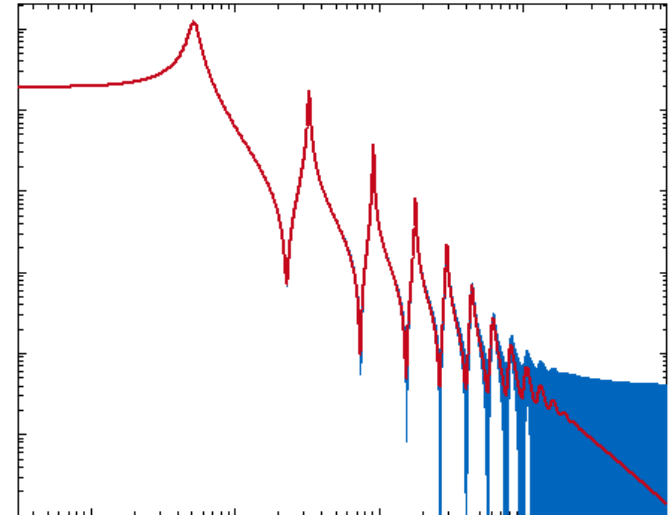


Adaptive Entwicklungspunktwahl und globale Fehlerschranken bei der Modellreduktion mittels Krylow-Unterraum-Verfahren

Heiko Panzer, Thomas Wolf, Boris Lohmann

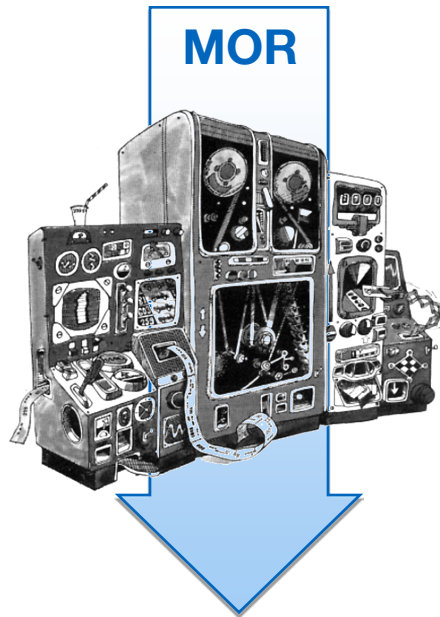


**48. Regelungstechnisches Kolloquium
Boppard
21. Februar 2014**

$$\mathbf{E} \dot{\mathbf{x}}(t) = \mathbf{A} \mathbf{x}(t) + \mathbf{B} \mathbf{u}(t)$$

$$\mathbf{y}(t) = \mathbf{C} \mathbf{x}(t)$$

$$\dim \mathbf{x} = N, \det \mathbf{E} \neq 0$$



Gesucht: Reduziertes Modell $G_r(s)$

- Getreue Nachbildung des Übertragungsverhaltens (mit Informationen über die Approximationsgüte)
- Erhaltung von Systemeigenschaften, insbesondere der Stabilität
- Numerisch effizientes Vorgehen

$$\mathbf{E}_r \dot{\mathbf{x}}_r(t) = \mathbf{A}_r \mathbf{x}_r(t) + \mathbf{B}_r \mathbf{u}(t),$$

$$\mathbf{y}(t) \approx \mathbf{y}_r(t) = \mathbf{C}_r \mathbf{x}_r(t)$$

$$\dim \mathbf{x}_r = n \ll N$$

Zur Bewertung der Reduktionsgüte betrachte das Fehlersystem (Ordnung: $N+n$)

$$\mathbf{G}_e(s) := \mathbf{G}(s) - \mathbf{G}_r(s)$$

Wir interessieren uns für die Systemnormen

$$\|\mathbf{G}\|_{\mathcal{H}_\infty} := \sup_{\omega \in \mathbb{R}} \|\mathbf{G}(i\omega)\|_2$$

$$\|\mathbf{G}\|_{\mathcal{H}_2} := \sqrt{\frac{1}{2\pi} \int_{-\infty}^{\infty} \text{tr} [\mathbf{G}^H(i\omega) \mathbf{G}(i\omega)] d\omega} = \sqrt{\text{tr} [\mathbf{B}^T \mathbf{Q} \mathbf{B}]} = \sqrt{\text{tr} [\mathbf{C} \mathbf{P} \mathbf{C}^T]}$$

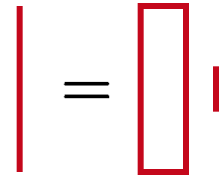
$$\mathbf{A}^T \mathbf{Q} \mathbf{E} + \mathbf{E}^T \mathbf{Q} \mathbf{A} + \mathbf{C}^T \mathbf{C} = \mathbf{0} \quad (\text{Beobachtbarkeits-Gramsche})$$

$$\mathbf{A} \mathbf{P} \mathbf{E}^T + \mathbf{E} \mathbf{P} \mathbf{A}^T + \mathbf{B} \mathbf{B}^T = \mathbf{0} \quad (\text{Steuerbarkeits-Gramsche})$$

$$\begin{aligned}\mathbf{E} \dot{\mathbf{x}}(t) &= \mathbf{A} \mathbf{x}(t) + \mathbf{B} \mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{C} \mathbf{x}(t)\end{aligned}$$

$\dim \mathbf{x} = N$

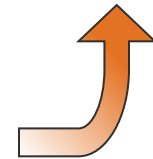
Galerkin-Approximation: $\mathbf{x}(t) \approx \hat{\mathbf{x}}(t) = \mathbf{V} \mathbf{x}_r(t)$



$$\mathbf{E} \mathbf{V} \dot{\mathbf{x}}_r(t) = \mathbf{A} \mathbf{V} \mathbf{x}_r(t) + \mathbf{B} \mathbf{u}(t) + \boldsymbol{\epsilon}(t)$$

$\dim \mathbf{x}_r = n \ll N$

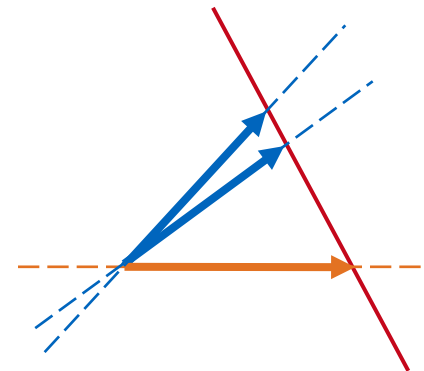
Aber:

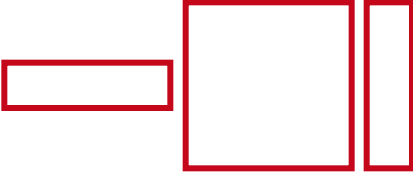
$$\begin{aligned}\mathbf{E} \mathbf{V} \dot{\mathbf{x}}_r(t) &\in \text{span}(\mathbf{E} \mathbf{V}) \\ \mathbf{A} \mathbf{V} \mathbf{x}_r(t) + \mathbf{B} \mathbf{u}(t) &\notin \text{span}(\mathbf{E} \mathbf{V})\end{aligned}$$


Idee: Gleichung auf $\text{span}(\mathbf{E} \mathbf{V})$ projizieren! (Petrov-Galerkin)

$$\text{Projektormatrix } \mathcal{P} = \mathbf{E} \mathbf{V} (\mathbf{W}^T \mathbf{E} \mathbf{V})^{-1} \mathbf{W}^T$$

$$\mathcal{P} \mathbf{E} \mathbf{V} \dot{\mathbf{x}}_r(t) = \mathcal{P} \mathbf{A} \mathbf{V} \mathbf{x}_r(t) + \mathcal{P} \mathbf{B} \mathbf{u}(t) + \mathcal{P} \boldsymbol{\epsilon}(t)$$





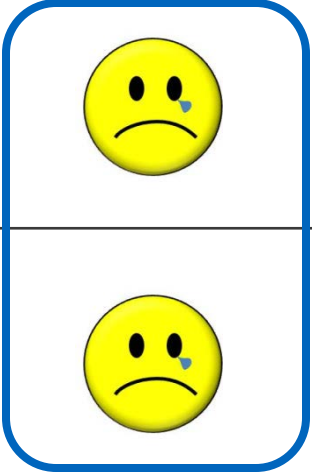






$$\begin{aligned}
 \mathbf{E} \mathbf{V} (\mathbf{W}^T \mathbf{E} \mathbf{V})^{-1} \mathbf{W}^T \mathbf{E} \mathbf{V} \dot{\mathbf{x}}_r(t) &= \mathbf{E} \mathbf{V} (\mathbf{W}^T \mathbf{E} \mathbf{V})^{-1} \mathbf{W}^T \mathbf{A} \mathbf{V} \mathbf{x}_r(t) \\
 &+ \mathbf{E} \mathbf{V} (\mathbf{W}^T \mathbf{E} \mathbf{V})^{-1} \mathbf{W}^T \mathbf{B} \mathbf{u}(t) \\
 &+ \mathbf{E} \mathbf{V} (\mathbf{W}^T \mathbf{E} \mathbf{V})^{-1} \mathbf{W}^T \boldsymbol{\epsilon}(t)
 \end{aligned}$$


Reduziertes Modell, Reduced Order Model (ROM)

$$\Leftrightarrow \mathbf{G}_r(s) : \begin{cases} \overbrace{\mathbf{W}^T \mathbf{E} \mathbf{V}}^{\mathbf{E}_r} \dot{\mathbf{x}}_r(t) = \overbrace{\mathbf{W}^T \mathbf{A} \mathbf{V}}^{\mathbf{A}_r} \mathbf{x}_r(t) + \overbrace{\mathbf{W}^T \mathbf{B}}^{\mathbf{B}_r} \mathbf{u}(t) \\ \mathbf{y}_r(t) = \underbrace{\mathbf{C} \mathbf{V}}_{\mathbf{C}_r} \mathbf{x}_r(t) \end{cases}$$

Aber wie wählen wir \mathbf{V} und \mathbf{W} ?!

Verfahren zur MOR von LZI Systemen

| | Modales Abschneiden | („exaktes“) Balanciertes Abschneiden | Krylow-Unterraum-Methoden |
|-------------------------------------|---|---|---|
| Güte und Fehlerschranken |  |  |  |
| Stabilitätserhaltung |  |  |  |
| Numerischer Aufwand und Machbarkeit |  |  |  |

Def.: Eingangs- und Ausgangs-Krylow-Unterraum zu Entwicklungspunkten σ_i :

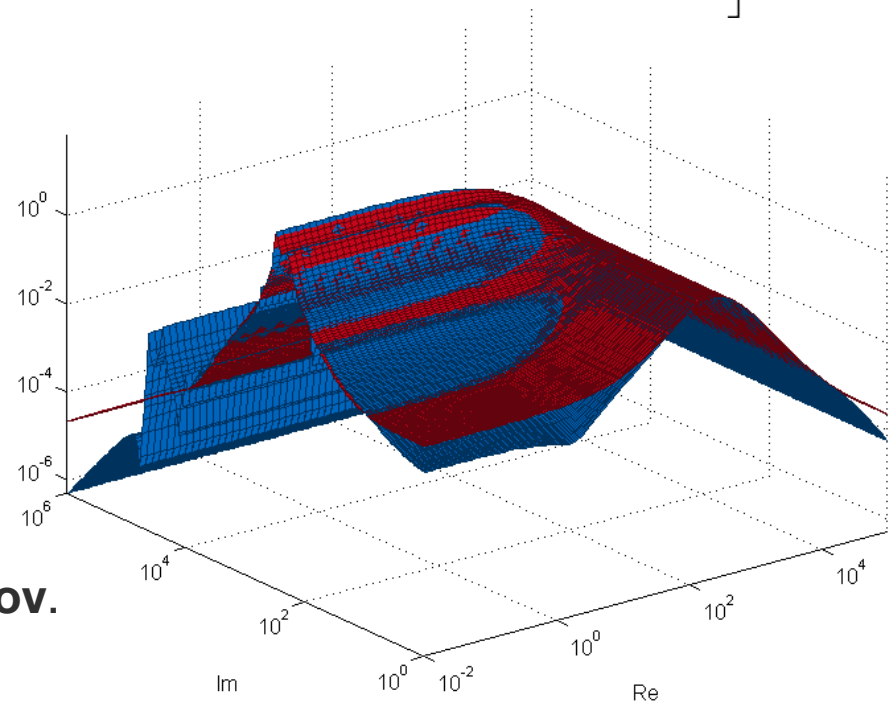
$$\mathbf{V} = \left[(\mathbf{A} - \sigma_1 \mathbf{E})^{-1} \mathbf{b}, \quad (\mathbf{A} - \sigma_2 \mathbf{E})^{-1} \mathbf{b}, \quad \dots, \quad (\mathbf{A} - \sigma_n \mathbf{E})^{-1} \mathbf{b} \right].$$

$$\mathbf{W} = \left[(\mathbf{A} - \sigma_1 \mathbf{E})^{-T} \mathbf{c}^T, \quad (\mathbf{A} - \sigma_2 \mathbf{E})^{-T} \mathbf{c}^T, \quad \dots, \quad (\mathbf{A} - \sigma_n \mathbf{E})^{-T} \mathbf{c}^T \right].$$

