Decoupling and stability of algorithms for boundary value problems

SIAM REVIEW
Vol. 27, No. 1, March 1985

DECOUPLING AND STABILITY OF ALGORITHMS FOR BOUNDARY VALUE PROBLEMS*

R. M. M. MATTHEIJ†

Abstract. The ordinary differential equations occurring in linear boundary value problems characteristically have both stable and unstable solution modes. Therefore a stable numerical algorithm should avoid both forward and backward integration of solutions on large intervals. It is shown that most methods (like multiple shooting, collocation, invariant imbedding and difference methods) derive their stability from the fact that they all decouple the continuous or the discrete problem sooner or later (for instance when solving a linear system). This decoupling is related to the dichotomy of the ordinary differential equations. In fact it turns out that the inherent initial value instability is an important prerequisite for a stable utilization of the decoupled representations from which the solutions are computed. How this stability is related to the use of the boundary conditions is also investigated.

Key words. boundary value problems, direct methods, singular perturbations

AMS (MOS) 1980 subject classifications. 65L10, 65F05, 34D15

CONTENTS

1. Introduction.

Consider the linear system

$$\dot{x}(t) = L(t)x(t) + r(t), \qquad \alpha \leqq t \leqq \beta, \tag{1.1}$$

where $L(t)$ is a continuous $n \times n$ matrix valued function. Let the following *boundary*

*Received by the editors April 12, 1983, and in revised form June 14, 1984.
†Mathematisch Instituut, Katholieke Universiteit, Toernooiveld, 6525 ED Nijmegen, the Netherlands.

condition (BC) be given:

$$M_\alpha x(\alpha)+M_\beta x(\beta)=b, \tag{1.2}$$

where M_α and M_β are $n\times n$ matrices. Assume the solution is unique.

Boundary value problems (BVPs) form an active area of research and there exists a large number of methods to compute solutions of such BVPs (1.1), (1.2), cf, [4], [11], [23], [53], [55] for some general references. Historically and conceptually, methods have had many different backgrounds. For example, multiple shooting [16], [17], [24], [51] was developed to improve the poor stability of single shooting. Collocation, was long considered too expensive (and hence not competitive) until a more rigorous investigation showed its usefulness [2], [3], [56]. It is interesting to realize, however, that a condensed form of collocation is more or less equivalent to multiple shooting with a shooting interval of only one integration (if the integration is based on Runge–Kutta formulae, this equivalence also follows from [30], [64]). It may also be equivalent to some difference methods, cf. [56].

Recently a relation between the box scheme and invariant imbedding has been established, cf. [26]. The latter paper and many others also show that sparse BVP matrix solvers are related to each other cf. [5], [8], [27], [37], [38], [66]. All these methods try to circumvent the inherent instability with respect to initial data that is so characteristic of BVPs. Indeed, the ordinary differential equation (ODE) system (1.1) usually has a dichotomy; that is, the solution space can be split into a subspace of solutions whose members do not decrease with increasing t (and often even decay) and a complementary solution space whose members do not decrease (and often decay for decreasing t). Properties of such ODEs are discussed, for example, in [15], [20], [47], [48], [58], [67].

As we shall see, the above-mentioned methods avoid this (initial value) instability via an appropriate decoupling of the dynamics, either analytically or numerically. By an appropriate analytic decoupling we mean that the system (1.1) is transformed such that the nonincreasing modes can be computed from a subsystem of lower dimensionality. If the system is first discretized by some numerical method, then an appropriate decoupling means that the resulting (discrete) system is analogously transformed. It will be shown that the idea of finding a decoupling is closely related to computing bases for the two aforementioned subspaces. In fact an important feature of a robust BVP algorithm should be the capability to find such transformations (implicitly or explicitly). These transformations are nothing but a (numerically well-conditioned) method of determining a geometric basis for the directions of the solutions. The computation of the solutions of the transformed system is then done in two sweeps. First, a suitable component is computed in a forward direction, after which the complementary component is found by integrating (or recurring) in a backward direction. By such an algorithm, one which finds the transformations and employs the forward and backward integration as indicated, we have a means to compute members of these two solution subspaces in a stable way. In order to indicate whether such a transformation may be successful, we introduce a consistency concept. Of great practical importance is the fact that certain BCs induce such decoupling transformations in a natural way. We shall show that they also induce *consistency*, which is why algorithms like invariant imbedding, or the Godunov–Conte algorithm, are stable. (It is curious that no explicit mention of this simple property seems to have been made before, cf. [35, Ex. 6.5]).

Roughly speaking, this paper consists of three parts. First, §§ 1 to 4 set the general framework. Section 2 summarizes several definitions and conventions, §3 describes the general decoupling algorithm and §4 shows why the BCs may imply consistency if the

problem is well-conditioned. The second part, §§5 and 6, mainly deals with analytic methods, most of them being some variant of the Riccati method. Because of a similarity with the problems of this paper, we have devoted §6 to decoupling of problems with two time scales. The last part, §§7 and 8, deals with discrete BVP methods. Section 7 considers multiple shooting and its variants and §8 the important linear algebraic systems that usually arise after discretizing a BVP.

2. Definitions and conventions.

2.1. Norms. We assume that the real n-dimensional space is provided with some Hölder norm, denoted by $\|\cdot\|$. This induces a *least upper bound* for matrices A

$$\operatorname{lub}(A) = \|A\| = \max_{\|x\|=1} \|Ax\|, \tag{2.1}$$

and a *greatest lower bound*

$$\operatorname{glb}(A) = \min_{\|x\|=1} \|Ax\|. \tag{2.2}$$

2.2. Partitioning. Any integer $k \leqq n$ induces a partitioning of the matrix A as follows

$$A = \begin{bmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{bmatrix}, \tag{2.3}$$

where A^{11} is a $k \times k$ matrix. By a block upper triangular matrix we mean such a matrix with $A^{21} = 0$. We also write

$$A = [A^1 | A^2], \tag{2.4}$$

where A^1 has k columns. Correspondingly we may partition a vector x as

$$x = \begin{pmatrix} x^1 \\ x^2 \end{pmatrix}, \tag{2.5}$$

where x^1 has k coordinates. By $\operatorname{span}(A^j)$ we mean the space spanned by the columns of A^j.

If we partition the rows of A in a similar way (i.e. into the first k rows and the last $n-k$ rows), we shall use the notation

$$A = \left[\begin{array}{c} {}^1A \\ \hline {}^2A \end{array}\right]. \tag{2.6}$$

2.3. Sums and products of matrices. We denote

$$\sum_{j=p}^{q} A_j = \begin{cases} A_p + \cdots + A_q & \text{if } q \geqq p, \\ 0 & \text{if } q < p, \end{cases} \tag{2.7a}$$

$$\prod_{j=p}^{q} A_j = \begin{cases} A_q \cdots A_p & \text{if } q \geqq p, \\ I & \text{if } q < p. \end{cases} \tag{2.7b}$$

2.4. Fundamental solutions: continuous case. If an $n \times n$ matrix function Φ satisfies

$$\dot{\Phi} = L(t)\Phi \tag{2.8}$$

and $\Phi(0)$ is nonsingular, then Φ is called a *fundamental solution* for (1.1). Continuity of L implies that $\Phi(t)$ is then nonsingular for all t. The linear space of solutions of (2.8) is

denoted by span(Φ). We now make

Assumption 2.9. Let the solution space have a *dichotomy*, i.e. suppose there exists a partitioning $\Phi=(\Phi^1|\Phi^2)$, fixed throughout $[\alpha,\beta]$, and a reasonably small constant $\bar{\kappa}$ ($\geqq 1$) the *dichotomy constant*, such that

$$\operatorname{lub}(\Phi^1(t))/\operatorname{glb}(\Phi^1(s))\leqq\bar{\kappa}, \qquad t\leqq s,$$
$$\operatorname{lub}(\Phi^2(t))/\operatorname{glb}(\Phi^2(s))\leqq\bar{\kappa}, \qquad t\geqq s.$$

We call span(Φ^1) the *unstable* and span(Φ^2) the *stable* solution space.

This notion is a slight generalization of exponential dichotomy, see [15], [31]. In [21] it has been shown that Assumption 2.9 holds *if the* BVP *is well-conditioned* (see §4). We realize that for finite intervals there always exists some constant $\bar{\kappa}$ for any fundamental solution splitting. However, the constant we use in Assumption 2.9 should not be something like $\exp(\mathscr{L}(\beta-\alpha))$, where $\mathscr{L}$ is some Lipschitz constant. Thus 2.9 should be interpreted in the proper spirit, that is for $\bar{\kappa}$ a moderate constant of order one. By doing so we do not have to complicate the subsequent analyses by performing rather obvious but tedious asymptotics.

For technical reasons we would like to have a normalized fundamental solution. Hence we also ask (cf. [35, Assumption 3.21]).

Assumption 2.10. Let Φ be normalized such that $\max_{t\in[\alpha,\beta]}\|\Phi(t)\|=1$ and $\forall\, p,q\leqq n \max_t\|\phi^p(t)\|=\max_s\|\phi^q(s)\|$, where $\phi^p(t)$ denotes the pth column of $\Phi(t)$.

2.5. Fundamental solutions: discrete case. The discrete BVP methods we will consider in the sequel can be thought of as one-step discretizations of the continuous problem. Consider the grid$\{t_0,\cdots,t_N\}\subset[\alpha,\beta]$, where $t_0=\alpha$ and $t_N=\beta$. Denoting a solution value of x at t_i by x_i ($=x(t_i)$), the resulting recursion should be

$$x_{i+1}=A_i x_i+f_i, \tag{2.11}$$

where (disregarding discretization errors)

$$A_i=\Phi(t_{i+1})\big[\Phi(t_i)\big]^{-1}=I+\int_{t_i}^{t_{i+1}} L(\tau)\Phi(\tau)\big[\Phi(t_i)\big]^{-1}d\tau, \tag{2.12a}$$

and

$$f_i=\int_{t_i}^{t_{i+1}} \Phi(t_{i+1})\Phi(s)^{-1}r(s)\,ds. \tag{2.12b}$$

Apparently, a discrete fundamental solution for (2.11) is given by $\{\Phi_i\}_{i=0}^N$, where

$$\Phi_i=\Phi(t_i). \tag{2.13}$$

This discrete solution space therefore has the same dichotomy as the continuous one.

2.6. Special boundary conditions. Quite often the BCs (1.2) have special form such as *separated* ones with

$$M_\alpha=\begin{bmatrix}\varnothing\\ {}^2M_\alpha\end{bmatrix}, \qquad M_\beta=\begin{bmatrix}{}^1M_\beta\\ \varnothing\end{bmatrix}. \tag{2.14}$$

If the number of nonzero rows in M_α and M_β is larger than n, but still smaller than $2n$, we have a *partially separated* BC for which, say,

$$M_\beta = \begin{bmatrix} {}^1M_\beta \\ \varnothing \end{bmatrix}, \tag{2.15}$$

where no special zero row structure for M_α is assumed.

3. Decoupling of the dynamics. We shall give the basic idea of decoupling both for a differential equation and for a difference equation. First we give a geometrical introduction which shows the basic principle. Then we treat the continuous and discrete cases.

3.1. Introduction. In order to understand more easily why transformations can produce a meaningful geometric basis for the directions of the solutions, we discuss a simple geometrical model first.

Let x and y be two independent vectors, such that $\|x\|_2 \gg \|y\|_2$. Let a and b be linear combinations of x and y, say

$$a = \alpha_1 x + \alpha_2 y, \qquad b = \beta_1 x + \beta_2 y, \tag{3.1}$$

where α_1, β_2 are $O(1)$ and such that a and b are independent as well. Although a and b span the same space as x and y, they are less attractive as a basis, since they generally enclose a small angle (see Fig. 3.1).

Given a and b and neglecting rounding errors, we can find a better basis as follows: Let γ_1 be of the order of $\|a\|_2$; then define

$$t_1 = \frac{a}{\gamma_1}. \tag{3.2}$$

Let $t_2 \in \text{span}(a,b)$ be a vector with $\|t_2\|_2 \approx \|t_1\|_2$ such that t_2 and t_1 do not enclose a small angle. Then for some γ_2, γ_3

$$b = \gamma_2 t_1 + \gamma_3 t_2. \tag{3.3}$$

We shall show that $\gamma_3 \|t_2\|_2$ is of the order of $\|y\|_2$ if we assume that the angle θ between x and y is not small.

Think of a situation where $\gamma_1 = \|a\|_2$ (so $\|t_1\|_2 = 1$), $(t_1, t_2) = 0$ and $\|t_2\|_2 = 1$. In such a case we just have a Gram–Schmidt orthogonalization, with

$$\gamma_2 = (b, t_1), \qquad \gamma_3 = (b, t_2). \tag{3.4}$$

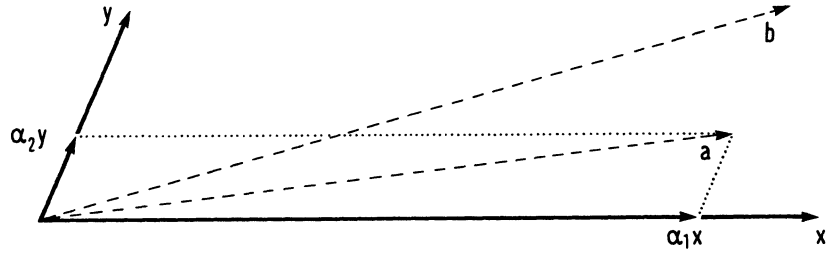

FIG. 3.1

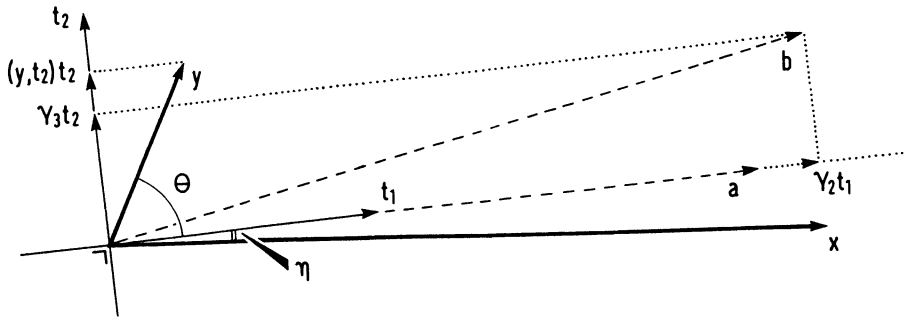

FIG. 3.2

Since a was assumed to have a significant component of x, the angle η between x and a is small, whence $|\gamma_1| \approx \|x\|_2$. It also follows that the angle between t_2 and x equals $\pi/2+\eta \approx \pi/2$. Now since b has a significant component of y (i.e. β_1/β_2 is not large), it can be seen from Fig. 3.2 that the projection of b on the t_2-axis and the projection of y on the t_2-axis have the same order of magnitude, namely $\|y\|_2$, (as θ was assumed not to be small). Formally, this process can be written as

$$(x|y)\begin{pmatrix}\alpha_1 & \beta_1\\ \alpha_2 & \beta_2\end{pmatrix} = (a|b) = (t_1|t_2)\begin{pmatrix}\gamma_1 & \gamma_2\\ 0 & \gamma_3\end{pmatrix}. \tag{3.5}$$

The third expression in (3.5), viz the QU-decomposition, therefore retrieves information about the original basis, that is $|\gamma_1| \approx \|x\|_2$ and $|\gamma_3 \approx \|y\|_2$ (where $\approx$ means "order of magnitude"). It should be realized that this simple but important phenomenon only occurs if t_1 and t_2 do not enclose a small angle. Now, if we generalize this to subspaces and let $(t_1|t_2)$ be a square matrix, we should expect such a factorization to produce magnitudes of blockvectors x and y, provided $(t_1|t_2)$ is *well-conditioned* (which implies that span(t_1) and span(t_2) do not enclose a small angle, cf. [32]).

3.2. The continuous case. Both analytically and computationally linear transformations of ODEs play an important role, cf. [5], [12], [20], [30], [32], [33], [39], [40], [41], [42], [47], [49], [62], [63], [68]. The most simple approach is to try to transform the system matrix $L(t)$ by a time-dependent matrix $T(t)$ such that $T(t)^{-1}L(t)T(t)$ has a *special* form, e.g. a diagonal or an upper triangular matrix. Unless L is constant or slowly varying, this does not necessarily lead to a system that has a special advantage over the original. To be more specific, let

$$\overline{W}(t) = T(t)^{-1}L(t)T(t). \tag{3.6}$$

Then by setting

$$x(t) = T(t)y(t), \tag{3.7}$$

we see that

$$\dot{y} = (T^{-1}LT - T^{-1}\dot{T})y = (\overline{W} - T^{-1}\dot{T})y =: Wy. \tag{3.8}$$

Hence in order for W to have the same *special* form as $\overline{W}$, $T^{-1}\dot{T}$ must have such a form too.

A better approach for obtaining a special form of W is to compute T from the *Lyapunov equation* (cf. (3.8))

$$\dot{T} = LT - TW, \tag{3.9}$$

where W may, for example, be block upper triangular. This has an important consequence for the solutions of the transformed system. To see this, let $T(\alpha)$ be some initial value, let K be a fundamental solution for (2.8) with $K(\alpha) = T(\alpha)$, and define

$$V(t) = T(t)^{-1}K(t). \tag{3.10}$$

Then V is a fundamental solution of

$$\dot{y} = Wy. \tag{3.11}$$

Since $V(\alpha) = I$, it now follows that $V(t)$ is block upper triangular for all t. Hence finding a block upper triangular form for W is equivalent to finding a matrix function T such that a fundamental solution K can be factorized as TV, with V in the same form as W.

We shall show that such a factorization gives rise to a splitting of the solution space into subspaces representing the growth classes of the dichotomy (cf. §2). In §3.1 in order to have $|\gamma_1| \approx \|x\|_2$, the vector a had to contain a (significant) component of the vector x. For similar reasons we now require that the matrix solution induced by the first k columns of $T(\alpha)$, viz. K^1, contains a nontrivial component of Φ^1. To this end we define the following important concept.

DEFINITION 3.12. Let the fundamental solution K be partitioned as $K = (K^1 | K^2)$. Then K is said to be *consistent* (with Φ as in Assumption 2.9) if $\operatorname{span}(K^1(\alpha)) \cap \operatorname{span}(\Phi^2(\alpha)) = \{0\}$.

