Probabilistic Model Uncertainty Bounding - an approach with finite time perspectives

Sippe Douma and Paul Van den Hof

14th IFAC Symposium on System Identification, 29-31 March 2006, Newcastle, Australia

Delft Center for Systems and Control

TU Delft

Delft University of Technology

Uncertainty bounding in PE identification

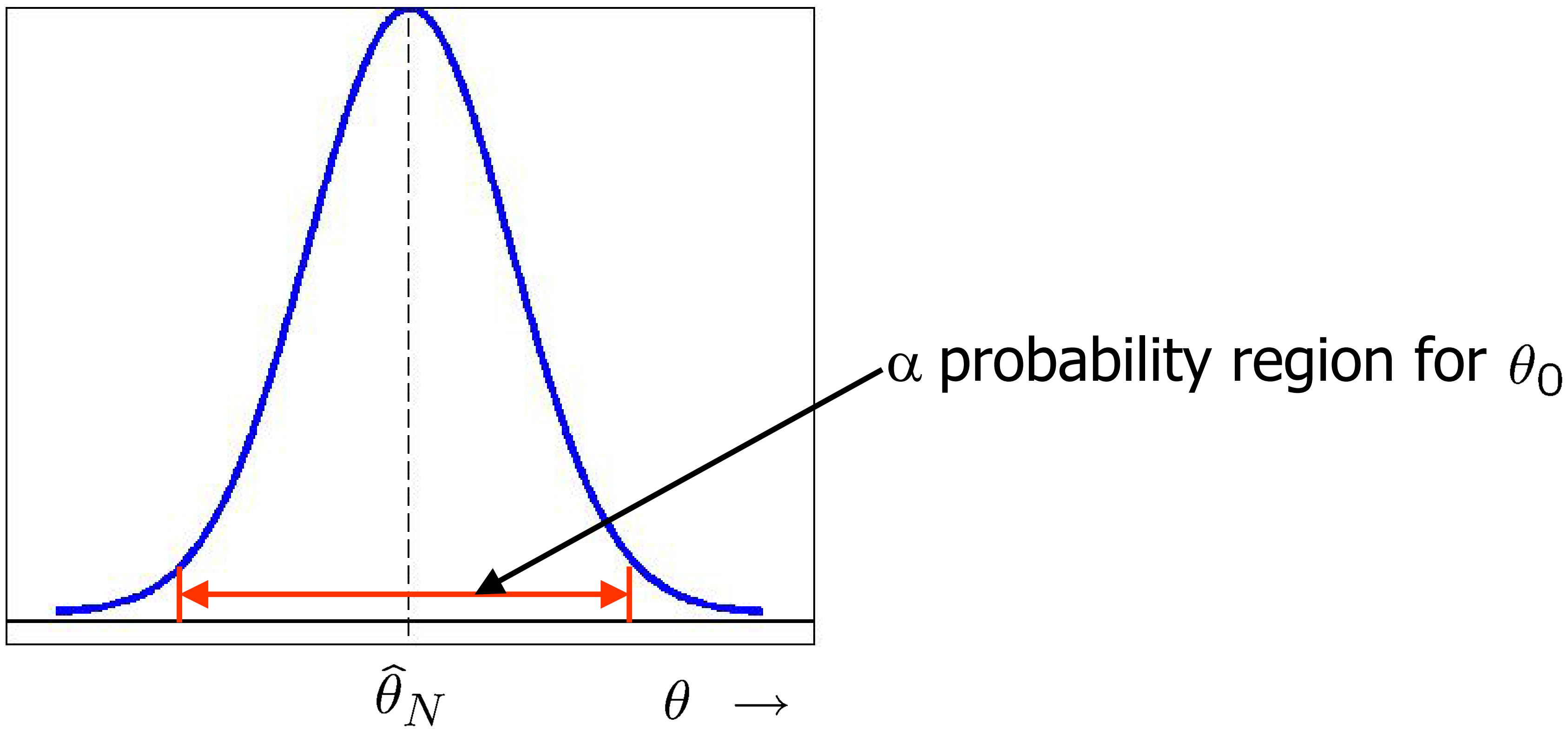
1. Analyse statistical properties of estimator $Z^N \rightarrow \hat{\boldsymbol{\theta}}_N$ through its pdf $p_{\boldsymbol{\theta}}(\theta)$

