

# Parametric uncertainty in system identification

***Citation for published version (APA):***

Falkus, H. M. (1994). *Parametric uncertainty in system identification*. [Phd Thesis 1 (Research TU/e / Graduation TU/e), Electrical Engineering]. Technische Universiteit Eindhoven. <https://doi.org/10.6100/IR420589>

***DOI:***

[10.6100/IR420589](https://doi.org/10.6100/IR420589)

***Document status and date:***

Published: 01/01/1994

***Document Version:***

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

***Please check the document version of this publication:***

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

***General rights***

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.tue.nl/taverne](http://www.tue.nl/taverne)

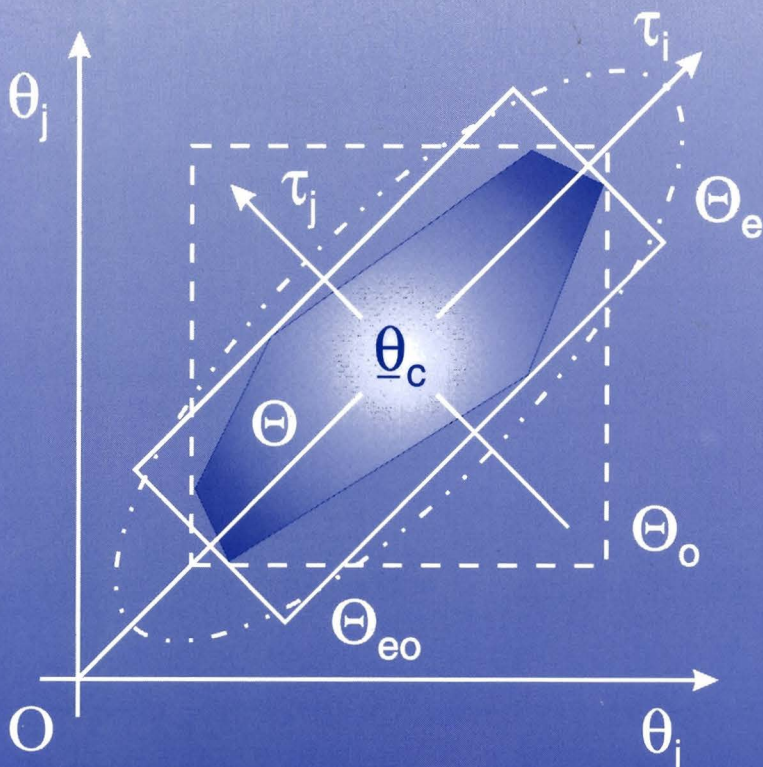
***Take down policy***

If you believe that this document breaches copyright please contact us at:

[openaccess@tue.nl](mailto:openaccess@tue.nl)

providing details and we will investigate your claim.

# ***Parametric Uncertainty in System Identification***



***Heinz Falkus***



# ***Parametric Uncertainty in System Identification***

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de  
Technische Universiteit Eindhoven,  
op gezag van de Rector Magnificus, prof.dr. J.H. van Lint,  
voor een commissie aangewezen door het College van  
Dekanen in het openbaar te verdedigen op  
donderdag 22 september 1994 om 16.00 uur

door

**Heinz Manfred Falkus**

geboren te Oldenburg



Dit proefschrift is goedgekeurd door de promotoren :

prof.dr.ir. A.C.P.M. Backx

en

prof.ir. O.H. Bosgra

Copromotor : dr.ir. A.A.H. Damen

CIP-DATA KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Falkus, Heinz Manfred

Parametric uncertainty in system identification /

Heinz Manfred Falkus. - Eindhoven :

Eindhoven University of Technology

Thesis Eindhoven. - With ref. - With summary in Dutch.

ISBN 90-386-0263-4

Subject headings : system identification.

# Voorwoord

---

Het puntje op de *i* zetten, kost meer tijd dan het schrijven van het hele proefschrift. Hoewel het schrijven gedurende de laatste 9 maanden van de promotie neerkomt op slechts een ½ woord per minuut, staat deze verhouding in geen enkele relatie met de **stress** van de laatste maanden, weken en dagen, die bij het afronden van de promotie onvermijdelijk is. Dit is echter een onontkoombaar feit, dat elke promovendus ondergaat en *de laatste loodjes wegen het zwaarst* is dan ook zeker van toepassing. Toch weegt dit niet op tegen de voldoening die er is, als het proefschrift eindelijk goedgekeurd wordt en naar de drukker kan.

Het onderzoek dat in dit proefschrift beschreven wordt, is verricht aan de Technische Universiteit Eindhoven bij de vakgroep Meten en Regelen. In de eerste instantie wil ik dan ook prof. Ton Backx, dr. Ad Damen en dr. Siep Weiland bedanken, voor de stimulerende discussies en hun adviezen, die een zeer positieve uitwerking hebben gehad op het onderzoek en de uiteindelijke resultaten. Verder wil ik prof. Okko Bosgra, prof. Paul v.d. Bosch en prof. Bart De Moor graag bedanken, voor het doorlezen van de conceptversie van dit proefschrift. Zij hebben in belangrijke mate bijgedragen, om de laatste puntjes op de *i* te zetten.

Het onderzoek zoals beschreven in dit proefschrift, maakte deel uit van een samenwerkingsverband met de universiteiten van Leuven (België), Leicester (Engeland) en Delft. De discussies tijdens de vele projectbijeenkomsten met de promovendi Peter



van Overschee, Ghassan Murrad, Richard Hakvoort en Erik van Bracht, vormden een belangrijke bron van inspiratie. In dit verband wil ik ook Jobert Ludlage van SETPOINT IPCOS bedanken, voor het aandragen van de praktijkproblemen, die uiteindelijk geresulteerd hebben in industrieel toepasbare technieken.

De sfeer in een vakgroep is mede bepalend voor de motivatie en daarmee natuurlijk ook van invloed op een succesvolle afronding van een promotie. De prettige samenwerking en vele discussies met mijn kamergenoten Marc Keulers en Robert-Jan Gorter en mijn collega AIO's Ton v.d. Boom, Leon Ariaans en Jozef Mazak, zijn belangrijk geweest bij het tot stand komen van dit proefschrift. De vele stagiaires en afstudeerders wil ik bedanken voor hun bijdrage in de verschillende deelproblemen. Het snelle oplossen van de vele computer problemen, verdient een woord van dank aan Udo Bartzke en Wim Beckers. Verder vormde het squashen met verschillende collegae, naast een gezellige avond uit, ook een ideale manier om even lekker uit te leven.

Tenslotte wil ik mijn vriendin Henny bedanken, voor haar begrip gedurende de laatste maanden. Zij zorgde er voor dat de stress beperkt bleef, voor veel afwisseling en voor een veilige landing als mijn gedachten weer eens te ver afdwaalden. Een bijzonder woord van dank verdienen uiteindelijk mijn ouders. Zij zijn altijd mijn steun en toeverlaat geweest tijdens mijn nu bijna vijftientig jaar durende studieloopbaan. Met hun voortdurende stimulatie hebben zij mede de basis gelegd voor deze promotie.

Eindhoven, 27 juni 1994

# ***Abstract***

---

Identification procedures in the time domain are studied for multivariable systems which yield a model with uncertainty bounds on its parameters. This parametric uncertainty can either be stochastic or deterministic depending on the assumptions which have been made with respect to the noise disturbing the system, i.e. statistical properties or bounded error characteristics.

The classical prediction error identification approach is briefly reviewed describing first the SISO model structures and the (pseudo-) canonical extensions to the multivariable case. These multivariable model structures, however, are not very suitable for process identification in practice, especially in an industrial environment because of the structure uncertainty. Therefore a minimal polynomial model structure is adopted which shows interesting similarities with the SISO case. The model estimation is solved by non-linear least-squares optimization minimizing the sum-squared prediction errors where the first- and second-order derivatives can be computed analytically in a very efficient way. In the classical approach the assumption is required that the process is contained in the defined model set in order to provide stochastic uncertainty bounds. In practice, however, only a low order process approximation is obtained. For the situation that the process is not contained in the model set, an expression for the parameter covariance matrix is presented taking the undermodelling explicitly into account, which can be used to compute stochastic uncertainty bounds on the parameters.

The bounded error identification or set estimation approach provides models with deterministic uncertainty on the model parameters. The bounded error assumption on the noise (prediction error) disturbing the process is translated into constraints in the parameter space which define the feasible parameter set. To guarantee that a convex and connected parameter set (polytope) in the parameter space is defined, the constraints have to be linear in the model parameters. For models which are not linear in their parameters, a linear approximation is given. The parameter set bounded by a polytope cannot be described exactly for high-dimensional identification problems due to exploding computational complexity. Therefore, simpler although approximate descriptions are preferred like ellipsoid or orthotopic bounding. The least conservative approximation of the parameter uncertainty set is obtained by ellipsoid-aligned orthotopic bounding which combines the advantages of both approaches. However, the application of this identification technique in practice is limited due to data outliers which violate the defined error bound. Furthermore, it is sensitive to overparametrization. Outliers may result from mistakes made during the acquisition and preprocessing of the data, but also from overoptimistic error bounds or unmodelled dynamics. To avoid that the parameter uncertainty set becomes empty, a robustification method for this identification approach is proposed where a reference model is used to detect data outliers in the constraints. The sensitivity to overparametrization results from the conservatism of the set estimation method itself and from the conservative approximation of the parameter uncertainty set which has been defined to solve the identification problem. If the parameter uncertainties are very large, the corresponding central estimate provides no accurate description of the process. This conservatism can be reduced by fixing the most uncertain parameters which also reduces the uncertainty of the remaining parameters. These difficulties limit the application of the bounded error identification approach in practice, especially in an industrial environment.

For connection with robust control design, the real parametric uncertainty in the polynomial models is transformed into a linear fractional transformation representation. The uncertainty is represented in a block-diagonal form and the nominal model together with the corresponding connection matrices (between the block-diagonal uncertainty and the nominal state-space model) are combined in an augmented state-space model.

The proposed identification methods are tested on various case studies ranging from SISO laboratory processes to multivariable industrial production processes. In these case studies, the main attention is focused on the application of the two identification methods in practice and to illustrate their advantages and disadvantages. Comparing both methods, the prediction error approach is preferred for the identification of processes in an industrial environment.



# Contents

---

	<b>Glossary</b>	<b>xi</b>
<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	General Introduction . . . . .	1
1.1.1	Dynamical Systems . . . . .	2
1.1.2	Disturbances . . . . .	3
1.1.3	Models and Modelling . . . . .	3
1.1.4	System Identification . . . . .	4
1.1.5	Practicalities . . . . .	5
1.2	Motivation . . . . .	7
1.2.1	Robust Control . . . . .	7
1.2.2	Uncertainties . . . . .	9
1.3	Problem Statement . . . . .	14
1.4	Scope of this Thesis . . . . .	17
1.5	Wat's New . . . . .	20
<b>2</b>	<b>Minimal Polynomial Parameter Estimation</b>	<b>23</b>
2.1	General Model Structure . . . . .	25
2.1.1	General Model Uniqueness . . . . .	26

2.1.2	Prediction Error . . . . .	27
2.2	Model Parametrization . . . . .	28
2.2.1	Scalar Model Structures . . . . .	28
2.2.2	Multivariable Model Structures . . . . .	29
2.3	Parameter Estimation . . . . .	34
2.3.1	Identification Criterion . . . . .	34
2.3.2	Order Selection . . . . .	36
2.3.3	Initial Estimates . . . . .	39
2.3.4	Numerical Minimization . . . . .	42
2.4	Asymptotic Properties . . . . .	44
2.4.1	Convergence . . . . .	44
2.4.2	Consistency . . . . .	45
2.4.3	Asymptotic Distribution . . . . .	46
2.4.4	Confidence Intervals . . . . .	49
<b>3</b>	<b>Bounded Error Identification</b>	<b>51</b>
3.1	Model Restrictions . . . . .	53
3.2	Outer Parameter Bounding . . . . .	62
3.2.1	Exact Polytope Bounding . . . . .	63
3.2.2	Ellipsoidal Bounding . . . . .	65
3.2.3	Orthotopic Bounding . . . . .	68
3.2.4	Ellipsoid-aligned Orthotopic Bounding . . . . .	69
3.3	Data Outliers . . . . .	71
3.4	Comparison of Stochastic and Deterministic Uncertainty . . . . .	75
<b>4</b>	<b>Robust Control Models</b>	<b>83</b>
4.1	General Block-diagonal Uncertainty . . . . .	87
4.2	Unstructured Uncertainty . . . . .	89
4.3	Structured Uncertainty . . . . .	89
4.3.1	General LFT Description for Parametric Uncertainty . . . . .	91
4.3.2	State-space Realization of Polynomial Models . . . . .	94
4.3.3	LFT Realization of Polynomial Process and Noise Models . . . . .	96
<b>5</b>	<b>Parsimonification</b>	<b>103</b>
5.1	Deterministic Parameter Uncertainty . . . . .	104
5.2	Stochastic Parameter Uncertainty . . . . .	107
5.3	Parameter Uncertainty Reduction . . . . .	108
<b>6</b>	<b>Case Studies</b>	<b>113</b>

---

6.1	A Watervessel Laboratory Process	114
6.1.1	A Theoretical Vessel Model	115
6.1.2	Bounds on the Disturbances	116
6.1.3	Reference Model	117
6.1.4	Bounded Error Estimates	118
6.2	A Fed-batch Fermentation Process	120
6.2.1	Yeast Production	121
6.2.2	The Laboratory Process	122
6.2.3	Nonlinear Compensation	124
6.2.4	Linear Identification	125
6.2.5	Set Estimation	129
6.3	A Distillation Column Simulation Process	131
6.3.1	Experiments and Model Structure Selection	133
6.3.2	Parametric Identification	135
6.4	A Glass Tube Production Process	141
6.4.1	Data Preprocessing	143
6.4.2	Model Estimation	144
<b>7</b>	<b>Conclusions and Remarks</b>	<b>151</b>
<b>A</b>	<b>Data Preprocessing</b>	<b>155</b>
A.1	Detrending	156
A.2	Peakshaving	160
A.3	Delay Correction	161
A.4	Data Decimation and Scaling	163
<b>B</b>	<b>Minimal Polynomial Identification</b>	<b>167</b>
B.1	Model Structure	167
B.2	Prediction Error	169
B.3	Pseudo-linear Regression Form	170
B.4	Parameter Optimization	172
B.5	Gradient Computation	174
B.6	Hessian Computation	177
<b>C</b>	<b>Parameter Covariance</b>	<b>187</b>
C.1	Covariance Approximation	188
C.2	Confidence Computation	193
<b>D</b>	<b>Parametric Uncertainty Descriptions</b>	<b>197</b>



---

D.1	Multivariable Bounded Error Models .....	197
D.2	Exact Polytope Updating .....	200
D.2.1	Initial Orthotope .....	200
D.2.2	Updating Vertex Set .....	202
D.2.3	Updating Vertex-plane Adjacency List .....	204
D.2.4	Updating Vertex-vertex Adjacency List .....	204
D.3	Ellipsoidal Parameter Bounding .....	205
D.4	Orthtopic Parameter Bounding .....	212
<b>References</b>		<b>215</b>
<b>Samenvatting</b>		<b>231</b>
<b>Curriculum Vitae</b>		<b>233</b>

# Glossary

## Symbols

$A(\underline{\theta})$ , $A^{co}(\underline{\theta})$ , $A^{ob}(\underline{\theta})$	State matrix of state-space model in controller or observer canonical form.
$A(z^{-1})$	Polynomial in prediction error model.
$B(\underline{\theta})$ , $B^{co}(\underline{\theta})$ , $B^{ob}(\underline{\theta})$	Input matrix of state-space model in controller or observer canonical form.
$B_{ij}(z^{-1})$ , $B(z^{-1})$	Polynomial in prediction error model.
$B_2(\underline{s}_\theta)$ , $B_{G,2}^{ob}(\underline{s}_\theta)$	Input connection matrix in LFT realization for a process model $G(z^{-1}, \underline{\theta})$ or a noise model $H(z^{-1}, \underline{\theta})$ in controller or observer canonical form.
$B_{G,2}^{ob}(\underline{s}_\theta)$ , $B_{H,2}^{co}(\underline{s}_\theta)$	
$B\Delta_{cn}$	Bounded or normalized set $\Delta_{cn}$ .
$B\Delta_{cm}$	Bounded or normalized set $\Delta_{cm}$ .
$B\Delta_{rn}$	Bounded or normalized set $\Delta_{rn}$ .
$C(\underline{\theta})$ , $C^{co}(\underline{\theta})$ , $C^{ob}(\underline{\theta})$	Output matrix of state-space model in controller or observer canonical form.
$C_{ij}(z^{-1})$ , $C(z^{-1})$	Polynomial in prediction error model.
$C_2(\underline{s}_\theta)$ , $C_{G,2}^{ob}(\underline{s}_\theta)$	Output connection matrix in LFT realization for a process model $G(z^{-1}, \underline{\theta})$ or a noise model $H(z^{-1}, \underline{\theta})$ in controller or
$C_{G,2}^{ob}(\underline{s}_\theta)$ , $C_{H,2}^{co}(\underline{s}_\theta)$	

$\text{Cov}(\underline{\theta})$	observer canonical form.
$\text{Cov}_n(\underline{\theta})$	Parameter covariance matrix.
$\mathbb{C}^n$	Normalized parameter covariance matrix.
$D(\underline{\theta})$ , $D^{\text{co}}(\underline{\theta})$ , $D^{\text{ob}}(\underline{\theta})$	Complex n-dimensional space.
$D(z^{-1})$	Direct feedthrough matrix of state-space model in controller or observer canonical form.
$D_{12}^{\text{ob}}(\underline{s}_{\theta})$ , $D_{G,12}^{\text{ob}}(\underline{s}_{\theta})$	Polynomial in prediction error model.
$D_{G,12}^{\text{co}}(\underline{s}_{\theta})$ , $D_{H,12}^{\text{co}}(\underline{s}_{\theta})$	Output feedthrough connection matrix in LFT realization for a process model $G(z^{-1}, \underline{\theta})$ or a noise model $H(z^{-1}, \underline{\theta})$ in controller or observer canonical form.
$D_{21}^{\text{ob}}(\underline{s}_{\theta})$ , $D_{G,21}^{\text{ob}}(\underline{s}_{\theta})$	Input feedthrough connection matrix in LFT realization for a process model $G(z^{-1}, \underline{\theta})$ or a noise model $H(z^{-1}, \underline{\theta})$ in controller or observer canonical form.
$D_{G,21}^{\text{co}}(\underline{s}_{\theta})$ , $D_{H,21}^{\text{co}}(\underline{s}_{\theta})$	
$\Delta$	General uncertainty block in robust control design.
$\Delta^c(j\omega)$	Normalized frequency dependent complex uncertainty matrix, $\sigma^{\max}(\Delta^c(j\omega)) \leq 1$ .
$\Delta_G^{\text{co}}$ , $\Delta_G^{\text{ob}}$ , $\Delta_H^{\text{co}}$	Parametric uncertainty block for a process model $G(z^{-1}, \underline{\theta})$ or a noise model $H(z^{-1}, \underline{\theta})$ in controller or observer canonical form.
$\Delta\Phi(\underline{\theta})$	Uncertainty regression matrix in bounded error description.
$\Delta\phi(k, \underline{\theta})$ , $\Delta\phi(\ell, \underline{\theta})$	Row of $\Delta\Phi(\underline{\theta})$ corresponding with time instant $k$ in the scalar case or index $\ell$ in the multivariable case.
$\Delta\phi_e(\underline{\theta})$	Matrix containing prediction error uncertainty $\underline{\delta}_e(k, \underline{\theta})$ .
$\Delta\phi_u$	Matrix containing input data uncertainty $\underline{\delta}_u(k)$ .
$\Delta\phi_v(\underline{\theta})$	Matrix containing auxiliary data uncertainty $\underline{\delta}_v(k, \underline{\theta})$ .
$\Delta\phi_w(\underline{\theta})$	Matrix containing auxiliary data uncertainty $\underline{\delta}_w(k, \underline{\theta})$ .
$\Delta\phi_y$	Matrix containing output data uncertainty $\underline{\delta}_y(k)$ .
$\Delta_{\text{cn}}$	Set of uncertainty blocks of complex-valued numbers.
$\Delta_{\text{cm}}$	Set of uncertainty blocks of complex frequency dependent matrices.
$\Delta_{\text{rn}}$	Set of uncertainty blocks of real-valued numbers.
$\mathbf{E}^k$	Bounding ellipsoid set obtained from first $k$ data.
$E$	Expectation operator.
$\emptyset$	Empty set.
$F(z^{-1})$	Polynomial in prediction error model.
$F_{\text{trend}}(z^{-1})$	Second order detrending filter.
$G(z^{-1}, \underline{\theta})$	Process model in polynomial model representation.
$\Gamma$	Gamma function.
$H(z^{-1}, \underline{\theta})$	Noise model in polynomial model representation.
$\mathbf{H}_1$ , $\mathbf{H}_2$	Hyperplanes defining parameter region $\mathbf{S}^k$ due to a new



$H_{aa}(\underline{\theta})$ , $H_{ab}(\underline{\theta})$ , $H_{ac}(\underline{\theta})$	observation at time instant $k$ .
$H_{ad}(\underline{\theta})$ , $H_{af}(\underline{\theta})$	Second-order derivative (Hessian) information of prediction error with respect to the minimal polynomial parameters $\{a, b, c, d, f\}$ , i.e. the Hessian tensor $H_{ab}(\underline{\theta})$ contains the information $\partial^2/\partial a \partial b$ .
$H_{bb}(\underline{\theta})$ , $H_{bc}(\underline{\theta})$ , $H_{bd}(\underline{\theta})$	
$H_{bf}(\underline{\theta})$ , $H_{cc}(\underline{\theta})$ , $H_{cd}(\underline{\theta})$	
$H_{cf}(\underline{\theta})$ , $H_{dd}(\underline{\theta})$ , $H_{df}(\underline{\theta})$	
$H_{ff}(\underline{\theta})$	
$I^n$	Identity matrix of dimension $n$ .
$J(\underline{\theta})$	Complete Jacobian matrix containing first-order derivatives of prediction error with respect to the parameter vector $\underline{\theta}$ .
$J_a(\underline{\theta})$ , $J_b(\underline{\theta})$ , $J_c(\underline{\theta})$	Jacobian matrix containing first-order derivatives of the prediction error with respect to the minimal polynomial parameters $\{a, b, c, d, f\}$ .
$J_d(\underline{\theta})$ , $J_f(\underline{\theta})$	
$K$	Controller block in robust control design.
$L(z^{-1})$	Prefilter for prediction error minimization.
$\Lambda(\underline{\theta})$	Covariance matrix of innovations.
$M(\underline{\theta}, \underline{y}(k), \underline{u}(k), \underline{\xi}(k), k)$	General model representation.
$M_a$	Augmented model in robust control design.
$M_c^{\text{co}}(\underline{\theta})$ , $M_c^{\text{ob}}(\underline{\theta})$	Nominal matrix of state-space model containing centre values of parameters where the superscript indicates a controller or observer canonical form.
$M_c^{\text{ob}}(\underline{\theta})$	
$M_{\text{delay}}$	Delay matrix between inputs and outputs.
$M_{\text{delay}}^r$	Realizable delay matrix.
$M_{\delta}^{\text{co}}(\underline{s}_{\theta})$ , $M_{\delta}^{\text{ob}}(\underline{s}_{\theta})$	Uncertainty matrix of state-space model in controller or observer canonical form.
$M_{ss}^{\text{co}}(\underline{\theta})$ , $M_{ss}^{\text{ob}}(\underline{\theta})$	Matrix representation of state-space model in controller or observer canonical form.
$M_{ss}^{\text{ob}}(\underline{\theta})$	
$M_{vs}$	Matrix containing vertex coordinates in EPU.
$M_{vp}$	Matrix containing vertex-plane adjacency information.
$M_{vv}$	Matrix containing vertex-vertex adjacency information.
$N$	Number of data samples.
$N_{\text{col}}$	Number of data samples in column form.
$\nabla_{\underline{\theta}} V_e(\underline{\theta}) =: G(\underline{\theta})$	Gradient of the loss-function $V_e(\underline{\theta})$ .
$\nabla_{\underline{\theta}\underline{\theta}}^2 V_e(\underline{\theta}) =: H(\underline{\theta})$	Hessian of the loss-function $V_e(\underline{\theta})$ .
$P^k$	Matrix specifying orientation and size of ellipsoid $E^k$ .
$\Phi(\underline{\theta})$	Regression matrix in pseudo-linear model description.
$\Psi(\underline{\theta})$	Regression matrix in bounded error identification.
$R(\underline{\theta})$	Part of Hessian matrix containing first-order derivative (Jacobian) information with respect to prediction error.
$R_e(\underline{\theta})$	Sample covariance matrix of prediction error $e(k, \underline{\theta})$ .

$\mathbb{R}^n$	Euclidean n-dimensional space.
$S(\underline{\theta})$	Part of Hessian matrix containing second-order derivative information with respect to prediction error.
$S_{aa}(\underline{\theta})$ , $S_{ab}(\underline{\theta})$ , $S_{ac}(\underline{\theta})$ $S_{ac}(\underline{\theta})$ , $S_{ad}(\underline{\theta})$ , $S_{af}(\underline{\theta})$ $S_{bb}(\underline{\theta})$ , $S_{bc}(\underline{\theta})$ , $S_{bd}(\underline{\theta})$ $S_{bf}(\underline{\theta})$ , $S_{cc}(\underline{\theta})$ , $S_{cd}(\underline{\theta})$ $S_{cf}(\underline{\theta})$ , $S_{dd}(\underline{\theta})$ , $S_{df}(\underline{\theta})$ $S_{ff}(\underline{\theta})$	Elements of $S(\underline{\theta})$ containing second-order derivative information with respect to the minimal polynomial model parameters $\{a, b, c, d, f\}$ multiplied with the prediction error, i.e. the matrix $S_{ab}(\underline{\theta})$ consists of the tensor $H_{ab}(\underline{\theta})$ multiplied with the prediction error.
$S_u$	Scaling matrix for input data, $U_{\text{scale}} = S_u U$ .
$S_y$	Scaling matrix for output data, $Y_{\text{scale}} = S_y Y$ .
$S^k$	Parameter region limited by two parallel hyperplanes when processing the $k^{\text{th}}$ data sample.
$S_t$	True system.
$T_{\text{sample}}$	Sampling time.
$T_{\text{trend}}$	Trend period.
$\Theta$	Parameter matrix.
$\Theta$	Parameter uncertainty set.
$\Theta_e$	Ellipsoidal bounding of the parameter uncertainty set $\Theta$ .
$\Theta_{eo}$	Ellipsoidal-aligned orthotopic bounding of the parameter uncertainty set $\Theta$ .
$\Theta_o$	Orthotopic bounding of the parameter uncertainty set $\Theta$ .
$\Theta_{pe}$	Parameter set over which $\underline{\theta}$ ranges in prediction error estimation.
$U$	Input data matrix of dimension $(N \times nu)$ .
$U_{\text{scale}}$	Scaled input data, $U_{\text{scale}} = S_u U$ .
$V_{E^k}$	Volume of ellipsoid $E^k$ .
$V_e(\underline{\theta})$	Loss-function for prediction error minimization.
$\Xi$	White noise matrix of dimension $(N \times ny)$ .
$Y$	Output data matrix of dimension $(N \times nu)$ .
$Y_{\text{scale}}$	Scaled output data, $Y_{\text{scale}} = S_y Y$ .
$a_p$	Coefficient of polynomial $A(z^{-1})$ .
$\alpha$	Step length in prediction error minimization.
$b_{ij,p}$ , $b_p$	Coefficient of polynomial $B_{ij}(z^{-1})$ , $B(z^{-1})$ .
$\underline{b}$	Vertex updating vector.
$c_{ii,p}$ , $c_p$	Coefficient of polynomial $C_{ii}(z^{-1})$ , $C(z^{-1})$ .
$c_1(\ell)$ , $c_2(\ell)$	Observation values in bounded error identification.
$d_p$	Coefficient of polynomial $D(z^{-1})$ .
$\delta^f$	Normalized real-valued uncertainty, $\delta^f \in [-1, 1]$ .
$\delta^c$	Normalized complex-valued uncertainty, $\delta^c \in \mathbb{C}$ , $ \delta^c  \leq 1$ .

---

$\underline{e}(k, \underline{\theta})$	Prediction error vector at time $k$ corresponding to $\underline{\theta}$ .
$\underline{\varepsilon}(k, \underline{\theta})$	Term in prediction error due to undermodelling.
$f_p$	Coefficient of polynomial $F(z^{-1})$ .
$\gamma_1, \gamma_2$	Tolerance margins for termination of prediction error minimization.
$j$	Complex number.
$k$	Integer-valued time variable for discrete-time models.
$k_F$	Gain of detrending filter $F_{\text{trend}}(z^{-1})$ .
$\ell$	Time and output dependent index in minimal polynomial model description.
$mc$	Number of complex parametric uncertainty blocks.
$mf$	Number of full matrix uncertainty blocks.
$mk$	Structure of general uncertainty block.
$mr$	Number of real parametric uncertainty blocks.
$n$	Maximum delay required for initialization of Jacobian (first-order derivative) matrix.
$na$	Number of parameters in polynomial $A(z^{-1})$ .
$nb_{ij}, nb$	Number of parameters in polynomial $B_{ij}(z^{-1}), B(z^{-1})$ .
$nc_{ji}, nc$	Number of parameters in polynomial $C_{ji}(z^{-1}), C(z^{-1})$ .
$nd$	Number of parameters in polynomial $D(z^{-1})$ .
$n\Delta_G^{\text{co}}, n\Delta_G^{\text{ob}}, n\Delta_H^{\text{co}}$	Dimension of parametric uncertainty block for a process model $G(z^{-1}, \underline{\theta})$ or a noise model $H(z^{-1}, \underline{\theta})$ in controller or observer canonical form.
$nf$	Number of parameters in polynomial $F(z^{-1})$ .
$nk_{ij}$	Delay in prediction error model from the input $u_j(k)$ to the output $y_i(k)$ .
$nn$	Array describing the minimal polynomial model structure.
$ns$	Maximum delay required for initialization of Hessian (second-order derivative) matrix.
$n\theta$	Dimension of parameter vector $\underline{\theta}$ .
$nu$	Number of inputs.
$ny$	Number of outputs.
$\omega$	Frequency.
$\omega_{\text{trend}}$	Trend frequency.
$p$	Pole in second order trend filter $F_{\text{trend}}(z^{-1})$ .
$\underline{p}$	Direction vector.
$\partial/\partial\theta_i$	Partial derivative with respect to $\theta_i$ .
$\Phi(k, \underline{\theta}), \Phi(\ell, \underline{\theta})$	Row of $\Phi(\underline{\theta})$ corresponding with time instant $k$ in the scalar case or index $\ell$ in the multivariable case.
$\Phi_e(\underline{\theta})$	Data matrix containing prediction error $\underline{e}(k, \underline{\theta})$ .

$\phi_u$	Data matrix containing input data $\underline{u}(k)$ .
$\phi_v(\underline{\theta})$	Data matrix containing auxiliary data $\underline{v}(k, \underline{\theta})$ .
$\phi_w(\underline{\theta})$	Data matrix containing auxiliary data $\underline{w}(k, \underline{\theta})$ .
$\phi_y$	Data matrix containing output data $\underline{y}(k)$ .
$\psi_1(\ell, \underline{\theta})$ , $\psi_2(\ell, \underline{\theta})$	Regression rows in bounded error identification.
$r_{uy}(\tau)$	Normalized cross-correlation function between the signals $\underline{u}$ and $\underline{y}$ for time lag $\tau$ .
$\rho$	Correlation coefficient.
$\rho_s$	Summed parameter correlations.
$\tau$	Time lag in cross-correlation function.
$\underline{\theta}$	Parameter vector of dimension $n\theta$ .
$\underline{u}(k)$	Input vector of dimension $n_u$ at time instant $k$ .
$\underline{v}(k, \underline{\theta})$	Auxiliary vector of dimension $n_y$ at time instant $k$ .
$\underline{w}(k, \underline{\theta})$	Auxiliary vector of dimension $n_y$ at time instant $k$ .
$\xi(k)$	White noise vector of dimension $n_y$ at time instant $k$ .
$\underline{y}(k)$	Output vector of dimension $n_y$ at time instant $k$ .
$\hat{\underline{y}}(k \underline{\theta})$	Predicted output at time $k$ using the model $M(\underline{\theta})$ .
$z^{-1}$	Backward shift operator.
$\zeta$	Damping ration in trend filter $F_{\text{trend}}(z^{-1})$ .

### Abbreviations

ARMAX	AutoRegressive Moving Average model with eXternal input.
ARX	AutoRegressive model with eXternal input.
BJ	Box-Jenkins model structure.
EE	Equation-Error model structure.
EOPB	Ellipsoidal-aligned Orthotopic Bounding algorithm.
EPB	Ellipsoidal Parameter Bounding algorithm.
EPC	Ellipsoidal bounding with Parallel Cuts algorithm.
EPU	Exact Polytope Updating algorithm.
FIR	Finite Impulse Response model.
LFT	Linear Fractional Transformation.
LP	Linear Programming.
MEB	Modified Ellipsoidal Bounding algorithm.
MFD	Matrix Fraction Description.
MIMO	Multi-Input Multi-Output system.
MISO	Multi-Input Single-Output system.
MPI	Minimal Polynomial Identification algorithm.

---

OPB	Orthotopic Parameter Bounding algorithm.
OE	Output-Error model structure.
OMNE	Outlier Minimal Number Estimator.
PRBNS	Pseudo Random Binary Noise Sequence.
SIMO	Single-Input Multi-Output system.
SISO	Single-Input Single-Output system.
SNR	Signal-to-Noise Ratio.

### **Notational Conventions**

$X$	Matrix (capital characters).
$\mathbf{X}$	Sets (bold characters).
$X^{-1}$	Inverse of a matrix $X$ .
$X^T$	Transpose of a matrix $X$ .
$X_{ij}$ , $X_{var,ij}$ , $[X_{var}]_{ij}$	$ij^{th}$ element of matrix $X$ and $X_{var}$ respectively.
$X_{i*}$	$i^{th}$ row of matrix $X$ , also denoted by $\underline{x}^T(i)$ .
$X_{*j}$	$j^{th}$ column of matrix $X$ .
$\underline{x}$	Vector (underlined characters).
$x$	Scalar (small characters).
$x_i$ , $x_{var,i}$ , $[x_{var}]_{ij}$	$i^{th}$ element of the vector $\underline{x}$ and $\underline{x}_{var}$ respectively.
$\underline{x} \in As \ N(m,C)$	Sequence of random variables $x$ converges in distribution to the normal distribution with mean $m$ and covariance matrix $C$ .
$\underline{x} \in As \ \chi^2(n)$	Sequence of random variables $x$ converges in distribution to the $\chi^2$ -distribution with $n$ degrees of freedom.
$x^k$	Value of $x$ at $k^{th}$ iteration.
$x_c$	Central value of set $x$ , $x_c = (x_{max} + x_{min})/2$ .
$\delta_x$	Uncertainty in the set $x$ , $\delta_x = (x_{max} - x_{min})/2$ .
$s_x$	Scaling factor to realize normalized uncertainty.
$x_{max}$	Maximum value of $x$ .
$x_{min}$	Minimum value of $x$ .
$\underline{x}_{col}$	A column vector formed by stacking the columns of the matrix $X$ on top of each other.
$\hat{x}$	Estimated value of $x$ .
$x_\infty$	Expected value of $x$ for $N \rightarrow \infty$ , $E\{x\}$ .
$x^*$	Limiting estimate of $x$ for $N \rightarrow \infty$ .
$\tilde{x}$	Weighted value of $x$ .
$x_t$	True value of $x$ .
$ x $	Absolute value of $x$ .

$\ \underline{x}\ _\infty$	$l_\infty$ -norm of vector $\underline{x}$ .
$\sigma_x$	Standard deviation of signal $\underline{x}$ .
$\sigma_x^2$	Variance of signal $\underline{x}$ .
$\text{adj}(X)$	Adjoint of a matrix $X$ , $\text{adj}(X) \triangleq X^{-1} \det(X)$ .
$\arg \min f(x)$	Value of $x$ that minimizes $f(x)$ .
$\det(X)$	Determinant of the matrix $X$ .
$\text{diag}(X)$	Vector containing the diagonal elements of the matrix $X$ .
$\text{sgn}(x)$	Sign value of $x$ .
$\text{tr}(X)$	Trace (sum of the diagonal elements) of the matrix $X$ .

# ***Introduction***

---

1.1	General Introduction	1.4	Scope of this Thesis
1.2	Motivation	1.5	What's New
1.3	Problem Statement		

---

Identification is a fascinating field of research and although a lot of work has been done in the past, many problems remain to be solved. But what is identification? Identification can be described as the scientific discipline of constructing a mathematical model of a dynamical system from observations and prior knowledge [Lju87, Nor86, SS89]. Many questions are raised by this definition. Why modelling? What are dynamical systems? What kind of observations and prior knowledge are used and finally how can these models be constructed? Important issues in identification when describing a process are complexity, accuracy, etc.

## ***1.1 General Introduction***

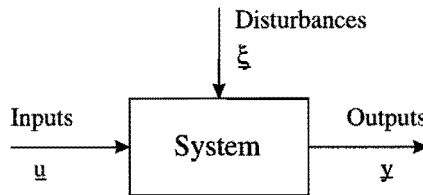
The aim of modelling is to describe the behaviour of a dynamical system or process by discovering the relations between the observations to understand the mechanisms driving

the process. Dynamical models used for control design, diagnosis, prediction, etc., have an extensive application area ranging from engineering, economics and medicine to ecology and agriculture.

The model of a process describes the dynamical behaviour in some way. In general, assumptions are made about the dynamical system, the disturbances affecting the process, the type of models selected to describe the process and finally the criterion used in the identification procedure. Some important distinctions can be made which significantly influence the model and its application in practice.

### 1.1.1 Dynamical Systems

A dynamical system can be described as depicted in Fig. 1.1



**Fig. 1.1** : Dynamical system.

The system is driven by external signals where the inputs  $\underline{u}$  can be manipulated and the disturbances  $\underline{\xi}$  can only be observed through their influence on the outputs  $\underline{y}$ . These vectors are time signals where each entry corresponds to a channel of the system. An essential property of dynamical systems is that the outputs at any time instant depend on the past and not just on the present inputs. This is related to the memory structure of the system. Hence, the past behaviour of a dynamical system influences the future as a result of the memory length and its initial conditions. Some main properties can be distinguished when describing a dynamical system.

- *Causal vs. noncausal* : The current behaviour of causal systems does not depend on the future but only on current and past information which is always the case for physical systems.
- *Time-invariant vs. time-varying* : A dynamical system is time-invariant if its properties do not change with time. A time-varying system, however, may be preferred if the time-varying properties are relevant for the process description.



- *Linear vs. nonlinear* : Linear systems have a nice superposition property which simplifies the mathematical analysis. Nevertheless, practical systems are in general nonlinear.
- *Continuous-time vs. discrete-time* : For many physical systems it is natural to work with a continuous-time representation. Often, however, since measurements are only available at discrete time instants, discrete-time models are used for many applications, e.g. control design.
- *Single-input single-output vs. multi-input multi-output* : The identification of SISO systems has been described extensively in many textbooks [Lju87, Nor86, SS89] and is very well known. Most practical systems however have a multivariable character for which the identification is much more complicated and less extensively explored.

### 1.1.2 Disturbances

Disturbances affecting the process will highly influence the modelling and therefore a priori assumptions on the noise are required in order to describe the process. Boundedness, stationarity, ergodicity, whiteness, Gaussianity, etc. are all properties which are used to characterize the noise together with the way the noise is assumed to affect the process.

- *Measurement noise* : The sensors which measure the signals are subject to noise and drift. Information to characterize this type of noise can be obtained by separate sensor analysis or by specific experiment design.
- *Uncontrollable inputs* : Signals which can be measured but not controlled by the user can be considered as disturbances.
- *Process noise* : Disturbances affecting the process internally are in general difficult to handle because no information is available besides the measured data.

### 1.1.3 Models and Modelling

The model of the dynamical system depicted in Fig. 1.1 should describe the behaviour of this process as accurately as possible within its intended use. The application of the model determines however, whether the internal behaviour or the input-output behaviour only should be described accurately and which measure can be used to define this accuracy.

- *Input-output models vs. state-space models* : State-space models contain state variables as intermediaries for the description of the relationships among inputs and outputs. The dynamics can be expressed as first-order differential equations in the state variable. These models have natural extensions to multivariable systems at the expense of more parameters compared to polynomial input-output models for which, however, the so called structure indices [HD88] in the multivariable canonical representation are difficult to estimate in practice.
- *Parametric vs. nonparametric* : In nonparametric models the dynamical behaviour is directly specified by the measured value, e.g. impulse-response. The benefit of parametric models are the relatively few parameters (parsimony) required to describe the dynamical behaviour.
- *Time-domain vs. frequency-domain* : Linear, time-invariant systems can either be described by differential equations in the time-domain or by transfer functions in the frequency domain.
- *Physical vs. black-box modelling* : Physical models are often based on first principles and therefore difficult to construct with the advantage that the parameters have a direct and physical interpretation. On the other hand, black-box models can be constructed relatively easily but with structure quantities that are not necessarily compatible with physical reality.

#### 1.1.4 System Identification

In the area of system identification, different approaches can be chosen to model the process :

- *Deterministic vs. Stochastic* : A variety of identification methods exist that range from treating the model coefficients as random variables and modelling the errors in some detail to describing the uncertainty in the measurements and the model by bounds using no statistics at all. The resulting models are defined respectively stochastic, i.e. probabilistic in time, and deterministic, i.e. certain.

This brief description of an immense variety in practical systems, disturbance and model properties together with different identification procedures clearly indicates the problem dependent character of modelling. It requires an "engineering approach" to obtain accurate and useful models. Within this variety, the construction of a model can roughly be divided in five phases :

- *The data* : The input-output data is recorded during specifically designed identification experiments where it has been defined a priori which input signals

will be manipulated and which output signals have to be measured. The object of experiment design is to include as much as possible dynamical information in the data within the physical constraints imposed by the process. The next step in practice is data preprocessing (detrending, peakshaving, delay correction and normalization) to obtain data which is suitable for identification.

- *The model set* : The intended use of the model together with available a priori knowledge and engineering intuition and insight have to be invoked to select a set of candidate models and an error criterion to define the accuracy of the models. These most difficult choices in the identification procedure determine in fact the achievable performance of the model.
- *Parametrization* : Within the selected model set, the parametrization involves issues like selecting a suitable model structure and an appropriate model order.
- *The optimization* : The identification method based on optimization defines the construction of the best model within the parametrized model set based on the available data and the chosen criterion which is the most straightforward phase.
- *The model validation* : The final judgment of the model quality is given in the model validation phase relating the model behaviour to the a priori knowledge, to observed data independent of the data set used for identification and to its intended use.

### 1.1.5 Practicalities

Although a nice description has been given of the various steps in identification, there are some severe limitations in practice which influence the maximum achievable performance of the model significantly :

- *System access* : The amount of data available for system identification will be limited, e.g. historical records in economics or operating records in industrial processes. In bulk production the access to the system could be rather restricted because of economic reasons.
- *Available time* : The duration of experiments in production processes depends strongly on its bandwidth which defines the time required to collect sufficient data ranging from milliseconds to days. So identification can be costly due to loss of production during experiments and data acquisition.
- *Sensors and actuators* : The nominal value, rate of change and smoothness of signals in the system to changes of setpoint inputs are limited by the input actuators and the system. The amplitude of a measured output response to setpoint inputs is normally limited by the process to ensure usable products and by the output sensors which have a restricted measurement range. The sampling

rate may be limited by the data acquisition equipment or by the time needed to collect data. These constitute important aspects in identification.

- *Computing facilities* : Although lack of computing power is not really a constraint anymore nowadays, on-line updating of the model as measurements are taken compared to off-line model optimization are still important and interesting issues.
- *Accuracy and effort* : The effort to obtain more and more accurate models of production processes is certainly limited by the question of the additional product improvement when controlling the system. Complexity of models increases as accuracy increases. Therefore, the economic benefit needs to be compared to the extra tools, experiment time and manpower (costs) required to achieve this goal.

All these aspects show that no straightforward procedure can be given for identification of practical processes. Only basic tools can be provided to solve the problems in the various identification steps. The different choices, however, in experiment design, model set and error criterion selection, parametrization and model validation have to be made by the user and will be highly problem dependent.

Within this broad setup, a general model of a dynamical system depicted in Fig 1.1 relating the inputs  $\underline{u}$ , the measured outputs  $\underline{y}$  and the disturbances  $\underline{\xi}$  in time  $t$ , can be described by :

$$M(\underline{\theta}(t), \underline{y}(t), \underline{u}(t), \underline{\xi}(t), t) = \underline{0} \quad (1.1)$$

where  $\underline{\theta}$  defines the parametrization of the model  $M$ . The modelling of dynamical systems however is problem dependent, too complex and too diverse for a general treatment. More general identification procedures can only be given if the dynamical system under study can be described by a restricted class of models, e.g. discrete, linear, and time-invariant. In this case the system description of Eq. 1.1 reduces to :

$$M(\underline{\theta}) \begin{pmatrix} \underline{y}(k) \\ \underline{u}(k) \\ \underline{\xi}(k) \end{pmatrix} = \underline{0} \quad \forall k \quad (1.2)$$

where the signals  $\{\underline{y}(k), \underline{u}(k), \underline{\xi}(k)\}$  at time instants  $k$  have been extracted in a linear, time-invariant way and  $M(\underline{\theta})$  defines now the model of the system which is somehow parametrized in  $\underline{\theta}$ . For example  $M(\underline{\theta})$  may be a polynomial model representation.

## 1.2 Motivation

Many different types of processes can be found in industry which all require some sort of control to meet the product specifications. To control the dynamics of a process knowledge is needed about the dynamical behaviour. The construction of a model by means of physical, mechanical, electrical and chemical laws and additional assumptions on the behaviour of the real process is often too complex for the process industry if not impossible. In general, this approach leads to very complex models with many unknown quantities which are difficult to estimate in practice. Therefore, these models are often of limited use for control design. On the other hand, black-box identification methods can be applied relatively easily to obtain accurate models which are suitable for control design. However, the parameters have no direct physical interpretation. Nevertheless, a better name for these identification techniques would be grey-box modelling because often an extensive use of physical knowledge has been included during the various identification steps.

A considerable amount of literature exists in this field [Lju87, Nor86, SS89, etc.] if the process is considered to be linear and time-invariant. This is not really a restriction in practice since a lot of processes can be described sufficiently accurate by linear time-invariant models, at least around an operating point. It should be noted, however, that most of these identification techniques are limited to SISO systems. For the purpose of analysis, the assumption is required that there exists a model which can describe the process exactly. In other words, it is assumed that the process is contained in the model set. This assumption is not valid in practice. Although a model may describe the physical system accurately, any model can at best be an approximation of the physical system. There is always some "uncertainty" present even when the underlying process would be essentially linear. In the last decade now, robust control theory has been a major research activity [DGK89, PD93]. Developments in robust control theory are providing the engineer with the capability for systematically handling models with increasingly sophisticated uncertainty descriptions. This offers the possibility to include system uncertainty into the control design.

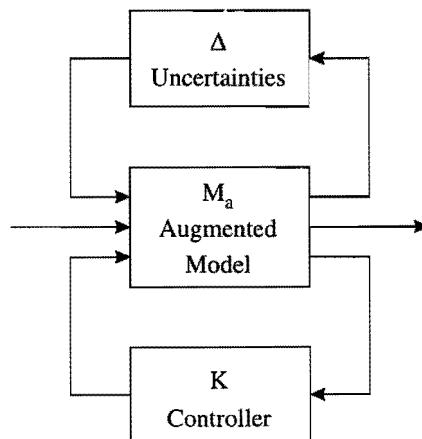
### 1.2.1 Robust Control

The general framework for robust control design ( $H_\infty$  and  $\mu$ -control) which is used in the literature is illustrated in Fig. 1.2. Any interconnection of inputs, commands, perturbations and a controller can be rearranged to match this figure. The exogenous input is a signal entering the system and is typically used to model disturbances, commands and noise. It is generally inadequate for control design to represent plant uncertainty only in the form of uncertain additive signals. The system model itself

typically has uncertainty which can have a significant impact on the system performance. This uncertainty is a consequence of unmodelled dynamics and parameter variations and is modelled as the perturbation  $\Delta$  to the nominal augmented model  $M_a$ . This augmented model contains the basic process to be controlled and the interconnection structure which has been defined to meet the control design requirements. This includes design filters which define performance specifications as well as the scaling factors for normalized block-diagonal perturbations which define the robustness specifications.

Note that the uncertainty modelled as  $\Delta$  has an effect very different from that of the exogenous inputs on the performance of the system. Perturbations  $\Delta$  can cause a nominally stable system to become unstable which an external signal cannot do.

At the heart of any theory about control are the assumptions made about the augmented model, the exogenous inputs and the perturbations, as well as the performance specifications on the outputs to be controlled. These assumptions determine the framework for the identification procedure (model structure selection, a priori knowledge) and the analysis methods which can be applied to obtain conclusions about the performance of the system under control.



**Fig. 1.2 :** General framework for robust control.

Nowadays, the synthesis and analysis theories can handle a rich class of uncertainty descriptions. However, it is still up to the designer to model the system appropriately in this more complex framework. Robust control design leads to controllers that have guaranteed performance and stability robustness with respect to all members of a limited

model set. The inclusion of physically unrealistic models in the model set can make the design conservative or even impossible because these models might determine the worst case system behaviour. To optimize the trade-off between necessary robustness and achievable performance the designer, therefore, requires a model set description which is as "tight" as possible, in that no physically unrealistic models are included, while describing all pertinent behaviours of the physical process.

### 1.2.2 Uncertainties

Many approaches can be taken to describe uncertainty associated with a multivariable model of a physical system. The uncertainty descriptions however, determine the trade-off between achievable performance and robustness of the control design. If a physical system is not within the set of plants described by the nominal and uncertainty models, the designed controller may cause instability or exhibit poor performance when implemented to the physical system. However, if the uncertainty descriptions are overly conservative, plants may be included in the set which limit the performance of the closed-loop system. Therefore it must be emphasized that no matter how attractive an uncertainty description may seem from a practical point of view, it is only useful if it permits the derivation of "tight" conditions for robust control.

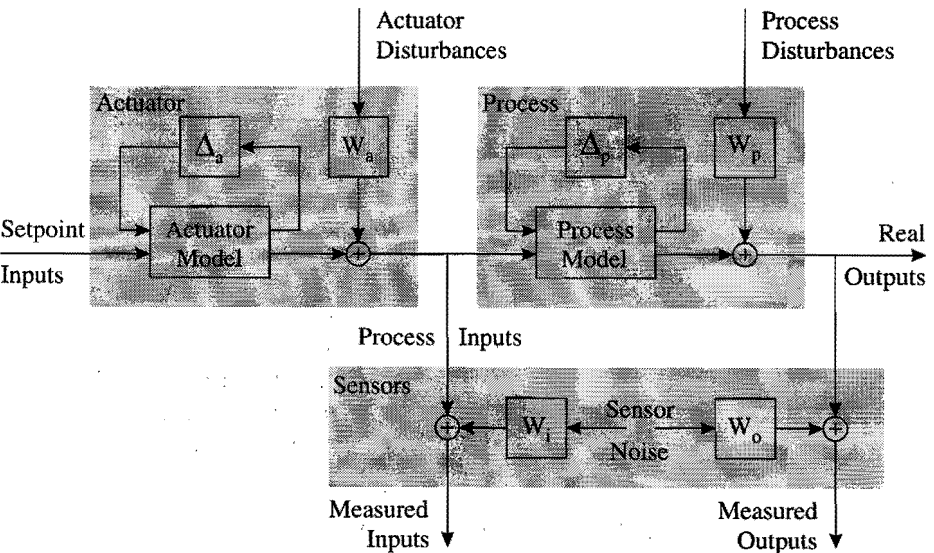


Fig. 1.3 : Modelling uncertainty and disturbances.

Basically, two types of uncertainty descriptions can be considered : "structured" and "unstructured" uncertainty.

- *Unstructured Uncertainty* : The uncertainty is expressed in terms of a single perturbation. It is assumed that the dynamic behaviour of a process is not described by a single linear time-invariant model but by a family of these models. One way of parametrizing this family are "Gershgorin bands" [Mac89] composed of "Gershgorin circles" of specific radius at each frequency in the Nyquist plane. The bounds derived using unstructured uncertainty [vdB93] are often conservative from a practical point of view since the actual uncertainty can rarely be lumped into a single norm bounded perturbation without including many more possible plants than actually needed. The problem of addressing uncertainty is especially present in multivariable systems. Generally, it is important to model uncertainty as detailed as possible where it occurs and not necessarily where it is convenient mathematically.
- *Structured Uncertainty* : The individual sources of uncertainty are identified and represented directly and independently. This leads to an uncertainty description with multiple perturbations which is in general non-conservative from a practical point of view. These perturbations may correspond to uncertainty in the model parameters, uncertainty with respect to manipulated variables (input or actuator uncertainty) and output variables (measurement uncertainty) etc. A schematic example is depicted in Fig. 1.3 where the highly structured uncertainty  $\Delta_a$  and  $\Delta_p$  should be represented independently in a block diagonal uncertainty when modelling the complete system. Including structure in the uncertainty description of the process will reduce the conservatism in the model set.

Parametrized plant perturbation sets can be used to model several different types of plant variations :

- *Component tolerances* (structured) : A controller  $K$  is to be designed for many more or less identical systems, for example, a mechanical system which consists of various components. The controller is designed on the basis of a nominal system and the parameter variations represent the differences in the individual manufactured components. Designing a controller that robustly achieves the design specifications avoids the need and cost of tuning each manufactured control system or allows larger variations in the individual manufactured components which can reduce production costs.
- *Component drift or aging* (structured) : A controller is designed for a system that is well modelled, but it is desired that the system should continue to work if the system to be controlled changes due to aging or drift in its components. This



avoids the need and cost of periodically re-tuning the control system.

- *Model parameter uncertainty* (structured) : In a model developed from physical principles, a parametrized plant model set may represent physical parameters such as lengths, masses, or heat conduction coefficients, and the bounds are the minimum and maximum values that could be expected to occur.
- *Externally induced changes* (unstructured) : The system to be controlled may be modelled as a linear time-invariant system that depends on an external operating condition, which changes slowly compared to the system dynamics, or is linearized around a specific operating point. If the variation range is not too large, designing a robust controller can avoid the need for a gain-scheduled or adaptive controller.

Uncertainty in the process can also be modelled as frequency dependent errors in the frequency responses of its entries. Such process perturbation sets can be used to account for :

- *Model uncertainty* (unstructured) : The plant transfer functions may inaccurately model the system to be controlled because of measurement or identification errors. For example, the transfer functions of the system to be controlled may have been measured at a finite number of frequencies with a limited accuracy.
- *Neglected dynamics* (unstructured) : If a simple model of the system to be controlled is needed, system dynamics which could in principle be modelled, are intentionally neglected. For example, a model of a mechanical system may be developed on the assumption that a drive train is rigid, an assumption that is good at low frequencies but poor at high frequencies. If the high frequency dynamics of this drive train could be accurately modelled or consistently measured, a more accurate and more complex model can be developed of the system to be controlled. However, it may be the case that these high frequency dynamics are very sensitive to minor physical variations in the system, like the ones that might be induced by temperature changes, bearing wear, load variations etc. In this case the drive train dynamics could be reasonably well modelled as an unknown transfer function that is close to one at low frequencies, and less close at high frequencies.
- *High frequency parasitic dynamics* (structured/unstructured) : The model of a system may become less accurate at high frequencies because of unknown or unmodelled parasitic dynamics. Moreover these parasitic dynamics may change with time or other physical dynamics. Consequently, they cannot be confidently modelled (e.g. elastic modes).

Uncertainty should not necessarily be described as rigorous or accurate as possible. Rather, an "engineering approach" should be taken to describe the uncertainty only as rigorous as necessary. This means, for example, that some sources of uncertainty (occurring in different places in the system) should be lumped into an "unstructured" perturbation, if this does not result in too much conservatism. This leads to a practical uncertainty description : some sources of uncertainty are described in a "structured" manner (e.g. parametric uncertainty), while the rest (usually uncertain high-frequency dynamics) are lumped into a single unstructured perturbation. This can be compared to determining the model order in standard identification techniques. The important question then is whether or not the improvement in the fit to the observed data is significant when the model order is increased.

### Example

To emphasize the conservatism which can occur in describing model uncertainty, the uncertainty of an actuator model [SDM87] will be used as an example. The actuator can be represented exactly as a system consisting of a single pole and a delay. So a nominal model can be defined as :

$$P = \frac{k e^{-j\omega\theta}}{1 + j\omega\tau} \quad (1.3)$$

For the structured actuator model set, the values of the gain, the pole position and the delay are assumed to be known within certain bounds :

$$k = k + s_k \delta_k \quad , \quad \theta = \theta + s_\theta \delta_\theta \quad , \quad \tau = \tau + s_\tau \delta_\tau \quad (1.4)$$

where  $s_i$  is a constant scaling factor and  $\delta_i$  is the normalized real-valued perturbation,  $\delta_i \in [-1,1]$ .