We have

PROPERTY 3.13. *K is consistent iff*

(i) $\operatorname{span}(K^1(t)) \cap \operatorname{span}(\Phi^2(t)) = \{0\}$, *for all t or*

(ii) $[[\Phi(\alpha)]^{-1}K(\alpha)]^{11}$ *is nonsingular.*

Proof. (i) is trivial. (ii): There exists some constant matrix H such that $K = \Phi H \Rightarrow K^1 = \Phi^1 H^{11} + \Phi^2 H^{21}$. Suppose (ii) is not true, i.e. H^{11} is singular; then there exists a vector $a \neq 0$ such that $H^{11}a = 0$. Since H is nonsingular, $H^{21}a$ must be nonzero, whence $\operatorname{span}(K^1) \cap \operatorname{span}(\Phi^2) \neq \{0\}$. Now, on the other hand, if (ii) is true, H^{11} is nonsingular and $K^1 v \in \operatorname{span}(\Phi^2)$ for some v, so $\Phi^1 H^{11} v = 0$ implies $v = 0$. □

From the foregoing we see that consistency of K means that the first k columns of K represent a basis for an *unstable solution subspace*.

If $T(t)$ in (3.10) is a well-conditioned matrix, i.e. $\|T(t)\|\|T^{-1}(t)\|$ is not large [71, p. 86ff] and K is consistent, then we can expect from our geometrical model that V^{11} and V^{22} represent the increments of the unstable and the stable modes of (3.11) respectively. This is quantified below.

THEOREM 3.14. *Let $K := \Phi H$. Let H^{11} be nonsingular, so K is consistent. Then $\tilde{\Phi} = (\tilde{\Phi}^1 | \tilde{\Phi}^2) = (\Phi^{11}H^{11} + \Phi^2 H^{21} | \Phi^2(H^{22} - H^{21}[H^{11}]^{-1}H^{12}))$ has a similar dichotomy to Φ, that is the dichotomy constants for Φ and $\tilde{\Phi}$ are of the same order. Moreover in the factorization $K = TV$ the following estimates hold*

$$\frac{1}{\|T\|} \leqq \frac{\|V^{11}\|}{\|\tilde{\Phi}^1\|}, \qquad \frac{\operatorname{glb}(V^{11})}{\operatorname{glb}(\tilde{\Phi}^1)} \leqq \|T^{-1}\|, \qquad \frac{\sin\theta}{\|T\|_2} \leqq \frac{\|V^{22}\|_2}{\|\tilde{\Phi}^2\|_2}, \qquad \frac{\operatorname{glb}_2(V^{22})}{\operatorname{glb}_2(\tilde{\Phi}^2)} \leqq \|T^{-1}\|_2,$$

where θ is the angle between $\tilde{\Phi}^1$ and $\tilde{\Phi}^2$ (cf. [32]).

Proof. Write

$$T^{-1} =: S = \begin{bmatrix} {}^1S \\ {}^2S \end{bmatrix},$$

where ${}^1S$ has k rows. From $V = T^{-1}K = S\Phi H$,

$$V^{11} = {}^1S(\Phi^1 H^{11} + \Phi^2 H^{21}) = {}^1S\tilde{\Phi}^1.$$

Hence: $\|V^{11}\| \leqq \|{}^1S\| \|\tilde{\Phi}^1\| \leqq \|T^{-1}\| \|\tilde{\Phi}^1\|$ and $\text{glb}(V^{11}) \leqq \|T^{-1}\| \text{glb}(\tilde{\Phi}^1)$. Since $\tilde{\Phi}_1 = T^1 V^{11}$, $\|\tilde{\Phi}^1\| \leqq \|T\| \|V^{11}\|$ and $\text{glb}(\tilde{\Phi}^1) \leqq \|T\| \text{glb}(V^{11})$. On the other hand,

$$0 = {}^2S(\Phi^1 H^{11} + \Phi^2 H^{21}) \Rightarrow {}^2S\Phi^1 = -{}^2S\Phi^2 H^{21}[H^{11}]^{-1}.$$

Combining this with $V^{22} = {}^2S(\Phi^1 H^{12} + \Phi^2 H^{22})$, we obtain $V^{22} = {}^2S\Phi^2(H^{22} - H^{21}[H^{11}]^{-1}H^{12}) = S^2\tilde{\Phi}^2$. The estimates for V^{22} now follow in a similar way to those for V^{11}, cf. [32, Thm. 5.9].

In order for $\tilde{\Phi}$ to have a similar dichotomy, it only remains to show that the Schur complement $(H^{22} - H^{21}[H^{11}]^{-1}H^{12})$ of E^{11} is nonsingular. Otherwise, an $a \neq 0$ belongs to its kernel, and

$$H \begin{bmatrix} [H^{11}]^{-1}H^{12}a \\ a \end{bmatrix} = 0,$$

contradicting the nonsingularity of H. □

Remark 3.15. It is fairly important that $\|[H^{11}]^{-1}\|$ and $\|[H^{22} - H^{21}[H^{11}]^{-1}H^{12}]^{-1}\|$ not be large to make sure that Φ^1 and Φ^2 resemble $\tilde{\Phi}^1$ and $\tilde{\Phi}^2$ respectively. Perhaps even more interesting is that Theorem 3.14 shows how the skewness of the transformation T affects the growth properties of $V^{11}(t)$ and $V^{22}(t)$ (for a more detailed discussion see [32]).

We now show how these results should be used. Return to the original inhomogeneous ODE (1.1) and define

$$s(t) = T^{-1}(t)r(t). \tag{3.16}$$

Transforming (1.1) via a solution T of (3.9) leads to the following *decoupled* ODE:

$$\dot{y}^1 = W^{11}y^1 + W^{12}y^2 + s^1, \tag{3.17a}$$

$$\dot{y}^2 = W^{22}y^2 + s^2. \tag{3.17b}$$

Recall that V^{22} and V^{11} satisfy $\dot{V}^{22} = W^{22}V^{22}$ and $\dot{V}^{11} = W^{11}V^{11}$. Since V^{22} and V^{11} resemble $\tilde{\Phi}^2$ and $\tilde{\Phi}^1$ respectively, in their growth behaviour, and the latter resemble Φ^2 and Φ^1 respectively, we can conclude that (3.17b) is stable for increasing t and (3.17a) is stable for decreasing t. These considerations lead to the following basic algorithm for computing solutions of conditionally stable problems.

(3.18) BASIC ALGORITHM (*continuous case*)

Step I. Compute a matrix function T, given some appropriate $T(\alpha)$, and a block upper triangular matrix function W satisfying the Lyapunov equation (3.9).

Step II. For appropriate initial conditions $y^2(\alpha)$ and terminal conditions $y^1(\beta)$, compute the particular solution y_p and a fundamental solution (or part of a fundamental solution) Y by employing the stable directions, i.e. integrate first (3.17b) for t increasing, and then (3.17a) for t decreasing.

Step III. Compute by superposition the transformed desired solution $y = y_p + Yc$, selecting the constant vector c to satisfy the BC $M_\alpha T^{-1}(\alpha) y(\alpha) + M_\beta T^{-1}(\beta) y(B) = b$. (This step is not needed if y happens to coincide with y_p.)

Step IV. Compute x as Ty.

Although our intuitive derivation of the stability of Step II made use of the fundamental solution K, we do not compute K in practice. Indeed, as we noted in §1, this is not possible in many practically relevant problems where the growth of solutions in span (Φ^1) causes serious rounding error problems. The trick in the algorithm above is that instead we compute a more convenient form for our ODE system-matrix and so avoid computations where these unstable modes may blur the results.

In §§5 and 6 we shall discuss a number of algorithms that fit into the framework of (3.18). An important point will be how these algorithms manage to keep T well-conditioned. (One should realize that (3.9) is not solved as an ODE in T, since W is also unknown!)

3.3. The discrete case. In any BVP algorithm we have to discretize the ODE sooner or later in order to find numerical solutions. In contrast to §3.2 we now assume that this discretization precedes other manipulations which are needed to compute an approximate solution. Therefore we consider the discrete problem setting of §2.5. The gridpoints $t_0, \cdots, t_N$ can be thought of as the points used for collocation, cf. [2], [56], [69], or the points where shooting is restarted, cf. [16], [24], [25], [42], [51], [60], or just the discretization points of a one step method, cf. [12], [26], [57], [65], [70]. As in the continuous case, the computation of a solution $\{x_i\}$ by using (2.10) in a forward direction is not meaningful if there exist strongly increasing modes (and x is not such a solution). Likewise the computation of a fundamental solution $\{K_i\}$ (that is a matrix solution of $K_{i+1} = A_i K_i$) would be unstable for the same reason. Recalling from §3.1 that the magnitudes of Φ^1 and Φ^2 might be retrieved from a factorization of K, cf. (3.10), we now investigate the possibility of a decomposition

$$K_i = T_i V_i, \qquad i = 0, \cdots, N, \tag{3.19}$$

where V_i is block upper triangular. Since we identified the discrete solution with appropriate continuous solution values, we can immediately use the same consistency concept here. Thus we say K is *consistent* (with Φ) if $\operatorname{span}(K_0^1) \cap \operatorname{span} \Phi^2(0) = \{0\}$. From this it follows (cf. Property 3.13) that $[[\Phi_0]^{-1}[K_0]]^{11}$ is nonsingular. Moreover, Theorem 3.14 carries over directly; the estimates should now hold for each index i. Thus, a consistent choice of T_0 gives rise to a V_i^{11} which has a similar magnitude (that is in terms of its lub and glb) to Φ_i^1 (and likewise for V_i^{22} and Φ_i^2). Again we should avoid direct computation of the fundamental solution, by using the following discrete version of the Lyapunov equation (3.9): Let $T_0 = K_0$ determine a consistent fundamental solution K. Then, for each $i = 0, \cdots, N-1$, we compute transformations $\{T_i\}$ and block upper triangular matrices $\{U_i\}$ such that

$$A_i T_i = T_{i+1} U_i \tag{3.20}$$

(cf. [34]).

Remark 3.21. In practice solving (3.20) utilizes QU- or LU-decompositions of $A_i T$ (i.e. T_{i+1} is orthogonal or (block) lower triangular).

By defining

$$V_i := \prod_{j=0}^{i-1} U_j, \tag{3.22}$$

we see that (3.20) actually gives (3.19) in factored form. It is important to realize that the recursive computation of the $\{T_i\}$ and $\{U_i\}$ produces errors in U_i of the order of $\xi\|A_i\|\|T_i\|\|T_{i+1}^{-1}\|$ only, where ξ is the machine constant. Moreover, because the U_i are expected to be properly decoupled incremental matrices with $\|[\Pi U_j^{11}]^{-1}\|$, $\|\Pi U_j^{22}\| = O(1)$, we see that, for example, $\Pi_{j=0}^{i} U_j^{22}$ is perturbed by errors of the order $\xi\Sigma_{j=0}^{i}\|A_j\|\|T_j\|\|T_{j+1}^{-1}\|$ only. Thus, it is important to have well-conditioned transformations in general. We now use the following decoupled recursion instead of (2.10):

$$y_{i+1}^1 = U_i^{11} y_i^1 + U_i^{12} y_i^2 + g_i^1, \tag{3.23a}$$

$$y_{i+1}^2 = U_i^{22} y_i^2 + g_i^2, \tag{3.23b}$$

where we have set

$$g_i := T_{i+1}^{-1} f_i. \tag{3.24}$$

A discrete version of (3.18) is then given by

(3.25) BASIC ALGORITHM (*discrete case*)

Step I. Compute a set of nonsingular matrices $\{T_i\}$, given some appropriate T_0, and a set of block upper triangular matrices $\{U_i\}$ which satisfy (3.20).

Step II. Choose appropriate initial conditions y_0^2 and terminal conditions y_N^1 to compute solutions of both the inhomogeneous recursions (3.23) (and generally also of the homogeneous parts of (3.23)) by employing the stable directions, i.e. by solving (3.23b) first and then (3.23a).

Step III, *Step* IV follow (3.18).

Finally, we wish to elaborate a bit on the relation between the continuous and the discrete Lyapunov equation.

THEOREM 3.26. *Let T, V satisfy the continuous Lyapunov equation* (3.9). *Define $T_i = T(t_i)$, and $U_i = I - \int_{t_i}^{t_{i+1}} V(\tau)[V(t_i)]^{-1} d\tau$, $i = 0, 1, \cdots, N$. Then $\{T_i\}$ and $\{U_i\}$ satisfy* (3.20) (*for A_i in* (2.12a)).

Proof.

$$A_i = I + \int_{t_i}^{t_{i+1}} L(\tau)\Phi(\tau)[\Phi(t_i)]^{-1} d\tau$$

$$= I + \int_{t_i}^{t_{i+1}} L(\tau)T(\tau)V(\tau)[V(t_i)]^{-1}[T(t_i)]^{-1} d\tau.$$

Hence

$$A_i T_i = T_i + \int_{t_i}^{t_{i+1}} \{(L(\tau)T(\tau) - T(\tau)W(\tau))V(\tau)[V(t_i)]^{-1} + T(\tau)W(\tau)V(\tau)[V(t_i)]^{-1}\} d\tau$$

$$= T_i + \int_{t_i}^{t_{i+1}} \{\dot{T}(\tau)V(\tau) + T(\tau)\dot{V}(\tau)\}[V(t_i)]^{-1} d\tau = T_i + \{T_{i+1}V_{i+1} - T_i V_i\} V_i^{-1}$$

$$= T_{i+1}V_{i+1}V_i^{-1} = T_{i+1}U_i. \qquad \square$$

Theorem 3.26 shows that the results for the continuous case carry over to the discrete case. Note, however, that in practice we have to reckon with discretization errors, cf. [36]. (For some more dramatic differences see, e.g., [33].) As one might expect

(and can simply verify), the converse of 3.26 does not hold. Indeed, given any pair of sets $\{T_i\}$ and $\{U_i\}$ that satisfies (3.20), we can find infinitely many pairs of functions T, such that $T(t_i)=T_i$. Of particular interest is the choice of T, where

$$\dot{T}(t_i)=0. \tag{3.27}$$

By interpolating T, we can define a sufficiently differentiable ("continuous") transformation. By interpolating the V_i, we have also constructed a pseudo decoupling transformation which yields a (continuous) fundamental solution that is (block) upper triangular on the grid. If the U_i exhibit a proper growth, so will the V_i and it follows that such a fundamental solution will very likely be directionally close to a (continuous) fundamental solution with the desired properties.

4. Well-conditioning of BVPs and its consequences. When solving a BVP, one should be aware that no numerical method can be held responsible for large errors if the problem is inherently unstable. Therefore it makes sense to investigate the behaviour of such methods for *well-conditioned problems* only. Fortunately most problems which actually describe physically realistic situations can be expected to be well-conditioned from physical considerations. Usually a similar well-conditioning carries over to the *discretized problem* (where we have perturbed solution approximates), see [7] and also [33], [36]. Although the conditioning of a problem deals with perturbation sensitivity with respect to all data, it was shown in [35] that if there is a dichotomy, it is mainly the sensitivity of the solution with respect to the BC that is of importance, see also [16], [17], [30]. A meaningful quantity to measure this conditioning is given by

$$\mathscr{CN} := \max_{t\in[\alpha,\beta]} \|\Phi(t)Q^{-1}\|, \tag{4.1}$$

where

$$Q := M_\alpha\Phi(\alpha)+M_\beta\Phi(\beta). \tag{4.2}$$

As was shown in [35], our normalization assumption (2.10) makes the quantity κ

$$\kappa := \|Q^{-1}\|, \tag{4.3}$$

a useful estimate of $\mathscr{CN}$. Further, we can obtain important information regarding the value of κ by inspecting the norms of a partitioned Q. For this, we partition Φ^1 into

$$\Phi^1=\left(\Phi^1_u|\Phi^1_m\right), \tag{4.4}$$

where Φ^1_u represents the rapidly increasing modes, and Φ^1_s the "moderate" modes (those which do not increase or decrease significantly on the inverval $[\alpha,\beta]$). In a similar way we partition Φ^2 into

$$\Phi^2=\left(\Phi^2_m|\Phi^2_s\right), \tag{4.5}$$

where Φ^2_s represents the rapidly decreasing modes and Φ^2_m the "moderate" modes. By our normalization assumption we then find

$$\|\Phi^1_u(\alpha)\| \text{ is small}, \qquad \|\Phi^1_u(\beta)\|=O(1), \tag{4.6a}$$

$$\|\Phi^1_m(\alpha)\|, \|\Phi^2_m(\alpha)\|, \|\Phi^1_m(\beta)\|, \|\Phi^2_m(\beta)\|=O(1), \tag{4.6b}$$

$$\|\Phi^2_s(\alpha)\|=O(1), \qquad \|\Phi^2_s(\beta)\| \text{ is small}. \tag{4.6c}$$

So

$$Q \approx \left[M_\alpha \Phi_u^1(\beta) | \hat{Q} | M_\alpha \Phi_s^2(\alpha) \right], \tag{4.7}$$

where $\hat{Q} = [M_\alpha(\Phi_m^1(\alpha)|\Phi_m^2(\alpha)) + M_\beta(\Phi_m^1(\beta)|\Phi_m^2(\beta))]$. We therefore have, cf. [30, Thm. 4.6]

PROPERTY 4.8. *If the problem is well-conditioned, that is* $\|Q^{-1}\| = O(1)$, *then* $\operatorname{rank}(M_\beta \Phi_u^1(\beta)) = \operatorname{rank}(\Phi_u^1(\beta))$ *and* $\operatorname{rank}(M_\alpha \Phi_s^2(\alpha)) = \operatorname{rank}(\Phi_s^2(\alpha))$.

In this way, well-conditioning gives natural constraints for the BC. In particular, no row vector of M_α can be orthogonal to $\operatorname{span}(\Phi_s^2(\alpha))$ and similarly no row vector of M_β can be orthogonal to $\Phi_u^1(\beta)$. In practice, near orthogonality should also be excluded. We omit a further quantification, however. We now have

PROPERTY 4.9. *Let the* BVP *be well-conditioned.* (i) *If for some solution* ϕ *of the homogeneous problem* $M_\alpha \phi(a) = 0$, *then* $\phi \notin \operatorname{span}(\Phi_s^2)$, *i.e.* ϕ *must be either a significantly growing or a moderately growing solution.* (ii) *If* $M_\beta \phi(\beta) = 0$, *then* $\phi \notin \operatorname{span}(\Phi_u^1)$, *i.e.* ϕ *must be either a significantly decaying or a moderately growing solution.*

The proof of this follows from Property 4.8 by contradiction (cf. also [35, Ex. 6.5]). For separated BCs we employ Property 4.9 to obtain an important tool for showing why certain algorithms are stable.

THEOREM 4.10. *Let the* BVP *be well-conditioned and have separated* BC. *If* ${}^2M_\alpha T^1(\alpha) = 0$, *then the fundamental solution* K, *with* $K(\alpha) = T(\alpha)$, *is consistent.*

Remark 4.11. The stability considerations in §3 only make sense if there is a dichotomy. As is shown in [21], however, well-conditioning of a BVP implies that there is a splitting of the solution space as assumed in Assumption 2.9, with a moderate $\bar{\kappa}$.

