CERN LIBRARIES, GENEVA

SCAN-9705050

Preprint

NTZ 13/1997

**Exact and approximate solutions
of some operator equations based
on the Cayley transform**

by

SW 9720

I. P. Gavriluk and V. L. Makarov

Die Preprintreihe des ZHS enthält Beiträge aus den Abteilungen Naturwissenschaftlich-Theoretisches Zentrum, Geistes- und Sozialwissenschaftliches Zentrum und Frankreich-Zentrum. Sie sind auch im Internet unter der Adresse <http://www.uni-leipzig.de/~zhs> abzurufen.

Für weitere Auskünfte und eine Liste der bereits erschienenen und geplanten Texte wenden Sie sich bitte an: Universität Leipzig, Zentrum für Höhere Studien, Augustusplatz 10/11, 04109 Leipzig.

Exact and approximate solutions of some operator equations based on the Cayley transform

Ivan P. Gavrilyuk

Fakultät für Mathematik und Informatik
Universität Leipzig
Augustusplatz 10
04109 Leipzig
GERMANY

Vladimir L. Makarov

Department of Cybernetics
Kiev University "T. Shevtshenko"
Prospect Glushkova 6
252127 Kiev - 127
UKRAINE

Abstract. We consider the operator equation $SX \equiv \sum_{j=1}^M U_j X V_j = Y$ where $\{U_j\}, \{V_j\}$ are some commutative sets of operators but in general $\{U_j\}$ need not commute with $\{V_j\}$. Particular cases of this equation are the Sylvester and Ljapunov equations. We give a new representation and an approximation of the solution which is suitable to perform it algorithmically. Error estimates are given which show exponential convergence for bounded operators and polynomial convergence for unbounded ones. Based on these considerations we construct an iterative process and give an existence theorem for the operator equation $Z^2 + A_1 Z + A_2 = 0$, arising for example when solving an abstract second order differential equation with non-commutative coefficients.

1. Introduction

Let $\mathcal{A}$ be a complex Banach algebra with identity e and $\mathcal{B}$ be a Banach algebra of operators on $\mathcal{A}$ considered as a Banach space. Given $\{U_j\}, \{V_j\} \subset \mathcal{A}$ let $S \in \mathcal{B}$ be defined by

$$SX = \sum_{j=1}^M U_j X V_j \quad (1)$$

where $\{U_j\}$ and $\{V_j\}$ are commutative subsets of $\mathcal{A}$ but $\{U_j\}$ need not commute with $\{V_j\}$. The operator S was studied in [24] where it was shown that if $\Sigma(X, \mathcal{A})$ denotes the spectrum of $X \in \mathcal{A}$, then

$$\Sigma(S, \mathcal{B}) \subset \sum_{j=1}^M \Sigma(U_j, \mathcal{A}) \times \Sigma(V_j, \mathcal{A}) .$$

Furthermore, if $f(\lambda)$ is holomorphic in a domain that contains $\sum_{j=1}^M \Sigma(U_j, \mathcal{A}) \times \Sigma(V_j, \mathcal{A})$ then there exist Cauchy domains $D_i^V \supset \Sigma(V_i, \mathcal{A})$, $D_i^U \supset \Sigma(U_i, \mathcal{A})$, $1 \leq i \leq M$ such that for any $X \in \mathcal{A}$

$$\begin{aligned} f(S)X &= \frac{1}{(2\pi i)^M} \int_{\partial D_1^V} \cdots \int_{\partial D_M^V} f \left(\sum_{j=1}^M \lambda_j U_j \right) X \prod_{j=1}^M (\lambda_j I - V_j)^{-1} d\lambda_j \\ &= \frac{(-1)^M}{(2\pi)^{2M}} \int_{\partial D_1^V} \cdots \int_{\partial D_M^V} \int_{\partial D_1^U} \int_{\partial D_M^U} f \left(\sum_{j=1}^M \lambda_j \mu_j \right) \prod_{j=1}^M (\mu_j I - U_j)^{-1} d\mu_j X \quad (2) \\ &\quad \times \prod_{j=1}^M (\lambda_j I - V_j)^{-1} d\lambda_j , \end{aligned}$$

where ∂D_i^V , ∂D_i^U denote the boundaries of D_i^V and D_i^U , respectively and I the unit operator.

Let us consider the following operator equation

$$SX = Y \quad (3)$$

and suppose that S^{-1} exists. Then using the formula (2) for the function

$$f \left(\sum_{j=1}^M \lambda_j \mu_j \right) \equiv \left(\sum_{j=1}^M \lambda_j \mu_j \right)^{-1} \quad (4)$$

we get

$$X = \frac{(-1)^M}{(2\pi)^{2M}} \int_{\partial D_1^V} \cdots \int_{\partial D_M^U} \left(\sum_{j=1}^M \mu_j \lambda_j \right)^{-1} \prod_{j=1}^M (\mu_j I - U_j)^{-1} d\mu_j Y \prod_{j=1}^M (\lambda_j I - V_j)^{-1} d\lambda_j . \quad (5)$$

The following special case of equation (3) was considered in [6]:

$$\sum_{j,k=0}^n c_{jk} A^j X B^k = Y , \quad (6)$$

where $c_{j,k} \in \mathbb{R}$, $B \in \mathcal{L}(B_1)$, $A \in \mathcal{L}(B_2)$, $Y \in \mathcal{L}(B_1, B_2)$ are given bounded operators, $X \in \mathcal{L}(B_1, B_2)$ is the operator to be found, B_1 , B_2 are Banach spaces, $\mathcal{L}(B_i)$ and $\mathcal{L}(B_1, B_2)$ denote the sets of linear bounded operators acting in B_i and from B_1 into B_2 respectively. It was shown that the unique solution is given by

$$X = - \frac{1}{4\pi^2} \oint_{\Gamma_A} \oint_{\Gamma_B} \frac{(A - \lambda I)^{-1} Y (B - \mu I)^{-1}}{P(\lambda, \mu)} d\lambda d\mu \quad (7)$$

provided that $P(\lambda, \mu) = \sum_{j,k=0}^n c_{jk} \lambda^j \mu^k \neq 0$ for $(\lambda, \mu) \in \Sigma(A) \times \Sigma(B)$, where $\Sigma(A)$, $\Sigma(B)$ are the spectral sets of A and B respectively, Γ_A and Γ_B denote pathes in the resolvent sets $\rho(A)$, $\rho(B)$ of A , B surrounding $\Sigma(A)$ and $\Sigma(B)$. The representations (5), (7) are also valid for unbounded operators under appropriate assumptions about the behaviour of their resolvents. They are very usefull in many applications (see, for example, [6]) but they are not suitable for algorithmical realizations.

Particular cases of the equations (3), (6) are the Sylvester equation

$$AX + XB = C$$

and the Lyapunov equation

$$AX + XA^T = C.$$

The Sylvester equation has applications, for example, in the control theory [2,3,7,9]. An application of Lyapunov's equation is involved while using Newton's method for Riccati matrix equations arising in optimal control problems [25]. One has to solve the Lyapunov equations

$$\begin{aligned} AX + XA^T &= -BB^T \\ A^T Y + YA &= -C^T C \end{aligned}$$

for constructing a near-optimal reduced-order model for a dynamical system

$$\begin{aligned} \dot{x} &= Ax + Bu, \\ y &= Cx \end{aligned} \quad (8)$$

with state $x \in \mathbb{R}^N$, input $u \in \mathbb{R}^p$ and output $y \in \mathbb{R}^q$ ($A \in \mathbf{Mat}_N(\mathbb{R})$, $B \in \mathbf{Mat}_{N,p}(\mathbb{R})$, $C \in \mathbf{Mat}_{N,q}(\mathbb{R})$). Many systems in large space structure control theory may be described by a system of partial differential equations [2] or abstract differential and operator equations [1]. The discretization of this partial differential operator in the space variables by finite element or finite difference methods leads to ordinary differential equations of the type (8) with a large and sparse matrix and finally to a Lyapunov equation. There are direct and iterative methods to solve this equation [4, 5, 7, 9, 10, 14 - 18, 21 - 23, 26 - 29]. The first ones are preferred when solving equations with matrices of moderate size. The direct methods are often based on various decompositions of the matrix (for example, the Schur decomposition) with complexity $O(N^3)$, which restricts their use to problems of relatively small N . The iterative methods (SOR, ADI and other) are used for the Lyapunov equation when the matrix A is large. In order to arrive the computational stability and to decrease the number of iteration steps various preconditioning techniques are used [16].

