Condition and conditioning, stability and stabilization

Eindhoven University of Technology
Department of Mathematics and Computing Science

RANA 89-01

January 1989

CONDITION AND CONDITIONING - STABILITY AND STABILIZATION

by

R.M.M. Mattheij

Reports on Applied and Numerical Analysis
Department of Mathematics and Computing Science
Eindhoven University of Technology
P.O. Box 513
5600 MB Eindhoven
The Netherlands

CONDITIONS AND CONDITIONING, STABILITY AND STABILIZATION [1)]

R.M.M. Mattheij
Department of Mathematics and Computing Science
Eindhoven University of Technology
P.O. Box 513
5600 MB Eindhoven
The Netherlands

ABSTRACT

An overview is given of several concepts that are useful when dealing with numerical BVP. In particular a generalisation of the well-conditioning concept for nonlinear problems is considered. This is done to be able to investigate a class of second order scalar nonlinear problems. A detailed study of the solution structure is made for this class. The results are applied to two specific problems (Korteweg-de Vries and Burgers) and a way is indicated to stabilize these (ill-posed) BVP.

1. INTRODUCTION

For a long time research on numerical aspects of boundary value problems (BVP) has been concentrated mostly on finding better and more sophisticated methods for solving them. It seems, however, that the subject has grown up now, as understanding of some fundamental aspects gets more and more attention. One such basic achievement is the knowledge that the presence of boundary conditions (BC) does not so much make a BVP more complicated than an initial value problem (IVP), as the former is necessary global in character, but rather that BC provide conditions that are crucially intertwined with the structure of the underlying ODE.

Consider

[1)] This paper is in final form and will not be submitted for publication elsewhere.

$$(1.1) \qquad \frac{d\mathbf{x}}{dt} = \mathbf{f}(t,x)\,, \quad \alpha \le t \le \beta$$

and the BC

$$(1.2) \qquad \mathbf{g}(\mathbf{x}(\alpha_1),\, \mathbf{x}(\alpha_2),\, \cdots,\, \mathbf{x}(\alpha_m)) = 0\,,$$

where **f** is an appropriate n-vector field and the *switch* points $\alpha_1, \cdots, \alpha_m$ are such that $\alpha_1 = \alpha$, $\alpha_m = \beta$ and $\alpha_j < \alpha_{j+1}$, $1 < j < m-1$. In the literature one usually takes $m = 2$ (or in principle reduces the considerations to this case). In case the vector field is linear various authors cf. [3,6,8] have shown that the two point BC effectively control the modes of (1.1); more specifically, if we assume a *dichotomy*, then decaying modes are controlled by initial conditions and increasing modes by terminal conditions, that is, assuming the problem is not sensitive to perturbations. Recently this idea was extended for the multipoint case, giving rise to a nontrivial generalisation of the dichotomy aspect [4,5].

In this paper we like to deal, more generally, with nonlinear problems and to investigate the sensitivity of the solution. To be not over-ambitious we shall only make an attempt to discuss a particular class of second order scalar ODE, which nevertheless contains many relevant problems from practice.

Our approach is based on quasi linearization and examining the solution space of the linearized problem. As will become clear in section 5 the underlying structure can be quite whimsical and sometimes appropriate analytical transformations or additional BC must be provided to "stabilize" the problem. In section 4 it is defined what we mean by stability. Before that we first introduce a more general concept of conditioning (section 2) and recall the notions of dichotomy and polychotomy in section 3.

2. CONDITIONING

An important practical question in dealing with BVP is the sensitivity of the solution with respect to perturbations in the data. From a numerical point of view the most important errors are those due to the discretisation of the ODE and the consequently imprecise solution of the BC (note that the latter fact makes BVP essentially more difficult than IVP). Hence it makes sense to consider the "perturbed" ODE

$$(2.1) \qquad \frac{d\bar{\mathbf{x}}}{dt} = \mathbf{f}(t,\bar{\mathbf{x}}) + \mathbf{d}(t)\,, \quad \alpha \le t \le \beta$$

(where d may depend on x) and the BC

$$(2.2) \qquad \mathbf{g}(\bar{\mathbf{x}}(\alpha_1),\, \cdots,\, \bar{\mathbf{x}}(\alpha_m)) = \mathbf{b}$$

where the "source term" **d** and the "residual" **b** are sufficiently small, to ensure $\bar{\mathbf{x}}$ to be an isolated solution of (2.1), (2.2), as we assume **x** is of (1.1) and (1.2).

Definition 2.3. Let ε_0 sufficiently small such that for all $\mathbf{d}$, $\mathbf{b}$ with $\| \mathbf{d}(\cdot) \|^{\dagger} \le \varepsilon_0$, $\| \mathbf{b} \| \le \varepsilon_0$, (2.1), (2.2) has an isolated solution, then *CN*, defined as the infimum of positive numbers κ for which $\| \mathbf{x}(\cdot) - \tilde{\mathbf{x}}(\cdot) \| \le \kappa\varepsilon$ for all $\mathbf{d}$ and $\mathbf{b}$ with $\| \mathbf{d}(\cdot) \| \le \varepsilon \le \varepsilon_0$, $\| \mathbf{b} \| \le \varepsilon \le \varepsilon_0$, is called the *condition number.*

If $CN\varepsilon$ is small we may use linearization in order to assess this conditioning. We obtain

$$\frac{d}{dt}(\tilde{\mathbf{x}} - \mathbf{x}) = J(\tilde{\mathbf{x}} - \mathbf{x}) + \mathbf{d}(t), \tag{2.4}$$

where J denotes the Jacobian at a suitable point not being far from (t,x)

$$\sum_{i=1}^{m} M_i \, [\tilde{\mathbf{x}}(\alpha_i) - \mathbf{x}(\alpha_i)] = \mathbf{b}, \tag{2.5}$$

where $M_i \approx \dfrac{\partial \mathbf{g}}{\partial \mathbf{x}(\alpha_i)}$.