5. Methods based on Riccati transformations. An important class of methods that decouple the system utilizes block lower triangular matrices T with diagonal blocks being identity matrices. This provides the normalization needed to make such a decoupling transformation meaningful (cf. Theorem 3.14). The equation to be satisfied by the remaining block of the T is a matrix *Riccati equation*. We shall first consider the most well-known member of this class, viz., *invariant imbedding*. Then we show that *order reduction* for scalar ODEs also belongs to this class and, finally, we briefly overview some algorithms for discrete scalar problems that are like Riccati transformation methods. In §6 we consider special Riccati transformations using power series as is natural for singularly perturbed problems.

5.1. Invariant imbedding. Although invariant imbedding can be introduced in many ways, cf. [1], [24], [43], [59], [61], we prefer to interpret the method as a linear transformation of a system to a nicer form, cf. [26], [39], [49]. This will enable us to use simple geometrical arguments to explain the possible blowup of the solution to the associated Riccati equation. We first describe the algorithm.

Consider the transformation

$$T(t) = \begin{bmatrix} I & \varnothing \\ P(t) & I \end{bmatrix}, \tag{5.1}$$

where $P(t)$ is an $(n-k)\times k$ matrix. Note that $T^{-1}(t)$ is obtained by replacing $P(t)$ by $-P(t)$ in (5.1). Substitution of T in the Lyapunov equation (3.9) gives

$$W = \begin{bmatrix} L^{11} + L^{12}P & L^{12} \\ \varnothing & -PL^{12} + L^{22} \end{bmatrix}, \tag{5.2}$$

provided P satisfies the *Riccati equation*

$$\dot{P} = L^{21} + L^{22}P - PL^{11} - PL^{12}P. \tag{5.3}$$

Originally invariant imbedding was advocated for its ability to transform a BVP into two IVPs, which could be solved by standard routines, when the BC are separated. The two ODEs for this purpose are (3.17a, b) (the latter being subject to *terminal* conditions). For P, one uses initial conditions.

We now show that this is not only a sensible, but also a consistent use of the BC. Consider the separated BC

$$\left(M_\alpha^{21} | M_\alpha^{22}\right) x(\alpha) = b^2, \tag{5.4a}$$

$$\left(M_\beta^{11} | M_\beta^{12}\right) x(\beta) = b^1. \tag{5.4b}$$

where M_α^{22} is nonsingular. Then

PROPERTY 5.5. *Let the* BVP *be well-conditioned and* M_α^{22} *be nonsingular. Define* $P(\alpha) = -[M_\alpha^{22}]^{-1} M_\alpha^{21}$. *Then the fundamental solution* K, *defined by* $K(\alpha) = T(\alpha)$ (*cf.* (5.1)) *is consistent.*

Proof. By this choice of $P(\alpha)$, we have

$$K^1(\alpha) = \begin{bmatrix} I \\ -\left[M_\alpha^{22}\right]^{-1} M_\alpha^{21} \end{bmatrix}.$$

Hence $M_\alpha K^1(\alpha) = 0$. Application of Theorem 4.10 completes the proof. □

PROPERTY 5.6. *If besides the previous assumptions,* $P(\beta)$ *also exists, then* $(M_\beta^{11} + M_\beta^{12} P(\beta))$ *is well-conditioned* (*and, a fortiori, nonsingular*).

Proof. (sketchily). Since K in Property 5.5 is consistent, the left upper block of $H = \Phi^{-1} K$ is nonsingular. Thus Theorem 3.14 implies that a properly scaled fundamental matrix $\hat{K}$, obtained from K by normalizing column solutions by their maximum value, should have similar growth properties to Φ. Now decompose $\hat{K}$ as $T\hat{V}$, where $\hat{V}$ is block upper triangular. Then $\hat{Q} = M_\alpha \hat{K}(\alpha) + M_\beta \hat{K}(\beta)$ satisfies:

$$\hat{Q} = \left[\begin{array}{c|c} \left[M_\beta^{11} + M_\beta^{12} P(\beta)\right] \hat{V}^{11}(\beta) & \left[M_\beta^{11} + M_\beta^{12} P(\beta)\right] \hat{V}^{11}(\beta) + M_\beta^{12} \hat{V}^{22}(\beta) \\ \hline \left[M_\alpha^{21} + M_\alpha^{22} P(\alpha)\right] \hat{V}^{11}(\alpha) & \left[M_\alpha^{21} + M_\alpha^{22} P(\alpha)\right] \hat{V}^{11}(\alpha) + M_\alpha^{22} \hat{V}^{22}(\alpha) \end{array}\right].$$

By construction $M_\alpha^{21} + M_\alpha^{22} P(\alpha) = 0$. Hence well-conditioning of $\hat{Q}$ means that both $[M_\alpha^{22} \hat{V}^{22}(\alpha)]$ and $[M_\beta^{11} + M_\beta^{12} P(\beta)]$ should be well-conditioned. □

By this choice of P (cf. (5.3)) to compute a T and W (cf. (5.1), (5.2)), we have performed Step I of the basic algorithm (3.18). The next step of the invariant imbedding method actually is to solve y from (3.17). By the clever choice of the initial value for $P(\alpha)$

$$\begin{aligned} y^2(\alpha) &= \left[T_{(\alpha)}^{-1} x(\alpha)\right]^2 = -P(\alpha) x^1(\alpha) + x^2(\alpha) \\ &= \left[M_\alpha^{22}\right]^{-1} M_\alpha^{21} x^1(\alpha) + x^2(\alpha) = \left[M_\alpha^{22}\right]^{-1} b^2. \end{aligned} \tag{5.7}$$

Moreover,

$$x^2(\beta) = y^2(\beta) + P(\beta) x^1(\beta), \tag{5.8}$$

which gives

$$M_\beta^{11}x^1(\beta)+M_\beta^{12}y^2(\beta)+M_\beta^{12}P(\beta)x^1(\beta)=b^2, \tag{5.9}$$

and so

$$y^1(\beta)=x^1(\beta)=\left[M_\beta^{11}+M_\beta^{12}P(\beta)\right]^{-1}\left(b^2-M_\beta^{12}y^2(\beta)\right). \tag{5.10}$$

From Properties 5.5 and 5.6 it follows that (5.7) and (5.10) give well defined and stably computed initial values to start the computation of (3.17a) and (3.17b). Hence we can perform Step II of (3.18). Since the BCs are used in such a special way, we can omit Step III. However, if we would not be able to use the special $P(\alpha)$ (see also below) or if we had more general BCs than (5.4), we could use (3.17) to compute both a fundamental solution and some particular solutions and determine the proper linear combination to satisfy the BC. For partially separated BCs, this leads to a generalization of invariant imbedding along the lines of the Godunov–Conte algorithm. In §7.3, we treat such variants in the framework of multiple shooting. Finally, Step IV is very simple, for back transformation is needed for x^2 only and

$$x^2(t)=y^2(t)+P(t)x^1(t). \tag{5.11}$$

We conclude this subsection with some marginal notes: The main part of the algorithm, in terms of the computational labour, seems to be the computation of the nonlinear matrix valued Riccati ODE. This may limit its use for higher order problems. Another problem is that P may become unbounded. It has sometimes been advocated then to try integration in the backward direction, but there is no guarantee that this will work. The Riccati transform, despite its mathematical elegance, is rather awkward from a geometrical point of view. Recall our derivation of the general algorithm in §3, where an appropriate decoupling takes place in terms of solution vectors. More precisely span($T^1(t)$) and span($\Phi^1(t)$) become closer as t increases. By our choice of

$$T^1(t)=\begin{bmatrix} I \\ P(t) \end{bmatrix}$$

for all t, we require the unstable part of the solution to have approximately the same direction as span ($T^1(t)$) and to have a nonsingular $\Phi^{11}(t)$ block. If the directions of the solutions are varying this may be no longer true. In [23] it is shown how the Riccati solution may be restarted at points where such a phenomenon threatens to take place; this leads to a strategy that resembles multiple shooting. (Then the nice choice of $P(\alpha)$ no longer has special advantages!) Unlike in multiple shooting, however, the new starting points are chosen according to the "speed" by which these directions change. This might sometimes make invariant imbedding a more powerful algorithm, especially if this "speed" is not too high compared to the activity of the unstable modes (as in some singular perturbation problems). In multiple shooting, if the integration works at all, one has to choose many shooting points in order to limit the solution growth on an interval (see §7). We illustrate this by an example.

Example 5.12. Consider the ODE

$$\frac{dx}{dt}=\begin{bmatrix} 1-\rho\cos 2\omega t & \omega+\rho\sin 2\omega t \\ -\omega+\rho\sin\omega t & 1+\rho\cos 2\omega t \end{bmatrix}x, \qquad 0\leqq t\leqq\pi. \tag{5.13}$$

A fundamental solution is given by

$$\Psi(t)=\left(\Psi^1(t)|\Psi^2(t)\right)=\begin{bmatrix}\sin\omega t & -\cos\omega t\\ \cos\omega t & \sin\omega t\end{bmatrix}\operatorname{diag}(e^{\rho t},e^{-\rho t}). \tag{5.14}$$

Since the general solution x is $x=\Psi\binom{a}{b}$ (for $a,b\in\mathbb{R}$),

$$P(t)=x^2(t)\left[x^1(t)\right]^{-1}=\frac{a\cos\omega te^{\rho t}+b\sin\omega te^{-\rho t}}{a\sin\omega te^{\rho t}-b\cos\omega te^{-\rho t}}. \tag{5.15}$$

If $a\neq 0$, which we need for consistency, P has poles at all points t where

$$\frac{b}{a}=e^{2\rho t}\operatorname{cotan}\omega t. \tag{5.16}$$

Hence even the most contrived choice of $P(\alpha)$ cannot prevent the need of $O(\omega)$ restarts for the Riccati equation. Away from these trouble spots $P(t)$ should be quite smooth and therefore a numerical integration of (5.3) is fairly simple even for extremely large ρ. On the other hand we will see in §7 that a multiple shooting type algorithm for (5.13), will be plagued by stiffness for larger values of ρ, and not so much by problems due to ω, unless ω becomes of the order of magnitude of ρ.

In control problems, the Riccati method has an inherent meaning cf. [48], [49]. Quite often those problems involve ODEs of high dimensions and with two (or more) time scales. Roughly speaking this means that the system matrix has large eigenvalues (in modulus) as well as moderately small eigenvalues. If one is interested in steady state solutions, then the Riccati method may provide a way to estimate the slow modes. This is not only convenient from a stiffness point of view but also because the reduced system is of lower dimension and hence more tractable. In the method suggested in [49], the W^{22} block contains the absolutely larger eigenvalues and W^{11} the absolutely smaller ones, so one tries to compute

$$P(\alpha)=Q^{21}(\alpha)\left[Q^{11}(\alpha)\right]^{-1} \tag{5.17}$$

where $\operatorname{span}(Q^1(\alpha))$ is the subspace corresponding to these small eigenvalues of $L(\alpha)$. This initial vector is used to start the integration of the Riccati equation. Unless the system has constant coefficients, such a procedure may only work if the large eigenvalues are all stable. Indeed in that case the fundamental solution K, induced by $T=\binom{I}{P}\binom{\varnothing}{I}$, is expected to be consistent. Small eigenvalues correspond to solutions that dominate the fast modes (here consistency provides *relative stability*). If the problem has unstable fast modes, we may have to take recourse to other approaches, like the ones described in §6.

5.2. Order reduction for scalar ODEs. A classical analytic tool to compute a solution to a linear ODE, given one (or more) solutions, is order reduction, cf. [13]. The reasons for considering this method here are threefold. First, some literature exists about this subject. Second, order reduction yields a system of which the dominant modes may be linked with subdominant modes of the original system. Third, it is the continuous analogue of a class of algorithms that play an important role in the computation of special functions (cf. §5.3). We shall restrict our discussion to the second order scalar case (for extensions see [64]). Consider the ODE

$$\ddot{u}(t)+p(t)\dot{u}(t)+q(t)u(t)=f(t),\qquad \alpha\leqq t\leqq\beta, \tag{5.18}$$

and let ϕ be some nonvanishing solution of the homogeneous equation.

Define a function ψ by

$$u = \phi\psi. \tag{5.19}$$

Substituting this in (5.18), we find a first order (*reduced*) equation for $\dot{\psi}$:

$$\ddot{\psi} = -\left(p + 2\frac{\dot{\phi}}{\phi}\right)\dot{\psi} + \frac{f}{\phi}. \tag{5.20}$$

Suppose we obtain $\dot{\psi}$ from (5.20); then u can be found from

$$\dot{u} = u\frac{\dot{\phi}}{\phi} + \phi\dot{\psi}. \tag{5.21}$$

If $\dot{u}(\alpha)$ is given, then an initial value for $\dot{\psi}(\alpha)$, needed to integrate (5.20), follows from (5.21). However, if we also know $u(\beta)$, then we can integrate (5.21) backward.

In order to show that this algorithm is just another implementation of the decoupling idea, we rewrite (5.18) as a linear system (1.1) with

$$x := \begin{pmatrix} u \\ \dot{u} \end{pmatrix}, \quad L := \begin{pmatrix} 0 & 1 \\ -p & -q \end{pmatrix}, \quad r := \begin{pmatrix} 0 \\ f \end{pmatrix}. \tag{5.22}$$

Now define the transformation matrix

$$T := \begin{pmatrix} 1 & 0 \\ \dot{\phi}/\phi & \phi \end{pmatrix}, \tag{5.23}$$

and y by

$$y := \begin{pmatrix} u \\ \dot{\psi} \end{pmatrix}. \tag{5.24}$$

Using this T in the Lyapunov equation yields a system $\dot{y} = Wy + s$, with

$$W = \begin{pmatrix} \dot{\phi}/\phi & \phi \\ 0 & +p - 2\dot{\phi}/\phi \end{pmatrix}. \tag{5.25}$$

In fact W^{21}, the (2,1) element in W, is given by

$$W^{21} = \frac{1}{\phi}\left[\frac{d}{dt}\left(\frac{\dot{\phi}}{\phi}\right) + p + \left(\frac{\dot{\phi}}{\phi}\right)^2 + q\frac{\dot{\phi}}{\phi}\right]. \tag{5.26}$$

Since ϕ is a homogeneous solution of (5.18), $W^{21} = 0$. Thus, (5.26) actually is a Riccati ODE for $\dot{\phi}/\phi$, the "direction" of the solution ϕ, and it is tempting to try to relate T in (5.23) to a Riccati transformation $\hat{T}$, with

$$\hat{T}(t) := \begin{pmatrix} 1 & \varnothing \\ z(t) & 1 \end{pmatrix}. \tag{5.27}$$

Defining

$$\hat{y} := \hat{T}^{-1}x, \tag{5.28}$$

we arrive at the triangular system

$$\dot{\hat{y}} = \begin{pmatrix} z & 1 \\ 0 & -z - p \end{pmatrix} y + \begin{pmatrix} 0 \\ f \end{pmatrix}, \tag{5.29}$$

provided

$$\dot{z}+p+z^2+qz=0. \tag{5.30}$$

Hence the ODE $W^{21}=0$ (cf. (5.20)) is *the* Riccati equation. Note that $\hat{T}$ is a scaled version of T; in fact they are related by

$$T=\hat{T}\begin{pmatrix}1 & 0\\ 0 & \phi\end{pmatrix}. \tag{5.31}$$

Apparently $\hat{y}^2=\phi$, which explains the different (2,2) blocks in (5.23) and (5.29). Having linked order reduction and invariant imbedding, we see that a consistent initialization of (5.20) implies stability of forward integration. We also find that backward integration of (5.21) must be stable. Therefore a method as given in [64] is a special implementation of the Riccati method, with all its vices and virtues. If instead of $\dot{u}(\alpha)$ we have $u(\alpha)$ as the given initial condition, the order reduction formulation does not offer any advantage, since we are obliged to compute a fundamental solution and a particular solution of the system $\dot{y}=Wy+s$ (cf. (3.18) Step II) and determine the proper linear combination (cf. (3.18) Step III), which is roughly three times as expensive as an invariant imbedding method applied to the system

$$\begin{pmatrix}\ddot{u}\\ \dot{u}\end{pmatrix}=\begin{pmatrix}-q & -p\\ 1 & 0\end{pmatrix}\begin{pmatrix}\dot{u}\\ u\end{pmatrix}-\begin{pmatrix}f\\ 0\end{pmatrix}, \tag{5.32}$$

where a natural initialization of the "direction" ODE and the stable part of the decoupled system follow, since

$$M_\alpha=\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}. \tag{5.33}$$

5.3. Riccati methods and scalar discrete problems. Besides BVPs that arise from discretizing ODEs, one also encounters discrete BVPs that are formulated for special functions, cf. [6], [34], [39].

They involve conditionally stable recurrence relations, i.e., BVPs on infinite intervals with the requirement that the solution should remain everywhere bounded which is more or less equivalent to a terminal condition. These recurrence relations plus initial and terminal conditions lead to a banded matrix which, in principle, can be solved by an LU-decomposition. A major problem in analyzing the stability of such an approach is that often little can be said about the conditioning of this matrix and hence a (global) error analysis is difficult. An additional problem is that often the exponential behaviour of the solution also makes such a global analysis less meaningful; one would like pointwise relative error estimates rather than a bound for the norm of the error vector (cf. [39]). We shall restrict ourselves to the second order scalar recursion

$$u_{i+1}=p_iu_i+q_iu_{i-1}+f_i, \qquad i=1,2,\cdots, \tag{5.34}$$

for which u_0 (and probably u_1) is given and where it is known that $\{u_i\}_{i\geqq 0}$ is dominated by some homogeneous solution. The discrete analogue of *order reduction* cf. [45], is as follows: Let $\{\phi_i\}$ be a solution of the homogeneous part of (5.34) which is nonvanishing (for $i\geqq 1$). Now define $\{\psi_i\}$ by

$$u_i=\phi_i\psi_i, \qquad i=0,1,\cdots. \tag{5.35}$$

Then we obtain from (5.34) and (5.35) ("Abel's summation trick"):

$$\psi_{i+1}\phi_{i+1}=\psi_i\phi_{i+1}+(\psi_i-\psi_{i-1})(-\phi_{i+1}+p_i\phi_i) \tag{5.36}$$
$$+\psi_{i+1}(-\phi_{i+1}+p_i\phi_i+q_i\phi_{i-1})+f_i.$$

Since $\{\phi_i\}$ satisfies the homogeneous part of (5.34), the *weighted differences*

$$\omega_i := (\psi_{i+1}-\psi_i)\rho_i, \qquad i=0,1,\cdots \tag{5.37}$$

(where ρ_i is some nonzero real number), must satisfy the reduced recurrence relation

$$\omega_i=-q_i\frac{\phi_{i-1}}{\phi_{i+1}}\frac{\rho_i}{\rho_{i-1}}\omega_{i-1}+\frac{\rho_i}{\phi_{i+1}}f_i, \qquad i=1,2,\cdots. \tag{5.38}$$

If we know u_0 and u_1, then we know ψ_0 and ψ_1, so ω_0. Hence we can use (5.38) to compute $\{\omega_i\}$ in forward direction. On the other hand, the relation (5.37) describes a one step recursion for $\{u_i\}$, viz.

$$u_i=\frac{\phi_i}{\phi_{i+1}}u_{i+1}-\frac{\omega_i}{\rho_i}\phi_i, \tag{5.39}$$

which can be used in backward direction once $\{\phi_i\}_{i=0}^N$ and $\{\omega_i\}_{i=0}^N$ have been computed. In the particular case where one chooses

$$\phi_0=0, \tag{5.40a}$$
$$\phi_1=1, \tag{5.40b}$$
$$\rho_i=-\phi_i\phi_{i+1}, \qquad i=1,\cdots, \tag{5.40c}$$

and moreover

$$\omega_0=\phi_1u_0=u_0, \tag{5.40d}$$

we obtain "Olver's algorithm", cf. [41]. Order reduction, including "Olver's algorithm" now fits in the general framework of (3.25). To see this, we form a matrix vector recursion that corresponds to (5.34) by

$$x_{i+1} := \begin{pmatrix} u_{i+2} \\ u_{i+1} \end{pmatrix} = \begin{pmatrix} p_i & q_i \\ 1 & 0 \end{pmatrix}\begin{pmatrix} u_{i+1} \\ u_i \end{pmatrix}+\begin{pmatrix} f_{i+1} \\ 0 \end{pmatrix}, \qquad i=0,1,\cdots. \tag{5.41}$$

If we define

$$T_i := \begin{bmatrix} 1 & 0 \\ \phi_i/\phi_{i+1} & -\phi_i/\rho_i \end{bmatrix}, \tag{5.42}$$

we obtain a decoupled recursion like (3.23):

$$\begin{bmatrix} u_{i+2} \\ \omega_{i+1} \end{bmatrix} = \begin{bmatrix} \dfrac{\phi_{i+2}}{\phi_{i+1}} & \dfrac{\phi_{i+2}\phi_i}{\phi_{i+1}\rho_{i+1}} \\ 0 & -q_{i+1}\dfrac{\phi_i\rho_{i+1}}{\phi_{i+2}\rho_i} \end{bmatrix}\begin{bmatrix} u_{i+1} \\ \omega_i \end{bmatrix}+\begin{bmatrix} f_{i+1} \\ \dfrac{\phi_{i+1}}{\phi_{i+2}}f_i \end{bmatrix}. \tag{5.43}$$

As noted before, the transformation matrix T_i may be very skew unless we take $\rho_i=\phi_i$ for all i. In that case we have a traditional Riccati transformation. More important, however, is the consistency question. Assume again that there is a dichotomy (cf. [6]).