The aim of this paper is to find explicit representations of the solutions of (3), (6) which are the basis for some constructive approximations and to give error estimates for these approximations. The computational costs of our approach are proportional to the costs of computing of $(\gamma_j I + U_j)^{-1}$, $(\delta_j I + V_j)^{-1}$, $\gamma_j > 0$, $\delta_j > 0$, $j = 1, \dots, M$ for the equation (3) and of $(\gamma I + A)^{-1}$, $(\delta I + B)^{-1}$, $\gamma > 0$, $\delta > 0$ for Sylvester equation. Furthermore, these results will be used in the second part of this article in order to solve a second order operator equation. Besides optimal control problems that equations arise in the theory of electrical networks [5] and in the theory of quadratic pencils of selfadjoint operators [20]. Using the Newton method for solving such equations one has to solve a Sylvester equation in each step of the iteration process [8]. We consider here also a relation between such equations and a second order operator differential equation with non-commutative coefficients. This relation allows one to represent the solution of the differential equation via some operator exponents.

2. Preliminary results

Let A be an operator in a Banach space E with the resolvent set $\rho(A)$ and the spectral set $\Sigma(A)$. The operator A is called strongly positive if there exists a closed path Γ in the complex plane which consists of two rays

$$\mathcal{R}(\pm\varphi) \equiv \mathcal{R}(\pm\varphi, A) = \{\rho e^{\pm i\varphi} : \underline{\gamma} \leq \rho \leq \bar{\gamma} \leq \infty\}$$

and of the circular arc

$$\underline{Ar}(\gamma) \equiv \underline{Ar}(\gamma, A) = \{z : |z| = \underline{\gamma}, \quad |\arg z| \leq \varphi\}$$

(for an unbounded A) or of two arcs $\underline{Ar}$ and $\overline{Ar} \equiv \overline{Ar}(\gamma, A) = \{z : |z| = \bar{\gamma}, \quad |\arg z| \leq \varphi\}$ (for a bounded A) surrounding a domain Ω_Γ such that $\varphi < \frac{\pi}{2}$, $\Sigma(A) \subset \Omega_\Gamma$ and the estimate

$$\|(z - A)^{-1}\| \leq \frac{M}{1 + |z|}$$

with a positive constant M holds on Γ and outside of Ω_Γ . By c, c_i , $i = 0, 1, \dots$ we will denote various constants independent of a parameter pointed out explicitly (e.g., $n, N, \dots$).

Next, we prove some lemmas which we will need in the sequel.

Lemma 1. *Let A be a densely defined, strongly positive operator and $T_\gamma = (\gamma I - A)(\gamma I + A)^{-1}$ its Cayley transform, then*

$$\|T_\gamma^n\| \leq c_0 q_\gamma^n,$$

provided that A is a bounded operator and

$$\|T_\gamma^n A^{-\sigma}\| \leq c_1 n^{-\sigma+\varepsilon}$$

provided that A is an unbounded operator, where $q_\gamma \in (0,1)$ and ε is an arbitrarily small positive number.

Proof. Let A be a bounded operator then we have (see Fig. 1)

Fig. 1

$$\begin{aligned}
 T_\gamma^n &= \frac{1}{2\pi i} \int_\Gamma \left(\frac{\gamma - z}{\gamma + z} \right)^n (z - A)^{-1} dz, \\
 \|T_\gamma^n\| &= \frac{1}{\pi} \left\| -\text{Im} e^{i\varphi} \int_{\underline{\gamma}}^{\bar{\gamma}} \left(\frac{\gamma - \rho e^{i\varphi}}{\gamma + \rho e^{i\varphi}} \right)^n (\rho e^{i\varphi} - A)^{-1} d\rho \right. \\
 &\quad \left. - \text{Re} i \underline{\gamma} \int_0^\varphi \left(\frac{\gamma - \underline{\gamma} e^{i\theta}}{\gamma + \underline{\gamma} e^{i\theta}} \right)^n (\underline{\gamma} e^{i\theta} - A)^{-1} d\theta \right. \\
 &\quad \left. + \text{Re} i \bar{\gamma} \int_0^\varphi \left(\frac{\gamma - \bar{\gamma} e^{i\theta}}{\gamma + \bar{\gamma} e^{i\theta}} \right)^n (\bar{\gamma} e^{i\theta} - A)^{-1} d\theta \right\|.
 \end{aligned}$$

Let $0 < \gamma < \underline{\gamma}$, then using the inequality

$$\left| \frac{\gamma - \rho e^{i\varphi}}{\gamma + \rho e^{i\varphi}} \right|^2 = \frac{\gamma^2 + \rho^2 - 2\gamma\rho \cos \varphi}{\gamma^2 + \rho^2 + 2\gamma\rho \cos \varphi} \leq \frac{\rho - \gamma \cos \varphi}{\rho + \gamma \cos \varphi}, \quad \rho \geq \gamma$$

we get

$$\begin{aligned}
\|T_\gamma^n\| &\leq \frac{M}{\pi} \left\{ \int_{\underline{\gamma}}^{\bar{\gamma}} \left(\frac{\rho - \gamma \cos \varphi}{\rho + \gamma \cos \varphi} \right)^{\frac{n}{2}} \frac{d\rho}{1 + \rho} \right. \\
&\quad \left. + \frac{\underline{\gamma}}{1 + \underline{\gamma}} \int_0^\varphi \left(\frac{\underline{\gamma} - \gamma \cos \theta}{\underline{\gamma} + \gamma \cos \theta} \right)^{\frac{n}{2}} d\theta + \frac{\bar{\gamma}}{1 + \bar{\gamma}} \int_0^\varphi \left(\frac{\bar{\gamma} - \gamma \cos \theta}{\bar{\gamma} + \gamma \cos \theta} \right)^{\frac{n}{2}} d\theta \right\} \\
&\leq \frac{M}{\pi} \left\{ \left(\frac{\bar{\gamma} - \gamma \cos \varphi}{\bar{\gamma} + \gamma \cos \varphi} \right)^{\frac{n}{2}} \ln \frac{1 + \bar{\gamma}}{1 + \underline{\gamma}} \right. \\
&\quad \left. + \frac{\underline{\gamma}\varphi}{1 + \underline{\gamma}} \left(\frac{\underline{\gamma} - \gamma \cos \varphi}{\underline{\gamma} + \gamma \cos \varphi} \right)^{\frac{n}{2}} + \frac{\bar{\gamma}\varphi}{1 + \bar{\gamma}} \left(\frac{\bar{\gamma} - \gamma \cos \varphi}{\bar{\gamma} + \gamma \cos \varphi} \right)^{\frac{n}{2}} \right\} \\
&\leq c_0 q_\gamma^n
\end{aligned}$$

with

$$\begin{aligned}
q_\gamma &= \left(\frac{\bar{\gamma} - \gamma \cos \varphi}{\bar{\gamma} + \gamma \cos \varphi} \right)^{\frac{1}{2}}, \\
c_0 &= \frac{M}{\pi} \left\{ \ln \frac{1 + \bar{\gamma}}{1 + \underline{\gamma}} + \frac{\underline{\gamma}\varphi}{1 + \underline{\gamma}} + \frac{\bar{\gamma}\varphi}{1 + \bar{\gamma}} \right\}.
\end{aligned}$$

