Acta Numerica (1996), *pp.* 259–307

Iterative solution of systems of linear differential equations

Ulla Miekkala and Olavi Nevanlinna
Helsinki University of Technology
Institute of Mathematics
Otakaari 1, 02150 Espoo, Finland
E-mail: Ulla.Miekkala@hut.fi, Olavi.Nevanlinna@hut.fi

CONTENTS

1. Introduction

Parallel processing has made iterative methods an attractive alternative for solving large systems of initial value problems. Iterative methods for initial value problems have a history of more than a century, and in the works of Picard (1893) and Lindelöf (1894) they were given a firm theoretical basis. In particular, the *superlinear* convergence on finite intervals is included in Lindelöf (1894).

In the early 1980s *waveform relaxation (WR)* was introduced for the simulation of electrical networks, by Lelarasmee, Ruehli and Sangiovanni-Vincentelli (1982). The methodology has been used in several application areas and has been extended to time-dependent PDEs. There are even books available: White and Sangiovanni-Vincentelli (1987) and Vandewalle (1993).

In this survey we shall only consider systems of ODEs, with some remarks on differential algebraic equations.

Practical problems are usually nonlinear, but it has been our working hypothesis that studying the linear case carefully, specifically introducing a clear notation and suitable concepts, might be what users really need. In applying the ideas to particular problems, including nonlinear mappings, it is often

relatively easy to make intelligent guesses, if one has a good understanding of the nonlinear problem at hand and of the behaviour of the method on linear problems. On papers dealing with strongly nonlinear problems we mention Nevanlinna and Odeh (1987), partly because it was Farouk Odeh who introduced the second author to waveform relaxation in 1983.

The iterative method has many names. To call it waveform relaxation is natural when the application area is electronics. To call it Picard–Lindelöf iteration is historically motivated, although 'block Picard–Lindelöf iteration' would perhaps be more accurate, if cumbersome. The names Picard and Lindelöf also occur in the analysis of the iteration: our convergence theory on finite windows is based on the theory of entire functions, to which both Picard and Lindelöf made important contributions.

We shall not discuss implementation issues at all, not because they are unimportant, but because they are well described in the literature. On linear PDEs we refer to Lubich and Ostermann (1987) and Vandewalle (1992), mentioning that a combination of multigrid in space and waveform relaxation in time is fast and parallelizes reasonably well.

2. Finite windows

2.1. Basic estimates

Let A be a constant d by d complex matrix. We want to solve the initial value problem

$$\dot{x} + Ax = f, \quad x(0) = x_0, \tag{2.1}$$

where the forcing function f generally depends on time t. The matrix A is decomposed as $A = M - N$, where M would typically contain the diagonal blocks of A, and N the off-diagonal couplings. We consider the iteration

$$\dot{x}^k + Mx^k = Nx^{k-1} + f, \quad x^k(0) = x_0. \tag{2.2}$$

If nothing better is available, one can take $x^0(t) = x_0$. In practice, equation (2.2) would be solved by high-quality software, within a (perhaps k-dependent) tolerance. Here we assume it to be solved exactly.

Introducing the following *iteration* operator

$$\mathcal{K}u(t) := \int_0^t e^{-(t-s)M} Nu(s)\,\mathrm{d}s \tag{2.3}$$

we can write (2.2) in the form

$$x^k = \mathcal{K}x^{k-1} + g, \tag{2.4}$$

where

$$g(t) := e^{-tN}x_0 + \int_0^t e^{-(t-s)M} f(s)\,\mathrm{d}s. \tag{2.5}$$

In what follows we shall assume that f is defined for all $t > 0$ and is locally in L_1, that is

$$\int_0^T |f(t)|dt < \infty \quad \text{for all} \quad T < \infty. \tag{2.6}$$

So, we in fact replace (2.1) by the fixed-point problem

$$x = \mathcal{K}x + g, \tag{2.7}$$

which then has a unique solution within continuous functions on $[0, \infty)$.

Proposition 1 Let f be absolutely integrable on bounded subsets of $[0, \infty)$, and let x_0 be given. Then there exists exactly one continuous solution x on $[0, \infty)$ satisfying (2.7). In addition, x is absolutely continuous and satisfies (2.1) almost everywhere.

Proof. To prove a result like this, one only has to show that $(1 - \mathcal{K})^{-1}$ is a bounded operator in $C[0, T]$ for all T. This is included in the estimates of the growth of the resolvent in Proposition 2. That x is absolutely continuous follows by differentiation. □

We shall use $|\cdot|$ to denote the Euclidean norm, and its induced matrix norm, throughout the paper. In $C[0, T]$ we shall then use the uniform norm

$$|x|_T := \sup_{0 \le t \le T} |x(t)|. \tag{2.8}$$

We shall also use $|\cdot|_T$ to denote the induced operator norm

$$|\mathcal{K}|_T := \sup_{|x|_T = 1} : \mathcal{K}x|_T.$$

Theorem 1 For $k \ge 1$, we have the bound

$$|\mathcal{K}^k|_T \le e^{T|M|} \frac{(T|N|)^k}{k!}. \tag{2.9}$$

Proof. The iterates $\mathcal{K}^k$ are integral operators whose kernels are k-fold convolutions of $e^{-sM}N$ satisfying

$$|(e^{-sM}N)^{*k}(t)| \le e^{t|M|}|N| \frac{(|N|t)^{k-1}}{(k-1)!}. \tag{2.10}$$

In fact, (2.10) is trivial for $k = 1$. For $k > 1$ we have

$$\mathcal{K}^k u(t) = \int_0^t e^{-(t-s)M} N \mathcal{K}^{k-1} u(s)\, ds, \tag{2.11}$$

and from $\mathcal{K}^{k-1}u(t) = (e^{-sM}N)^{*(k-1)} * u(t)$ we obtain the induction step needed to conclude (2.10). Then (2.9) follows from (2.10) as

$$|\mathcal{K}^k|_T \le \int_0^T |(e^{-sM}N)^{*k}(t)|\, dt. \tag{2.12}$$

□

To obtain an explicit formula for the resolvent operator

$$R(\lambda, \mathcal{K}) := (\lambda - \mathcal{K})^{-1},$$

consider the following problem for $\lambda \neq 0$:

$$\lambda u - \mathcal{K}u = g. \tag{2.13}$$

Assuming g is smooth, differentiate (2.13) to obtain

$$\dot{u} + (M - \frac{1}{\lambda}N)u = \frac{1}{\lambda}(\dot{g} + Mg). \tag{2.14}$$

Thus the only solution of (2.13) is given by

$$u(t) = R(\lambda, \mathcal{K})g(t) = \frac{1}{\lambda}g(t) + \frac{1}{\lambda^2}\int_0^t e^{-(t-s)(M-\frac{1}{\lambda}N)}Ng(s)\,\mathrm{d}s. \tag{2.15}$$

Proposition 2 The resolvent $R(\lambda, \mathcal{K})$ mapping g to u is given by (2.15) for $\lambda \neq 0$ and it satisfies

$$|R(\lambda, \mathcal{K})|_T \leq \frac{1}{|\lambda|} + \frac{1}{|\lambda|^2}e^{T|M-\frac{1}{\lambda}N|}|N|T. \tag{2.16}$$

Proof. For smooth g, (2.16) follows in the same way as (2.9). As (2.15) only deals with values of g, the bound (2.16) holds as such for all $g \in C[0,T]$. □

2.2. Quasinilpotency, order and type

Bounded operators with spectrum equalling the origin are called quasinilpotent. Their resolvents are *entire* functions in $1/\lambda$ whose growth can be used to bound the powers of the operators. From (2.16) we see that $R(\lambda, \mathcal{K})$ is an entire function in $1/\lambda$ and that it essentially grows like $\exp(T|N|/|\lambda|)$ as $\lambda \to 0$. This means that $R(\lambda, \mathcal{K})$ is of at most *order* 1, and if the order is 1, then the *type* satisfies $\tau \leq T|N|$. These concepts are important because the growth of the resolvent as $\lambda \to 0$ and the decay of the powers are intimately related.

Hadamard (1893) used the maximum modulus

$$M(r, f) := \sup_{|z|=r} |f(z)| \tag{2.17}$$

of an entire function f to define the order

$$\omega := \limsup_{r\to\infty} \frac{\log\log M(r, f)}{\log r}. \tag{2.18}$$

In our case $R(\lambda, \mathcal{K})$ is an operator valued entire function in $1/\lambda$ and likewise we set, following Miekkala and Nevanlinna (1992, page 207),

$$\omega := \limsup_{r\to 0} \frac{\log\log(\sup_{|\lambda|=r}|R(\lambda, \mathcal{K})|_T)}{\log\frac{1}{r}}. \tag{2.19}$$

In general ω could be any nonnegative number, but it follows immediately from (2.16) that $0 \leq \omega \leq 1$. In Miekkala and Nevanlinna (1992) we proved that ω must be a rational number with denominator not exceeding the dimension d of the vectors. Its value depends on the 'graph properties' of M and N only, and in particular is independent of the window size T.

By definition, the order ω is an *asymptotic* concept. Together with the order, one often talks about the type τ of an entire function. This is also an asymptotic concept, which here takes the following form.

If $R(\lambda, \mathcal{K})$ is of positive order ω in $1/\lambda$ then we say that it is of type τ where

$$\tau := \limsup_{r \to 0} r^{\omega} \log(\sup_{|\lambda|=r} |R(\lambda, \mathcal{K})|_T). \tag{2.20}$$

While the order ω is independent of T, the type is of the form $\tau = cT$, where c is a positive constant.

Thus, if $R(\lambda, \mathcal{K})$ is of order $\omega > 0$ and type cT, then

$$\sup_{|\lambda|=r} |R(\lambda, \mathcal{K})|_T \sim e^{cT/r^{\omega}}, \quad \text{as} \quad r \to 0, \tag{2.21}$$

and, in particular, we have for any $\varepsilon > 0$ a constant C such that

$$|R(\lambda, \mathcal{K})|_T \leq \frac{C}{|\lambda|} e^{(1+\varepsilon)cT/|\lambda|^{\omega}} \tag{2.22}$$

holds for all $\lambda \neq 0$. To see how the growth of the resolvent is connected with the decay of the powers of the operator we state the following result.

Theorem 2 Let $\mathcal{A}$ be a bounded linear operator on a Banach space. If $R(\lambda, \mathcal{A})$ is entire in $1/\lambda$ and satisfies

$$\sup_{|\lambda|=r} \|R(\lambda, \mathcal{A})\| \leq \frac{C}{r} e^{\tau/r^{\omega}} \tag{2.23}$$

for all $r > 0$, then

$$\|\mathcal{A}^k\| \leq C(\frac{\tau e \omega}{k})^{k/\omega}, \quad k = 1, 2, 3, \ldots. \tag{2.24}$$

Conversely, if (2.24) holds, then for $0 < \alpha \leq 1/2$ and $r > 0$ we have

$$\sup_{|\lambda|=r} \|R(\lambda, \mathcal{A})\| \leq \frac{1}{r}(1 + \frac{13}{\alpha} C\omega e^{(1+\alpha)\tau/r^{\omega}}). \tag{2.25}$$

Proof. To obtain (2.24) from (2.23) write

$$\mathcal{A}^n = \frac{1}{2\pi \mathrm{i}} \int_{|\lambda|=r} \lambda^n R(\lambda, \mathcal{A}) \, \mathrm{d}\lambda \tag{2.26}$$

and substitute $r^{\omega} = \frac{\tau\omega}{n}$. The reverse direction is also standard in spirit but the actual constants needed in (2.25) require some care. Here we refer to the proof of Theorem 5.3.4 in Nevanlinna (1993). □

2.3. Characteristic polynomial and computation of order and type

As the iteration operator is given by convolution with a matrix valued kernel, it is possible to analyse the growth properties of its resolvent using the Laplace transform.

Consider $u = \mathcal{K}g$ where, say, $|g(t)| \leq Ce^{\alpha t}$ for some positive constants C and α. Taking the Laplace transform we obtain

$$\hat{u}(z) = (z + M)^{-1} N \hat{g}(z) \tag{2.27}$$

and in particular $\hat{u}$ is analytic for sufficiently large $\operatorname{Re} z$. Here $(z + M)^{-1}N$ is the *symbol* of $\mathcal{K}$, denoted by $K(z)$. Analogously, the resolvent operator $R(\lambda, \mathcal{K})$ has the symbol

$$\frac{1}{\lambda}[1 + (z + M - \frac{1}{\lambda}N)^{-1}\frac{1}{\lambda}N]. \tag{2.28}$$

Definition 1 We shall call

$$P(z, \frac{1}{\lambda}) := \det(z + M - \frac{1}{\lambda}N) \tag{2.29}$$

the *characteristic polynomial* of the iteration operator $\mathcal{K}$.

In Section 3 we shall see how the zeros of P determine the spectrum of the operator $\mathcal{K}$ when considered on the infinite time interval $[0, \infty)$: one looks at the supremum of all roots $|\lambda|$ when z travels in a right half plane. Here the properties of $\mathcal{K}$ on the finite interval $[0, T]$ are explained in terms of growth of $|z|$ as λ decays to zero.

Expanding the determinant yields the following result.

Proposition 3 We have

$$P(z, \mu) = \sum_0^d q_j(\mu) z^j, \tag{2.30}$$

where q_j is a polynomial of degree at most $d - j$ and $q_d = 1$.

The equation $P(z(1/\lambda), 1/\lambda) = 0$ determines an algebraic function $z = z(1/\lambda)$ which is d-valued. We need to study the behaviour of $z(1/\lambda)$ as $\lambda \to 0$.

Let z_j denote the branches of z. If z_j is not independent of λ we define ω_j by

$$z_j(\frac{1}{\lambda}) = c_j(\frac{1}{\lambda})^{\omega_j} + o((\frac{1}{\lambda})^{\omega_j}) \quad \text{as} \quad \lambda \to 0 \tag{2.31}$$

(with $c_j \neq 0$). If z_j is independent of λ then we define $\omega_j = 0$. Further we set $\omega := \max \omega_j$.

Lemma 1

$$\omega \in \{\frac{m}{k} \ : \ k, m \quad \text{integers}, \quad 0 \leq m \leq k \leq d, k \neq 0\} \tag{2.32}$$

Proof. The ω_js are computed from the Newton diagram, which is explained in Section 2.5. Theorem 7 implies the claim. □

We can show that ω is the order of the iteration operator $\mathcal{K}$. Consider $\omega > 0$. Let $c := \max |c_j|$ where j runs over those indices for which $\omega_j = \omega$. Then we can formulate our result as follows.

Theorem 3 The iteration operator $\mathcal{K}$ of (2.3) is in $C[0,T]$ of order $\omega = \frac{m}{n}$ with some integers $0 \leq m \leq n \leq d$, independently of T. If $\omega > 0$ then there exists a positive c ($c = \max |c_j|$ as above) such that for all T the type is $\tau = cT$. If $\omega = 0$ then the operator is nilpotent with index $n \leq d$, independently of T. Furthermore, $\mathcal{K}$ is nilpotent if and only if the characteristic polynomial P is independent of λ.

Proof. This is Theorem 4.6 in Miekkala and Nevanlinna (1992).

Since ω can take only a finite number of rational values for a d-dimensional problem, it should not be surprising that ω depends on graph properties of M and N, but not on the values of their elements. This topic is discussed further in Section 2.5.

For $\omega < 1$ we have the following characterization.

Theorem 4 $R(\lambda, \mathcal{K})$ is of the order $\omega < 1$ in $C[0,T]$ if and only if N is nilpotent.

Proof. This is included in Section 2.5. □

Finally, if M and N commute, the analysis of convergence is easy.

Theorem 5 If M and N commute, then either N is nilpotent and then $\mathcal{K}$ is nilpotent too, or the order $\omega = 1$ and the type $\tau = \rho(N)T$.

