# DISTRIBUTION-FREE SYSTEM IDENTIFICATION: EXACT-, NON-ASYMPTOTIC CONFIDENCE REGIONS

**Balázs Csanád Csáji**

The University of Melbourne, Australia

Hungarian Academy of Sciences, Hungary

joint work with:

Marco Campi & Erik Weyer

16 July 2012, University of Paderborn, Germany

# Structural Overview

PART I. Introduction and Preliminaries
(System Identification, Motivations, General Framework)

PART II. Confidence Regions for FIR Systems
(LS Estimate, Asymptotic and Non-Asymptotic Regions)

PART III. Confidence Regions for ARX Systems
(Difficulties, Alternative Outputs, Perturbed Regressors)

PART IV. Confidence Regions for General Linear Systems
(PEM Estimate, ARMAX, Box-Jenkins Models)

PART V. Summary and Concluding Remarks
(Conceptual Summary, Highlights, Further Directions)

# PART I.

## Introduction and Preliminaries

# Scope and Motivations

- System Identification (SI) aims at building “efficient” mathematical models of dynamical systems from experimental data
- These models are widely used for prediction and control purposes
- Classical problems: parametrized and nonparametrized estimation, experiment design, model reduction, model validation, etc.
- Our scope: selecting “certified” model(s) from a parameterized class
- Standard methods (LS, IV, PEM, etc.) typically provide point estimates
- Strict safety, stability, or quality requirements? $\Rightarrow$ confidence regions
- In practice, our samples are always finite $\Rightarrow$ non-asymptotic methods
- Wide and off-the-shelf applicability $\Rightarrow$ distribution-free approaches

# Parameter Estimation

- We observe a (discrete-time) dynamical system, i.e., stochastic process

$$Y_t \triangleq f_t(\theta^*, \mathbb{Y}_{t-1}, \mathbb{U}_{t-1}, \mathbb{N}_{t-1}),$$

where $Y_t$ — output, $U_t$ — input, $N_t$ — noise at $t$, $\theta^*$ — true parameter

$$\begin{aligned} \mathbb{Y}_{t-1} &\triangleq (Y_{t-1}, Y_{t-2}, \dots), \\ \mathbb{U}_{t-1} &\triangleq (U_{t-1}, U_{t-2}, \dots), \\ \mathbb{N}_{t-1} &\triangleq (N_{t-1}, N_{t-2}, \dots), \end{aligned}$$

- Data: we are given a realization $D_n = (y_n, y_{n-1}, \dots, u_n, u_{n-1}, \dots)$
- Point estimate: find a parameter value $\hat{\theta}_n$ with $\hat{\theta}_n = \arg\min_\theta \varepsilon(\theta, D_n)$
- Confidence region: find a set $\Theta_n^p$ with $\mathbb{P}(\theta^* \in \Theta_n^p) \geq p$, given $p \in [0, 1]$

# Parameter Estimation

- The cost or loss is typically measured by the prediction error,

$$\hat{\varepsilon}(\theta, D_n) \triangleq \sum\nolimits_t \hat{\varepsilon}_t(\theta, D_t) = \sum\nolimits_t w_t\, d(y_t, f_t(\theta, D_t)),$$

where $w_t$ — weight, $d(\cdot,\cdot)$ — metric $\Rightarrow$ often $\|y_t - f_t(\theta, D_t)\|_2^2$

- $\hat{\theta}_{\mathrm{PEM}} \triangleq \arg\min_\theta \hat{\varepsilon}(\theta, D_n)$ is called the prediction error estimate (PEM)

- Our goal: to construct exact, finite sample, quasi distribution-free confidence regions around the PEM estimate, i.e., build a set $\Theta_n^p$ such that

$$\mathbb{P}(\hat{\theta}_{\mathrm{PEM}} \in \Theta_n^p) = 1 \qquad \text{and} \qquad \mathbb{P}(\theta^* \in \Theta_n^p) = p,$$

where the confidence probability $p \in [0,1]$ is user-chosen

- In this talk, we focus on linear time-invariant (LTI) systems

# Sign-Perturbed Sums: Outline

- Sign-Perturbed Sums (SPS): a new finite sample identification algorithm
- It is based on re-randomization of the normal equations
- It builds non-asymptotic confidence regions for dynamical systems
- Under mild assumptions on the noise, e.g., symmetry about zero
- The inclusion of the prediction error (PEM) estimate is always guaranteed
- The true parameters are included with a user-chosen probability
- The confidence probability is exact even for multiple parameters
- SPS is presented for linear systems systems (e.g., FIR, ARX, Box-Jenkins)

# Part II.

## Confidence Regions for FIR Systems

# First-Order FIR Systems

- System: first-order FIR, i.e.

$$Y_t \triangleq b_1^* U_{t-1} + N_t,$$

$Y_t$ — output, $U_t$ — input, $N_t$ — noise (at time $t$)