A theory for dealing with this linear(ized) problem can be found in [4.9] and in particular for the multipoint case in [5]. We only need some relevant conclusions, which can be stated in terms of fundamental solutions. So, let Φ be a fundamental solution of (2.4), i.e.

$$\frac{d}{dt}\Phi = J\,\Phi. \tag{2.6}$$

It is not restrictive to assume that

$$\sum_{i=1}^{m} M_i \, \Phi(\alpha_i) = I. \tag{2.7}$$

This induces a Green's function $G(t,s)$ satisfying

$$G(t,s) = \Phi(t) \sum_{j=1}^{i} M_j \, \Phi(\alpha_j) \, \Phi^{-1}(s), \; \alpha_i < s < \alpha_{i+1}, \; t > s \tag{2.8a}$$

$$G(t,s) = -\Phi(t) \sum_{j=i+1}^{m} M_j \, \Phi(\alpha_j) \, \Phi^{-1}(s), \; \alpha_i < s < \alpha_{i+1}, \; t < s. \tag{2.8b}$$

So the "error" $\tilde{\mathbf{x}} - \mathbf{x}$ can be written as

$$\tilde{\mathbf{x}}(t) - \mathbf{x}(t) = \Phi(t)\,\mathbf{b} + \int_{\alpha}^{\beta} G(t,s)\,\mathbf{d}(s)\,ds. \tag{2.9}$$

Therefore we expect a sharp bound for *CN* to be given by

$$CN \le \| \Phi(\cdot) \| + (\beta - \alpha) \, \| G(.\,,.) \|. \tag{2.10}$$

Remark: Note that the factor $(\beta - \alpha)$ naturally disappears if $\mathbf{d}(s)$ is an appropriately scaled local discretisation error.

† Here $\| \cdot \|$ denotes a Hölder norm and $\| \mathbf{d}(\cdot) \|$ an associated function space norm.

3. POLYCHOTOMY

Although nonlinear BC do occur in practice, linear BC are more common. Even more common are so called *separated* BC, a situation we shall study in this section. In such a case we have effectively the BC

$$(3.1)\qquad \mathbf{g}(\mathbf{x}(\alpha_1), \cdots, \mathbf{x}(\alpha_m)) = \sum_{i=1}^{m} M_i \mathbf{x}(\alpha_i) - \mathbf{b} ,$$

where $\mathbf{b}$ is some n-vector, M_i are some $n \times n$ matrices, such that $\sum_{i=1}^{m} \text{rank}(M_i) = n$ and moreover such that M_i has systematically zero rows but for precisely $\text{rank}(M_i)$.

This BC naturally induces projections P_i by

$$(3.2)\qquad P_i := M_i \Phi(\alpha_i)$$

(where Φ is normalized as in (2.7)). In particular, given a bound CN on the Green's function $G(t,s)$, we derive from (2.8)

$$(3.3a)\qquad \| \Phi(t) \sum_{j=1}^{i} P_j \Phi^{-1}(s) \| \le CN , \quad \alpha_i < s < \alpha_{i+1} , \quad t > s ,$$

$$(3.3b)\qquad \| \Phi(t) \sum_{j=i+1}^{m} P_j \Phi^{-1}(s) \| \le CN , \quad \alpha_i < s < \alpha_{i+1} , \quad t < s .$$

Because of the nice form of the matrices P_j, we see that $\sum_{j=1}^{i} P_j$ is a projection and so is $\sum_{j=i+1}^{m} P_j = I - \sum_{j=1}^{i} P_j$. Apparently for each i, (3.3) describes a suitable partitioning of the fundamental solution with respect to the growth. If $m = 2$, we call Φ *dichotomic* (with *dichotomy constant CN*), otherwise *polychotomic* (with a *polychotomic constant CN*).

Remark: Dichotomy means that we can find suitable fundamental modes that either do not grow or do not decay. Polychotomy is more complicated. Either a fundamental mode grows, or it decays, or it grows first and decays then (only after passing a switch point α_i!).

Given a dichotomic (or polychotomic) fundamental solution and a well-posed problem (i.e. such that $\| \Phi(\cdot) \|$ is not unacceptably large), well-conditioning can easily be shown. Conversely, given well-conditioning (i.e. CN is "not large"), we apparently show dichotomy or polychotomy with a "not large" constant.

Using the linearization, carried out in section 2, it only seems reasonable to expect well-conditioning of a nonlinear problem (i.e. with $\mathbf{f}$ nonlinear) to be related to the growth properties of the fundamental solution of the linear system. (2.6). In dealing with a process for computing the solution x of (1.1), (1.2) one necessarily linearizes the problem, thus solving a set of linear problems in an iterative setting. At each iteration step one should hope the linear problem under consideration to be fairly well-conditioned (say such that rounding errors do not blur the result more than an amount allowed by the required tolerance). From the foregoing it follows that this can only be the case if there is an appropriate dichotomy or polychotomy (where the constant is related to the well-conditioning as in (3.3)). In other words lack of sufficient dichotomy say, implies a nearly singular Jacobian in a Newton process. The next, somewhat artificial, example shows that such a phenomenon can occur quit unexpectedly.

