Computing matrix functions of infinite quasi-Toeplitz matrices

D.A. Bini
Università di Pisa

dario.bini@unipi.it

Joint work with
Stefano Massei (SNS Pisa) and Beatrice Meini (Università di Pisa)

Luminy 2016

The problem and its motivation

Certain mathematical models from the applications lead to solve computational problems of the kind

- Compute the "minimal" solution of the matrix equation

$$A_2X^2 + A_1X + A_0 = 0$$

- Compute the function

$$\exp(A) = \sum_{i=0}^{\infty} \frac{1}{i!} A^i$$

where A_0, A_1, A_2 and A are **infinite matrices** of the kind

$$T + E, \quad T = (t_{i,j})_{i,j \in \mathbb{Z}^+}, \quad E = (e_{i,j})_{i,j \in \mathbb{Z}^+}$$

with

$$t_{i,j} = a_{j-i}, \quad \sum_{k \in \mathbb{Z}} |a_k| < \infty, \sum_{i,j \in \mathbb{Z}^+} |e_{i,j}| < \infty$$

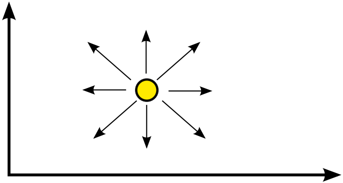
The problem and its motivation

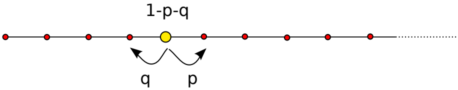
That is, they are the sum of an **infinite Toeplitz matrix** $T(a)$ associated with a sequence a_k and of an infinite matrix E having a finite sum of the moduli of its entries

A typical example originated in the analysis of **random walks on the quarter plane**, or along a half-line is

$$T = \begin{bmatrix} a_0 & a_1 & & & \\ a_{-1} & a_0 & a_1 & & \\ & a_{-1} & a_0 & a_1 & \\ & & \ddots & \ddots & \ddots \end{bmatrix}, \quad E = \begin{bmatrix} b_0 & b_1 & 0 & \dots \\ 0 & 0 & 0 & \dots \\ \vdots & \vdots & \ddots & \dots \end{bmatrix}$$

where $A_0 + A_1 + A_2 - I$, and A are stochastic or sub-stochastic



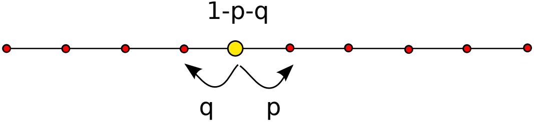


The problem and its motivation

Another situation is the case where the matrices have finite (large) size and still have the form Toeplitz + correction

$$T = \begin{bmatrix} a_0 & a_1 & & & & \\ a_{-1} & a_0 & a_1 & & & \\ & a_{-1} & a_0 & a_1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & a_{-1} & a_0 & a_1 \\ & & & & a_{-1} & a_0 \end{bmatrix} + \begin{bmatrix} b_0 & b_1 & & & & \\ 0 & 0 & 0 & & & \\ & 0 & 0 & \ddots & & \\ & & \ddots & \ddots & 0 & \\ & & & 0 & 0 & 0 \\ & & & & c_{-1} & c_0 \end{bmatrix}$$

This case is encountered for instance in the analysis of bidimensional random walks in a semi-infinite stripe



Some related literature

Recently, some interest has been addressed to this kind of problems under different forms

- D. Kressner and R. Luce (2016) analyze the case of the exponential function for finite Toeplitz matrices
- D.A.B., S. Dendievel, G. Latouche, B. Meini (2015), analyze the case of the exponential function for block triangular block Toeplitz matrices

Applications to Markov chains and queueing models are given by

- M. Miyazawa (2011) for infinite random walks
- M. Kobayashi and M. Miyazawa (2012) for double QBDs
- S. Dendievel and G. Latouche (2014) for fluid queues

Some definitions and properties

- **Wiener class**: $\mathcal{W} = \{a(z) = \sum_{i\in\mathbb{Z}} a_i z^i : \quad \sum_{i\in\mathbb{Z}} |a_i| < \infty\}$
- **Toeplitz matrix associated with** $a(z) \in \mathcal{W}$: $\quad T(a) = (t_{i,j})$, $t_{i,j} = a_{j-i}$, $i, j \in \mathbb{Z}^+$

