ are less than 1, and that, via the transformation $\kappa = 1/(1-\lambda)$, the eigenvalue problem $\Phi^{-1}L[u] = \lambda u$ is equivalent to

$$-\Delta u + \sigma u = \kappa\left(\sigma - \frac{\partial f}{\partial y}(\,\cdot\,,\omega)\right)u \tag{50}$$

(to be understood as an equation in $H^{-1}(\Omega)$), which is furthermore equivalent to the eigenvalue problem for the selfadjoint operator $R := (I_{H_0^1(\Omega)} - \Phi^{-1}L)^{-1}$. Thus, *defining* the essential spectrum of problem (50) to be the one of R, we find that it is bounded from below by $1/(1-s_0)$ if $s_0 < 1$, and is empty if $s_0 = 1$. In particular, its infimum is larger than 1, since $s_0 > 0$ by assumption. Therefore, the eigenvalue enclosure methods mentioned briefly below (which are applicable to eigenvalues *below* the essential spectrum; see [39]) can be used to enclose the eigenvalue(s) of problem (50) neighboring 1 (if they exist), whence by the transformation $\kappa = 1/(1-\lambda)$ we obtain enclosures for the eigenvalue(s) of $\Phi^{-1}L$ neighboring 0 (if they exist). Taking also s_0 into account, we can now easily compute the desired constant K via (47). (Note that $K = s_0^{-1}$ can be chosen if no eigenvalues below the essential spectrum exist.)

The desired *eigenvalue bounds* for problem (50) can be obtained by computer-assisted means of their own. For example, *upper* bounds to $\kappa_1, \ldots, \kappa_N$ (with $N \in \mathbb{N}$ given) are easily and efficiently computed by the *Rayleigh–Ritz* method [33]:

Let $\tilde{\varphi}_1, \ldots, \tilde{\varphi}_N \in H_0^1(\Omega)$ denote linearly independent trial functions, for example approximate eigenfunctions obtained by numerical means, and form the matrices

$$A_1 := (\langle \tilde{\varphi}_i, \tilde{\varphi}_j \rangle_{H_0^1})_{i,j=1,\ldots,N}, \quad A_0 := \left(\int_\Omega \left(\sigma - \frac{\partial f}{\partial y}(\,\cdot\,,\omega)\right)\tilde{\varphi}_i\tilde{\varphi}_j\, dx\right)_{i,j=1,\ldots,N}.$$

Then, with $\Lambda_1 \leq \cdots \leq \Lambda_N$ denoting the eigenvalues of the matrix eigenvalue problem

$$A_1 x = \Lambda A_0 x$$

(which can be enclosed by means of verifying numerical linear algebra; see [3]), the Rayleigh–Ritz method gives

$$\kappa_i \leq \Lambda_i \quad \text{for } i = 1, \ldots, N.$$

However, also *lower* eigenvalue bounds are needed, which constitute a more complicated task than upper bounds. The most accurate method for this purpose has been proposed by Lehmann [20], and improved by Goerisch concerning its range of applicability [4]. Its numerical core is again (as in the Rayleigh–Ritz method) a matrix eigenvalue problem, but the accompanying analysis is more involved. In particular, in order to compute lower bounds to the first N eigenvalues, a *rough*

lower bound to the $(N + 1)$-st eigenvalue must be known already. This a priori information can usually be obtained via a *homotopy method* connecting a simple "base problem" with known eigenvalues to the given eigenvalue problem, such that all eigenvalues increase (index-wise) along the homotopy; see [30] or [6] for details on this method, a detailed description of which would be beyond the scope of this article. In fact, [6] contains the newest version of the homotopy method, where only very small (2×2 or even 1×1) matrix eigenvalue problems need to be treated rigorously in the course of the homotopy.

