Matrix Analytic Methods for reflected random walks with stochastic restarts

Leonardo Robol, UniPI

<leonardo.robol@unipi.it>

joint work with:

D. A. Bini, S. Massei, B. Meini

MAM10, Feb 2019

Problem of interest

We consider a QBD with infinite levels; in particular:

- It is possible to move from level i to $i-1, i+1$;
- Phase transitions are level independent.

Problem of interest

We consider a QBD with infinite levels; in particular:

- It is possible to move from level i to $i-1, i+1$;
- Phase transitions are level independent.

In this case the probability transition matrix has the form

$$P = \begin{bmatrix} \widehat{A}_0 & A_1 & & \\ A_{-1} & A_0 & A_1 & \\ & \ddots & \ddots & \ddots \end{bmatrix},$$

Computing the steady state probability vector can be recasted to solving the quadratic matrix equations:

$$A_{-1} + A_0 G + A_1 G^2 = G, \qquad R^2 A_{-1} + R A_0 + A_1 = R$$

The infinite phase case

We would like to consider an infinite number of phases; then

- The matrices A_i are semi-infinite matrices;
- We would like to avoid truncating them;
- Therefore, we need approximability with a finite number of parameters.

The infinite phase case

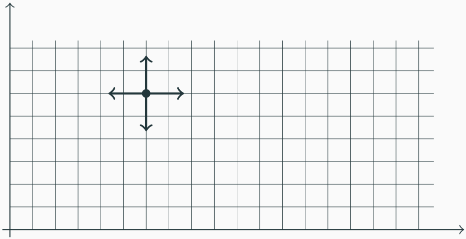
We would like to consider an infinite number of phases; then

- The matrices A_i are semi-infinite matrices;
- We would like to avoid truncating them;
- Therefore, we need approximability with a finite number of parameters.

Our hypothesis: the phase transitions only depend on $j - i$, up to some border conditions. Moreover, these probabilities decay to zero as $|j - i| \to \infty$.

Random walks on a quadrant (double QBD)

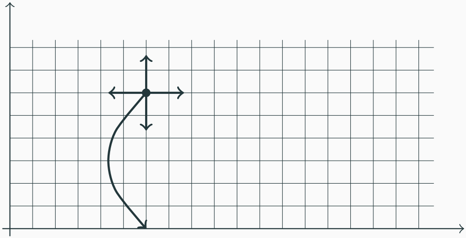
Representative example: random walk on $\mathbb{N} \times \mathbb{N}$



- Allowed moves only to adjacent states.
- Probabilities of going up/down/left/right are eventually independent of position (i, j).
- The problem can be stated also on the infinite strip $\{1, \dots, m\} \times \mathbb{N}$.

Random walks on a quadrant (reset events)

We may allow reset-like events.



- At any time, with some probability ρ, the phase may be reset to 0;
- This can be combined with moving right/left by 1 unit.

The link with Toeplitz matrices

Several queuing problems can be recasted in this framework.

The simpler 1D case considers movements on a line:

The probability transition matrix is as follows:

$$P = \begin{bmatrix} \rho_0 + \rho_{-1} & \rho_1 & & \\ \rho_{-1} & \rho_0 & \rho_1 & \\ & \ddots & \ddots & \ddots \end{bmatrix},$$

where ρ_{-1} is the probability of moving left, ρ_1 right, and $\rho_0 = 1 - \rho_{-1} - \rho_1$: a quasi-Toeplitz semi-infinite matrix.

The restarts

If we allow restarts, or resets, with probability ρ, then the 1D probability transition matrix takes the form

$$P = \begin{bmatrix} 1-\rho_1 & \rho_1 & & & \\ \rho+\rho_{-1} & \rho_0 & \rho_1 & & \\ \rho & \rho_{-1} & \rho_0 & \rho_1 & \\ \vdots & & \ddots & \ddots & \ddots \end{bmatrix},$$

- Not Toeplitz anymore, but only for the first column and some top-left correction.
- The non-Toeplitz part does not have finite support.
- We may consider also resets to a finite number of starting states.

