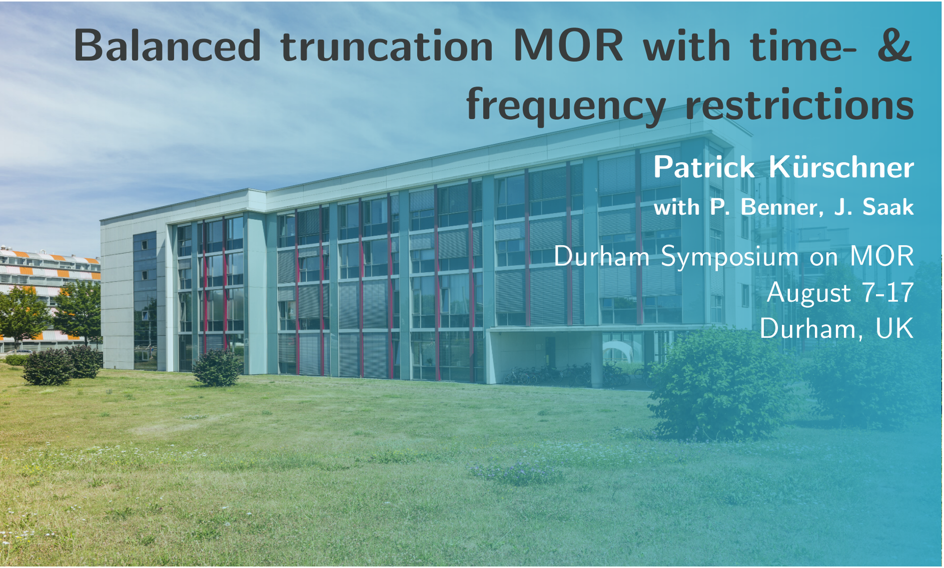
MAX PLANCK INSTITUTE
FOR DYNAMICS OF COMPLEX
TECHNICAL SYSTEMS
MAGDEBURG

CSC
COMPUTATIONAL METHODS IN
SYSTEMS AND CONTROL THEORY



Balanced truncation MOR with time- & frequency restrictions

Patrick Kürschner

with P. Benner, J. Saak

Durham Symposium on MOR
August 7-17
Durham, UK

Introduction

Linear, time-invariant systems

$$\Sigma: \begin{cases} \dot{x}(t) &= Ax(t) + Bu(t), \ x(0) = 0, \\ y(t) &= Cx(t) \end{cases} \quad \leftrightsquigarrow \quad \mathbf{G}(s) = C(sI - A)^{-1}B$$

with

- $A \in \mathbb{R}^{n \times n}$, large, stable $(\mathrm{Re}(\lambda(A)) < 0)$,
- $B \in \mathbb{R}^{n \times m}, \ C \in \mathbb{R}^{p \times n}, \ m, p \ll n.$

Recall Tatjana Stykel's lectures:
Balanced truncation (BT) MOR requires (factors of) the *infinite* reachability & observability Gramians $P_\infty, \ Q_\infty$, i.e. the solutions of the algebraic Lyapunov equations

$$AP_\infty + P_\infty A^T = -BB^T, \quad A^T Q_\infty + Q_\infty A = -C^T C.$$

Lyapunov equation for reachability Gramian $AP_\infty + P_\infty A^T = -BB^T$
Integral representations w.r.t. time / frequencies:

$$P_\infty = \int_0^\infty \mathrm{e}^{At} BB^T \mathrm{e}^{A^T t} \mathrm{d}t = \frac{1}{2\pi} \int_{-\infty}^{\infty} (\mathrm{i}\omega I - A)^{-1} BB^T (\mathrm{i}\omega I - A)^{-H} \mathrm{d}\omega$$

Here: *time- / frequency-limited* Gramians & BT [Gawronski/Juang '90]

$$P_{T_f} := \int_0^{T_f} \mathrm{e}^{At} BB^T \mathrm{e}^{A^T t} \mathrm{d}t, \quad T_f < \infty,$$

