Numer. Math. 49, 275-290 (1986)

Numerische
Mathematik

A Convergence Analysis of Hopscotch Methods for Fourth Order Parabolic Equations

E. Jan W. ter Maten and Gerard L.G. Sleijpen

Mathematisch Instituut, Rijksuniversiteit Utrecht, Budapestlaan 6, 3508 TA Utrecht, The Netherlands

Summary. Consider the ODE (ordinary differential equation) that arises from a semi-discretization (discretization of the spatial coordinates) of a first order system form of a fourth order parabolic PDE (partial differential equation). We analyse the stability of the finite difference methods for this fourth order parabolic PDE that arise if one applies the hopscotch idea to this ODE.

Often the error propagation of these methods can be represented by a three terms matrix-vector recursion in which the matrices have a certain anti-hermitian structure. We find a (uniform) expression for the stability bound (or error propagation bound) of this recursion in terms of the norms of the matrices. This result yields conditions under which these methods are strongly asymptotically stable (i.e. the stability is uniform both with respect to the spatial and the time stepsizes (tending to 0) and the time level (tending to infinity)), also in case the PDE has (spatial) variable coefficients. A convergence theorem follows immediately.

Subject Classifications: AMS(MOS): 65M10, 65M20; CR: G1.8.

1. Introduction

Consider a family $\mathcal{F}$ of pairs (B, C) of real anti-hermitian matrices. In each pair the matrices B and C have the same size, but this size may differ from pair to pair. We are interested in the stability of the recursions

$$(I+C^*)\,U_{n+1}=BU_n+(I+C)\,U_{n-1} \qquad (n\in\mathbf{N}) \tag{1}$$

in which (U_n) is a sequence of real vectors of appropriate size.

In this paper we derive conditions on $\mathcal{F}$ for which there is a bound $\mathscr{C}$ on the euclidean norm $\|U_m\|$ of U_m that is uniform with respect to all m in $\mathbf{N}$, to all sequences (U_n) that satisfy (1) and for which $\|U_0\|^2+\|U_1\|^2=1$ and to all (B, C) in $\mathcal{F}$.

In applications B and C depend on the mesh-widths in space and time employed in the discretization of a PDE (partial differential equation); the recursions (1) appear in the stability analysis of certain finite difference methods.

Let $\mathbf{A}$ denote the companion matrix (see (13)) of (1). We actually obtain a bound $\mathscr{C}$ of $\sup\{\|\mathbf{A}^m\| \mid m \in \mathbf{N}\}$ that is uniform with respect to this family $\mathscr{F}$. One may assume that $\tilde{\mathscr{C}} \leqq \mathscr{C} \leqq \sqrt{2}\tilde{\mathscr{C}}$.

Such a uniform bound $\mathscr{C}$ only exists if the $\mathbf{A}$ have the following property (a).

(a) The spectral radius of $\mathbf{A}$ does not exceed 1 and all eigenvalues λ of $\mathbf{A}$ with $|\lambda| = 1$ are semisimple.

However, as is well-known, this property (a) does not guarantee the existence of this uniform bound. Assume that all $\mathbf{A}$ have property (a). From the fact that $|\det(\mathbf{A})| = 1$ one sees that the $\mathbf{A}$ are diagonizable:

there are non-singular matrices $\mathbf{X}$ and diagonal matrices Δ such that

$$\mathbf{A} = \mathbf{X}\Delta\mathbf{X}^{-1} \quad \text{and} \quad |\Delta| = I. \tag{2}$$

$\mathscr{C}$ is bounded by the supremum of the condition numbers of $\mathbf{X}$ (where the supremum is taken over $\mathscr{F}$). The $\mathbf{X}$ are such that any column vector of $\mathbf{X}$ has norm 1. Here, we obtain conditions on $\mathscr{F}$ for which these condition numbers have a uniform bound.

In our previous paper [6], among other things we proved that the companion matrices $\mathbf{A}$ have property (a) whenever $\|B\| < 2$.

Concerning our applications, the condition "$\|B\| < 2$" can considered to be sharp (see the discussion in [6, (6.4)]). Here we can bound the condition number of $\mathbf{X}$ under a somewhat more restrictive condition. Our main result runs as follows.

Main-Theorem. *Assume that $\|B\| < 2$. Let $\tau_0, \tau_1 \in (0, 1]$ be such that*

$$\left(\tfrac{1}{2}\|B\| + \tau_i \|C\|\right)^2 = 2^i(1 - \tau_i^2) \qquad \textit{for } i = 0, 1.$$

Then $t_0 := \inf\{|\mathrm{Re}(\lambda)| \mid \lambda$ eigenvalue of $\mathbf{A}\} \in [\tau_0, 1]$. Put $\beta := \|BC + CB\|/(4t_0)$ and $\mu := 2(1 - \tau_1^2)$. If $\beta < 1$ and $\tau_1 > \frac{1}{2}\sqrt{2}$ then $\mu < 1$ and

$$\|\mathbf{X}\| \, \|\mathbf{X}^{-1}\| \leqq 2(1 + \|C\|^2) \frac{1+\beta}{1-\beta} \frac{1}{\tau_1(1-\mu)}.$$

In our applications the above stability question arises as follows.

Let Ω be a region in $\mathbf{R}^M$. Let $\mathscr{L}$ be a second order partial differential operator from $C^{(2)}(\Omega)$ into $C(\Omega)^K$. Consider functions $v, w_1, \ldots, w_K$ on $[0, \infty) \times \Omega$ that are sufficiently smooth, that are the solutions of the following PDE (3) on $[0, \infty) \times \Omega$ and that satisfy some IC (initial conditions) and BC (boundary conditions)

$$\frac{\partial}{\partial t}(v, w_1, \ldots, w_K) = (-\mathscr{L}^*(w_1, \ldots, w_K), \mathscr{L}(v)). \tag{3}$$

Here, $\mathscr{L}^*$ is the formal adjoint of $\mathscr{L}$. $\mathscr{L}$ and the IC and BC are such that the solutions are unique. Consider the ODE (ordinary differential equation) that arises from a semi-discretization (a discretization with spatial stepsize $\Delta = (\Delta x_1, \ldots, \Delta x_M)$ of the spatial coordinates) of the above PDE (3). Recursion (1) appears in a stability analysis of the finite difference method that arises if, with time stepsize Δt, one applies the hopscotch idea to this ODE (for details, see § 2).

Now B and C are real anti-hermitian $N \times N$-matrices; N is proportional to $\prod_{k=1}^{M} 1/\Delta x_k$ and U_n represents the propagated error at the spatial grid at time level $n\Delta t$.

Although both B and C depend on Δ and Δt, quantities like $\|B\|$ and $\|C\|$ depend essentially only on $r_1, \ldots, r_M$, where $r_k = \Delta t/\Delta x_k^2$ $(k=1, \ldots, M)$. Therefore, the main-theorem gives conditions on $r_1, \ldots, r_M$ under which the error amplification is bounded uniformly both with respect to the time level $n\Delta t$ and to the stepsize $(\Delta t, \Delta)$ (provided that $\Delta t = r_k \Delta x_k^2$). This stability statement applies to many problems with variable coefficients and mixed derivatives. In particular, it also yields a convergence statement (with respect to the discrete L^2-norm on the spatial coordinates), and it tells us how the method can be used stably in a variable stepsize procedure.

