Frequency-Domain Identification of Discrete-Time Systems using Sum-of-Rational Optimization

Mohamed Abdalmoaty, **Jared Miller**, Mingzhou Yin, and Roy Smith

MTNS 2024, Cambridge, UK

Main Ideas

Frequency-domain sysid is nonconvex

Can be cast as sum-of-rationals problem

Get global bounds on suboptimality using moment-SOS

The Sysid Problem

System Identification Problem

Want to recover true SISO model $(a, b) \in \mathbb{R}^{2n}$

$$G(z; a, b) = \frac{B(z; b)}{1 + A(z; a)} = \frac{\sum_{k=1}^{n} b_k z^{-k}}{1 + \sum_{k=1}^{n} a_k z^{-k}}$$

But we only have noisy data $\mathcal{D} = \{(\omega_f, G_f),\ f = 1..N_f\}$

$$G_f = G_\circ(e^{j\omega_f}) + \eta_f$$

Optimize to minimize residuals $\{\eta_f\}$

Sysid Estimator

Least squares estimator for $\theta = (a, b)$ given weights $\{W_f\}$:

$$\theta^\star := \operatorname*{arg\,min}_{\theta \in \mathbb{K}} \mathcal{J}_{\mathcal{D}}(\theta)$$

$$\mathcal{J}_{\mathcal{D}}(\theta) = \sum_{f=1}^{N_f} \left| W_f \big(G_f - G(e^{j\omega_f}; \theta) \big) \right|^2$$

W_f	user-defined weights
$\mathbb{K} \subset \mathbb{R}^{2n}$	compact feasibility set

Challenge: global optimization over stable rational models!

Prior work

- Classical methods: mainly *local search techniques*
    + several heuristic initialization methods

  [Söderström & P. Stoica 1989], [Ljung 1999], [Pentilon & Schoukens 2012], ...

- Numerical linear algebra method

  e.g. [Agudelo et al. 2021], [Lagauw et al. 2023], ...

- Method based on the Moments-SOS hierarchy;

  e.g. [Rodrigues et al. 2019], [Rodrigues et al. 2020], [Vuillemin 2014]

Our work

Main contributions of this work:

- Use sum-of-rational structure
- infinite-dimensional LP, solved by hierarchy of *finite* SDPs.
- convergence in objective with increasing degree
- stability enforced using Polynomial Matrix Inequality (PMI)
- *Global optimally certified a-posteriori*: (rank)

Sum-of-Rationals Reformulation

Sum-of-Rational Formulation: cost function

The SysID problem: $P^\star = \min_{\theta \in \mathbb{K}} \mathcal{J}_\mathcal{D}(\theta)$

$$\mathcal{J}_\mathcal{D}(\theta) = \sum_{f=1}^{N_f} \left| W_f \big(G_f - G(e^{j\omega_f}; \theta) \big) \right|^2, \qquad G(z;\theta) = \frac{B(z;b)}{1 + A(z;a)}$$

Sum-of-Rational Formulation: cost function

$$\boxed{\text{The SysID problem:} \quad P^\star = \min_{\theta \in \mathbb{K}} \mathcal{J}_{\mathcal{D}}(\theta)}$$

$$\mathcal{J}_{\mathcal{D}}(\theta) = \sum_{f=1}^{N_f} \left| W_f\big(G_f - G(e^{j\omega_f};\theta)\big)\right|^2, \qquad G(z;\theta) = \frac{B(z;b)}{1 + A(z;a)}$$

Expand out each frequency data-point:

$$\begin{aligned} W_f(G_f - G(e^{j\omega_f};\theta)) &= W_f\left(G_f - \frac{B(e^{-j\omega_f};b)}{1 + A(e^{-j\omega_f};a)}\right) \\ &= \frac{W_f\left(G_f(1 + A(e^{-j\omega_f};a)) - B(e^{-j\omega_f};b)\right)}{1 + A(e^{-j\omega_f};a)} \end{aligned}$$

Sum-of-Rational Formulation: cost function

Define polynomials p_f, q_f for each datapoint f

$$\frac{p_f(\theta)}{q_f(a)} := \frac{\left|W_f\left(G_f(1 + A(e^{-j\omega_f}; a)) - B(e^{-j\omega_f}; b)\right)\right|^2}{\left|1 + A(e^{-j\omega_f}; a)\right|^2}$$

$p_f(\theta)$ and $q_f(a)$ quadratic polynomials in θ and a respectively.

$$\boxed{\sum_{f=1}^{N_f} \left|W_f\left(G_f - G(e^{j\omega_f}; \theta)\right)\right|^2 = \min_{\theta \in \mathbb{K}} \sum_{f=1}^{N} \frac{p_f(\theta)}{q_f(\theta)}}$$

Constraint term $\mathbb{K} = \mathbb{K}_0 \cap \mathcal{S}_\delta$

- $\mathbb{K}_0$: prior information on (a, b) range
- $\mathcal{S}_\delta$: Enforce that (a, b) is a stable model