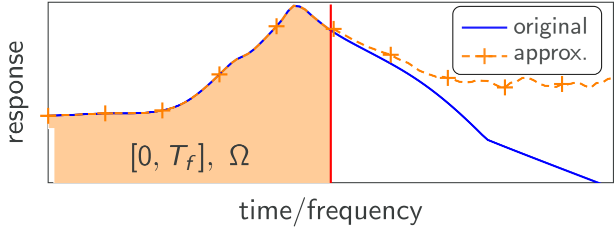
$$P_\Omega := \frac{1}{2\pi} \int_\Omega (\mathrm{i}\omega I - A)^{-1} BB^T (\mathrm{i}\omega I - A)^{-H} \mathrm{d}\omega, \ \Omega = -\Omega \subset \mathbb{R}$$

completely analogue for observability Gramians

Introduction

Why bother with time/frequency-limited BT?

- Compared to normal BT, we want **higher accuracies** ($\mathbf{G}(i\omega) \approx \mathbf{G_r}(i\omega)$, $y(t) \approx y_\mathbf{r}(t)$) in **given** Ω, $[0, T_f]$ with reduced system of the **order r**,
- or, **comparable accuracies** in Ω, $[0, T_f]$ with **smaller reduced system**,
- but allow arbitrary inaccuracies elsewhere:



- **Also of interest:** extra / reduced computational effort coming from restrictions.

Restricted BT

Lyapunov equations of restricted Gramians

[Gawronski/Juang '90, Petersson/Löfberg '14]

Time-limited reachability Gramian for $[0, T_f]$, $T_f < \infty$:

$$AP_{T_f} + P_{T_f}A^T = -BB^T + f(A)BB^Tf(A)^T, \quad f(A) = \mathrm{e}^{AT_f}.$$

Frequency-limited reachability Gramian

for $\Omega := [-\omega_2, \ -\omega_1] \cup [\omega_1, \ \omega_2], 0 \leq \omega_1 < \omega_2 < \infty$:

$$AP_\Omega + P_\Omega A^T = -f(A)BB^T - BB^Tf(A)^T,$$

$$f(A) = \mathrm{Re}\left(\tfrac{\mathrm{i}}{\pi} \log\left((A + \mathrm{i}\omega_1 I)^{-1}(A + \mathrm{i}\omega_2 I)\right)\right).$$

completely analog for observability Gramians $Q_\infty \ \rightsquigarrow \ Q_{T_f}, \ Q_\Omega$.

Restricted BT

Time-limited Balanced Truncation (TLBT)

1. Get $f(A)B, Cf(A),\ f(A) = \mathrm{e}^{AT_f}$.
2. Compute Cholesky factors of the Gramians $P_{T_f} = Z_{T_f} Z_{T_f}^T, \quad Q_{T_f} = Y_{T_f} Y_{T_f}^T$ s.t.
$$\mathbf{A P_{T_f} + P_{T_f} A^T + B B^T - f(A) B B^T f(A)^T = 0},$$
$$\mathbf{A^T Q_{T_f} + Q_{T_f} A + C^T C - f(A)^T C^T C f(A) = 0}.$$
3. SVD: $Y_{T_f}^T Z_{T_f} = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} S_1 & \\ & S_2 \end{bmatrix} \begin{bmatrix} R_1^T \\ R_2^T \end{bmatrix}$
4. Define $W := Y_{T_f} R_1 S_1^{-1/2}$, $V := Z_{T_f} U_1 S_1^{-1/2}$.
5. $\Sigma_{\mathbf{r}} \ : \ (A_{\mathbf{r}} := W^T A V, B_{\mathbf{r}} := W^T B, C_{\mathbf{r}} := CV)$.

