SET MEMBERSHIP IDENTIFICATION WITH UNMODELED DYNAMICS

Michele TARAGNA

Dipartimento di Automatica e Informatica

Politecnico di Torino

michele.taragna@polito.it

III Level Course 01LCPIU

“Experimental modeling: model building from experimental data”

Summary

- Motivations for the identification of dynamic uncertainty models
- Model set identification theory
- Identification errors and optimality concepts
- Assumptions on system and noise and their validation
- Norms for measuring the identification error: $\mathcal{H}_\infty$, ℓ_1, $\mathcal{H}_2$
- $\mathcal{H}_\infty$ identification:
  - validation methods
  - linear, “two-stage”, interpolatory algorithms
  - optimal and “nearly-optimal” algorithms
  - reduced order uncertainty models
  - (identification-control interaction)

Motivations

Problem: make *inferences* on an actual dynamic system S^o starting from general information on it and from a finite number N of noisy measurements

$$y^N = F_N\left(S^o\right) + e^N$$

measurement vector (known)	"information operator" (known)	measurement noise (unknown)

Examples of inference:
- prediction of the future response of S^o
- control design of the future response of S^o
- fault detection
- diagnosis
- . . .

- **Experimental information**
  - *Time-domain measurements*: N samples of the output y of the system S^o, initially at rest, fed by a known one-sided input u ($u_\ell = 0$ for $\ell < 0$, $u_0 \neq 0$)

$$y_\ell = \sum_{k=0}^{\ell} h_k^{S^o} u_{\ell-k} + e_\ell, \quad \ell = 0, \ldots, N-1$$

$$\Downarrow$$

information operator : $F_N(S) = F_N h^S$

with $F_N := \begin{bmatrix} U_N & 0_{N\times\infty} \end{bmatrix} \in \mathbb{R}^{N\times\infty}$, $U_N := \begin{bmatrix} u_0 & 0 & \cdots & 0 \\ u_1 & u_0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ u_{N-1} & u_{N-2} & \cdots & u_0 \end{bmatrix} \in \mathbb{R}^{N\times N}$,

$h^S := \{h_0^S, h_1^S, \ldots\}$ = impulse response of S

– *Frequency-domain measurements*: real and imaginary parts of $N/2$ samples of the system frequency response $S^o(\omega_k) = \sum_{\ell=0}^{\infty} h_\ell^{S^o} e^{-j\ell\omega_k}$, $k = 1, \ldots, N/2$:

$$\begin{cases} y_{2k-2} &= \Re e\left(S^o\left(\omega_k\right)\right) + e_{2k-2}, \\ y_{2k-1} &= \Im m\left(S^o\left(\omega_k\right)\right) + e_{2k-1}, \end{cases} \qquad k = 1, \ldots, N/2$$

$$\Downarrow$$

information operator : $F_N\left(S\right) = F_N h^S$

$$F_N := \left[\begin{array}{ccc} \Omega^T(\omega_1) & \cdots & \Omega^T(\omega_{N/2}) \end{array}\right]^T \in \mathbb{R}^{N\times\infty}$$

$$\Omega(\omega) := \left[\begin{array}{c} \Omega_1(\omega) \\ \Omega_2(\omega) \end{array}\right] = \left[\begin{array}{c} \Re e\left(\Psi(\omega)\right) \\ \Im m\left(\Psi(\omega)\right) \end{array}\right] \in \mathbb{R}^{2\times\infty}$$

$$\Psi(\omega) := \left[\begin{array}{cccc} 1 & e^{-j\omega} & e^{-j2\omega} & \ldots \end{array}\right] \in \mathbb{C}^{1\times\infty}$$

- **Approach:**
  - estimate a model $\hat{M}$
  - make the inference on the model $\hat{M}$
  - compute the *inference error* $E(\hat{M})$ due to the fact that the inference is made on $\hat{M}$ instead of S^o

- **Problems:**
  - given a model $\hat{M}$, compute the error $E(\hat{M})$
  - find a model $\hat{M}$ that gives a "small" error $E(\hat{M})$
  - analyze the asymptotic behavior of the error $E(\hat{M})$ for $N \to \infty$
  - design the experiment in order to have a "small" error $E(\hat{M})$

- With no assumption on S^o and on e^N, the error $E(\hat{M})$ is unbounded

$$\Downarrow$$

which kind of "a priori" information on S^o and e^N can be reasonably assumed?

Classical statistical identification theory

- **Assumptions**:
  - $S^o \in K = \{M_n(p),\ p \in \mathbb{R}^n\}$: set of parametric dynamic models
    (for example, models with rational transfer function of a given order)
  - e^N: stochastic noise with known probability density function
    (possibly filtered by a parametric noise model)

$$\Downarrow$$

statistical parametric estimation problem:

$$y^N = F_N(M_n(p)) + e^N$$

- If $S^o \notin K$, as it happens in most practical applications, then

$$y^N = F_N(M_n(p)) + e^N + \text{unmodeled dynamics}$$

$$\Downarrow$$

how to account for the unmodeled dynamics?

Model set identification theory

- **Assumptions**:
  - $S^o \in K =$ nonfinitely parameterized subset $\mathcal{S}$ of dynamic systems
    (e.g., systems with an exponentially decaying impulse response)
  - $e^N \in \mathcal{B}_e =$ bounded set
- **Feasible Systems Set**: set of all the systems consistent with the assumptions and the experimental data (*unfalsified systems set*)

$$FSS\left(K, \mathcal{B}_e, F_N, y^N\right) := \left\{S \in K : y^N = F_N\left(S\right) + \tilde{e}^N,\ \ \tilde{e}^N \in \mathcal{B}_e\right\}$$

  - the FSS summarizes the overall information on the system to be identified (assumptions on S^o and e^N, data y^N)
  - $S^o \in FSS$ if the assumptions are "true" $\Rightarrow$
    validating the assumptions $\Leftrightarrow$ checking that the FSS is not void
  - if the assumptions are too weak $\Rightarrow$ the FSS is huge or even unbounded

- Problem:

  in most cases, the FSS is not represented in a form suitable to be used $\Rightarrow$ model sets with a *suitable* form and *including the* FSS are looked for

$$\Downarrow$$

- **Model set**: a set of models $\mathcal{M} \subseteq \mathcal{S}$ is called a model set for S^o if

$$\mathcal{M} \supseteq FSS \ni S^o$$

- *Additive* model sets are considered, of the form*:*

$$\mathcal{M} = \left\{ \hat{M} + \Delta : ||\Delta||_{\mathcal{S}} \leq \gamma \right\}$$

  $\hat{M} \in \mathcal{S}$: nominal model

  $\Delta = S^o - \hat{M}$: model error (unmodeled dynamics)

  $||\cdot||_{\mathcal{S}}$: norm in the space of the dynamic systems $\mathcal{S}$

- **Identification algorithm ϕ**:

$$\phi\left(K, \mathcal{B}_e, F_N, y^N\right) = \hat{M}$$

- Measure of the error due to the inference made on $\hat{M}$ instead of S^o:

$$\delta(\hat{M}) = \|S^o - \hat{M}\|_{\mathcal{S}} \leq \sup_{S \in FSS} \|S - \hat{M}\|_{\mathcal{S}}$$

it is the tightest upper bound and it is bounded thanks to the assumptions on S^o

$$\Downarrow$$

given a model $\hat{M}$, the "smallest" additive model set is:

$$\hat{\mathcal{M}} = \left\{ \hat{M} + \Delta : ||\Delta||_{\mathcal{S}} \leq \sup_{S \in FSS} \|S - \hat{M}\|_{\mathcal{S}} \right\}$$

- **Identification error of an algorithm ϕ, identification error of a model set $\hat{\mathcal{M}}$:**

$$E(\phi) = E(\hat{\mathcal{M}}) := \sup_{S \in FSS(K, \mathcal{B}_e, F_N, y^N)} \| S - \phi\left(K, \mathcal{B}_e, F_N, y^N\right) \|_{\mathcal{S}}$$

  - it is the minimal inference error guaranteed by the model $\hat{M}$, identified by the algorithm ϕ applied to the available information $\left(K, \mathcal{B}_e, F_N, y^N\right)$
  - it is not conservative with respect to the noise, since it accounts for the effective noise realization during the experiment