- Assumptions: $(N_t)_t$ is independent, symmetric $\big[$and continuous$\big]$
  $(N_t)_t \perp\!\!\!\perp (U_t)_t$

  $\big[$No information on stationarity, variances, particular distributions, etc.$\big]$

- Data: $(Y_t)_{t=1}^{n}$ and $(U_t)_{t=0}^{n-1}$, i.e., a finite sample

- Goal: to construct a confidence interval around the LS estimate, the "true" parameter $b_1^*$ should be included with a guaranteed probability

# Least-Squares Estimate

- Linear regression form — $\varphi_t \triangleq [\, U_{t-1} \,],\;\; \theta^* \triangleq [\, b_1^* \,]$

$$Y_t \;=\; \varphi_t^{\mathrm{T}} \theta^* + N_t,$$

- The least-squares (LS) estimate, with $\hat{Y}_t(\theta) \;\triangleq\; b_1 U_{t-1} = \varphi_t^{\mathrm{T}} \theta$

$$\hat{\theta}_{\mathrm{LS}} \triangleq \arg\min_{\theta} \sum_{t=1}^{n} (Y_t - \hat{Y}_t(\theta))^2 = \arg\min_{\theta} \sum_{t=1}^{n} (Y_t - \varphi_t^{\mathrm{T}} \theta)^2,$$

can be found by solving the normal equation

$$\sum_{t=1}^{n} U_{t-1}(Y_t - b_1 U_{t-1}) = \sum_{t=1}^{n} \varphi_t (Y_t - \varphi_t^{\mathrm{T}} \theta) = 0.$$

# Asymptotic Confidence Regions

- The LS estimate, $\hat{\theta}_n$, is asymptotically normal, i.e.

$$\sqrt{n}\,(\hat{\theta}_n - \theta^*) \xrightarrow{d} N(0, \Gamma) \quad \text{as} \quad n \to \infty,$$

$$\Gamma \triangleq \sigma_0^2 \left(\mathbb{E}\,[\,\varphi_0 \varphi_0^{\mathrm{T}}\,]\right)^{-1},$$

where $\sigma_0^2$ is the variance of the noise — stationarity assumption

- An approximate confidence region can be build as

$$\Theta_n^{\mu} \triangleq \left\{\theta \in \mathbb{R}^d : \|\,\hat{\theta}_n - \theta\,\|_{\Phi_n}^2 \leq \mu\,\hat{\sigma}_n^2/n\right\},$$

$$\Phi_n \triangleq \frac{1}{n}\sum_{t=1}^{n} \varphi_t \varphi_t^{\mathrm{T}}, \quad \text{and} \quad \hat{\sigma}_n^2 \triangleq \frac{1}{n}\sum_{t=1}^{n} \hat{N}_t^2(\hat{\theta}_n).$$

- $\mathbb{P}(\theta^* \notin \Theta_n^{\mu})$ can be computed from the $\mu$-level of the $\chi^2$ distribution

# Non-Asymptotic Confidence Regions

- Let us introduce a reference sum (cf. normal equation)

$$S_0(\theta) \triangleq \sum_{t=1}^{n} \varphi_t (Y_t - \varphi_t^{\mathrm{T}} \theta),$$

and further introduce $m-1$ sign-perturbed sums

$$S_i(\theta) \triangleq \sum_{t=1}^{n} \varphi_t \, \alpha_{i,t} (Y_t - \varphi_t^{\mathrm{T}} \theta),$$

for $i = 1, \dots, m-1$, where $\alpha_{i,t}$ ($t = 1, \dots, n$) are i.i.d. random signs, that is $\alpha_{i,t} = \pm 1$ with probability 1/2 each

- Parameter $m$ is user-chosen (positive integer, explained later)
- Observe that the LS estimates is a root of $S_0$, i.e., $S_0(\hat{\theta}_{\mathrm{LS}}) = 0$

# Non-Asymptotic Confidence Regions

- Observe that, in case $\theta = \theta^*$, we have ($i = 1, \dots, m-1$)

$$S_0(\theta^*) = \sum_{t=1}^{n} \varphi_t N_t,$$

$$S_i(\theta^*) = \sum_{t=1}^{n} \varphi_t \, \alpha_{i,t} N_t,$$

- Consider the ordering $S_i(\theta^*)_{i_0} < \cdots < S_i(\theta^*)_{i_{m-1}}$
- Claim: $\boxed{\text{All orderings are equally probable!}}$
- Note: random tie-breaking may be needed
- Hint: $(N_t)_{t=1}^{n}$ and $(\alpha_{i,t} N_t)_{t=1}^{n}$ have the same symmetric distribution
- This is even the case if we order $Z_i(\theta^*) \triangleq S_i^2(\theta^*)$ variables

