Model Checking Markov Chains using Krylov Subspace Methods

An Experience Report

Falko Dulat   Joost-Pieter Katoen   Viet Yen Nguyen

Software Modelling and Verification Group
RWTH Aachen University, Germany

European Performance Engineering Workshop
September 24th, 2010

Funded by ESA/ESTEC under Contract No. 21171/07/NL/JD

Continuous Time Markov Chains

A labeled CTMC is a tuple $(S, \mathbf{Q}, L)$ where

- S is a finite set of states,
- $\mathbf{Q} : S \times S \to \mathbb{R}$ is a generator matrix,
- $L : S \to 2^{AP}$ is a labeling fuction.

Probability of one-step transition from state s to s' in t time units is

$$1 - e^{-q_{s,s'}t}$$

Transient distribution, probability to be in s_i with $i < |S|$ at time t is

$$\pi(t) = (\ldots, \pi_i(t), \ldots) = e^{\mathbf{Q}t} \cdot \pi(0)$$

Continuous Time Markov Chains

A labeled CTMC is a tuple $(S, \mathbf{Q}, L)$ where

- S is a finite set of states,
- $\mathbf{Q} : S \times S \to \mathbb{R}$ is a generator matrix,
- $L : S \to 2^{AP}$ is a labeling fuction.

Probability of one-step transition from state s to s' in t time units is

$$1 - e^{-q_{s,s'} t}$$

Transient distribution, probability to be in s_i with $i < |S|$ at time t is

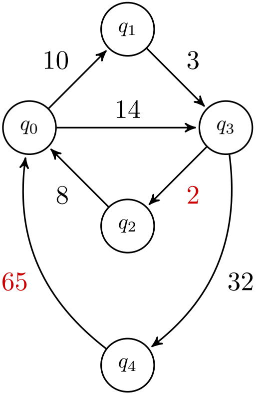
$$\pi(t) = (\ldots, \pi_i(t), \ldots) = e^{\mathbf{Q}t} \cdot \pi(0)$$

Applications

System biology	(e.g. probability to full reaction of molecules)
Networking	(e.g. probability of package loss)
Mission critical systems	(e.g. probability of system failure)

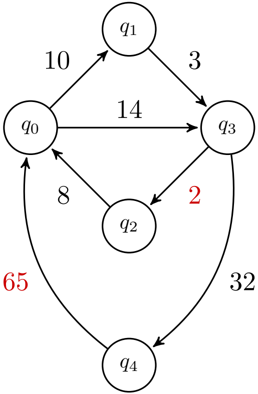
Stiffness

Non-Stiff

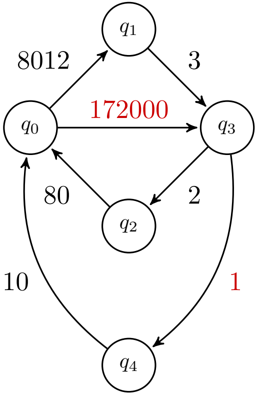


Stiffness

Non-Stiff



Stiff



Stiffness is the degree between the largest and biggest rates.

Techniques for Effective Transient Analysis

i.e. $\pi(t) = e^{\mathbf{Q}t}\pi(0)$

Uniformization

Use a uniformization rate $\Lambda \geq max_{i \in S}|q_{i,i}$ to rewrite $\mathbf{Q} = \Lambda \cdot (\mathbf{P} - I)$

$$\pi(t) = (\sum_{n=0}^{\infty} e^{-\Lambda t}\frac{(\Lambda t)^n}{n!}\mathbf{P}^n) \cdot \pi(0)$$

Sum is truncated given error ϵ. Truncation bound is in order of $O(\Lambda t)$.

Techniques for Effective Transient Analysis

i.e. $\pi(t) = e^{\mathbf{Q}t}\pi(0)$

Uniformization

Use a uniformization rate $\Lambda \geq max_{i \in S}|q_{i,i}$ to rewrite $\mathbf{Q} = \Lambda \cdot (\mathbf{P} - I)$

$$\pi(t) = (\sum_{n=0}^{\infty} e^{-\Lambda t}\frac{(\Lambda t)^n}{n!}\mathbf{P}^n) \cdot \pi(0)$$

Sum is truncated given error ϵ. Truncation bound is in order of $O(\Lambda t)$.