Proof. For $z \notin \sigma(-M)$, we have

$$K(z)^k = (z + M)^{-k} N^k,$$

which gives

$$\rho(K(z)) \leq \frac{1}{\rho(z + M)} \rho(N).$$

On the other hand, from

$$N^k = (z + M)^k K(z)^k$$

we obtain

$$\rho(N) \leq \rho(z + M) \rho(K(z))$$

and thus

$$\rho(K(z)) = \frac{1}{\rho(z + M)} \rho(N) = (1 + o(1)) \frac{1}{|z|} \rho(N)$$

as $z \to \infty$. The claim follows; see, for example, the proof of Theorem 6, equation (2.37). □

2.4. Norm estimates

While the order and type are asymptotic concepts relating the decay of $|\mathcal{K}^n|_T$ to the behaviour of $z_j = z_j(1/\lambda)$ as $\lambda \to 0$, it is also possible to relate $|\mathcal{K}^n|_T$ to the decay of the symbol as $\operatorname{Re} z \to \infty$. Results of this nature are given in Nevanlinna (1989*b*), and here we present the following basic version.

As in the proof of Theorem 1 of Section 2.1 the claim takes a somewhat better form if formulated for the iterated kernels $(e^{-sM}N)^{*k}(t)$ pointwise in t rather than for the operator norm.

Let m, k be positive integers with $m \geq k$. Consider an estimate of the form

$$|(e^{-sM}N)^{*k}(t))| \leq Be^{\gamma t}\frac{(Bt)^{m-1}}{(m-1)!}, \quad \text{for} \quad t > 0. \tag{2.33}$$

Since $K(z)^k$ is the Laplace transform of $(e^{-sM}N)^{*k}$ we have

$$|K(z)^k| \leq B\int_o^\infty e^{-(\operatorname{Re} z-\gamma)t}\frac{(Bt)^{m-1}}{(m-1)!}\,\mathrm{d}t = (\frac{B}{\operatorname{Re} z-\gamma})^m \quad \text{for} \quad \operatorname{Re} z > +\gamma. \tag{2.34}$$

Thus, an estimate for the iterated kernel implies an estimate for the power of the symbol. The nontrivial fact is that the reverse conclusion also holds.

Theorem 6 Suppose that there are positive integers k and m and positive constants B and γ such that (2.34) holds. Then, for all $j = 1, 2, \ldots$, we have

$$|\mathcal{K}^{jk}|_T \leq \frac{k}{m}de^{\gamma T}(\frac{BeT}{jm})^{jm}. \tag{2.35}$$

Proof. Theorem 2.4.1. in Nevanlinna (1989*b*) is slightly more general but formulated for the iterated kernels. Integrating the kernel estimate gives (2.35) but for a factor of 2. This can be dropped because Spijker (1991) has since proved a sharp version of a lemma by LeVeque and Trefethen. □

Just for comparison, write $\omega := k/m$. Then for $n = jk$, $j = 1, 2, \ldots$ (2.35) reads

$$|\mathcal{K}^n|_T \leq d\omega e^{\gamma T}(\frac{Be\omega T}{n})^{n/\omega}, \tag{2.36}$$

which should be compared with (2.24) and with Theorem 3.

To make this connection explicit observe that (2.31) implies

$$\rho(K(z)) = (1 + o(1))(\frac{c}{|z|})^{\frac{1}{\omega}} \tag{2.37}$$

as $z \to \infty$. In fact, since

$$\lambda - K(z) = \lambda(z+M)^{-1}(z + M - \frac{1}{\lambda}N), \tag{2.38}$$

the eigenvalues of $K(z)$ are obtained from solving $P(z, 1/\lambda) = 0$ and (2.37) follows by 'inverting' (2.31). Thus (2.34) can be viewed as a 'norm' version of (2.37), with $k/m = \omega$ and $B \geq c$.

2.5. Computing order from the graph of A

It was explained in Section 2.3 that the order ω of the iteration operator $\mathcal{K}$ can be computed by solving $z = z(\lambda)$ from the characteristic polynomial $P(z, \frac{1}{\lambda}) = 0$ near $\lambda = 0$. The different branches are of the form (2.31) and the order ω is then the largest ω_j in (2.31). Before finding these solutions we need some background connecting graphs to matrices.

Let $G(B)$ be the directed graph associated with a $d \times d$-matrix B. $G(B)$ contains d vertices v_i. Each nonzero element B_{ij} of B corresponds to an edge of $G(B)$ with weight $B_{i,j}$ directed from v_j to v_i. By a *circuit* of $G(B)$ we mean a subgraph of $G(B)$ which consists of one or more nonintersecting loops. A circuit is denoted by $\mathcal{C}_j$ and its length (or number of edges) by $l(\mathcal{C}_j)$. Further, the product of the weights of the edges is called the *weight of the circuit* and is denoted by $w(\mathcal{C}_j, B)$. The second argument refers to $\mathcal{C}_j$ being a circuit in $G(B)$. Finally, j_{ev} means the number of components of even length in the circuit $\mathcal{C}_j$.

The coefficients b_i in the following expansion of the determinant by diagonal elements

$$\det(zI + B) = z^d + b_1 z^{d-1} + \ldots + b_d \tag{2.39}$$

can be linked to $G(B)$ by noticing first that each b_i is a sum of all principal minors of order i in $\det(B)$. From the definition of the determinant one can then show that these sums have the following graph interpretation (Chen 1976, Theorem 3.1):

$$b_i = \sum_{l(\mathcal{C}_j)=i} (-1)^{j_{ev}} w(\mathcal{C}_j, B), \quad i = 1, \ldots, d, \tag{2.40}$$

where the sum is taken over all circuits of length i in the digraph of B.

Now back to solving for $z(\lambda)$ from $P(z, 1/\lambda) = 0$, or rather from $p(\lambda, z) = 0$, where

$$p(\lambda, z) = \lambda^d P(z, \frac{1}{\lambda}) = \det(z\lambda I + \lambda M - N). \tag{2.41}$$

Now

$$p(\lambda, z) = \sum_{r=0}^{d} p_r(\lambda) z^r,$$

where each p_r is a polynomial in λ. To compute the solution

$$z = c_1\lambda^{\varepsilon_1} + c_2\lambda^{\varepsilon_2} + \cdots, \qquad \varepsilon_1 < \varepsilon_2 < \cdots, \tag{2.42}$$

near $\lambda = 0$, we use the Newton diagram; see, for instance, Vainberg and Trenogin (1974).

For the diagram one needs to determine the smallest power of λ occurring in $p_r(\lambda)$, provided $p_r(\lambda) \not\equiv 0$. If it is denoted by s_r, then the Newton diagram consists of points $\{(r, s_r) : r = 0, \ldots, d\}$ and line segments between the points such that all points are either above or on the line segments. The slopes of the line segments then give the smallest exponent in the expansion (2.42) in such a way that descending line segments correspond to positive ε_1, ascending segments to negative ε_1 and horizontal segments to $\varepsilon_1 = 0$; for details see Vainberg and Trenogin (1974).

Expanding the determinant in (2.41) by diagonal elements, and using (2.40), results in

$$(z\lambda)^d + \sum_{l(\mathcal{C}_j)=1} (-1)^{j_{ev}} w(\mathcal{C}_j, \lambda M - N)(z\lambda)^{d-1} + \cdots +$$

$$\sum_{l(\mathcal{C}_j)=k} (-1)^{j_{ev}} w(\mathcal{C}_j, \lambda M - N)(z\lambda)^{d-k} + \cdots + \det(\lambda M - N) = 0. \tag{2.43}$$

Each weight $w(\mathcal{C}_j, \lambda M - N)$ is a polynomial in λ, since the weights of the edges contained in circuit $\mathcal{C}_j$ are now of the form $\lambda M_{ij} - N_{ij}$. To draw the Newton diagram we need to know the smallest power of λ occurring in the coefficient polynomial of each z^{d-k}.

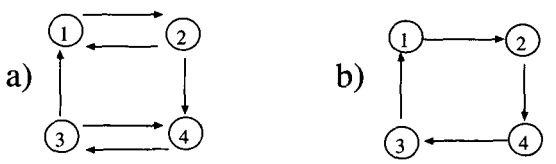

Fig. 1.

Example 1 Let the graph of Figure 1a) represent the matrix A decomposed into $M - N$. $G(A)$ contains four different circuits. Three of them should be obvious, the fourth contains the two small loops as its components. Consider the circuit $\mathcal{C}_1$ of $G(A)$ in Figure 1b). The weight of $\mathcal{C}_1$ becomes

$$w(\mathcal{C}_1, \lambda M - N) =$$

$$(\lambda M_{21} - N_{21})(\lambda M_{42} - N_{42})(\lambda M_{34} - N_{34})(\lambda M_{13} - N_{13}). \tag{2.44}$$

The smallest power of λ in this polynomial clearly depends on how many elements N_{ij} vanish.

The following definition is useful.

Definition 2 Let m_j be the number of nonzero elements of M in the weight $w(\mathcal{C}_j, M-N)$ of the circuit $\mathcal{C}_j$. Similarly, let n_j be the number of nonzero elements of N belonging to circuit $\mathcal{C}_j$.

Clearly $0 \le m_j,\ n_j \le l(\mathcal{C}_j)$ and $l(\mathcal{C}_j) \le m_j + n_j \le 2l(\mathcal{C}_j)$.

The lowest order of λ in the polynomial $w(\mathcal{C}_j, \lambda M - N)$ is $l(\mathcal{C}_j) - n_j$ (and the highest order m_j). Let us denote $n(k) := \max\{n_j : l(\mathcal{C}_j) = k\}$. If no cancellation of terms occurs, then the coefficient of z^{d-k} in (2.43) is a multiple of the polynomial

$$\lambda^{d-k}(\lambda^{k-n(k)} + \text{higher order terms in } \lambda).$$

This means that the $d+1$ points in the Newton diagram of (2.43) are

$$(d,d) \quad (\text{for } z^d) \text{ and } (d-k, d-n(k)), 1 \le k \le d \quad (\text{for } z^{d-k}).$$

The largest slope in the diagram is given by

$$\max_k \frac{d-(d-n(k))}{d-(d-k)} = \max_k \frac{n(k)}{k} = \max_j \frac{n_j}{l(\mathcal{C}_j)}.$$

This corresponds to the solution of (2.43) near $\lambda = 0$

$$z = c\Big(\frac{1}{\lambda}\Big)^{\max_j n_j/l(\mathcal{C}_j)} + \cdots, \tag{2.45}$$

which should be compared with (2.42). If there is cancellation of terms in the coefficients of (2.43), then it is possible that $c = 0$ in (2.45) and the first nonzero term in the expansion $z = c_1\lambda^{\varepsilon_1} + \cdots$ satisfies $\varepsilon_1 > \max_j n_j/l(\mathcal{C}_j)$. The exponent is a rational number yet to be found by the Newton diagram. By Section 2.3, $-\varepsilon_1$ gives the order ω.

Theorem 7 The order of the iteration operator $\mathcal{K}$ defined in (2.3) can be computed from the digraph of $G(M-N)$ and

$$\omega = \max\{\frac{m}{k} : \text{ there exists } \mathcal{C}_j \text{ such that } l(\mathcal{C}_j) = k \text{ and } n_j = m, \text{ and } \sum_{l(\mathcal{C}_j)=k, n_j=m} (-1)^{j_{ev}} w(\mathcal{C}_j, M-N) \neq 0\}$$

where each $\mathcal{C}_j$ is a circuit in the digraph $G(M-N)$ and n_j is given in Definition 2.

Corollary 1 The order of the iteration operator K has the upper bound

$$\omega \le \max_j \frac{n_j}{l(\mathcal{C}_j)}.$$

Since $l(\mathcal{C}_j) \le d$ and $n_j \le l(\mathcal{C}_j)$, it is easy to see that Lemma 1 of Section 2.3 holds.

We will now prove Theorem 4, which states that $\omega < 1$ if and only if N is nilpotent.

Proof of Theorem 4. We prove the complement: $\omega = 1$ if and only if N is not nilpotent. Since the Newton diagram always contains the point (d, d), we obtain $\omega = 1$ if and only if it also contains the point $(d-k, d-k)$ for some $k \in \{1, \ldots, d\}$. From (2.43), we conclude that this can happen if and only if, for some $k \in \{1, \ldots, d\}$, the polynomial $\sum_{l(\mathcal{C}_j)=k}(-1)^{j_{ev}} w(\mathcal{C}_j, \lambda M - N)$ contains a nonzero constant term. Such a constant term is a weight of a circuit containing only elements of N, whence

$$\omega = 1 \quad \text{if and only if} \quad \sum_{l(\mathcal{C}_j)=k} (-1)^{j_{ev}} w(\mathcal{C}_j, N) \neq 0 \text{ for some } k = 1, \ldots, d.$$

By (2.40), this implies that at least one coefficient in expansion (2.39) for N is nonzero. But this happens if and only if N is not nilpotent. □

For the Gauss–Seidel iteration, A is decomposed so that N contains the upper triangular part of A.

Corollary 2 The order of the iteration operator corresponding to Gauss–Seidel iteration is always < 1.

The Newton diagram can also be used to compute the type of the iteration operator $\mathcal{K}$. For the derivation of the following result we refer to Miekkala and Nevanlinna (1992).

Theorem 8 Let C_{max} denote the set of all circuits yielding the maximum quotient in Theorem 7. Then c in the expression of the type $\tau = cT$ of the iteration operator $\mathcal{K}$ is the largest root in absolute value of the equation

$$c^d + \sum_{C_{max}} \left(\sum_{l(\mathcal{C}_j)=k \& n_j=m} (-1)^{j_{ev}} w(\mathcal{C}_j, M - N) \right) c^{d-k} = 0.$$

If there is only one circuit, C_m say, giving ω, then c satisfies the equation

$$c^k + (-1)^{m_{ev}} w(\mathcal{C}_m, M - N) = 0,$$

where $k = l(\mathcal{C}_m)$ and m_{ev} is the number of even components in $\mathcal{C}_m$.

We have shown how the order of the iteration operator can be computed from the graph $G(M-N)$. For large systems this may be a very large graph. It is possible to construct smaller graphs from $G(M-N)$ still containing the essential information for computing ω. Two such graphs are defined in Miekkala and Nevanlinna (1992). One gives ω exactly, the other, based on block partitioning of A, gives an upper bound on ω.

Quite often the dependencies between subsystems are modelled by constructing a graph G_S, where one vertex corresponds to one subsystem, and there is an edge from vertex v_i to v_j if and only if there is at least one connection from some of the vertices belonging to subsystem i in the original graph to some of the vertices belonging to subsystem j. This does give an upper bound for $\max_j n_j/l(\mathcal{C}_j)$, but it may be pessimistic.

Our results are related to the accuracy increase studied by Juang (1990). Assuming Taylor series expansions near time $t = 0$ for the iterates and the exact solution, the accuracy of the iterate is defined to be the number of correct terms in its Taylor series. The classical Picard–Lindelöf iteration ($N = -A$) increases accuracy by at least one at every iteration. Juang showed how the accuracy increase of block Gauss–Seidel can be studied by examining circuits in the dependency graph G_S. If all subsystems contain only one vertex (pointwise Gauss–Seidel), then $\min_j l(\mathcal{C}_j)/n_j$ equals the lower bound for accuracy increase proved by Juang. In general, we have the inequalities

$$\omega \leq (\text{accuracy increase})^{-1} \leq \max_{G_S} \frac{n_j}{l(\mathcal{C}_j)}.$$

2.6. Other remarks

From the basic estimate of the form

$$|\mathcal{K}^n|_T \leq C(\frac{BeT\omega}{n})^{n/\omega}, \tag{2.46}$$

we see that, for relatively long windows, we reach the 'superlinear era' after $\mathcal{O}(T)$ sweeps, when $n > Be\omega T$. Simultaneously, the error at $t \gg \mathcal{O}(T)$ still decays at most linearly, if at all.