With probability one, $\binom{0}{1}$ is the initial value of an unstable mode of (5.41). Hence we see that the first column of T_i has the direction of such a mode, i.e. we should expect consistency. We conclude that computing $\{\phi_i\}_{i=0}^{N}$ (for some N) effectively corresponds to Step I of algorithm (3.25). If u_0 is given and we use Olver's implementation of order reduction (cf. (5.40)), we obtain an initial value to compute $\{\omega_i\}_{i=1}^{N}$ from (5.38); by setting $u_N=0$, we can also compute $\{u_i\}_{i=0}^{N}$ backward from (5.39). Note that a homogeneous solution here is given by $\{\phi_i\}_{i\geq 0}$; this constitutes Step II of (3.25). As with invariant imbedding, we do not need to compute a fundamental system and a particular solution with the choice of (5.40), i.e. Step III is void. For different choices of the ρ_i, however, such a step is required. Finally, whatever choice we made for the ρ_i, back transformation is not needed, since the first coordinate of the solution coincides in the original and transformed recursions. This algorithm may fail if the Lyapunov equation or some similar relation does not have a solution, in particular if $\phi_i=0$ for some $i>0$. In such a case, we need to restart this computation. However, this is not likely to happen too frequently. Indeed, one should realize that scalar recursions give rise to very special matrix vector recursions (and the same is true with scalar ODEs).

For most of the well-known orthogonal functions it is a powerful algorithm, however.

6. Transformations based on power series. One of the basic problems in decoupling an ODE is how to find reasonably well-conditioned matrices T and block upper triangular matrices W *at the same time*. An intuitively simple idea to achieve this is to try a good guess for W (like $\overline{W}$ in (3.6)) and then correct this so that the result satisfies the Lyapunov equation (3.9), at least better than $\overline{W}$ does. If the system matrix has a power series expansion, we can utilize this to obtain better and better approximations in terms of a power series. The classical work in this area is Wasow's book [68]. In this section, we are particularly interested in problems which depend on t and in a singular way on some (small) parameter ε. Such problems have been investigated in detail, cf. [3], [20], [27], [28], [40], [41], [47], [48], [49], [58]. The combination of analytic techniques with appropriate numerical tools may provide a powerful method to solve singular perturbation problems. We shall restrict ourselves to a brief discussion (for an extensive list of references, see [10]) and mainly treat this technique from a stability viewpoint.

Consider the system

$$\dot{x}(t)=\mathscr{L}(t,\varepsilon)x(t)+r(t,\varepsilon), \qquad \varepsilon \text{ small}, \tag{6.1}$$

where $\mathscr{L}(t,\varepsilon)$ has a power series expansion

$$\mathscr{L}(t,\varepsilon)=\frac{1}{\varepsilon}L(t,\varepsilon)=\frac{1}{\varepsilon}L_0(t)+L_1(t)+\varepsilon L_2(t)+\cdots. \tag{6.2}$$

We assume that the coefficient matrices L_i are smooth. If the matrix L_0 is singular, then the system in (6.1) has two time scales, viz. fast modes with $O(1/\varepsilon)$ derivatives and slow modes with $O(1)$ derivatives. We first restrict ourselves to the case where $L_0(t)$ is nonsingular for all t, so that the homogeneous part of (6.1) has only fast modes. Now if we define

$$x(t)=T(t,\varepsilon)y(t), \tag{6.3}$$

where T is a (smooth) matrix function, we hope to transform (6.1) to obtain a decoupled ODE

$$\varepsilon\dot{y}=W(t,\varepsilon)y+\varepsilon s(t,\varepsilon) \tag{6.4}$$

via the Lyapunov equation

$$\varepsilon \dot{T} = LT - TW. \tag{6.5}$$

As an Ansatz we set

$$T(t,\varepsilon) = T_0(t) + \varepsilon T_1(t) + \cdots, \tag{6.6a}$$

$$W(t,\varepsilon) = W_0(t) + \varepsilon W_1(t) + \cdots. \tag{6.6b}$$

This leads to the recursive relations (cf. [68, p. 141])

$$L_0 T_0 - T_0 W_0 = 0, \tag{6.7a}$$

$$L_0 T_l - T_l W_0 = \sum_{s=l}^{l-1} [T_s W_{l-s} - L_{l-s} T_s] + \dot{T}_{l-1}, \qquad l \geqq 1. \tag{6.7b}$$

(Note that (6.7b) is not to be regarded as ODE for T_{l-1}, but a relation for T_l in terms of T_s, $s \leqq l-1$.) Further assume that the signs of the eigenvalues of $L_0(t)$ are independent of t throughout $[\alpha, \beta]$; so we do not have turning points. Since in (6.7a) the matrix $W_0(t)$ is similar to $L_0(t)$, we use a canonical form of $W_0(t)$. In contrast to the use of Jordan forms in [68], we prefer $W_0(t)$ that are (block) upper triangular since they are fairly easy to compute in practice. Indeed, by subspace iteration or an extended form of the QR algorithm, cf. [27], [41] one can compute an orthogonal matrix T_0 and a (block) upper triangular W_0 such that

$$L_0 T_0 = T_0 W_0, \tag{6.8}$$

where W_0 has an ordered diagonal with the positive real part eigenvalues appearing first, or—in the block upper triangular case—diagonal blocks which are similar to these W_0. (Note that appropriate transformations can bring the stable eigenvalues into the upper left block.) While proceeding with t, one hopes that the eigensystem T_0 is only slowly varying, thus making $T_0(t_i)$ a good starting guess for the iterative computation of $T_0(t_{i+1})$. From [41] we derive the following adapted basic theorem.

THEOREM 6.9. *Let the eigenvalues of $L_0(t)$ be distinct for all t. Assume $T(t,\varepsilon)$ has an asymptotic expansion on $[\alpha,\beta]$. Then there exists a fundamental solution $\Phi(t,\varepsilon)$ such that*

$$\Phi(t,\varepsilon) \sim T_0(t) \exp\left[\frac{1}{\varepsilon}\int_0^t W_0(s)\,ds\right], \qquad \varepsilon \to 0.$$

Hence the choice in (6.8) induces the proper dichotomy provided ε is small enough. In [41] a description is given of how to compute subsequent W_s and T_s form (6.7b). It should be realized, however, that it will often be quite satisfactory to restrict oneself to the first order term, since ε^2 will often be smaller than the required numerical tolerance. In fact, in [27] only the zeroth order terms are taken into account. After discretization, we obtain a sequence of transformation matrices $\{T_0(t_i) + \varepsilon T_1(t_i)\}$ and block upper triangular matrices $\{W_0(t_i) + \varepsilon W_1(t_i)\}$ for a first order approximation, corresponding to Step I of (3.25). The remaining three steps are exactly as in the basic algorithm. It goes without saying that consistency and hence stability is assured if ε is small enough (cf. Theorem 6.9).

Quite often the matrix L_0 is singular, so there are also slow modes. If, in particular, $L_0^{22}(t) = 0$ for all t, then the homogeneous system (6.1) can more conveniently be

written upon rescaling x^2 as

$$\varepsilon \dot{x}^1 = L^{11}x^1 + L^{12}x^2, \tag{6.10a}$$

$$\dot{x}^2 = L^{21}x^1 + L^{22}x^2, \tag{6.10b}$$

see [41]. In such a system it is reasonable to assume that $L^{11}(t)$ is invertible for all t. Although decoupling of two time scales is a somewhat different problem than decoupling increasing and decreasing modes (which seems to be imperative in order to have a stable algorithm), we shall still treat it here because of the nice similarity with our problem setting. Consider the Riccati transformation

$$T(t,\varepsilon) = \begin{bmatrix} I & \varnothing \\ \varepsilon P(t,\varepsilon) & I \end{bmatrix}; \tag{6.11}$$

then we obtain a time scale decoupled block upper triangular system

$$\varepsilon \dot{y}^1 = W^{11}(t,\varepsilon)\, y^1 + W^{12}(t,\varepsilon)\, y^2, \tag{6.12a}$$

$$\dot{y}^2 = W^{22}(t,\varepsilon)\, y^2, \tag{6.12b}$$

provided P satisfies the (*singularity perturbed*) Riccati equation (cf. (5.3))

$$\varepsilon \dot{P} = -PL^{11} + L^{21} + \varepsilon L^{22}P - \varepsilon PL^{12}P. \tag{6.13}$$

If we try a power series expansion for P

$$P(t,\varepsilon) = P_0(t) + \varepsilon P(t) + \cdots, \tag{6.14}$$

we have

$$P_0 = L^{21}[L^{11}]^{-1}, \tag{6.15}$$

and

$$\dot{P}_0 = L^{22}P_0 - P_1L^{11} - P_0L^{12}P_0. \tag{6.16}$$

Since $\dot{P}_0$ can be found from differentiation in (6.15), i.e.,

$$\dot{P}_0 = \dot{L}_{21}[L^{11}]^{-1} - L^{21}[L^{11}]^{-1}\dot{L}^{11}[L^{11}]^{-1}, \tag{6.17}$$

we thus can derive P_1 analytically. If we are again satisfied with first order approximations, we may use

$$W(t,\varepsilon) \doteq \begin{bmatrix} L^{11} + L^{412}(P_0 + \varepsilon P_1) & L^{12} \\ \varnothing & -(P_0 + \varepsilon P_1)L^{12} + L^{22}. \end{bmatrix}. \tag{6.18}$$

Such a decoupling has been suggested in [3], [40]. If we would relate consistency of a fundamental solution of (6.1) to the fast mode part, then it can be seen that this Riccati transformation should produce a consistent transformation for ε small enough. In fact, we can expect

$$\lim_{\varepsilon \downarrow 0} T(t,\varepsilon) = I. \tag{6.19}$$

This shows that the Riccati equation is always solvable, because we do not have significant rotation of the fast mode part.

7. Special implementations of multiple shooting. In this section we consider three special implementations of multiple shooting. The *global strategy* of this method will be described first. Then we consider the Godunov–Conte variant for *separated* BCs and its generalization to *partially separated* BCs. Finally, it is shown how *general* BCs can be handled by a decoupling algorithm.

7.1. Multiple shooting. Multiple shooting is an adaptation of single shooting, since it solves a number of initial value problems in order to make the interval of integration, and therefore the error growth, smaller. Although this method is quite well-known, cf. [16], [17], [23], [24], [42], [51], we give a short description for better understanding of the subsequent implementations. Let the interval $[\alpha,\beta]$ be divided into N subintervals $[t_i,t_{i+1}]$, $i=0,\cdots,N-1$. Then over each such subinterval both some fundamental solution F_i and some particular solution p_i are computed. The desired solution x can be found from superposition of these solutions, i.e., for $i=0,\cdots,N-1$ there exists a vector v_i such that

$$x(t)=F_iv_i+p_i(t) \quad \text{on } [t_i,t_{i+1}]. \tag{7.1}$$

By matching at the "shooting points" t_i we obtain the recursion

$$F_{i+1}(t_{i+1})v_{i+1}=F_i(t_{i+1})v_i+p_i(t_{i+1})-p_{i+1}(t_{i+1}), \qquad i=0,\cdots,N-1. \tag{7.2}$$

In order to obtain the sequence $\{v_i\}$ the following BC must be satisfied

$$M_\alpha F_0(t_0)v_0+M_\beta F_{N-1}(t_N)v_{N-1}=\tilde{b}:=b-M_\alpha p_0(t_0)-M_\beta p_{N-1}(t_N). \tag{7.3}$$

Sometimes it is preferable to write (7.2), (7.3) as the linear system

$$\begin{bmatrix} F_0(t_1) & -F_1(t_1) & & & \\ & F_1(t_2) & -F_2(t_2) & & \varnothing \\ \varnothing & & \ddots & \ddots & \\ & & & F_{N-2}(t_{N-1}) & -F_{N-1}(t_{N-1}) \\ M_\alpha F_0(t_0) & & & & M_\beta F_{N-1}(t_N) \end{bmatrix} \begin{bmatrix} v_0 \\ \vdots \\ \\ \\ \\ v_{N-1} \end{bmatrix} = \begin{bmatrix} p_1(t_1)-p_0(t_1) \\ p_2(t_2)-p_1(t_2) \\ \vdots \\ \\ \\ p_{N-1}(t_{N-1})-p_{N-2}(t_N) \\ \tilde{b} \end{bmatrix}. \tag{7.4}$$

It is worth noting that the system (7.4) also describes the discrete problem for most other methods including so-called "direct" and "global" methods. Those methods mainly differ from multiple shooting in two ways. First, they need some global strategy to determine the grid before discretization takes place. Second, the system generally contains many more unknowns since the grid $\{t_0,\cdots,t_N\}$ coincides more or less with the entire mesh where the solution is discretized (for collocation this is true for the matching points between elements).

7.2. The Godunov–Conte algorithm. If the BCs are separated, cf. (2.14), some savings in CPU time can be gained by integrating only suitable parts of the fundamental solutions. This idea was first suggested by Godunov and later developed by Conte, cf. [14]. Part of its popularity is due to a FORTRAN implementation described in [60]. Analyses of this algorithm, as well as the version given in the next section, can be found in [33], [42], [52]. We first describe the algorithm: Let F_0^1 be a homogeneous matrix solution, where $F_0^1(t)$ consists of k columns such that

$$ {}^2M_\alpha F_0^1(t_0)=0, \tag{7.5} $$

and $F_0^1(t_0)$ has orthonormal columns. At the same time let p_0 be a particular solution of (1.1) which satisfies

$$ {}^2M_\alpha p_0(t_0)=b^2. \tag{7.6} $$

If, for some reason, the designer of a code decides to stop the integration of p_0 and F_0^1 at some point t_1 (in order, say, to restrict the error growth) a new particular solution p_1 and a new partial fundamental solution F_1 are computed, starting at t_1. The important special feature now is to ensure that $\operatorname{span}(F_1^1(t_1))=\operatorname{span}(F_0^1(t_0))$ and that $\operatorname{span}(F_0^1(t_1))\oplus p_0(t_1)=\operatorname{span}(F_1^1(t_1))\oplus p_1(t_1)$. This is done by a QU-decomposition of $F_0^1(t_1)$, viz.,

$$ F_0^1(t_1)=F_1^1(t_1)B_0, \tag{7.7} $$

where $F_1^1(t_1)$ has orthonormal columns and B_0 is upper triangular. At the same time $p_0(t_1)$ is reduced by subtracting its projection on span $(F_1^1(t_1))$, thus resulting in

$$ p_1(t_1)=p_0(t_1)-F_1^1(t_1)\left[F_1^1(t_1)\right]^T p_0(t_1), $$

which is orthogonal to $\operatorname{span}(F_1^1(t_1))$. In this way we proceed until we reach $t_N:=\beta$. In doing so we have produced a recursive relation between the successive fundamental solutions as follows:

$$ F_i^1(t_{i+1})=F_{i+1}^1(t_{i+1})B_i, \tag{7.8} $$

where $F_{i+1}^1(t_{i+1})$ has k orthogonal columns and B_i is upper triangular. The particular solutions are related by

$$ p_{i+1}(t_{i+1})=p_i(t_{i+1})-F_{i+1}^1(t_{i+1})\left[F_{i+1}^1(t_{i+1})\right]^T p_i(t_{i+1}). \tag{7.9} $$

By matching as in (7.2) we find a recursion for the vectors v_i^1, defined by

$$ x(t)=F_i^1(t)v_i^1+p_i(t), \qquad i=0,\cdots,N-1. \tag{7.10} $$

The vector v_{N-1}^1 can be found from the BC by solving

$$ \left[{}^1M_\beta F_{N-1}^1(t_N)\right]v_{N-1}^1=b^1-{}^1M_\beta p_{N-1}(t_n). \tag{7.11} $$

Once v_{N-1}^1 is know, we can compute $v_{N-2}^1,\cdots,v_0^1$ from

$$ \begin{aligned} v_{i+1}^1&=B_i v_i^1+\left[F_{i+1}^1(t_{i+1})\right]^T\left(p_i(t_{i+1})-p_{i+1}(t_{i+1})\right)\\ &=B_i v_i^1+\left[F_{i+1}^1(t_{i+1})\right]^T p_i(t_{i+1}). \end{aligned} \tag{7.12} $$

We now show that this is a decoupling algorithm which is stable because of consistency. If we formally complete the solutions F_i^1 to a full fundamental solution F_i by requiring

that span$(F_i^2(t_i))$ be orthogonal to span$(F_i^1(t_i))$, we obtain the factorization

$$F_i(t_{i+1}) = F_{i+1}(t_{i+1})U_i, \tag{7.13}$$

so U_i is block upper triangular. We also introduce vectors

$$g_i := [F_{i+1}(t_{i+1})]^T(p_i(t_{i+1}) - p_{i+1}(t_{i+1})) \tag{7.14}$$

and

$$y_i := [F_i(t_i)]^{-1}x_i \tag{7.15}$$

so that (7.2) implies

$$y_{i+1} = U_i y_i + g_i. \tag{7.16}$$

From this we derive

PROPERTY 7.17. *Assume the problem is well-conditioned. Let* $[F_i^1(t_i)]^T F_i^2(t_i) = 0$. *Then*

(i) ${}^2M_\alpha F_0^2(t_0)$ *is nonsingular.*

(ii) $y_0^2 = [{}^2M_\alpha F_0^2(t_0)]^{-1}b^2$, *and*
$y_i^2 = [[F_i(t_i)]^{-1}p_i(t_i)]^2 = [[F_i(t_i)]^{-1}p_{i-1}(t_i)]^2$, $i = 0, \cdots, N-1$.