In [1, 2], and [3], for second order parabolic PDEs, one showed that the approximates (obtained by a hopscotch method) converge with respect to an inner-product norm that is defined by some positive definite matrix H (depending on Δ and Δt) arising from the PDE. However, in the situation of the PDE (3), their approach has two disadvantages: in general, the matrix H can only be hermitian in case $\mathscr{L}$ is a constant coefficient operator and, moreover, glb(H) is proportional to $\Delta t^{\frac{1}{2}}$ (if, for some $r_1, \ldots, r_M$ $\Delta t = r_k \Delta x_k^2$, $k=1, \ldots, M$). Consequently, in this way, as far as convergence with respect to the discrete L^2-norm concerns, one should expect an additional loss in the convergence order of $\Delta t^{\frac{1}{2}}$.

In § 2, we explain how recursion (1) appears in the stability analysis of hopscotch methods applied to the PDE (3). For a special class of PDEs we interprete the stability result in the main-theorem as a stability theorem for the hopscotch methods. In § 3, 4, 5, we concentrate on the proof of our main-theorem. Although this paper is a natural sequel to our previous one [6], it can be read independently. In § 3, we collect the facts from [6] that we need here. Furthermore, in this section, we give a detailed description of the problem. The main part of this paper can be found in § 4. There, we give a purely algebraic analysis of matrices that are given by some matrix equation arising from (1). In § 5, we combine the results from § 4 and obtain our main-theorem.

2. A Stability Analysis of Hopscotch Methods for Fourth Order Parabolic Equations

In this section, we briefly recall the hopscotch methods (see (2.3)). Further, in (2.1), we sketch the PDEs for which our main-theorem gives stability results and in (2.4) we explain how recursion (1) appears in a stability analysis of the hopscotch methods applied to these PDEs. Finally, in (2.5), for some special PDEs and a special hopscotch method we explicitly formulate a stability theorem.

In order to avoid too much detail and since this section only serves as an illustration of the applicability of our results, we restrict ourselves to a model setting whenever it is convenient. For more details we refer the interested reader to [6].

(2.1) Let Ω be a region in $\mathbf{R}^M$. Let $\mathscr{L}$ be a second order partial differential operator from $C^{(2)}(\Omega)$ into $C(\Omega)^K$. $\mathscr{L}^*$ is the formal adjoint of $\mathscr{L}$ and operates from $C^{(2)}(\Omega)^K$ into $C(\Omega)$. Consider real-valued functions u and v and $\mathbf{R}^K$-valued functions $w=(w_1, \dots, w_K)^T$ on $[0, \infty) \times \Omega$ that are sufficiently smooth, that are the solutions of the following PDEs on $[0, \infty) \times \Omega$ and that satisfy some IC and BC

$$\begin{array}{ll} \text{PDE (4)} & u_{tt} = -\mathscr{L}^* \circ \mathscr{L}(u) \quad \text{on} \quad [0, \infty) \times \Omega \\ & u \text{ satisfies some IC (4) and BC (4)} \end{array} \tag{4}$$

and

$$\begin{array}{l} \text{PDE (5)} \quad \dfrac{\partial}{\partial t}\begin{pmatrix} v \\ w \end{pmatrix} = \begin{pmatrix} 0 & -\mathscr{L}^* \\ \mathscr{L} & 0 \end{pmatrix} \begin{pmatrix} v \\ w \end{pmatrix} \quad \text{on } [0, \infty) \times \Omega \\ (v, w_1, \dots, w_K)^T \text{ satisfies some IC (5) and BC (5).} \end{array} \tag{5}$$

$\mathscr{L}$ and the IC and BC are such that the solutions are unique.

For an extensive class of BC one can give a correspondence between BC (4) and BC (5) (and IC (4) and IC (5)) under which the problems (4) and (5) are equivalent (i.e. either $v=u_t$, $w=\mathscr{L}(u)$ or, integrated, $v=u$, $w_t=\mathscr{L}(u)$. See [5, Chap. I]). Here, we are interested in the methods that are induced by hopscotch methods for the (semi-discretization of the) PDE (5). However, in case the problems are equivalent methods and stability results can be translated (see [5, Chap. I] and [6, (4.5)]).

(2.2) *Example.* (The bending beam equation.) Let $\Omega=[0,1]^M$. For some positive coefficient functions $a_1, \dots, a_M$ in $C^{(2)}(\Omega)$ let $\mathscr{L}$: $C^{(2)}(\Omega) \to C(\Omega)^M$ be given by

$$\mathscr{L}(v) = \left(a_1 \frac{\partial^2 v}{\partial x_1}, \dots, a_M \frac{\partial^2 v}{\partial x_M}\right)^T \quad \text{for all } v \in C^{(2)}(\Omega).$$

Then

$$\mathscr{L}^*(w) = \sum_{k=1}^{M} \frac{\partial^2}{\partial x_k^2} a_k w_k \quad \text{for all } w=(w_1, \dots, w_M)^T \in C^{(2)}(\Omega)^M.$$

Consider the following BC and IC

$$\begin{array}{llll} \text{BC (4)} \; u=\varphi, \; \mathscr{L}(u)=\psi & \text{BC (5)} \; v=\varphi_t, \; w=\psi & & \text{on } [0, \infty) \times \partial\Omega \\ \text{IC (4)} \; u=f, \; u_t=g & \text{IC (5)} \; v=g, \; w=\mathscr{L}(f) & & \text{on } \{0\} \times \Omega \end{array}$$

in which, with $\psi=(\psi_1, \dots, \psi_M)^T$, $\varphi, \psi_1, \dots, \psi_M$ are smooth functions on $[0, \infty) \times \Omega$ and f, g on $\{0\} \times \Omega$.

Now $v = u_t$ and $w = \mathscr{L}(u)$ if u is the solution of (4) and $\begin{pmatrix} v \\ w \end{pmatrix}$ is the solution of (5).

(2.3) We now sketch the hopscotch method for problem (5).
We assume model BC (as, for instance in (2.2)). For simplicity and since our interest concerns stability, we also assume that the BC are homogeneous.

Let $\Delta = (\Delta x_1, \dots, \Delta x_M)$ be a spatial stepsize such that with $\kappa_k := -1 + 1/\Delta x_k$ we have that $\kappa_k \in \mathbf{N}$ $(k = 1, \dots, M)$. With

$$Y(t) = (V(t), W_1(t), \dots, W_K(t))^T \quad \text{and} \quad \mathbf{L} = \begin{pmatrix} 0 & -L^* \\ L & 0 \end{pmatrix}$$

consider the ODE

$$Y'(t) = \mathbf{L}\,Y(t), \qquad t \geqq 0 \tag{6}$$

that arises by the finite difference standard discretization of (5) on the spatial grid $\mathbf{Z}(\Delta) := \{(j_1, \dots, j_M) \in \mathbf{Z}^M \mid (j_1 \Delta x_1, \dots, j_M \Delta x_M) \in \text{interior}(\Omega)\}$; for $t \geqq 0$, $V(t), W_1(t), \dots, W_K(t)$ are real-valued functions on the grid $\mathbf{Z}(\Delta)$, $Y(t)$ is a $\mathbf{R}^{K+1}$-valued function on $\mathbf{Z}(\Delta)$ and $\mathbf{L}$ is the finite difference operator induced by $\mathscr{L}$ and the BC.