We shall see below that the convergence on $[0, \infty)$ is of the following form. If $\mathcal{K}^d \neq 0$, then the spectral radius $\rho(\mathcal{K})$ is positive and

$$|\mathcal{K}^n|_\infty^{1/n} \to \rho(\mathcal{K}). \tag{2.47}$$

Combined with the superlinear estimate, the convergence can be bounded by

$$|\mathcal{K}^n|_T \leq \min\{C(\frac{Be\omega T}{n})^{n/\omega}, C_\varepsilon(\rho(\mathcal{K}) + \varepsilon)^n\}. \tag{2.48}$$

For sufficiently small $\rho(\mathcal{K})$ and large T, superlinear convergence does not occur for practical tolerances.

Since $|\mathcal{K}^n|^{1/n} \to 0$ as $n \to \infty$ but $|\mathcal{K}^n|_\infty^{1/n} \to \rho(\mathcal{K}) > 0$, the spectrum is not continuous as $T \to \infty$. Trefethen (1992) has defined the pseudospectrum for an operator $\mathcal{A}$ on a Banach space by

$$\Lambda_\varepsilon(\mathcal{A}) := \{\lambda \in \mathbb{C} : \|R(\lambda, \mathcal{A})\| \geq 1/\varepsilon\}. \tag{2.49}$$

Here it is understood that $\|R(\lambda, \mathcal{A})\| = \infty$ if and only if $\lambda \in \sigma(\mathcal{A})$. Notice that $\sigma(\mathcal{A}) = \bigcap_{\varepsilon>0} \Lambda_\varepsilon(\mathcal{A})$.

Lumsdaine and Wu (1995) have shown that even though the spectrum is not continuous at $T = \infty$, we do obtain

$$\lim_{T\to\infty} \Lambda_\varepsilon(\mathcal{K}_T) = \Lambda_\varepsilon(\mathcal{K}_\infty) \tag{2.50}$$

for $\varepsilon > 0$, where $\mathcal{K}_T$ and $\mathcal{K}_\infty$ denote the operator $\mathcal{K}$ acting on $L_2[0, T]$ and $L_2[0, \infty)$ respectively.

3. Infinite windows

3.1. Definition of spaces

In practice one would not usually iterate on $[0, \infty)$, but the infinitely long window is a natural setup for stiff problems and for DAEs: for the fast transients a finite window $[0, T]$ can be regarded as infinitely long.

The exceptional situation with stiff problems would appear if couplings are very small, that is, $T|N| = \mathcal{O}(1)$, then by the discussion of Section 2 we would have superlinear convergence. For $T|N| \gg 1$ we typically obtain only linear convergence, and one of the first interesting things is that the linear rate given by the spectral radius is very insensitive to the choice of norm.

Let X be a Banach space of functions $x : [0, \infty) \to \mathbb{C}^d$ such that the following conditions hold:

(i) $e^{\lambda t}c$ with $\operatorname{Re}\lambda > 0$ and $0 \neq c \in \mathbb{C}^d$ is not in X;

(ii) $e^{\lambda t}p(t)$, where p is a $\mathbb{C}^d$-valued polynomial and $\operatorname{Re}\lambda < 0$, is in X;

(iii) $x \mapsto \int_0^t e^{(s-t)B}Cx(s)\,ds$, where B, C are constant matrices and the eigenvalues of B have positive real parts, are bounded operators in X;

(iv) test functions C_0^∞ are dense in X.

Let $\|\cdot\|$ denote the norm in X. By (iii), $\sigma(M) \subset \mathbb{C}_+$ implies that $\mathcal{K}$ is a bounded operator in X. In order to formulate our results we recall the definition of the symbol

$$K(z) := (z + M)^{-1}N \tag{3.1}$$

The basic property of X is that it is 'unweighted' in exponential scales. However, such scaling is trivial and simply translates the imaginary axis: requirements on real parts being positive would become positive lower bounds on real parts.

The properties (i) and (iii) imply that if we try to solve our initial value problem in X we must require that all eigenvalues of A have positive real parts. Furthermore the following holds.

Theorem 9 If all eigenvalues of A have positive real parts, then $\mathcal{K}$ is a bounded operator in X if and only if the eigenvalues of M have positive real parts.

Proof. The sufficiency part is of course obvious, while the necessity needs a small discussion. In Miekkala and Nevanlinna (1987*a*) the result is proved for L_p spaces.

Assume that $\mathcal{K}$ is bounded in X and that μ is an eigenvalue of M with

nonpositive real part. Let J denote the Jordan block

$$J = \begin{pmatrix} \mu & 1 & & \\ & \mu & \ddots & \\ & & \ddots & 1 \\ & & & \mu \end{pmatrix},$$

associated with this eigenvalue, and let S be a similarity transform such that $\tilde{M} = S^{-1}MS$ is of the form

$$\tilde{M} = \begin{pmatrix} J & 0 \\ 0 & M_0 \end{pmatrix}.$$

Put $\tilde{N} = S^{-1}NS$ and let the corresponding operator be denoted by $\tilde{K}$. Since multiplication by a constant matrix is bounded in X, and $\tilde{K} = S^{-1}KS$, $\tilde{K}$ is bounded in X. Let the block structure induced by $\tilde{M}$ be denoted by

$$\tilde{N} = \begin{pmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{pmatrix}.$$

We claim that there exists $c \in \mathbb{C}^d$ such that

$$\tilde{N}c = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix},$$

with $b_1 \neq 0$. Indeed, if $N_{11} \neq 0$, then this is trivial, while if $N_{11} = 0$, then the claim follows from $N_{12} \neq 0$. N_{11} and N_{12} cannot simultaneously vanish because every eigenvalue of A has positive real part.

Let c be as above and let λ be any complex number such that $\operatorname{Re}\lambda < 0$. Then $u := e^{\lambda t}c$ is an element of X, whence $\tilde{K}u \in X$. Let k be the largest index i for which $b_{1i} \neq 0$, where $b_1 = (b_{11}, b_{12}, \ldots)^T$. Then the kth component of the vector $\tilde{K}u$ satisfies

$$\left(\tilde{K}u\right)_k(t) = \int_0^t e^{\mu(s-t)}e^{\lambda s}\,\mathrm{d}s \quad b_{1k}e_k,$$

where $e_k \in \mathbb{C}^d$ denotes the usual cordinate vector.

For any $f \in X$, define

$$\mathcal{L}f := \left(e_k, \tilde{K}f\right)c,$$

where $(\cdot,\cdot)$ denotes the usual inner product in $\mathbb{C}^d$. Since $\mathcal{L}f$ can also be written as $C\tilde{K}$, with $C = ce_k^T$, we see that $\mathcal{L}$ is bounded in X. By construction, $\mathcal{L}u = e^{-\mu t} * e^{\lambda t}c$ and, since $u = e^{\lambda t}c \in X$, we have $e^{-\mu t} * e^{\lambda t}c \in X$. This implies $e^{-\mu t}c \in X$, which, by (i), implies $\operatorname{Re}\mu \geq 0$. By assumption, $\operatorname{Re}\mu \leq 0$ and we conclude that $\mu = \mathrm{i}\xi$ and $\nu := e^{-\mathrm{i}\xi t}c \in X$.

It is easily checked that

$$\mathcal{L}^n v = \frac{t^n}{n!}v,$$

for all $n \geq 0$. But, since $\mathcal{L}$ is bounded, we may put

$$w := \left(\sum_{n \geq 0} \frac{(t/\|2\mathcal{L}\|)^n}{n!} \right) v = \exp(t/\|2\mathcal{L}\|)v$$

and obtain the contradiction

$$w = \exp\left((\|2\mathcal{L}\|^{-1} - \mathrm{i}\xi)t \right) c \in X.$$

□

3.2. The spectrum and the spectral radius

From now on we assume that the eigenvalues of M have positive real parts. The main result here may be found in Nevanlinna (1990*a*).

Theorem 10 In every space X, we have $\sigma(\mathcal{K}) = \mathrm{cl} \bigcup_{\mathrm{Re}\, z \geq 0} \sigma(K(z))$.

We state some consequences before embarking on the proof.

Corollary 3 $\rho(\mathcal{K}) = \max\{K(\mathrm{i}\xi) : \xi \in \mathbb{R}\}$.

Proof. Since the eigenvalues of M have positive real parts, $K(z)$ is analytic in the closed right half plane. The claim follows from Theorem 10 using the maximum principle, on a Riemann surface corresponding to the algebraic function formed by the eigenvalues of $K(z)$, and the fact that all eigenvalues of $K(z)$ vanish at infinity. Alternatively, we may apply the maximum principle directly to the spectral radius of $K(z)$, because it is a subharmonic function; see Theorem 3.4.7 in Aupetit (1991). □

Corollary 4 $\sigma(\mathcal{K})$ is compact and connected, and $0 \in \sigma(\mathcal{K})$.

Proof. All branches of the algebraic function vanish at infinity. Thus all components of $\sigma(\mathcal{K})$ contain the origin, which implies connectedness. Compactness is obvious, remembering that K is bounded by assumption (iii). □

Corollary 5 $\rho(\mathcal{K}) = 0$ if and only if there exists $m \leq d$ such that $\mathcal{K}^m = 0$.

Proof. If $\rho(\mathcal{K}) = 0$ then $K(z)$ is nilpotent for all z in the right half plane. Thus there exists $m \leq d$ such that $K(z)^m \equiv 0$, for all z. Now $\mathcal{K}^m$ applied to, say, test functions can be written using the inverse Laplace transform in terms of $K(z)^m$. Since test functions are dense in X and $\mathcal{K}^m$ is continuous, $\mathcal{K}^m$ must vanish in all of X. The converse is trivial. □

Comparing this corollary with Theorem 3 in Section 2 we see that $\rho(\mathcal{K}) > 0$ in X if and only if $\mathcal{K}$ is of positive order in $C[0, T]$.

Proof of Theorem 10. The formula

$$R(\lambda, \mathcal{K})f(t) = \frac{1}{\lambda}f(t) + \frac{1}{\lambda^2}\int_0^t e^{-(t-s)(M-\lambda^{-1}N)}Nf(s)\,ds \tag{3.2}$$

is valid for all $t > 0$ and $\lambda \neq 0$, and at least for smooth f (see Section 2.1). This implies immediately that $R(\lambda, \mathcal{K})$ is bounded in X by (iii) and (iv), provided that all eigenvalues of $M - \lambda^{-1}N$ have positive real parts.

On the other hand, suppose that $M - \lambda^{-1}N$ has an eigenvalue, μ say, with negative real part. Denoting the corresponding eigenvector by b, we choose f to be the solution of $\dot{f} + Mf = 0$ with $f(0) = b$. Thus f is in X by (ii). However, from (2.14) we see that $u(t) = e^{-\mu t}b$, and thus $u \notin X$ by (i). Finally, suppose that $M - \lambda^{-1}N$ has a purely imaginary nonzero eigenvalue λ_0. Then, since $M - \lambda^{-1}N$ is analytic in λ, $M - \lambda^{-1}N$ would have at least one eigenvalue with negative real part near λ_0, unless the eigenvalue is constant. In that case M would also have a purely imaginary eigenvalue, contradicting our hypothesis. Since the spectrum is closed, such a λ_0 does belong to the spectrum, and we can conclude that $\lambda \in \sigma(\mathcal{K}) \setminus \{0\}$ if and only if $\det(M - \lambda^{-1}N - \mu)$ vanishes for some μ with nonpositive real part.

Writing $z = -\mu$ and recalling that $z + M$ is invertible for $\operatorname{Re} z \geq 0$, we deduce that $\det((\lambda - z + M)^{-1}N)$ vanishes for some z with non-negative real part.

Thus $0 \neq \lambda \in \sigma(\mathcal{K})$ if and only if there is a z, $\operatorname{Re} z \geq 0$, such that $\lambda \in \sigma(K(z))$. Since $K(z) \to 0$ as $z \to \infty$ and $\sigma(\mathcal{K})$ is closed, we deduce $0 \in \sigma(\mathcal{K})$, and the claim follows. □

Since the spectral radius $\rho(\mathcal{K})$ is independent of the space X we may loosely say that the iteration converges on $[0, \infty)$ if and only if $\rho(\mathcal{K}) < 1$. Note that

$$\|\mathcal{K}^n\|^{1/n} \to \rho(\mathcal{K}), \tag{3.3}$$

so that, for any $\varepsilon > 0$, there exists $C < \infty$ for which

$$\|\mathcal{K}^n\| \leq C(\rho(\mathcal{K}) + \varepsilon)^n, \tag{3.4}$$

but that in general C depends on both ε and on the ambient norm.

The formula $\rho(\mathcal{K}) = \max_\xi \rho(K(\mathrm{i}\xi))$ is very easy to use in practice. For example, in several special cases one has

$$\max_\xi \rho(K(\mathrm{i}\xi)) = \rho(K(0)), \tag{3.5}$$

which simply means that the convergence is dominated by the speed of convergence of the iteration

$$Mx^{k+1} = Nx^k + b \tag{3.6}$$

for $Ax = b$. Such situations can occur, for instance in Jacobi splittings of linearized versions of parabolic equations. Results of this form have been

discussed in Miekkala and Nevanlinna (1987*a*). This paper also contains results where $\rho(\mathcal{K}) > \rho(K(0))$. As an example we mention SOR for consistently ordered matrices. In this case the rate of convergence ($\rho(K(0))$) is known for iteration (3.6). It turns out that for consistently ordered matrices, we obtain $\rho(\mathcal{K}) = \rho(K(0))$ for small values of the overrelaxation parameter ω. However, when ω is close to 2, we have $\rho(\mathcal{K}) > \rho(K(0))$ and the iteration can diverge. For the precise result, see Theorem 4.1 in Miekkala and Nevanlinna (1987*a*).

In comparing two splittings it is important to notice that a splitting that looks favourable on $[0, \infty)$ may look inferior on $[0, T]$ and vice versa. For example, by Theorem 4 of Section 2, the order of $R(\lambda, \mathcal{K})$ is always less than 1 for Gauss–Seidel splitting on $C[0, T]$, whilst for overrelaxation splittings the order equals 1 if the diagonal does not vanish. On the other hand, for consistently ordered matrices, for instance, $\rho(\mathcal{K})$ initially decreases as the overrelaxation parameter increases from 1. So the Gauss–Seidel splitting provides ultimately the fastest convergence rate on finite windows, but on the infinite interval creates *propagating error waves*, which are best damped with a modest amount of overrelaxation. Overrelaxing too much will in turn cause growing error waves, making the process diverge on the infinite window.

3.3. On generalizing the theory for DAE systems

Let us change the model problem to

$$B\dot{x} + Ax = f \tag{3.7}$$

with consistent initial values for x, where B may be singular and f is sufficiently smooth (the required smoothness depends on the index of the system). The boundedness assumption for continuous solutions of (3.7) on the infinite time interval becomes

$$\det(zB + A) \neq 0, \quad \operatorname{Re} z \geq 0.$$

To see this and for the whole analysis of Miekkala (1989), one needs to use the Kronecker Canonical Form (KCF) of the DAE (Gantmacher 1959). Decompositions of the matrices $B = M_B - N_B$ and $A = M_A - N_A$ define the dynamic iteration for (3.7)

$$M_B\dot{x}^n + M_A x^n = N_B\dot{x}^{n-1} + N_A x^{n-1} + f, \quad n = 1, 2, \ldots \tag{3.8}$$

with consistent initial values for x. The iteration operator can now be written after transformation of (3.8) into KCF form, and constitutes two parts, one being an integral operator and the other a sum of matrix multiplication and differentiation operators. The basic difference to the ODE case is that, in order to guarantee boundedness of iteration (3.8), one needs to preserve the structure of the DAE while decomposing B and A in (3.7). Essentially, we mean that the index is preserved and the state variables and algebraic

variables are preserved. The condition is formulated for the KCF of (3.8), but in Miekkala (1989) there is an error, corrected in Miekkala (1991). The space where x is iterated by (3.8) is chosen in Miekkala (1989) to be that of continuously differentiable functions with appropriate norm, but one could equally well use the space of continuous functions with the uniform norm. The smoothness requirement for f is essential since, for high index DAE systems, some components of the solution of (3.7) depend on derivatives of f. For index one (or zero) systems one might consider iteration (3.8) in L^p-space (both f and x^n in L^p), as in the ODE-case; the results of Miekkala (1989) still hold. For high index DAEs the space has to be modified so that the components corresponding to high index algebraic variables have different requirements from the other components. In general it would be difficult to recognize these components, but in applications it is sometimes possible. In Section 7 this kind of modified L^p-space formulation is used for the index two case.