Example 3.4. Consider the ODE

(3.5a) $$u'' + (u^2)' = 0, \quad 0 < t < \infty$$

(3.5b) $$u(0) = 1, \quad u(\infty) = 0.$$

By writing $\mathbf{x} = (u, u')^T = (u, v)^T$, we obtain

(3.6a) $$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}) = \begin{bmatrix} v \\ -2uv \end{bmatrix},$$

(3.6b) $$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \mathbf{x}(0) + \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \mathbf{x}(\infty) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

(the latter being a separated two point BC).
It can simply be checked that the solution of (3.5), $\hat{u}$ say, is given by

(3.7) $$\hat{u}(t) = \frac{1}{1+t},$$

so $\hat{v}(t) := \frac{d}{dt}\hat{u}(t) = -\frac{1}{(1+t)^2}$.
Hence if we linearize (3.6b) we find the ODE

(3.8a) $$\frac{d\hat{\mathbf{y}}}{dt} = \hat{J}\,\hat{\mathbf{y}},$$

with

(3.8b) $$\hat{J} := \begin{bmatrix} 0 & 1 \\ -2\hat{v} & -2\hat{u} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \frac{1}{(1+t)^2} & -\frac{1}{1+t} \end{bmatrix}.$$

Simple calculation shows that (3.8) has a fundamental solution

(3.9) $$\hat{\Phi}(t) = \begin{bmatrix} 1+t & \frac{1}{(1+t)^2} \\ 1 & \frac{-2}{(1+t)^3} \end{bmatrix}.$$

Hence, defining (3.8) on the interval $(0, L)$, we see that we have a dichotomic solution space with one increasing and one decreasing mode and with a dichotomy constant $O(1)$, if $L \to \infty$. This in in agreement with the (well-) conditioning of the problem with ODE (3.8) and BC (3.6b).

Given the BC (3.6) the following solution would be a very acceptable candidate as a first approximation (for starting a Newton process say):

$$\text{(3.10)} \qquad \begin{bmatrix} \tilde{u} \\ \tilde{v} \end{bmatrix} = \begin{bmatrix} \dfrac{1}{1+t} - \dfrac{1}{1+t^3} \\ \dfrac{3t+6t^2}{(1+t)(1+t^3)} - \dfrac{1}{(1+t)^2} \end{bmatrix} .$$

Note that $(\tilde{u}(0),\tilde{v}(0))^T = (\hat{u}(0),\hat{v}(0))^T$ and $(\tilde{u}(\infty),\tilde{v}(\infty))^T = (\hat{u}(\infty),\hat{v}(\infty))^T$. If we linearize (3.6a) at $(\tilde{u},\tilde{v})^T$ we find an ODE similar to (3.8), i.e.

$$\text{(3.11a)} \qquad \frac{d\tilde{y}}{dt} = \tilde{J}\,\tilde{y} ,$$

with

$$\text{(3.11b)} \qquad \tilde{J} := \begin{bmatrix} 0 & 1 \\ -2\tilde{v} & -2\tilde{u} \end{bmatrix} ,$$

where $\tilde{u}$, $\tilde{v}$ are defined by (3.10). Straightforward calculation shows that (3.11) has a fundamental solution

$$\text{(3.12)} \qquad \tilde{\Phi}(t) = \begin{bmatrix} \dfrac{1+t}{1+t^3} & \dfrac{1}{(1+t)^2\,(1+t^3)} \\ \dfrac{1-3t^2-3t^3}{(1+t^3)^2} & \dfrac{-2-3t^2-5t^3}{(1+t)^3\,(1+t^3)^2} \end{bmatrix} .$$

Therefore we see that (3.11) has decaying modes only. In order to have *a well-posed* "BVP" we need to specify *two independent initial conditions*! Thus we can conclude that a linearization of (3.6) at (3.10) leads to an ill-conditioned problem. In fact it is not difficult to see that a sharp bound for the Green's function on the interval $(0, L\,]$ is $O(L^2)$.

4. STABILITY AND STABILIZATION

Like the notion of well-conditioning, usage of the word *stability* enjoys a widespread popularity among numerical analysts. The concepts are in fact related, and, having defined a condition number in section 2, we can precisely define what we mean by stability. Intuitively, stability of a process indicates that we can trust the results, only if we have been willing to invest sufficiently. The two control parameters to measure this are the required accuracy tolerance, TOL, and the (given) machine constant, ε_M. If we associate with this process (e.g. an algorithm) a condition number *CN* then we can say

Definition 4.1. Given a positive number ε_M, a process is called *G-acceptable* if $CN\ \varepsilon_M \le G$.

Definition 4.2. Given a tolerance TOL for the required accuracy, a process is called *stable* if it is

G-acceptable with $G \leq$ TOL.

Returning to the example in the previous section we see that linearization of (3.6) in the neighbourhood of the exact solution and truncating the interval to $[0, L]$ is G-acceptable with $G \approx \varepsilon_M$ only if this neighbourhood is small enough. If we e.g. linearize at $(\tilde{u},\tilde{v})^T$, see (3.10), then this together with the BC (3.6b) will be $\approx L^2\, \varepsilon_M$ acceptable. For any finite tolerance however, we can find an L such that we may call this linearization *unstable.*

If we examine Example 3.4 with the choices $(\hat{u},\hat{v})^T$ and $(\tilde{u},\tilde{v})^T$ more closely then it appears that the afore mentioned instability is caused by the lack of appropriate BC. This leads us to try to *stabilize* the linearization by providing a sufficient BC. There may be various possibilities, but a very simple one is the following: Consider

$$(4.3)\qquad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{x}(0) + \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \mathbf{x}(\infty) = \begin{bmatrix} 1 \\ -1 \end{bmatrix} .$$

Note that we have just added information about the derivative of the solution u of (3.5a) and that (4.3) contains a full rank set of initial conditions for an "initial value stable" problem (and likewise for a "terminal value stable problem", though this is a less likely situation to occur).
In order to assess the stability of this choice we first compute the solution $\hat{y}$ satisfying (3.8) and

$$(4.4)\qquad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \hat{\mathbf{y}}(0) + \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \hat{\mathbf{y}}(L) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$(4.5a)\qquad \hat{\mathbf{y}}(t) = \hat{\Phi}(t)\, \hat{Q}_L^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} =: \Phi_1(t) \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

where

$$(4.5b)\qquad \hat{Q}_L = \begin{bmatrix} 2 & -1 + \dfrac{2}{(1+L)^3} \\ \\ 2+L & -2 + \dfrac{1}{(1+L)^2} \end{bmatrix} .$$

Hence:

$$(4.5c)\qquad \Phi_1(t) = \hat{\Phi}(t) \left\{ \begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix} + R(t) \right\},\quad \| R(t) \| = O\left(\frac{1}{L}\right)$$

which shows that $\| \Phi_1(t) \| = O(1)$, uniformly in t and L, $t \leq L$.

