*Dario A. Bini, Stefano Massei, Leonardo Robol*

# Efficient cyclic reduction for QBDs with rank structured blocks: algorithm and analysis

University of Pisa

Scuola Normale Superiore

KU Leuven

stefano.massei@sns.it

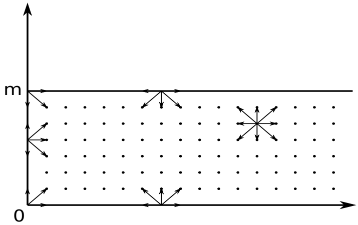
Budapest, 28 June 2016

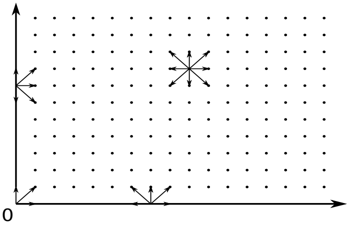
# Quasi-birth-death processes

A QBD process, in discrete time, is a bidimensional Markov chain whose transition probability matrix has the tridiagonal block Toeplitz structure

$$P = \begin{pmatrix} B_0 & B_1 & & & \\ A_{-1} & A_0 & A_1 & & \\ & A_{-1} & A_0 & A_1 & \\ & & A_{-1} & A_0 & \ddots \\ & & & \ddots & \ddots \end{pmatrix},$$

with $A_i, B_i \in \mathbb{R}^{m\times m}$ ($m \in \mathbb{N} \cup \{+\infty\}$) non negative and $P$ stochastic.



# The main problem

Suppose $m < \infty$ and let the matrix $P$ be irreducible and nonperiodic.
We consider the computation of the stationary distribution of the QBD, i.e. an infinite vector $\boldsymbol{\pi}$ such that

$$\boldsymbol{\pi}^T P = \boldsymbol{\pi}^T, \quad \boldsymbol{\pi} \geq 0, \quad \text{and} \quad \| \boldsymbol{\pi} \|_1 = 1.$$

Due to the matrix-geometric property of $\boldsymbol{\pi}$, a crucial step consists in finding the minimal non negative solution $G$ of the quadratic matrix equation:

$$\boxed{X = A_{-1} + A_0 X + A_1 X^2, \quad X \in \mathbb{R}^{m \times m}.}$$

Many numerical methods have been proposed to address the problem and most of them are designed to deal with the general case where the block coefficients $A_{-1}$,$A_0$ and $A_1$ have no particular structure.

# Cyclic Reduction

The method on which we are going to focus is the Cyclic Reduction.
Its iterative scheme requires the computation of four sequences of matrices, $A_i^{(h)}$, $i = -1, 0, 1$ and $\hat{A}_0^{(h)}$, which follow the recurrence relations:

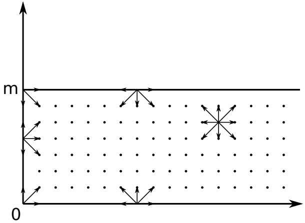
$$\boxed{\begin{aligned} A_1^{(h+1)} &= A_1^{(h)} \, (I - A_0^{(h)})^{-1} \, A_1^{(h)}, \\ A_0^{(h+1)} &= A_0^{(h)} + A_1^{(h)} \, (I - A_0^{(h)})^{-1} \, A_{-1}^{(h)} \; + \; A_{-1}^{(h)} \, (I - A_0^{(h)})^{-1} \, A_1^{(h)}, \\ A_{-1}^{(h+1)} &= A_{-1}^{(h)} \, (I - A_0^{(h)})^{-1} \, A_{-1}^{(h)}, \\ \hat{A}_0^{(h+1)} &= \hat{A}_0^{(h)} + A_1^{(h)} \, (I - A_0^{(h)})^{-1} \, A_{-1}^{(h)}. \end{aligned}}$$

with $A_i^{(0)} = A_i$, $i = -1, 0, 1$ and $\hat{A}_0^{(0)} = A_0$.
After each step, an approximation of the matrix $G$ is provided by

$$(I - \hat{A}_0^{(h)})^{-1} A_{-1}.$$

Under mild hypothesis applicability and quadratic convergence are guaranteed.
The cost of each iteration is $\mathcal{O}(m^3)$ because it involves four matrix multiplications and the resolution of $2m$ linear systems of size $m$.

For example, consider the case in which $A_i$ is finite tridiagonal for $i = -1, 0, 1$ (Double Quasi-Birth and Death).