Finding a base problem for problem (50), and a suitable homotopy connecting them, is often possible along the following lines. If Ω is a bounded rectangle (whence the eigenvalues of $-\Delta$ on $H_0^1(\Omega)$ are known), we choose a constant lower bound c_0 for $(\partial f/\partial y)(\,\cdot\,,\omega)$ on Ω, and the coefficient homotopy

$$c_s(x) := (1 - s)c_0 + s\frac{\partial f}{\partial y}(x, \omega(x)) \quad (x \in \Omega,\ 0 \le s \le 1).$$

Then, the family of eigenvalue problems

$$-\Delta u + \sigma u = \kappa^{(s)}(\sigma - c_s(x))u$$

connects the explicitly solvable constant-coefficient base problem ($s = 0$) to problem (50) ($s = 1$), and the eigenvalues increase in s, since the Rayleigh quotient does, by Poincaré's min-max principle.

If Ω is bounded but no rectangle (or ball), we can first choose a rectangle Ω_0 containing Ω, and a domain deformation homotopy between Ω_0 and Ω, to enclose the (first M) eigenvalues of $-\Delta$ on $H_0^1(\Omega)$, see e.g., [32]. Then, the above coefficient homotopy is applied in a second step.

If Ω is unbounded, for example the whole of $\mathbb{R}^n$, we choose a *piecewise* constant lower bound c_0 for $(\partial f/\partial y)(\,\cdot\,,\omega)$ (because a constant lower bound would usually lower the essential spectrum too much), and proceed as before. If, e.g., c_0 is chosen constant on (finitely many) spherical shells, the base problem can be solved using polar coordinates and matched Bessel functions.

3.4. Local Lipschitz bound g for $\mathcal{F}'$

In the setting (21), (29), condition (9) now reads

$$\left| \int_\Omega \left[\frac{\partial f}{\partial y}(x, \omega(x) + u(x)) - \frac{\partial f}{\partial y}(x, \omega(x)) \right] v(x)\varphi(x)\,dx \right| \le g(\|u\|_{H_0^1})\|v\|_{H_0^1}\|\varphi\|_{H_0^1}$$

$$(51)$$

for all $u, v, \varphi \in H_0^1(\Omega)$. Here, we have assumed that the Fréchet derivative of $\mathcal{G}$ (defined in (30)) is given by (35), which is true, e.g., under the growth conditions (31)–(34), but also in the exponential case (with $n = 2$ and Ω bounded) mentioned in Remark 2 c). We will now concentrate on the case where (31)–(34) hold true. For the exponential case, we refer to [32] and to the second example in Section 4.

We start with a monotonically non-decreasing function $\tilde{g}\colon [0,\infty) \to [0,\infty)$ satisfying

$$\left| \frac{\partial f}{\partial y}(x, \omega(x) + y) - \frac{\partial f}{\partial y}(x, \omega(x)) \right| \le \tilde{g}(|y|) \quad \text{for all } x \in \Omega, \ y \in \mathbb{R}, \tag{52}$$

and $\tilde{g}(t) \to 0$ as $t \to 0+$. We require in addition that $\tilde{g}\bigl(t^{\frac{1}{r}}\bigr)$ is a *concave* function of t. Here, $r := r_2$ is the (larger) exponent in (33).

In practice, $\tilde{g}$ can often be put up in the form

$$\tilde{g}(t) = \sum_{j=1}^{N} a_j t^{\mu_j} \quad (0 \le t < \infty),$$

where $a_1, \ldots, a_N > 0$ and $\mu_1, \ldots, \mu_N \in (0, r]$ are arranged in order to satisfy (52).

Now defining $\psi(t) := \tilde{g}\bigl(t^{\frac{1}{r}}\bigr)$, the left-hand side of (51) can be bounded by (note (52))

$$\int_{\Omega} \tilde{g}(|u(x)|)|v(x)\varphi(x)|\, dx = \int_{\Omega} \psi(|u(x)|^r)|v(x)\varphi(x)|\, dx. \tag{53}$$

Without loss of generality we may assume that $v\varphi$ does not vanish identically (almost everywhere) on Ω (otherwise, (51) is trivial because the left-hand side is zero). Since $v\varphi \in L^1(\Omega)$ and hence $|v(x)\varphi(x)|\, dx$ induces a *finite* measure, and since ψ is concave, Jensen's inequality [2] shows that

$$\frac{\int_{\Omega} \psi(|u(x)|^r)|v(x)\varphi(x)|\, dx}{\int_{\Omega}|v(x)\varphi(x)|\, dx} \le \psi\left(\frac{\int_{\Omega}|u(x)|^r|v(x)\varphi(x)|\, dx}{\int_{\Omega}|v(x)\varphi(x)|\, dx} \right). \tag{54}$$