Since we have a dichotomic solution space a similar bound on $\| G(.,.) \|$ follows from [4, Thm 5.2].

Let us now consider the linearization using the choice $(\tilde{u},\tilde{v})^T$ and the BC (4.4) m.m. We then obtain

(4.6a) $$\tilde{y}(t) = \tilde{\Phi}(t)\, \tilde{Q}_L^{-1} \begin{bmatrix} 1 \\ -1 \end{bmatrix} =: \Phi_2(t) \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

where

(4.6b) $$\tilde{Q}_L = \begin{bmatrix} 1 + \dfrac{1-3L^2-3L^3}{(1+L^3)^2} & 1 - \dfrac{2+3L^2+5L^3}{(1+L)^3\,(1+L^3)^2} \\ 1 + \dfrac{1+L}{1+L^3} & -2 + \dfrac{1}{(1+L)^2\,(1+L^3)} \end{bmatrix}.$$

Hence

(4.6c) $$\Phi_2(t) = \tilde{\Phi}(t) \left\{ \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} \end{bmatrix} + S(t) \right\}, \text{ where } \| S(t) \| = O\left(\frac{1}{L^2}\right).$$

which gives that $\| \Phi_2(t) \| = O(1)$, uniformly in t and L, $t \le L$. Since the terminal condition does not play a role at all, we conclude that we can effectively say that $G(t,s) = \Phi(t)\, \Phi^{-1}(s)$, which is easily seen to be uniformly bounded for $t > s$.

For both linearizations we thus have shown that the condition numbers when considering (4.3) are $O(1)$ and so they are G-acceptable for G up to $\approx \varepsilon_M$, the machine accuracy; i.e. for $\mathrm{TOL} \ge \varepsilon_M$ we have a stable problem.

We are not aware of any general recipe for stabilizing a process. However, for two point BVP it is likely that providing a BC consisting of full rank matrices is definitely a way to explore. We like to emphasize that this kind of stabilization only makes sense when there is a dichotomy (which includes the cases $P_1 = 0$ and $P_1 = I$, $m = 2$ in (3.3)).

In the next section we consider a class of problems with a more difficult solution structure.

5. A CLASS OF NONDICHOTOMIC PROBLEMS

Consider the second order scalar ODE

(5.1) $$u'' = p(t,u,u'), \quad -\infty < t < \infty,$$

where the solution u is such that

(5.2) $$u(\pm\infty) = 0.$$

For clarity of notation let us again denote the exact solution of (5.1), (5.2) by $\hat{u}$ and its derivative by $\hat{v}$. By linearizing (5.1) at this solution we find the ODE

(5.3) $$u'' = \frac{\partial p}{\partial u'}(t,\hat{u},\hat{v})\, u' + \frac{\partial p}{\partial u}(t,\hat{u},\hat{v})\, u.$$

We like to estimate suitable fundamental modes of (5.3) in order to assess conditioning etc. of the resulting linearized BVP as we did in section 3. First we give some preliminary results dealing with growth properties on a semi-infinite interval. Let us denote

(5.4a) $$a(t) := \frac{\partial p}{\partial u'}(t,\hat{u},\hat{v})$$

(5.4b) $$b(t) = \frac{\partial p}{\partial u}(t,\hat{u},\hat{v}).$$

Since our results are related to and/or based on positivity we shall assume the following

Assumption 5.5. Let $b(t) > 0$ for $t \le t_1 \le 0$ and $t \ge t_2 \ge 0$.

We can associate to (5.3) the equation

(5.6) $$\lambda^2 = a(t)\,\lambda + b(t).$$

Because of assumption 5.5 both roots of (5.6) are real and nonzero for $t \le t_1$, $t \ge t_2$. Hence we may assume that there is a continuous $\lambda_1(t)$, with $\lambda_1(t) > 0$ and a continuous $\lambda_2(t)$, with $\lambda_2(t) < 0$ for all $t \le t_1$ and likewise for $t \ge t_2$. We shall refer to $\lambda_1(t)$ and $\lambda_2(t)$ as the *local eigenvalues* of (5.3).