- **Optimal algorithm, optimal model set**

  i) a model set $\mathcal{M}^*$ is optimal if it minimizes the inference error for the *actual* set of available information $\left(K, \mathcal{B}_e, F_N, y^N\right)$:

$$E(\mathcal{M}^*) = \inf_{\hat{\mathcal{M}} \subseteq \mathcal{S}} E(\hat{\mathcal{M}}) = r\left(K, \mathcal{B}_e, F_N, y^N\right)$$

$$\begin{aligned} r\left(K, \mathcal{B}_e, F_N, y^N\right) &:= \inf_{\phi} \sup_{S \in FSS} \left\| S - \phi\left(K, \mathcal{B}_e, F_N, y^N\right) \right\|_{\mathcal{S}} \\ &= \text{radius of information} \end{aligned}$$

  ii) an algorithm ϕ^* is optimal if it minimizes the inference error for *any* possible set of information $\left(K, \mathcal{B}_e, F_N, y^N\right)$:

$$E\left(\phi^*\right) = r\left(K, \mathcal{B}_e, F_N, y^N\right), \quad \forall \left(K, \mathcal{B}_e, F_N, y^N\right)$$

– the *central algorithm* is optimal, defined as:

$$\phi^c\left(y^N\right) = M^c = \underbrace{\arg \inf_{M \in \mathcal{S}} \sup_{S \in FSS} \|S - M\|_{\mathcal{S}}}_{\text{Chebicheff center of the } FSS}$$

$\Rightarrow$ the corresponding *central model set* is optimal, defined as:

$$\mathcal{M}^c = \{M^c + \Delta : ||\Delta||_{\mathcal{S}} \leq r\}$$

– in many cases, the central model M^c cannot be computed

$\Downarrow$

suboptimal algorithms are considered

- **Suboptimal algorithms**
  - * linear (parametric) algorithms
  - * two-steps nonlinear (nonparametric) algorithms

    Helmicki-Jacobson-Nett: *TAC* 1991; Partington: *IJC* 1991;

    Gu-Khargonekar: *Automatica* 1992, *TAC* 1992; . . .
  - * interpolatory nonlinear (nonparametric) algorithms

    Chen-Nett-Fan: *ACC* 1992, *TAC* 1995; Bai-Raman: *ACC* 1992;

    Chen-Nett: *CDC* 1993, *TAC* 1995; Gu-Xiong-Zhou: *S&CL* 1993; . . .
  - * mixed parametric - nonparametric algorithms

    Wahlberg-Ljung: *TAC* 1992; Kosut-Boyd: *TAC* 1992; Younce-Rohrs: *TAC* 1992;

    Zhou-Kimura: *TAC* 1995; Giarré-Milanese-Taragna: *TAC* 1997; . . .

- **Suboptimal algorithms**
  - they are computationally *simpler* than the central algorithm,
    but they require to accept a *deterioration* in the identification accuracy
  - the nominal model can be in a form not suitable for the following use
    (for example, too high order, ...) $\Rightarrow$
    algorithms are looked for that provide model sets with *low order* nominal models,
    but they imply a further *deterioration* in the identification quality

$$\Downarrow$$

how to measure such a deterioration?

- **Optimality level of an algorithm ϕ:**

  an algorithm ϕ is α-optimal if $\exists \alpha \in \mathbb{R}$ such that

$$E\left(\phi\left(K,\mathcal{B}_e,F_N,y^N\right)\right) \leq \alpha \cdot r\left(K,\mathcal{B}_e,F_N,y^N\right),\ \forall\left(K,\mathcal{B}_e,F_N,y^N\right)$$

  - it measures the algorithm quality with respect to the optimal algorithm in the *worst case*, considering any possible information (assumptions and data)
  - the interpolatory nonlinear algorithms are 2-optimal for any norm and for this reason they are called “almost optimal”

- **Optimality level of a model set $\hat{\mathcal{M}}$:**

$$\alpha(\hat{\mathcal{M}}) = E(\hat{\mathcal{M}}) \Big/ r \geq 1$$

  - it measures the *actual* quality of the identified model set with respect to the optimal one, considering the available information only (assumptions and data)
  - the exact evaluation of $\alpha(\hat{\mathcal{M}})$ is difficult, because in general the computation of $E(\hat{\mathcal{M}})$ and r is a very hard problem

- **Convergence**:

  an algorithm ϕ is termed *convergent* if

$$\lim_{\substack{N\to\infty \\ \mathcal{B}_e\to\emptyset}} E(\phi) = 0$$

- **Robust convergence**:

  an algorithm ϕ is termed *robustly convergent* if it is convergent

  independently of the a priori information on the system and on the noise

Classification

- **Norm $||\cdot||_{\mathcal{S}}$ to measure the identification error**

$$\begin{aligned} \|S(z)\|_{\infty}^{W} &:= \sup_{0\le\omega\le 2\pi} \left|W^{-1}(\omega)S(\omega)\right| && \Rightarrow \mathcal{H}_{\infty} \text{ identification} \\ \|S(z)\|_{1} &:= \sum_{k=-\infty}^{\infty} \left|h_k^S\right| && \Rightarrow \ell_1 \text{ identification} \\ \|S(z)\|_{2} &:= \sqrt{\sum_{k=-\infty}^{\infty} \left|h_k^S\right|^2} && \Rightarrow \mathcal{H}_2 \text{ identification} \end{aligned}$$

- **"A priori" information on the measurement noise e^N**

$$e^N \in \mathcal{B}_{e,\infty} := \left\{\tilde{e}^N \in \mathbb{R}^N : ||A\tilde{e}^N||_{\infty}^{W_e} \le \varepsilon\right\}$$

$$e^N \in \mathcal{B}_{e,2} := \left\{\tilde{e}^N \in \mathbb{R}^N : ||\tilde{e}^N||_2 \le \varepsilon\right\}$$

$\mathcal{B}_{e,\infty}$ is more suitable to deterministically characterize the measurement noise, similarly to what happens with the statistical identification

- **"A priori" information on the system S^o**

  $\mathcal{S}$: Banach space of dynamic, SISO, discrete-time, causal, LTI, BIBO-stable, possibly infinite dimensional systems, with

  $h^S := \left\{h_k^S\right\}_{k=0}^{\infty}$: impulse response of $S \in \mathcal{S}$

  $S(z) := \sum_{k=0}^{\infty} h_k^S z^k$: transfer function of S (or Λ-transform of h^S)

  $S^o \in \mathcal{H}_\infty(\mathbb{D})$: actual system to be identified, partially known, with

$$\mathcal{H}_\infty(\mathbb{D}) := \left\{ f : \mathbb{D} \to \mathbb{C} \mid f \text{ analytic in } \mathbb{D},\ \|f\|_\infty := \operatorname*{ess\,sup}_{z \in \mathbb{D}} |f(z)| < \infty \right\}$$

$$\mathbb{D} := \{z \in \mathbb{C} : |z| < 1\} \text{ : open unitary circle}$$

1. S^o is an exponentially stable system, with a known maximum steady-state gain $M > 0$ with respect to an input $\rho^{-k} e^{i\omega k}$, $\rho > 1$:

$$S^o \in K^{(1)}_{\rho,M} := \{S \in \mathcal{H}_\infty(\mathbb{D}) : S(z) \in \mathcal{H}_{\rho,M}(\mathbb{D})\}$$

$$\mathcal{H}_{\rho,M}(\mathbb{D}) := \left\{ f : \mathbb{D} \to \mathbb{C} \mid f \text{ analytic in } \mathbb{D}_\rho,\ \|f\|_{\infty,\rho} := \sup_{z \in \mathbb{D}_\rho} |f(z)| \leq M \right\}$$

$$\mathbb{D}_\rho := \{z \in \mathbb{C} : |z| \leq \rho\} : \text{closed circle of radius } \rho$$

[Helmicki-Jacobson-Nett: *TAC* '91; Partington: *IJC* '91; Gu-Khargonekar: *TAC* '92; Gu-Xiong-Zhou: *SCL* '93; Chen-Nett-Fan: *TAC* '95; . . .]

2. S^o is an exponentially stable system, with an envelope $M\rho^{-k}$ on the impulse response, being $M > 0$ and $\rho > 1$ known constants:

$$S^o \in K^{(2)}_{\rho,M} := \left\{S \in \mathcal{H}_\infty(\mathbb{D}) : \left|h^S_k\right| \leq M\rho^{-k},\ \forall k \geq 0\right\}$$

[Giarré-Milanese-Taragna: *TAC* '97; Milanese-Taragna: *TAC* '02; . . .]