Krylov Subspaces

Given a $m < dim(\mathbf{Q})$ to rewrite by decomposition $\mathbf{Q}t = V_m H_m V_m^T$

$$\pi(t) = V_m e^{H_m} e_1 ||\pi(0)||_2$$

Time complexity is in order of m and $dim(\mathbf{Q})$.

Approximating the Matrix Exponential via Krylov Subspaces

$$\underbrace{Iv + Av + \frac{1}{2}A^2v + \frac{1}{6}A^3v + \ldots}_{e^A v} \in \underbrace{span\{v, Av, A^2v, A^3v, \ldots, A^{m-1}v\}}_{K_m(A,v)}$$

(Taylor-MacLaurin series expansion)

(m-order Krylov subspace)

Approximating the Matrix Exponential via Krylov Subspaces

$$\underbrace{Iv + Av + \frac{1}{2}A^2v + \frac{1}{6}A^3v + \dots}_{e^Av} \in \underbrace{span\{v, Av, A^2v, A^3v, \dots, A^{m-1}v\}}_{K_m(A,v)}$$

(Taylor-MacLaurin series expansion)

(m-order Krylov subspace)

Finding $e^{Av} \in K_m(A, v)$

1. Perform Hessenberg decomposition using Arnoldi iteration.
2. Compute the matrix exponential over the Hessenberg matrix.
3. Project the result from the Krylov subspace to “normal” space.

Approximating the Matrix Exponential via Krylov Subspaces

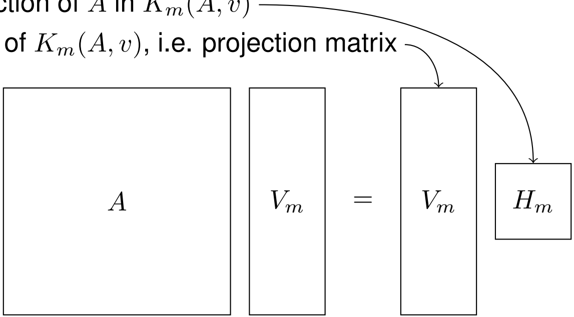
Algorithm 1 Arnoldi Iteration

$v_1 = v/||v||_2$
for $j = 1, 2, \ldots, m$ **do**
  $w = Av_j$
  for $i = 1, 2, \ldots, j$ **do**
    $h_{i,j} = (w, v_i)$
    $w = w - h_{i,j} v_i$
  end for
  $h_{j+1,j} = ||w||_2$
  $v_{j+1} = w/h_{j+1,j}$
end for

Performs Hessenberg decomposition, result is $AV_m = V_m H_m$

Approximating the Matrix Exponential via Krylov Subspaces

- Projection of A in $K_m(A, v)$
- Basis of $K_m(A, v)$, i.e. projection matrix



Large matrix A is transformed to the much smaller H_m

Approximating the Matrix Exponential via Krylov Subspaces

Theorem (Saad's Matrix Exponential)

Given the Hessenberg decomposition $AV_m = V_m H_m$ in m-order Krylov subspace $K_m(A, v)$ then

$$e^A v = V_m e^{H_m} e_1 ||v||_2$$

where e_1 is the first vector of I and $||v||_2$ is the Euclidean norm.

Approximating the Matrix Exponential via Krylov Subspaces

Theorem (Saad's Matrix Exponential)

Given the Hessenberg decomposition $AV_m = V_m H_m$ in m-order Krylov subspace $K_m(A, v)$ then

$$e^A v = V_m e^{H_m} e_1 ||v||_2$$

where e_1 is the first vector of I and $||v||_2$ is the Euclidean norm.

Proof:

$$e^A \approx e^{V_m H_m V_m^T} \quad \text{(by application of exponential)}$$

$$e^A \approx I + V_m H_m V_m^T + \frac{1}{2}(V_m H_m V_m^T)^2 + \ldots \quad \text{(by series expansion)}$$

$$e^A \approx V_m (I + H_m + \frac{1}{2} H_m^2 + \ldots) V_m^T \quad \text{(by } I = V_m^T V_m = V_m V_m^T\text{)}$$