2. Under (asymptotic) assumptions and no bias
$$p_{\boldsymbol{\theta}}(\theta) \sim \mathcal{N}(\theta_0, P/N)$$
with an analytical expression for covariance matrix P

3. For a given single realization $\hat{\theta}_N$ determine a set of θ for which, within a probability level of α%, holds that
$$\hat{\theta}_N \in \mathcal{N}(\theta, P/N)$$

4. This set is identical to the set of θ that forms an α% probability set of the pdf $\mathcal{N}(\hat{\theta}_N, P/N)$

$\mathcal{N}(\hat{\theta}_N, P/N)$



$$\theta_0 \in \left\{ \theta \mid (\theta - \hat{\theta}_N) P^{-1} (\theta - \hat{\theta}_N) \leq c_\chi(\alpha, n)/N \right\} \quad \text{w.p.} \quad \alpha$$

- Approximations have to be made for obtaining computable expressions:
  - Employing asymptotic Gaussian distribution of pdf
  - Assumption $\mathcal{S} \in \mathcal{M}$ (no bias)
  - Obtaining P through Taylor approximation (OE/BJ)
  - Replacing covariance matrix P by estimate

The message

- *Estimator* statistics are not *necessary* for obtaining probabilistic parameter uncertainty regions;
  There are alternatives with attractive properties

Contents

- Example for illustration
- Uncertainty bounding in ARX models
- Extension towards finite time (?)
- Summary

Example

Data generating system: $\mathrm{y} = \theta_0 \boldsymbol{x}_1 + \boldsymbol{x}_2$

$\boldsymbol{x}_2 \in \mathcal{N}(0,2)$; $\boldsymbol{x}_1$ correlated with $\boldsymbol{x}_2$; 1 data point (x_1, y)

Estimator: $\boldsymbol{\theta} = \mathrm{y}/\boldsymbol{x}_1 = \theta_0 + \boldsymbol{x}_2/\boldsymbol{x}_1$

pdf of $\boldsymbol{\theta}$ is very hard to analyze

However: $\boldsymbol{x}_1(\boldsymbol{\theta} - \theta_0) = \boldsymbol{x}_2 \in \mathcal{N}(0,2)$

After one experiment we have realizations: $x_1, \widehat{\theta}$ of $\boldsymbol{x}_1, \boldsymbol{\theta}$
Then $x_1(\widehat{\theta} - \theta_0)$ is a realization of $\boldsymbol{x}_2 \in \mathcal{N}(0,2)$.
Based on test statistic $x_1(\widehat{\theta} - \tilde{\theta})$ we select all $\tilde{\theta}$ that are within the α-probability level of $\boldsymbol{x}_2$:

$$\theta_0 \in \left\{\theta \mid (\widehat{\theta}-\theta)x_1^2(\widehat{\theta}-\theta) \leq 2c_\chi(\alpha,1)\right\} \quad \text{w.p.} \quad \alpha$$

Probabilistic parameter bounding without pdf of estimator

Employ statistical properties of random variable

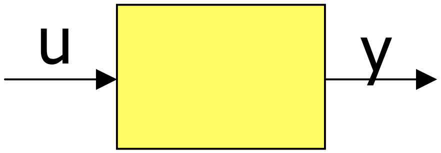
$$\boldsymbol{x}_1(\boldsymbol{\theta}-\theta_0) = \boldsymbol{x}_2 \in \mathcal{N}(0,2)$$

rather than those of

$$\boldsymbol{\theta}-\theta_0$$

Benefit = simplicity of expression / analysis

Issue in i/o dynamical systems:



Question whether to consider measured input u as deterministic or stochastic in variance analysis of estimator

Prior or posterior variance

However, the issue goes beyond the role of u, and also incorporates the role of y.

Uncertainty bounding in ARX models

$$\hat{y}(t|t-1;\theta) = \varphi^T(t)\theta$$

With

$$\Phi = \begin{pmatrix} \varphi^T(1) \\ \vdots \\ \varphi^T(N) \end{pmatrix} \text{ and } \mathbf{y} = [y(1)\cdots y(N)]^T$$

$$\hat{\theta}_N = (\Phi^T\Phi)^{-1}\Phi^T\mathbf{y}$$

If $\mathcal{S} \in \mathcal{M}$: $\quad \mathbf{y} = \Phi\theta_0 + \mathbf{e}$

$$\boxed{\hat{\theta}_N - \theta_0 = (\Phi^T\Phi)^{-1}\Phi^T\mathbf{e}}$$

ARX modelling

If $\mathcal{S} \in \mathcal{M}$: $\hat{\theta}_N - \theta_0 = (\Phi^T \Phi)^{-1} \Phi^T \mathrm{e}$