(iii) ${}^1M_\beta F_{N-1}^1(t_N)$ *is nonsingular.*

(iv) $y_N^1 = [{}^1M_\beta F_{N-1}^1(t_N)]^{-1}\{b^1 - {}^1M_\beta F_{N-1}^2(t_N)y_N^2\}$ *and* $y_i^1 = v_i^1$, $i = 0, \cdots, N$.

Proof. Assertion (i) follows since ${}^2M_\alpha$ has full row rank and the space spanned by *its* rows must be identical to span$(F_0^2(t_0))$. For the other assertions, it is useful to realize that

$$[F_i(t_i)]^{-1} = \begin{pmatrix} [F_i^1(t_i)]^T \\ [G_i^2]^T \end{pmatrix},$$

where span(G_i^2) = span$(F_i^2(t_i))$. Hence

$$y_i^2 = [G_i^2]^T x(t_i) = [G_i^2]^T\{F_i^1(t_i)v_i^1 + p_i(t_i)\} = [G_i^2]^T p_i(t_i) = [[F_i(t_i)]^{-1}p_i(t_i)]^2.$$

Similarly, $y_i^2 = [G_i^2]^T\{F_{i-1}(t_i)v_{i-1}^1 + p_{i-1}(t_i)\} = [G_i^2]^T p_{i-1}(t_i) = [[F_i(t_i)]^{-1}p_{i-1}(t_i)]^2$. Moreover, it is straightforward to see that y_0^2 should satisfy $[{}^2M_\alpha F_0^2(t_0)]y_0^2 = b^2$. Assertion (iii) can be proven similar to Property 5.6 using the well-conditioning. For the first components y_i^1, we realize that $g_i^1 = [F_{i+1}^1(t_{i+1})]^T(p_i(t_{i+1}) - p_{i+1}(t_{i+1}))$ so that $\{y_i^1\}$ and $\{v_i^1\}$ satisfy the same recursion. Finally it is straightforward to check the terminal condition for y_N^2. □

As can be seen from Property 7.17, the Godunov–Conte algorithm fits nicely in the framework of (3.25). The clever point is that the transformation $T_i = F_i(t_i)$ is not computed completely. Indeed, the actual choice of the last $n-k$ columns of T_i does not matter as long as they are taken orthogonal to $F_i^1(t_i)$. Here consistency follows from the well-conditioning and the use of *separated* BCs (cf. Theorem 4.10). *In particular this implies that the backward recursion* (7.12) *should be stable* (cf. [35, Ex. 6.5]).

Finally note that the use of the BC and a consequence the resulting consistency are almost identical to what we found for invariant imbedding.

7.3. Orthonormalization and partially separated BCs. The important point in the strategy of §7.2 was using the zero rows in M_β in order to find an explicit initial value for a particular solution p_0 being in the same k-dimensional linear variety as the desired solution x. It is not surprising that this idea can be generalized to BCs with a few zero rows in M_β, not necessarily complementary to the zero rows in M_α. Keller [23] called such BCs *partially separated* BCs. Descriptions and an error analysis of such algorithms can be found in [52]. Because of its similarity with the Godunov–Conte algorithm we do not elaborate here but only note that the row partitioning should now be such that

$$^2M_\beta = 0. \tag{7.18}$$

The notations of §7.2 immediately carry over, except the BCs now contain values of the transformed solution at both ends. Instead at (7.11) we now have

$$^1M_\alpha F_0^1(t_0)v_0^1 + {}^1M_\beta F_{N-1}^1(t_N)v_{N-1}^1 = b^1 - {}^1M_\alpha p_0(t_0) - {}^1M_\beta p_{N-1}(t_N). \tag{7.11}*$$

Obviously, we cannot use Theorem 4.10 to check consistency. In fact we do not have such a property in general.

Example 7.19. Consider the ODE

$$\dot{x} = \begin{pmatrix} -10 & 0 & 0 \\ 0 & -20 & 0 \\ -20 & 0 & 10 \end{pmatrix} x + \begin{pmatrix} 10 \\ 20 \\ 10 \end{pmatrix} \tag{7.20}$$

and the BC

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} x(0) + \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} x(0) = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}, \tag{7.21}$$

with solution $x(t) \equiv (1,1,1)^T$. It is simple to see that (7.20) has an (unnormalized) fundamental solution:

$$\Psi(t) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix} \operatorname{diag}(e^{-20t}, e^{-10t}, e^{10t}). \tag{7.22}$$

Moreover, the problem is well-conditioned. If we use the max-norm, then the conditioning matrix

$$Q \approx \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}; \tag{7.23}$$

so κ (cf. (4.3)) is approximately 2. If we take

$$F_0^1(0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix},$$

$\operatorname{span}(F_0^1(0))$ is orthogonal to the rows of $^2M_0$. Nevertheless the second column of $F_0^1(0)$ is the initial value of a mode that decreases like $\exp(-20t)$, whereas the first one is the initial value of a solution that grows like $\exp(-10t)+\exp(10t)$. Hence we certainly do not have consistency. Therefore, if no errors are made, we should expect the upper triangular matrices B_i to have a $(1,1)$ element equal to $\exp(-20(t_{i+1}-t_i))$, thus making backward recursion an unstable affair.

One should realize that there is no good reason to use (7.12) directly as we have to satisfy the BC (7.11). *The gain in this approach is that we have reduced the order of the blocks* in the recurrence relation for the v_i to l, the number of coordinates of v_i^1 compared to "complete" multiple shooting. Rather than (7.4), we now have the simpler linear system

$$\begin{bmatrix} B_0 & -I & & & & \\ & B_1 & -I & & & \\ & & \ddots & \ddots & & \\ & & & B_{N-2} & -I & \\ \overline{M}_\alpha & & & & \overline{M}_\beta & \end{bmatrix} \begin{bmatrix} v_0^1 \\ \vdots \\ \\ v_{N-1}^1 \end{bmatrix} = \begin{bmatrix} g_0^1 \\ \vdots \\ g_{N-2}^1 \\ \bar{b}^1 \end{bmatrix}, \tag{7.24}$$

where $\overline{M}_\alpha = {}^1M_\alpha F_0^1(t_0)$, $\overline{M}_\beta = {}^1M_\beta F_{N-1}^1(t_N)$ and $\bar{b}^1 - {}^1M_\alpha p_0(t_0) - {}^1M_\beta p_{N-1}(t_N)$ (cf. (7.11)). Questions concerning the stability of solving (7.24) will be dealt with in §8. We now want to show that well-conditioning implies stability of the above strategy. The crucial point is that any possible unstable component in the (orthogonal) complementary part of $F_i^1(t)$ also occurs in $F_i^1(t)$. For if this were not the case, 7.17 implies that y_i^2 would contain unstable modes (projection onto the "dominant" space would not "remove" these components). Thus we have the following generalization of Theorem 4.10.

PROPERTY 7.25. *Let the* BVP *be well-conditioned, let* ${}^2M_\beta = 0$, *and let* ${}^2M_\alpha F_0^1(t_0) = 0$. *Setting* $F_0(t_0) = \Phi(t_0)H$, *let* $\hat{H}$ *be the* $l \times l$ *principal submatrix of* H *having the order* l *of the column rank of* Φ_u^1 (*cf.* (4.4)). *Then* $\hat{H}$ *is nonsingular.* (Note that $l < k$.)

Proof. (Sketchily). If rank $(\hat{H}) < l$, $F_0^1(t)$ consists of less than l unstable solutions (whether or not polluted by components of "moderate" or "stable" solutions, cf. (4.4)). It is no restriction then to suppose that there exists some basis solution in span (Φ_u^1), but not in span(F_0^1), whence span(F_0^1) contains at most an $(l-1)$-dimensional unstable solutions space and there are $n-(l-1)$ moderate or stable solutions in span(F_0^1). But it follows from the well-conditioning that no vector in the orthogonal complement of span$(F_0^1(t_0))$, viz., span$(F_0^2(t_0))$, can be almost orthogonal to an $(n-k)$-dimensional subspace of the $(n-l)$-dimensional space of initial values of moderate and stable solutions. This implies that we would have a subspace of such moderate and stable solutions of dimension $n-l+1$, which contradicts the assumption. □

The complicated argumentation in Property 7.25 indicates that the stability of this generalized Godunov–Conte algorithm is a delicate matter. It shows a generalization of the consistency concept: *we always compute the unstable solution space at least*! Moreover, this strategy excludes the growth of the orthogonal complement of the basis solutions F_i^1. In fact, it is sufficient to show that $\{y_i^2\}$ (see also Property 7.17(ii)) does not grow like an unstable solution:

PROPERTY 7.26. *Let* $F_i(t_i)$ *satisfy* $[F_i^1(t_i)]^T F_i^2(t_i) = 0$, *thereby inducing* U_i *as in* (7.13). *If the* BVP *is well-conditioned, there exists a moderate constant* γ *such that* $\|\prod_{j=0}^{i} U_j^{22}\| \leqq \gamma$, *for all* i.

Proof. (cf. Theorem 3.14). Let $F_0(t_0) = \Phi(t_0)H$ as in Property 7.25. Since $\hat{H}$ has full rank, by a suitable permutation of columns of $(\Phi_m^1 | \Phi_m^2 | \Phi_s^2)$, we can make sure that H^{11} is nonsingular. Arguing as for Theorem 3.14 but replacing span(Φ^1) by span(Φ_u^1) plus some suitable "not unstable" space and span(Φ^2) by some complementary "unstable" space, say, span(Φ_s^2), then gives $\|\prod_{j=0}^{i-1} U_j^{22}\| \leqq \gamma \|\Phi_s^2(t_i)\|$ for some γ. □

COROLLARY 7.27. *The computation of the* $p_i(t_i)$ *is stable* (*i.e.* $\{p_i(t_i)\}$ *does not grow like an unstable solution*).

Proof. Define

$$[F_{i+1}(t_{i+1})]^{-1} = \begin{pmatrix} [F_{i+1}^1(t_{i+1})]^T \\ [G_{i+1}^2]^T \end{pmatrix}.$$

Then from (7.9), Property 7.17

$$[F_{i+1}(t_{i+1})]^{-1} p_{i+1}(t_{i+1}) = \begin{pmatrix} [F_{i+1}^1(t_{i+1})]^T p_{i+1}(t_{i+1}) \\ y_{i+1}^2 \end{pmatrix}$$

$$= \begin{pmatrix} [F_{i+1}^1(t_{i+1})]^T p_i(t_{i+1}) \\ y_{i+1}^2 \end{pmatrix} - \begin{pmatrix} [F_{i+1}^1(t_{i+1})]^T p_i(t_{i+1}) \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ y_{i+1}^2 \end{pmatrix}.$$

Hence $\| p_{i+1}(t_{i+1})\|_2 = \|F_{i+1}(t_{i+1})\|_2 \|y_{i+1}^2\|$. It is not restrictive to choose $F_{i+1}(t_{i+1})$ orthogonal. Now since $\{y_i^2\}$ satisfies a stable recursion (cf. Property 7.26) it follows that $\{p_i(t_i)\}$ does not grow faster than the particular solution of the lower right block of the recursion (7.16). □

Remark 7.28. In [60] the authors raise the question whether or not one should normalize the particular solution $p_i(t_i)$ as well. Corollary 7.27 shows that there is no need for this if the problem is well-conditioned.

7.4. General BCs. In [42] a method is suggested to employ decoupling and a special recursion technique to solve a BVP via multiple shooting. Omitting the details about how the integration is performed and the shooting points are selected, it can be seen as a step further in the Godunov–Conte algorithm. Theoretically it is based on finding a suitable sequence of matrices which transform the incremental matrices $F_{i+1}(t_{i+1})[F_{i+1}(t_i)]^{-1}$ onto upper triangular form. Again, the upper triangular matrix is obtained directly, i.e., the untransformed increment never appears.

The method starts off with some initial transformation matrix Q_0. This is used to generate a fundamental solution F_0 with

$$F_0(t_0) := Q_0. \tag{7.29}$$

At the next shooting point we decompose

$$F_0(t_1) =: Q_1 U_0, \tag{7.30}$$

where Q_1 is orthogonal and U_0 upper triangular. This is done using elementary hermitians (Householder's method). Now it is most important to have an ordered diagonal in U_0. Therefore, if the diagonal elements of U_0 do not appear in decreasing modulus from above to below, an appropriate permutation matrix P_0 is constructed which premultiplies Q_0. The result $U_0 P_0$ is then decomposed again as

$$U_0 P_0 =: P_1 \hat{U}_0. \tag{7.31}$$

Suppose $\hat{U}_0$ is in order; then the fundamental solution on the next shooting interval should satisfy

$$F_1(t_1) = P_1 Q_1 \tag{7.32}$$

with $P_1 = I$ if $U_0 = \hat{U}_0$. Since there usually is no idea about a meaningful Q_0, take as a first guess $Q_0 = I$. Without loss of generality, let $Q_0 = I$ be correct. Then this recursive

computation of the $\{Q_i\}$ and $\{U_i\}$ gives a decoupling algorithm like (3.25). As was shown in [32], if there is a dichotomy, this strategy gives an upper triangular recursion where the upper left blocks correspond to the unstable and the lower right blocks correspond to the stable incremental values. This is now used to compute a fundamental solution and a particular solution in a stable way. (Step II of (3.25)). To start with, let the particular solution $\{z_i\}_{i=0}^N$ satisfy the BC

$$z_0^2 = 0, \qquad z_N^1 = 0. \tag{7.33}$$

For the fundamental solution, $\{\Omega_i\}_{i=0}^N$, we choose

$$\Omega_0^2 = (\varnothing \mid I_{n-k}), \qquad \Omega_N^1 = (I_n \mid \varnothing). \tag{7.34}$$

Apparently $\Omega_i^{21} = 0$ for all i, while Ω_N^{22} is found by forward recursion via the homogeneous part of (3.23b), and Ω_0^{11} and Ω_0^{12} are found by using the homogeneous part of (3.23a) in the backward direction. For some fixed vector a, we should have

$$Q_i^{-1} x(t_i) = z_i + \Omega_i a. \tag{7.35}$$

In Step III of (3.25), we compute a from

$$\left[M_\alpha Q_0 \Omega_0 + M_\beta Q_N \Omega_N\right] a = b - M_\alpha Q_0 z_0 - M_\beta Q_N z_N, \tag{7.36}$$

with the matrix appearing in (7.36) as well-conditioned as the problem itself, cf. [42, Thm. 4.3]:

PROPERTY 7.37. $\|[M_\alpha Q_0 \Omega_0 + M_\beta Q_N \Omega_N]^{-1}\| \leqq 2\mathscr{K}$.

In §8.3 we shall return to this algorithm. We conclude this section by giving a numerical example to demonstrate the impact of conditioning on consistency and, hence, stability.

Example 7.38. Consider the ODE (cf. [42, Ex. 5.1])

$$\frac{dx}{dt} = \begin{bmatrix} 1-19\cos 2t & 0 & 1+19\sin 2t \\ 0 & 19 & 0 \\ 1+19\sin 2t & 0 & 1+19\cos 2t \end{bmatrix} x + f(t), \tag{7.39}$$

where $f(t) = e^t(-1+10(\cos 2t - \sin 2t),\ -18,\ 1-19\ (\cos 2t + \sin 2t))^T$. A fundamental solution is given by

$$\Psi(t) = \begin{bmatrix} \sin t & 0 & -\cos t \\ 0 & 1 & 0 \\ \cos t & 0 & \sin t \end{bmatrix} \operatorname{diag}(e^{20t}, e^{19t}, e^{-18t}), \tag{7.40}$$

and a particular solution by

$$x(t) = e^t(1,1,1)^T. \tag{7.41}$$

Let the BC be

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} x(0) + \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} x(\pi) = \begin{pmatrix} 1+e^{-\pi} \\ 1+e^{-\pi} \\ 1 \end{pmatrix}. \tag{7.42}$$

It is straightforward to see that the condition number $\kappa = 1.000$ (for κ cf. (4.3)). We computed the solution x at 10 equally spaced points with our code MUTS, cf. [42]. The algorithm on which this code is based uses a special implementation of the method outlined in this subsection. The code automatically computes the conditioning matrix

$[M_\alpha Q_0 \Omega_0 + M_\beta Q_N \Omega_N]^{-1}$. We found for any tolerance up to 10^{-10}, errors bounded by this tolerance and a numerical estimate for κ equal to 1 (up to the tolerance). We also used a modified version capable of employing a zero row of structure in M_α or M_β. The results were similar.

We also tested the BC

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} x(0) + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} x(\pi) = \begin{pmatrix} 1+e^{\pi} \\ 1+e^{\pi} \\ 1 \end{pmatrix}. \tag{7.43}$$

Here, the matrix M_β does not control the mode growing like e^{20t}, and we have a condition number $\approx e^{20\pi} = 1.9\ 10^{27}$. MUTS gave an error message that the matrix in (7.36) was numerically singular. The choice

$$F_0^1(0) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{7.44}$$

induces a consistent fundamental solution. However, the ill-conditioning is inherited by the reduced problem since the partial terminal matrix does not control the most unstable mode.

Finally, we tested the BC

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} x(\pi) + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} x(0) = \begin{pmatrix} 1+e^{\pi} \\ 1+e^{\pi} \\ 0 \end{pmatrix}. \tag{7.45}$$

(Mathematically and numerically, interchanging the initial and terminal points does not matter.) As can be seen now the unstable mode (in backward direction) is not controlled by this BC; hence $\kappa \approx e^{18\pi} = 3.6 \cdot 10^{24}$. Again MUTS gave an error message. The adapted version for partially separated BCs did give numerical results, though with $O(1)$ errors. The explanation for this is simple, but extremely instructive. At the "initial" point $t_0 = \pi$ the columns of

$$F_0(t_0) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{7.46}$$

are the "initial" values ($t_0 = \pi$) of the stable solutions, growing like e^{-20t} and e^{-19t} (as t runs from π to 0). Necessarily the particular solution induced by $p_0(t_0)$ contains a component of the unstable mode which grows like e^{+18t}. This instability illustrates the need for the well-conditioning assumption in Property 7.6. Because we actually deal with a problem with perturbations of the order of the tolerance (tol), we should expect that after the point $\tilde{t}$ such that $\exp(18\tilde{t}) \approx 1/\text{tol}$, the numerically computed fundamental solutions $F_i^1(t)$ with be again "consistent". Instead of a theoretical condition number $\exp(18(\pi - 0))$, we have a numerically relevant condition number $\exp(18(\pi - \tilde{t}))$. As a consequence, the local errors are amplified by $(\text{tol})^{-1}$, which explains the $O(1)$ errors. This phenomenon is closely related to the observations in [42, Ex. 5.4].