Furthermore, for $\lambda \in (0,1]$ and $t \in [0,\infty)$, $\psi(\lambda t) = \psi(\lambda t + (1-\lambda)0) \ge \lambda \psi(t) + (1-\lambda)\psi(0) = \lambda\psi(t)$, i.e., $\psi(t) \le \lambda^{-1}\psi(\lambda t)$. By Cauchy–Schwarz and the embedding $H_0^1(\Omega) \hookrightarrow L^2(\Omega)$,

$$\lambda := \frac{\int_{\Omega}|v(x)\varphi(x)|\, dx}{C_2^2\|v\|_{H_0^1}\|\varphi\|_{H_0^1}} \in (0,1],$$

whence the right-hand side of (54) is bounded by

$$\frac{C_2^2\|v\|_{H_0^1}\|\varphi\|_{H_0^1}}{\int_{\Omega}|v(x)\varphi(x)|\, dx} \cdot \psi\left(\frac{\int_{\Omega}|u(x)|^r|v(x)\varphi(x)|\, dx}{C_2^2\|v\|_{H_0^1}\|\varphi\|_{H_0^1}} \right). \tag{55}$$

According to (34), we can find some

$$q \in (1,\infty) \text{ if } n = 2, \quad q \in \left[\frac{n}{2}, \infty\right) \text{ if } n \ge 3, \tag{56}$$

such that qr is in the range (27). Since (56) implies that also $p := 2q/(q-1)$ is in the range (27), both the embeddings $H_0^1(\Omega) \hookrightarrow L^{qr}(\Omega)$ and $H_0^1(\Omega) \hookrightarrow L^p(\Omega)$ are bounded. Furthermore, $q^{-1} + p^{-1} + p^{-1} = 1$, whence the generalized Hölder inequality gives

$$\int_\Omega |u(x)|^r |v(x)\varphi(x)| \, dx \leq \|u\|_{L^{qr}}^r \|v\|_{L^p} \|\varphi\|_{L^p} \leq C_{qr}^r C_p^2 \|u\|_{H_0^1}^r \|v\|_{H_0^1} \|\varphi\|_{H_0^1}.$$

Using this estimate in (55), and combining it with (54) and (53), we find that the left-hand side of (51) is bounded by

$$C_2^2 \|v\|_{H_0^1} \|\varphi\|_{H_0^1} \cdot \psi \big(C_{qr}^r (C_p/C_2)^2 \|u\|_{H_0^1}^r \big).$$

Since $\psi(t) = \tilde{g}\big(t^{\frac{1}{r}}\big)$, (51) therefore holds for

$$g(t) := C_2^2 \cdot \tilde{g}\big(C_{qr}(C_p/C_2)^{\frac{2}{r}} t\big) \quad (0 \leq t < \infty), \tag{57}$$

which also satisfies (10) and is non-decreasing. Note that (upper bounds to) the embedding constants can be computed using Lemma 1.

4.   Examples

In our first example, we consider the problem of finding nontrivial solutions to

$$-\Delta u + V(x)u - u^2 = 0 \quad \text{on } \Omega := \mathbb{R}^2, \tag{58}$$

where $V(x) = A + B \sin(\pi(x_1 + x_2)) \sin(\pi(x_1 - x_2))$, with real parameters A and B. The results presented here have been obtained in joint work with B. Breuer and P.J. McKenna.

We are interested only in solutions which are symmetric with respect to reflection about both coordinate axes. Thus, we include these symmetries into all function spaces used, and into the numerical approximation spaces.