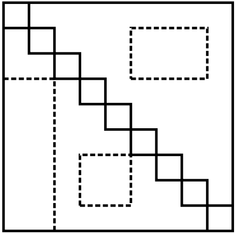

The band structure is lost immediately when applying CR due to the inversions in its iteration scheme.

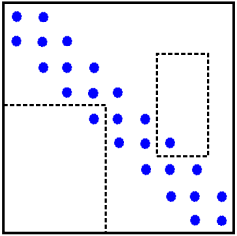
**Goal:** Find an alternative structure to exploit for speeding up the cyclic reduction.

# Quasiseparable rank

**Definition**

*$A \in \mathbb{R}^{m \times m}$ has quasiseparable rank $k$ if the maximum rank among the off diagonal submatrices of $A$ is $k$.*





**Properties:**

(i) $q_{rank}(A + B) \leq q_{rank}(A) + q_{rank}(B)$

(ii) $q_{rank}(A \cdot B) \leq q_{rank}(A) + q_{rank}(B)$

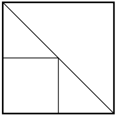
(iii) $q_{rank}(A) = q_{rank}(A^{-1})$

Vandebril, Van Barel and Mastronardi. Matrix computations and semiseparable matrices. Johns Hopkins University Press, 2008.

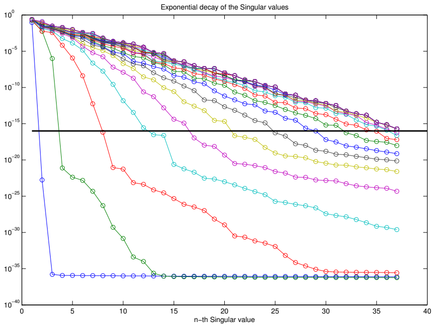
# Experimental evidences

Example:

Cyclic reduction with starting points $A_i \in \mathbb{R}^{1000\times 1000}$ tridiagonal.
Distribution of the singular values of the sub-block

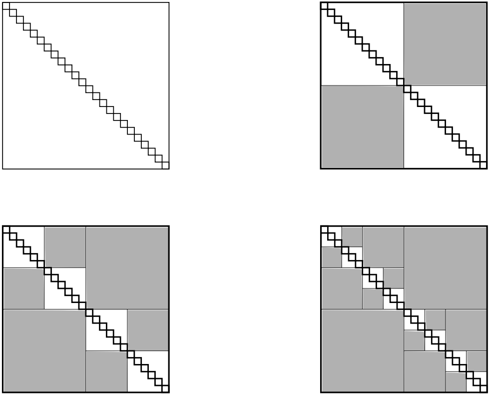


in $A_0^{(h)}$ during the iteration.



# Hierarchical matrices

Strategy: Partition the row and column indices recursively and - at each step - represent the new off-diagonal blocks as low-rank outer products. Stop when the diagonal blocks reach a small dimension and represent them as full matrices.



Matrix operations:

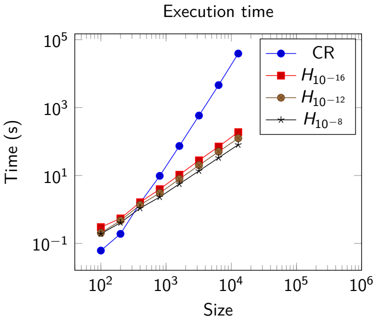
| | |
|---|---|
| Addition | $\mathcal{O}(m\log(m))$ |
| Multiplication | $\mathcal{O}(m\log(m)^2)$ |
| Lin. System | $\mathcal{O}(m\log(m)^2)$ |
| Storage | $\mathcal{O}(m\log(m))$ |

Börm, Grasedyck and Hackbusch. Hierarchical matrices. Lecture notes, 2003.
Hackbusch. Hierarchical Matrices: Algorithms and Analysis. Springer Berlin, 2016.