$$e^A \approx V_m e^{H_m} V_m^T \quad \text{(by series de-expansion)}$$

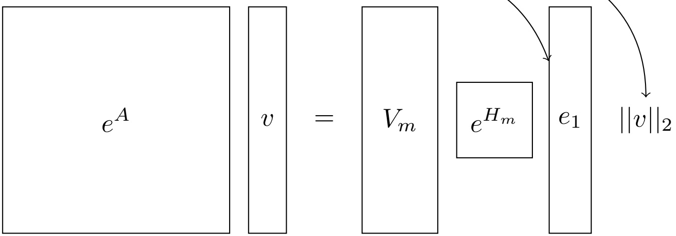
$$e^A V_m \approx V_m e^{H_m} \quad \text{(by multiplication with } V_m\text{)}$$

$$e^A v_1 \approx V_m e^{H_m} e_1 \quad \text{(by } v_1 = V_m e_1\text{)}$$

$$e^A v \approx V_m e^{H_m} e_1 ||v||_2 \quad \text{(by } v_1 = v/||v||_2\text{)}$$

Approximating the Matrix Exponential via Krylov Subspaces

- Euclidian norm of v
- First vector of I



Compute the matrix exponential in the Krylov subspace and project the result back to “normal” space

Quality of Approximation by Krylov Subspaces

Schwerdtfeger's formula uses *only* eigenpairs to express e^A

- Multiplicity of j^{th} eigenvalue
- Number of distinct eigenvalues

$$e^A = \sum_{j=1}^{t} A_j \sum_{k=0}^{s_j - 1} \frac{e^{\mu_j}}{k!} (A - \mu_j I)^k$$

- Frobenius covariant of j^{th} eigenvalue
- j^{th} distinct eigenvalue

Quality of Approximation by Krylov Subspaces

Schwerdtfeger's formula uses *only* eigenpairs to express e^A

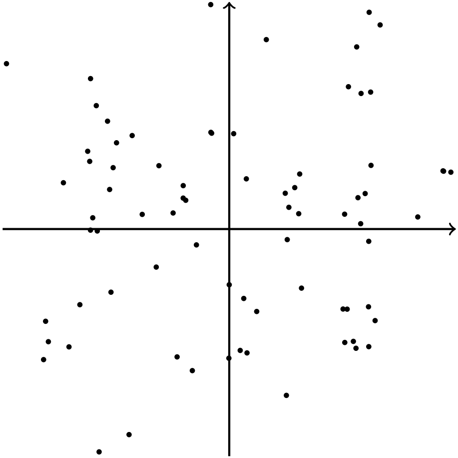
- Multiplicity of j^{th} eigenvalue
- Number of distinct eigenvalues

$$e^A = \sum_{j=1}^{t} A_j \sum_{k=0}^{s_j-1} \frac{e^{\mu_j}}{k!}(A - \mu_j I)^k$$

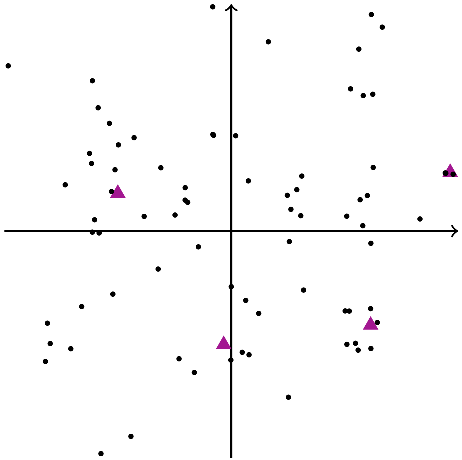
- Frobenius covariant of j^{th} eigenvalue
- j^{th} distinct eigenvalue