3. S^o has a known upper bound $\gamma > 0$ on the absolute value of the transfer function derivative:

$$S^o \in K_\gamma^{(3)} \doteq \left\{ S \in \mathcal{H}_\infty(\mathbb{D}) : \sup_{z \in \mathbb{D}} \left| \frac{dS(z)}{dz} \right| \leq \gamma \right\}$$

[Glaum-Lin-Zames: *CDC*'96]

4. S^o has a known upper bound $\gamma > 0$ on the absolute value of the real and imaginary parts of the transfer function derivative:

$$S^o \in K_\gamma^{(4)} \doteq \left\{ S \in \mathcal{H}_\infty(\mathbb{D}) : \sup_{\omega \in [0,2\pi]} \left| \frac{dS_{R/I}(\omega)}{d\omega} \right| \leq \gamma \right\}$$

[Milanese-Novara-Taragna: *ECC*'01]

$$K_{\rho,M}^{(1)} \subset K_{\rho,M}^{(2)} \subset K_\gamma^{(3)} \subset K_\gamma^{(4)}, \text{ with } \gamma = M\rho \Big/ (\rho - 1)^2$$

$\mathcal{H}_\infty$ identification

- **Norm $||\cdot||_{\mathcal{S}}$ to measure the identification error**

$$||S||_{\mathcal{S}} = \|S(z)\|_\infty^W = \sup_{0\leq\omega\leq 2\pi} \left|W^{-1}(\omega)S(\omega)\right|$$

$W^{-1}(\omega)$: weight function *suitably* chosen according to the use of the model (for example, for the H_∞ robust control)

Linear algorithms

- The linear algorithms work linearly on the experimental data
- The linear algorithm structure is extremely simple, being possibly weighted least-squares algorithms
- The linear algorithms can be classified in two different categories:
  - the “untuned” linear algorithms
  - the “tuned” linear algorithms

  according to the fact that they are independent or not on the available a priori information on the system and the noise
- The linear algorithm limitations justify the investigation of more sophisticated nonlinear algorithms

Fundamental limitations of linear algorithms

- In the $\mathcal{H}_\infty$ identification case, no α-optimal linear algorithm exists for any (whatever huge) value α
- No linear and robustly convergent algorithm exists, mainly because the projection operator is unbounded
- To guarantee at least the convergence, a linear algorithm has to be “tuned”, i.e., it has to explicitly account for the available a priori information

"Tuned" linear algorithm for frequency-domain measurements, assuming that

- the system $S^o \in K^{(1)}_{\rho,M}$, with $\rho > 1$ and $M \geq 0$
- the noise $e^N \in \mathcal{B}_{e,\infty}$, with $\varepsilon \geq 0$ and $A = W_e = I_{N\times N}$

$$\hat{M}(z) = \sum_{k=0}^{n-1} q_k^* \, z^k$$

where

$$q^* = [q_k^*] = \left[\frac{c_k\left(y^N\right)}{1 + \left(\frac{\varepsilon + M\rho^{-n}}{M}\right)^2 \rho^{2k}} \right]$$

being $c_k\left(y^N\right)$ the inverse DFT coefficients of y^N

$$c_k\left(y^N\right) = \frac{1}{N} \sum_{i=0}^{N-1} y_i \left(e^{j2\pi/N}\right)^{-ik}$$

“Two-stage” nonlinear algorithms

- To overcome the convergence limitations of linear algorithms, “untuned” nonlinear algorithms are looked for, performing the following two-step procedure:
  - step 1: a preliminary noncausal model $\hat{M}^{(0)} \in \mathcal{L}_\infty$ is built through an “untuned” linear algorithm that performs a bilateral interpolation

$$\hat{M}^{(0)}(z) = \sum_{k=-n+1}^{n-1} w_{k,n}\, c_k(y)\, z^k$$

    where $\{w_{k,n}\}$ is a sequence of weights called *window*
  - step 2: the identified model is chosen as the best approximation in $\mathcal{H}_\infty(\mathbb{D})$ of $\hat{M}^{(0)}$ (Nehari nonlinear approximation problem)

$$\hat{M}(z) = \arg\min_{M \in \mathcal{H}_\infty(\mathbb{D})} \|\hat{M}^{(0)} - M\|_\infty$$

The two-stage algorithms proposed in literature mainly differ in step 1:

“Two-stage” nonlinear algorithm #1:

- step 1: approximation through linear splines and truncation

“Two-stage” nonlinear algorithm #2:

- step 1: approximation through weighting sequences $\{w_{k,n}\}$ that are symmetrical even with respect to k (e.g., triangular, sinc, cos or trapezoidal windows), truncated for $k \geq n$.

Properties and limitations of the “two-stage” nonlinear algorithms

- They are robustly convergent if the weighting sequence $\{w_{k,n}\}$ is independent of the a priori information
- The identification error and the identified model order depend on the kind of window used
- They maintain a rather simple structure
- Their optimality level is unknown
- The identified model may not belong to the set FSS of the systems consistent with the overall information on the system to be identified

Interpolatory algorithms

- They provide identified models that belong to the FSS:

$$\phi^I\left(K,\mathcal{B}_e,F_N,y^N\right)=\hat{M}^I\in FSS$$

- They are nonlinear and “tuned” algorithms
- They are able to approximately interpolate the experimental data, taking explicitly into account the available a priori information
- In general, they follow a two-step procedure:
  - step 1: validation of the a priori information
  - step 2: construction of a model $\hat{M}^I\in FSS$ through nonlinear interpolation techniques

Validation problem:

given the a priori information and the experimental data, is the FSS empty or not?

Result #1 (Nevanlinna-Pick interpolation, for frequency-domain measurements)

The a priori information

- on the system: $S^o \in K^{(1)}_{\rho,M}$, with $\rho > 1$ and $M \geq 0$
- on the noise: $e^N \in \mathcal{B}_{e,\infty}$, with $\varepsilon \geq 0$ and $A = W_e = I_{N\times N}$

are consistent with the experimental data vector $y^N \in \mathbb{C}^N$ (i.e., $FSS \neq \emptyset$)

if and only if there exists a vector $\eta^N \in \mathcal{B}_{e,\infty}$ such that the Pick matrix:

$$P_N := \left[\frac{1 - \left(1/M^2\right)\left(y_i - \eta_i\right)\left(\overline{y}_j - \overline{\eta}_j\right)}{1 - \left(1-\rho^2\right) z_i \overline{z}_j}\right]$$

is nonnegative definite, i.e., $P_N \geq 0$.

Such a consistency problem can be rewritten in LMI form and can be solved if and only if there exists a vector $\eta^N \in \mathcal{B}_{e,\infty}$ such that:

$$A\left(\eta^N\right) := \begin{bmatrix} -Q & -\frac{1}{M}\left(D_y - D_\eta\right) \\ -\frac{1}{M}\left(D_y - D_\eta\right)^H & -Q^{-1} \end{bmatrix} \leq 0$$

with $D_y := \mathrm{diag}\left(y_0, y_1, \cdots, y_{N-1}\right)$, $D_\eta := \mathrm{diag}\left(\eta_0, \eta_1, \cdots, \eta_{N-1}\right)$ e $Q := \left[\frac{1}{1-(1-\rho^2)z_i\overline{z}_j}\right]$

[Chen-Nett-Fan, *TAC* 1995]

Interpolatory algorithm for frequency-domain data:

- Step 1: find the solution $\eta^N = \eta^*$ to the consistency problem provided by the previous LMI, such that

$$P_N^* = \left[\frac{1 - \left(1/M^2\right)\left(y_i - \eta_i^*\right)\left(\overline{y}_j - \overline{\eta}_j^*\right)}{1 - \left(1 - \rho^2\right) z_i \overline{z}_j} \right] \geq 0$$

- Step 2: by applying the Pick algorithm, build a function that approximately interpolates $\tilde{y}^N := y^N + \eta^N$ and use it as identified model

The interpolatory algorithm is very cumbersome from the computational viewpoint when the number of data is huge

The Pick matrix may easily result to be ill-conditioned

Result #2 (Carathéodory-Fejér interpolation, for time-domain measurements)

The a priori information

- on the system: $S^o \in K^{(1)}_{\rho,M}$, with $\rho > 1$ and $M \geq 0$
- on the noise: $e^N \in \mathcal{B}_{e,\infty}$, with $\varepsilon \geq 0$ and $A = W_e = I_{N \times N}$

are consistent with the experimental data vector $y^N \in \mathbb{R}^N$ (i.e., $FSS \neq \emptyset$)
if and only if there exists a vector $\eta^N \in \mathcal{B}_{e,\infty}$ such that:

$$(T_y - T_\eta)^T D_\rho^2 (T_y - T_\eta) \leq M^2 T_u^T D_\rho^2 T_u$$

with $D_\rho := \mathrm{diag}\left(1, \rho, \cdots, \rho^{N-1}\right)$ and T_u, T_y e T_η given by the lower triangular Toeplitz matrices associated to u^N, y^N and η^N, respectively

