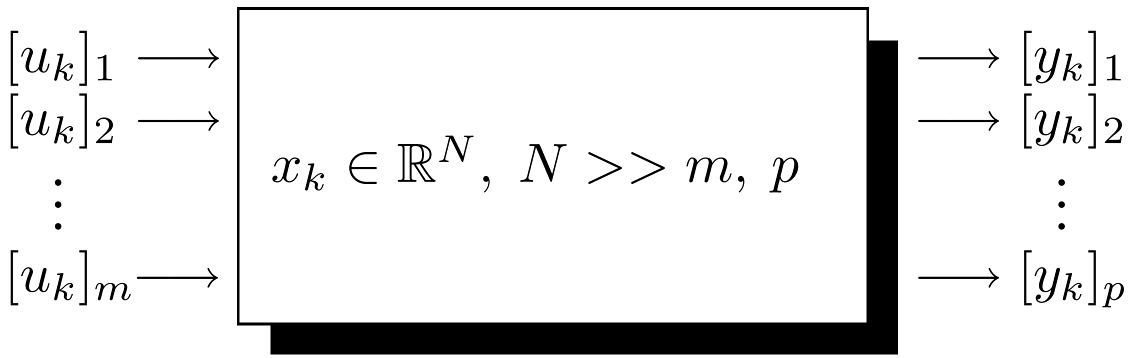
Multivariable $\mathcal{H}_2$-optimal approximation

P. Van Dooren, K.A. Gallivan, P.A. Absil

CESAME, Université Catholique de Louvain, Belgium
Department of Mathematics, Florida State University

Continuous- and Discrete-time State-space Models



(explicit) continuous time-invariant

$$\begin{cases} \dot{x} = Ax + Bu \\ y = Cx \end{cases}$$

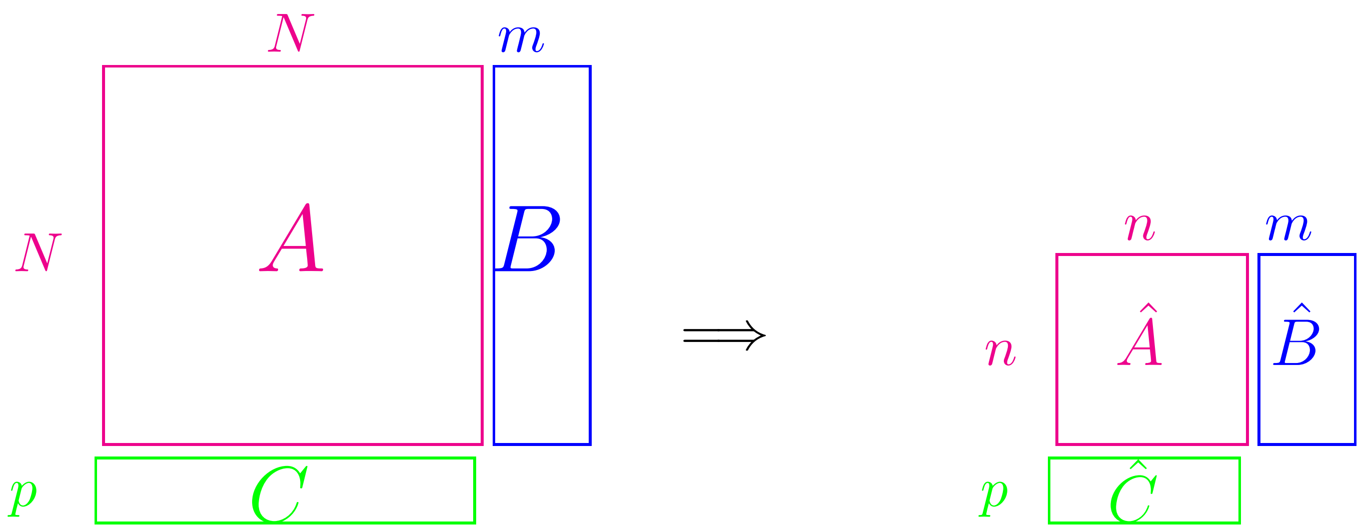
(explicit) discrete time-invariant

$$\begin{cases} x_{k+1} = Ax_k + Bu_k \\ y_k = Cx_k \end{cases}$$

(explicit) discrete time-varying

$$\begin{cases} x_{k+1} = A_k x_k + B_k u_k \\ y_k = C_k x_k \end{cases}$$

Time invariant model reduction idea



where $\quad n \ll N, \quad \hat{A} = W^T A V, \quad \hat{B} = W^T B, \quad \hat{C} = CV$

$P = VW^T$ is a projector for $W^T V = I_n$

Error model (CT)