Properties

- $\mathcal{W}$ with the norm $\|a\|_{\mathcal{W}} = \sum_{i\in\mathbb{Z}} |a_i|$ is a **Banach algebra**, in particular,

$$a(z), b(z) \in \mathcal{W} \;\Rightarrow\; c(z) := a(z)b(z) \in \mathcal{W}, \quad \|c\|_{\mathcal{W}} \le \|a\|_{\mathcal{W}} \cdot \|b\|_{\mathcal{W}}$$

- $\|T(a)\|_p \le \|a\|_{\mathcal{W}}$, for any $p \ge 1$, included $p = \infty$

[Boettcher and Grudsky]

Some properties

Theorem [Boettcher and Grudsky] For $a(z), b(z) \in \mathcal{W}$, let $c(z) := a(z)b(z)$, then

$$T(a)T(b) = T(c) - H(a_-)H(b_+),$$

where $H(a_-) = (a_{-i-j+1})_{i,j\in\mathbb{Z}^+}$, $H(b_+) = (b_{i+j-1})_{i,j\in\mathbb{Z}^+}$ are **Hankel**

$$H(a_-) = \begin{bmatrix} a_{-1} & a_{-2} & a_{-3} & \dots \\ a_{-2} & a_{-3} & \iddots & \iddots \\ a_{-3} & \iddots & \iddots & \iddots \\ \vdots & \iddots & \iddots & \iddots \end{bmatrix}, \quad H(b_+) = \begin{bmatrix} b_1 & b_2 & b_3 \dots & \\ b_2 & b_3 & \iddots & \dots \\ b_3 & \iddots & \iddots & \iddots \\ \vdots & \iddots & \iddots & \iddots \end{bmatrix}$$

Moreover

$$\|H(a_-)\|_p \le \|a\|_{\mathcal{W}}, \quad \|H(b_+)\|_p \le \|b\|_{\mathcal{W}}$$

In words:

The product of two infinite Toeplitz matrices differs from a Toeplitz matrix by a correction which is located mostly in the upper left corner

New results

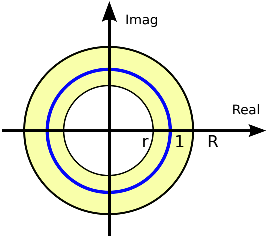
For $E = (e_{i,j})_{i,j\in\mathbb{Z}^+}$ define $\|E\|_{\mathcal{F}} := \sum_{i,j\in\mathbb{Z}^+} |e_{i,j}|$, and

$$\mathcal{F} = \{F = (f_{i,j})_{i,j\in\mathbb{Z}^+} : \quad \|F\|_{\mathcal{F}} < +\infty\}$$

Define **Quasi-Toeplitz (QT)** a matrix of the form

$$A = T(a) + E, \quad a(z) \in \mathcal{W}, \quad \|E\|_{\mathcal{F}} < \infty$$

Define **Analitically Quasi-Toeplitz (AQT)** a QT matrix where $a(z) \in \mathcal{W}$ is analytic in some annulus $\mathbb{A}(r, R) = \{z \in \mathbb{C} : \quad r < |z| < R\}$ where $r < 1 < R$.



New results

Theorem 1

The set $\mathcal{QT}$ of QT matrices is a Banach space with the norm

$$\|A\|_{\mathcal{QT}} = \|a\|_{\mathcal{W}} + \|E\|_{\mathcal{F}}, \qquad \text{where } A = T(a) + E$$

Theorem 2

The set $\mathcal{AQT}$ of AQT matrices is a normed matrix algebra with the norm

$$\|A\|_{\mathcal{AQT}} = \|a\|_{\mathcal{W}} + \|a'\|_{\mathcal{W}} + \|E\|_{\mathcal{F}}, \qquad \text{where } A = T(a) + E$$

Moreover $\|AB\|_{\mathcal{AQT}} \le \|A\|_{\mathcal{AQT}} \cdot \|B\|_{\mathcal{AQT}}$

More precisely if $A = T(a) + E_a$ and $B = T(b) + E_b$ are AQT matrices, then $C = AB$ is an AQT matrix, that is

$$C = T(c) + E_c, \quad c(z) = a(z)b(z), \quad \sum_{i,j \in \mathbb{Z}^+} |e_{i,j}^{(c)}| < \infty$$