We treated the particular case $A = 6$, $B = 2$. On a "computational" domain $\Omega_0 := (-l, l) \times (-l, l)$, we computed an approximation $\omega_0 \in H^2(\Omega_0) \cap H_0^1(\Omega_0)$ of the differential equation in (58), with Dirichlet boundary conditions on $\partial\Omega_0$, as a finite Fourier series with 80×80 trigonometric basis functions. For finding ω_0, we started with a nontrivial approximate solution for Emden's equation (which is (58) with $A = B = 0$) on Ω_0, and performed a path following Newton method, deforming (A, B) from $(0, 0)$ into $(6, 2)$.

In the single Newton steps, we used a collocation method with equidistant collocation points. By increasing the sidelength of Ω_0 in an additional path following, we found that the approximation ω_0 remains "stable," with rapidly decreasing normal derivative $\partial\omega_0/\partial\nu_0$ (on $\partial\Omega_0$), as l increases; this gives rise to some hope that a "good" approximation ω for problem (58) is obtained in the

form (39). For $l = 8$, $\|\partial\omega_0/\partial\nu_0\|_{L^2(\partial\Omega_0)}$ turned out to be small enough compared with $\|-\Delta\omega_0 + V\omega_0 - \omega_0^2\|_{L^2(\Omega_0)}$, and we computed a defect bound δ (satisfying (40)) via (43) as

$$\delta = 0.7102 \cdot 10^{-2}; \tag{59}$$

note that the embedding constants $C_2 = \sigma^{-\frac{1}{2}}$, and $C_{\mathrm{tr}} = \sigma^{-\frac{1}{2}}[l^{-1} + \sqrt{l^{-2} + 2\sigma}]^{\frac{1}{2}}$ can be used; compare (44). Moreover, (49) requires $\sigma > A + B = 8$ (since ω turns out to be non-negative). Choosing $\sigma := 9$, we obtain $C_2 \leq 0.3334$ and $C_{\mathrm{tr}} \leq 0.6968$.

Since condition (48) holds for $c_0 = A - B = 4$ (and $\rho^* = 0$), the arguments following (48) give the lower bound $s_0 := 4/9 \geq 0.4444$ for the essential spectrum of $\Phi^{-1}L$, and hence the lower bound $1/(1 - s_0) = 1.8$ for the essential spectrum of problem (50).

By the eigenvalue enclosure methods mentioned at the end of Subsection 3.3, we were able to compute the bounds

$$\kappa_1 \leq 0.5293, \quad \kappa_2 \geq 1.1769$$

for the first two eigenvalues of problem (50), which by (47) leads to the constant

$$K = 6.653 \tag{60}$$

satisfying (45).

For computing g satisfying (9) or (51), we first note that (52) holds for

$$\tilde{g}(t) := 2t,$$

and (33) for $r_1 = r_2 = 1$, whence the additional concavity condition is satisfied. Choosing $q := 2$ we obtain $qr = 2$ and $p = 4$ in the arguments following (56), whence (57) gives

$$g(t) = 2C_2 C_4^2 t = \frac{1}{9}t \tag{61}$$

since $2C_2 C_4^2 = \sigma^{-1}$ by Lemma 1.

Using (59)–(61), we find that (14) and (15) hold for $\alpha = 0.04811$, whence Theorem 1 implies the existence of a solution $u^* \in H_0^1(\mathbb{R}^2) = H^1(\mathbb{R}^2)$ to problem (58) such that

$$\|u^* - \omega\|_{H_0^1} \leq 0.04811. \tag{62}$$

It is easy to check on the basis of the numerical data that $\|\omega\|_{H_0^1} > 0.04811$, whence (62) shows in particular that u^* is non-trivial.

We wish to remark that it would be of great interest to achieve such results also for cases where $0 < A < B$ in the potential V, because V is then no longer non-negative, which excludes an important class of purely analytical approaches to prove existence of a nontrivial solution. So far, we were not successful with

such cases due to difficulties in the homotopy method which has to be used for our computer-assisted eigenvalue enclosures (see the brief remarks at the end of Subsection 3.3); note that these difficulties occur on a rather "technical" level. We were however able to compute an (apparently) "good" approximation ω, e.g., in the case $A = 6$, $B = 26$ (using $l = 8$ again).