Sum-of-Rational Formulation: feasibility set

Constraint term $\mathbb{K} = \mathbb{K}_0 \cap \mathcal{S}_\delta$

- $\mathbb{K}_0$: prior information on (a, b) range
- $\mathcal{S}_\delta$: Enforce that (a, b) is a stable model

$$\mathcal{S}_\delta = \left\{\theta \in \mathbb{R}^{2n} \ : \ \Xi(a) \succeq \delta I\right\}$$ Polynomial Matrix Inequality

with Hermite matrix $\Xi(a) := \Theta(a)^T\Theta(a) - \tilde{\Theta}(a)^T\tilde{\Theta}(a)$,

$$\Theta(a) = \begin{bmatrix} 1 & a_1 & a_2 & \dots \\ 0 & 1 & a_1 & \dots \\ 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \quad \tilde{\Theta}(a) = \begin{bmatrix} a_n & a_{n-1} & a_{n-2} & \dots \\ 0 & a_n & a_{n-1} & \dots \\ 0 & 0 & a_n & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad \in \mathbb{R}^{n \times n}$$

[Barnet 1983], [Henrion & Lasserre 2006]

Wrapping up the formulation

$$\boxed{\mathcal{J}_{\mathcal{D}} = \min_{\theta \in \mathbb{K}} \sum_{f=1}^{N} \frac{p_f(\theta)}{q_f(\theta)}}$$

Objective from data $\mathcal{D}$, weights W

Constraint set $\mathbb{K}$ from prior info, stability

… now to solve it

Bounds for Global Optimization

Transformation of Optimization

Original continuous optimization problem

$$f^* = \inf_{x \in C} f(x)$$

same objective as infinite-dimensional LP

$$\begin{aligned} f^* = \sup_{\gamma \in \mathbb{R}} \, & \gamma \\ & f(x) - \gamma \geq 0 \qquad \forall x \in C \end{aligned}$$

Hard problem: certification of nonnegativity over C

Ways to Discretize

Infinite-dimensional LP must be discretized for computation

More complexity: more accurate solutions

Method	Increasing Complexity
Gridding (MDP)	# Grid Points
Basis Functions (ADP)	# Functions
Random Sampling	# Samples
Sum-of-Squares	Polynomial Degree
Neural Nets (FOSSIL)	Width and Depth
Your Favorite Method	Some Accuracy Parameter

Runtime usually exponential in dimension, complexity

Sum-of-Squares Tightening

Truncation when f, C have polynomial structure

$$f^* = \inf_{x \in C} f(x)$$

finite-degree d truncation

$$f^*_{\mathrm{d}} = \sup_{\gamma \in \mathbb{R}} \gamma, \qquad f(x) - \gamma \in \Sigma[C]_{\leq \mathrm{2d}}$$

Quadratic Module formed by constraint description

$$C = \{x \mid g_k(x) \geq 0 \; \forall k \in 0..N_c\}, \; g_0(x) = 1$$

$$\Sigma[C] = \left\{ p \in \mathbb{R}[x] \mid p = \sum_{k=0}^{N_c} g_k(x)\sigma_k(x), \; \deg(\sigma_k g_k) \leq 2d, \quad \sigma_k \in \Sigma[x] \right\}$$

Sum-of-Rationals Tightening

SOS only admissible for polynomial structure.

Options for sum-of-rational optimization:

- Add new states
- Clear to common denominators
- Epigraph approach [Jibetean, de Klerk 2003]
- Absolute-Continuity/Sandwich [Bugarin, Henrion, Lasserre 2016]

Sandwich approach

Sum-of-rationals optimization

$$P^* = \min_{\theta \in \mathbb{K}} \sum_{f=1}^{N_f} \frac{p_f(\theta)}{q_f(\theta)}$$

express as nonnegativity

$$= \max_{\gamma \in \mathbb{R}} \gamma, \ \sum_{f=1}^{N_f} \frac{p_f(\theta)}{q_f(\theta)} - \gamma \geq 0 \qquad \forall \theta \in \mathbb{K}$$

Assume all $q_f(\theta) > 0$ over $\mathbb{K}$ ($e^{\pm j\omega_f}$ not a true pole)

Sandwich approach

Introduce new functions $\zeta \in \mathbb{R}[\theta]$:

$$\frac{p_f(\theta)}{q_f(\theta)} \geq \zeta_f(\theta) \implies p_f(\theta) - \zeta_f(\theta) q_f(\theta) \geq 0$$

Substitute ζ into objective

$$\max_{\gamma \in \mathbb{R}} \gamma, \ \sum_{f=1}^{N_f} \zeta_f(\theta) - \gamma \geq 0 \ \forall \theta \in \mathbb{K}$$

