

# Numerical analysis of a singularly perturbed nonlinear reaction-diffusion problem with multiple solutions <sup>\*</sup>

Natalia Kopteva<sup>†</sup> and Martin Stynes<sup>‡</sup>

July 5, 2004

## Abstract

A nonlinear reaction-diffusion two-point boundary value problem with multiple solutions is considered. Its second-order derivative is multiplied by a small positive parameter  $\varepsilon$ , which induces boundary layers. Using dynamical systems techniques, asymptotic properties of its discrete sub- and super-solutions are derived. These properties are used to investigate the accuracy of solutions of a standard three-point difference scheme on layer-adapted meshes of Bakhvalov and Shishkin types. It is shown that one gets second-order convergence (with, in the case of the Shishkin mesh, a logarithmic factor) in the discrete maximum norm, uniformly in  $\varepsilon$  for  $\varepsilon \leq CN^{-1}$ , where  $N$  is the number of mesh intervals. Numerical experiments are performed to support the theoretical results.

*AMS Subject Classification:* Primary 65L10, Secondary 65L12, 65L50.

---

<sup>\*</sup>This research was supported by the Boole Centre for Research in Informatics, National University of Ireland, Cork, Ireland

<sup>†</sup>Department of Computational Mathematics and Cybernetics, Moscow State University, Vorob'ovy Gory, 119899 Moscow, Russia (kopteva@cs.msu.su)

<sup>‡</sup>Department of Mathematics, National University of Ireland, Cork, Ireland (m.stynes@ucc.ie)

# 1 Introduction

Consider the singularly perturbed semilinear reaction-diffusion boundary-value problem

$$F_\varepsilon u(x) \equiv -\varepsilon^2 u''(x) + b(x, u) = 0 \quad \text{for } x \in X := (0, 1), \quad (1.1a)$$

$$u(0) = g_0, \quad u(1) = g_1, \quad (1.1b)$$

where  $\varepsilon$  is a small positive parameter,  $b \in C^\infty(X \times \mathcal{R}^1)$ , and  $g_0$  and  $g_1$  are given constants. Related problems arise in the modelling of many biological processes [12, §14.7]. We shall examine solutions of (1.1) that exhibit boundary layer behaviour. In general, solutions of (1.1) may also have interior transition layers, which we will consider in a future paper.

The reduced problem of (1.1) is defined by formally setting  $\varepsilon = 0$  in (1.1.a), viz.,

$$b(x, u_0(x)) = 0 \quad \text{for } x \in X. \quad (1.2)$$

A solution  $u_0$  of (1.2) does not in general satisfy either of the boundary conditions in (1.1b).

In the literature it is often assumed that

$$b_u(x, u) > \gamma^2 > 0 \quad \text{for all } (x, u) \in X \times \mathcal{R}^1, \quad (1.3)$$

for some positive constant  $\gamma$ . Under this condition the reduced problem has a unique solution  $u_0 \in C^\infty(X)$ , as can be seen by using the implicit function theorem and the compactness of  $X$ . The condition (1.3) is nevertheless rather restrictive since it is assumed to hold true even at points that are far from any solution of (1.1). Consequently we shall examine (1.1) under weaker local hypotheses that permit (1.2) to have more than one solution.

Following Fife [4], D'Annunzio [1], Howes [6], O'Malley [14] and Sun and Stynes [17], we consider (1.1) under the assumptions that:

(i) it has a *stable reduced solution*, i.e., there exists a solution  $u_0 \in C^\infty(X)$  of (1.2) such that

$$b_u(x, u_0) > \gamma^2 > 0 \quad \text{for all } x \in X; \quad (1.4a)$$

(ii) the boundary conditions satisfy

$$\int_0^v b(0, u_0(0) + s) ds > 0 \quad \text{for all } v \in (0, g_0 - u_0(0)]', \quad (1.4b)$$

and

$$\int_0^v b(1, u_0(1) + s) ds > 0 \quad \text{for all } v \in (0, g_1 - u_0(1)]'; \quad (1.4c)$$

here the notation  $(0, a]'$  is defined to be  $(0, a]$  when  $a > 0$  and  $[a, 0)$  when  $a < 0$ . If  $u_0(0) = g_0$  or  $u_0(1) = g_1$ , then the corresponding zero-order boundary-layer term is not needed in the asymptotic expansion of  $u$ , so (1.4b) or (1.4c) is not needed. The formulation of (1.4b, 1.4c) in [1, 4, 14, 17] is equivalent. Note that (1.3) implies (1.4). Tobiska [18, Chapter 3] considers a generalization of (1.1) where  $u$  is a function of  $n$  variables and  $u''$  is replaced by a much more general elliptic operator; he assumes (1.4a) and a condition on the boundary layers whose generality lies between (1.3) and (1.4b,c). Fife [4] considers a version of (1.1) and (1.4) for functions of  $n$  variables.

The papers [1, 4, 14, 17] all use the techniques of classical asymptotic analysis for singularly perturbed differential equations to construct asymptotic expansions of solutions  $u$  of (1.1). Conditions (1.4b,c), which are used in all these papers, might seem obscure but in fact they are quite natural. Their precise relevance will be revealed by a detailed analysis of a certain nonlinear differential equation (see (2.2) below) that aims to define the zero-order boundary layer terms in an expansion of  $u$ .

In the present paper, unlike [1, 4, 14, 17], we work in the framework of dynamical systems, as used for example in [19, 20]. This makes the conditions (1.4b,c) much easier to understand and greatly simplifies the presentation. Two by-products of this new approach are a shortening of the proof of the main result of [17] and an amendment to a gap in its argument (see our Remark 2.3). Furthermore, this framework seems suitable for the extension of the present results to systems of nonlinear reaction-diffusion systems; we shall investigate this in a subsequent paper.

Vasil'eva and Butuzov [19, 20] (see also [7]) use dynamical systems to find asymptotic expansions for solutions of (1.1). We take their process further by constructing discrete sub- and super-solutions for (1.1) on special layer-adapted meshes. This is a discrete analogue of the differential inequalities approach; see for example [6, 13]. Invoking the theory of  $Z$ -fields then yields existence of discrete solutions and enables us to prove sharp bounds on the error in the computed solution.

The paper is organized as follows. In §2 asymptotic properties of solutions of (1.1) are derived using techniques from dynamical systems. In particular we construct sub- and super-solutions for each solution of (1.1). In §3 a difference scheme for solving (1.1) is described, and discrete analogues of the sub- and super-solutions are used to obtain tight upper and lower bounds on the computed solutions. Precise convergence results for the numerical method are then derived on Bakhvalov and Shishkin meshes. Finally, in §4, numerical results illustrate the sharpness of our convergence

bounds.

We make three simplifying assumptions to facilitate our presentation.

**Assumption 1.** For convenience in our theoretical analysis, assume without loss of generality that  $u_0(1) = g_1$ , as the construction of the layer terms at each end of the interval  $X$  is carried out independently of the layer terms at the other end.

**Assumption 2.** To avoid considering cases, assume that  $u_0(0) < g_0$ .

**Assumption 3.** Throughout our analysis take

$$\varepsilon \leq CN^{-1}, \tag{1.5}$$

where  $N$  is the number of mesh intervals. This is not a practical restriction, and from a theoretical viewpoint the analysis of a nonlinear problem such as (1.1) would be very different if  $\varepsilon$  were not small.