In Miekkala (1989), assuming that the algebraic part of the iteration operator is bounded, the other results are analogous to the ODE case. For example,

$$\det(zM_B + M_A) \neq 0, \quad \operatorname{Re} z \geq 0, \tag{3.9}$$

is needed for boundedness of the iteration. The convergence rate is given by the 'Laplace transform' of (3.8),

$$\rho(\mathcal{K}_{DAE}) = \sup_{\operatorname{Re} z \geq 0} \rho((zM_B + M_A)^{-1}(zN_B + N_A)). \tag{3.10}$$

Convergence results, like those for consistently ordered matrices, can be generalized to special index one systems.

4. Acceleration techniques

We can accelerate waveform relaxation in two ways: we can try to get the error to decrease more rapidly per iteration, or spend less time integrating the early sweeps. The latter strategy is outlined in connection with the discretization, while here we address the former possibility.

4.1. The speed of optimal Krylov methods

Suppose the initial value problem has been transformed into the fixed-point problem

$$x = \mathcal{K}x + g. \tag{4.1}$$

Instead of iterating as usual, that is

$$x^{k+1} := \mathcal{K}x^k + g, \tag{4.2}$$

we could in principle keep all the vectors $\{x^k\}$ in memory and try to find as good a linear combination of these as possible.

We outline first the abstract Krylov subspace method approach; see Nevanlinna (1993). Let $\mathcal{A}$ be a bounded operator in a Banach space and b a vector in that space. Put

$$K_n(\mathcal{A}, b) := \text{span}\{\mathcal{A}^j b\}_0^{n-1} \tag{4.3}$$

and

$$K(\mathcal{A}, b) := \text{cl span}\{\mathcal{A}^j b\}_0^{\infty}. \tag{4.4}$$

Thus, $K(\mathcal{A}, b)$ is the smallest closed invariant subspace of $\mathcal{A}$ that contains b. In fact either $\dim K_n(\mathcal{A}, b) = n$ or there exists $m < n$ such that, for all $k > m$,

$$K_k(\mathcal{A}, b) = K_m(\mathcal{A}, b). \tag{4.5}$$

If we are given a fixed point problem

$$x = \mathcal{A}x + b, \tag{4.6}$$

such that $1 \notin \sigma(\mathcal{A})$, then clearly $x = (1 - \mathcal{A})^{-1} b$. Consider the following simple embedding:

$$x_\lambda = \frac{1}{\lambda}\mathcal{A}x_\lambda + b, \tag{4.7}$$

and assume that $\sigma(\mathcal{A})$ does not separate 1 from ∞. Then there exists a path $\lambda(s) : \lambda(1) = 1, \quad \lambda(\infty) = \infty$ such that (4.7) has a solution x_λ and clearly this solution is continuous along the path. Trivially, the Krylov subspace of $\lambda^{-1}\mathcal{A}$ equals that of $\mathcal{A}$ for nonzero λ. For $|\lambda| > \|\mathcal{A}\|$ we have

$$x_\lambda = \sum_0^\infty (\lambda^{-1}\mathcal{A})^k b, \tag{4.8}$$

which shows that $x_\lambda \in K(\mathcal{A}, b)$. By continuity, as $\lambda(s) \to 1$ and because $K(\mathcal{A}, b)$ is a closed set, we have $x \in K(\mathcal{A}, b)$, and $K(\mathcal{A}, b)$ is invariant for $(1 - \mathcal{A})^{-1}$ as well.

We assume in the following that $1 \notin \sigma(\mathcal{A})$ and that $\sigma(\mathcal{A})$ does not separate 1 from ∞. That the latter must be assumed is clear from the maximum principle, but can be understood immediately from the following example.

If $\mathcal{A} := \rho S$ where $\rho > 1$ and $S : e_j \mapsto e_j + 1$ is the unitary shift in $\ell_2(\mathbb{Z})$, then $\sigma(\mathcal{A})$ is the circle centred at the origin of radius ρ and 1 is separated from ∞. If $b := e_0$, then $K(\mathcal{A}, e_0) = \text{cl span}\{e_j\}_0^\infty$ whilst the solution $x \notin K(\mathcal{A}, e_0)$. In fact, $x = \sum_{-\infty}^0 \rho^{1-j} e_j$.

Every vector in $K_n(\mathcal{A}, b)$ is of the form $q_{n-1}(\mathcal{A})b$ for some polynomial q_{n-1} of degree $n - 1$. Let us write

$$p(\lambda) := 1 - (1 - \lambda)q(\lambda), \tag{4.9}$$

where q is a given polynomial; then $q(\mathcal{A})$ approximates $(1 - \mathcal{A})^{-1}$ well if and

only if $p(\mathcal{A})$ is small. In fact we have the following result (Nevanlinna 1993, Proposition 1.6.1).

Proposition 4 $$\frac{1}{\|1-\mathcal{A}\|}\|p(\mathcal{A})\| \le \|(1-\mathcal{A})^{-1}-q(\mathcal{A})\| \le \|(1-\mathcal{A})^{-1}\|\|p(\mathcal{A})\|.$$

Now, it is of interest to ask how small $p(\mathcal{A})$ can be. Therefore set $b_n(\mathcal{A}) := \inf \|p(\mathcal{A})\|$ where the infimum is taken over all polynomials of degree at most n, satisfying $p(1) = 1$ (see (4.9)).

Definition 3 (Nevanlinna 1990a, and Definition 3.3.1 in Nevanlinna 1993). Given a bounded $\mathcal{A}$, define

$$\eta(\mathcal{A}) := \inf_n b_n(\mathcal{A})^{1/n}.$$

We call $\eta(\mathcal{A})$ the *optimal reduction factor* of $\mathcal{A}$.

The main properties of $\eta(\mathcal{A})$ are collected in the following theorem.

Theorem 11

(i) $\eta(\mathcal{A}) < 1$ if and only if $1 \notin \sigma(\mathcal{A})$ and $\sigma(\mathcal{A})$ does not separate 1 from ∞;
(ii) if $\eta(\mathcal{A}) < 1$ then $\eta(\mathcal{A}) = 0$ if and only if $\mathrm{cap}(\sigma(\mathcal{A})) = 0$;
(iii) $0 < \eta(\mathcal{A}) < 1$, then the value of $\eta(\mathcal{A})$ only depends on $\sigma(\mathcal{A})$ and is given by $\eta(\mathcal{A}) = e^{-g(1)}$, where g is the (extended) Green's function, satisfying

- g is harmonic in the unbounded component G of $\mathbb{C}\backslash\sigma(\mathcal{A})$;
- $g(\lambda) = \log|\lambda| + \mathcal{O}(1)$ as $\lambda \to \infty$;
- $g(\lambda) \to 0$ as $\lambda \to \zeta$ from G, for every $\zeta \in \partial G \subset \partial\sigma(\mathcal{A})$.

Proof. These are covered by Theorem 3.3.4 and Theorem 3.4.9 in Nevanlinna (1993)

Operators $\mathcal{A}$ for which $\mathrm{cap}(\sigma(\mathcal{A})) = 0$ are *quasialgebraic*. In such a case $\sigma(\mathcal{A})$ cannot separate 1 from ∞, so that we can combine (i) and (ii) in the statement: The optimal reduction factor vanishes exactly for quasialgebraic operators with $1 \notin \sigma(\mathcal{A})$.

This is analogous to the vanishing of the spectral radius for quasinilpotent operators.

We shall say that $\mathcal{A}$ is *algebraic* if $q(\mathcal{A}) = 0$ for some polynomial q. Thus nilpotent operators form a subclass of algebraic operators.

4.2. Finite windows

Consider $\mathcal{K}$ in $C[0,T]$. From Section 2 we know that

$$|\mathcal{K}^n|_T \sim (\frac{cTe\omega}{n})^{n/\omega} \tag{4.10}$$

as $n \to \infty$. We do not know sharp lower bounds for

$$b_n(\mathcal{K}) = \inf_{\deg p \le n,\, p(1)=1} |p(\mathcal{K})|_T \tag{4.11}$$

for general $\mathcal{K}$, but we give an illustrative example instead. Consider the following operator $\mathcal{V}$,

$$\mathcal{V}u(t) := \int_0^t u(s)\,\mathrm{d}s \tag{4.12}$$

or $M = 0$ and $N = 1$. Then clearly

$$|\mathcal{V}^n|_T = \frac{T^n}{n!}, \tag{4.13}$$

so that $\omega = 1$, $\tau = T$. The following result shows that, when it is optimally accelerated, we obtain a speed of convergence in which the order is still 1 but the type is lowered from T to $T/4$.

Theorem 12 Let $\mathcal{V} = \int_0^t$ operate in $C[0,T]$. Then for $n \ge 2$

$$e^{-T}\frac{(T/4)^n}{n!} \le b_n(\mathcal{V}) \le 8(1+T)e^T\frac{(T/4)^n}{(n-1)!} \tag{4.14}$$

Proof. This is proposition 5.2.5 in Nevanlinna (1993).

Thus, the speed can be accelerated, but not dramatically.

4.3. Infinite windows

Let X be any space considered in Section 3.1. The first result says that acceleration is possible, but then we shall see that the acceleration is often only of modest nature.

Theorem 13

(i) $\eta(\mathcal{K}) = 0$ only if $\rho(\mathcal{K}) = 0$.
(ii) If $0 < \eta(\mathcal{K}) < 1$ then $\eta(\mathcal{K}) < \rho(\mathcal{K})$.

Proof. This is Theorem 4 in Nevanlinna (1990*a*). □

Recall that $\rho(\mathcal{K}) = 0$ implies that $\mathcal{K}$ is nilpotent, so that the interesting case is (ii). By Theorem 11 $\eta(\mathcal{K}) < 1$ if and only if $1 \notin \sigma(\mathcal{K})$ and 1 is not separated from ∞ by $\sigma(\mathcal{K})$. In this setup, the case $1 \notin \sigma(\mathcal{K})$ can occur, so that the fixed point problem

$$x = \mathcal{K}x + g \tag{4.15}$$

would as such be well posed in X, but for all normalized polynomials p we would have

$$\|p(\mathcal{K})\| \ge 1. \tag{4.16}$$

In fact, since $\sigma(\mathcal{K})$ is connected and contains the origin, we require real M and N such that, if $\sigma(\mathcal{K})$ is symmetric over the real axis, then $1 \notin \sigma(\mathcal{K})$ but $\alpha \in \sigma(\mathcal{K})$ for some $\alpha > 1$.

To see how much smaller $\eta(\mathcal{K})$ can be compared with $\rho(\mathcal{K})$ consider the following simple example. Let

$$\mathcal{L}u(t) = \rho \int_0^t e^{-(t-s)} u(s)\, \mathrm{d}s, \tag{4.17}$$

with $\rho > 0$. Then

$$\sigma(\mathcal{L}) = \{\lambda : |\lambda - \rho/2| \leq \rho/2\}. \tag{4.18}$$

Thus $\rho(\mathcal{L}) = \rho$ while $\eta(\mathcal{L}) = \min\{\frac{\rho}{|2-\rho|}, 1\}$. On the other hand, for the operator $-\mathcal{L}$ we obtain

$$\rho(-\mathcal{L}) = \rho \quad \text{and} \quad \eta(-\mathcal{L}) = \frac{\rho}{2+\rho}, \tag{4.19}$$

(Nevanlinna 1990*a*, page 155). In particular, if $\rho = 1 - \varepsilon$ with a small $\varepsilon > 0$ then $\eta(\mathcal{L}) \approx 1 - 2\varepsilon$, and this is only a 'modest' improvement, while $\eta(-\mathcal{L}) \approx \frac{1}{3}$, in which case we would speak about 'dramatic' improvement. More generally, if $\rho(\mathcal{K}) \in \sigma(\mathcal{K})$, with $\rho(\mathcal{K}) = 1 - \varepsilon$, then there cannot be any dramatic improvement for the following reason: the boundary of $\sigma(\mathcal{K})$ must be smooth near $\rho(\mathcal{K})$ (by Proposition 2 in Nevanlinna (1990*a*) and the Green's function $g(\lambda) \sim \mathcal{O}(\mathrm{dist}(\lambda, \sigma(\mathcal{K})))$ and $\eta(\mathcal{K}) = e^{-g(1)} = 1 - \mathcal{O}(\varepsilon)$. This should be contrasted with the situation for self-adjoint operators $\mathcal{A}$, for which the spectrum would be contained in an interval. Near the end point the corresponding Green's function would stretch the distance like the square root function and one would have $\eta(\mathcal{A}) = 1 - \mathcal{O}(\sqrt{\varepsilon})$, a well known effect of the conjugate gradient method. Finally, if $\rho(\mathcal{K}) \ll 1$, then it is possible to bound $\rho(\mathcal{K})$ in terms of $\eta(\mathcal{K})$. In fact, since $\sigma(\mathcal{K})$ is connected and contains both 0 and $\rho(\mathcal{K})e^{\mathrm{i}\theta}$, for some θ, one has

$$\rho(\mathcal{K}) \geq \mathrm{cap}(\sigma(\mathcal{K})) \geq \tfrac{1}{4}\rho(\mathcal{K}). \tag{4.20}$$

This allows us to formulate the following theorem.

Theorem 14 For every $\mathcal{K}$ we have as $\varepsilon \to 0$,

$$\eta(\varepsilon\mathcal{K}) \geq (\tfrac{1}{4} + o(1))\rho(\varepsilon\mathcal{K}). \tag{4.21}$$

Proof. If g_ε is the Green's function for the outside of $\sigma(\varepsilon\mathcal{K})$, then

$$\eta(\varepsilon\mathcal{K}) = e^{-g_\varepsilon(1)} = (1 + o(1))\varepsilon \mathrm{cap}(\sigma(\mathcal{K})), \quad \text{as} \quad \varepsilon \to 0. \tag{4.22}$$

□

To summarize: Krylov subspace acceleration is always possible, but dramatic improvement is obtained only if the distance between $\sigma(\mathcal{K})$ and 1 is essentially larger than $1 - \rho(\mathcal{K})$.

4.4. Time-dependent linear combinations

By a *subspace method* we mean any Krylov-subspace method that takes linear combinations of sweeps. To generalize this, we may think of processing the sweeps with some other operation. We outline here an approach of Lubich (1992). The basic special assumption here is that one decomposes $A = mI - (mI - A)$, so that the unaccelerated version would be

$$\dot{x}^{k+1} + mx^{k+1} = Nx^k + f, \quad x^{k+1}(0) = x_0. \tag{4.23}$$

Observe that multiplication with m commutes with N.

The accelerated version is as follows. Given x^k, solve

$$\dot{u}^k + mu^k = Nx^k + f, \quad u^k(0) = x_0, \tag{4.24}$$

set

$$v^k := u^k - x^k \tag{4.25}$$

and solve again for w^k from

$$\dot{w}^k + \lambda_k w^k = v^k, \quad w^k(0) = 0. \tag{4.26}$$

Finally, set

$$x^{k+1} := u^k + \alpha_k v^k + \beta_k w^k. \tag{4.27}$$

Note that all equations and substitutions (4.24)–(4.27) are on the component level, apart from the evaluation of Nx^k in (4.24) – in this sense the extra work is small compared with (4.23). The parameters α_k, β_k and λ_k can now be chosen so that the error reduction in $L_2(\mathbb{R}_+)$ is the same as that of Chebyshev acceleration of Richardson's iteration

$$x^{k+1} = x^k - \frac{1}{m}Ax^k + \frac{1}{m}b$$

for the static linear system $Ax = b$. To see that this is possible, compute the Laplace transform of the iteration error, and require this to be the Chebyshev acceleration of the Laplace transform of the iteration error of the basic scheme (4.23) for every z.

Related ideas are also discussed in Skeel (1989) and Reichelt, White and Allen (1995).