The difference of the systems

$$\begin{cases} \dot{x} = Ax + Bu \\ y = Cx \end{cases} \quad \text{and} \quad \begin{cases} \dot{\hat{x}} = \hat{A}\hat{x} + \hat{B}u \\ \hat{y} = \hat{C}\hat{x} \end{cases}$$

is the error model, where $e := y - \hat{y}$

$$\begin{cases} \dot{\tilde{x}} = A_e\tilde{x} + B_e u \\ e = C_e\tilde{x} \end{cases} \qquad E(s) := H(s) - \hat{H}(s) = C_e(sI - A_e)^{-1}B_e$$

with

$$(A_e, B_e, C_e) := \left(\begin{bmatrix} A & \\ & \hat{A} \end{bmatrix}, \begin{bmatrix} B \\ \hat{B} \end{bmatrix}, \begin{bmatrix} C & -\hat{C} \end{bmatrix} \right),$$

Reduced Model via $\|E(s)\|_{\mathcal{H}_2}$ (CT)

$$\mathcal{J} = \|E(s)\|_{\mathcal{H}_2} = \mathrm{tr}\left(C_e P_e C_e^T\right) = \mathrm{tr}\left(B_e^T Q_e B_e\right)$$

where gramians P_e and Q_e solve the Lyapunov equations

$$A_e P_e + P_e A_e^T + B_e B_e^T = 0, \quad Q_e A_e + A_e^T Q_e + C_e^T C_e = 0$$

One can also partition

$$P_e := \begin{bmatrix} P & X \\ X^T & \hat{P} \end{bmatrix}, \quad Q_e := \begin{bmatrix} Q & Y \\ Y^T & \hat{Q} \end{bmatrix}$$

Reduced Model via $\|E(z)\|_{\mathcal{H}_2}$ (CT)

Solve partitioned equations for gramians

$$\begin{bmatrix} A & \\ & \hat{A} \end{bmatrix} \begin{bmatrix} P & X \\ X^T & \hat{P} \end{bmatrix} + \begin{bmatrix} P & X \\ X^T & \hat{P} \end{bmatrix} \begin{bmatrix} A^T & \\ & \hat{A}^T \end{bmatrix} + \begin{bmatrix} B \\ \hat{B} \end{bmatrix} \begin{bmatrix} B^T & \hat{B}^T \end{bmatrix} = 0,$$

and

$$\begin{bmatrix} A^T & \\ & \hat{A}^T \end{bmatrix} \begin{bmatrix} Q & Y \\ Y^T & \hat{Q} \end{bmatrix} + \begin{bmatrix} Q & Y \\ Y^T & \hat{Q} \end{bmatrix} \begin{bmatrix} A & \\ & \hat{A} \end{bmatrix} + \begin{bmatrix} C^T \\ -\hat{C}^T \end{bmatrix} \begin{bmatrix} C & -\hat{C} \end{bmatrix} = 0.$$

$$\mathcal{J} = \mathrm{tr}\left(B^T Q B + 2 B^T Y \hat{B} + \hat{B}^T \hat{Q} \hat{B} \right) = \mathrm{tr}\left(C P C^T - 2 C X \hat{C}^T + \hat{C} \hat{P} \hat{C}^T \right),$$

Gradients (CT)

Theorem 1.1 (Wilson 70, but rederived several times)**.**

$$\nabla_{\hat{A}}\mathcal{J} = 2(\hat{Q}\hat{P} + Y^T X), \quad \nabla_{\hat{B}}\mathcal{J} = 2(\hat{Q}\hat{B} + Y^T B), \quad \nabla_{\hat{C}}\mathcal{J} = 2(\hat{C}\hat{P} - CX),$$

where

$$A^T Y + Y\hat{A} - C^T\hat{C} = 0, \quad \hat{A}^T\hat{Q} + \hat{Q}\hat{A} + \hat{C}^T\hat{C} = 0,$$
$$X^T A^T + \hat{A}^T X^T + \hat{B}B^T = 0, \quad \hat{P}\hat{A}^T + \hat{A}\hat{P} + \hat{B}\hat{B}^T = 0.$$

Imposing zero gradients yields non-minimal optimality conditions

Tangential Interpolation

Given $(\Sigma_\sigma \in \mathbb{R}^{n\times n}, \;\; R \in \mathbb{R}^{m\times n})$ and $(\Sigma_\mu \in \mathbb{R}^{n\times n}, \;\; L \in \mathbb{R}^{p\times n})$

Solve two Sylvester equations:

$$AV - V\Sigma_\sigma + BR = 0,$$
$$W^T A - \Sigma_\mu^T W^T + L^T C = 0,$$

Define:

$$\{\hat{A}, \hat{B}, \hat{C}\} := \{(W^T V)^{-1} W^T A V, (W^T V)^{-1} W^T B, CV\},$$

If $W^T V$ is invertible, this uniquely determines the projected system $\{\hat{A}, \hat{B}, \hat{C}\}$.