Wir können insgesamt $2n$ Shifts wählen.
Dies bewirkt sog. **Moment Matching**:

$$G_r(s) = G(s) \quad \forall s \in \{\sigma_1, \dots, \sigma_{2n}\}$$

Das Verfahren heißt daher auch
Rational Interpolation oder **Rational Krylov**.



Aber welche (und wie viele) Shifts führen auf ein „gutes“ reduziertes Modell?!
Wie garantiert man Erhaltung der Stabilität? Wie groß ist der resultierende Fehler?

Vereinfachte MATLAB Implementierung von IRKA [Gugercin et al. 2008]

```
function [V, W, s0] = IRKA(A,B,C,E,n)

% compute initial shifts
s0 = -eigs(A,E,n,0);

while (1)
    % compute Krylov subspaces
    [V, W] = RationalKrylov(A,B,C,E,s0);

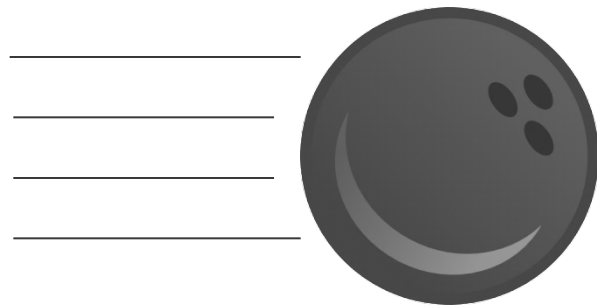
    % compute new shifts from eigendecomposition of ROM
    s0_new = -eig(W.'*A*V, W.'*E*V);

    % test for convergence
    if norm((s0-s0_new)./s0) < 1e-4, break, end

    % mirror shifts with negative real part
    s0 = s0_new.*sign(real(s0_new));
end
```

Liefert bei Konvergenz ein H_2 -optimales ROM.

Aber: Wahl der reduzierten Ordnung unklar.
Konvergenz nicht gesichert. (Stabilitätsverlust möglich.)
Resultierender Fehler unbekannt.



Fehler-Faktorisierung
Krylow-Unterraum-
Methoden



Kumulative
Reduktion
und Wahl
der red.
Ordnung



Trust Region
basierte
 H_2 -optimale
MOR



Globale
 H_2 - und H_∞ -
Fehler-
schranken

Wichtige Beobachtung: Eingangs-Krylow-Räume erfüllen die Sylvester-Gleichung

$$\mathbf{AV} - \mathbf{EVS}_V - \mathbf{B}\tilde{\mathbf{C}}_r = \mathbf{0}$$

$$\Rightarrow (\mathbf{I} - \mathcal{P})\mathbf{AV} - (\mathbf{I} - \mathcal{P})\mathbf{EVS}_V - (\mathbf{I} - \mathcal{P})\mathbf{B}\tilde{\mathbf{C}}_r = \mathbf{0}$$

$$\Rightarrow \mathbf{AV} - \mathbf{EVE}_r^{-1}\mathbf{A}_r - \mathbf{B}_\perp\tilde{\mathbf{C}}_r = \mathbf{0}$$

Wolf/P. 2011

Ist die Projektionsmatrix \mathbf{V} die Basis eines Eingangs-Krylow-Unterraums, so erfüllt das resultierende Fehlersystem:

$$\mathbf{G}_e(s) = \mathbf{G}(s) - \mathbf{G}_r(s) = \underbrace{\left[\begin{array}{c|c} \mathbf{E}, \mathbf{A} & \mathbf{B}_\perp \\ \hline \mathbf{C} & \mathbf{0} \end{array} \right]}_{\mathbf{G}_\perp(s)} \cdot \underbrace{\left[\begin{array}{c|c} \mathbf{E}_r, \mathbf{A}_r & \mathbf{B}_r \\ \hline \tilde{\mathbf{C}}_r & \mathbf{I}_n \end{array} \right]}_{\tilde{\mathbf{G}}_r^R(s)}$$

Systemmatrizen
von $\mathbf{G}(s)$

leicht zu berechnen

$$\mathbf{B}_\perp = \mathbf{B} - \mathbf{EVE}_r^{-1}\mathbf{B}_r$$

- geringe Ordnung
- Entwicklungspunkte bestimmen Nullstellen
- Allpass?!

$$G(s) = G_r(s) + G_e(s)$$

V Basis eines Krylow-Unterraums

$$= G_r(s) + G_{\perp}(s) \cdot \tilde{G}_r(s)$$

2. Reduktion

$$= G_r(s) + \left[G_{r,2}(s) + G_{\perp,2}(s) \cdot \tilde{G}_{r,2}(s) \right] \cdot \tilde{G}_r(s)$$

$$= q + \left[q_2 + n \quad q_2 \right] q$$

$$= q + q_2 q + n q_2 q$$

3. Reduktion

$$= q + q_2 q + \left[q_3 + n \quad q_3 \right] q_2 q$$

$$= q + q_2 q + q_3 q_2 q + n q_3 q_2 q$$

$$= \underbrace{q + q_2 q + q_3 q_2 q}_{\sum q_i} + n \underbrace{q_3 q_2 q}_{\sum q_i}$$