4.5. Overlapping splittings

If M in the splitting $A = M - N$ is chosen to be a block diagonal of A then the iteration (2.2) can clearly be computed in parallel for each small subsystem corresponding to one block of A. This is known as block Jacobi iteration. If the order of the original system was d and we use s subsystems (blocks) we only need to solve systems of order d/s in parallel. The reduction of work (and time) is so large that one might as well increase the size of subsystems

with a few components without losing this gain. The idea of overlapping was introduced by Jeltsch and Pohl (1995) in order to accelerate convergence of WR iteration. For block Jacobi iteration it can be best explained by an example.

Example 2 Let

$$A = \begin{pmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 2 \end{pmatrix}, \quad \mathbf{x} = (x_1\ x_2\ x_3\ x_4)^T$$

and we use two subsystems of the same size, so that

$$A = M - N = \begin{pmatrix} 2 & -1 & & \\ -1 & 2 & & \\ & & 2 & -1 \\ & & -1 & 2 \end{pmatrix} - \begin{pmatrix} 0 & 0 & & \\ 0 & 0 & 1 & \\ & 1 & 0 & 0 \\ & & 0 & 0 \end{pmatrix}.$$

Unknowns x_1, x_2 are solved from the first subsystem S_1 and x_3, x_4 from the second S_2. The idea of overlapping is that some components of the unknown vector are assigned to several subsystems, for instance x_3 in this example. Then, in (2.1), we obtain

$$\tilde{A} = \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & & -1 \\ & -1 & & 2 & -1 \\ & & & -1 & 2 \end{pmatrix}, \quad \tilde{\mathbf{x}} = (x_1\ x_2\ x_{3.1}\ x_{3.2}\ x_4)^T$$

and

$$\tilde{M} - \tilde{N} = \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & & \\ & & & 2 & -1 \\ & & & -1 & 2 \end{pmatrix} - \begin{pmatrix} 0 & 0 & 0 & & \\ 0 & 0 & 0 & & \\ 0 & 0 & 0 & 0 & 1 \\ & 1 & 0 & 0 & 0 \\ & & & 0 & 0 \end{pmatrix}.$$

We use the first system to find $\{x_1, x_2, x_{3.1}\}$ and the second to calculate $\{x_{3.2}, x_4\}$. The value used for x_3 in the next iteration is taken as the linear combination $x_3 = \alpha x_{3.1} + (1 - \alpha)x_{3.2}$

The number of overlapping components between subsystems was first one, then two, in this example. This number is called the *overlap* and we denote it by o.

In general it is reasonable to assume that if we have $s > 2$ subsystems then

(A1): The overlapping components are assigned to at most two common subsystems.

The overlap o can be defined as the maximum number of overlapping components in the intersections $S_j \cap S_k$.

Jeltsch and Pohl (1995) formulated overlapping splittings for block Jacobi iteration at the subsystem level, and showed that a convergence analysis similar to that of basic WR can be carried out. Their numerical results suggested that overlapping accelerates the convergence of WR. In order to explain when and why this happens we describe the process for the whole system. Let us assume that the splitting $M - N$ corresponds to block Jacobi iteration. Thus the components of $\mathbf{x}$ corresponding to each subsystem must be numbered consecutively, and M must be block diagonal.

When we have chosen the overlapping components of $\mathbf{x}$ we modify (2.1) as follows:

- If x_i is copied from subsystem k_1 to subsystem k_2, then rename it $x_{i.k_1}$, and add a new component $x_{i.k_2}$ to subsystem k_2.
- Duplicate the ith row of system (2.1), $\dot{x}_i + \sum_{j=1}^{d} A_{ij}x_j = f_i$, and add the duplicated row to the row corresponding to index $i.k_2$.

Hence each overlapping component increments the dimension of A by one by duplicating a row and adjoining a new column to the new duplicated component. Between the integration sweeps, each overlapped component x_j of $\mathbf{x}$ is postprocessed by replacing both copies with a linear combination of the overlapped components. The effect for the whole solution can be viewed as a multiplication by a constant matrix E. From the iteration's perspective, the iteration matrix $(zI + M)^{-1}N$ is replaced by $(zI + M)^{-1}NE$. Since the overlapped components of $\mathbf{x}$ are in the nullspace of N, the graphs of NE and $G(N)$ are identical.

We shall show that overlapping can accelerate convergence by decreasing the order ω of the iteration operator $\mathcal{K}$. The order can be computed from the directed graph of the matrix A as stated in Section 2.5.

4.6. How overlapping decreases the order

The graph $G(\tilde{A})$ is formed from $G(A)$ by making the following modifications to $G(A)$.

- Duplicate the vertices of $G(A)$ corresponding to the overlapped components.
- Duplicate the edges coming into the vertices corresponding to the overlapped components.

The latter statement demands some explanation. If a vertex v is duplicated from subsystem k_1 to subsystem k_2, then for the copy in k_1, draw all edges incident to v in $G(A)$ not linked to subsystem k_2. Similarly, for the copy of v in subsystem k_2, draw the edges incident to v not intersecting subsystem k_1.

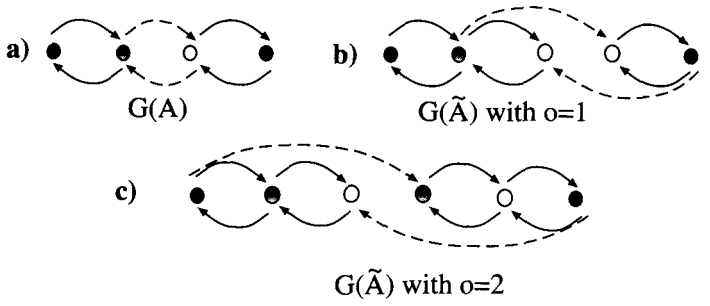

Fig. 2. Directed graphs of Example 2. Edges belonging to $G(N)$ are denoted by dashed lines. Duplicated vertices are recognized by shading.

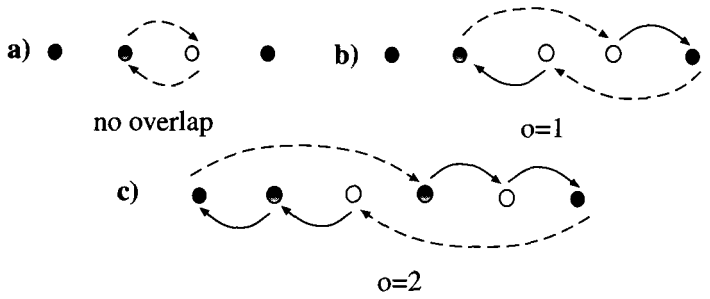

Fig. 3. Critical cycles of Example 2. Edges belonging to $G(N)$ are denoted by dashed lines.

All these edges were also in $G(A)$. The new edges are copies of the incoming edges of v.

Example 2 continued. The directed graphs of A and $\tilde{A}$ are given in Fig. 2. The cycles giving $\max(n_j/l(\mathcal{C}_j))$ in $G(M-N)$ or $G(\tilde{M}-\tilde{N})$ are given in Figure 3.

This example is summarized in Table 1, and overlapping decreases ω_{graph}, and hence the order of the iteration operator.

Table 1.

o	ω_{graph}
0	1
1	$2/4 = 1/2$
2	$2/6 = 1/3$

Fig. 4. Graph structure corresponding to a band matrix with $b = 2$. The edges with arrows in both heads are abbreviations of two edges connecting the vertices both ways.

In general, assuming condition (A1) given in Section 4.5, and that only adjoining subsystems (S_l and S_{l+1}) overlap, we can derive our next result.

Theorem 15 Using overlapping in block Jacobi iteration never increases the ratio ω_{graph}.

The proof is based on showing that the only new circuits created in $G(\tilde{A})$ when compared to the original graph $G(A)$ are such that the maximum ratio of $n_j/l(\mathcal{C}_j)$ is smaller than in $G(A)$. Simultaneously, for the old circuits remaining also in $G(\tilde{A})$, the ratio $n_j/l(\mathcal{C}_j)$ may decrease to 0 if the overlapping is such that the circuit stays inside one subsystem in the new graph $G(\tilde{A})$. The detailed proof is given in Miekkala (1996); the following result is a direct consequence.

Corollary 6 If ω_{graph} is determined only by the cycle $\mathcal{C}_1$, and $\mathcal{C}_1$ can be contained into one of the subsystems using overlapping, then ω_{graph} decreases.

This result tells us how overlapping should be used to accelerate convergence of the iteration. Indeed, the cycle (or cycles) attaining $\max(n_j/l(\mathcal{C}_j))$ in $G(M - N)$ should first be located, and then the subsystems overlapped in such a way that this cycle stays inside one of the enlarged subsystems.

The matrix A in Example 2 was the so-called Laplacian matrix, a band matrix. We will now show how overlapping decreases the order for general band matrices. The band width is denoted by $2b + 1$, where b is the smallest integer for which $A_{ij} = 0$ whenever $|i - j| > b$.

Once again we need only study overlapping between two consecutive subsystems, say S_1 and S_2. In graph theoretic language, $b = 1$ means that every pair of adjacent vertices is connected by a loop of length two; for $b = n$, every pair of vertices at mutual distance at most n is connected by a loop of length two. Figure 2 shows the case $b = 1$ and Figure 4 case $b = 2$; the general case should be obvious (if too messy to draw).

We have already analysed overlapping vertices for $b = 1$, in Example 2. The interface between the subsystems has o overlapping vertices; thus the coupling edge entering one subsystem from another has to skip o vertices. From Figures 2 and 3, we conclude that the cycle between the subsystems satisfies $l(\mathcal{C}_j) = o + o + 2 = 2o + 2$ and $n_j = 2$. Therefore $\omega_{\text{graph}} = (o + 1)^{-1}$.

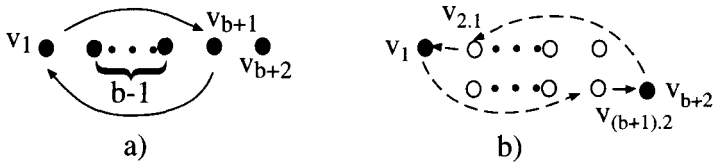

Fig. 5. a) Using overlap $o < b$ does not change the indicated loop between the subsystems. b) A critical circuit for $o = b$.

In the general case, the length of the critical circuit is $l(\mathcal{C}_j) = 2 + 2\lfloor o/b \rfloor$. Since the graph of a band matrix contains loops as in Figure 5a, it is clear that these loops remain in $G(\tilde{A})$ for $o \in \{1, \ldots, b-1\}$. Hence $\omega_{\text{graph}} = 1$ for these values of o. If $o = b$, then we duplicate b subsequent vertices as in Figure 5b and the critical circuit

$$v_1 \to v_{(b+1).2} \to v_{b+2} \to v_{2.1} \to v_1$$

has length 4. If overlap $o \in \{b+1, \ldots, 2b-1\}$, then we still get a circuit of length 4 and $n_j = 2$, that is

$$v_1 \to v_{(b+1).2} \to v_{o+2} \to v_{(o-b+2).1} \to v_1.$$

When $o = 2b$, the second edge of this circuit cannot occur because $o + 2 - (b+1) > b$ and the length of the critical circuit increases to 6. The general result should now be obvious.

Theorem 16 Let A be a band matrix of band width $2b+1$ and use block Jacobi iteration with overlap o in (2.2). Then

$$\omega \leq \frac{1}{\lfloor o/b \rfloor + 1}.$$

5. Discretized iterations

The results of the previous sections have analogues for discretized equations. We briefly discuss these analogues and then look at the new phenomena that arise when several grids are used during the calculation. Also, we mention some interesting step size control problems.

5.1. Discretization methods

The most natural approach to 'continuous time iteration' is simply to apply reliable software to integrate the associated equations. The process is sufficiently robust for results on the continuous version to describe what happens in practice, as long as the iteration errors are larger than the discretization errors. This robustness can be seen very well from an exact analysis of the

discretized equations. Here we consider linear multistep methods with a constant time step h:

$$\frac{1}{h}\sum_{j=0}^{k}\alpha_j x_{n+j}^{\nu}+\sum_{j=0}^{k}\beta_j M x_{n+j}^{\nu}=\sum_{j=0}^{k}\beta_j(N x_{n+j}^{\nu-1}+f_{n+j}). \tag{5.1}$$

As is customary, we use operator notation for the linear multistep methods. In order to avoid confusion with the spectral radius and the spectrum, we set

$$a(\zeta):=\sum_{j=0}^{k}\alpha_j\zeta^j,\quad b(\zeta):=\sum_{j=0}^{k}\beta_j\zeta^j. \tag{5.2}$$

We normalize $b(1)=1$, require that the order of consistency satisfies $p\geq 1$, and assume that $a(\zeta)$ and $b(\zeta)$ have no common factors. We abbreviate $\{\sum_{j=0}^{k}\alpha_j v_{n+j}\}$ to av.

In this notation, (5.1) reads

$$\frac{1}{h}ax^{\nu}+bMx^{\nu}=bNx^{\nu-1}+bf. \tag{5.3}$$

As in the continuous case it is advantageous to introduce a linear operator $\mathcal{K}_h$ and write the solution of the difference equation (5.3) in the form

$$x^{\nu}=\mathcal{K}_h x^{\nu-1}+\varphi_h. \tag{5.4}$$

Here $\mathcal{K}_h$ is well defined provided we understand the sequences to vanish for negative indices and

$$\frac{\alpha_k}{\beta_k}\notin\sigma(-hM). \tag{5.5}$$

In what follows we shall always assume that (5.5) holds. The role of the Laplace transform is played by the 'ζ-transform'.

If X_h denotes $\mathbb{C}^d$-valued sequences, then we write

$$\tilde{v}(\zeta):=\sum_{n=0}^{\infty}\zeta^{-k}v_k,\qquad v\in X_h,$$

and this leads to the following expression for the *symbol* of $\mathcal{K}_h$:

$$K_h(\zeta):=(\frac{1}{h}a(\zeta)+b(\zeta)M)^{-1}b(\zeta)N. \tag{5.6}$$

In particular, $K_h(\zeta)=K(a(\zeta)/hb(\zeta))$. It is also useful to write $v_n=v(nh)$ for $v\in X_h$. We shall need standard terminology to describe the stability properties of the method (a,b).

Definition 4 The *stability region* S consists of those $\mu\in\mathbb{C}\cup\{\infty\}$ for which the polynomial $a(\zeta)-\mu b(\zeta)$ (around ∞ consider $\mu^{-1}a-b$) satisfies the root condition. The method is called *strongly stable* if all roots of $a(\zeta)/(\zeta-1)$ are

less than one in modulus. The method is called *A-stable* if S contains the closed left half plane.

5.2. Finite windows

Consider bounding $\mathcal{K}_h^\nu$ on the 'window' $0 \le j \le T/2$. We identify the vectors $v \in X_h$ by sequences indexed over $\mathbb{Z}$ with v vanishing for negative indices and

$$|v|_T = \max_{0 \le h_j \le T} |v_j|. \tag{5.7}$$

Theorem 17 If $\alpha_k/h\beta_k \notin \sigma(-M)$, then, in the window $0 \le j \le T/h$ $\mathcal{K}_h$ has the spectral radius

$$\rho(\mathcal{K}_h) = \rho(K(\alpha_k/h\beta_k)). \tag{5.8}$$

This is Theorem 4.1 in Nevanlinna (1989*c*). Comparing this with (5.6), observe that $z \to \infty$ corresponds to $\zeta \to \infty$ and $\lim_{\zeta \to \infty} a(\zeta)/hb(\zeta) = \alpha_k/h\beta_k$. It should be noticed that, unlike the infinite window case, we do *not* obtain a result of the form $\rho(\mathcal{K}_h) = \rho(\mathcal{K}) + \mathcal{O}(h^p)$ with p related to the accuracy of the discretization method. However, the following holds.

Corollary 7 Under the assumptions of Theorem 17 we have

$$\rho(\mathcal{K}_h) = (1 + o(1))(\frac{\beta_k}{\alpha_k} ch)^{1/\omega}, \quad \text{as } h \to 0, \tag{5.9}$$

if $\mathcal{K}$ is of order ω and of type $\tau = cT$.