Such a consistency problem can be rewritten in LMI form and can be solved if and only if there exists a vector $\eta^N \in \mathcal{B}_{e,\infty}$ such that:

$$A\left(\eta^N\right) := \begin{bmatrix} -T_u^T D_\rho^2 T_u & -\left(T_y - T_\eta\right)^T \\ -\left(T_y - T_\eta\right) & -M^2 D_\rho^{-2} \end{bmatrix} \leq 0$$

Starting from the solution η^* of this LMI, an interpolatory model is obtained through the Carathéodory-Fejér procedure

[Chen-Nett, *TAC* 1995]

Interpolatory algorithm for time-domain data

- Step 1: find the solution $\eta^N = \eta^* \in \mathcal{B}_{e,\infty}$ to the consistency problem provided by the previous LMI.
  If $A\left(\eta^N\right) > 0$, stop; otherwise, compute the matrix

$$H := \frac{1}{M} D_\rho \left(T_y - T_{\eta^*}\right) T_u^{-1} D_\rho^{-1} = \begin{bmatrix} h_0 & 0 & \cdots & 0 \\ h_1 & h_0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h_{N-1} & h_{N-2} & \cdots & h_0 \end{bmatrix}$$

- Step 2: choose an integer $n \leq N$ and take as identified model:

$$\hat{M}(z) := M \sum_{k=0}^{n-1} \frac{1 - \rho^{2(k-n)}}{\rho^k} h_k \, z^k$$

Properties of the interpolatory algorithms

- They are convergent, but not robustly convergent
- They are 2-optimal and for this reason are called “almost optimal”.

Fundamental limitations of the interpolatory algorithms

- The obtained models are of high order
- Only upper bounds on the identification error are known: $\left\| E\left(\phi^I\right) \right\| < B$
- Despite the fact that the interpolatory algorithms are 2-optimal, the previous upper bounds on the identification error can be not tight and in particular it may turns out that $B/r \gg 2$
- Reduced order models are obtained, whose optimality level is however unknown

"Nearly-optimal" interpolatory algorithm for time and/or frequency domain data

Given a system belonging to the FSS, a corresponding feasible transfer function exists, represented on the polar plane (or $\mathbb{R}^2$) for any $\omega \in [0, 2\pi] \Rightarrow$

- **Value set** $V(\omega)$: given $\omega \in [0, 2\pi]$,

$$V(\omega) := \left\{ v \in \mathbb{R}^2 : v = \begin{bmatrix} \Re e\,(S(\omega)) \\ \Im m\,(S(\omega)) \end{bmatrix},\ S \in FSS \right\} \subset \mathbb{R}^2$$

- Properties:
  - $V(\omega)$ is a polytope, convex and bounded $\forall \omega$
  - *Chebicheff center* $c_2[V(\omega)]$ in Euclidean norm:

$$c_2[V(\omega)] = \arg \inf_{s \in \mathbb{R}^2} \sup_{v \in V(\omega)} ||s - v||_2$$

  - *Chebicheff radius* $r_2[V(\omega)]$ in Euclidean norm:

$$r_2[V(\omega)] = \inf_{s \in \mathbb{R}^2} \sup_{v \in V(\omega)} ||s - v||_2$$

Problems:

- the exact description of $V(\omega)$ cannot be performed
- the exact computations of $c_2[V(\omega)]$ and $r_2[V(\omega)]$ cannot be performed

$$\Downarrow$$

convergent approximations of $V(\omega)$ are computed, based on the construction of two polytopes in $\mathbb{R}^2$ with m vertices:

- the polytope $\overline{VO}_m^{\nu}(\omega)$, with vertices $\overline{v_k}(\omega)$, $k=1,\ldots,m$, "contains" $V(\omega)$
- the polytope $\underline{VI}_m^{\nu}(\omega)$, with vertices $\underline{v_k}(\omega)$, $k=1,\ldots,m$, is contained in $V(\omega)$

these polytopes are obtained through linear programming techniques

Result #1: **convergence of the approximations of** $V(\omega)$

[Milanese-Taragna, *TAC* 2002]

Let $\overline{\overline{VO}}_m^{\nu}(\omega) := \left\{ s \in \Re^2 : s = \overline{s} + \tilde{s},\ \overline{s} \in \overline{VO}_m^{\nu}(\omega),\ \|\tilde{s}\|_2 \leq \delta_\nu \right\}, \delta_\nu := \dfrac{L\rho^\nu}{1-\rho}.$

Fixed $\nu \in \mathbb{Z}$ and $m \in \mathbb{Z}$,

$$\underline{VI}_m^{\nu}(\omega) \subseteq V(\omega) \subseteq \overline{\overline{VO}}_m^{\nu}(\omega)$$

$$r_2\left[\underline{VI}_m^{\nu}(\omega)\right] \leq r_2\left[V(\omega)\right] \leq r_2\left[\overline{\overline{VO}}_m^{\nu}(\omega)\right] \leq r_2\left[\overline{VO}_m^{\nu}(\omega)\right] + \delta_\nu$$

$$\lim_{\nu,m\to\infty} r_2\left[\underline{VI}_m^{\nu}(\omega)\right] = \lim_{\nu,m\to\infty} \left\{ r_2\left[\overline{VO}_m^{\nu}(\omega)\right] + \delta_\nu \right\} = r_2\left[V(\omega)\right]$$

$$\lim_{\nu,m\to\infty} c_2\left[\underline{VI}_m^{\nu}(\omega)\right] = \lim_{\nu,m\to\infty} c_2\left[\overline{VO}_m^{\nu}(\omega)\right] = c_2\left[V(\omega)\right]$$

Result 2: computation of the identification error $E(\hat{\mathcal{M}})$

[Milanese-Taragna, *TAC* 2002]

Given a model $\hat{M}$, fixed $\nu \in \mathbb{Z}$ and $m \in \mathbb{Z}$:

$$\underline{E}_m^{\nu}(\hat{\mathcal{M}}) \leq E(\hat{\mathcal{M}}) \leq \overline{E}_m^{\nu}(\hat{\mathcal{M}})$$

$$\lim_{\nu,m\to\infty} \overline{E}_m^{\nu}(\hat{\mathcal{M}}) = \lim_{\nu,m\to\infty} \underline{E}_m^{\nu}(\hat{\mathcal{M}}) = E(\hat{\mathcal{M}})$$

$$\underline{E}_m^{\nu}(\hat{\mathcal{M}}) = \sup_{0\leq\omega\leq 2\pi} \max_{k=1,\ldots,m} \left\| v_k\left(\underline{VI}_m^{\nu}(\omega)\right) - \hat{M}(\omega) \right\|_2$$

$$\overline{E}_m^{\nu}(\hat{\mathcal{M}}) = \sup_{0\leq\omega\leq 2\pi} \max_{k=1,\ldots,m} \left\| v_k(\overline{VO}_m^{\nu}(\omega)) - \hat{M}(\omega) \right\|_2 + \delta_{\nu}$$

Result 3: computation of the optimality level $\alpha(\hat{\mathcal{M}})$

[Milanese-Taragna, *TAC* 2002]

Given a model set $\hat{\mathcal{M}}$, fixed $\nu \in \mathbb{Z}$ and $m \in \mathbb{Z}$:

$$\underline{\alpha}_m^\nu(\hat{\mathcal{M}}) \leq \alpha(\hat{\mathcal{M}}) \leq \overline{\alpha}_m^\nu(\hat{\mathcal{M}})$$

$$\lim_{\nu,m\to\infty} \overline{\alpha}_m^\nu(\hat{\mathcal{M}}) = \lim_{\nu,m\to\infty} \underline{\alpha}_m^\nu(\hat{\mathcal{M}}) = \alpha(\hat{\mathcal{M}}).$$

$$\underline{\alpha}_m^\nu(\hat{\mathcal{M}}) = \max\left\{1,\ \underline{E}_m^\nu(\hat{\mathcal{M}}) \Big/ \left[\sup_{0\leq\omega\leq 2\pi} r_2[\overline{VO}_m^\nu(\omega)] + \delta_\nu\right]\right\}$$

$$\overline{\alpha}_m^\nu(\hat{\mathcal{M}}) = \overline{E}_m^\nu(\hat{\mathcal{M}}) \Big/ \sup_{0\leq\omega\leq 2\pi} r_2\left[\underline{VI}_m^\nu(\omega)\right].$$