Model uncertainty bounding requires:

- (asymptotic) normality of $(\Phi^T \Phi)^{-1} \Phi^T \mathrm{e}$
- Replacement of P_{arx} :

$$P_{arx} = (\mathbb{E}[\frac{1}{N}\Phi^T \Phi])^{-1} \cdot \sigma_e^2$$

by an estimate $\hat{P}_{arx}$

$$\hat{P}_{arx} = (\frac{1}{N}\Phi^T \Phi)^{-1} \hat{\sigma}_e^2$$

leading to ellipsoid determined by $\hat{P}_{arx}$

Alternative

$$\hat{\theta}_N - \theta_0 = (\Phi^T \Phi)^{-1} \Phi^T \mathrm{e}$$

$$\frac{1}{\sqrt{N}} \Phi^T \Phi (\hat{\theta}_N - \theta_0) = \frac{1}{\sqrt{N}} \Phi^T \mathrm{e}.$$

Model uncertainty bounding requires:

- (asymptotic) normality of $\frac{1}{\sqrt{N}} \Phi^T \mathrm{e}$ with covariance Q
- Replacement of

$$Q = \mathbb{E}[\frac{1}{N} \Phi^T \Phi] \cdot \sigma_e^2$$

by an estimate

$$\frac{1}{N} \Phi^T \Phi \hat{\sigma}_e^2$$

leading to ellipsoid determined by $\hat{P}_{arx,n} = (\frac{1}{N} \Phi^T \Phi)^{-1} \hat{\sigma}_e^2$

Same as before but conditions are relaxed

ARX modelling

Note that in its implemented form, using

$$\hat{P}_{arx,n} = (\frac{1}{N}\Phi^T\Phi)^{-1}\hat{\sigma}_e^2$$

result is related to likelihood method, determined by

$$\left\{\theta \mid V_N(\theta) - V_N(\hat{\theta}_N) \leq c_\chi(\alpha, n)/N\right\}$$

level sets of cost function

(Donaldson & Schnabel, 1987)

Conclusion

Classical results with P_{arx} approximated by sample estimates, requires

$$\frac{1}{\sqrt{N}}\Phi^T \mathrm{e}$$

to become (asymptotically) Gaussian rather than

$$(\Phi^T\Phi)^{-1}\Phi^T \mathrm{e}$$

Benefit: relaxation of conditions for normality

One step further

$$(\Phi^T\Phi)(\hat{\theta}_N - \theta_0) \;=\; \Phi^T\mathbf{e}$$

With svd: $\Phi^T = U\Sigma V^T$ it follows that

$$\Sigma^{-1}U^T(\Phi^T\Phi)(\hat{\theta}_N - \theta_0) = V^T\mathbf{e}$$

Lemma:
If V^T unitary and random, and $\mathbf{e}$ Gaussian with cov($\mathbf{e}$)=σ^2 I, and V^T and $\mathbf{e}$ independent, then $V^T\mathbf{e}$ is Gaussian with Cov = σ^2 I.

This would suggest that $V^T\mathbf{e}$ is Gaussian for any value of N.

Corresponding uncertainty

Starting from:

$$\Sigma^{-1}U^T(\Phi^T\Phi)(\widehat{\theta}_N - \theta_0) \in \mathcal{N}(0, \sigma_e^2 I)$$

It follows that (for finite N):

$$\theta_0 \in \left\{\theta \mid (\widehat{\theta}_N - \theta)P^{-1}(\widehat{\theta}_N - \theta) \leq c_\chi(\alpha, n)/N\right\} \quad \text{w.p.} \quad \alpha$$

with $$P = (\frac{1}{N}\Phi^T\Phi)^{-1}\sigma_e^2$$

Only σ_e^2 needs to be replaced by an estimated expression

Only problem: condition of independent V^T and $\mathbf{e}$ is not satisfied for ARX models

However: this does not appear devastating in practice!

Simulation example:

First order ARX system:

$$y(t) = \frac{0.5}{1 + 0.9q^{-1}} u(t) + \frac{1}{1 + 0.9q^{-1}} e(t)$$

identified with 1st order ARX model.