If $\underline{\gamma} < \gamma < \bar{\gamma}$ then

$$\begin{aligned}
\|T_\gamma^n\| &\leq \frac{M}{\pi} \left\{ \int_{\underline{\gamma}}^\gamma \left(\frac{\gamma - \rho \cos \varphi}{\gamma + \rho \cos \varphi} \right)^{\frac{n}{2}} \frac{d\rho}{1 + \rho} + \int_\gamma^{\bar{\gamma}} \left(\frac{\rho - \gamma \cos \varphi}{\rho + \gamma \cos \varphi} \right)^{\frac{n}{2}} \frac{1}{1 + \rho} d\rho \right. \\
&\quad \left. + \frac{\underline{\gamma}\varphi}{1 + \underline{\gamma}} \left(\frac{\gamma - \underline{\gamma} \cos \varphi}{\gamma + \underline{\gamma} \cos \varphi} \right)^{\frac{n}{2}} + \frac{\bar{\gamma}\varphi}{1 + \bar{\gamma}} \left(\frac{\bar{\gamma} - \gamma \cos \varphi}{\bar{\gamma} + \gamma \cos \varphi} \right)^{\frac{n}{2}} \right\} \\
&\leq \frac{M}{\pi} \left\{ \left(\frac{\gamma - \underline{\gamma} \cos \varphi}{\gamma + \underline{\gamma} \cos \varphi} \right)^{\frac{n}{2}} \ln \frac{1 + \gamma}{1 + \underline{\gamma}} + \left(\frac{\bar{\gamma} - \gamma \cos \varphi}{\bar{\gamma} + \gamma \cos \varphi} \right)^{\frac{n}{2}} \ln \frac{1 + \bar{\gamma}}{1 + \gamma} \right. \\
&\quad \left. + \frac{\underline{\gamma}\varphi}{1 + \underline{\gamma}} \left(\frac{\gamma - \underline{\gamma} \cos \varphi}{\gamma + \underline{\gamma} \cos \varphi} \right)^{\frac{n}{2}} + \frac{\bar{\gamma}\varphi}{1 + \bar{\gamma}} \left(\frac{\bar{\gamma} - \gamma \cos \varphi}{\bar{\gamma} + \gamma \cos \varphi} \right)^{\frac{n}{2}} \right\} \\
&\leq c_0 q_\gamma^n,
\end{aligned}$$

where

$$q_\gamma = \max \left\{ \left(\frac{\gamma - \underline{\gamma} \cos \varphi}{\gamma + \underline{\gamma} \cos \varphi} \right)^{\frac{1}{2}}, \left(\frac{\bar{\gamma} - \gamma \cos \varphi}{\bar{\gamma} + \gamma \cos \varphi} \right)^{\frac{1}{2}} \right\}.$$

Finally, when $\bar{\gamma} < \gamma$ then

$$\|T_\gamma^n\| \leq \frac{M}{\pi} \left\{ \left(\frac{\gamma - \underline{\gamma} \cos \varphi}{\gamma + \underline{\gamma} \cos \varphi} \right)^{\frac{n}{2}} \ln \frac{1 + \bar{\gamma}}{1 + \underline{\gamma}} + \frac{\underline{\gamma} \varphi}{1 + \underline{\gamma}} \left(\frac{\gamma - \underline{\gamma} \cos \varphi}{\gamma + \underline{\gamma} \cos \varphi} \right)^{\frac{n}{2}} + \frac{\bar{\gamma} \varphi}{1 + \underline{\gamma}} \left(\frac{\gamma - \bar{\gamma} \cos \varphi}{\gamma + \bar{\gamma} \cos \varphi} \right)^{\frac{n}{2}} \right\} \leq c_0 q_\gamma^n$$

where

$$q_\gamma = \left(\frac{\gamma - \underline{\gamma} \cos \varphi}{\gamma + \underline{\gamma} \cos \varphi} \right)^{\frac{1}{2}}.$$

It is easy to find that

$$g_\gamma \geq q_* = \left(\frac{\sqrt{\bar{\gamma}} - \sqrt{\underline{\gamma}} \cos \varphi}{\sqrt{\bar{\gamma}} + \sqrt{\underline{\gamma}} \cos \varphi} \right)^{\frac{1}{2}} \quad \forall \gamma \in (0, \infty).$$

If A is unbounded then (see Fig. 2)

Fig. 2

$$T_\gamma^n A^{-\sigma} = \frac{1}{2\pi i} \int_\Gamma \left(\frac{\gamma - z}{\gamma + z} \right)^n z^{-\sigma} (z - A)^{-1} dz,$$

$$\|T_\gamma^n A^{-\sigma}\| = \frac{1}{\pi} \left\| -\text{Im} e^{i\varphi} \int_{\underline{\gamma}}^{\infty} \left(\frac{\gamma - \rho e^{i\varphi}}{\gamma + \rho e^{i\varphi}} \right)^n \times (\rho e^{i\varphi})^{-\sigma} (\rho e^{i\varphi} - A)^{-1} d\rho - \text{Re} i \underline{\gamma} \int_0^\varphi \left(\frac{\gamma - \underline{\gamma} e^{i\theta}}{\gamma + \underline{\gamma} e^{i\theta}} \right)^n (\underline{\gamma} e^{i\theta})^{-\sigma} (\underline{\gamma} e^{i\theta} - A)^{-1} e^{i\theta} d\theta \right\|.$$

In the case $0 < \gamma < \underline{\gamma}$ we have

$$\|T_{\gamma}^n A^{-\sigma}\| \leq \frac{M}{\pi} \left\{ \int_{\underline{\gamma}}^{\infty} \left(\frac{\rho - \gamma \cos \varphi}{\rho + \gamma \cos \varphi} \right)^{\frac{n}{2}} \rho^{-\sigma} \frac{1}{1 + \rho} d\rho + \frac{\gamma^{1-\sigma} \varphi}{1 + \underline{\gamma}} \left(\frac{\underline{\gamma} - \gamma \cos \varphi}{\underline{\gamma} + \gamma \cos \varphi} \right)^{\frac{n}{2}} \right\}.$$

The function

$$\Psi(\tilde{\rho}) = \left(\frac{\tilde{\rho} - \tilde{\gamma} \cos \varphi}{\tilde{\rho} + \tilde{\gamma} \cos \varphi} \right)^k \tilde{\rho}^{-\tau}, \quad \tilde{\rho} \geq \tilde{\gamma}, \quad \tau > 0$$

satisfies

$$\max_{\tilde{\rho} \in [\tilde{\gamma}, \infty)} \Psi(\tilde{\rho}) = \left(\frac{\tau^{-1}(k + \sqrt{k^2 + \tau^2}) - 1}{\tau^{-1}(k + \sqrt{k^2 + \tau^2}) + 1} \right)^k \times \left(\frac{\gamma \cos \varphi}{\tau} (k + \sqrt{k^2 + \tau^2}) \right)^{-\tau} \leq ck^{-\tau},$$

where

$$c \equiv c(\tilde{\gamma}, \tau, \varphi) = \left(\frac{\tau}{\tilde{\gamma} \cos \varphi} \right)^{\tau}.$$

Thus, we get

$$\begin{aligned} \|T_{\gamma}^n A^{-\sigma}\| &\leq \frac{M}{\pi} \left\{ c \left(\frac{n}{2} \right)^{-\sigma+\varepsilon} \int_{\underline{\gamma}}^{\infty} \rho^{-1-\varepsilon} d\rho + \frac{\gamma^{1-\sigma} \varphi}{1 + \underline{\gamma}} \left(\frac{\underline{\gamma} - \gamma \cos \varphi}{\underline{\gamma} + \gamma \cos \varphi} \right)^{\frac{n}{2}} \right\} \\ &= \frac{M}{\pi} \left\{ \frac{c(\underline{\gamma}, \sigma, \varphi) 2^{\sigma-\varepsilon}}{\varepsilon \underline{\gamma}^{\varepsilon}} n^{-\sigma+\varepsilon} + \frac{\gamma^{1-\sigma} \varphi}{1 + \underline{\gamma}} \left(\frac{\underline{\gamma} - \gamma \cos \varphi}{\underline{\gamma} + \gamma \cos \varphi} \right)^{\frac{n}{2}} \right\} \end{aligned}$$

and for n large enough

$$\|T_{\gamma}^n A^{-\sigma}\| \leq c_{11} n^{-\sigma+\varepsilon},$$

where

$$c_{11} \equiv c_{11}(\underline{\gamma}, \sigma, \varphi) = \frac{M}{\pi} \left\{ \frac{c 2^{\sigma-\varepsilon}}{\varepsilon \underline{\gamma}^{\varepsilon}} + \frac{\varphi \underline{\gamma}^{1-\sigma}}{1 + \underline{\gamma}} \right\}.$$