One may identify $Y(t)$ with a vector in $\mathbf{R}^N$, where $N := \left(\prod_{k=1}^{M} \kappa_k\right)(K+1)$, and $\mathbf{L}$ with an $N \times N$-matrix.

Let Δt be a time stepsize. The hopscotch method produces a sequence (Y_n) of $\mathbf{R}^{K+1}$-valued functions Y_n on $\mathbf{Z}(\Delta)$ as follows. Let $r \subset \mathbf{Z}(\Delta)$ and put $b := \mathbf{Z}(\Delta) \backslash r$. $\{r, b\}$ is a partitioning of the spatial grid in say red and black points. For any function Z on the grid $\mathbf{Z}(\Delta)$ let the grid-functions Z_r and Z_b be such that $Z_r = Z$, $Z_b = 0$ on r and $Z_r = 0$, $Z_b = Z$ on b. For each $n \in \mathbf{N}$, the grid-function operators J_{2n-1} and J_{2n} are given by $J_{2n} Z = Z_r$ and $J_{2n-1} Z = Z_b$. Now, the hopscotch method produces the sequence (Y_n) by

$$(Y_{n+1} - Y_n)/\Delta t = J_{n+1} \mathbf{L} Y_{n+1} + J_n \mathbf{L} Y_n \qquad (n \in \mathbf{N}) \tag{7}$$

(see e.g. [3], also [6, § 2] and [5, Chap. III]).

Depending on the particular choice of the "red-black partitioning" of $\mathbf{Z}(\Delta)$, (7) is an efficient scheme that is second order accurate with respect to Δt.

(2.4) With $C = \Delta t [J_{n+1} \mathbf{L} J_{n+1} + J_n \mathbf{L} J_n]$ and $B = 2\Delta t [J_n \mathbf{L} J_{n+1} + J_{n+1} \mathbf{L} J_n]$ consider also the recursion

$$(I - C)\, Y_{n+2} = B Y_{n+1} + (I + C)\, Y_n \qquad (n \in \mathbf{N}) \tag{8}$$

(see [6, § 2] and [5, Chap. III]).

The recursions (7) and (8) are equivalent in the following sense. If (Y_n) satisfies (7) then (Y_n) also satisfies (8). If (Y_n) satisfies (8) then both $([J_n Y_n + \frac{1}{2} J_{n+1}(Y_{n+1} + Y_{n-1})])$ and $([J_{n+1} Y_n + \frac{1}{2} J_n (Y_{n-1} + Y_{n+1})])$ satisfy (7).

In particular this implies that (7) and (8) have equivalent stability properties.

Essentially, the above construction only requires the ordinary differential equation (6). Such an equation also arises by a semi-discretization of the simple heat flow equation. In this case (8) corresponds to the Du Fort-Frankel scheme and the scheme is equivalent to a hopscotch method.

(2.5) Consider the situation as in (2.2). Let the red-black partitioning of $\mathbf{Z}(\Delta)$ be the checker board one. Then the matrices B and C are given by

$$C:=\begin{bmatrix} 0 & 2r_1 D_1 \dots 2r_M D_M \\ -2r_1 D_1 & 0 \quad \dots \quad 0 \\ \vdots & \vdots \qquad\quad \vdots \\ -2r_M D_M & 0 \quad \dots \quad 0 \end{bmatrix}$$

and

$$B:=\begin{bmatrix} 0 & -2r_1 S_1 D_1 & \dots & -2r_M S_M D_M \\ 2r_1 D_1 S_1 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 2r_M D_M S_M & 0 & \dots & 0 \end{bmatrix}.$$

Here D_k is the diagonal matrix that corresponds to the grid-function operator

$$(U(j))_{j\in\mathbf{Z}(\Delta)}\mapsto(a_k(j_1\Delta x_1,\dots,j_M\Delta x_M)\,U(j))_{j\in\mathbf{Z}(\Delta)}$$

S_k is the matrix that corresponds to the grid-function operator

$$(U(j))_{j\in\mathbf{Z}(\Delta)}\mapsto(U(j_1,\dots,j_{k-1},j_k+1,j_{k+1},\dots j_M)+U(j_1,\dots,j_{k-1},j_k-1,j_{k+1},\dots,j_M))_j$$

and $r_k=\Delta t/\Delta x_k^2$ (see also example (7.2) in [6]). Put

$$\rho:=(\rho_1^2+\dots+\rho_M^2)^{\frac{1}{2}} \quad \text{with} \quad \rho_k:=2r_k\{\max a_k(x)\,|\,x\in\Omega\}. \tag{9}$$

Then $\|C\|\leqq\rho$ and $\|B\|\leqq 2\rho$.

Now, from the main-theorem one can easily deduce the following result.

Theorem. *Let* $r_1,\dots,r_M\in(0,\infty)$ *be such that* $\rho<2-\sqrt{2}$ *(with* ρ *as in* (9)). *Then there is a* $\mathscr{C}\in\mathbf{R}$ ($\mathscr{C}\leqq 56/(2-\sqrt{2}-\rho)$) *such that for all* Δt, $\Delta=(\Delta x_1,\dots,\Delta x_M)$ *with* $1/\Delta x_k\in\mathbf{N}$, $\Delta t/\Delta x_k^2\leqq r_k$ ($k=1,\dots,M$) *and for all* $m\in\mathbf{N}$ *we have that* $\|Y_m\|\leqq\mathscr{C}$ *whenever* (Y_n) *satisfies* (8) *and* $\|Y_0\|^2+\|Y_1\|^2\leqq 1$ *(with* B *and* C *as in* (2.5)). □

By means of this theorem, one can easily prove that the method converges; this is left to the reader.

3. Notations, Conventions and Basic Facts

(3.1) On $\mathbf{C}^N$, $(.,.)$ denotes the standard inner product and $\|.\|$ is the associated Euclidean norm. With respect to the standard basis $e_1,\dots,e_N$ in $\mathbf{C}^N$ we identify $N\times N$-matrices A with linear maps A from $\mathbf{C}^N$ into $\mathbf{C}^N$ (the (k,l)-matrix entry A_{kl} of A is equal to (Ae_l,e_k)). The spectrum of A is denoted by $\sigma(A)$. The spectral norm $\sqrt{\max|\sigma(A^*A)|}$ of A is also denoted by $\|A\|$. If A is invertible then the condition number $\|A\|\,\|A^{-1}\|$ of A is denoted by $\mathscr{C}(A)$.

(3.2) Let B and C be anti-hermitian $N \times N$-matrices.

