Eidgenössische Technische Hochschule Zürich
Ecole polytechnique fédérale de Zurich
Politecnico federale di Zurigo
Swiss Federal Institute of Technology Zurich

Bivariate matrix functions

D. Kressner

Research Report No. 2010-22
August 2010

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Bivariate Matrix Functions

Daniel Kressner*

August 29, 2010

Abstract

A definition of bivariate matrix functions is introduced and some theoretical as well as algorithmic aspects are analyzed. It is shown that our framework naturally extends the usual notion of (univariate) matrix functions and allows to unify existing results on linear matrix equations and derivatives of matrix functions.

1 Introduction

Given a square matrix A and a *univariate* scalar function $f(z)$ defined on the spectrum of A, the matrix function $f(A)$ is again a square matrix of the same size. Well-known examples include the matrix inverse A^{-1}, the matrix exponential $\exp(A)$ and the matrix logarithm $\log(A)$, see the recent monograph by Higham [11] for an excellent overview on the analysis and computation of such matrix functions.

This paper is concerned with the following question. What is an appropriate *bivariate* extension of matrix functions? More specifically, given two square matrices along with a bivariate scalar function $f(x, y)$, is there a sensible way of "evaluating f at these matrices"? Implicitly, as will be seen in the course of this note, this question has been considered many times in the literature for particular classes of bivariate functions. However, to the best of our knowledge, the most general case has not been put in a unified mathematical framework, with minimal assumptions on f and A, B. The main contribution of this note is to provide such a unification, covering several existing results and hopefully leading to new insights.

Given an $m \times m$ matrix A and an $n \times n$ matrix B, the bivariate matrix function $f\{A, B\}$ proposed in this note is not a matrix but a linear operator on the set of $m \times n$ matrices. In Section 2, three equivalent characterizations are provided, based on bivariate Hermite interpolation, an explicit expression, and a Cauchy integral formulation. In Section 3, it is shown that $f\{A, B\}$ nicely extends some well-known properties of univariate matrix functions. For example, the eigenvalues of $f\{A, B\}$ are the values of f at the eigenvalues of A and B. Another useful property is $(g \circ f)\{A, B\} = g\big(f\{A, B\}\big)$, allowing to succinctly express compositions of bivariate with univariate functions. For example, this shows that the solution to the matrix Sylvester equation $AX - XB^T = C$ can be written as $X = f\{A, B\}(C)$ with $f(x, y) = 1/(x - y)$. Section 4 presents another important special case of bivariate matrix functions: The Fréchet derivative of a univariate function f at a matrix A is shown to admit the expressions $f^{[1]}\{A, A^T\}$ with $f^{[1]}(x, y) = (f(x) - f(y))/(x - y)$. Section 5 sketches a general algorithm for computing bivariate matrix functions. However, it should be stressed that this

*Seminar for applied mathematics, ETH Zurich, kressner@math.ethz.ch

algorithm can be expected to be inferior in terms of efficiency and robustness compared to more specialized algorithms covering the special cases mentioned above. Some further developments that could offspring from the framework developed in this note are outlined in the concluding section.

2 Definition and Basic Properties

Before defining bivariate matrix functions, we briefly recall the definition of a univariate matrix function. Let $A \in \mathbb{C}^{m\times m}$ have the pairwise distinct eigenvalues $\lambda_1, \ldots, \lambda_s$. Then we let $\mathrm{ind}_{\lambda_i}(A)$ denote the index of λ_i, i.e., the size of the largest Jordan block associated with λ_i. Then the matrix function associated with a univariate scalar function f is defined as $f(A) := p(A)$, where $p(z)$ is the unique Hermite interpolating polynomial of degree less than $\sum_{i=1}^{s} \mathrm{ind}_{\lambda_i} A$ satisfying

$$\frac{\partial^g}{\partial z^g} p(\lambda_i) = \frac{\partial^g}{\partial z^g} f(\lambda_i), \quad g = 0, \ldots, \mathrm{ind}_{\lambda_i} A - 1, \quad i = 1, \ldots, s. \tag{1}$$

It is assumed that f is defined on the spectrum of A in the sense of [11], which in particular means that all required derivatives of f exist. Following [11], we say that *p interpolates f at A* if (1) is satisfied.

2.1 Definition via Hermite Interpolation

To extend the definition above from the univariate to the bivariate setting, we first consider the case of polynomials.

Definition 2.1 *Let $A \in \mathbb{C}^{m\times m}, B \in \mathbb{C}^{n\times n}$ and consider a bivariate polynomial $p(x, y) = \sum_{i=1}^{s} \sum_{j=1}^{t} p_{ij} x^i y^j$ with $p_{ij} \in \mathbb{C}$. Then $p\{A, B\} : \mathbb{C}^{m\times n} \to \mathbb{C}^{m\times n}$ is defined by*

$$p\{A, B\}(C) := \sum_{i=1}^{s} \sum_{j=1}^{t} p_{ij} A^i C (B^T)^j. \tag{2}$$

Note that the transposition of B in (2) is purely a matter of convention, which has the main advantage that it allows for a non-ambiguous extension to multivariate functions, see Section 6.

As in the univariate case, we will approach a general bivariate function by means of Hermite interpolation. This is only possible if the function is defined on the spectra of A and B in the following sense.

Definition 2.2 *Let $A \in \mathbb{C}^{m\times m}$ have pairwise distinct eigenvalues $\lambda_1, \ldots, \lambda_s$ and let $B \in \mathbb{C}^{n\times n}$ have pairwise distinct eigenvalues $\mu_1, \ldots, \mu_t$. Then a bivariate function $f(x, y)$ is* defined on the spectra of A and B *if the following mixed partial derivatives exist and are continuous:*

$$\frac{\partial^{g+h}}{\partial x^g y^h} f(\lambda_i, \mu_j), \qquad \begin{array}{ll} g = 0, \ldots, \mathrm{ind}_{\lambda_i}(A) - 1, & i = 1, \ldots, s, \\ h = 0, \ldots, \mathrm{ind}_{\mu_j}(B) - 1, & j = 1, \ldots, t. \end{array}$$

Bivariate Hermite interpolation on tensor grid data is well-understood and can be easily performed by tensorized univariate Hermite interpolation, see, e.g., [1, 13]. In particular, for any $f(x,y)$ defined on the spectra of A and B there is a bivariate polynomial $p(x,y)$ satisfying

$$\frac{\partial^{g+h}}{\partial x^g y^h} p(\lambda_i, \mu_j) = \frac{\partial^{g+h}}{\partial x^g y^h} f(\lambda_i, \mu_j), \qquad \begin{array}{ll} g = 0, \dots, \mathrm{ind}_{\lambda_i}(A) - 1, & i = 1, \dots, s, \\ h = 0, \dots, \mathrm{ind}_{\mu_j}(B) - 1, & j = 1, \dots, t. \end{array} \tag{3}$$

The choice of $p(x,y)$ is unique if it has degree less than $\sum_{i=1}^{s} \mathrm{ind}_{\lambda_i}(A)$ in x and degree less than $\sum_{j=1}^{t} \mathrm{ind}_{\mu_j}(B)$ in y. As in the univariate case, we say that *p interpolates f at $\{A,B\}$* if (3) is satisfied.