# Non-Asymptotic Confidence Regions

- $R_0(\theta)$ is the rank of $Z_0(\theta)$ in the ordering $\big[$recall $Z_i(\theta) \triangleq S_i^2(\theta)\big]$

$$Z_i(\theta)_{i_0} < \cdots < Z_i(\theta)_{i_{m-1}}$$

- The confidence region is then

$$\boxed{\Theta_m^q \triangleq \{\, \theta \in \mathbb{R} : R_0(\theta) \geq q \,\},}$$

where $q$ is a user-chosen parameter (nonnegtive integer)

- Claim: $\boxed{\mathbb{P}(\theta^* \in \Theta_m^q) = 1 - q/m}$
- This confidence probability is exact! $\Longrightarrow$ no conservativism introduced
- $q$ and $m$ are user-chosen $\Longrightarrow$ the confidence probability is under control

# Non-Asymptotic Confidence Regions

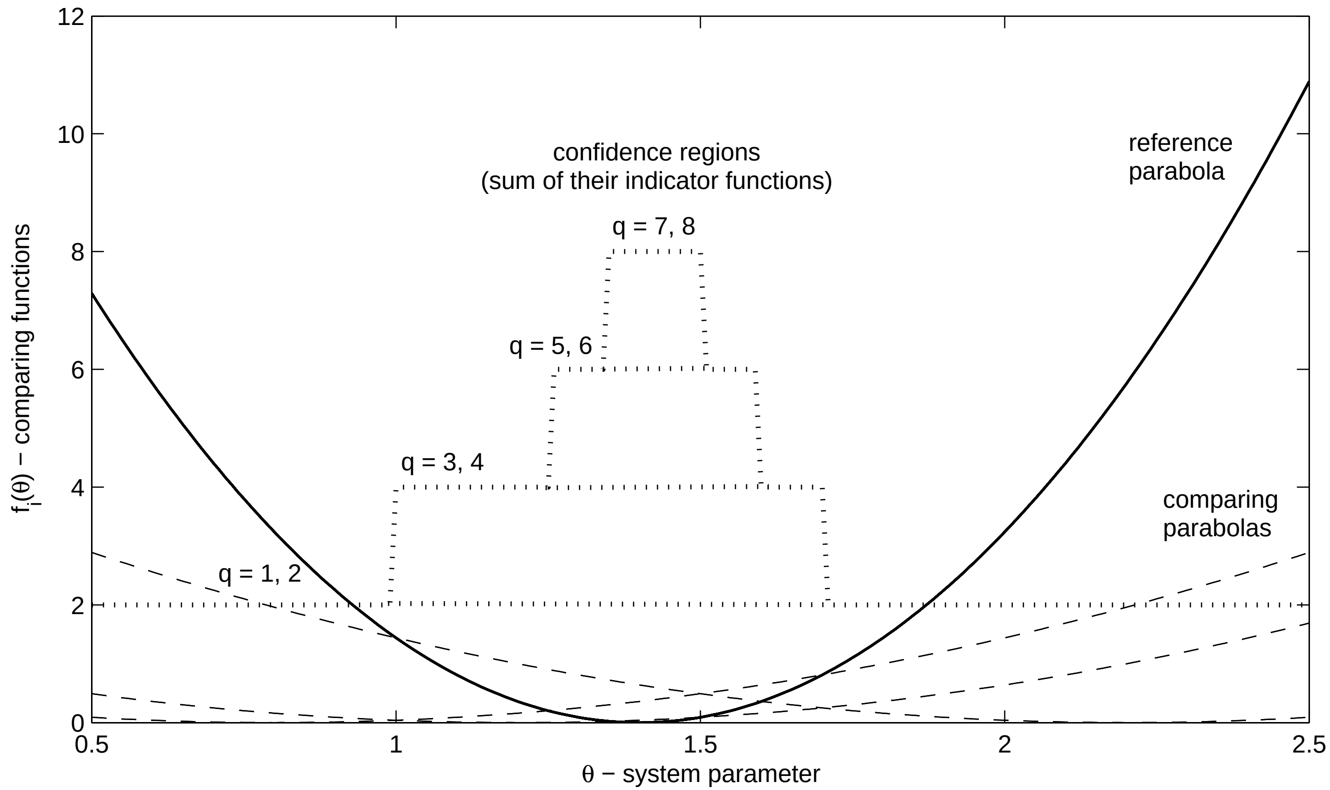
- $S_0^2(\hat{\theta}_{\mathrm{LS}}) = 0 \Longrightarrow \hat{\theta}_{\mathrm{LS}} \in \Theta_m^q$ — the LS estimate is included
- When $\theta \neq \theta^*$ we have that

$$Z_0(\theta) = \left[\sum_{t=1}^{n} (\theta^* - \theta)\, U_{t-1}^2 + U_{t-1}\, N_t \right]^2,$$

$$Z_i(\theta) = \left[\sum_{t=1}^{n} \alpha_{i,t}\, (\theta^* - \theta)\, U_{t-1}^2 + \alpha_{i,t}\, U_{t-1}\, N_t \right]^2.$$