The following unstructured multiplicative output uncertainty description of the actuator transfer function (Eq. 1.3) is adopted :

$$P = \left\{ P ( I + W \Delta ) \mid \sigma_{\max}(\Delta) \leq 1 \right\} \quad (1.5)$$

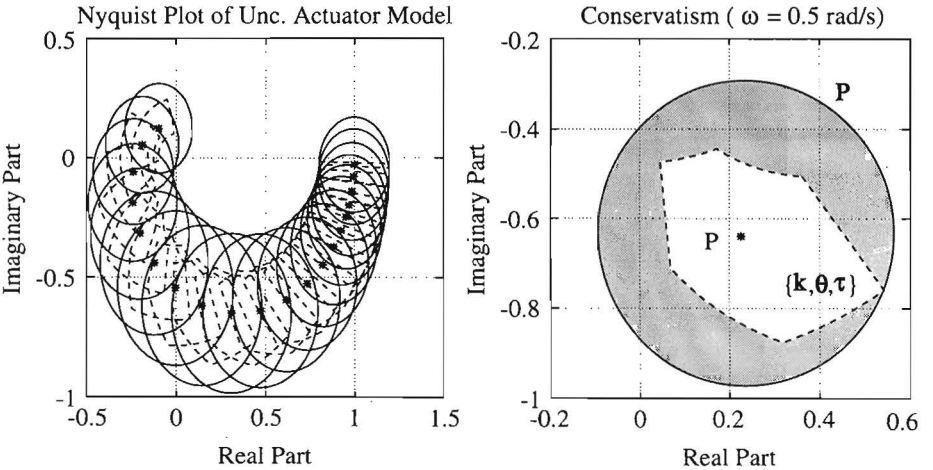
where the set  $P$  describes all plants,  $P$  is the nominal plant,  $W$  is a frequency dependent output uncertainty weight which should be taken as a stable and minimum phase rational transfer and  $\sigma_{\max}$  indicates the maximum singular value. The parametric uncertainty can then be represented as a single complex block  $\Delta$  while the corresponding stable, minimum phase and rational frequency weight  $W$  should fulfil the requirement :

$$\forall \omega : |W| \geq \sup_{\delta_k, \delta_\theta, \delta_\tau} \left| \left( \frac{k + s_k \delta_k}{k} \right) \left( \frac{1 + j\omega\tau}{1 + j\omega(\tau - s_\tau \delta_\tau)} \right) e^{j\omega s_\theta \delta_\theta} - 1 \right| \quad (1.6)$$

The following values of the actuator parameters have been taken :  $k = 1$ ,  $s_k = 0.2$ ,  $\theta = 0.75$ ,  $s_\theta = 0.25$ ,  $\tau = 2$ ,  $s_\tau = 0.5$ . If the delay is approximated by a Padé approximation the weight function  $W$  will become unstable. The unstable pole can be removed by multiplication by the appropriate allpass function. Using the bounds in Eq. 1.4 and a second order Padé approximation gives :

$$P = \frac{0.5s^2 - 2.67s + 7.11}{s^3 + 5.83s^2 + 16.89s + 7.11}, \quad W = \frac{0.6s^3 + 41.73s^2 + 100.26s + 17.07}{s^3 + 16.67s^2 + 138.66s + 85.33} \quad (1.7)$$

The Nyquist plot of this uncertain actuator model is depicted in Fig. 1.4. The shaded area in this figure indicates the conservatism in the Nyquist plane for a single frequency ( $\omega = 0.5$  rad/s) between the true structured uncertainty due to the parametric uncertainty  $\{k, \theta, \tau\}$  and the multiplicative unstructured approximation (Eq. 1.7). The conservatism is due to physically unrealistic models. This simple example illustrates the importance of including available physical knowledge.



**Fig. 1.4 :** Conservatism in uncertainty approximation.

Much of recent work on robust control is based on assumptions about uncertainties around the nominal model which are quite different from the type of information that is classically available about an identified process. A robust control model consisting of a nominal and an uncertainty model is more complex than the standard linear system transfer function model which has been generated from experimental input-output data, because the structure of the uncertainty as well as bounds on its size must be specified. The gap between a priori known descriptions of model uncertainties or model errors on which robust control theory is built and the failure of identification theory to deliver accurate uncertainty bounds of unmodelled dynamics and parametric variations, show the need for a better understanding of the interaction between both theories. This includes the development of new identification techniques which are capable of delivering accurate a priori knowledge needed for robust control.

### **1.3 Problem Statement**

Over the past years a start has been made to close the gap between identification and control design. Identification for robust control has become a very active research area aiming at the identification of models with error bounds suitable for robust control design [SD94]. As discussed before, the identification of a robust control model is more complex than finding standard linear system transfer functions because the structure of the uncertainty as well as bounds on its size must be specified. An engineer wishing to model a physical system within this framework is immediately faced with some problems :

- How to first select the system structure (e.g. specify the inputs, outputs and connections of the robust control framework which are necessary and/or available to solve the control design problem).
- How to specify bounds on the perturbations, characterize inputs by weighting filters and specify meaningful weightings for the output sets in order to meet the control design requirements.

Therefore, an identification methodology is required such that, given a priori (physical) knowledge, input-output experiments and some explicit assumptions on the system, the methodology gives a class of models with the disturbances defined in some class of specified uncertainties (see Fig. 1.2) which will lead to satisfactory control design.

Several methods have been developed to derive unstructured model error bounds either using a stochastic approach [GS90, Gev91, Zhu89] or a deterministic approach assuming bounded noise in the time domain [WL91, HJN91] or bounded noise in the frequency domain [LVS91, GK92, vdB93]. Set estimation methods provide structured uncertainty

bounds in the form of parametric uncertainty based on a deterministic approach with bounded noise in the time domain [Mil89, WPL90]. Note that classical identification techniques also provide structured parametric uncertainty information in a stochastic setting (Cramér-Rao bound, Fisher Information matrix), although various restrictive assumptions have to be satisfied (Gaussian distributed white noise with zero mean, uncorrelated with the input, process in the model set and the number of data samples tends to infinity) which are often violated in practice.

All these methods have different starting points, advantages, drawbacks and limitations, but they all require that the process is linear and time-invariant in order to solve the identification problem. This is not really a restriction in practice since a lot of processes can be described accurately by linear time-invariant models, at least around an operating point. Moreover, the immense variety in systems and models is too large for a general treatment of the identification problem, as stated in the general introduction. Therefore, the type of processes as well as the type of models which will be considered has to be restricted. The type of models which will be used to describe the linear, time-invariant dynamical systems, has been restricted to polynomial input-output models. These type of models are well known in the field of identification and provide sufficient flexibility in order to describe different possible systems.

For any given data set, a priori knowledge and a selected identification criterion, a large set of polynomial models can represent the dynamical system which is not falsified by the available information. The measure of suitability of these models will depend strongly on the design performance objectives. Given any input-output data, it is possible to attribute the discrepancies between a nominal model and the observed behaviour, for example, entirely to additive output noise or entirely to unstructured perturbations. In this context ambiguity enters easily the description due to uncertainty about the uncertainty. While the goal of good experiment design should be to reduce this ambiguity in the modelling process by identifying the individual sources of uncertainty and representing them directly, it is practically not possible to remove it entirely. Although it has been emphasized to include as much as possible a priori knowledge, this information must be available or has to be extracted from the physical system in some way. Since for industrial processes this specific information is usually not available, additional experiments have to be performed to obtain the required knowledge. In practice, however, this is often very difficult due to limited process access and available time. Therefore, in the case that no or not all required structure and uncertainty information can be obtained, necessarily the remaining uncertainty has to be lumped together.

In industrial applications, an important aspect is the required computing time. Although the computing power has been increased significantly the last few years, the required time to obtain an accurate model of the physical system by using an identification methodology has to be restricted to become industrially applicable. Due to economic

reasons, the introduction of new techniques depends strongly on the required computing time and the resulting benefit compared to existing techniques.

Combining all these aspects, the following problem statement can be formulated :

### **General Problem Statement**

*The development of industrially applicable identification techniques  
for multivariable systems resulting in models which are suitable  
for the design of robustly performing control systems.*

Taking the large variety in dynamical systems, disturbances, models, system identification and practicalities into account as described in this introduction, restrictions have to be imposed on this general problem statement in order to limit the area of research. Although in general processes are nonlinear, accurate approximations can be obtained by linear models describing the process around an operating point. Further, it will be assumed that the systems under consideration can be described as time-invariant, which is valid for a wide range of processes in industry. Combined with the fact that measurements are often only available at discrete time instants, the identification of dynamical systems in this thesis will be restricted to discrete, linear and time-invariant models. The identification will be performed in the time domain since the data is directly available in this form from the experiments and no transformations have to be performed to apply for example identification in the frequency domain. As model structure, a polynomial model representation will be adopted because of its simple and direct representation of dynamical systems in the time domain.

In identification for robust control design, the model uncertainty description is of crucial importance, especially for multivariable systems. The bounds derived using unstructured uncertainty are often conservative from a practical point of view since the actual uncertainty can rarely be combined into a single norm bounded perturbation without including many physically unrealistic models. Using structured uncertainty, i.e. model uncertainty as detailed as possible where it occurs, will reduce the conservatism. In polynomial models, the model uncertainty can be included as structured parametric uncertainty. This parametric uncertainty can either be stochastic or deterministic. Both approaches to obtain these types of parametric uncertainty will be compared.

The attention is especially focused on solving the multivariable identification problem and the circumstances under which reliable models and error bounds can be obtained. It should be noted that the uncertainty due to disturbances affecting the process,

measurement noise and the uncertainty due to approximate modelling in practice, will be lumped together into parametric uncertainty of the identified model. Although it would be better to separate these sources of uncertainty as discussed before, in general this goal cannot be realized because of limited a priori information, limited system access for additional experiments, etc.

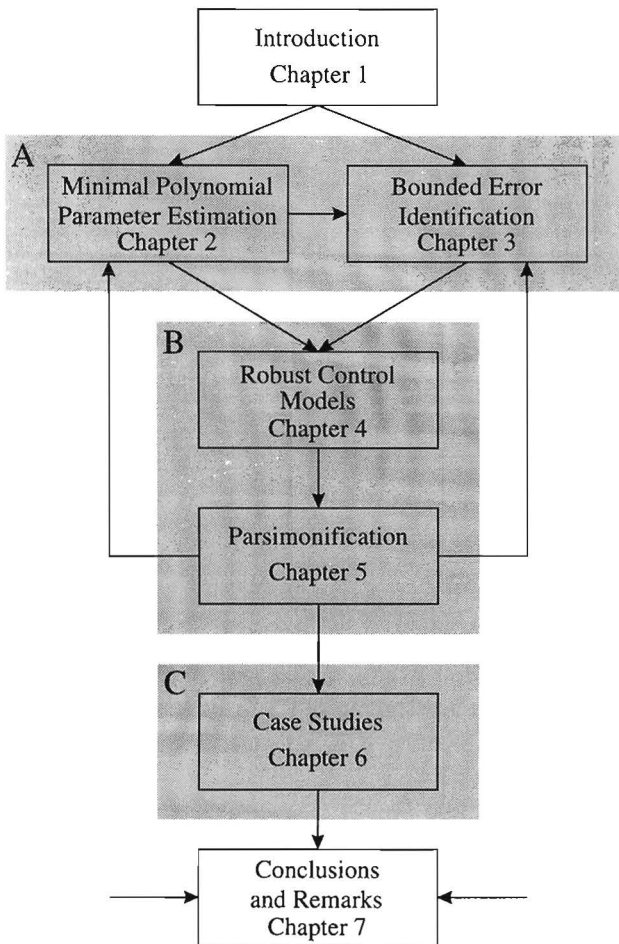
Linearity plays a crucial role in the stochastic and deterministic identification approaches which will be described in this thesis. The processes which can be modelled have been restricted to linear systems. Linearity in this context for the system setup depicted in Fig. 1.1, can be described as the linear mapping from the input and the disturbance samples  $\{\underline{u}(k), \xi(k)\}$  to the output samples  $\underline{y}(k)$  for a given parameter vector  $\underline{\theta}$  (see Fig. 1.5a), i.e.  $\{\underline{u}(k), \xi(k)\} \xrightarrow{\underline{\theta}} \underline{y}(k)$ . In the deterministic identification approach, it will additionally be required that the models are linear in their parameters. This type of linearity can be described as the linear mapping from the parameter vector  $\underline{\theta}$  to the disturbance samples  $\xi(k)$  for a given data set  $\{\underline{u}(k), \underline{y}(k)\}$  (see Fig. 1.5b), i.e.  $\underline{\theta} \xrightarrow{\{\underline{u}(k), \underline{y}(k)\}} \xi(k)$ . Note that basically a linear model can be nonlinear in its parameters, while a model which is linear in its parameters can describe nonlinear system behaviour.



**Fig. 1.5 :** Linearity ; a) Linear model  $\{\underline{u}(k), \xi(k)\} \xrightarrow{\underline{\theta}} \underline{y}(k)$  , b) Linear in the model parameters  $\underline{\theta} \xrightarrow{\{\underline{u}(k), \underline{y}(k)\}} \xi(k)$ .

## 1.4 Scope of this Thesis

This thesis consists basically of three parts (Fig. 1.6) : identification of multivariable systems resulting in polynomial models with parametric uncertainty (A), the transformation of these models into representations which are suitable for robust control design (B), and the application of the developed identification techniques to various systems (C). The symbols, abbreviations and notational conventions used throughout this thesis are defined in the glossary.



**Fig. 1.6 :** Scope of this thesis.

## **Chapter 2 : Minimal Polynomial Parameter Estimation**

In Chapter 2 the classical identification approach is described briefly. It is based on statistical assumptions of the noise disturbing the process. The standard polynomial model identification of SISO systems based on prediction error minimization is reviewed together with the basic problems related to extending this identification method to multivariable systems. An extended minimal polynomial model representation



is introduced for multivariable systems which explores the advantage of prediction error minimization and avoids the problem of structure identification. Appendix B provides a detailed description of the solution to this minimal polynomial identification problem. The covariance matrix of the parameter estimates can be approximated without the assumption that the process should be contained in the model set, as described in Appendix C. This quality measure of the estimated model can be used to determine confidence intervals of the corresponding parameters in a stochastic setting.

### ***Chapter 3 : Bounded Error Identification***

A pure deterministic approach based on a bounded error description of the uncertainty is described in Chapter 3 for SISO systems. The corresponding feasible parameter set is convex and connected if the polynomial models can be described as linear in their parameters. In this case the set is bounded by linear constraints. For models nonlinear in their parameters, e.g. output-error models, the parameter uncertainty set is in general non-convex and non-connected. However, linearization of the bounds describing the parameter uncertainty set can overcome this problem thereby avoiding excessive computational complexity but at the expense of introducing conservatism. Although the parameter uncertainty set is bounded now by linear constraints, exact computation of the feasible parameter set is only possible for systems which can be described by a few parameters. To further reduce the computational complexity, approximate descriptions of the parameter uncertainty set have to be used to determine the parameter bounds. For practical applications, however, special precautions are necessary because these set estimators are not robust against outliers, i.e. data points which are not consistent with the specified assumptions. The extension to multivariable models together with the various parameter uncertainty descriptions of the feasible parameter set is presented in Appendix D.

### ***Chapter 4 : Robust Control Models***

A minimal polynomial model representation has been used to describe multivariable dynamical systems. For robust control design, however, state-space models are required. Once a polynomial model with parametric uncertainty has been obtained, the corresponding state-space representation can be derived relating the identification method to the robust control framework. This is subject of discussion in Chapter 4.

### ***Chapter 5 : Parsimonification***

When applying identification to multivariable processes, the number of parameters which are required to describe the system grows rapidly, especially for increasing

number of inputs and outputs. For robust control design, however, the number of uncertain parameters should be restricted. In Chapter 5 the attention is focused on reducing the number of uncertain parameters by lumping the uncertainty together into a manageable amount of parameters.

## **Chapter 6 : Case Studies**

Chapter 6 covers the evaluation of the developed identification techniques. For single-input single-output systems a watervessel laboratory process and a fed-batch fermentation process for yeast production have been used as test cases. Since identification becomes more complex for multivariable systems, a simulation model of a distillation column gives a better understanding of the relation between signal-to-noise ratios, model complexity and parameter uncertainty. For testing the identification techniques in a multivariable industrial environment, the methods have been applied to a glass tube production process. The necessary preprocessing of the practical data is described in Appendix A.

## **Chapter 7 : Conclusions and Remarks**

Finally, conclusions and remarks are given in Chapter 7. Several identification techniques have been extended and applied in practice based on statistical or bounded error assumptions of the disturbances. The corresponding parametric uncertainty description will therefore have either a stochastic or a deterministic character. The circumstances, however, under which reliable identification results can be obtained for each of the applied methods, is an important issue.

### **1.5 What's New**

Anticipating on the chapters to come, a few statements will be given which describe the new elements in this thesis work :

- The minimal polynomial output-error model structure for multivariable systems which has been introduced in [Bac87], is extended to the general family of prediction error models [Lju87]. This multivariable model structure shows interesting similarities with the SISO case. The outstanding performance of the identification algorithms compared to existing implementations using (pseudo-) canonical model structures, is illustrated by industrial case studies.

- To provide a quality measure of the model (Cramér-Rao bound), it is a fundamental requirement in current prediction error identification that the process is contained in the model set. This requirement is relieved in the present work. If the process is not contained in the model set, which is always the case in practice, an estimation of the parameter covariance matrix is presented which can be used to compute stochastic uncertainty bounds on the parameters.
- The basic idea of bounded error identification is that the process behaviour can be described by a model set where the data does not conflict the bounded error assumption. It will be shown that for application of this identification technique in practice, proper precautions as described in this thesis are required to avoid an empty or unbounded parameter uncertainty set.
- Parameters in black-box models are treated as independent variables. Modelling the dynamics of a process, several parameters can contribute to the same input/output behaviour. However, related parameters in the polynomial model will increase the corresponding parametric uncertainty. Therefore, parsimonification in the number of free parameters is required. This is realized, counter intuitively, by fixing the most uncertain parameters. This paradoxical approach ensures that the dynamical input/output behaviour of the model set is least affected and that the uncertainty in the remaining free parameters reduced.

In the context of *Parametric Uncertainty in System Identification*, all these aspects will be treated in the following chapters and compared to existing techniques which have been described in the literature.



# ***Minimal Polynomial Parameter Estimation***

---

2.1	General Model Structure	2.3	Parameter Estimation
2.2	Model Parametrization	2.4	Asymptotic Properties

---

The system identification problem is to estimate a model of a system based on observed input-output data. In general, it is impossible and often even undesirable to obtain an exact mathematical description of the properties and the behaviour of a real system. The reason is that such a description leads to very complex mathematical structures. Taking this consideration into account, the problem of system identification is primarily an approximate modelling problem on the basis of experimental data. Therefore, it will be of major importance to specify in which way the identification procedure affects the final estimate.

Confining the identification problem to discrete, linear, time-invariant systems described by polynomial models in the time-domain as motivated in Chapter 1, approximate modelling can roughly be defined as specifying a modelling error or model residual, obtained by performing specific model dependent operations on the experimental data, being a measure of the misfit between the model and the data and making this residual "small" in some sense, e.g. by minimizing with respect to the model a certain scalar measure of the residual. The choice of this model residual and this scalar measure

should reflect the important aspects of the system to be approximated and the intended goal of the model.

A comprehensive treatment of prediction error methods which denote a wide class of identification methods, like equation-error, output-error, etc., is given in [Lju87, SS89]. An extensive amount of literature exists in this field, although most of it is focused on scalar systems. The extension to the multivariable case is often regarded as difficult, because, unlike the scalar case, additional freedom in the choice of model structures and minimization criteria increases the complexity of the identification problem. Other important aspects that influence the selection of model structures are :

- *Flexibility* : A model structure should be selected which can describe the dynamical behaviour of a process accurately and is suitable for its purpose, for example control design. Both the number of free parameters and the way they enter into the model are important. Increasing flexibility with respect to the model structure, however, results in increasing complexity of the model representation.
- *Parsimony* : The process description should be parsimonious meaning that the model should contain the smallest number of free parameters required to describe the system adequately, thereby realizing a minimal parametrization.
- *Algorithm complexity* : Identification methods, like e.g. a prediction error method, can be applied to a variety of model parametrizations. However, the form of the selected structure can considerably influence the required amount of computation.
- *Criterion function* : The asymptotic properties of prediction error method estimates depend crucially on the criterion function. The existence of local as well as non-unique global minima is highly dependent on the selected model structure.

The class of models to describe dynamical systems which will be considered in this thesis, has been restricted to discrete, linear and time-invariant representations. The general model structure which has been adopted to describe the input-output behaviour of a dynamical system consists of a separate process and noise model and will be presented in Section 2.1. In this structure, all disturbances are lumped together under the assumption that they can be modelled as a white noise sequence which is filtered by some noise dynamics and affects the process output in an additive way. More specific polynomial structures of the process and the noise part can be specified describing various types of models. A brief review of these polynomial structures, their parametrizations for the scalar prediction error method and the straightforward extension to the multivariable case are given in Section 2.2. In order to obtain useful algorithms for approximate modelling of industrial processes, a specific model structure has been

selected for the multivariable case : the minimal polynomial representation. This minimal polynomial model structure shows large similarities with the SISO case which is very well known from literature. The model structure definition has been reduced to an order selection problem in contradiction to the difficult estimation of structure indices which are required for multivariable (pseudo-) canonical polynomial models. In addition, less parameters have to be estimated. Further, a reduction of freedom in the minimal polynomial parametrization will be discussed to simplify the computational complexity without loss of generality together with some assumptions to solve the multivariable identification problem which motivates the selection of this specific model structure. The solution to the identification problem, the parameter optimization resulting from minimizing the prediction error for this model structure, is subject of discussion in Section 2.3. It will be shown that modelling multivariable systems using this minimal polynomial representation instead of the (pseudo-) canonical forms which have been proposed in literature, makes the prediction error approach a feasible method for identification of industrial processes. Finally, the asymptotic properties will be discussed in Section 2.4 where the Cramér-Rao inequality is briefly reviewed indicating a lower bound for the variance of the parameter estimate. In literature the assumption is required that the process is contained in the model set in order to apply this inequality for the computation of stochastic parameter bounds. In general, however, this assumption can never be realized in practice due to the low order approximation of the real process. This undermodelling will be taken into account resulting in an estimation of the parameter covariance matrix which can be used to compute stochastic parameter uncertainties.

## 2.1 General Model Structure

Under the assumption that the dynamical system consists of process as well as noise dynamics, the general form of the corresponding model structure is defined by :

$$\begin{aligned} M(\underline{\theta}) : \underline{y}(k) &= \left[ \sum_{k=0}^{\infty} g(k) z^{-k} \right] \underline{u}(k) + \left[ \sum_{k=0}^{\infty} h(k) z^{-k} \right] \underline{\xi}(k) \\ &= G(z^{-1}, \underline{\theta}) \underline{u}(k) + H(z^{-1}, \underline{\theta}) \underline{\xi}(k) \end{aligned} \quad (2.1)$$

where the noise disturbing the process satisfies :  $E \{ \underline{\xi}_t(k) \underline{\xi}_t^T(s) \} = \Lambda_t \delta_{k,s}$ . Further,  $\underline{y}(k)$  is a  $n_y$ -dimensional output at time instant  $k$ ,  $\underline{u}(k)$  a  $n_u$ -dimensional input and  $\underline{\xi}(k)$  a  $n_y$ -dimensional sequence of independent random variables with zero mean value which is referred to as white noise. The filters in Eq. 2.1,  $G(z^{-1}, \underline{\theta})$  and  $H(z^{-1}, \underline{\theta})$ , represent

$(n_y \times n_u)$ - and  $(n_y \times n_y)$ -dimensional bounded impulse responses of infinite length describing stable systems. The argument  $z^{-1}$  denotes the backward shift operator, so  $z^{-1}u(k) = u(k-1)$ , etc. When restricting the filters  $G(z^{-1}, \underline{\theta})$  and  $H(z^{-1}, \underline{\theta})$  to a finite number of taps, they can be described by rational transfer functions in  $z^{-1}$ . The model of Eq. 2.1 is depicted in Fig. 2.1.

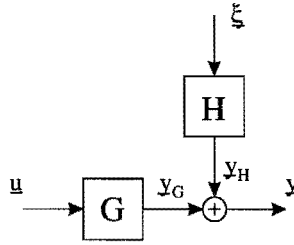


Fig. 2.1 : Block diagram of general model structure.

### 2.1.1 General Model Uniqueness

For analysis purposes, general uniqueness considerations of the models used for identification are often introduced. This concerns the problem of adequately and uniquely describing or representing a given system within a certain model structure. Therefore some assumptions on the true system are required, but the application of identification techniques is not dependent on the validity of such assumptions. Assume that the true system  $S_t$  is linear, discrete, time-invariant, and that the disturbances have rational spectral density. Then it can be described as :

$$S_t : \underline{y}(k) = G_t(z^{-1}) \underline{u}(k) + H_t(z^{-1}) \xi_t(k) \quad , \quad E \left\{ \xi_t(k) \xi_t^T(s) \right\} = \Lambda_t \delta_{k,s} \quad (2.2)$$

Introducing the set :

$$\Theta_t = \left\{ \underline{\theta} \mid G_t(z^{-1}) = G(z^{-1}, \underline{\theta}) \quad , \quad H_t(z^{-1}) = H(z^{-1}, \underline{\theta}) \quad , \quad \Lambda_t = \Lambda(\underline{\theta}) \right\} \quad (2.3)$$

The set  $\Theta_t$  consists of those parameter vectors  $\underline{\theta}$  for which the model  $M$  gives a perfect description of the true system  $S_t$  :

- The set  $\Theta_t$  may be empty, which is referred to as underparametrization.
- The set  $\Theta_t$  consists of one element  $\underline{\theta}_t$  which is called the true parameter vector.
- The set  $\Theta_t$  may consist of several elements which is called overparametrization.



Note that in practice underparametrization will always be the case since the model set can never be large enough to capture the real system as pointed out in the introduction of this chapter. That is, approximate modelling is the best one can hope for.

### 2.1.2 Prediction Error

The filters  $G(z^{-1}, \underline{\theta})$  and  $H(z^{-1}, \underline{\theta})$  as well as the noise covariance matrix  $\Lambda(\underline{\theta})$  are functions of the parameter vector  $\underline{\theta}$  that ranges over a subset  $\Theta_{pe}$  of  $\mathbb{R}^{n\theta}$ , where  $n\theta$  is the dimension of  $\underline{\theta}$  :

$$\underline{\theta} \in \Theta_{pe} \subset \mathbb{R}^{n\theta} \quad (2.4)$$

This set is given by :

$$\Theta_{pe} = \{ \underline{\theta} \mid G(z^{-1}, \underline{\theta}), H(z^{-1}, \underline{\theta}) \text{ and } H^{-1}(z^{-1}, \underline{\theta}) \text{ are asymptotically stable, } H(0, \underline{\theta}) = I, \Lambda(\underline{\theta}) \text{ is nonnegative definite} \} \quad (2.5)$$

Under the assumption that  $\underline{\theta} \in \Theta_{pe}$  and also that the input  $\underline{u}(k)$  and the noise  $\xi(s)$  are uncorrelated for  $k < s$ , which holds if the system operates in open loop or in closed-loop with no direct feedthrough, the one-step-ahead prediction  $\hat{y}(k|\underline{\theta})$  and the corresponding prediction error  $e(k, \underline{\theta})$  are denoted by :

$$\begin{aligned} \hat{y}(k|\underline{\theta}) &= H^{-1}(z^{-1}, \underline{\theta}) G(z^{-1}, \underline{\theta}) \underline{u}(k) + \left[ I - H^{-1}(z^{-1}, \underline{\theta}) \right] y(k) \\ e(k, \underline{\theta}) &= y(k) - \hat{y}(k|\underline{\theta}) \\ &= H^{-1}(z^{-1}, \underline{\theta}) \left[ y(k) - G(z^{-1}, \underline{\theta}) \underline{u}(k) \right] \end{aligned} \quad (2.6)$$

Note that the assumption  $H(0, \underline{\theta}) = I$  and hence  $H^{-1}(0, \underline{\theta}) = I$ , means that the predictor  $\hat{y}(k|\underline{\theta})$  depends not on  $y(k)$  but only on previous output values (i.e.  $y(k-1)$ ,  $y(k-2)$  ...). Further,  $\underline{\theta}$  is restricted to those values for which the model as well as the predictor are asymptotically stable.

Before applying prediction error identification, the following choices have to be made :

- **Model Parametrization** : This concerns the parametrization of  $G(z^{-1}, \underline{\theta})$ ,  $H(z^{-1}, \underline{\theta})$  and  $\Lambda(\underline{\theta})$  as functions of  $\underline{\theta}$  and is subject of discussion in Section 2.2.
- **Identification Criterion** : This concerns the scalar measure of the prediction errors. This criterion will be minimized with respect to  $\underline{\theta}$  to choose the "best" predictor in the selected model structure which is described in Section 2.3.

## 2.2 Model Parametrization

The most immediate way of parametrizing the filters  $G(z^{-1}, \underline{\theta})$  and  $H(z^{-1}, \underline{\theta})$  is to represent them as rational functions and let the parameters be the numerator and denominator coefficients. In this section various ways of carrying out such parametrizations will be described. The general family of model structures for scalar systems will be reviewed briefly and an extension of this family to multivariable systems is described. Parameter identifiability problems due to the non-unique parametrization of these multivariable systems will be discussed resulting in the selection of the minimal polynomial model structure.

### 2.2.1 Scalar Model Structures

For single-input single-output (SISO) systems, the general family of model structures for prediction error methods, introduced in [Lju87], is described by :

$$A(z^{-1}) y(k) = \frac{B(z^{-1})}{F(z^{-1})} u(k) + \frac{C(z^{-1})}{D(z^{-1})} \xi(k) \quad (2.7)$$

where the polynomials are defined as :

$$\begin{aligned} A(z^{-1}) &= 1 + a_1 z^{-1} + \dots + a_{na} z^{-na} \\ B(z^{-1}) &= b_1 + b_2 z^{-1} + \dots + b_{nb} z^{-nb+1} \\ C(z^{-1}) &= 1 + c_1 z^{-1} + \dots + c_{nc} z^{-nc} \\ D(z^{-1}) &= 1 + d_1 z^{-1} + \dots + d_{nd} z^{-nd} \\ F(z^{-1}) &= 1 + f_1 z^{-1} + \dots + f_{nf} z^{-nf} \end{aligned} \quad (2.8)$$

resulting in the following parameter vector  $\underline{\theta}$  :

$$\underline{\theta} = [a_1 \dots a_{na} \ b_1 \dots b_{nb} \ c_1 \dots c_{nc} \ d_1 \dots d_{nd} \ f_1 \dots f_{nf}]^T \quad (2.9)$$

To ensure a unique model representation, the polynomials  $F(z^{-1})$  and  $D(z^{-1})$  should have no common roots. Possible additional delays of  $nk$  samples from the input  $u(k)$  to the output  $y(k)$  have not been included for ease of notation. However, this constitutes no limitation, since  $u(k)$  can always be replaced by a shifted value  $u(k-nk+1)$ , where  $nk > 0$  denotes the delay. Note that the degrees of freedom for delay correction in a multivariable system are limited (see Section A.3).

### 2.2.2 Multivariable Model Structures

The extension to multivariable model structures is often regarded as difficult. There exists an extensive literature concerning this generalization. Representation techniques for multivariable transfer functions are more involved because, unlike the scalar case, there is no unique choice of system representation. It can be easily shown that different multivariable models can represent the same input-output behaviour. The model structure must therefore be restricted to make it uniquely identifiable.

#### General Polynomial Form

A straightforward extension of Eq. 2.7 to the multivariable case is defined by :

$$A(z^{-1})\underline{y}(k) = F^{-1}(z^{-1})B(z^{-1})\underline{u}(k) + D^{-1}(z^{-1})C(z^{-1})\underline{\xi}(k) \quad (2.10)$$

where  $A(z^{-1})$ ,  $B(z^{-1})$ ,  $C(z^{-1})$ ,  $D(z^{-1})$  and  $F(z^{-1})$  are polynomial matrices with dimension  $(n_y \times n_y)$ ,  $(n_y \times n_u)$ ,  $(n_y \times n_y)$ ,  $(n_y \times n_y)$  and  $(n_y \times n_y)$  respectively :

$$\begin{aligned} A(z^{-1}) &= I + A_1 z^{-1} + \dots + A_{na} z^{-na} \\ B(z^{-1}) &= B_1 + B_2 z^{-1} + \dots + B_{nb} z^{-nb+1} \\ C(z^{-1}) &= I + C_1 z^{-1} + \dots + C_{nc} z^{-nc} \\ D(z^{-1}) &= I + D_1 z^{-1} + \dots + D_{nd} z^{-nd} \\ F(z^{-1}) &= I + F_1 z^{-1} + \dots + F_{nf} z^{-nf} \end{aligned} \quad (2.11)$$

defining the parameter matrix analog to Eq. 2.9 as :

$$\Theta = [A_1 \dots A_{na} \ B_1 \dots B_{nb} \ C_1 \dots C_{nc} \ D_1 \dots D_{nd} \ F_1 \dots F_{nf}]^T \quad (2.12)$$

Further, the degree indices, i.e.  $na$ ,  $nb$ , etc., of the polynomial matrices, i.e.  $A(z^{-1})$ ,  $B(z^{-1})$ , etc., are now matrices of the same dimension where every entry indicates the degree of the corresponding scalar polynomial in the polynomial matrix, as shown for  $F(z^{-1})$  as example :

$$F(z^{-1}) = \begin{bmatrix} F_{11}(z^{-1}) & F_{12}(z^{-1}) & \dots & F_{1n_y}(z^{-1}) \\ F_{21}(z^{-1}) & F_{22}(z^{-1}) & \dots & F_{2n_y}(z^{-1}) \\ \vdots & \vdots & & \vdots \\ F_{n_y1}(z^{-1}) & F_{n_y2}(z^{-1}) & \dots & F_{n_y n_y}(z^{-1}) \end{bmatrix} = I + F_1 z^{-1} + \dots + F_{nf} z^{-nf} \quad (2.13)$$

where

$$F_{ij}(z^{-1}) = \delta_{ij} + f_{ij,1} z^{-1} + \dots + f_{ij,nf_{ij}} z^{-nf_{ij}} \quad \begin{cases} \delta_{ij} = 1 & \text{if } i=j \\ \delta_{ij} = 0 & \text{if } i \neq j \end{cases} \quad (2.14)$$

Table 2.1 describes some common black-box models as special cases of the general form presented in Eq. 2.8. The factorization in terms of two polynomials as indicated in Table 2.1 for the process and the noise model, is also called (left) matrix fraction description (MFD). A detailed treatment of such descriptions can be found in [Kai80]. Note that the scalar case is of course included in this description.

Table 2.1 : Some common models.

Polynomials	Model name	$G(z^{-1}, \underline{\theta})$	$H(z^{-1}, \underline{\theta})$
B	FIR (finite impulse response)	B	I
A, B	ARX (equation-error)	$A^{-1}B$	$A^{-1}$
A, B, C	ARMAX	$A^{-1}B$	$A^{-1}C$
B, F	OE (output-error)	$F^{-1}B$	I
B, F, C, D	BJ (Box-Jenkins)	$F^{-1}B$	$D^{-1}C$

To have a unique representation of a given MIMO process and to achieve parameter identifiability, some structure conditions should be imposed on the multivariable model. Several techniques have been proposed to represent multivariable systems in a uniquely identifiable way as an extension to the scalar case :

- Canonical structure realization [Gui75, HD88].
- Pseudo-canonical or overlapping parametrizations [GL78, GW84, vOL82].
- Diagonal structure realization [GL78, ZBE91].
- Minimal polynomial parametrization [HK65, Bac87].

Brief comments on these structures will be given in the following subsections.

### Canonical Form

The structure of a canonical process representation is determined by a set of "structural invariants" (e.g. the Kronecker indices) which defines the model uniquely. The selection procedure results in a minimum number of parameters to be estimated. The disadvantage of using canonical representations of this form is that the estimation of these indices is very critical if not impossible and the parameter estimates are not

consistent if the indices have been wrongly estimated. That is, the model cannot describe the input-output behaviour of the process.

### **Pseudo-canonical Form**

The advantage of overlapping parametrizations is that less structural indices have to be determined at the expense of a few additional parameters to be estimated. A large set of canonical structures can be described in a single overlapping model structure. The structure for overlapping parametrizations where the different columns are given different orders (i.e.  $nf_{ij} = nf_j$ ,  $\forall i$ ) is discussed in [GL78, GW84]. The additional conditions on the polynomial matrices which are required to realize identifiability for the multivariable structures given in Table 2.1 are described in [Lju87, App. 4A].

A more detailed overview of conditions and properties of these canonical and pseudo-canonical model representations is described in [Jan88]. The conditions to achieve parameter identifiability for these (pseudo-) canonical model structures, e.g. by selecting Kronecker indices, are difficult problems in practice and therefore other model structures are preferred.

### **Diagonal Form**

Another natural way to realize the multivariable case is described in [ZBE91]. It is assumed that the  $ny$ -dimensional white noise vector  $\xi(k)$  has mutually independent entries which results in a covariance matrix :

$$\Lambda(\underline{\theta}) = \begin{bmatrix} \sigma_{\xi_1}^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{\xi_{ny}}^2 \end{bmatrix} \quad (2.15)$$

where  $\sigma_{\xi_i}^2$  is the variance of  $\xi_i(k)$ . This imposes no restriction since a diagonal  $\Lambda(\underline{\theta})$  can always be realized by a suitable transformation of the noise model. The assumption that the output disturbances are mutually independent implies that  $A(z^{-1})$ ,  $C(z^{-1})$  and  $D(z^{-1})$  are diagonal matrices. Further, to obtain a unique representation of a given multivariable process,  $F(z^{-1})$  is also assumed to be diagonal. The diagonalization of these matrices decomposes the multivariable process into  $ny$  MISO (multi-input single-output) sub-processes where the  $i^{\text{th}}$  sub-process is defined as :

$$A_{ii}(z^{-1})y_i(k) = F_{ii}^{-1}(z^{-1})B_{i*}(z^{-1})u(k) + D_{ii}^{-1}(z^{-1})C_{ii}(z^{-1})\xi_i(k) \quad (2.16)$$

Although no structure indices have to be estimated for this diagonal form, it is preferred

to identify the process in a true multivariable form.

### Minimal Polynomial Form

Avoiding the estimation of structural indices and deriving a true multivariable description of the process can be obtained by introducing the minimal polynomial model structure which was first proposed by [HK65] and successfully applied in practice by [Bac87]. The basic idea of introducing this model structure is that a dynamical system has a true multivariable character containing common modes which can be controlled by all inputs and can be observed through all outputs. The size of the minimal polynomial model set to describe dynamical systems is large compared to the model set of (pseudo-) canonical models with a given order and structure. In the set of minimal polynomial models of a given degree, several canonical model sets of the same order with different structures are contained. As shown in [Bac87] the minimal polynomial model set of degree  $r$  contains models of order  $n$ , where  $n \leq r \cdot \min(ny, nu)$ , at the expense of multiplicity of poles. This multiplicity of poles results from the property of the minimal polynomial that its roots are all distinct eigenvalues of the model with multiplicity one only. In general, the order of the model and the degree of the minimal polynomial are different. The relation between the order of a system in minimal state-space representation  $n$  defined by its characteristic polynomial and the degree of the minimal polynomial  $r$  is :  $n = r \cdot \min(ny, nu)$ , resulting in a multiplicity of  $\min(ny, nu)$  for all poles.

Generalizing this minimal polynomial model structure to the general model structure depicted in Fig. 2.1 and assuming that the entries of the noise vector  $\xi(k)$  are mutually independent, the polynomial matrices of Eq. 2.11 reduce to scalar (minimal) polynomials  $A(z^{-1})$ ,  $D(z^{-1})$  and  $F(z^{-1})$ , a full polynomial matrix  $B(z^{-1})$  and a diagonal polynomial matrix  $C(z^{-1})$ , both of dimension  $(ny \times ny)$  :

$$\begin{aligned}
 A(z^{-1}) &= (1 + a_1 z^{-1} + \dots + a_{na} z^{-na}) I^{ny} \\
 B_{ij}(z^{-1}) &= b_{ij,1} + b_{ij,2} z^{-1} + \dots + b_{ij,nb_{ij}} z^{-nb_{ij}+1} \\
 C_{ii}(z^{-1}) &= 1 + c_{ii,1} z^{-1} + \dots + c_{ii,nc_{ii}} z^{-nc_{ii}} \\
 D(z^{-1}) &= 1 + d_1 z^{-1} + \dots + d_{nd} z^{-nd} \\
 F(z^{-1}) &= 1 + f_1 z^{-1} + \dots + f_{nf} z^{-nf}
 \end{aligned} \tag{2.17}$$

where the parameter vector  $\theta$  contains again all polynomial coefficients of Eq. 2.17 :

$$\underline{\theta} = [ a_1 \dots a_{na} \ b_{11,1} \dots b_{11,nb_{11}} \ b_{12,1} \dots b_{nynu,nb_{nynu}} \ c_{11,1} \dots c_{11,nc_{11}} \ c_{22,1} \dots c_{nyny,nc_{nyny}} \ d_1 \dots d_{nd} \ f_1 \dots f_{nf} ]^T \quad (2.18)$$

Note that prediction error estimation of dynamical systems when adopting the diagonal model form of [ZBE91] presented in Eq. 2.16, can be computed using this minimal polynomial model structure by identifying every output separately.

This minimal polynomial model structure provides a true multivariable description of a dynamical system with a unique representation. The corresponding predictor  $\hat{y}(k|\underline{\theta})$  and prediction error  $\underline{e}(k,\underline{\theta})$  can be easily derived by substituting Eq. 2.17 into Eq. 2.6 resulting in :

$$\begin{aligned} \hat{y}(k|\underline{\theta}) &= D(z^{-1}) C^{-1}(z^{-1}) \frac{B(z^{-1})}{F(z^{-1})} \underline{u}(k) + [I - D(z^{-1}) C^{-1}(z^{-1}) A(z^{-1})] \underline{y}(k) \\ \underline{e}(k,\underline{\theta}) &= D(z^{-1}) C^{-1}(z^{-1}) \left[ A(z^{-1}) \underline{y}(k) - \frac{B(z^{-1})}{F(z^{-1})} \underline{u}(k) \right] \end{aligned} \quad (2.19)$$

Although the diagonalization of the noise polynomial matrix  $C(z^{-1})$  results from the assumption that the white noise vector  $\xi(k)$  consists of mutually independent entries, it simplifies the calculation of the prediction error  $\underline{e}(k,\underline{\theta})$  and the parameter estimation, which is subject of discussion in Section 2.3, significantly. According to Eq. 2.19 polynomial matrix inversion,  $C^{-1}(z^{-1}) = \text{adj}\{C(z^{-1})\}/\det\{C(z^{-1})\}$ , is required to calculate the predictor. For a full polynomial matrix  $C(z^{-1})$ , this can become very complicated with exploding computational complexity especially for increasing number of outputs. However, when  $C(z^{-1})$  is restricted to have diagonal entries only, the polynomial matrix inversion becomes very simple,  $C^{-1}(z^{-1}) = \text{diag}\{1/C_{11}(z^{-1}) \dots 1/C_{nyny}(z^{-1})\}$ . The inversion reduces to a simple scalar operation applied to the diagonal entries  $C_{ii}(z^{-1})$ . This also provides the possibility to compute analytically the first- and second-order derivatives which are required for parameter optimization (see Section 2.3.4 and Appendix B).

The advantages of this minimal polynomial model structure are various and can be summarized as follows :

- No estimation of structural indices required.
- The structure of the parameter vector is uniquely identifiable for the multivariable case.
- The number of parameters to be estimated is generally lower compared to (pseudo-) canonical model structures.
- Stable predictions  $\hat{y}(k|\underline{\theta})$  for  $\underline{\theta} \in \Theta_{pe}$  (Eq. 2.5) can be easily guaranteed because

the corresponding polynomials  $F(z^{-1})$  and  $C_{ii}(z^{-1})$  are scalars.

- Restricting  $C(z^{-1})$  to have diagonal entries only, reduces the computational complexity for the parameter estimation and the prediction error significantly (see Section 2.3).
- Initial estimates for the parameter optimization can be derived similar to the scalar case because of the minimal polynomial model structure (see Section 2.3.3).

Of course, there are also some disadvantages :

- The diagonalization of  $C(z^{-1})$  reduces the degrees of freedom and eliminates in fact the true multivariable character of the noise model.
- The corresponding state-space representation will have a multiplicity of the poles corresponding to  $\min(n_y, n_u)$  as described in [Bac87] and an additional model reduction is required to get rid of this high order representation.

## 2.3 Parameter Estimation

In the previous section the structure and parametrization of multivariable models have been discussed. There is another important aspect in parameter estimation : the criterion which will be minimized with respect to  $\underline{\theta}$  to choose the "best" predictor in the selected model structure. In contradiction to SISO identification problems, the selection of an identification criterion for multivariable systems is more involved.

### 2.3.1 Identification Criterion

The criterion which maps the sequence of prediction errors into a scalar can be chosen in many ways. Define the sample covariance matrix as :

$$R_e(\underline{\theta}) = \frac{1}{N-n} \sum_{k=n+1}^N \underline{e}(k, \underline{\theta}) \underline{e}^T(k, \underline{\theta}) \quad (2.20)$$

where  $N$  denotes the length of the data set and  $n$  the number of data samples which are required for initialization (see Section B.2). For scalar systems,  $\underline{e}(k, \underline{\theta})$  (Eq. 2.6) is a scalar and so is  $R_e(\underline{\theta})$  which can then be taken directly as a criterion to be minimized. In the multivariable case,  $R_e(\underline{\theta})$  is a positive definite matrix. Then the criterion  $V(R_e)$  is a scalar-valued function defined on the set of positive definite matrices  $R_e(\underline{\theta})$ . Some possible choices of  $V(R_e)$  are :



- $V(R_e) = \det(R_e(\underline{\theta}))$ , which corresponds to the Maximum Likelihood method if the process is contained in the model set for Gaussian distributed disturbances.
- $V(R_e) = \text{tr}(\Lambda^{-1}(\underline{\theta}) R_e(\underline{\theta}))$ , which has some nice features from a computational point of view, but requires scaling of the data with the inverse of the covariance matrix.

Proofs of these assertions with respect to the various identification criteria are given in [SS89, Section 7.5]. Since the covariance matrix  $\Lambda(\underline{\theta})$  of the prediction error is very seldom known in practice, selecting the trace as scalar valued identification criterion seems not very useful in practical situations. However, an accurate estimate of  $\Lambda(\underline{\theta})$  can be derived from a high-order ARX model, because for sufficiently high-order as the number of data samples  $N$  tends to infinity, an ARX model is capable of approximating any linear system arbitrarily well [Lju85, ZBE91]. This provides the possibility to scale the data correctly in advance (see Appendix A, Section A.4). Because  $\text{tr}(\Lambda^{-1}(\underline{\theta}) R_e(\underline{\theta}))$  requires less calculations for evaluation of the loss-function compared to  $\det(R_e(\underline{\theta}))$  and for the scaling only a least-squares problem has to be solved, the trace criterion will be used for multivariable system identification.

Under the assumption that the elements of the white noise vector  $\xi(k)$  are mutually independent and if the data is properly scaled in advance, the covariance matrix reduces to :

$$\Lambda(\underline{\theta}) = \begin{bmatrix} \sigma_{\xi_1}^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{\xi_{ny}}^2 \end{bmatrix} = \sigma_{\xi}^2 \mathbf{I}^{ny} \quad (2.21)$$

where  $\sigma_{\xi}^2$  defines the variance of  $\xi(k)$  because of  $\sigma_{\xi_1}^2 = \sigma_{\xi_2}^2 = \dots = \sigma_{\xi_{ny}}^2$ . Consequently, the identification criterion for multivariable system becomes :

$$V(R_e) = \text{tr}(\Lambda^{-1}(\underline{\theta}) R_e(\underline{\theta})) = \frac{1}{\sigma_{\xi}^2 (N-n)} \sum_{k=n+1}^N \underline{e}^T(k, \underline{\theta}) \underline{e}(k, \underline{\theta}) \quad (2.22)$$

Furthermore, the minimal polynomial structure definition for multivariable systems introduces a special column format for the prediction error :

$$\underline{e}_{\text{col}}(\underline{\theta}) = [\underline{e}_1(n+1, \underline{\theta}) \dots \underline{e}_1(N, \underline{\theta}) \underline{e}_2(n+1, \underline{\theta}) \dots \underline{e}_{ny}(N, \underline{\theta})]^T \quad (2.23)$$

where the subscript col in  $\underline{e}_{\text{col}}(\underline{\theta})$  indicates a column vector formed by stacking the columns of the prediction error matrix  $E(\underline{\theta})$  on top of each other. The new index  $\ell$  has been introduced because the  $N_{\text{col}} = ny(N-n)$  elements of  $\underline{e}_{\text{col}}(\underline{\theta})$  do not correspond

anymore with the time instants  $k$  (see Eq. 2.22). A more detailed description of this minimal polynomial prediction error is given in Appendix B, Minimal Polynomial Identification. Multiplying the identification criterion  $V(R_e)$  in Eq. 2.22 by a constant factor illustrates the large similarity with the SISO case :

$$V_e(\underline{\theta}) = \frac{\sigma_\xi^2}{ny} V(R_e) = \frac{1}{N_{col}} \sum_{\ell=1}^{N_{col}} e_{col}^2(\ell, \underline{\theta}) = \frac{1}{N_{col}} \underline{e}_{col}^T(\underline{\theta}) \underline{e}_{col}(\underline{\theta}) \quad (2.24)$$

which is frequently called the loss-function. When minimizing the prediction errors, this loss-function  $V_e(\underline{\theta})$  will converge to its minimum value  $\sigma_\xi^2$  which is identical to the minimum value of the SISO identification criterion. Therefore,  $V_e(\underline{\theta})$  has been selected as multivariable prediction error identification criterion.

A disadvantage of the identification criterion as defined in Eq. 2.24, is its relative high sensitivity to large prediction errors which can result from outliers in the data. To limit the influence of bad data using robust estimation techniques, the prediction error can be filtered first through a stable linear filter  $L(z^{-1})$  :

$$\tilde{e}_i(k, \underline{\theta}) = L_i(z^{-1}) e_i(k, \underline{\theta}) \quad (2.25)$$

instead of using the prediction error  $e(k, \underline{\theta})$  directly. The corresponding criterion function  $V_{\tilde{e}}(\underline{\theta})$  can be influenced by prefiltering the prediction error  $e(k, \underline{\theta})$ . The relative importance of specific frequency regions can be enhanced or suppressed with a proper choice of  $L(z^{-1})$ . The parameter estimate  $\hat{\underline{\theta}}$  can now be determined as the argument corresponding with the (global) minimum value of the loss-function  $V_{\tilde{e}}(\underline{\theta})$  :

$$\hat{\underline{\theta}} = \arg \min_{\underline{\theta} \in \Theta_{pe}} V_{\tilde{e}}(\underline{\theta}) \quad (2.26)$$

Although the identification criterion has been defined, the selection of a specific model structure, i.e. one of the models described in Table 2.1 or a combination of them, and the model parametrization are still open problems which will now be treated.

### 2.3.2 Order Selection

The structure and parametrization of the models described in Section 2.2 provide sufficient flexibility to identify a wide range of dynamical systems. This flexibility, however, introduces problems as well. The choice of a typical model structure (see Table 2.1) is determined by the type of process to be identified and the purpose of the

model, e.g. control design. The selection of a model structure which is most suitable to describe the process under study accurately, is based on experience for which no general selection rules can be given.

Furthermore, the parametrization of the process model  $G(z^{-1}, \underline{\theta})$  and the noise model  $H(z^{-1}, \underline{\theta})$  is a major problem in practice. Under the assumption that  $G(z^{-1}, \underline{\theta})$  and  $H(z^{-1}, \underline{\theta})$  are fully parametrized, i.e. proper transfer functions in the polynomial model representation, the parametrization problem reduces to the order selection of the process model  $G(z^{-1}, \underline{\theta})$  and the noise model  $H(z^{-1}, \underline{\theta})$ . These orders correspond to the degrees of the minimal polynomials.

The order selection rules in classical identification are based on comparison of models with increasing order. When the model structure is expanded so that more parameters are included in the parameter vector, the minimal value of the loss-function  $V_e(\underline{\theta})$  naturally decreases since new degrees of freedom have been added to the optimization problem, or, in other words, the set over which the optimization is done has been enlarged. The comparison of model orders can then be interpreted as a test for a significant decrease in the minimal values of the loss-function associated with the (nested) model orders. Classical selection criteria are Akaike's information criterion (AIC) and the final prediction error criterion (FPE) [Lju87] :

$$AIC = N_{col} \log V_e(\hat{\underline{\theta}}) + 2n\theta$$

$$FPE = \left( \frac{1 + \frac{n\theta}{N_{col}}}{1 - \frac{n\theta}{N_{col}}} \right) V_e(\hat{\underline{\theta}}) \quad (2.27)$$

which define a trade-off between a decreasing loss-function  $V_e(\underline{\theta})$  and an increasing number of parameters  $n\theta$  for  $N_{col} = ny(N-n)$  data samples. The minimal value of the selection criteria defines then the optimal order. A major disadvantage of this approach is the time consuming and computational complex procedure since the selection criteria have to be computed for all orders until the criterion value increases. In addition, no clear distinction can be made between the order of the process and noise model. This seems not very useful in practice.

An alternative approach has been proposed in [Bac87] where a FIR model has been used to construct a Hankel matrix of Markov parameters. If the process is contained in the model set and no noise disturbs the data, the singular values of this Hankel matrix obtained by singular value decomposition will become zero if the process order is exceeded. The last nonzero singular value defines then the order of the minimal polynomial. In practice, however, all singular values will be larger than zero due to the

noise present on the estimated Markov parameters. A significant decrease of the singular values indicates then the order of the model. This approach can be extended to order estimation of the process model  $G(z^{-1}, \underline{\theta})$  as well as the noise model  $H(z^{-1}, \underline{\theta})$ . In fact, any model of the process, e.g. FIR or high-order equation-error model, which approximates the process accurately, can be used to compute the Markov parameters and to construct the Hankel matrix. Of course, models which can be computed easily are preferred to reduce the required computation time for this order selection.

Suppose an accurate polynomial model of sufficiently high order has been estimated and that the corresponding state-space representations are given by :

$$\begin{aligned}\underline{x}_G(k+1) &= A_G(\underline{\theta}) \underline{x}_G(k) + B_G(\underline{\theta}) \underline{u}(k) \\ \underline{y}_G(k) &= C_G(\underline{\theta}) \underline{x}_G(k) + D_G(\underline{\theta}) \underline{u}(k) \\ \underline{x}_H(k+1) &= A_H(\underline{\theta}) \underline{x}_H(k) + B_H(\underline{\theta}) \underline{\xi}(k) \\ \underline{y}_H(k) &= C_H(\underline{\theta}) \underline{x}_H(k) + D_H(\underline{\theta}) \underline{\xi}(k)\end{aligned}\tag{2.28}$$

for the process model  $G(z^{-1}, \underline{\theta})$  and the noise model  $H(z^{-1}, \underline{\theta})$  respectively. The Markov parameters of the process model for example, can be obtained from :

$$M_{G,i} = \begin{cases} D_G(\underline{\theta}) & i = 0 \\ C_G(\underline{\theta}) A_G^{i-1}(\underline{\theta}) B_G(\underline{\theta}) & i > 0 \end{cases}\tag{2.29}$$

based on the equivalence relation :

$$\begin{aligned}\underline{y}_G(k) &= \sum_{i=0}^m M_{G,i} \underline{u}(k-i) \\ \underline{y}_G(k) &= C_G(\underline{\theta}) A_G^m(\underline{\theta}) \underline{x}_G(0) + \sum_{i=1}^m C_G(\underline{\theta}) A_G^{i-1}(\underline{\theta}) B_G(\underline{\theta}) \underline{u}(k-i) \\ &\quad + D_G(\underline{\theta}) \underline{u}(k)\end{aligned}\tag{2.30}$$

under the assumption that the initial states of the process are equal to zero. To determine the degree of the minimal polynomial the number of independent block rows/columns of the largest possible Hankel matrix has to be found. The Markov parameters can be written in column vector form to find an appropriate estimate of this degree which is defined by the dependencies of the column vectors. The Hankel matrix built from the Markov parameters in column form can be constructed according to :

$$H_G = \begin{bmatrix} \text{col}(M_{G,1}) & \text{col}(M_{G,2}) & \dots & \text{col}(M_{G,j}) \\ \text{col}(M_{G,2}) & \text{col}(M_{G,3}) & \dots & \text{col}(M_{G,j+1}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{col}(M_{G,i}) & \text{col}(M_{G,i+1}) & \dots & \text{col}(M_{G,i+j-1}) \end{bmatrix} \quad (2.31)$$

where

$$\text{col}(M_{G,i}) = [ [M_{G,i}]_{11} [M_{G,i}]_{21} \dots [M_{G,i}]_{ny1} [M_{G,i}]_{12} \dots [M_{G,i}]_{nynu} ]^T \quad (2.32)$$

The column vector of the Markov parameters is formed by stacking the columns of the Markov matrix (Eq. 2.26) on top of each other.

The singular value decomposition of the block Hankel matrix can be written as :

$$H_G = U_G S_G V_G^T = \sum_{i=1}^m s_{G,i} U_{G,i} V_{G,i}^T \quad (2.33)$$

where  $U_G$  is an orthogonal matrix containing the left singular vectors,  $S_G$  a diagonal matrix containing the singular values  $s_{G,i}$  and  $V_G$  an orthogonal matrix containing the right singular vectors. Because all singular vectors are orthogonal, each singular value defines a weighting factor in the direction of the corresponding singular vector. A singular value which is almost equal to zero indicates redundant information. The order of the minimal polynomial is selected from the number of singular values which differ significantly from zero.

Although this approach of order selection reduces the amount of computations significantly, the correctness of the orders for the process and the noise model should always be verified afterwards.

### 2.3.3 Initial Estimates

After selecting a specific model structure and order, the parameter vector  $\theta$  has to be estimated. In the special case where  $\underline{g}(k, \theta)$  depends linearly on  $\theta$ , the minimization of the loss-function  $V_e(\theta)$  (Eq. 2.24) can be done analytically. In most cases however, the minimum of  $V_e(\theta)$  cannot be found analytically. For such cases the minimization must be performed using numerical search routines. There is an extensive literature on such numerical problems, see [Lue73, GMW81, Sca85]. A typical property of this numerical minimization is that convergence will be achieved. Note however, that the parameter optimization can converge to a local minimum of the loss-function  $V_e(\theta)$ . To find the

global solution, there is usually no other way than to start the iterative minimization routine at different feasible initial values and compare the results. Due to the possible occurrence of undesired local minima of the criterion, it is worthwhile to spend some effort on preliminary estimation procedures to produce good initial values for the minimization.