8. Solution of linear algebraic systems. We saw that multiple shooting gives rise to a linear system with a special sparsity structure. A similar system is found for many other methods, e.g. collocation. The special block structure has inspired a number of algorithms that aim at an LU-decomposition without loss of zero blocks, cf. [8], [18], [25], [37], [38], [66]. We shall show that such a strategy and, in fact, any successful

partial pivoting strategy is related to an appropriate decoupling method. Before that, however, we consider a simpler matrix structure, viz., a (block) tridiagonal form such as appears when we use a three point difference scheme for a second order ODE.

8.1. Block tridiagonal matrices. If we have a scalar second order ODE with Dirichlet BC, we may use a central difference scheme to find an approximate solution. The resulting three point recursion plus the BC give rise to a tridiagonal system. We now consider, more generally, such block systems

$$\mathbf{A}\mathbf{x}=\mathbf{b}, \tag{8.1a}$$

where

$$\mathbf{A}=\begin{bmatrix} C_0 & D_0 & & & \varnothing \\ B_1 & C_1 & D_1 & & \\ & \ddots & \ddots & \ddots & \\ & & & & D_{N-1} \\ \varnothing & & & B_N & C_N \end{bmatrix}. \tag{8.1b}$$

Assmption 8.2. Let the blocks D_i in (8.1b) be nonsingular.

(This assumption is trivially fulfilled for the central difference scheme.) Suppose we can decompose $\mathbf{A}$ as

$$\mathbf{A}=\mathbf{L}\mathbf{U}, \tag{8.3}$$

where

$$\mathbf{L}:=\begin{bmatrix} I & & & & \varnothing \\ L_1 & I & & & \\ & L_2 & I & & \\ & & \ddots & \ddots & \\ \varnothing & & & L_N & I \end{bmatrix}, \tag{8.4}$$

$$\mathbf{U}:=\begin{bmatrix} U_0 & D_0 & & & \varnothing \\ & U_1 & D_1 & & \\ & & \ddots & \ddots & \\ & & & & D_{N-1} \\ \varnothing & & & & U_N \end{bmatrix}. \tag{8.5}$$

Then the following relations must hold (cf. [37]).

$$U_0=C_0, \tag{8.6a}$$

$$L_iU_{i-1}=B_i, \qquad i=1,\cdots,N, \tag{8.6b}$$

$$L_iD_{i-1}+U_i=C_i, \qquad i=1,\cdots,N. \tag{8.6c}$$

Note that if $\mathbf{A}$ is nonsingular and $\mathbf{L}$ and $\mathbf{U}$ can be found, then all U_i are nonsingular. Hence we obtain for the pivot sequence $\{U_i\}$,

$$U_0=C_0, \tag{8.7a}$$

$$U_i=C_i-B_iU_{i-1}^{-1}D_{i-1}, \qquad i=1,\cdots,N. \tag{8.7b}$$

We shall show that the computation of **L** and **U** and the resulting forward and backward substitution can be viewed as a decoupling algorithm. To this end, we write

$$A_i := \begin{pmatrix} -D_{i+1}^{-1}C_{i+1} & -D_{i+1}^{-1}B_{i-1} \\ I & \varnothing \end{pmatrix}, \qquad i=0,\cdots,N-2, \tag{8.8}$$

$$X_i := \begin{pmatrix} x_{i+1} \\ x_i \end{pmatrix}, \quad i=0,\cdots,N-1, \qquad f_i := \begin{pmatrix} D_{i+1}^{-1}b_{i+1} \\ \varnothing \end{pmatrix}, \quad i=0,\cdots,N-2 \tag{8.9}$$

so that

$$X_{i+1}=A_iX_i+f_i. \tag{8.10}$$

Using the transformation matrix

$$T_i := \begin{bmatrix} I & \varnothing \\ -U_i^{-1}D_i & U_i^{-1} \end{bmatrix}, \tag{8.11}$$

we obtain (cf. (3.20))

$$T_{i+1}^{-1}A_iT_i=\begin{bmatrix} -D_{i+1}^{-1}\left(C_{i+1}-B_{i+1}U_i^{-1}D_i\right) & -D_{i+1}^{-1}B_iU_i^{-1} \\ -C_{i+1}+B_{i+1}U_i^{-}D_i & -B_{i+1}U_i^{-1} \end{bmatrix}, \tag{8.12}$$

so, using (8.7b),

$$T_{i+1}^{-1}A_iT_i=\begin{bmatrix} -D_{i+1}^{-1}U_{i+1} & -D_{i+1}^{-1}L_{i+1} \\ \varnothing & -L_{i+1} \end{bmatrix}. \tag{8.13}$$

Hence, if we define

$$Y_i=T_i^{-1}X_i, \tag{8.14}$$

we have the recursion,

$$Y_{i+1}^2=-L_{i+1}Y_i^2+b_{i+1}, \qquad i=0,2,\cdots,N-1, \tag{8.15}$$

which is just forward substitution on the lower triangular system. On the other hand,

$$\begin{aligned} Y_{i+1}^1 &= -D_{i+1}^{-1}U_{i+1}Y_i^1-D_{i+1}^{-1}L_{i+1}Y_i^2+D_{i+1}^{-1}b_{i+1} \\ &= -D_{i+1}^{-1}U_{i+1}Y_i^1-D_{i-+1}^{-1}Y_i^2. \end{aligned} \tag{8.16}$$

Hence we can find $Y_i^1=X_i^1=x_{i+1}$ by backward recursion

$$X_i^1=-U_{i+1}^{-1}\left[D_{i+1}X_{i+1}^1+Y_i^2\right], \qquad i=N-2,\cdots,0. \tag{8.17}$$

Concluding, the LU-decomposition is equivalent to (3.25), Step I. By this special choice of T_i, we have a suitable initial condition for Y_0^2 and a suitable terminal condition for $Y_{N-1}^1=x_N$. Hence Step II and Step III of (3.25) are just (8.15) and (8.17), while Step IV can be omitted. It should be clear that this LU-decomposition is closely related to a Riccati method. Indeed, if we perform such an algorithm for (8.10), with a transformation matrix

$$\hat{T}_i := \begin{pmatrix} I & \varnothing \\ -U_i^{-1}D_i & I \end{pmatrix}, \tag{8.18}$$

(cf. §5.3), we could also study its stability and consistency.

Recall that we tacitly assumed that no pivoting was necessary in the LU-decomposition. This may be restrictive. As is shown in [26], a privoting strategy is closely related to eliminating the blow-up of the Riccati solution. This is also obvious from (8.11), where the "unstable directions" cannot be represented meaningfully if U_i^{-1} becomes large, i.e., if pivoting becomes necessary.

8.2. Intermezzo: Generalized decoupling transformations. Thinking of the more general setting of the recursion relation in multiple shooting, cf. (7.2), it makes sense to consider decoupling transformations for

$$\hat{A}_i X_{i+1} = A_i X_i + f_i. \tag{8.19}$$

In order to have a decoupling we may use two sets of nonsingular matrices $\{S_i\}$ and $\{T_i\}$, such that

$$S_{i+1}\hat{A}_i T_{i+1} T_{i+1}^{-1} X_{i+1} = (S_{i+1} A_i T_i) T_i^{-1} X_i + S_i f_i \tag{8.20}$$

with

(8.21a) $S_{i+1} A_i T_i$ block upper triangular

(8.21b) $S_{i+1}\hat{A}_i T_{i+1}$ block diagonal, with nonsingular lower right block.

Then we can use the lower (decoupled) recursion in the forward direction and the upper (coupled) one in backward direction. If we already have a matrix T_i at the ith stage, we may perform a QU-decomposition for $A_i T_i$, giving an orthogonal S_{i+1}^{-1}, cf. (8.21a). Then, provided $\hat{A}_i$ is nonsingular, T_{i+1}^{-1} is found from (8.21b). However, unless $\hat{A}_i$ is orthogonal, it will require a matrix inversion to compute T_{i+1}. We therefore see that we could as well invert $\hat{A}_i$ first and apply a decoupling to the sequence of matrices $\{\hat{A}_i^{-1} A_i\}$. For some problems we may have recursions like (8.19) where the $\hat{A}_i$ are singular. Then we have even more freedom to choose T_{i+1}. In particular, suppose the recursion is implicitly defined by (8.1), but now with D_i singular. Written in matrix vector notation, cf. (8.9), this results in a relation like (8.19) with

$$\hat{A}_i = \begin{pmatrix} D_{i+1} & \varnothing \\ \varnothing & I \end{pmatrix}, \qquad A_i = \begin{pmatrix} -C_{i+1} & -B_{i+1} \\ I & \varnothing \end{pmatrix}. \tag{8.22}$$

For these special matrices we propose using

$$S_i = \begin{pmatrix} I & \varnothing \\ I & U_i \end{pmatrix}, \qquad T_i^{-1} = \begin{pmatrix} I & \varnothing \\ D_i & U_i \end{pmatrix}, \tag{8.23}$$

where the nonsingular U_i's are still to be chosen. For the transformed $\{Y_i\}$ (i.e. $Y_i = T_i^{-1} X_i$) we then find the following recursions

$$D_{i+1} Y_{i+1}^1 = \left(-C_{i+1} + B_{i+1} U_i^{-1} D_i\right) Y_i^1 - B_{i+1} U_i^{-1} Y_i^2 + b_{i+1}, \tag{8.24a}$$

$$Y_{i+1}^2 = \left(-C_{i+1} + B_{i+1} U_i^{-1} D_i + U_{i+1}\right) Y_i^1 - B_{i+1} U_i^{-1} Y_i^2 + b_{i+1}. \tag{8.24b}$$

Formally, we may apply the same procedure as in §7.1 to find the desired decoupled form, by requiring $\{U_i\}$ to satisfy (8.7). As a consequence, (8.15) and (8.16) are also valid without Assumption 8.2. Unfortunately, the simple relation to the Riccati method suggested in §8.1 does not imply stability. The singularity of the matrices $\hat{A}_i$ does not allow the consistency argument, but the stability of (8.15) and (8.16) follows from the more general result of the next section.

8.3. Matrices arising from BVPs with general BCs. As was already noted, the multiple shooting system (7.4) arises in quite a few other algorithms. Without affecting the generality, we may assume that the $F_i(t_i)=I$. This is automatically true for the methods in §§7.2–7.4, since we can always postmultiply the matrix in (7.4) by $\operatorname{diag}(I,[F_1(t_1)]^{-1},\cdots,[F_{N-1}(t_{N-1})]^{-1})$. For convenience we also replace the last two block rows in (7.4) by

$$F_{N-1}(t_N)v_{N-1}-v_N=-p_{N-1}(t_N), \tag{8.25a}$$

$$M_\alpha v_0+M_\beta v_N=b-M_\alpha p_0(t_0). \tag{8.25b}$$

(For orthogonalization methods this corresponds to a QU-decomposition of "$F_{N-1}(t_N)$"). If we moreover use the notation

$$A_i := F_i(t_{i+1}), \tag{8.26a}$$

$$M_0 := M_\alpha, \tag{8.26b}$$

$$M_N := M_\beta, \tag{8.26c}$$

we obtain the matrix

$$\mathbf{A}=\begin{bmatrix} A_0 & -I & & & \\ & A_1 & -I & & \\ & & \ddots & \ddots & \\ & & & A_{N-1} & -I \\ M_0 & & & & M_N \end{bmatrix}. \tag{8.27}$$

Of course there exist many other forms of such a multiple shooting matrix. Of particular interest are matrices $\bar{\mathbf{A}}$ that can be found from $\mathbf{A}$ by writing some of the BC first. Formally this can be described using the permutation matrix

$$\begin{bmatrix} & & & & I_{n-k} \\ & & & & \\ I_n & & & & \\ & \ddots & & & \\ & & I_n & & \\ & & & I_k & \end{bmatrix}, \tag{8.28}$$

so that

$$\bar{\mathbf{A}} := \mathbf{PA}=\begin{bmatrix} {}^2M_0 & & & & {}^2M_N \\ A_0 & -I & & & \\ & A_1 & -I & & \\ & & \ddots & \ddots & \\ & & & A_{N-1} & -I \\ {}^1M_0 & & & & {}^1M_N \end{bmatrix}. \tag{8.29a}$$

By repartitioning $\bar{\mathbf{A}}$ into a matrix of $n \times n$ blocks, we obtain an almost tridiagonal block matrix. We shall use the notation

$$\bar{\mathbf{A}} = \begin{bmatrix} C_0 & D_0 & & & H_N \\ B_1 & C_1 & D_1 & & \\ & \ddots & \ddots & \ddots & \\ & & B_{N-1} & C_{N-1} & D_{N-1} \\ H_1 & & & B_N & C_N \end{bmatrix}. \tag{8.29b}$$

In (8.29b) we see that H_N and the B_i systematically have zeros in their last k rows; likewise H_1 and the D_i have zeros in their first $(n-k)$ rows. Specifically, if the BC are separated, then (8.29) may be chosen as a block tridiagonal matrix. The aim of many methods is to obtain an LU-decomposition where the upper and lower matrices contain codiagonal blocks with a similar zero row structure to $\bar{\mathbf{A}}$ (cf. [8], [18], [25], [37], [66] . To preserve this structure, one may only allow row permutations within the rows $n-k+1+in$ to $n-k+(i+1)n$, where i is an integer, $0 \leqq i \leqq N-1$. Similarly one may allow permutations in the columns $1+jn$ to $(j+1)n$, $0 \leqq j \leqq N$. However, this means that we may permute rows and columns of $\mathbf{A}$ by pre and postmultiplication by a block diagonal matrix. Thus, in its most general form, cf. [37], we compute an LU-decomposition of

$$\bar{\mathbf{A}} = \mathbf{PRAE}, \tag{8.30}$$

where

$$\mathbf{R} = \operatorname{diag}(R_0, \cdots, R_N), \tag{8.31a}$$

$$\mathbf{E} = \operatorname{diag}(E_0, \cdots, E_N). \tag{8.31b}$$

So we have

$$\bar{\mathbf{A}} = \mathbf{LU}, \tag{8.32a}$$

where

$$\mathbf{L} = \begin{bmatrix} Y_0 & & & & \\ L_1 & Y_1 & & \varnothing & \\ & \ddots & \ddots & & \\ & & L_{N-1} & Y_{N-1} & \\ S_1 & & S_{N-2} & S_{N-1} & Y_N \end{bmatrix}, \tag{8.32b}$$

$$\mathbf{U} = \begin{bmatrix} U_0 & Y_0^{-1}D_0 & & & Z_0 \\ & \ddots & \ddots & & \\ & & Y_{N-2}^{-1}D_{N-2} & & \\ & \varnothing & U_{N-1} & & Z_{N-1} \\ & & & & U_N \end{bmatrix}. \tag{8.32c}$$

Remark 8.33. The L_i, S_i and Z_i have the same systematic zero rows as B_0, H_1 and H_N respectively. More commonly, one chooses the matrices Y_i to be identity matrices. For other choices, however, see [23], [25].

In order to investigate the existence and stability of this decomposition, we compare it to a simple LU-decomposition that results from a decoupled recursion, cf. [37]. First we have:

PROPERTY 8.34. *Let R_N be the permutation matrix in* (8.31) *which permutes the rows of the BC. Assume that the last $(n-k)$ rows of $R_N M_0$ are independent. Then there exists an orthogonal matrix Q_0 such that $\tilde{M}_0 := R_N M_0 Q_0$ is upper triangular with a nonsingular lower right block.*

Proof. By bottom-up Gram–Schmidt orthogonalization of the row vectors of $R_N M_0$ completed to a full basis, we find Q_0 and $\tilde{M}_0$. □

Now use this Q_0 to generate transformations $\{Q_i\}$ which decouple the A_i into Δ_i as in the basis algorithm, cf. (3.25). We now have

$$A_i Q_i = Q_{i+1}\Delta_i, \qquad i=0,\cdots,N-1, \tag{8.35a}$$

$$\tilde{M}_N = R_N M_N Q_N. \tag{8.35b}$$

Then premultiplying $\mathbf{A}$ by $\tilde{\mathbf{P}}\tilde{\mathbf{R}}$, where $\tilde{\mathbf{P}}$ equals $\mathbf{P}$ (see (8.28)) but for a minus sign in front of I_{n-k}, and

$$\tilde{\mathbf{R}} = \operatorname{diag}\left(Q_1^{-1},\cdots,Q_N^{-1},R_N\right), \tag{8.36}$$

and postmultiplying $\mathbf{A}$ by

$$\tilde{\mathbf{E}} = \operatorname{diag}(Q_0,\cdots,Q_N), \tag{8.37}$$

we obtain the matrix

$$\tilde{\mathbf{A}} = \tilde{\mathbf{P}}\tilde{\mathbf{R}}\mathbf{A}\tilde{\mathbf{E}}, \tag{8.38}$$

$$\tilde{\mathbf{A}} = \begin{bmatrix}
\varnothing & -\tilde{M}_0^{22} & \varnothing & \varnothing & & & & & & -\tilde{M}_N^{21} & -\tilde{M}_N^{22} \\
\Delta_0^{11} & \Delta_0^{12} & -I & \varnothing & & & & & & \varnothing & \varnothing \\
\varnothing & \Delta_0^{22} & \varnothing & -I & & & & & & & \\
\varnothing & \varnothing & \Delta_1^{11} & \Delta_1^{12} & & & & & & & \\
 & & & & \ddots & & & & & & \\
 & & & & & \varnothing & -I & \varnothing & \varnothing & & \\
 & & & & & \Delta_{N-1}^{11} & \Delta_{N-1}^{12} & -I & \varnothing & & \\
\varnothing & \varnothing & & & & \varnothing & \Delta_{N-1}^{22} & \varnothing & -I & & \\
\tilde{M}_0^{11} & \tilde{M}_0^{12} & & & & \varnothing & \varnothing & \tilde{M}_N^{11} & \tilde{M}_N^{12} & & \\
\end{bmatrix}. \tag{8.39}$$

This matrix $\tilde{\mathbf{A}}$ is interesting because one can quite easily find an explicit LU-decomposition

$$\tilde{\mathbf{A}} = \tilde{\mathbf{L}}\tilde{\mathbf{U}}, \tag{8.40a}$$

with lower block $\tilde{\mathbf{L}}$ and upper block $\tilde{\mathbf{U}}$ as in (8.32b,c) where, (providing elements of these matrices with a tilde) we have

$$\tilde{Y}_i = I_n, \qquad i=0,\cdots,N, \tag{8.40b}$$

$$\tilde{U}_0 = \begin{bmatrix} \varnothing & -\tilde{M}_0^{22} \\ \Delta_0^{11} & \Delta_0^{12} \end{bmatrix}, \quad \tilde{U}_i = \begin{bmatrix} \varnothing & -I_n \\ \Delta_i^{11} & \Delta_i^{12} \end{bmatrix}, \qquad i=1,\cdots,N-1, \tag{8.40c}$$

$$\text{(8.40d)}\quad \tilde{L}_1=\begin{bmatrix} -\Delta_0^{22}\left[\tilde{M}_0^{22}\right]^{-1} & \varnothing \\ \varnothing & \varnothing \end{bmatrix},\quad \tilde{L}_i=\begin{bmatrix} -\Delta_{i-1}^{22} & \varnothing \\ \varnothing & \varnothing \end{bmatrix},\qquad i=2,\cdots,N-1,$$

$$\text{(8.40e)}\quad \tilde{D}_i=\begin{bmatrix} \varnothing & \varnothing \\ -I_k & \varnothing \end{bmatrix},\qquad i=0,\cdots,N-2.$$

The expressions for the other blocks are somewhat more complicated. However, the main insight into the stability of the problem can already be obtained from (8.40b–e), by comparing them to (8.32). We have

THEOREM 8.41. (i) *Solving the homogeneous part of* $L_i z_{i-1}+Y_i z_i=0$ *is equivalent to solving* $y_i^2=\Delta_{i-1}^{22}y_{i-1}^2$, (*cf.* (3.23b).