*Notation.* Throughout this paper we let  $C$  denote a generic positive constant that may take different values in different formulas, but is always independent of  $N$  and  $\varepsilon$ . A subscripted  $C$  (e.g.,  $C_1$ ) denotes a positive constant that is independent of  $N$  and  $\varepsilon$  and takes a fixed value. For any function  $w \in C(\bar{X})$ , the notation  $w_i$  means  $w(x_i)$ , where  $x_i$  is a mesh point.

## 2 Asymptotic expansions and dynamical systems

### 2.1 Zero-order asymptotic expansion

Rewrite (1.1a) as the system

$$\varepsilon u' = U, \tag{2.1a}$$

$$\varepsilon U' = b(x, u). \tag{2.1b}$$

When  $\varepsilon = 0$ , (1.2) and (1.4) imply that (2.1) has an isolated root  $(u_0(x), 0)$ . We seek a solution of (2.1) that, away from the boundary of  $X$ , is close to  $u_0$ . Thus, take  $u_0$  as the *zero-order smooth component* in an asymptotic expansion of  $u$ . Introduce the stretched variable  $\xi := x/\varepsilon$ , where  $0 < \xi < \infty$ . Use a dot to denote differentiation with respect to  $\xi$ . Then the autonomous nonlinear system for the *zero-order boundary-layer* term  $v(x) \equiv \tilde{v}(\xi)$  that is

associated with the endpoint  $x = 0$  is

$$\dot{\tilde{v}} = \tilde{V}, \quad (2.2a)$$

$$\dot{\tilde{V}} = \tilde{b}(\tilde{v}) := b(0, u_0(0) + \tilde{v}) \quad \text{for } 0 < \xi < \infty, \quad (2.2b)$$

with boundary conditions

$$\tilde{v}(0) = g_0 - u_0(0), \quad \tilde{v}(\infty) = \tilde{V}(\infty) = 0. \quad (2.2c)$$

Does (2.2) have a solution? Consider the phase plane for (2.2a) and (2.2b). First, (1.2) implies that (2.2) has a fixed point at  $(0,0)$ . To find a solution of (2.2), we need a trajectory that leaves the point  $(\tilde{v}(0), \tilde{V}(0))$ , where  $\tilde{V}(0)$  is unknown, and enters  $(0,0)$  as  $\xi \rightarrow \infty$ .

The existence of this trajectory can be deduced from an argument of Vasil'eva and Butuzov [19], [20, §2.3.1] that we now describe. The Jacobian matrix associated with the right-hand side of (2.1) is

$$\begin{pmatrix} 0 & 1 \\ b_u & 0 \end{pmatrix},$$

and its eigenvalues are  $\pm\sqrt{b_u}$ . Now (1.4a) implies that these eigenvalues are real and of opposite sign at the fixed point  $(0,0)$ , so  $(0,0)$  is a saddle point for (2.2). Thus four separatrices meet at  $(0,0)$  and *two of them enter* this saddle point as  $\xi \rightarrow \infty$ . In the phase plane, the condition that

$$\text{the straight line } \tilde{v}(0) = g_0 - u_0(0) \text{ intersects one of these two separatrices} \quad (2.3)$$

is necessary and sufficient for this separatrix to be a solution curve of (2.2).

We now show that our condition (1.4b) is precisely what is needed to ensure that (2.3) is satisfied. Set  $\gamma_0 = \sqrt{b'(0)} = \sqrt{b_u(0, u_0(0))}$ , so  $\gamma_0 > \gamma > 0$  by (1.4a).

**Lemma 2.1.** *Under our hypothesis (1.4b), problem (2.2) has a solution  $\tilde{v}(\xi)$ , and for each  $\delta \in (0, \gamma_0)$ , there exists a positive constant  $C_\delta$  such that*

$$|\tilde{v}^{(k)}(\xi)| \leq C_\delta e^{-(\gamma_0 - \delta)\xi} \quad (2.4)$$

for  $0 \leq \xi < \infty$  and  $k = 0, 1, \dots, 4$ .

*Proof.* From the discussion above, to obtain existence of a solution  $\tilde{v}$  of (2.2), all that remains is to check that (2.3) holds true. From (2.2) one has  $\tilde{V}d\tilde{V} = \tilde{b}(\tilde{v})d\tilde{v}$ , so as the separatrices entering  $(0,0)$  satisfy the boundary condition (2.2c) at infinity, we can integrate and solve to obtain

$$\tilde{V} = \pm \sqrt{2 \int_0^{\tilde{v}} \tilde{b}(s) ds}. \quad (2.5)$$

To choose here a trajectory that leaves  $(\tilde{v}(0), \tilde{V}(0))$  and decays to  $(0, 0)$  as  $\xi \rightarrow \infty$ , recall from Assumption 2 that  $\tilde{v}(0) > 0$ ; thus one should choose the negative root in (2.5). Now (1.4b) ensures that (2.5) defines  $\tilde{V}$  along this separatrix from  $\tilde{v}(0)$  to  $\tilde{v}(\infty)$ , i.e., (2.3) is satisfied. Hence (2.2) has a solution.

In the case where  $\tilde{b}$  is linear, the separatrix that enters  $(0, 0)$  is of the form  $Ce^{-\gamma_0\xi}$ . In the more general nonlinear case, there is the following similar result. Write  $\phi(\cdot)$  for the time-one mapping [15, p.165] of the flow associated with (2.2); thus  $\phi(\tilde{v}(\xi), \tilde{V}(\xi)) = (\tilde{v}(\xi + 1), \tilde{V}(\xi + 1))$ . Then  $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  is smooth. The Stable Manifold Theorem [15, p.182] for hyperbolic fixed points of differentiable mappings states that, along the separatrix that enters  $(0, 0)$ , iterates of  $\phi$  decay geometrically to  $(0, 0)$ , and gives an upper bound for this decay rate. Rewriting this result in terms of  $(\tilde{v}, \tilde{V})$  yields precisely (2.4) for the cases  $k = 0$  and  $k = 1$ .

Write (2.2) in the form

$$\ddot{v} = \tilde{b}(\tilde{v}). \quad (2.6)$$

Then (1.2) implies that  $\ddot{v} = \tilde{v} b_u(0, u_0(0) + \hat{v})$  for some  $\hat{v}$  between 0 and  $\tilde{v}$ . From the case  $k = 0$  of (2.4) it follows that  $\hat{v}$  is bounded and  $|\ddot{v}(\xi)| \leq C|\tilde{v}(\xi)| \leq Ce^{-(\gamma_0 - \delta)\xi}$ , which proves (2.4) in the case  $k = 2$ .

Differentiating (2.6), the case  $k = 3$  follows from the bounds obtained already. Differentiate again to get the result for  $k = 4$ . ■

**Remark 2.1.** *Vasil'eva and Butuzov do not specify precisely the decay constant in the exponential of (2.4), but for the numerical analysis of §3 we need this information.*

**Remark 2.2.** *For the interested reader, an Appendix to the paper gives an alternative self-contained proof of (2.4) in the cases  $k = 0, 1$ , but this argument would be more difficult to extend to generalizations of (1.1) such as a system of reaction-diffusion equations.*

## 2.2 A generalization of (2.2)

To obtain tight control on the solutions of (1.1), one needs to construct certain sub- and super-solutions. The first step in this direction is an examination of a generalization of (2.2). Let  $p$  be a small constant that will be

chosen later. Consider

$$\dot{\tilde{v}}(\xi, p) = \tilde{V}, \quad (2.7a)$$

$$\dot{\tilde{V}}(\xi, p) = \tilde{b}(\tilde{v}, p) := b(0, u_0(0) + \tilde{v}) - p\tilde{v} \quad \text{for } 0 < \xi < \infty \quad (2.7b)$$

with boundary conditions

$$\tilde{v}(0, p) = g_0 - u_0(0), \quad \tilde{v}(\infty, p) = \tilde{V}(\infty) = 0. \quad (2.7c)$$

By Assumption 2,  $\tilde{v}(0, p) = \tilde{v}(0) > 0$ .