If $\gamma > \underline{\gamma} > 0$ then analogously

$$\begin{aligned} \|T_{\gamma}^n A^{-\sigma}\| &\leq \frac{M}{\pi} \left\{ \int_{\underline{\gamma}}^{\gamma} \left(\frac{\gamma - \rho \cos \varphi}{\gamma + \rho \cos \varphi} \right)^{\frac{n}{2}} \rho^{-\sigma} \frac{1}{1 + \rho} d\rho \right. \\ &\quad \left. + \int_{\gamma}^{\infty} \left(\frac{\rho - \gamma \cos \varphi}{\rho + \gamma \cos \varphi} \right)^{\frac{n}{2}} \rho^{-\sigma} \frac{1}{1 + \rho} d\rho + \frac{\gamma^{1-\sigma} \varphi}{1 + \underline{\gamma}} \left(\frac{\gamma - \underline{\gamma} \cos \varphi}{\gamma + \underline{\gamma} \cos \varphi} \right)^{\frac{n}{2}} \right\} \\ &\leq \frac{M}{\pi} \left\{ \left(\frac{\gamma - \underline{\gamma} \cos \varphi}{\gamma + \underline{\gamma} \cos \varphi} \right)^{\frac{n}{2}} \left(\frac{1}{\sigma \underline{\gamma}^{\sigma}} - \frac{1}{\sigma \gamma^{\sigma}} \right) + \frac{c(\gamma, \sigma, \varphi) 2^{\sigma-\varepsilon}}{\varepsilon \gamma^{\varepsilon}} n^{-\sigma+\varepsilon} \right. \\ &\quad \left. + \frac{\gamma^{1-\sigma} \varphi}{1 + \underline{\gamma}} \left(\frac{\gamma - \underline{\gamma} \cos \varphi}{\gamma + \underline{\gamma} \cos \varphi} \right)^{\frac{n}{2}} \right\} \\ &\leq c_{12} n^{-\sigma+\varepsilon}, \end{aligned}$$

$$c_{12} = \frac{M}{\pi} \left\{ \frac{1}{\sigma \underline{\gamma}^\sigma} - \frac{1}{\sigma \gamma^\sigma} + \frac{c 2^{\sigma-\varepsilon}}{\varepsilon \gamma^\varepsilon} + \frac{\gamma^{1-\sigma} \varphi}{1 + \underline{\gamma}} \right\}.$$

Choosing $c_1 = \max_i c_{1i}$ we get the assertion of the lemma.

The next lemma shows that the estimates of Lemma 1 can be slightly improved if A is a densely defined, selfadjoint positive definite operator in a Hilbert space.

Lemma 2. *Let A be a densely defined, positive definite operator in a Hilbert space H , then*

$$\|T_\gamma^n\| \leq q_{\text{opt}}^n \leq q_\gamma^n, \quad q_\gamma \equiv q(\gamma) = \max \left\{ \left| \frac{\gamma - \lambda_0}{\gamma + \lambda_0} \right|, \left| \frac{\Lambda_0 - \gamma}{\Lambda_0 + \gamma} \right| \right\},$$

$$q_{\text{opt}} = \inf_{\gamma > 0} q(\gamma) = q(\lambda_{\text{opt}}) = \frac{\sqrt{\Lambda_0} - \sqrt{\lambda_0}}{\sqrt{\Lambda_0} + \sqrt{\lambda_0}}, \quad \lambda_{\text{opt}} = \sqrt{\lambda_0 \Lambda_0}$$

provided that A is a bounded operator, $\Sigma(A) \in [\lambda_0, \Lambda_0]$, $\lambda_0 > 0$, and

$$\|T_\gamma^n A^{-\sigma}\| \leq c_2 n^{-\sigma}, \quad \sigma > 0$$

provided that A is an unbounded operator, $\Sigma(A) \in [\lambda_0, \infty)$.

Proof. Let A be bounded. Then, using the operator calculus we get

$$\|T_\gamma^n\| = \left\| \int_{\lambda_0}^{\Lambda_0} \left(\frac{\gamma - \lambda}{\gamma + \lambda} \right)^n dE_\lambda \right\| \leq q_{\text{opt}}^n \leq q^n(\gamma).$$

Here, E_λ denotes the spectral family corresponding to A .

If A is unbounded then for n large enough

$$\|T_\gamma^n A^{-\sigma}\| = \left\| \int_{\lambda_0}^{\infty} \left(\frac{\gamma - \lambda}{\gamma + \lambda} \right)^n \lambda^{-\sigma} dE_\lambda \right\| \leq \max_{\lambda \in [\lambda_0, \infty)} \left| \frac{\gamma - \lambda}{\gamma + \lambda} \right|^n \lambda^{-\sigma} \leq c_2 n^{-\sigma},$$

where

$$c_2 = \left(\frac{\sigma}{\lambda_0} \right)^\sigma.$$

Lemma 3. *Let α_i , $1 \leq i \leq n$ and p be natural numbers, and $\sigma > 1$ be a real number. Then it holds*

$$\sum_{\alpha_1 + \dots + \alpha_n = p} (\alpha_1 \cdots \alpha_n)^{-\sigma} \leq [2^\sigma \zeta(\sigma)]^{n-1} p^{-\sigma} \quad \forall p \geq n$$

where $\zeta(\sigma)$ is Riemann's zeta-function.

Proof. We shall prove the inequality by induction over n . The statement is obviously true for $n = 1$. Let it hold for $n = k$, then we get for $n = k + 1$

$$\begin{aligned}
\sum_{\alpha_1 + \dots + \alpha_{k+1} = p} (\alpha_1 \dots \alpha_{k+1})^{-\sigma} &= \sum_{\alpha_{k+1}=1}^{p-k} \alpha_{k+1}^{-\sigma} \sum_{\alpha_1 + \dots + \alpha_k = p - \alpha_{k+1}} (\alpha_1 \dots \alpha_k)^{-\sigma} \\
&\leq \sum_{\alpha_{k+1}=1}^{p-k} \alpha_{k+1}^{-\sigma} [2^\sigma \zeta(\sigma)]^{k-1} (p - \alpha_{k+1})^{-\sigma} \\
&= [2^\sigma \zeta(\sigma)]^{k-1} p^{-\sigma} \sum_{\alpha_{k+1}=1}^{p-k} [\alpha_{k+1}^{-1} + (p - \alpha_{k+1})^{-1}]^\sigma
\end{aligned}$$

Using Höldes inequality one gets

$$\begin{aligned}
\sum_{\alpha_{k+1}=1}^{p-k} [\alpha_{k+1}^{-1} + (p - \alpha_{k+1})^{-1}]^\sigma &\leq \sum_{\alpha_{k+1}=1}^{p-k} 2^{\sigma-1} [\alpha_{k+1}^{-\sigma} + (p - \alpha_{k+1})^{-\sigma}] \\
&= 2^\sigma \sum_{\alpha_{k+1}=1}^{p-k} \alpha_{k+1}^{-\sigma} \leq 2^\sigma \zeta(\sigma),
\end{aligned}$$

from where the statement of the lemma follows.

3. Linear equation, exact solution and its approximation

In what follows we assume that the operators U_j, V_j in (3) are strongly positive. In order to get an explicit representation of the solution of (3) we set in (4)

$$\mu_j = \gamma_j \frac{1 + z_j}{1 - z_j}, \quad \lambda_j = \delta_j \frac{1 + \zeta_j}{1 - \zeta_j}, \quad (9)$$

where $\gamma_j > 0, \delta_j > 0$ are arbitrary numbers. One gets from (9)

$$z_j = -\frac{\gamma_j - \mu_j}{\gamma_j + \mu_j}, \quad \zeta_j = -\frac{\delta_j - \lambda_j}{\delta_j + \lambda_j}. \quad (10)$$

We suppose that the following expansion is valid:

$$P^{-1}(\lambda, \mu) \equiv \left(\sum_{j=1}^M \mu_j \lambda_j \right)^{-1} = \sum_{|m|, |n|=0}^{\infty} d_{mn} z^m \zeta^n, \quad (11)$$

where $m = (m_1, \dots, m_M)$, $n = (n_1, \dots, n_M)$, $|m| = \sum_{i=1}^M m_i$, $|n| = \sum_{i=1}^M n_i$, $z^m = z_1^{m_1} \dots z_M^{m_M}$; $|z_j| \leq 1$, $|\zeta_j| \leq 1$,

$$d_{mn} = \left(\prod_{i=1}^M (m_i)!(n_i)! \right)^{-1} \frac{\partial^{|m|+|n|}}{\partial z^m \partial \zeta^n} \left(\sum_{j=1}^M \mu_j \lambda_j \right)^{-1} \Big|_{z=0, \zeta=0}.$$

Note that the numbers d_{mn} can be found also by comparison of the coefficients in (11). Substituting (11) into (5) and taking into account (9), (10) we get

$$X = \sum_{|m|, |n|=0}^{\infty} (-1)^{|m|+|n|} d_{mn} \prod_{j=1}^M T_{\gamma_j}^{m_j}(U_j) Y \prod_{j=1}^M T_{\delta_j}^{n_j}(V_j) \quad (12)$$

where

$$T_{\gamma_j}(U_j) = (\gamma_j I - U_j)(\gamma_j I + U_j)^{-1}$$

$$T_{\delta_j}(V_j) = (\delta_j I - V_j)(\delta_j I + V_j)^{-1}$$

are the Cayley transforms of the operators U_j and V_j respectively.