New results

Remark: $\mathcal{AQT}$ is not complete (not Banach). In fact, there are Cauchy sequences $A_k \in \mathcal{AQT}$ which do not converge in $\mathcal{AQT}$

However, since $\mathcal{AQT} \subset \mathcal{QT}$ and since $\|\cdot\|_{\mathcal{QT}} \le \|\cdot\|_{\mathcal{AQT}}$, then

$$A_k \text{ Cauchy in } \mathcal{AQT} \quad \Rightarrow \quad A_k \text{ Cauchy in } \mathcal{QT}$$

Thus, since $\mathcal{QT}$ is Banach then $\lim_k A_k \in \mathcal{QT}$

Theorem 3

Any Cauchy sequence in $\mathcal{AQT}$ has a limit in $\mathcal{QT}$ in the form $T(c) + E_c$, where $c(z) \in \mathcal{W}$ is not necessarily analytic

There are very interesting **computational consequences**

Computational consequences

- Any AQT matrix can be approximated to any desired precision with a **finite set of parameters**
- An approximated arithmetic can be defined for AQT matrices, it formally behaves like the floating point arithmetic
- Computing the exponential of AQT matrices
- Solving matrix equations
- Computing matrix functions by means of power series
- Computing matrix functions by means of integration
- All the above stuff for **finitely large Toeplitz matrices**

Computing the exponential of a Toeplitz matrix

Let $A = T(a)$, consider

$$\exp(A) = \sum_{i=0}^{\infty} \frac{1}{i!} A^i$$

We have the following properties concerning the powers of a Toeplitz matrix A where $E_1 = 0$

$$A^k = T(a^k) + E_k$$
$$E_k = T(a)E_{k-1} + H(a_-)H((a^{k-1})_+), \quad k \geq 2$$

$$\|E_k\|_p \leq (k-1)\|a\|_{\mathcal{W}}^k, \qquad \|E_k\|_{\mathcal{F}} \leq \frac{k(k-1)}{2}\|a'\|_{\mathcal{W}}^2 \|a\|_{\mathcal{W}}^{k-2}$$

Moreover,

$$S_k := \sum_{i=0}^{k} \frac{1}{i!} A^k = T(\sum_{i=0}^{k} \frac{1}{i!} a^i) + \sum_{i=0}^{k} \frac{1}{i!} E_i =: T(s_k) + F_k$$

Computing the exponential of a Toeplitz matrix

Consequently, we may prove the following

Theorem 4

The sequence $S_k = \sum_{i=0}^{k} \frac{1}{i!} A^i$ is a **Cauchy sequence**. There exists

$$\exp(A) = \lim_k S_k = \sum_{i=0}^{\infty} \frac{1}{i!} A^i \in \mathcal{AQT}$$

Moreover, $\exp(A) = T(\exp(a)) + E_{\exp}, \quad \|E_{\exp}\|_{\mathcal{F}} \le \frac{1}{2} \|a'\|_{\mathcal{W}}^2 \exp(\|a\|_{\mathcal{W}})$

The function $\exp(a)$ as well as the matrix $E_{\exp}$ are numerically computable with a finite number of ops. The algorithm computes at each step both

$$s_k(z) = \sum_{i=0}^{k} \frac{1}{i!} z^i, \quad F_k = \sum_{i=0}^{k} \frac{1}{i!} E_i$$

Complexity analysis

band $\cdot$ log(band): for updating $s_k(z)$ given $s_{k-1}(z)$
band $\cdot$ rank $\cdot$ log(band): for updating E_k given E_{k-1}
band $\cdot$ (rank)^2: for compression

where

band is the maximum among the band-width in $T(a)$ and in $T(s_k)$, and the size of E_k
rank is the maximum rank of E_k

Their values depend on the decay of the Fourier coefficients of $\exp(a(z))$ and of $a(z)$, and on the decay of the entries of $E_{\exp}$.

We have provided a Matlab implementation of the algorithm

Computing the exponential of an AQT-matrix

A similar result can be proved for an AQT-matrix $A = T(a) + E$.

We have the following properties concerning the powers of A

$$
\begin{aligned}
A^k &= T(a^k) + E_k \\
E_k &= T(a)E_{k-1} + H(a_-)H((a^{k-1})_+) + ET(a^{k-1}), \quad k \geq 2
\end{aligned}
$$

with $E_1 = E$.