Also: Wir müssen nicht in einem Schritt reduzieren, sondern können mehrmals auf sehr kleine Ordnung (z.B. $n=2$) reduzieren. Innerhalb der „Salamischeiben“ können wir z.B. IRKA nutzen.

```

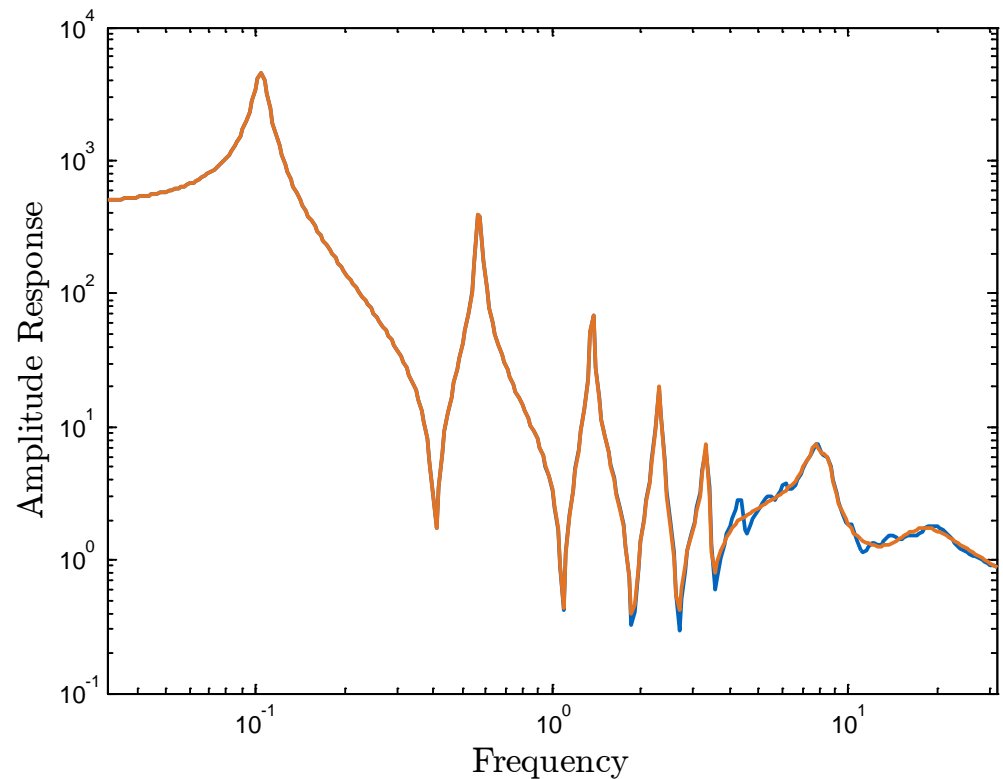
Bbot = B;

for i=1:15
    [V,W,Crt] = IRKA(A,Bbot,C,E,2);

    Ar = W'*A*V;
    Er = W'*E*V;
    Br = W'*Bbot;
    Cr = C*V;
    Bbot = Bbot - E*V*(Er\Br);

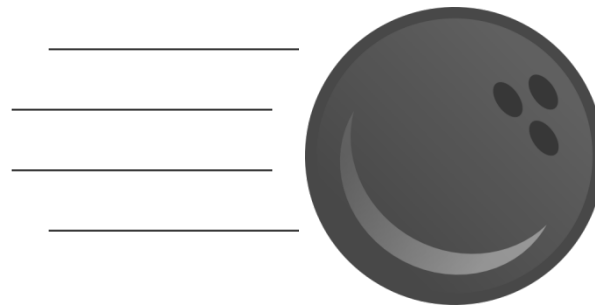
    Ar_ = [Ar_,      Br_*[0 0]; ...
           Br_*Crt_, Ar];
    Er_ = blkdiag(Er_, Er);
    Br_ = [Br_; Br];
    Cr_ = [Cr_, Cr];
    Crt_ = [Crt_, Crt];
end

sysr = dss(Ar_,Br_,Cr_, 0,Er_);
sysrt = dss(Ar_,Br_,Crt_,1,Er_);
    
```



Die Wahl der reduzierten Systemordnung ist damit adaptiv möglich!

Weiterhin offen: Konvergenzprobleme in IRKA. Resultierender Fehler?



Fehler-Faktorisierung
Krylow-Unterraum-
Methoden

Kumulative
Reduktion
und Wahl
der red.
Ordnung

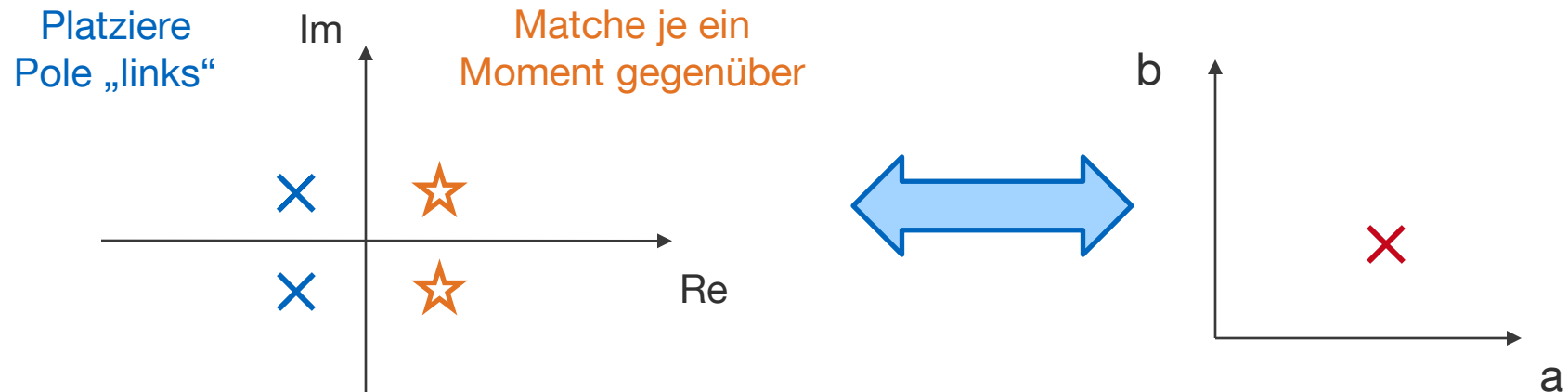


Trust Region
basierte
 H_2 -optimale
MOR



Globale
 H_2 - und H_∞ -
Fehler-
schranken

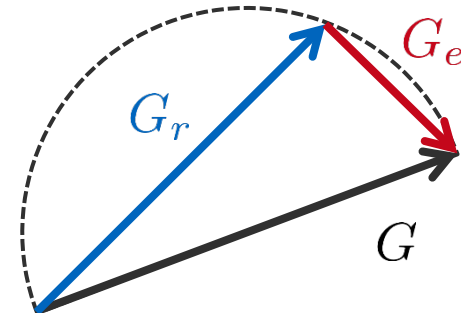
Idee: Grenze Suche nach H₂-Optimum auf H₂-Pseudo-Optima ein!