Can specify $n(m+p)$ parameters which is good for generic strictly proper transfer functions Rat_{pm}^n.

Tangential Interpolation

Eigenvalues of Σ_σ and Σ_μ and columns of R and L specify the interpolation.

If realization is real then at most $n(m+p)$ real conditions due to conjugate pairs.

For $\lambda \in \mathbb{R}$

$$[H(s) - \hat{H}(s)] \sum_{i=1}^{k} r_i (s-\lambda)^{i-1} = O(s-\lambda)^k, \quad r_1^T r_1 = 1$$

$$\sum_{i=1}^{k} \ell_i^T (s-\lambda)^{i-1} [H(s) - \hat{H}(s)] = O(s-\lambda)^k, \quad \ell_1^T \ell_1 = 1$$

where $\{r_1, \cdots, r_k\}$ are columns from R, $\{\ell_1, \cdots, \ell_k\}$ are columns from L corresponding to the Jordan blocks.

Tangential Interpolation

For $\lambda,\ \bar{\lambda} \in \mathbb{C}$

$$[H(s) - \hat{H}(s)] \sum_{i=1}^{k} (r_{2i-1} + j r_{2i})(s-\lambda)^{i-1} = O(s-\lambda)^k,$$

$$\sum_{i=1}^{k} (\ell_{2i-1} + j\ell_{2i})^T (s-\lambda)^{i-1} [H(s) - \hat{H}(s)] = O(s-\lambda)^k,$$

$$r_1^T r_2 = 0,\ r_1^T r_1 + r_2^T r_2 = 1,\ \ell_1^T \ell_2 = 0,\ \ell_1^T \ell_1 + \ell_2^T \ell_2 = 1.$$

where $\{r_1, \cdots, r_{2k}\}$ are columns from R or $\{\ell_1, \cdots, \ell_{2k}\}$ are columns from L, corresponding each of the pair of Jordan blocks.

Stationary Points (CT)

At stationary points A, B, C and $\hat{A}, \hat{B}, \hat{C}$ are related via:

Theorem 1.2. *At every stationary point of $\mathcal{J}$ where $\hat{P}$ and $\hat{Q}$ are invertible, we have the following identities*

$$\hat{A} = W^T AV, \quad \hat{B} = W^T B, \quad \hat{C} = CV, \quad W^T V = I_n$$

where $W := -Y\hat{Q}^{-1}$, $V := X\hat{P}^{-1}$ and X, Y, $\hat{P}$ and $\hat{Q}$ satisfy the Sylvester equations

$$A^T Y + Y\hat{A} - C^T\hat{C} = 0, \quad \hat{A}^T\hat{Q} + \hat{Q}\hat{A} + \hat{C}^T\hat{C} = 0,$$

$$X^T A^T + \hat{A}^T X^T + \hat{B}B^T = 0, \quad \hat{P}\hat{A}^T + \hat{A}\hat{P} + \hat{B}\hat{B}^T = 0.$$

Stationary Points (CT)

Rewriting the Sylvester equations shows the relation with the tangential interpolation equations:

Stationary point:

$$W^T A + (\hat{Q}^{-1}\hat{A}\hat{Q})^T W^T + (\hat{C}\hat{Q}^{-1})C = 0,$$
$$AV + V(\hat{P}\hat{A}^T\hat{P}^{-1}) + B(\hat{B}^T\hat{P}^{-1}) = 0,$$

T.I. problem:

$$W^T A - \Sigma_\mu^T W^T + L^T C = 0,$$
$$AV - V\Sigma_\sigma + BR = 0,$$

$\hat{A} = -\Sigma_\sigma = -\Sigma_\mu$ implies tangential interpolation at the mirror images of the poles of $\hat{H}(s)$. (independent of order of the poles)

Distinct Poles (CT)

$$\hat{A}s_i = \hat{\lambda}_i s_i, \quad \hat{C}s_i = \hat{c}_i, \quad t_i^H \hat{A} = \hat{\lambda}_i t_i^H, \quad t_i^H \hat{B} = \hat{b}_i^H.$$