Definition 2.3 *The* bivariate matrix function *associated with a bivariate scalar function $f(x,y)$ defined on the spectra of A and B is defined by $f\{A,B\} := p\{A,B\}$, where $p(x,y)$ is the bivariate polynomial of minimal degree interpolating f at $\{A,B\}$.*

In the following, we discuss some basic properties of bivariate matrix functions.

Lemma 2.4 *Under the conditions of Definition 2.3,*

$$f\{A,B\}(C) = P\Big(f\{P^{-1}AP, Q^{-1}BQ\}(P^{-1}CQ^{-T})\Big)Q^T \tag{4}$$

for any two invertible matrices P,Q of matching size.

Proof. It is straightforward to verify this statement for polynomials f, which concludes the proof by definition. □

Lemma 2.5 *Consider any two polynomials p_1, p_2 satisfying the interpolation conditions (3). Then $p_1\{A,B\} = p_2\{A,B\}$.*

Proof. By Lemma 2.4, we can assume A, B to be in Jordan canonical form. Since the evaluation of bivariate matrix polynomials decouples for block diagonal matrices A and B (see also Lemma 2.6 below), it suffices to prove the statement for $\lambda_i I + N_A$, and $\mu_j I + N_B$, where N_A, N_B are nilpotent matrices of index $\mathrm{ind}_{\lambda_i}(A)$ and $\mathrm{ind}_{\mu_j}(B)$, respectively. Set $e := p_1 - p_2$. Then

$$\frac{\partial^{g+h}}{\partial x^g y^h} e(\lambda_i, \mu_j) = 0, \quad g = 0, \dots, \mathrm{ind}_{\lambda_i}(A) - 1, \quad h = 0, \dots, \mathrm{ind}_{\mu_j}(B) - 1$$

and hence e takes the form

$$e(x,y) = \sum_{\substack{g \geq \mathrm{ind}_{\lambda_i}(A) \\ h \geq \mathrm{ind}_{\mu_i}(B)}} e_{gh}(x - \lambda_i)^g (y - \mu_j)^h$$

for some coefficients e_{gh}. By Definition 2.1 and the nilpotency of N_A, N_B, this implies $e\{\lambda_i I + N_A, \mu_i I + N_B\} = 0$. □

Lemma 2.5 has the convenient consequence that *any* polynomial satisfying the appropriate Hermite interpolation conditions can be used for defining a bivariate matrix function.

Lemma 2.6 *Consider block (diagonal) matrices*

$$A = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} B_{11} & 0 \\ 0 & B_{22} \end{bmatrix}, \quad C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix},$$

where A_{ii}, B_{jj} are square and C is partitioned conformally with A and B. Then

$$f\{A,B\}(C) = \begin{bmatrix} f\{A_{11},B_{11}\}(C_{11}) & f\{A_{11},B_{22}\}(C_{12}) \\ f\{A_{22},B_{11}\}(C_{21}) & f\{A_{22},B_{22}\}(C_{22}) \end{bmatrix}, \tag{5}$$

holds for any $f(x,y)$ defined on the spectra of A, B,

Proof. Clearly, any polynomial p interpolating f at $\{A,B\}$ also interpolates f at $\{A_{ii}, B_{jj}\}$ for $i \in \{1,2\}$, $j \in \{1,2\}$. Thus, by Lemma 2.5, it suffices to establish (5) for the polynomial p, which is straightforward to verify. □

Lemma 2.6 extends in a direct manner to block diagonal matrices A, B with arbitrarily many square diagonal blocks.

2.2 An explicit expression

The aim of this section is to characterize $f\{A,B\}(C)$ in terms of the Jordan structure of A and B. First, let us briefly consider the special case that A and B happen to be of the form

$$A = \lambda I + N_A, \quad B = \mu I + N_B,$$

where N_A, N_B are nilpotent matrices of index $\tilde{m}, \tilde{n}$, respectively. Then an interpolating polynomial is given by the truncated Taylor expansion

$$p(x,y) = \sum_{g=0}^{\tilde{m}-1} \sum_{h=0}^{\tilde{n}-1} \frac{1}{g!h!} \frac{\partial^{g+h}}{\partial^g x \partial^h y} f(\lambda,\mu)(x-\lambda)^g (y-\mu)^h.$$

According to Definition 2.3,

$$f\{A,B\}(C) = \sum_{g=0}^{\tilde{m}-1} \sum_{h=0}^{\tilde{n}-1} \frac{1}{g!h!} \frac{\partial^{g+h}}{\partial x^g y^h} f(\lambda,\mu) N_A^g C (N_B^T)^h. \tag{6}$$

This can be used to derive an explicit expression based on the Jordan canonical forms of A and B:

$$\begin{array}{ll} A = PJ_AP^{-1}, & J_A = \operatorname{diag}\big(J_A(\lambda_1), J_A(\lambda_2), \dots, J_A(\lambda_s)\big), \\ B = QJ_BQ^{-1}, & J_B = \operatorname{diag}\big(J_B(\mu_1), J_B(\mu_2), \dots, J_B(\mu_t)\big), \end{array} \tag{7}$$

where $J_A(\lambda_i)$ contains all Jordan blocks belonging to the eigenvalue λ_i of A, and analogously $J_B(\mu_j)$.

Lemma 2.7 *Let A, B have the Jordan canonical forms (7) and partition*

$$P^{-1}CQ^{-T} = \begin{bmatrix} C_{11} & \cdots & C_{1t} \\ \vdots & & \vdots \\ C_{s1} & \cdots & C_{st} \end{bmatrix}$$

conformally. Then

$$f\{A,B\}(C) = P \begin{bmatrix} F_{11} & \cdots & F_{1t} \\ \vdots & & \vdots \\ F_{s1} & \cdots & F_{st} \end{bmatrix} Q^T$$

with

$$F_{ij} = \sum_{g=0}^{\mathrm{ind}_{\lambda_i}(A)-1} \sum_{h=0}^{\mathrm{ind}_{\mu_j}(B)-1} \frac{1}{g!h!} \frac{\partial^{g+h}}{\partial x^g y^h} f(\lambda_i,\mu_j)\big(J_A(\lambda_i)-\lambda_i I\big)^g C_{ij} \big(J_B^T(\mu_j)-\mu_j I\big)^h \qquad (8)$$

for $i = 1,\dots,s$ and $j = 1,\dots,t$.