Dies ermöglicht die Verwendung des sehr einfachen Gütefunktional

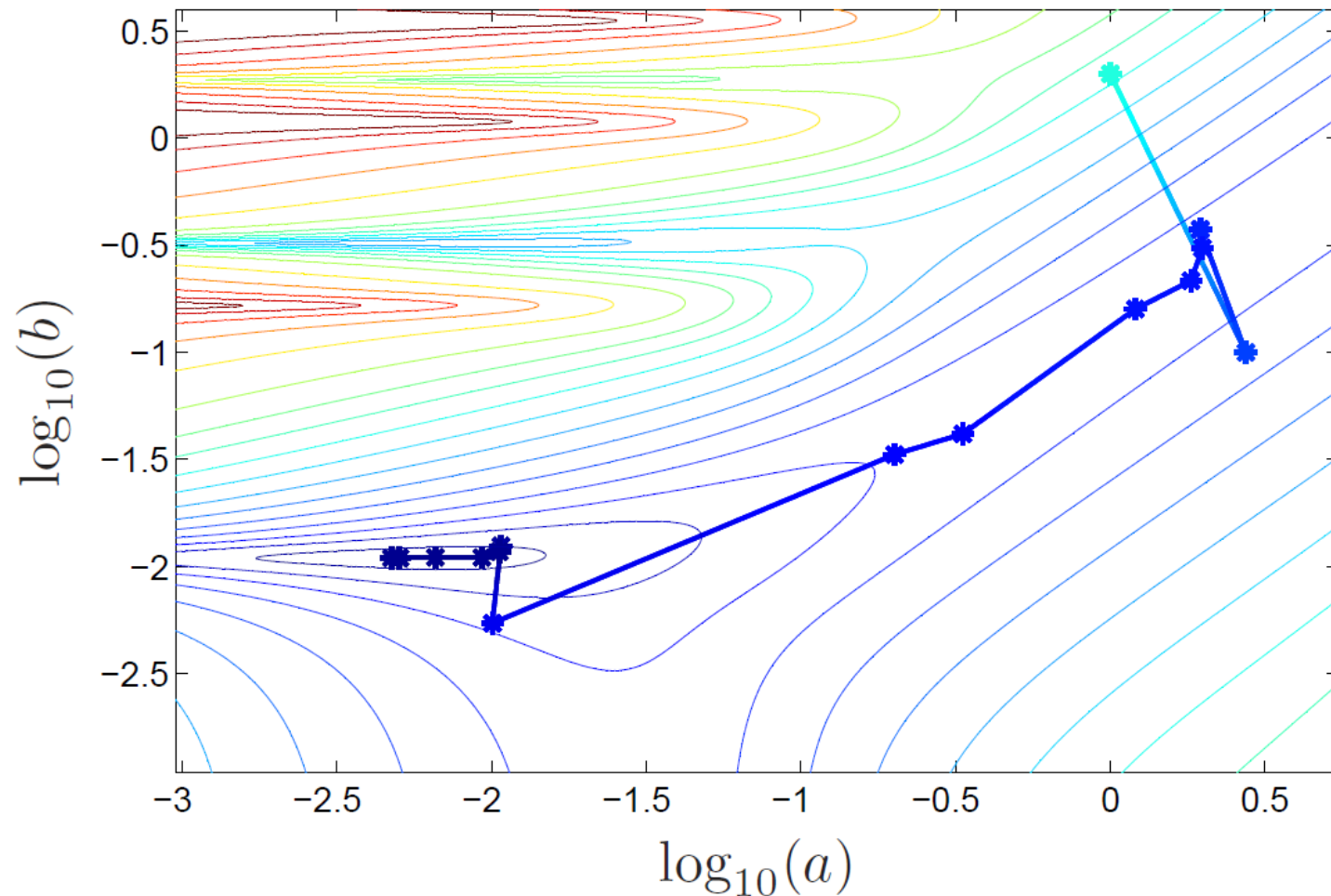
$$\mathcal{J} := - \|G_r\|_{\mathcal{H}_2}^2$$

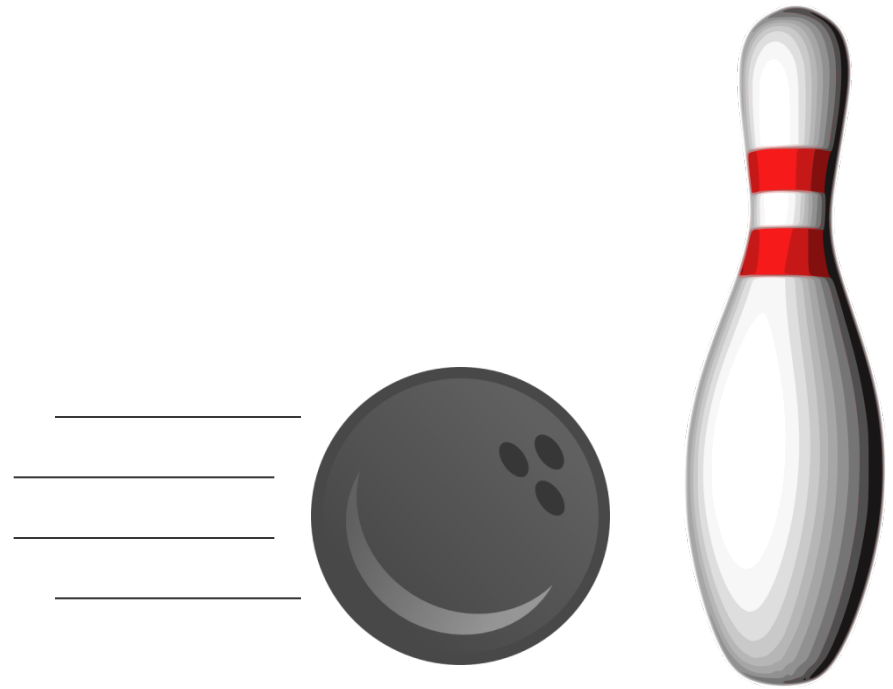
und damit analytische Ausdrücke für J, den Gradienten und die Hessematrix.