For such a pair (B, C) of matrices, we consider $N \times N$-matrices S, T and X that have the following properties.

$$\begin{gathered} S \text{ and } T \text{ are real diagonal matrices,} \\ \sigma(T) \subseteq (0, 1] \text{ and } S^2 + T^2 = I. \end{gathered} \tag{10}$$

$$\begin{gathered} X \text{ is non-singular, complex and} \\ BX + 2CXT = 2iXS \end{gathered} \tag{11}$$

$$\|X e_k\| = 1 \text{ and } (X e_k, X e_l) = 0 \text{ if } k \neq l \text{ and } T_{kk} = T_{ll}. \tag{12}$$

Put

$$\Delta := T - iS.$$

(3.3) **Lemma.** *If* $\|B\| < 2$ *then the matrices* S, T *and* X *with the above properties* (10), (11) *and* (12) *exist and, moreover,*

$$(I + C^*) X \bar{\Delta} = BX + (I + C) X \Delta. \quad \square$$

For a proof of the above lemma, we refer to the results in § 5 and § 6 of [6] (see (5.11) and (6.1)).

(3.4) The matrices S, T and X which have the above properties (10), (11) and (12) that correspond to the pair $(B, -C)$ of matrices are denoted by $\tilde{S}$, $\tilde{T}$ and $\tilde{X}$ respectively.

We put $D := -\tilde{T} - i\tilde{S}$ and $t_0 := \min(\sigma(T), \sigma(\tilde{T}))$ (>0). With

$$\mathbf{A} := \left[\begin{array}{c:c} I + C^* & 0 \\ \hdashline 0 & I \end{array}\right]^{-1} \left[\begin{array}{c:c} B & I + C \\ \hdashline I & 0 \end{array}\right] \tag{13}$$

$$\mathbf{X} := \tfrac{1}{2}\sqrt{2} \left[\begin{array}{c:c} X\bar{\Delta} & \tilde{X} \\ \hdashline X & \tilde{X}D \end{array}\right] \quad \text{and} \quad \boldsymbol{\Delta} := \left[\begin{array}{c:c} \bar{\Delta} & 0 \\ \hdashline 0 & \bar{D} \end{array}\right], \tag{14}$$

we have the following theorem; for a proof, we refer to the proof of (5.6) and to (5.11) and (6.1) in [6].

(3.5) **Theorem.** *Assume that* $\|B\| < 2$. *Then*

(a) $\mathbf{X}$ *is a non-singular matrix in which each column is a vector with norm equal to* 1.

(b) $$\mathbf{AX} = \mathbf{X}\boldsymbol{\Delta}, \quad |\boldsymbol{\Delta}| = I. \quad \square$$

(3.6) If B and C are as in (1), then obviously $\mathbf{A}$ is the companion matrix of this recursion (1). We are interested in an upper-bound of the condition number $\mathscr{C}(\mathbf{X})$. In § 4, we give such an upper-bound. Our estimate is based on the following proposition.

(3.7) **Proposition.** *Put* $\mathscr{C} := [\max(\|X\|, \|\tilde{X}\|)]\,[\max(\|X^{-1}\|, \|\tilde{X}^{-1}\|)]$.

(a) *If*

$$\mathscr{C}\sqrt{1 - t_0} < 1 \tag{15}$$

then

$$\mathscr{C}(\mathbf{X}) \leqq \sqrt{2}\,\mathscr{C}/(1-\mathscr{C}\sqrt{1-t_0}).$$

(b) *With* $\mathscr{S} := \{v \in \mathbf{C}^N \mid \|v\| = 1\}$, *put*

$$\alpha := \min(\inf\{\mathrm{Re}(X\bar{\Delta}X^{-1}v, v) \mid v \in \mathscr{S}\},\ \inf\{\mathrm{Re}(-\tilde{X}D\tilde{X}^{-1}v, v) \mid v \in \mathscr{S}\}).$$

If

$$\alpha > 0 \tag{16}$$

then $\mathscr{C}(\mathbf{X}) \leqq 2\mathscr{C}/\alpha$.

Proof. One easily shows that $\|\mathbf{X}\| \leqq \sqrt{2}\max(\|X\|, \|\tilde{X}\|)$. With

$$\tilde{\mathbf{X}} := \begin{bmatrix} X\bar{\Delta}X^{-1} & -I \\ I & -\tilde{X}D\tilde{X}^{-1} \end{bmatrix} \quad \text{we have that} \quad \mathbf{X} = \tfrac{1}{2}\sqrt{2}\,\tilde{\mathbf{X}} \begin{bmatrix} X & 0 \\ 0 & -\tilde{X} \end{bmatrix}.$$

Therefore $\|\mathbf{X}^{-1}\| \leqq \sqrt{2}\,\|\tilde{\mathbf{X}}^{-1}\| \max(\|X^{-1}\|, \|\tilde{X}^{-1}\|)$ and

$$\mathscr{C}(\mathbf{X}) \leqq 2\,\|\tilde{\mathbf{X}}^{-1}\|\,\mathscr{C} \text{ whenever } \tilde{\mathbf{X}} \text{ is invertible.}$$

(a) Note that $\|X\bar{\Delta}X^{-1} - I\| \leqq \mathscr{C}\,\|T - I + iS\| \leqq \mathscr{C}\sqrt{2(1-t_0)}$ and likewise $\|\tilde{X}D\tilde{X}^{-1} + I\| \leqq \mathscr{C}\sqrt{2(1-t_0)}$. Since, for any matrix A we have that $\|(I+A)^{-1}\| \leqq (1 - \|A\|)^{-1}$ whenever $\|A\| < 1$, and since we also have that

$$\tilde{\mathbf{X}} = \left[\begin{array}{c:c} I & -I \\ \hdashline I & I \end{array}\right] \left[\left[\begin{array}{c:c} I & 0 \\ \hdashline 0 & I \end{array}\right] + \frac{1}{2}\left[\begin{array}{c:c} I & I \\ \hdashline -I & I \end{array}\right] \left[\begin{array}{c:c} X\bar{\Delta}X^{-1} - I & 0 \\ \hdashline 0 & -\tilde{X}D\tilde{X}^{-1} - I \end{array}\right]\right]$$

it follows that

$$\|\tilde{\mathbf{X}}^{-1}\| \leqq \tfrac{1}{2}\sqrt{2}(1 - \tfrac{1}{2}\sqrt{2}\,\mathscr{C}\sqrt{2(1-t_0)})^{-1} \quad \text{if (15) holds.}$$

(b) Let $w \in \mathbf{C}^{2N}$, $\|w\| = 1$ be such that $\mathrm{glb}(\tilde{\mathbf{X}}) = \|\tilde{\mathbf{X}}w\|$. Since

$$\mathrm{glb}(\tilde{\mathbf{X}}) \geqq \mathrm{Re}(\tilde{\mathbf{X}}w, w) = \tfrac{1}{2}([\tilde{\mathbf{X}} + \tilde{\mathbf{X}}^*]\,w, w)$$

and

$$\tilde{\mathbf{X}} + \tilde{\mathbf{X}}^* = \left[\begin{array}{c:c} X\bar{\Delta}X^{-1} + [X\bar{\Delta}X^{-1}]^* & 0 \\ \hdashline 0 & -\tilde{X}D\tilde{X}^{-1} - [\tilde{X}D\tilde{X}^{-1}]^* \end{array}\right]$$

it follows that $\mathrm{glb}(\tilde{\mathbf{X}}) \geqq \alpha$.