Computing the invariant vector

The steady state probability vector solving $\pi^T P = \pi^T$ can be recovered by finding the minimal non-negative solutions to

$$A_{-1} + A_0 G + A_1 G^2 = G, \qquad R^2 A_{-1} + R A_0 + A_1 = R$$

- Cyclic reduction (CR) is a matrix iteration that converges to the correct solution.
- Needs to perform multiplications, sum, and inversions.
- Complexity $\mathcal{O}(m^3)$ if the matrices A_i are $m \times m$ — for instance a random walk on $\{1, \dots, m\} \times \mathbb{N}$.

A quick look at the matrix iteration

$$\begin{cases} A_0^{(h+1)} = A_0^{(h)} + A_{-1}^{(h)} S^{(h)} A_1^{(h)} + A_1^{(h)} S^{(h)} A_{-1}^{(h)} \\ A_{-1}^{(h+1)} = A_{-1}^{(h)} S^{(h)} A_{-1}^{(h)} \\ A_1^{(h+1)} = A_1^{(h)} S^{(h)} A_1^{(h)} \\ S^{(h)} = (I - A_0^{(h)})^{-1} \end{cases}$$

The initial matrices are $A_i^{(0)} := A_i$, so they are Toeplitz. Is the structure preserved?

A quick look at the matrix iteration

$$\begin{cases} A_0^{(h+1)} = A_0^{(h)} + A_{-1}^{(h)} S^{(h)} A_1^{(h)} + A_1^{(h)} S^{(h)} A_{-1}^{(h)} \\ A_{-1}^{(h+1)} = A_{-1}^{(h)} S^{(h)} A_{-1}^{(h)} \\ A_1^{(h+1)} = A_1^{(h)} S^{(h)} A_1^{(h)} \\ S^{(h)} = (I - A_0^{(h)})^{-1} \end{cases}$$

The initial matrices are $A_i^{(0)} := A_i$, so they are Toeplitz. Is the structure preserved?

- Experimentally, approximately, up to a top-left correction.
- Non-trivial to exploit in practice. In fact, typical implementation ignore the structure and require $\mathcal{O}(m^3)$ flops $\implies$ what about $m = \infty$?.

Basic facts about Toeplitz matrices

$$T(a(z)) = \begin{bmatrix} a_0 & a_1 & a_2 & \dots \\ a_{-1} & a_0 & a_1 & \ddots \\ a_{-2} & a_{-1} & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix}, \qquad a(z) = \sum_{j \in \mathbb{Z}} a_j z^j$$

- $a(z)$ is in the Wiener algebra $\mathcal{W}$ if

$$\sum_{j \in \mathbb{Z}} |a_j| < \infty$$

Hankel matrices

Given a power series $f(z) = \sum_{j=0}^{\infty} f_j z^j$, we have

$$H(f(z)) = \begin{bmatrix} f_1 & f_2 & f_3 & \dots \\ f_2 & f_3 & & \\ f_3 & & \ddots & \\ \vdots & & & \end{bmatrix}$$

the Hankel matrix defined by $f(z)$. Often, we use the notation

$$a^+(z) = \sum_{j\geq 1} a_j z^j, \qquad a^-(z) = \sum_{j\geq 1} a_{-j} z^j.$$

Toeplitz matrices are not an algebra

The power series in $\mathcal{W}$ form an algebra. However,

$$T(a(z)) \cdot T(b(z)) \neq T(c(z)), \qquad c(z) = a(z)b(z).$$

The link between Toeplitz matrices and Laurent series is indeed an isomorphism for bi-infinite matrices.

Toeplitz matrices are not an algebra

The power series in $\mathcal{W}$ form an algebra. However,

$$T(a(z)) \cdot T(b(z)) \neq T(c(z)), \qquad c(z) = a(z)b(z).$$

The link between Toeplitz matrices and Laurent series is indeed an isomorphism for bi-infinite matrices.