Die Parametrierung erfolgt über zwei positive reelle Zahlen a, b (Koeffizienten des charakteristischen Polynoms).

Benchmark-Modell eines Kragbalkens (SLICOT)
 Startwerte für die Entwicklungspunkte: $1 \pm j$
 Trust Region Algorithmus (fmincon in MATLAB)





Fehler-Faktorisierung
Krylow-Unterraum-
Methoden

Kumulative
Reduktion
und Wahl
der red.
Ordnung

Trust Region
basierte
 H_2 -optimale
MOR

Globale
 H_2 - und H_∞ -
Fehler-
schranken

Ist V die Basis eines Eingangs-Krylow-Unterraums, gilt die Zerlegung

$$\mathbf{G}_e(s) = \underbrace{\begin{bmatrix} \mathbf{E}, \mathbf{A} & | & \mathbf{B}_\perp \\ \hline \mathbf{C} & | & \mathbf{0} \end{bmatrix}}_{\mathbf{G}_\perp(s)} \cdot \underbrace{\begin{bmatrix} \mathbf{E}_r, \mathbf{A}_r & | & \mathbf{B}_r \\ \hline \tilde{\mathbf{C}}_r & | & \mathbf{I}_n \end{bmatrix}}_{\tilde{\mathbf{G}}_r^R(s)}$$

Diese ermöglicht die Angabe oberer Schranken für die Fehlernormen

$$\|\mathbf{G}_e\|_{\mathcal{H}_2} \leq \|\mathbf{G}_\perp\|_{\mathcal{H}_2} \cdot \|\tilde{\mathbf{G}}_r^R\|_{\mathcal{H}_\infty}$$

$$\|\mathbf{G}_e\|_{\mathcal{H}_\infty} \leq \|\mathbf{G}_\perp\|_{\mathcal{H}_\infty} \cdot \|\tilde{\mathbf{G}}_r^R\|_{\mathcal{H}_\infty}$$

Wir brauchen obere Schranken für die Normen von $\mathbf{G}_\perp(s)$.

Im Folgenden nehmen wir hierzu an:

$$\begin{aligned} \mathbf{E} = \mathbf{E}^T > \mathbf{0} \\ \mathbf{A} + \mathbf{A}^T < \mathbf{0} \end{aligned} \quad \Rightarrow \quad \mu_{\mathbf{E}}(\mathbf{A}) = \max_i \lambda_i \left(\frac{\mathbf{A} + \mathbf{A}^T}{2}, \mathbf{E} \right) < 0$$

Typische Systeme zweiter Ordnung (z.B. aus der Strukturmechanik) lassen sich leicht entsprechend modellieren. [P. et al. 2012/2013]

Liegt das Originalmodell in strikt dissipativer Realisierung vor, dann ist $\mathbf{S} := -\mathbf{A} - \mathbf{A}^T > \mathbf{0}$ symmetrisch positiv definit und es gilt:

$$\|\mathbf{G}_e\|_{\mathcal{H}_\infty} \leq \left[\|\mathbf{C}\mathbf{S}^{-1}\mathbf{B}_\perp\|_2 + \sqrt{\|\mathbf{B}_\perp^T\mathbf{S}^{-1}\mathbf{B}_\perp\|_2 \|\mathbf{C}\mathbf{S}^{-1}\mathbf{C}^T\|_2} \right] \cdot \|\tilde{\mathbf{G}}_r\|_{\mathcal{H}_\infty}$$

```
function bnd = bndHinf(A,B,C)
```

```
L_S = chol(-A-A');
```

```
B_S = L_S' \ B;
```

```
C_S = C / L_S;
```

```
bnd = norm(full(C_S*B_S)) + ...  
      norm(full(B_S))*norm(full(C_S));
```

Geringe Ordnung.
Oft Allpass!

Es sei $\hat{\mathbf{Q}} = \mathbf{Z}\mathbf{Z}^T$ eine beliebige positiv semidefinite Matrix und \mathbf{R}_O das zugehörige Residuum, so dass

$$\mathbf{R}_O := \mathbf{A}^T \hat{\mathbf{Q}} \mathbf{E} + \mathbf{E}^T \hat{\mathbf{Q}} \mathbf{A} + \mathbf{C}^T \mathbf{C}.$$

Liegt das Originalmodell in strikt dissipativer Realisierung vor, dann gilt:

$$\|\mathbf{G}_\perp\|_{\mathcal{H}_2}^2 \leq \underbrace{\text{tr} \left[\mathbf{B}_\perp^T \hat{\mathbf{Q}} \mathbf{B}_\perp \right]}_{k_1} + \underbrace{\frac{1}{-2\mu_{\mathbf{E}}(\mathbf{A})}}_{>0} \cdot \underbrace{\max_i |\lambda_i(\mathbf{R}_O, \mathbf{E})|}_{k_2} \cdot \underbrace{\|\mathbf{B}_\perp^T \mathbf{E}^{-1} \mathbf{B}_\perp\|_F}_{k_3}$$

```
function bnd = BoundH2(A,B,C,E,mu,Z)
```

```
    k_1 = norm(Z'*B, 'fro')^2;
```