Consequently

$$
\begin{aligned}
&\exp(A) = T(\exp(a)) + E_{\exp}, \\
&\|E_{\exp}\|_{\mathcal{F}} \leq (\frac{1}{2}\|a'\|^2_{\mathcal{W}} + \|E\|_{\mathcal{F}}) \exp(\|a\|_{\mathcal{W}} + \|E\|_{\mathcal{F}})
\end{aligned}
$$

The finite case

Similar results can be proved for finite Toeplitz matrices where

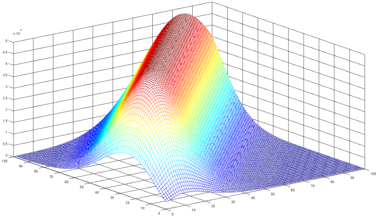
$$
\begin{aligned}
A^k &= T(a^k) + E_k + F_k \\
E_k &= T(a)E_{k-1} + H(a_-)H((a^{k-1})_+), \quad k \geq 2 \\
F_k &= T(a)^T F_{k-1} + JH(a_+)J \cdot JH((a^{k-1})_-)J, \quad k \geq 2
\end{aligned}
$$

where $J = (\delta_{i,n-j+1})$ is the **flip matrix** and

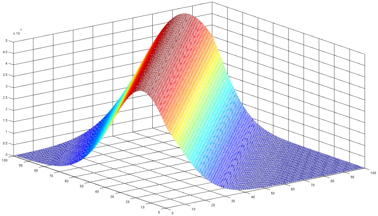
$$
E_k = \begin{bmatrix} * & * & * & & & \\ * & * & & & & \\ * & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \end{bmatrix}, \quad F_k = \begin{bmatrix} & & & & & \\ & & & & & \\ & & & & & \\ & & & & & * \\ & & & & * & * \\ & & & * & * & * \end{bmatrix}
$$

Some numerical experiments

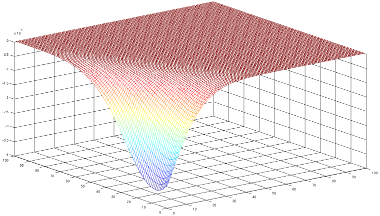
$$a(z) = \sum_{i=1}^{10} z^i + \sum_{i=0}^{7} z^{-i}, \quad A = T(a), \quad \exp(T(a)) = T(\exp(a)) + E_{\exp}$$



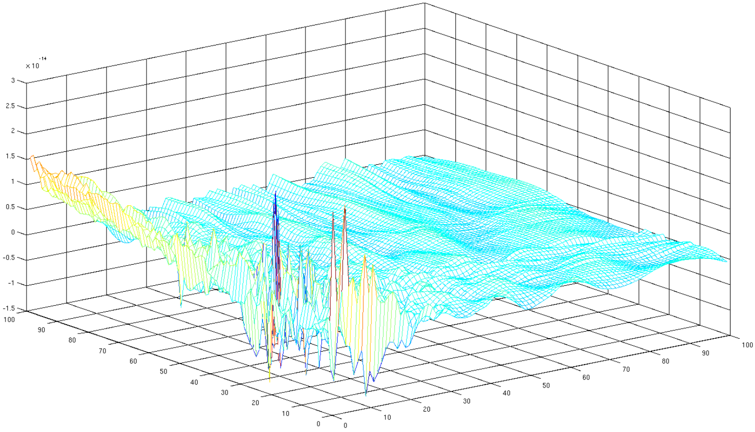
$=$



$+$



Some numerical experiments



Some numerical experiments

$a(z) = \sum_{i=1}^{4} rand\, z^i + \sum_{i=0}^{k} rand\, z^{-i}, \quad A = T(a),$