Restricted BT

Time-limited Balanced Truncation (TLBT) with low-rank factors

1. Get $f(A)B, Cf(A),\ f(A) = \mathrm{e}^{AT_f}$.
2. Compute **low-rank** factors of the Gramians $P_{T_f} \approx Z_{T_f} Z_{T_f}^T, \quad Q_{T_f} \approx Y_{T_f} Y_{T_f}^T$ s.t.
$$\mathbf{A P_{T_f} + P_{T_f} A^T + B B^T - f(A) B B^T f(A)^T = 0},$$
$$\mathbf{A^T Q_{T_f} + Q_{T_f} A + C^T C - f(A)^T C^T C f(A) = 0}.$$
3. SVD: $Y_{T_f}^T Z_{T_f} = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} S_1 & \\ & S_2 \end{bmatrix} \begin{bmatrix} R_1^T \\ R_2^T \end{bmatrix}$
4. Define $W := Y_{T_f} R_1 S_1^{-1/2}$, $V := Z_{T_f} U_1 S_1^{-1/2}$.
5. $\Sigma_\mathbf{r} \ : \ (A_\mathbf{r} := W^T A V, B_\mathbf{r} := W^T B, C_\mathbf{r} := CV)$.

Numerical methods for computing low-rank factors:
Low-rank ADI, Rational Krylov subspace methods,

[E.G., Penzl '98, Li/White '02, Druskin/Simoncini '11]

Restricted BT

Frequency-limited Balanced Truncation (FLBT) with low-rank factors

1. Get $f(A)B, Cf(A),\ f(A) = \operatorname{Re}\left(\frac{\mathrm{i}}{\pi}\log\left((A + \mathrm{i}\omega_1 I)^{-1}(A + \mathrm{i}\omega_2 I)\right)\right)$.
2. Compute **low-rank** factors of the Gramians
   $P_\Omega \approx Z_\Omega Z_\Omega^T, \quad Q_\Omega \approx Y_\Omega Y_\Omega^T$ s.t.
   $$\mathbf{A P_\Omega + P_\Omega A^T + f(A) B B^T - B B^T f(A)^T = 0},$$
   $$\mathbf{A^T Q_\Omega + Q_\Omega A + f(A)^T C^T C - C^T C f(A) = 0}.$$
3. SVD: $Y_\Omega^T Z_\Omega = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} S_1 & \\ & S_2 \end{bmatrix} \begin{bmatrix} R_1^T \\ R_2^T \end{bmatrix}$
4. Define $W := Y_\Omega R_1 S_1^{-1/2}$, $V := Z_\Omega U_1 S_1^{-1/2}$.
5. $\Sigma_{\mathbf{r}}\ :\ (A_{\mathbf{r}} := W^T A V, B_{\mathbf{r}} := W^T B, C_{\mathbf{r}} := CV)$.

Numerical methods for computing low-rank factors:
Low-rank ADI, Rational Krylov subspace methods,

[E.g., Penzl '98, Li/White '02, Druskin/Simoncini '11]

Restricted Gramians

Consider reachability Gramians P, P_Ω, P_{T_f}.

For low-rank approximation, a rapid decay of the singular values is required. [E.g., Antoulas/Sorenson/Zhou '02, Grasedyck '04]

Lyapunov equations for infinite, time-/frequency-limited Gramians:

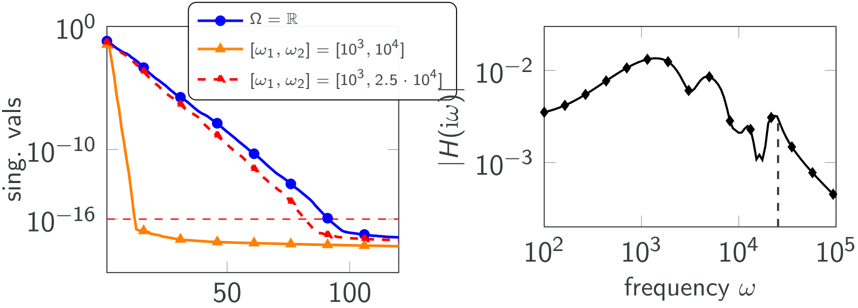
$$AP + PA^T = \begin{cases} -BB^T & \text{(BT)} \\ -BB^T + f(A)BB^T f(A)^T & \text{(TLBT)} \\ -f(A)BB^T - BBf(A)^T & \text{(FLBT)} \end{cases}$$

$\Rightarrow$ rank(rhs$_{\text{BT}}$)$= 2m = 2\cdot$ rank(rhs$_{\text{FLBT,TLBT}}$)$= m$

Expectation: slower singular value decay $\Rightarrow$ larger low-rank factors required for $P_{T,\Omega}$, $Q_{T,\Omega}$ compared P_∞, Q_∞.