For sufficiently small  $|p|$ , the next Lemma shows that the appropriate analogue of (1.4b) is valid for (2.7).

**Lemma 2.2.** *Set  $g(s) = b(0, u_0(0) + s)$ . Then  $g(0) = 0$ ,  $g'(0) > 0$ , and there exists  $p_0 \in (0, g'(0)) = (0, \gamma_0^2)$  such that the inequality*

$$\int_0^{\tilde{v}} \tilde{b}(s, p) ds = \int_0^{\tilde{v}} \{b(0, u_0(0) + s) - ps\} ds > 0 \quad (2.8)$$

holds true for  $|p| \leq p_0$  and all  $\tilde{v} \in (0, \tilde{v}(0)]$ .

*Proof.* First,  $g(0) = 0$  and  $g'(0) = \gamma_0^2 > 0$  follow immediately from (1.2) and (1.4a). Initially suppose that  $p_0 \leq g'(0)/2$  (it will be specified precisely later) and assume that  $|p| \leq p_0$ .

Note that  $g(s) - ps \geq [g'(0) - p_0]s - C_0s^2$  for all  $s \in [0, \tilde{v}(0)]$  and some constant  $C_0$ . Hence  $g(s) - ps > 0$  for  $0 < s < s_0 := g'(0)/(2C_0)$ . Consequently (2.8) holds true for  $0 < \tilde{v} \leq s_0$ . Suppose that  $s_0 < \tilde{v}(0)$ , as otherwise we are done. To handle  $s_0 \leq \tilde{v} \leq \tilde{v}(0)$ , let

$$m = \min_{\tilde{v} \in [s_0, \tilde{v}(0)]} \int_0^{\tilde{v}} g(s) ds.$$

Now (1.4b) implies that  $m > 0$ . Set  $p_0 = \min\{m/\tilde{v}^2(0), g'(0)/2\}$ . Then

$$\int_0^{\tilde{v}} ps ds = \frac{p\tilde{v}^2}{2} \leq \frac{p_0\tilde{v}^2(0)}{2} \leq \frac{m}{2} < \int_0^{\tilde{v}} g(s) ds,$$

which yields (2.8) for  $v_0 \in [s_0, \tilde{v}(0)]$ . ■

**Lemma 2.3.** *Suppose that  $|p| \leq p_0$ . Let  $\delta \in (0, \sqrt{b_u(0, u_0(0)) - p})$  be arbitrary but fixed. Then (2.7) has a solution  $\tilde{v}(\xi, p)$  and there exists a positive constant  $C_\delta$  such that*

$$\left| \frac{\partial^k \tilde{v}(\xi, p)}{\partial \xi^k} \right| \leq C_\delta e^{-(\gamma_0 - \delta)\xi} \quad (2.9)$$

and

$$0 \leq \frac{\partial \tilde{v}(\xi, p)}{\partial p} \leq C_\delta e^{-(\gamma_0 - \delta)\xi} \quad (2.10)$$

for  $0 \leq \xi < \infty$  and  $k = 0, 1, \dots, 4$ .

*Proof.* Lemma 2.2 shows that  $\tilde{b}(\tilde{v}, p)$  in (2.7) satisfies the same conditions as  $\tilde{b}(\tilde{v})$  in (2.2). Hence, the conclusions of §2.1 are also applicable to (2.7), which yields (2.9).

Now  $\tilde{v}(0, p) = \tilde{v}(0) > 0$ , so analogously to (2.5) we have

$$\dot{\tilde{v}}(\xi, p) = -\sqrt{2 \int_0^{v_0(\xi, p)} \{b(0, u_0(0) + s) - ps\} ds}. \quad (2.11)$$

Suppose that  $p_2$  and  $\tilde{p}_2$  satisfy  $|p_2| \leq p_0$  and  $|\tilde{p}_2| \leq \tilde{p}_0$ , with  $p_2 < \tilde{p}_2$ . For  $0 \leq \xi_0 < \infty$ , if  $\tilde{v}(\xi_0, p_2) = \tilde{v}(\xi_0, \tilde{p}_2)$ , then (2.11) implies that  $\dot{\tilde{v}}(\xi_0, p_2) < \dot{\tilde{v}}(\xi_0, \tilde{p}_2)$ . Consider the function  $z(\xi) := \tilde{v}(\xi, \tilde{p}_2) - \tilde{v}(\xi, p_2)$ . Then  $z(0) = 0$ , and if  $z(\xi_0) = 0$  for some  $\xi_0 \geq 0$ , then  $\dot{z}(\xi_0) > 0$ . It follows that  $z(\xi) \geq 0$  for all  $\xi \geq 0$ . This implies the first inequality in (2.10).

If  $\xi$  is bounded, then it is trivial to choose  $C_\delta$  so that the second inequality in (2.10) holds true. Thus it suffices to prove this inequality for  $\xi$  sufficiently large. Set  $\theta_p(\xi, p) = \partial \tilde{v}(\xi, p) / \partial p$ . Rewriting (2.7) as  $\ddot{\tilde{v}}(\xi, p) = \tilde{b}(\tilde{v}, p)$  then differentiating with respect to  $p$  yields

$$-\varepsilon^2 \ddot{\theta}_p + b_u(0, u_0(0) + \tilde{v}) \theta_p = \tilde{v}, \quad \theta_p(0, p) = \theta_p(\infty, p) = 0. \quad (2.12)$$

But  $b_u(0, u_0(0)) = \gamma_0^2$ , so for  $\xi$  sufficiently large, say  $\xi \geq \xi_1$ , inequality (2.9) implies that  $b_u(0, u_0(0) + \tilde{v}) \geq \gamma_0^2/2 > 0$ . One can apply a maximum principle to (2.12) on  $[\xi_1, \infty)$  to complete the argument. ■

### 2.3 Sub- and super-solutions, first-order asymptotic expansion

Let  $v_0(x, p)$  and  $v_1(x)$  be defined by

$$\begin{aligned} -\varepsilon^2 v_0'' + b(0, u_0(0) + v_0) &= p v_0, & (2.13) \\ -\varepsilon^2 v_1'' + b_u(0, u_0(0) + v_0(x, 0)) v_1 &= -\frac{x}{\varepsilon} \frac{d}{dx} b(x, u_0(x) + t) \Big|_{x=0, t=v_0(x, 0)} & (2.14) \end{aligned}$$

with boundary conditions

$$v_0(0, p) = g_0 - u_0(0), \quad v_1(0) = 0, \quad v_0(\infty, p) = v_1(\infty) = 0.$$

From Lemma 2.3 we know that (2.13) and its boundary conditions have a solution when  $|p| \leq p_0$ . In [20, §2.3.1] it is stated that the linear problem (2.14) has a solution  $v_1(x)$  and that this solution satisfies the following analogue of (2.4):

$$|v_1^{(k)}(\xi)| \leq C_\delta e^{-(\gamma_0 - \delta)\xi} \quad (2.15)$$

for  $0 \leq \xi < \infty$  and  $k = 0, 1, \dots, 4$ . See [19] for details of the proof or [4] for an alternative argument.