“Nearly optimal” interpolatory algorithm

Starting from the knowledge of centers and radii of the value sets $V(\omega)$ for a frequency set, a “nearly optimal” model set $\hat{\mathcal{M}}^{no}$ is computed that approximates the central model set $\hat{\mathcal{M}}^{c}$

- the nominal model $\hat{M}^{no}$ is solution of a linear programming problem
- the algorithm $\phi^{no}\left(y^N\right) = \hat{M}^{no}$ is $\sqrt{2}$-optimal
- the identified nominal model is typically of high order

$$\Downarrow$$

model sets with reduced order nominal models can be derived through order reduction of the “nearly optimal” model set

Reduced order model sets

Model sets with reduced order nominal models can be derived through order reduction of the “nearly optimal” model set

$\Downarrow$

model set order reduction with minimal inclusion and α-optimality computation

$\Downarrow$

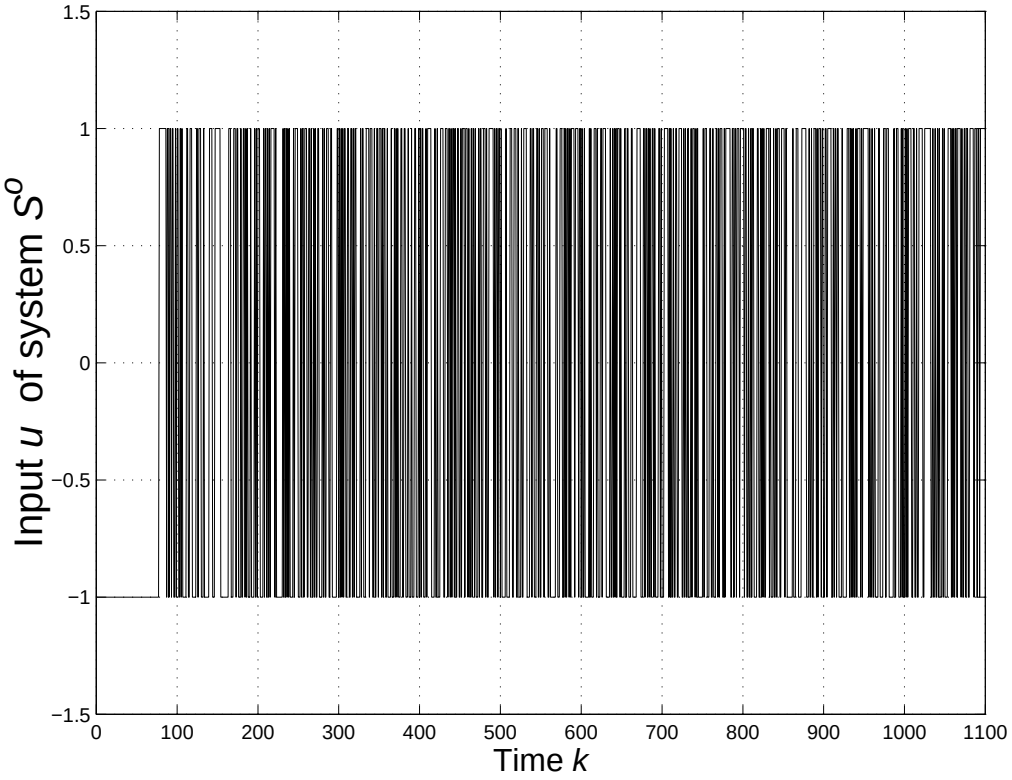
choice of the model set order as trade-off between α-optimality level and model set complexity