It follows from Lemma 1 that for bounded operators U_j , V_j there exist positive constants $q_j < 1$, $r_j < 1$ such that

$$\|T_{\gamma_j}^{m_j}(U_j)\| \leq c q_j^{m_j}, \quad \|T_{\delta_j}^{n_j}(V_j)\| \leq c r_j^{n_j}, \quad 1 \leq j \leq M. \quad (13)$$

As an approximation for the exact solution (12) we take the truncated sum

$$X_N = \sum_{p=0}^N (-1)^p \sum_{|m|+|n|=p} d_{mn} \prod_{j=1}^M T_{\gamma_j}^{m_j}(U_j) Y \prod_{j=1}^M T_{\delta_j}^{n_j}(V_j). \quad (14)$$

Using (13) and the estimate $|d_{mn}| \leq D \forall m, n$ one gets the inequality

$$\|X - X_N\| \leq cD \|Y\| \sum_{p=N+1}^{\infty} \sum_{|m|+|n|=p} \prod_{j=1}^M q_j^{m_j} r_j^{n_j}.$$

Let $q = \max_j(q_j, r_j)$, then we get

$$\|X - X_N\| \leq c(M) \|Y\| \sum_{p=N+1}^{\infty} p^{2M-1} q^p.$$

The function $f(x) = x^{2M-1} q^x$ decreases for $x > N$ large enough, hence

$$\sum_{p=N+1}^{\infty} p^{2M-1} q^p < \int_N^{\infty} x^{2M-1} q^x dx \equiv I_{2M-1}.$$

Using partial integration one gets

$$I_n = -N^{-n} \frac{q^N}{\ln q} - \frac{n}{\ln q} I_{n-1}, \quad I_0 = -\frac{q^N}{\ln q}$$

what yields

$$I_n \leq cN^n q^N,$$

where c is a constant independent of N .

Hence, we get

$$\|X - X_N\| \leq c\|Y\|N^{2M-1}q^N \quad (15)$$

with a constant c independent of N .

We have proved the following statement.

Theorem 1. *Let $\{U_j\}$, $\{V_j\}$ be commutative subsets of densely defined, bounded, strongly positive operators in a Banach space E and assume that the expansion (11) is valid. Then, the exact solution of the equation (3) for an arbitrary Y can be represented by the formula (12), and for the approximate solution (14) the estimate (15) holds.*

For unbounded operators U_j , V_j the following statement holds true.

Theorem 2. *Let $\{U_j\}$, $\{V_j\}$ be commutative subsets of unbounded, densely defined, strongly positive operators in a Banach space E and assume that the expansion (11) is valid. Then, the exact solution of the equation (3) can be represented by the formula (12) and for the approximate solution (14) the estimate*

$$\|X - X_N\| \leq cN^{-\sigma+1+\varepsilon}\|\tilde{Y}\| \quad (16)$$

holds provided that $\tilde{Y} \equiv (\prod_{j=1}^M U_j^{\sigma_j})Y(\prod_{j=1}^M V_j^{\theta_j})$ is a bounded operator for some positive σ_j , θ_j , where $\sigma = \min(\sigma_j, \theta_j)$, c is a constant independent of N , ε is an arbitrarily small positive number.

Proof. Due to Lemma 1 we get

$$\begin{aligned} \|X - X_N\| &= \left\| \sum_{p=N+1}^{\infty} (-1)^p \sum_{|m|+|n|=p} d_{mn} \left(\prod_{j=1}^M T_{\gamma_j}^{m_j}(U_j) U_j^{-\sigma_j} \right) \left(\prod_{j=1}^M U_j^{\sigma_j} \right) Y \right. \\ &\quad \times \left. \left(\prod_{j=1}^M V_j^{\theta_j} \right) \left(\prod_{j=1}^M T_{\delta_j}^{n_j}(V_j) V_j^{-\theta_j} \right) \right\| \\ &\leq c\|\tilde{Y}\| \sum_{p=N+1}^{\infty} \sum_{|m|+|n|=p} \prod_{j=1}^M \left(m_j^{-\sigma_j+\varepsilon} n_j^{-\theta_j+\varepsilon} \right) \\ &\leq c\|\tilde{Y}\| \sum_{p=N+1}^{\infty} \sum_{|m|+|n|=p} \prod_{j=1}^M \left(m_j n_j \right)^{-\sigma+\varepsilon}, \end{aligned}$$

where $\sigma = \min_j(\sigma_j, \theta_j)$. Using Lemma 3 one gets

$$\|X - X_N\| \leq c \|\tilde{Y}\| [2^\sigma \zeta(\sigma)]^{2M-1} \sum_{p=N+1}^{\infty} p^{-\sigma+\varepsilon} \leq cN^{-\sigma+1+\varepsilon} \|\tilde{Y}\|,$$

q.e.d.

The estimate (16) can be improved for densely defined, selfadjoint positive definite operators U_j, V_j .

Theorem 3. *Let $\{U_j\}, \{V_j\}$ be commutative subsets of unbounded, densely defined, selfadjoint, positive definite operators in a Hilbert space H . Then, the exact solution of the equation (3) can be represented by the formula (12), and for the approximate solution (14) the estimate*

$$\|X - X_N\| \leq cN^{-\sigma+1} \quad (17)$$

holds provided that $\tilde{Y} \equiv (\prod_{j=1}^M U_j^{\sigma_j}) Y (\prod_{j=1}^M V_j^{\theta_j})$ is a bounded operator for some positive σ_j, θ_j , where $\sigma = \min(\sigma_j, \theta_j)$, c is a constant independent of N .

The Proof of Theorem 3 is analogous to that of Theorem 2 using Lemma 2 instead of Lemma 1.

Example 1. Let us consider the equation

$$UX + XV = Y, \quad (18)$$

where $U : B_2 \rightarrow B_2$, $V : B_1 \rightarrow B_1$ are known strongly positive operators, Y is a given operator from B_1 into B_2 , B_i ($i = 1, 2$) are some Banach spaces and $X : B_1 \rightarrow B_2$ is the operator to be found. In this case we have

$$\begin{aligned} P^{-1}(\lambda, \mu) &= (\lambda + \mu)^{-1} \\ &= \left(\gamma_1 \frac{1+z_1}{1-z_1} + \gamma_2 \frac{1+z_2}{1-z_2} \right)^{-1} \\ &= \frac{1 - z_1 - z_2 + z_1 z_2}{(\gamma_1 + \gamma_2) + (\gamma_1 - \gamma_2)z_1 + (\gamma_2 - \gamma_1)z_2 - (\gamma_1 + \gamma_2)z_1 z_2}, \end{aligned}$$

where $\gamma_1 > 0$, $\gamma_2 > 0$, $|z_1| < 1$, $|z_2| < 1$. Choosing $\gamma_1 = \gamma_2$ we get

$$\begin{aligned} P^{-1}(\lambda, \mu) &= \frac{1 - z_1 - z_2 + z_1 z_2}{2\gamma(1 - z_1 z_2)} \\ &= \frac{1}{2\gamma} \left\{ \sum_{k=0}^{\infty} (z_1 z_2)^k - \sum_{k=0}^{\infty} z_1^{k+1} z_2^k - \sum_{k=0}^{\infty} z_1^k z_2^{k+1} + \sum_{k=0}^{\infty} (z_1 z_2)^{k+1} \right\} \\ &= \sum_{p=0}^{\infty} \sum_{m+n=p} d_{mn} z_1^m z_2^n. \end{aligned} \quad (19)$$

It is easy to see that $d_{00} = 1$, $d_{10} = d_{01} = -1$, $d_{m0} = d_{0n} = 0$ ($m, n > 1$), $d_{mn} = d_{m-1, n-1}$ ($m > 1, n > 1$), $\sum_{m+n=p} d_{mn} = 2(-1)^p$, $p \geq 1$. It is convenient to arrange d_{mn} as the following table:

		p $\longrightarrow$					
			0	1	2	3	4
n \ m		0	1	2	3	4	
0		1	-1	0	0	0	
1		-1	2	-1	0	0	
2		0	-1	2	-1	0	
3		0	0	-1	2	-1	
4		0	0	0	-1	2	

Table 1: The coefficients $d_{mn} \cdot 2\gamma$.