Proof. By Lemma 2.4 and Lemma 2.6, we have $F_{ij} = f\{J_A(\lambda_i), J_B(\mu_j)\}(C_{ij})$. The formula (8) therefore follows directly from (6). □

Lemma 2.7 extends a similar expression given in [20, Sec. 10] for the solution of matrix Sylvester equations, which will be seen below to correspond to the case $f(x,y) = 1/(x-y)$. More specifically, in [20, Sec. 10], a more compact formulation has been attained by defining matrices of the form

$$\begin{aligned} V_{ig} &= P\,\mathrm{diag}\big(0,\dots,0,\big(J_A(\lambda_i)-\lambda_i I\big)^g,0,\dots,0\big)P^{-1}, \\ W_{jh} &= Q\,\mathrm{diag}\big(0,\dots,0,\big(J_B(\mu_j)-\mu_j I\big)^h,0,\dots,0\big)Q^{-1} \end{aligned} \qquad (9)$$

for $g = 0,\dots,\mathrm{ind}(\lambda_i)-1,\ i = 1,\dots,s$ and $h = 0,\dots,\mathrm{ind}(\mu_j)-1,\ j = 1,\dots,t$. Equivalently,

$$V_{ig} = \frac{1}{2\pi \mathrm{i}} \oint_{\Gamma(\lambda_i)} (x-\lambda_i)^g (zI-A)^{-1}\mathrm{d}x, \quad W_{jh} = \frac{1}{2\pi \mathrm{i}} \oint_{\Gamma(\mu_j)} (z-\mu_j)^h (yI-B)^{-1}\mathrm{d}y,$$

where $\Gamma(\lambda_i)$ and $\Gamma(\mu_j)$ are sufficiently small circles surrounding λ_i and μ_j, respectively. By Lemma 2.7,

$$f\{A,B\}(C) = \sum_{i,j}\sum_{g,h} \frac{1}{g!h!}\frac{\partial^{g+h}}{\partial x^g y^h} f(\lambda_i,\mu_j) V_{ig} C W_{jh}^T. \qquad (10)$$

2.3 A Cauchy integral representation

For holomorphic f, the expression (10) leads to a a Cauchy integral representation of $f\{A,B\}(C)$. We refer to [17] for an introduction to multivariate holomorphic functions.

Theorem 2.8 *Let $\Omega_A, \Omega_B \subset \mathbb{C}$ be open sets containing the eigenvalues of A and B, respectively, such that f is holomorphic on $\Omega_A \times \Omega_B$ and continuous on $\overline{\Omega_A \times \Omega_B}$. Then*

$$f\{A,B\}(C) = -\frac{1}{4\pi^2} \oint_{\Gamma_A}\oint_{\Gamma_B} f(x,y)(xI-A)^{-1}C(yI-B)^{-1}\mathrm{d}y\,\mathrm{d}x, \qquad (11)$$

where Γ_A, Γ_B are the contours of Ω_A, Ω_B.

Proof. By changing the path of integration, the right-hand side of (11) can be replaced by

$$-\frac{1}{4\pi^2}\sum_{i,j} \oint_{\Gamma(\lambda_i)}\oint_{\Gamma(\mu_j)} f(x,y)(xI-A)^{-1}C(yI-B)^{-1}\mathrm{d}y\,\mathrm{d}x, \qquad (12)$$

with sufficiently small circles $\Gamma(\lambda_i), \Gamma(\mu_j)$ surrounding λ_i, μ_j. The matrices V_{ig}, W_{jh} introduced in (9) allow for the decompositions

$$(xI - A)^{-1} = \sum_i \sum_{g=0}^{\mathrm{ind}_{\lambda_i}(A)-1} \frac{1}{(x-\lambda_i)^{g+1}} V_{ig}, \quad (yI - B)^{-1} = \sum_j \sum_{h=0}^{\mathrm{ind}_{\mu_j}(B)-1} \frac{1}{(y-\mu_j)^{h+1}} W_{jh}.$$

Inserting this into (12) gives

$$-\frac{1}{4\pi^2} \sum_{i,j} \sum_{g,h} \oint_{\Gamma(\lambda_i)} \oint_{\Gamma(\mu_j)} \frac{f(x,y)}{(x-\lambda_i)^{g+1}(y-\mu_j)^{h+1}} V_{ig} C W_{jh}^T \mathrm{d}y\, \mathrm{d}x. \tag{13}$$

Using

$$\frac{\partial^{g+h}}{\partial x^g y^h} f(\lambda_i, \mu_j) = -\frac{g!h!}{4\pi^2} \oint_{\Gamma(\lambda_i)} \oint_{\Gamma(\mu_j)} \frac{f(x,y)}{(x-\lambda_i)^{g+1}(y-\mu_j)^{h+1}} \mathrm{d}y\, \mathrm{d}x,$$

we thus obtain that the right-hand side of (11) is identical with the expression (10) for $f\{A,B\}(C)$. ◻

For the case $f(x,y) = 1/p(x,y)$ with an arbitrary bivariate polynomial p the result of Theorem 2.8 is attributed in [20] to Krein [18].

In most practically relevant instances of the univariate case, the eigenvalues of A are contained in a domain of holomorphy of f and therefore Ω_A is connected. This appears to happen less frequently for bivariate holomorphic functions. As a typical example, consider $f(x,y) = 1/(x-y)$ and let the eigenvalues of A be $1/2, 1+1/2, \ldots, n+1/2$ while the eigenvalues of B are $1, 2, \ldots, n$. Then any $\Omega_A \times \Omega_B$ in the sense of Theorem 2.8 consists of at least n^2 connected components, see also Figure 1.

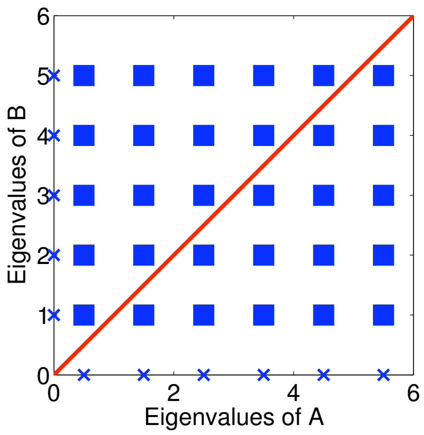

Figure 1: *Red line:* Singularities of $f(x,y) = 1/(x-y)$. *Blue squares:* Set $(\Omega_A \cap \mathbb{R}) \times (\Omega_B \cap \mathbb{R})$ for which $\Omega_A \times \Omega_B$ satisfies the requirements of Theorem 2.8

3 Spectral properties and composition of functions

The eigenvalues of the linear operator $f\{A,B\}$ are scalars $\lambda \in \mathbb{C}$ for which there is a nonzero $C \in \mathbb{C}^{m\times n}$ such that $f\{A,B\}(C) = \lambda C$. Equivalently, these are the eigenvalues of the

$mn \times mn$ matrix $\mathcal{M}(f\{A,B\})$, where $\mathcal{M}$ denotes the natural isomorphism between linear operators on $\mathbb{C}^{m\times n}$ and $mn \times mn$ matrices.