Results to obtain initial estimates available for the scalar black-box structure (Eq. 2.8) can be used as well for the minimal polynomial model structure (Eq. 2.17) because of the large similarity between the scalar and the selected multivariable case. The discussion here will be restricted to the approach based on estimating high-order ARX models of the system and applied to the model structures described in Table 2.1. Suppose the true system is defined by Eq. 2.3 and an ARX model structure :

$$A_h(z^{-1}) \underline{y}(k) = B_h(z^{-1}) \underline{u}(k) + \underline{\xi}(k) \quad (2.34)$$

of order  $h$  is used where  $h$  defines the order of the minimal polynomial  $A_h(z^{-1})$  in a fully parametrized model. Then it can be shown, see [Lju85, ZBE91], that if the number of data samples and the model order tend to infinity ( $N \gg h$ ), then the polynomial model  $(\hat{A}_h, \hat{B}_h)$  will converge to the true system in the following sense :

$$\frac{\hat{B}_h(e^{i\omega})}{\hat{A}_h(e^{i\omega})} \rightarrow G_t(e^{i\omega})$$

uniformly in  $\omega$  as  $N \gg h \rightarrow \infty$  (2.35)

$$\frac{1}{\hat{A}_h(e^{i\omega})} \rightarrow H_t(e^{i\omega})$$

This means that a high-order ARX model is capable of approximating any linear system arbitrarily well. For an ARX model in the proposed general prediction model set the loss-function  $V_e(\underline{\theta})$  is a quadratic function of the parameters  $\underline{\theta}$  which guarantees that there is only one minimum, the global minimum. It is of course desirable to reduce this high-order model to a lower order model within the structures described in Table 2.1. However, the high-order model (Eq. 2.34) yields a first indication of the maximal achievable performance while also the scaling of the data can be verified to ensure equal weighting of the prediction errors during the parameter optimization (see Appendix. A.4, Data Decimation and Scaling).

Before describing a procedure to obtain initial parameter estimates for several model structures, define the following signals :

$$\begin{aligned}\hat{y}_h(k) &= \frac{\hat{B}_h(z^{-1})}{\hat{A}_h(z^{-1})} u(k) \\ \hat{e}_h(k) &= \hat{A}_h(z^{-1}) y(k) - \hat{B}_h(z^{-1}) u(k)\end{aligned}\quad (2.36)$$

where  $\hat{y}_h(k)$  denotes the simulated output and  $\hat{e}_h(k) \approx \xi(k)$  the prediction error of the high-order ARX model.

The proposed procedures to obtain appropriate initial estimates for the various model structures are then as follows :

- *ARMAX model* : Use the following model structure :

$$A(z^{-1}) y(k) = B(z^{-1}) u(k) + [C(z^{-1}) - I] \hat{e}_h(k) + \xi(k) \quad (2.37)$$

to estimate the polynomials A, B and C. Since  $\hat{e}_h(k)$  is a known sequence, this model has an ARX structure with ny additional inputs, and consequently the estimates can be determined by the least-squares method [MF82].

- *OE model* : For the output error model structure, a modified Steiglitz-McBride method [SM65] has been adopted. This iterative scheme :

$$(\hat{F}^{i+1}, \hat{B}^{i+1}) = \arg \min_{(F,B)} \sum_{k=1}^N \left\{ F(z^{-1}) \left[ \frac{1}{\hat{F}^i(z^{-1})} \hat{y}_h(k) \right] - B(z^{-1}) \left[ \frac{1}{\hat{F}^i(z^{-1})} u(k) \right] \right\}^2 \quad (2.38)$$

is based on successive data filtering,

$$y_f^i(k) = \frac{1}{\hat{F}^i(z^{-1})} \hat{y}_h(k) \quad , \quad u_f^i(k) = \frac{1}{\hat{F}^i(z^{-1})} u(k) \quad i = 1, 2, \dots \quad (2.39)$$

and least-squares estimation for determining B(q) and F(q) :

$$F(z^{-1}) y_f^i(k) = B(z^{-1}) u_f^i(k) + \xi(k) \quad (2.40)$$

The iteration is initialized by applying Eq. 2.34 first for  $y_f^0(k) = \hat{y}_h(k)$ . Note that in contradiction to the original method and other references [Lju87, SS89], the simulated output of the high-order model  $\hat{y}_h(k)$  has been used instead of the measured output  $y(k)$ . In practice, this approach converges to a more accurate initial estimate because the output data  $\hat{y}_h(k)$  is generated by a linear

(simulation) process and contains no noise where the measured data  $\underline{y}(k)$  in practice often is corrupted with a lot of noise or even with spikes and nonlinearities. Although in general no convergence to a minimum point of the output error loss-function can be guaranteed, this approach gives a sufficiently accurate initialization estimate for output error optimization.

- *BJ model* : For the Box-Jenkins model structure the initial estimates of the process and the noise dynamics will be derived separately. The process model :

$$G(z^{-1}, \underline{\theta}) = \frac{B(z^{-1})}{F(z^{-1})} \quad (2.41)$$

can be initialized using the Steiglitz-McBride iteration described for the output error model structure resulting in  $\{\hat{B}(z^{-1}), \hat{F}(z^{-1})\}$ . To determine the noise dynamics, an estimate of the noise sequence can be constructed according to :

$$\hat{\underline{v}}(k) = \underline{y}(k) - \frac{\hat{B}(z^{-1})}{\hat{F}(z^{-1})} \underline{u}(k) \quad (2.42)$$

This noise can then be regarded as a measured signal for which an ARMA model can be determined in a separate step. One approach is to apply a high-order AR model to  $\hat{\underline{v}}_i(k)$  and form estimates of the innovations  $\hat{\underline{e}}_i(k)$ . The polynomials  $D(z^{-1})$  and  $C(z^{-1})$  can then be estimated with the standard least-squares method from the ARX model :

$$D(z^{-1}) \hat{\underline{v}}(k) = [C(z^{-1}) - I] \hat{\underline{e}}(k) + \underline{\xi}(k) \quad (2.43)$$

with  $\hat{\underline{v}}(k)$  as output and  $\hat{\underline{e}}(k)$  as input.

The methods to obtain initial estimates for parameter optimization which have been described briefly and only for the most common model structures are similar to the scalar case. Other approaches to derive initial estimates for general model structures can be found in the literature.

### 2.3.4 Numerical Minimization

After selecting a correct model structure and order, the parameters can be estimated by minimizing the prediction error using the initial estimate as starting value. A commonly used approach is based on Newton algorithms [Lue73, GMW81, Sca85] using values of the function  $V_e(\underline{\theta})$  to be minimized, of its gradient  $\nabla_{\underline{\theta}} V_e(\underline{\theta})$ , and of its Hessian



$\nabla_{\theta\theta}^2 V_e(\underline{\theta}) :$

$$\hat{\underline{\theta}}^{i+1} = \hat{\underline{\theta}}^i - \alpha^i \left[ \nabla_{\theta\theta}^2 V_e(\hat{\underline{\theta}}^i) \right]^{-1} \nabla_{\theta} V_e(\hat{\underline{\theta}}^i) \quad (2.44)$$

Here  $\nabla_{\theta}$  and  $\nabla_{\theta\theta}^2$  denote respectively the first- and second-order derivatives with respect to  $\underline{\theta}$ ,  $\hat{\underline{\theta}}^i$  indicates the  $i^{\text{th}}$  iteration estimate in the optimization and  $\alpha^i$  is a variable step length which is often necessary in practice. The positive constant can be determined so that an maximal decrease in the value of  $V_e(\underline{\theta})$  is obtained :

$$\alpha^i = \arg \min_{\alpha} V_e \left( \hat{\underline{\theta}}^i - \alpha \left[ \nabla_{\theta\theta}^2 V_e(\hat{\underline{\theta}}^i) \right]^{-1} \nabla_{\theta} V_e(\hat{\underline{\theta}}^i) \right) \quad (2.45)$$

Straightforward differentiation of the loss-function  $V_e(\underline{\theta})$  with respect to  $\underline{\theta}$  results in the following gradient expression :

$$\frac{\partial V_e(\underline{\theta})}{\partial \underline{\theta}} = \frac{2}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} e_{\text{col}}(\ell, \underline{\theta}) \frac{\partial e_{\text{col}}(\ell, \underline{\theta})}{\partial \underline{\theta}} \quad (2.46)$$

Differentiating of Eq. 2.46 again with respect to  $\underline{\theta}$  gives the Hessian matrix :

$$\frac{\partial^2 V_e(\underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}^T} = \frac{2}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \left( \frac{\partial e_{\text{col}}(\ell, \underline{\theta})}{\partial \underline{\theta}} \frac{\partial e_{\text{col}}(\ell, \underline{\theta})}{\partial \underline{\theta}^T} + e_{\text{col}}(\ell, \underline{\theta}) \frac{\partial^2 e_{\text{col}}(\ell, \underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}^T} \right) \quad (2.47)$$

In most literature it is assumed that the prediction error  $e_{\text{col}}(\ell, \underline{\theta})$  is white and therefore the second-order derivative of the prediction error  $\nabla_{\theta\theta}^2 e_{\text{col}}(\ell, \underline{\theta})$  will be independent of  $e_{\text{col}}(\ell, \underline{\theta})$  making the second term in Eq. 2.47 zero. This approximation of the Hessian matrix  $\nabla_{\theta\theta}^2 V_e(\underline{\theta})$  is then by construction guaranteed to be positive definite resulting in a decreasing loss-function at every iteration if  $\alpha^i$  is appropriately chosen. Although the amount of computations reduces significantly since only the first order derivatives of the prediction error  $\nabla_{\theta} e_{\text{col}}(\ell, \underline{\theta})$  need to be evaluated, a white prediction error  $e_{\text{col}}(\ell, \underline{\theta})$  is hard to realize in practice and the results will be less accurate when ignoring the second derivative term completely. However, when taking the second order derivative of the prediction error  $\nabla_{\theta\theta}^2 e_{\text{col}}(\ell, \underline{\theta})$  into account, proper precautions are necessary because the Hessian matrix might become non positive definite in a point far from the optimum resulting in an incorrect search direction and no convergence of the parameter estimate.

In Appendix B describing the Minimal Polynomial Identification in more detail, it is shown that the first- and second-order derivatives of the prediction error for the minimal polynomial model structure can be computed analytically in an efficient way by simply filtering the data sequences by polynomial transfer functions. The predictor  $\hat{y}(k|\underline{\theta})$  (Eq. 2.18) is reformulated into a pseudo-linear regression form and expressions of the

gradient  $\nabla_{\theta} V_e(\underline{\theta})$  together with the Hessian  $\nabla_{\theta\theta}^2 V_e(\underline{\theta})$  are given in matrix notation. The parameter vector  $\underline{\theta}$  can then be optimized by iteratively calculating the derivatives and solving a least-squares problem until convergence is achieved.

## 2.4 Asymptotic Properties

In analyzing the asymptotic properties of the identified models, determined by Eq. 2.26, the limit properties of the loss-function  $V_e(\underline{\theta})$  will be described as the number of data samples ( $N$ ) tends to infinity. In the following,  $\hat{\underline{\theta}}$  denotes the parameter estimate which minimizes the loss-function  $V_e(\underline{\theta})$ .

For the theoretical analysis the following basic assumptions are made in standard literature :

- 1) The data  $\{u(k), y(k)\}$  are stationary processes.
- 2) The input is persistently exciting.
- 3) The Hessian  $\nabla_{\theta\theta}^2 V_e(\underline{\theta})$  is nonsingular at least locally around the minimum value of  $V_e(\underline{\theta})$ .
- 4) The filters  $G(z^{-1}, \underline{\theta})$ , the process model, and  $H(z^{-1}, \underline{\theta})$ , the noise model, are smooth (differentiable) functions of the parameter vector  $\underline{\theta}$ .
- 5) The set  $\Theta_t$  consists of precisely one element  $\underline{\theta}_t$ .

The assumptions 1-4 are fairly weak and can be mostly satisfied in practice. However, as stated before, the last assumption can never be fulfilled in practice. This, because in practice no perfect description of the system can be obtained. Hence, in general the true model set  $\Theta_t$  is empty. Despite the fact that assumption 5 is always violated in practice, most results in standard literature are based on this assumption. For comparison and to emphasize the consequences for the asymptotic properties of the estimates, the results for both situations :

- $\Theta_t = \{ \underline{\theta}_t \}$
- $\Theta_t = \emptyset$

will be described.

### 2.4.1 Convergence

When  $N$  tends to infinity, the sample covariance matrix (Eq. 2.20) converges to the

corresponding expected values. Since the scalar-valued function  $V(R_e)$  is assumed to be continuous, it follows that the loss-function :

$$V_e(\underline{\theta}) = \frac{\sigma_e^2}{ny} V(R_e) \rightarrow \frac{\sigma_e^2}{ny} V(R_{e_\infty}) \triangleq V_{e_\infty}(\underline{\theta}) \quad , \quad N \rightarrow \infty \quad (2.48)$$

where

$$R_{e_\infty}(\underline{\theta}) = E \left\{ \underline{e}(k, \underline{\theta}) \underline{e}^T(k, \underline{\theta}) \right\} \quad (2.49)$$

It has been shown in [Lju78] that the convergence is uniform on compact (i.e. closed and bounded) sets in the parameter space. For a uniform convergence, it follows that  $\hat{\underline{\theta}}$  converges to a minimum point of  $V_{e_\infty}(\underline{\theta})$  denoted by  $\underline{\theta}^*$  :

$$\hat{\underline{\theta}} \rightarrow \underline{\theta}^* \triangleq \arg \min_{\underline{\theta} \in \Theta_{pe}} V_{e_\infty}(\underline{\theta}) \quad , \quad N \rightarrow \infty \quad (2.50)$$

Note that for this convergence result it is not required that the model structure has to be large enough to cover the true system. Consequently the estimate  $\hat{\underline{\theta}}$  converges to the best possible approximation of the system within the selected model set  $\Theta_{pe}$ . The approximation is in fact the most reasonable. The parameter vector  $\underline{\theta}^*$  is by definition such that the prediction error  $\underline{e}(k, \underline{\theta})$  has a variance as small as possible.

### 2.4.2 Consistency

Formally, the consistency property of the estimates implies that asymptotically with probability 1 the true system is identified. Parameter identifiability, or consistency of  $\hat{\underline{\theta}}$ , can only be guaranteed when assumption 5 is satisfied which is never the case in practice. In general for the situation of underparametrization, no consistency results are known.

Only when the true system is given by Eq. 2.2 and under the assumptions that the model structure is independently parametrized :

$$\underline{y}(k) = G(z^{-1}, \underline{\theta}_1) \underline{u}(k) + H(z^{-1}, \underline{\theta}_2) \underline{e}(k) \quad , \quad E\{\underline{e}(k) \underline{e}^T(k)\} = \Lambda(\theta_3) \quad (2.51)$$

and that there exists a parameter vector  $\underline{\theta}_1$  such that the process model can be described exactly within the model set :

$$G_t(z^{-1}) \equiv G(z^{-1}, \underline{\theta}_1) \quad (2.52)$$

it can be shown [SS89] that  $\underline{\theta}_1^* = \underline{\theta}_1$ . In other words, a perfect description of the process dynamics  $G_t(z^{-1})$  can be obtained even though the true noise dynamics  $H_t(z^{-1})$  may not

be adequately parametrized. Output-error and Box-Jenkins models are special cases where these assumptions are applicable.

Note however, in general Eq. 2.52 is not satisfied in practice because process and noise dynamics can hardly be separated. For practical problems, it is simply not known whether dynamical behaviour belongs to the process or the noise system.

### 2.4.3 Asymptotic Distribution

For the practical situation where the system is more complex than the model structure, it follows from the asymptotic properties that the parameter estimates will converge to a minimum value of the loss-function as defined in Eq. 2.50. An interesting analysis is to study how fast these asymptotic properties can be obtained.

Complete knowledge of the parameter vector  $\hat{\underline{\theta}}$  is present in its probability density function. However, no explicit expressions for this function can be found for a finite number of data samples. Instead, it can be shown [Lju87, SS89] that the asymptotic probability density function of  $(\hat{\underline{\theta}} - \underline{\theta}^*)$  converges to a Gaussian distribution according to :

$$\sqrt{N} (\hat{\underline{\theta}} - \underline{\theta}^*) \in As N(0, \text{Cov}(\underline{\theta}^*)) \quad (2.53)$$

where

$$\text{Cov}(\underline{\theta}^*) = \left[ \nabla_{\underline{\theta}\underline{\theta}}^2 V_{e_{\infty}}(\underline{\theta}^*) \right]^{-1} \left[ \lim_{N \rightarrow \infty} N E \{ \nabla_{\underline{\theta}} V_e(\underline{\theta}^*) \nabla_{\underline{\theta}} V_e(\underline{\theta}^*)^T \} \right] \left[ \nabla_{\underline{\theta}\underline{\theta}}^2 V_{e_{\infty}}(\underline{\theta}^*) \right]^{-1} \quad (2.54)$$

To evaluate the parameter covariance matrix  $\text{Cov}(\hat{\underline{\theta}})$  it is necessary to use the properties of the data as well as the model structure. Knowing this asymptotic distribution, conclusions can be drawn on the accuracy of the parameter estimates, for example confidence intervals concerning the estimated parameters can specify the distance to the real parameters.

Under the assumption that the set  $\Theta_t$  consists precisely of one element  $\underline{\theta}_t$  and that the prediction error  $\underline{e}(k, \underline{\theta}_t)$  is white, the asymptotic covariance matrix for the multivariable case is given by [SS89] :

$$\begin{aligned} \text{Cov}(\underline{\theta}_t) &= [ E \{ \nabla_{\underline{\theta}} \underline{e}^T(k, \underline{\theta}_t) \nabla_{R_e} V(R_e) \nabla_{\underline{\theta}}^T \underline{e}(k, \underline{\theta}_t) \} ]^{-1} * \\ &\quad [ E \{ \nabla_{\underline{\theta}} \underline{e}^T(k, \underline{\theta}_t) \nabla_{R_e} V(R_e) \wedge \nabla_{R_e} V(R_e) \nabla_{\underline{\theta}}^T \underline{e}(k, \underline{\theta}_t) \} ] * \quad (2.55) \\ &\quad [ E \{ \nabla_{\underline{\theta}} \underline{e}^T(k, \underline{\theta}_t) \nabla_{R_e} V(R_e) \nabla_{\underline{\theta}}^T \underline{e}(k, \underline{\theta}_t) \} ]^{-1} \end{aligned}$$

which reduces for the minimal polynomial model structure to :

$$\text{Cov}(\underline{\theta}_t) = \sigma_e^2 [E\{\nabla_{\theta} e_{\text{col}}(\ell, \underline{\theta}_t) \nabla_{\theta}^T e_{\text{col}}(\ell, \underline{\theta}_t)\}]^{-1} \quad (2.56)$$

when taking Eq. 2.24 into account. This expression of the parameter covariance matrix, however, cannot be evaluated in practice because :

- Contrary to the assumption  $\Theta_t = \underline{\theta}_t = \underline{\theta}^*$ , approximate modelling in practice (underparametrization) results in  $\Theta_t = \emptyset$ .
- Contrary to the assumption that  $\underline{e}(k, \hat{\theta})$  is white, a white prediction error can often not be realized for industrial processes. The prediction error  $\underline{e}(k, \hat{\theta})$  is in general correlated up to time  $t-\delta$  where  $\delta$  indicates the correlation depth which is problem dependent. Further, the whiteness depends also on the accuracy of  $\hat{\theta}$  with respect to  $\underline{\theta}_t$ .
- The expected value  $E\{.\}$  cannot be calculated because only one manifestation of a stochastically assumed variable is available.

In addition, evaluating  $\text{Cov}(\underline{\theta}^*)$  (Eq. 2.56) at  $\hat{\theta}$  gives trivially  $\text{Cov}(\hat{\theta})=0$ , since  $\nabla_{\theta} V_e(\hat{\theta})=0$ , which is a useless estimate. Therefore, the question is how to calculate the best possible approximation of the parameter covariance matrix in practice. Assume now that the prediction error  $\underline{e}(k, \theta)$  is decomposed in a parameter dependent term,  $\underline{g}(k, \theta)$ , due to the undermodelling and a white noise term,  $\xi(k)$ , affecting the true system :

$$\underline{e}(k, \underline{\theta}) = \underline{g}(k, \underline{\theta}) + \xi(k) \quad (2.57)$$

Further, assume that the elements of the white noise sequence  $\xi(k)$  are mutually independent and that the input/output data has been scaled properly in advance to satisfy Eq. 2.21. Under these conditions the asymptotic parameter covariance matrix can be rewritten as :

$$\begin{aligned} \text{Cov}(\underline{\theta}^*) = \sigma_{\xi}^2 & \left[ \nabla_{\theta\theta}^2 V_e(\underline{\theta}^*) \right]^{-1} * \\ & \left[ \frac{4}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \nabla_{\theta} \varepsilon_{\text{col}}(\ell, \underline{\theta}^*) \nabla_{\theta}^T \varepsilon_{\text{col}}(\ell, \underline{\theta}^*) \right] * \\ & \left[ \nabla_{\theta\theta}^2 V_e(\underline{\theta}^*) \right]^{-1} \end{aligned} \quad (2.58)$$

where  $V_e(\underline{\theta}^*)$  defines the contribution of the term  $\underline{g}(k, \underline{\theta}^*)$  in the loss-function  $V_e(\underline{\theta}^*)$

(see Section C.1). This expression of the parameter covariance matrix  $\text{Cov}(\underline{\hat{\theta}}^*)$  can be approximated in practice as follows :

$$\text{Cov}(\underline{\hat{\theta}}) = \hat{\sigma}_\varepsilon^2 \left[ \frac{2}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \nabla_{\theta} e_{\text{col}}(\ell, \underline{\hat{\theta}}) \nabla_{\theta}^T e_{\text{col}}(\ell, \underline{\hat{\theta}}) + e_{\text{col}}(\ell, \underline{\hat{\theta}}) \nabla_{\theta\theta}^2 e_{\text{col}}(\ell, \underline{\hat{\theta}}) \right]^{-1} * \left[ \frac{4}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \nabla_{\theta} e_{\text{col}}(\ell, \underline{\hat{\theta}}) \nabla_{\theta}^T e_{\text{col}}(\ell, \underline{\hat{\theta}}) \right] * \left[ \frac{2}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \nabla_{\theta} e_{\text{col}}(\ell, \underline{\hat{\theta}}) \nabla_{\theta}^T e_{\text{col}}(\ell, \underline{\hat{\theta}}) + e_{\text{col}}(\ell, \underline{\hat{\theta}}) \nabla_{\theta\theta}^2 e_{\text{col}}(\ell, \underline{\hat{\theta}}) \right]^{-1} \quad (2.59)$$

The main difference between the expressions of the parameter covariance matrix for both situations, exact (Eq. 2.56) and undermodelling (Eq. 2.59), is basically the evaluation of the second-order derivative for the prediction error  $\nabla_{\theta\theta}^2 e_{\text{col}}(\ell, \underline{\hat{\theta}})$ . Note also the large similarity for the covariance matrix expression between the scalar [SS89, Section 7.5] and the multivariable case using the minimal polynomial model structure when exact modelling is considered. A detailed explanation and the proof of Eq. 2.59 can be found in Section C.1, Covariance Approximation.

When considering the situation of exact modelling, i.e.  $\Theta_t = \{ \theta_t \}$ , and realizing a white prediction error with Gaussian distribution, i.e. satisfying the assumptions 1-5 described in the introduction of this section, a well known result is that the lower bound on the covariance matrix is given by the Cramér-Rao inequality :

$$\text{Cov}(\underline{\theta}_t) \geq [ E \{ \nabla_{\theta} e(k, \underline{\theta}_t) \Lambda_t \nabla_{\theta}^T e(k, \underline{\theta}_t) \} ]^{-1} \quad (2.60)$$

The equality in Eq. 2.60 holds if the data has been properly scaled in advance with the true inverse covariance matrix  $\Lambda_t^{-1}$  resulting in Eq. 2.56 for the minimal polynomial model structure. This is a nice theoretical result, but system identification in practice can be described in general as approximate modelling because the estimated model is always a low order approximation of the real process. Therefore, the Cramér-Rao bound is in principle not valid. However, the parameter covariance matrix as defined in Eq. 2.59 provides a reasonable approximation. This expression can be considered as a modified Cramér-Rao bound in the situation that the process is not contained in the model set taking the fact of undermodelling explicitly into account. Therefore,

confidence intervals based on  $\text{Cov}(\hat{\underline{\theta}})$  can still be used to derive bounds on the parameters in this stochastic setting.

#### 2.4.4 Confidence Intervals

In spite of the fact that there exists no true parameter vector for practical identification problems, i.e.  $\Theta_t = \emptyset$ , the parameter estimate  $\hat{\underline{\theta}}$  converges to the best possible approximation denoted by  $\underline{\theta}^*$  and the asymptotic probability density function of  $(\hat{\underline{\theta}} - \underline{\theta}^*)$  converges to a Gaussian distribution as given in Eq. 2.53.

Finally remark, that if a random vector  $\underline{\eta}$  has a Gaussian distribution :

$$\underline{\eta} \in N(0, \text{Cov}) \quad (2.61)$$

then the scalar :

$$z_{\underline{\eta}} = \underline{\eta}^T \text{Cov}^{-1} \underline{\eta} \quad (2.62)$$

has a  $\chi^2$ -distribution with  $\dim(\underline{\eta}) = n\theta$  degrees of freedom,

$$z_{\underline{\eta}} \in \chi^2(n\theta) \quad (2.63)$$

From Eq. 2.55 and 2.56, the conclusion can be drawn that :

$$z_{\underline{\theta}} = N_{\text{col}} \left( \hat{\underline{\theta}} - \underline{\theta}^* \right)^T \text{Cov}^{-1}(\hat{\underline{\theta}}) \left( \hat{\underline{\theta}} - \underline{\theta}^* \right) \in \text{As } \chi^2(n\theta) \quad (2.64)$$

This implies that  $z_{\underline{\theta}}$  converges in distribution to the  $\chi^2(n\theta)$ -distribution as  $N_{\text{col}}$  tends to infinity. Using either  $\chi^2$ -tables or direct computation of the  $\chi^2$ -distribution as described in Section C.2, confidence intervals for  $z_{\underline{\theta}}$  can be derived which correspond to confidence ellipsoids for  $\hat{\underline{\theta}}$ .

The bounds of the parameters derived in this stochastic setting will be compared to the deterministic approach described in Chapter 3. In this comparison of stochastic and deterministic parameter uncertainty, the main attention will be focused on the estimated uncertainty bounds which will be obtained under various noise conditions for several types of models.

Although the prediction error identification is well known from literature, a major contribution in this section is the proposed minimal polynomial model structure for multivariable systems. The motivation for this model structure is that the complex and high interaction of dynamics in industrial processes makes the estimation of structural indices for (pseudo-) canonical models very difficult due to the structural uncertainty.

The parameter estimation when using multivariable (pseudo-) canonical models suffers from the fundamental difficulty that for this structure no convergence can be achieved for the nonlinear multivariable prediction error minimization problem. This problem is avoided by adopting the minimal polynomial model structure which results therefore in a very effective identification algorithm for multivariable systems.

Further, a modified Cramér-Rao bound is derived for which the assumption that the process must be contained in the model set can be relieved. The fact that every model is only a low order approximation of the real process is explicitly taken into account. The corresponding parameter uncertainty bounds provide a quality measure of the identified model in a stochastic setting.



## ***Bounded Error Identification***

---

3.1	Model Restrictions	3.3	Data Outliers
3.2	Outer Parameter Bounding	3.4	Comparison of Stochastic and Deterministic Uncertainty

---

In classical identification methods a dynamical model of a process is estimated using measured data and usually stochastic assumptions on the noise corrupting the data. The noise is statistically modelled by an at least partially known or specified distribution. This approach has been discussed in Chapter 2 for multivariable systems using minimal polynomial models. However, there are many situations where the main contribution to the error may not be of a random nature and therefore cannot be suitably described by random noise. For example, the industrial process from which the data is collected may be very complex so that only simplified models can be used in the identification procedure. Then, the residuals of the estimated model have a component due to deterministic structural errors, i.e. unmodelled dynamics, which may lead to unsatisfactory results when treating them as purely random variables.

An alternative approach, initiated by Witsenhausen and Schweppe [Sch68, Sch73, Wit68], is based on a much simpler representation of the prediction error which defines the misfit between the data and the model due to noise disturbing the process,

undermodelling etc. The unknown but bounded error assumption is used in the sense that each error sample of output  $i$ , is known to be bounded by a given value :

$$|e_i(k, \underline{\theta})| \triangleq |y_i(k) - \hat{y}_i(k|\underline{\theta})| \leq \delta_{e,i}(k) \quad \forall k, i \quad (3.1)$$

Frequently, the  $l_\infty$ -norm is adopted to simplify this noise bound :

$$\|E_{*i}(\underline{\theta})\|_\infty \triangleq \max_k |e_i(k, \underline{\theta})| \leq \delta_{e,i} \quad (3.2)$$

Motivation for this kind of error approach is the fact that, in many practical cases, the unknown but bounded error assumption is closer to the actually available information on, for example, measurement noise, truncation, round-off errors etc. In this context the identification procedure consists of finding the complete set of all admissible parameter values which is consistent with the measurements, the model and the error description. A general overview can be found in [Mil89, WPL90].

To obtain generally useful algorithms for parameter set modelling of multivariable systems, the identification will be restricted to models parametrized by linear difference equations with their coefficients defined as parameters. This is subject of discussion in Section 3.1. For models which are linear in their parameters, the feasible parameter region is described by a set of linear inequalities, which define a convex set in the parameter space with possibly very complicated bounds. It is therefore convenient to look for simpler although approximate descriptions like ellipsoidal or orthotopic bounding. As extension to the set estimation methods which have been described in the literature, it will be shown that for ellipsoidal bounding repeated processing of the set of inequalities is required to obtain convergence to the final ellipsoid of for example minimum volume. Furthermore, it will be motivated that the most accurate approximation of the parameter uncertainty set can be obtained by combining the features of ellipsoidal and orthotopic bounding resulting in ellipsoid-aligned orthotopic bounding. The different bounding types of the feasible parameter set are described in Section 3.2. For practical applications of bounded error identification, it should be noted that set estimators are not robust to outliers, i.e. data values which are not consistent with the specified assumptions. Such outliers may result from mistakes made during the acquisition and preprocessing of the data, but also from overoptimistic error bounds or unmodelled dynamics. As described in the introduction, this is always the case in practice and therefore bounded error identification will result in an empty parameter set if no proper precautions are taken. This problem, despite its severe consequences, is hardly mentioned or even not recognized in the literature. One way to circumvent an empty parameter set due to data outliers is outlined in Section 3.3. Finally, an extensive example is treated in Section 3.4, where the parameter uncertainty obtained by both stochastic and deterministic identification is compared under various noise conditions for different types of models. In this comparison of stochastic and deterministic

parameter uncertainty the main attention will be focused on the estimated uncertainty bounds and the consequences for these bounds if theoretical assumptions are violated. For simplicity of notation, only the SISO case will be described in this chapter. The extension to multivariable systems using the minimal polynomial model structure introduced in Chapter 2 is described in Appendix D, Parametric Uncertainty Descriptions.

### 3.1 Model Restrictions

As mentioned in Chapter 1, a general model representation of a SISO dynamical system relating the input, the measured output and the disturbances in discrete time, can be described by :

$$M(\underline{\theta}, y(k), u(k), \xi(k), k) = 0 \quad \forall k \quad (3.3)$$

Under the assumption that the process is contained in the model set, the prediction error  $e(k, \underline{\theta})$  which describes the misfit between the data and the model of a dynamical system :

$$e(k, \underline{\theta}) \triangleq y(k) - \hat{y}(k | \underline{\theta}) \quad (3.4)$$

is an estimate of the noise  $\xi(k)$  disturbing the process (see Eq. 2.57) for a specific parameter vector  $\underline{\theta}$ . When now the unknown but bounded error formulation is adopted (Eq. 3.1 and Eq. 3.2), the parameter uncertainty set  $\Theta$ , containing all the parameter values which are not falsified by the data, the process and the error model (Eq. 3.3), is defined by :

$$\Theta = \left\{ \underline{\theta} \in \mathbb{R}^{n_\theta} \mid M(\underline{\theta}, y(k), u(k), e(k, \underline{\theta}), k) = 0, \quad |e(k, \underline{\theta})| \leq \delta_e \right\} \quad (3.5)$$

If there exists a true parameter vector  $\underline{\theta}_t$ , i.e. the process is contained in the model set, and  $e(k, \underline{\theta}_t)$  satisfies the bounded error assumption (Eq. 3.1 and Eq. 3.2), then the parameter uncertainty set  $\Theta$  contains  $\underline{\theta}_t$ .

In the unknown but bounded error identification approach, estimating the parameter uncertainty set  $\Theta$  (Eq. 3.5) for time-invariant systems and adopting the setup as depicted in Fig. 1.1 where the disturbances  $\xi(k)$  are modelled by the prediction error  $e(k, \underline{\theta})$ , basically two types of models can be distinguished :

- Nonlinear in the parameters ; given the data  $\{u(k), y(k)\}$ , the mapping from the parameters  $\underline{\theta}$  to the prediction error  $e(k, \underline{\theta})$  is nonlinear.

- Linear in the parameters ; the model can be described as a linear mapping from the parameter vector  $\underline{\theta}$  to the prediction error  $e(k, \underline{\theta})$  for a given data set  $\{u(k), y(k)\}$ , i.e.  $\underline{\theta} \xrightarrow{\{u(k), y(k)\}} e(k, \underline{\theta})$ , or, in other words, the model is affine in  $\underline{\theta}$ .

**Table 3.1 : Linear vs. nonlinear parameter models.**

Quantity	Linear in $\underline{\theta}$	Nonlinear in $\underline{\theta}$
Parameter Uncertainty Set $\Theta$ .	Convex polytope.	Set that in general can be non-convex and non-connected.
Boundary of $\Theta$ in parameter space $\mathbb{R}^{n_\theta}$ .	$(n_\theta-1)$ -dimensional linear hyperplanes.	Hypersurfaces of any complex form.
Central estimate $\underline{\theta}_c$ (centre of minimum and maximum value).	It exists, is unique and belongs to the parameter uncertainty set $\Theta$ .	It may not be unique and may not belong to the parameter uncertainty set $\Theta$ .

The consequences of these model types for the bounded error identification are quite significant and have been summarized in Table 3.1 (see [BBC88]). To illustrate the main differences between the linear and the nonlinear parameter models in bounded error identification, two simple examples have been selected.

### Example

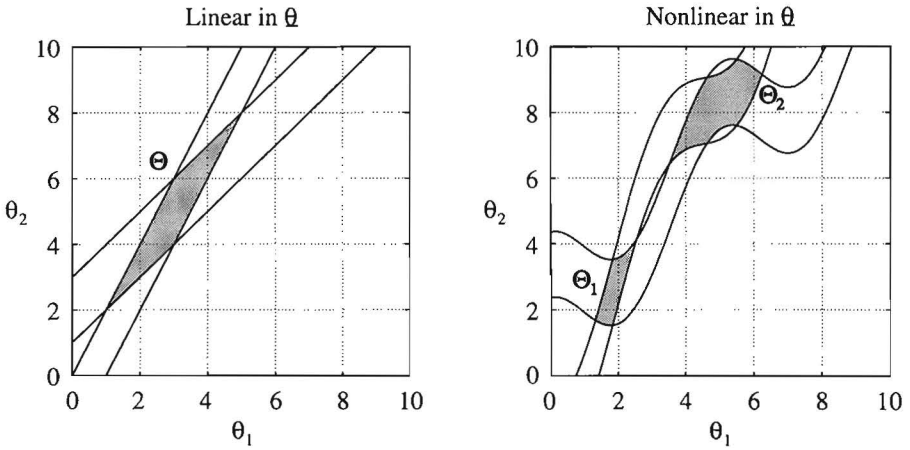
Suppose the true dynamical systems are defined by :

- *Linear in  $\underline{\theta}$*  :  $M = y(k) + k \theta_1 - \theta_2 + \xi(k) = 0$   
 $\Rightarrow \hat{y}(k|\underline{\theta}) = -k \theta_1 + \theta_2$
- *Nonlinear in  $\underline{\theta}$*  :  $M = y(k) + k \theta_1 - \theta_2 + 1.5 \sin (1.2 \theta_1 + 2 k) + \xi(k) = 0$   
 $\Rightarrow \hat{y}(k|\underline{\theta}) = -k \theta_1 + \theta_2 - 1.5 \sin (1.2 \theta_1 + 2 k)$

where only the sine term has been incorporated for the nonlinear case. These expressions have been evaluated at time instants  $k \in \{1, 2\}$  for the measured values  $y(k) = \{2, -1\}$ . Further, assume that measurement noise  $\xi(k)$  additive to the output of the true system disturbs the data which is bounded by  $\|\xi\|_\infty \leq 1$ . When the process is in the model set, the prediction error satisfies :  $|e(k, \underline{\theta})| = |y(k) - \hat{y}(k|\underline{\theta})| \leq \delta_e = 1$ . The corresponding boundaries derived from :

- *Linear in  $\underline{\theta}$*  :  $1 \leq -\theta_1 + \theta_2 \leq 3$   
 $-2 \leq -2\theta_1 + \theta_2 \leq 0$
- *Nonlinear in  $\underline{\theta}$*  :  $1 \leq -\theta_1 + \theta_2 - 1.5 \sin(1.2\theta_1 + 2) \leq 3$   
 $-2 \leq -2\theta_1 + \theta_2 - 1.5 \sin(1.2\theta_1 + 4) \leq 0$

have been depicted in Fig 3.1 for both models, linear and nonlinear in the parameters respectively. The shaded areas in both figures define the parameter uncertainty sets  $\Theta$ . These examples show clearly the advantages of the linear in the parameter model description. The parameter uncertainty set  $\Theta$  is convex and connected due to the linear constraints which define the boundary. This also guarantees that the centre value  $\underline{\theta}_c$  which is defined as the centre of the extreme values, i.e. minimum and maximum value, is an element of the parameter uncertainty set  $\Theta$ . All these advantages will be lost whenever the model is nonlinear in its parameters.



**Fig. 3.1** : Parameter uncertainty sets  $\Theta$  ; a) Linear in  $\underline{\theta}$ , b) Nonlinear in  $\underline{\theta}$ .

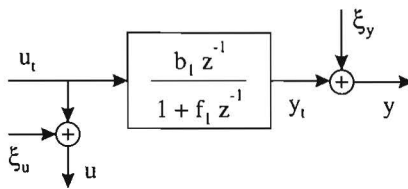
Most models used in practice are based on difference equations which are nonlinear in their parameters. For nonlinear parameter models the hyperplanes of the linear case are replaced by more general hypersurfaces, as shown in Fig. 3.1. The major problems of these nonlinear parameter models are that the corresponding parameter uncertainty sets  $\Theta_i$  are non-convex and possibly non-connected. Several approaches have been proposed to describe these parameter uncertainty sets  $\Theta_i$ . When there are up to three parameters and an analytical expression for the model is available, the hypersurfaces limiting the feasible parameter region can be drawn and visualized to see which observations define

the parameter uncertainty sets [Nor86]. Other approaches [Smi83, WPL86] use global optimization methods based on random search to construct the boundary of the parameter uncertainty sets. Projections onto one- or two-dimensional spaces are computed and a systematic scan of the projection space is performed.

Whenever the hypersurfaces can be described as explicit polynomial functions of positive parameters, the bounded error identification problem can be solved by signomial programming [MV89]. When  $\underline{\theta}$  cannot be assumed to be positive, a suitable transformation of the parameter space can be performed to obtain positive parameters. These signomial problems are in general non-convex and may thus exhibit local solutions. However, iterative algorithms exist which can evaluate at each iteration the lower and upper bounds of the sets which are guaranteed to converge monotonically to the global solution [Fal73, Eck80]. Thus the method provides orthotopic (box) bounds evaluating the extremes of the parameter uncertainty sets.

All these methods however, suffer from exploding computational complexity for increasing dimensionality which motivates the interest in looking for less detailed but more easily computable approximations of the feasible parameter regions. The most interesting solution to the nonlinear bounded error identification problem is to linearize the model around some value of the parameters to obtain linear hyperplanes. Using this approach, the advantages of the linear case are obtained again in the form of a convex and connected parameter uncertainty set  $\Theta$ . A simple example of an error-in-variables model will be used to illustrate this linearization.

### Example



**Fig. 3.2** : Errors-in-variables model.

The errors-in-variables model depicted in Fig. 3.2 is nonlinear in its parameters :  $\underline{\theta} = [b_1 \ f_1]^T$ . In this simple example it will be shown how linearized constraints can be derived describing the parameter uncertainty set  $\Theta$  in the parameter space. Suppose the true dynamical system is defined by :

$$y_t(k) = b_1 u_t(k-1) - f_1 y_t(k-1) \quad (3.6)$$

where the measured data is obtained according to :

$$\begin{aligned} y(k) &= y_t(k) + \xi_y(k) \\ u(k) &= u_t(k) + \xi_u(k) \end{aligned} \quad (3.7)$$

and the measurement noise satisfies the bounded error conditions :

$$|\xi_y(k)| \leq \delta_y(k) \quad , \quad |\xi_u(k)| \leq \delta_u(k) \quad (3.8)$$

The linear inequalities which define the boundary of the parameter uncertainty set  $\Theta$  can be determined by deriving bounds on each of the components of Eq. 3.6. Because of the bounded error conditions, every true output sample is limited by :

$$y(k) - \delta_y(k) \leq y_t(k) \leq y(k) + \delta_y(k) \quad (3.9)$$

For unknown parameters, this lower and upper bound can be adopted as follows :

$$\begin{aligned} f_1 > 0 : f_1 [y(k-1) - \delta_y(k-1)] &\leq f_1 y_t(k-1) \leq f_1 [y(k-1) + \delta_y(k-1)] \\ f_1 < 0 : f_1 [y(k-1) + \delta_y(k-1)] &\leq f_1 y_t(k-1) \leq f_1 [y(k-1) - \delta_y(k-1)] \end{aligned} \quad (3.10)$$

resulting in :

$$f_1 y_t(k-1) \begin{cases} \geq f_1 [y(k-1) - \text{sgn}(f_1) \delta_y(k-1)] \\ \leq f_1 [y(k-1) + \text{sgn}(f_1) \delta_y(k-1)] \end{cases} \quad (3.11)$$

where  $\text{sgn}(\cdot)$  denotes the sign function of its argument. Changing the sign :

$$-f_1 y_t(k-1) \begin{cases} \geq -f_1 [y(k-1) + \text{sgn}(f_1) \delta_y(k-1)] \\ \leq -f_1 [y(k-1) - \text{sgn}(f_1) \delta_y(k-1)] \end{cases} \quad (3.12)$$

results then in bounds for the second term on the right hand side of Eq. 3.6. Similarly, bounds can be derived for the first term :

$$b_1 u_t(k-1) \begin{cases} \geq b_1 [u(k-1) - \text{sgn}(b_1) \delta_u(k-1)] \\ \leq b_1 [u(k-1) + \text{sgn}(b_1) \delta_u(k-1)] \end{cases} \quad (3.13)$$

Substituting now Eq. 3.9, 3.12 and 3.13 into Eq. 3.6 results in the following bounds for the parameter uncertainty set  $\Theta$  :

$$\begin{aligned}
 \mathbf{H}_1 &: [\phi(k, \underline{\theta}) + \Delta\phi(k, \underline{\theta})] \underline{\theta} \geq y(k) - \delta_y(k) \\
 \mathbf{H}_2 &: [\phi(k, \underline{\theta}) - \Delta\phi(k, \underline{\theta})] \underline{\theta} \leq y(k) + \delta_y(k)
 \end{aligned}
 \tag{3.14}$$

where

$$\begin{aligned}
 \phi(k, \underline{\theta}) &= [u(k-1) \quad -y(k-1)] \\
 \Delta\phi(k, \underline{\theta}) &= [\text{sgn}(b_1) \delta_u(k-1) \quad \text{sgn}(f_1) \delta_y(k-1)] \\
 \underline{\theta} &= [b_1 \quad f_1]^T
 \end{aligned}
 \tag{3.15}$$

Note that the sign of all parameters must be known in order to construct these constraints. For this specific example only the sign information of the parameters is needed. In the more general case, however, the regression vector in the constraints of Eq. 3.14 becomes a function of  $\underline{\theta}$ . Further, these linearized conditions for the boundary of the parameter uncertainty set  $\Theta$  are necessary but not sufficient. Conservatism is introduced by the fact that due to a sign change of the parameters the corresponding bounds,  $\delta_u(k)$  and/or  $\delta_y(k)$  in Eq. 3.15, change in sign as well. This is physically unrealistic because the sign of this measurement noise at a certain time instant  $k$  is fixed and cannot change within time. Therefore, more freedom is allowed in the uncertainty than physically can occur resulting in only necessary conditions.

The type of models which can be described in a linear regression form will be reviewed briefly where the transformation of the models nonlinear in their parameters to a pseudo-linear regression form will be described in more detail. The bounding constraints  $\{\mathbf{H}_1, \mathbf{H}_2\}$  of Eq. 3.14 will be extended to more general models. Whenever a model is linear in the parameters, the data regression vector  $\phi(k, \underline{\theta})$  will become independent of the parameters  $\underline{\theta}$  and the bounding regression vector  $\Delta\phi(k, \underline{\theta})$  will reduce to zero. For models which are nonlinear in their parameters, an approximation is necessary resulting in a pseudo-linear regression form. Depending on the term  $\Delta\phi(k, \underline{\theta})$ , the bounding constraints  $\{\mathbf{H}_1, \mathbf{H}_2\}$  will represent parallel or non-parallel hyperplanes in the parameter space respectively. All parameter values  $\underline{\theta}$  which satisfy these conditions are considered as feasible solutions of the identification problem and define the parameter set  $\Theta$ . Note that by linearization of models which are nonlinear in their parameters, conservatism has been introduced extending the feasible parameter set  $\Theta$  with physically unrealistic models. In the trade-off between complex but accurate parameter uncertainty sets in the general case and the conservatism added by linearization of nonlinear models resulting in easy computable parameter bounds, the latter approach has been selected to derive deterministic parameter uncertainty bounds.



- *General prediction error models :*

$$A(z^{-1})y(k) = \frac{B(z^{-1})}{F(z^{-1})} u(k) + \frac{C(z^{-1})}{D(z^{-1})} \xi(k) \quad (3.16)$$

Let :

$$\begin{aligned} w(k, \underline{\theta}) &= \frac{B(z^{-1})}{F(z^{-1})} u(k) \\ v(k, \underline{\theta}) &= A(z^{-1}) y(k) - w(k, \underline{\theta}) \end{aligned} \quad (3.17)$$

Then, the prediction error  $e(k, \underline{\theta})$  for this general model representation which describes the noise sequence  $\xi(k)$  based on the data  $\{u(k), y(k)\}$  for a specific parameter vector  $\underline{\theta}$ , is defined by :

$$e(k, \underline{\theta}) = \frac{D(z^{-1})}{C(z^{-1})} v(k, \underline{\theta}) \quad (3.18)$$

This general prediction error model can be rewritten in the following pseudo-linear regression form (App. B.3) :

$$\hat{y}(k|\underline{\theta}) = [ \underbrace{\phi_y(k)}_a \quad \underbrace{\phi_u(k)}_b \quad \underbrace{\phi_e(k, \underline{\theta})}_c \quad \underbrace{\phi_v(k, \underline{\theta})}_d \quad \underbrace{\phi_w(k, \underline{\theta})}_f ] \underline{\theta} \quad (3.19)$$

← parameters

where <sup>1)</sup>

$$\begin{aligned} \phi_y(k) &= [ -y(k-1) \dots -y(k-na) ] \\ \phi_u(k) &= [ u(k) \dots u(k-nb+1) ] \\ \phi_e(k, \underline{\theta}) &= [ e(k-1, \underline{\theta}) \dots e(k-nc, \underline{\theta}) ] \\ \phi_v(k, \underline{\theta}) &= [ -v(k-1, \underline{\theta}) \dots -v(k-nd, \underline{\theta}) ] \\ \phi_w(k, \underline{\theta}) &= [ -w(k-1, \underline{\theta}) \dots -w(k-nf, \underline{\theta}) ] \end{aligned} \quad (3.20)$$

which is clearly parameter dependent. To construct the constraints  $\{\mathbf{H}_1, \mathbf{H}_2\}$  which describe the parameter uncertainty set  $\Theta$ , lower and upper bounds for the data  $\{u(k), y(k)\}$ , the prediction error  $e(k, \underline{\theta})$  and the auxiliary variables

---

<sup>1)</sup>  $k$  indicates a row in the regression matrix  $\Phi(\underline{\theta})$  (see Eq. B.17). In the multivariable case the time index  $k$  is replaced by  $\ell$  indicating time as well as channels.

$\{w(k, \underline{\theta}), v(k, \underline{\theta})\}$  must be determined. To achieve this, the following procedure is required :

- 1) Suppose an initial parameter uncertainty set  $\Theta^0$  is available which has been determined from a priori information.
- 2) For all parameters  $\underline{\theta} \in \Theta^0 \subset \mathbb{R}^{n\theta}$ , lower and upper bounds of all variables, i.e.  $u(k)$ ,  $y(k)$ ,  $e(k, \underline{\theta})$ ,  $v(k, \underline{\theta})$  and  $w(k, \underline{\theta})$  must be computed, for example by simulating these data sequences for all elements of  $\Theta^0$ . The corresponding lower and upper bounds, e.g.  $w_{\min}(k)$  and  $w_{\max}(k)$  :

$$w_{\min}(k) = \min_{\forall \underline{\theta} \in \Theta^0} w(k, \underline{\theta}) \quad , \quad w_{\max}(k) = \max_{\forall \underline{\theta} \in \Theta^0} w(k, \underline{\theta}) \quad (3.21)$$

can then be used to compute a centre value  $w_c(k)$  and the data uncertainty  $\delta_w(k)$  according to :

$$\begin{aligned} w_c(k) &= \frac{1}{2} (w_{\max}(k) + w_{\min}(k)) \\ \delta_w(k) &= \frac{1}{2} (w_{\max}(k) - w_{\min}(k)) \end{aligned} \quad \forall k \quad (3.22)$$

In the bounded error identification context, the constraints  $\{\mathbf{H}_1, \mathbf{H}_2\}$  describing the boundary of the feasible parameter set based on these centre values and uncertainty information, are then defined by :

$$\begin{aligned} \phi(k, \underline{\theta}) &= [\phi_y(k) \quad \phi_u(k) \quad \phi_e(k, \underline{\theta}) \quad \phi_v(k, \underline{\theta}) \quad \phi_w(k, \underline{\theta})] \\ \Delta \phi(k, \underline{\theta}) &= [\Delta \phi_y(k) \quad \Delta \phi_u(k) \quad \Delta \phi_e(k, \underline{\theta}) \quad \Delta \phi_v(k, \underline{\theta}) \quad \Delta \phi_w(k, \underline{\theta})] \end{aligned} \quad (3.23)$$

where the regression vector  $\phi(k, \underline{\theta})$  is defined similar to Eq. 3.20 thereby replacing the various data samples, e.g.  $w(k-1, \underline{\theta})$ , by the corresponding centre values, e.g.  $w_c(k-1)$ , and :

$$\begin{aligned} \Delta \phi_y(k) &= [0 \quad \dots \quad 0] \\ \Delta \phi_u(k) &= [0 \quad \dots \quad 0] \\ \Delta \phi_e(k, \underline{\theta}) &= [\text{sgn}(c_1) \delta_e(k-1) \quad \dots \quad \text{sgn}(c_{nc}) \delta_e(k-nc)] \\ \Delta \phi_v(k, \underline{\theta}) &= [\text{sgn}(d_1) \delta_v(k-1) \quad \dots \quad \text{sgn}(d_{nd}) \delta_v(k-nd)] \\ \Delta \phi_w(k, \underline{\theta}) &= [\text{sgn}(f_1) \delta_w(k-1) \quad \dots \quad \text{sgn}(f_{nf}) \delta_w(k-nf)] \end{aligned} \quad (3.24)$$

- 3) Compute for the constraints defined in Eq. 3.14 using the regression matrices in Eq. 3.23, a new parameter uncertainty set  $\Theta^1 \subseteq \Theta^0$ .
- 4) Repeat step 2) and 3) until no significant reduction of the parameter uncertainty set can be obtained anymore, i.e.  $\Theta^{i+1} \approx \Theta^i$ .

Besides the rather complex procedure to compute the parameter uncertainty set  $\Theta$  together with the centre values and uncertainty bounds (Eq. 3.22) for all data variables over all time instants, this approach may yield very pessimistic parameter bounds and is therefore not recommended. For special cases of the general model structure (Eq. 3.16), this iterative procedure can be avoided.

- *Equation-error models :*

$$A(z^{-1}) y(k) = B(z^{-1}) u(k) + \xi(k) \quad (3.25)$$

This type of models is essentially linear in the parameters and can therefore be rewritten in a straightforward way into a linear regression form where the regression vector is defined by :

$$\begin{aligned} \phi_{ee}(k) &= [\phi_y(k) \phi_u(k)] \\ \Delta \phi_{ee}(k) &= 0 \end{aligned} \quad (3.26)$$

- *Output-error models :*

$$y(k) = \frac{B(z^{-1})}{F(z^{-1})} u(k) + \xi(k) \quad (3.27)$$

The approximation of the nonlinear part  $F(z^{-1})$  [CG88] has already been illustrated for the error-in-variables model in the preliminary example. For output-error models,  $w_c(k, \underline{\theta})$  and  $\delta_w(k)$  are equivalent to  $y(k)$  and  $\delta_e(k)$ , respectively, which yields a pair of pseudo-linear bounds modifying the regression vector into :

$$\begin{aligned} \phi_{oe}(k) &= [\phi_u(k) \phi_y(k)] \\ \Delta \phi_{oe}(k, \underline{\theta}) &= [\Delta \phi_u(k) \Delta \phi_w(k, \underline{\theta})] \end{aligned} \quad (3.28)$$

- *ARMAX models :*

$$A(z^{-1}) y(k) = B(z^{-1}) u(k) + C(z^{-1}) \xi(k) \quad (3.29)$$

Similar to the linearization of the error-in-variables model, an approximation of the boundary can be derived. Assuming a prediction error  $e(k, \underline{\theta})$  with zero mean for which  $e_c(k)$  reduces to zero and a known bound  $\delta_e(k)$ , the pseudo-linear regression vector for ARMAX models can be modified into :

$$\begin{aligned}\phi_{\text{armax}}(k) &= [\phi_y(k) \phi_u(k) 0 \dots 0] \\ \Delta\phi_{\text{armax}}(k, \underline{\theta}) &= [\Delta\phi_y(k) \Delta\phi_u(k) \Delta\phi_e(k, \underline{\theta})]\end{aligned}\quad (3.30)$$

- *Errors-in-variables models :*

$$\begin{aligned}y_t(k) &= b_1 u_t(k-1) - f_1 y_t(k-1) \\ y(k) &= y_t(k) + \xi_y(k) \\ u(k) &= u_t(k) + \xi_u(k)\end{aligned}\quad (3.31)$$

The boundary of these models (see Fig. 3.2) has been derived in the preliminary example and can be generalized for higher order models by the following regression vectors :

$$\begin{aligned}\phi_{\text{eiv}}(k) &= [\phi_u(k) \phi_y(k)] \\ \Delta\phi_{\text{eiv}}(k, \underline{\theta}) &= [\text{sgn}(b_1) \delta_u(k) \dots \text{sgn}(b_{nb}) \delta_u(k - nb + 1) \Delta\phi_w(k)]\end{aligned}\quad (3.32)$$

where  $\delta_w(k)$  in  $\Delta\phi_w(k, \underline{\theta})$  is equivalent with  $\delta_y(k)$  in Eq. 3.8.

Due to the linearization of models which are nonlinear in their parameters, the signs of the corresponding parameters must be known. The parameter uncertainty set  $\Theta$  can then be evaluated from a priori knowledge or from classical identification techniques assuming that the estimated parameters are elements of the feasible set and that the corresponding signs will be fixed to either positive or negative values. Whenever a priori information about the signs of the parameters is not available,  $2^q$  identification problems have to be solved, where  $q$  defines the number of parameters for which the sign must be known, evaluating all possible parameter combinations in order to find the parameter uncertainty set  $\Theta$ . Because of the exploding computational complexity for an increasing number of parameters, this latter approach is not recommended.

### 3.2 Outer Parameter Bounding

Applying bounded error identification to obtain complete and least conservative parameter bounds, requires a tight outer bound description of the parameter uncertainty

set  $\Theta$ . To ensure that all parameters will be identified which are consistent with the measurements, the model and the error description, i.e. the complete set, an outer bound approach is required. This outer parameter bounding should of course be as accurate as possible to reduce the conservatism when approximating the parameter uncertainty set  $\Theta$ . Outer bound approximations often turn out to be rather pessimistic descriptions of the parameter uncertainty set  $\Theta$ . It may therefore seem of interest to compute maximal inner bounds for  $\Theta$ . Such inner bounds may provide a useful indication of the tightness of the approximate outer bounds. However, simulation examples have shown that these inner bounds often shrink quickly to zero due the approximation which is involved when computing the inner bounds.

Whenever the models are linear in their parameters or can be approximated by a pseudo-linear regression form as described in the previous section, the complete least conservative parameter uncertainty set  $\Theta$  consists of a convex and connected set which is bounded by linear hyperplanes in the parameter space [BBC88]. Note however, that in this bounded error identification approach conservatism is introduced, i.e.  $\Theta$  might contain physically unrealistic models, due to the fact that all constraints  $\{\mathbf{H}_1, \mathbf{H}_2\}$  are processed independently without taking the relation between each equation at time instant  $k$  into account.

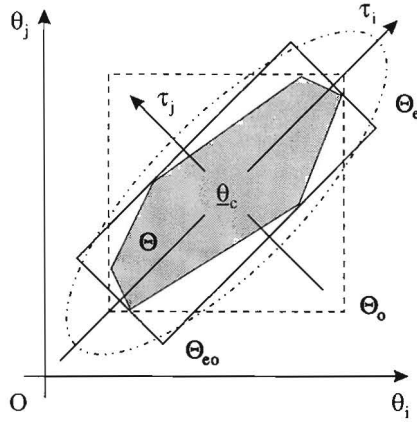
In this section several techniques will be reviewed to describe or approximate this ideal parameter uncertainty set  $\Theta$  thereby emphasizing the exchange between accuracy and complexity in computing time. To visualize the several outer bounding techniques :

- Exact (ideal) polytope bounding ( $\Theta$ )
- Ellipsoidal bounding ( $\Theta_e$ )
- Orthotopic bounding ( $\Theta_o$ )
- Ellipsoid-aligned orthotopic bounding ( $\Theta_{eo}$ )

a convex region for a two-dimensional example together with all outer parameter bounding descriptions has been shown in Fig. 3.3.