Thus, we obtain the following representation of the unique solution of (18):

$$X = \frac{1}{2\gamma} \left\{ Y + \hat{Y} + 2T_\gamma(U)YT_\gamma(V) + T_\gamma(U)\hat{Y}T_\gamma(V) + 2T_\gamma^2(U)YT_\gamma^2(V) + T_\gamma^2(U)\hat{Y}T_\gamma^2(V) + \dots \right\}, \quad (20)$$

where $\hat{Y} = T_\gamma(U)Y + YT_\gamma(V)$.

Let X_N be the truncated sum consisting of the $(N + 1)$ first terms. Then in the case $U \in \mathcal{L}(B_2)$, $V \in \mathcal{L}(B_1)$, $Y \in \mathcal{L}(B_1, B_2)$ we get

$$\|X - X_N\| \leq 2\|Y\| (q^{N+1} + q^{N+2} + \dots) = 2\|Y\| \frac{q^{N+1}}{1 - q} \quad (21)$$

where $q = \max\{\|T_\gamma(U)\|, \|T_\gamma(V)\|\} < 1$.

Let us consider an example of the equation (18), where U, V are unbounded operators in $L^2(0, 1)$ defined by

$$D(U) = D(V) = \{w : w \in H^2(0, 1) \cap H_0^1(0, 1)\},$$

$$Uw = -\frac{d^2w(s)}{ds^2} + q_1(s)w(s), \quad Vw = -\frac{d^2w(s)}{ds^2} + q_2(s)w(s) \quad \forall w \in D(U)$$

with various non-negative functions q_1, q_2 so that U and V do not commute. We denote by $G(s, t; q_i)$, $i = 1, 2$ the Green functions of U and V respectively and suppose that $U^\sigma Y V^\sigma$ is bounded in $L^2(0, 1)$ (for example, if we define Y by

$$D(Y) = D(U) = D(V) ,$$

$$\begin{aligned} Yz = & \int_0^1 G(s, t_1; q_1) \int_0^1 \left[G(t_1, t_2; 0) \int_0^1 G(t_2, t_3; q_2) + G(t_1, t_2; q_1) \int_0^1 G(t_2, t_3; 0) \right] \\ & \times \int_0^1 G(t_3, t_4; q_2) z(t_4) dt_4 dt_3 dt_2 dt_1 \end{aligned}$$

then

$$U^2 Y V^2 w = 2w(s) + q_1 \int_0^1 G(s, t_1; 0) w(t_1) dt_1 + \int_0^1 G(s, t_1; 0) q_2(t_1) w(t_1) dt_1$$

is bounded, i.e. $\sigma = 2$. The only solution X of (18) in this case with $D(X) = D(Y)$ is the operator defined by

$$\begin{aligned} Xw = & \int_0^1 G(s, t_1; q_1) \int_0^1 G(t_1, t_2; q_1) \int_0^1 G(t_2, t_3; 0) \int_0^1 G(t_3, t_4; q_2) \\ & \times \int_0^1 G(t_4, t_5; q_2) w(t_5) dt_5 dt_4 dt_3 dt_2 dt_1 . \end{aligned}$$

The solution of (18) can be represented by (20), where

$$T_\gamma(U)w(s) = -w(s) + \gamma \int_0^1 G(s, t; q_1 + \gamma) w(t) dt ,$$

$$T_\gamma(V)w(s) = -w(s) + \gamma \int_0^1 G(s, t; q_2 + \gamma) w(t) dt .$$

If we choose, for example,

$$q_1(s) \equiv 0, \quad q_2(s) = \begin{cases} c_1^2, & 0 < s < 0,5 \\ c_2^2, & 0,5 < s < 1, \end{cases} \quad c_i > 0, \quad c_1 \neq c_2$$

then the Green functions mentioned above are

$$G(s, t; q_1) = \begin{cases} (1-s)t, & t \leq s \\ (1-t)s, & t \geq s, \end{cases}$$

$$G(s, t; q_1 + \gamma) = -\frac{1}{\sqrt{\gamma} \sinh \sqrt{\gamma}} \begin{cases} \sinh \sqrt{\gamma} s \cdot \sinh \sqrt{\gamma} (1-t), & s \leq t \\ \sinh \sqrt{\gamma} t \cdot \sinh \sqrt{\gamma} (1-s), & s \geq t, \end{cases}$$

$$G(s, t; q_2) = \frac{1}{B \sinh c_2 + C \cosh c_2} \begin{cases} v_1(s)v_2(t), & s \leq t \\ v_1(t), v_2(t), & s \geq t, \end{cases}$$

$$v_1(s) = \begin{cases} D \sinh c_1 s, & 0 < s < 0.5 \\ B \sinh c_2 s + C \cosh c_2 s, & 0.5 < s < 1, \end{cases}$$

$$B = -\frac{1}{c_1 c_2} \left(c_2 \sinh \frac{c_2}{2} \sinh \frac{c_1}{2} - c_1 \cosh \frac{c_2}{2} \cosh \frac{c_1}{2} \right),$$

$$C = \frac{1}{c_1 c_2} \left(c_2 \cosh \frac{c_2}{2} \sinh \frac{c_1}{2} - c_1 \cosh \frac{c_1}{2} \sinh \frac{c_2}{2} \right),$$

$$D = c_1^{-1},$$

$$v_2(s) = \begin{cases} c_2^{-1} \sinh c_2 (1-s), & 0.5 < s < 1 \\ E \sinh c_1 s + F \cosh c_1 s, & 0 < s < 0.5, \end{cases}$$

$$E = -\frac{1}{c_1 c_2} \left(c_1 \sinh \frac{c_1}{2} \sinh \frac{c_2}{2} - c_2 \cosh \frac{c_1}{2} \cosh \frac{c_2}{2} \right),$$

$$F = \frac{1}{c_1 c_2} \left(c_1 \cosh \frac{c_1}{2} \sinh \frac{c_2}{2} - c_2 \sinh \frac{c_1}{2} \cosh \frac{c_2}{2} \right).$$

Theorem 3 yields in our case:

$$\|X - X_N\| \leq cN^{-\sigma+1}. \quad (22)$$

Note, that the terms of the series (20) can be computed by computer algebra tools if the analytical expressions for the Green functions exist.

4. A quadratic operator equation

In this section we consider the operator equation

$$Z^2 + A_1 Z + A_2 = 0 \quad (23)$$

in a complex Hilbert space H where the given operators A_1, A_2 are densely defined, closed and satisfy

$$A_1 = A_1^* \geq 0, \quad A_2 = A_2^* \quad (24)$$

$$4A_2 - A_1^2 > 4\omega^2 > 0 \quad (25)$$

(for unbounded operators (23) means that $Z^2 x + A_1 Z x + A_2 x = 0 \forall x \in D(A_2) \cap D(Z)$).