Theorem (Gohberg–Feldman)
Let $a(z), b(z) \in \mathcal{W}$. Then,

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

$H(f)$ Hankel matrix. The correction $H(a^-)H(b^+)$ is a compact operator on ℓ^2.

Toeplitz matrices are not an algebra

The power series in $\mathcal{W}$ form an algebra. However,

$$T(a(z)) \cdot T(b(z)) \neq T(c(z)), \qquad c(z) = a(z)b(z).$$

The link between Toeplitz matrices and Laurent series is indeed an isomorphism for bi-infinite matrices.

Theorem (Gohberg–Feldman)
Let $a(z), b(z) \in \mathcal{W}$. Then,

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

$H(f)$ Hankel matrix. The correction $H(a^-)H(b^+)$ is a compact operator on ℓ^2.

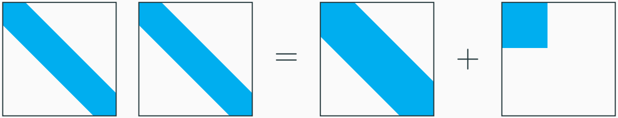
Toeplitz matrices are an algebra up to compact corrections.

The Hankel correction

Since $a(z), b(z)$ are in $\mathcal{W}$, then $a_j, b_j \to 0$ for $j \to \infty$, so:

- $T(a)$, $T(b)$ are numerically banded.
- $H(a^-), H(b^+)$ have the support in the top-left corner.

Therefore, $T(a) \cdot T(b) = T(c) - H(a^-)H(b^+)$ is numerically:



A new class of matrices

Definition
$A = T(a(z)) + E_a$ is quasi-Toeplitz (QT), if:

- $a(z) \in \mathcal{W}$.
- E_a is compact and satisfies $\|E_a\|_\infty < \infty$.

We define

$$\|T(a) + E_a\|_{QT} := \gamma \|a(z)\|_{\mathcal{W}} + \|E_a\|_\infty, \qquad \gamma := \frac{1+\sqrt{5}}{2}.$$

A new class of matrices

Definition
$A = T(a(z)) + E_a$ is quasi-Toeplitz (QT), if:

- $a(z) \in \mathcal{W}$.
- E_a is compact and satisfies $\|E_a\|_\infty < \infty$.

We define

$$\|T(a) + E_a\|_{QT} := \gamma \|a(z)\|_{\mathcal{W}} + \|E_a\|_\infty, \qquad \gamma := \frac{1+\sqrt{5}}{2}.$$

- QT-matrices form a (computationally-friendly) algebra.
- With this norm, QT is a Banach algebra.

Numerical representation of a QT matrix

We can represent QT matrices with a finite number of parameters at arbitrary accuracy.

If $A = T(a) + E_a$:

- $T(a)$ can be stored by a (truncated) approximation of its symbol, since $\sum_{|i|>j} |a_i| < \infty$;
- $E_a \approx UV^T$ by means of SVD; U, V have compact support and $\sigma_j(E_a) \to 0$, since E_a is compact and so "numerically low-rank".

We need a decay in the entries of E_a as we move away from $(1, 1)$, to truncate the vectors U, V to finite storage: is this always guaranteed?

QT is an algebra

We can check that:

- $A, B \in \mathcal{QT} \implies A + B \in \mathcal{QT}$:

$$T(a) + E_a + T(b) + E_b = T(a + b) + (E_a + E_b).$$

QT is an algebra

We can check that:

- $A, B \in \mathcal{QT} \implies A + B \in \mathcal{QT}$:

$$T(a) + E_a + T(b) + E_b = T(a+b) + (E_a + E_b).$$

- $A, B \in \mathcal{QT} \implies AB \in \mathcal{QT}$:

$$(T(a) + E_a) \cdot (T(b) + E_b) = T(ab) + (T(a)E_b + E_a T(b) + E_a E_b - H(a^-)H(b^+))$$

- The inverse can be computed similarly.

Solution of QBD problems

- QT is an algebra – and we have fast and accurate implementation of all the arithmetic operations.
- We can use the well-known iterations from the finite dimensional setting, and solve infinite-dimensional problems! Cyclic reduction, functional iterations, ...
- "Most" of the theory carries over.