That's not what happens!

$$AP_\Omega + P_\Omega A^T + f(A)BB^T + BB^T f(A)^T = 0,$$
$$f(A) = \mathrm{Re}\left(\tfrac{\mathrm{i}}{\pi}\log\left((A+\mathrm{i}\omega_1 I)^{-1}(A+\mathrm{i}\omega_2 I)\right)\right)$$

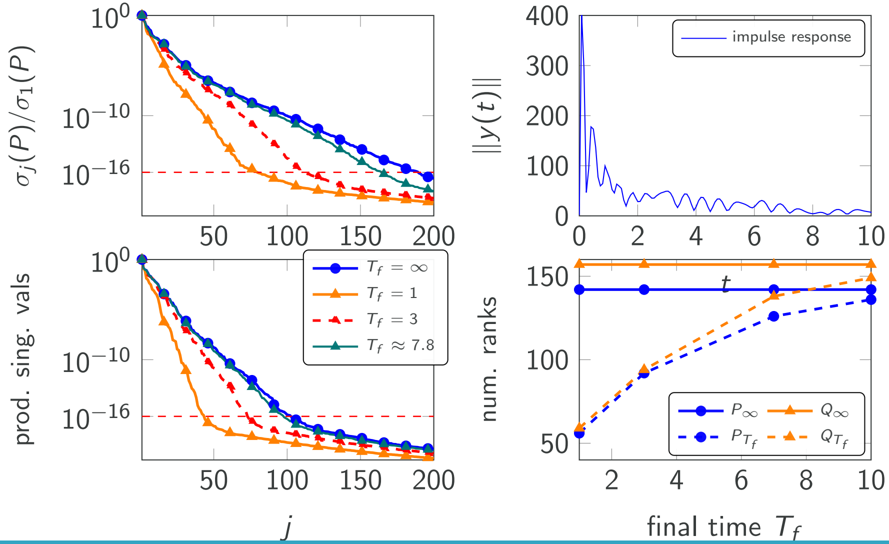


Singular values of frequency-limited Gramians [Benner/K./Saak '16]

If $|\omega_2 - \omega_1| < \rho(A)$, and ω_1, $\omega_2 \not\approx |\lambda(A)|$ with large $\left|\frac{\mathrm{Im}(\lambda(A))}{\mathrm{Re}(\lambda(A))}\right|$,
then $\sigma_i(P_\Omega) < \sigma_i(P_\infty)$.

Singular Value Decay of P_{T_f}

$$AP_{T_f} + P_{T_f}A^T + BB^T - f(A)BB^Tf(A)^T = 0,\; f(A) = e^{AT_f}$$



Singular Value Decay of P_{T_f}

Observation: The higher T_f, the slower the singular value decay of P_{T_f} (and of Q_{T_f}, $\sqrt{\lambda(P_{T_f}Q_{T_f})}$)
$\rightsquigarrow$ the smaller the differences to P_∞, Q_∞.
Rough explanation: $P_{T_f} = P_\infty - \Delta(T_f)$, $\quad \Delta(t) := \mathrm{e}^{At} P_\infty \mathrm{e}^{A^T t} \succeq 0$
the difference $\Delta(t)$ and the impulse-to-state map

$$x_\delta(t) = \mathrm{e}^{At} B\mathbf{v} \quad (u(t) = \mathbf{v}\delta(t)).$$

decay at a similar speed [K.' 17].

$\rightsquigarrow$ TLBT might be only worthwhile for

- small final times T_f, s.t., $x_\delta(t)$, $y_\delta(t)$ haven't reached stationary phase,
- and/or weakly damped systems.