Lemma 3.1 *Let $\lambda_1,\dots,\lambda_s$ and $\mu_1,\dots,\mu_t$ denote the eigenvalues of A and B, respectively. Then the eigenvalues of $f\{A,B\}$ are given by $f(\lambda_i,\mu_j)$ for $i=1,\dots,s,\ j=1,\dots,t$.*

Proof. By Definition 2.3, $f\{A,B\} = p\{A,B\}$, where the polynomial $p(x,y) = \sum_{ij} p_{ij} x^i y^j$ interpolates f at $\{A,B\}$. By a direct extension of the usual argument for linear matrix equations (see, e.g., [12, Thm 4.45]), the eigenvalues of

$$\mathcal{M}(p\{A,B\}) = \sum_{ij} p_{ij}\big(B^j \otimes A^i\big).$$

are given by $p(\lambda_i,\mu_j) = f(\lambda_i,\mu_j)$. ∎

To discuss the index of $f(\lambda_i,\mu_j)$ as an eigenvalue of $f\{A,B\}$, the following result will turn out to be useful.

Lemma 3.2 *Let f_1, f_2 be holomorphic functions in the vicinity of the spectra of A,B. Then $f_1\{A,B\}\circ f_2\{A,B\} = \pi\{A,B\}$ with $\pi(x,y) = f_1(x,y)f_2(x,y)$.*

Proof. By an appropriate choice of contours Γ_A and Γ_B, Theorem 2.8 implies

$$\begin{aligned}\big(f_1\{A,B\}\circ f_2\{A,B\}\big)(C) &= \frac{1}{16\pi^4}\oint_{\Gamma_A}\oint_{\Gamma_B}\oint_{\Gamma_A}\oint_{\Gamma_B} f_1(x_1,y_1)f_2(x_2,y_2)(x_1I-A)^{-1}(x_2I-A)^{-1}\cdots\\ &\qquad \cdots C(y_1I-B)^{-1}(y_2I-B)^{-1}\mathrm{d}y_2\mathrm{d}x_2\mathrm{d}y_1\mathrm{d}x_1.\end{aligned}$$

Using

$$\begin{aligned}-\frac{1}{4\pi^2}\oint_{\Gamma_A}\oint_{\Gamma_A} & f_1(x_1,y_1)f_2(x_2,y_2)(x_1I-A)^{-1}(x_2I-A)^{-1}\mathrm{d}x_2\mathrm{d}x_1\\ &= \frac{1}{2\pi\mathrm{i}}\oint_{\Gamma_A} f_1(x,y_1)f_2(x,y_2)(xI-A)^{-1}\mathrm{d}x,\end{aligned}$$

see, e.g., [20, Sec 7], and an analogous formula for $(y_1-B)^{-1}(y_2-B)^{-1}$, we obtain

$$\begin{aligned}\big(f_1\{A,B\}\circ f_2\{A,B\}\big)(C) &= -\frac{1}{4\pi^2}\oint_{\Gamma_B}\oint_{\Gamma_A} f_1(x,y)f_2(x,y)(xI-A)^{-1}C(yI-B)^{-1}\mathrm{d}y\,\mathrm{d}x\\ &= \pi\{A,B\}(C).\end{aligned}$$

∎

An immediate consequence of Lemma 3.2, bivariate matrix functions evaluated at the same arguments commute: $f_1\{A,B\}\circ f_2\{A,B\} = f_2\{A,B\}\circ f_1\{A,B\}$. Another consequence is the power rule

$$\underbrace{f\{A,B\}\circ\cdots\circ f\{A,B\}}_{d \text{ times}} = f^d\{A,B\}. \tag{14}$$

Corollary 3.3 *Let σ be an eigenvalue of $f\{A,B\}$. Then*

$$\mathrm{ind}_\sigma f\{A,B\} \le \max\big\{\mathrm{ind}_\lambda A + \mathrm{ind}_\mu B - 1 :\ \sigma = f(\lambda,\mu), \lambda\in\Lambda(A),\ \mu\in\Lambda(B)\big\}, \tag{15}$$

where Λ denotes the set of eigenvalues of a matrix.

Proof. From Lemma 2.4 it is clear that we can assume without loss of generality that A and B are already in Jordan canonical form (7). Then the matrix representation of $f\{A,B\}$ becomes block diagonal with diagonal blocks

$$\mathcal{M}(f\{J_A(\lambda),J_B(\mu)\}) = \sum_{i=0}^{\mathrm{ind}_\lambda(A)-1}\sum_{j=0}^{\mathrm{ind}_\mu(B)-1} p_{ij}\big(J_B^T(\mu)-\mu I\big)^j \otimes \big(J_A(\lambda)-\lambda I\big)^i,$$

for some coefficients p_{ij} with $p_{00} = f(\lambda,\mu) =: \sigma$, see (8). Defining the polynomial $p(x,y) = \sum_{i+j\geq 1} p_{ij}x^iy^j$, we have

$$f\{J_A(\lambda),J_B(\mu)\} - \sigma I = p\{J_A(\lambda)-\lambda I, J_B(\mu)-\mu I\}$$

and therefore, by (14),

$$\big(f\{J_A(\lambda),J_B(\mu)\} - \sigma I\big)^d = p^d\{J_A(\lambda)-\lambda I, J_B(\mu)-\mu I\}.$$

By the binomial theorem,

$$p^d\{J_A(\lambda)-\lambda I, J_B(\mu)-\mu I\} = \sum_{i+j\geq d} q_{ij}\big(J_B^T(\mu)-\mu I\big)^j \otimes \big(J_A(\lambda)-\lambda I\big)^i$$

for some coefficients q_{ij}. A term in this sum becomes zero if $i \geq \mathrm{ind}_\lambda(A)$ or $j \geq \mathrm{ind}_\lambda(B)$, which will always be the case if $d = \mathrm{ind}_\lambda(A) + \mathrm{ind}_\mu(B) - 1$. Hence,

$$\mathrm{ind}_\sigma f\{J_A(\lambda),J_B(\mu)\} \leq \mathrm{ind}_\lambda A + \mathrm{ind}_\mu B - 1.$$

This shows the result by taking the maximum over all eigenvalue pairs λ,μ that satisfy $\sigma = f(\lambda,\mu)$. □

In most cases of practical interest, we expect that equality holds in (15). However, there are obvious exceptions, as the trivial example $f(x,y)\equiv 0$ demonstrates.