The following Fig. 1 shows plots of ω for the successful case $A = 6$, $B = 2$, and for the non-successful case $A = 6$, $B = 26$.
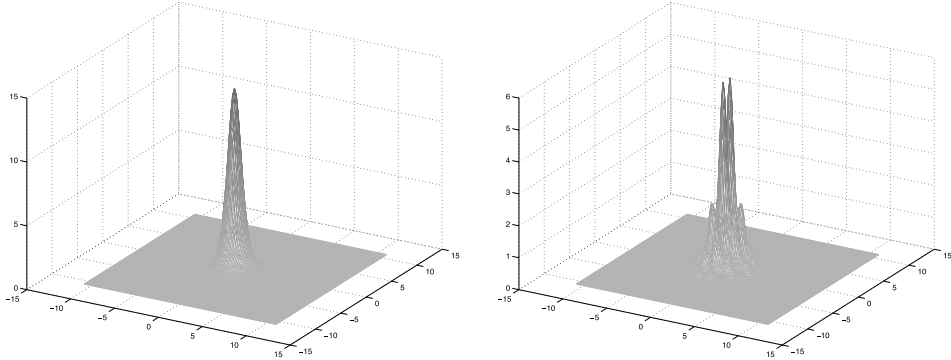

Fig. 1.   Example (58); $A = 6$, $B = 2$ (left) and $A = 6$, $B = 26$ (right).

In our second example, we consider the *Gelfand* equation

$$-\Delta u = \lambda e^u \text{ on } \Omega, \quad u = 0 \text{ on } \partial\Omega, \tag{63}$$

depending on a real parameter λ. We are interested in parameter values $\lambda \geq 0$ only; negative values of λ are less important. The results reported on here are joint work with C. Wieners and published in [32].

It is known that, on "simple" domains Ω like the unit square or the unit ball, problem (63) has a "nose"-shaped branch (λ, u) of solutions, starting in $(\lambda = 0, u \equiv 0)$, going up to some maximal value of λ where the branch has a turning point, and then returning to $\lambda = 0$ but with $\|u\|_\infty$ tending to ∞ as $\lambda \to 0$. Moreover, there are no other solutions (on these "simple" domains).

Here (and in [32]) we are concerned with a special *non-convex* domain $\Omega \subset \mathbb{R}^2$ plotted in Fig. 2. (For an exact quantitative definition of Ω, see [32].) Ω is symmetric with respect to the x_1-axis but not quite symmetric with respect to the x_2-axis; it is a bit shorter on the left-hand side than on the right. Starting at $(\lambda = 0, u \equiv 0)$, and performing numerical branch following, we obtained the usual "nose"-shaped branch (of approximate solutions) plotted in Fig. 3; the plot consists in fact of an interpolation of many grid points.

Obviously, the approximations develop substantial unsymmetries along the branch. In order to find new (approximate) solutions, we *reflected* such an unsymmetric approximation about the x_2-axis, re-arranged the boundary values (which is necessary but easily possible due to the slight unsymmetry of Ω), and
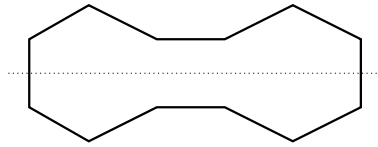
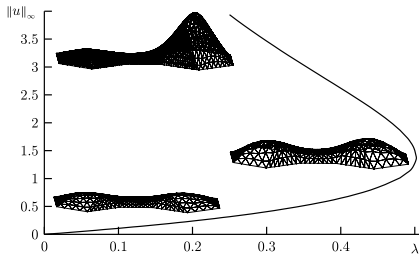Fig. 2.   Domain Ω for example (63).

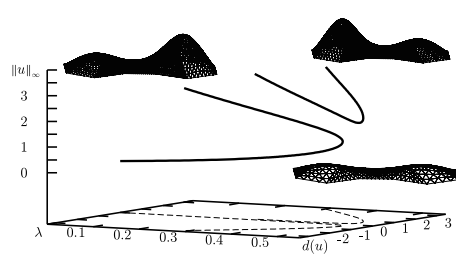Fig. 3.   Main branch of (approximate) solutions for problem (63).