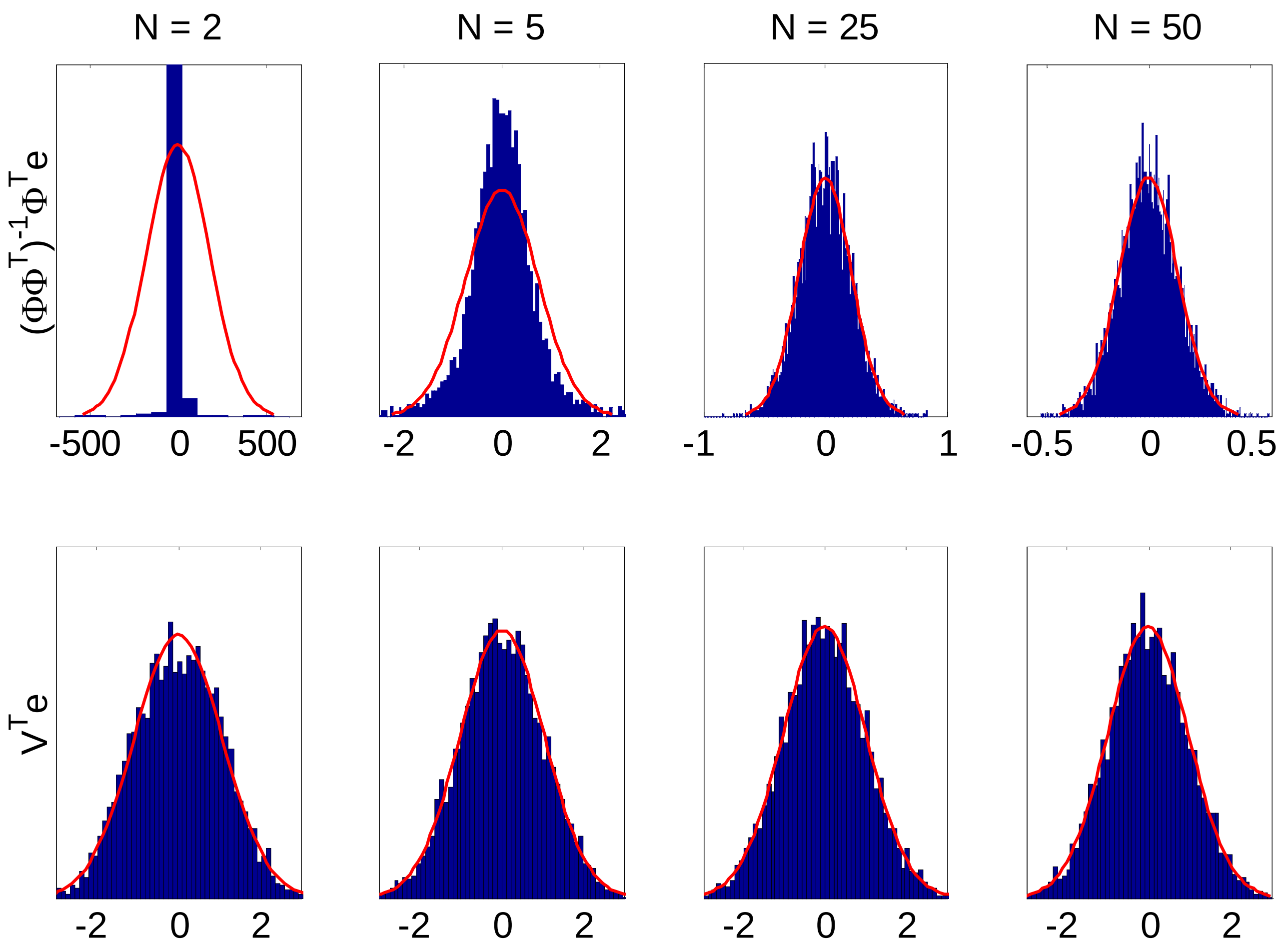
Compare empirical distributions of

$$\hat{\theta}_N - \theta_0 = (\Phi^T \Phi)^{-1} \Phi^T \mathbf{e}$$

and $$V^T \mathbf{e}$$

for different values of N,
on the basis of 5000 Monte Carlo simulations

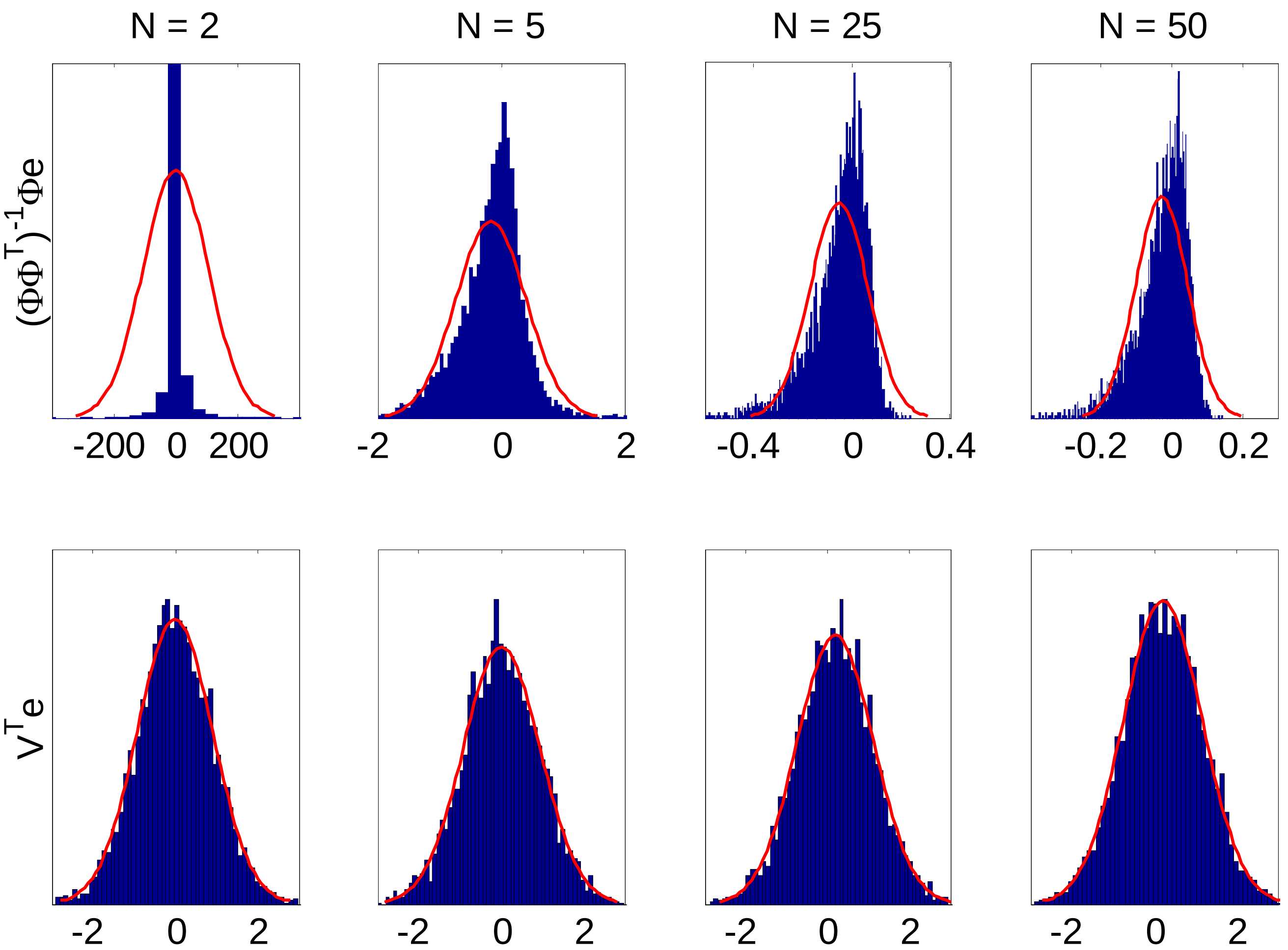
Component related to numerator parameter:



Delft Center for Systems and Control

TUDelft

Component related to denominator parameter:



Summary

- There are alternatives for parameter uncertainty bounding, without constructing pdf of estimator
- Applicable to ARX, OE and also BJ models (Douma & VdHof, CDC/ECC-2005, ACC2006)
- Leading to simpler and less approximative expressions, remarkably robust w.r.t. finite time properties
- Relation with (empirical) likelihood based uncertainty intervals

DCSC
Delft Center for Systems and Control

TU Delft