$B = \exp(A) = T(\exp(a)) + E_{\exp}$

$m = \max(\text{bandwidth, correction size})$ of B

A_{2m} : $2m \times 2m$ leading principal submatrix of A

$C = \mathtt{expm}(A_{2m})$: computed by Matlab

$err = \| B_m - C_m \|_\infty / \|B_m\|_\infty$

t_{AQT}: CPU time in seconds of our algorithm

$t_{\mathtt{expm}}$ CPU time in seconds of the Matlab function `expm`

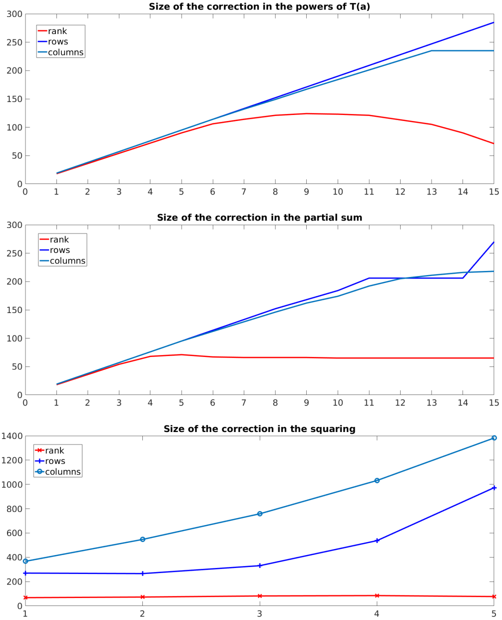
rank: rank of the correction $E_{\exp}$

Experiments performed on a processor i3 with Matlab 8.6.0.267246 (R2015b)

Some numerical experiments

k	t_{AQT}	$t_{\texttt{expm}}$	err	band	rows	cols	rank
10	0.07	0.10	1.5e-14	311	253	177	21
20	0.06	2.49	8.1e-14	812	758	195	19
30	0.08	12.26	2.0e-13	1501	1448	213	15
40	0.11	50.10	2.5e-13	2358	2306	211	10
50	0.14	102.41	4.0e-13	3375	3319	210	10
60	0.15	*	-	4545	4489	164	9
70	0.15	*	-	5863	6594	153	9
80	0.21	*	-	7325	7276	61	9
90	0.25	*	-	8928	8878	51	8
100	0.28	*	-	10671	10622	33	8

Size growth of the correction: $a(z) = \sum_{i=-20}^{20} rand * z^i$



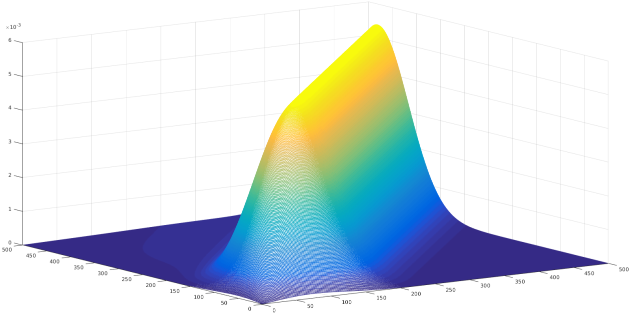
Some numerical experiments

Toeplitz matrices with a finite size

Problem: Option pricing using the Merton model (Kressner-Luce)

size	1024	2048	4096	8192
time	0.5	1.2	3.2	11.3
time (`expm`)	1.0	4.6	35.8	*

rank of the correction : $18 + 18$



Extension to other functions

Recall that $T(a)^k = T(a^k) + E_k$

From the inequality $\|E_k\|_{\mathcal{F}} \le \frac{k(k-1)}{2}\|a'\|_{\mathcal{W}}^2\|a\|_{\mathcal{W}}^{k-2}$ it follows

Theorem.

If $f(x) = \sum_{i=0}^{+\infty} f_i z^i$ is analytic for $|x| \le \rho$ and if $a(z)$ is analytic and $\|a\|_{\mathcal{W}} \le \rho$ then $f(a(z))$ is analytic and

$$f(T(a)) = T(f(a)) + F,$$
$$F = \sum_{k=0}^{+\infty} f_k E_k, \quad \|F\|_{\mathcal{W}} < +\infty$$

A similar result holds for $f(A)$, $A = T(a) + E$

AQT matrix arithmetic

Since the class of AQT matrices is an algebra, we may implement an infinite matrix arithmetic in this class so that any algorithm performing arithmetic operations between matrices can in principle be implemented for AQT matrices.