Singular Value Decay

Under some conditions on Ω, T_f, the restricted Gramians often exhibit

- faster singular value decay
- $\rightsquigarrow$ smaller numerical rank than the infinite ones

$\Rightarrow$ $\exists$ low-rank approximations

with $\quad \mathrm{rank}(Z_\infty) = k \geq \ell = \mathrm{rank}(Z_\Omega), \quad (k, \ell \ll n)$

s.t. $\quad \|P_\infty - Z_\infty Z_\infty^T\| \approx \|P_\Omega - Z_\Omega Z_\Omega^T\|.$

The Matrix Function

Before starting with the low-rank Gramian factors, we need $B_f = f(A)B$!

Galerkin projection for $B_f := f(A) \times B$ onto range$(V_k) \subset \mathbb{R}^n$:

$$B_f \approx V_k f(\underbrace{V_k^T A V_k}_{=:H_k \in \mathbb{R}^{k\times k}}) V_k^T B, \quad V_k \in \mathbb{R}^{n\times k},\ V_k^T V_k = I_k,\ k \ll n.$$

Here: use of rational Krylov subspace range$(V_k) = \mathcal{RK}_k$,

$$\mathcal{RK}_k = \text{span}\left\{ B, (A-\xi_2 I)^{-1}B, \ldots, \prod_{i=1}^{k}(A-\xi_i I)^{-1}B \right\},$$

shifts/poles $\xi \in \mathbb{C}$ generated adaptively $\xi_{k+1} = \text{argmax}_{s\in\mathbb{S}_k} |r_k(s)|^{-1}$,

- TLBT $f = \mathrm{e}^{zT_f}$, $\mathbb{S}_k = \partial\text{conv}[\Lambda(H_k)]$ or $[\lambda_{\min}, \lambda_{\max}]$
  [Druskin/Lieberman/Zaslavsky '10, Druskin/Simoncini '11, Güttel '13]
- FLBT $f = \text{Re}\left(\frac{\mathrm{i}}{\pi}\log\frac{z+\mathrm{i}\omega_2}{z+\mathrm{i}\omega_1}\right)$, $\mathbb{S}_k = \mathrm{i}[\omega_1, \omega_2]$ [Benner/K./Saak '16]

Restricted Gramians

Big advantage of using $\mathcal{RK}$ approximation: we can reuse this subspace for computing the low-rank Gramian approximation!

Rational Krylov method for restricted Gramians [Benner/K./Saak '16]

1. Approximation $\tilde{B}_f \approx B_f := f(A)B$ by (block) rational Krylov method. $H_k = V_k^T(AV_k)$, $V_k^T V_k = I_k$ (range$(V_k) = \mathcal{RK}_k(A, B, \xi)$).
2. Galerkin for (time-lim.) Lyapunov equation: $H_k \hat{P}_{T_f} + \hat{P}_{T_f} H_k^T - (V_k^T \tilde{B}_f)(\tilde{B}_f^T V_k) + (V_k^T B)(B^T V_k) = 0.$ $\Rightarrow P_{T_f} \approx \tilde{P}_{T_f} := V_k \hat{P}_{T_f} V_k^T.$
3. Continue (rarely required) rational Krylov method until, e.g., $\|A\tilde{P}_{T_f} + \tilde{P}_{T_f} A^T - \tilde{B}_f \tilde{B}_f^T + BB^T\| < \mathrm{tol}.$

Restricted Gramians

Big advantage of using $\mathcal{RK}$ approximation: we can reuse this subspace for computing the low-rank Gramian approximation!