We are now ready to construct sub- and super-solutions  $\alpha$  and  $\beta$  of (1.1). Set

$$\begin{aligned} \alpha(x) &= u_0(x) + [v_0(x, -p_2) + \varepsilon v_1(x)] - p_1, \\ \beta(x) &= u_0(x) + [v_0(x, p_2) + \varepsilon v_1(x)] + p_1, \end{aligned}$$

The values  $p_1$  and  $p_2$  in the definitions of  $\alpha(x)$  and  $\beta(x)$  are small positive numbers that will be chosen later and are typically  $o(N^{-1})$ ; thus for  $N$  sufficiently large,  $\alpha$  and  $\beta$  are well defined.

**Lemma 2.4.** *Assume that  $0 < p_1$  and  $0 < p_2 \leq p_0$  so that  $\alpha$  and  $\beta$  are well defined. Then  $\alpha(x) \leq u_0(x) + v_0(x, 0) + \varepsilon v_1(x) \leq \beta(x)$  for all  $x \in X$ .*

*Proof.* This follows from Lemma 2.3. ■

In fact  $u_0(x) + v_0(x, 0) + \varepsilon v_1(x)$  is a standard first-order asymptotic expansion of  $u(x)$ ; see [4]. It is slightly modified by using the small parameters  $p_1$  and  $p_2$  to construct sub- and super-solutions.

**Theorem 2.1.** *[4, Theorem 2.3][20, Theorem 2.4] Under our hypotheses, for sufficiently small  $\varepsilon$  there exists a unique solution  $u(x)$  of (1.1) in a neighbourhood of the zero-order asymptotic expansion  $u_0(x) + v_0(x, 0)$ . Furthermore,*

$$|u(x) - [u_0(x) + v_0(x, 0) + \varepsilon v_1(x)]| \leq C\varepsilon^2 \quad \text{for all } x \in X.$$

**Remark 2.3.** *Some of the results of this section appeared in [4, 17] but our results are more general and our presentation is more accessible. The proof of [17, Lemma 2.4] that  $\alpha$  and  $\beta$  are super- and sub-solutions of (1.1) is incomplete, as it leaves unclear how [17, (2.28)] can be obtained for large values of  $\eta$ .*

### 3 Analysis of the numerical method

#### 3.1 The discrete problem

For a given positive integer  $N$ , we denote by  $X^N$  an arbitrary mesh

$$0 = x_0 < x_1 < \cdots < x_{N-1} < x_N = 1,$$

with  $h_i = x_i - x_{i-1}$ , for  $i = 1, \dots, N$ , and  $\bar{h}_i = (h_i + h_{i+1})/2$ , for  $i = 1, \dots, N-1$ .

The 3-point central difference scheme used to solve (1.1) is

$$F^N(u^N)_i := -\varepsilon^2 \delta^2 u_i^N + b(x_i, u_i^N) = 0 \quad \text{for } i = 1, \dots, N-1, \quad u_0^N = g_0, \quad u_N^N = g_1, \quad (3.1)$$

where the computed solution is  $u^N = \{u_i^N\}_{i=0}^N$  and

$$\delta^2 u_i^N = \frac{1}{\bar{h}_i} \left( \frac{u_{i+1}^N - u_i^N}{h_{i+1}} - \frac{u_i^N - u_{i-1}^N}{h_i} \right)$$

is the standard difference approximation of  $u''(x_i)$ .

Is (3.1) guaranteed to have at least one solution  $u^N$ ? How accurate is this solution? Our answers to these questions uses the material of §3.2, which comes from [10].

#### 3.2 $Z$ -fields

An operator  $H : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{n+1}$  is a  $Z$ -field if for all  $i \neq j$  the mapping

$$x_j \mapsto (H(x_0, x_1, \dots, x_n))_i$$

is a monotonically decreasing function from  $\mathbb{R}$  to  $\mathbb{R}$  when  $x_0, \dots, x_{j-1}, x_{j+1}, \dots, x_n$  are fixed. If  $H$  is differentiable, then  $H$  is a  $Z$ -field if and only if its Jacobian matrix has non-negative off-diagonal entries.

**Remark 3.1.** *The mapping  $(x_0, x_1, \dots, x_{N-1}, x_N) \mapsto (g_0, F^N(u^N)_1, \dots, F^N(u^N)_{N-1}, g_1)$  is clearly a  $Z$ -field.*

We give the proof of the following unpublished result of Lorenz for completeness.

**Lemma 3.1.** *[10] Let  $H : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{n+1}$  be continuous and a  $Z$ -field. Let  $r \in \mathbb{R}^{n+1}$  be given. Assume that there exist  $\alpha, \beta \in \mathbb{R}^{n+1}$  such that  $\alpha \leq \beta$  and  $H\alpha \leq r \leq H\beta$ . (The inequalities are understood to hold true component-wise.) Then the equation  $Hy = r$  has a solution  $y \in \mathbb{R}^{n+1}$  with  $\alpha \leq y \leq \beta$ .*

Proof. If  $Hv \leq r$  and  $Hw \leq r$  for  $v, w \in \mathbb{R}^{n+1}$ , then on defining  $\zeta \in \mathbb{R}^{n+1}$  by  $\zeta_i = \max\{v_i, w_i\}$  for  $i = 0, \dots, n$ , one can see that  $H\zeta \leq r$ . Let  $S = \{v \in \mathbb{R}^{n+1} : Hv \leq r, \alpha \leq v \leq \beta\}$  be the set of all lower solutions that lie between  $\alpha$  and  $\beta$ . Then  $S$  is non-empty since  $\alpha \in S$ . Define  $y \in \mathbb{R}^{n+1}$  by  $y_i = \sup\{v_i : v \in S\}$  for  $i = 0, \dots, n$ . Clearly  $\alpha \leq y \leq \beta$  and  $Hy \leq r$ .

We claim that  $Hy = r$ . Fix  $i \in \{0, \dots, N\}$ . There are two cases. First, suppose that  $y_i = \beta_i$ . Then  $(Hy)_i \geq (H\beta)_i$  since  $y_k \leq \beta_k$  for all  $k \neq i$  and  $H$  is a  $Z$ -field. But  $H\beta \geq r$ , so we must have  $(Hy)_i = r_i$ . Second, suppose that  $y_i < \beta_i$ . If  $(Hy)_i < r_i$ , since  $H$  is a  $Z$ -field one can increase  $y_i$  slightly while retaining the properties that  $\alpha \leq y \leq \beta$  and  $Hy \leq r$ —but this contradicts the original definition of  $y$ . Thus  $(Hy)_i = r_i$  in this case also. ■

**Remark 3.2.** *The functions  $\alpha$  and  $\beta$  of Lemma 3.1 are called sub- and super- solutions of the discrete problem  $Hy = r$ .*

### 3.3 Properties of discrete sub- and super-solutions

Lemma 3.1 implies that one can show existence of a solution to (3.1) by proving that the mesh functions  $\alpha(x_i)$  and  $\beta(x_i)$  from §2.3 really are sub- and super-solutions in the sense of Remark 3.2, i.e., that  $\alpha(x_i) \leq \beta(x_i)$  and  $F^N \alpha_i \leq 0 \leq F^N \beta_i$  for all  $i$ .