Proof. This follows immediately from the defining relation of ω and c (see Section 2.1, line (2.37)) by choosing $z = z(h) = \alpha_k/h\beta_k$. □

As this corollary shows, the decay of $|\mathcal{K}_h^k|_T$ as $k \to \infty$ is again related to the order and type of the original resolvent operator. Upper bounds again hold, analogous to those in Section 2.4 for $|\mathcal{K}^k|_T$, but here we just refer to the original paper by Nevanlinna (1989*c*).

5.3. Infinite windows

We now look at the usual ℓ_2-space of square summable $\mathbb{C}^d$-valued sequences (but again the spectral radius of $\mathcal{K}_h$ would be the same for a large class of 'unscaled' norms). Now the local solvability condition $\alpha_k/h\beta_k \notin \sigma(-M)$ is still needed for $\mathcal{K}_h$ to be well defined, but another condition is needed to guarantee that $\mathcal{K}_h$ is bounded in ℓ_2. In the continuous case this was simply the condition $\sigma(-M) \subset \mathbb{C}_-$. Now the role of $\mathbb{C}_-$ is played by the stability region below.

Proposition 5 If $\sigma(-M) \subset h^{-1}\text{int}S$, then $\mathcal{K}$ is bounded in ℓ_2.

Proof. This is Lemma 3.1 in Miekkala and Nevanlinna (1987*b*). □

The correspondence of $\mathbb{C}_+$ with $h^{-1}(\mathbb{C} \setminus S)$, best seen in the fact $K_h(\zeta) = K(a(\zeta)/hb(\zeta))$, immediately explains this Proposition and related claims. For example, the boundaries of these sets osculate at the origin, displaying the *order of accuracy.* However, in order to obtain the exact formulation we need the following concept.

Definition 5 A multistep method (a, b) has *order of amplitude fitting* q if the principal root $\zeta_1(\mu)$ of $a(\zeta) - \mu b(\zeta) = 0$ (the zero for which $\zeta_1(\mu) - e^{\mu} = \mathcal{O}(\mu^{p+1})$) satisfies $|\zeta_1(\,it)| - 1 = \mathcal{O}(t^{q+1})$, for small real t.

Thus $q \geq p$ if p is the usual discretization order, and, with the trapezoidal rule for instance, we have $q = \infty$, while $p = 2$.

Theorem 18 Assume that the multistep method is of amplitude fitting order q and is strongly stable. Then, for all sufficiently small h, $\mathcal{K}_h$ is a bounded operator,

$$\|\mathcal{K}_h\| = \|\mathcal{K}\|[1 + \mathcal{O}(h^q)] \tag{5.10}$$

and

$$\rho(\mathcal{K}_h) = \rho(\mathcal{K})[1 + \mathcal{O}(h^q)]. \tag{5.11}$$

Theorem 19 Assume that the multistep method is A-stable. Then, for all $j \geq 1$,

$$\|\mathcal{K}_h^j\| \leq \|\mathcal{K}^j\| \tag{5.12}$$

and

$$\rho(\mathcal{K}_h) \leq \rho(\mathcal{K}). \tag{5.13}$$

Both Theorem (18) and (19) are proved in Nevanlinna (1990*b*), using results of Miekkala and Nevanlinna (1987*b*).

For Krylov acceleration it is interesting to know something of the spectrum.

Theorem 20 (Miekkala and Nevanlinna 1987*b*)

$$\sigma(\mathcal{K}_h) = \mathrm{cl} \bigcup_{|\zeta| \geq 1} \sigma(K(a(\zeta)/hb(\zeta))).$$

Corollary 8 For A-stable methods we have $\sigma(\mathcal{K}_h) \subset \sigma(\mathcal{K})$.

Proof. By A-stability, $\bigcup_{|\zeta| \geq 1} \{a(\zeta)/hb(\zeta)\}$ is a subset of the closed right half plane. □

Corollary 9 $\sigma(\mathcal{K}_h)$ consists of at most d components, each containing eigenvalues of $K(\alpha_k/h\beta_k)$.

Proof. By letting $\zeta \to \infty$, we see that the eigenvalues of $K(\alpha_k/h\beta_k)$ belong to $\sigma(\mathcal{K}_h)$. Each eigenvalue, or, rather, branch of the algebraic function, can be continued to $|\zeta| \geq 1$, giving at most d components. □

For example, if we take the implicit Euler method and relatively large step size, then $\sigma(\mathcal{K})$ and $\sigma(\mathcal{K}_h)$, and in particular the corresponding optimal reduction factors, may differ considerably. Lumsdaine and White (1995) give an example of this nature.

5.4. Multigrid in time

The effective use of 'multigrids' in our setup simply consists of balancing the iteration error and discretization error. Thus one moves only towards ever finer grids. The computational goal is to be able to compute or simulate the full system with an amount of work W which is a modest multiple of the work W_0, say, needed to compute the 'uncoupled' system

$$\dot{u} + Mu = Nx + f(t), \quad u(0) = x_0, \tag{5.14}$$

to the same tolerance, where x denotes the solution of

$$\dot{x} + Mx = Nx + f(t), \quad x(0) = x_0. \tag{5.15}$$

The ideas in setting up such a computational strategy, or 'tolerance game', have been discussed in Nevanlinna (1989*c*) and Nevanlinna (1990*b*). We shall not go into such a discussion here but rather concentrate on two issues that might cause difficulties if the implementation is careless. It may not be evident that it is possible to arrange for $W = \mathcal{O}(W_0)$ to hold. Two extremes are possible. First, the step size selection routine is extremely stupid and the step is constant on the grid. Second, the step size selection process is extremely clever and the step changes with the smoothness of the solution, so rapidly reducing the step size when the solution is rough, but increases the step size stably when the solution becomes smooth. The potential danger to be avoided is this: in solving stiff problems, it is to be expected that the solution at the end of the window is smooth. However, on the next window, say $[mT, (m+1)T]$, the initial guess $x^0(t) \equiv x(mT)$ introduces an error which causes, in exact computation, a travelling error wave, which, however, has very small support. Roughly speaking, with fixed step strategy the error wave cannot be supported at all, while integration with automatic software has to be done with a good step size routine so that not too many time points are wasted at the thinly supported rough parts. Here we discuss the constant step case and in the next section the latter one.

For simplicity, let the grid at the ν_{th} iteration level be $\{jh_\nu\}_j$ where the time step $h_\nu = 2^{-n(\nu)}h_0$ and $n(\nu)$ is nondecreasing and unbounded. A detailed analysis of this is given in Section 3.2. of Nevanlinna (1990*b*). In the iteration process, whenever the grid is refined ($n(\nu) > n(\nu-1)$), we need to be able to *extend* a grid function $v_{\nu-1} = \{v(jh_{\nu-1})\}$ to a grid function on $v_\nu = \{v(jh_\nu)\}$. Thus we have the *prolongation operator*

$$\mathcal{P}_\nu : \quad v_{\nu-1} \to v_\nu,$$

computed with accuracy matching the integration method, but the important point is that there is also a stability property to be satisfied. In fact, we want the overall process to decay with a rate essentially equalling $\rho(\mathcal{K})$, for all refinement sequences $\{n(\nu)\}$, and this is possible if the prolongation operators $\{\mathcal{P}_\alpha\}$ are stable: there exists a C such that

$$\|\Pi_1^n \mathcal{P}_{\alpha(j)}\| \le C, \quad \mathcal{P}_{\alpha(j)} \in \{\mathcal{P}_\alpha\}.$$

The norms here are the naturally induced operator norms; grid functions $v_\nu = \{v(jh_\nu)\}$ are normed as follows:

$$\|v_\nu\| := \{h_\nu \sum_j |v(jh_\nu)|^2\}^{1/2}.$$

It turns out that there are arbitrary high-order stable prolongations but that the information should in general be collected from both sides of the grid points. A symbolic calculus for stepwise translation invariant prolongations was developed in Nevanlinna (1990*b*). The crucial dilation process here is quite similar to the subdivision algorithm in CAD or in wavelets and this eventually led Eirola (1992) to study the obtainable smoothness of wavelets.

For the error analysis the main result is the following theorem.

Let

$$B_{\mu\mu} := \quad \text{identity on grid functions on} \quad \{jh_\mu\}$$

$$B_{\nu\mu} := \mathcal{K}_{h_\nu} \mathcal{P}_\nu B_{\nu-1,\mu}, \quad \nu \ge \mu + 1.$$

Theorem 21 (Nevanlinna 1990*b*) Assume that the multistep method is strongly stable and we are given a stable set of prolongations. Let h_0 be small enough so that

$$\sigma(-M) \subset \frac{1}{h} \text{int } S$$

holds for all $h \le h_0$. Given $\varepsilon > 0$ and $h_\nu = 2^{-n(\nu)} h_0$ with $n(\nu)$ nondecreasing and unbounded, there exists a C such that

$$\|B_{\nu\mu}\| \le C(\rho(\mathcal{K}) + \varepsilon)^{\nu-\mu}, \quad \nu \ge \mu \ge 0.$$

This is the key result needed to show that the 'tolerance game' is possible.

5.5. A difficulty due to stiffness

Consider solving

$$\dot{x} + Ax = f \tag{5.16}$$

in a window where the solution is already smooth, that is, the transient has died out in the earlier window.

Usually one takes the initial function to be identically the initial value and it is no longer clear whether the iterates will stay smooth. A naive application of

Picard–Lindelöf iteration might then spend a lot of time in the early iterations, because any local error estimator would require tiny steps compared with the smoothness of the limit function.

We present here a model analysis of the smoothness of the iterates x^j, following Nevanlinna (1989*a*). We assume that the iterates x^j are computed exactly from

$$\begin{aligned} \dot{x}^{j+1} + Mx^{j+1} &= Nx^j + f, \\ x^{j+1}(0) &= x_0 \equiv x^0, \end{aligned} \tag{5.17}$$

but we measure the 'cost of integration' as if we were using high quality software, based on first-order local error estimation: at time t the time step $h = h(t)$ would satisfy

$$h(t)|\ddot{x}^j(t)| = \epsilon_j. \tag{5.18}$$

This corresponds to the criterion of error per unit step; calculation for criterion of error per step is analogous.

Thus the relevant measure for the cost or for the total number of time points is proportional to $\sum \frac{1}{\epsilon_j} \int |\ddot{x}^j|$. An efficient implementation of Picard–Lindelöf iteration would gradually decrease the tolerance ϵ_j. Here we shall not discuss the choice of ϵ but focus on estimating $\int |\ddot{x}^j|$.

We put $T = 1$ and assume that x is so smooth that it can be represented as a convergent power series

$$x(t) = \sum t^i x_i. \tag{5.19}$$

Since we are interested in the second derivatives, we measure smoothness on the window [0,1] by

$$\|x\| := |x_0| + |x_1| + \sum_{i=2}^{\infty} i(i-1)|x_i|. \tag{5.20}$$

If $e^j := x - x^j$ denotes the iteration error, then

$$\dot{e}^{j+1} + Me^{j+1} = Ne^j, \ e^j(0) = 0. \tag{5.21}$$

Introducing

$$k(t) := e^{-tM}N,$$

and setting

$$k^{*j} = k * k^{*(j-1)},$$

we have

$$e^j = k^{*j} * (x - x_0). \tag{5.22}$$

Substituting (5.19) into (5.22) yields

$$e^j(t) = \sum_{i=1}^{\infty} \int_0^t (t-s)^i k^{*j}(s) x_i ds, \tag{5.23}$$

and hence

$$\ddot{e}^j(t) = k^{*j}(t)x_1 + \sum_{i=2}^{\infty} i(i-1) \int_0^t (t-s)^{i-2} k^{*j}(s) x_i ds. \tag{5.24}$$

In order to estimate this we introduce the following bound:

$$C := \sup_j \sup_{|a|=1} \int_0^1 |k^{*j}(s)a| ds. \tag{5.25}$$

Theorem 22 We have

$$\int_0^1 |\ddot{x} - \ddot{x}^j| \leq C\|x - x_0\|. \tag{5.26}$$

Furthermore, for any given splitting M, N there exists $x_0 \neq 0$ and f such that $x(t) = (1+t)x_0$, $\ddot{x} \equiv 0$ and for some j,

$$\int_0^1 |\ddot{x}^j| = C\|x - x_0\|. \tag{5.27}$$

Proof. The definition of C immediately gives (5.26). Since $\int |k^{*j}|$ tends to zero and a in (5.25) runs over a compact set, there exist an integer j and a unit vector x_0 such that

$$C = \int_0^1 |k^{*j}(s)x_0| \, ds.$$

If $f(t) = x_0 + (1+t)Ax_0$ then $x(t) = (1+t)x_0$ is the solution of (5.16). Since $x_1 = x_0$, and $\ddot{e}^j = -\ddot{x}^j$, we obtain

$$\int_0^1 |\ddot{x}^j| = C = C\|x - x_0\|.$$

from (5.24). □

We conclude from Theorem 22 that the important quantity $\int |\ddot{x}^j|$ stays small for smooth limit functions x if and only if C is of moderate size.

It is important to note that, even if C is of moderate size, the smallest 'steps' can be very small, since we may have $\sup_{[0,1]} |\ddot{x}^j| \gg C$, while $\int_0^1 |\ddot{x}^j| \leq C$. Nevanlinna (1989*a*) contains an example of this.

Recall from Section 2 (proof of Theorem 5) that if

$$|K(z)| \leq \frac{B}{\operatorname{Re} z - \gamma} \quad \operatorname{Re} z > \gamma,$$

then we obtain a bound for the norms of the iterated kernels and hence the

upper bound:

$$C \leq e^{\gamma} d \max_j (\frac{Be}{j})^j.$$

6. Periodic problems

We discuss briefly the iterative solution for periodic problems. We shall see that the speed of the basic iteration is related to the speed of the 'corresponding' initial value problem in the *infinite* window, while the speed after optimal Krylov acceleration is related to the speed obtained for initial value problems on the *finite* window.

6.1. The problem and the iteration operator

Consider solving the periodic boundary value problem

$$\dot{x} + Ax = f, \quad x(0) = x(T), \tag{6.1}$$

when f is a continuous function of period T. Splitting $A = M - N$ as usual leads to an iteration of the form

$$x^k = \mathcal{F} x^{k-1} + g \tag{6.2}$$

provided the *solvability condition* holds:

$$\frac{2\pi\, in}{T} \notin \sigma(-M) \quad \text{for all } n \in \mathbb{Z}. \tag{6.3}$$

Here the integral operator $\mathcal{F}$ can be written in convolution form

$$\mathcal{F}x(t) = \int_0^T \varphi(t-s)x(s)\,\mathrm{d}s, \tag{6.4}$$

where the kernel φ is periodic and

$$\varphi|_{[0,T)}(t) = e^{-tM}(1 - e^{-TM})^{-1}N.$$

For more details, see Vandewalle (1992).

6.2. Spectrum and consequences

Computation of the Fourier coefficients of φ gives $\hat{\varphi}(n) = K(\frac{2\pi\, in}{T})$, where $K(z) = (z + M)^{-1}N$ is the symbol of the Volterra operator $\mathcal{K}$. This leads to the following result.