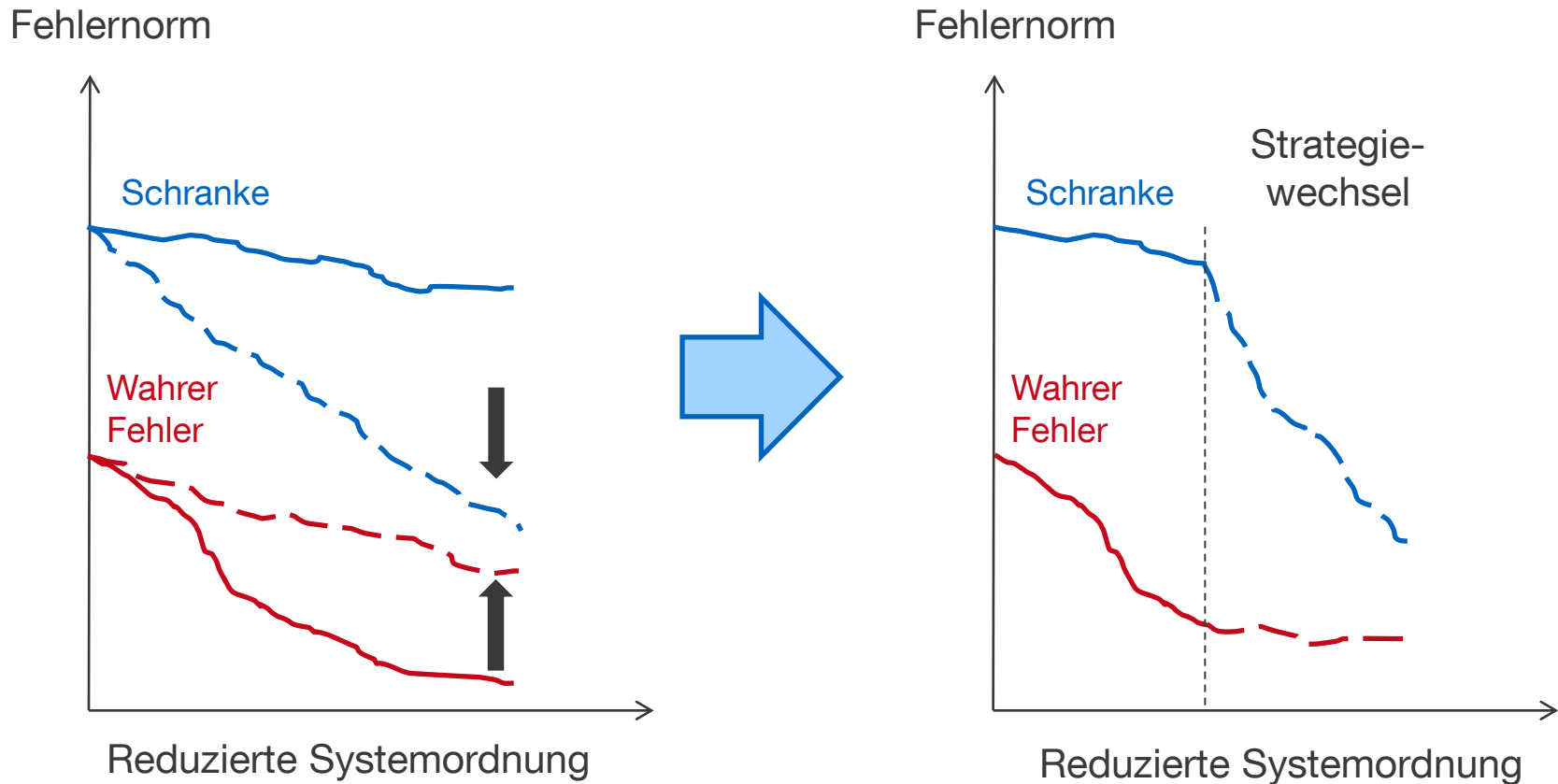
```
    R_O = @(x) (E*(Z*(Z'*(A*x)))) + (x'*E*Z*Z'*A)' + C'*(C*x));
```

```
    k_2 = abs(eigs(R_O, size(A,1), E, 1, 'LM'));
```

```
    k_3 = norm(B/E*B, 'fro');
```

```
    bnd = sqrt(k_1 + k_2*k_3/(-2*mu));
```

Paradigmenwechsel: Fehlerkontrollierte Modellreduktion



Lege Reduktionsvorgang darauf aus, die **Schranke** zu verkleinern, ohne Berücksichtigung des **tatsächlichen Fehlers**!

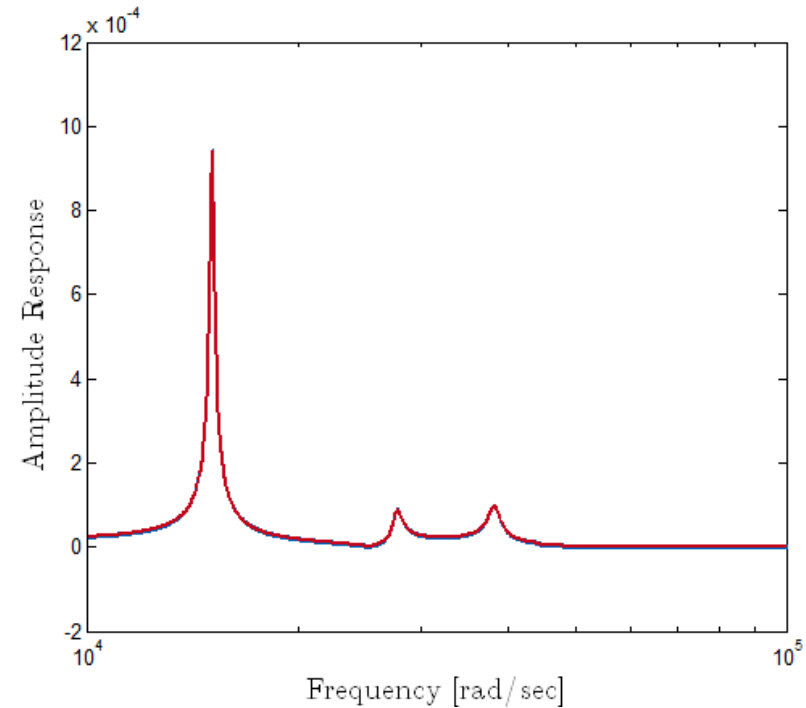
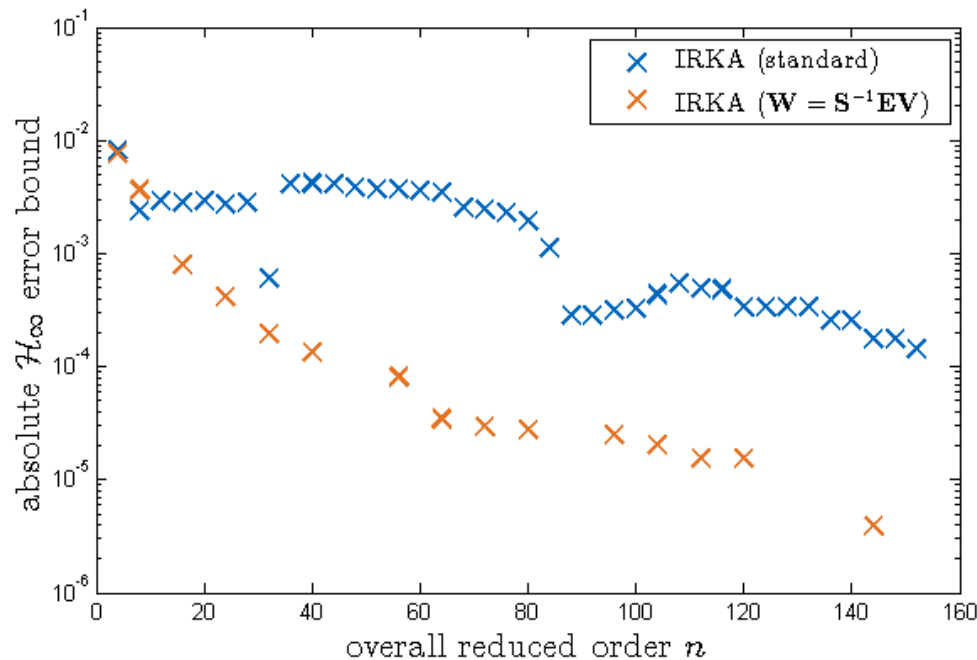
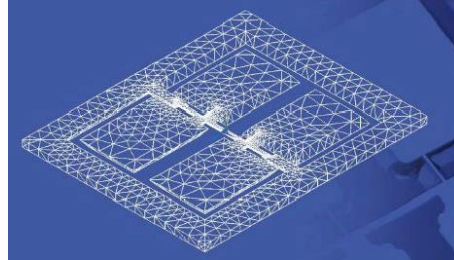
Numerisches Beispiel: Butterfly-Gyroskop