We first prove a result that is valid on arbitrary meshes. Later, layer-adapted meshes will be used to control the truncation error term  $\varepsilon^2[\delta^2 \bar{w}_i - \bar{w}_i'']$  in (3.2).

**Lemma 3.2.** *Assume that  $0 < p_1$  and  $0 < p_2 \leq p_0$  so that  $\alpha$  and  $\beta$  are well defined. Let  $\bar{w}(x) := v_0(x, p_2) + \varepsilon v_1(x)$ . Then there exists constants  $C_1$  and  $C_2$  such that the choice  $p_2 = C_1 p_1$  yields*

$$F^N \beta_i \geq -\varepsilon^2[\delta^2 \bar{w}_i - \bar{w}_i''] + p_1 \gamma^2 - C_2(\varepsilon^2 + p_1^2) \quad (3.2)$$

for  $i = 1, \dots, N - 1$ .

*Proof.* Set  $\bar{v}_0(x) = v_0(x, p_2)$ . To make the presentation more readable, we sometimes write  $\bar{v}_0$  and  $u_0$  instead of  $\bar{v}_0(x)$  and  $u_0(x)$ . For  $i = 1, \dots, N - 1$ ,

$$F^N \beta_i = -\varepsilon^2 \delta^2 \beta_i + b(x_i, \beta_i) = -\varepsilon^2 u_0''(\theta_i) - \varepsilon^2[\delta^2 \bar{w}_i - \bar{w}_i''] + [-\varepsilon^2 \bar{w}''(x_i) + b(x_i, \beta(x_i))], \quad (3.3)$$

where  $\theta_i \in [x_{i-1}, x_{i+1}]$ . Now

$$\begin{aligned} -\varepsilon^2 \bar{w}''(x) + b(x, \beta(x)) &= -\varepsilon^2[\bar{v}_0'' + \varepsilon v_1''] + b(x, u_0 + \bar{v}_0 + \varepsilon v_1 + p_1) \\ &= p_2 \bar{v}_0 - \varepsilon^3 v_1'' + [b(x, u_0 + \bar{v}_0 + \varepsilon v_1 + p_1) - b(x, u_0 + \bar{v}_0)] \\ &\quad + [b(x, u_0 + \bar{v}_0) - b(0, u_0(0) + \bar{v}_0)]. \end{aligned} \quad (3.4)$$

The most important term here is

$$\begin{aligned}
b(x, u_0 + \bar{v}_0 + \varepsilon v_1 + p_1) - b(x, u_0 + \bar{v}_0) &\geq b_u(x, u_0 + \bar{v}_0)[\varepsilon v_1 + p_1] - C(\varepsilon^2 + p_1^2) \\
&\geq p_1[b_u(x, u_0) - C_1 \bar{v}_0] + \varepsilon v_1[b_u(0, u_0(0) + \bar{v}_0) - O(x)] - C(\varepsilon^2 + p_1^2) \\
&\geq p_1[\gamma^2 - C_1 \bar{v}_0] + \varepsilon v_1 b_u(0, u_0(0) + \bar{v}_0) - C(\varepsilon^2 + p_1^2),
\end{aligned}$$

where  $C_1$  and  $C$  are some constants, by (2.15). Note that  $b(x, u_0 + \bar{v}_0) - b(0, u_0(0) + \bar{v}_0) = F(x, \bar{v}_0) - F(0, \bar{v}_0)$ , where  $F(x, t) := b(x, u_0(x) + t)$ . Thus

$$F(x, t) - F(0, t) = xF_x(0, t) + (x^2/2)F_{xx}(\tilde{x}(t), t)$$

for some  $\tilde{x}$  between 0 and  $x$ . But  $F(x, 0) = 0$ , so  $F_{xx}(x, 0) = 0$  and  $|F_{xx}(x, t)| \leq C|t|$ . This implies that  $F(x, t) - F(0, t) \geq xF_x(0, t) - Cx^2|t|$ , whence

$$b(x, u_0 + \bar{v}_0) - b(0, u_0(0) + \bar{v}_0) \geq xF_x(0, \bar{v}_0) - Cx^2|\bar{v}_0| \geq xF_x(0, \bar{v}_0) - C\varepsilon^2,$$

where Lemma 2.3 was used to bound the final term.

Substituting these bounds into (3.4), one obtains

$$\begin{aligned}
-\varepsilon^2 w''(x) + b(x, \beta(x)) &\geq p_2 \bar{v}_0 - \varepsilon^3 v_1'' + p_1[\gamma^2 - C_1 \bar{v}_0] + \varepsilon v_1 b_u(0, u_0(0) + \bar{v}_0) \\
&\quad + [xF_x(0, \bar{v}_0) - C\varepsilon^2] - C(\varepsilon^2 + p_1^2).
\end{aligned}$$

Now (2.14) yields  $-\varepsilon^3 v_1'' + \varepsilon b_u(0, u_0(0) + v_0(x, 0))v_1 + xF_x(0, v_0(x, 0)) = 0$ , which implies that

$$-\varepsilon^3 v_1'' + \varepsilon b_u(0, u_0(0) + \bar{v}_0)v_1 + xF_x(0, \bar{v}_0) \geq -C\varepsilon p_2 (1+x/\varepsilon) \max_{0 \leq p \leq p_2} \frac{\partial v_0}{\partial p}(x, p) \geq -C\varepsilon p_2,$$

where we used Lemma 2.3. Hence

$$-\varepsilon^2 \bar{w}''(x) + b(x, \beta(x)) \geq p_1 \gamma^2 + [p_2 - C_1 p_1] \bar{v}_0 - C(\varepsilon^2 + p_1^2 + p_2^2).$$

Returning finally to (3.3), we now have

$$F^N \beta_i \geq -\varepsilon^2 [\delta^2 \bar{w}_i - \bar{w}_i''] + p_1 \gamma^2 + [p_2 - C_1 p_1] \bar{v}_0 - C(\varepsilon^2 + p_1^2 + p_2^2).$$

Choose  $p_2 = C_1 p_1$  to finish the proof.  $\blacksquare$

Naturally  $F^N \alpha$  satisfies an inequality analogous to (3.2).

The next Theorem sheds light on the properties required of the mesh and of  $p_1$  to ensure existence and accuracy of the discrete solution.