A bivariate matrix function $f\{A,B\}$ can be composed with a univariate function $u(z)$ by applying the usual definition of matrix function to the matrix representation $\mathcal{M}(f\{A,B\})$. Formally, we let

$$u(f\{A,B\}) := \mathcal{M}^{-1}(u(\mathcal{M}(f\{A,B\}))).$$

This definition assumes u to be defined on the spectrum of $f\{A,B\}$, for which a sufficient condition in terms of the Jordan structures of A and B can be easily derived from (3.3): The derivatives

$$u^{(g)}\big(f(\lambda_i,\mu_j)\big), \qquad \begin{array}{l} i=1,\dots,s,\ j=1,\dots,t, \\ g=0,\dots,\mathrm{ind}_{\lambda_i}A + \mathrm{ind}_{\mu_j}B - 2, \end{array} \tag{16}$$

are assumed to exist.

Theorem 3.4 *Consider a bivariate function $f(x,y)$ defined on the spectra of square matrices A,B and a univariate function $u(z)$ for which the derivatives (16) exist. Then*

$$u(f\{A,B\}) = (u\circ f)\{A,B\}. \tag{17}$$

Proof. Since both sides of (17) are linear in u, the power rule (14) implies that the statement of the lemma holds for any polynomial u. Now, let the polynomial p_f interpolate f at $\{A, B\}$, and let p_u be a Hermite interpolation of u satisfying $p_u^{(g)}\big(f(\lambda_i, \mu_j)\big) = u^{(g)}\big(f(\lambda_i, \mu_j)\big)$ for all i, j, g as in (16). Then

$$u(f\{A, B\}) = p_u(p_f\{A, B\}) = (p_u \circ p_f)\{A, B\}.$$

The proof is concluded if we can show that $p_u \circ p_f$ interpolates $u \circ f$ at $\{A, B\}$. By Faà di Bruno's chain rule, the mixed derivative

$$\frac{\partial^{g+h}}{\partial x^g y^h} p_u\big(p_f(x, y)\big)$$

can be expressed in terms of derivatives of p_u up to order $g + h$ and mixed partial derivatives of p_f up to order g, h in x, y. Applying this chain rule to the conditions (3) for $u \circ f$, all the resulting derivatives of p_u and p_f are found to match those of u and f, respectively. Hence, (3) is satisfied; $p_u \circ p_f$ indeed interpolates $u \circ f$ at $\{A, B\}$. $\square$

To give some examples of Theorem 3.4, consider first the Sylvester equation $AX - XB^T = C$ or, equivalently, $f\{A, B\}(X) = C$ for $f(x, y) = x - y$. Provided that A and B have disjoint spectra, Theorem 3.4 implies that the solution $X = f\{A, B\}^{-1}(C)$ can be written as

$$X = f_{\mathsf{sylv}}\{A, B\}(C) \quad \text{with} \quad f_{\mathsf{sylv}}(x, y) = \frac{1}{x - y}.$$

Similarly, the solution to the Stein equation $X - AXB^T = C$, if it exists and is unique, can be written as

$$X = f_{\mathsf{stein}}\{A, B\}(C) \quad \text{with} \quad f_{\mathsf{stein}}(x, y) = \frac{1}{1 - xy}.$$

As a last example, Theorem 3.4 implies the identity

$$\big(I \otimes A + B \otimes I\big)^{-1/2}\mathrm{vec}(C) = f_{\mathsf{isqr}}\{A, B\}(C) \quad \text{with} \quad f_{\mathsf{isqr}}(x, y) = (x + y)^{-1/2}, \tag{18}$$

which allows for the application of the inverse matrix square root of $I \otimes A + B \otimes I$ without having to form this matrix explicitly, see Section 5 for a more detailed discussion.

4 Fréchet derivatives of univariate matrix functions

Given a sufficiently often differentiable univariate function $f(x)$, the Fréchet derivative of f at a matrix A in direction C is defined as

$$Df\{A\}(C) := \lim_{h \to 0} \frac{1}{h}\big(f(A + hC) - f(A)\big).$$

The following result shows that $Df\{A\}(C)$ can be interpreted as a bivariate matrix function representing the finite difference evaluated at A.

Theorem 4.1 *Let A be a square matrix and let f be $2 \cdot \mathrm{ind}_\lambda A - 1$ times continuously differentiable at λ for every $\lambda \in \Lambda(A)$. Then*

$$Df\{A\}(C) = f^{[1]}\{A, A^T\}(C), \quad \textit{with} \quad f^{[1]}(x, y) := f[x, y] = \begin{cases} \frac{f(x) - f(y)}{x - y}, & \textit{for } x \neq y, \\ f'(x), & \textit{for } x = y. \end{cases}$$

Proof. For $f(x) = x^k$, it is well known (and easy to see) that

$$Df\{A\}(C) = \sum_{i=1}^{k} A^{k-i} C A^{i-1} = f^{[1]}\{A, A^T\}(C)$$

with $f^{[1]}(x,y) = \sum_{i=1}^{k} x^{k-i} y^{i-1} = (x^k - y^k)/(x-y)$ for $x \neq y$ and $f^{[1]}(x,x) = f'(x)$. Because of linearity, this shows the statement of the theorem for every polynomial. For the general case of a function f satisfying the assumptions, let p be an interpolating polynomial matching the first $2 \cdot \mathrm{ind}_\lambda A - 1$ derivatives of f at every eigenvalue λ of A. Consider any pair of eigenvalues λ, μ of A, and let

$$T = \begin{bmatrix} \tau_0 & 1 & & \\ & \tau_1 & \ddots & \\ & & \ddots & 1 \\ & & & \tau_{g+h+1} \end{bmatrix}, \quad \tau_0 = \cdots = \tau_g = \lambda, \quad \tau_{g+1} = \cdots = \tau_{g+h+1} = \mu.$$