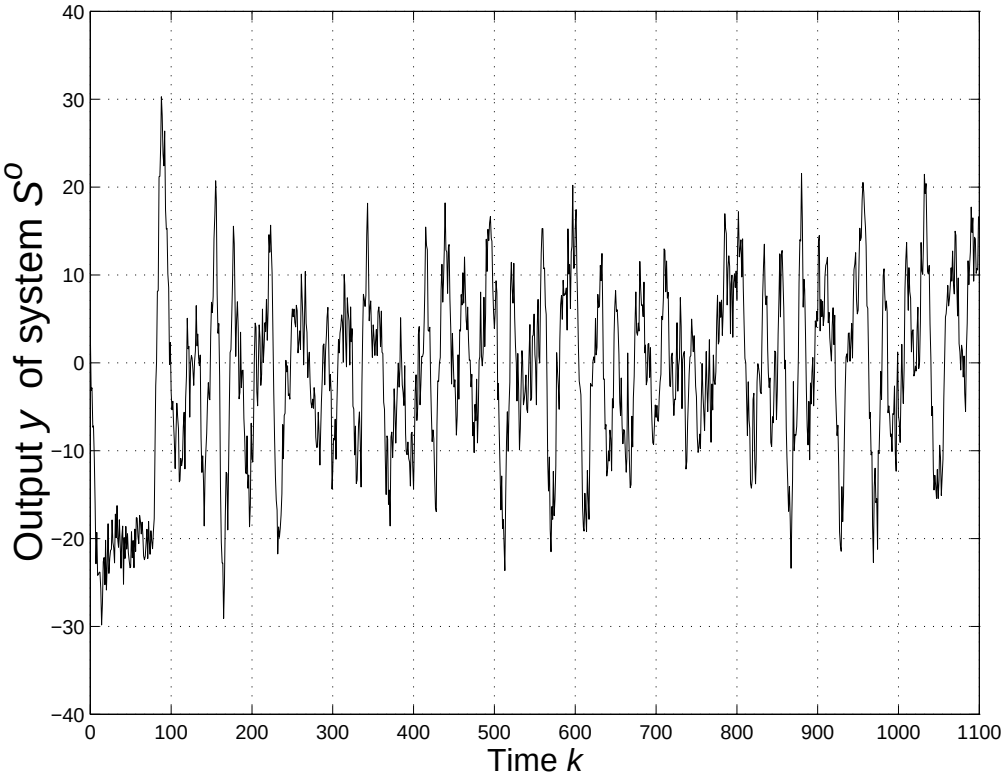
Numerical example

- $N = 1100$ samples of the response of the system

$$S^o(z) = \frac{z + 0.5z^2}{1 - 2.2z + 2.42z^2 - 1.87z^3 + 0.7225z^4}$$

to a PRBS (pseudo random binary signal) of unitary amplitude as input

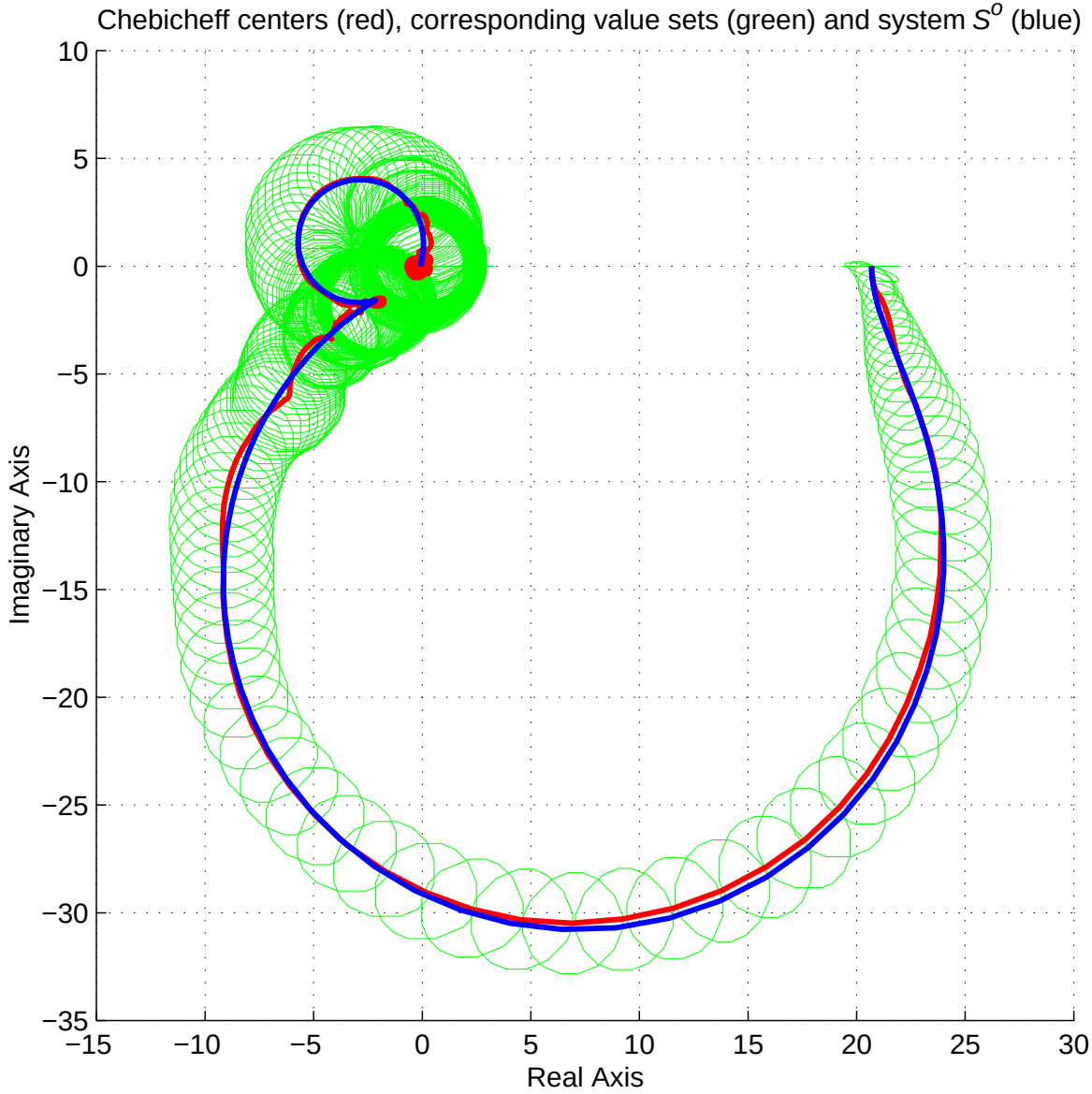




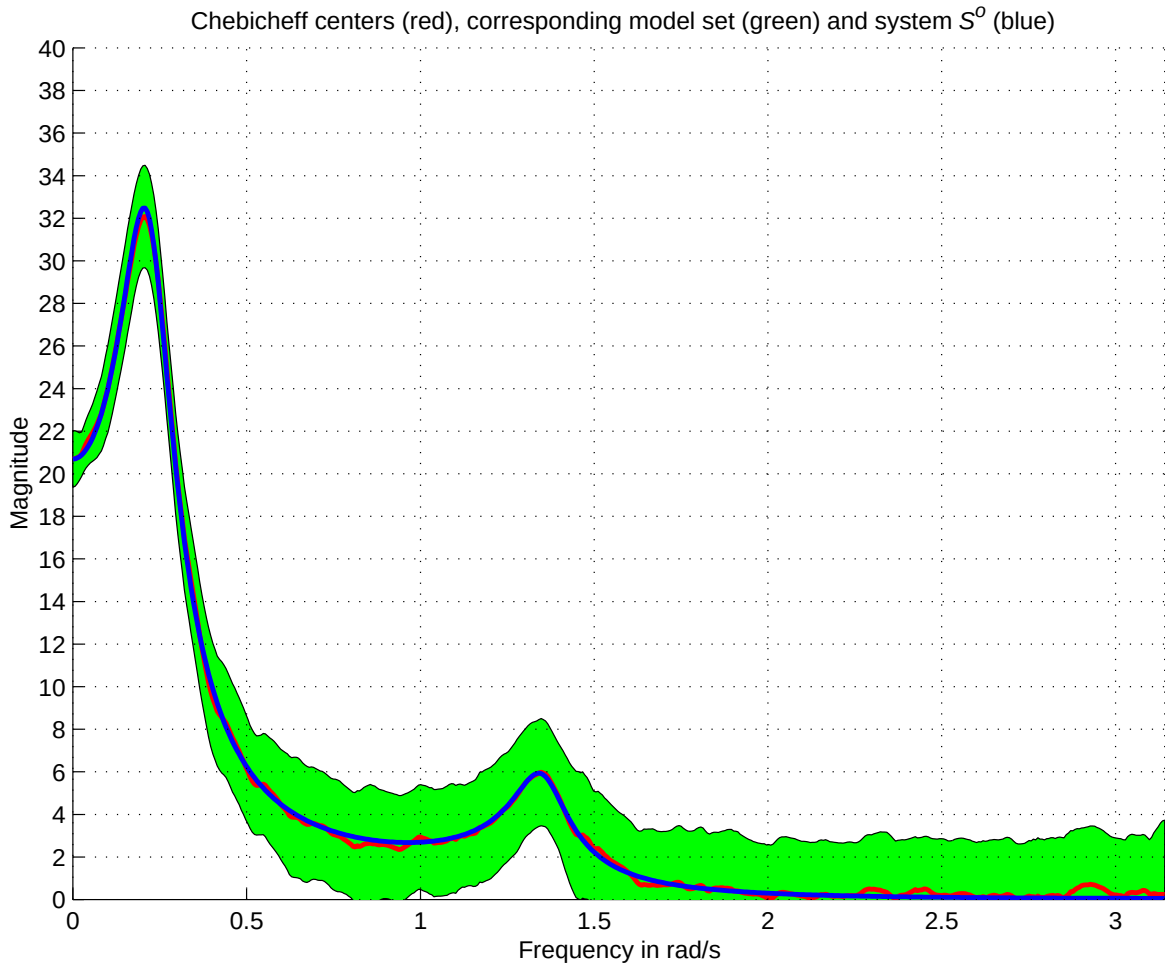
- "A priori" information on the system: $S^o \in K^{(2)}_{\rho,M}$, with: $M = 6$, $\rho = 1/0.93$
- "A priori" information on the noise: $e^N \in \mathcal{B}_{e,\infty}$, with $\varepsilon = 4$, $A = W_e = I_{N\times N}$

  (M, ρ, ε) validated according to [Milanese-Taragna: *SYSID* '00]

- Value set $V(\omega)$: computed in 500 values of ω equispaced in $[0, \pi]$, with $\nu = 150$, $m = 16$

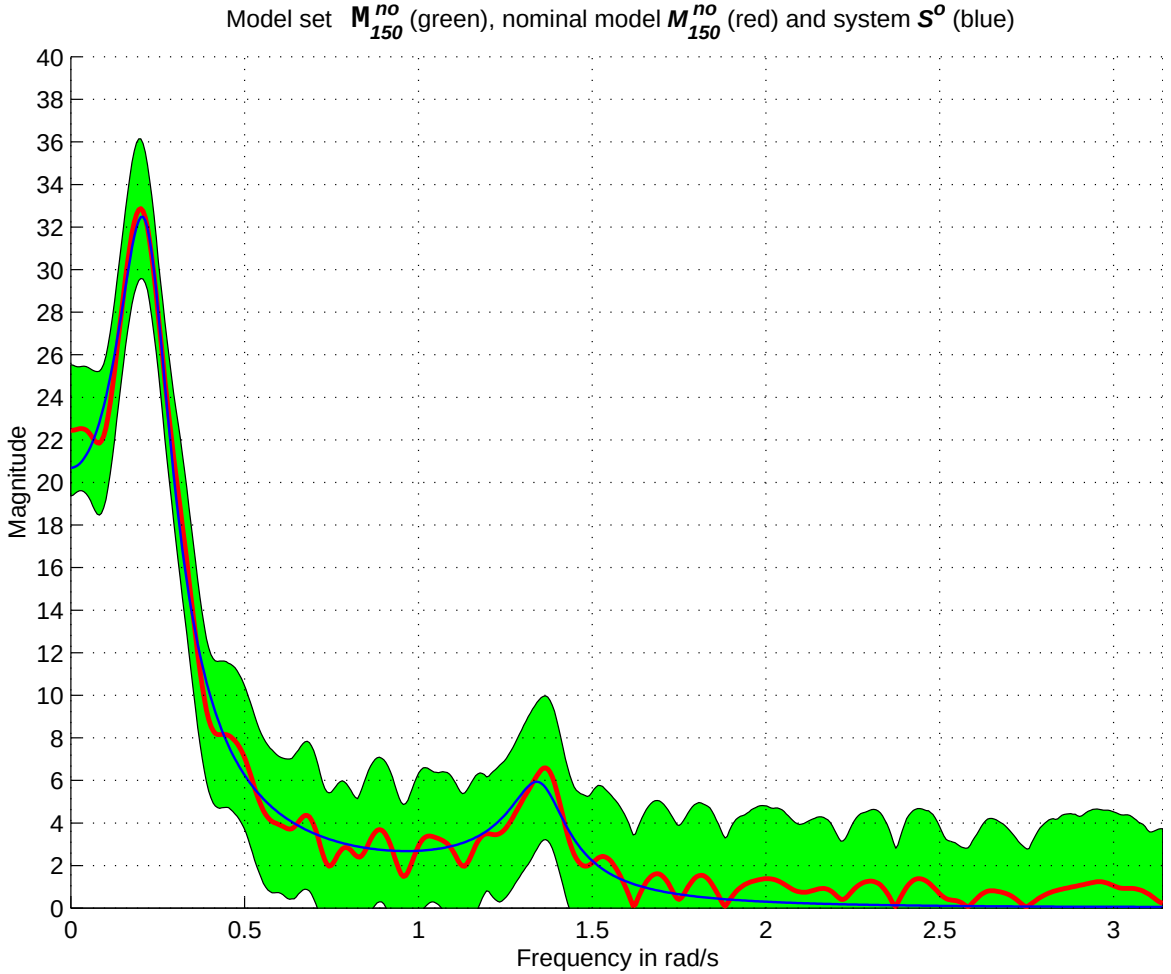


- "Central" model set obtained by considering:
  - as nominal model, the Chebicheff centers of the value sets
  - as perturbation, the Chebicheff radii of the value sets