Converges to 0 for small eigenvalues

Eigenvalues of H_m converge to Extreme Eigenvalues of A

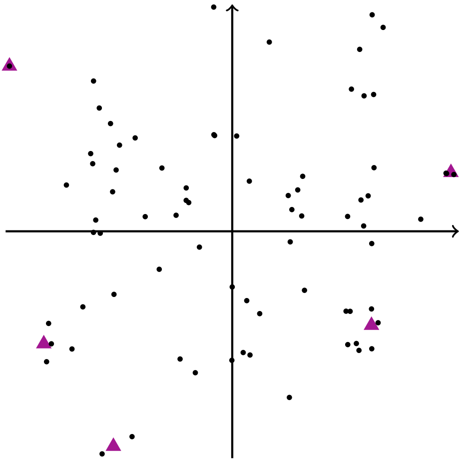


Eigenvalues of H_m converge to Extreme Eigenvalues of A



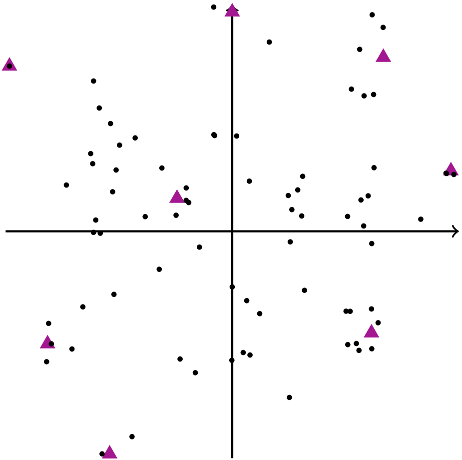
at $m = 5$

Eigenvalues of H_m converge to Extreme Eigenvalues of A



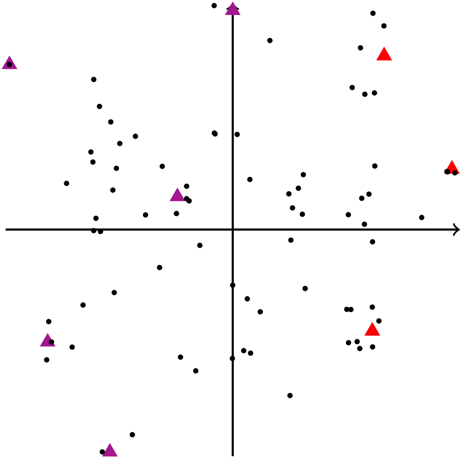
at $m = 8$

Eigenvalues of H_m converge to Extreme Eigenvalues of A



at $m = 10$

Eigenvalues of H_m converge to Extreme Eigenvalues of A



at $m = 10$

Only right extreme eigenvalues of A are useful for e^A

Error Margins

A Priori (Error Bounds)

- Saad [1992] shows given $\rho = ||A||_2$ the error is

$$||v||_2 \cdot \frac{\rho^m e^\rho}{m!}$$

- Hochbruck and Lubich [1997] shows given the numerical radius $\rho = max\{|z| : z \in \{x^* Ax : x^* x = 1\}\}$ and timebound τ the error is

$$12e^{-\rho\tau} \left(\frac{e\rho\tau}{m}\right)^m$$

Error Margins

A Priori (Error Bounds)

- Saad [1992] shows given $\rho = ||A||_2$ the error is

$$||v||_2 \cdot \frac{\rho^m e^\rho}{m!}$$

- Hochbruck and Lubich [1997] shows given the numerical radius $\rho = max\{|z| : z \in \{x^* Ax : x^* x = 1\}\}$ and timebound τ the error is

$$12e^{-\rho\tau} \left(\frac{e\rho\tau}{m}\right)^m$$

A Posteriori (Error Estimates)

- Saad [1992] shows that $h_{m+1,m}$ is a good estimator for the error

$$h_{m+1,m} \left| e_m^T e^{H_m} ||v||_2 e_1 v_{m+1} \right|$$

Experimental Setup

Configuration

Implementation in MRMC and running on 2.33 GHz computers with 16 GB RAM.

- Uniformization
- Uniformization + steady state detection
- Krylov

Model	Description	States	Transitions	Stiffness
CSPS	Cyclic server polling system	3072	14848	1600
TQN	Tandem queuing network	861	2859	400
PTP	Simple peer-to-peer protocol	1024	5121	0.5
ER	Enzymatic reaction	4011	11431	4000000
WGC	Wireless group communication	1329669	9624713	6164

Results: Non-Stiff Case Studies

Model	Time-bound	Terms UNI	Terms UNI-S	m KRY	Time [ms] UNI	Time [ms] UNI-S	Time [ms] KRY	Probability
CSPS	10	545	653	109	190	360	2677	0.6524983
	20	769	922	132	340	680	5633	0.8982785
	30	941	1129	147	490	980	9158	0.9708183
	40	1086	1303	155	640	1280	11249	0.9916387
	50	1214	1456	157	780	1580	11933	0.9976044
	60	1330	1595	162	940	1860	14133	0.9993137
	70	1436	1722	162	1070	2170	13992	0.9998034
	80	1535	1841	162	1230	2470	13811	0.9999437
	90	1627	1952	162	1380	2760	13651	0.9999839
	100	1715	2058	162	1530	3060	14038	0.9999954
PTP	1	163	192	20	10	10	18	0.3892596
	2	177	212	23	10	20	23	0.9055015
	3	184	220	25	10	20	28	0.987485
	4	190	228	25	10	20	28	0.9983193

...