- $\sum_t \alpha_{i,t}(\theta^* - \theta)U_{t-1}^2$ grows slower than $\sum_t(\theta^* - \theta)U_{t-1}^2$ in $Z_0(\theta)$
- As $\|\theta - \theta^*\| \to \infty$ we have $R_0(\theta) \to m - 1$ (i.e., the largest)
- Hence, values different from $\theta^*$ will be eventually excluded from $\Theta_m^q$

# Example Confidence Intervals




- $n = 3$ observations $\Longrightarrow 8$ sign sequences $\Longrightarrow 4$ parabolas
- $x$-axis: value of $\theta$; $y$-axis: $Z_i(\theta)$; indicator functions of $\Theta_m^q$ (sum)
- Solid line: reference parabola $[Z_0(\cdot)]$; dashed lines: other parabolas

# General FIR Systems

- System: general FIR, i.e.

$$Y_t = \sum_{k=1}^{d} b_k^* \, U_{t-k} + N_t = \varphi_t^{\mathrm{T}} \theta^* + N_t,$$

  $Y_t$ — output, $U_t$ — input, $N_t$ — noise (at time $t$)

  $\varphi_t \triangleq [\, U_{t-1}, \dots, U_{t-d} \,]^{\mathrm{T}}, \;\; \theta^* \triangleq [\, b_1^*, \dots, b_d^* \,]^{\mathrm{T}}$

- Assumptions: $(N_t)_t$ is independent, symmetric [and continuous]

  $(N_t)_t \perp\!\!\!\perp (U_t)_t$

- Data: $(Y_t)_{t=1}^{n}$ and $(U_t)_{t=1-d}^{n-1}$, i.e., a finite sample

- Goal: to construct a confidence region around the LS estimate, the “true” parameter $\theta^*$ should be included with a guaranteed probability

# Non-Asymptotic Confidence Regions

- We again define a reference sum and $m-1$ sign-perturbed sums

$$S_0(\theta) \triangleq \Phi_n^{-\frac{1}{2}} \sum_{t=1}^{n} \varphi_t (Y_t - \varphi_t^{\mathrm{T}} \theta),$$

$$S_i(\theta) \triangleq \Phi_n^{-\frac{1}{2}} \sum_{t=1}^{n} \varphi_t \, \alpha_{i,t} (Y_t - \varphi_t^{\mathrm{T}} \theta),$$

for $i = 1, \dots, m-1$, where $\alpha_{i,t}$ ($t = 1, \dots, n$) are again i.i.d. random signs, that is $\alpha_{i,t} = \pm 1$ with probability 1/2 each

- Matrix $\Phi_n$ is a covariance estimate $\Longrightarrow$ shapes the region

$$\Phi_n \triangleq \frac{1}{n} \sum_{t=1}^{n} \varphi_t \varphi_t^{\mathrm{T}}.$$

# Non-Asymptotic Confidence Regions

- $S_0(\theta), \ldots, S_{m-1}(\theta)$ are now vector valued ! $\Longrightarrow$ ordering?
- Let's use a norm: $Z_i(\theta) \triangleq ||S_i(\theta)||_2^2$, for $i = 0, \ldots, m-1$
- $R_0(\theta)$ is again the rank of $Z_0(\theta)$ in the ordering

$$Z_i(\theta)_{i_0} < \cdots < Z_i(\theta)_{i_{m-1}}$$

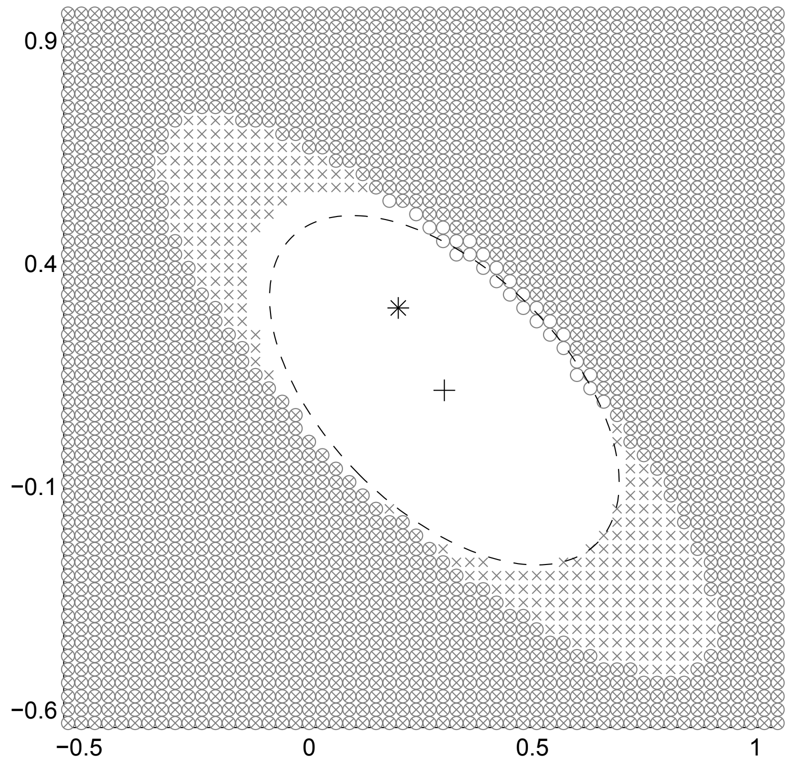
- Similarly as before, the confidence region is