Fig. 4.   Main and new branch for problem (63).

re-started the Newton iteration. Fortunately, it "converged" to a new approximation, and by branch following we could detect a new branch of approximate solutions plotted (together with the "old" one) in Fig. 4; in order to obtain a nicely visible separation of the two branches, we introduced the difference $d(u)$ between the two peak values of each approximation as a third dimension in the bifurcation diagram.

In oder to *prove* the existence of a new solution branch, we performed the computer-assisted method described above for the selected value $\lambda = 15/32$. Here, our "new" approximation ω was computed with 65536 quadratic triangular finite elements, corresponding to 132225 unknowns.

For calculating a defect bound δ (satisfying (40)), we used essentially (up to some technical refinements) the estimate (42), where the approximation $\rho \in H(\mathrm{div}, \Omega)$ to $\nabla \omega$ was computed by linear Raviart–Thomas elements. The result is

$$\delta = 0.8979 \cdot 10^{-2}. \tag{64}$$

Since $(\partial f / \partial y)(x, y) = -\lambda e^y < 0$ here, condition (49) is satisfied for $\sigma = 0$; indeed, this choice is allowed because Ω is bounded (see Remark 2 a)). We computed eigenvalue bounds for problem (50) by the Rayleigh–Ritz and the Lehmann–Goerisch method, exploiting symmetry properties, with the final result that (45) holds for

$$K = 3.126; \tag{65}$$

note that problem (50) has no essential spectrum here since Ω is bounded.

For proving that $\mathcal{G}$ defined in (30) is Fréchet differentiable and for computing a function g satisfying (9) or (51), we make essential use of the Trudinger–Moser inequality (38) (note that Lemma 2 does not apply here due to the exponential nonlinearity). For each $u \in H_0^1(\Omega) \setminus \{0\}$,

$$4|u(x)| = 2 \cdot 2\|u\|_{H_0^1} \cdot \frac{|u(x)|}{\|u\|_{H_0^1}} \leq 4\|u\|_{H_0^1}^2 + \left(\frac{|u(x)|}{\|u\|_{H_0^1}}\right)^2,$$

whence (38) (with $c := 1$) gives, since $[4\pi/(4\pi - 1)]^{1/4} \leq 1.03$,

$$\|\exp(|u|)\|_{L^4} \leq 1.03 \operatorname{meas}(\Omega)^{\frac{1}{4}} \exp\left(\|u\|_{H_0^1}^2\right). \tag{66}$$

For all $u_0, u, v, \varphi \in H_0^1(\Omega)$, the generalized Hölder inequality and (66) imply

$$\int_\Omega |e^{u_0+u} - e^{u_0}| \, |v| \, |\varphi| \, dx \leq \int_\Omega e^{u_0} e^{|u|} |u| \, |v| \, |\varphi| \, dx$$
$$\leq \|e^{u_0}\|_{L^4} \|e^{|u|}\|_{L^4} \|u\|_{L^6} \|v\|_{L^6} \|\varphi\|_{L^6}$$
$$\leq \|e^{u_0}\|_{L^4} \cdot 1.03 \operatorname{meas}(\Omega)^{\frac{1}{4}} \exp\left(\|u\|_{H_0^1}^2\right) C_6^3 \|u\|_{H_0^1} \|v\|_{H_0^1} \|\varphi\|_{H_0^1}. \tag{67}$$

By an argument similar to the abstract estimate (18), (19), we obtain the desired Fréchet differentiability from (67). Furthermore, for $u_0 := \omega$, (67) shows that (51) holds for

$$g(t) = \gamma t e^{t^2}, \quad \text{where} \quad \gamma := \|\lambda e^\omega\|_{L^4} \cdot 1.03 \operatorname{meas}(\Omega)^{\frac{1}{4}} C_6^3, \tag{68}$$

and thus $G(t) = \int_0^t g(s) \, ds = \frac{1}{2}\gamma(\exp(t^2) - 1) \leq \frac{1}{2}\gamma t^2 \exp(t^2)$. From the numerical data, Lemma 1, and the result $\rho^* \geq 1.4399$ (obtained by eigenvalue bounds), we obtain that $\gamma \leq 5.62$. Together with (64), (65), (68), we obtain that (14) and (15) hold for $\alpha := 0.05066$, whence Theorem 1 gives the existence of a solution $u^* \in H_0^1(\Omega)$ of problem (63) (with $\lambda = 15/32$) such that

$$\|u^* - \omega\|_{H_0^1} \leq 0.05066. \tag{69}$$

(It should be remarked that we could do without condition (15) being satisfied, since Ω is bounded and hence we could use compactness properties, and Schauder's instead of Banach's fixed point theorem.)