Theorem 23 (Vandewalle 1992) Let the solvability condition (6.3) hold. Then $\mathcal{F}$ is a compact operator in $C[0,T]$ with the spectrum

$$\sigma(\mathcal{F}) = \mathrm{cl} \bigcup_{n \in \mathbb{Z}} \sigma(K(\frac{2\pi\, in}{T})), \tag{6.5}$$

and spectral radius

$$\rho(\mathcal{F}) = \max_{n\in\mathbb{Z}} \rho(K(\frac{2\pi\, in}{T})). \tag{6.6}$$

Corollary 10 Assume that eigenvalues of M have positive real parts. Then

$$\rho(\mathcal{F}) \le \rho(\mathcal{K}) \le \rho(\mathcal{F}) + \mathcal{O}\left(\frac{1}{T^2}\right). \tag{6.7}$$

Proof. Here $\mathcal{F}$ is considered in $C[0,T]$ while $\mathcal{K}$ is considered on the infinite window (any space X of Section 3). The first inequality is immediately verified because

$$\rho(\mathcal{F}) = \max_n \rho(K(\frac{2\pi\, in}{T})) \le \max_\theta \rho(K(\,\mathrm{i}\theta)) = \rho(\mathcal{K}).$$

The second inequality follows from the fact that the boundary $\partial\sigma(\mathcal{K})$ is locally analytic at points $\lambda \in \sigma(\mathcal{K})$ where $|\lambda| = \rho(\mathcal{K})$; see Proposition 2 in Nevanlinna (1990*a*). Sampling this with density $\mathcal{O}(T^{-1})$ provides the maximum within tolerance $\mathcal{O}(T^{-2})$. □

Observe in particular that even if the solvability condition holds (and in particular it holds for all T if M has eigenvalues of positive real parts) we can have $\rho(\mathcal{F}) > 1$, so that the iteration would diverge. However, Krylov acceleration would always work.

Corollary 11 Let the local solvability condition (6.3) hold and suppose $1 \notin \sigma(\mathcal{F})$. Then the optimal reduction factor vanishes: $\eta(\mathcal{F}) = 0$.

Proof. Since $\sigma(\mathcal{F})$ is countable, it is of zero capacity, and the claim follows from Theorem 11 in Section 4.1. □

Since $\eta(\mathcal{F}) = 0$, we consider the superlinear decay of $b_n(\mathcal{F})$. Again the answer can be related to the corresponding initial value problem. Namely, Piirilä (1993) has shown that the order of decay for $b_n(\mathcal{F})$ equals that of $|\mathcal{K}^n|_T$, that is, if $R(\lambda,\mathcal{K})$ is of order ω, then

$$\limsup \frac{n\log n}{\log(1/b_n(\mathcal{F}))} = \omega.$$

7. A case study: linear RC-circuits

Linear RC circuits can be modelled in several ways. The sparse tableau model contains all the equations governing a circuit and results in a large DAE system. Nodal formulation results in a substantially smaller system and then equations are written for nodal voltages with the aid of so-called *stamps*. Using nodal formulation it was already shown in the earliest waveform relaxation paper of Lelarasmee et al. (1982) that waveform relaxation converges if the cutting is done only across such capacitors that there is a path connecting

them to the ground involving only capacitors. A simple model problem where splitting is done across a capacitor not obeying this rule was considered by Miekkala, Nevanlinna and Ruehli (1990), showing that in this case waveform relaxation still converges, but convergence is sublinear. This result can be generalized to all splittings of linear RC circuits. waveform relaxation always converges, but convergence may be slow, like $\mathcal{O}(k^{-r})$, where k is the iteration index and r a small number. This result was proved by Nevanlinna (1991).

We consider here as a case study applying waveform relaxation for the sparse tableau formulation of linear RC circuits, following closely the treatment of Leimkuhler, Miekkala and Nevanlinna (1991). The system is a DAE of index one or two. We describe a splitting strategy that allows us to break the circuit into subcircuits only across resistors. This strategy leads to convergence that is shown using mainly Laplace transforms.

7.1. RC-circuit equations

The system of equations for an RC-circuit is

$$\begin{pmatrix} C\dot{v}_C \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & -I & 0 \\ 0 & -R & 0 & 0 & A_R \\ 0 & 0 & 0 & 0 & A_E \\ -I & 0 & 0 & 0 & A_C \\ 0 & A_R^T & A_E^T & A_C^T & 0 \end{pmatrix} \begin{pmatrix} v_C \\ i_R \\ i_E \\ i_C \\ v_N \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ E(t) \\ 0 \\ 0 \end{pmatrix}. \quad (7.1)$$

The unknown vector contains voltages across capacitors (v_C), currents through resistors (i_R), voltage sources (i_E) and capacitors (i_C), and nodal voltages (v_N). The matrices in (7.1) satisfy

$$\begin{aligned} R &: \text{a positive, diagonal } n_R \times n_R \text{ matrix;} \\ C &: \text{a positive, diagonal } n_C \times n_C \text{ matrix;} \\ A_R &: \text{an } n_R \times N \text{ incidence matrix;} \\ A_E &: \text{an } n_E \times N \text{ incidence matrix;} \\ A_C &: \text{an } n_C \times N \text{ incidence matrix.} \end{aligned}$$

Here an *incidence matrix* is a matrix whose elements belong to the set $\{-1, 0, 1\}$ and whose rows contain either two nonzeros $\{1, -1\}$ or one nonzero. The usual definition does not allow the latter case, which arises because we have eliminated the ground node from the circuit (which is a directed graph). So N is the number of nodes in the circuit after a reference (ground) node has been fixed and n_R, n_C and n_E are the number of resistors, capacitors and voltage sources in the circuit, respectively. The appropriate sizes of the variable vectors should be apparent.

The problem is well posed if

$$\begin{aligned} n_R + n_E + n_C &\geq N + 1 \quad \text{and} \\ A := \begin{pmatrix} A_R \\ A_E \\ A_C \end{pmatrix} &\quad \text{has full rank.} \end{aligned} \tag{7.2}$$

Another basic assumption is that

$$A_E \quad \text{has linearly independent rows.} \tag{7.3}$$

This only means that there cannot be two voltage sources in parallel.

Proposition 6 Assume that assumptions (7.2) and (7.3) hold. Then the DAE (7.1) has index one if

$$\begin{pmatrix} A_E \\ A_C \end{pmatrix} \quad \text{has linearly independent rows.}$$

Otherwise it has index two.

Index two occurs if there are loops containing only capacitors and voltage sources. For the proof of the proposition and a discussion of the assumptions (7.2) and (7.3) see Manke et al. (1979).

We discuss the initial conditions for (7.1) after applying the Laplace transform to (7.1). The Laplace transform $\hat{x}$ of x is given by $\hat{x}(z) = \int_0^\infty e^{-zt}x(t)dt$. The transformed system becomes

$$\begin{aligned} zC\hat{v}_C &= \hat{i}_C + Cv_C(0), \\ R\hat{i}_R &= A_R\hat{v}_N, \\ A_E\hat{v}_N &= \hat{E}, \\ \hat{v}_C &= A_C\hat{v}_N, \\ \text{and } A_R^T\hat{i}_R + A_E^T\hat{i}_E + A_C^T\hat{i}_C &= 0. \end{aligned} \tag{7.4}$$

Eliminating $\hat{i}_R$, $\hat{v}_C$ and $\hat{i}_C$ gives

$$\begin{pmatrix} 0 & A_E \\ A_E^T & A_R^T R^{-1} A_R + z A_C^T C A_C \end{pmatrix} \begin{pmatrix} \hat{i}_E \\ \hat{v}_N \end{pmatrix} = \begin{pmatrix} \hat{E} \\ A_C^T C A_C v_N(0) \end{pmatrix}, \tag{7.5}$$

where we have used $v_C(0) = A_C v_N(0)$. Equation (7.5) can be solved for $\begin{pmatrix} \hat{i}_E \\ \hat{v}_N \end{pmatrix}$ if the coefficient matrix is nonsingular. This can be shown by an indirect proof (Leimkuhler et al. 1991) for $\operatorname{Re} z \geq 0$ and $z \neq 0$. If $z = 0$, (7.5) can still be solved if

$$\begin{pmatrix} A_R \\ A_E \end{pmatrix} \quad \text{has linearly independent columns.} \tag{7.6}$$

When solving the Laplace-transformed system (7.4), we find that if the transform of the input function $\hat{E}$ stays bounded as z grows, then $\hat{v}_C$, $\hat{i}_R$, $\hat{i}_E$, $\hat{i}_C$ and $\hat{v}_N$ are also bounded, if we have an index one DAE system. However, if index two occurs, then some components of $\hat{i}_C$ and $\hat{i}_E$ may grow linearly with z even when $\hat{E}$ is bounded (Leimkuhler et al. 1991).

Now let us discuss the initial values for (7.1). The form of the equation suggests that one can assign arbitrary initial values to the state variables v_{C_i}. However, if we study the Laplace transform of (7.1) we see from (7.5) that one may as well assume arbitrary initial values for all nodal voltages v_{N_i}. Not all of them will have any effect on the solution but only $A_C v_N(0)$, that is, those $v_{N_i}(0)$ corresponding to nodes adjacent to capacitors. Notice that although the solution for $\hat{v}_N$ is continuous for any initial values $v_N(0)$ there will, in general, be a discontinuity in the time domain solution because at $t > 0$ the algebraic equations in (7.1) determine $v_N(t)$, which may jump from the arbitrary $v_N(0)$. If one wants to avoid this discontinuity at the initial point, one should at least choose $v_N(0)$ consistent with $E(0)$ by the third equation of (7.1):

$$A_E v_N(0) = E(0).$$

Since A_E has independent rows by (7.3), the number of independent initial values that are used to obtain the solution of (7.5) is

$$\operatorname{rank} \begin{pmatrix} A_E \\ A_C \end{pmatrix} - n_E.$$

By Proposition 6, this equals n_C for index one and, for index two, it is at most n_C.

The results motivate us to assume all the bounded components are in an α-weighted L_2 space, but the 'index two' variables lie in a larger space, say Y_α. As shown above, the index two variables consist of some components of i_E and i_C. Since it is difficult to identify these particular components, we assume all components of i_E and i_C are elements of Y_α. The α-norm is defined by

$$|x|_\alpha^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\hat{x}(\alpha + \mathrm{i}\xi)|^2 d\xi, \quad \alpha > 0,$$

and the Y_α-norm by

$$|y|_{Y_\alpha}^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\hat{y}(\alpha + \mathrm{i}\xi)|^2 \left(1 + \alpha^2 + |\xi|^2\right)^{-1} \mathrm{d}\xi,$$

corresponding to the loss of one derivative. We may now take the space X_α to consist of elements $x^T = (v_C^T \;\; i_R^T \;\; i_E^T \;\; i_C^T \;\; v_N^T)$ where v_C, i_R and $v_N \in L_2^\alpha$ and i_E and $i_C \in Y_\alpha$. The norm in X_α is defined by

$$|x|_\alpha^2 = |v_C|_\alpha^2 + |i_R|_\alpha^2 + |i_E|_{Y_\alpha}^2 + |i_C|_{Y_\alpha}^2 + |v_N|_\alpha^2.$$

Remark 1 In the index one case we can simply take $X_\alpha = L_2^\alpha$ for all components of x, since the i_E and i_C have a special behaviour only in the index two case. Then one should replace $|\cdot|_{Y_\alpha}$ in the preceding norm definition by $|\cdot|_\alpha$.

From the input function $E(t)$ we assume

$$E \in L_2^\alpha. \tag{7.7}$$

Theorem 24 Assume (7.2), (7.3) and (7.7). Then (7.1) has a unique solution in X_α for all $\alpha > 0$.

Remark 2 In the classical treatment of DAEs, smoothness for the high index variables is guaranteed by requiring the input function ($E(t)$ in our application) to have as many derivatives as needed.

7.2. Splittings

For large-scale circuits it is sometimes natural to write (7.1) in the permuted form where the RC-circuit equations are repeated for each subcircuit. One tries to choose the subcircuits in such a way that there are as few connections, or couplings, to other subcircuits as possible. The resulting permuted form of (7.1) will have a block structure

$$\begin{pmatrix} \square & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \square \end{pmatrix} \frac{d}{dt} \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} + \begin{pmatrix} \square & \star & \star \\ \star & \ddots & \star \\ \star & \star & \square \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} = \begin{pmatrix} f_1 \\ \vdots \\ f_k \end{pmatrix}, \tag{7.8}$$

where the coefficient matrix of $\dot{x}$ is still diagonal ≥ 0 and the coefficient matrix of x has nonzero elements mainly on its diagonal blocks, but also elsewhere because of the couplings. We will show that if the subcircuits are chosen in such a way that the *subcircuits are coupled solely through resistors*, then the waveform relaxation method converges linearly. One should duplicate each interface branch equation and assign the involved resistor current variable to both subsystems connected through this branch. Applying dynamic block Jacobi iteration to (7.8) after these modifications will always converge, as we will show in the next section.

As mentioned above, our rule is that when splitting (7.1) we only cut through resistors. So the equations that are possibly affected by this relaxation are those that contain i_R:

$$\begin{aligned} -Ri_R + A_R v_N &= 0 \\ \text{and} \quad A_R^T i_R + A_E^T i_E + A_C^T i_C &= 0. \end{aligned}$$

For one specific resistor r_i the first one is

$$-r_i i_{r_i} + v_{k_i} - v_{l_i} = 0, \tag{7.9}$$

where r_i is between the nodes k_i and l_i. If we cut through r_i it is not obvious to which subcircuit the above equation and variable i_{r_i} should be assigned. In order to preserve symmetry in the flow of information between subcircuits, it seems that both subcircuits should 'see' r_i in the same way; this means (7.9) should be assigned to both of them. To do that we have to duplicate equation (7.9) and also the variable i_{r_i}, in the sense that we associate equation (7.9) with $i_{r_i}^+$ to the first subcircuit and with $i_{r_i}^-$ to the second:

$$-r_i i_{r_i}^+ + v_{k_i} - v_{l_i} = 0 \qquad \text{and} \qquad -r_i i_{r_i}^- + v_{k_i} - v_{l_i} = 0 \tag{7.10}$$

The relaxation is now defined by the iteration we apply to all pairs of equations involving 'cut resistors' as (7.10):

$$-r_i i_{r_i}^{+\,k} + v_{k_i}^k - v_{l_i}^{k-1} = 0 \qquad \text{and} \qquad -r_i i_{r_i}^{-\,k} + v_{k_i}^{k-1} - v_{l_i}^k = 0. \tag{7.11}$$

All components of the unknown x other than those v_i occurring in the 'cut equations' (7.11) are treated at the new iteration index; thus the mentioned v_i are the only coupling terms.

There is of course no duplication of the KCL equations: the number of nodes does not change. The only change is that in the KCL equations corresponding to the nodes k_i and l_i (refer to (7.11)) we must use $i_{r_i}^+$ and $i_{r_i}^-$, respectively.

Next we want to describe the splitting process in equation form.

Let L_R be the set of indices of those resistors that are cut in the relaxation process. Then the resistor current variable i_R is modified so that each i_{r_i}, $i \in L_R$, is replaced by the pair of variables $i_{r_i}^+$ and $i_{r_i}^-$:

$$i_R = (i_{r_1} \dots i_{r_i} \dots i_{r_{nR}})^T \mapsto \bar{i}_R = (i_{r_1} \dots i_{r_i}^+ i_{r_i}^- \dots i_{r_{nR}})^T \text{ for all } i \in L_R.$$

Also, those rows $(A_R)_{i\cdot}$ of the incidence matrix A_R for which i is a member of L_R are duplicated, and the resulting new matrix $\bar{A}_R$ is split in the way suggested by (7.11)

$$\bar{A}_R = A_M - A_N,$$

where A_N has nonzero elements only on the pairs of rows corresponding to the duplicated equations. On those rows the splitting is

$$\bar{A}_R = \begin{pmatrix} \dots & \dots \\ 1 & -1 \\ 1 & -1 \\ \dots & \dots \end{pmatrix} = \begin{pmatrix} \dots & \dots \\ 1 & 0 \\ 0 & -1 \\ \dots & \dots \end{pmatrix} - \begin{pmatrix} \dots & \dots \\ 0 & 1 \\ -1 & 0 \\ \dots & \dots \end{pmatrix}$$
$$=: A_M - A_N.$$

If we modify the diagonal matrix of resistors R in the same way as i_R, that is

$$R = \operatorname{diag}(r_1, \dots r_i, \dots r_{n_R}) \mapsto \bar{R} = \operatorname{diag}(r_1, \dots r_i, r_i, \dots r_{n_R}),$$

then we obtain the iterative system

$$\begin{pmatrix} C\frac{d}{dt}v_C^k \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & -I & 0 \\ 0 & -\bar{R} & 0 & 0 & A_M \\ 0 & 0 & 0 & 0 & A_E \\ -I & 0 & 0 & 0 & A_C \\ 0 & A_M^T & A_E^T & A_C^T & 0 \end{pmatrix} \begin{pmatrix} v_C \\ i_R \\ i_E \\ i_C \\ v_N \end{pmatrix}^k = \begin{pmatrix} 0 \\ A_N v_N^{k-1} \\ E(t) \\ 0 \\ 0 \end{pmatrix}, \tag{7.12}$$

with initial values $v_N^k(0) = v_N(0)$. The first observation of (7.12) is that its left-hand side has exactly the same symmetric structure as (7.1). In fact, if we can show that assumption (7.2) holds when A_R has been replaced by A_M, then we can immediately use the results of Section 7.1 to show that (7.12) can be solved in X_α for $\alpha > 0$.