Let us now recall a well-known monotonicity result (cf. [12, CH I])

Lemma 5.7. *Let u be a solution of (5.3) and let w be such that $w'' - aw' - bw \le 0$ on (γ, δ) for some γ, δ $(\gamma < \delta)$. If $u(\gamma) \le w(\gamma)$ and $u(\delta) \le w(\delta)$ then $u(t) \le w(t)$, for $t \in (\gamma, \delta)$.* □

Then we can show:

Property 5.8. *Let t_2 be as before. Define $\lambda_1 := \inf_{t \ge t_2} \lambda_1(t)$ and $\lambda_2 := \sup_{t \ge t_2} \lambda_2(t)$. Then (5.3) has solutions z_1, z_2 defined on (t_2, ∞) such that for t, $s \ge t_2$*

$$0 < \frac{z_1(t)}{z_1(s)} \le \exp\left[\lambda_1(t-s)\right], \quad t < s$$

$$0 < \frac{z_2(t)}{z_2(s)} \le \exp\left[\lambda_2(t-s)\right], \quad t > s$$

Proof: Define $w_1(\gamma,\delta,t) := \exp[\lambda_1(t-t_2)]$.
Then $w_1'' - aw_1' - bw_1 = [\lambda_1^2 - a\lambda_1 - b]\exp[\lambda_1(t-t_2)]$.
Since $\lambda_2(t) < \lambda_1 \le \lambda_1(t)$ we thus have $w_1'' - aw_1' - bw_1 \le 0$. If we define a solution $u_1(\gamma,\delta,t)$ of (5.3) by $u_1(\gamma,\delta,\gamma) = w_1(\gamma,\delta,\gamma)$, $u_1(\gamma,\delta,\delta) = w_1(\gamma,\delta,\delta)$, then we find from Lemma 5.7 that $u_1(\gamma,\delta,t) \le w_1(\gamma,\delta,t)$, $t \in (\gamma,\delta)$. Hence

(5.8a) $$\frac{u_1(\gamma,\delta,t)}{u_1(\gamma,\delta,\delta)} \leq \exp[\lambda_1(t-\delta)].$$

By taking $\gamma = t_2$ and letting $\delta \to \infty$ we obtain the existence of a solution $z_1(t)$ with $z_1(t_2) = 1$, which is bounded away from zero. At any point s, consider the scaled solutions $\tilde{u}_1(t_2,\delta,t) := z_1(t) \dfrac{w_1(t_2,\delta,s)}{\tilde{u}_1(t_2,\delta,s)}$, $\delta > s$. From the monotonicity it follows that $\tilde{u}_1(t_2,\delta,t) < z_1(t)$, $t < s$. A limit argument then gives the estimate for z_1.

The proof for z_2 is analogous and will be omitted. □

Corollary 5.9. *For the solutions z_1, z_2 in Property 5.8 we also have*

$$\frac{d}{dt} z_1(t) \geq \lambda_1 z_1(t)$$

$$\frac{d}{dt} z_2(t) \leq \lambda_2 z_2(t).$$

Proof: Let $h > 0$. $\dfrac{z_1(t+h)-z_1(t)}{z_1(t+h} = 1 - \dfrac{z_1(t)}{z_1(t+h)} \geq 1 - e^{-\lambda_1 h} \geq +\lambda_1 h$. Letting $h \to 0$ shows the first estimate. Moreover $\dfrac{z_2(t) - z_2(t+h)}{z_2(t)} = 1 - \dfrac{z_2(t+h)}{z_2(t)} \geq 1 - e^{\lambda_2 h} \geq -\lambda_2 h$. □

There also exists a natural counter part of Lemma 5.7, Property 5.8 and Corollary 5.9, of which we only give the result

Corollary 5.10. *Define* $\bar{\lambda}_1(s) = \sup_{t_2 \leq t \leq s} \lambda_1(t)$ *and* $\bar{\lambda}_s(t)$ *and* $\bar{\lambda}_2(s) = \inf_{t_2 \leq t \leq s} \lambda_2(t)$.

Then $\dfrac{d}{dt} z_1(t) \leq \bar{\lambda}_1(t)\, z_1(t)$

$\dfrac{d}{dt} z_2(t) \geq \bar{\lambda}_2(t)\, z_2(t)$. □

If we now associate to (5.3) a matrix vector ODE via the dependent variable $x := \begin{bmatrix} u \\ u' \end{bmatrix}$, say (cf. section 3)

(5.11) $$\frac{d}{dt} \mathbf{x} = J(\mathbf{u})\, \mathbf{x},$$

then we can show

Theorem 5.12. *Let (5.3) be such that for some $\rho > 0$, $\rho < b(t) < \infty$, $t \geq t_2$, $t \leq t_1$. Let $\|\cdot\| = \|\cdot\|_1$. Then both on the interval $[t_2, \infty]$ and on $(-\infty, t_1]$ there exists a dichotomic fundamental solution of (5.11), with dichotomy constant*

$$\kappa := \max \left[\max_t \frac{1+\bar{\lambda}_1(t)}{1+\lambda_1} , \max_t \frac{1+|\bar{\lambda}_2(t)|}{1+|\lambda_2|} , \frac{1+|\lambda_1 \lambda_2|}{\lambda_1 - \lambda_2} \right] ,$$

where $\bar{\lambda}_1(t)$ etc. is now also defined for relevant negative t-values.

Proof: For

$$t \geq s , \left| \begin{matrix} z_1(t) \\ z_1'(t) \end{matrix} \right| \left| \begin{matrix} z_1(s) \\ z_1'(s) \end{matrix} \right|^{-1} \leq \frac{1+\bar{\lambda}_1(t)}{1+\lambda_1} \frac{|z_1(t)|}{|z_1(s)|} \leq \frac{1+\bar{\lambda}_1(t)}{1+\lambda_1} \exp[\lambda_1(t-s)].$$

Similarly, for

$$t > s , \left| \begin{matrix} z_2(t) \\ z_2'(t) \end{matrix} \right| \left| \begin{matrix} z_2(s) \\ z_2'(s) \end{matrix} \right|^{-1} \leq \frac{1+|\bar{\lambda}_2(t)|}{1+|\lambda_2|} \frac{|z_2(t)|}{|z_2(s)|} \leq \frac{1+|\bar{\lambda}_2(t)|}{1+|\lambda_2|} \exp[\lambda_2(t-s)].$$

From [4] we deduce that the dichotomy constant bounds the cotg of the angle θ between the two subspaces induced by the projection P. In [7] it is shown that this quantity, together with the growth as above gives an estimate of the dichotomy constant.