# A bunch of applications

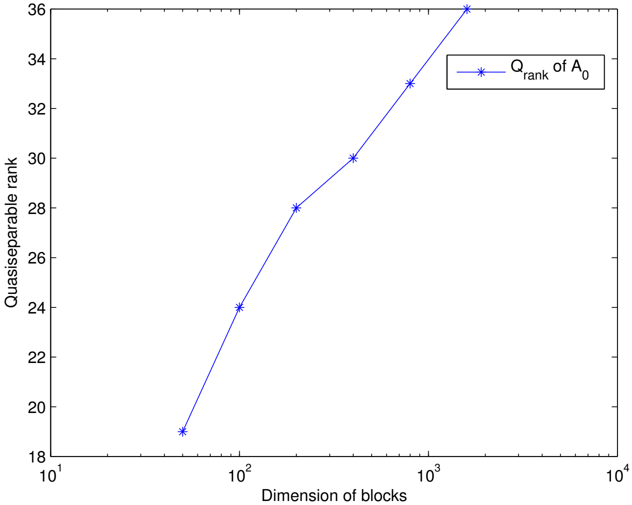
- $\mathcal{H}$-matrix approximation of BEM matrices. [Hackbusch, Sauter,. . . 1990s]
- Matrix sign function iteration in $\mathcal{H}$-arithmetic for solving matrix Lyapunov and Riccati equations. [Grasedyck,Hackbusch,Khoromskij 2004]
- Contour integral + $\mathcal{H}$-matrices for matrix functions [Gavrilyuk et al. 2002].
- $\mathcal{H}$-matrix based preconditioning for FE discretization of 3D Maxwell [Ostrowski et al. 2010].
- Block low-rank approximation of kernel matrices [Si, Hsieh, Dhillon 2014, Wang et al. 2015].
- Clustered low-rank approximation of graphs [Savas, Dhillon 2011].
- Cyclic reduction + $\mathcal{H}$-matrices for quadratic matrix equations with quasiseparable coefficients. [Bini, M., Robol 2016]
- . . .

# Numerical Results/ Tridiagonal: Size VS Execution Time

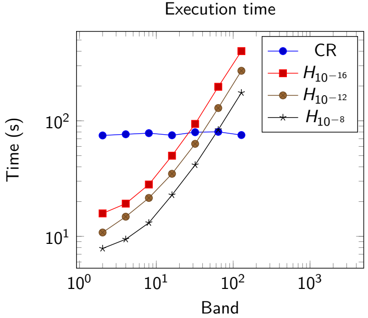



| | CR | | $H_{10^{-16}}$ | | $H_{10^{-12}}$ | | $H_{10^{-8}}$ | |
|---|---|---|---|---|---|---|---|---|
| Size | Time (s) | Residue | Time (s) | Residue | Time (s) | Residue | Time (s) | Residue |
| 100 | $6.04e-02$ | $1.91e-16$ | $2.21e-01$ | $1.79e-15$ | $2.04e-01$ | $8.26e-14$ | $1.92e-01$ | $7.40e-10$ |
| 200 | $1.88e-01$ | $2.51e-16$ | $5.78e-01$ | $1.39e-14$ | $5.03e-01$ | $1.01e-13$ | $4.29e-01$ | $2.29e-09$ |
| 400 | $1.61e+01$ | $2.09e-16$ | $3.32e+00$ | $1.41e-14$ | $2.60e+00$ | $1.33e-13$ | $1.98e+00$ | $1.99e-09$ |
| 800 | $2.63e+01$ | $2.74e-16$ | $4.55e+00$ | $1.94e-14$ | $3.49e+00$ | $2.71e-13$ | $2.63e+00$ | $2.69e-09$ |
| 1600 | $8.12e+01$ | $3.82e-12$ | $1.18e+01$ | $3.82e-12$ | $8.78e+00$ | $3.82e-12$ | $6.24e+00$ | $3.39e-09$ |
| 3200 | $6.35e+02$ | $5.46e-08$ | $3.12e+01$ | $5.46e-08$ | $2.21e+01$ | $5.46e-08$ | $1.51e+01$ | $5.43e-08$ |
| 6400 | $5.03e+03$ | $3.89e-08$ | $7.83e+01$ | $3.89e-08$ | $5.38e+01$ | $3.89e-08$ | $3.58e+01$ | $3.87e-08$ |
| 12800 | $4.06e+04$ | $1.99e-08$ | $1.94e+02$ | $1.99e-08$ | $1.29e+02$ | $1.99e-08$ | $8.37e+01$ | $1.97e-08$ |