### 3.2.1 Exact Polytope Bounding

Several algorithms exist to derive an exact description of the parameter uncertainty set  $\Theta$  based on work of [MRTT53]. Two basic algorithms, exact polytope updating [Mo89] and exact cone updating [PLW88], which update the feasible set recursively show large similarities. As will be shown later, the exact bounding technique becomes inapplicable for high dimensional problems. Therefore, only the basic procedure of exact polytope updating will be described here. A more detailed description of the algorithm can be found in Appendix D.

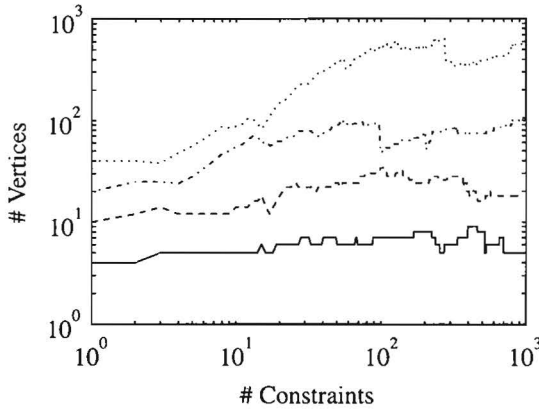


**Fig. 3.3 :** Parameter set estimation ; Ideal  $\Theta$  (shaded area) ; Best possible set for given linear constraints, Orthotope  $\Theta_o$  (dotted), Ellipse  $\Theta_e$  (dashed), Ellipsoid-aligned orthotope  $\Theta_{eo}$  (solid).

The parameter uncertainty set  $\Theta$  (Fig. 3.3) can be represented by its vertices and its edges. When a new constraint is processed, the intersection with the existing polytope is computed. The new vertices can be computed from the intersection of the new constraint and the edges of the existing polytope. Old vertices outside the new polytope will be eliminated and the edges must be updated. For the updating of the vertex set  $M_{vs}$ , the adjacent vertices and the neighbouring constraint hyperplanes of  $M_{vs}$  must be identified. In exact polytope updating, the vertex-vertex adjacency list and the vertex-plane adjacency list are kept and updated as part of the polytope updating procedure whenever a new constraint is added to the set. An initial polytope  $\Theta^0$  must be specified before updating can start. If there is no a priori information about the parameter uncertainty set,  $\Theta^0$  can be of any polytope shape as long as its boundary does not influence the resulting feasible parameter set.

A great disadvantage of exact polytope updating however is its exploding computational complexity for increasing number of parameters which makes this approach inapplicable in practice for multivariable systems. This becomes obvious when an example is considered where an initial polytope  $\Theta^0$  is for example defined as a rectangular box consisting of  $2^{n\theta}$  vertices (1024 vertices for  $n\theta = 10$ ) which increases when updating the polytope. This has been visualized in Fig. 3.4 where the number of vertices versus an increasing amount of constraints is shown for simulation examples ranging from two to five dimensional parameter sets. When a sufficiently large number of constraints has been processed, saturation in the number of parameters becomes visible, i.e. the number

of new vertices added to the vertex-list is approximately equal to the number of old vertices which have been eliminated. The dips in Fig. 3.4 indicate that more vertices have been eliminated than added when updating the vertex list. The final number of vertices, however, is highly problem dependent. In this specific example, the logarithmic scale in Fig. 3.4 indicates approximately an exponential increase in the number of vertices as a function of  $n\theta$ . In this case the updating of the vertex-vertex and vertex-plane adjacency lists becomes very slow (days of computing time) and the resulting matrices cannot be handled anymore for a large number of parameters. Therefore approximate parameter set descriptions are often preferred.



**Fig. 3.4** : Exploding complexity of exact polytope updating for output-error models with uniform distributed noise ; # Parameters : 2 (solid), 3 (dashed), 4 (dash-dot) and 5 (dotted).

### 3.2.2 Ellipsoidal Bounding

This approach is based on a recursive construction of ellipsoidal sets [Sch68, FH82] enclosing the parameter uncertainty set  $\Theta$  (Eq. 3.5). It is numerically simple and computationally fast. In its original form, two parallel hyperplanes  $\{\mathbf{H}_1, \mathbf{H}_2\}$  are required to update the ellipsoid. This implies that the term  $\Delta\phi(k, \underline{\theta})$  in Eq. 3.14 should be zero, which is only valid for equation-error models.

Let  $\mathbf{E}^{k-1}$  be a bounding ellipsoid obtained from the first  $(k-1)$  entries of the data vector :

$$\mathbf{E}^{k-1}(\underline{\theta}_c^{k-1}, \mathbf{P}^{k-1}) = \left\{ \underline{\theta} \in \mathbb{R}^{n\theta} : \left( \underline{\theta} - \underline{\theta}_c^{k-1} \right)^T [\mathbf{P}^{k-1}]^{-1} \left( \underline{\theta} - \underline{\theta}_c^{k-1} \right) \leq 1 \right\} \quad (3.33)$$

where  $\underline{\theta}^{k-1}$  defines the centre of  $\mathbf{E}^{k-1}$  and  $[\mathbf{P}^{k-1}]^{-1}$  a matrix specifying its orientation and size ( $\mathbf{P}^{k-1}$  is symmetrical positive definite). Let  $\mathbf{S}^k$  be the region which is defined by two parallel hyperplanes in the parameter space associated with the new measurement  $y(k)$  :

$$\mathbf{S}^k = \left\{ \underline{\theta} \in \mathbb{R}^{n\theta} : (y(k) - \phi(k, \underline{\theta}))^2 \leq \delta_e^2(k) \right\} \quad (3.34)$$

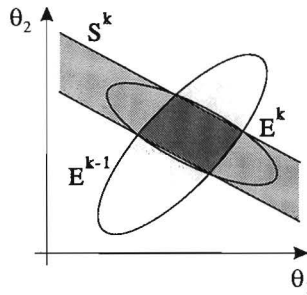


Fig. 3.5 : Ellipsoidal bounding.

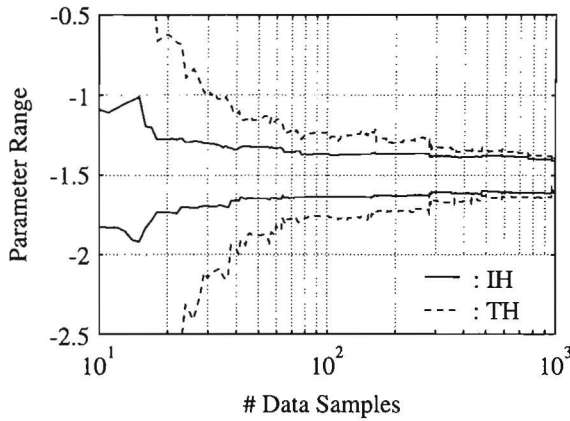
$\mathbf{E}^k$  is now defined as the minimal volume ellipsoid containing the intersection of  $\mathbf{E}^{k-1}$  and  $\mathbf{S}^k$  (Fig. 3.5), i.e.  $\mathbf{E}^{k-1} \cap \mathbf{S}^k \subset \mathbf{E}^k$ . The volume is proportional to the determinant of  $\mathbf{P}^k$ . This ellipsoidal approximation of  $\Theta$  yields the smallest possible ellipsoidal parameter set  $\Theta_e$ . Another family of ellipsoids is defined by the minimal trace of  $\mathbf{P}^k$  [FH82, AFA91]. This approach minimizes the sum of the parameter uncertainties which results in an ellipsoidal parameter set  $\Theta_e$  with minimal axes.

For models, however, where the parameter uncertainty set  $\Theta$  is described by linearized constraints (e.g. output-error models), the parameter space in Eq. 3.14 defines a subspace limited by two non-parallel hyperplanes  $\{\mathbf{H}_1, \mathbf{H}_2\}$ . In this case, parallel hyperplanes have to be constructed, before ellipsoidal bounding can be applied. In [CG90] a two step algorithm is proposed by constructing hyperplanes tangent to  $\mathbf{E}^{k-1}$  and parallel to the hyperplanes  $\{\mathbf{H}_1, \mathbf{H}_2\}$  to approximate  $\mathbf{E}^k$  from the parameter space  $\mathbf{S}^k$  and the ellipsoid  $\mathbf{E}^{k-1}$ . An ellipsoid of significant smaller volume can be obtained however, by constructing parallel hyperplanes through the intersection points of  $\mathbf{E}^{k-1}$  and  $\{\mathbf{H}_1, \mathbf{H}_2\}$  as has been proposed in [Fal92]. A more detailed description of this approach is given in Appendix D.3 (compare Fig. D.2 and Fig. D.4). This new approach ensures that the parallel hyperplanes which have been constructed for the intersection of the parameter space  $\mathbf{S}^k$  and the ellipsoid  $\mathbf{E}^{k-1}$ , are as tight as possible.

The approach of tangent hyperplanes (TH) makes the parameter uncertainty interval



unnecessarily large if no a priori knowledge of the parameter uncertainty set  $\Theta$  is available, so the initial ellipsoid  $P^0$  has to be chosen large. Using intersecting hyperplanes (IH), a significant reduction of the parameter uncertainty intervals can be achieved especially for short data sequences as shown in Fig. 3.6 for a simple simulation example.



**Fig. 3.6** : Parameter range  $f_1$  for intersecting (solid) and tangent (dashed) hyperplanes ; A 2<sup>nd</sup> order output error model,  $x(k) = 1.5 x(k-1) - 0.7 x(k-2) + u(k-1) + 0.5 u(k-2)$ ,  $y(k) = x(k) + \xi(k)$ , has been simulated using a PRBN sequence of  $N = 1023$  data samples as input signal,  $u(k) \in [-1,1]$ , with uniform distributed noise  $|\xi(k)| \leq 0.1$ .

In order to obtain a final ellipsoid of minimal volume, all constraints have to be processed repeatedly. This is required because ellipsoidal bounding computes recursively a conservative approximation of the parameter uncertainty set  $\Theta$ . Therefore, the result becomes dependent of the initialization and the order in which the constraints are processed. Sequential processing of all constraints where the ellipsoid of the previous iteration step is used as new initial condition, will decrease this dependence.

When minimizing the volume  $V_E$  of the ellipsoid, there holds :

$$0 < V_{E^k} \leq V_{E^{k-1}} \quad (3.35)$$

It has been shown in [FH82, VN91] that the parameter uncertainty set  $\Theta$  converges to a single parameter vector  $\theta_*$  if the number of data samples tends to infinity, provided that the noise sequence is at sufficiently many time instants close to the specified noise bounds.

The advantage of this method is the recursive and easy computable approximation  $\Theta_e$  of the parameter uncertainty set  $\Theta$  together with the ability to detect a possible orientation in the parameter space. The disadvantage, on the other hand, is that the resulting parameter set  $\Theta_e$  can be rather conservative compared to the actual parameter uncertainty set  $\Theta$ .

### 3.2.3 Orthotopic Bounding

An alternative description of the parameter uncertainty set  $\Theta$  can be obtained by orthotopic (box) bounding based on linear programming. This characterization competes with ellipsoidal bounding in terms of the number of parameters required to approximate the set  $\Theta$  and the computational complexity. For the ellipsoidal bounding  $n\theta$  parameters are needed for the centre of the ellipsoid, i.e.  $\underline{\theta}_c$ , and  $\frac{1}{2}n\theta(n\theta+1)$  for the orientation and size of the ellipsoid, i.e.  $P$ . For the orthotopic bounding just a minimum and a maximum value are required for each parameter leading to a total of  $2n\theta$  parameters. The parameter set  $\Theta_o$  is consequently defined by :

$$\Theta_o = \left\{ \underline{\theta} \in \mathbb{R}^{n\theta} : \theta_i = \theta_{c,i} + s_i \delta_i, \delta_i \in [-1, 1], i = 1 \dots n\theta \right\} \quad (3.36)$$

where

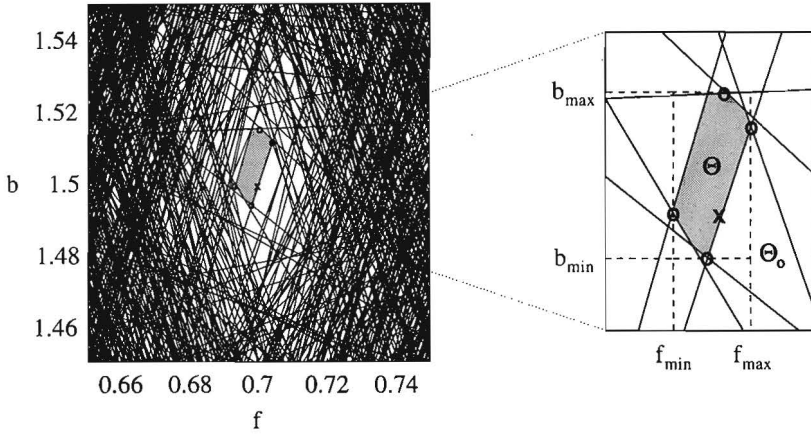
$$\theta_{c,i} = \frac{1}{2} (\theta_{\max,i} + \theta_{\min,i}) \quad , \quad s_{\theta,i} = \frac{1}{2} (\theta_{\max,i} - \theta_{\min,i}) \quad (3.37)$$

and

$$\theta_{\min,i} = \min_{\theta_i \in \Theta} \theta_i \quad , \quad \theta_{\max,i} = \max_{\theta_i \in \Theta} \theta_i \quad (3.38)$$

Thus, the computation of the orthotopic bounding (Eq. 3.38) requires the solution to  $2n\theta$  linear programming problems each with  $2(N-n)$  linear inequalities, i.e. the constraints  $H_i$  defined in Eq. 3.14, where  $N$  indicates the length of the data set and  $n$  denotes the number of samples required for initialization. Fig. 3.7 shows clearly that only a small set of constraints define the extreme parameter values,  $\theta_{\min,i}$  and  $\theta_{\max,i}$  respectively. In these vertices of the parameter uncertainty set  $\Theta$ , the inequalities  $\{H_1, H_2\}$  which define the boundary, become equalities. This set of equality constraints is called the active set of the parameter vector  $\underline{\theta}$ . Since equality constraints are easier to treat than inequalities which suffer from the fundamental difficulty that the active set is unknown, is a motivation to use active set methods [GMW91] to solve the linear programming problem. The presence of linear equality constraints actually reduces the dimensionality in which the optimization occurs. The idea of active set methods is to treat inequality constraints by developing a prediction (working set) of the constraints active in the solution. This can be achieved by optimizing the objective function along

the boundary of the parameter uncertainty set  $\Theta$ . The working set may change at each iteration and constraints in the working set are (temporarily) treated as equality constraints during a given iteration. The solution to the problem can now be derived in two steps. First an initial feasible solution together with the corresponding initial working set has to be derived. The final solution can then be found by solving the  $2n\theta$  linear programming problems (Eq. 3.38) for the constraints defined in Eq. 3.14 using the initial solution. A more detailed description is given in Appendix D.4 or can be found in various textbooks on optimization [GMW81, GMW91].



**Fig. 3.7 :** Parameter space  $\Theta$  ; Constraints of  $x(k) = -0.7x(k-1) + 1.5u(k)$ ,  $y(k) = x(k) + \xi(k)$ , which has been simulated using a PRBN sequence of  $N = 1023$  data samples as input signal,  $u(k) \in [-1,1]$  and uniformly distributed noise  $|\xi(k)| \leq 0.1$ . The true parameter vector  $\theta_t = [0.7, 1.5]^T$  and the extreme values  $\theta_{\min,i}$  and  $\theta_{\max,i}$  of the orthotopic parameter set  $\Theta_0$  are indicated by  $x$  and  $o$  respectively.

Although the extreme points (Eq. 3.38) in orthotopic bounding can be computed exactly (Fig. 3.7), it is obvious that whenever the orientation of the parameter uncertainty set  $\Theta$  is not aligned with the coordinate axis, the parameter set  $\Theta_0$  (Fig. 3.3) can become very conservative.

### 3.2.4 Ellipsoid-aligned Orthotopic Bounding

Considering the conservatism which can result when the parameter uncertainty set  $\Theta$  is

approximated by ellipsoidal or orthotopic outer bounding descriptions, the objective of this method is to obtain an orthotope of minimal volume enclosing  $\Theta$ , where the orthotope may rotate in all directions to obtain this minimality. Explicit calculation with volume minimization as criterion [Kee92], however, will become an unsolvable problem because of exploding computational complexity for an increasing number of parameters. An implicit solution where the advantages of both the ellipsoidal and the orthotopic bounding method are combined is proposed in [Mo89]. The ellipsoidal bounding algorithm is used as a preprocessing tool to obtain the orientation of the parameter uncertainty set  $\Theta$ . The new parameter bounds can then be computed by solving the following linear programming problems :

$$\tau_{\min,i} = \min_{\underline{\theta} \in \Theta} U_{*i}^T \underline{\theta} \quad , \quad \tau_{\max,i} = \max_{\underline{\theta} \in \Theta} U_{*i}^T \underline{\theta} \quad (3.39)$$

Then  $\underline{\tau} = U^T \underline{\theta}$  defines a new orthonormal basis derived from  $[P^k]^{-1}$ . This approach can reduce the parameter set significantly (compare  $\Theta_o$  and  $\Theta_{eo}$  in Fig. 3.3). Note, however, that there is no guarantee that the resulting orthotope is the smallest one enclosing the exact parameter uncertainty set  $\Theta$ . This is due to the fact that the final ellipsoid enclosing  $\Theta$  depends highly on the initial ellipsoid  $P^o$  and the order in which the constraints (Eq. 3.14) are processed.

Therefore this approach has to be modified where a further improvement is proposed in [FD93] which reduces the initialization and constraint dependency by repeated ellipsoidal bounding. Suppose sequential processing of the constraints results in ellipsoidal matrices  $P_u$  and  $P_v$  respectively, where the result of the previous iteration step has been used as new initial conditions. Orthonormal ellipsoid-aligned bases,  $U$  and  $V$  can be found by singular value decomposition of the inverse of the corresponding ellipsoidal matrices  $P_u$  and  $P_v$ . The columns  $U_{*i}$  and  $V_{*i}$  of these orthonormal bases are ordered according to decreasing singular values. The angles  $\chi_i$  between these bases can therefore be derived from  $\cos(\chi_i) = U_{*i}^T V_{*i}$ . The stop criterion for iteratively processing the constraints then becomes :

$$1 - \|\underline{x}\|_1 / n\theta \leq \text{tol} \quad (3.40)$$

where  $\underline{x} = \text{diag}(U^T V)$ , i.e. the diagonal terms of  $U^T V$  placed in a vector  $\underline{x}$ , and 'tol' is a user defined tolerance margin which defines an upper bound for the maximum angular change between these bases. Proceeding in this way ensures that no change in basis occurs when processing the constraints again which defines therefore an optimal basis for orthotopic bounding. The resulting ellipsoid-aligned orthotope will then be of minimum volume.

An alternative approach to obtain a basis for aligned orthotopic bounding is classical identification. For the general minimal polynomial model structure which has been

described extensively in Appendix B, a nominal model can be computed using prediction error minimization. The corresponding parameter covariance matrix for which an approximation has been derived in Appendix C, defines a quality measure of the parameter estimates. This positive definite covariance matrix describes, similar as  $P^k$  in ellipsoidal bounding, the orientation and the stochastic parameter uncertainty (see Section 2.4) in the parameter space and can therefore be used to determine a basis for orthotopic bounding.

### 3.3 Data Outliers

Parameter set estimation methods are based on the assumption that the process is consistent with the presumed error bounds and model structure. However, in practice this assumption will often be violated due to data outliers. Such outliers may result from mistakes introduced during the acquisition and the preprocessing of the data, or alternatively, from overoptimistic upper bounds of the disturbances or unmodelled dynamics. Proper precautions are necessary to obtain a non-empty parameter set  $\Theta$ . This problem cannot be solved by just increasing the model order because this approach results in unnecessarily high model orders. Alternatively, increasing the error bounds until  $\Theta$  becomes non-empty will result in identifiability problems.

The outliers considered in this section consist of a relative small amplitude compared to the amplitude of the data itself. Severe outliers which are clearly visible in the data can be eliminated by proper data preprocessing (Appendix A). For each time instant  $k$ , Eq. 3.14 defines a feasible subset  $\Theta^k$  of the parameter space which is compatible with the data for the presumed error bounds and model structure. Suppose now that a parameter vector  $\underline{\theta} \in \mathbb{R}^{n\theta}$  has been selected. Then, by definition a constraint is called violating with respect to  $\underline{\theta}$  if  $\underline{\theta} \notin \Theta^k$ , i.e. the model is not compatible with the constraint.

Outlier Minimal Number Estimator (OMNE) provides a solution for this problem which has been proposed in [LWG87]. OMNE consists of two steps assuming that  $\Theta = \emptyset$  :

- The first step is used to obtain a model which is violated by a minimum number of constraints. Parameter vectors  $\underline{\theta}$  are selected by global optimization using random search. For each  $\underline{\theta}$ , the number of violating constraints is determined and that parameter vector is selected which violates the least number of constraints.
- The second step aims at exploring the boundary of the corresponding parameter set when skipping the violating constraints.

A disadvantage of this method is that there is no guarantee to find the optimal solution

due to the global optimization procedure of the first step. The basic idea in this approach is that the model, i.e.  $\underline{\theta}$ , is violated by some constraints.

An alternative method will be proposed based on the idea that some constraints wrongly violate the model, e.g. due to outliers in the data. Assume now that an initial model is available, for example obtained by prediction error identification, which defines an element  $\underline{\theta}_{\text{ref}}$  of the parameter uncertainty set  $\Theta$ . It is then possible to apply data correction for detected violating constraints instead of just skipping these constraints. Basically, this method consists of three steps :

- Select a parameter vector  $\underline{\theta}_{\text{ref}}$ .
- Construct the constraints  $\{\mathbf{H}_1, \mathbf{H}_2\}$  of Eq. 3.14 for all  $k$ .
- A constraint is violating, if this constraint makes  $\underline{\theta}_{\text{ref}}$  invalid. All violating constraints which have been detected will be modified by data correction.

First an extended parameter and constraint set is defined by :

$$\begin{aligned} \left[ \begin{array}{cc} y(k) - \delta_e(k) & -\phi(k, \underline{\theta}_{\text{ref}}) - \Delta\phi(k, \underline{\theta}_{\text{ref}}) \end{array} \right] \left[ \begin{array}{c} 1 \\ \underline{\theta}_{\text{ref}} \end{array} \right] &\leq 0 \\ \left[ \begin{array}{cc} -y(k) - \delta_e(k) & \phi(k, \underline{\theta}_{\text{ref}}) - \Delta\phi(k, \underline{\theta}_{\text{ref}}) \end{array} \right] \left[ \begin{array}{c} 1 \\ \underline{\theta}_{\text{ref}} \end{array} \right] &\leq 0 \end{aligned} \quad \forall k \quad (3.41)$$

where  $\phi(k, \underline{\theta}_{\text{ref}})$  and  $\Delta\phi(k, \underline{\theta}_{\text{ref}})$  are defined according to Eq. 3.14 for  $\underline{\theta} = \underline{\theta}_{\text{ref}}$ . The distance between the left hand side of Eq. 3.41 and 0 is a measure for the allowed parameter variations. In this context, a constraint violates the reference model  $\underline{\theta}_{\text{ref}}$  if the inequalities of Eq. 3.41 become larger than zero. Depending on the selected model structure, e.g. the general prediction error models or special cases like the output-error model, the error-in-variables model etc., it can be concluded from Eq. 3.14 that several measured (input and/or output) data samples can be modified to correct a violating constraint, i.e. to enforce the validation of Eq. 3.41 with modified data. Denote such a measured data sample by  $x$ , i.e. some  $u(k)$  or  $y(k)$  being an entry of a violating constraint in Eq. 3.41, then this violating constraint can be rewritten as :

$$\alpha_i^x x - \beta_i^x > 0 \quad x \in \{u(k), y(k)\} \quad (3.42)$$

where both  $\alpha_i^x$  and  $\beta_i^x$  contain the remaining data samples and the parameters  $\underline{\theta}_{\text{ref}}$ . In fact, the function of known and unknown variables has been interchanged for this data correction compared to the identification procedure, i.e. all parameters which have to be identified are now assumed to be known ( $\underline{\theta}_{\text{ref}}$ ) and a single data sample will become a

free variable where the measured value is considered as a data outlier. Then  $x$  can be each data sample involved in a violating constraint. Each particular  $x$ , however, appears in more constraints and consequently the index  $i$  ranges over all constraints in which  $x$  appears. The corrected sample  $x_{\text{cor}}$  minimizes the maximum value of the constraint expression  $(\alpha_i^x x - \beta_i^x)$ . So, compute by linear programming :

$$x_{\text{cor}} = \min_x \left\{ \max_i (\alpha_i^x x - \beta_i^x) \right\} \quad (3.43)$$

where  $x_{\text{cor}}$  is the argument  $x$  of the minimization of Eq. 3.43. If this minimum is less than zero, the obtained corrected value  $x_{\text{cor}}$  is a candidate for correction. Next all possible choices for  $x$  may be processed in this way (i.e. all possible corrupted data samples involved in the violating constraint). Finally, that particular corrected  $x_{\text{cor}}$  which leaves the parameter space least affected should be selected. Each possible corrected  $x_{\text{cor}}$ , however, appears in a different set of constraints. Therefore, the union of all sets is taken and consequently that corrected  $x_{\text{cor}}$ , which leads to the least maximum of Eq. 3.41 in this full set, is selected.

### Example

A simple example will be considered to illustrate this data correction procedure in more detail. Suppose a first order output-error model with bounded noise has been selected :

$$y(k) = \frac{b_1 z^{-1}}{1 + f_1 z^{-1}} u(k) + \xi(k) \quad , \quad e(k, \underline{\theta}) = y(k) - \hat{y}(k | \underline{\theta}) \leq \delta_e \quad (3.44)$$

Constructing the extended constraint set according to Eq. 3.41 gives :

time	regression vectors	index
$k-1$	$y(k-1) - \delta_e \quad -u(k-2) \quad y(k-2) - \text{sgn}(f_1)\delta_e$	1
	$-y(k-1) - \delta_e \quad u(k-2) \quad -y(k-2) - \text{sgn}(f_1)\delta_e$	2
$k$	$y(k) - \delta_e \quad -u(k-1) \quad y(k-1) - \text{sgn}(f_1)\delta_e$	3
	$-y(k) - \delta_e \quad u(k-1) \quad -y(k-1) - \text{sgn}(f_1)\delta_e$	4
$k+1$	$y(k+1) - \delta_e \quad -u(k) \quad y(k) - \text{sgn}(f_1)\delta_e$	5
	$-y(k+1) - \delta_e \quad u(k) \quad -y(k) - \text{sgn}(f_1)\delta_e$	6

(3.45)

indicates clearly that whenever the data is corrupted with outliers, several constraints

will be affected modifying thereby the parameter uncertainty set  $\Theta$ . Suppose now that a reference estimate  $\underline{\theta}_{\text{ref}}$  is available (e.g. obtained by classical identification techniques) and that the third constraint violates this  $\underline{\theta}_{\text{ref}}$ . Then either  $y(k)$  or  $y(k-1)$  violates the bounded error assumption. Therefore consider the following :

$$\begin{array}{ccc}
 y(k) & & y(k-1) \\
 \Downarrow \text{violates} & & \Downarrow \text{violates} \\
 \alpha_3^{y(k)} y(k) - \beta_3^{y(k)} > 0 & & \alpha_3^{y(k-1)} y(k-1) - \beta_3^{y(k-1)} > 0 \\
 \Downarrow \text{optimize} & & \Downarrow \text{optimize} \\
 \min_{y(k)} \max_j \left( \alpha_j^{y(k)} y(k) - \beta_j^{y(k)} \right)_{j=3 \dots 6} & & \min_{y(k-1)} \max_j \left( \alpha_j^{y(k-1)} y(k-1) - \beta_j^{y(k-1)} \right)_{j=1 \dots 4} \\
 \Downarrow \text{correction} & & \Downarrow \text{correction} \\
 y_{\text{cor}}(k) & & y_{\text{cor}}(k-1)
 \end{array}$$

where for example  $\alpha_j^{y(k)}$  and  $\beta_j^{y(k)}$  for  $j = 3 \dots 6$  are defined by :

$$\begin{aligned}
 \alpha_3^{y(k)} &= \theta_{\text{new},1} & \beta_3^{y(k)} &= \begin{bmatrix} \delta_e & u(k-1) & -y(k-) + \text{sgn}(f_1) \delta_e \end{bmatrix} \underline{\theta}_{\text{new}} \\
 \alpha_4^{y(k)} &= \theta_{\text{new},1} & \beta_4^{y(k)} &= \begin{bmatrix} -\delta_e & u(k-1) & -y(k-) - \text{sgn}(f_1) \delta_e \end{bmatrix} \underline{\theta}_{\text{new}} \\
 \alpha_5^{y(k)} &= \theta_{\text{new},3} & \beta_5^{y(k)} &= \begin{bmatrix} -y(k+1) + \delta_e & u(k) & \text{sgn}(f_1) \delta_e \end{bmatrix} \underline{\theta}_{\text{new}} \\
 \alpha_6^{y(k)} &= \theta_{\text{new},3} & \beta_6^{y(k)} &= \begin{bmatrix} -y(k+1) - \delta_e & u(k) & -\text{sgn}(f_1) \delta_e \end{bmatrix} \underline{\theta}_{\text{new}}
 \end{aligned}$$

After correction of the possible data outliers,  $y(k)$  and  $y(k-1)$ , resulting in  $y_{\text{cor}}(k)$  and  $y_{\text{cor}}(k-1)$  respectively, the "optimal" correction can be selected by computing the maximum of Eq. 3.41 over all constraints which have been considered during the data correction, i.e.  $j = 1 \dots 6$ , for each of the corrected data samples. The minimum of these maximum values defines then the optimal data correction which leaves the parameter uncertainty set  $\Theta$  most unaffected.

An advantage of this procedure is that violating constraints are eliminated by real data correction. A disadvantage however is that it highly depends on  $\underline{\theta}_{\text{ref}}$ . Nevertheless, straightforward analysis using simulation examples shows that for decreasing error bounds and correct  $\underline{\theta}_{\text{ref}}$ , i.e. the true parameter vector  $\underline{\theta}_t$  in the simulation examples, single data outliers converge to the true data sample. Once all outliers have been corrected, standard set estimation methods can again be used to determine the parameter uncertainty set  $\Theta$ .

Theoretically it is possible that correction of only one data outlier does not yield a solution, i.e. correction of a violating constraint with respect to  $\underline{\theta}_{\text{ref}}$  for a specified error

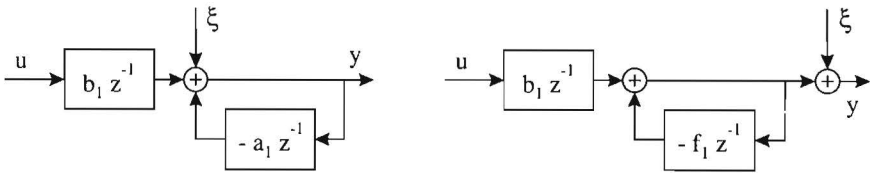


bound. In this case more outliers close in time have to be considered. The proposed method of data correction, however, becomes then unsuitable because of an exploding amount of possible data combinations which qualify for correction when the number of outliers increases.

Nevertheless, all constraints containing data samples of the violating constraint have to be considered whenever a violating constraint is detected with respect to  $\Theta_{\text{ref}}$  for a specified error bound, either in the data correction or when simply skipping constraints. This is necessary, since a data outlier appearing in a non-violating constraint may modify the parameter uncertainty set  $\Theta$  significantly resulting in an incorrect estimate of the central parameter values and the corresponding error bounds.

### 3.4 Comparison of Stochastic and Deterministic Uncertainty

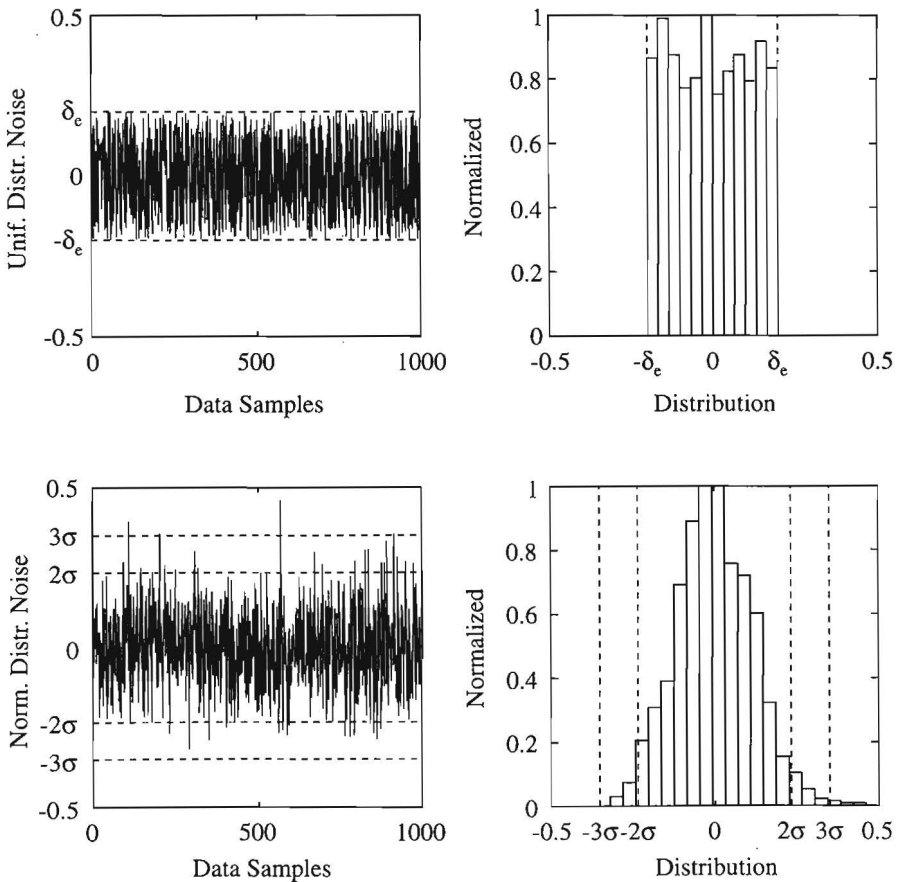
Data outliers in bounded error identification can influence the estimate significantly in contradiction to classical identification where a single outlier will hardly change the estimate due to the averaging effect of the prediction error minimization. When using practical data in bounded error identification, the effect of violating error bounds can never be excluded completely since the error bound should be as tight as possible to obtain the smallest possible parameter uncertainty set. From a theoretical point of view a guaranteed upper bound is required while from a practical point of view the error bound should not be determined by outliers. Even when proper precautions are taken, not all outliers may be detected. Consequently, the final estimate may be influenced.



**Fig. 3.8** : Simulation examples ; a) Equation-error model, b) Output-error model.

To ensure that the set estimation approach with an upper bound characterization of the error describes the process accurately resulting in parameter sets that are not too conservative, the error should preferably have an approximately uniform distribution so that errors at the edge of the distribution have a fair chance to occur, as these give most information. In classical prediction error identification, however, the noise is often

assumed to be normal distributed. For comparison of stochastic and deterministic parameter uncertainty, simple simulation examples will be used to illustrate the main consequences under various conditions. These conditions range from an equation-error model with uniformly distributed noise (optimal conditions for bounded error identification) to an output-error model with normally distributed noise for increasing number of data samples (optimal conditions for prediction error identification). The two different types of models, equation-error and output-error model respectively with nominal parameter values :  $a_1 = f_1 = -0.7$ ,  $b_1 = 1.5$ , are depicted in Fig. 3.8.



**Fig. 3.9 :** Uniformly and normally distributed white noise sequences for  $N = 1000$  data samples together with their distribution. The dashed lines indicate the corresponding noise bounds,  $\delta_e$ ,  $2\sigma$  and  $3\sigma$ , which have been selected for bounded error identification.

A uniformly distributed white noise sequence has been used as input signal,  $u(k) \in [-1,1]$ . In all simulation examples, uniformly as well as normally distributed white noise  $\xi(k)$  has been applied. The signal-to-noise ratio in  $y(k)$  is approximately 20 dB. To achieve the asymptotic properties of the prediction error identification approach, it is required that the number of data samples  $N$  tends to infinity. However, in practice this cannot be realized. Therefore, to observe the differences in parametric uncertainty for a realistic length of the data set and an almost infinite number of data samples, the simulation examples have been performed for  $N = 1000$  and  $N = 10000$  data samples. The uniformly and normally distributed white noise sequences for  $N = 1000$  data samples together with their distribution have been depicted in Fig. 3.9.

The parametric uncertainties have been computed for various simulation conditions :

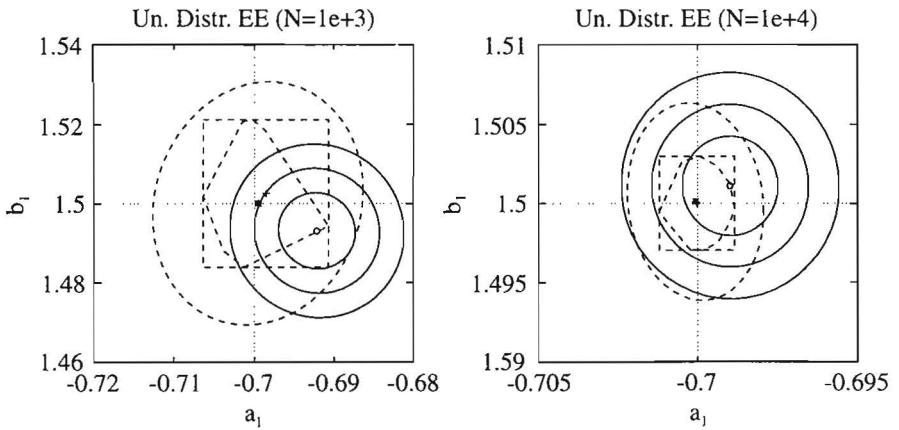
- Prediction error and bounded error identification.
- Equation-error and output-error model structure.
- Uniformly and normally distributed noise.
- 1000 and 10000 data samples.

Note that from a theoretical point of view not all combinations of these simulation conditions will provide correct parametric uncertainty bounds. In particular, prediction error estimation has been defined in a setting where the noise distribution must be continuously differentiable, which is not the case for uniformly distributed noise. Therefore, the corresponding stochastic parametric uncertainties should be treated with care. On the other hand, when applying bounded error identification to systems with normally distributed noise, the  $l_\infty$ -norm bound of the noise becomes infinity large when the number of data samples tends to infinity. To obtain parameter uncertainties that are not too conservative, the upper bound  $\delta_e$  of the noise should be approximated in this case by either a  $2\sigma$  or  $3\sigma$  bound. The corresponding parameter bounds, however, must be treated with care, because the final parameter uncertainty set might be smaller than the true set due to undetected data outliers. Nevertheless, comparing the results of these extreme simulation examples will help to evaluate the reliability of the parameter bounds obtained by different identification methods under various conditions. This exercise is motivated by the fact that when applying these methods in practice, where in general the theoretical assumptions cannot always be satisfied, it should be known how violating theoretical assumptions affect the computed parameter uncertainty set  $\Theta$ .

The parametric uncertainty obtained by prediction error and bounded error identification for uniform as well as normal distributed noise with 1000 and 10000 data samples have been depicted in Fig. 3.10, 3.11 3.12 and 3.13 respectively. Classical prediction error identification has been applied to all simulation examples. For these estimates (indicated by 'o'), confidence ellipsoids (solid lines) have been computed using the modified Cramér-Rao bound (see App. C) which correspond to  $1\sigma$ ,  $2\sigma$  and  $3\sigma$  uncertainty

intervals, or, in other words, 68%, 95% and 99.7% confidence respectively. In the simulation examples with uniformly distributed noise, the deterministic parameter uncertainty has been derived according to several approaches discussed in Section 3.2, i.e. exact (dashed polytope  $\Theta$ ), ellipsoid (dashed ellipsoid  $\Theta_e$  with central estimate '\*') and ellipsoid-aligned orthotopic (dashed box  $\Theta_{eo}$  with central estimate '+') parameter bounding. The upper bound  $\delta_e$  of the normal distributed noise, required for bounded error identification, has been approximated by  $2\sigma$  and  $3\sigma$  bounds, as depicted in Fig. 3.9. The parameter estimates obtained by prediction error identification have been used to detect and eliminate violating constraints in the bounded error context. The various outer bound descriptions are indicated by dashed and dash-dotted figures for the  $\delta_e = 2\sigma$  and the  $\delta_e = 3\sigma$  bounds respectively.

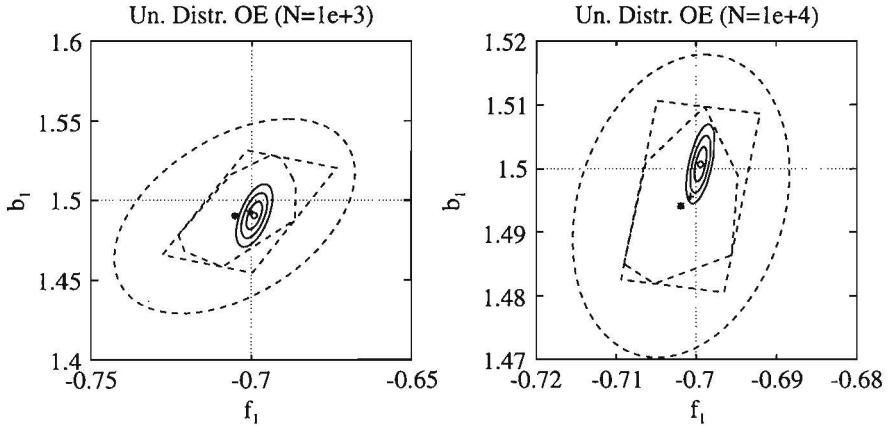
Equation-error models with uniformly distributed noise fit perfectly in the bounded error context. Fig. 3.10 shows the accuracy of the corresponding set estimates which clearly outperform the prediction error results. For an increasing number of data samples the deterministic uncertainty sets decrease rapidly in contrast with the stochastic uncertainty ellipsoids which decrease approximately with  $\sqrt{N}$ .



**Fig. 3.10** : Equation-error model with uniformly distributed noise for  $N = 1000$  and  $N = 10000$  data samples. Solid ellipsoids with centre 'o' indicate stochastic parametric uncertainty with 68%, 95% and 99.7% confidence. The dashed polytope ( $\Theta$ ), dashed ellipsoid ( $\Theta_e$  with central estimate '\*') and dashed ellipsoid-aligned orthotopic ( $\Theta_{eo}$  with central estimate '+') indicate the deterministic outer bound descriptions.

For output-error models the parametric uncertainty set  $\Theta$  can only be approximated

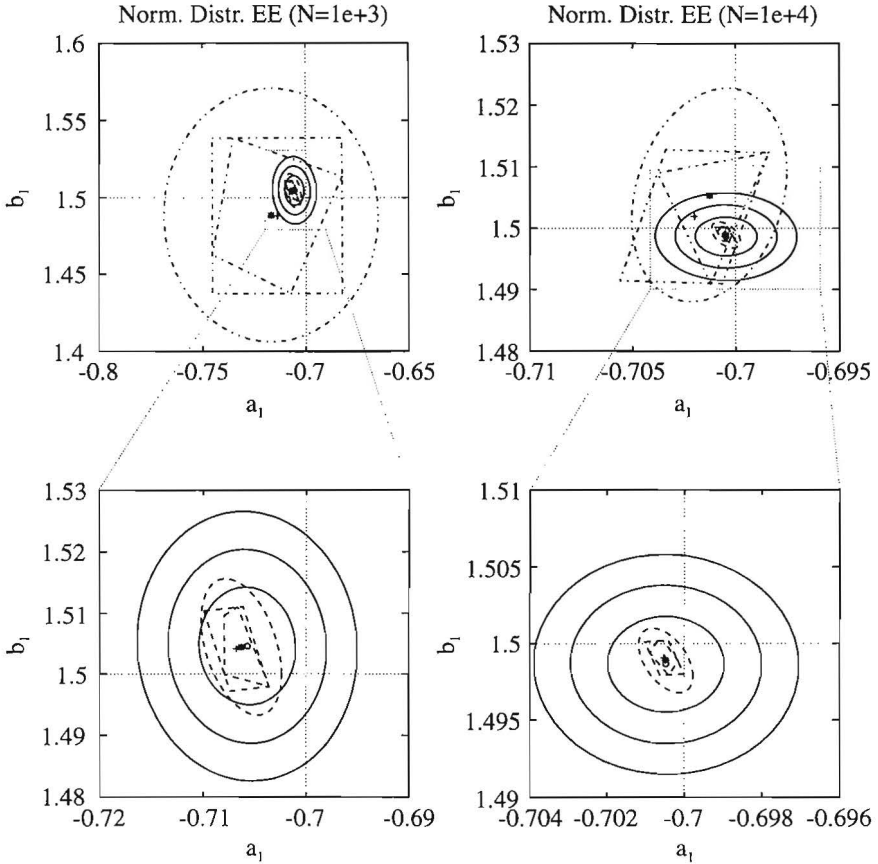
thereby introducing conservatism (compare the deterministic uncertainty sets of Fig. 3.10 and Fig. 3.11). The conservatism introduced by this approximation is quite significant. In this case the prediction error approach with confidence ellipsoids seems to be more accurate.



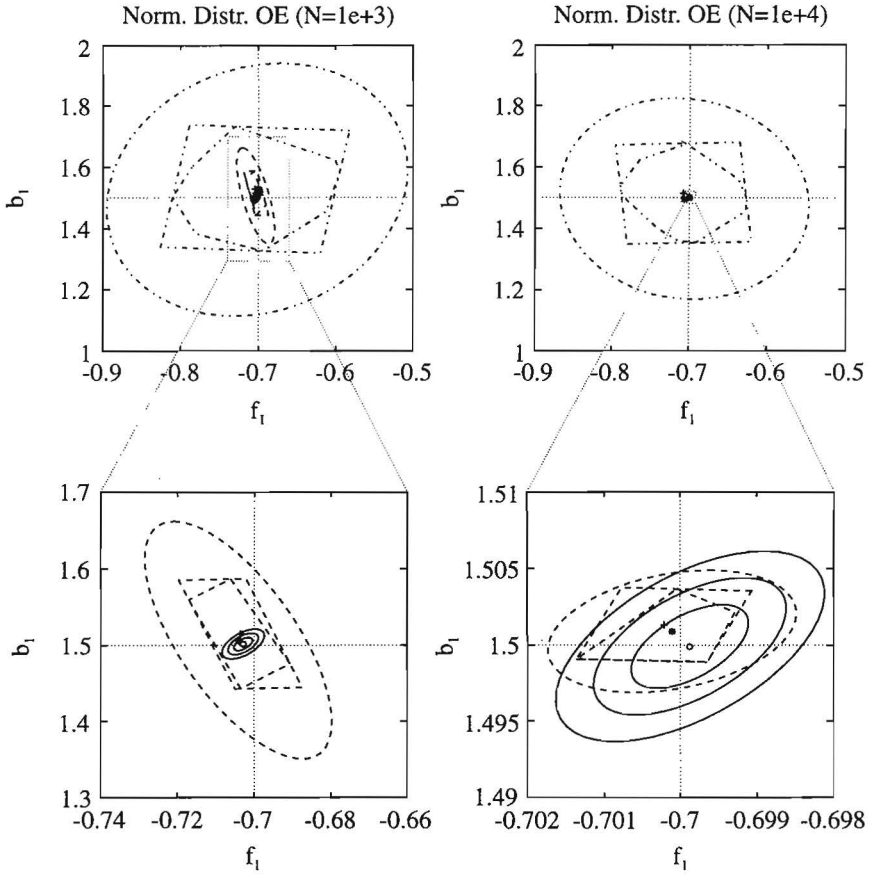
**Fig. 3.11** : Output-error model with uniformly distributed noise for  $N = 1000$  and  $N = 10000$  data samples. Solid ellipsoids with centre 'o' indicate stochastic parametric uncertainty with 68%, 95% and 99.7% confidence. The dashed polytope ( $\Theta$ ), dashed ellipsoid ( $\Theta_e$  with central estimate '\*') and dashed ellipsoid-aligned orthotopic ( $\Theta_{eo}$  with central estimate '+') indicate the deterministic outer bound descriptions.

In identification problems with normally distributed noise, the selection of the noise bound is a difficult problem. To avoid unnecessary large parameter sets, this noise bound has to be approximated as accurate as possible, e.g. by selecting  $\delta_e = 2\sigma$  or  $\delta_e = 3\sigma$  noise bounds. This approximation, however, will also result in violating constraints which have to be eliminated to avoid an empty parameter set. Because the prediction error estimates have been used to eliminate these violating constraints, it cannot be guaranteed that the true parameter vector  $\theta_t$  will be an element of the parameter uncertainty set  $\Theta$ . This is clearly visible in the equation-error example with normally distributed noise which is shown in Fig. 3.12. Not all data outliers,  $\xi(k) > 2\sigma$ , with respect to the  $\delta_e = 2\sigma$  bound have been detected in the form of violating constraints, which decreases the parameter uncertainty set  $\Theta$  more than actually is allowed because the true parameter vector  $\theta_t$  is not included anymore. Because  $\delta_e = 3\sigma$  is quite conservative as upper bound approximation of the noise, the stochastic

confidence ellipsoids and the parameter sets obtained by  $\delta_e = 2\sigma$  have been depicted in a separate picture. For normally distributed noise in the equation-error approach, the parameter uncertainty sets obtained by the  $\delta_e = 3\sigma$  noise bound are much larger than the confidence ellipsoids obtained by prediction error estimation, where on the other hand the parameter sets obtained by a  $\delta_e = 2\sigma$  noise bound are significantly smaller.



**Fig. 3.12** : Equation-error model with normally distributed noise for  $N = 1000$  and  $N = 10000$  data samples. Solid ellipsoids with centre 'o' indicate stochastic parametric uncertainty with 68%, 95% and 99.7% confidence. The dashed polytope ( $\Theta$ ), dashed ellipsoid ( $\Theta_e$  with central estimate '\*'') and dashed ellipsoid-aligned orthotope ( $\Theta_{eo}$  with central estimate '+') indicate the deterministic outer bound descriptions for  $\delta_e = 2\sigma$ . Similar, the dash-dotted polytope, ellipsoid and ellipsoid-aligned orthotope describe the deterministic parameter bounds for  $\delta_e = 3\sigma$ .



**Fig. 3.13** : Output-error model with normally distributed noise for  $N = 1000$  and  $N = 10000$  data samples. Solid ellipsoids with centre 'o' indicate stochastic parametric uncertainty with 68%, 95% and 99.7% confidence. The dashed polytope ( $\Theta$ ), dashed ellipsoid ( $\Theta_e$  with central estimate '\*') and dashed ellipsoid-aligned orthotope ( $\Theta_{eo}$  with central estimate '+') indicate the deterministic outer bound descriptions for  $\delta_e = 2\sigma$ . Similar, the dash-dotted polytope, ellipsoid and ellipsoid-aligned orthotope describe the deterministic parameter bounds for  $\delta_e = 3\sigma$ .

Output-error models with normally distributed noise fit perfectly in the prediction error context. Fig 3.13 shows clearly the accuracy of the estimates, especially for  $N = 10000$  data samples for which the conditions to realize the asymptotic properties of prediction error identification are almost satisfied. However, when applying bounded error

identification, conservatism is introduced twice, resulting in very conservative parameter sets. First of all, the noise bound has to be approximated, similar to the equation-error example. Secondly, the linearization of the output-error model in order to obtain linear constraints describing the parameter uncertainty set  $\Theta$  in the parameter space, introduces additional conservatism (see Fig. 3.13 for  $\delta_e = 3\sigma$ , dash-dotted polytope).

These examples point out the problems which can occur when applying bounded error identification to different types of models under various noise conditions. In practice the error will neither have a typical normal nor uniform distribution. This shows clearly that a balance has to be found between the chosen upper bound, the number of violating constraints and the conservatism of the resulting parameter set. Further, these examples show clearly that the ellipsoid-aligned orthotope is the most accurate approximation of the parameter uncertainty set  $\Theta$  compared to the ellipsoidal outer bound description. In all simulation examples the set  $\Theta_e$  is much larger than the set  $\Theta_{eo}$ . The number of violating constraints which can be expected should roughly correspond with the error bound which has been defined, i.e. 5% and 0.3% violating constraints for a  $2\sigma$  and  $3\sigma$  bound respectively. In practice, however, less violating constraints are often detected due to the conservative description of the parameter uncertainty set  $\Theta$  by linearized constraints. A significant difference between the expected and the detected number of violating constraints indicates that the estimated parameter uncertainty bounds might be smaller than actually allowed.

The stochastic confidence intervals are based on the asymptotic properties of prediction error identification when the number of data samples  $N$  tends to infinity. This cannot be realized in practice. The problem raises the question for which amount of data samples these asymptotic properties can approximately be obtained. The simulation examples indicate that a  $2\sigma$  confidence ellipsoid is required to contain the true parameter vector  $\theta_t$  for  $N = 1000$  samples, where for  $N = 10000$  data samples a  $1\sigma$  confidence ellipsoid is sufficient. In these examples, the stochastic parameter bounds obtained by prediction error estimation seem to provide accurate uncertainty sets. Therefore, these results may be used as an indication of the conservatism often introduced in bounded error identification. Whenever the parameter uncertainty bounds obtained by both identification methods, stochastic as well as deterministic, contain approximately the same set of feasible models, the identification results can be considered as reliable. For major differences, however, the violating theoretical assumptions and their consequences for the parameter uncertainty set  $\Theta$  should be analyzed in more detail.

It has been illustrated in this chapter that ellipsoid-aligned orthotopic bounding provides the least conservative approximation of the parameter uncertainty set  $\Theta$ . Further, the prediction error method can be used as initialization for this bounded error approach. The nominal model then defines a reference model to detect violating constraints and the stochastic parameter bounds can be used as comparison for the deterministic bounds.



## ***Robust Control Models***

---

4.1	General Block-diagonal Uncertainty	4.2	Unstructured Uncertainty
		4.3	Structured Uncertainty

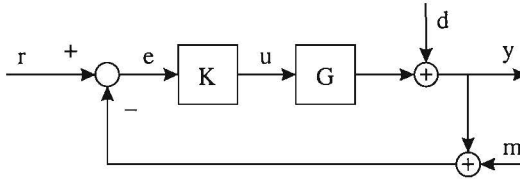
---

No mathematical model can exactly describe a physical process. Some uncertainty is always present, both in the environment of the process, e.g. it is not known in advance which disturbances and noise signals will affect the process, and in the behaviour of the process, e.g. the models of the process obtained either by physical modelling or by using identification methods are not perfect. In this chapter a brief overview of general uncertainty descriptions will be given. The attention, however, is especially focused on the inclusion of parametric uncertainties in robust control design which can be obtained by using the identification methods described in the Chapters 2 and 3 resulting in stochastic and deterministic parametric uncertainty respectively.

Methods for designing robust controllers which take explicitly the discrepancies between the model and the real process into account, range from root locus techniques, controller design with a high stability margin to feedback design in the presence of unknown but bounded uncertainties. An overview can be found for example in [Lun88]. Recent developments in modern feedback control design like  $H_\infty$  and structured singular value ( $\mu$ ) robust control, provide the possibility to include bounded uncertainty in the control

design process to achieve guaranteed robustness with respect to stability and performance of the closed-loop system. Because a diverse range of uncertainties can be handled quantitatively in  $H_\infty$  and  $\mu$  control design, the attention in this chapter will deliberately be restricted to these methods.

In this context, a control system is robust with respect to stability if it remains stable in the presence of model uncertainty. Likewise, it is said to be robust with respect to performance if its behaviour is satisfactory under the influence of uncertain disturbances and model uncertainty. The basic ideas of robust control design and the relation to identification will be explained by considering the control structure depicted in Fig. 4.1.



**Fig. 4.1** : Control system.

For this system , the fundamental input-output relation is given by :

$$y = S d + [I - S] r - [I - S] m \quad (4.1)$$

where  $S = [I + GK]^{-1}$  is the sensitivity function, or alternatively :

$$y = [I - T] d + T r - T m \quad (4.2)$$

where  $T = SGK$  is the complimentary sensitivity function. In general, the objective in a control system is to make some output, say  $y$ , behave in a desired way by manipulating some control input, say  $u$ . A simple objective might be to keep  $y$  small, i.e. a regulator problem, or to keep  $y-r$  small for a reference or command signal  $r$ , i.e. a tracking problem. In addition, there might be the constraint of keeping  $u$  itself small as well, because it might be constrained, e.g. the flow rate through a valve has a maximum value, or it might be too expensive to use a large input. So, the performance of a closed-loop system can be specified in different ways. In the case of exact modelling, the nominal performance corresponds to the performance of the closed-loop system under the assumption that the controller  $K$  stabilizes the system.

These nominal models can be obtained for example by the traditional approach to system identification with a stochastic problem formulation. This has led to fairly well developed theory of system identification and identification algorithms. The issue of obtaining bounds on the corresponding model error has not been a point of research. In modern robust control, however, the starting point for control system analysis and design is a nominal model and (norm) bounds on the model uncertainty. The intensive work that is presently going on in the general area of identification in connection with robust control finds its origin in the understanding that there is a wide gap between the assumptions on which robust control design is built and the tools and results that classical identification theory is able to deliver. Robust control theory requires a priori hard bounds on the model error, whereas classical identification theory delivers at best soft bounds.

Despite the debate which identification method can be used best in practice, the main attention should be focused on the combination of identification and robust control design to achieve the requirements in an industrial environment. Control engineers in industry have to deal with constrained optimization problems to obtain the "best" possible performance with guaranteed stability under hard physical constraints such as actuator or sensor saturation, limitations on feedback due to plant uncertainty, time delays, etc. To emphasize the difficulties which can arise from all these conflicting constraints and objectives, some closed-loop requirements will be summarized for the control system depicted in Fig. 4.1 which illustrates these difficulties even for a simple example.

- Disturbance rejection ( $d \rightarrow y$ ) : keep the sensitivity function  $S$  as small as possible, i.e.  $\sigma_{\max}[(I+GK)^{-1}]$  small.
- Measurement noise rejection ( $m \rightarrow y$ ) : keep the complimentary sensitivity function  $T$  as small as possible, i.e.  $\sigma_{\max}[I-(I+GK)^{-1}]$  small. This conflicts with the disturbance rejection.
- Tracking of a reference signal  $r$  : keep  $\sigma_{\min}[I-(I+GK)^{-1}] \approx 1$ . and  $\sigma_{\max}[I-(I+GK)^{-1}] \approx 1$ . This coincides with the disturbance rejection but conflicts with the measurement noise rejection.
- Minimization of the control energy  $u$  : keep  $\sigma_{\max}(K)$  as small as possible. This will conflict with the disturbance rejection and the reference tracking.