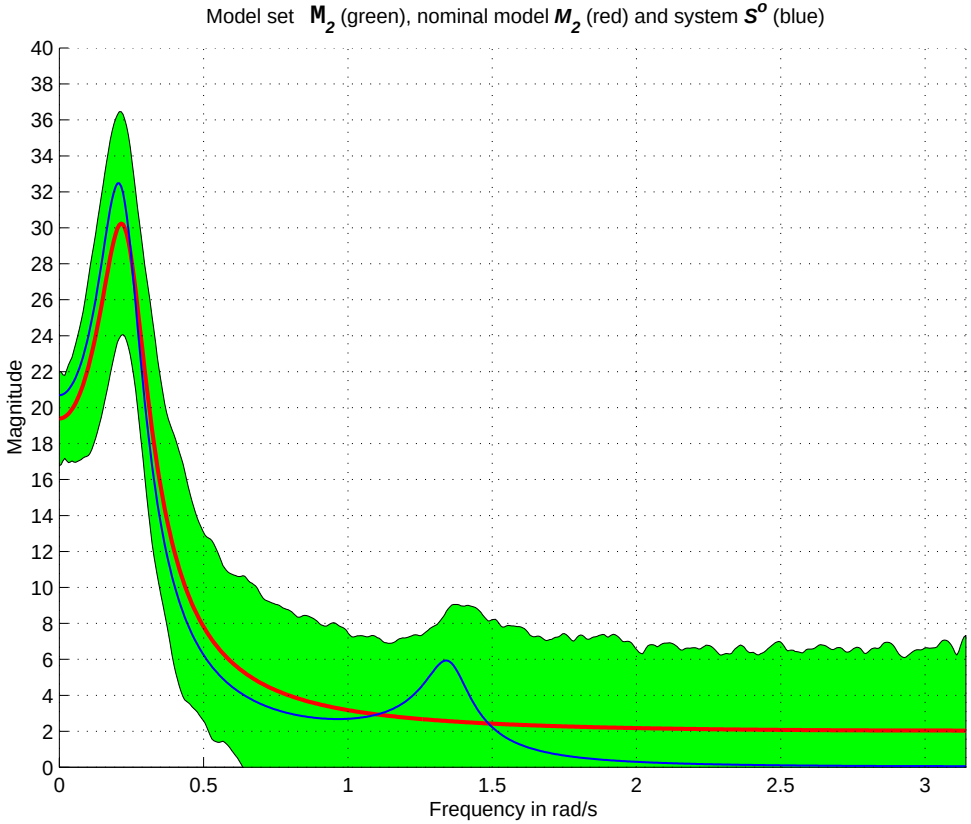


- Nominal models and corresponding model sets:

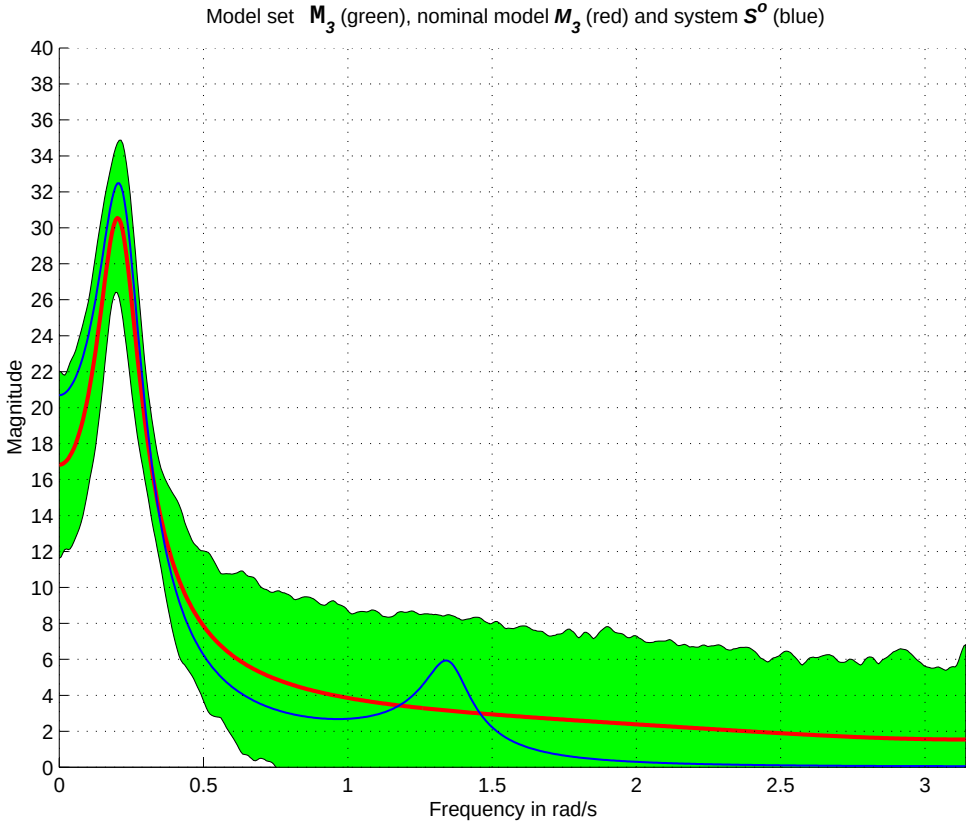
  1) FIR $\hat{M}_{150}^{no}$ of order 150 obtained through the "nearly optimal" algorithm using the Chebicheff radii and centers of the value sets $V(\omega)$ computed in 500 values of ω equispaced in $[0, \pi]$



2) approximations of $\hat{M}_{150}^{no}$ of order $2 \div 5$ ($\hat{M}_2 \div \hat{M}_5$) obtained through the reduction methods (optimal Hankel norm approximation, balanced truncation) available under MATLAB