$$\boxed{\Theta_m^q \triangleq \{\, \theta \in \mathbb{R}^d : R_0(\theta) \geq q \,\}\,,}$$

where $q$ and $m$ are user-chosen parameters

- Claim: $\boxed{\mathbb{P}(\theta^* \in \Theta_m^q) = 1 - q/m}$ $\Longrightarrow$ no conservativism !

# Numerical Example: Second-Order FIR




- $99\,\%$ confidence regions $(n = 200,\; m = 100,\; q = 1)$;   inputs: AR(1)
- "$\times$": confidence region with weights $\Phi_n^{-\frac{1}{2}}$;   "$\circ$": without weights
- "$*$": position of $\theta^*$;  '$+$": LS estimate;  dashed ellipsoid: asymptotic set

# PART III.

## Confidence Regions for ARX Systems

# General ARX Systems

- System: general ARX, i.e.

$$Y_t + a_1^* Y_{t-1} + \cdots a_{n_a}^* Y_{t-n_a} \triangleq b_1^* U_{t-1} + \cdots b_{n_b}^* U_{t-n_b} + N_t,$$

which can be written in linear regression form as

$$\begin{aligned} Y_t &= \varphi_t^{\mathrm{T}} \theta^* + N_t, \\ \varphi_t &\triangleq [\,-Y_{t-1}, \dots, -Y_{t-n_a}, U_{t-1}, \dots, U_{t-n_b}\,]^{\mathrm{T}}, \\ \theta^* &\triangleq [\,a_1^*, \dots, a_{n_a}^*, b_1^*, \dots, b_{n_b}^*\,]^{\mathrm{T}}. \end{aligned}$$

- Assumptions: $(N_t)_t$ is independent, each $N_t$ is symmetric about zero
  $(N_t)_t \perp\!\!\!\perp (U_t)_t$

- Data: $(Y_t)_{t=1-n_a}^{n}$ and $(U_t)_{t=1-n_b}^{n-1}$, i.e., a finite sample

- Goal: guaranteed confidence region around the LS estimate

# Difficulties in the Generalization

- The direct application of the previous approach does not work
- For example, in the AR(1) case, we have (assuming $\theta = \theta^*$)

$$S_0(\theta^*) \;=\; \Phi_n^{-\frac{1}{2}} \sum_{t=1}^{n} \varphi_t (Y_t - \varphi_t^{\mathrm{T}} \theta) = \Phi_n^{-\frac{1}{2}} \sum_{t=1}^{n} \begin{bmatrix} -Y_{t-1} \\ U_{t-1} \end{bmatrix} N_t$$

$$S_i(\theta^*) \;=\; \Phi_n^{-\frac{1}{2}} \sum_{t=1}^{n} \varphi_t \, \alpha_{i,t} (Y_t - \varphi_t^{\mathrm{T}} \theta) = \Phi_n^{-\frac{1}{2}} \sum_{t=1}^{n} \begin{bmatrix} -Y_{t-1} \\ U_{t-1} \end{bmatrix} \alpha_{i,t} N_t,$$

for $i = 1, \dots, m-1$, where $(\alpha_{i,t})_{i,t}$ are i.i.d. random signs

- But now $Y_{t-1}$ depends on $N_{t-1}, N_{t-2}, \dots$, thus
- $S_0(\theta^*)$ and $S_i(\theta^*)$, $i \neq 0$, does not have the same distribution

# Restoring the Symmetry

- Idea: construct alternative outputs using sign-perturbed prediction errors

- The alternative outputs can be constructed as

$$\bar{Y}_t(\theta, \alpha_i) = -a_1 \bar{Y}_{t-1}(\theta, \alpha_i) - \cdots - a_{n_a} \bar{Y}_{t-n_a}(\theta, \alpha_i) +$$

$$+ b_1 U_{t-1} + \cdots + b_{n_b} U_{t-n_b} + \alpha_{i,t} \hat{N}_t(\theta),$$

  where $\hat{N}_t(\theta) \triangleq Y_t - \varphi_t^{\mathrm{T}} \theta$ are prediction errors