The goal in robust control design is now to realize a robustly performing closed-loop system taking model uncertainty into account and defining the trade-offs in control constraints and objectives by weighting filters in the frequency domain. In this way, the importance of certain constraints and/or objectives in specific frequency ranges can be enhanced or suppressed (see for example [DFB94, FDB92]). A more detailed overview

of the conditions and requirements for robust control design of uncertain systems can be found in several textbooks [Mac89, MZ89, DFT92] and tutorial papers [DGKF89, PD93]. A general framework to represent uncertain systems for robust control has been depicted in Fig. 1.2. The augmented model  $M_a$  represents the model of the process to be controlled and  $\Delta$  is an uncertainty block belonging to some class  $\Delta$  of norm bounded transfer functions. Further,  $M_a$  contains the structure and the weighting functions that reflect how the uncertainty affects the system. With  $M_a$  partitioned in the obvious way, there holds that :

$$\mathcal{F}_u(M_a, \Delta) \triangleq M_{a,22} + M_{a,21} \Delta (I - M_{a,11} \Delta)^{-1} M_{a,12} \quad (4.3)$$

where  $\Delta \in \Delta$  denotes the class of uncertain transfer functions of interest. The uncertainty block  $\Delta$  has been included in an upper loop, resulting in an upper linear fractional transformation (LFT) form. This representation is depicted in Fig. 4.2 where the augmented model  $M_a$  has been partitioned according to the defined inputs and outputs. This LFT representation of uncertain systems for robust control design indicates that identification should provide models in the form of  $\mathcal{F}_u(M_a, \Delta)$  as defined in Eq. 4.3.

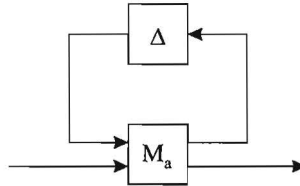


Fig. 4.2 : Upper LFT.

The advantage of using linear fractional transformation representations of the model structure is that series, parallel and feedback connections result in new LFT's with a block-diagonal uncertainty structure [LTBS92]. The complete model set is defined by  $\mathcal{F}_u(M_a, \Delta)$  with  $\Delta \in \Delta$  and the nominal model is defined by  $\mathcal{F}_u(M_a, 0)$ .

The general representation of structured and unstructured uncertainties which can be included in the theory of  $\mu$ -control [Doy82] will be discussed briefly in Section 4.1. When considering unmodelled dynamics, the perturbation  $\Delta$  is unstructured and can be handled by  $H_\infty$ -control design [DGKF89]. In Section 4.2 several approaches to model unstructured uncertainty will be described shortly. The main attention however, will be focused on the description of real parametric uncertainty. This type of uncertainty is obtained from polynomial identification methods, either as stochastic or as deterministic parameter uncertainty (see Chapter 2 and Chapter 3). In Section 4.3, a general LFT

description for parametric uncertainty modelling will be briefly reviewed. The extraction of this parametric model uncertainty into the LFT representation of Fig. 4.2 is based on a state-space representation of the model. Therefore, a state-space realization of the polynomial models will be derived. If the state-space realization of the process part contains no direct-feedthrough term, general expressions can be provided to obtain LFT representations for both the process and the noise model which have been estimated in minimal polynomial form. These expressions can only be derived if the parameters enter the state-space representation in a linear way. If nonlinear combinations of the parameters appear in the state-space representation, the LFT realization becomes highly problem dependent and no general expressions can be derived.

#### 4.1 General Block-diagonal Uncertainty

Since in practice uncertainty includes parameter variations in the model representation as well as unmodelled dynamics, in general both real and complex uncertainties have to be taken into account. This will be formalized as follows :

- A scalar real-valued bounded perturbation  $\Delta_{rn}$  consists of real numbers  $\delta^r$  which are bounded in magnitude by some real number  $\varepsilon_r \in \mathbb{R}^+$  :

$$\Delta_{rn} = \{ \delta^r \mid \delta^r \in [-\varepsilon_r, +\varepsilon_r] \} \quad (4.4)$$

The normalized set is :

$$\mathbf{B} \Delta_{rn} = \{ \delta^r \mid \delta^r \in [-1, +1] \} \quad (4.5)$$

- A scalar complex-valued bounded perturbation  $\Delta_{cn}$  consists of complex numbers  $\delta^c$  which are bounded in magnitude by some real number  $\varepsilon_r \in \mathbb{R}^+$  :

$$\Delta_{cn} = \{ \delta^c \mid \delta^c \in \mathbb{C}, \quad |\delta^c| \leq \varepsilon_r \} \quad (4.6)$$

The normalized set is :

$$\mathbf{B} \Delta_{cn} = \{ \delta^c \mid \delta^c \in \mathbb{C}, \quad |\delta^c| \leq 1 \} \quad (4.7)$$

- A complex-valued norm bounded perturbation matrix  $\Delta_{cm}$  consists of frequency dependent complex matrices  $\Delta^c$  which have a point wise norm-bound  $\varepsilon_c(\omega) \in \mathbb{R}^+$ , i.e. for which  $\sigma_{\max}(\Delta^c) \leq \varepsilon_c(\omega)$ . Precisely,

$$\Delta_{\text{cm}} = \{ \Delta^c \mid \sigma_{\max}(\Delta^c(j\omega)) \leq \varepsilon_c(\omega), \forall \omega \} \quad (4.8)$$

where  $\sigma_{\max}$  denotes the maximum singular value. The normalized set with a uniform upper bound function is :

$$\mathbf{B}\Delta_{\text{cm}} = \{ \Delta^c \mid \sigma_{\max}(\Delta^c(j\omega)) \leq 1, \forall \omega \} \quad (4.9)$$

The complete uncertainty matrix  $\Delta$  consists of real and complex-valued entries, as defined in the following general block structure. Given three non-negative integers  $m_r$ ,  $m_c$  and  $m_f$ , define a vector  $\underline{k}$  of length  $m_k := m_r + m_c + m_f$  as :

$$\underline{k} = (k_1, \dots, k_{m_r}, k_{m_r+1}, \dots, k_{m_r+m_c}, k_{m_r+m_c+1}, \dots, k_{m_k})^T \quad (4.10)$$

where each entry is a non-negative integer. The associated block-diagonal uncertainty  $\Delta_b$  is then defined by the set :

$$\Delta_b = \{ \Delta \mid \Delta = \text{diag}(\delta_i^r I^{k_i}, \delta_j^c I^{k_j}, \Delta_k^c I^{k_k}) \} \quad (4.11)$$

for  $i = 1 \dots k_{m_r}$ ,  $j = k_{m_r+1} \dots k_{m_r+m_c}$  and  $k = k_{m_r+m_c+1} \dots k_{m_r+m_c+m_f}$ , where  $\delta_i^r \in \Delta_{\text{rm}}$  (real parametric uncertainty),  $i = 1 \dots m_r$ ,  $\delta_j^c \in \Delta_{\text{cn}}$  (complex parametric uncertainty),  $i = m_r+1 \dots m_r+m_c$  and  $\Delta_i^c \in \Delta_{\text{cm}}$  (unmodelled dynamics),  $i = m_r+m_c+1 \dots m_k$ . The normalized block-diagonal uncertainty set is denoted as  $\mathbf{B}\Delta_b$  with  $\delta_i^r \in \mathbf{B}\Delta_{\text{rm}}$ ,  $\delta_j^c \in \mathbf{B}\Delta_{\text{cn}}$  and  $\Delta_i^c \in \mathbf{B}\Delta_{\text{cm}}$ . The structural parameters  $m_r$ ,  $m_c$ ,  $m_f$  and  $\underline{k}$  define the uncertainty structure information (real/complex, repeatedness).

Note that for analysis and synthesis purposes, the structured singular value ( $\mu$ ) approach cannot handle yet all types of uncertainties as described in the general block-diagonal uncertainty  $\Delta$ . For complex uncertainties ( $m_r = 0$ ), algorithms exist which can compute an upper bound for  $\mu$ . For simple, low dimensional uncertainty problems (no scalar blocks,  $m_c = 0$ , and only full matrix blocks,  $m_f \leq 3$ ) the upper bound can be computed exactly in contrast to more complex analysis and synthesis problems containing unstructured complex matrix uncertainty combined with structured complex and real parametric uncertainty, where only approximations of the upper bound can be computed. Especially the computation of  $\mu$  with real perturbations is a major problem. It has been shown [BKST89] that these so called real  $\mu$  problems can be discontinuous in the problem data which increases the computational difficulties significantly. It has been shown that the computation of  $\mu$  with real perturbations suffers from exploding computational complexity which has not been solved yet. This probably requires a more practical approach for all problems which arise in engineering applications. A more detailed overview with respect to the computation of  $\mu$  can be found in [PD93].

## 4.2 Unstructured Uncertainty

Classical-feedback design deals with the problem of process uncertainty by prescribing stability margins, by means of specified gain or phase margin. These margins are based on a rather crude model of the uncertainty present in the process. In addition, to maintain stability, only the phase margin at the cross-over frequency is considered.

Using unstructured uncertainty in robust control design a much broader frequency range, not only the cross-over frequency but the whole frequency range which is of interest for control design, can be included in the uncertainty model. Such a description of uncertainty is unstructured in the sense that the magnitude of possible perturbations is bounded, but the sources of the perturbations to specific elements of the process are not specified.

The model error  $\Delta$  can be considered as the unknown transfer function which indicates the difference between the true process  $M_t$  and the model  $M$ . The model error can be included in a LFT using various structures. The most common model error structures are presented briefly in Table 4.1. A more detailed description is given in [vdB93a] where more references can be found.

Of course many other configurations are possible, but the model error structures which have been reviewed briefly are the most common model error descriptions for unstructured uncertainties used in the literature and extensions to other configurations are straightforward.

## 4.3 Structured Uncertainties

The use of unstructured uncertainty descriptions generally leads to compensator designs which are unnecessarily conservative, because they have to perform satisfactorily even for perturbations which will never occur in practice. Therefore, whenever knowledge is available about the structure of the uncertainty, this information should be used in the robust control design to reduce the conservatism instead of lumping uncertainty together. Of course, it is not always possible to assign uncertainty to its source, i.e. measurement noise, unmodelled dynamics etc., but including uncertainty structure will generally decrease the conservatism.

For the polynomial models identified in Chapter 2 (stochastic parameter uncertainty) and Chapter 3 (deterministic parameter uncertainty), the uncertainty is structured in the sense that the coefficients of the polynomial models are not known exactly but can vary within a certain interval. These coefficients will be denoted by  $\theta_i$  assuming that the real value  $\theta_i$  can vary between  $\theta_{\min,i}$  and  $\theta_{\max,i}$  :

**Table 4.1** : Overview of model error structures.

Model error structure	Block diagram
<p>Additive model error <math>\Delta_a</math> :</p> $M_t = M + \Delta_a \quad , \quad \Delta_a \in \Delta_{cm}$ $M_{a,11} = 0, M_{a,12} = I,$ $M_{a,21} = I, M_{a,22} = M.$	
<p>Multiplicative input model error <math>\Delta_{mi}</math> :</p> $M_t = M ( I + \Delta_{mi} ) \quad , \quad \Delta_{mi} \in \Delta_{cm}$ $M_{a,11} = 0, M_{a,12} = I,$ $M_{a,21} = M, M_{a,22} = M.$	
<p>Multiplicative output model error <math>\Delta_{mo}</math> :</p> $M_t = ( I + \Delta_{mo} ) M \quad , \quad \Delta_{mo} \in \Delta_{cm}$ $M_{a,11} = 0, M_{a,12} = M,$ $M_{a,21} = I, M_{a,22} = M.$	
<p>Inverse mult. output model error <math>\Delta_{imo}</math> :</p> $M_t = ( I - \Delta_{imo} )^{-1} M \quad , \quad \Delta_{imo} \in \Delta_{cm}$ $M_{a,11} = I, M_{a,12} = M,$ $M_{a,21} = I, M_{a,22} = M.$	
<p>Left coprime factor model error <math>\Delta_{lcf}</math> :</p> $M_t = ( P + \Delta_p )^{-1} ( Q + \Delta_Q ) ,$ $\Delta_{lcf} = [ \Delta_Q \ -\Delta_p ] \in \Delta_{cm}$ <p>where <math>\Delta_p = P_t - P</math> and <math>\Delta_Q = Q_t - Q</math> define the coprime factors for the model : <math>M = P^{-1}Q</math> and the true process : <math>M_t = P_t^{-1}Q_t</math>.</p> $M_{a,11} = [ 0 \ P^{-1} ]^T, M_{a,12} = [ I \ M ]^T,$ $M_{a,21} = P^{-1}, M_{a,22} = M.$	



$$\theta_i \in [\theta_{\min,i}, \theta_{\max,i}] \quad (4.12)$$

for which the following values are defined :

$$\theta_{c,i} = 1/2 (\theta_{\max,i} + \theta_{\min,i}) \quad , \quad s_{\theta,i} = 1/2 (\theta_{\max,i} - \theta_{\min,i}) \quad (4.13)$$

resulting in :

$$\theta_i = \theta_{c,i} + s_{\theta,i} \delta_i^r \quad , \quad \delta_i^r \in \mathbf{B} \Delta_{rn} \quad (4.14)$$

This polynomial model representation with parametric uncertainty sets will be transformed into a LFT representation where the parametric uncertainty is described in a block-diagonal form.

#### 4.3.1 General LFT Description for Parametric Uncertainty

For the general LFT description with parametric uncertainty, the following state-space representation is considered :

$$\begin{bmatrix} \underline{x}(k+1) \\ \underline{y}(k) \end{bmatrix} = \mathbf{M}_{ss}(\underline{\theta}) \begin{bmatrix} \underline{x}(k) \\ \underline{u}(k) \end{bmatrix} \quad (4.15)$$

which can be rewritten as :

$$\begin{bmatrix} \underline{x}(k+1) \\ \underline{y}(k) \end{bmatrix} = \mathbf{M}_c(\underline{\theta}_c) \begin{bmatrix} \underline{x}(k) \\ \underline{u}(k) \end{bmatrix} + \mathbf{M}_\delta(\underline{s}_\theta) \begin{bmatrix} \underline{x}(k) \\ \underline{u}(k) \end{bmatrix} \quad (4.16)$$

when including the parameter uncertainty sets, where  $\mathbf{M}_c(\underline{\theta}_c)$  contains the nominal centre values of the parameters and  $\mathbf{M}_\delta(\underline{s}_\theta)$  the parametric uncertainty.

To obtain a linear fractional transformation by extracting  $\Delta$ , a new input vector  $\underline{u}_2(k)$  and a new output vector  $\underline{y}_2(k)$  will be defined which depend on the structure of  $\Delta$  (Eq. 4.11) and how the individual entries affect the nominal model. The output  $\underline{y}_2(k)$  is then fed back to the input  $\underline{u}_2(k)$  through a diagonal uncertainty block  $\Delta$  which contains the real parametric uncertainty  $(\delta_1^r \dots \delta_{nt}^r)$ . Furthermore, connection matrices  $\mathbf{B}_2(\underline{s}_\theta)$ ,  $\mathbf{C}_2(\underline{s}_\theta)$ ,  $\mathbf{D}_{12}(\underline{s}_\theta)$ ,  $\mathbf{D}_{21}(\underline{s}_\theta)$  and  $\mathbf{D}_{22}(\underline{s}_\theta)$  must be defined which contain information on how the nominal model  $\mathbf{M}_c(\underline{\theta}_c)$  is affected by the uncertainty  $\underline{\delta}^r$ .



$$\begin{bmatrix} \underline{x}(k+1) \\ \underline{y}(k) \end{bmatrix} = M_c(\underline{\theta}_c) \begin{bmatrix} \underline{x}(k) \\ \underline{u}(k) \end{bmatrix} + \begin{bmatrix} B_2(\underline{s}_\theta) \\ D_{12}(\underline{s}_\theta) \end{bmatrix} \underline{u}_2(k) \quad (4.19)$$

$$\underline{y}_2(k) = \begin{bmatrix} C_2(\underline{s}_\theta) & D_{21}(\underline{s}_\theta) \end{bmatrix} \begin{bmatrix} \underline{x}(k) \\ \underline{u}(k) \end{bmatrix} + D_{22}(\underline{s}_\theta) \underline{u}_2(k)$$

and closing the loop around  $\Delta$  yields for Eq. 4.19 :

$$\begin{bmatrix} \underline{x}(k+1) \\ \underline{y}(k) \end{bmatrix} = M_c(\underline{\theta}_c) \begin{bmatrix} \underline{x}(k) \\ \underline{u}(k) \end{bmatrix} + \begin{bmatrix} B_2(\underline{s}_\theta) \\ D_{12}(\underline{s}_\theta) \end{bmatrix} \left( I - \Delta D_{22}(\underline{s}_\theta) \right)^{-1} \Delta \begin{bmatrix} C_2(\underline{s}_\theta) & D_{21}(\underline{s}_\theta) \end{bmatrix} \begin{bmatrix} \underline{x}(k) \\ \underline{u}(k) \end{bmatrix} \quad (4.20)$$

which must be equivalent to Eq. 4.16. Therefore, finding a linear fractional transformation description requires finding a solution to :

$$M_\delta(\underline{s}_\theta) = \begin{bmatrix} B_2(\underline{s}_\theta) \\ D_{12}(\underline{s}_\theta) \end{bmatrix} \left( I - \Delta D_{22}(\underline{s}_\theta) \right)^{-1} \Delta \begin{bmatrix} C_2(\underline{s}_\theta) & D_{21}(\underline{s}_\theta) \end{bmatrix} \quad (4.21)$$

It has been shown in [LTBS92] that a solution to the problem of transforming a state-space model with parametric uncertainty into a LFT exists, if the state-space matrices can be given as real rational transfer functions in the parameters. A realization algorithm based on the properties of the interconnection of LFT's, i.e. series, parallel and feedback connections of LFT's result in a new LFT, is proposed to solve Eq. 4.21. A LFT is constructed for all individual terms in the uncertain state-space matrices  $\{ A_\delta(\underline{s}_\theta), B_\delta(\underline{s}_\theta), C_\delta(\underline{s}_\theta), D_\delta(\underline{s}_\theta) \}$ , which are then combined by series, parallel or feedback connections resulting in a complete description of all uncertain entries specified in  $M_{ss}(\underline{\theta})$  in the form of LFT's. Combining then all LFT's of the state-space matrices in  $M_\delta(\underline{s}_\theta)$  and rearranging all uncertainties into the real-valued repeated scalar block structure results in a high dimensional uncertainty block  $\Delta$  where possibly the individual repeated blocks can be replaced by smaller blocks to reduce the dimension of  $\Delta$ . Minimality with respect to the dimension of  $\Delta$  of the obtained LFT cannot be guaranteed for the general case in which the parameters enter the state-space matrices in any polynomial way.

However, if the parameters enter the state-space matrices in a linear way, the term  $D_{22}(\underline{s}_\theta)$  becomes zero. In this case it has been shown in [Ter90] that the LFT realization problem of Eq. 4.21 reduces to :

$$M_{\delta}(\underline{s}_{\theta}) = \begin{bmatrix} B_2(\underline{s}_{\theta}) \\ D_{12}(\underline{s}_{\theta}) \end{bmatrix} \Delta \begin{bmatrix} C_2(\underline{s}_{\theta}) & D_{21}(\underline{s}_{\theta}) \end{bmatrix} = \sum_{q=1}^{nt} \delta_q^r M_{\delta}(s_{\theta_q}) \quad (4.22)$$

where the smallest  $\Delta$  for which a solution exists has dimension :

$$\sum_{q=1}^{nt} \text{rank} \left( M_{\delta}(s_{\theta_q}) \right) \quad (4.23)$$

The LFT realization of the polynomial models obtained by system identification, can now be divided into two steps :

- Conversion of the polynomial model into a state-space representation.
- Obtaining a linear fractional transformation.

#### 4.3.2 State-space Realization of Polynomial Models

Several approaches can be selected to convert proper minimal polynomial models into state-space representations [Kai80]. To illustrate this conversion the following SISO process model will be considered :

$$G(z^{-1}, \underline{\theta}) = \frac{b_1 + b_2 z^{-1} + b_3 z^{-2}}{1 + f_1 z^{-1} + f_2 z^{-2}} \quad (4.24)$$

Two typical state-space representations are the controller and the observer canonical form which have the following form for the model  $G(z^{-1}, \underline{\theta})$  defined in Eq. 4.24 :

$$M_{ss}^{co}(\underline{\theta}) = \begin{bmatrix} A^{co}(\underline{\theta}) & B^{co}(\underline{\theta}) \\ C^{co}(\underline{\theta}) & D^{co}(\underline{\theta}) \end{bmatrix} = \left[ \begin{array}{cc|c} -f_1 & -f_2 & 1 \\ 1 & 0 & 0 \\ \hline b_2 - b_1 f_1 & b_3 - b_1 f_2 & b_1 \end{array} \right] \quad (4.25)$$

$$M_{ss}^{ob}(\underline{\theta}) = \begin{bmatrix} A^{ob}(\underline{\theta}) & B^{ob}(\underline{\theta}) \\ C^{ob}(\underline{\theta}) & D^{ob}(\underline{\theta}) \end{bmatrix} = \left[ \begin{array}{cc|c} -f_1 & 1 & b_2 - b_1 f_1 \\ -f_2 & 0 & b_3 - b_1 f_2 \\ \hline 1 & 0 & b_1 \end{array} \right] \quad (4.26)$$

where  $A^{ob}(\underline{\theta}) = [A^{co}(\underline{\theta})]^T$ ,  $B^{ob}(\underline{\theta}) = [C^{co}(\underline{\theta})]^T$ ,  $C^{ob}(\underline{\theta}) = [B^{co}(\underline{\theta})]^T$ ,  $D^{ob}(\underline{\theta}) = D^{co}(\underline{\theta})$ .

For multivariable systems which have been modelled using a minimal polynomial structure, the smallest possible state-space realization corresponds to  $n \cdot \min(ny, nu)$  for order  $n$  of the minimal polynomial [Bac87]. Therefore, a controller canonical form should be constructed for  $ny \geq nu$  and an observer canonical form otherwise. The multiplicity of the poles due to the minimal polynomial description corresponds then to  $nu$  and  $ny$  respectively. The multivariable state-space realization  $M_{ss}^{co}(\underline{\theta})$  or  $M_{ss}^{ob}(\underline{\theta})$  of minimal polynomial models corresponds in fact to connecting the state-space matrices of SIMO models  $[M_{ss}^{co}(\underline{\theta})]_{*j}$  for all inputs  $j$  or MISO models  $[M_{ss}^{ob}(\underline{\theta})]_{i*}$  for all outputs  $i$ . For each of these SIMO state-space realizations in controller canonical form, there holds for the corresponding state matrices that  $A_{*p}^{co}(\underline{\theta}) = A_{*q}^{co}(\underline{\theta})$  which will be denoted as  $A^{co}(\underline{\theta})$ . Similar, the input matrix can be denoted by  $B^{co}(\underline{\theta})$  because of  $B_{*p}^{co}(\underline{\theta}) = B_{*q}^{co}(\underline{\theta})$ . For each of the MISO state-space realizations in observer canonical form, the state and output matrix will be denoted by  $A^{ob}(\underline{\theta})$  and  $C^{ob}(\underline{\theta})$  respectively, because of  $A_{p*}^{ob}(\underline{\theta}) = A_{q*}^{ob}(\underline{\theta})$  and  $C_{p*}^{ob}(\underline{\theta}) = C_{q*}^{ob}(\underline{\theta})$ . This results in the state-space representations defined in Eq. 4.27 and Eq. 4.28.

The state-space realization of a noise model  $H(z^{-1}, \underline{\theta})$  can be performed identically to the realization of  $G(z^{-1}, \underline{\theta})$ . Note, however, that the resulting state-space realization will have a controller canonical diagonal form for the minimal polynomial structure defined in Eq. 2.17. The diagonal form results from the fact that the noise model can have diagonal entries only and a controller canonical form has been realized because the noise models have an equal number of inputs and outputs ( $ny = n\xi$ ).

$$M_{ss}^{co}(\underline{\theta}) = \begin{bmatrix} A^{co}(\underline{\theta}) & 0 & \dots & 0 & | & B^{co}(\underline{\theta}) & 0 & \dots & 0 \\ 0 & A^{co}(\underline{\theta}) & \ddots & \vdots & | & 0 & B^{co}(\underline{\theta}) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 & | & \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & A^{co}(\underline{\theta}) & | & 0 & \dots & 0 & B^{co}(\underline{\theta}) \\ \hline C_{11}^{co}(\underline{\theta}) & C_{12}^{co}(\underline{\theta}) & \dots & C_{1nu}^{co}(\underline{\theta}) & | & D_{11}^{co}(\underline{\theta}) & D_{12}^{co}(\underline{\theta}) & \dots & D_{1nu}^{co}(\underline{\theta}) \\ C_{21}^{co}(\underline{\theta}) & C_{22}^{co}(\underline{\theta}) & & \vdots & | & D_{21}^{co}(\underline{\theta}) & D_{22}^{co}(\underline{\theta}) & & \vdots \\ \vdots & & \ddots & \vdots & | & \vdots & & \ddots & \vdots \\ C_{ny1}^{co}(\underline{\theta}) & \dots & \dots & C_{nynu}^{co}(\underline{\theta}) & | & D_{ny1}^{co}(\underline{\theta}) & \dots & \dots & D_{nynu}^{co}(\underline{\theta}) \end{bmatrix} \quad (4.27)$$

$$M_{ss}^{ob}(\underline{\theta}) = \begin{bmatrix} A^{ob}(\underline{\theta}) & 0 & \dots & 0 & | & B_{11}^{ob}(\underline{\theta}) & B_{12}^{ob}(\underline{\theta}) & \dots & B_{1nu}^{ob}(\underline{\theta}) \\ 0 & A^{ob}(\underline{\theta}) & \ddots & \vdots & | & B_{21}^{ob}(\underline{\theta}) & B_{22}^{ob}(\underline{\theta}) & & \vdots \\ \vdots & \ddots & \ddots & 0 & | & \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & A^{ob}(\underline{\theta}) & | & B_{ny1}^{ob}(\underline{\theta}) & \dots & \dots & B_{nynu}^{ob}(\underline{\theta}) \\ \hline C^{ob}(\underline{\theta}) & 0 & \dots & 0 & | & D_{11}^{ob}(\underline{\theta}) & D_{12}^{ob}(\underline{\theta}) & \dots & D_{1nu}^{ob}(\underline{\theta}) \\ 0 & C^{ob}(\underline{\theta}) & & \vdots & | & D_{21}^{ob}(\underline{\theta}) & D_{22}^{ob}(\underline{\theta}) & & \vdots \\ \vdots & & \ddots & 0 & | & \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & C^{ob}(\underline{\theta}) & | & D_{ny1}^{ob}(\underline{\theta}) & \dots & \dots & D_{nynu}^{ob}(\underline{\theta}) \end{bmatrix} \quad (4.28)$$

Considering now the state-space realization of minimal polynomial models, Eq. 4.25 and Eq. 4.26 show that the parameters enter the state-space matrices linearly except the term  $b_1$  representing the direct feedthrough term from input to output. For process models  $G(z^{-1}, \underline{\theta})$  this direct feedthrough term can be eliminated by demanding that the models should contain at least one sample delay. This is not really a restriction in practice since physical systems show a strictly proper behaviour, i.e.  $b_1 = 0$  in discrete process models. Therefore, no direct feedthrough term will appear in the model and the identified parameters enter the state-space matrices in a linear way. For noise models  $H(z^{-1}, \underline{\theta})$  the LFT realization is always linear since the first coefficient of the  $C_{ii}(z^{-1})$  polynomial has been defined to be one (Eq. 2.17) and consequently the parameters enter the state-space matrices in a linear way.

### 4.3.3 LFT Realization of Polynomial Process and Noise Models

In this section general expressions will be derived for the LFT realization of the polynomial process model  $G(z^{-1}, \underline{\theta})$  and noise model  $H(z^{-1}, \underline{\theta})$ . A simple example is used to illustrate the general LFT realization.

Suppose a 2-input 2-output multivariable process has been modelled as a 2<sup>nd</sup> order minimal polynomial Box-Jenkins model with 1 sample delay. The state-space representation  $M_{ss}^{co}(\underline{\theta})$  of the process model  $G(z^{-1}, \underline{\theta})$  is then defined by a nominal part :

$$M_c^{co}(\underline{\theta}) = \left[ \begin{array}{cccc|cc} -f_1 & -f_2 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -f_1 & -f_2 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hline & & & & + & - \\ b_{11,2} & b_{11,3} & b_{12,2} & b_{12,3} & 0 & 0 \\ b_{21,2} & b_{21,3} & b_{22,2} & b_{22,3} & 0 & 0 \end{array} \right] \quad (4.29)$$

and an uncertain part :

$$M_\delta^{co}(\underline{s}_\theta) = \left[ \begin{array}{cccc|cc} -s_{f_1} \delta_{f_1}^r & -s_{f_2} \delta_{f_2}^r & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -s_{f_1} \delta_{f_1}^r & -s_{f_2} \delta_{f_2}^r & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hline & & & & + & - \\ s_{b_{11,2}} \delta_{b_{11,2}}^r & s_{b_{11,3}} \delta_{b_{11,3}}^r & s_{b_{12,2}} \delta_{b_{12,2}}^r & s_{b_{12,3}} \delta_{b_{12,3}}^r & 0 & 0 \\ s_{b_{21,2}} \delta_{b_{21,2}}^r & s_{b_{21,3}} \delta_{b_{21,3}}^r & s_{b_{22,2}} \delta_{b_{22,2}}^r & s_{b_{22,3}} \delta_{b_{22,3}}^r & 0 & 0 \end{array} \right] \quad (4.30)$$

To solve the LFT realization defined in Eq. 4.22, the minimal dimension of the uncertainty block  $\Delta$  can be determined according to Eq. 4.23 :

$$\begin{aligned} \text{rank} \left( M_\delta^{co}(s_{b_{ij,q}}) \right) &= 1 \quad \forall i,j,q \\ \text{rank} \left( M_\delta^{co}(s_{f_q}) \right) &= 2 \quad \forall q \end{aligned} \quad (4.31)$$

resulting in :

$$\Delta_G^{co} = \text{diag} \left\{ \delta_{b_{11,2}}, \delta_{b_{11,3}}, \delta_{b_{12,2}}, \delta_{b_{12,3}}, \delta_{b_{21,2}}, \delta_{b_{21,3}}, \delta_{b_{22,2}}, \delta_{b_{22,3}}, \delta_{f_1} I^2, \delta_{f_2} I^2 \right\} \quad (4.32)$$

A solution for the connection matrices of the controller canonical LFT realization can be found by solving the equivalence of Eq. 4.22 :

$$\begin{bmatrix} B_{G,2}^{co} \\ D_{G,12}^{co} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ - & - & - & - & - & - & - & - & - & - & - & - & - \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (4.33)$$

$$\begin{bmatrix} C_{G,2}^{co}(\underline{s}_\theta) & D_{G,12}^{co} \end{bmatrix} = \begin{bmatrix} s_{b_{11,2}} & 0 & 0 & 0 & | & 0 & 0 \\ 0 & s_{b_{11,3}} & 0 & 0 & | & 0 & 0 \\ 0 & 0 & s_{b_{12,2}} & 0 & | & 0 & 0 \\ 0 & 0 & 0 & s_{b_{12,3}} & | & 0 & 0 \\ s_{b_{21,2}} & 0 & 0 & 0 & | & 0 & 0 \\ 0 & s_{b_{21,3}} & 0 & 0 & | & 0 & 0 \\ 0 & 0 & s_{b_{22,2}} & 0 & | & 0 & 0 \\ 0 & 0 & 0 & s_{b_{22,3}} & | & 0 & 0 \\ s_{f_1} & 0 & 0 & 0 & | & 0 & 0 \\ 0 & 0 & s_{f_1} & 0 & | & 0 & 0 \\ 0 & s_{f_2} & 0 & 0 & | & 0 & 0 \\ 0 & 0 & 0 & s_{f_2} & | & 0 & 0 \end{bmatrix} \quad (4.34)$$

Note that this solution is only one possible LFT realization for this example because there are more degrees of freedom to construct the connection matrices  $\{B_2(\underline{s}_\theta), C_2(\underline{s}_\theta), D_{12}(\underline{s}_\theta), D_{21}(\underline{s}_\theta)\}$  than equivalence relations which have been defined in Eq. 4.22. For example, interchanging the 1's in Eq. 4.33 with the scaling factors  $\underline{s}_\theta$  in Eq. 4.34 indicates already various alternative realizations.

General expressions for LFT realizations of a polynomial process model  $G(z^{-1}, \underline{\theta})$  and noise model  $H(z^{-1}, \underline{\theta})$  can be derived in a similar way :



- Controller canonical LFT realization of process model  $G(z^{-1}, \underline{\theta})$  :

In general the dimension of the uncertainty block  $\Delta$  for a minimal polynomial process model  $G(z^{-1}, \underline{\theta})$  of order  $n$  :

$$\Delta_G^{\text{co}} = \text{diag} \left\{ \delta_{b_{11,2}}, \dots, \delta_{b_{11,nb}}, \delta_{b_{12,2}}, \dots, \delta_{b_{1nu,nb}}, \dots, \delta_{b_{nynu,nb}}, f_1 I^{nu}, \dots, f_n I^{nu} \right\} \quad (4.35)$$

where  $nb = n+1$ , which can be transformed into a state-space representation with linear parameters is given by :

$$\begin{array}{ccc} n\Delta_G^{\text{co}} = & n.nu.ny & + & n.nu \\ & b - \text{parameters} & & f - \text{parameters} \end{array} \quad (4.36)$$

The connection matrices for the general case can be constructed according to :

$$B_{G,2}^{\text{co}} = \begin{bmatrix} 0^{(n.nu \times n.nu.ny)} & -1^{(1 \times nu)} \otimes I^{nu} \otimes I^{(n \times 1)} \end{bmatrix} \quad (4.37)$$

$$D_{G,12}^{\text{co}} = \begin{bmatrix} I^{ny} \otimes I^{(1 \times n.nu)} & 0^{(ny \times n.nu)} \end{bmatrix} \quad (4.38)$$

$$C_{G,2}^{\text{co}}(\underline{s}_{\theta}) = \begin{bmatrix} \text{diag}(\underline{s}_{b_{1x}}) \\ \vdots \\ \text{diag}(\underline{s}_{b_{nyx}}) \\ s_{f_1} I^{nu} \otimes \begin{bmatrix} 0^{(1 \times 0)} & I^{(1 \times n)} \end{bmatrix} \\ \vdots \\ s_{f_n} I^{nu} \otimes \begin{bmatrix} 0^{(1 \times n-1)} & I^{(1 \times 1)} \end{bmatrix} \end{bmatrix} \quad (4.39)$$

$$D_{G,21}^{\text{co}} = 0^{(n\Delta \times nu)} \quad (4.40)$$

where  $\otimes$  denotes the Kronecker matrix product,  $0^{p \times q}$ ,  $1^{p \times q}$  and  $I^{p \times q}$  define zero, one and identity matrices respectively of dimension  $p \times q$ , and

$$\underline{s}_{b_{ix}} = \begin{bmatrix} s_{b_{i1,2}}, \dots, s_{b_{i1,nb}}, s_{b_{i2,2}}, \dots, s_{b_{inu,nb}} \end{bmatrix}^T \quad (4.41)$$

- Observer canonical LFT realization of process model  $G(z^{-1}, \underline{\theta})$  :

Similar to a controller canonical LFT, an observer canonical LFT realization can be constructed. To show, however, the same similarity between controller and observer connection matrices when realizing a LFT that exists between controller and observer state-space matrices (compare Eq. 4.25 and Eq. 4.26), the order of the b-parameters in the uncertainty block has been changed to :

$$\Delta_G^{\text{ob}} = \text{diag} \left\{ \delta_{b_{11,2}}, \dots, \delta_{b_{11,nb}}, \delta_{b_{21,2}}, \dots, \delta_{b_{ny1,nb}}, \dots, \delta_{b_{nynu,nb}}, f_1 I^{ny}, \dots, f_n I^{ny} \right\} \quad (4.42)$$

with dimension :

$$\begin{array}{ccc} n\Delta_G^{\text{ob}} & = & n.nu.ny \quad + \quad n.ny \\ & & \text{b - parameters} \quad \quad \text{f - parameters} \end{array} \quad (4.43)$$

This change in order of the b-parameters results from the fact that the order in which the inputs and outputs are processed to construct multivariable controller and observer canonical state-space matrices respectively from minimal polynomial models, has been interchanged.

The corresponding connection matrices for observer canonical state-space realizations are given by :

$$\begin{aligned} B_{G,2}^{\text{ob}}(\underline{s}_{\theta}) &= \left[ \text{diag}(\underline{s}_{b_{x1}}), \dots, \text{diag}(\underline{s}_{b_{xnu}}), \right. \\ &\quad \left. s_{f_1} I^{ny} \otimes \begin{bmatrix} 0^{(0 \times 1)} \\ I^{(n \times 1)} \end{bmatrix}, \dots, s_{f_n} I^{ny} \otimes \begin{bmatrix} 0^{(n-1 \times 1)} \\ I^{(1 \times 1)} \end{bmatrix} \right] \end{aligned} \quad (4.44)$$

$$D_{G,12}^{\text{ob}} = 0^{(ny \times n\Delta)} \quad (4.45)$$

$$C_{G,2}^{\text{ob}} = \begin{bmatrix} 0^{(n.nu.ny \times n.ny)} \\ -1^{(ny \times 1)} \otimes I^{ny} \otimes I^{(1 \times n)} \end{bmatrix} \quad (4.46)$$

$$D_{G,21}^{\text{ob}} = \begin{bmatrix} I^{nu} \otimes 1^{(n.ny \times 1)} \\ 0^{(n.ny \times nu)} \end{bmatrix} \quad (4.47)$$

where

$$\underline{s}_{b_{xi}} = \left[ s_{b_{li,2}}, \dots, s_{b_{li,nb}}, s_{b_{2i,2}}, \dots, s_{b_{nyi,nb}} \right]^T \quad (4.48)$$

Comparing now Eq. 4.37 to 4.40 and Eq. 4.44 to 4.47, the similarity objective between controller and observer connection matrices has clearly been realized :  $B_2^{co} = [C_2^{ob}]^T$ ,  $D_{12}^{co} = [D_{21}^{ob}]^T$ ,  $C_2^{co}(\underline{s}_\theta) = [B_2^{ob}(\underline{s}_\theta)]^T$ ,  $D_{21}^{co} = [D_{12}^{ob}]^T$  when thereby interchanging the indices of the b-parameters and the number of inputs and outputs, nu and ny respectively.

- Controller canonical LFT realization of noise model  $H(z^{-1}, \underline{\theta})$  :

This LFT realization of the noise model  $H(z^{-1}, \underline{\theta})$  is almost the same as for the process model  $G(z^{-1}, \underline{\theta})$ . Only less c-parameters are involved due to the diagonalization of the polynomial matrix  $C(z^{-1})$  and the d-parameters of the minimal polynomial appear in more entries of  $M_{ss}^{co}(\underline{\theta})$  but without increasing the rank of  $M_{\delta}^{co}(s_{dq})$ . Therefore, the same multiplicity is maintained compared to  $f_q$  (Eq. 4.29) resulting in :

$$\Delta_H^{co} = \text{diag} \left\{ \delta_{c_{11,l}}, \dots, \delta_{c_{11,n}}, \delta_{c_{22,l}}, \dots, \delta_{c_{ny,ny,n}}, d_1 I^{ny}, \dots, d_n I^{ny} \right\} \quad (4.49)$$

which has dimension :  $n\Delta_H^{co} = 2.n.ny$ . Note that the state-space realization of the noise model  $H(z^{-1}, \underline{\theta})$  is always square, i.e.  $n\xi = ny$ .

The corresponding connection matrices are given by :

$$B_{H,2}^{co} = \begin{bmatrix} 0^{(n.ny \times n.ny)} & -I^{(1 \times ny)} \otimes I^{ny} \otimes I^{(n \times 1)} \end{bmatrix} \quad (4.50)$$

$$D_{H,12}^{co} = \begin{bmatrix} I^{ny} \otimes I^{(1 \times n)} & -I^{(1 \times ny)} \otimes I^{ny} \end{bmatrix} \quad (4.51)$$

$$C_{H,2}^{co}(\underline{s}_\theta) = \begin{bmatrix} \text{diag}(\underline{s}_c) \\ s_{d_1} I^{ny} \otimes \begin{bmatrix} 0^{(1 \times 0)} & I^{(1 \times n)} \end{bmatrix} \\ \vdots \\ s_{d_n} I^{ny} \otimes \begin{bmatrix} 0^{(1 \times n-1)} & I^{(1 \times 1)} \end{bmatrix} \end{bmatrix} \quad (4.52)$$

$$D_{H,21}^{co} = 0^{(n\Delta \times ny)} \quad (4.53)$$

where

$$\underline{s}_c = \left[ s_{c_{11,1}}, \dots, s_{c_{11,n}}, s_{c_{22,1}}, \dots, s_{c_{nny,n}} \right]^T \quad (4.54)$$

It has been shown that LFT realizations of the minimal polynomial process models  $G(z^{-1}, \underline{q})$  and the minimal polynomial noise models  $H(z^{-1}, \underline{q})$  can be easily obtained imposing only a slight restriction on the process models, i.e. no direct feedthrough. Whenever other type of models than Box-Jenkins are selected, e.g. ARMAX, the polynomials  $F(z^{-1})$  and  $D(z^{-1})$  are of course replaced by  $A(z^{-1})$ . Note however, that for models with combined  $A(z^{-1})$  and  $F(z^{-1})$  polynomials the LFT cannot be realized using the proposed method because the parameters enter the state-space matrices as multiplications.

The same linear approach for the LFT construction can be followed if coordinate transformation is involved as for example in ellipsoid-aligned orthotopic bounding (see Section 3.2.4). If no direct feed-through term is included in the process model, the parameter uncertainty in the ellipsoid-aligned coordinate system can still be extracted in a linear way from the state-space realization to construct a LFT representation. As a consequence, the dimension of the uncertainty block  $\Delta$  (Eq. 4.18) will increase because of increasing rank of the uncertainty matrix  $M_\delta(s_\theta)$  (Eq. 4.23).

In general, as long as the parameters enter the state-space matrices in a linear way, the LFT can be constructed fairly easy. However, when the state-space entries contain nonlinear parameter combinations, the multiplicity of the scalar uncertainty blocks for the parameters is not known in advance and matrix inversion is required because  $D_{22} \neq 0$  (Eq. 4.21) which makes the LFT construction very complicated and highly problem dependent. In this case a general solution cannot be given and the LFT realization has to be solved for every problem separately.

This approach of first constructing a state-space representation of the polynomial models and then extracting the parametric uncertainty into  $\Delta$  has been selected because realizing LFT representations from state-space models with parametric uncertainty is very well from literature. Of course, it is possible to extract the parametric uncertainty directly from the polynomial models and to derive afterwards a state-space representation of the augmented model  $M_a$ . The dimension of the corresponding state-space model, however, will be significantly higher than  $n \cdot \min(ny, nu)$  of the controller/observer canonical state-space representation and therefore this direct LFT realization has not been included in this chapter.

For multivariable systems of increasing order the dimension of  $\Delta$  can become rather large due to the rapidly growing number of parameters. However, it will be shown in Chapter 5 and 6 that the dimension of  $\Delta$  can be reduced by fixing the parameters which have the least contribution to the input/output behaviour, to predefined values with no uncertainty.

# ***Parsimonification***

---

5.1	Deterministic Parameter Uncertainty	5.2	Stochastic Parameter Uncertainty
		5.3	Parameter Uncertainty Reduction

---

Many problems can occur when applying set estimation techniques in practice. In principle, it is assumed that the process is contained in the model set and that a guaranteed upper bound of the noise disturbing the process, e.g. the  $l_\infty$ -norm of the prediction error, is known. Both conditions however, are hard to realize in practice. First of all, in practice the process is never in the model set because a low order, accurate but approximate model is preferred for control design resulting in undermodelling of the process. Secondly, upper bounds of the noise cannot be guaranteed because in general this noise does not meet the assumptions on the probability distributions, e.g. a uniform distribution for bounded error identification (see Chapter 3). In addition, outliers in the data set, introduced during data acquisition or data preprocessing, together with the undermodelling would require a severe overestimation of the noise bounds for the computation of guaranteed upper bounds. Clearly, this results in very conservative (wide) bounds on the parameters. However, to obtain useful models for robust control design, the (parameter) bounds on the models have to be as tight as possible so that minimal conservatism is introduced to realize closed-loop systems with acceptable

performance. It cannot be expected that for a model with very conservative bounds a robust controller can be designed achieving reasonable performance.

Therefore, conservatism in parameter uncertainty has to be reduced as much as possible. In Section 5.1, various sources of conservatism in bounded error identification will be discussed. Most of these sources, like noise bound, linear constraint approximation and parameter uncertainty set approximation result from the defined solution to the identification problem. Another source of conservatism is the interdependence of parameters. Mutual relations between the parameters in the polynomial models, however, are not taken into account. The interdependence of parameters results from the fact that several parameters can describe the same dynamical input/output behaviour. It will be shown that for interdependent parameters in the polynomial models which are treated as independent variables, the corresponding parameter uncertainty will increase. This type of conservatism can be reduced by fixing parameters to predefined values which will be illustrated in Section 5.1. Although the consequences of this increased parameter uncertainty due to interdependent parameters are more severe for bounded error identification, this effect also appears in prediction error identification. This method, however, is less sensitive for this type of conservatism compared to bounded error identification. This provides the possibility, however, as will be shown in Section 5.2, to determine the correlations between the parameters in the stochastic setting and select an order in which parameters can be fixed in the bounded error approach. Finally, this interplay between prediction en bounded error identification to reduce the conservatism in the parameter bounds will be described in Section 5.3.

## 5.1 *Deterministic Parameter Uncertainty*

In bounded error identification or set estimation several effects can be distinguished which increase the conservatism on the parameter bounds (see Section 3.4). This conservatism results from several choices which have been made (see Chapter 3) to solve the identification problem. The most important sources will be described briefly :

- *Noise Bound* : Frequently, a  $l_\infty$ -bound will be used in set estimation to obtain guaranteed parameter estimates. Accurate results without introducing too much conservatism, however, can only be obtained if the noise has a uniform distribution. In practice, this is only the case for very specific examples (see [BBC88] where the error in the A/D converter of a digital voltmeter is considered). It has been shown in Section 3.4 that whenever the noise has a normal distribution,  $l_\infty$ -bounds on the prediction error give conservative results. For normally distributed noise,  $2\sigma$  or  $3\sigma$  bounds should be used to obtain reasonable results.

In practice, the noise will neither have a pure uniform distribution nor a pure normal distribution. Then, the noise bounds should be defined without introducing too much conservatism (e.g. a  $2\sigma$  bound). For this specific choice, it is known that not all noise samples will be within the selected bound resulting in an empty parameter set if no proper precautions are taken. An empty parameter set can be avoided by eliminating all constraints that violate a reference model, as described in Section 3.3.

- *Linear Constraint Approximation* : The parameter uncertainty set is guaranteed convex and connected if the constraints are linear in the parameters (see Fig. 3.1). For these type of models the constraints ( $n\theta-1$  dimensional linear hyperplanes) in bounded error identification are necessary and sufficient (e.g. EE models). In other type of models (e.g. OE, EIV models etc.), however, the parameters appear in a nonlinear form in the model representation. To avoid the problem of non-convex and possibly non-connected parameter uncertainty sets, the constraints need to be linearized. As described in Section 3.1, however, these linearized constraints are only necessary conditions, but not sufficient. This linearization to obtain convex and connected parameter uncertainty sets introduces conservatism because the sufficient conditions are not satisfied. This additional conservatism has been visualized in the example of Section 3.4, where EE and OE models are considered for uniformly distributed noise.
- *Parameter Uncertainty Set Approximation* : Several methods have been proposed in Section 3.2 to describe the parameter uncertainty set :
  - Exact polytope updating.
  - Ellipsoidal parameter bounding.
  - Orthotopic parameter bounding
  - Ellipsoid-aligned orthotopic parameter bounding.

Of all these methods, only the exact polytope updating describes the parameter uncertainty set exactly. However, this method can only be applied to low dimensional identification problems. For higher dimensional problems, the exploding computational complexity makes this method useless for multivariable identification in practice. Consequently, approximate parameter uncertainty set descriptions are preferred based on ellipsoidal, orthotopic or combined bounding techniques which compete with respect to computational complexity and accuracy in the approximation of the parameter uncertainty set. Of all approximate methods, the ellipsoid-aligned orthotopic parameter bounding adds the least conservatism at the cost of additional computational effort when

computing repeatedly the ellipsoidal and the corresponding ellipsoid-aligned orthotopic parameter uncertainty set.

- *Parameter Relation* : In a noisy environment with a limited number of data samples the identified parameters will be interdependent. This relation can result from the fact that in the chosen model structure a functional relationship between the parameters determines the dynamical input-output behaviour. This interdependence of parameters is not taken into account because the parameters in a black-box model are treated as independent variables. Clearly related parameters can be an indication for overparametrization in the black-box model. This will increase the parameter uncertainty bounds and therefore the conservatism of the resulting model set when applying bounded error identification techniques. Parsimonification of parameters, i.e. sparsity in the number of parameters to be estimated, is required to reduce this conservatism.

The amount of conservatism introduced by set estimation as described before results from the various steps in the identification procedure. Incorporating a priori knowledge such as noise bounds, model structures (linear constraint approximation) and the choice of identification (parameter uncertainty set approximation) and evaluation (parameter correlation) methods will be highly problem dependent.

When applying set estimation, the linear constraint method and parameter uncertainty set approximation have been introduced to solve the identification problem. The conservatism introduced in these two steps has to be taken for granted. The basis for accurate set estimation is uniformly distributed noise disturbing the process. In general, this is not the case in practice, but using the actual  $l_\infty$ -bound will give useless parameter estimates. This conservatism can be reduced significantly by defining a  $2\sigma$  bound as upper bound, but this alternative method requires a reference model to avoid an empty parameter set. It should be noted however, that because not all noise samples will be covered by a  $2\sigma$  bound, the parameter uncertainty set might not contain the true model (see Section 3.4).

The results of bounded error identification depend highly on the signal-to-noise ratio. For industrial processes, this ranges approximately from 10 to 20 dB. This low signal-to-noise ratio allows that even a low order model is not falsified by the measured data. So, in general, parameter set estimation tends to favour lower order models compared to classical prediction error identification. This is especially true for small signal-to-noise ratios. The reduced parameter identifiability in set estimation with decreasing signal-to-noise ratio results in a less accurate dynamical description of the process. However, the unmodelled dynamics might be relevant for control design.

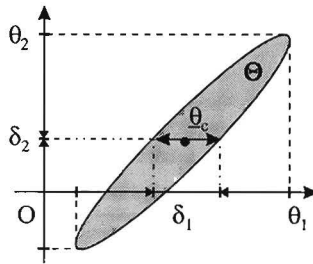
As indicated in Chapter 3, a reference model is required in practice to detect data outliers. This model can be obtained for example by classical identification methods. As



an alternative to model order reduction, the parameter uncertainty can be considered component wise. The uncertainty of several parameters is combined by fixing the related parameters except one in a group to the parameter values of the reference model. The idea of this approach is that no relevant dynamical information is lost but at the same time the dimension of the set estimation problem is reduced because less parameters are taken into account. In this way, the uncertainty present in the data is combined and translated into less free parameters. The value of the parameters which have been fixed is not set to zero, because fixing parameters in a polynomial to zero can change the roots significantly and thereby also the dynamical behaviour compared to the reference model.

The parameters which will be difficult to estimate and are therefore most uncertain, should be fixed to their reference value. This approach may seem counter intuitive. The idea, however, is that these parameters have the least contribution to the input/output behaviour and if such a parameter has been fixed to an incorrect value, the dynamical behaviour of the model will hardly be affected. On the other hand, if a parameter for which the input/output behaviour is very sensitive, is fixed to an incorrect value, the dynamical behaviour of the model will be influenced significantly.

This uncertainty reduction by fixing parameters has been illustrated in Fig. 5.1. The uncertainty of two highly related parameters can become very large. The uncertainty of  $\theta_2$  is even larger than 100%. However, if this parameter is fixed, i.e.  $\delta_2 = 0$ , the uncertainty  $\delta_1$  in the other parameter  $\theta_1$  reduces significantly.



**Fig. 5.1** : Reducing uncertainty in two highly related parameters by fixing one of the parameters. The shaded area indicates the parameter uncertainty set  $\Theta$ .

## 5.2 Stochastic Parameter Uncertainty

This effect of increased parameter uncertainty due to interdependent parameters is also of influence in prediction error identification. In this stochastic setting, the variance of

the parameters will increase if several parameters contribute to the same input/output behaviour. However, because the prediction error identification method is less sensitive to this type of conservatism compared to the bounded error identification, the parameter covariance matrix can be used to select the most correlated parameters.

The order in which the parameters should be fixed depends on the relative correlation between the parameters. This can be derived from the estimated parameter covariance matrix when applying prediction error identification. The parameter covariance matrix is a symmetric positive definite matrix of dimension  $n\theta$  where the diagonal entries describe the parameter variance and the off-diagonal elements indicate the parameter correlation. Whenever the parameters are independent the corresponding off-diagonal entry will be zero. An ordering of the correlations among the parameters can be established by computing the corresponding correlation coefficients [MS88]. The correlation coefficient  $\rho$  for two random variables  $x_1$  and  $x_2$  is defined by :

$$\rho = \frac{\text{Cov}(x_1, x_2)}{\sigma_1 \sigma_2} \quad -1 \leq \rho \leq 1 \quad (5.1)$$

where  $\sigma_1$  and  $\sigma_2$  are the standard deviations of  $x_1$  and  $x_2$  respectively. The normalized parameter covariance matrix is defined as follows :

$$\text{Cov}_{n,ij}(\hat{\underline{\theta}}) = \frac{\text{Cov}_{ij}(\hat{\underline{\theta}})}{\sqrt{\text{Cov}_{ii}(\hat{\underline{\theta}}) \text{Cov}_{jj}(\hat{\underline{\theta}})}} \quad (5.2)$$

and the absolute values of the entries in every column of  $\text{Cov}_n(\hat{\underline{\theta}})$  are summed :

$$\rho_{s,j} = \frac{1}{n\theta} \sum_{i=1}^{n\theta} \left| \text{Cov}_{n,ij}(\hat{\underline{\theta}}) \right| \quad (5.3)$$

Note that  $\text{Cov}_n(\hat{\underline{\theta}})$  equals 1 for all diagonal entries. The decreasing order of these summed parameter correlations ( $\rho_s$ ) indicates a decreasing dependence of one parameter to the other parameters. This will be translated into the order of the parameters which will be given a fixed value in the parameter uncertainty reduction procedure. Following this approach the parameter uncertainty can be reduced which results in a decrease of the conservatism introduced by interdependent parameters. An alternative to the 1-norm in Eq. 5.3 is for example the maximum value or  $l_\infty$ -norm in every column of  $\text{Cov}_n(\hat{\underline{\theta}})$ .

### 5.3 Parameter Uncertainty Reduction

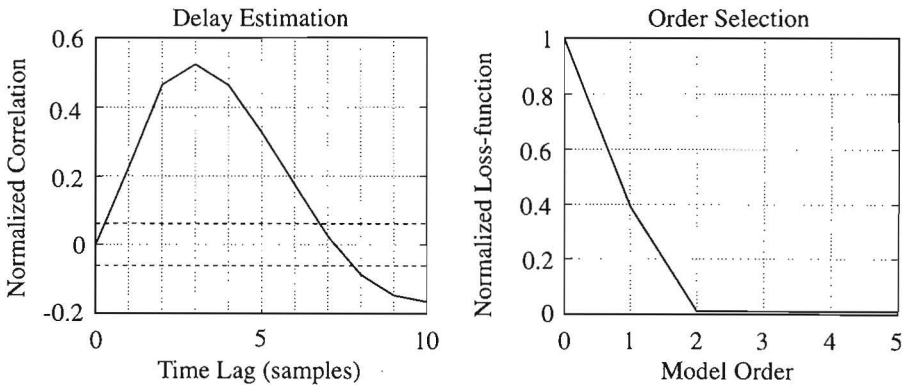
In general, when applying identification methods, almost unobservable parameters in a

model will be eliminated. In set estimation, however, parameters which are almost unobservable will be given a fixed value, for example obtained by classical identification methods, because their estimate is unreliable. This paradoxical approach implies that *uncertainty in fixed parameters is eliminated and therefore the uncertainty in the other parameters reduces as well*.

This effect of reduced parameter identifiability in set estimation for data with low signal-to-noise ratios and the proposed parameter uncertainty reduction approach will be illustrated by a simulation example. The true system in this simulation example is defined by the following OE-model :

$$M(\underline{\theta}) : y(k) = \frac{1 z^{-1} + 0.5 z^{-2}}{1 - 1.5 z^{-1} + 0.7 z^{-2}} u(k) + \xi(k) \quad (5.4)$$

As input signal  $u(k)$ , a PRBN sequence has been applied ( $u(k) \in [-1,1] \forall k$  and  $N = 1023$  samples) to simulate the output signal  $y(k)$  ( $\sigma_y = 4.343$ ). To illustrate the reduced parameter identifiability in set estimation, a uniform white noise sequence ( $\sigma_\xi = 4.343e-1$ ) has been generated realizing a signal-to-noise ratio of 20 dB. A uniform distribution of the noise has been chosen to provide a guaranteed upper bound of the noise,  $\|\xi(k)\|_\infty = 8.025e-1$ . Therefore, the only conservatism introduced in this example will be the linear constraint approximation of the output error model and a possible parameter correlation induced by the identification procedure.



**Fig. 5.2 :** a) Delay Estimation using input-output correlation with 95% confidence interval (dashed lines) ; b) Order Selection.

Given now the input and disturbed output data only, delay estimation and classical

prediction error identification can be applied to determine the model structure. The normalized correlation in Fig. 5.2a indicates a delay of one sample, since a clear correlation outside the 95% confidence interval starts at time lag one. Taking this delay into account, several models have been estimated for increasing model order for which the corresponding loss-function (sum of squared prediction errors, see Section 2.3.1) is depicted in Fig. 5.2b. As could be expected for this simulation example, the order selection indicates a second order model.