Now can truncate into degree-$\leq 2d$ SOS

Recovery of Global Optima

SOS program is **dual** to LMI in moments

$$\left.\begin{aligned} P_d^\star = \min_{y,y^f} \quad & \sum_{f=1}^{N} \mathbb{L}_{y^f}(p_f(\theta)) \\ & y_0 = 1 \\ & \mathbb{L}_{y^f}(\theta^\alpha q_f(a)) = y_\alpha, \quad \forall \alpha \in \mathbb{N}_d^{2n} \\ & \mathbb{M}_d(\mathbb{K}y) \succeq 0, \quad \mathbb{M}_{d-1}(\mathbb{K}y^f) \succeq 0 \quad \forall f \end{aligned}\right\} \quad \text{SDP}(d)$$

Read off candidate solutions from order-1 moments of θ

Global optima if $\mathbb{M}_d(\mathbb{K}y)$, $\mathbb{M}_{d-1}(\mathbb{K}y^f)$ satisfy rank conditions

Computational Complexity and Convergence

Scaling linear in N_f and exponential in d (for fixed n)

Measure	Matrix	Size ($\mathbb{K}_0$)	Size ($\mathbb{K}_s$)	Mult.
μ	$\mathbb{M}_d[\mathbb{K}y]$	$\binom{2n+d}{d}$	$n\binom{2n+d-1}{d-1}$	1
ν_f	$\mathbb{M}_{d+1}[\mathbb{K}y^f]$	$\binom{2n+d+1}{d+1}$	$n\binom{2n+d}{d}$	N_f

Can exploit symmetry, term sparsity to reduce cost further

Convergence Result:

- when $\mathbb{K} = \mathbb{K}_0$, $P^\star_d \leq P^\star_{d+1}$ for all d
- when $\mathbb{K} = \mathbb{K}_s$, $P^\star_d \uparrow P^\star$

Numerical Examples

Numerical Examples

MATLAB / julia

Data sets:

- $\omega_f = \frac{(f-1)\pi}{10}, \qquad f = 1, \ldots, 11$
- $G_f = G_\circ(e^{j\omega_f}) + \eta_f,$
- $\eta_f \sim \begin{cases} \mathcal{N}(0, 0.3^2), & f = 1, 11 \\ \mathcal{CN}(0, 0.3^2), & \text{otherwsie} \end{cases}$

$\equiv$ ETFE with periodic input & transient effects ignored.

True systems with $\forall f : W_f = 1$:

- Case 1: $G_\circ(z) = \frac{2z^{-1} - z^{-3}}{1 - 0.18z^{-1} - 0.134z^{-2} - 0.637z^{-3}}$
- Case 2: 14 randomly generated Schur-stable second-order systems

Numerical Example 1

Stability is enforced: $\mathbb{K}_S = \{\|\theta\|_\infty \leq 2\} \cap \mathcal{S}_{10^{-4}}$

Method	SDP(1)	n4sid	oe	fmincon(n4sid)	tfest
$\mathcal{J}_{\mathcal{D}}(\theta^*)$	0.3173	0.3368	0.3224	0.3173	0.3173

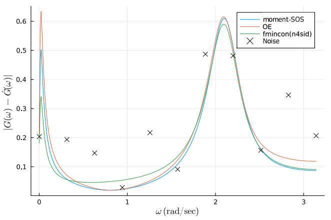
θ^* is extracted from the moment matrix $\mathbb{M}_1[y^\star]$

$$\theta_\circ = [-0.18 \quad -0.134 \quad -0.637 \quad 2 \quad 0 \quad -1]^T$$

$$\theta^* = [-0.1415 \quad -0.2016 \quad -0.6107 \quad 1.978 \quad 0.0835 \quad -1.109]^T$$

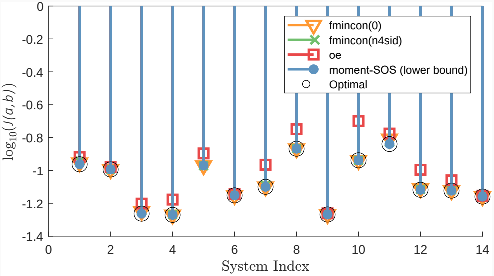
$$0.3173 = \mathcal{J}_{\mathcal{D}}(\theta^*) \geq P^\star \geq P_1^\star = 0.3173$$

$$\implies P^\star = P_1^\star$$



Stability NOT enforced: $\mathbb{K} = \{\|\theta\|_\infty \leq 2\} \leq 1.6$

- In all cases $d = 3$ is used.
- almost all systems have global optimality certificate
- no solution is extracted for System 5



Take-aways

Conclusions

An approach for frequency-domain identification of parameteric models

- tractable SDP + a certificate of global optimally
- stability enforced using Hermite criterion (PMI)

Future work includes:

- application to real systems (e.g. battery health)
- homogenization and improved conditioning
- comparisons with Sparse-POP , and similar methods.