Then $f(T)$ is defined and equals $p(T)$ as long as $0 \le g + h \le \mathrm{ind}_\lambda A + \mathrm{ind}_\mu A - 1$. A result by Opitz [24] shows that the upper triangular entries of $f(T)$ are the divided differences of f. In particular, the entry in the upper right corner equals

$$f[\underbrace{\lambda, \ldots, \lambda}_{g+1 \text{ times}}, \underbrace{\mu, \ldots, \mu}_{h+1 \text{ times}}] = \begin{cases} \frac{1}{g!h!} \frac{\partial^{g+h}}{\partial x^g y^h} f[x,y] \Big|_{x=\lambda, y=\mu} & \text{for } \lambda \neq \mu, \\ \frac{1}{(g+h)!} \frac{\partial^{g+h}}{\partial x^{g+h}} f(x) \Big|_{x=\lambda} & \text{for } \lambda = \mu, \end{cases}$$

which, together with $f(T) = p(T)$, shows

$$\frac{\partial^{g+h}}{\partial x^g y^h} p^{[1]}(\lambda, \mu) = \frac{\partial^{g+h}}{\partial x^g y^h} f^{[1]}(\lambda, \mu)$$

for all $0 \le g+h \le \mathrm{ind}_\lambda A + \mathrm{ind}_\mu A - 1$. Hence, $p^{[1]}$ satisfies the required interpolation conditions and

$$Df\{A\}(C) = Dp\{A\}(C) = p^{[1]}\{A, A^T\} = f^{[1]}\{A, A^T\}$$

concludes the proof. □

Using Lemma 3.2 and Theorem 4.1, the trivial relation $f[x,y]x - f[x,y]y = f[x,y](x-y) = f(x) - f(y)$, gives – expressed in terms of the matrix A – the commutator relations

$$A\, Df\{A\}(C) - Df\{A\}(C)\, A = Df\{A\}(AC - CA) = f(A)C - Cf(A), \quad \forall C \in \mathbb{C}^{n \times n} \tag{19}$$

for any holomorphic function f, see also [3, Thm 2.1]. Najfeld and Havel [23, Thm 4.4] have obtained an expression for $Df\{A\}$ from (19) for functions f that admit a power series with convergence radius ρ and $\|A\| < \rho$. In [23], this expression is called "generalized divided difference matrix", which coincides with the matrix representation of $f\{A, A\}$ and thus matches the statement of Theorem 4.1. Note, however, that Theorem 4.1 imposes much weaker conditions on f and A.

Theorem 4.1 together with Lemma 3.1 reconfirm the well-known fact that the eigenvalues of $Df\{A\}$ are given by $f[\lambda, \mu]$ for all pairs $\lambda, \mu \in \Lambda(A)$, see also Theorem 3.9 in [11]. Lemma 2.7 and (10) yield explicit expressions for $Df\{A\} = f^{[1]}\{A, A^T\}$ that recover an expression of Horn and Johnson stated in [12, Thm 6.6.14] under stronger assumptions on the differentiability of f. Note that all these explicit expressions coincide with a formula by Daleckiĭ and Kreĭn [5, 4] in the special case that A is diagonalizable, see also (21) below and Theorem 3.11 in [11].

Finally, we demonstrate the versatility of the framework of bivariate matrix functions by showing the well-known relation

$$f\left(\begin{bmatrix} A & C \\ 0 & A \end{bmatrix}\right) = \begin{bmatrix} A & Df\{A\}(C) \\ 0 & A \end{bmatrix} \tag{20}$$

under minimal conditions on f, A.

Theorem 4.2 *Equation (20) holds under the assumptions of Theorem 4.1.*

Proof. It is well known (and easy to show) that (20) holds for any polynomial. Setting $M = \begin{bmatrix} A & C \\ 0 & A \end{bmatrix}$, we clearly have $\mathrm{ind}_\lambda M \le 2 \cdot \mathrm{ind}_\lambda A$ for every $\lambda \in \Lambda(A)$. Hence, for a polynomial p interpolating the first $2 \cdot \mathrm{ind}_\lambda A - 1$ derivatives of f at λ, we have $p(M) = f(M)$. By the argument used in the proof of Theorem 4.1, $p^{[1]}$ interpolates $f^{[1]}$ at $\{A, A^T\}$. This concludes the proof:

$$f(M) = p(M) = \begin{bmatrix} p(A) & p^{[1]}\{A, A^T\}(C) \\ 0 & p(A) \end{bmatrix} = \begin{bmatrix} f(A) & f^{[1]}\{A, A^T\}(C) \\ 0 & f(A) \end{bmatrix}.$$

□

In comparison, Theorem 2.1 in [22] shows (20) only under the stronger assumption that f is $m-1$ times continuously differentiable at *every* eigenvalue of A, where $m = \max\{\mathrm{ind}_\lambda A : \lambda \in \Lambda(A)\}$.

By applying the Cauchy integral representation for holomorphic matrix functions, Equation (20) implies a well-known integral representation for $Df(A)$, see [27]. It is instructive to rederive this representation from the Cauchy integral formulation of Theorem 2.8 applied to $f^{[1]}\{A, A^T\}$.

5 Computation of bivariate matrix functions

The purpose of this section is to provide a rather informal discussion of possible algorithms for computing $f\{A, B\}(C)$ for medium-sized matrices A, B.

Diagonalization of A and B. Suppose that A, B are diagonalizable:

$$P^{-1}AP = \mathrm{diag}(\lambda_1, \dots, \lambda_m), \quad Q^{-1}BQ = \mathrm{diag}(\mu_1, \dots, \mu_n),$$

and let $\tilde{C} = P^{-1}CQ^{-T}$. Then Lemma 2.7 implies

$$f\{A, B\}(C) = P\big(\tilde{F} \circ \tilde{C}\big)Q^T \quad \text{with} \quad \tilde{f}_{ij} = f(\lambda_i, \mu_j), \tag{21}$$

where "∘" denotes the Hadamard product. This expression is well-suited for (nearly) normal matrices A and B but can be expected to run into numerical instabilities when P and/or Q are ill-conditioned.

Diagonalization of B only. The above approach can be modified if only one of the matrices, say B, is known to admit a well-conditioned basis of eigenvectors. Let $Q^{-1}BQ = \operatorname{diag}(\mu_1, \dots, \mu_n)$ and partition $\tilde{C} = CQ^{-T} = \big[c_1, \dots, c_n\big]$. Then Lemma 2.4 and Lemma 2.6 imply

$$f\{A,B\}(C) = \big[y_1, \dots, y_n\big]Q^T \quad \text{with} \quad y_j = f\{A, \mu_j\}(c_j). \tag{22}$$

Note that $f\{A, \mu_j\} = f_{\mu_j}(A)$ is a univariate matrix function for $f_{\mu_j}(x) = f(x, \mu_j)$. Having a stable procedure for evaluating/applying $f_{\mu_j}(A)$ at hand, this approach can be expected to be significantly more robust than (21). An analogous, row-wise procedure can be performed if A is known to admit a well-conditioned basis of eigenvectors.