We set

$$B = \frac{i}{2} \sqrt{4A_2 - A_1^2}, \quad i = \sqrt{-1}, \quad -iB \geq -\omega I, \quad \omega < 0, \quad (26)$$

and look for solutions of (23) of the form

$$Z = -\frac{1}{2}A_1 + B + C \quad (27)$$

where C is an unknown bounded operator. Substituting (27) in (23) we find that C has to satisfy the equation

$$C^2 + UC + CV + W = 0, \quad (28)$$

where

$$\begin{aligned} U &= \frac{1}{2}A_1 + B, \quad V = -\frac{1}{2}A_1 + B, \\ W &= \frac{1}{2}[A_1, B] = \frac{1}{2}(A_1 B - B A_1). \end{aligned} \quad (29)$$

Let us transform (28) to the form

$$F(D) = 0 \quad (30)$$

with

$$F(D) = D^2 - iUD - DiV - W, \quad D = -iC \quad (31)$$

and we use the modified Newton iteration procedure

$$F'(0)\Delta_n = -F(D_n) \quad (32)$$

where

$$\begin{aligned} \Delta_n &= D_{n+1} - D_n, \\ F'(0)\Delta_n &\equiv -L_{iU, iV} \Delta_n = -iU \Delta_n - \Delta_n iV. \end{aligned} \quad (33)$$

We get from (32)

$$L_{iU, iV} D_{n+1} = E_n, \quad (34)$$

where $E_n = D_n^2 - W$. It is easy to see that iA_1 is a skew-hermitian operator and iB is self-adjoint and positive definite. Using (24), (25) we get

$$\operatorname{Re}(iUx, x) = \operatorname{Re} \left(-\sqrt{4A_2 - A_1^2} x, x \right) \leq \omega \|x\|^2, \quad (35)$$

$$\operatorname{Re}(iVx, x) \leq \omega \|x\|^2, \quad \omega < 0$$

and the operators iU , iV generate the contractive semigroups $T_{iU}(t)$, $T_{iV}(t)$ such that [19]

$$\|T_{iU}(t)\| \leq e^{\omega t}, \quad \|T_{iV}(t)\| \leq e^{\omega t}. \quad (36)$$

The inverse operator $L_{iU, iV}^{-1}$ is defined by

$$L_{iU, iV}^{-1} D = - \int_0^\infty T_{iU}(t) D T_{iV}(t) dt, \quad (37)$$

and we get from (36)

$$\|L_{iU, iV}^{-1}\| \leq \frac{1}{2|\omega|}. \quad (38)$$

The next lemma states some properties of the solutions of (30), (32).

Lemma 4. *Let $D_0 = D_0^*$ be an arbitrary operator and C be the solution of (30), then $D_n = D_n^*$ for all $n = 1, 2, \dots$ and $\lim_{n \rightarrow \infty} iD_n = C$ provided that the sequence $\{D_n\}$ converges.*

Proof. Let us assume that $D_n = D_n^*$. Then it follows from (34)

$$(L_{iU, iV} D_{n+1})^* = -D_{n+1}^* iU^* - iV^* D_{n+1}^* = E_n^*.$$

We have $W^* = W$,

$$U^* = \frac{1}{2}A_1 + B^* = \frac{1}{2}A_1 - B = -V,$$

$$V^* = -\frac{1}{2}A_1 + B^* = -\frac{1}{2}A_1 - B = -U,$$

$$E_n^* = (D_n^*)^2 - W^* = E_n.$$

Then it holds

$$D_{n+1}^* iV + iU D_{n+1}^* = E_n,$$

i.e. $D_{n+1} = D_{n+1}^*$. Passing to the limit in (34) as $n \rightarrow \infty$ we obtain

$$UC + CV = -C^2 - W,$$

what coincides with (28) and completes the proof of Lemma 4.

Note that the modified Newton method is equivalent to the fixed point iteration for the equation

$$D = \Phi_{iU,iV}(D), \quad (39)$$

where

$$\Phi_{iU,iV}(D) = L_{iU,iV}^{-1}(D^2 - W). \quad (40)$$

Sufficient conditions for the convergence of the fixed point iteration

$$D_{n+1} = \Phi_{iU,iV}(D_n) \quad (41)$$

are

$$\|\Phi_{iU,iV}(D^{(1)}) - \Phi_{iU,iV}(D^{(2)})\| \leq q \|D^{(1)} - D^{(2)}\|, \quad 0 < q < 1 \quad (42)$$

for any $D^{(1)}, D^{(2)}$ from some closed ball $R_r \equiv \{D : \|D\| \leq r\}$ and

$$\|D_n\| \leq r \quad \forall n = 0, 1, 2, \dots \quad (43)$$

Using (38), we obtain

$$\begin{aligned} & \left\| \Phi(D^{(1)}) - \Phi(D^{(2)}) \right\| \\ & \leq \frac{1}{2|\omega|} \left\| (D^{(1)})^2 - (D^{(2)})^2 \right\| \\ & = \frac{1}{2|\omega|} \left\| (D^{(1)} + D^{(2)}) (D^{(1)} - D^{(2)}) - D^{(1)} (D^{(1)} - D^{(2)}) - (D^{(2)} - D^{(1)}) D^{(1)} \right\| \\ & \leq \frac{1}{2|\omega|} (3\|D^{(1)}\| + \|D^{(2)}\|) \|D^{(1)} - D^{(2)}\| \end{aligned} \quad (44)$$

what yields that conditions (42), (43) hold true if we choose $r = \frac{|\omega|q}{2}$ with an arbitrary $q \in (0, 1)$ and if

$$\frac{1}{2|\omega|} \|D^2 - W\| \leq \frac{|\omega|q}{2} \quad \forall D \in R_{\frac{|\omega|q}{2}}.$$

The last inequality is fulfilled if

$$\frac{1}{2|\omega|} \left(\frac{|\omega|^2 q^2}{4} + \frac{1}{2} \|[A_1, B]\| \right) \leq \frac{|\omega|q}{2}$$

what leads to the following sufficient condition on the commutator $[A_1, B]$:

$$d \equiv \|[A_1, B]\| \leq 2\omega^2 q \left(1 - \frac{q}{4} \right) \equiv d_0. \quad (45)$$

Thus, we have proved the following result.

Theorem 4. *If A_1, A_2 satisfy assumptions (24), (25), (45) then for $D_0 = D_0^*$ it holds $D_n = D_n^* \forall n = 1, 2, \dots$, and the equation (23) has a root of the form (27) with $C = iD$, and $D = \lim_{n \rightarrow \infty} D_n$ lies in the closed ball $R_{\frac{|\omega|q}{2}}$.*

We look for another solution of (23) in the form

$$\hat{Z} = -\frac{1}{2}A_1 - B - \hat{C}, \quad (46)$$

where $\hat{C}$ satisfies the equation

$$\hat{C}^2 + V\hat{C} + \hat{C}U - \frac{1}{2}[A_1, B] = 0. \quad (47)$$

To solve (47) we use the modified Newton method

$$\hat{F}'(0)\hat{\Delta}_n = -\hat{F}(\hat{D}_n), \quad (48)$$

where $\hat{D}_n = -i\hat{C}_n$, $\hat{\Delta}_n = \hat{D}_{n+1} - \hat{D}_n$, $\hat{F}'(0)\hat{\Delta}_n \equiv -L_{iV, iU}\hat{\Delta}_n = -iV\hat{\Delta}_n - \hat{\Delta}_n iU$, $\hat{F}(\hat{D}_n) = \hat{D}_n^2 - L_{iV, iU}D_n + W$.

In order to find $\hat{D}_{n+1}$ one has to solve the following equation in each step of the iteration process

$$L_{iV, iU}\hat{D}_{n+1} = \hat{E}_n, \quad (49)$$

where

$$\hat{E}_n = \hat{D}_n^2 + W.$$

Proceeding as done for the proof of Theorem 4 we arrive at the following statement.

Theorem 5. *Under the assumptions of Theorem 4 the equation (23) has a root of the form (46) with $\hat{C} = i\hat{D}$, where $\hat{D} = \lim_{n \rightarrow \infty} \hat{D}_n$ ($\hat{D}_0 = \hat{D}_0^*$) lies in the closed ball $R_{\frac{|\omega|q}{2}}$.*

5. An abstract second order differential equation with non-commutative coefficients

Let us consider the problem

$$\frac{d^2 u}{dt^2} + A_1 \frac{du}{dt} + A_2 u = 0, \quad u(0) = u_0, \quad u'(0) = u_1 \quad (50)$$

where $u : \mathbb{R}_+ \rightarrow H$ is a vector-valued function, A_1, A_2 are densely defined, closed operators in a Hilbert space which satisfy

$$\begin{aligned} D(A_1) &\subseteq D(A_2), \\ A_1 &= A_1^* \geq 0, \quad A_2 = A_2^*, \quad B^2 = A_1^2 - 4A_2 < -4\omega^2 < 0. \end{aligned} \quad (51)$$

For $A_1 = 0$, $A_2 > 0$ the problem (50) was considered in [12, 13] where the explicit solution and an algorithmical approximation has been found.