# Quasiseparable rank's growth



# Numerical Results/ Size=1600: Band VS Execution Time




| | CR | | $H_{10^{-16}}$ | | $H_{10^{-12}}$ | | $H_{10^{-8}}$ | |
|---|---|---|---|---|---|---|---|---|
| Band | Time (s) | Residue | Time (s) | Residue | Time (s) | Residue | Time (s) | Residue |
| 2 | $7.47e+01$ | $2.11e-16$ | $1.58e+01$ | $6.95e-15$ | $1.08e+01$ | $2.62e-13$ | $7.86e+00$ | $2.57e-09$ |
| 4 | $7.65e+01$ | $1.66e-16$ | $1.92e+01$ | $4.88e-15$ | $1.48e+01$ | $2.36e-13$ | $9.44e+00$ | $3.15e-09$ |
| 8 | $7.82e+01$ | $1.48e-16$ | $2.81e+01$ | $6.11e-15$ | $2.15e+01$ | $2.08e-13$ | $1.31e+01$ | $2.10e-09$ |
| 16 | $7.50e+01$ | $1.35e-16$ | $4.99e+01$ | $4.98e-15$ | $3.48e+01$ | $2.29e-13$ | $2.28e+01$ | $2.08e-09$ |
| 32 | $7.97e+01$ | $1.33e-16$ | $9.40e+01$ | $5.79e-15$ | $6.32e+01$ | $2.01e-13$ | $4.15e+01$ | $2.28e-09$ |
| 64 | $8.03e+01$ | $1.31e-16$ | $1.97e+02$ | $6.79e-15$ | $1.29e+02$ | $1.99e-13$ | $8.37e+01$ | $2.01e-09$ |
| 128 | $7.53e+01$ | $1.28e-16$ | $4.01e+02$ | $5.89e-15$ | $2.71e+02$ | $2.02e-13$ | $1.75e+02$ | $2.15e-09$ |

# Theoretical analysis/ Functional interpretation of CR

We associate at each step of the CR the matrix Laurent polynomial

$$\varphi^{(h)}(z) := -z^{-1} \cdot A_{-1}^{(h)} + (I - A_0^{(h)}) - z \cdot A_1^{(h)}, \qquad \varphi(z) := \varphi^{(0)}(z),$$

and the matrix function defined by recurrence

$$\begin{cases} \psi^{(0)}(z) := \varphi(z)^{-1} \\ \psi^{(h+1)}(z^2) := \frac{1}{2}(\psi^{(h)}(z) + \psi^{(h)}(-z)) \end{cases} \Rightarrow \psi^{(h)}(z^{2^h}) = \frac{1}{2^h} \sum_{j=0}^{2^h-1} \psi^{(0)}(\zeta_j z)$$

**Theorem (Bini,Meini)**

*Let* $z \in \mathbb{C} \setminus \{0\}$ *be such that* $\det(\varphi^{(i)}(z)) \neq 0 \ \forall i = 0, \dots, h$ *and let* $\det(I - A_0^{(i)}) \neq 0 \ \forall i = 0, \dots, h-1$ *then*

$$\varphi^{(i)}(z) = \psi^{(i)}(z)^{-1}, \qquad i = 0, \dots, h.$$

In particular $\varphi^{(h)}(z^{2^h}) = \psi^{(h)}(z^{2^h})^{-1} = \left(\frac{1}{2^h} \sum_{j=0}^{2^h-1} \psi^{(0)}(\zeta_j z)\right)^{-1}$.

**Goal:** Show that the off-diagonal singular values in $A_i^{(h)}$ decay fast.

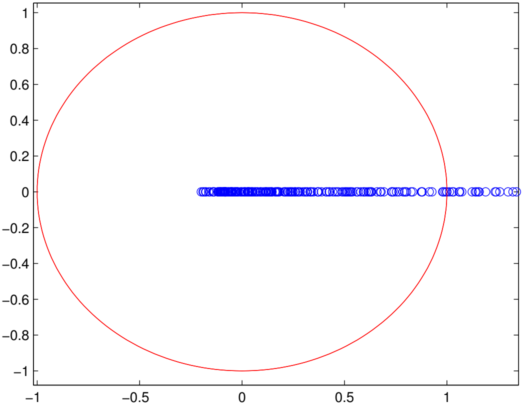
**First approach:**

- Since $\varphi^{(h)}(z) = \psi^{(h)}(z)^{-1}$ — by means of interpolation techniques — we can reformulate the problem in proving the property for $\psi^{(h)}(z)$ with $z$ on the unit circle.
- Using the formula $\psi^{(h)}(z^{2^h}) = \left(\frac{1}{2^h}\sum_{j=0}^{2^h-1}\psi^{(0)}(\zeta_j z)\right)$ and the decay in the Laurent coefficients of $\psi^{(0)}$ we get the property for the latter.

**Pros:** We get explicit exponentially decaying bounds for the singular values of a generic off-diagonal block.
**Cons:** The bounds depend on the gap between the eigenvalues of $\varphi(z)$ which lie inside the unit disc and those outside.