Theorem 1.3. *Let* $\hat{H}(s) = \sum_{i=1}^n \hat{c}_i \hat{b}_i^H/(s - \hat{\lambda}_i)$ *have distinct first order poles* $\hat{\lambda}_i$ *where* $(\hat{\lambda}_i, \hat{b}_i, \hat{c}_i),\ i = 1, \ldots, n$ *is self-conjugate, and let* $-\hat{\lambda}_i$ *not be poles of* $H(s)$*. Then*

$$\begin{aligned}
\frac{1}{2}(\nabla_{\hat{B}}\mathcal{J})^T s_i &= [H^T(-\hat{\lambda}_i) - \hat{H}^T(-\hat{\lambda}_i)]\hat{c}_i \\
\frac{1}{2}t_i^H(\nabla_{\hat{C}}\mathcal{J})^T &= \hat{b}_i^H[H^T(-\hat{\lambda}_i) - \hat{H}^T(-\hat{\lambda}_i)] \\
\frac{1}{2}t_i^H(\nabla_{\hat{A}}\mathcal{J})^T s_i &= \hat{b}_i^H \frac{d}{ds}\left[H^T(s) - \hat{H}^T(s)\right]\Big|_{s=-\hat{\lambda}_i} \hat{c}_i \\
\frac{1}{2}t_i^H(\nabla_{\hat{A}}\mathcal{J})^T s_j &= \frac{1}{2(\hat{\lambda}_i - \hat{\lambda}_j)}[\hat{b}_i^H(\nabla_{\hat{B}}\mathcal{J})^T s_j - t_i^H(\nabla_{\hat{C}}\mathcal{J})^T \hat{c}_j],\ i \neq j,
\end{aligned}$$

Distinct Poles (CT)

Corollary 1.4. *If* $(\nabla_{\hat{B}}\mathcal{J})^T = 0$, $(\nabla_{\hat{C}}\mathcal{J})^T = 0$ *and* $diag\ S^{-1}(\nabla_{\hat{A}}\mathcal{J})^T S = 0$ *then* $\nabla_{\hat{A}}\mathcal{J} = 0$ *and the following tangential interpolation conditions are satisfied for all* $\hat{\lambda}_i, i = 1, \ldots, n$ *:*

$$[H^T(-\hat{\lambda}_i) - \hat{H}^T(-\hat{\lambda}_i)]\hat{c}_i = 0, \quad \hat{b}_i^H[H^T(-\hat{\lambda}_i) - \hat{H}^T(-\hat{\lambda}_i)] = 0,$$

$$\hat{b}_i^H \frac{d}{ds}\left[H^T(s) - \hat{H}^T(s)]\right|_{s=-\hat{\lambda}_i} \hat{c}_i = 0. \tag{1}$$

Equivalently, in terms of the Taylor expansion of $H(s) - \hat{H}(s)$ *:*

$$[H^T(s) - \hat{H}^T(s)]\hat{c}_i = O(s + \hat{\lambda}_i), \quad \hat{b}_i^H[H^T(s) - \hat{H}^T(s)] = O(s + \hat{\lambda}_i),$$

$$\hat{b}_i^H[H^T(s) - \hat{H}^T(s)]\hat{c}_i = O(s + \hat{\lambda}_i)^2.$$

This yields $n(m + p)$ conditions.