Rational Krylov method for restricted Gramians [Benner/K./Saak '16]

1. Approximation $\tilde{B}_f \approx B_f := f(A)B$ by (block) rational Krylov method. $H_k = V_k^T(AV_k)$, $V_k^T V_k = I_k$ $(\text{range}(V_k) = \mathcal{RK}_k(A, B, \xi))$.
2. Galerkin for (time-lim.) Lyapunov equation: $H_k \hat{P}_{T_f} + \hat{P}_{T_f} H_k^T - (V_k^T \tilde{B}_f)(\tilde{B}_f^T V_k) + (V_k^T B)(B^T V_k) = 0.$ $\Rightarrow P_{T_f} \approx \tilde{P}_{T_f} := V_k \hat{P}_{T_f} V_k^T.$
3. Continue (rarely required) rational Krylov method until, e.g., $\|A\tilde{P}_{T_f} + \tilde{P}_{T_f} A^T - \tilde{B}_f \tilde{B}_f^T + BB^T\| < \text{tol}.$

$\Rightarrow$ numerical effort of Gramian approximation and, hence, executing time / frequency-limited BT
(almost entirely) shifted to approximation of $f(A)B$!

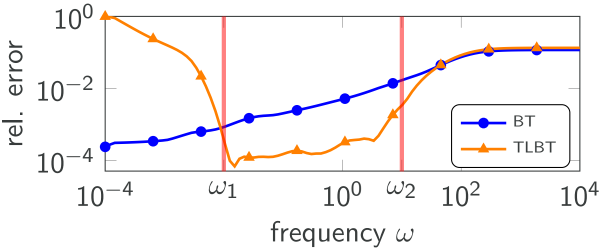
BT vs. freq.-limited BT

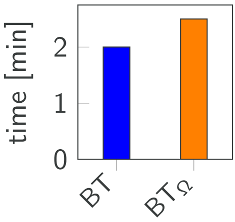
- computation of low-rank factors of Gramians P_∞, Q_∞, P_Ω, Q_Ω by rational Krylov subspace method.

  [Druskin/Simoncini '11, Benner/K./Saak '16]

- stopping criterion $\|\text{Lyap. residual}\| \leq 10^{-8}\|\text{rhs}\|$.
- accuracy assessment via relative transfer function error $\frac{\|\mathbf{G}(\mathrm{i}\omega)-\mathbf{G_r}(\mathrm{i}\omega)\|}{\|\mathbf{G}(\mathrm{i}\omega)\|}$

  Example `rail`[a]: $m = 7, p = 6$ reduction from $n = 79,841$ to order $\mathbf{r} = 50$



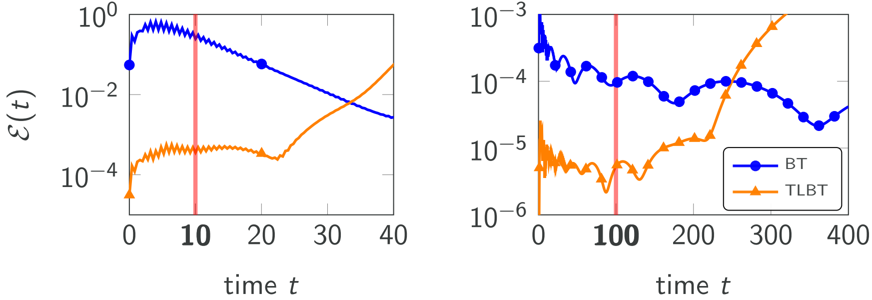


[a] Oberwolfach MOR Benchmark collection

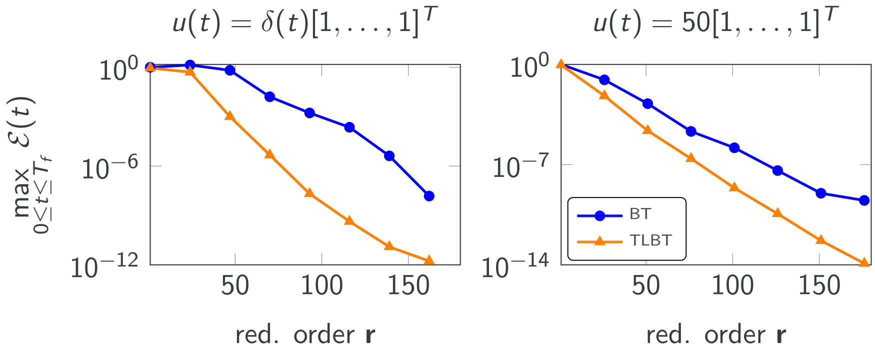
BT vs. time-limited BT

- **Example** `rail` with $n = 79,841$,
- point wise relative output error $\mathcal{E}(t) := \frac{\|y(t)-y_{\mathbf{r}}(t)\|}{\|y(t)\|}$,
- computing times: BT 65 sec., TLBT 130 sec.