- This leads to the perturbed regressors ($i = 0, \dots, m-1$)

$$\bar{\varphi}_t(\theta, \alpha_i) = [\, -\bar{Y}_{t-1}(\theta, \alpha_i), \dots, -\bar{Y}_{t-n_a}(\theta, \alpha_i), U_{t-1}, \dots, U_{t-n_b} \,]^{\mathrm{T}}.$$

- This can restore distributional symmetry to the sign-perturbed sums

# Refined Sign-Perturbed Sums

- Refined reference sum and $m-1$ sign-perturbed sums

$$S_0(\theta) \triangleq \Phi_n^{-\frac{1}{2}} \sum_{t=1}^{n} \varphi_t \hat{N}_t(\theta),$$

$$S_i(\theta) \triangleq \bar{\Phi}_n^{-\frac{1}{2}}(\theta, \alpha_i) \sum_{t=1}^{n} \alpha_{it}\, \bar{\varphi}_t(\theta, \alpha_i) \hat{N}_t(\theta),$$

for $i = 0, \dots, m-1$, where $\bar{\Phi}_n(\theta, \alpha_i)$ are perturbed covariances:

$$\bar{\Phi}_n(\theta, \alpha_i) \triangleq \frac{1}{n} \sum_{t=1}^{n} \bar{\varphi}_t(\theta, \alpha_i) \bar{\varphi}_t^{\mathrm{T}}(\theta, \alpha_i),$$

- Claim: $\boxed{\mathbb{P}(\theta^* \in \Theta_m^q) = 1 - q/m}$ $\Longrightarrow$ exact, no conservativism

# PART IV.

## Confidence Regions for General Linear Systems

# General Linear Systems

- System: general linear, i.e.

$$A(z^{-1})\, Y_t = \frac{B(z^{-1})}{F(z^{-1})}\, U_t + \frac{C(z^{-1})}{D(z^{-1})}\, N_t,$$

where $Y_t$ — output, $U_t$ — input, $N_t$ — noise,
$A$, $B$, $C$, $D$ and $F$ are (finite) polynomials in $z^{-1}$, i.e., $z^{-1}Y_t = Y_{t-1}$

- The coefficients of $A$, $B$, $C$, $D$ and $F$ are $(a_k^*)_{k=1}^{n_a}$, $(b_k^*)_{k=1}^{n_b}$, $(c_k^*)_{k=1}^{n_c}$, $(d_k^*)_{k=1}^{n_d}$ and $(f_k^*)_{k=1}^{n_f}$, respectively. We use

$$\theta^* = (a_1^*, \dots, a_{n_a}^*, b_1^*, \dots, b_{n_b}^*, c_1^*, \dots, c_{n_c}^*, d_1^*, \dots, d_{n_d}^*, f_1^*, \dots, f_{n_f}^*)^{\mathrm{T}}.$$

- Data: $(Y_t)_{t=1}^{n}$ and $(U_t)_{t=1}^{n-1}$, i.e., a finite sample

- Goal: guaranteed confidence region around the PEM estimate

# System Assumptions

- The system can be written using (rational) transfer functions:

$$Y_t \triangleq G(z^{-1};\theta^*)\,U_t + H(z^{-1};\theta^*)\,N_t,$$

**Assumption 1.** *The true system is in the model class, the orders are known*

**Assumption 2.** *The transfer function $H(z^{-1};\theta^*)$ has a (stable) inverse, as well as $G(0;\theta^*) = 0$ and $H(0;\theta^*) = 1$*

**Assumption 3.** $(N_t)_{t=1}^n$ *is independent, symmetrically distributed about zero*

**Assumption 4.** $(U_t)_{t=1}^{n-1}$ *is observed and independent of* $(N_t)_{t=1}^n$

**Assumption 5.** *Initialization: for all $t \leq 0$, we have $Y_t = N_t = U_t = 0$*

# Prediction Error Estimate

- The prediction errors can be calculated as

$$\hat{N}_t(\theta) \triangleq H^{-1}(z^{-1};\theta)(Y_t - G(z^{-1};\theta)\, U_t),$$

which defines the "reconstructed" noise at $t$ for $\theta$

- Note that $\hat{N}_t(\theta^*) = N_t$, thus, the reconstruction is accurate for $\theta = \theta^*$

- The prediction error estimate (PEM) for model class $\mathcal{M}$ is

$$\hat{\theta}_{\mathrm{PEM}} \triangleq \arg\min_{\theta \in \mathcal{M}} \sum_{t=1}^{n} \hat{N}_t^2(\theta).$$

- In general, there is no closed-form solution for PEM

# Prediction Error Estimate

- The PEM estimate can be found, e.g., by using the equation

$$\sum_{t=1}^{n} \psi_t(\hat{\theta}_{\mathrm{PEM}})\hat{N}_t(\hat{\theta}_{\mathrm{PEM}}) \,=\, 0,$$

  where $\psi_t(\theta)$ is the negative gradient of the prediction error,

$$\psi_t(\theta) \,\triangleq\, -\frac{d}{d\theta}\hat{N}_t(\theta).$$

- In case of ARX systems $\psi_t(\theta)$ is simply $\varphi_t$
- These gradients can be directly calculated in terms of the defining polynomials $A$, $B$, $C$, $D$ and $F$