Antoulas, Beattie, Gugercin; Bunse Gerstner et al.; Van Dooren et al,

A Second Order Case

Assume a real second order reduced model

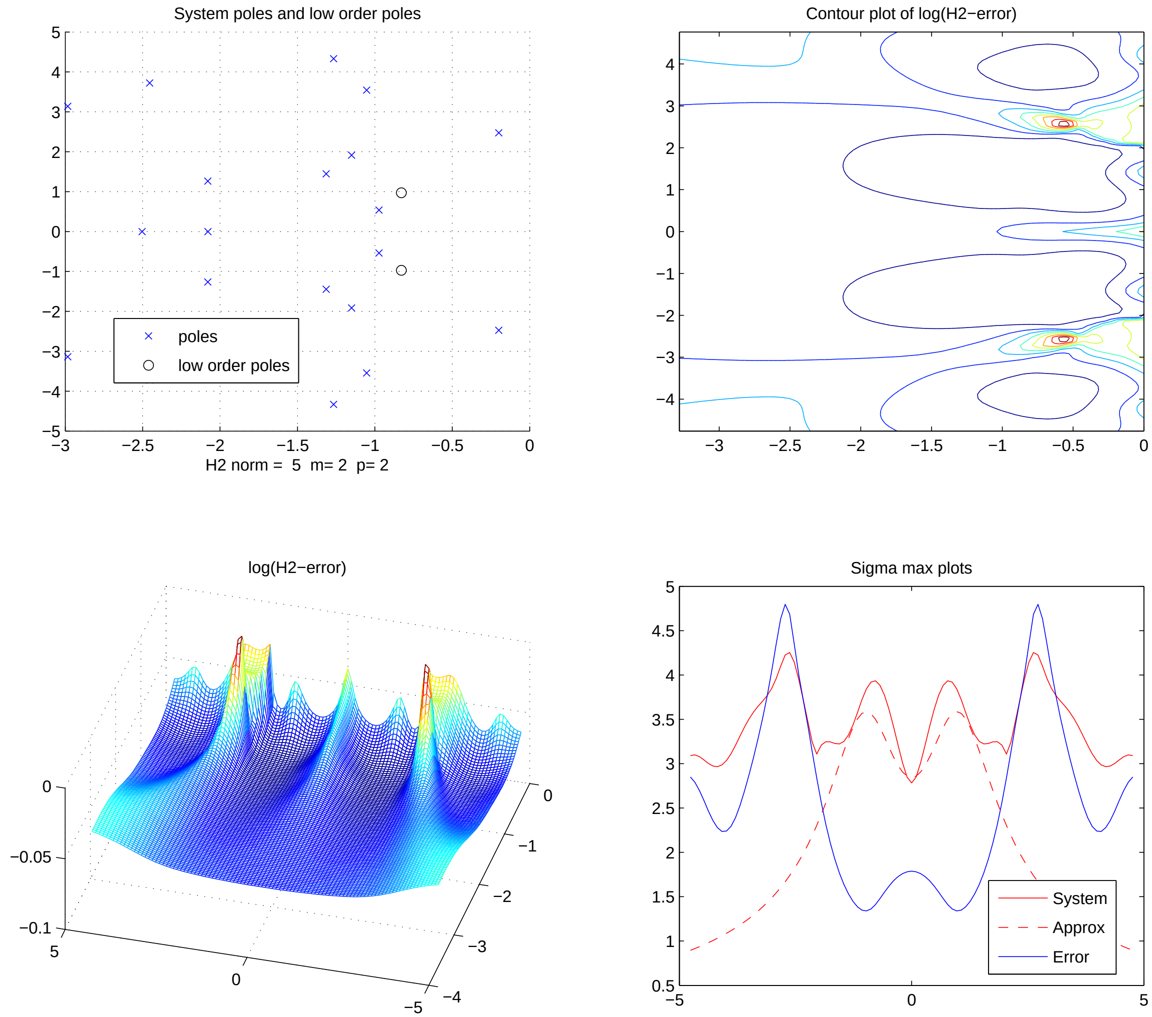
$$\hat{H}(s) = \frac{cb^H}{s-\lambda} + \frac{\overline{c}\overline{b}^H}{s-\overline{\lambda}}$$

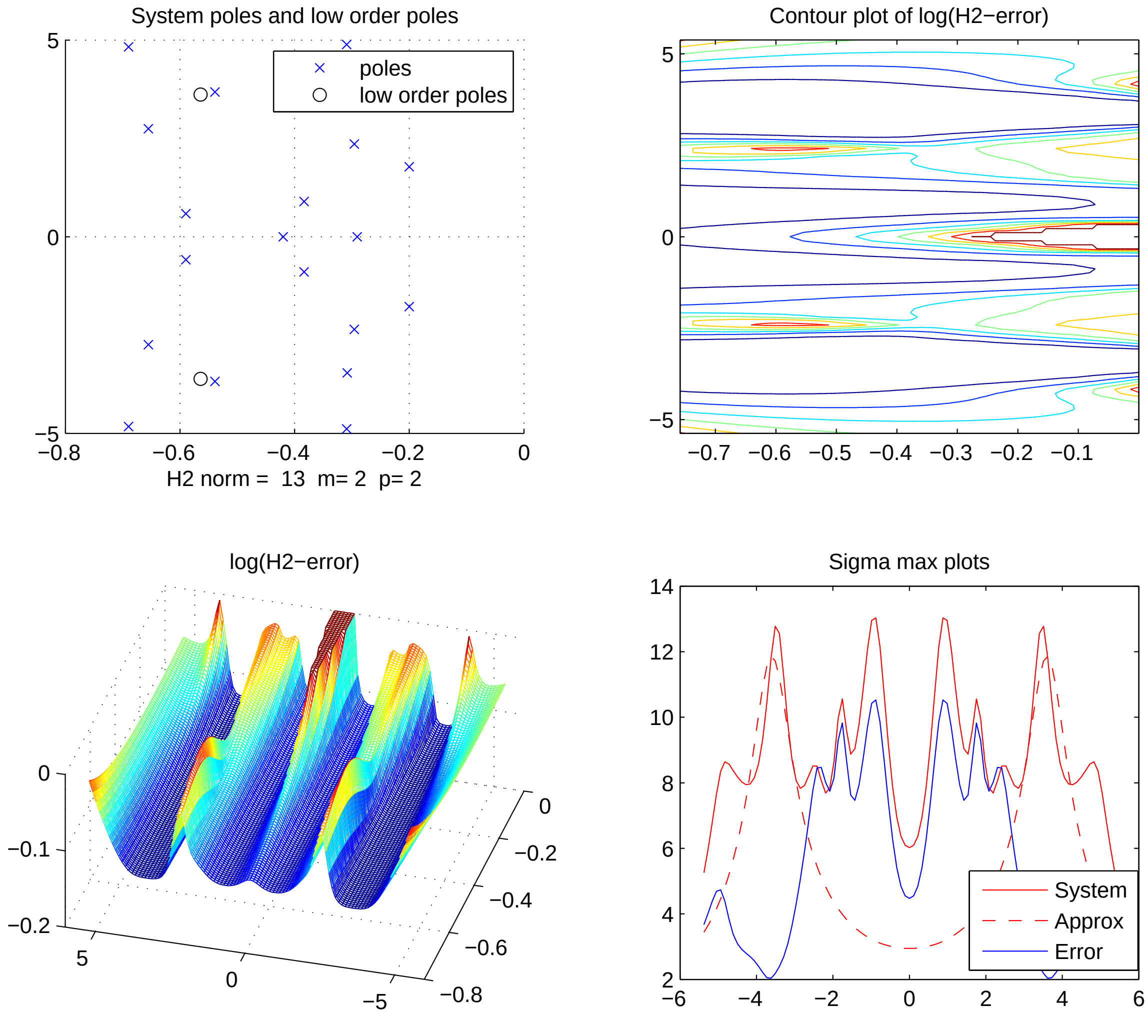
At every stationary point of $\mathcal{J}$:

$$H^T(-\lambda)c = -b\frac{c^H c}{2\lambda}, \quad b^H H^T(-\lambda) = -c^H \frac{b^H b}{2\lambda},$$
$$b^H \frac{d}{ds} H^T(s)c|_{s=-\lambda} = -\frac{b^H b c^H c}{4\lambda^2}.$$

$\therefore$ b and c must be the dominant singular vectors of $H^T(-\lambda)$ and can be eliminated from the optimization problem

Plot error $\|E(s)\|_{\mathcal{H}_2}$ as a function of interpolation point λ


System poles and low order poles
poles
low order poles
H2 norm = 5 m= 2 p= 2
Contour plot of log(H2−error)
log(H2−error)
Sigma max plots
System
Approx
Error


System poles and low order poles
poles
low order poles
H2 norm = 13 m= 2 p= 2
Contour plot of log(H2−error)
log(H2−error)
Sigma max plots
System
Approx
Error

Higher Order Poles (CT)

$$\hat{A}S_i = S_i\hat{A}_i, \quad \hat{C}S_i = \hat{C}_i, \quad T_i^H\hat{A} = \hat{A}_iT_i^H, \quad T_i^H\hat{B} = \hat{B}_i^H, \quad T_i^HS_i = I_k.$$

Theorem 1.5. *Let $\hat{H}(s) = \sum_{i=1}^{\ell}\hat{C}_i(sI - \hat{A}_i)^{-1}\hat{B}_i^H$ where $\hat{A}_i$ is a Jordan block of size k_i for $\hat{\lambda}_i$, and where $-\hat{\lambda}_i$ is not a pole of $H(s)$ or $\hat{H}(s)$. Then with*

$$\psi_{\hat{\lambda}_i}(s) := \begin{bmatrix}(s+\hat{\lambda}_i)^{k_i-1} & \dots & (s+\hat{\lambda}_i) & 1\end{bmatrix}, \;\; \hat{b}_i^H(s) := \psi_{\hat{\lambda}_i}(s)\hat{B}_i^H$$

$$\phi_{\hat{\lambda}_i}(s) := \begin{bmatrix}1 & (s+\hat{\lambda}_i) & \dots & (s+\hat{\lambda}_i)^{k_i-1}\end{bmatrix}^T, \;\; \hat{c}_i(s) := \hat{C}_i\phi_{\hat{\lambda}_i}(s)$$

we have

$$\frac{1}{2}(\nabla_{\hat{B}}\mathcal{J})^TS_i\phi_{\hat{\lambda}_i}(s) = [H^T(s) - \hat{H}^T(s)]\hat{c}_i(s) + O(s+\hat{\lambda}_i)^{k_i},$$

$$\frac{1}{2}\psi_{\hat{\lambda}_i}(s)T_i^H(\nabla_{\hat{C}}\mathcal{J})^T = \hat{b}_i^H(s)[H^T(s) - \hat{H}^T(s)] + O(s+\hat{\lambda}_i)^{k_i}$$

T.I. for Higher Order Poles (CT)

We do not have expressions for $T_i^H(\nabla_{\hat{A}}\mathcal{J})^T S_j$ that are clean extensions of the first order poles. We have:

Theorem 1.6. *With usual assumptions. if* $\nabla_{\hat{B}}\mathcal{J} = 0$, $\nabla_{\hat{C}}\mathcal{J} = 0$ *and* $\nabla_{\hat{A}}\mathcal{J} = 0$*, then the following tangential interpolation conditions are satisfied for* $i = 1, \ldots, \ell$*:*

$$[H^T(s) - \hat{H}^T(s)]\hat{c}_i(s) = O(s + \hat{\lambda}_i)^{k_i},$$

$$\hat{b}_i(s)^H[H^T(s) - \hat{H}^T(s)] = O(s + \hat{\lambda}_i)^{k_i},$$

$$\hat{b}_i(s)^H[H^T(s) - \hat{H}^T(s)]\hat{c}_i(s) = O(s + \hat{\lambda}_i)^{2k_i},$$

where $\hat{b}_i^H(s) := \psi_{\hat{\lambda}_i}(s)\hat{B}_i^H$ *and* $\hat{c}_i(s) := \hat{C}_i\phi_{\hat{\lambda}_i}(s)$.

These conditions do not follow easily from earlier tangential interpolation conditions of Vandendorpe et al.

Algorithms for minimizing $\|E(z)\|_{\mathcal{H}_2}$

Define $(X, Y, \hat{P}, \hat{Q}) = F(\hat{A}, \hat{B}, \hat{C})$ where

$$A^T Y \hat{A} - C^T \hat{C} = Y, \quad \hat{A}^T \hat{Q} \hat{A} + \hat{C}^T \hat{C} = \hat{Q},$$

$$\hat{A} X^T A^T + \hat{B} B^T = X^T, \quad \hat{A} \hat{P} \hat{A}^T + \hat{B} \hat{B}^T = \hat{P}$$

and then compute$(\hat{A}, \hat{B}, \hat{C}) = G(X, Y, \hat{P}, \hat{Q})$ from

$W := -Y\hat{Q}^{-1}, V := X\hat{P}^{-1}$ $\hat{A} = W^T A V, \hat{B} = W^T B, \hat{C} = CV,$

The fixed point of $(\hat{A}, \hat{B}, \hat{C}) = G(F(\hat{A}, \hat{B}, \hat{C}))$ are also stationary points of $\|E(z)\|_{\mathcal{H}_2}$ and satisfy the interpolation conditions

Simpler forms of $\hat{H}(s)$ make more efficient algorithms

One can also define a CG-like method or a Newton-like method

see Antoulas,Sorenson; Beattie-Gugercin

Theorem 1.7. *Let $\hat{H}(s) = \hat{C}(sI_n - \hat{A})^{-1}\hat{B}$ be a given stable n-th degree transfer function, then there always exists a stable N-th degree transfer function $H(s) = C(sI_N - A)^{-1}B$ with $N > n$, for which $\hat{H}(s)$ is a stationary point of the $\mathcal{H}_2$ error function.*

Theorem 1.8. *Let $\hat{H}(s) = \hat{C}(sI_n - \hat{A})^{-1}\hat{B}$ and $H(s) = C(sI_N - A)^{-1}B$ be stable and minimal transfer functions such that $\hat{H}(s)$ is a stationary point of the error function $J = \|H(s) - \hat{H}(s)\|_{\mathcal{H}_2}$. Then every (sufficiently) nearby transfer function $\hat{H}_\Delta(s)$ is a stationary point of a nearby system $H_\Delta(s)$. The same holds for every nondegenerate local minimum.*

$$\hat{H}(s) = 1/(s-a)^2, \quad a = -1$$

$$\hat{A} = \begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}, \hat{B} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \hat{C} = \begin{pmatrix} 1 & 0 \end{pmatrix}$$

$$H(s) = (0.25s^2 - 0.5s + 9.25)/(s^3 + 7s^2 + 19s + 9)$$

$$A = \begin{pmatrix} a & 1 & d \\ 0 & a & e \\ e & d & f \end{pmatrix}, B = \begin{pmatrix} 0 \\ 1 \\ g \end{pmatrix}, C = \begin{pmatrix} 1 & 0 & g \end{pmatrix}$$

$$f = -5, \quad g = .5, \quad d = 4ag, \quad e = 4a^2g$$

are stable and satisfies the stationarity conditions.

$\hat{H}$ is also a local minimum of $\|H - \hat{H}\|_{\mathcal{H}_2}$

- We have a completely general expression of the stationarity conditions of the $\mathcal{H}_2$ error function with $\hat{H}$ in Jordan canonical form and its relation to T.I.
- Using first order form of $\hat{H}(s)$ should be numerically sensitive if minimum is near a higher order pole minimum.
- For large scale problems, is this problem observed?
- If not common are there important classes of problems with higher order pole minima?
- Over-parameterization of $\hat{A},\ \hat{B},\ \hat{C}$ in optimization $\Leftrightarrow$ efficiency?