- It is natural to represent $A = T(a) + E$ by means of the function $a(z)$ and the matrix E
- Since $a(z)$ is represented by a bi-infinite sequence $\{a_k\}_{k\in\mathbb{Z}}$ having decay of the coefficients, we may represent $a(z)$ with a **finite sequence** $a = (a_{n_-}, \dots, a_0, \dots a_{n_+})$ up to an arbitrarily small error.
- The matrix E is represented by means of the pair (U, V) such that $E = UV^T$, U and V have a finite number of columns given by the numerical rank of E.
- The matrices U and V can be truncated to a finite number of rows

AQT matrix arithmetic

- The arithmetic operations of addition and multiplication of two matrices A and B can be reduced to the corresponding operations between the functions $a(z)$ and $b(z)$ and in terms of the matrices U_a, V_a and U_b, V_b
- For instance, for $C = A + B$ one has

$$c(z) = a(z) + b(z)$$
$$U_c = [U_a \ , \ U_b], \quad V_c = [V_a \ , \ V_b]$$

- Similarly we can do for $C = AB$
- An SVD-based compression technique is introduced in order to keep small the numerical rank
- A more complicate situation is encountered for infinite matrix inversion. We rely on the Wiener-Hopf factorization

Inverting an AQT matrix: the Wiener-Hopf factorization

Given a function $a(z) \in \mathcal{W}$ there exist $u(z) = \sum_{i=0}^{\infty} u_i z^i$ and $\ell(z) = \sum_{i=0}^{\infty} \ell_i z^i$ in $\mathcal{W}$ and an integer κ such that

$$a(z) = u(z) z^{\kappa} \ell(z^{-1}) \quad \textbf{Wiener-Hopf factorization}$$

where $u(z) \neq 0$, $\ell(z) \neq 0$ for $|z| < 1$ If $\kappa = 0$ the factorization is said **canonical**

Matrix version of the canonical factorization

$$T(a) = \begin{bmatrix} u_0 & u_1 & u_2 & \dots & \\ & u_0 & u_1 & u_2 & \ddots \\ & & \ddots & \ddots & \ddots \end{bmatrix} \begin{bmatrix} \ell_0 & & & \\ \ell_1 & \ell_0 & & \\ \ell_2 & \ell_1 & \ell_0 & \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix} = T(u) T(\ell)^T$$

Inverting an AQT matrix: the Wiener-Hopf factorization

It turns out that

$$T(a)^{-1} = T(\ell^{-1})^T T(u^{-1})$$

This provides an algorithm for computing the AQT matrix $T(a)^{-1}$ for $a(z)$ analytic in $\mathbb{T}(r, R)$

1. compute the canonical factorization of $a(z)$
2. compute the power series of $1/u(z)$ and $1/\ell(z)$ by using evaluation and interpolation
3. compute the product $c(z)$ of the two power series
4. represent $T(a)^{-1}$ as $T(c) + E$, where E is product of two Hankel matrices

General properties of AQT arithmetic

A matrix A in $\mathcal{AQT}$ can be represented as

$$A = T(a) + E_a + \mathcal{E}_a =: \widehat{A} + \mathcal{E}_a$$

where $a(z)$ is a Laurent polynomial, and E_a has a finite number of nonzero entries.

$\widehat{A}$ is the finite representation of A

$\mathcal{E}_a$ is the representation error

If $A = \widehat{A} + \mathcal{E}_a$ and $B = \widehat{B} + \mathcal{E}_b$ are represented this way, define $\widehat{C} = \text{trunc}(\widehat{A}\widehat{B})$ so that

$$\widehat{A}\widehat{B} = \widehat{C} + \mathcal{E}, \quad \text{where } \mathcal{E} \text{ is the representation error}$$

then

$$C = \widehat{C} + \mathcal{E} + \widehat{A}\mathcal{E}_b + \mathcal{E}_1\widehat{B} + \mathcal{E}_a\mathcal{E}_b$$

In blue the inherent error

Similar relations hold for the other operations

Application to solving a matrix equation

Equations of the kind

$$BX^2 + AX + C = 0$$

encountered in queuing models can be solved with Cyclic Reduction

$$\begin{aligned}
B_{k+1} &= -B_k A_k^{-1} B_k \\
C_{k+1} &= -C_k A_k^{-1} C_k \\
A_{k+1} &= A_k - B_k A_k^{-1} C_k - C_k A_k^{-1} B_k \\
\widehat{A}_{k+1} &= A_k - B_k A_k^{-1} C_k
\end{aligned}$$

where $A_0 = \widehat{A}_0 = A,\ B_0 = B,\ C_0 = C$

For the minimal nonnegative solution G it holds

$$\lim_k \widehat{A}_k^{-1} B = G$$

moreover convergence is quadratic

Application to solving a matrix equation

Relying on the AQT matrix arithmetic we can solve semi-infinite QBD problems in few seconds