Now we have, using the inner product definition

$$\cos\theta \leq \frac{1+|\lambda_1 \lambda_2|}{\sqrt{(1+\lambda_1^2)(1+\lambda_2^2)}} .$$

Hence

$$\text{cotg}\,\theta \leq \frac{1+|\lambda_1 \lambda_2|}{\lambda_1 - \lambda_2} .$$

□

The intriguing question now arises whether both dichotomic parts of the solution space join up to a globally dichotomic solution or not. The answer is negative as is shown by

Theorem 5.13. *The ODE (5.11) has a fundamental solution $\Phi = (\Phi^1 \ \Phi^2)$, with*

$$\|\Phi^1(0)\| = 1 , \ \|\Phi^1(t)\| \to \infty , \ \pm t \to \infty$$

$$\|\Phi^2(0)\| = 1 , \ \|\Phi^2(t)\| \to 0 , \ \pm t \to \infty .$$

Proof: Let u be the solution of (5.1) and let w_ε be defined by $w_\varepsilon(t) := u(t+\varepsilon)$. Then w_ε satisfies the same ODE and the same BC as u. If ε is small enough, then $\mathbf{x} := \begin{bmatrix} u \\ u' \end{bmatrix} - \begin{bmatrix} w_\varepsilon \\ w_\varepsilon' \end{bmatrix}$ satisfies $\frac{d\mathbf{x}}{dt} = \tilde{J}(\mathbf{u})\mathbf{x}$, with $\tilde{J}(\mathbf{u}) \approx J(\mathbf{u})$ as in (5.11).

Since for ε small enough the assumption for the coëfficiënt b in J will also hold for its counterpart in $\tilde{J}$, we see that it is not restrictive to identify J and $\tilde{J}$. Hence we conclude that (5.11) must have a solution Φ^1 say, that disappears for $\pm t \to \infty$. Consequently it follows from Theorem 5.12 that there must exist a complementary solution that grows unbounded for $\pm t \to \infty$. □

The result of the preceding analysis thus far is rather disappointing. Apparently we do not have dichotomy but also no "polychotomy".

5.2. A stabilization of the ODE: Kortweg-de Vries

The system (5.1), (5.2) plays an important role in the study of solitary waves, cf. [11]. Normally (5.1) appears in an eigenvalue formulation, where the unknown eigenvalue c is (a normalized form of) the wave speed; in particular consider

(5.14a) $\quad u'' = q(t,u,u') + cu.$

We can augment the ODE to a third order system by adding

(5.14b) $\quad c' = 0.$

Accordingly, we need a third relation to have a full set of BC. In order to find out how to choose this, let us consider a special ODE,

(5.15) $\quad u'' = cu - \frac{3}{2} u^2,$

which is a dimensionless formulation for the solitary wave of the Kortweg-de Vries equation, cf. [13]. We have chosen this equation because we happen to know "the" solution; in fact, there exists a family of solutions

(5.16) $\quad \hat{u}(t) = \hat{c} \cosh^{-2} (\tfrac{1}{2}\sqrt{\hat{c}}\, t), \ \hat{c} > 0.$

This then enables to analyze the linearized version of the augmented system in some detail. We find for J in (5.11)

(5.17) $$J = \begin{bmatrix} 0 & 1 & 0 \\ \hat{c} - 3\hat{u} & 0 & \hat{u} \\ 0 & 0 & 0 \end{bmatrix},$$

where we have introduced $x := (u, u', c)^T$.
In order to analyze the solution space of (5.11), (5.17) we concentrate on the ODE

(5.18) $\quad u'' = (\hat{c} - 3\hat{u})u + \hat{u}c.$

Note that for $c = 0$ we obtain an ODE like (5.3), (5.4) with $b(t) \to \hat{c}$. Hence we can apply Theorem 5.13, which gives us a characterization for two fundamental modes, viz Φ^1, Φ^2 with an additional zero third coordinate. To be more specific, let $\hat{c} = 1$, then one can easily deduce that

$$(5.19a)\qquad \begin{bmatrix} \Phi^1 \\ 0 \end{bmatrix} =: \Psi^1 \sim \begin{cases} e^t \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, & t \to \infty, \\ e^{-t} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, & t \to -\infty, \end{cases}$$

$$(5.19b)\qquad \begin{bmatrix} \Phi^2 \\ 0 \end{bmatrix} =: \Psi^2 \sim \begin{cases} e^{-t} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, & t \to \infty, \\ e^t \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, & t \to -\infty. \end{cases}$$

We have used a routine from BOUNDPAK [10], to compute the fundamental modes of (5.18). The graphs below were derived using an absolute error tolerance of 10^{-12} and with $L = 15$ (that is the interval was truncated to [-15,15], see Figs. 5.1, 5.2, 5.3.