# Gradients of the Prediction Error

$$\frac{\partial}{\partial a_k}\hat{N}_t(\theta) = \frac{D(z^{-1})}{C(z^{-1})}\,Y_{t-k},$$

$$\frac{\partial}{\partial b_k}\hat{N}_t(\theta) = \frac{D(z^{-1})}{C(z^{-1})F(z^{-1})}\,U_{t-k},$$

$$\frac{\partial}{\partial c_k}\hat{N}_t(\theta) = \frac{D(z^{-1})B(z^{-1})}{C(z^{-1})C(z^{-1})F(z^{-1})}\,U_{t-k} - \frac{D(z^{-1})A(z^{-1})}{C(z^{-1})C(z^{-1})}\,Y_{t-k},$$

$$\frac{\partial}{\partial d_k}\hat{N}_t(\theta) = \frac{A(z^{-1})}{C(z^{-1})}\,Y_{t-k} - \frac{B(z^{-1})}{C(z^{-1})F(z^{-1})}\,U_{t-k},$$

$$\frac{\partial}{\partial f_k}\hat{N}_t(\theta) = \frac{D(z^{-1})B(z^{-1})}{C(z^{-1})F(z^{-1})F(z^{-1})}\,U_{t-k}.$$

# Perturbed Gradients

- We again apply perturbed versions of the (reconstructed) outputs

$$\bar{Y}_t(\theta, \alpha_i) \triangleq G(z^{-1}; \theta)\, U_t + H(z^{-1}; \theta)\, (\alpha_{it} \hat{N}_t(\theta)),$$

where $(\alpha_{it})_{i,t}$ are random signs, as previously

- As we saw, $\psi_t(\theta)$ is a linear filtered version of $(Y_t)_t$ and $(U_t)_t$,

$$\psi_t(\theta) = W_0(z^{-1}; \theta)\, Y_t + W_1(z^{-1}; \theta)\, U_t,$$

where $W_0$ and $W_1$ are vector-valued, and $\psi_t(\theta) \in \mathbb{R}^d$

- We use these to define perturbed versions of the gradient, $\psi_t(\theta)$,

$$\bar{\psi}_t(\theta, \alpha_i) \triangleq W_0(z^{-1}; \theta)\, \bar{Y}_t(\theta, \alpha_i) + W_1(z^{-1}; \theta)\, U_t,$$

where the novelty is that now we filter the perturbed outputs

# General Sign-Perturbed Sums

- Finally, the sign-perturbed sums for $\theta$ are $(0 \leq i \leq m-1)$

$$S_0(\theta) \triangleq \Psi_n^{-\frac{1}{2}}(\theta) \sum_{t=1}^{n} \psi_t(\theta) \hat{N}_t(\theta),$$

$$S_i(\theta) \triangleq \bar{\Psi}_n^{-\frac{1}{2}}(\theta, \alpha_i) \sum_{t=1}^{n} \alpha_{it}\, \bar{\psi}_t(\theta, \alpha_i) \hat{N}_t(\theta),$$

- Where $\Psi_n$ and $\bar{\Psi}_n(\theta, \alpha_i)$, the perturbed covariances, are

$$\Psi_n(\theta) \triangleq \frac{1}{n} \sum_{t=1}^{n} \psi_t(\theta) \psi_t^{\mathrm{T}}(\theta),$$

$$\bar{\Psi}_n(\theta, \alpha_i) \triangleq \frac{1}{n} \sum_{t=1}^{n} \bar{\psi}_t(\theta, \alpha_i) \bar{\psi}_t^{\mathrm{T}}(\theta, \alpha_i).$$

# Non-Asymptotic Confidence Regions

- $R_0(\theta)$ is again the rank of $\|S_0(\theta)\|_2^2$ among $(\|S_i(\theta)\|_2^2)_{i=0}^{m-1}$

- As before, the confidence region is defined as

$$\Theta_m^q \triangleq \{\, \theta \in \mathbb{R}^d \,:\, R_0(\theta) \geq q \,\},$$

  where $q \in \{0, \dots, m-1\}$ and $m$ are user-chosen parameters

- The PEM estimate, $\hat{\theta}_{\mathrm{PEM}}$, satisfies $S_0(\hat{\theta}_{\mathrm{PEM}}) = 0$, which guarantees the inclusion of $\hat{\theta}_{\mathrm{PEM}}$ in $\Theta_m^q$

- For general linear systems, under Assumptions 1-5, the confidence probability that $\theta^*$ is in $\Theta_m^q$ is exactly $1 - q/m$