$u(t) = \delta(t)[1,\ldots,1]^T, \mathbf{r} = 50$ $\qquad$ $u(t) = 50[1,\ldots,1]^T, \mathbf{r} = 50$

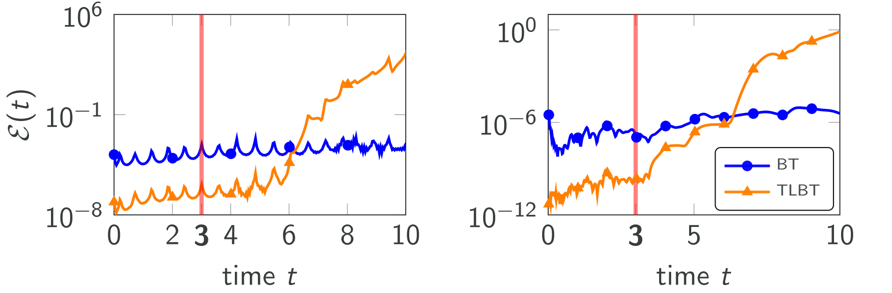


- **Example** `rail` with $n = 79,841$,
- point wise relative output error $\mathcal{E}(t) := \frac{\|y(t)-y_{\mathbf{r}}(t)\|}{\|y(t)\|}$,
- computing times: BT 65 sec., TLBT 130 sec.



- **Example** `bips`[a], index-1 descriptor system with $n = 21128$ and $\hat{n} = 3078$ dynamic states, $m = p = 4$,
- computing times: BT 12.5 sec., TLBT 28.5 sec.

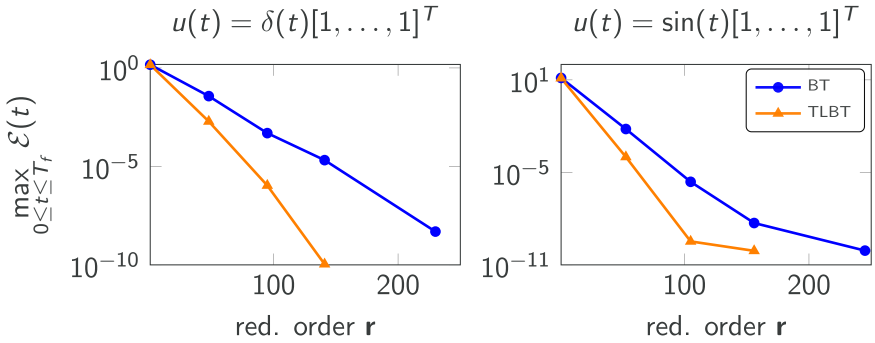
$u(t) = \delta(t)[1, \ldots, 1]^T$, $\mathbf{r} = 30$ $\quad$ $u(t) = \sin(t)[1, \ldots, 1]^T$, $\mathbf{r} = 100$



[a] Available from `https://sites.google.com/site/rommes/software`

BT vs. time-limited BT

- **Example** `bips`[a], index-1 descriptor system with $n = 21128$ and $\hat{n} = 3078$ dynamic states, $m = p = 4$,
- computing times: BT 12.5 sec., TLBT 28.5 sec.



[a] Available from `https://sites.google.com/site/rommes/software`

Originally, [GAWRONSKI/JUANG '90] also proposed TLBT for time-intervals $\mathcal{T} := [T_s,\ T_f]$ with $0 < T_s < T_f < \infty$.