Taylor expansion. In the extreme case that all eigenvalues of B are nearly identical, the diagonalization of B is clearly not the preferred option but this situation can be exploited as well. Following the approach for univariate matrix functions proposed by Kågström [16], see also [6], we let $\mu = \operatorname{trace}(B)/n$ and consider the truncated Taylor expansion

$$f(x,y) \approx f(x,\mu) + (y-\mu)\frac{\partial}{\partial y} f(x,\mu) + \cdots + \frac{1}{k!}(y-\mu)^k \frac{\partial^k}{\partial y^k} f(x,\mu). \tag{23}$$

This yields an approximation for the bivariate matrix function in terms of the univariate functions $f_\mu^{(0)}(x) := f(x,\mu), f_\mu^{(1)}(x) := \frac{\partial}{\partial y} f(x,\mu), \dots, f_\mu^{(k)}(x) := \frac{\partial^k}{\partial y^k} f(x,\mu)$:

$$f\{A,B\}(C) \approx f_\mu^{(0)}(A)C + f_\mu^{(1)}(A)C(B-\mu I) + \cdots + \frac{1}{k!} f_\mu^{(k)}(A)C(B-\mu I)^k. \tag{24}$$

Since B has all eigenvalues close to μ, one can expect that k need not be chosen very large to obtain good accuracy, see [6, 11, 16, 21] for discussions. Compared to (22), formula (24) has the requirement that not only $f_\mu^{(0)}$ but also the derivatives $f_\mu^{(1)}, \dots, f_\mu^{(k)}$ need to be evaluated at A. Without going into implementation details, we only mention that if the Schur-Parlett algorithm [6] is used for this purpose then the mixed partial derivatives of f need to be available, which could be considered a not too unreasonable requirement.

Block diagonalization of B only. Diagonalization and Taylor expansion can be combined in an obvious manner by considering a block diagonalization of B:

$$Q^{-1}BQ = \operatorname{diag}\big(B_{11}, \dots, B_{tt}\big) \tag{25}$$

such that Q is well-conditioned and the eigenvalues of each diagonal block B_{jj} are nearly identical. Methods for performing such a decomposition reliably are a subtle matter and have been discussed, e.g., in [6, 9, 25]. Assuming (25) is available,

$$f\{A,B\}(C) = \big[Y_1, \dots, Y_n\big]Q^T \quad \text{with} \quad Y_j = f\{A, B_{jj}\}(C_j),$$

where $\tilde{C} = CQ^{-T} = \big[C_1, \dots, C_n\big]$ is partitioned in accordance with $Q^{-1}BQ$. In effect, each block column Y_j can be computed by means of (24).

An analogous, block row-wise procedure can be derived if it is preferable to block diagonalize A.

Summary. As noted in [6, 11], algorithms for evaluating univariate matrix functions based on block diagonalization have their deficiencies. In particular, to obtain a very well-conditioned Q, the spectrum of the diagonal blocks can often not be chosen very narrow. Consequently, to yield good accuracy, a large value of k needs to be chosen in the truncated Taylor expansion (24). The Schur-Parlett algorithm [6, 25] has been demonstrated to allow for more narrow block diagonal spectra and is therefore preferred over block diagonalization. It would be desirable to have a bivariate analogue of this algorithm, which ideally would reduce to the well-known Bartels-Stewart algorithm for Sylvester matrix equations if applied to $f(x,y) = 1/(x-y)$. Unfortunately, the derivation of such an analogue appears to be difficult.

It should be stressed that there are far better algorithms available for the two most important special cases of matrix functions; we refer to [15] for linear matrix equations and to [11] for matrix Fréchet derivatives.

6 Extension to multivariate functions

For the sake of clarity, the focus of this paper has been on bivariate matrix functions. However, the extension to arbitrary multivariate functions is rather simple.

First, consider a d-variate polynomial

$$p(x_1,\dots,x_d) = \sum_{i_1=1}^{s_1}\cdots\sum_{i_d=1}^{s_d} p_{i_1,\dots,i_d} x_1^{i_1}\cdots x_d^{i_d} = \sum_{i\in\mathcal{I}} p_i x^i,$$

where we used the usual multiindex notation and $\mathcal{I} = [1,s_1]\times\cdots\times[1,s_d]$. For the evaluation of p at d matrices $A_1\in\mathbb{C}^{n_1\times n_1},\dots,A_d\in\mathbb{C}^{n_d\times n_d}$, we propose to define

$$p\{A_1,\dots,A_d\}(C) := \sum_{i\in\mathcal{I}} p_i\, C\times_1 A_1^{i_1}\times_2 A_2^{i_2}\cdots\times_d A_d^{i_d},$$

where $C\in\mathbb{C}^{n_1\times n_2\times\cdots\times n_d}$ is a tensor of order d and $\times_j$ denotes the j-mode multiplication of a tensor with a matrix [7, 2]. This matches (2) for $d=2$ since $A_1^{i_1}C(A_2^T)^{i_2} = C\times_1 A_1^{i_1}\times_2 A_2^{i_2}$.

For a general function $f(x_1,\dots,x_d)$, tensor Hermite interpolation yields a polynomial $p(x_1,\dots,x_d)$ satisfying

$$\frac{\partial^{|g|}}{\partial x^g}p(\lambda_1,\dots,\lambda_d) = \frac{\partial^{|g|}}{\partial x^g}f(\lambda_1,\dots,\lambda_d), \qquad \begin{array}{l} g=(g_1,\dots,g_d), \\ g_k = 0,\dots,\mathrm{ind}_{\lambda_k}(A_k)-1,\ k=1,\dots,d, \end{array} \tag{26}$$

for every tuple of eigenvalues $\lambda_1\in\Lambda(A_1),\dots,\lambda_d\in\Lambda(A_d)$, provided of course that all required mixed derivatives of f exist and are continuous. The d-variate matrix function associated with f can then be defined as $f\{A_1,\dots,A_d\} := p\{A_1,\dots,A_d\}$, which is a linear operator on $\mathbb{C}^{n_1\times n_2\times\cdots\times n_d}$.

Mutatis mutandis, all results presented for bivariate matrix functions can be expected to admit d-variate extensions. For example, if f is holomorphic on an open set $\Omega = \Omega_1\times\cdots\times\Omega_d$, with $\Lambda(A_k)\subset\Omega_k$, and continuous on $\overline{\Omega}$, the d-variate analogue of the Cauchy integral formula of Theorem 2.8 becomes

$$f\{A_1,\dots,A_d\}(C) = \frac{1}{(2\pi\mathrm{i})^d}\oint_{\Gamma_1}\cdots\oint_{\Gamma_d} f(x_1,\dots,x_d)C\times_1(x_1I-A_1)^{-1}\cdots\times_d(x_dI-A_d)^{-1}\mathrm{d}x,$$

where Γ_k is the contour of Ω_k.

The multivariate matrix function for $f(x_1, \ldots, x_d) = 1/(x_1 + \cdots + x_d)$ can be used to solve discretizations of separable partial differential equations, see [8, 19]. We are not aware of any other applications.

7 Conclusions and Outlook

The definition of bivariate matrix function proposed in this paper has resulted in the unification and (mild) improvements of some existing results for linear matrix equations and matrix Fréchet derivatives. It remains to be seen whether other applications fit into our framework.