# Example: ARMA(1, 1)

- Consider the following ARMA process

$$Y_t + a^* Y_{t-1} = N_t + c^* N_{t-1},$$

where the "true" parameter is $\theta^* = (a^*,\, c^*)$

- The inverse filter of $C(z^{-1}; \theta)\, N_t = N_t + c\, N_{t-1}$ is

$$C^{-1}(z^{-1}; \theta) \,=\, \sum_{k=0}^{\infty} (-1)^k\, c^k\, z^{-k}$$

- We can use this to define the prediction errors as

$$\hat{N}_t(\theta) = C^{-1}(z^{-1}; \theta)\,(Y_t + a\, Y_{t-1}),$$

for all $t$, where $\theta = (a,\, c)$ is a given parameter

# Example: ARMA(1, 1)

- Then, the reconstructed outputs are defined as

$$\bar{Y}_t(\theta, \alpha_i) \;=\; -a\,\bar{Y}_{t-1}(\theta, \alpha_i) + \alpha_{i,t}\,\hat{N}_t(\theta) + c\,\alpha_{i,t-1}\hat{N}_{t-1}(\theta),$$

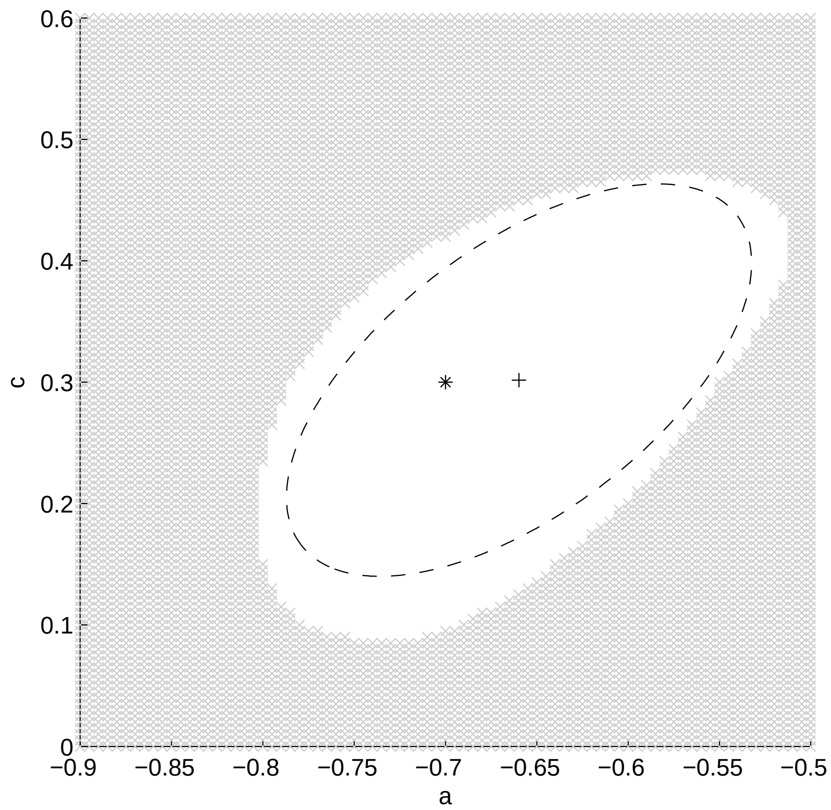
for $1 \leq i \leq m$ and $1 \leq t \leq n$, where $(\alpha_{it})_{i,t}$ are random signs

- Finally, we can calculate the negative gradient, $\bar{\psi}_t(\theta, \alpha_i)$,

$$\bar{\psi}_t(\theta, \alpha_i) \;=\; \begin{bmatrix} -C^{-1}(z^{-1};\theta)\,\bar{Y}_t(\theta, \alpha_i) \\ C^{-1}(z^{-1};\theta)\,\alpha_{i,t}\,\hat{N}_t(\theta) \end{bmatrix},$$

which can be used to define the sign-perturbed sums

# Simulation: ARMA(1, 1)




- $99\,\%$ confidence regions $(n = 500,\ m = 100,\ q = 1)$
- "$\times$": confidence region based on the SPS algorithm
- "$\ast$": position of $\theta^*$; '$+$": LS estimate; dashed ellipsoid: asymptotic set

# PART V.

## Summary and Concluding Remarks

# Summary and Concluding Remarks

- A new finite sample identification algorithm has been introduced
- It constructs confidence regions centered around the PEM estimate
- Only mild statistical assumptions are needed, e.g., symmetry about zero
- No information on stationarity, variances, particular distributions, etc.
- For any (rational) confidence probability, a region can be constructed
- The confidence probability is exact, even for multiple parameters
- Core idea: comparing carefully chosen sign-perturbed sums
- The algorithm has been presented for general linear systems
- The fundamental ideas work for much more general systems

# Thank you for your attention!

bcsaji@unimelb.edu.au