Therefore $\|\tilde{\mathbf{X}}^{-1}\| \leqq 1/\alpha$ if $\alpha > 0$. □

In (4.1–6), we deduce a lower and an upper bound for $\sigma(X^*X)$ (and $\sigma(\tilde{X}^*\tilde{X})$; see (4.5)) and we comment on these results (in (4.6)). In (4.7–11), we concentrate on an estimate for a lower bound for α (see (4.10)). A combination of these results in § 4 with those in (3.7) yields our main-theorem. Since, by (3.7.a), an estimate for $\sigma(X^*X)$ and $\sigma(\tilde{X}^*\tilde{X})$ is sufficient to have a result of the announced type, we should justify our additional analysis in (4.7–11): our estimate τ_0 of t_0 ($\tau_0 \leqq t_0$), $\tilde{\mathscr{C}}$ of $\mathscr{C}$ ($\mathscr{C} \leqq \tilde{\mathscr{C}}$ in (4.6)) and $\tilde{\alpha}$ of α ($\alpha \geqq \tilde{\alpha}$ in (4.10)) are such that $\tilde{\alpha} > 0$ (see (15)) whenever $\tilde{\mathscr{C}}\sqrt{1-\tau_0} < 1$ (see (15); for a proof of this claim, see (4.11.b)). Therefore, our result based on (3.7.b) gives a better estimate than the one that is based on (3.7.a). For instance in the realistic situation where $\rho := \frac{1}{2}\|B\| = \|C\|$ (see § 2), we have that

$$\tilde{\alpha} > 0 \text{ if } \rho \leqq 0.589 \text{ (see (4.11.c)) while}$$
$$\tilde{\mathscr{C}}\sqrt{1-\tau_0} < 1 \text{ only if } \rho < 0.436 \text{ (see (4.6.c)).}$$

In § 4, and § 5 all the matrices and quantities are as in (3.2) and (3.4). Moreover, B is such that $\|B\| < 2$.

4. On The Equation $BX + 2CXT = 2iXS$

In (4.5), we obtain a bound for the spectrum $\sigma(X^*X)$. In order to prove this result, associated to the diagonal matrix T, we introduce the following two matrix operations.

(4.1) *Notations.* Let F be an $N \times N$-matrix.
$¶(F)$ is the $N \times N$-matrix in which the (k, l)-entry is equal to 0 if $T_{kk} \neq T_{ll}$ and equal to F_{kl} if $T_{kk} = T_{ll}$ ($¶(F)$ may considered to be a block-diagonal matrix). For any $n \in \mathbf{N}$,

$$F^{\sharp n} := \sum_{j=1}^{n} T^{n-j} F T^{j-1}.$$

A number of elementary properties of these operations will be used in the proof of (4.3). These properties are listed in the following lemma; its proofs are left to the reader (in the proof of (e) and (f), one may apply the theorem of Courant-Fischer and the relations in (h)).

(4.2) **Lemma.** *Let F be a hermitian $N \times N$-matrix*

(a) *If $TF = FT$ then $F = ¶(F)$.*

(b) $T^n F - F T^n = T F^{\sharp n} - F^{\sharp n} T$.

(c) *Both $F^{\sharp n}$ and $¶(F)$ are hermitian.*

(d) $¶(F^{\sharp n}) = n T^{n-1} ¶(F)$.

(e) $\|¶[(X^* F X)^{\sharp n}]\| \leqq n \|F\|$.

(f) *If F is positive definite then $¶(F^{\sharp n})$ is positive definite.*

(g) $\|(X^* F X)^{\sharp n}\| \leqq n \|F\| \, \|X\|^2$.

(h) $¶(X^* X) = I$ *and* $¶(TX^* X + X^* X T) = 2T$. □

(4.4) **Proposition.**

(a) *Let $D_0, D_1, D_2, \ldots$ be hermitian $N \times N$-matrices such that*

$$\gamma := \frac{1}{2t_0} \sum_{n=1}^{\infty} n \|D_n\| < 1$$

and

$$\sum_{n=0}^{\infty} D_n X T^n = X T^2.$$

Put $\beta := \dfrac{1}{2t_0}\|D_1\|$ *and* $\rho := \gamma - \beta$. *Then*

$$\left(1-\gamma-\rho\frac{1+\gamma}{1-\gamma}\right)\Big/\ (1+\beta) \leqq [\mathrm{glb}(X)]^2 \leqq \|X\|^2 \leqq (1+\gamma)/(1-\gamma).$$

(b) *Let* $\tilde{A}$, $\tilde{B}$ *and* $\tilde{C}$ *be hermitian* $N\times N$-*matrices such that*

$$\beta := \|\tilde{B}\|/(2t_0) < 1$$

and

$$\tilde{A}X + \tilde{B}XT + \tilde{C}^2XT^2 = -XT^2.$$

Then

$$\frac{1}{1+\|\tilde{C}\|^2}\,\frac{1-\beta}{1+\beta} \leqq [\mathrm{glb}(X)]^2 \leqq \|X\|^2 \leqq \frac{1+\beta+\|\tilde{C}\|^2}{1-\beta}$$

and

$$\mathscr{C}(X) \leqq (1+\|\tilde{C}\|^2)\frac{1+\beta}{1-\beta}.$$

Proof. (a) Put $W := \sum\limits_{n=0}^{\infty} D_n X T^n - XT^2$. Then, by our assumption and Lemma (4.2.b), we have that

$$\begin{aligned}0 = W^*X - X^*W = T\sum_{n=1}^{\infty}(X^*D_nX)^{\sharp n}\\ -\sum_{n=1}^{\infty}(X^*D_nX)^{\sharp n}T - T(TX^*X + X^*XT) + (TX^*X + X^*XT)T.\end{aligned}$$

Hence, by (a) and (h) of (4.2)

$$\sum_{n=1}^{\infty}(X^*D_nX)^{\sharp n} - [TX^*X + X^*XT] = \P\left[\sum_{n=1}^{\infty}(X^*D_nX)^{\sharp n}\right] - 2T.$$

Consider a $\lambda \in \sigma(X^*X)$ and a $v \in \mathbf{C}^N$ for which $(X^*X-\lambda)v = 0$ and $\|v\| = 1$. Then

$$(T[X^*X-\lambda]v, v) = 0 \quad \text{and} \quad ([X^*X-\lambda]Tv, v) = 0.$$

Therefore,

$$\begin{aligned}2\lambda(Tv, v) &= ([TX^*X + X^*XT]v, v)\\ &= \left(\sum_{n=1}^{\infty}(X^*D_nX)^{\sharp n}v, v\right) + 2(Tv, v) - \left(\P\sum_{n=1}^{\infty}(X^*D_nX)^{\sharp n}v, v\right).\end{aligned}$$

Now, note that

$$2(\lambda-1)(Tv, v) \in \{s \in \mathbf{R} \mid s = 2(\lambda-1)t,\ t \in [t_0, 1]\};$$

$$\left|\left(\sum_{n=1}^{\infty}(X^*D_nX)^{\sharp n}v, v\right)\right| \leqq 2t_0\gamma;$$

$$\left|\left(\sum_{n=2}^{\infty}(X^*D_nX)^{\sharp n}v, v\right)\right| \leqq 2t_0\rho\|X\|^2;$$

$$|(X^*D_1Xv, v)| = \left|\frac{(D_1Xv, Xv)}{(Xv, Xv)}\lambda\right| \leqq \|D_1\|\lambda = 2t_0\beta\lambda.$$

Therefore, if $\lambda > 1$ then

$$2(\lambda-1)\,t_0 \leqq 2t_0\beta\lambda + 2t_0\rho\,\|X\|^2 + 2t_0\gamma,$$

and $\lambda \leqq (1+\rho\,\|X\|^2+\gamma)/(1-\beta)$.