Uniformization is very efficient for non-stiff models

Results: Stiff Case Studies

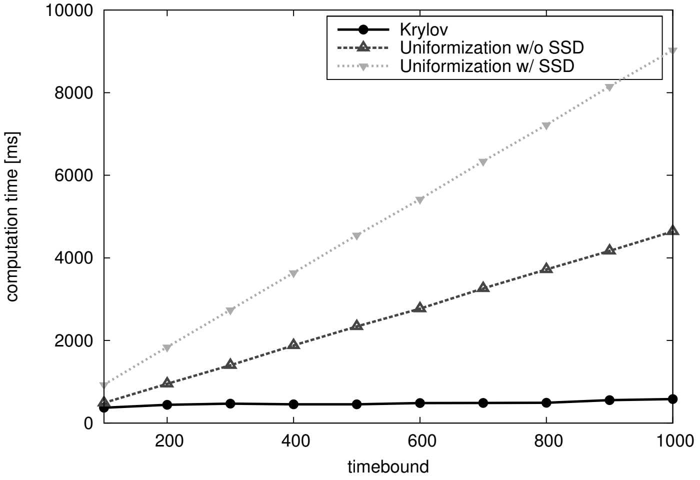


Figure: Verification times of $\mathcal{P}_{=?}(\overset{[0,t]}{\Diamond} Pr = 4)$ with increasing timebounds t on the ER model.

Results: Stiff Case Studies

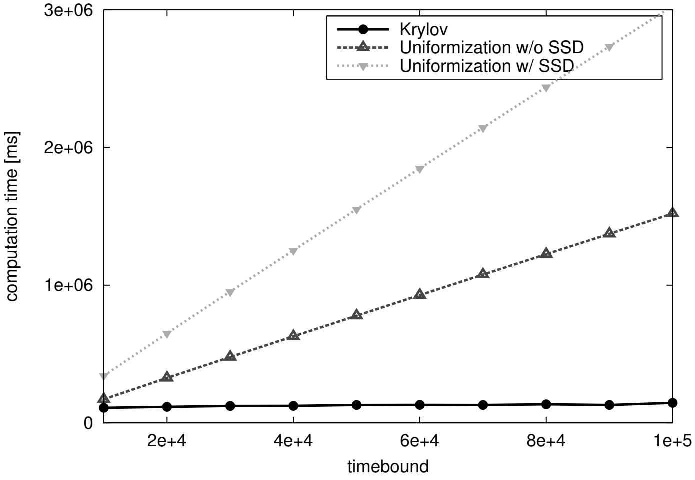
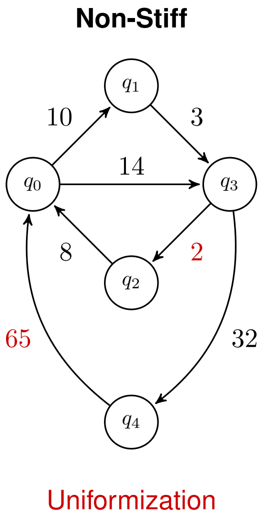
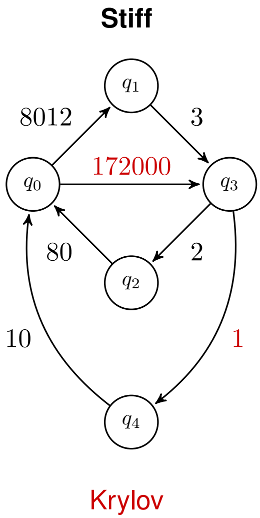


Figure: Verification times of $\mathcal{P}_{=?}(\overset{[0,t]}{\Diamond}$ *fail*) with increasing timebounds t on the WGC model.

Krylov-Based Method Performs Well on Stiff Markov Chains





Conclusions

Main Observation

Experimental results from several case studies reval that the Krylov-based transient is a magnitude faster than uniformization on stiff Markov models.

Future Work

- Study the clustering of eigenvalues w.r.t. benefits of the Krylov-based method.
- Connect Schwerdtfeger's formula to Arnoldi lemiscates.
- The usual: improve error bounds, reduce memory consumption, ...

Availability

Implementation is under GPL and is available upon request.