Computational remarks

We can operate on quasi-Toeplitz matrices combining:

- Operations/Factorizations on power series (FFT-based);
- Toeplitz-vector multiplications (again FFT-based);
- Compression of low-rank matrices of the form $H(a^-)H(b^+)$.

For the last item, fast matvec is available, so we can run either Lanczos or random sampling to compress the products of Hankel matrices.

Computational remarks

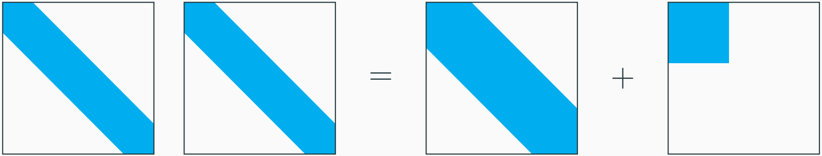
We can operate on quasi-Toeplitz matrices combining:

- Operations/Factorizations on power series (FFT-based);
- Toeplitz-vector multiplications (again FFT-based);
- Compression of low-rank matrices of the form $H(a^-)H(b^+)$.

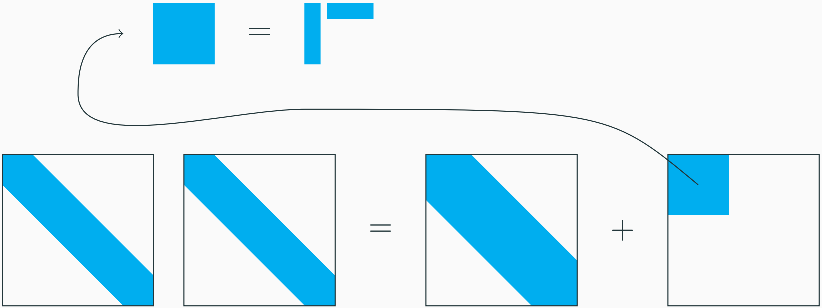
For the last item, fast matvec is available, so we can run either Lanczos or random sampling to compress the products of Hankel matrices.

Ranks of the corrections are small in practice: 10–20 is a typical value.

Keeping the rank low (pictorial version)

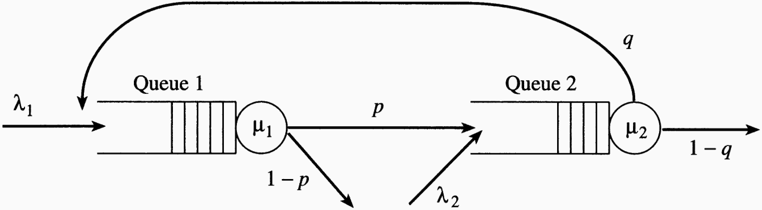


Keeping the rank low (pictorial version)



An example

Consider a tandem Jackson queue[1].



[1] Motyer, A.J. and Taylor, P.G., 2006. Decay rates for quasi-birth-and-death processes with countably many phases and tridiagonal block generators. Advances in applied probability, 38(2), pp.522-544.

Explicit solution

We can compute the steady state vector π using CR:

Problem	Time (s)	Residue $\|\cdot\|_\infty$	Band	Support	Rank
1	2.61	$2.02 \cdot 10^{-13}$	561	541	8
2	2.91	$9.09 \cdot 10^{-13}$	561	555	8
3	0.28	$2.02 \cdot 10^{-13}$	143	89	8
4	2.32	$1.77 \cdot 10^{-13}$	463	481	9
5	0.47	$1.93 \cdot 10^{-13}$	233	148	9
6	7.96	$1.16 \cdot 10^{-12}$	455	462	10
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

The residue is $\|\pi^T P - \pi^T\|_\infty$, and each problem corresponds to different rates and parameters. Ranks and support refer to the corrections in G, R, and the band to their Toeplitz part.