We look for the solutions of (50) in the form

$$u_1(t) = \exp Zt, \quad u_2(t) = \exp \hat{Z}t, \quad (52)$$

where $Z = -\frac{1}{2}A_1 + B + C$, $\hat{Z} = -\frac{1}{2}A_1 - B - \hat{C}$ and C , $\hat{C}$ are bounded, skew-symmetric operators. Substituting (52) into (50) we get the operator equations

$$C^2 + UC + CV + \frac{1}{2}[A_1, B] = 0,$$

$$\hat{C}^2 + VC + \hat{C}U - \frac{1}{2}[A_1, B] = 0,$$

where

$$U = \frac{1}{2}A_1 + B, \quad V = -\frac{1}{2}A_1 + B, \quad [A_1, B] = A_1B - BA_1.$$

It follows from Theorem 4 and Theorem 5 that these equations have solutions in the ball $R_{\frac{|\omega|q}{2}}$, $0 < q < 1$ provided that $\|[A_1, B]\| \leq 2\omega^2q(1 - \frac{q}{4})$ which can be found by the method described in Section 4. It is easy to see that if the operators A_1, A_2 commute, i.e. $[A_1, B] = 0$, then $C = \hat{C} = 0$ and we get the solution of (50) in the form

$$u(t) = \exp\left(-\frac{1}{2}A_1t\right) \left\{ \left(\cos iBt - \frac{i}{2}B^{-1}A_1 \sin iBt\right) u_0 - iB^{-1} \sin iBtu_1 \right\}. \quad (53)$$

There are algorithmical representations for fractional powers of operators, for the operator exponential, cos-, and sin-functions and their approximations with accuracy depending on the regularity of u_0 and u_1 [11, 12, 13].

If A_1 and A_2 do not commute then we look for the solution in the form

$$u(t) = (\exp Zt)a + (\exp \hat{Z}t)b \quad (54)$$

where a, b are unknown vectors yet. From the initial conditions (50) we get the following system of operator equations

$$\begin{cases} a + b = u_0 \\ Za + \hat{Z}b = u_1, \end{cases} \quad (55)$$

what yields

$$(2B + C + \hat{C})a = u_1 + \left(\frac{1}{2}A_1 + B + \hat{C}\right)u_0, \quad b = u_0 - a.$$

Since the operator iB is positive definite and iC , $i\hat{C}$ are selfadjoint operators one gets by using Theorems 4,5

$$\begin{aligned} (2iBx, x) + (iCx, x) + (i\hat{C}x, x) &\geq (2|\omega| - \|C\| - \|\hat{C}\|)\|x\|^2 \\ &\geq 2\left(|\omega| - \frac{|\omega|q}{2}\right)\|x\|^2, \end{aligned}$$

i.e. the operator $2iB + iC + i\hat{C}$ is positive definite and the equations (55) have a unique solution

$$a = (2B + C + \hat{C})^{-1} \left(u_1 + \left(\frac{1}{2}A_1 + B + \hat{C} \right) u_0 \right) \quad (56)$$

$$b = u_0 - a .$$

Thus, we have proved the following statement.

Theorem 6. *Let the assumptions of Theorem 4 hold. Then the problem (50) possesses a unique solution of the kind (54) where a, b are given by (56).*

References

- [1] Arov, D.Z.: *Passive linear steady-state dynamical systems*. Siberian Math. J. 20(1979) 2, 149 - 162.
- [2] Balas, M. J.: *Trends in large space structure control theory: Fondest hopes, wildest dreams*. IEEE Trans. Automat. Control 27(1982), 522 - 535.
- [3] Barnett, S.: *Introduction to Mathematical Control Theory*. Oxford: Clarendon Press 1975.
- [4] Bartels, R. and G. W. Stewart: *Algorithm 432: Solution of the matrix equation $AX + XB = C$* . Comm. ACM 15(1972), 820 - 826.
- [5] Bellman, R.: *Introduction to Matrix Analysis*. New York - Toronto - London: McGraw-Hill Book Company Inc. 1960.
- [6] Dalezkiĭ, Ju. L. and M. G. Krein: *Stability of Solutions of Differential Equations in Banach Space*. Moscow: Nauka 1970 (in Russian)
- [7] Datta, B. N. and K. Datta: *Theoretical and computational aspects of some linear algebra problems in control theory*. In: C. I. Byrnes and A. Lindquist (eds.): *Computational and Combinatorial Methods in System Theory*. Amsterdam: Elsevier 1986, pp. 201 - 212.
- [8] Davis, G. J.: *Numerical solution of a quadratic matrix equation*. SIAM J. Sci. Stat. Comp. 2(1981)2, 164 - 175.
- [9] Gajić, Z. and M.T.J. Qureshi: *Lyapunov Matrix Equation in System Stability and Control*. San Diego: Academic Press 1995.
- [10] Gantmacher, F. R.: *Matrizentheorie*. Berlin: Deutscher Verlag der Wissensch. 1986.
- [11] Gavriluk, I. P.: *An algorithmic representation of fractional powers of positive operators*. Numerical Functional Analysis and Optimization 17(1996)4, 293 - 305.
- [12] Gavriluk, I. P. and V. L. Makarov: *Representation and approximation for the solution of an initial value problem for a linear second order differential equation in Hilbert space*. Universität Leipzig, NTZ, Preprint-Nr. 6/95.
- [13] Gavriluk, I. P. and V. L. Makarov: *Explicit and approximate solutions of second order evolution differential equations in Hilbert space*. (submitted)

- [14] Golub, G. H., Nash, S. and C. Van Loan: *A Hessenberg-Schur method for the problem $AX + XB = C$* . IEEE Trans. Automat. Cont. 24(1979)6, 909 - 913.
- [15] Hammarling, S. J.: *Numerical solution of the stable, non-negative definite Lyapunov equation*. IMA J. Numer. Anal. 2(1982), 303 - 323.
- [16] Hochbruck, M. and G. Starke: *Preconditioned Krylov subspace methods for Lyapunov matrix equations*. SIAM J. Matrix Anal. Appl. 16(1995)1, 156 - 171.
- [17] Hoskins, W., Meek, D. and D. Walton: *The numerical solution of $A^T Q + QA = -C$* . IEEE Trans. Automatic Control 22(1977), 882 - 883.
- [18] Ikramov, H. D.: *The Numerical Solution of Matrix Equations*. Moscow: Nauka 1984 (in Russian).
- [19] Krein, S.G.: *Linear Differential Operators in Banach Spaces* (Transl. Amer. Math. Soc.: vol. 29). New York: Amer. Math. Soc., 1971.
- [20] Krein, M. G. and H. K. Langer: *On the theory of quadratic pencil of selfadjoint operators*. Soviet Mathematics Doklady 154(1964)6, 1258 - 1261 (in Russian).
- [21] Lancaster, P.: *Explicit solutions of linear matrix equations*. SIAM Review 12(1970), 544 - 566.
- [22] Lu, A. and E. L. Wachspress: *Solution of Lyapunov equations by ADI iteration*. Comp. Math. Appl. 21(1991), 43 - 58.
- [23] Lu, C.: *Solution of the matrix equation $AX + XB = C$* . Electronic Letters 7(1971), 185 - 186.
- [24] Lumer, G. and M. Rosenbloom: *Linear operator equations*. Proc. of the AMS, 10(1959), 32 - 41.
- [25] Mehrmann, V.: *The Autonomous Linear Quadratic Control Problem*. Lecture Notes in Control and Information Sciences 163(1991). Berlin - Heidelberg: Springer 1991.
- [26] Rosenbloom, M.: *On the operator equation $BX - XA = Q$* . Duke Math. J. 23(1956), 263 - 269.
- [27] Starke, G.: *SOR-line methods for Lyapunov matrix equation*. In: Proc. IMACS Internat. Symp. Iterative Methods in Linear Algebra. Amsterdam: North-Holland 1992, pp. 233 - 240.
- [28] Wachspress, E. L.: *Iterative solution of the Lyapunov matrix equation*. Appl. Math. Lett. 1(1988)87 - 90.
- [29] Young, D. M.: *Iterative Solution of Large Linear Systems*. New York - London: Academic Press 1971.