(ii) *Solving the homogeneous part of* $U_i w_i+D_i w_{i+1}=0$ *is equivalent to solving* $y_{i+1}^1=\Delta_i^{11}y_i^1$, (*cf.* (3.23a) .

Proof. By our special choice of $\mathbf{R},\tilde{\mathbf{R}}$ and $\mathbf{P}$, we see that

$$\text{(a)}\qquad \mathbf{P}\mathbf{R}\tilde{\mathbf{R}}^{-1}\mathbf{P}^{-1}=\operatorname{diag}(I_{n-k},R_0,Q_1,\cdots,R_{N-1}Q_N,I_k).$$

From (8.30) and (8.38)

$$\mathbf{A}=\mathbf{R}^{-1}\mathbf{P}\hat{\mathbf{A}}\mathbf{E}^{-1}=\tilde{\mathbf{R}}^{-1}\mathbf{P}\tilde{\mathbf{A}}\tilde{\mathbf{E}}^{-1}.$$

Using (8.32) and (8.40), we get

$$\text{(b)}\qquad \mathbf{U}\mathbf{E}^{-1}\tilde{\mathbf{E}}\tilde{\mathbf{U}}^{-1}=\mathbf{L}^{-1}\mathbf{P}\mathbf{R}\tilde{\mathbf{R}}^{-1}\mathbf{P}^{-1}\tilde{\mathbf{L}}.$$

Since the left-hand side in (b) is block upper triangular and the right-hand is block lower triangular (cf. (a)), but with a different staircase (cf. [18]), we can express L_i and U_i in terms of $\tilde{L}_i$ and $\tilde{U}_i$, respectively. To this end, let $Y_i=I$ for all i (not essential, but convenient) and write

$$\text{(c)}\qquad \tilde{\mathbf{L}}^{-1}\mathbf{P}\tilde{\mathbf{R}}^{-1}\mathbf{P}^{-1}\mathbf{L}=\begin{bmatrix} F_0 & K_0 & & \\ & \ddots & \ddots & \\ & & \ddots & K_{N-1} \\ & & & F_N \end{bmatrix},$$

where, upon denoting

$$Q_i^{-1}R_{i-1}=:P_i=\begin{bmatrix} P_i^{11} & P_i^{12} \\ P_i^{21} & P_i^{22} \end{bmatrix},$$

$$L_i:=\begin{bmatrix} L_i^{22} & L_i^{21} \\ \varnothing & \varnothing \end{bmatrix},$$

we have

$$F_i=\left[\begin{array}{c|c} P_i^{22} & \varnothing \\ \hline P_{i+1}^{12}L_{i+1}^{22} & P_{i+1}^{11}+P_{i+1}^{12}L_{i+1}^{21} \end{array}\right]$$

and

$$K_i=\begin{bmatrix} \varnothing & \varnothing \\ P_{i+1}^{12} & \varnothing \end{bmatrix}.$$

By comparing the blocks in **L** with the values obtained from (c) we find, denoting $\tilde{P}_i = P_i^{-1}$,

$$L_{i+1}^{22} = \tilde{P}_{i+1}^{21} P_{i+1}^{12} L_{i+1}^{22} - \tilde{P}_{i+1}^{22} \Delta_i^{22} P_i^{22} \Rightarrow \tilde{P}_{i+1}^{22} P_{i+1}^{22} L_{i+1}^{22} = -\tilde{P}_{i+1}^{22} \Delta_i^{22} P_i^{22},$$

$$L_{i+1}^{21} = \tilde{P}_{i+1}^{21} P_{i+1}^{11} + \tilde{P}_{i+1}^{21} P_{i+1}^{12} L_{i+1}^{21} \Rightarrow \tilde{P}_{i+1}^{22} P_{i+1}^{22} L_{i+1}^{21} = \tilde{P}_{i+1}^{22} P_{i+1}^{21}.$$

Now we can interchange the roles of $\mathbf{L}, \mathbf{R}$ and $\tilde{\mathbf{L}}, \tilde{\mathbf{R}}$ respectively in (c), giving a staircase matrix with diagonal blocks $\tilde{F}_i$ which should be nonsingular. In particular, we deduce from [37, Prop. 2.13] that $\tilde{P}_{i+1}^{11}$ should be nonsingular, whence $\tilde{P}_i^{22}$ (from the orthogonality). Finally, the vector z_i in $L_{i+1} z_i + z_{i+1}$ should be partitioned like

$$z_i = \begin{bmatrix} z_i^2 \\ z_{i+1}^1 \end{bmatrix}.$$

Defining

$$\begin{bmatrix} y_i^1 \\ y_i^2 \end{bmatrix} = P_i \begin{bmatrix} z_i^1 \\ z_i^2 \end{bmatrix},$$

we thus obtain

$$\begin{aligned} L_{i+1}^{22} z_i^2 + L_{i+1}^{21} z_{i+1}^1 + z_{i+1}^2 = 0 &\Leftrightarrow P_{i+1}^{22} L_{i+1}^{22} z_i^2 + P_{i+1}^{22} L_{i+1}^{21} z_{i+1}^1 + P_{i+1}^{22} z_{i+1}^2 = 0 \\ &\Leftrightarrow -\Delta_i^{22} P_i^{22} z_i^2 + y_{i+1}^2 = 0 \\ &\Leftrightarrow -\Delta_i^{22} y_i^2 + y_{i+1}^2 + \Delta_i^{22} P_i^{21} z_i^1 = 0. \end{aligned}$$

Since the z_i^1 are mere inhomogeneities, part (a) is proven. Part (b) follows similarly. □

From Theorem 8.41 we see that the stability of forward and backward substitution (which is controlled by the underlying homogeneous recursion) is similar to the one that would be found from transforming the incremental recursion onto upper triangular form and applying algorithm (3.5). Hence we have shown stability of block LU-decomposition if we can show consistency. To this end we combine Theorem 4.10 and Property 8.34 to obtain

PROPERTY 8.42. *If the* BC *are separated (so* $\tilde{M}_0^{22}$ *has full rank), then* Q_0 *induces a consistent fundamental solution.*

COROLLARY 8.43. *If in* (8.29a) ${}^2\overline{M}_N = 0$ *and* ${}^1\overline{M}_0 = 0$, *then a block* LU*-decomposition, found from restricted pivoting, cf.* [32], *is stable.*

One should realize, that if ${}^2\overline{M}_N$ and ${}^1\overline{M}_0$ are nontrivial matrices, a result like Corollary 8.43 is, in general, not true. For partially separated BC it might be tempting to choose the initial permutation **P** such that either ${}^2\overline{M}_N = 0$ *or* ${}^1\overline{M}_0 = 0$ in order to have either the **U** or **L** block bidiagonal. This may lead to a dramatic instability as can be seen from the following:

Example 8.44. Consider

$$\dot{x} = \begin{bmatrix} -20 & 30 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & -10 \end{bmatrix} x + \begin{bmatrix} -10 \\ 10 \\ 10 \end{bmatrix}$$

and

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} x(0) + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} x(10) = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}.$$

A fundamental solution is given by

$$\Psi(t) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \operatorname{diag}(e^{-20t}, e^{+10t}, e^{-10t}).$$

If we write the last row of the BC first and the first two rows last in $\bar{\mathbf{A}}$ (cf. (8.29a)), then we obtain a bidiagonal block $\mathbf{U}$. If we use shooting on an interval of length δ, we obtain the incremental matrix

$$A_i = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \operatorname{diag}(e^{-20\delta}, e^{10\delta}, e^{-10\delta}).$$

Since A_i is upper triangular, we may identify A_i and Δ_i. So we immediately see (cf. (8.40)):

$$\Delta_i = \begin{bmatrix} 0 & 0 & -1 \\ e^{-20\delta} & e^{10\delta} & e^{-10\delta} \\ 0 & e^{10\delta} & 0 \end{bmatrix}.$$

Obviously

$$\left[\Delta_i^{11}\right]^{-1} = \begin{bmatrix} e^{20\delta} & -e^{20\delta} \\ 0 & e^{-10\delta} \end{bmatrix},$$

implying that backward substitution is as unstable as single shooting for this BVP, starting at $t = 10$. The important conclusion to be drawn is that the zero row structure in partially separated BC matrices should *not* be employed to save memory (viz., of the last block columns in $\mathbf{U}$ or the last block row in $\mathbf{L}$) but should be used to compute a special subspace of solutions containing at least the unstable ones, as in the generalized Godunov–Conte algorithm (see §7.3). Assuming that the BVP is well-conditioned, it should be expected that a classical partial pivoting strategy is stable for any BC. Therefore we conclude this section by showing that such a strategy can also be viewed as a decoupling method. Such a partial pivoting process can be described as

$$\mathbf{L}_{N-1}\mathbf{D}_{N-1} \cdots \mathbf{L}_0\mathbf{D}_0\mathbf{A} = \mathbf{U}. \tag{8.45}$$

The matrices $\mathbf{L}_i$ are lower triangular ("generalized") elementary matrices cf. [71, p. 44], and the $\mathbf{D}_i$ are permutation matrices such that $\mathbf{L}_i\mathbf{D}_i$ actually describes the elimination of elements in the ith block column. Inspired by the preceding results we look for $\mathbf{D}_i$ of a special form. Since we can only permute within blocks and moreover between the pivotal block and the block in the last row, we typically have

$$\mathbf{D}_0 = \mathbf{R}_0\mathbf{P}_0\mathbf{O}_0, \tag{8.46a}$$

where

$$\mathbf{R}_0 = \operatorname{diag}(R_0, I, \cdots, I, \hat{R}_0), \qquad \mathbf{O}_0 = \operatorname{diag}(O_0, I, \cdots, \hat{O}_0), \tag{8.46b}$$

and for some suitable number k_0

$$\mathbf{P}_0 = \begin{bmatrix} \ddots & & & & \ddots & \\ & \varnothing & \varnothing & & \varnothing & I_{n-k_0} \\ & I_{k_0} & \varnothing & & \varnothing & \varnothing \\ & & & I_n & & \\ & & & & \ddots & \\ & & & I_n & & \\ & \varnothing & I_{n-k_0} & & \varnothing & \varnothing \\ & \varnothing & \varnothing & & I_{k_0} & \varnothing \\ & & & & & \ddots \end{bmatrix}. \tag{8.46c}$$

The matrix $\mathbf{L}_0$ typically looks like

$$\mathbf{L}_0 = \begin{bmatrix} L_0^{00} & & & & \\ & I_n & & \varnothing & \\ & & \ddots & & \\ & \varnothing & & & \\ L_0^{N0} & & & & I_N \end{bmatrix}. \tag{8.46d}$$

In a similar way we have

$$\mathbf{D}_1 = \mathbf{R}_1 \mathbf{P}_1 \mathbf{O}_1, \tag{8.47a}$$

where

$$\mathbf{R}_1 = \operatorname{diag}(I_n, R_1, I_n, \cdots, I_n, \hat{R}_1), \qquad \hat{\mathbf{O}}_1 = \operatorname{diag}(I_n O_1, I_n, \cdots, I_n, \hat{O}_1) \tag{8.47b}$$

and $\mathbf{P}_1$ as a matrix like $\mathbf{P}_0$, but now permuting rows in the second and Nth "block" position, with $k = k_1$. We can rewrite

$$\mathbf{L}_1 \mathbf{D}_1 \mathbf{L}_0 \mathbf{D}_0 = \operatorname{diag}(R_0, R_1, I_n, \cdots, I_n) \hat{\mathbf{L}}_1 \mathbf{P}_1 \hat{\mathbf{L}}_0 \mathbf{P}_0 \operatorname{diag}(O_0, O_1, I_n, \cdots, I_n), \tag{8.48}$$

where $\hat{\mathbf{L}}_0$ and $\hat{\mathbf{L}}_1$ have a zero block structure similar to $\mathbf{L}_0$ and $\mathbf{L}_1$ respectively. In general there exist $\hat{\mathbf{L}}_i$ such that

$$\mathbf{L}_{N-1} \mathbf{D}_{N-1} \cdots \mathbf{L}_0 \mathbf{D}_0 = \operatorname{diag}(R_i) \hat{\mathbf{L}}_{N-1} \mathbf{P}_{N-1} \cdots \hat{\mathbf{L}}_0 \mathbf{P}_0 \operatorname{diag}(O_i), \tag{8.49}$$

where R_n and O_N are orthogonal. Now use the matrix O_N in Property 8.34 to generate sequences of orthogonal $\{Q_i\}$ and upper triangular $\{\Delta_i\}$. This then induces a sparse matrix $\tilde{\mathbf{R}}\mathbf{A}\tilde{\mathbf{E}}$ (cf. (8.38), (8.39)). We may use the same permutation matrices $\mathbf{P}_i$ in order to obtain a block factorization as in (8.45), i.e.,

$$\tilde{\mathbf{L}}_{N-1} \mathbf{P}_{N-1} \cdots \tilde{\mathbf{L}}_0 \mathbf{P}_0 [\tilde{\mathbf{R}}\mathbf{A}\tilde{\mathbf{E}}] = \tilde{\mathbf{U}}_{N-1}. \tag{8.50}$$

As a first step,

$$\tilde{\mathbf{L}}_0 \mathbf{P}_0 [\tilde{\mathbf{R}}\mathbf{A}\tilde{\mathbf{E}}] = \tilde{\mathbf{U}}_0, \tag{8.51a}$$

where

$$\tilde{\mathbf{L}}_0 = \begin{bmatrix} I_n & & \varnothing \\ & \ddots & \\ \varnothing & & I_n \\ \tilde{L}_{N0} & & \end{bmatrix}, \qquad \tilde{\mathbf{U}}_0 = \begin{bmatrix} \tilde{U}_0 & C_0 & \varnothing & Z_0 \\ & \hat{U}_1 & & \\ \varnothing & & \text{▨} & \\ & \tilde{\hat{U}}_1 & & \end{bmatrix}. \tag{8.51b}$$

The shaded area in $\tilde{\mathbf{U}}_0$ is equal to the corresponding part of $(\tilde{\mathbf{P}}^{-1}\mathbf{A})$. We obtain

$$\text{(8.52a)} \qquad \tilde{U}_0 = \begin{pmatrix} \Delta_0^{11} & \Delta_0^{12} \\ \varnothing & \tilde{M}_0^{22} \end{pmatrix}, \quad C_0 = \begin{pmatrix} -I & \varnothing \\ \varnothing & \varnothing \end{pmatrix}, \quad Z_0 = \begin{pmatrix} \varnothing & \varnothing \\ \tilde{M}_N^{21} & \tilde{M}_N^{22} \end{pmatrix},$$

$$\text{(8.52b)} \qquad \tilde{L}_{N0} = -\begin{pmatrix} \tilde{M}_0^{11} & \tilde{M}_0^{12} \\ \varnothing & \Delta_0^{22} \end{pmatrix} \left(\begin{array}{c:c} [\Delta_0^{11}]^{-1} & [\Delta_0^{11}]^{-1}\Delta_0^{12} \\ \hdashline \varnothing & [\tilde{M}_0^{22}]^{-1} \end{array}\right),$$

$$\text{(8.52c)} \qquad \hat{\tilde{U}}_1 = \begin{pmatrix} \tilde{M}_0^{11}[\Delta_0^{11}]^{-1} & \varnothing \\ \varnothing & -I_{n-k} \end{pmatrix}, \qquad \hat{U}_1 = \Delta_1.$$

The partitioning in (8.52) is after k_0 rows and columns. In order to understand the arguments, we perform one more "column elimination", viz.,

$$\text{(8.53)} \qquad \tilde{\mathbf{L}}_1 \mathbf{P}_1 \tilde{\mathbf{U}}_0 = \tilde{\mathbf{U}}_1,$$

where

$$\text{(8.54)} \qquad \tilde{\mathbf{L}}_1 \begin{bmatrix} I_n & & \varnothing \\ & I_n & \\ \varnothing & \ddots & \\ & \tilde{L}_{N1} & I_n \end{bmatrix}, \qquad \tilde{\mathbf{U}}_1 = \begin{bmatrix} \tilde{U}_0 & C_0 & & \varnothing & Z_0 \\ & \tilde{U}_1 & C_1 & & Z_1 \\ \varnothing & & \hat{U}_2 & \blacksquare & \\ & & \hat{\hat{U}}_2 & & \end{bmatrix}.$$

Basically the blocks in $\tilde{\mathbf{L}}_1$ and $\tilde{\mathbf{U}}_1$ can be found as in (8.52). We are mainly interested in $\tilde{U}_1$ and $\tilde{L}_{N1}$. We distinguish three cases

$$\text{(8.55a)} \quad k_1 < k_0\text{: } \tilde{U}_1 = \begin{bmatrix} \overline{\Delta_1^{11}} & \overline{\Delta_1^{12}} \\ \overline{\tilde{M}_0^{11}[\Delta_0^{11}]^{-1}} & \varnothing \\ \varnothing & -I_{n-k_1} \end{bmatrix}, \qquad H_1 = \begin{bmatrix} \overline{\tilde{M}_0^{11}[\Delta_0^{11}]^{-1}} & \varnothing \\ \overline{\Delta_1^{11}} & \overline{\Delta_1^{12}} \\ \varnothing & \Delta_1^{22} \end{bmatrix},$$

$$\text{(8.55b)} \quad k_1 = k_0\text{: } \tilde{U}_1 = \begin{bmatrix} \Delta_1^{11} & \Delta_1^{12} \\ \varnothing & -I_{n-k_1} \end{bmatrix}; \qquad H_1 = \begin{bmatrix} \tilde{M}_0^{11}[\Delta_0^{11}]^{-1} & \varnothing \\ \varnothing & \Delta_1^{22} \end{bmatrix},$$

$$\text{(8.55c)} \quad k_1 > k_0\text{: } \tilde{U}_1 = \begin{bmatrix} \Delta_1^{11} & \Delta_1^{12} \\ \varnothing & \overline{\Delta_1^{22}} \\ \varnothing & -I_{n-k_1} \end{bmatrix}; \qquad H_1 = \begin{bmatrix} \tilde{M}_0^{11}[\Delta_0^{11}]^{-1} & \varnothing \\ \varnothing & -I_{k_1-k_0} \\ \varnothing & \overline{\Delta_1^{22}} \end{bmatrix}.$$

The bar above a matrix denotes that $|k_0 - k_1|$ rows are left out. The matrix H_1 temporarily appears in the $(N,1)$ block position after permutation; it determines the matrix:

$$\text{(8.55d)} \qquad \tilde{L}_{N1} = H_1[\tilde{U}_1]^{-1}.$$

These expressions are less frightening than they look at first sight. To start with, we have

PROPERTY 8.56. *Let $k_i = k$ for all i (so the permutation blocks in the $\mathbf{P}_i$ are the same). Then we may identify $\tilde{\mathbf{P}}$ with $\mathbf{P}_{N-1}\cdots\mathbf{P}_0$, $\tilde{\mathbf{L}}$ with $\tilde{\mathbf{L}}_{N-1}^{-1}\cdots\tilde{\mathbf{L}}_0^{-1}$ and $\tilde{\mathbf{U}}$ with $\tilde{\mathbf{U}}_{N-1}$, cf.* (8.38), (8.40) *and* (8.50).