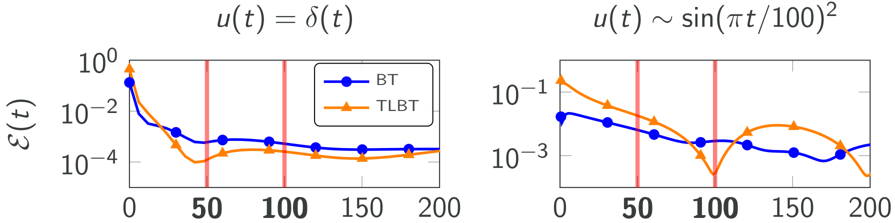
Gramian: $P_{\mathcal{T}} := \int_{T_s}^{T_f} \mathrm{e}^{At} BB^T \mathrm{e}^{A^T t} \mathrm{d}t$

Lyapunov equations:

$AP_{\mathcal{T}} + P_{\mathcal{T}}A^T + \mathrm{e}^{AT_s}\, BB^T\, \mathrm{e}^{AT_s} - \mathrm{e}^{AT_f}\, BB^T\, \mathrm{e}^{A^T T_f} = 0.$

$\rightsquigarrow$ no problem for low-rank factor computation!

However, TLBT in $[T_s,\ T_f]$ delivers **only** accurate reduced systems for the impulse input!

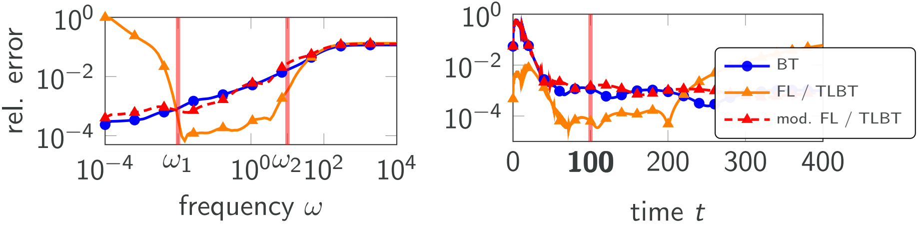


Bad news:

Time- / frequency limited BT are in general not stability preserving!

Modified TLBT / FLBT by [Gugercin/Antoulas '04] are, but

- sacrifices high accuracy of TLBT, FLBT in the relevant regions,
- numerically more expensive.



⇝ **our recommendation:** if stability preservation is crucial, stick to unrestricted BT!

Time-limited Gramians $P_{T_f},\ Q_{T_f}$ still exist for unstable systems, provided $\Lambda(A) \cap \Lambda(-A) = \emptyset$.

Moreover, they still solve the time-limited Lyapunov equations. [K '17]

Idea:

Use (low-rank factors of) time-limited Gramians as basis for BT of unstable systems. [Antoulas '05]

Efficient computation of low-rank factors by the same algorithms.

Might be only practicable for trajectories $x(t),\ y(t)$ that do not drastically blow up in $[0, T_f]$.

Summary

- Faster singular value decay of frequency- / time-limited Gramians.
- Rational Krylov subspace methods enable the approximation of $f(A)B$ and Gramians in a single algorithm (subspace recycling).
- Frequency- / time-limited BT now applicable for large-scale systems (almost) with the same ease as standard BT.
  - higher accuracy in Ω / $[0, T_f]$ compared to standard BT,
  - comparable numerical effort.

Further research:

- Alternative (better?) methods fo e^{At},
- Discrete-time systems. (✓) [Gawronski/Juang '90, Petersson '13]
- Descriptor systems.

Thank you for your attention!

Some literature

P. Benner, P. Kürschner, J. Saak, Frequency-Limited Balanced Truncation with Low-Rank Approximations, SIAM J. Sci. Comput. 38(1), pp. A471–A499, 2016.

P. Kürschner, Balanced truncation model order reduction in limited time intervals for large systems, arXiv:1707.02839 , 2017.

W. Gawronski and J. Juang, *Model reduction in limited time and frequency intervals*, Int. J. Syst. Sci. 21(2), pp. 349–376, 1990.

D. Petersson, *A Nonlinear Optimization Approach to H2-Optimal Modeling and Control*, PhD thesis, Linköping University, 2013.