The identification results obtained from prediction error estimation, the parameter estimate  $\hat{\theta}$  together with a  $3\sigma$  uncertainty bound (corresponding to 99.7 % confidence) derived from the estimated parameter covariance matrix  $\text{Cov}(\hat{\theta})$ , are presented in Table 5.1. For comparison with set estimation, the results obtained by orthotopic parameter bounding, the central estimate  $\underline{\theta}_c$  and the parameter uncertainty  $\delta_\theta$ , are presented as well.

**Table 5.1** : Comparison of PE and OPB estimation.

	PE ( $\sigma_{OE} = 4.336e-1$ )		OPB ( $\sigma_{OE} = 5.261e-1$ )	
$\underline{\theta}_t$	$\hat{\theta}$	$3\sigma$ (%)	$\underline{\theta}_c$	$\delta_\theta$ (%)
1.000e+0	1.005e+0	4.75	1.027e+0	43.50
5.000e-1	5.010e-1	13.08	5.221e-1	108.56
-1.500e+0	-1.498e+0	0.67	-1.515e+0	12.40
7.000e-1	6.984e-1	1.17	7.136e-1	30.08

The comparison of the prediction error estimation and the orthotopic parameter bounding in Table 5.1, shows clearly the reduced parameter identifiability of set estimation for low signal-to-noise ratios. However, in practice the signal-to-noise ratio is often even worse and 20 dB SNR for industrial processes is in many cases the best one can hope for. To determine now the order in which the parameters should be fixed, the parameter covariance matrix has been estimated :

$$\text{Cov}(\hat{\theta}) = \begin{bmatrix} 1.43e-1 & -1.67e-1 & -1.02e-2 & 4.48e-3 \\ -1.67e-1 & 2.69e-1 & 2.86e-2 & -1.82e-2 \\ -1.02e-2 & 2.86e-2 & 6.28e-3 & -4.88e-3 \\ 4.48e-3 & -1.82e-2 & -4.88e-3 & 4.18e-3 \end{bmatrix} \quad (5.5)$$

which can be normalized according to Eq. 5.2 :

$$\text{Cov}_n(\hat{\underline{\theta}}) = \begin{bmatrix} 1 & -8.52\text{e-}1 & -3.40\text{e-}1 & 1.83\text{e-}1 \\ -8.52\text{e-}1 & 1 & 6.96\text{e-}1 & -5.42\text{e-}1 \\ -3.40\text{e-}1 & 6.96\text{e-}1 & 1 & -9.53\text{e-}1 \\ 1.83\text{e-}1 & -5.42\text{e-}1 & -9.53\text{e-}1 & 1 \end{bmatrix} \quad (5.6)$$

These results obtained by prediction error estimation indicate a clear correlation between the parameters, especially between the "b" and between the "f" parameters (-8.52e-1 and -9.53e-1 respectively). The corresponding parameter correlation values (Eq. 5.3) are given by :

$$\underline{\rho}_s = [ 5.94\text{e-}1 \quad 7.73\text{e-}1 \quad 7.47\text{e-}1 \quad 6.70\text{e-}1 ] \quad (5.7)$$

which results in : 2, 3, 4 and 1 respectively as the order of the parameter components to be fixed. In addition, an uncertainty  $\delta_{\theta}$  of more than 100% for  $\theta_2$  in Table 5.1 computed by orthotopic parameter bounding indicates already that this parameter should be fixed because of its reduced identifiability. The results of orthotopic parameter bounding for fixing 1 ( $\theta_2$ ) and 2 ( $\theta_2$  &  $\theta_3$ ) parameters respectively according to derived order are presented in Table 5.2.

**Table 5.2 :** Fixing parameters for OPB estimation.

	OPB <sub>1</sub> ( $\sigma_{OE} = 8.014\text{e-}1$ )		OPB <sub>2</sub> * ( $\sigma_{OE} = 4.378\text{e-}1$ )	
$\underline{\theta}_i$	$\underline{\theta}_c$	$\delta_{\theta}$ (%)	$\underline{\theta}_c$	$\delta_{\theta}$ (%)
1.000e+0	1.017e+0	42.73	9.997e-1	1.80
5.000e-1	5.010e-1	0	5.010e-1	0
-1.500e+0	-1.543e+0	9.73	-1.498e+0	0
7.000e-1	7.269e-1	21.21	6.961e-1	0.63

- 1) Parameter  $\theta_2$  has been fixed.
- 2) Parameters  $\theta_2$  and  $\theta_3$  have been fixed.
- \*) Violating constraints have been detected (9.79%) and eliminated (25.76%).

The parameter uncertainty reduction obtained by fixing the most correlated parameters is substantially. The parameter  $\theta_2$  is fixed to its prediction estimate because this is the most reliable information of  $\theta_2$  that is available. This reduces the parameter uncertainty in the other parameters only a few percent, but is a clear indication for the conservatism

present in the original set estimation problem (OPB estimation in Table 5.1) due to the linear constraint approximation of the OE-model and the interdependence of the parameters. The corresponding central estimate which has been computed from the extreme parameter values according to Eq. 3.37, however, differs significantly from the true system parameters resulting in a rather poor performance of the nominal model (compare the simulated output error,  $\sigma_{OE} = 8.014e-1$ , with the original noise,  $\sigma_E = 4.343e-1$ ).

Fixing the parameters  $\theta_2$  and  $\theta_3$ , reduces the parameter uncertainty significantly and provides an accurate nominal model ( $\sigma_{OE} = 4.378e-1$ ). It should be noted however that violating constraints with respect to the prediction error estimate  $\hat{\theta}$  have been detected and all constraints containing the same data samples have been eliminated to avoid an empty parameter set. This indicates that the resulting parameter uncertainty set is smaller than actually allowed, i.e. the parameter uncertainty set does not contain the true model:  $\theta_t \notin \Theta$ .

Of course, it can be argued that this approach of parameter uncertainty reduction forces the central estimate  $\theta_c$  in set estimation towards the estimate  $\hat{\theta}$  obtained by prediction error estimation which is then considered as the "true" system. However, it is the only way to reduce the conservatism induced by set estimation as described in this thesis to obtain reasonable estimates. The main problem in bounded error identification is the worst case approach, taking no averaging effect into account together with the several approximations in the set estimation. The parameter uncertainty set is determined only by a few constraints neglecting the information present in the other constraints. Because only the extreme parameter values are determined, the corresponding central estimate  $\theta_c$  can differ significantly from the prediction error estimate  $\hat{\theta}$  and is therefore of limited use when the bounds are relatively large. Bounded error models in set estimation with an accurate nominal model can only be obtained whenever the parameter uncertainty is small. This requires in general a high signal-to-noise ratio which can often not be realized in practice.

The ultimate goal in identification for robust control is to provide models with minimal model uncertainty. A conservative model uncertainty description will include physically unrealistic models in the model set. However, all elements of the model set have to be taken into account when designing a robust controller with guaranteed stability and performance. Of course, this will decrease the maximum achievable performance. Therefore, the conservatism has to be reduced as much as possible. The proposed method in this chapter, i.e. fixing interdependent parameters, can reduce the conservatism in the parameter uncertainty as will be illustrated in the case studies of Chapter 6.

## ***Case Studies***

---

6.1	A Watervessel Laboratory Process	6.3	A Distillation Column Simulation Process
6.2	A Fed-batch Fermentation Process	6.4	A Glass Tube Production Process

---

In this chapter, the prediction and bounded error identification methods which have been described in Chapter 2 and 3 respectively, are applied to various processes ranging from a SISO laboratory process to a multivariable industrial production process. The processes are described as detailed as necessary to provide a basic understanding of the dynamic behaviour of these processes. The experiment design and the data preprocessing (Appendix A) are described briefly where it is relevant for the identification methods. The main attention will be focused on the performance of the identification methods in terms of the accuracy of the resulting models and the computational complexity of the algorithms.

### 6.1 A Watervessel Laboratory process

The watervessel process is a laboratory process which is used for practical training purposes. This SISO process consists of three vessels which have been placed above each other. A rollerpump with rotation speed detector pumps the water from a supply vessel into the top vessel after which it flows via the middle one into the lowest vessel as shown in Fig. 6.1. The taps in the interconnection parts between the vessels act as flow resistances.

The water flow is the system input while the height of the water column in the lower vessel is the system output. The water flow which is used as model input cannot be measured directly in an accurate way without expensive sensors. The rotation speed of the rollerpump, however, can be measured easily and is a sufficiently accurate indication of the water flow. The height of the water column is measured by a level sensor consisting of two parallel plates at a constant distance which are isolated from each other. An AC-voltage is applied to the plates. The impedance between the plates is inversely proportional to the water level. So a constant AC-voltage on these plates will result in an AC-current, which is proportional to the water level. This smoothly nonlinear SISO process can be approximated by a linear model by studying the dynamic behaviour around an operating point.

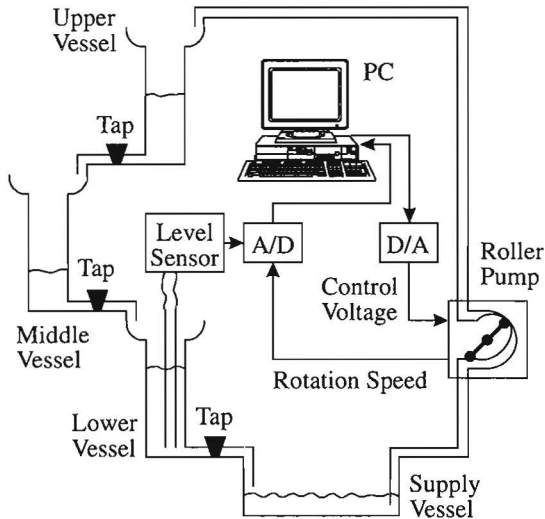


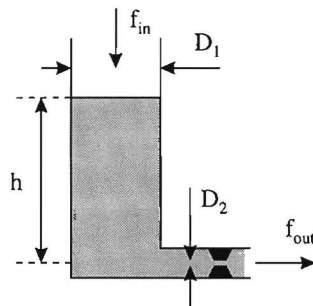
Fig. 6.1 : The watervessel process.



The identification results which will be described in this section have been published in [FD93]. The identification of this process using frequency domain methods can be found in [Lie91, vdB93a]. The actuators and sensors of this process have been studied extensively in [Lie91].

### 6.1.1 A Theoretical Vessel Model

A theoretical model of one vessel of the process can be derived using Bernouilli's law. Basically this law is an energy balance between two points in a process. If there is no pressure difference between those points and under the assumptions that the flow is laminar and the velocity uniform, the change in energy is assigned to varying kinetic and potential energy.



**Fig. 6.2** : A schematic view of a single watervessel.

Consider the variables indicated in Fig. 6.2. Here, the variables denote :

$D_1$	=	Diameter vessel	(m <sup>2</sup> )
$D_2$	=	Diameter tap	(m <sup>2</sup> )
$h$	=	Difference in height	(m)
$f_{in}$	=	Flow into vessel	(m <sup>3</sup> /s)
$f_{out}$	=	Flow out of hose	(m <sup>3</sup> /s)

As shown in [vdB93a], the nonlinear dynamic behaviour of a single water vessel can be described by :

$$f_{\text{out}} = A_2 \sqrt{2 h g} \quad (6.1)$$

$$\dot{h} = \frac{1}{A_1} (f_{\text{in}} - f_{\text{out}})$$

where  $A_1 = \frac{1}{4} \pi D_1^2$ ,  $A_2 = \frac{1}{4} \pi D_2^2$ . This model can be linearized by considering the behaviour of a vessel around a working point  $h = h_0$  resulting in :

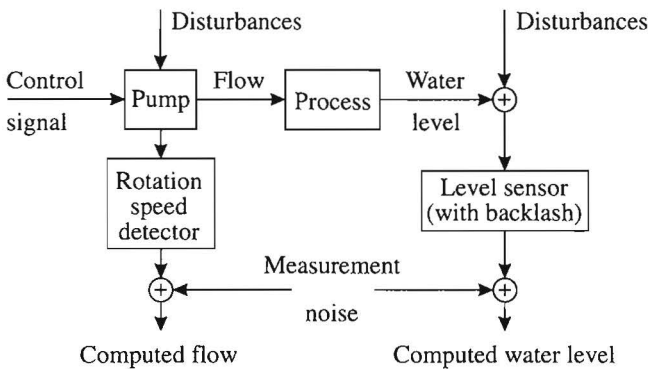
$$\dot{h} = \alpha f_{\text{out}} + \beta f_{\text{in}} \quad (6.2)$$

where

$$\alpha = -\beta = \frac{A_2}{A_1} \sqrt{\frac{g}{2 h_0}} \quad (6.3)$$

### 6.1.2 Bounds on the Disturbances

Bounded error identification is based on the fact that upper bounds on the disturbances which affect the process are available. All disturbances are indicated in Fig. 6.3.



**Fig. 6.3** : Disturbances on the process.

Basically there are three effects that may cause disturbances and therefore uncertainty in the determination of the water level (output disturbances) :

- Errors in the level sensor, i.e. the inaccuracy of the output sensor voltage to represent the water level in the vessel.
- Backlash effects in the level measurements due to the adhesion between the water and the parallel plates of the level sensor.
- Disturbances of the water surface in the vessel caused by the water flow.

The upper bound on the total output error is approximately 2.25 mm where 0.25 mm is due to the measurement errors and 2 mm because of backlash and disturbances [Lie91]. It has been shown in [Lie91] that the dynamics of the roller pump can be neglected because of its small time-constant compared to the time-constant of the watervessel process. In addition, fluctuations in the water flow caused by the roller pump can be neglected because it contains frequencies which are far beyond the bandwidth of the watervessel process. The water flow can be described most accurately by the average rotation speed of the pump. Variations in the rotation speed results in a flow error of  $8 \cdot 10^{-7} \text{ m}^3/\text{s}$  (input disturbance).

### 6.1.3 Reference Model

A priori information about the structure of the model as well as the signs of the parameters is necessary to solve the bounded error identification problem. Therefore an initial model is required. Further, to be able to make the parameter set identification robust against outliers (see Section 3.4), it is assumed that this initial model defines an element of the final feasible parameter uncertainty set  $\Theta$ .

Because the input of the process (the water flow) is not exactly known, prediction error identification cannot be used to estimate an initial model. Therefore, the Total Least Squares (TLS) estimation method has been adopted which is described in [vHV91]. The TLS problem considers an overdetermined set of linear equations  $A \hat{\theta} = b$ , obtained for example from Eq. B.15 containing scaled input and output data, where both the data matrix  $A$  as well as the observation vector  $b$  are inaccurate. The scaling is based on the bounds of the input/output noise. If the perturbations  $D$  on the data  $[A \ b]$  have zero mean and their covariance matrix  $E(D^T D)$ , with  $E$  the expected value operator, equals identity up to an unknown scaling factor (e.g. when all errors are independent and equally sized), then a strongly consistent estimate of the true solution of the unperturbed set can be computed where  $\|\Delta \hat{A} \ \Delta \hat{b}\|_F = \|[A \ b] - [\hat{A} \ \hat{b}]\|_F$  is minimized. Surely this condition cannot be fulfilled completely in practice, but it is reasonable to assume that the resulting initial model is a member of the feasible model set.

A Pseudo Random Binary Noise Sequence (PRBNS) is used to excite the process resulting in a data set of 1100 samples. After signal preprocessing as described in Appendix A, the input and output data,  $u(k)$  and  $y(k)$  respectively, has been normalized

using the standard deviations of the corresponding error signals. Normalization to equal sized errors is required to obtain a consistent initial model using the TLS estimation method. This results in :

$$y(k) = -f_1 y(k-1) - f_2 y(k-2) + b_1 u(k) + b_2 u(k-1) + b_3 u(k-2) \quad (6.4)$$

$$1.09 y(k-1) - 0.29 y(k-2) + 0.21 u(k) + 0.27 u(k-1) + 0.07 u(k-2)$$

Because of the first order nature of each vessel (Eq. 6.2), a third order model can be expected on physical considerations. However, a second order model turned out to be sufficiently accurate because the sampling time of this data set is approximately 1.5 times the time constant of the upper vessel which has therefore almost no contribution to the input-output behaviour resulting in a second order model. A higher sampling frequency is required to estimate a third order model.

#### 6.1.4 Bounded Error Estimates

For this errors-in-variables model a constraint set has been generated according to Eq. 3.14 and Eq. 3.32 using the signs and the structure of the TLS estimate in Eq. 6.4 together with the amplitude bound  $\delta_u = 8 \cdot 10^{-7} \text{ m}^3/\text{s}$  on the input disturbances, and the amplitude bound  $\delta_y = 2.25 \cdot 10^{-3} \text{ m}$  on the output disturbances. Under the assumption that this reference model :  $\hat{\theta} = [2.06\text{e-}1, 2.71\text{e-}1, 6.67\text{e-}2, -1.09, 2.91\text{e-}1]^T$ , is an element of the parameter uncertainty set  $\Theta$ , 50 ( $\approx 2.3\%$ ) violating constraints have been detected. After outlier correction which has been described in Section 3.3, the modified data set can be used for parameter set estimation.

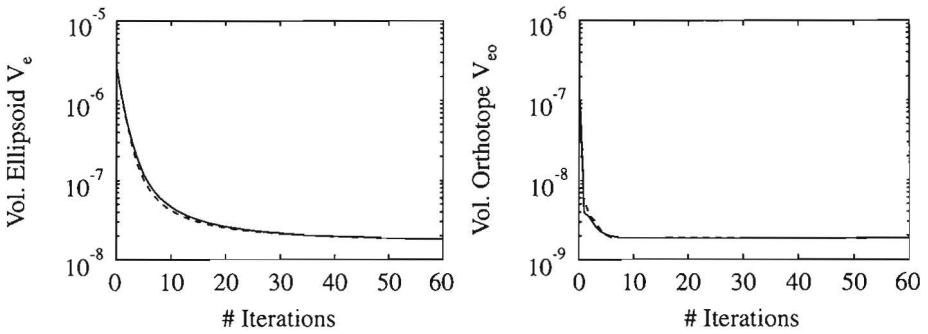
**Table 6.1** : Orthotopic bounding.

Par.	$\hat{\theta}$	$\theta_c$	$s_\theta$	%
$b_1$	2.06e -1	2.05e -1	8.70e-3	4.3
$b_2$	2.71e -1	2.71e -1	1.48e-2	5.4
$b_3$	6.67e -2	5.81e -1	1.88e-2	32.3
$f_1$	-1.09e+0	-1.11e+0	4.92e-2	4.5
$f_2$	2.91e -1	3.07e -1	4.30e-2	14.0

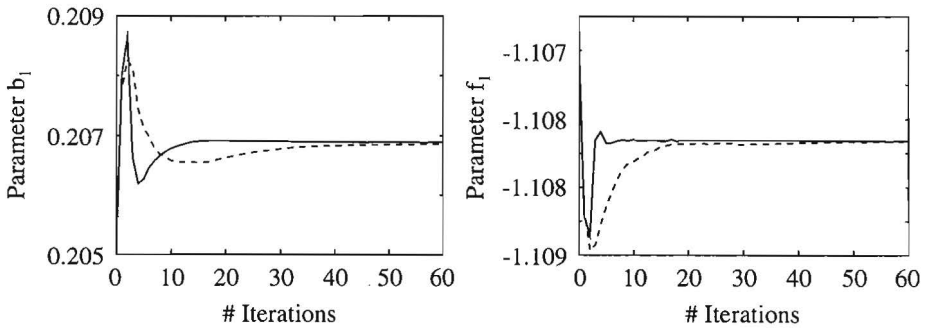
Orthotopic bounding ( $\Theta_o$ ) results in the feasible parameter set described in Table 6.1 where the central estimate  $\theta_c$  and the parameter uncertainty  $s_\theta$  is computed according to Eq. 3.37. The volume of this initial orthotope is given by :

$$V_o = \prod_{i=1}^5 2 s_{\theta,i} = 1.73 \cdot 10^{-7} \quad (6.5)$$

As initial ellipsoid a hypersphere of minimal volume, defined by :  $P^o = r^2 I$ , is used enclosing the initial orthotope where the radius is defined by :  $r = \sqrt{s_{\theta}^T s_{\theta}}$ .



**Fig. 6.4** : Parameter uncertainty reduction in set estimation by repeated constraint processing ; a) Ellipsoidal volume reduction and b) Ellipsoid-aligned orthotopic volume reduction for normal (solid) and reverse constraint processing (dashed).



**Fig. 6.5** : Behaviour of central parameter estimates during repeated constraint processing ; a)  $b_1$  and b)  $f_1$  for normal (solid) and reverse (dashed) constraint processing.

Repeated constraint processing has been applied to eliminate the influence of the initial ellipsoid and the sequence in which the constraints are processed (see Section 3.2.2). Fig. 6.4 and 6.5 show now after every iteration the volume reduction of the ellipsoid and the ellipsoid-aligned orthotope together with the behaviour of the central parameter estimates  $b_1$  and  $f_1$  respectively. To show that the final solution is indeed independent of the sequence in which the constraints are processed, the solutions after every iteration step of both normal and reverse constraint processing are depicted in the figures. The oscillation of the centre value during the first iterations is caused by small variations of the coordinate system obtained from the ellipsoidal bounding method and disappears when the ellipsoidal volume converges to its final value. After satisfying the stop criterion, orthotopic bounding with coordinate transformation then results in an ellipsoid-aligned orthotope  $\Theta_{eo}$  with volume :  $V_{eo} = 1.86 \cdot 10^{-9}$ , which is a significant reduction compared to  $V_o$ .

The application of the bounded error identification method to this watervessel laboratory process shows clearly that ellipsoidal bounding can be used as a preprocessing tool to derive an optimal orthonormal base for orthotopic bounding as motivated in Section 3.2.4. The resulting ellipsoid-aligned orthotope has a significant smaller volume. Convergence has been achieved independently of the sequence in which the constraints are processed and initial conditions, i.e  $P^0$  (see Fig. 6.4 and Fig. 6.5). All models in the parameter uncertainty sets  $\Theta_o$  and  $\Theta_{eo}$  are also stable because these sets are within the stability triangle of a second order system which is defined by the inequalities :  $f_2 < 1$ ,  $1 + f_1 + f_2 > 0$  and  $1 + f_1 + f_2 > 0$  (see [GS84] ).

## 6.2 A Fed-batch Fermentation Process

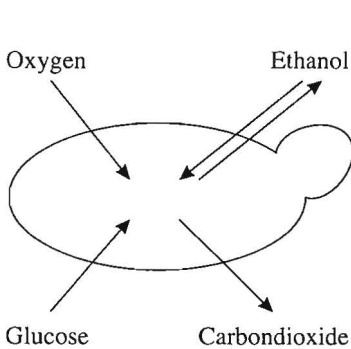
The application of bounded error identification methods to a fed-batch fermentation process for the production of bakers' yeast (*Saccharomyces cerevisiae*) describes a possible approach to obtain the required a priori information from a single data set.

Nonlinear behaviour is the rule, rather than the exception, in the dynamical behaviour of biotechnical processes. The culture of *Saccharomyces cerevisiae* grown in a fed-batch reactor is a highly nonlinear and time-varying process due to its changing reactions and process nonlinearities. The identification of the overall process can be divided roughly into 3 phases where only specific operation conditions will be considered to obtain a time-invariant behaviour. First the data is compensated for assumed static nonlinearities which will be explained later. Applying classical prediction error identification techniques to the resulting "linear" data will provide the required a priori knowledge for parameter set estimation. Combined ellipsoidal/orthotopic bounding (see Section 3.2.4) can then be used to obtain minimal parameter uncertainties. The accuracy of the central estimates will be compared to the nominal model obtained by classical prediction error

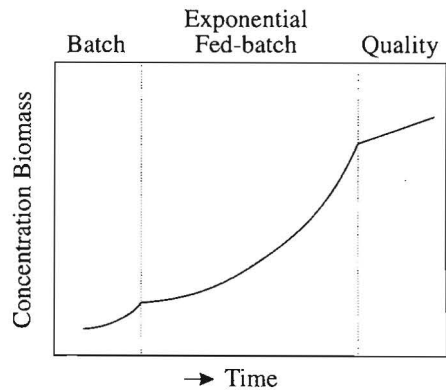
identification. These identification results have been published in [FK94]. The following process description is taken from [Keu93].

### 6.2.1 Yeast Production

The production of yeast is closely related with the yeast (cell) itself and with its biochemical reactions (dynamics). A basic description of yeast is thus relevant to understand the specifications and limitations for production. A graphical representation of a yeast cell is shown in Fig. 6.6. The yeast cell grows through cell division, also called budding (the small knot at the right hand side in Fig. 6.6). For this growth, the cell needs building material and energy. The energy and building material are derived from carbon sources, e.g. glucose. Beside the carbon sources the cell needs various other substances, e.g. minerals, vitamins, etc. Energy is most effectively generated with the aid of oxygen, the so called oxidative phosphorylation. If no oxygen is present, the yeast cell can grow as well, but less efficient. The energy is derived through production of ethanol. However, high concentrations of ethanol are toxic to the yeast (wine and beer production). As ethanol is a carbon source as well, the cell can use the ethanol for growth. However, yeast cells have a preference of sugars over ethanol. During growth, the yeast cell produces carbon dioxide, whether the cell grows on glucose or ethanol.



**Fig. 6.6** : A schematic representation of a yeast cell and its main inputs and outputs.



**Fig. 6.7** : The concentration of biomass as a function of time during the three phases of the laboratory process.

A physiological model of the yeast is made to understand the cell dynamics and its

nonlinearities better. The model is based on the oxygen limited capacity of the yeast cell [SK86] combined with a regulated enzyme production of the cell and a self containing term [Kou93]. The hypothesis of [SK86] divides the growth of yeast into three different pathways :

- Under normal conditions the yeast consumes glucose together with oxygen, without producing ethanol.
- If the glucose consumption is larger than the oxidative capacity of the cells, a reductive reaction occurs producing ethanol.
- Ethanol can be consumed oxidatively if the glucose consumption is less than the oxidative capacity.

The yeast cell metabolizes glucose and ethanol according to one or a combination of the three above mentioned pathways. Switching between the pathways is regulated through specific enzyme production. A basic nonlinearity is found in the budding. Cell concentration equations are exponential at constant growth (Eq. 6.6). The switching between pathways causes nonlinear effects as the growth efficiency is higher on glucose than on ethanol. The time-varying effects are caused by the specific enzyme production which regulates the switching between the pathways.

Bakers' yeast is produced on a industrial scale. A "translation" is made from this industrial scale to a laboratory process. The laboratory process has three distinct phases (Fig. 6.7) :

- A batch phase in which the start concentration of the yeast is cultivated.
- A fed-batch phase in which the maximum yield of yeast is the main objective.
- A quality phase in which the yeast is cultivated according to certain quality constraints.

Together they present the seed and propagation stages of the industrial process. Here the attention is focused on the fed-batch phase in the production of bakers' yeast. The fed-batch phase is normally operated in such a way that a minimum amount of glucose is needed for a maximum yield (concentration) of yeast cells. This under the constraints of no ethanol production and a maximum available oxygen supply.

### **6.2.2 The Laboratory Process**

The laboratory process is carried out in an aerated, continuously stirred fermenter. This 10 litre fermenter (Fig 6.8) is the core of the fermentation system. The glasswall vessel



is fitted with an airtight rubber ring on both bottom and top. The top plate contains a number of ports that are used to support a pH probe, a dissolved oxygen probe, an anti-foam sensor, an outlet air-cooling device, and several ports to add or withdraw chemicals. The bottom plate contains ports for a temperature probe, a heating and cooling element, a combined air and feed inlet, a sample needle and a waste outlet. Agitation is provided by a magnetically-coupled motor with adjustable speed. An impeller and four baffles provide good mixing. Air is blown into the fermenter through a small pipe ending just below the impeller, providing constant mixing of the air. The fermenter is held sterile during the fermentation in order to avoid contamination. The laboratory process is performed at favourable environmental conditions, i.e. temperature at  $30^{\circ}\text{C}$  and pH at 5.

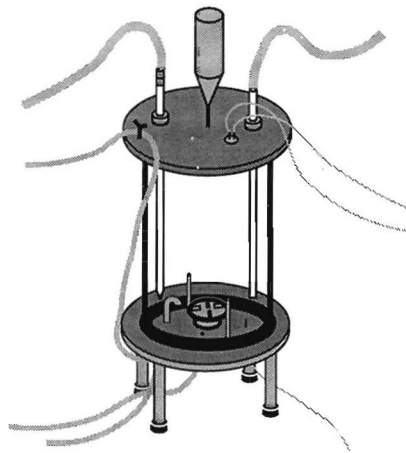
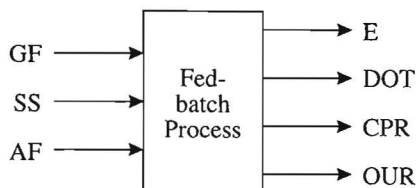


Fig. 6.8 : The fermenter.

Looking at the process from a control engineering point of view, it can be seen as a three input, four output MIMO system (Fig. 6.9). The inputs are : stirrer speed (SS), air flow (AF) and glucose flow (GF). The outputs are : ethanol concentration (E), dissolved oxygen tension (DOT), the carbon-dioxide production rate (CPR) and the oxygen uptake rate (OUR). In practice the production requirements placed upon the controlled process are : a maximal yeast cell production with negligible ethanol production. Measuring the cell concentration on-line is not always feasible. Consequently, an auxiliary signal has to be found to control the cell production. This auxiliary signal is the oxygen uptake rate. The oxygen uptake rate depends on all inputs. However, the dependence on the stirrer speed and air flow is physically and occurs only if the oxygen is limited. The

glucose flow has a pure biochemical influence on the cells. The appropriate combination for controlling the cell concentration production is thus the glucose flow and the oxygen uptake rate. Here, only the oxidative growth on glucose will be considered. This limitation is chosen because of the severe nonlinearities and time-varying behaviour of the yeast production when all three pathways occur. It is also practical, as industrial production is focused around oxidative growth on glucose because of the high yield. Note that the dynamics of the carbondioxide production rate are comparable to the oxygen uptake rate dynamics, only the gain is different.



**Fig. 6.9** : A simplified view of the fed-batch process.

### 6.2.3 Nonlinear Compensation

As mentioned before, only oxidative growth on glucose will be considered. This part of the process, if no ethanol is present, can be regarded as linear and time-invariant. However, the process itself remains highly nonlinear due to the exponential character of the cell growth. This nonlinearity associated with the cell growth will be compensated in the sense that assuming the amount of cells and overall growth rate is known, all input signals can be compensated with the estimated growth of the cells, an exponential function. The initial cell concentration and the expected growth rate are obtained from information supplied by estimators, dry-weight and optical density measurements. The nonlinear compensation is simply related to the cell concentration (Eq. 6.6) as an exponential increasing function upon the growth rate ( $\mu$ ) of the cells :

$$V_x(t) = V_x(0) e^{\mu(t)t} \quad (6.6)$$

where  $V_x$  = amount of biomass, and  $\mu$  = specific growth rate. Considering the input signals, the glucose flow (GF) is directly related to the modelled growth :

$$GF(t) = \frac{\mu(t)}{Y_{x|s} S_o} V_x(0) e^{\mu(t)t} \quad (6.7)$$

where  $Y_{x|s}$  = yield of cells on glucose,  $S_o$  = glucose concentration in the feed. The stirrer speed and the air flow are also directly correlated with the modelled growth. With the proper nonlinear compensation, the expected dissolved oxygen concentration is at constant level and is large enough to guarantee that no oxygen limitation occurs. The carbondioxide production rate and the oxygen uptake rate are expected to increase with the glucose flow as they are directly related to the amount of biomass in the fermenter. Fig. 6.10 and 6.11 depict the measured glucose flow and oxygen uptake rate (a) together with the compensated data (b) respectively. The unknown parameters of the nonlinear exponential compensation (Eq. 6.7) have been tuned to achieve the best possible fit between the measured data ( $y_m$ ) and the assumed nonlinear compensation ( $y_o e^{\mu t}$ ) :

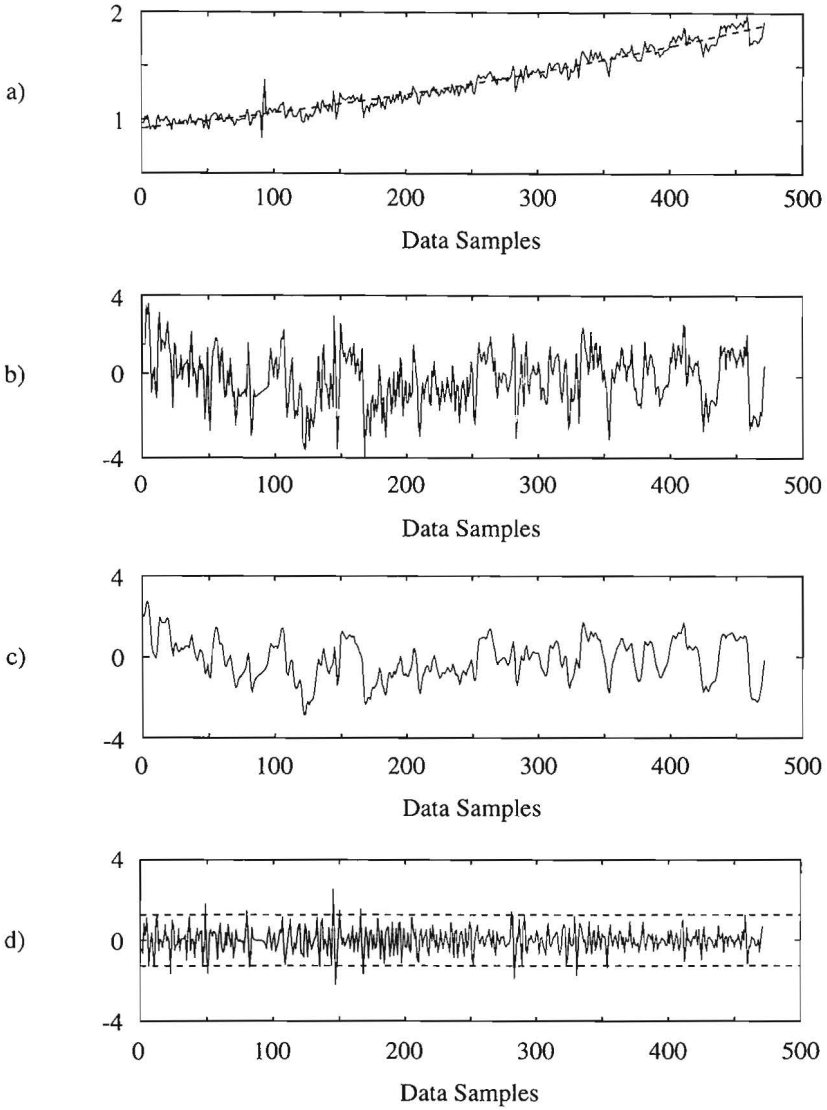
$$\min_{y_o, \mu} \left\| y_m - y_o e^{\mu t} \right\|_{L_2[0, T]} \quad (6.8)$$

### 6.2.4 Linear Identification

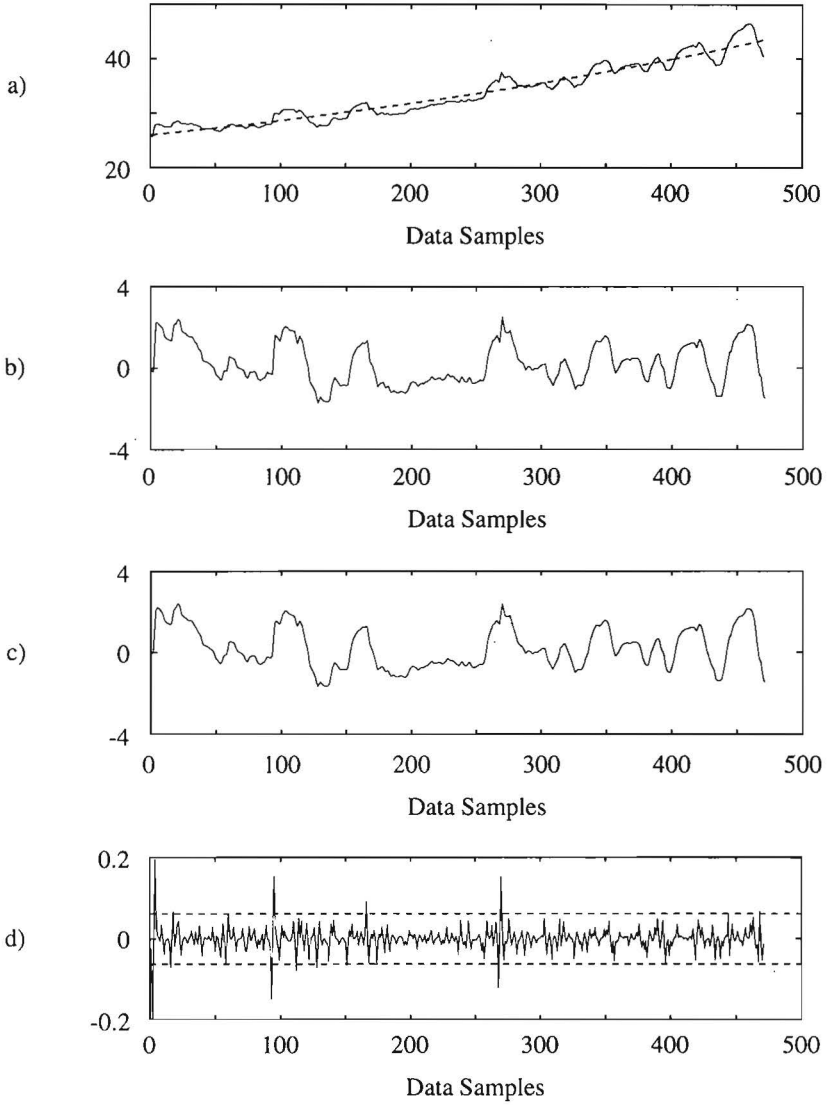
Before set estimation techniques can be applied, a priori knowledge, like model structure, signs of the parameters and noise bounds on input and output, is required. Because only the current data set is available without any extra noise information, the data has been filtered resulting in a low-frequent (LF) and a high-frequent (HF) component of both input and output signals. (Fig. 6.10 and 6.11 c,d). The separation frequency for the filters has been chosen in such a way that the auto-correlation of the HF-component (noise) is as white as possible. This noise component can then be used to determine the bounds on the input/output noise. The remaining LF-component is suitable to apply classical identification techniques.

**Table 6.2** : Order selection.

$n\theta$	$V_e(\theta)$	AIC	FPE
2	2.72e-1	7.27e+2	2.74e-1
3	2.62e-1	7.12e+2	2.65e-1
4	2.51e-1	6.93e+2	2.55e-1
5	2.50e-1	6.94e+2	2.56e-1
6	2.51e-1	6.97e+2	2.57e-1



**Fig. 6.10** : Glucose Flow (GF) ; a) Measured GF (solid) and nonlinear compensation (dashed), b) Compensated GF, c) LF-component GF, d) HF-component GF with bound  $\delta_u = 1.27$  (dashed).

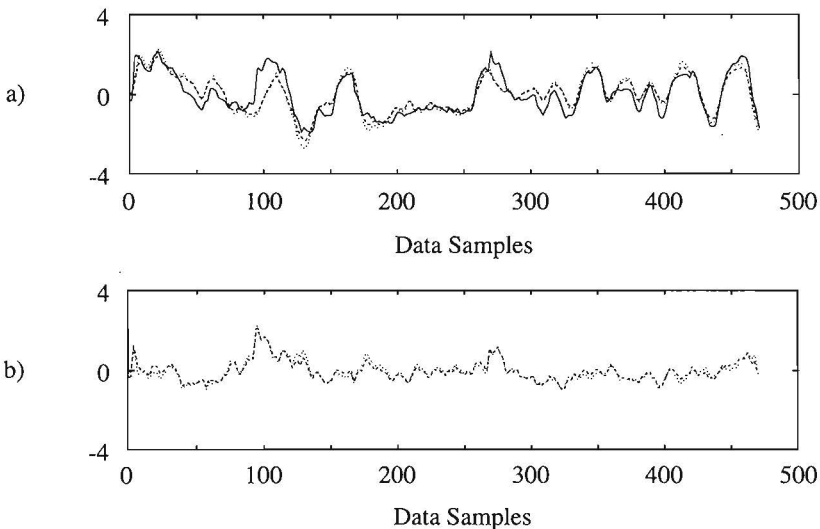


**Fig. 6.11 :** Oxygen Uptake Rate (OUR) ; a) Measured OUR (solid) and nonlinear compensation (dashed), b) Compensated OUR, c) LF-component OUR, d) HF-component OUR with bound  $\delta_y = 6.22e-2$  (dashed).

To obtain a model which is suitable for long horizon control, output-error identification has been applied to the filtered data. Because the order of the model is unknown, the loss-function, the AIC and FPE criterion (Eq. 2.27) have been determined for increasing order. The results are given in Table 6.2. The models have been fully parametrized with or without one delay to achieve a continuous increasing number of parameters ( $n\theta$ ). Theoretically, the loss-function  $V_e(\theta)$  decreases continuously for increasing model order. However, due to overparametrization, the optimization might not converge correctly resulting in a larger value of the loss-function compared to a model of lower order.

According to the AIC and the FPE criterion a second order model with one delay ( $n\theta = 4$ ) is optimal. The corresponding model simulation together with the output-error are depicted in Fig. 6.12. The relative variance of the output-error is 26%. The remaining misfit between the measured oxygen uptake rate and the simulated model output cannot be reduced significantly by increasing the model order. This misfit is mainly due to the nonlinearities still present in the data [KAG94].

The parameter estimates and the corresponding stochastic parameter uncertainty set  $\Theta_{pe}$  are given in Table 6.3. These bounds are computed from the covariance matrix of the parameter estimates (see Appendix C) where a  $2\sigma$  bound has been adopted which corresponds to 95% confidence.



**Fig. 6.12 :** a) Measured oxygen uptake rate (solid), Simulation output-error model (dashed) and central estimate of set model (dotted), b) Error ; Output-error model (dashed) and central estimate of set model (dotted).

### 6.2.5 Set Estimation

To obtain an uncertainty model which is suitable for robust control design, set estimation can be used resulting in hard bounds on the parameters. The required a priori information has been derived in the previous section. Consider now the model structure of the optimal output-error model :

$$y(k) = \frac{b_1 z^{-1} + b_2 z^{-2}}{1 + f_1 z^{-1} + f_2 z^{-2}} u(k) + \xi(k) \quad (6.9)$$

Contrary to the output-error identification where the filtered data has been used, the fact is included that both input and output are corrupted by noise using the original compensated data where only a bounded-error assumption will be made on the input/output noise (see Eq. 3.7 and Eq. 3.8). The noise bounds have been determined in the previous section from the high-frequent components of the oxygen uptake rate ( $\delta_y = 6.22e-2$ ) and the glucose flow ( $\delta_u = 1.27$ ) respectively.

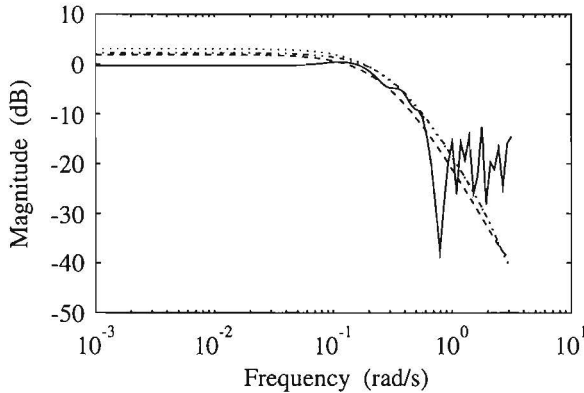
**Table 6.3 :** Parameter identification of a fed-batch fermentation process.

Par.	$\Theta_{pe}$		$\Theta_o$		$\Theta_{eo}$	
	$\hat{\theta}$	% (2 $\sigma$ )	$\theta_c$	%	$\theta_c$	%
$b_1$	5.40e -2	7.7	6.32e -2	38.2	6.30e -1	(10.0)*
$b_2$	1.97e -2	34.6	2.64e -2	>100	3.50e -2	(10.5)*
$f_1$	-1.50e+0	1.2	-1.45e+0	7.2	-1.45e+0	(3.3)*
$f_2$	5.55e -1	2.9	5.23e -1	19.3	5.18e -1	(6.4)*

\*) Relative uncertainty in ellipsoid-aligned coordinate system.

The constraints for bounded error identification are constructed according to Eq. 3.14. The figures 6.10d and 6.11d indicate clearly that outliers in the data violate the defined error bounds. To avoid that the parameter uncertainty set  $\Theta$  becomes empty the output-error model obtained by classical identification has been used as reference model (see Section 3.4) to eliminate all violating constraints ( $\approx 5.5\%$ ). Ellipsoid-aligned orthotopic bounding with repeated constraint processing has been applied to approximate the parameter uncertainty set  $\Theta$ . To illustrate the achievable reduction of the parameter uncertainty taking the orientation of the parameter set into account, Table 6.3 shows the

results of the orthotopic set estimation, with and without orientation. The simulation results using the central estimate of the ellipsoid-aligned orthotope  $\Theta_{eo}$  are depicted in Fig. 6.12. Comparing Fig. 6.12b and Fig. 6.11d indicates that the error bound is defined too optimistic because no model errors are taken into account. The variance of the output-error (28%) is only slightly worse compared to the output-error model obtained by classical identification (26%). For comparison of the models in Table 6.3, the spectrum analysis of the input/output data together with the amplitude plots of the central estimates have been depicted in Fig. 6.13. Due to limitations in the available measurement equipment the sampling time which can be realized is in fact too low and has caused aliasing in the data.



**Fig. 6.13 :** Magnitude plots of fed-batch fermentation models ; Spectrum analysis of input/output data (solid), prediction error estimate  $\hat{\theta}$  (dashed), central estimate  $\hat{\theta}_c$  of  $\Theta_o$  (dash-dotted) and central estimate  $\hat{\theta}_c$  of  $\Theta_{eo}$  (dotted).

This example illustrates that even if the amount of information is limited to a single, relative short data set, set estimation can provide reliable models. The output-error estimate and the central estimates obtained by bounded error identification are not significantly different. However, the bounded error approach requires more effort to obtain accurate models. Additional information like bounds on the disturbances, outlier correction to avoid an empty parameter set and repeated constraint processing in ellipsoid-aligned orthotopic bounding to obtain a minimum uncertainty set  $\Theta_{eo}$ , make this identification method more complex from a computational point of view.



### 6.3 A Distillation Column Simulation Process

The ever increasing automation in industry and the application of multivariable and adaptive control systems has led to the development of a specific software package by STAMICARBON (an engineering company of DSM) called GIDS : Graphical Interactive Dynamical Simulation system. GIDS is a real-time computer simulation system which has been developed for the training of operators, instrumentation technicians in the process industry and students in technical education. GIDS contains a simulation of a chemical process, i.e. a binary distillation column, which is specifically designed to teach some essential aspects in the control of a physical process and to study the dynamical behaviour of this process. The education objectives are :

- To operate a chemical process by simulated instrumentation and control.
- To obtain insight in the dynamical behaviour of a chemical process from a theoretical as well as a practical point of view, without danger for people, environment, equipment and/or economic risks.
- To startup control loops without disturbing the process.
- To study and/or optimize the applied control strategies and to work with advanced process control.

A typical two-product distillation column is shown in Fig. 6.14. This simulation process is based on algebraic and nonlinear differential equations which describe the physical behaviour of the chemical process.

The distillation column separates a mixture of hexane ( $C_6H_{14}$ ) and heptane ( $C_7H_{16}$ ). The feed is preheated in a heat exchanger and the mixture is fed into the column as a saturated liquid. The overhead vapor stream is cooled and completely condensed after which it flows into the reflux tank. The cooling of the overhead vapor is accomplished with cooling water. The liquid from the reflux tank is partly pumped back into the column and is partly removed as the distillate or top product. At the base of the distillation column, a liquid product stream (the bottom product) is partly removed and partly recirculated after it has been heated with steam in the reboiler. More detailed information in the field of distillation dynamics and control can be found in various textbooks, e.g. [Shi84]. A more recent overview is given in [Sko92].

The objective of this section is to show that set estimation identification can be applied to multivariable systems. An indication of the signal-to-noise ratio is required to guarantee reasonable estimates. In addition, it will be shown that fixing the value of parameters which have almost no contribution to the input/output behaviour can reduce the uncertainty in the remaining parameters significantly.

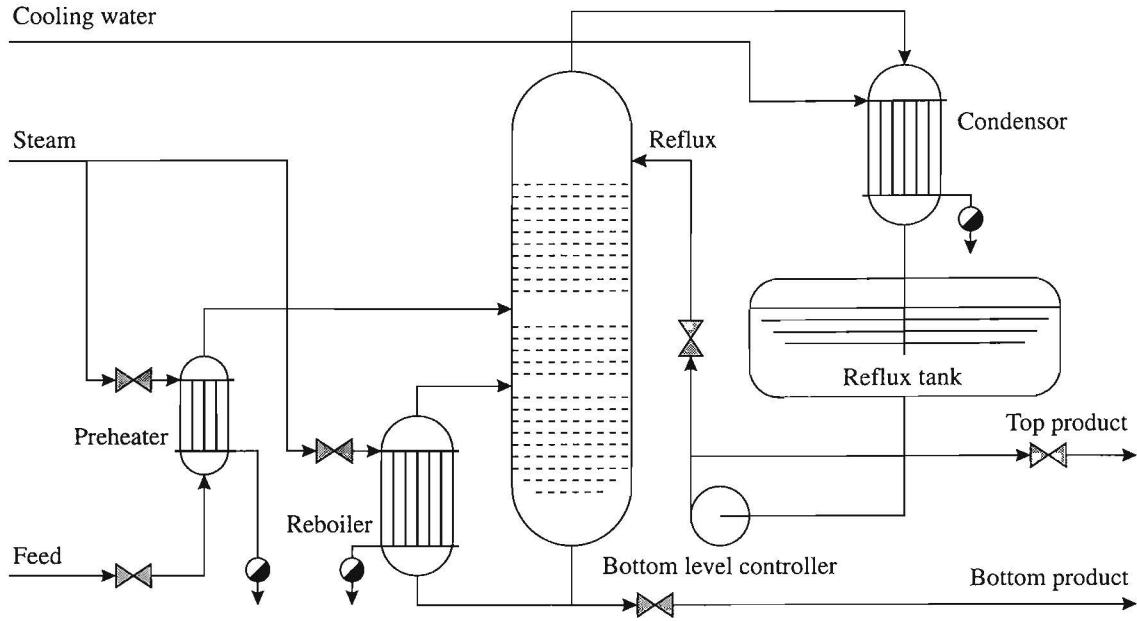


Fig. 6.14 : Schematic diagram of a distillation column.

### 6.3.1 Experiments and Model Structure Selection

Although this simulation process is typically nonlinear, the dynamical behaviour around an operating point can be described sufficiently accurate by a linear model. The inputs and outputs which have been selected to model this chemical process are described in Table 6.4. For multivariable identification, independent PRBNS signals have been applied to the 3 inputs simultaneously. The excitation in the operating point has been maximized maintaining linear process behaviour. The quality of the top (hexane,  $C_6H_{14}$ ) and bottom (heptane,  $C_7H_{16}$ ) product, the system outputs, are measured as impurities of heptane and hexane respectively. The output range is also given in Table 6.4.

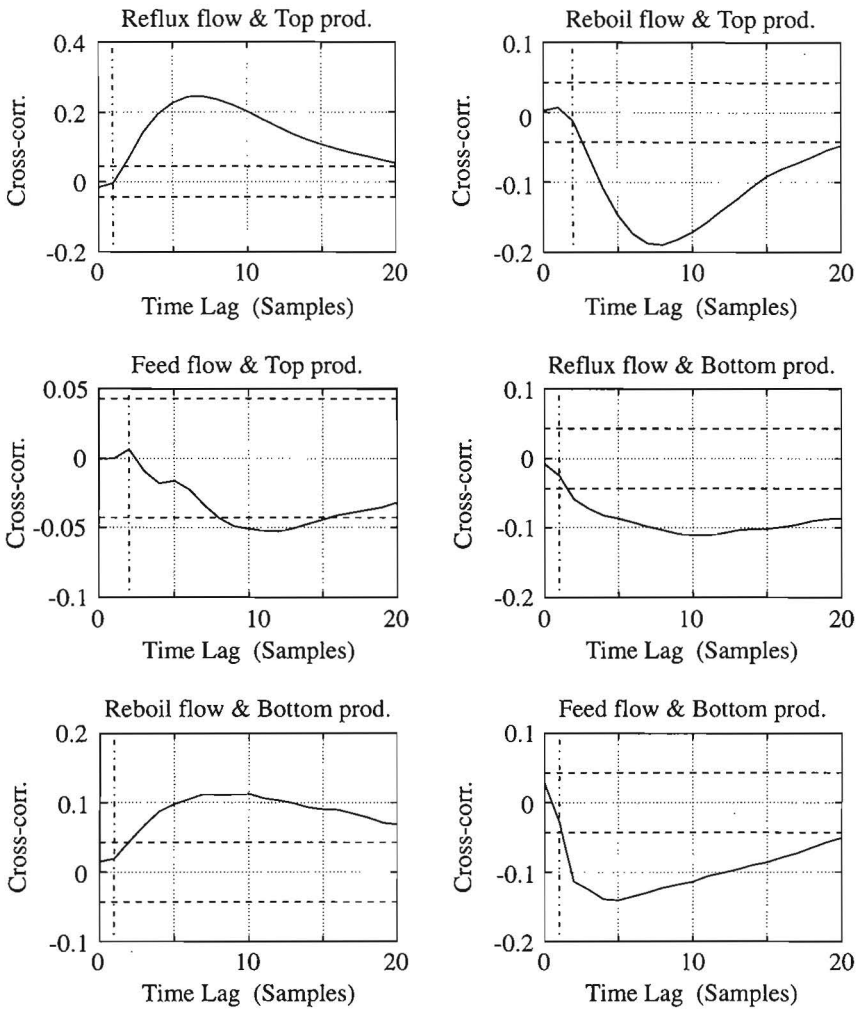
**Table 6.4 :** Identification experiment.

Inputs	Excitation (ton/hr)		
Reflux flow	$29.7 \pm 2.0$		
Reboil flow	$7.0 \pm 0.4$		
Feed flow	$15.8 \pm 5.0$		
Outputs	Min. (%)	Mean (%)	Max (%)
Top product, impurity $C_7H_{16}$	1.38e-2	2.80e-2	5.45e-2
Bottom product, impurity $C_6H_{14}$	0.69	5.92	15.17

Because this is a simulation process and no additional disturbances, like process or output noise, have been applied during the identification experiment, the data preprocessing (Appendix A) reduces to mean (offset) correction and scaling of the data to unit variance. This approach guarantees a maximum achievable signal-to-noise ratio. A high signal-to-noise ratio is preferred for set estimation to obtain reasonable parameter uncertainties. For delay estimation, the cross-correlations between all inputs and outputs are depicted in Fig. 6.15 (see Section A.3). The delay matrix (Eq. A.12) determined from Fig. 6.15 is given by :

$$M_{\text{delay}} = \begin{bmatrix} 1 & 2 & 2 \\ 1 & 1 & 1 \end{bmatrix} \quad (6.10)$$

As expected for a physical process, at least one sample delay is present in each transfer function which will be corrected for identification. Note however, that the additional delay in the first output cannot be corrected and will therefore be included in the model.



**Fig. 6.15** : Delay estimation for the distillation column simulation process ; Dashed lines indicate the 95% confidence intervals and the dash-dotted lines define the estimated delay in each transfer.