This paper has only discussed basic results and briefly touched computational aspects. There is evidence that the concept of bivariate matrix functions may offer a more abstract view and possibly new insights for a variety of other, more advanced results. First, existing Krylov subspace methods for Lyapunov matrix equations [14, 26] could be extended and viewed as bivariate polynomial matrix approximations. Second, an analogue of Theorem 4.1 for Fréchet derivatives of bivariate matrix functions could lead to a more efficient way to compute condition numbers for linear matrix equations, cf. [10, Sec. 16.3].

References

[1] A. C. Ahlin. A bivariate generalization of Hermite's interpolation formula. *Math. Comp.*, 18:264–273, 1964.

[2] B. W. Bader and T. G. Kolda. Algorithm 862: MATLAB tensor classes for fast algorithm prototyping. *ACM Trans. Math. Software*, 32(4):635–653, 2006.

[3] R. Bhatia and K. B. Sinha. Derivations, derivatives and chain rules. *Linear Algebra Appl.*, 302/303:231–244, 1999.

[4] Ju. L. Daleckiĭ. Differentiation of non-Hermitian matrix functions depending on a parameter. *Amer. Math. Soc. Transl., Series* 2, 47:73–87, 1965.

[5] Ju. L. Daleckiĭ and S. G. Kreĭn. Integration and differentiation of functions of Hermitian operators and applications to the theory of perturbations. *Amer. Math. Soc. Transl., Series* 2, 47:1–30, 1965.

[6] P. I. Davies and N. J. Higham. A Schur–Parlett algorithm for computing matrix functions. *SIAM J. Matrix Anal. Appl.*, 25(2):464–485, 2003.

[7] L. De Lathauwer, B. De Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.*, 21(4):1253–1278, 2000.

[8] L. Grasedyck. Existence and computation of low Kronecker-rank approximations for large linear systems of tensor product structure. *Computing*, 72(3-4):247–265, 2004.

[9] M. Gu. Finding well-conditioned similarities to block-diagonalize nonsymmetric matrices is NP-hard. *Journal of Complexity*, 11(3):377–391, September 1995.

[10] N. J. Higham. *Accuracy and Stability of Numerical Algorithms.* SIAM, Philadelphia, PA, second edition, 2002.

[11] N. J. Higham. *Functions of matrices*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008.

[12] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, Cambridge, 1991.

[13] E. Isaacson and H. B. Keller. *Analysis of numerical methods*. Dover Publications Inc., New York, 1994. Corrected reprint of the 1966 original.

[14] I. M. Jaimoukha and E. M. Kasenally. Krylov subspace methods for solving large Lyapunov equations. *SIAM J. Numer. Anal.*, 31:227–251, 1994.

[15] I. Jonsson and B. Kågström. Recursive blocked algorithm for solving triangular systems. I. one-sided and coupled Sylvester-type matrix equations. *ACM Trans. Math. Software*, 28(4):392–415, 2002.

[16] B. Kågström. Numerical computation of matrix functions. Report UMINF-58.77, Department of Information Processing, University of Umeå, Sweden, July 1977.

[17] S. G. Krantz. *Function Theory of Several Complex Variables*. John Wiley & Sons Inc., New York, 1982.

[18] M. G. Kreĭn. *Lektsii po teorii ustoichivosti reshenii differentsialnykh uravnenii v Banakhovom prostranstve*. Izdat. Akad. Nauk Ukrain. SSR, Kiev, 1964.

[19] D. Kressner and C. Tobler. Krylov subspace methods for linear systems with tensor product structure. *SIAM J. Matrix Anal. Appl.*, 31(4):1688–1714, 2010.

[20] P. Lancaster. Explicit solutions of linear matrix equations. *SIAM Rev.*, 12:544–566, 1970.

[21] R. Mathias. Approximation of matrix-valued functions. *SIAM J. Matrix Anal. Appl.*, 14(4):1061–1063, 1993.

[22] R. Mathias. A chain rule for matrix functions and applications. *SIAM J. Matrix Anal. Appl.*, 17(3):610–620, 1996.

[23] I. Najfeld and T. F. Havel. Derivatives of the matrix exponential and their computation. *Advances in Applied Mathematics*, 16:321–375, 1995.

[24] G. Opitz. Steigungsmatrizen. *Z. Angew. Math. Mech.*, 44:T52–T54, 1964.

[25] B. N. Parlett and K. C. Ng. Development of an accurate algorithm for $\exp(Bt)$. Technical Report PAM-294, Center for Pure and Applied Mathematics, University of California, Berkeley, August 1985.

[26] Y. Saad. Numerical solution of large Lyapunov equations. In *Signal processing, scattering and operator theory, and numerical methods (Amsterdam, 1989)*, volume 5 of *Progr. Systems Control Theory*, pages 503–511. Birkhäuser Boston, Boston, MA, 1990.

[27] E. Stickel. On the Fréchet derivative of matrix functions. *Linear Algebra Appl.*, 91:83–88, 1987.

Research Reports

No.	Authors/Title
10-22	*D. Kressner* Bivariate matrix functions
10-21	*C. Jerez-Hanckes and J.-C. Nédélec* Variational forms for the inverses of integral logarithmic operators over an interval
10-20	*R. Andreev* Space-time wavelet FEM for parabolic equations
10-19	*V.H. Hoang and C. Schwab* Regularity and generalized polynomial chaos approximation of parametric and random 2nd order hyperbolic partial differential equations
10-18	*A. Barth, C. Schwab and N. Zollinger* Multi-Level Monte Carlo Finite Element method for elliptic PDE's with stochastic coefficients
10-17	*B. Kågström, L. Karlsson and D. Kressner* Computing codimensions and generic canonical forms for generalized matrix products
10-16	*D. Kressner and C. Tobler* Low-Rank tensor Krylov subspace methods for parametrized linear systems
10-15	*C.J. Gittelson* Representation of Gaussian fields in series with independent coefficients
10-14	*R. Hiptmair, J. Li and J. Zou* Convergence analysis of Finite Element Methods for $H(\mathrm{div};\Omega)$-elliptic interface problems
10-13	*M.H. Gutknecht and J.-P.M. Zemke* Eigenvalue computations based on IDR
10-12	*H. Brandsmeier, K. Schmidt and Ch. Schwab* A multiscale hp-FEM for 2D photonic crystal band
10-11	*V.H. Hoang and C. Schwab* Sparse tensor Galerkin discretizations for parametric and random parabolic PDEs. I: Analytic regularity and gpc-approximation
10-10	*V. Gradinaru, G.A. Hagedorn, A. Joye* Exponentially accurate semiclassical tunneling wave functions in one dimension