In particular, for $\lambda = \|X\|^2$, we have that

$$\|X\|^2 \leqq (1+\gamma)/(1-\gamma).$$

If $\lambda \leqq 1$ then

$$2(\lambda-1)\,t_0 \geqq -2t_0\beta\lambda - 2t_0\rho\,\|X\|^2 - 2t_0\gamma \geqq -2t_0\left(\beta\lambda+\gamma+\rho\frac{1+\gamma}{1-\gamma}\right).$$

(b) Obviously, the setting in (b) is a particular one of (a). However, here we have the additional information that $D_2 = -\tilde{C}^2$ is negative definite. One can use this additional information to improve the result a little.

Proceed as above in order to find that

$$TX^*(1+\tilde{C}^2)\,X + X^*(1+\tilde{C}^2)\,XT = 2T - X^*\tilde{B}X + \P(X^*\tilde{B}X) + 2T\,\P(X^*\tilde{C}^2X).$$

Now, consider a $\lambda\in\sigma(X^*(1+\tilde{C}^2)\,X)$ and a $v\in\mathbf{C}^N$, $\|v\|=1$ for which

$$[X^*(1+\tilde{C}^2)\,X-\lambda]\,v=0.$$

Observe that

$$|(X^*\tilde{B}X\,v,v)| \leqq \left|\frac{(\tilde{B}X\,v, X\,v)}{(X\,v, X\,v)}\,\frac{(X\,v, X\,v)}{([1+\tilde{C}^2]\,X\,v, X\,v)}\,\lambda\right| \leqq \|\tilde{B}\|\,\lambda,$$

and

$$0 \leqq (T\,\P(X^*\tilde{C}^2X)\,v,v) \leqq \frac{(\P(X^*\tilde{C}^2X)\,T^{\frac{1}{2}}v,\,T^{\frac{1}{2}}v)}{(T^{\frac{1}{2}}v,\,T^{\frac{1}{2}}v)} \quad (Tv,v) \leqq \|\tilde{C}\|^2\,(Tv,v).$$

By a reasoning as in (a), using these observations, we see that

$$\sigma(X^*(1+\tilde{C}^2)\,X) \subseteq \left[\frac{1-\beta}{1+\beta}, \frac{1+\beta+\|\tilde{C}\|^2}{1-\beta}\right].$$

We also have that

$$\sigma(X^*X) \subseteq \left\{([1+\tilde{C}^2]\,X\,v, X\,v)\frac{(X\,v, X\,v)}{([1+\tilde{C}^2]\,X\,v, X\,v)}\,\middle|\;\|v\|=1\right\},$$

and

$$\frac{1}{1+\|\tilde{C}\|^2} \leqq \frac{(X\,v, X\,v)}{([1+\tilde{C}^2]\,X\,v, X\,v)} \leqq 1.$$

Now, the statement in (b) of the proposition follows easily. □

(4.4) *Remark.* In our estimate of $\sigma(X^*X)$, we actually only need the result in (b) of (4.3). However, since the result in (a) has some interest on its own account, we also give this general formulation in (a). For instance, let $\zeta\mapsto\Gamma(\zeta)$ be an analytic map from a neighbourhood of $[0, 1]$ (in $\mathbf{C}$) into the space of $N\times N$-matrices such that

$$\Gamma(\zeta)^* = \Gamma(\bar{\zeta}) \quad \text{for all } \zeta$$

and for some $t_0 \in (0, 1]$ we also have that

$$\sigma(\Gamma(t)) \subseteq [t_0^2, 1] \quad \text{for all } t \in [t_0, 1].$$

Then (see [4, Chap. III, §§6.1–2] and [6, (5.3)]), there are $t_1, \ldots, t_N$ in $[t_0, 1]$ and $x_1, \ldots, x_N$ in $\mathbf{C}^N$ such that for each $k=1, \ldots, N$

$$\Gamma(t_k)x_k = t_k^2 x_k, \quad \|x_k\| = 1, \quad (x_k, x_l) = 0 \quad \text{if } k \neq l \text{ and } t_k = t_l.$$

Now, the result in (a) of (4.3) may be used to estimate the condition number of the matrix with k-th column equal to x_k.

(4.5) **Theorem.** *Put* $\beta := \frac{1}{4t_0}\|BC + CB\|$. *If* $\beta < 1$ *then*

$$\sigma(X^* X) \subseteq \left[\frac{1-\beta}{(1+\|C\|^2)(1+\beta)}, \frac{1+\beta+\|C\|^2}{1-\beta}\right] \quad \textit{and} \quad \mathscr{C}(X) \leqq (1+\|C\|^2)\frac{1+\beta}{1-\beta}.$$

Proof. Since $BX + 2CXT = 2iXS$, we also have that

$$B^2 X + 2(BC + CB)XT + 4C^2 XT^2 = -4XS^2.$$

Therefore, by (10),

$$\tfrac{1}{4}(B^2 + I)X + \tfrac{1}{2}(BC + CB)XT + C^2 XT^2 = XT^2.$$

Now, we may apply (4.3.b) with

$$\tilde{A} = -\tfrac{1}{4}(B^2 + I), \quad \tilde{B} = -\tfrac{1}{2}(BC + CB) \quad \text{and} \quad \tilde{C} = iC. \quad \square$$

(4.6) *Remark.* (a) If $t_m := \frac{1}{2}\|B\|\,\|C\|$ then $\beta < 1$ whenever $t_0 > t_m$. In particular $\beta < 1$ if $4(1 - t_m^2) > (\|B\| + 2t_m\|C\|)^2$ or, equivalently, if $-1 + 4(\|B\|)^{-1} > \|C\|^2(3 + \|C\|^2)$.

(b) Let $\tau_0 \in (0, 1]$ be such that

$$(\tfrac{1}{2}\|B\| + \tau_0\|C\|)^2 = (1 - \tau_0^2). \tag{17}$$

Then $t_0 \geqq \tau_0$. Put

$$\tilde{\beta} := \|B\|\,\|C\|/(2\tau_0) \quad \text{and} \quad \tilde{\mathscr{C}} := (1-\tilde{\beta})^{-1}[(1+\|C\|^2)(1+\tilde{\beta})(1+\|C\|^2+\tilde{\beta})]^{\frac{1}{2}}.$$

Then $\beta \leqq \tilde{\beta}$ and $\mathscr{C}(X) \leqq \tilde{\mathscr{C}}$.