Some references

- $\mathcal{H}_2$ model reduction admits for efficient optimization
  Antoulas,Beattie,Gugercin
- Stationary points of time-invariant case amounts to interpolation
  Wilson; Antoulas,Beattie,Gugercin; Bunse Gerstner et al.; vDooren-Gallivan-Absil
- Higher order case and T.I.-based projection methods
  Gallivan,Vandendorpe,vDooren; vDooren-Gallivan-Absil
- Extension to discrete time-varying systems
  vDooren,Gallivan,Absil

Time-varying case

$$\begin{cases} x_{k+1} &= A_k x_k + B_k u_k \\ y_k &= C_k x_k \end{cases} \qquad \begin{cases} \hat{x}_{k+1} &= \hat{A}_k \hat{x}_k + \hat{B}_k u_k \\ \hat{y}_k &= \hat{C}_k \hat{x}_k \end{cases}$$

$$e_k := y_k - \hat{y}_k, \quad \mathcal{E} := \begin{cases} x^e_{k+1} &= A^e_k x^e_k + B^e_k u_k \\ e_k &= C^e_k x^e_k \end{cases}$$

where $A^e_k := \begin{bmatrix} A_k & \\ & \hat{A}_k \end{bmatrix}, \quad B^e_k = \begin{bmatrix} B_k \\ \hat{B}_k \end{bmatrix}, \quad C^e_k = \begin{bmatrix} C_k & -\hat{C}_k \end{bmatrix}$

$$x^e_{k_0} = 0 \Rightarrow x^e_k = \sum_{i=k_0}^{k-1} \Phi^e_{k,i+1} B^e_i u_i, \quad \Phi^e_{k+1,i} = A^e_k \Phi^e_{k,i} \ (k \geq i), \quad \Phi^e_{k,k} = I$$

The "stacked" error system response is $\tilde{e} = E\tilde{u}$ and the cost function to minimize are given by

$$\|\mathcal{E}\|^2_{\mathcal{H}_2} := \mathcal{J}(k_0, k_f) := \mathrm{tr}(E^T E) = tr(EE^T)$$

It follows that

$$\mathcal{J}(k_0, k_f) := \mathrm{tr} \sum_{k=k_0+1}^{k_f+1} C_k^e P_k^e C_k^{e^T} = \mathrm{tr} \sum_{k=k_0}^{k_f} B_k^{e^T} Q_k^e B_k^e$$

where

$$P_{k+1}^e = \begin{bmatrix} A_k & \\ & \hat{A}_k \end{bmatrix} P_k^e \begin{bmatrix} A_k^T & \\ & \hat{A}_k^T \end{bmatrix} + \begin{bmatrix} B_k \\ \hat{B}_k \end{bmatrix} \begin{bmatrix} B_k^T & \hat{B}_k^T \end{bmatrix}, \quad P_{k_0}^e = 0$$

$$Q_{k-1}^e = \begin{bmatrix} A_k^T & \\ & \hat{A}_k^T \end{bmatrix} Q_k^e \begin{bmatrix} A_k & \\ & \hat{A}_k \end{bmatrix} + \begin{bmatrix} C_k^T \\ \hat{C}_k^T \end{bmatrix} \begin{bmatrix} C_k & \hat{C}_k \end{bmatrix}, \quad Q_{k_f+1}^e = 0$$

Gradients are given by

$$\nabla_{\hat{A}_k}\mathcal{J} = 2(\hat{Q}_k\hat{A}_k\hat{P}_k + Y_k^T A_k X_k),$$

$$\nabla_{\hat{B}_k}\mathcal{J} = 2(\hat{Q}_k\hat{B}_k + Y_k^T B_k),$$

$$\nabla_{\hat{C}_k}\mathcal{J} = 2(\hat{C}_k\hat{P}_k - C_k X_k)$$

Updating rules and fixed point results are as before

$$W_k := Y_k\hat{Q}_k^{-1}, V_k = X_k\hat{P}_k^{-1}$$

$$(A_k^e, B_k^e, C_k^e) := (W_k^T A_k V_k, W_k^T B_k, C_k V_k).$$

where $X_k, \quad Y_k, \quad \tilde{P}_k, \quad \tilde{Q}_k$ satisfy Stein like recurrences

$$X_{k+1} = A_k X_k \hat{A}_k^T + B_k\hat{B}_k^T, \quad X_{k_0} = 0$$

$$\hat{P}_{k+1} = \hat{A}_k\hat{P}_k\hat{A}_k^T + \hat{B}_k\hat{B}_k^T, \quad \hat{P}_{k_0} = 0$$

$$Y_{k-1} = A_k^T Y_k \hat{A}_k^T - C_k^T\hat{C}_k, \quad Y_{k_f+1} = 0$$

$$\hat{Q}_{k-1} = \hat{A}_k^T Q_k \hat{A}_k + \hat{C}_k^T\hat{C}_k, \quad \hat{Q}_{k_f+1} = 0$$