Handling restarts

We can consider an extension of the QBD setting. In 1D:

- It is possible to move from state i to $i-1$ and $i+1$
- At any moment, it is possible that a "reset event" fires, and we go back to state 1 with probability ρ.

Recall that the probability transition matrix has the form

$$P = \begin{bmatrix} 1-\rho_1 & \rho_1 & & & \\ \rho+\rho_{-1} & \rho_0 & \rho_1 & & \\ \rho & \rho_{-1} & \rho_0 & \rho_1 & \\ \vdots & & \ddots & \ddots & \ddots \end{bmatrix},$$

Simililarly, in the 2D setting, we can allow to go back to state 1 from any other phase in all the levels.

Some interesting facts

- Matrices of this kind have the form $A = A_0 + ev^T$, with A_0 QT, e the vector of all ones, and $v \in \ell^1$. In the previous case, $v = \rho e_1$.
- The arithmetic can be extended. For instance,

$$(A + ev_A^t)(B + ev_B^T) = AB + Aev_B^T + e(B + v_A^T B),$$

  and $Ae = ew^T + C$, with C compact and (numerically) finite support.
- The same algorithms can be used in this extended setting.

Conclusions and future outlook

- Computing with infinite matrices might be easy, if you have the right structure. We can compute several quantities without truncating to finite sections.
- Finite case handled as well.
- A MATLAB toolbox is available[2] for you to try: feedback is very welcome:

  `http://github.com/numpi/cqt-toolbox/`

[2] Bini, D. A., Massei, S., & Robol, L. *"Quasi-Toeplitz matrix arithmetic: a MATLAB toolbox"*. to appear in Num. Algo, 2019

Thank you for your attention!

Keeping the rank bounded

When we perform arithmetic operations, the rank of the correction increases in the representation. This can be kept low by recompression.

Keeping the rank bounded

When we perform arithmetic operations, the rank of the correction increases in the representation. This can be kept low by recompression.

Assume $E_a = U_a V_a^T$, U_a, V_a with k columns.

- $[Q_U, R_U] = \mathtt{qr}(U_a)$, and $[Q_V, R_V] = \mathtt{qr}(V_a)$.

Keeping the rank bounded

When we perform arithmetic operations, the rank of the correction increases in the representation. This can be kept low by recompression.

Assume $E_a = U_a V_a^T$, U_a, V_a with k columns.

- $[Q_U, R_U] = \texttt{qr}(U_a)$, and $[Q_V, R_V] = \texttt{qr}(V_a)$.
- $[U_1, \Sigma_1, V_1] \approx \texttt{svd}(R_U R_V^T)$ (truncated SVD).

Keeping the rank bounded

When we perform arithmetic operations, the rank of the correction increases in the representation. This can be kept low by recompression.

Assume $E_a = U_a V_a^T$, U_a, V_a with k columns.

- $[Q_U, R_U] = \texttt{qr}(U_a)$, and $[Q_V, R_V] = \texttt{qr}(V_a)$.
- $[U_1, \Sigma_1, V_1] \approx \texttt{svd}(R_U R_V^T)$ (truncated SVD).
- $E_a \approx Q_U U_1 \sqrt{\Sigma_1} \cdot (Q_V V_1 \sqrt{\Sigma_1})$.

The new representation has $k' < k$ columns.

Keeping the rank bounded

When we perform arithmetic operations, the rank of the correction increases in the representation. This can be kept low by recompression.

Assume $E_a = U_a V_a^T$, U_a, V_a with k columns.

- $[Q_U, R_U] = \mathtt{qr}(U_a)$, and $[Q_V, R_V] = \mathtt{qr}(V_a)$.
- $[U_1, \Sigma_1, V_1] \approx \mathtt{svd}(R_U R_V^T)$ (truncated SVD).
- $E_a \approx Q_U U_1 \sqrt{\Sigma_1} \cdot (Q_V V_1 \sqrt{\Sigma_1})$.

The new representation has $k' < k$ columns.

The cost of the compression is $\mathcal{O}(nk^2 + k^3)$ flops, where n is the support of E_a.