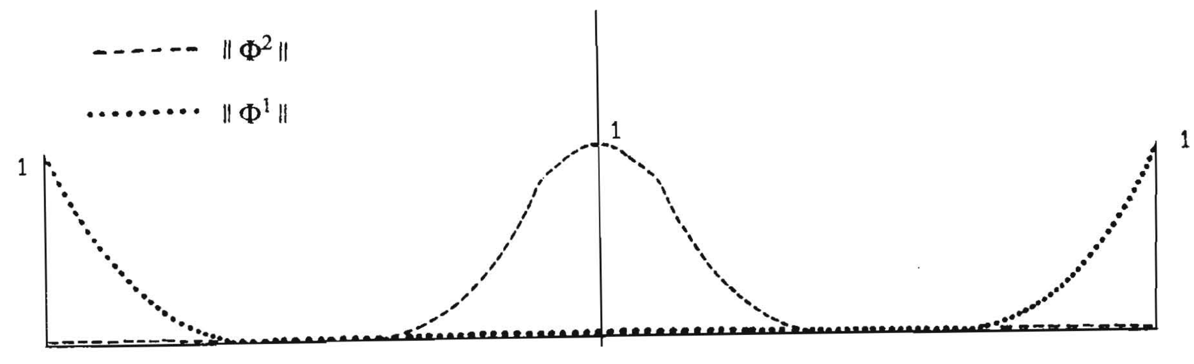

Fig. 5.1

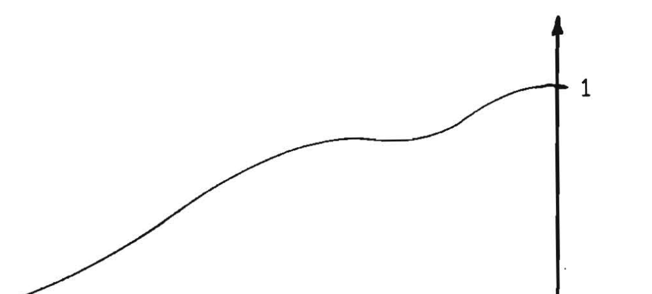

Fig. 5.2. $\| \Phi^2 \|$ different scale on $[-u\,, 0]$

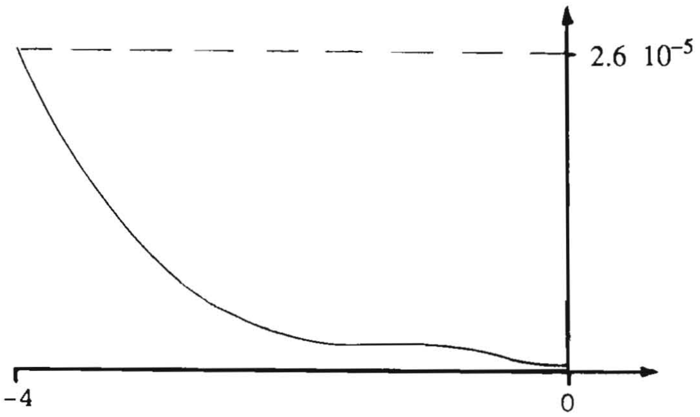

Fig. 5.3. $\| \Phi^1 \|$ different scale on $[-u , 0]$

In order to find a third fundamental mode we compute a particular solution of (5.18), e.g.

$$\Phi^3(t) := c \int_0^t \Phi(t)\,\Phi^{-1}(s) \begin{bmatrix} 0 \\ \hat{u}(s) \end{bmatrix} ds\,, \quad c \neq 0. \tag{5.20}$$

Property 5.21. $\| \Phi^3(t) \|$ *is uniformly bounded.*

Proof. Let $t > 0$. We can write

$$\Phi(t) = [\Phi^1(t)\,\Phi^2(t)\,] = \begin{bmatrix} e^t & e^{-t} \\ e^t & -e^{-t} \end{bmatrix} [I + R(t)]\,, \; t \to \infty\,,$$

where $\| [I + R(t)]^{-1} \|$ is uniformly bounded on $[0, \infty)$ and $\| R(t) \| \to 0$ for $t \to \infty$. Hence

$$\Phi^{-1}(s) = -\tfrac{1}{2} \begin{bmatrix} e^{-s} & e^{-s} \\ e^s & -e^{-s} \end{bmatrix} [I + S(s)\,, \text{ where } \| S(s) \| \to 0\,, \; s \to \infty.$$

It can be shown that $\int_0^\infty \| S(s) \|\, ds < \infty$, so

$$| \Phi^{-1}(s) \begin{bmatrix} 0 \\ \hat{u}(s) \end{bmatrix} | < \begin{bmatrix} e^{-2}s(1 + O(1)) \\ 1 + \varepsilon(s) \end{bmatrix}\,, \text{ with } \int_0^\infty | \varepsilon(s) |\, ds < \infty.$$

From this it can simply be checked that $\| \int_0^t \Phi(t)\,\Phi^{-1}(s) \begin{bmatrix} 0 \\ \hat{u}(s) \end{bmatrix} ds \| < \infty$ for all t.

A similar estimate can be given for $t < 0$. □

We thus conclude that a third fundamental mode of (5.11), (5.17) is given by

$$\begin{bmatrix} \Phi^3 \\ 1 \end{bmatrix} =: \psi^3 \sim \begin{bmatrix} \mu_1(t) \\ \mu_2(t) \\ 1 \end{bmatrix}\,, \; | \mu_1(t) |\,, \; | \mu_2(t) | < \infty \text{ for all it.} \tag{5.19c}$$

We therefore see that (5.11), (5.17) can not have a dichotomic solution space and not even a polychotomic one. That means that it is impossible to stabilize the eigenvalue problem (5.14), (5.15) by choosing appropriate (multipoint) BC.

However, for this problem, and quite a few other similar situations, often some analytic behaviour of the desired solution is available, like its asymptotic growth. Here we know, with $\hat{c} = 1$, that $\hat{u}(t) \sim e^{-t}$, $t \to \infty$ and $\hat{u}(t) \to e^{t}$, $t \to -\infty$. Therefore we may try to reformulate (5.15) via the transformation

(5.22a) $\quad v(t) := u(t)\, e^{-t}$,

which gives

(5.22b) $\quad v'' = (c-1)\, v - \frac{3}{2} v^2 e^{t} - 2v'.$

By adding

(5.22) $\quad c' = 0$

we obtain a linearized system with Jacobian (at the solution $\hat{v} := \hat{u} e^{-t}$, $\hat{c} = 1$)

(5.23) $$J := \begin{bmatrix} 0 & 1 & 0 \\ -3\hat{u} & -2 & \hat{u} e^{-t} \\ 0 & 0 & 0 \end{bmatrix} .$$

As before it is clear that the behaviour of the fundamental modes can be deduced from

(5.24) $\quad u'' = -2u' = 3\hat{u}\, u.$

Although Theorem 5.12 is not applicable here ("$\mathbf{b}(t) \to 0$") it is not difficult to see, using Lemma 5.7, that there exists a solution $z_1(t)$ with $z_1(0) = 1$ and on $[0, \infty)$ uniformly bounded, and at the same time a solution $z_2(t)$, with $z_2(0) = 1$, that "grows" at most like e^{-2t}. On $(-\infty, 0]$ there exist likewise solutions being uniformly bounded.