**Theorem 3.1.** *Assume that  $0 < p_1$  and  $0 < p_2 \leq p_0$  so that  $\alpha$  and  $\beta$  are well defined. Let  $w(x) = v_0(x, p) + \varepsilon v_1(x)$  and  $p_2 = C_1 p_1$  as in Lemma 3.2. Suppose that the mesh is such that*

$$|-\varepsilon^2[\delta^2 w_i - w_i'']| \leq p_1 \gamma^2 / 2 \quad \text{for } i = 1, \dots, N-1 \text{ and } |p| \leq p_0. \quad (3.5)$$

*Suppose also that  $C_2(\varepsilon^2 + p_1^2) \leq p_1 \gamma^2 / 2$ . Then there exists a solution  $u^N$  of (3.1) and a constant  $C$  such that for  $N$  sufficiently large,*

$$|u(x_i) - u_i^N| \leq C(p_1 + \varepsilon^2) \leq C(p_1 + N^{-2}) \quad \text{for } i = 0, \dots, N,$$

*where  $u$  is the solution of (1.1) guaranteed by Theorem 2.1.*

*Proof.* Lemma 2.4 shows that  $\alpha \leq \beta$ . For  $N$  sufficiently large, Lemma 3.2 and our hypotheses yield  $F^N \alpha_i \leq 0 \leq F^N \beta_i$  for  $i = 1, \dots, N-1$ . Define  $F^N \alpha_0 = g_0$  and  $F^N \alpha_N = g_1$  (and similarly for  $\beta$ ). Lemma 3.1 can now be invoked to give existence of a discrete solution  $u^N$  of (3.1), with moreover  $\alpha_i \leq u_i^N \leq \beta_i$  for  $i = 0, \dots, N$ . Combining this bound with Lemma 2.3, Lemma 2.4 and Theorem 2.1, we are done. ■

**Remark 3.3.** *Under the hypotheses of Theorem 3.1, each solution  $u$  described by Theorem 2.1 has a neighbourhood that contains a corresponding solution  $u^N$  of (3.1). Thus the discrete problem (3.1) may have multiple solutions.*

**Remark 3.4.** *The Z-field technique used here to prove existence of a discrete solution is much simpler than the topological degree theory used in [17].*

### 3.4 Existence and accuracy on layer-adapted meshes

The results of this Section also hold true on more general layer-adapted meshes such as those considered in [9], but for clarity we shall discuss only the two main representatives of this class: Bakhvalov and Shishkin meshes. These meshes are presented for the general case where the solution of (1.1) has boundary layers at both ends of the interval  $X$ . For convenience, we nevertheless continue our analysis under Assumption 2.

#### 3.4.1 Bakhvalov mesh

The Bakhvalov mesh is graded and is therefore more complicated than the piecewise uniform Shishkin mesh, but it often yields more accuracy in computed solutions. It first appeared in [2], and has (with its variants) been

combined with various difference schemes in numerous papers; see for example [8, 9, 16].

The mesh points  $x_i$  are defined as  $x_i = x(t_i)$ , with  $t_i = i/N$ , where the mesh-generating function  $x(t) \in C[0, 1]$  is defined by

$$x(t) = \begin{cases} -(2\varepsilon/\gamma) \ln(1 - 4t) & \text{for } 0 \leq t \leq \theta, \\ 1/2 - d(1/2 - t) & \text{for } \theta < t \leq 1/2, \end{cases} \quad (3.6)$$

with  $\theta = 1/4 - C_3\varepsilon$  for some positive constant  $C_3$ , and  $x(t) = 1 - x(1 - t)$  for  $1/2 < t \leq 1$ . Here  $d = [1/2 + (2\varepsilon/\gamma) \ln(1 - 4\theta)]/(1/2 - \theta)$  is chosen so that  $x(t)$  is continuous at  $t = \theta$ . This definition is valid only for  $\varepsilon \leq \min\{\gamma/8, 1/(8C_3)\}$ , which is not a practical restriction. For a certain choice of the arbitrary constant  $C_3$  one obtains the original Bakhvalov mesh, for which  $x(t) \in C^1[0, 1]$ .

**Lemma 3.3.** *Let  $w(x) = v_0(x, p) + \varepsilon v_1(x)$ , where  $|p| \leq p_0$ . Then on the Bakhvalov mesh, there exists a constant  $C$  such that the truncation error of the function  $w(x)$  satisfies*

$$|-\varepsilon^2[\delta^2 w_i - w_i'']| \leq CN^{-2} \quad \text{for } i = 1, \dots, N - 1.$$

*Proof.* By symmetry we need only consider the case  $t_i \leq 1/2$ . Using Taylor series expansions one can easily check that

$$|-\varepsilon^2[\delta^2 w_i - w_i'']| \leq C\varepsilon^2 \min \left\{ |h_{i+1} - h_i| M_i^{(3)} + \bar{h}_i^2 M_i^{(4)}, M_i^{(2)} \right\},$$

where

$$M_i^{(k)} := \max_{x \in [x_{i-1}, x_{i+1}]} |w^{(k)}(x)| \leq C\varepsilon^{-k} e^{-\gamma x_{i-1}/\varepsilon},$$

by Lemma 2.3 and (2.15). Hence

$$|-\varepsilon^2[\delta^2 w_i - w_i'']| \leq C \min \left\{ |h_{i+1} - h_i|/\varepsilon + \bar{h}_i^2/\varepsilon^2, 1 \right\} e^{-\gamma x_{i-1}/\varepsilon}.$$

Let  $\bar{i}$  satisfy  $t_{\bar{i}} \leq \theta < t_{\bar{i}+1}$ . Now  $1/4 - CN^{-1} \leq t_{\bar{i}-2} < \theta$ , so  $e^{-\gamma x_{\bar{i}-2}/\varepsilon} \leq CN^{-2}$ ; hence for all  $i \geq \bar{i} - 1$  we also have  $e^{-\gamma x_{i-1}/\varepsilon} \leq CN^{-2}$ . Thus  $|-\varepsilon^2[\delta^2 w_i - w_i'']| \leq CN^{-2}$  when  $i \geq \bar{i} - 1$ . For  $i \leq \bar{i} - 2$  we have  $e^{-\gamma x_{i-1}/\varepsilon} = (1 - 4t_{i-1})^2$ , while  $|h_{i+1} - h_i| \leq N^{-2} x''(t_{i+1})$  and  $\bar{h}_i \leq N^{-1} x'(t_{i+1})$ , since the functions  $x'(t) = \varepsilon(8/\gamma)(1 - 4t)^{-1}$  and  $x''(t) = \varepsilon(8/\gamma)(1 - 4t)^{-1}$  are increasing on  $[0, \theta]$ . Consequently in this case also,

$$|-\varepsilon^2[\delta^2 w_i - w_i'']| \leq CN^{-2} \left( \frac{1 - 4t_{i-1}}{1 - 4t_{i+1}} \right)^2 = CN^{-2} \left( 1 + \frac{4(t_{i+1} - t_{i-1})}{1 - 4t_{i+1}} \right)^2 \leq CN^{-2},$$

as  $1 - 4t_{i+1} \geq 1 - 4t_{\bar{i}-1} \geq 4N^{-1}$ . ■

Choosing  $p_1 = CN^{-2}$  in Theorem 3.1, we obtain the following result.

**Theorem 3.2.** *There exists a solution  $\{u_i^N\}_{i=0}^N$  of (3.1) on the Bakhvalov mesh (3.6) such that for  $N$  sufficiently large,*

$$|u(x_i) - u_i^N| \leq CN^{-2} \quad \text{for } i = 0, \dots, N.$$

### 3.4.2 Shishkin mesh

This mesh is discussed at length in [3, 11, 16]. It is constructed as follows. Let  $N$  be an even integer. Set

$$\sigma = \min \{(2\varepsilon/\gamma) \ln N, 1/4\}.$$

Divide the intervals  $[0, \sigma]$ ,  $[\sigma, 1 - \sigma]$  and  $[1 - \sigma, 1]$  into  $N/4$ ,  $N/2$  and  $N/4$  equidistant subintervals respectively. In practice one usually has  $\sigma \ll 1$ , so the mesh is coarse on  $[\sigma, 1 - \sigma]$  and fine otherwise.