For the classical prediction error identification approach the realization of a white prediction error implies in practice that a noise model must be included. For set estimation a linear constraint approximation is preferred to guarantee a convex and connected parameter uncertainty set  $\Theta$  (see Section 3.1) which simplifies the solution of

the identification problem. Therefore, taking these considerations into account, an ARMAX model has been adopted :

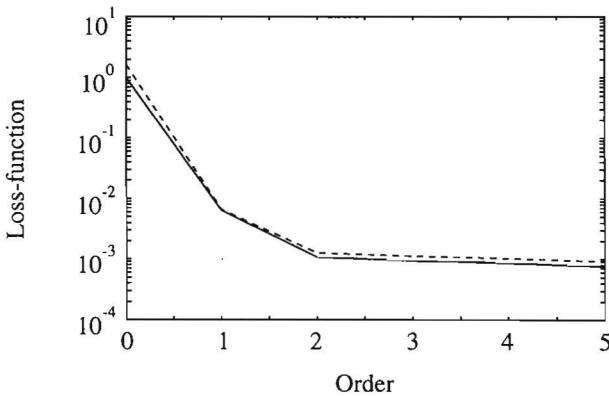
$$A(z^{-1}) \underline{y}(k) = B(z^{-1}) \underline{u}(k) + C(z^{-1}) \underline{\xi}(k) \quad (6.11)$$

where  $\underline{u}$  and  $\underline{y}$  denote the inputs and outputs with entries as indicated in Table 6.4. The structure selection has been reduced to an order selection by assuming proper process and noise transfer functions. Taking now the delay matrix of Eq. 6.10 into account, the model structure (Eq. B.4) for order  $mo$  of the ARMAX model is defined by :

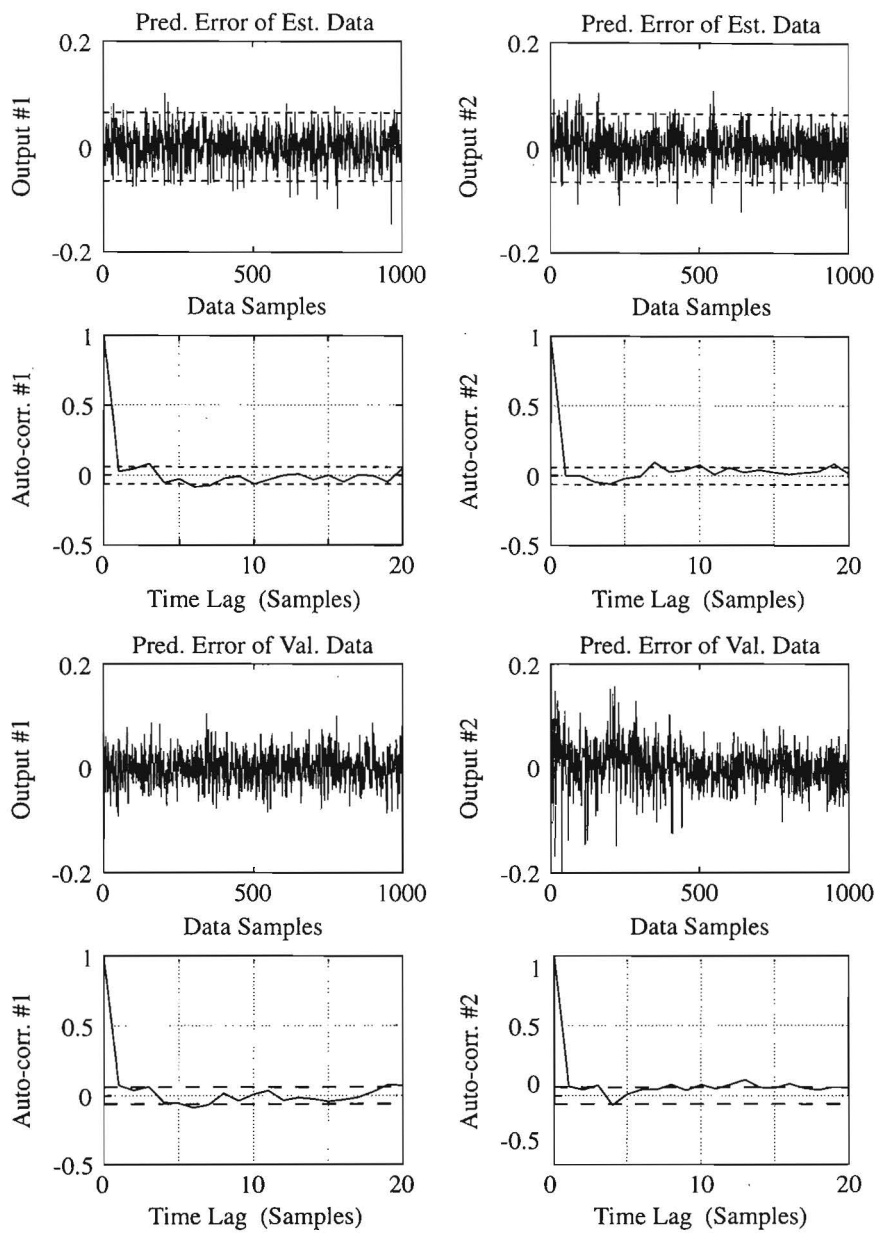
$$\begin{aligned} nn &= [ na \mid nb_{11} \ nb_{12} \ nb_{13} \ nb_{21} \ nb_{22} \ nb_{23} \mid nc_{11} \ nc_{22} \mid \\ &\quad nd \mid nf \mid nk_{11} \ nk_{12} \ nk_{13} \ nk_{21} \ nk_{22} \ nk_{23} ] \\ &= [ mo \mid mo+1 \ mo \ mo \ mo+1 \ mo+1 \ mo+1 \mid mo \ mo \mid \\ &\quad 0 \mid 0 \mid 0 \ 1 \ 1 \ 0 \ 0 \ 0 ] \end{aligned} \quad (6.12)$$

### 6.3.2 Parametric Identification

The total data set has been divided into an estimation and a validation data set of  $N = 1042$  data samples each. To select a correct model order, prediction error identification has been applied for increasing model order  $mo$ . The value of the loss-function (Eq. 2.24) for both data sets is depicted in Fig. 6.16. This figure indicates clearly a 2<sup>nd</sup> order model.



**Fig. 6.16** : Order selection of ARMAX model ; Loss-function of estimation (solid) and validation (dashed) data set.



**Fig. 6.17 :** Prediction error of the 2<sup>nd</sup> order ARMAX model for both estimation and validation data set. The auto-correlations of the prediction errors indicate the accuracy of the estimated model.

The prediction errors of this estimated 2<sup>nd</sup> order ARMAX model for both the estimation and the validation data are depicted in Fig. 6.17. The auto-correlations of these prediction errors show that the residual is almost white. Increasing the model order will not improve the whiteness of the residuals significantly. Therefore this model describes the distillation column simulation process in the selected operating point sufficiently accurate. The signal-to-noise ratio which has been realized for both outputs is approximately 30 dB.

This prediction error identification has been performed to obtain the a priori knowledge which is required for set estimation. The prediction errors depicted in Fig. 6.17 provide information about the upper bounds of the noise (Eq. 3.1) which have been defined as  $2\sigma$  bounds resulting in  $\hat{\delta}_e = [6.51e-2, 6.53e-2]^T$ . The prediction error estimate  $\hat{\theta}$  which can be used as reference model to select the constraints which violate the defined upper bounds of the noise, is shown in Table 6.5 together with the corresponding stochastic parameter uncertainties which have been computed from the covariance matrix of the estimate  $\text{Cov}(\hat{\theta})$  (see Section C.1). It has been shown in Section 3.4 that for  $N = 1000$  data samples, a  $2\sigma$  bound which corresponds to a confidence of 95% gives accurate uncertainty bounds. These results show that there are several parameters with more than 100% uncertainty which have almost no influence on the input/output behaviour.

It can be expected that the deterministic uncertainty bounds are equal but probably larger than the stochastic uncertainty bounds due to the conservatism which is introduced by the way the bounded error identification problem is solved (see Chapter 5). Therefore, the parameters with more than 100% uncertainty ( $b_{22,3}$ ,  $c_{11,2}$ ,  $c_{22,1}$  and  $c_{22,2}$ ) have been fixed to the value of the prediction error estimate because these parameters cannot be estimated accurately by set estimation. The constraints for bounded error identification can be constructed according to Eq. 3.14 for the defined upper bounds of the prediction error  $\hat{\delta}_e$  and the parameters  $b_{22,3}$ ,  $c_{11,2}$ ,  $c_{22,1}$  and  $c_{22,2}$  fixed. Under the assumption that the prediction error estimate can be used as reference model which defines an element of the parameter uncertainty set  $\Theta$ , 3.2% of the constraints violate the reference model where 4.4% have been eliminated for bounded error identification (see Section 3.3). More constraints have been eliminated because the possible data outliers in the violating constraints are also included in other constraints which should therefore be eliminated as well.

The identification results of ellipsoid-aligned orthotopic parameter bounding (EOPB<sup>1</sup>) are given in Table 6.5. This approach has been selected because as shown in the previous SISO case studies, the ellipsoid-aligned orthotopic parameter bounding provides the most accurate description of the parameter uncertainty set  $\Theta$ . It should be noted that the parameter uncertainty corresponds to the ellipsoid-aligned coordinate system while for comparison the central estimate  $\hat{\theta}_c$  has been transformed back to the original coordinate system. This predictor is even unstable because of the large parameter uncertainties. Whenever this uncertainty is close to 100% the corresponding

central estimate is not reliable. For example, the parameter  $c_{11,1}$  differs significantly from the prediction error estimate and makes therefore the predictor unstable.

**Table 6.5** : Parametric uncertainty in ARMAX model of the distillation column. The stochastic parameter uncertainty of the prediction error (PE) estimate corresponds to a  $2\sigma$  bound (95% confidence). The deterministic parameter uncertainty has been computed by ellipsoid-aligned orthotopic parameter bounding (EOPB).

Par.	PE		EOPB <sup>1</sup>		EOPB <sup>2</sup>		EOPB <sup>3</sup>	
	$\hat{\theta}$	%	$\theta_c$	%	$\theta_c$	%	$\theta_c$	% <sup>*)</sup>
$a_1$	-1.72e+0	2.7	-1.72e+0	(0.3)*	-1.73e+0	(0.7)*	-1.73e+0	(0.8)*
$a_2$	7.37e-1	6.3	7.40e-1	(0.7)*	7.43e-1	(0.2)*	7.42e-1	(0.2)*
$b_{11,1}$	7.94e-3	73.7	1.77e-2	(1.6)*	8.14e-3	(0.8)*	7.80e-3	(2.2)*
$b_{11,2}$	6.28e-2	10.7	5.69e-2	(0.7)*	6.25e-2	(1.6)*	6.24e-2	(1.2)*
$b_{11,3}$	3.18e-2	19.2	5.89e-2	(10.8)*	3.09e-2	(1.8)*	3.12e-2	(17.5)*
$b_{12,2}$	-2.33e-2	27.6	-2.69e-2	(4.5)*	-2.28e-2	(1.6)*	-2.27e-2	(1.8)*
$b_{12,3}$	-2.75e-2	24.1	-5.81e-2	(8.1)*	-2.78e-2	(>100)*	<b>-2.75e-2</b>	x
$b_{13,2}$	1.03e-2	56.4	2.24e-2	(5.5)*	1.06e-2	(2.3)*	1.08e-2	(2.3)*
$b_{13,3}$	-2.23e-2	31.5	-4.07e-2	(4.1)*	-2.26e-2	(1.4)*	-2.26e-2	(2.7)*
$b_{21,1}$	-1.69e-2	34.5	-1.69e-2	(3.9)*	-1.76e-2	(1.3)*	-1.76e-2	(0.9)*
$b_{21,2}$	-2.93e-2	20.5	-2.93e-2	(>100)*	-2.97e-2	(2.3)*	-2.95e-2	(2.1)*
$b_{21,3}$	1.42e-2	44.0	1.54e-2	(>100)*	1.57e-2	(22.6)*	1.57e-2	(17.1)*
$b_{22,1}$	8.69e-3	67.3	9.03e-3	(>100)*	8.76e-3	(2.5)*	8.64e-3	(2.4)*
$b_{22,2}$	3.11e-2	18.9	3.05e-2	(>100)*	2.97e-2	(4.6)*	2.97e-2	(2.4)*
$b_{22,3}$	8.39e-4	>100	<b>8.39e-4</b>	x	<b>8.39e-4</b>	x	<b>8.39e-4</b>	x
$b_{23,1}$	-6.68e-2	8.7	-6.58e-2	(>100)*	-6.61e-2	(4.0)*	-6.61e-2	(4.8)*
$b_{23,2}$	-5.34e-2	14.3	-5.36e-2	(>100)*	-5.32e-2	(7.8)*	-5.34e-2	(4.4)*
$b_{23,3}$	6.12e-2	13.6	6.20e-2	(>100)*	6.23e-2	(0.7)*	6.24e-2	(0.7)*
$c_{11,1}$	5.29e-1	42.2	2.66e+1	(99.0)*	<b>5.29e-2</b>	x	<b>5.29e-1</b>	x
$c_{11,2}$	2.39e-1	>100	<b>2.39e-1</b>	x	<b>2.39e-1</b>	x	<b>2.39e-1</b>	x
$c_{22,1}$	1.21e-2	>100	<b>1.21e-2</b>	x	<b>1.21e-2</b>	x	<b>1.21e-2</b>	x
$c_{22,2}$	1.72e-2	>100	<b>1.72e-2</b>	x	<b>1.72e-2</b>	x	<b>1.72e-2</b>	x

1) Parameters  $b_{22,3}$ ,  $c_{11,2}$ ,  $c_{22,1}$  and  $c_{22,2}$  have been fixed to their prediction error estimate.

2) Also parameter  $c_{11,1}$  has been fixed.

3) Next parameter  $b_{12,3}$  has been fixed.

\*) Parameter uncertainty in ellipsoid-aligned coordinate system.



To reduce these large parameter uncertainties, an additional parameter will be fixed as described in Chapter 5. The order in which the parameters will be fixed can be computed from the summed parameter correlations (Eq. 5.3). If accordingly the parameter  $c_{11,1}$  is fixed as well, 6.6% of the constraints violate the reference model. Note that because all c-parameters have been fixed, the bounded error identification of the ARMAX model reduces in fact to a pure equation-error problem. The identification results of the ellipsoid-aligned orthotopic parameter bounding (EOPB<sup>2</sup>) are given in Table 6.5. The parameter uncertainty of EOPB<sup>2</sup> compared to EOPB<sup>1</sup> has been reduced significantly by fixing this additional parameter  $c_{11,1}$ . Only the parameter  $b_{12,3}$  has still an uncertainty of more than 100%. If this parameter is also fixed, the estimates of EOPB<sup>3</sup> are obtained. The parameter uncertainty however reduces hardly, and the different uncertainties are mainly caused by a different orientation of the ellipsoid-aligned coordinate system.

To illustrate the accuracy of the estimated models, the magnitude plots of the transfers are compared with the spectra of the input/output data in Fig. 6.18. Because of the large similarity of the models, only the magnitude plots of the prediction error estimate are shown. The differences in the low frequency range is caused by the ARMAX model structure selection which emphasizes the higher frequency range.

For comparison of the different models, the standard deviations of the prediction error for each output and for both the estimation and the validation data set have been computed (Table 6.6) using the prediction error estimate  $\hat{\theta}$  and the central estimates  $\theta_c$  obtained by bounded error identification. This quality measure indicates that all models, except the estimate EOPB<sup>1</sup>, are almost the same.

**Table 6.6** : Comparison of prediction error and bounded error estimates.

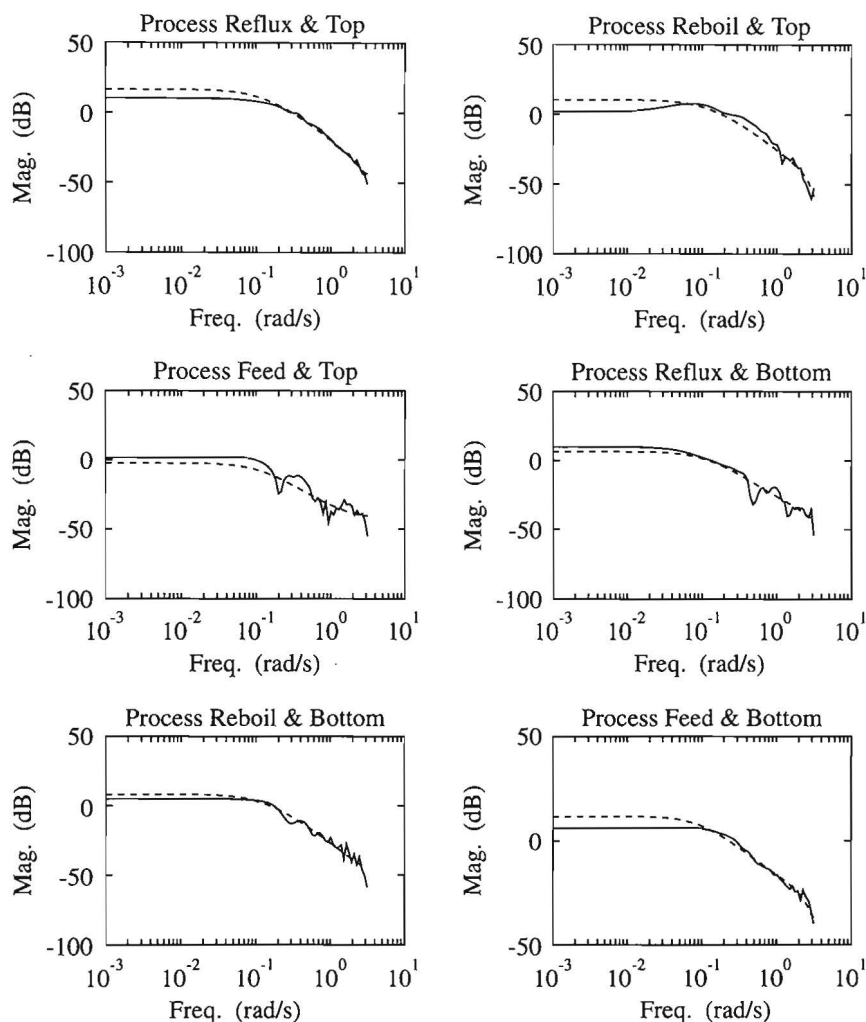
Type	Estimation		Validation	
	Output #1	Output #2	Output #1	Output #2
PE	3.256e-2	3.265e-2	3.095e-2	3.936e-2
EOPB <sup>1</sup>	5.965e-2*	3.267e-2	5.919e-2*	3.939e-2
EOPB <sup>2</sup>	3.242e-2	3.291e-2	3.082e-2	3.977e-2
EOPB <sup>3</sup>	3.240e-2	3.293e-2	3.078e-2	3.978e-2

1) Parameters  $b_{22,3}$ ,  $c_{11,2}$ ,  $c_{22,1}$  and  $c_{22,2}$  have been fixed to their prediction error estimate.

2) Also parameter  $c_{11,1}$  has been fixed.

3) Next parameter  $b_{12,3}$  has been fixed.

\*) Polynomial  $C_{11}(z^{-1})$  has been stabilized by mirroring the unstable roots inside the unit circle which causes an increased error function.



**Fig. 6.18** : Magnitude plots of transfer functions in distillation column ; Spectra of input/output data (solid) and process transfer functions of the prediction error estimate (dashed).

From a computational point of view, however, the prediction error identification approach is preferred because it is less time consuming and more accurate. The corresponding stochastic parameter uncertainty provides a reliable description of the

parameter uncertainty set  $\Theta$ . Solving the ellipsoid-aligned orthotopic parameter bounding problem is computationally time consuming because of the repeated constraint processing. Especially for increasing number of parameters, more iterations are required before the ellipsoid converges to its minimum volume. The main problem, however, is that set estimation is very sensitive to overparametrization, i.e. including parameters which hardly influence the input/output behaviour. Repeated ellipsoid-aligned orthotopic parameter bounding is required where the most uncertain parameters will be fixed, in order to obtain reasonable central estimates  $\theta_c$ .

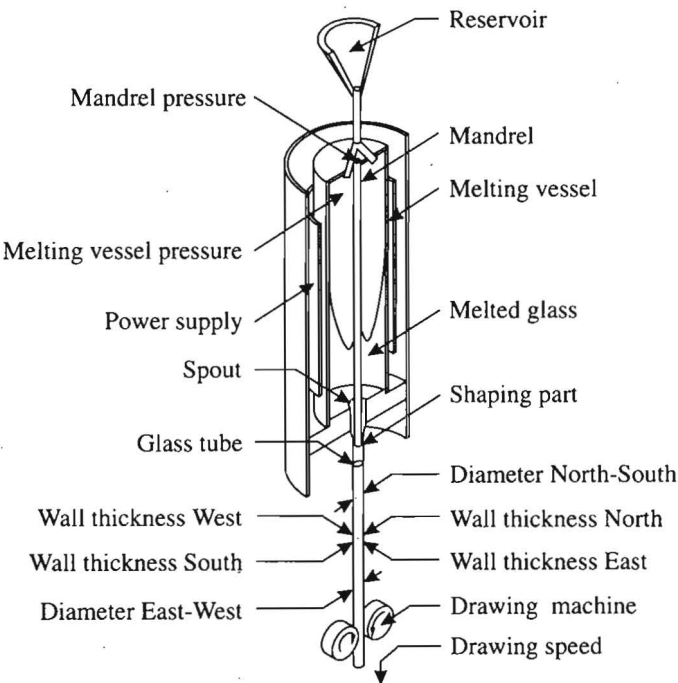
Although a signal-to-noise ratio of 30 dB has been studied, which is large compared to an industrial production environment, this simulation example of a distillation column shows clearly the bottlenecks of the bounded error identification approach :

- Data outliers which violate the defined upper bounds of the errors.
- Sensitivity to overparametrization.

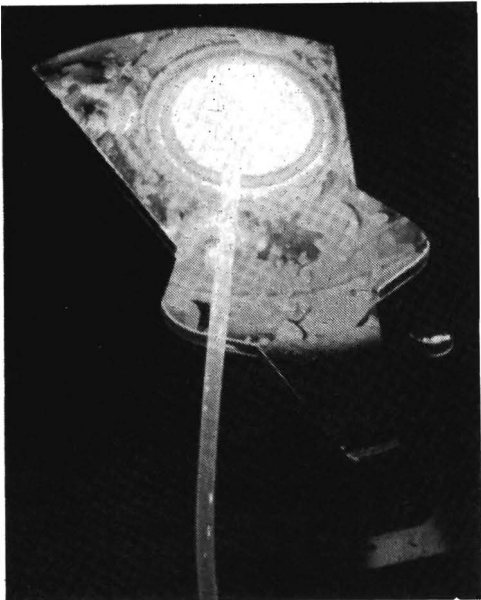
## 6.4 A Glass Tube Production Process

The identification of an industrial multivariable manufacturing process will be considered which can be approximated around an operating point as being linear and time-invariant. Because the signal-to-noise ratio which can be realized for this process is limited, only the prediction error identification approach will be described in this section. This example will illustrate that the prediction error identification algorithms adopting a minimal polynomial model structure can be applied to industrial processes.

Figure 6.19 depicts a schematic outline of the most important parts of the tube glass production process. The raw material pours down through a feeder into the furnace vessel. This vessel is heated electromagnetically which yields sintering at the transition between the sand and the melted glass. At the bottom of the furnace the melted glass pours out through a spout where a mandrel is accurately positioned. Shaping of the tube takes place at and just below the mandrel. The shape of the tube is determined by two output variables : the wall thickness ( $W$ ) and the diameter ( $D$ ) as function of time. In the production process, 4 wall thickness (North, South, East and West) and 2 diameter (North/South and East/West) sensors measure the tube quality. The two process parameters that can be influenced most easily and affect the shape of the tube most directly with the shortest time delay over the largest frequency range, are the mandrel pressure ( $MP$ ), i.e. the pressure of the gas led into the mandrel at the top, and the drawing speed ( $DS$ ) [Bac87]. Increasing the mandrel pressure results in an increase of the diameter and, simultaneously, in a decrease of the wall thickness, while increasing the drawing speed results in a decrease of both the diameter and the wall thickness.



**Fig. 6.19 :** The glass tube production process.



**Fig. 6.20 :** Bottom view of a glass furnace.

### 6.4.1 Data Preprocessing

For model identification a pseudo random binary noise experiment has been carried out resulting in a data set of 18000 samples where the two inputs have been excited simultaneously. An oversampling ratio of 10 is chosen for proper data preprocessing. The 4 wall thickness and 2 diameter measurements must be preprocessed to construct an average wall thickness and diameter signal.

A small trend present in the measured output signals corresponds to very slowly varying process behaviour which is not the objective in the identification of a dynamic model and has therefore been removed from the signals (see for example Fig. A.1). Further, peaks in the wall thickness measurements due to sensor failures which have been detected are eliminated by peakshaving (see for example Fig. A.2). An average wall thickness signal can be constructed by simply adding the 4 measurements since all wall thickness sensors are installed at the same height (Fig. 6.19) and therefore no relative delay exists between the 4 signals. Since the diameter sensors are installed at different heights, the relative delay between the two diameter measurements has to be derived first before an average diameter signal can be constructed. The time lag which corresponds to the maximum value in the cross-correlation between the North/South and East/West diameter signals defines this relative delay.

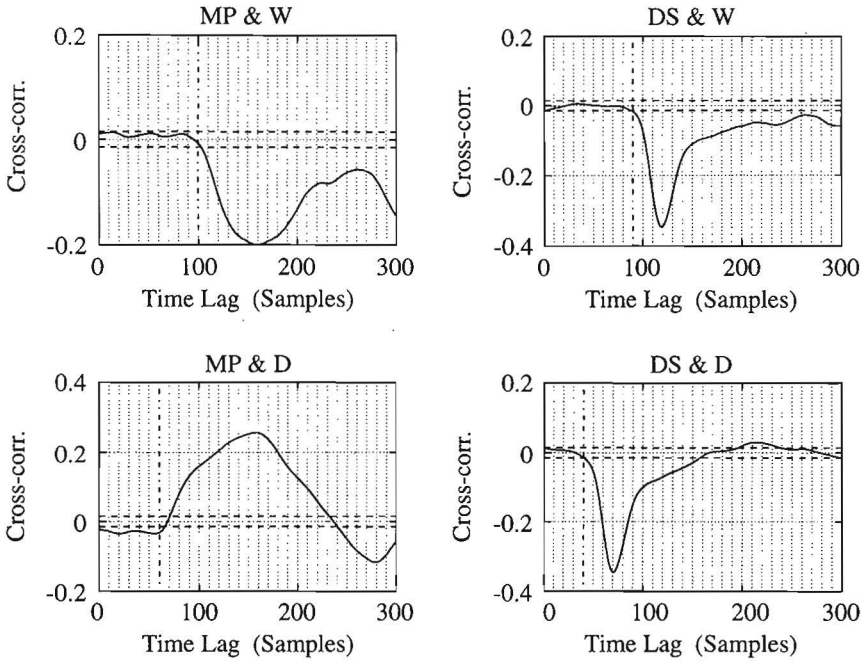
The next step in the data preprocessing after constructing average wall thickness and diameter signals, is the estimation of the process delays. To determine these delays, the cross-correlations between all inputs and outputs have been computed which are depicted in Fig. 6.21. To make the delay estimation easier, a vertical line at every 10<sup>th</sup> sample is plotted which corresponds to the defined oversampling ratio. The process delays which have been determined from Fig. 6.21 are given in Table 6.7.

**Table 6.7** : Process delays.

	MP	DS
W	100	90
D	60	40

Since in a 2-input 2-output system only 3 degrees of freedom are available, it cannot be guaranteed that the defined delay matrix can be realized. For this problem, the realizable delay (Eq. A.14) matrix is :

$$M_{\text{delay}}^r = \begin{bmatrix} 100 & 90 \\ 50 & 40 \end{bmatrix} \quad (6.13)$$

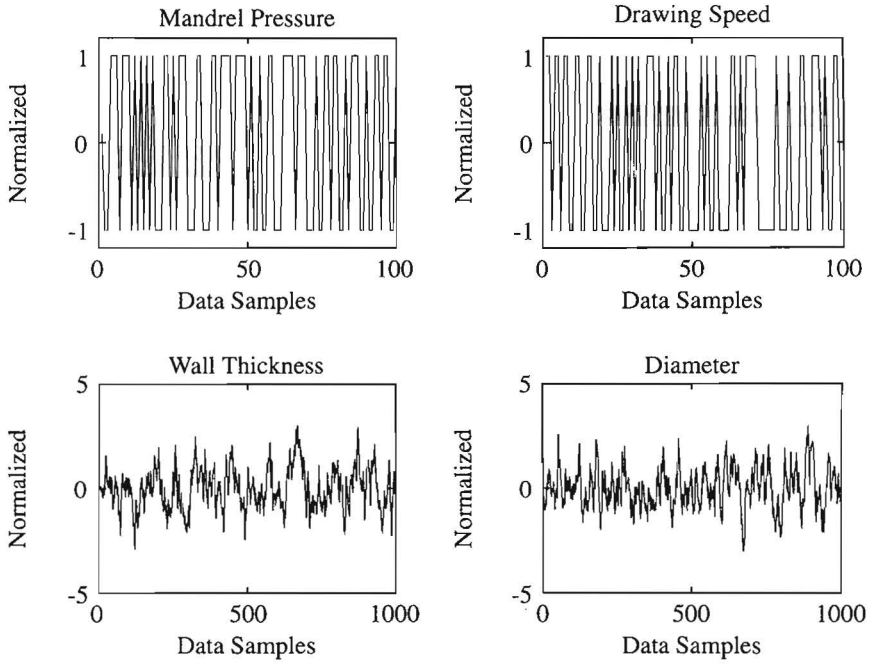


**Fig. 6.21** : Estimating process delays where the dashed lines indicate the 95% confidence intervals and the dash-dotted lines indicate the defined process delays.

This implies that with an oversampling ratio of 10, 1 sample delay from the mandrel pressure to the diameter cannot be corrected. Finally, for model identification the first 1000 samples have been selected for the estimation data set and the remaining data samples for the model validation. The input/output estimation data has been scaled to unit standard deviation and is depicted in Fig. 6.22.

### 6.4.2 Model Estimation

For minimal polynomial model identification it is required that the input/output data is scaled in such a way that the white noise sequences disturbing the process have equal variance. Otherwise, a relative weighting is introduced during the prediction error minimization. Of course, these variances are not known a priori in practice. An accurate estimate however, can be obtained by first estimating a high-order equation-error model which can approximate any linear system arbitrary well for sufficiently high-order.



**Fig. 6.22** : Estimation data.

The standard deviation of the prediction error computed from a 50<sup>th</sup> order ARX-model,  $\sigma_{pe} = [1.19e-1 \ 7.43e-2]$ , can then be used to rescale the output data (Eq. A.18) :

$$S_y = \begin{bmatrix} 6.24e-1 & 0 \\ 0 & 1 \end{bmatrix} \quad (6.14)$$

This scaling matrix indicates that the noise disturbing the wall thickness measurements is larger compared to the diameter measurements.

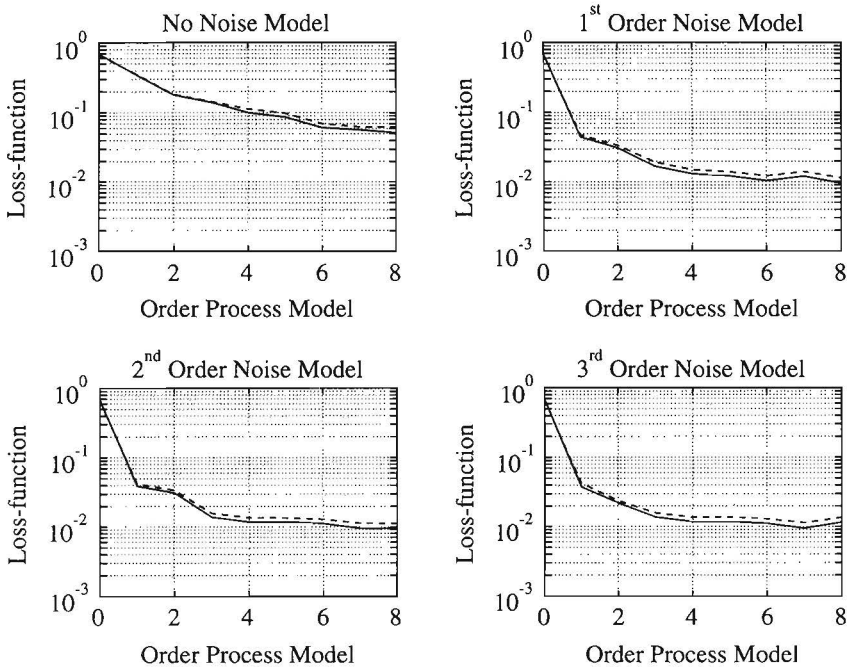
To obtain a model which is suitable for controller design and to obtain a white prediction error, a Box-Jenkins model structure has been adopted :

$$\underline{y}(k) = \frac{B(z^{-1})}{F(z^{-1})} \underline{u}(k) + \frac{C(z^{-1})}{D(z^{-1})} \underline{\xi}(k) \quad (6.15)$$

by selecting the polynomials  $B(z^{-1})$ ,  $F(z^{-1})$  and  $C(z^{-1})$ ,  $D(z^{-1})$  in the general model

definition. The 1 sample delay which could not be corrected in the data preprocessing phase, will be included in the model. Because the order of the process and the noise dynamics are not known, the identification has been performed for several orders of the process (po) and the noise part (no), resulting in the following model structure :

$$\begin{aligned} nn &= [ na \mid nb_{11} \ nb_{12} \ nb_{21} \ nb_{22} \mid nc_{11} \ nc_{22} \mid nd \mid nf \mid nk_{11} \ nk_{12} \ nk_{21} \ nk_{22} ] \\ &= [ 0 \mid po+1 \ po+1 \ po \ po+1 \mid no \ no \mid no \mid no \mid 0 \ 0 \ 1 \ 0 ] \end{aligned} \quad (6.16)$$



**Fig. 6.23** : Process and noise order selection ; Loss-function of estimation data set (solid) and validation data set (dashed).

The loss-functions of the estimated models for the estimation and the validation data sets are depicted in Fig. 6.23 where the process order ranges from 0 to 8 and the noise order from 0 to 3. Theoretically, the loss-function  $V_e(\theta)$  decreases continuously for increasing model order. However, due to overparametrization, the optimization might not converge correctly resulting in a larger value of the loss-function compared to a model of lower order.



From Fig. 6.23 it follows that the loss-function for the output-error models, i.e. no noise model, is significantly larger compared to the estimates including a noise model. On the other hand, a 3<sup>rd</sup> order noise model does not result in a significant lower value of the loss-function. By selecting a 1<sup>st</sup> order noise model, the process order would be 6 ( $\theta_{61}$ ), while selecting a 2<sup>nd</sup> order noise model would indicate a 4<sup>th</sup> order process model ( $\theta_{42}$ ) when comparing just the values of the loss-function. Because it is difficult to decide which model should be selected considering the values of the loss-function and the number of parameters only, the auto-correlation of the prediction error for every output has been computed. The results are depicted in Fig. 6.24. It is obvious from these figures, that the prediction error of the estimate  $\theta_{61}$  with the 1<sup>st</sup> order noise model is still clearly correlated, while the objective was to obtain a white prediction error. This will never be completely possible in practice, but the prediction error of the estimate  $\theta_{42}$  with the 2<sup>nd</sup> order noise model is almost completely within the confidence intervals. For this reason, the combination of a 4<sup>th</sup> order process model and a 2<sup>nd</sup> order noise model has been selected.

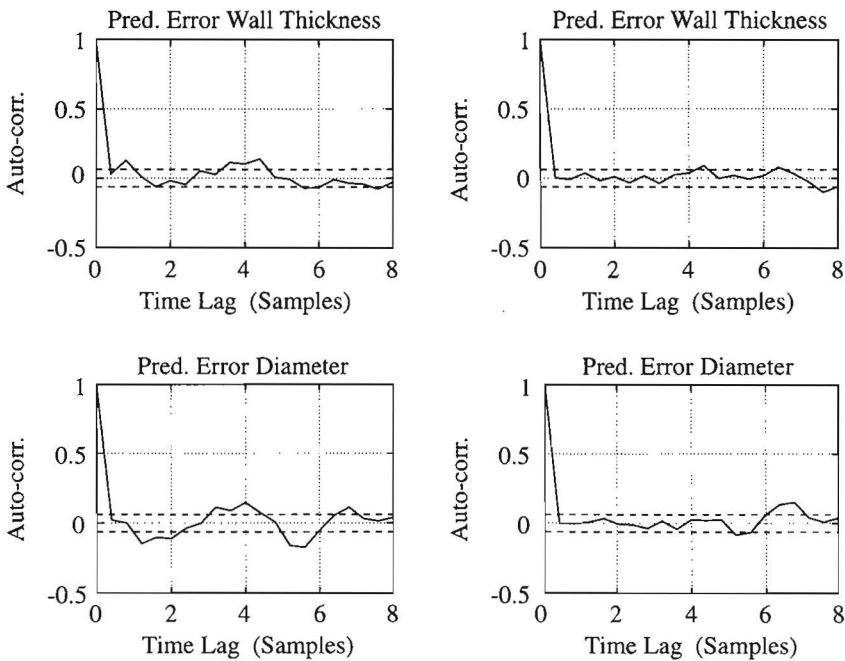


Fig. 6.24 : Auto-correlations of the prediction errors ;  $\theta_{61}$  (left) and  $\theta_{42}$  (right).

To realize a correct identification criterion in the prediction error method, the estimated prediction error should be white and of equal variance for each output. If the scaling is not correct, an implicit weighting is introduced during the optimization. The whiteness of the prediction error is an indication of the remaining dynamical information which has not been estimated. In practice, however, a complete white prediction error can never be realized because the model will always be a low order process approximation. Applying an AR-model to the prediction errors can be used to validate the scaling of the input/output data (see Section C.1, Eq. C.21) and provides information about the whiteness of the prediction error. The variances of the prediction error ( $\sigma_{pe}^2$ ) and the whitened prediction error ( $\sigma_{\xi}^2$ ) for both outputs are given in Table 6.8. For both outputs the variance is approximately of the same size which shows that the scaling of the data based on the prediction error estimate of an high-order ARX-model is a fast way to scale the data properly a priori.

**Table 6.8** : Validation of the data scaling.

Variance	$\sigma_{ARX}^2$	$\sigma_{pe}^2$	$\sigma_{\xi}^2$
Wall thickness	1.42e-2*	1.10e-2	1.05e-2
Diameter	5.52e-3	1.25e-2	1.15e-2

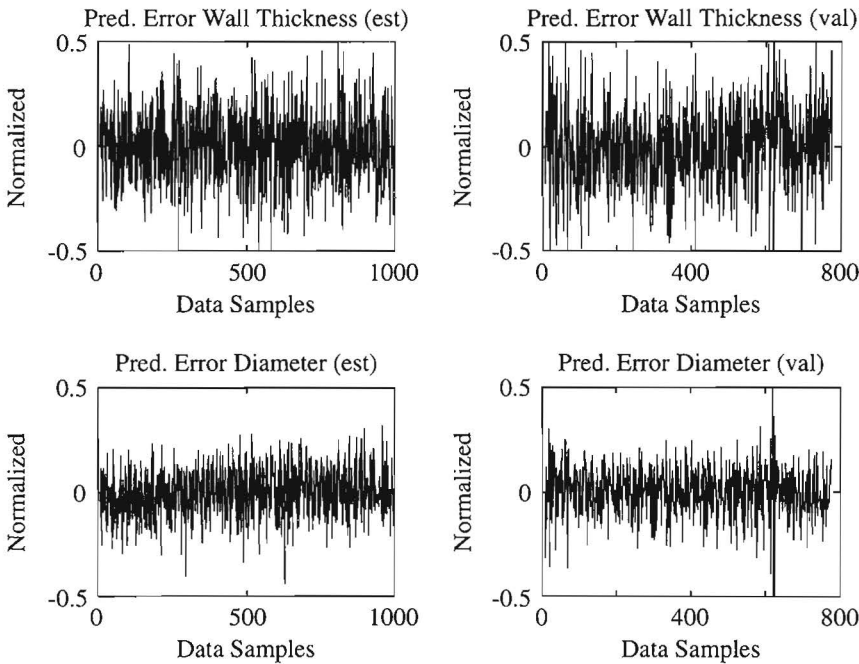
\*) Unscaled data. If the scaling matrix Eq. 6.14 is included, this value reduces to the diameter variance.

For model validation, the prediction errors for both data sets, estimation and validation respectively, are depicted in Fig. 6.25. Note that these prediction errors have been scaled to a normalized value by dividing by the standard deviation of the corresponding output. The standard deviations of these normalized prediction errors for both the estimation ( $\sigma_e$ ) and the validation ( $\sigma_v$ ) data set are given in Table 6.9.

**Table 6.9** : Model validation.

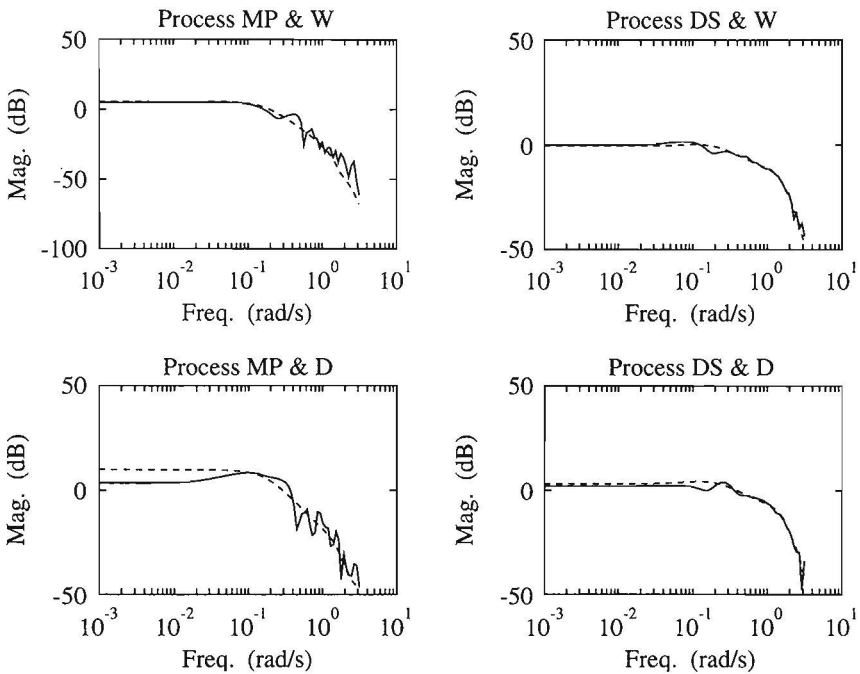
Standard deviation	$\sigma_e$	$\sigma_v$
Wall thickness	1.68e-1	2.01e-1
Diameter	1.12e-1	1.21e-1

This validation of the estimate  $\theta_{42}$  using an independent data set shows clearly that an accurate model of the glass tube production process has been obtained in the defined operating point. The signal-to-noise ratios which have been realized with this model for the wall thickness and diameter outputs are 15.5 dB and 19.0 dB respectively. These rather low signal-to-noise ratios make this example not suitable for bounded error identification.



**Fig. 6.25** : Prediction error for estimation (left) and validation (right) data set.

A further indication of the model accuracy is shown in Fig. 6.26. In this figure the magnitude plots of the process model are compared with the estimated spectra of the input/output data for each transfer. The 4<sup>th</sup> order process model can describe the dynamical behaviour of the tube glass production process accurately. The remaining misfit can only be reduced if a model of significantly higher order is estimated. However, according to Fig. 6.23, the quality of the model in terms of prediction errors will hardly improve.



**Fig. 6.26** : Magnitude plots of transfer functions in the glass tube production process ; Spectra of input/output data (solid) and process transfer functions of prediction error estimate (dashed).

These case studies show clearly that the prediction error identification approach using a minimal polynomial model structure provides fast and very efficient identification algorithms to model multivariable processes accurately. This approach definitely outperforms existing multivariable identification procedures using (pseudo-) canonical model structures. A library of identification and data preprocessing routines is one of the deliverables of this thesis work.

To apply bounded error identification methods to multivariable processes, an extremely high quality of the data is required which cannot always be realized in an industrial environment. In addition, this bounded error identification method is more sensitive with respect to violating theoretical assumptions than the prediction error identification approach. Therefore, depending on the character of the noise, prediction error identification is preferred to model industrial processes because of its robustness and relative low computational complexity compared to bounded error identification.

## ***Conclusions and Remarks***

---

In this thesis procedures have been described for the identification of multivariable linear time-invariant systems in the time domain which yield a nominal model with uncertainty bounds on its parameters. This parametric uncertainty can either be stochastic or deterministic depending on the assumptions which have been made with respect to the noise which disturbs the system, i.e. a statistical or a bounded error description. The analyses made can be summarized in the following statements :

- The extension of SISO process descriptions in prediction error identification to multivariable systems using (pseudo-) canonical model structures have been studied extensively in the literature [Lju87, SS89] but are of limited use in practice. This limitation is caused by the fact that structural indices, which are required for these descriptions, are hard to estimate for industrial processes. Adopting a minimal polynomial model structure [Bac87] with an extension to the general family of polynomial model structures (Eq. 2.17), avoids the problem of structural indices and is therefore more suitable for modelling multivariable industrial processes (Chapter 2).
- If the data is properly scaled, the minimal polynomial model structure for multivariable systems shows interesting similarities with the SISO case (e.g.

Eq. 2.24). The first- and second-order derivatives which are required for solving the parameter estimation problem by non-linear least-squares optimization, can be computed analytically by simply filtering data sequences (Appendix B). This results in fast and very efficient identification algorithms. A library of identification and data preprocessing routines is one of the deliverables of this thesis work.

- In order to provide a quality measure of the model (Cramér-Rao bound), it is a fundamental requirement in current prediction error identification that the process is (asymptotically) contained in the model set. This requirement has been relieved in the present work. If the process is not contained in the model set, which is always the case in practice, an estimation of the parameter covariance matrix has been derived which can be used to compute stochastic uncertainty bounds on the parameters (Appendix C).
- The basic idea of bounded error identification is that the process behaviour can be described by a model from a given model set where the data set does not conflict the bounded error assumption (Section 3.2). In practice, however, the interplay between model structure (with linearized constraints in the parameter space), error distribution, signal-to-noise ratio, length of the data set and data outliers, requires proper precautions like outlier correction and/or fixing specific parameters (Section 3.3 and Chapter 5). This, to avoid an empty or unbounded parameter uncertainty set.
- The least conservative approximation of the parameter uncertainty set in bounded error identification can be obtained by ellipsoid-aligned orthotopic bounding. This approach combines the features of ellipsoidal and orthotopic bounding and treats the parameter space as a coordinate free Euclidean space.
- In identification, sensitivity to overparametrization can be reduced by fixing those parameter values which have only a small contribution to the input/output behaviour (Chapter 5). These values can be fixed to predefined values for example of a reference model. By fixing the parameters which are difficult to estimate, the uncertainty in the remaining parameters will be reduced. This approach may seem counter intuitive. Indeed, one may argue that it is logical to fix system parameters with small uncertainty and which have a relatively large contribution to the system behaviour. However, with the proposed method an error in a fixed parameter which has a large uncertainty and a small dynamic contribution has little influence on the set of identified models. These parameters may be viewed intuitively as poorly observable parameters of the model set

(Section 5.3). This strategy can be applied both in deterministic model set estimation and in conventional stochastic prediction error identification. The criteria for bad observability then is respectively sign uncertainty in deterministic identification and standard deviation exceeding the absolute value of the expectation in the stochastic case.

- Parameter bounds which have been computed either by stochastic prediction or by bounded error identification can be considered as reliable if the uncertainty bounds obtained by both methods are of equal size. So both identification methods have to be applied to obtain a measure of the quality of the parameter uncertainty (Section 3.4).

From practical experience, the following conclusions can be drawn :

- The case studies (Chapter 6) have illustrated that from a computational point of view prediction error identification is less complex compared to bounded error identification. The latter approach requires repeated constraint processing in ellipsoid-aligned orthotopic bounding (Section 3.2) to approximate the parameter uncertainty set as accurate as possible. In addition, the prediction error identification approach is more robust with respect to violating the theoretical assumptions which is often the case in practice.
- In order to apply bounded error identification methods to multivariable industrial processes, an extremely high quality of the data is required (Section 6.3).
- The accuracy of a model which describes a process is determined by the quality and quantity of the data. Once the data has been preprocessed correctly, the application of the identification algorithms is straightforward and automated and will certainly provide accurate models for the given data set. However, the accuracy of this model with respect to the actual process will highly depend on the quality and the quantity of the data. Therefore, system access, experiment design, process excitation and the duration of the experiments in production processes will mainly determine the achievable model quality (Chapter 6).

Identification and robust control design based on a worst case approach will in practice not result in a closed-loop behaviour of acceptable performance. Treating all models in the model set equivalent, will make the control design very conservative. It is therefore required to include stochastic information in future research and to develop identification and robust control design methods which can handle the models in the model set with different probability. Consequently both the bounded error identification

and the stochastic prediction error estimation will then be indispensable to provide together the necessary information for control design.



## *Data Preprocessing*

---

A.1	Detrending	A.3	Delay Correction
A.2	Peakshaving	A.4	Data Decimation and Scaling

---

The quality (accuracy) of models obtained by identification techniques will in practice highly depend on the quality of the data which has been offered to the identification algorithms. Disturbances which will always be present in industrial process data range from noise and spikes to drifts and offsets. These disturbances will decrease the resulting model quality. Therefore, their influence should be diminished by dedicated signal processing techniques [Bac87].

For parameter identification techniques, the models describing the system dynamics should be parsimonious reducing the number of parameters as much as possible. For increasing number of parameters the bias of the estimate will reduce while the variance will increase approximately by  $n\theta/N$ . Overparametrization, however, decreases the model quality because the additional parameters will hardly improve the identification criterion while thereby the higher dimensional identification problem is numerically harder to solve. In particular, time delays in the measured process data should be compensated in advance as much as possible before identifying the dynamical system. If no correction is applied, many extra moving average parameters have to be estimated,

which deteriorate the model quality.

Finally, the scaling of the measured data is the last step in the data preprocessing which influences the model quality. In industrial practice not all inputs and outputs have the same order of magnitude. Additionally, to realize a correct identification criterion for prediction error minimization of multivariable systems, it is required that the variances of the prediction error for the several outputs are of equal size.

In this appendix, each step of the data preprocessing :

- *Detrending* to remove drifts and offsets,
- *Peakshaving* to eliminate data outliers,
- *Delay correction* to reduce the required number of parameters,
- *Data decimation and scaling* of input and output sequences to realize a correct identification criterion,

will be described briefly. For industrial practice in particular, all these steps have to be processed to obtain useful data for identification.

It is assumed that the process signals have been sampled at a frequency higher than the frequency actually needed for identification. This oversampling rate should be chosen 5 to 10 times higher than the bandwidth of the process. The redundancy (due to the oversampling) present in the test signals which have been defined for the identification experiments, can be used for proper preprocessing of the recorded data.

## A.1 Detrending

The purpose of identification is to describe the dynamical behaviour of the process accurately. In industrial practice, however, recorded process data often shows a low frequent drift which has been induced by external changes, e.g. temperature variations. This low frequent information present in the measured output data should be considered as a disturbance because no process dynamics are described and therefore this drift has to be eliminated. Detrending of the input signals consists only of offset correction since these signals are assumed to be exact, e.g. setpoint inputs.

A measured signal can be divided into a process signal with zero mean and a trend signal :

$$\underline{x} = \underline{x}_{\text{process}} + \underline{x}_{\text{trend}} \quad (\text{A.1})$$

To determine the trend signal, various types of detrending can be distinguished depending on the best description of the drift which should be selected by visualization of the data :

- *Mean* : the mean value is removed from the data signal. In fact no drift is present but only an offset.
- *Linear* : linear detrending removes the best straight-line fit from the data signal.
- *Filter* : the drift can be described best by a low frequent signal which has to be eliminated by filtering.

Note, that whenever spikes (data outliers) disturb the recorded data, e.g. sensor failures, data acquisition errors etc., the detrending will be influenced significantly by these outliers. Therefore, the signal should be clipped before applying any detrending to reduce the influence of these outliers. Formally, for any measured data value the clipping is defined by :

$$x_{\text{clip}} = \begin{cases} x_{\text{max}} & x > x_{\text{max}} \\ x & x_{\text{min}} \leq x \leq x_{\text{max}} \\ x_{\text{min}} & x < x_{\text{min}} \end{cases} \quad (\text{A.2})$$

where  $x_{\text{min}}$  and  $x_{\text{max}}$  indicate the lower and upper bound respectively.

Mean and linear detrending are standard available (MATLAB identification toolbox), but these types are often not sufficient for accurate detrending of low frequent drifts. To extract these low frequent drifts, the clipped process signal is filtered with a low-pass second order trend filter :

$$F_{\text{trend}}(z^{-1}) = \frac{k_F z^{-1}}{(1 - p z^{-1})^2} \quad (\text{A.3})$$

where the gain is defined by  $k_F = (1-p)^2$  realizing a unit transfer at frequency zero. The pole  $p$  of the filter can be derived from the trend period ( $T_{\text{trend}}$  in samples) defining the highest frequency which the trend signal will contain. This trend period has to be determined by hand from the visualized signal. Further, the damping of the filter has been set to 1 dB which corresponds to a damping ratio  $\zeta = 0.89$  in order to realize a detrending filter with no overshoot and oscillation in its step response. The pole  $p$  of the filter can then be computed from [HSS82] :

$$\begin{aligned}
 p &= \exp\left(-\zeta \omega T_{\text{sample}}^*\right) = \exp\left(\frac{-\zeta \omega_{\text{trend}} T_{\text{sample}}^*}{\sqrt{1-\zeta^2}}\right) \\
 &= \exp\left(\frac{-2\pi\zeta T_{\text{sample}}^*}{T_{\text{trend}}^* \sqrt{1-\zeta^2}}\right) = \exp\left(\frac{-2\pi\zeta}{T_{\text{trend}}^* \sqrt{1-\zeta^2}}\right)
 \end{aligned} \tag{A.4}$$

with  $\omega_{\text{trend}} = \omega \sqrt{1-\zeta^2}$  where  $\omega$  is the undamped natural frequency,  $T_{\text{sample}}^*$  the sampling time (seconds),  $\omega_{\text{trend}}$  the trend frequency and  $T_{\text{trend}}^*$  the trend period (seconds). Note that the superscript \* has been used to change the units of the time variables from samples to seconds.

Before filtering, the beginning and ending data sequences of the signal have to be deflected with matching slope to minimize startup and ending transients of the filter. The length of these data sequences can be derived from the settling time  $T_{\text{settling}}^*$  of a unit step response of the second order filter. The time to reach its steady state value is approximately given by :  $T_{\text{settling}}^* \approx 8/\zeta\omega$ . Therefore the required length of the data sequences to eliminate transient effects of the filter is given by :

$$\begin{aligned}
 T_{\text{settling}} &= \frac{8}{\zeta \omega T_{\text{sample}}^*} = \frac{8\sqrt{1-\zeta^2}}{\zeta \omega_{\text{trend}} T_{\text{sample}}^*} \\
 &= \frac{8 T_{\text{trend}}^* \sqrt{1-\zeta^2}}{2\pi\zeta T_{\text{sample}}^*} = \frac{8 T_{\text{trend}}^* \sqrt{1-\zeta^2}}{2\pi\zeta}
 \end{aligned} \tag{A.5}$$

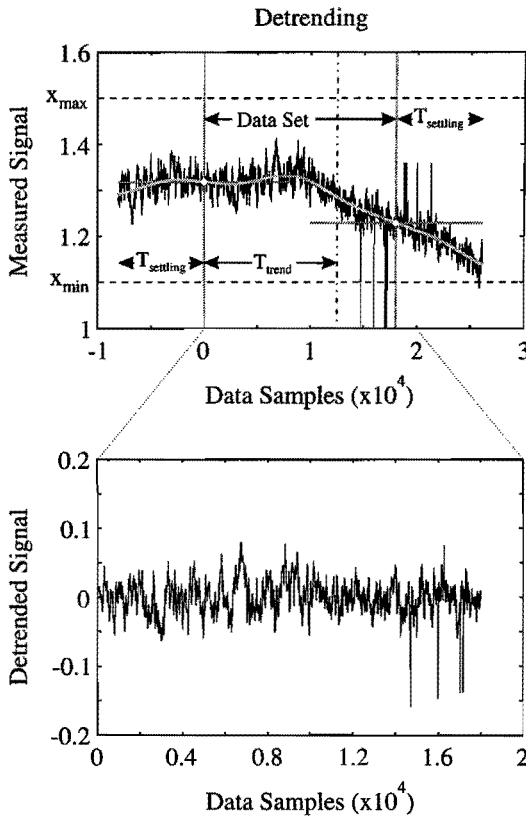
To avoid that the connections of the deflected beginning and ending data sequences is based on extreme values or outliers, a mean connection value is computed taking  $T_{\text{settling}}/10$  samples into account (see Fig. A.1). These mean connection values of the beginning and ending data, are added to the original data sequence and used to position fictitious x-y axes. A matching slope can then be realized by mirroring the data selected by  $T_{\text{settling}}$  around the origins of these fictitious x-y axes. Further, to realize zero-phase distortion, the data signal has to be filtered causal and anti-causal with matching initial conditions. When deriving in this way the trend signal :

$$x_{\text{trend}}(k) = F_{\text{trend}}(z) F_{\text{trend}}(z^{-1}) x_{\text{clip}}(k) \tag{A.6}$$

the detrended process signal can simply be obtained by subtracting the trend signal from the clipped signal :

$$\underline{x}_{\text{process}} = \underline{x}_{\text{clip}} - \underline{x}_{\text{trend}} \quad (\text{A.7})$$

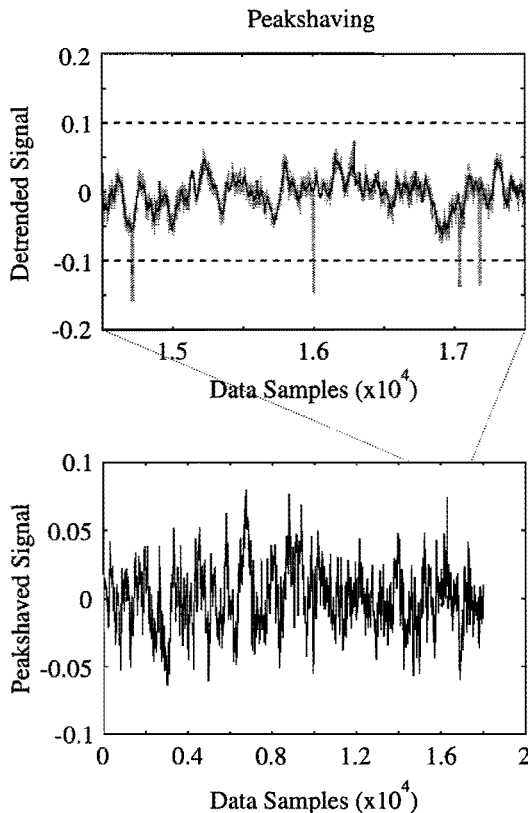
The trend period  $T_{\text{trend}}$  should be chosen carefully and as low as possible because all frequencies below this trend frequency will be filtered out and hence reducing the frequency contents of the signals available for identification. An example of detrending where clearly filtering is required is depicted in Fig. A.1.



**Fig. A.1 :** Detrending and clipping in data preprocessing ; a) A trend period of  $T_{\text{trend}} = 12500$  samples has been selected for a measured (solid) signal of  $N = 18000$  samples. The white dots indicate the mean connection values where fictitious x-y axes have been positioned to deflect the beginning and ending data of  $T_{\text{settling}} = 8098$  samples. The grey line shows the computed trend signal and the horizontal (dashed) lines indicate the bounds which have been used for clipping the data signal ; b) Detrended (solid) signal.

## A.2 Peakshaving

Although the influence of spikes in the recorded signals has been limited by clipping during the detrending phase, the remaining amplitudes of the spikes are still expected to be larger than the signal variations. If these spikes are not removed from the process signals, the corresponding energy might be large compared to the noise energy. Therefore these disturbances can have a considerable influence on the estimated model whereas they may not be caused by the dynamics of the process. A correction of these data samples is therefore necessary.



**Fig. A.2 :** Peakshaving in data preprocessing ; a) The horizontal (dashed) lines indicate the newly defined upper and lower bounds for peakshaving. All data values of the detrended (grey) signal outside these bounds are considered as outliers and have been corrected by interpolation (solid signal) ; b) Peakshaved (solid) signal.

The data correction of these spikes can be realized by linear interpolation. Since the duration of outliers can range from one sample to several samples, the first sample before and after an outlier has to be determined. Outliers are selected by defining new upper and lower bounds for the detrended signal. In fact, all those data samples are considered to belong to the selected outlier which are outside this band. Once these indices are known, the data samples belonging to a outlier can be corrected by linear interpolation. This correction has to be performed for all detected outliers. The peakshaving in the data preprocessing has been illustrated in Fig. A.2 where the clipped and detrended signal of Fig. A.1 is further processed.