Hence, in whatever way the fundamental modes to the left are linked up with those to the right, we have two *initial value stable modes*! In a similar way as Property 5.21 we can show that (5.11), (5.23) also has an independent mode that is uniformly bounded. Thus we conclude that the eigenvalue problem (5.22b), (5.22c) is well-conditioned if we impose a BC like

(5.24) $\quad v(-\infty) = 1 \,,\; v'(-\infty) = 0 \,,\; v(0) = 1.$

Note that we effectively have a problem on the half interval $(-\infty, 0)$ only; not surprisingly in view of the symmetric character of the original ODE with respect to 0.

5.3. Stabilization of the BC: Burgers equation

As a final example, consider the ODE (shock form for Burgers equation)

(5.25a) $\quad u'' = uu' + cu'$

where u satisfies

(5.25b) $\quad u(-\infty) = 1\,,\ u(\infty) = -1.$

With $\hat{c} = 0$, an exact solution is given by

$$\hat{u}(t) = -1 + \frac{2}{1+e^t}. \tag{5.26}$$

The linearized ODE is

$$u'' = \hat{u}\,u' + \hat{u}'u\,, \tag{5.27}$$

which can be shown to have, for $t > 0$, solutions $z_1(t)$, $z_2(t)$, with $z_1(t) \sim e^{-t}$, $t \to \infty$ and $z_2(t) \sim 1$, $t \to \infty$, and similarly, for $t < 0$, solutions $z_1^*(t)$, $z_2^*(t)$ with $z_1^*(t) \sim e^t$, $t \to \infty$ and $z_2^*(t) \sim 1$, $t \to \infty$.
The proof is similar to that of Theorem 5.13 but has to employ a different version of Property 5.8 since assumption 5.5 does not hold.

As for KdV we can argue that the linearized system should be such that (5.27) has a fundamental mode that goes to zero when $\pm t \to \infty$. Consequently, we find that there is a fundamental mode $u_1(t)$, such that

$$u_1(t) \sim \begin{cases} z_1(t)\,, & t \to \infty\,, \\ z_1^*(t)\,, & t \to -\infty\,, \end{cases} \tag{5.28a}$$

and a fundamental mode $u_2(t)$, such that

$$u_2(t) \sim \begin{cases} z_2(t)\,, & t \to \infty\,, \\ z_2^*(t)\,, & t \to -\infty\,. \end{cases} \tag{5.28b}$$

Hence we have polychotomy. Therefore we can stabilize the eigenvalue BVP (5.25) by choosing an additional BC at $t = 0$, e.g.

(5.25c) $\quad u(0) = 0.$

If we would choose any other value for $u(0)$, as long as it is between -1 and 1, we have a well-conditioned BVP with a solution $\hat{u}(t+\rho)$, ρ being some suitable shift.

REFERENCES

[1] U.M. Ascher, R.M.M. Mattheij, R.D. Russell, Numerical Methods for Boundary Value Problems of Ordinary Differential Equations, Prentice-Hall, Englewood Cliffs (1988).

[2] U.M. Ascher, R.D. Russell, Numerical Boundary Value ODE's, Birkhäuser, Boston (1985).

[3] F.R. de Hoog, R.M.M. Mattheij, The role of conditioning in shooting techniques, in [2], pp. 21-54.

[4] F.R. de Hoog, R.M.M. Mattheij, On dichotomy and well conditioning in BVP, SIAM J. Numer. Anal. **24** (1987), 89-105.

[5] F.R. de Hoog, R.M.M. Mattheij, On the conditioning of Multipoint Boundary Value Problems, SIAM J. Math. Anal. 1989.

[6] M. Lentini, M.R. Osborne, R.D. Russell, The close relationships between methods for solving two-point boundary value problems, SIAM J. Numer. Anal. **22** (1985), pp. 280-309.

[7] P.M. van Loon, Continuous decoupling transformations for linear boundary value problem, Ph.D. Thesis, Eindhoven 1987.

[8] R.M.M. Mattheij, The conditioning of linear boundary value problems, SIAM J. Numer. Anal. **19** (1982), pp. 963-978.

[9] R.M.M. Mattheij, Stable Solution of Boundary Value Problems of Ordinary Differential Equations, in Proceedings of CTAC-85 (eds. J. Noye, R. May) North-Holland, Amsterdam (1986), pp. 51-70.

[10] R.M.M. Mattheij, G.W.M. Staarink, BOUNDPAK, Package for Solving Boundary Value Problem, User's Manual Chapter I, report WD 85-03, Nijmegen (1985).

[11] A.C. Newell, Solitons in Mathematics and Physics, CBMS-NSF Regional Conference Series in Applied Mathematics 48, SIAM, Philadelphia (1985).

[12] M.H. Protter, H.F. Weinberger, Maximum Principles in Differential Equations, Prentice-Hall, Englewood Cliffs (1967).

[13] G.B. Witham, Linear and Nonlinear Waves, John Wiley & Sons, New York (1974).