Imitating Lemma 3.3 and choosing  $p_1 = CN^{-2} \ln^2 N$  in Theorem 3.1, we obtain the following result.

**Theorem 3.3.** *There exists a solution  $\{u_i^N\}_{i=0}^N$  of (3.1) on the Shishkin mesh such that for  $N$  sufficiently large,*

$$|u(x_i) - u_i^N| \leq CN^{-2} \ln^2 N \quad \text{for } i = 0, \dots, N.$$

See also [17] for a similar result.

## 4 Numerical results

Consider the following problem of Herceg [5]:

$$-\varepsilon^2 u'' + (u^2 + u - 0.75)(u^2 + u - 3.75) = 0 \quad \text{for } x \in X, \quad u(0) = u(1) = 0. \quad (4.1)$$

Here  $b_u(x, u) = (2u+1)(2u^2+2u-4.5)$  and the reduced problem  $b(x, u_0) = 0$  has four (constant) solutions:  $u_1 = -2.5$ ,  $u_2 = -1.5$ ,  $u_3 = 0.5$  and  $u_4 = 1.5$ , with

$$b_u(x, u_1) = -12, \quad b_u(x, u_2) = 6, \quad b_u(x, u_3) = -6 \quad \text{and} \quad b_u(x, u_4) = 12.$$

By (1.4a),  $u_1$  and  $u_3$  are not stable reduced solutions. A calculation shows that  $u_2$  and  $u_4$  satisfy all the conditions (1.4). We shall present numerical results only for the solution of (1.1) near  $u_2$ ; the results for the solution near  $u_4$  are similar.

To solve the discrete nonlinear problem we use Newton's method with the initial guess equal to  $u_2$  at all mesh nodes.

## 4.1 Bakhvalov mesh

Numerical results for the Bakhvalov mesh are given in Table 4.1.

Table 4.1: Bakhvalov mesh. Computational rates  $r$  in  $(N^{-1})^r$  and maximum nodal errors;  $u_2(x) = -1.5$ ,  $\gamma := \sqrt{b_u(u_2)}/1.3$ ,  $\theta = 1/4 - (2/\gamma)\varepsilon$

$N$	$\varepsilon = 1$	$\varepsilon = 10^{-1}$	$\varepsilon = 10^{-2}$	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-8}$	$\max_{\varepsilon}$
64	2.00	2.00	2.00	2.00	2.00	2.00
128	2.00	2.00	2.00	2.00	2.00	2.00
256	2.00	2.00	2.00	2.00	2.00	2.00
512	2.00	2.00	2.00	2.00	2.00	2.00
64	2.65e-5	1.74e-3	1.90e-3	1.90e-3	1.90e-3	1.90e-3
128	6.64e-6	4.36e-4	4.76e-4	4.76e-4	4.76e-4	4.76e-4
256	1.66e-6	1.09e-4	1.19e-4	1.19e-4	1.19e-4	1.19e-4
512	4.15e-7	2.73e-5	2.97e-5	2.97e-5	2.97e-5	2.97e-5
1024	1.04e-7	6.82e-6	7.42e-6	7.42e-6	7.42e-6	7.42e-6

The exact solution  $u(x)$  is unknown but we want to estimate errors in the computed solution and rates of convergence, so we assume that

$$u_i^N - u_i \approx C_4 N^{-r}$$

for some  $r > 0$ .

- Compute  $u^N$  and  $u^{2N}$  for the Bakhvalov meshes with  $N = 64, 128, \dots, 1024$ .

Then

$$u_i^{2N} - u_i \approx C_4 (2N)^{-r}.$$

- Hence

$$\|u^N - u\| \approx (1 - 2^{-r})^{-1} \|u^{2N} - u^N\|, \quad (4.2)$$

where  $\|\cdot\|$  denotes the discrete maximum norm, and

$$r = \log_2 \frac{\|u^N - u\|}{\|u^{2N} - u\|} \approx \log_2 \frac{\|u^N - u^{2N}\|}{\|u^{2N} - u^{4N}\|}.$$

This formula, which also appears in [3, p.108 and Chapter 8], enables us to compute approximate values of  $r$  by using known values from our computations; these values of  $r$  are presented in the upper part of Table 4.1.

- It is clear from our numerical results that the method yields  $r = 2$ . Consequently we set  $r = 2$  in (4.2) to compute the approximate value

of the errors  $\|u^N - u\|$ , which are presented in the lower part of Table 4.1.

Our results confirm the sharpness of the bound of Theorem 3.2.

## 4.2 Shishkin mesh

The numerical results for the Shishkin mesh are given in Table 4.2.

Table 4.2: Shishkin mesh. Computational rates  $r$  in  $(N^{-1} \ln N)^r$  and maximum nodal errors;  $u_2(x) = -1.5$ ,  $\sigma = \min\{(2.2/\sqrt{b_u(u_2)})\varepsilon \ln N, 1/4\}$

$N$	$\varepsilon = 1$	$\varepsilon = 10^{-1}$	$\varepsilon = 10^{-2}$	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-8}$	$\max_{\varepsilon}$
64	2.57	2.48	1.98	1.98	1.98	1.98
128	2.48	2.47	1.89	1.89	1.89	1.89
256	2.41	2.40	1.98	1.98	1.98	1.98
512	2.36	2.36	2.00	2.00	2.00	2.00
64	2.65e-5	2.73e-3	5.98e-3	5.98e-3	5.98e-3	5.98e-3
128	6.64e-6	7.17e-4	2.05e-3	2.05e-3	2.05e-3	2.05e-3
256	1.66e-6	1.80e-4	7.12e-4	7.12e-4	7.12e-4	7.12e-4
512	4.15e-7	4.53e-5	2.27e-4	2.27e-4	2.27e-4	2.27e-4
1024	1.04e-7	1.13e-5	7.02e-5	7.02e-5	7.02e-5	7.02e-5

The exact solution  $u(x)$  is unknown, so we assume that

$$u_i^N - u_i \approx C_5(h/\varepsilon)^r = C_6(N^{-1} \ln N)^r$$

and proceed as follows:

- Compute  $u^N$  for the Shishkin meshes with  $N = 64, 128, \dots, 1024$ .
- For each  $N$ , let  $\bar{u}^{2N}$  be the computed solution on the bisected Shishkin mesh (i.e., the transition point is the same as for  $u^N$  but each mesh interval is bisected). Then

$$\bar{u}_i^{2N} - u_i \approx C_5((h/2)/\varepsilon)^r = C_6(N^{-1} \ln N)^r 2^{-r}.$$

- Hence

$$\|u^N - u\| \approx (1 - 2^{-r})^{-1} \|\bar{u}^{2N} - u^N\| \quad (4.3)$$

and as in [17] we have

$$r = k^{-1}(N) \ln \frac{\|u^N - u\|}{\|u^{2N} - u\|} \approx k^{-1}(N) \ln \frac{\|u^N - \bar{u}^{2N}\|}{\|u^{2N} - \bar{u}^{4N}\|}, \quad \text{where } k(N) = \ln \frac{2 \log_2 N}{\log_2 N + 1}.$$

This formula enables us to compute approximate values of  $r$ , which are presented in the upper part of Table 4.2.