Since $0 \leqq (\|B\| - 2\tau_0\|C\|)^2 = 4(1 - \tau_0^2) - 16\tau_0^2\tilde{\beta}$, we have that

$$\tilde{\beta} \leqq (1 - \tau_0^2)/(4\tau_0^2).$$

In particular, we have that $\beta < 1$ whenever

$$\tau_0 > \tfrac{1}{5}\sqrt{5} \approx 0.447 \quad (\text{or } \|B\| < \tfrac{2}{5}\sqrt{5}(2 - \|C\|)).$$

(c) Suppose that

$$\rho := \tfrac{1}{2}\|B\| = \|C\|.$$

(In a stability analysis of hopscotch methods this often occurs see [6, § 7] and (2.5).) Then $\tau_0=(1-\rho^2)/(1+\rho^2)$ and $\tilde{\beta}=\rho^2/\tau_0$.

Now, $\tilde{\beta}<1$ if and only if $\rho<(\sqrt{2}-1)^{\frac{1}{2}}\approx 0.644$. Since

$$\tilde{\mathscr{C}}\sqrt{1-\tau_0}\geqq\sqrt{1+\rho^2}\frac{(1+\tilde{\beta})}{(1-\tilde{\beta})}\sqrt{1-\tau_0}=\sqrt{2}\rho\frac{1+\tilde{\beta}}{1-\tilde{\beta}},$$

we have that $\tilde{\mathscr{C}}\sqrt{1-\tau_0}>1$ whenever $\rho\geqq 0.420$. Therefore, we can only hope that the estimate as in (3.7.a) can be used if at least $\rho<0.420$.

(4.7) Put $\tilde{T}:=XTX^{-1}$ and for any $\tau\in(0,1]$, $E(\tau):=-I+\tilde{T}/\tau$ and $e(\tau):=\|E(\tau)\|$. In (4.8-9), we obtain an upper bound for $e(\tau)$. This bound will be used as follows. Let

$$\alpha_0:=\inf\{\operatorname{Re}(X\Delta X^{-1}v,v)\mid v\in\mathbf{C}^N,\ \|v\|=1\}\qquad(\text{see (3.7.b)}).$$

In our estimate for the error propagation bound, it is important to have an α_0 that is as large as possible; at least we should have that $\alpha_0>0$ (see (3.7.b)).

Since $BX+2CXT=2iXS$ we have that

$$B+2C\tilde{T}=2iXSX^{-1}$$

and consequently

$$X\Delta X^{-1}=\tau+\tfrac{1}{2}B+\tau C+\tau(I+C)E(\tau)\qquad(\text{for all }\tau\in(0,1]).$$

By the fact that the B and C are anti-hermitian, this implies that

$$\alpha_0\geqq\tau[1-\|(I+C)E(\tau)\|].\tag{18}$$

We will use an upper bound for $e(\tau)$ to obtain an upper bound for $\|(I+C)E(\tau)\|$ and consequently a lower bound for α_0. The parameter τ will be used to optimalize this lower bound.

(4.8) **Lemma.** *For $\tau\in[-1,+1]$, put $W(\tau):=(\tfrac{1}{2}B+\tau C)^2+(I-\tau^2)$ and*

$$w(\tau):=\max(\|W(\tau)\|,\|W(-\tau)\|).$$

Assume that there is a $\tau\in(\tfrac{1}{2},1]$ such that $w(\tau)\leqq\tau^2$. Then

$$e(\tau)\leqq w(\tau)/(2\tau^2-w(\tau))\tag{19}$$

and

$$\|(I+C)E(\tau)\|\leqq w(\tau)/(2\tau^2-w(\tau)).\tag{20}$$

Proof. From the identity

$$(\tfrac{1}{2}B)^2+\tfrac{1}{2}(BC+CB)\tilde{T}+C^2\tilde{T}^2=\tilde{T}^2-I$$

one may verify that

$$W(\tau)(I+\tilde{T}/\tau)+W(-\tau)(I-\tilde{T}/\tau)=2(I-C^2)(\tilde{T}^2-\tau^2).$$

In other words

$$2\tau^2(I-C^2)E(\tau)(2+E(\tau))=W(\tau)(2+E(\tau))-W(-\tau)E(\tau).\tag{21}$$

Now, suppose that

$$e(\tau) \leqq 1. \tag{22}$$

Then (21), and the fact that $\|(I-C^2)^{-1}\| \leqq 1$ (C^2 is negative definite) and also $\|(2+E(\tau))^{-1}\| \leqq (2-e(\tau))^{-1}$ imply that

$$2\tau^2 e(\tau)(2-e(\tau)) \leqq w(\tau)(2-e(\tau)) + w(\tau)\, e(\tau) = 2w(\tau).$$

If $w(\tau) \leqq \tau^2$, we now see that

$$e(\tau) \leqq w(\tau)/(2\tau^2 - w(\tau)) \leqq w(\tau)/\tau^2. \tag{23}$$

Since $(I-C^2) = (I+C)(I-C)$ and $\|(I-C)^{-1}\| \leqq 1$, from (21), we also may conclude that

$$2\tau^2 \|(I+C)E(\tau)\| (2-e(\tau)) \leqq 2w(\tau)$$

whenever (22) is correct. If $w(\tau) \leqq \tau^2$, we may use (23) in order to find that (20) holds.

Finally, we should show that our assumption (22) is correct whenever $w(\tau) \leqq \tau^2$. For this purpose, for $\rho \in [0,1]$, consider the relation $\rho BX(\rho) + 2\rho CX(\rho)T(\rho) = 2iX(\rho)S(\rho)$. Using the results in [4, Chap. III, §§ 6.1-2] (see also (3.2)), one can see that for the pair $(\rho B, \rho C)$ there is a choice $S(\rho)$, $T(\rho)$ and $X(\rho)$ (and $X(\rho)^{-1}$) that continuously depend on ρ. With $e_\rho(\tau) := \left\| I - \frac{1}{\tau} X(\rho) T(\rho) X(\rho)^{-1} \right\|$, $e_\rho(\tau)$ continuously depends on ρ. Since $e_0(\tau) = |1 - 1/\tau| < 1$ and $e_1(\tau) = e(\tau)$, a continuity argument and the result (23) imply (22). □

(4.9) **Lemma.** *Let* $\tau_1 \in (0,1]$ *be such that*

$$(\|\tfrac{1}{2}B\| + \tau_1 \|C\|)^2 = 2(1-\tau_1^2).$$

Then $w(\tau_1) \leqq \frac{1}{2}(\|\frac{1}{2}B\| + \tau_1\|C\|)^2 = (1-\tau_1^2)$. *If* $\tau_1 \geqq \frac{1}{2}\sqrt{2}$ *then* $w(\tau_1) \leqq \tau_1^2$ *and*

$$e(\tau_1) \leqq (\|\tfrac{1}{2}B\| + \tau_1\|C\|)^2$$
$$\|(I+C)E(\tau_1)\| \leqq (\|\tfrac{1}{2}B\| + \tau_1\|C\|)^2.$$

Proof. Since $(\frac{1}{2}B + \tau_1 C)^2$ is negative definite and since $\|(\frac{1}{2}B + \tau_1 C)^2\| \leqq (\|\frac{1}{2}B\| + \tau_1\|C\|)^2 = 2(1-\tau_1^2)$, we have that $\|W(\tau_1)\| \leqq (1-\tau_1^2)$. The other results easily follow from the preceding lemma. □

From (18) and the result in (4.9), one immediately deduces the following proposition.