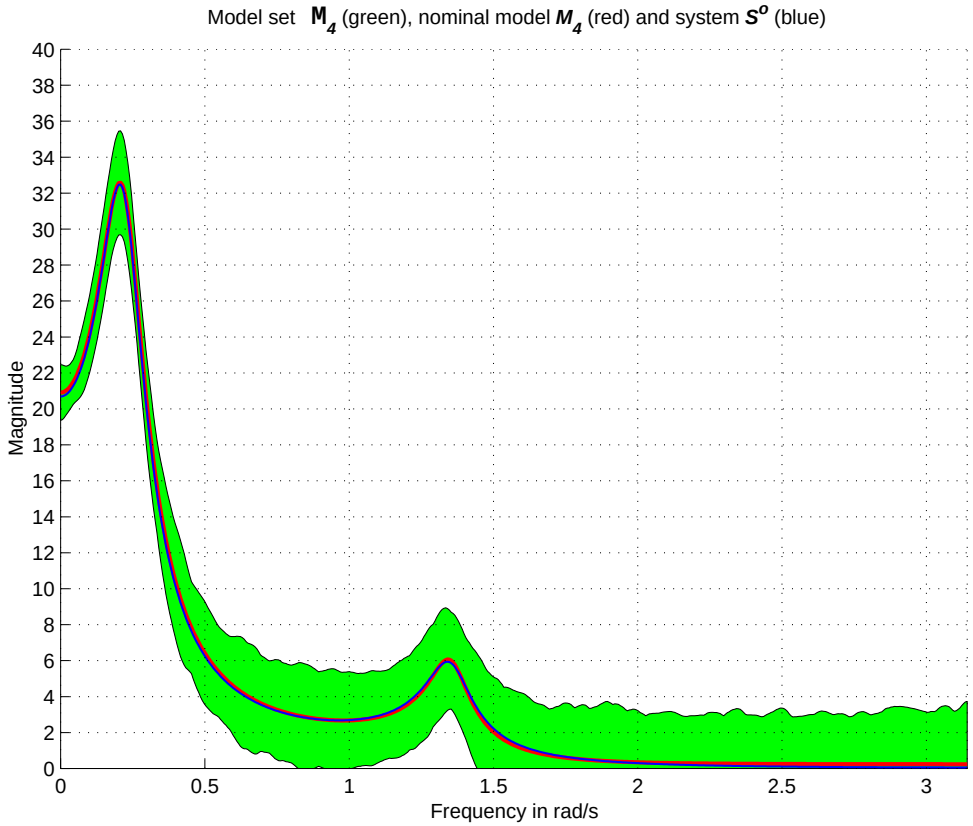


model set $\hat{\mathcal{M}}_2$

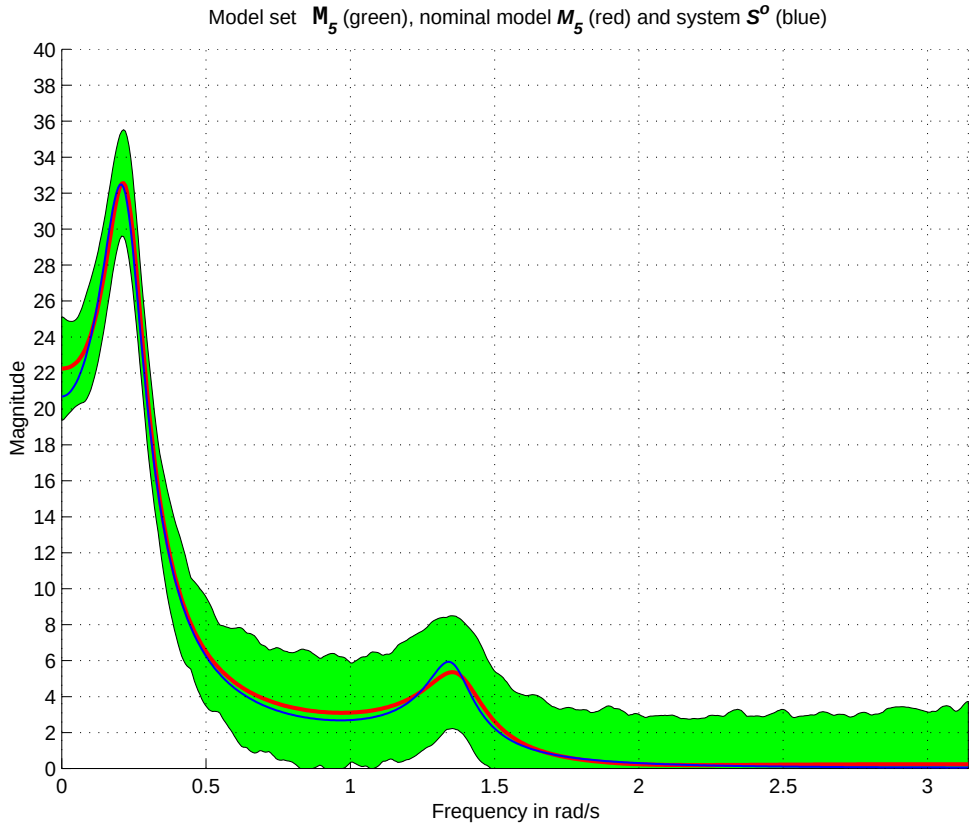


model set $\hat{\mathcal{M}}_3$

2) approximations of $\hat{M}_{150}^{no}$ of order $2 \div 5$ ($\hat{M}_2 \div \hat{M}_5$) obtained through the reduction methods (optimal Hankel norm approximation, balanced truncation) available under MATLAB



model set $\hat{\mathcal{M}}_4$



model set $\hat{\mathcal{M}}_5$

Identification errors and
optimality level of the identified model sets

$\hat{\mathcal{M}}$	$\underline{E}(\hat{\mathcal{M}})$	$\overline{E}(\hat{\mathcal{M}})$	$\underline{\alpha}_m^\nu(\hat{\mathcal{M}})$	$\overline{\alpha}_m^\nu(\hat{\mathcal{M}})$
$\hat{\mathcal{M}}_{150}^{no}$	3.57	3.61	1.00	1.04
$\hat{\mathcal{M}}_2$	6.49	6.52	1.77	1.87
$\hat{\mathcal{M}}_3$	5.31	5.34	1.45	1.53
$\hat{\mathcal{M}}_4$	3.48	3.49	1.00	1.01
$\hat{\mathcal{M}}_5$	3.48	3.49	1.00	1.01

Bounds on the error $E(\hat{\mathcal{M}})$ proposed in [Chen-Nett, *TAC* 95]:

$$\begin{cases} E(\hat{\mathcal{M}}) < 167.43 & \text{(Th. 4.2)} \\ E(\hat{\mathcal{M}}) < 16.99 & \text{(Th. 4.3)} \end{cases}$$

References

J. Chen and G. Gu, *Control-Oriented System Identification: An H_∞ Approach*. New York: John Wiley & Sons, Inc., 2000.

J. Chen, C. N. Nett and M. K. H. Fan, "Worst case system identification in H_∞: validation of *a priori* information, essentially optimal algorithms, and error bounds," *IEEE Transactions on Automatic Control*, vol. AC-40, no. 7, pp. 1260–1265, 1995.

A. Garulli, A. Tesi and A. Vicino, eds., *Robustness in Identification and Control*, vol. 245 of *Lecture Notes in Control and Information Sciences*. Godalming, UK: Springer-Verlag, 1999.

L. Giarré and M. Milanese, "SM identification of approximating models for H_∞ robust control," *International Journal of Robust and Nonlinear Control*, vol. 9, pp. 319–332, 1999.

L. Giarré, M. Milanese and M. Taragna, "H_∞ identification and model quality evaluation," *IEEE Transactions on Automatic Control*, vol. AC-42, no. 2, pp. 188–199, 1997.

M. Glaum, L. Lin and G. Zames, "Optimal H_∞ approximation by systems of prescribed order using frequency response data," in *Proc. of the 35th IEEE Conference on Decision and Control*, (Kobe, Japan), pp. 2318–2321, 1996.

G. Gu and P. P. Khargonekar, "A class of algorithms for identification in H_∞," *Automatica*, vol. 28, no. 2, pp. 299–312, 1992.

G. Gu, D. Xiong and K. Zhou, "Identification in H_∞ using Pick's interpolation," *Systems & Control Letters*, vol. 20, pp. 263–272, 1993.

A. J. Helmicki, C. A. Jacobson and C. N. Nett, "Control oriented system identification: a worst-case/deterministic approach in H_∞," *IEEE Transactions on Automatic Control*, vol. AC-36, no. 10, pp. 1163–1176, 1991.

M. Milanese, J. Norton, H. Piet-Lahanier and É. Walter, eds., *Bounding Approaches to System Identification*. New York: Plenum Press, 1996.

M. Milanese, C. Novara and M. Taragna, ""Fast" Set Membership H_∞ identification from frequency-domain data," in *Proc. of European Control Conference ECC 2001*, (Porto, Portugal), pp. 1698–1703, 2001.

M. Milanese and M. Taragna, "Optimality, approximation, and complexity in Set Membership H_∞ identification," *IEEE Transactions on Automatic Control*, vol. AC-47, no. 10, pp. 1682-1690, 2002.

M. Milanese and M. Taragna, "Nearly optimal model sets in H_∞ identification," in *Proc. of the European Control Conference 2001*, (Porto, Portugal), pp. 1704–1709, 2001.

M. Milanese and A. Vicino, "Optimal estimation theory for dynamic systems with set membership uncertainty: an overview," *Automatica*, vol. 27, no. 6, pp. 997–1009, 1991.

B. Ninness and G. C. Goodwin, "Estimation of model quality," *Automatica*, vol. 31, no. 12, pp. 1771–1797, 1995.

J. R. Partington, *Interpolation, Identification, and Sampling*, vol. 17 of *London Mathematical Society Monographs New Series*. New York: Clarendon Press - Oxford, 1997.

T. Zhou and H. Kimura, "Structure of model uncertainty for a weakly corrupted plant," *IEEE Transactions on Automatic Control*, vol. AC-40, no. 4, pp. 639–655, 1995.