Example: Ten instances of the two-node Jackson network from [Motyer and Taylor 2006]: problems 1–10

$$A_{-1} = \begin{bmatrix} (1-q)\mu_2 & q\mu_2 & & \\ & (1-q)\mu_2 & q\mu_2 & \\ & & \ddots & \ddots \end{bmatrix},$$

$$A_0 = \begin{bmatrix} -(\lambda_1+\lambda_2+\mu_2) & \lambda_1 & & \\ (1-p)\mu_1 & -(\lambda_1+\lambda_2+\mu_1+\mu_2) & \lambda_1 & \\ & \ddots & \ddots & \ddots \end{bmatrix},$$

$$A_1 = \begin{bmatrix} \lambda_2 & & \\ p\mu_1 & \lambda_2 & \\ & \ddots & \ddots \end{bmatrix},$$

Application to solving a matrix equation

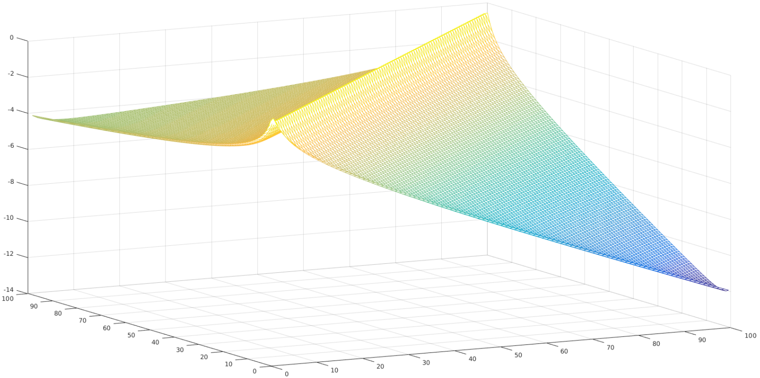
Problem	λ_1	λ_2	μ_1	μ_2	p	q
1	1	0	1.5	2	1	0
2	1	0	2	1.5	1	0
3	0	1	1.5	2	0	1
4	0	1	2	1.5	0	1
5	1	1	2	2	0.1	0.8
6	1	1	2	2	0.8	1
7	1	1	2	2	0.4	0.4
8	1	1	10	10	0.5	0.5
9	1	5	10	15	0.4	0.9
10	5	1	15	10	0.9	0.4

Application to solving a matrix equation

Problem	CPU	Res	Band	Rows	Columns	Rank
1	2.61	$8.6 \cdot 10^{-16}$	561	541	138	8
2	2.91	$1.5 \cdot 10^{-15}$	561	555	145	8
3	0.29	$1.1 \cdot 10^{-16}$	143	89	66	8
4	2.32	$6.8 \cdot 10^{-16}$	463	481	99	9
5	0.48	$1.2 \cdot 10^{-15}$	233	108	148	9
6	7.96	$1.9 \cdot 10^{-14}$	455	462	153	10
7	29.00	$4.3 \cdot 10^{-15}$	1423	1543	247	13
8	1.01	$1.1 \cdot 10^{-15}$	366	348	40	6
9	0.31	$5.4 \cdot 10^{-16}$	157	81	86	8
10	1.25	$1.1 \cdot 10^{-15}$	268	241	107	8

Application to solving a matrix equation

Log plot of a portion of the solution G



Once again general matrix functions

Matrix functions can be expressed either in terms of power series expansions

$$f(A) = \sum_{i=0}^{\infty} f_i z^i$$

or in terms of numerical integration

$$f(A) = \int_{\gamma} f(z)(zI - A)^{-1} dz$$

where γ is a Jordan curve

We may rely on these expressions and on AQT matrix arithmetic and extend, under suitable conditions, $f(z)$ to $f(A)$ by providing effective computational methods

Conclusions

- The class of AQT matrices has been introduced; we proved that it is a normed matrix algebra contained in a Banach space
- A general framework to compute the matrix exponential and any other matrix function of an AQT matrix expressed by means of power series has been introduced
- Quasi Toeplitz matrix arithmetic has been introduced and implemented
- Applications to solving quadratic matrix equations have been given
- Applications are shown to compute matrix functions through Cauchy integrals
- Extensions to finitely large Toeplitz matrices are given

Thank you