- It is clear that  $r = 2$  for this numerical method. Consequently we set  $r = 2$  in (4.3) to compute the approximate value of the errors  $\|u^N - u\|$ , which are presented in the lower part of Table 4.2.

Our results confirm the sharpness of the bound of Theorem 3.3.

## A Appendix: Alternative proof of exponential decay of $\tilde{v}$

By Assumption 2,  $\tilde{v}(0) > 0$ . Then  $\tilde{V} = -\sqrt{2 \int_0^{\tilde{v}} \tilde{b}(s) ds}$ . While  $\tilde{v}(\xi) > 0$  we have  $\tilde{V}(\xi) < 0$ , so  $\tilde{v}(\xi)$  is a decreasing function.

Let  $\delta \in (0, \gamma_0)$  be arbitrary but fixed. Since  $\tilde{b}(0) = 0$ , there exists  $s_0 > 0$  such that

$$(\gamma_0 - \delta)^2 s \leq \tilde{b}(s) \leq (\gamma_0 + \delta)^2 s \quad \text{for } 0 \leq s \leq s_0. \quad (\text{A.1})$$

If  $\tilde{v}(\xi) > s_0$  for all  $\xi > 0$ , then  $\tilde{V}(\xi) \leq -C_7$  for some positive constant  $C_7$ , which is impossible. Thus  $\tilde{v}(\xi_0) \leq s_0$  for some  $\xi_0$ , which yields  $\tilde{v}(\xi) \leq s_0$  for all  $\xi \geq \xi_0$ .

Fix  $\xi_0$  with  $0 < \tilde{v}(\xi_0) \leq s_0$ . When  $\tilde{v}(\xi) \in (0, s_0]$ , from (2.5) we have

$$-\sqrt{2 \int_0^{\tilde{v}(\xi)} (\gamma_0 + \delta)^2 s ds} \leq \tilde{V}(\xi) \leq -\sqrt{2 \int_0^{\tilde{v}(\xi)} (\gamma_0 - \delta)^2 s ds},$$

i.e.,

$$-(\gamma_0 + \delta)\tilde{v}(\xi) \leq \tilde{V}(\xi) \leq -(\gamma_0 - \delta)\tilde{v}(\xi). \quad (\text{A.2})$$

Rewriting this as  $-(\gamma_0 + \delta) \leq \tilde{v}'(\xi)/\tilde{v}(\xi) \leq -(\gamma_0 - \delta)$  and integrating from  $\xi_0$  to  $\xi \geq \xi_0$  yields, after some rearranging,

$$\tilde{v}(\xi_0)e^{(\gamma_0 + \delta)\xi_0}e^{-(\gamma_0 + \delta)\xi} \leq \tilde{v}(\xi) \leq \tilde{v}(\xi_0)e^{(\gamma_0 - \delta)\xi_0}e^{-(\gamma_0 - \delta)\xi}. \quad (\text{A.3})$$

In particular this implies that  $\tilde{v}(\xi) \rightarrow 0$  as  $\xi \rightarrow \infty$ .

The above analysis shows that there exists a positive constant  $C_\delta$  such that

$$C_\delta^{-1}e^{-(\gamma_0 + \delta)\xi} \leq \tilde{v}(\xi) \leq C_\delta e^{-(\gamma_0 - \delta)\xi} \quad \text{for } 0 < \xi < \infty, \quad (\text{A.4})$$

as (A.4) is equivalent to  $s_0 \leq \tilde{v}(\xi) \leq \tilde{v}(0)$  when  $0 \leq \xi \leq \xi_0$ .

Using (A.2), we see that  $|\tilde{V}(\xi)|$  satisfies a similar inequality.

## References

- [1] C.M. D'Annunzio, *Numerical analysis of a singular perturbation problem with multiple solutions*, Ph.D. Dissertation, University of Maryland at College Park, 1986 (unpublished).
- [2] N. S. Bakhvalov, On the optimization of methods for solving boundary value problems with boundary layers, *Zh. Vychisl. Mat. Mat. Fis.* 9 (1969), 841–859 (in Russian).
- [3] P. A. Farrell, A. F. Hegarty, J. J. H. Miller, E. O'Riordan and G. I. Shishkin, *Robust Computational Techniques for Boundary Layers*, Chapman & Hall / CRC, Boca Raton, 2000.
- [4] P.C. Fife, Semilinear elliptic boundary value problems with small parameters, *Arch. Rational Mech. Anal.* 52 (1973), 205–232.
- [5] D. Herceg, Uniform fourth order difference scheme for a singular perturbation problem, *Numer. Math.* 56 (1990), 675–693.
- [6] F.A. Howes, *Boundary-interior layer interactions in nonlinear singular perturbation theory*, *Mem. Amer. Math. Soc.* 15 (1978), no. 203.
- [7] W.L. Kath, Slowly varying phase planes and boundary-layer theory, *Stud. Appl. Math.* 72 (1985), 221–239.
- [8] N. Kopteva, Uniform pointwise convergence of difference schemes for convection-diffusion problems on layer-adapted meshes, *Computing* 66 (2001), 179–197.
- [9] T. Linß, Sufficient conditions for uniform convergence on layer-adapted grids, *Appl. Numer. Math.* 37 (2001), 241–255.
- [10] J. Lorenz, Nonlinear singular perturbation problems and the Enquist-Osher scheme, Report 8115, Mathematical Institute, Catholic University of Nijmegen, 1981 (unpublished).
- [11] J. J. H. Miller, E. O'Riordan and G. I. Shishkin, *Fitted Numerical Methods for Singular Perturbation Problems: Error Estimates in the Maximum Norm for Linear Problems in One and Two Dimensions*, World Scientific, Singapore, 1996.
- [12] J.D. Murray, *Mathematical Biology*. Second corrected edition, Springer-Verlag, Berlin, 1993.

- [13] N.N. Nefedov, The method of differential inequalities for some classes of nonlinear singularly perturbed problems with internal layers, *Differ. Equ.* 31 (1995), 1077–1085.
- [14] R.E. O’Malley, *Singular perturbation methods for ordinary differential equations*, Springer–Verlag, New York, 1991.
- [15] C. Robinson, *Dynamical systems. Stability, symbolic dynamics, and chaos*. CRC Press, Boca Raton, 1995.
- [16] H.-G. Roos, M. Stynes and L. Tobiska, *Numerical Methods for Singularly Perturbed Differential Equations*, Volume 24, Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 1996.
- [17] G. Sun and M. Stynes, A uniformly convergent method for a singularly perturbed semilinear reaction–diffusion problem with multiple solutions, *Math. Comp.* 65 (1996), 1085–1109.
- [18] L. Tobiska, *Die asymptotische Lösung singular gestörter elliptischer Randwertaufgaben*, 2nd Doctoral Thesis, Technische Hochschule Otto von Guericke, Magdeburg, 1983.
- [19] A.B. Vasil’eva and V.F. Butuzov, *Asimptoticheskie razlozheniya reshenii singularno-vozmushchennykh uravnenii*. (Russian) [Asymptotic expansions of the solutions of singularly perturbed equations] Izdat. Nauka, Moscow, 1973.
- [20] A.B. Vasil’eva, V.F. Butuzov and L.V. Kalachev, *The boundary function method for singular perturbation problems*. With a foreword by Robert E. O’Malley, Jr. SIAM Studies in Applied Mathematics, 14. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1995.