Proof. Because of the special form of the $\tilde{\mathbf{L}}_i$ and $\tilde{\mathbf{P}}_i$ for $i>j$: $\mathbf{P}_1\tilde{\mathbf{L}}_j=\tilde{\tilde{\mathbf{L}}}_j\mathbf{P}_i$, where $\tilde{\tilde{\mathbf{L}}}_j$ is a "block" elementary matrix with the identity matrix I_n as (j,j) block. Hence, $\tilde{\mathbf{L}}_{N-1}\mathbf{P}_{N-1}\cdots\tilde{\mathbf{L}}_0\mathbf{P}_0=\tilde{\tilde{\mathbf{L}}}_{N-1}\cdots\tilde{\tilde{\mathbf{L}}}_0\mathbf{P}_{N-1}\cdots\mathbf{P}_0$. It is easy to see that $\mathbf{P}_{N-1}\cdots\mathbf{P}_0=\mathbf{P}$ (cf. (8.28)), which was equal to $\tilde{\mathbf{P}}$ apart from a minus sign in the first $(n-k)$ rows. The rest is trivial. □

We now obtain the following interesting result:

THEOREM 8.57. *If we solve a system with a matrix like* (8.27) *by partial pivoting such that for each j,* $(0\leqq j\leqq N-1)$ *the number of rows exchanged between positions* $jn+1$ *until* $jn+n$ *and the last n positions, is constant, then the solution of such a system is mathematically equivalent to decoupling the recursion* (*and solving it in a stable way*).

Proof. From (8.45) and (8.50), we obtain

$$\mathbf{A}=\mathbf{D}_0^{-1}\mathbf{L}_0^{-1}\cdots\mathbf{D}_{N-1}^{-1}\mathbf{P}_{N-1}^{-1}\mathbf{U}=\tilde{\mathbf{R}}^{-1}\mathbf{P}_0^{-1}\tilde{\mathbf{L}}_0^{-1}\cdots\mathbf{P}_{N-1}^{-1}\tilde{\mathbf{L}}_{N-1}^{-1}\tilde{\mathbf{U}}_{N-1}\tilde{\mathbf{E}}^{-1}.$$

From this via (8.49):

$$\text{(a)}\quad \mathbf{U}\tilde{\mathbf{E}}\tilde{\mathbf{U}}_{N-1}^{-1}=\operatorname{diag}(R_i)\hat{\mathbf{L}}_{N-1}\mathbf{P}_{N-1}\cdots\hat{\mathbf{L}}_0\mathbf{P}_0\operatorname{diag}(0_i)\tilde{\mathbf{R}}^{-1}\mathbf{P}_0^{-1}\mathbf{L}_0^{-1}\cdots\mathbf{P}_{N-1}^{-1}\tilde{\mathbf{L}}_{N-1}^{-1}.$$

As in Property 8.56, the right-hand side can be written as

$$\operatorname{diag}(R_i)\mathbf{L}_1\mathbf{P}\operatorname{diag}(O_i)\tilde{\mathbf{R}}^{-1}\mathbf{P}^{-1}\mathbf{L}_2,$$

where $\mathbf{L}_1$ and $\mathbf{L}_2$ are block lower triangular. Since the last block in $\tilde{\mathbf{R}}$ equals O_N, we find that $\mathbf{P}\operatorname{diag}(O_i)\tilde{\mathbf{R}}^{-1}\mathbf{P}^{-1}$ is a block diagonal matrix as in 8.41(a) (Proof). We can now use the same arguments as were used in the proof of 8.41 to show that a property like Theorem 8.41(i) or (ii) holds. Moreover, due to the partial pivoting no inhomogeneous term in those recursions will be comparatively large, as can happen in [37, Thm. 3.25]. □

Although we may have k_i different at each block, this is not very likely to happen if we have an exponential dichotomy (i.e, only decaying and increasing and no moderately growing solutions). This can be seen from (8.52) and (8.55). Indeed, in such a case, it is very reasonable to suppose that $\operatorname{glb}(\Delta_0^{22})\leqq\operatorname{lub}(H_0)$. This implies that $\|\tilde{L}_N\|$ and $\|\hat{\tilde{U}}_1\|$ are 0(1) and that we must have (8.55b) (both (8.55a) and (8.55b) would give "larger" $\|H_1\|$ and $\|\tilde{L}_{N1}\|$). For this reason and to avoid fairly messy notation, we simply note that when the k_i are not constant we still have a kind of decoupled recursion. If we restrict ourselves to the "backsolving", the expression $\tilde{U}_0x_0+C_0x_1+Z_0x_N$ has a "homogeneous part" (cf. Theorem 8.41(ii)) $\Delta_0^{11}x_0^1(0)-x_1^1(0)$, where $x_i^1(0)$ has k_0 elements. In fact we see in (8.54) that the expression $\tilde{U}_1x_1+C_1x_2+Z_1x_N$ has a "homogeneous part" $\Delta_1^{11}(1)x_2^1(1)-x_3^1(1)$, where $\Delta_1^{11}(1)$ is the $k_1\times k_1$ left upper block in $\tilde{U}_1$ and $x_j^1(1)$ has k_1 element. Hence, if k_s is the smallest of all k_i, back solving involves a recursion for $\{x_j^1(k_s)\}$, which should be stable in backward direction.

Acknowledgments. The author is greatly indebted to Prof. R. E. O'Malley, Jr. for the stimulating discussions regarding the topics of this paper. He is also grateful for the excellent working conditions when visiting the Department of Mathematical Sciences at Rensselaer Polytechnic Institute. Finally he thanks the referees for their constructive criticism.

REFERENCES

[1] R. C. ALLEN AND G. M. WING, *An invariant imbedding approach for the solution of inhomogeneous linear two point boundary value problems*. J. Comp. Phys., 14 (1974), pp. 40–58.

[2] U. ASCHER, J. CHRISTIANSEN AND R. D. RUSSELL, *A collocation solver for mixed order systems of boundary value problems*, Math. Comp., 33 (1979), pp. 659–674.

[3] U. ASCHER AND R. WEISS, *Collocation for singular perturbation problems* II. *Linear first order systems without turning points*, Math. Comp., 43 (1984), pp. 157–187.

[4] A. K. AZIZ, ed., *Numerical Solutions of Boundary Value Problems for Ordinary Differential Equations*, Academic Press, New York, 1965.

[5] I. BABUŠKA AND V. MAJER, *The factorization method for two point boundary value problems for* ODE's *and its relation to the finite difference method*, Technical note BN-1021, Univ. Maryland, College Park, 1984.

[6] C. DE BOOR, *Dichotomies for band matrices*, SIAM J. Numer. Anal., 17 (1980), pp. 894–907.

[7] C. DE BOOR, F. DE HOOG AND H. B. KELLER, *The stability of one-step schemes for first-order two-point boundary value problems*, SIAM J. Numer. Anal., 20 (1983), pp. 1139–1146.

[8] C. DE BOOR AND R. WEISS, SOLVE BLOK: *A package for solving almost block diagonal linear systems*, ACM Trans. Math. Software, 6 (1980), pp. 80–87.

[9] W. E. BOYCE AND R. D. DIPRIMA, *Elementary Differential Equations*, Wiley, New York, 1969.

[10] D. L. BROWN, *Solution adaptive mesh procedures for the numerical solution of singular perturbation problems*, Ph. D. Thesis, California Institute of Technology, Pasadena, 1982.

[11] B. CHILDS, ed., *Codes for Boundary-Value Problems in ordinary differential equations*, Lecture Notes in Computer Science 76, Springer, Berlin, 1978.

[12] R. C. Y. CHIN, G. W. HEDSTROM AND L. THIGPEN, *A fast, accurate method for computing evanescent waves*, report, Lawrence Livermore Laboratory, 1982.

[13] E. A. CODDINGTON AND N. LEVINSON, *Theory of Ordinary Differential Equations*, McGraw-Hill, New York, 1955.

[14] S. D. CONTE, *The numerical solution of linear* BVP, this Review, 8 (1966), pp. 309–321.

[15] W. A. COPPEL, *Dichotomics in Stability Theory*, Lecture Notes in Mathematics 629, Springer-Verlag, New York, 1978.

[16] P. DEUFLHARD, *Recent advances in multiple shooting techniques*, in Computational Techniques for Ordinary Differential Equations, Gladwell and Sayers, eds., Academic Press, London, 1980, pp. 217–272.

[17] P. DEUFLHARD AND G. BADER, *Multiple shooting techniques revisited*, in Numerical Treatment of Inverse Problems in Differential and Integral Equations, Deuflhard and Hairer, eds., Birkhaüser, Boston, 1983, pp. 74–122.

[18] R. FOURER, *Staircase matrices and systems*, this Review, 26 (1984), pp. 1–70.

[19] W. GAUTSCHI, *Computational aspects of three-term recurrence relations*, this Review, 9 (1967), pp. 24–82.

[20] W. A. HARRIS, JR., *Singularly perturbed boundary value problems revisited*, Lecture Notes in Mathematics 312, Springer-Verlag, Berlin, 1973, pp. 54–64.

[21] F. R. DE HOOG AND R. M. M. MATTHEIJ, *On dichotomy and well-conditioning in* BVP, report 8356, Kath. Univ. Nijmegen, 1983.

[22] F. R. DE HOOG AND R. WEISS, *An approximation theory for boundary value problems on infinite intervals*, Computing, 24 (1980), pp. 227–239.

[23] H. B. KELLER, *Numerical Solution of Two Point Boundary Value Problems*, CBMS Regional Conference Series in Applied Mathematics 24, Society for Industrial and Applied Mathematics, Philadelphia, 1976.

[24] _____, *Shooting and embedding for two-point boundary value problems*, J. Math. Anal. Appl., 36 (1971), pp. 599–610.

[25] _____, *Accurate difference methods for nonlinear two point boundary value problems*, SIAM J. Numer. Anal., 11 (1974), pp. 305–320.

[26] H. B. KELLER AND M. LENTINI, *Invariant imbedding, the box scheme and an equivalence between them*, SIAM J. Numer. Anal., 19 (1982), pp. 942–962.

[27] B. KREISS AND H. O. KREISS, *Numerical methods for singular perturbation problems*, SIAM J. Numer. Anal., 18 (1981), pp. 262–276.

[28] H. O. KREISS, *Difference approximations for boundary and eigenvalue problems for ordinary differential equations*, Math. Comp., 26 (1972), pp. 605–624.

[29] M. LENTINI AND H. B. KELLER, *Boundary value problems on semi-infinite intervals and their numerical solution*, SIAM J. Numer. Anal., 17 (1980), pp. 577–604.

[30] M. LENTINI, M. R. OSBORNE AND R. D. RUSSELL, *The close relationships between methods for solving two-point boundary value problems*, SIAM J. Numer. Anal., 22 (1985), pp. 280–309.

[31] J. L. MASSARA AND J. J. SCHÄFFER, *Linear Differential Equations and Function Spaces*, Academic Press, New York, 1966.

[32] R. M. M. MATTHEIJ, *Characterization of dominant and dominated solutions of linear recursions*, Numer. Math., 35 (1980), pp. 421–442.

[33] R. M. M. MATTHEIJ, *Estimates for the errors in the solution of linear boundary value problems due to perturbations*, Computing, 27 (1981), pp. 299–213.

[34] _____, *Stable computation of solutions of unstable linear initial value recursions*, BIT, 22 (1982), pp. 79–93.

[35] _____, *The conditioning of linear boundary value problems*, SIAM J. Numer. Anal., 19 (1982), pp. 963–978.

[36] _____, *Estimates for the fundamental solutions of discrete* BVP, J. Math. Anal. Appl., 101 (1984), pp. 444–464.

[37] _____, *The stability of LU-decompositions of block tridiagonal systems*, Bull. Austral. Math. Soc., 29 (1984), pp. 177–205.

[38] _____, *Stability of block LU-decompositions of matrices arising from* BVP, SIAM J. Alg. Discr. Meth., 5 (1984), pp. 314–331.

[39] _____, *On decoupling of linear recursions*, report Dept. of Math. Sci., Bull. Austral. Math. Soc., 27 (1983), pp. 347–360.

[40] _____, *Riccati type transformations and decoupling of singularly perturbed* ODE, in Stiff Computation, R. C. Aiken, ed., Oxford Univ. Press, Cambridge, 1985.

[41] R. M. MATTHEIJ AND R. E. O'MALLEY, JR., *Decoupling of boundary value problems for two-time-scale systems*, in Stiff Computation, R. C. Aiken, ed., Oxford Univ. Press, Cambridge, 1985.

[42] R. M. M. MATTHEIJ AND G. W. M. STAARINK, *An efficient algorithm for solving general linear two point* BVP, SIAM J. Sci. Stat. Comp., 5 (1984), pp. 745–763.

[43] G. H. MEYER, *Initial Value Methods for Boundary Value Problems*, Academic Press, New York, 1973.

[44] W. L. MIRANKER, *Numerical Methods for Stiff Equations*, D. Reidel, Dordrecht, 1981.

[45] N. E. NÖRLUND, *Vorlesungen über Differenzenrechnung*, Verlag von Julius Springer, Berlin, 1924.

[46] F. W. J. OLVER, *Numerical solution of second-order linear difference equations*, J. Res. National Bureau of Standards, 71B (1967), pp. 111–129.

[47] R. E. O'MALLEY, JR., *Boundary value problems for linear systems of ordinary differential equations involving small parameters*, J. Math. Mech., 18 (1969), pp. 835–856.

[48] _____, *Singular Perturbations and Optimal Control*, Lecture Notes in Mathematics 680, Springer-Verlag, Berlin, 1978, pp. 170–218.

[49] R. E. O'MALLEY, JR., AND L. R. ANDERSON, *Time-scale decoupling systems*, Optimal Control Appl. and Methods, 3 (1982), pp. 133–153.

[50] R. E. O'MALLEY, AND J. E. FLAHERTY, *Analytical and numerical methods for nonlinear singularly perturbed initial value problems*, SIAM J. Appl. Math., 38 (1980), pp. 225–248.

[51] M. R. OSBORNE, *On shooting methods for boundary value problems*, J. Math. Anal. Appl., 27 (1969), pp. 417–433.

[52] _____, *The stabilized march is stable*, SIAM J. Numer. Anal., 16 (1979), pp. 923–933.

[53] G. W. REDDIEN, *Projection methods for two point* BVP, this Review, 22 (1980), pp. 156–171.

[54] N. ROBERTSON, *The linear two-point boundary value problem on an infinite interval*, Math. Comp., 25 (1971), pp. 475–483.

[55] S. M. ROBERTS AND J. S. SHIPMAN, *Two Point Boundary Value Methods: Shooting Methods*, American Elsevier, New York, 1972.

[56] R. D. RUSSELL, *A comparison of collocation and finite differences for two-point* BVP, this Review, 14 (1977), pp. 19–39.

[57] R. D. RUSSELL AND J. M. VARAH, *A comparison of global methods for two-point boundary value problems*, Math. Comp., 29 (1975), pp. 1007–1009.

[58] R. J. SACKER AND G. R. SELL, *Singular perturbation and conditional stability*, J. Math. Anal. Appl., 76 (1980), pp. 406–431.

[59] M. R. SCOTT, *Invariant Imbedding and its Applications to Ordinary Differential Equations*, Addison-Wesley, Reading, MA, 1973.

[60] M. R. SCOTT AND H. A. WATTS, *Computational solution of linear two point boundary value problems via orthonormalization*, SIAM J. Numer. Anal., 14 (1977), pp. 40–70.

[61] D. M. SLOAN, *Eigenfunctions of systems of linear ordinary differential equations with separated* BC, *using Riccati transformations*, J. Comp. Phys., 24 (1977), pp. 320–330.

[62] A. VAN DER SLUIS, *Estimating the solutions of slowly varying recursions*, SIAM J. Math. Anal., 7 (1976), pp. 662–695.

[63] _____, *Estimating the Solutions of Slowly Varying Differential Equations*, Rep., Mathematisch Instituut, Rijksuniversiteit Utrecht, 1980.

[64] TH. L. STIJN AND A. I. VAN DE VOOREN, *An accurate method for solving Orr-Sommerfeld equations*, J. Engrg. Math., 14 (1980), pp. 17–26.

[65] J. M. VARAH, *A comparison of some numerical methods for two point* BVP, Math. Comp., 28 (1974), pp. 743–755.

[66] _____, *Alternate row and column elimination for solving certain linear systems*, SIAM J. Numer. Anal., 13 (1976), pp. 71–75.

[67] M. VAN VELDHUIZEN, *D-Stability*, SIAM J. Numer. Anal., 18 (1981), pp. 45–64.

[68] W. WASOW, *Asymptotic Expansions for Ordinary Differential Equations*, John Wiley, New York, 1965.

[69] R. WEISS, *The application of implicit Runge–Kutta and collocation methods to* BVP, Math. Comp., 28 (1974), pp. 449–464.

[70] _____, *An analysis of the box and trapezoidal schemes for linear singularly perturbed boundary value problems*, Math. Comp., 42 (1984), pp. 41–67.

[71] J. H. WILKINSON, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, 1965.