In the same way, we also obtained existence results with H_0^1-error bounds for two solutions of (63) on the "old" (nose-shaped) branch, again for $\lambda = 15/32$. From the numerical data, and all three error bounds, we can easily deduce that the three solutions are pairwise different, whence u^* established above lies on a new independent solution branch; the implicit function theorem (plus some perturbation type argument showing that $-\Delta - \lambda e^{u^*}: H_0^1(\Omega) \to H^{-1}(\Omega)$ is one-to-one and onto) shows that indeed a solution *branch* through $(\lambda = 15/32, u^*)$ exists.

References

[1] R.A. Adams, Sobolev Spaces. Academic Press, New York, 1975.

[2] H. Bauer, Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie, 3rd edition. de Gruyter, Berlin, 1978.

[3] H. Behnke, Inclusion of eigenvalues of general eigenvalue problems for matrices. Scientific Computation with Automatic Result Verification, U. Kulisch and H.J. Stetter (eds.), Computing Suppl., **6** (1987), 69–78.

[4] H. Behnke and F. Goerisch, Inclusions for eigenvalues of selfadjoint problems. Topics in Validated Computations, J. Herzberger (ed.), Series Studies in Computational Mathematics, North-Holland, Amsterdam, 1994, 277–322.

[5] B. Breuer, P.J. McKenna and M. Plum, Multiple solutions for a semilinear boundary value problem: a computational multiplicity proof. J. Differential Equations, **195** (2003), 243–269.

[6] B. Breuer, J. Horak, P.J. McKenna and M. Plum, A computer-assisted existence and multiplicity proof for travelling waves in a nonlinearly supported beam. J. Differential Equations, **224** (2006), 60–97.

[7] Y.S. Choi and P.J. McKenna, A mountain pass method for the numerical solutions of semilinear elliptic problems. Nonlinear Anal. Theory Methods Appl., **20** (1993), 417–437.

[8] L. Collatz, Aufgaben monotoner Art. Arch. Math., **3** (1952), 366–376.

[9] L. Collatz, The Numerical Treatment of Differential Equations. Springer, Berlin-Heidelberg, 1960.

[10] S. Day, Y. Hiraoka, K. Mischaikow and T. Ogawa, Rigorous numerics for global dynamics: a study of the Swift–Hohenberg equation. SIAM J. Appl. Dynamical Systems, **4** (2005), 1–31.

[11] B. Fazekas, M. Plum and Ch. Wieners, Enclosure for biharmonic equation. Dagstuhl Online Seminar Proceedings 05391, 2005, `http://drops.dagstuhl.de/portals/05391/`.

[12] A. Friedman, Partial differential equations. Holt, Rinehart and Winston, New York, 1969.

[13] D. Gilbarg and N.S. Trudinger, Elliptic Partial Differential Equations of Second Order, 2nd edition. Springer, Berlin-Heidelberg, 1983.

[14] P. Grisvard, Elliptic Problems in Nonsmooth Domains. Pitman, Boston, MA, 1985.

[15] Y. Hiraoka, Topological verification in infinite dimensional dynamical systems. Doctoral dissertation, Department of Informatics and Mathematical Science, Graduate School of Engineering Science, Osaka University, 2004.

[16] T. Kato, Perturbation Theory for Linear Operators. Springer, New York, 1966.

[17] R. Klatte, U. Kulisch, C. Lawo, M. Rausch and A. Wiethoff, C-XSC-A C++ Class Library for Extended Scientific Computing. Springer, Berlin, 1993.