Bini, M. and Robol. Efficient cyclic reduction for Quasi-birth and death problems with rank structured blocks. Applied Numerical Mathematics, to appear in 2016.

# Example: tridiagonal blocks, eigenvalues of $\varphi(z)$




**Blocks dimension:** 200 **Gap:** $2.50e-03$

# Theoretical analysis/ Alternative approach

$$\varphi(z) = \begin{bmatrix} A(z) & B(z) \\ C(z) & D(z) \end{bmatrix} \qquad \psi(z) = \begin{bmatrix} * & * \\ \tilde{C}(z) & * \end{bmatrix} \qquad \psi^{(h)}(z) = \begin{bmatrix} * & * \\ \tilde{C}^{(h)}(z) & * \end{bmatrix}$$

$$\tilde{C}^{(h)}(z^{2^h}) = \frac{1}{2^h} \sum_{j=0}^{2^h-1} \tilde{C}(\zeta_j z)$$

Retrieve directly the Laurent expansion of the generic off-diagonal block $\tilde{C}^{(h)}(z)$ using linear algebra techniques and the Wiener Hopf factorization

$$\varphi(z) = (I - z\ R) \cdot U \cdot (I - z^{-1} G).$$

It turns out the following displacement rank property for $\tilde{C}^{(h)}(z)$:

$$\tilde{C}^{(h)}(z) = X^{(h)}(z) + Y^{(h)}(z), \qquad \Pi = \begin{bmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{bmatrix},$$

$$\operatorname{rank}(\Pi X^{(h)}(z) - X^{(h)}(z) G^t) = 1, \qquad \operatorname{rank}(\Pi Y^{(h)}(z) - Y^{(h)}(z) R^t) = 1.$$

# Theoretical analysis/ Alternative approach

**Theorem (Beckermann)**

*Let $G = WDW^{-1}$ be diagonalizable, call $E, F \subset \mathbb{C}$ be the spectrum of $G$ and $\Pi$ respectively. Let $X$ be a matrix such that $\mathrm{rank}(\Pi X - XG^t) = 1$.*
*Then, the singular values of $X$ can be bounded by*

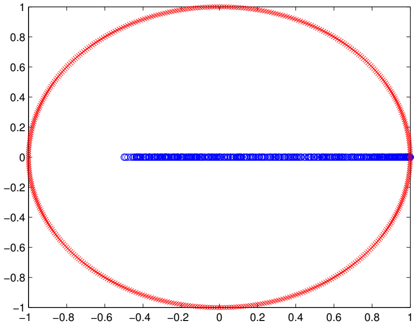
$$\sigma_l(X) \leq \kappa(W) \cdot \|X\|_2 \cdot Z_l(E, F), \qquad Z_l(E, F) = \min_{deg(r)=(l,l)} \frac{\sup_E r(z)}{\inf_F r(z)}.$$

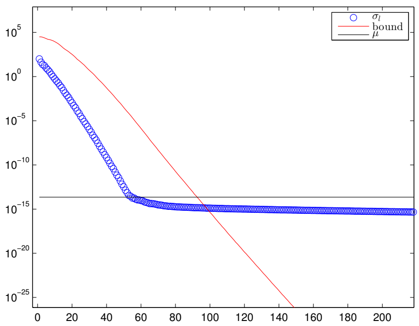
In our framework $F$ is the set of the $2^h$-th roots of the unit while $E$ is the set of eigenvalues of $\varphi(z)$ inside $S^1$ or the reciprocal of those outside.

**Cons:** Explicit general estimates for $Z_l(E, F)$ are not available.
**Pros:** It is possible to find good numerical bounds for the singular values of $X$, even if $E$ gets arbitrarily close to $F$.

# Example of Zolotarev estimates

Example: Eigenvalues of $\varphi(z)$ (tridiagonal blocks) and singular values of $X^{(20)}(i)$.




**Blocks dimension:** 1000 **Gap:** $3.16e-05$

# Conclusions and research lines

- CR numerically preserves the quasiseparable structure and implementing the $\mathcal{H}$-matrix representation we can significantly speed up the algorithm.
- A strong splitting property (wide gap) — for the eigenvalues of $\varphi(z)$ — implies this phenomenon but it is not necessary.
- The decay in the off-diagonal singular values seems to be better described with the quality of some discrete rational approximation problems.
- Test other kind of partitioning — in the hierarchical representation — with respect to different sparsity patterns.
- Extend the analysis and formulate algorithms for infinite phase scenario.