$$\begin{aligned} \mathbf{M}\ddot{\mathbf{z}}(t) + \mathbf{D}\dot{\mathbf{z}}(t) + \mathbf{K}\mathbf{z}(t) &= \mathbf{F}\mathbf{u}(t), \\ \mathbf{y}(t) &= \mathbf{S}\mathbf{z}(t) \end{aligned}$$

[Moosmann/Greiner 2004]

$$N = 34.722$$

$$nnz(A) = 4.084.636$$



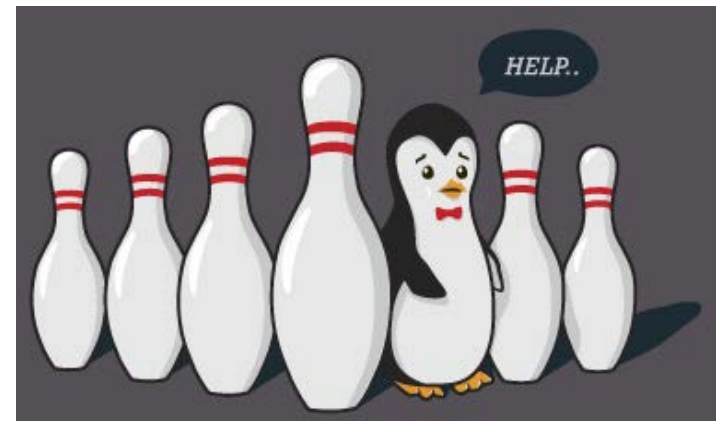
- Krylow-Unterraum-Verfahren erlauben die Faktorisierung des Fehlersystems.
- Dies ermöglicht die adaptive Wahl von Entwicklungspunkten sowie die
- Herleitung globaler Fehlerschranken bzgl. H_2 - und H_∞ -Norm für Systeme mit der Eigenschaft $\mathbf{E}=\mathbf{E}^T>\mathbf{0}$ und $\mathbf{A}+\mathbf{A}^T<\mathbf{0}$.
(Viele Systeme zweiter Ordnung lassen sich entsprechend modellieren.)
- Die Auswertung der Schranken erfordert geringen Aufwand.
- Der Reduktionsvorgang muss geeignet gestaltet werden, um akzeptable Überschätzung des wahren Fehlers zu erzielen.
(Hierauf konzentrieren sich unsere derzeitigen Bemühungen.)
- Die reduzierte Systemordnung kann dann automatisch angepasst werden.

Vielen Dank für Ihre Aufmerksamkeit!

Dipl.-Ing. Heiko K. F. Panzer

panzer@tum.de

<http://www.rt.mw.tum.de>



<http://de.lafraise.com/Article/index/id/1097>

A.C. Antoulas: *Approximation of Large-Scale Dynamical Systems*. SIAM, 2004.

S. Gugercin; A.C. Antoulas; C. Beattie: *H_2 model reduction for large-scale linear dynamical Systems*. SIAM Journal on Matrix Analysis and Applications, Vol. 30, 2008

K. Gallivan; A. Vandendorpe; P. Van Dooren: *Sylvester equations and projection-based model reduction*. Journal of Computational and Applied Mathematics, 2004.

H. Panzer; S. Jaensch; T. Wolf; B. Lohmann: *A Greedy Rational Krylov Method for H_2 -Pseudoptimal Model Order Reduction with Preservation of Stability*. American Control Conference, 2013.

H. Panzer; T. Wolf; B. Lohmann: *H_2 and H_∞ Error Bounds for Model Order Reduction of Second Order Systems by Krylov Subspace Methods*. European Control Conference, 2013.

H. Panzer; T. Wolf; B. Lohmann: *A Strictly Dissipative State Space Representation of Second Order Systems*. In at-Automatisierungstechnik, 2012.

H. Panzer; B. Kleinherne; B. Lohmann: *Analysis, Interpretation and Generalization of a Strictly Dissipative State Space Formulation of Second Order Systems*. Lohmann/Roppenecker (Hrsg.): Methoden und Anwendungen der Regelungstechnik, 2013.