[18] O.A. Ladyzhenskaya and N.N. Ural'tseva, Linear and Quasilinear Elliptic Equations. Academic Press, New York, 1968.

[19] J.-R. Lahmann and M. Plum, A computer-assisted instability proof for the Orr–Sommerfeld equation with Blasius profile. ZAMM, **84** (2004), 188–204.

[20] N.J. Lehmann, Optimale Eigenwerteinschließungen. Numer. Math., **5** (1963), 246–272.

[21] J. Moser, A sharp form of an inequality by N. Trudinger. Indiana Univ. Math. J., **20** (1971), 1077–1092.

[22] K. Nagatou, M.T. Nakao and N. Yamamoto, An approach to the numerical verification of solutions for nonlinear elliptic problems with local uniqueness. Numer. Funct. Anal. Optim., **20** (1999), 543–565.

[23] M.T. Nakao, Solving nonlinear elliptic problems with result verification using an H^{-1} type residual iteration. Computing Suppl., **9** (1993), 161–173.

[24] M.T. Nakao and N. Yamamoto, Numerical verifications for solutions to elliptic equations using residual iterations with higher order finite elements. J. Comput. Appl. Math., **60** (1995), 271–279.

[25] M.T. Nakao, M. Plum and Y. Watanabe, A computer-assisted instability proof for the Orr–Sommerfeld problem with Poiseuille flow. ZAMM, **89** (2009), 5–18, DOI: 10.1002/zamm.200700158.

[26] M. Plum, Explicit H_2-estimates and pointwise bounds for solutions of second-order elliptic boundary value problems. J. Math. Anal. Appl., **165** (1992), 36–61.

[27] M. Plum, Enclosures for solutions of parameter-dependent nonlinear elliptic boundary value problems: Theory and implementation on a parallel computer. Interval Computations, **3** (1994), 106–121.

[28] M. Plum, Existence and enclosure results for continua of solutions of parameter-dependent
      nonlinear boundary value problems. J. Comput. Appl. Math., **60** (1995), 187–200.

[29] M. Plum, Enclosures for two-point boundary value problems near bifurcation points. Sci-
      entific Computing and Validated Numerics, G. Alefeld, A. Frommer and B. Lang (eds.),
      Akademie Verlag, 1996, 265–279.

[30] M. Plum, Guaranteed numerical bounds for eigenvalues. Spectral Theory and Computational
      Methods of Sturm–Liouville Problems, D. Hinton and P.W. Schaefer (eds.), Marcel Dekker,
      New York, 1997, 313–332.

[31] M. Plum, Existence and multiplicity proofs for semilinear elliptic boundary value problems
      by computer assistance. DMV Jahresbericht JB, **110** (2008), 19–54.

[32] M. Plum and Ch. Wieners, New solutions of the Gelfand problem. J. Math. Anal. Appl.,
      269 (2002), 588–606.

[33] K. Rektorys, Variational Methods in Mathematics, Science and Engineering, 2nd edition.
      Reidel Publ. Co., Dordrecht, 1980.

[34] S.M. Rump, INTLAB—INTerval LABoratory, a Matlab toolbox for verified computations,
      version 4.2.1. Inst. Informatik, TU Hamburg-Harburg, 2002, `http://www.ti3.tu-harburg.de/`
      `rump/intlab/`.

[35] J. Schröder, Vom Defekt ausgehende Fehlerabschätzungen bei Differentialgleichungen. Arch.
      Rat. Mech. Anal., **3** (1959), 219–228.

[36] J. Schröder, Operator Inequalities. Academic Press, New York, 1980.

[37] J. Schröder, Operator inequalities and applications. Inequalities, Fifty Years on from Hardy,
      Littlewood and Polya, W.N. Everitt (ed.), Marcel Dekker Inc., 1991, 163–210.

[38] W. Walter, Differential and Integral Inequalities. Springer, Berlin-Heidelberg, 1970.

[39] S. Zimmermann and U. Mertins, Variational bounds to eigenvalues of self-adjoint eigenvalue
      problems with arbitrary spectrum. Z. Anal. Anwendungen, **14** (1995), 327–345.