(4.10) **Proposition.** *Let* $\tau_1 \in [0,1]$ *be such that*

$$(\|\tfrac{1}{2}B\| + \tau_1\|C\|)^2 = 2(1-\tau_1^2).$$

If $\tau_1 > \frac{1}{2}\sqrt{2}$ *then*

$$\mathrm{Re}(X\varDelta X^{-1}v, v) \geqq \tau_1[1 - (\|\tfrac{1}{2}B\| + \tau_1\|C\|)^2] > 0 \quad \textit{for all } v \in \mathbf{C}^N,\ \|v\| = 1.$$

(4.11) *Remark.* (a) Obviously, one may improve the result in (4.9) be finding a τ_1 for which

$$\max(\|\tfrac{1}{2}B+\tau_1 C\|^2, \|\tfrac{1}{2}B-\tau_1 C\|^2)=2(1-\tau_1^2).$$

Also by using an expression like the one in (20) (or (19)), one may improve the bound in (4.10).

(b) Let τ_0, $\tilde{\beta}$ and $\tilde{\mathscr{C}}$ be as in (4.6.b). Let τ_1 be as in (4.10). In order to justify our analysis in (4.7–10), we will show now that we have the following property.

Property.

$$\tau_1>\tfrac{1}{2}\sqrt{2} \textit{ whenever } \tilde{\mathscr{C}}\sqrt{1-\tau_0}\leqq 1 \tag{24}$$

(*compare with the conditions* (15) *and* (16) *in* (3.7)).

Proof. One easily verifies that $\tau_1^2\geqq 2\tau_0^2/(1+\tau_0^2)$. Therefore,

$$\tau_1>\tfrac{1}{2}\sqrt{2} \quad \text{whenever} \quad \tau_0>\tfrac{1}{3}\sqrt{3}\approx 0.577. \tag{25}$$

Since

$$\tau_0^{-2}[1-\tfrac{1}{4}\|B\|^2]=1+\|C\|^2+(\|B\|\;\|C\|)/\tau_0\leqq 1+\|C\|^2+\frac{2\tilde{\beta}}{1-\tilde{\beta}}\leqq\tilde{\mathscr{C}},$$

the assumptions

$$\tau_0\in(0,\tfrac{1}{2}\sqrt{2}] \quad \text{and} \quad \tilde{\mathscr{C}}\leqq(1-\tau_0)^{-\frac{1}{2}}$$

imply that $\tau_0\geqq\frac{1}{2}\sqrt{2}[1-\frac{1}{4}\|B\|^2]^{\frac{1}{2}}$ and

$$\|B\|\;\|C\|\leqq[(1-\tau_0)^{-\frac{1}{2}}-1]\,\tau_0\leqq\sqrt{2}\tau_0^2.$$

Hence

$$\begin{aligned}(\tfrac{1}{2}\|B\|+\tfrac{1}{2}\sqrt{2}\|C\|)^2&=1-\tau_0^2+(\tfrac{1}{2}\sqrt{2}-\tau_0)\,\|B\|\;\|C\|+(\tfrac{1}{2}-\tau_0^2)\,\|C\|^2\\&\leqq 1-\tau_0^2+(\tfrac{1}{2}\sqrt{2}-\tau_0)\,\|B\|\;\|C\|+\tfrac{1}{8}\|B\|^2\,\|C\|^2\leqq 1.\end{aligned}$$

A combination of (25) and this last inequality implies (24). □

(c) Consider again the case where $\rho:=\frac{1}{2}\|B\|=\|C\|$ (see (4.6.c)). Then $\tau_1=(2-\rho^2)/(2+\rho^2)$. Now, we have that $\tau_1>\frac{1}{2}\sqrt{2}$ if

$$\rho<2-\sqrt{2}\approx 0.585.$$

5. The Main-Theorem

A combination of the results in (3.7.b), Theorem (4.5) and Proposition (4.10) gives an upper bound for $\mathscr{C}(\mathbf{X})$ (see (3.6)).

(5.1) **Main-Theorem.** *Put* $\beta:=\|BC+CB\|/(4t_0)$ *and let* $\tau_1\in[0,1]$ *be such that* $\mu:=(\|\frac{1}{2}B\|+\tau_1\|C\|^2)=2(1-\tau_1^2)$. *If*

$$\beta<1 \quad \textit{and} \quad \tau_1>\tfrac{1}{2}\sqrt{2} \tag{26}$$

then $\mu<1$ *and*

$$\mathscr{C}(\mathbf{X})\leqq 2(1+\|C\|)^2\,\frac{1+\beta}{1-\beta}\,\frac{1}{\tau_1(1-\mu)}\leqq\frac{4\sqrt{2}(1+\|C\|^2)}{(1-\beta)\,(1-\mu)}. \quad \square$$

(5.2) *Remark.* (a) One may improve the result in (5.1) a little by exploiting the observation in (4.11.a).

(b) Let τ_0 be such that $(\|\frac{1}{2}B\|+\tau_0\|C\|)^2=(1-\tau_0^2)$ (as in (4.6.b)). If $\tau_0>\frac{1}{3}\sqrt{3}$ then (26) holds (see (4.6.b) and (4.11.b)), and $\beta<\frac{1}{2}$ (see (4.6.b)). In this case (see (4.11.b)),

$$\mathscr{C}(\mathbf{X})\leqq 6\sqrt{2}\,\frac{1+\|C\|^2}{1-\mu}\leqq 6\sqrt{2}(1+\|C\|^2)(\tau_0^2+1)/(3\tau_0^2-1).$$

(c) If $\rho:=\|\frac{1}{2}B\|=\|C\|$ then (26) holds if

$$\rho<2-\sqrt{2}\approx 0.585 \qquad \text{(see (4.6.c) and (4.11.c)).}$$

In this case

$$\beta\approx 0.702 \quad \text{and} \quad \mathscr{C}(\mathbf{X})\leqq 56/(2-\sqrt{2}-\rho).$$

Acknowledgment. The authors are grateful to Prof. dr. A. van der Sluis for stimulating discussions and helpful comments.

References

1. Evans, D.J., Danaee, A.: A new group hopscotch method for the numerical solution of partial differential equations. SIAM J. Numer. Anal. **19**, 588–598 (1982)
2. Gane, C.R., Gourlay, A.R.: Block hopscotch procedures for second order parabolic differential equations. JIMA **19**, 205–216 (1977)
3. Gourlay, A.R.: Hopscotch: A fast second-order partial differential equation solver. JIMA **6**, 375–390 (1970)
4. Kato, T.: Perturbation theory for linear operators. Berlin, Heidelberg, New York: Springer 1966
5. ter Maten, E.J.W.: Stability analysis of finite difference methods for fourth order parabolic partial differential equations. Thesis, University of Utrecht, Utrecht 1984
6. ter Maten, E.J.W., Sleijpen, G.L.G.: Hopscotch methods for fourth order parabolic equations I: stability results for fixed stepsizes. Preprint 275, Mathematical Institute, University of Utrecht, Utrecht 1983

Received January 19, 1984/February 28, 1986