The following lemma is proved in Leimkuhler et al. (1991).

Lemma 2

$$\text{If} A = \begin{pmatrix} A_R \\ A_E \\ A_C \end{pmatrix} \quad \text{has full rank, then} \quad \begin{pmatrix} A_M \\ A_E \\ A_C \end{pmatrix} \quad \text{has full rank.}$$

Let $\mathcal{K}: X_\alpha \to X_\alpha$ be the iteration operator of equation (7.12), and let

$$x^k = \mathcal{K}x^{k-1} + \varphi. \tag{7.13}$$

The Laplace transform of the iteration equation is then seen to be

$$\hat{x}^k = K(z)\hat{x}^{k-1} + \hat{\varphi},$$

where

$$K(z) = \begin{pmatrix} Cz & 0 & 0 & -I & 0 \\ 0 & -\bar{R} & 0 & 0 & A_M \\ 0 & 0 & 0 & 0 & A_E \\ -I & 0 & 0 & 0 & A_C \\ 0 & A_M^T & A_E^T & A_C^T & 0 \end{pmatrix}^{-1} \begin{pmatrix} & & & & \\ & & & & A_N \\ & & & & \end{pmatrix} \tag{7.14}$$

and $\hat{\varphi}$ is obvious from (7.12) because it does not depend on k.

The operator norm for $\mathcal{K}$ induced by $|\cdot|_\alpha$ is defined by

$$|\mathcal{K}|_\alpha = \sup_{|x|_\alpha = 1} |\mathcal{K}x|_\alpha.$$

By Lemma 1 and the preceding analysis we now obtain

Theorem 25 Assume (7.2) and (7.3) and apply the described splitting process. Then $|\mathcal{K}|_\alpha$ is finite for all positive α.

7.3. Spectrum of $\mathcal{K}$

We can see from (7.12) that all the other components of x are in ker $\mathcal{K}$ than v_N. If we define the projection operators

$$\mathcal{P}\begin{pmatrix} v_C \\ \vdots \\ v_N \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ v_N \end{pmatrix} \quad \text{and} \quad P\begin{pmatrix} \hat{v}_C \\ \vdots \\ \hat{v}_N \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \hat{v}_N \end{pmatrix},$$

we can deduce that the nontrivial part of the spectrum $\sigma(\mathcal{K})$ is actually $\sigma(\mathcal{P}\mathcal{K}\mathcal{P})$, which can be computed as in Leimkuhler et al. (1991):

$$\sigma(\mathcal{P}\mathcal{K}\mathcal{P}) = \text{cl} \bigcup_{\text{Re}\, z \geq \alpha} \sigma(PK(z)P).$$

As in Section 7.1 the equation $\hat{x}^k = K(z)\hat{x}^{k-1}$ can easily be manipulated to yield

$$\begin{pmatrix} 0 & A_E \\ A_E^T & B_M + zB_C \end{pmatrix}\begin{pmatrix} \hat{i}_E^k \\ \hat{v}_N^k \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & B_N \end{pmatrix}\begin{pmatrix} 0 \\ \hat{v}_N^{k-1} \end{pmatrix}, \tag{7.15}$$

where

$$B_M = A_M^T \bar{R}^{-1} A_M, \qquad B_C = A_C^T C A_C \qquad \text{and} \qquad B_N = A_M^T \bar{R}^{-1} A_N.$$

The solution of this equation clearly satisfies $\hat{v}_N^k \in \text{Ker} A_E$. Because of the incidence matrix structure of A_E (each row has at most two nonzero elements ± 1) we can easily eliminate the $\hat{i}_E$ and n_E components of $\hat{v}_N$ from (7.15), ending up with the equation

$$(\tilde{B}_M + z\tilde{B}_C)\hat{v}_P^k = \tilde{B}_N \hat{v}_P^{k-1}, \tag{7.16}$$

where $\hat{v}_P$ is a part of $\hat{v}_N$ with $N - n_E$ components.

The elimination described in Leimkuhler et al. (1991) is the same as 'shorting the edges' in graph theory: we short all edges containing voltage sources, and simultaneously the nodes adjacent to these edges are pairwise combined.

The computation of the spectrum relies on the following properties of the $\tilde{B}$-matrices. They are all symmetric, $\tilde{B}_N$ is nonnegative and $\tilde{B}_M$ and $\tilde{B}_C$ are positive semidefinite matrices with nonpositive off-diagonal elements. These facts imply that the splitting in iteration (7.16) is a regular splitting of the matrix $\tilde{B}_M + z\tilde{B}_C - \tilde{B}_N$, if $z \in (0, \infty)$. By the convergence theorem for regular splittings, iteration (7.16) then converges for $z \in (0, \infty)$, since $\tilde{B}_M + z\tilde{B}_C - \tilde{B}_N$ is a nonsingular M-matrix. The spectral radius $\rho((\tilde{B}_M + z\tilde{B}_C)^{-1}\tilde{B}_N) = r < 1$ for $z \in \mathbb{R}_+$, and, by the Perron–Frobenius theorem, r is also an eigenvalue. For nonreal z with $\text{Re}\, z > 0$, taking quadratic forms in the eigenvalue equation provides

$$(\tilde{B}_M + z\tilde{B}_C)^{-1}\tilde{B}_N x = \lambda x,$$

that is, following Leimkuhler et al. (1991),

$$|\mathrm{Re}\,\frac{1}{\lambda}| \geq \inf_{|x|=1} \frac{x^*\tilde{B}_M x + \mathrm{Re}\, z x^*\tilde{B}_C x}{|x^*\tilde{B}_N x|}.$$

The vector giving the minimum can be directly computed and is, of course, an eigenvector, so that

$$|\mathrm{Re}\,\frac{1}{\lambda}| \geq \frac{1}{r} > 1.$$

This inequality can be restated as

$$(\mathrm{Re}\,\lambda \mp \frac{r}{2})^2 + (\mathrm{Im}\,\lambda)^2 \leq (\frac{r}{2})^2,$$

where the negative sign is used for $\mathrm{Re}\,\lambda > 0$ and the positive sign for $\mathrm{Re}\,\lambda < 0$. So the spectrum of $PK(z)P$ lies in the closed circles of Figure 6 for $\mathrm{Re}\,z \geq \alpha > 0$.

Theorem 26 Let $\alpha > 0$. Assume (7.2), (7.3) and (7.7), and apply the described splitting process that only allows cutting through resistors. Then $\sigma(\mathcal{K})$ lies in the set $\mathcal{D}_\alpha$ of Figure 6. In particular, $\rho(\mathcal{K}) < 1$ and the iteration (7.12) converges in X_α.

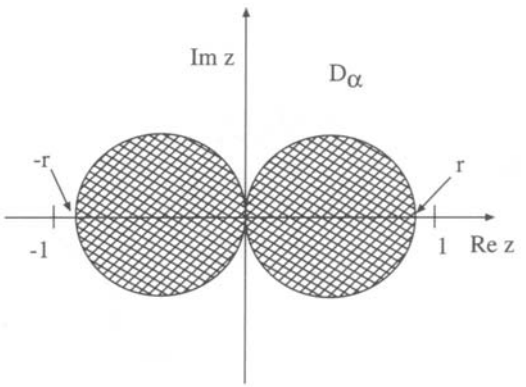

Fig. 6.

As mentioned in Section 7.1, for those circuits satisfying (7.6), the Laplace transform of (7.1) may also be boundedly solved for $z = 0$. Since

$$(7.6) \text{ implies } \begin{pmatrix} A_M \\ A_E \end{pmatrix} \quad \text{has linearly independent columns,}$$

this means that (7.12) can also be solved in the space X_0 without exponential weighting. In particular, for an index one system, $\mathcal{K}$ is a continuous operator in the ordinary L_2-space.

Theorem 27 Assume (7.2), (7.3), (7.6) and (7.7), and apply the described splitting process that only allows cutting through resistors. Then $\mathcal{K}$ is continuous in X_0 and $\rho(\mathcal{K}) < 1$, that is, the iteration (7.12) converges in X_0.

As stated in the beginning of Section 7.2 we can always permute the circuit equations and variables to a block form, where the blocks correspond to different subsystems. In that formulation, our iteration scheme is dynamic block Jacobi iteration and the corresponding iteration matrix clearly has zero trace. It has the same eigenvalues as $PK(z)P$ defined by (7.12), so we deduce that the trace of $K(z)$ vanishes.

REFERENCES

B. Aupetit (1991), *A Primer on Spectral Theory*, Springer, Berlin.

W. K. Chen (1976), *Applied Graph Theory, graphs and electrical networks*, North-Holland, Amsterdam.

T. Eirola (1992), 'Sobolev characterization of solutions of dilation equations', *SIAM J. Math. Anal.* **23**(4), 1015–1030.

F. R. Gantmacher (1959), *Matrizenrechnung II*, VEB Deutscher Verlag der Wissenschaften, Berlin.

J. Hadamard (1893), 'Etude sur les propriétés des fonctions entières et en particulier d'une fonction considérée par Riemann', *J. de Math. Pures et Appl.* **9**, 171–215.

R. Jeltsch and B. Pohl (1995), 'Waveform relaxation with overlapping splittings', *SIAM J. Sci. Comput.* **16**(1), 40–49.

F. Juang (1990), Waveform Methods for Ordinary Differential Equations, PhD thesis, University of Illinois at Urbana-Champaign, Dept of Computer Science. Report No. UIUCDCS-R-90-1563.

B. Leimkuhler, U. Miekkala and O. Nevanlinna (1991), 'Waveform relaxation for linear RC circuits', *Impact of Computing in Science and Engineering* **3**, 123–145.

E. Lelarasmee, A. Ruehli and A. Sangiovanni-Vincentelli (1982), 'The waveform relaxation method for time-domain analysis of large scale integrated circuits', *IEEE Trans. CAD* **1**(3), 131–145.

E. Lindelöf (1894), 'Sur l'application des méthodes d'approximations successives à l'Etude des intégrales réelles des Equations différentielles ordinaires', *J. de Math. Pures et Appl., 4e Série* **10**, 117–128.

C. Lubich (1992), 'Chebyshev acceleration of Picard–Lindelöf iteration', *BIT* **32**, 535–538.

C. Lubich and A. Ostermann (1987), 'Multigrid dynamic iteration for parabolic equations', *BIT* **27**, 216–234.

A. Lumsdaine and J. White (1995), 'Accelerating waveform relaxation methods with application to parallel semiconductor device simulation', *Numerical functional analysis and optimization* **16**(3,4), 395–414.

A. Lumsdaine and D. Wu (1995), Spectra and pseudospectra of waveform relaxation operators, Technical Report CSE-TR-95-14, Department of Computer Science and Engineering, University of Notre Dame, IN.

J. W. Manke, B. Dembart, M. A. Epton, A. M. Erisman, P. Lu, R. F. Sincovec and E. L. Yip (1979), *Solvability of Large Scale Descriptor Systems*, Boeing Computer Services Company, Seattle, WA.

U. Miekkala (1989), 'Dynamic iteration methods applied to linear DAE systems', *J. Comp. Appl. Math.* **25**, 133–151.

U. Miekkala (1991), Theory for iterative solution of large dynamical systems using parallel computations, PhD thesis, Helsinki University of Technology.

U. Miekkala (1996), Remarks on WR method with overlapping splittings. In preparation.

U. Miekkala and O. Nevanlinna (1987*a*), 'Convergence of dynamic iteration methods for initial value problems', *SIAM J. Sci. Stat. Comput.* **8**(4), 459–482.

U. Miekkala and O. Nevanlinna (1987*b*), 'Sets of convergence and stability regions', *BIT* **27**, 554–584.

U. Miekkala and O. Nevanlinna (1992), 'Quasinilpotency of the operators in Picard–Lindelöf iteration', *Numer. Funct. Anal. and Optimiz.* **13**(1,2), 203–221.

U. Miekkala, O. Nevanlinna and A. Ruehli (1990), Convergence and circuit partitioning aspects for waveform relaxation, in *Proceedings of the Fifth Distributed Memory Computing Conference, Charleston, South Carolina* (D. Walker and Q. Stout, eds), IEEE Computer Society Press, Los Alamitos, CA, pp. 605–611.

O. Nevanlinna (1989*a*), A note on Picard–Lindelöf iteration, in *Numerical Methods for Ordinary Differential Equations, Proceedings of the Workshop held in L'Aquila (Italy), Sept. 16-18, 1987. Vol. 1386 of Lecture Notes in Mathematics* (A. Bellen, C. W. Gear and E. Russo, eds), Springer.

O. Nevanlinna (1989*b*), 'Remarks on Picard–Lindelöf iteration, PART I', *BIT* **29**, 328–346.

O. Nevanlinna (1989*c*), 'Remarks on Picard Lindelöf iteration, PART II', *BIT* **29**, 535–562.

O. Nevanlinna (1990*a*), 'Linear acceleration of Picard–Lindelöf iteration', *Numer. Math.* **57**, 147–156.

O. Nevanlinna (1990*b*), 'Power bounded prolongations and Picard–Lindelöf iteration', *Numer. Math.* **58**, 479–501.

O. Nevanlinna (1991), Waveform relaxation always converges for RC-circuits, in *Proc. of NASECOD VII, held in April 8-12, 1991, Colorado*, Front Range Press, Colorado.

O. Nevanlinna (1993), *Convergence of iterations for linear equations*, Lectures in Mathematics, ETH Zürich, Birkhäuser, Basel.

O. Nevanlinna and F. Odeh (1987), 'Remarks on the convergence of waveform relaxation method', *Numer. Funct. Anal. and Optimiz.* **9**(3,4), 435–445.

E. Picard (1893), 'Sur l'application des méthodes d'approximations successives à l'étude de certaines équations différentielles ordinaires', *J. Math. Pures. Appl., 4e série* **9**, 217–271.

O. Piirilä (1993), 'Questions and notions related to quasialgebraicity in Banach algebras', *Annales Academia Scientiarum Fennica, Series A, Mathematica Dissertationes.* Helsinki.

M. Reichelt, J. White and J. Allen (1995), 'Optimal convolution SOR acceleration of waveform relaxation with application to parallel simulation of semiconductor devices', *SIAM J. Sci. Comput.* **16**(5), 1137–1158.

R. Skeel (1989), 'Waveform iteration and the shifted Picard splitting', *SIAM J. Sci. Stat. Comput.* **10**(4), 756–776.

M. N. Spijker (1991), 'On a conjecture by LeVeque and Trefethen related to the Kreiss matrix theorem', *BIT* **31**, 551–555.

L. N. Trefethen (1992), Pseudospectra and matrices, in *Numerical Analysis* (D. F. Griffiths and G. A. Watson, eds), Longman, Harlow, UK, pp. 234–266.

M. M. Vainberg and V. A. Trenogin (1974), *Theory of branching of solutions of non-linear equations*, Noordhoff International Publishing, Leyden.

S. Vandewalle (1992), The Parallel Solution of Parabolic Partial Differential Equations by Multigrid Waveform Relaxation Methods, PhD thesis, Katholieke Universiteit Leuven, Belgium.

S. Vandewalle (1993), *Parallel Multigrid Waveform Relaxation for Parabolic Problems*, B.G. Teubner, Stuttgart.

J. White and A. Sangiovanni-Vincentelli (1987), *Relaxation Techniques for the Simulation of VLSI Circuits*, Kluwer, Boston.