### A.3 Delay Correction

Time delays in industry are often introduced because the required information is not instantaneously available (e.g. due to delay in data acquisition) and/or can only be measured at delayed time instants (e.g. due to high temperature which sensors cannot tolerate). If measured process responses contain time delays and if the data is not corrected before using it for process identification, these time delays necessarily have to be estimated as part of the process model. Although the time delays constitute an essential part of the process, they can be treated separately in the identification process so as to avoid unnecessary complex (i.e. high order) models to describe the dynamical behaviour of the process under study. Estimation of the process delays based on the data during the identification by means of polynomial models, increases the number of parameters to be estimated significantly. The model quality will reduce because of the effect of overparametrization. Therefore, time delays should be estimated separately and treated as a priori information for the process identification.

Estimation of time delays in process transfers can be done using correlation techniques. When analyzing the cross-correlation between input and output, the process is assumed to be ergodic and the input signal applied to the dynamical system is assumed to be a stationary, white, zero mean noise sequence.

The normalized cross-correlation function for time lag  $\tau$  when  $N$  tends to infinity between an input signal  $u$  and an output signal  $y$  is defined by :

$$r_{uy}(\tau) = \frac{\frac{1}{N} \sum_{k=1}^{N-\max(0,\tau)} u(k) y(k+\tau)}{\sqrt{r_u(0) r_y(0)}} \quad (\text{A.8})$$

where

$$r_u(0) = \frac{1}{N} \sum_{k=1}^N u^2(k) \quad , \quad r_y(0) = \frac{1}{N} \sum_{k=1}^N y^2(k) \quad (\text{A.9})$$

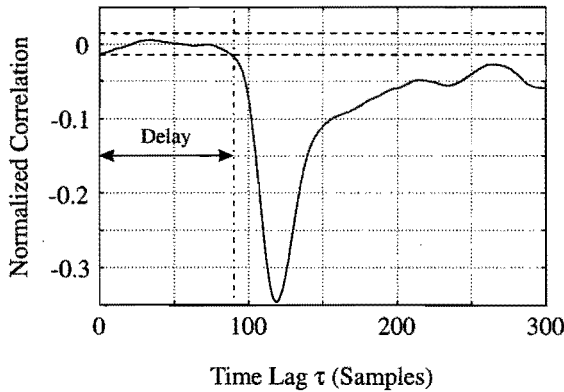
The advantage of computing the normalized cross-correlation function is that confidence intervals can be provided. It has been shown in [SS89] that when the number of data samples  $N$  tends to infinity and  $r_{uy}(\infty) = 0$ , the normalized cross-correlation function converges in distribution to a normal distribution :

$$\sqrt{N} \, r_{uy}(\tau) \xrightarrow{\text{dist}} N(0, 1) \quad (\text{A.10})$$

and a 95% confidence interval indicating no correlation between the signals (null hypothesis) is given by :

$$|r_{uy}(\tau)| \leq \frac{1.96}{\sqrt{N}} \quad (\text{A.11})$$

The time delays can be found by selecting the beginning of the cross-correlation values which differ significantly from the confidence interval. An example of estimating the time delay from the cross-correlation function using industrial data, is depicted in Fig. A.3.



**Fig. A.3 :** Delay estimation. The dashed lines indicate the 95% confidence interval. The input has been excited with a PRBN sequence and the delay obtained from this figure is about 90 samples.



Note that for a process with  $nu$  inputs and  $ny$  outputs typically ( $nu \times ny$ ) different time delays have to be corrected. The correction of the time delays can be realized by shifting the signals in time with respect to each other. Suppose now that for a multivariable system a delay matrix,  $M_{\text{delay}} \in \mathbb{R}^{ny \times nu}$ , has been determined from the various cross-correlation functions. The output delay,  $d_y \in \mathbb{R}^{ny \times 1}$ , has to be computed first by selecting the minimum delay for every output :

$$\begin{aligned} d_{y,i} &= \min(M_{\text{delay},i*}) \\ M_{\text{delay},i*} &:= M_{\text{delay},i*} - d_{y,i} \end{aligned} \quad (\text{A.12})$$

The remaining delays have to be corrected by the inputs,  $d_u \in \mathbb{R}^{1 \times nu}$  :

$$d_{u,j} = -\min(M_{\text{delay},*j}) \quad (\text{A.13})$$

and the process delay matrix which can be realized,  $M_{\text{delay}}^r \in \mathbb{R}^{ny \times nu}$ , is then given by :

$$M_{\text{delay},ij}^r = d_{y,i} - d_{u,j} \quad (\text{A.14})$$

Note, that this delay correction depends on the order in which the input and output delays are processed. Further, not all delays in a multivariable system which have been estimated, for example from the cross-correlation functions (Fig. A.3), can always be corrected because the degrees of freedom for input/output delay correction are limited. For example in a 2-input 2-output system, 3 delays can be corrected which define automatically the remaining delay. In general,  $nu+ny-1$  delays can be corrected in a multivariable system with  $nu$  inputs and  $ny$  outputs.

## A.4 Data Decimation and Scaling

The final step in the preprocessing to obtain data suitable for identification, is the data decimation and scaling. Since it has been assumed that the data is recorded using a higher sampling frequency than needed for identification, the decimation of the data can be realized by simply selecting every  $k^{\text{th}}$  sample corresponding to the oversampling ratio.

The most important step however is the data scaling. As described in Chapter 2 to realize a correct identification criterion for prediction error minimization of multivariable systems, the data should be scaled in such a way that the covariance matrix satisfies :

$$\Lambda(\hat{\underline{\theta}}) = \begin{bmatrix} \sigma_{e_1}^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{e_{ny}}^2 \end{bmatrix} = \sigma_e^2 \mathbf{I} \quad (\text{A.15})$$

Of course, this scaling is not known a priori and an iterative procedure to satisfy Eq. A.15 seems not very useful in practice. Therefore an alternative approach will be proposed.

Because the data will be recorded probably in different units resulting in large differences between the signals with respect to their amplitude, the first step in the scaling is approximately unifying the signal amplitude to equal standard deviation :

$$\begin{aligned} \underline{U}_{\text{scale}} &= \left[ \text{diag}(\underline{\sigma}_u) \right]^{-1} \underline{U} \\ \underline{Y}_{\text{scale}} &= \left[ \text{diag}(\underline{\sigma}_y) \right]^{-1} \underline{Y} \end{aligned} \quad (\text{A.16})$$

where  $\underline{\sigma}_u \in \mathbb{R}^{1 \times nu}$  and  $\underline{\sigma}_y \in \mathbb{R}^{1 \times ny}$  define the standard deviations for all inputs and outputs respectively.

Applying prediction error minimization for multivariable systems, Eq. A.16 shows that the input/output data has to be scaled in such a way that the supposed white noise sequences disturbing the process have equal variance. Otherwise a relative weighting is introduced during the prediction error minimization reducing the model quality. Of course, these variances are not known a priori in practice. An accurate estimate however, can be obtained by first estimating a high order equation error model :

$$\underline{A}_h(z^{-1}) \underline{y}_{\text{scale}}(k) = \underline{B}_h(z^{-1}) \underline{u}_{\text{scale}}(k) + \underline{\xi}(k) \quad (\text{A.17})$$

which can be computed analytically and can approximate any linear system arbitrary well for sufficiently high order as  $N$  tends to infinity (see Section 2.3.3). Using the standard deviation of the prediction error,  $\underline{\sigma}_e \in \mathbb{R}^{1 \times ny}$ , computed from this high order ARX-model as new scaling factors, the rescaled output data to realize Eq. A.15 is given by :

$$\underline{Y}_{\text{scale}} := \min(\underline{\sigma}_e) \left[ \text{diag}(\underline{\sigma}_e) \right]^{-1} \underline{Y}_{\text{scale}} \quad (\text{A.18})$$

Note that only rescaling of the output is sufficient since in prediction error minimization it is assumed that the inputs are exactly known. This results in the following overall

scaling factors for the original data :

$$\begin{aligned} S_u &= \left[ \text{diag}(\underline{\sigma}_u) \right]^{-1} \\ S_y &= \min(\underline{\sigma}_e) \left[ \text{diag}(\underline{\sigma}_y \underline{\sigma}_e) \right]^{-1} \end{aligned} \quad (\text{A.19})$$

The scaling of the data using this approach is based on the assumption that the final low order model obtained by identification describes the process accurately realizing a white prediction error. Because a white prediction error cannot always be realized in an industrial environment, the correctness of the scaling should always be verified after completing the model estimation.



# B

## ***Minimal Polynomial Identification***

---

B.1	Model Structure	B.4	Parameter Optimization
B.2	Prediction error	B.5	Gradient Computation
B.3	Pseudo-linear Regression Form	B.6	Hessian Computation

---

This appendix describes briefly the basic ideas of the minimal polynomial identification procedure based on numerical minimization of the prediction error, the identification criterion. The prediction error approach will be reviewed briefly and the minimal polynomial model structure is rewritten into a pseudo-linear regression form. The solution to the identification problem is described in parameter optimization using analytic expressions for the Gradient and the Hessian, the first and second order derivatives of the prediction error, respectively.

### ***B.1 Model Structure***

The input/output data of a multivariable system with  $n_u$  inputs and  $n_y$  outputs consisting of  $N$  data samples together with the independent white noise sequences disturbing the process, are defined as :

$$Y = \begin{bmatrix} \underline{y}^T(1) \\ \vdots \\ \underline{y}^T(N) \end{bmatrix} \in \mathbb{R}^{N \times ny}, \quad U = \begin{bmatrix} \underline{u}^T(1) \\ \vdots \\ \underline{u}^T(N) \end{bmatrix} \in \mathbb{R}^{N \times nu}, \quad \Xi = \begin{bmatrix} \underline{\xi}^T(1) \\ \vdots \\ \underline{\xi}^T(N) \end{bmatrix} \in \mathbb{R}^{N \times ny} \quad (B.1)$$

For identification of multivariable systems, the minimal polynomial model structure which has been introduced in Section 2.2.2, is given by :

$$A(z^{-1}) \underline{y}(k) = \frac{B(z^{-1})}{F(z^{-1})} \underline{u}(k) + \frac{C(z^{-1})}{D(z^{-1})} \underline{\xi}(k) \quad (B.2)$$

where the polynomials are defined as :

$$\begin{aligned} A(z^{-1}) &= \left( 1 + a_1 z^{-1} + \dots + a_{na} z^{-na} \right) I^{ny} \\ B_{ij}(z^{-1}) &= z^{-nk_{ij}} \left( b_{ij,1} + b_{ij,2} z^{-1} + \dots + b_{ij,nb_{ij}} z^{-nb_{ij}+1} \right) \\ C_{ii}(z^{-1}) &= 1 + c_{ii,1} z^{-1} + \dots + c_{ii,nc_{ii}} z^{-nc_{ii}} \\ D(z^{-1}) &= 1 + d_1 z^{-1} + \dots + d_{nd} z^{-nd} \\ F(z^{-1}) &= 1 + f_1 z^{-1} + \dots + f_{nf} z^{-nf} \end{aligned} \quad \begin{array}{l} i = 1 \dots ny \\ j = 1 \dots nu \end{array} \quad (B.3)$$

In this model structure,  $A(z^{-1})$ ,  $D(z^{-1})$  and  $F(z^{-1})$  are described by scalar polynomials while  $B(z^{-1})$  and  $C(z^{-1})$  are matrix polynomials of size  $(ny \times nu)$  and  $(ny \times ny)$  respectively. Under the assumption that the influences of the noise sequences are mutually independent,  $C(z^{-1})$  can be restricted to have diagonal entries only (off-diagonal polynomials are zero). This reduces the degrees of freedom and eliminates in fact the true multivariable character of the noise model, but the computational complexity for the parameter estimation reduces significantly. The complete structure of the multivariable minimal polynomial model is defined by the structure array :

$$\begin{aligned} nn &= \begin{bmatrix} na & nb_{11} \dots nb_{1nu} & nb_{21} \dots nb_{nynu} & nc_{11} \dots nc_{nyny} \\ nd & nf & nk_{11} \dots nk_{1nu} & nk_{21} \dots nk_{nynu} \end{bmatrix} \end{aligned} \quad (B.4)$$

where the entries define the orders of the various polynomials and where  $nk_{ij}$  indicates the additional delay from input  $j$  to output  $i$ . All polynomial coefficients of Eq. B.3 are combined in a parameter vector,  $\underline{\theta} \in \Theta_{pe} \subset \mathbb{R}^{n\theta}$ , defined by :

$$\underline{\theta} = \left[ a_1 \dots a_{na} b_{11,1} \dots b_{11,nb_{11}} b_{12,1} \dots b_{ny nu, nb_{ny nu}} c_{11,1} \dots c_{11,nc_{11}} c_{22,1} \dots c_{ny ny, nc_{ny ny}} d_1 \dots d_{nd} f_1 \dots f_{nf} \right]^T \quad (B.5)$$

Note that the parameter vector  $\underline{\theta}$  will be restricted to those values for which the model as well as the predictor are asymptotically stable (see Eq. 2.5).

## B.2 Prediction Error

The prediction error is defined by :

$$\underline{e}(k, \underline{\theta}) = \underline{y}(k) - \hat{\underline{y}}(k | \underline{\theta}) \quad (B.6)$$

where  $\underline{\theta}$  is defined in Eq. B.5 and the predictor is defined as :

$$\hat{\underline{y}}(k | \underline{\theta}) = D(z^{-1}) C^{-1}(z^{-1}) \frac{B(z^{-1})}{F(z^{-1})} \underline{u}(k) + \left[ 1 - D(z^{-1}) C^{-1}(z^{-1}) A(z^{-1}) \right] \underline{y}(k) \quad (B.7)$$

The prediction error :

$$\underline{e}(k, \underline{\theta}) = D(z^{-1}) C^{-1}(z^{-1}) \left[ A(z^{-1}) \underline{y}(k) - \frac{B(z^{-1})}{F(z^{-1})} \underline{u}(k) \right] \quad (B.8)$$

can be simplified by introducing the following auxiliary variables :

$$\underline{w}(k, \underline{\theta}) = \frac{B(z^{-1})}{F(z^{-1})} \underline{u}(k) \quad , \quad \underline{v}(k, \underline{\theta}) = A(z^{-1}) \underline{y}(k) - \underline{w}(k, \underline{\theta}) \quad (B.9)$$

where similar to the data definition in Eq. B.1 :

$$W(\underline{\theta}) = \begin{bmatrix} \underline{w}^T(1, \underline{\theta}) \\ \vdots \\ \underline{w}^T(N, \underline{\theta}) \end{bmatrix} \in \mathbb{R}^{N \times ny}, \quad V(\underline{\theta}) = \begin{bmatrix} \underline{v}^T(1, \underline{\theta}) \\ \vdots \\ \underline{v}^T(N, \underline{\theta}) \end{bmatrix} \in \mathbb{R}^{N \times nu}, \quad E(\underline{\theta}) = \begin{bmatrix} \underline{e}^T(1, \underline{\theta}) \\ \vdots \\ \underline{e}^T(N, \underline{\theta}) \end{bmatrix} \in \mathbb{R}^{N \times ny} \quad (B.10)$$

Then, the prediction error  $\underline{e}(k, \underline{\theta})$  reduces to :

$$\underline{e}(k, \underline{\theta}) = D(z^{-1}) C^{-1}(z^{-1}) \underline{v}(k, \underline{\theta}) \quad (B.11)$$

Since  $C(z^{-1})$  is restricted to have diagonal entries only, the polynomial matrix inversion

in Eq. B.11 becomes very simple,  $C^{-1}(z^{-1}) = \text{diag}\{1/C_{11}(z^{-1}) \dots 1/C_{nyny}(z^{-1})\}$ . The inversion reduces to simple scalar operations applied to the diagonal entries  $C_{ii}(z^{-1})$ . Note that for identification, it is required that the process under study is stable. Therefore, model stability which is obtained by stability of the polynomials  $A(z^{-1})$ ,  $D(z^{-1})$  and  $F(z^{-1})$ , i.e. having roots inside the unit circle, should be imposed by the data. For prediction error estimation however, the predictor has to be stable as well in order to compute stable predictions. This requires stability of the polynomials  $C(z^{-1})$ , or, in other words, the noise model has to be minimum phase. Because of the allpass character of this noise model, due to monic polynomials  $C(z^{-1})$  and  $D(z^{-1})$ , stable predictors can be realized by multiplying the noise model with an appropriate allpass filter, without changing the estimate. In practice, this corresponds to deflecting possible unstable roots of the polynomials  $C(z^{-1})$  into the unit circle. This approach guarantees a stable estimate of the multivariable system on the long run.

### B.3 Pseudo-linear Regression Form

To derive a pseudo-linear regression form of the minimal polynomial model structure, the scalar case will be considered first :

$$\left\{ \begin{array}{l} A(z^{-1}) y(k) = \underbrace{\frac{B(z^{-1})}{F(z^{-1})} u(k)}_{w(k, \underline{\theta})} + \frac{C(z^{-1})}{D(z^{-1})} \xi(k) \\ \left[ \underbrace{A(z^{-1}) y(k) - w(k, \underline{\theta})}_{v(k, \underline{\theta})} \right] = \frac{C(z^{-1})}{D(z^{-1})} \xi(k) \end{array} \right. \quad (\text{B.12})$$

The prediction error  $e(k, \underline{\theta})$  which describes in fact the noise sequence  $\xi(k)$  disturbing the process, can be written as :

$$\begin{aligned} -e(k, \underline{\theta}) &= [C(z^{-1}) - 1] e(k, \underline{\theta}) - [D(z^{-1}) - 1] v(k, \underline{\theta}) - v(k, \underline{\theta}) \\ &= [C(z^{-1}) - 1] e(k, \underline{\theta}) - [D(z^{-1}) - 1] v(k, \underline{\theta}) - [A(z^{-1}) - 1] y(k) \\ &\quad - y(k) - w(k, \underline{\theta}) \end{aligned} \quad (\text{B.13})$$

where  $A(z^{-1})$ ,  $C(z^{-1})$ ,  $D(z^{-1})$  and  $F(z^{-1})$  are monic polynomials. This results in the



following expression of the predictor :

$$\begin{aligned}
 \hat{y}(k|\underline{\theta}) - y(k) &= [C(z^{-1}) - 1] e(k, \underline{\theta}) - [D(z^{-1}) - 1] v(k, \underline{\theta}) - [A(z^{-1}) - 1] y(k) \\
 &\downarrow \\
 &= -y(k) - [F(z^{-1}) - 1] w(k, \underline{\theta}) + B(z^{-1}) u(k) \\
 \hat{y}(k|\underline{\theta}) &= [A(z^{-1}) - 1] y(k) + B(z^{-1}) u(k) + [C(z^{-1}) - 1] e(k, \underline{\theta}) \\
 &\quad - [D(z^{-1}) - 1] v(k, \underline{\theta}) - [F(z^{-1}) - 1] w(k, \underline{\theta})
 \end{aligned} \tag{B.14}$$

which can then be rewritten in the required pseudo-linear regression form :

$$\hat{\underline{y}}(\underline{\theta}) = \left[ \underbrace{\phi_y}_{a} \quad \underbrace{\phi_u}_{b} \quad \underbrace{\phi_e(\underline{\theta})}_{c} \quad \underbrace{\phi_v(\underline{\theta})}_{d} \quad \underbrace{\phi_w(\underline{\theta})}_{f} \right] \underline{\theta} \quad \leftarrow \text{Parameters} \tag{B.15}$$

where  $\underline{\theta}$  is defined in Eq. B.5. The extension to the multivariable case for the minimal polynomial model structure is straightforward and results in :

$$\hat{\underline{y}}_{\text{col}}(\underline{\theta}) = \text{col}(\hat{\underline{Y}}(\underline{\theta})) = \Phi(\underline{\theta}) \underline{\theta} \tag{B.16}$$

where

$$\Phi(\underline{\theta}) = \begin{bmatrix} \phi_{y_1} & | & \phi_{u_{11}} & \dots & \phi_{u_{1nu}} & 0 & \dots & 0 & \dots & 0 & | \\ \phi_{y_2} & | & 0 & \dots & 0 & \phi_{u_{21}} & \dots & \phi_{u_{2nu}} & \ddots & & | \\ \vdots & | & \vdots & & & 0 & \dots & 0 & \ddots & & | \\ \vdots & | & \vdots & & & & & & \ddots & 0 & \dots & 0 & | \\ \phi_{y_{ny}} & | & 0 & & \dots & & & & 0 & \phi_{u_{ny1}} & \dots & \phi_{u_{nynu}} & | \end{bmatrix} \tag{B.17}$$

$$\begin{bmatrix} \phi_{e_1}(\underline{\theta}) & 0 & \dots & 0 & | & \phi_{v_1}(\underline{\theta}) & | & \phi_{w_1}(\underline{\theta}) \\ 0 & \phi_{e_2}(\underline{\theta}) & \ddots & \vdots & | & \phi_{v_2}(\underline{\theta}) & | & \phi_{w_2}(\underline{\theta}) \\ \vdots & 0 & \ddots & \vdots & | & \vdots & | & \vdots \\ \vdots & & \ddots & 0 & | & \vdots & | & \vdots \\ 0 & \dots & 0 & \phi_{e_{ny}}(\underline{\theta}) & | & \phi_{v_{ny}}(\underline{\theta}) & | & \phi_{w_{ny}}(\underline{\theta}) \end{bmatrix}$$

and

$$\hat{\underline{y}}_{\text{col}}(\underline{\theta}) = \left[ \hat{y}_1(n+1|\underline{\theta}) \dots \hat{y}_1(N|\underline{\theta}) \hat{y}_2(n+1|\underline{\theta}) \dots \hat{y}_{ny}(N|\underline{\theta}) \right]^T \tag{B.18}$$

The column representation of the predictor  $\hat{\underline{y}}_{\text{col}}(\underline{\theta})$  results from the "common"

parameters in all outputs due to the minimal polynomial model structure. The data samples of all outputs in the predictor matrix  $\hat{Y}(\underline{\theta})$  have to be stacked on top of each other when rewriting the minimal polynomial model into a pseudo-linear regression form. The several entries of the regression matrix  $\Phi(\underline{\theta})$  describe the regression sub-matrices which have been built from the input/output data and the auxiliary variables (Eq. B.9) :

$$\begin{aligned} \phi_{y_i} &= \begin{bmatrix} -y_i(n) & \dots & -y_i(n-na+1) \\ \vdots & & \vdots \\ -y_i(N-1) & \dots & -y_i(N-na) \end{bmatrix}, \quad \phi_{u_{ij}} = \begin{bmatrix} u_j(n-nk_{ij}+1) & \dots & u_j(n-nk_{ij}-nb_{ij}+2) \\ \vdots & & \vdots \\ u_j(N-nk_{ij}) & \dots & u_j(N-nk_{ij}-nb_{ij}+1) \end{bmatrix} \\ \phi_{e_i}(\underline{\theta}) &= \begin{bmatrix} e_i(n, \underline{\theta}) & \dots & e_i(n-nc_{ii}+1, \underline{\theta}) \\ \vdots & & \vdots \\ e_i(N-1, \underline{\theta}) & \dots & e_i(N-nc_{ii}, \underline{\theta}) \end{bmatrix}, \quad \phi_{v_i}(\underline{\theta}) = \begin{bmatrix} -v_i(n, \underline{\theta}) & \dots & -v_i(n-nd+1, \underline{\theta}) \\ \vdots & & \vdots \\ -v_i(N-1, \underline{\theta}) & \dots & -v_i(N-nd, \underline{\theta}) \end{bmatrix} \\ \phi_{w_i}(\underline{\theta}) &= \begin{bmatrix} -w_i(n, \underline{\theta}) & \dots & -w_i(n-nf+1, \underline{\theta}) \\ \vdots & & \vdots \\ -w_i(N-1, \underline{\theta}) & \dots & -w_i(N-nf, \underline{\theta}) \end{bmatrix} \end{aligned} \quad (B.19)$$

where  $n$  defines the maximum delay required for initialization :

$$n = \max_{i,j} \{ [na \ nb_{ij} + nk_{ij} - 1 \ nc_{ii} \ nd \ nf] \} \quad (B.20)$$

## B.4 Parameter Optimization

Methods for numerical minimization of the weighted loss-function  $V_{\tilde{e}}(\underline{\theta})$  :

$$V_{\tilde{e}}(\underline{\theta}) = \frac{1}{N_{col}} \sum_{\ell=1}^{N_{col}} \tilde{e}_{col}^2(\ell, \underline{\theta}) \quad (B.21)$$

where

$$\tilde{e}_i(k, \underline{\theta}) = L_i(z^{-1}) e_i(k, \underline{\theta}) \quad (B.22)$$

update the estimate iteratively. This is usually done according to :

$$\hat{\underline{\theta}}^{i+1} = \hat{\underline{\theta}}^i + \alpha^i \underline{p}^i \quad (\text{B.23})$$

where  $\underline{p}^i$  is a search direction based on information about  $V_{\tilde{\epsilon}}(\underline{\theta})$  acquired at previous iterations, and  $\alpha^i$  is a positive constant determined so that an appropriate decrease in the value of  $V_{\tilde{\epsilon}}(\underline{\theta})$  is obtained. Numerical minimization methods using values of the function, of its gradient (the first order derivative), and of its Hessian (the second order derivative) correspond to Newton algorithms, where the correction is chosen in the 'Newton' direction :

$$\underline{p}^i = - \left[ \nabla_{\underline{\theta}\underline{\theta}}^2 V_{\tilde{\epsilon}}(\hat{\underline{\theta}}^i) \right]^{-1} \nabla_{\underline{\theta}} V_{\tilde{\epsilon}}(\hat{\underline{\theta}}^i) \quad (\text{B.24})$$

To obtain an expression for the gradient vector of  $V_{\tilde{\epsilon}}(\underline{\theta})$  note that the first partial derivative with respect to  $\theta_j$  is given by :

$$\frac{\partial V_{\tilde{\epsilon}}(\underline{\theta})}{\partial \theta_j} = \frac{2}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \tilde{\epsilon}_{\text{col}}(\ell, \underline{\theta}) \frac{\partial \tilde{\epsilon}_{\text{col}}(\ell, \underline{\theta})}{\partial \theta_j} \quad (\text{B.25})$$

Define now the Jacobian matrix  $J(\underline{\theta})$  by :

$$J(\underline{\theta}) = \begin{bmatrix} \frac{\partial \tilde{\epsilon}_{\text{col}}(1, \underline{\theta})}{\partial \theta_1} & \dots & \frac{\partial \tilde{\epsilon}_{\text{col}}(1, \underline{\theta})}{\partial \theta_{n\theta}} \\ \vdots & & \vdots \\ \frac{\partial \tilde{\epsilon}_{\text{col}}(N_{\text{col}}, \underline{\theta})}{\partial \theta_1} & \dots & \frac{\partial \tilde{\epsilon}_{\text{col}}(N_{\text{col}}, \underline{\theta})}{\partial \theta_{n\theta}} \end{bmatrix} \quad (\text{B.26})$$

then the gradient vector can be written as :

$$\nabla_{\underline{\theta}} V_{\tilde{\epsilon}}(\underline{\theta}) \triangleq G(\underline{\theta}) = \frac{2}{N_{\text{col}}} J^T(\underline{\theta}) \tilde{\epsilon}_{\text{col}}(\underline{\theta}) \quad (\text{B.27})$$

Differentiating Eq. B.25 with respect to  $\theta_i$  gives the  $ij^{\text{th}}$ -element of the Hessian matrix :

$$\frac{\partial^2 V_{\tilde{\epsilon}}(\underline{\theta})}{\partial \theta_i \partial \theta_j} = \frac{2}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \left\{ \frac{\partial \tilde{\epsilon}_{\text{col}}(\ell, \underline{\theta})}{\partial \theta_i} \frac{\partial \tilde{\epsilon}_{\text{col}}(\ell, \underline{\theta})}{\partial \theta_j} + \tilde{\epsilon}_{\text{col}}(\ell, \underline{\theta}) \frac{\partial^2 \tilde{\epsilon}_{\text{col}}(\ell, \underline{\theta})}{\partial \theta_i \partial \theta_j} \right\} \quad (\text{B.28})$$

Let  $H_{\ell}(\underline{\theta})$  be the Hessian matrix of  $\tilde{\epsilon}_{\text{col}}(\ell, \underline{\theta})$  :

$$H_\ell(\underline{\theta}) = \nabla_{\theta\theta}^2 \tilde{\epsilon}_{\text{col}}(\ell, \underline{\theta}) \quad (\text{B.29})$$

then the complete Hessian matrix of  $V_{\tilde{\epsilon}}(\underline{\theta})$  can be written as :

$$\begin{aligned} \nabla_{\theta\theta}^2 V_{\tilde{\epsilon}}(\underline{\theta}) \triangleq H(\underline{\theta}) &= \underbrace{\frac{2}{N_{\text{col}}} \mathbf{J}^T(\underline{\theta}) \mathbf{J}(\underline{\theta})}_{\mathbf{R}(\underline{\theta})} + \underbrace{\frac{2}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \tilde{\epsilon}_{\text{col}}(\ell, \underline{\theta}) H_\ell(\underline{\theta})}_{\mathbf{S}(\underline{\theta})} \\ &=: \mathbf{R}(\underline{\theta}) + \mathbf{S}(\underline{\theta}) \end{aligned} \quad (\text{B.30})$$

Apparently, the derivatives of a sum-of-squares function display considerable structure. This structure can be explored even more when considering the special construction of the minimal polynomial model which is clearly visible in the pseudo-linear regression form (see e.g. regression matrix  $\Phi(\theta)$  of Eq. B.17 and the appendices B.5 and B.6). The iterative search for the optimal parameter estimates is terminated when at iteration step  $i$  no improvement of the loss-function  $V_{\tilde{\epsilon}}(\underline{\theta})$  in the calculated search direction  $\underline{p}^i$  can be found after the step length  $\alpha^i$  has been bisected 10 times starting with  $\alpha^{i,0} = 1$ , or when :

$$\| \nabla_{\theta} V_{\tilde{\epsilon}}(\underline{\theta}) \|_2 = \| \mathbf{G}(\underline{\theta}) \|_2 \leq \gamma_1 \quad \text{and} \quad \| (\hat{\underline{\theta}}^{i+1} - \hat{\underline{\theta}}^i) / \alpha^i \| = \| \underline{p}^i \|_2 \leq \gamma_2 \quad (\text{B.31})$$

are satisfied where a tolerance margin  $\gamma_1$  is defined for the maximum norm of the gradient and a tolerance margin  $\gamma_2$  for the minimal parameter improvement. To update the parameter vector during the optimization correctly towards the minimum of the criterium, it is required that the Hessian matrix  $H(\underline{\theta})$  is positive definite (all eigenvalues larger than zero). When starting the optimization, possibly far from the optimum, the term  $\mathbf{S}(\underline{\theta})$  in Eq. B.30, consisting of the second order derivatives with respect to the prediction error, can make the Hessian non-positive definite. In this case, to guarantee convergence to a (local) minimum, the Hessian matrix will be approximated by  $\mathbf{R}(\underline{\theta})$  which is positive definite by construction. To terminate the optimization successfully, the complete Hessian matrix  $H(\underline{\theta})$  has to be positive definite and the tolerance margins defined in Eq. B.31 have to be satisfied. This problem of a non-positive definite Hessian matrix  $H(\underline{\theta})$  can be reduced by providing accurate initial estimates of  $\underline{\theta}$  (see Subsection 2.3.3).

## B.5 Gradient Computation

In this section, expressions will be provided to compute the gradient (first-order

derivatives) of the weighted prediction error (Eq. B.22) :

$$\tilde{\mathbf{e}}(\underline{\theta}) = \mathbf{L}(\mathbf{z}^{-1}) \mathbf{C}^{-1}(\mathbf{z}^{-1}) \mathbf{D}(\mathbf{z}^{-1}) \left[ \mathbf{A}(\mathbf{z}^{-1}) \mathbf{y}(\mathbf{k}) - \frac{\mathbf{B}(\mathbf{z}^{-1})}{\mathbf{F}(\mathbf{z}^{-1})} \mathbf{u}(\mathbf{k}) \right] \quad (\text{B.32})$$

with respect to the parameters  $\underline{\theta}$  in an efficient way. The complete Jacobian matrix  $\mathbf{J}(\underline{\theta}) \in \mathbb{R}^{ny(N-n) \times n\theta}$  (Eq. B.26) is defined by :

$$\mathbf{J}(\underline{\theta}) = \begin{bmatrix} \mathbf{J}_{a,1}^o(\underline{\theta}) & | & \mathbf{J}_{b,1}^o(\underline{\theta}) \dots \mathbf{J}_{b,nu}^o(\underline{\theta}) & 0 & \dots & 0 & \dots & 0 & | \\ \mathbf{J}_{a,2}^o(\underline{\theta}) & | & 0 & \dots & 0 & \mathbf{J}_{b,nu+1}^o(\underline{\theta}) \dots \mathbf{J}_{b,2nu}^o(\underline{\theta}) & \dots & \vdots & | \\ \vdots & | & \vdots & & & 0 & \dots & 0 & \vdots & | \\ \vdots & | & \vdots & & & \vdots & & 0 & \dots & 0 & | \\ \mathbf{J}_{a,ny}^o(\underline{\theta}) & | & 0 & \dots & \dots & 0 & \mathbf{J}_{b,(ny-1)nu+1}^o(\underline{\theta}) \dots \mathbf{J}_{b,ny}^o(\underline{\theta}) & | \\ \\ \mathbf{J}_{c,1}^o(\underline{\theta}) & 0 & \dots & 0 & | & \mathbf{J}_{d,1}^o(\underline{\theta}) & | & \mathbf{J}_{f,1}^o(\underline{\theta}) \\ 0 & \mathbf{J}_{c,2}^o(\underline{\theta}) & \dots & \vdots & | & \mathbf{J}_{d,2}^o(\underline{\theta}) & | & \mathbf{J}_{f,2}^o(\underline{\theta}) \\ \vdots & 0 & \dots & \vdots & | & \vdots & | & \vdots \\ \vdots & \dots & \dots & 0 & | & \vdots & | & \vdots \\ 0 & \dots & 0 & \mathbf{J}_{c,ny}^o(\underline{\theta}) & | & \mathbf{J}_{d,ny}^o(\underline{\theta}) & | & \mathbf{J}_{f,ny}^o(\underline{\theta}) \end{bmatrix} \quad (\text{B.33})$$

which shows the same structure as the regression matrix  $\Phi(\underline{\theta})$ . The derivatives can be distinguished with respect to the parameters of the several polynomials and the outputs of the dynamical system, e.g. :

$$\frac{\partial \tilde{e}_i(\mathbf{k}, \underline{\theta})}{\partial a_n} = \mathbf{L}_i(\mathbf{z}^{-1}) \frac{\mathbf{D}(\mathbf{z}^{-1})}{\mathbf{C}_{ii}(\mathbf{z}^{-1})} y_i(\mathbf{k}-n) \quad (\text{B.34})$$

This expression shows clearly that the derivatives of all parameters  $a_n$  for the  $i^{\text{th}}$  output can be computed by filtering  $y_i$  correctly. To reduce data storage, a matrix  $\mathbf{J}_a(\underline{\theta})$  is constructed :

$$\mathbf{J}_a(\underline{\theta}) = [ \mathbf{J}_{a,*1}(\underline{\theta}) \dots \mathbf{J}_{a,*ny}(\underline{\theta}) ] \in \mathbb{R}^{N \times ny} \quad (\text{B.35})$$

where the  $i^{\text{th}}$  column contains the derivative information of the  $a$ -parameters for output  $i$ , i.e. :

$$\mathbf{J}_{a,*i}(\underline{\theta}) \triangleq \mathbf{L}_i(\mathbf{z}^{-1}) \frac{\mathbf{D}(\mathbf{z}^{-1})}{\mathbf{C}_{ii}(\mathbf{z}^{-1})} \mathbf{Y}_{*i} \quad (\text{B.36})$$

This is not the true Jacobian matrix, but a matrix that contains this information in a compact way. The entries of the Jacobian matrix, e.g.  $J_{a,i}^0(\underline{\theta})$ , can then be constructed by selecting the correct data sequence in the derivative matrix  $J_a(\underline{\theta})$  and shifting this data sequence according the delays which correspond to the selected parameters :

$$J_{a,i}^0(\underline{\theta}) = \begin{bmatrix} [J_a(\underline{\theta})]_{ni} & \dots & [J_a(\underline{\theta})]_{(n-na+1)i} \\ \vdots & \ddots & \vdots \\ [J_a(\underline{\theta})]_{(N-1)i} & \dots & [J_a(\underline{\theta})]_{(N-na)i} \end{bmatrix} \quad (B.37)$$

$\underbrace{\hspace{10em}}_{a_1} \quad \dots \quad \underbrace{\hspace{10em}}_{a_{na}}$

The derivatives of the prediction error with respect to the parameters of the polynomials  $B_{ij}(z^{-1})$ ,  $C_{ii}(z^{-1})$ ,  $D(z^{-1})$  and  $F(z^{-1})$ , can be derived in a similar way :

$$\frac{\partial \tilde{e}_i(k, \underline{\theta})}{\partial b_{ij,n}} = L_i(z^{-1}) \frac{-D(z^{-1})}{C_{ii}(z^{-1}) F(z^{-1})} u_j(k - nk_{ij} - n + 1) \quad (B.38)$$

$$\Rightarrow J_{b,*(i-1)nu+j}(\underline{\theta}) \triangleq L_i(z^{-1}) \frac{-D(z^{-1})}{C_{ii}(z^{-1}) F(z^{-1})} U_{*j} \quad , \quad J_b(\underline{\theta}) \in \mathbb{R}^{N \times (ny, nu)} \quad (B.39)$$

$$\begin{aligned} \frac{\partial \tilde{e}_i(k, \underline{\theta})}{\partial c_{ii,n}} &= L_i(z^{-1}) \frac{-D(z^{-1})}{C_{ii}^2(z^{-1})} \left( A(z^{-1}) y_i(k-n) - \sum_{j=1}^{nu} \frac{B_{ij}(z^{-1})}{F(z^{-1})} u_j(k - nk_{ij} - n) \right) \\ &= L_i(z^{-1}) \frac{-1}{C_{ii}(z^{-1})} e_i(k-n, \underline{\theta}) \end{aligned} \quad (B.40)$$

$$\Rightarrow J_{c, *i}(\underline{\theta}) \triangleq L_i(z^{-1}) \frac{-1}{C_{ii}(z^{-1})} E_{*i}(\underline{\theta}) \quad , \quad J_c(\underline{\theta}) \in \mathbb{R}^{N \times ny} \quad (B.41)$$

$$\begin{aligned} \frac{\partial \tilde{e}_i(k, \underline{\theta})}{\partial d_n} &= L_i(z^{-1}) \frac{1}{C_{ii}(z^{-1})} \left( A(z^{-1}) y_i(k-n) - \sum_{j=1}^{nu} \frac{B_{ij}(z^{-1})}{F(z^{-1})} u_j(k - nk_{ij} - n) \right) \\ &= L_i(z^{-1}) \frac{1}{C_{ii}(z^{-1})} v_i(k-n, \underline{\theta}) \end{aligned} \quad (B.42)$$

$$\Rightarrow J_{d,*i}(\underline{\theta}) \triangleq L_i(z^{-1}) \frac{1}{C_{ii}(z^{-1})} V_{*i}(\underline{\theta}) \quad , \quad J_d(\underline{\theta}) \in \mathbb{R}^{N \times ny} \quad (B.43)$$

$$\begin{aligned} \frac{\partial \tilde{e}_i(k, \underline{\theta})}{\partial f_n} &= L_i(z^{-1}) \frac{D(z^{-1})}{C_{ii}(z^{-1}) F^2(z^{-1})} \sum_{j=1}^{nu} B_{ij}(z^{-1}) u_j(k - nk_{ij} - n) \\ &= L_i(z^{-1}) \frac{D(z^{-1})}{C_{ii}(z^{-1}) F(z^{-1})} w_i(k - n, \underline{\theta}) \end{aligned} \quad (B.44)$$

$$\Rightarrow J_{f,*i}(\underline{\theta}) \triangleq L_i(z^{-1}) \frac{D(z^{-1})}{C_{ii}(z^{-1}) F(z^{-1})} W_{*i}(\underline{\theta}) \quad , \quad J_f(\underline{\theta}) \in \mathbb{R}^{N \times ny} \quad (B.45)$$

It has been shown that the first-order derivatives of the prediction error with respect to the parameters,  $\nabla_{\underline{\theta}} \tilde{e}(k, \underline{\theta})$ , can be computed by a single filtering procedure for all the parameters of one polynomial, i.e.  $A(z^{-1})$ ,  $B_{ij}(z^{-1})$ ,  $C_{ii}(z^{-1})$ ,  $D(z^{-1})$  and  $F(z^{-1})$ . It is therefore sufficient and efficient to store the first order derivative information in data matrices :  $J_a(\underline{\theta})$ ,  $J_b(\underline{\theta})$ ,  $J_c(\underline{\theta})$ ,  $J_d(\underline{\theta})$  and  $J_f(\underline{\theta})$  where the subscripts indicate the corresponding polynomials.

## B.6 Hessian Computation

In this section the elements of the Hessian matrix will be computed (the second-order derivatives) of the prediction error with respect to the parameters. The derivatives can be distinguished with respect to the parameters of the several polynomials similar to the Gradient computation (Section B.5), e.g. :

$$\begin{aligned} \frac{\partial^2 \tilde{e}_i(k, \underline{\theta})}{\partial a_m \partial c_{ii,n}} &= \frac{\partial}{\partial a_m} \left[ L_i(z^{-1}) \frac{-D(z^{-1})}{C_{ii}^2(z^{-1})} \left( A(z^{-1}) y_i(k-n) - \sum_{j=1}^{nu} \frac{B_{ij}(z^{-1})}{F(z^{-1})} u_j(k - nk_{ij} - n) \right) \right] \\ &= L_i(z^{-1}) \frac{-D(z^{-1})}{C_{ii}^2(z^{-1})} y_i(k-n-m) \end{aligned} \quad (B.46)$$

This shows that also the second-order derivatives can be computed by appropriate

filtering of data sequences. A matrix  $H_{ac}(\underline{\theta})$  is constructed :

$$H_{ac}(\underline{\theta}) = [ H_{ac,*1}(\underline{\theta}) \dots H_{ac,*ny}(\underline{\theta}) ] \in \mathbb{R}^{N \times ny} \quad (B.47)$$

for compact data storage, where the  $i^{\text{th}}$  column contains the second order derivative information of all (a,c)-parameters for output  $i$ , i.e. :

$$H_{ac,*i}(\underline{\theta}) \triangleq \frac{-1}{C_{ii}(z^{-1})} J_{a,*i}(\underline{\theta}) \quad (B.48)$$

Similar to the computation of the Jacobian matrix  $J(\underline{\theta})$  (Eq. B.33), the tensor elements of the Hessian matrix  $H_{\theta}(\underline{\theta})$  (Eq. B.29) can be computed efficiently by filtering data sequences. In addition, this Hessian matrix for the minimal polynomial model structure shows considerable structure where only a few elements have to be computed.

Note however, that the construction of the complete tensor  $\nabla_{\theta\theta}^2 \underline{e}_{col}(\underline{\theta}) \in \mathbb{R}^{n\theta \times n\theta \times N_{col}}$  should be avoided to reduce data storage combined with memory limitations. It is therefore much more efficient to compute instead of this tensor, the term  $S(\underline{\theta}) \in \mathbb{R}^{n\theta \times n\theta}$  (Eq. B.30) in the Hessian matrix which is clearly of much lower dimension. The structure of this term  $S(\underline{\theta})$  is depicted in Table B.1.

The entries in Table B.1 are matrices constructed by multiplying the shifted data sequences  $H_{ac}(\underline{\theta})$  with the weighted prediction error  $\tilde{E}(\underline{\theta})$ . For example, the entry  $S_{ac,1}(\underline{\theta}) \in \mathbb{R}^{na \times nc_{11}}$  is constructed by multiplying the weighted prediction error  $\tilde{E}_{*1}(\underline{\theta})$  with the data sequence  $H_{ac,*1}(\underline{\theta})$  and thereby shifting  $H_{ac,*1}(\underline{\theta})$  over  $na$  samples in vertical and over  $nc_{11}$  samples in horizontal direction :

$$S_{ac,1}(\underline{\theta}) = \begin{bmatrix} \frac{1}{N-ns} \sum_{k=ns+1}^N \tilde{e}_1(k, \underline{\theta}) H_{ac,(k-2)_1}(\underline{\theta}) & \dots & \frac{1}{N-ns} \sum_{k=ns+1}^N \tilde{e}_1(k, \underline{\theta}) H_{ac,(k-1-nc_{11})_1}(\underline{\theta}) \\ \vdots & \ddots & \vdots \\ \frac{1}{N-ns} \sum_{k=ns+1}^N \tilde{e}_1(k, \underline{\theta}) H_{ac,(k-na-1)_1}(\underline{\theta}) & \dots & \frac{1}{N-ns} \sum_{k=ns+1}^N \tilde{e}_1(k, \underline{\theta}) H_{ac,(k-na-nc_{11})_1}(\underline{\theta}) \end{bmatrix} \quad (B.49)$$

where  $ns$  defines the maximum delay required for initialization of this Hessian term :

$$ns = \max_{i,j} \left( \left[ na \ nb_{ij} + nk_{ij} - 1 \ nc_{ii} \ nd \ nf \right]^T * \right. \\ \left. \left[ na \ nb_{ij} + nk_{ij} - 1 \ nc_{ii} \ nd \ nf \right] \right) \quad (B.50)$$

Since by construction the Hessian term  $S(\underline{\theta})$  has to be symmetric, see Table B.1, it is sufficient to construct the upper triangular part only.



$\underline{\theta}$	na	nb <sub>11</sub>	...	nb <sub>1nu</sub>	nb <sub>21</sub>	...	nb <sub>nyny</sub>	nc <sub>11</sub>	nc <sub>22</sub>	...	nc <sub>nyny</sub>	nd	nf
na	0	0	...	0	0	...	0	S <sub>ac,1</sub>	S <sub>ac,2</sub>	...	S <sub>ac,ny</sub>	S <sub>ad</sub>	0
nb <sub>11</sub>	0	0	...	0	0	...	0	S <sub>bc,1</sub>	0	...	0	S <sub>bd,1</sub>	S <sub>bf,1</sub>
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
nb <sub>1nu</sub>	0	0	...	0	0	...	0	S <sub>bc,nu</sub>	0	...	0	S <sub>bd,nu</sub>	S <sub>bf,nu</sub>
nb <sub>21</sub>	0	0	...	0	0	...	0	0	S <sub>bc,nu+1</sub>	...	0	S <sub>bd,nu+1</sub>	S <sub>bf,nu+1</sub>
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
nb <sub>nyny</sub>	0	0	...	0	0	...	0	0	0	...	S <sub>bc,nyny</sub>	S <sub>bd,nyny</sub>	S <sub>bf,nyny</sub>
nc <sub>11</sub>	S <sub>ca,1</sub>	S <sub>cb,1</sub>	...	S <sub>cb,nu</sub>	0	...	0	S <sub>cc,1</sub>	0	...	0	S <sub>cd,1</sub>	S <sub>cf,1</sub>
nc <sub>22</sub>	S <sub>ca,2</sub>	0	...	0	S <sub>cb,nu+1</sub>	...	0	0	S <sub>cc,2</sub>	...	0	S <sub>cd,2</sub>	S <sub>cf,2</sub>
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
nc <sub>nyny</sub>	S <sub>ca,ny</sub>	0	...	0	0	...	S <sub>cb,nyny</sub>	0	0	...	S <sub>cc,ny</sub>	S <sub>cd,ny</sub>	S <sub>cf,ny</sub>
nd	S <sub>da</sub>	S <sub>db,1</sub>	...	S <sub>db,nu</sub>	S <sub>db,nu+1</sub>	...	S <sub>db,nyny</sub>	S <sub>dc,1</sub>	S <sub>dc,2</sub>	...	S <sub>dc,ny</sub>	0	S <sub>df</sub>
nf	0	S <sub>fb,1</sub>	...	S <sub>fb,nu</sub>	S <sub>fb,nu+1</sub>	...	S <sub>fb,nyny</sub>	S <sub>fc,1</sub>	S <sub>fc,2</sub>	...	S <sub>fc,ny</sub>	S <sub>fd</sub>	S <sub>ff</sub>

Table B.1 : Structure Hessian term  $S(\underline{\theta})$ .

The other entries can be constructed in a similar way where the dimensions are defined by the number of parameters of the corresponding polynomials.

$$\frac{\partial^2 \tilde{\epsilon}_i(\mathbf{k}, \underline{\theta})}{\partial a_m \partial a_n} = \frac{\partial}{\partial a_m} \left[ L_i(z^{-1}) \frac{D(z^{-1})}{C_{ii}(z^{-1})} y_i(\mathbf{k}-n) \right] = 0 \quad (\text{B.51})$$

$$\Rightarrow H_{aa}(\underline{\theta}) \triangleq 0 \quad (\text{B.52})$$

$$\frac{\partial^2 \tilde{e}_i(k, \underline{\theta})}{\partial a_m \partial b_{ij,n}} = \frac{\partial}{\partial a_m} \left[ L_i(z^{-1}) \frac{-D(z^{-1})}{C_{ii}(z^{-1}) F(z^{-1})} u_j(k - nk_{ij} - n + 1) \right] = 0 \quad (\text{B.53})$$

$$\Rightarrow H_{ab}(\underline{\theta}) \triangleq 0 \quad (\text{B.54})$$

$$\begin{aligned} \frac{\partial^2 \tilde{e}_i(k, \underline{\theta})}{\partial a_m \partial d_n} &= \frac{\partial}{\partial a_m} \left[ L_i(z^{-1}) \frac{1}{C_{ii}(z^{-1})} \left( A(z^{-1}) y_i(k-n) - \sum_{j=1}^{nu} \frac{B_{ij}(z^{-1})}{F(z^{-1})} u_j(k - nk_{ij} - n) \right) \right] \\ &= L_i(z^{-1}) \frac{1}{C_{ii}(z^{-1})} y_i(k-n-m) \end{aligned} \quad (\text{B.55})$$

$$\Rightarrow H_{ad,*i}(\underline{\theta}) \triangleq L_i(z^{-1}) \frac{1}{C_{ii}(z^{-1})} Y_{*i}, \quad H_{ad}(\underline{\theta}) \in \mathbb{R}^{N \times ny} \quad (\text{B.56})$$

$$\frac{\partial^2 \tilde{e}_i(k, \underline{\theta})}{\partial a_m \partial f_n} = \frac{\partial}{\partial a_m} \left[ L_i(z^{-1}) \frac{D(z^{-1})}{C_{ii}(z^{-1}) F^2(z^{-1})} \sum_{j=1}^{nu} B_{ij}(z^{-1}) u_j(k - nk_{ij} - n) \right] = 0 \quad (\text{B.57})$$

$$\Rightarrow H_{af}(\underline{\theta}) \triangleq 0 \quad (\text{B.58})$$

$$\frac{\partial^2 \tilde{e}_i(k, \underline{\theta})}{\partial b_{pq,m} \partial b_{ij,n}} = \frac{\partial}{\partial b_{pq,m}} \left[ L_i(z^{-1}) \frac{-D(z^{-1})}{C_{ii}(z^{-1}) F(z^{-1})} u_j(k - nk_{ij} - n + 1) \right] = 0 \quad (\text{B.59})$$

$$\Rightarrow H_{bb}(\underline{\theta}) \triangleq 0 \quad (\text{B.60})$$

$$\frac{\partial^2 \tilde{e}_i(k, \underline{\theta})}{\partial b_{pq,m} \partial c_{ii,n}} = \frac{\partial}{\partial b_{pq,m}} \left[ L_i(z^{-1}) \frac{-D(z^{-1})}{C_{ii}^2(z^{-1})} \left( A(z^{-1}) y_i(k-n) - \sum_{j=1}^{nu} \frac{B_{ij}(z^{-1})}{F(z^{-1})} u_j(k-nk_{ij}-n) \right) \right] \quad (B.61)$$

$$= \begin{cases} L_i(z^{-1}) \frac{D(z^{-1})}{C_{ii}^2(z^{-1}) F(z^{-1})} u_q(k-nk_{iq}-n-m+1) & p = i \\ 0 & p \neq i \end{cases}$$

$$\Rightarrow H_{bc,*(i-1)nu+j}(\underline{\theta}) \triangleq \frac{-1}{C_{ii}(z^{-1})} J_{b,*(i-1)nu+j}(\underline{\theta}) \quad , \quad H_{bc}(\underline{\theta}) \in \mathbb{R}^{N \times (ny, nu)} \quad (B.62)$$

$$\frac{\partial^2 \tilde{e}_i(k, \underline{\theta})}{\partial b_{pq,m} \partial d_n} = \frac{\partial}{\partial b_{pq,m}} \left[ L_i(z^{-1}) \frac{1}{C_{ii}(z^{-1})} \left( A(z^{-1}) y_i(k-n) - \sum_{j=1}^{nu} \frac{B_{ij}(z^{-1})}{F(z^{-1})} u_j(k-nk_{ij}-n) \right) \right] \quad (B.63)$$

$$= L_i(z^{-1}) \frac{-1}{C_{ii}(z^{-1}) F(z^{-1})} u_q(k-nk_{iq}-n-m+1)$$

$$\Rightarrow H_{bd,*(i-1)nu+j}(\underline{\theta}) \triangleq L_i(z^{-1}) \frac{-1}{C_{ii}(z^{-1}) F(z^{-1})} U_{*j} \quad , \quad H_{bd}(\underline{\theta}) \in \mathbb{R}^{N \times (ny, nu)} \quad (B.64)$$

$$\begin{aligned} \frac{\partial^2 \tilde{e}_i(k, \underline{\theta})}{\partial b_{pq,m} \partial f_n} &= \frac{\partial}{\partial b_{pq,m}} \left[ L_i(z^{-1}) \frac{D(z^{-1})}{C_{ii}(z^{-1}) F^2(z^{-1})} \sum_{j=1}^{nu} B_{ij}(z^{-1}) u_j(k - nk_{ij} - n) \right] \\ &= \begin{cases} L_i(z^{-1}) \frac{D(z^{-1})}{C_{ii}(z^{-1}) F^2(z^{-1})} u_q(k - nk_{iq} - n - m + 1) & p = i \\ 0 & p \neq i \end{cases} \end{aligned} \quad (B.65)$$

$$\Rightarrow H_{bf,*(i-1)nu+j}(\underline{\theta}) \triangleq \frac{-1}{F(z^{-1})} J_{b,*(i-1)nu+j} \quad , \quad H_{bf}(\underline{\theta}) \in \mathbb{R}^{N \times (ny, nu)} \quad (B.66)$$

$$\begin{aligned} \frac{\partial^2 \tilde{e}_i(k, \underline{\theta})}{\partial c_{pp,m} \partial c_{ii,n}} &= \frac{\partial}{\partial c_{pp,m}} \left[ L_i(z^{-1}) \frac{D(z^{-1})}{C_{ii}^2(z^{-1})} \left( A(z^{-1}) y_i(k-n) - \sum_{j=1}^{nu} \frac{B_{ij}(z^{-1})}{F(z^{-1})} u_j(k - nk_{ij} - n) \right) \right] \\ &= \begin{cases} L_i(z^{-1}) \frac{2D(z^{-1})}{C_{ii}^3(z^{-1})} \left( A(z^{-1}) y_i(k-n-m) - \sum_{j=1}^{nu} \frac{B_{ij}(z^{-1})}{F(z^{-1})} u_j(k - nk_{ij} - n - m, 1) \right) & p = i \\ 0 & p \neq i \end{cases} \end{aligned} \quad (B.67)$$

$$\Rightarrow H_{cc,*(i)}(\underline{\theta}) \triangleq \frac{-2}{C_{ii}(z^{-1})} J_{c,*(i)}(\underline{\theta}) \quad , \quad H_{cc}(\underline{\theta}) \in \mathbb{R}^{N \times ny} \quad (B.68)$$

$$\begin{aligned}
 \frac{\partial^2 \tilde{e}_i(k, \underline{\theta})}{\partial c_{pp,m} \partial d_n} &= \frac{\partial}{\partial c_{pp,m}} \left[ L_i(z^{-1}) \frac{1}{C_{ii}(z^{-1})} \left( A(z^{-1}) y_i(k-n) - \sum_{j=1}^{nu} \frac{B_{ij}(z^{-1})}{F(z^{-1})} u_j(k-nk_{ij}-n) \right) \right] \\
 &= \begin{cases} L_i(z^{-1}) \frac{-1}{C_{ii}^2(z^{-1})} \left( A(z^{-1}) y_i(k-n-m) - \sum_{j=1}^{nu} \frac{B_{ij}(z^{-1})}{F(z^{-1})} u_j(k-nk_{ij}-n-m) \right) & p = i \\ 0 & p \neq i \end{cases} \quad (B.69)
 \end{aligned}$$

$$\Rightarrow H_{cd,*i}(\underline{\theta}) \triangleq \frac{-1}{C_{ii}(z^{-1})} J_{d,*i}(\underline{\theta}) \quad , \quad H_{cd}(\underline{\theta}) \in \mathbb{R}^{N \times ny} \quad (B.70)$$

$$\begin{aligned}
 \frac{\partial^2 \tilde{e}_i(k, \underline{\theta})}{\partial c_{pp,m} \partial f_n} &= \frac{\partial}{\partial c_{pp,m}} \left[ L_i(z^{-1}) \frac{D(z^{-1})}{C_{ii}(z^{-1}) F^2(z^{-1})} \sum_{j=1}^{nu} B_{ij}(z^{-1}) u_j(k-nk_{ij}-n) \right] \\
 &= \begin{cases} L_i(z^{-1}) \frac{-D(z^{-1})}{C_{ii}^2(z^{-1}) F^2(z^{-1})} \sum_{j=1}^{nu} B_{ij}(z^{-1}) u_j(k-nk_{ij}-n-m) & p = i \\ 0 & p \neq i \end{cases} \quad (B.71)
 \end{aligned}$$

$$\Rightarrow H_{cf,*i}(\underline{\theta}) \triangleq \frac{-1}{C_{ii}(z^{-1})} J_{f,*i}(\underline{\theta}) \quad , \quad H_{cf}(\underline{\theta}) \in \mathbb{R}^{N \times ny} \quad (B.72)$$

$$\frac{\partial^2 \tilde{e}_i(\underline{\theta})}{\partial d_m \partial d_n} = \frac{\partial}{\partial d_m} \left[ L_i(z^{-1}) \frac{1}{C_{ii}(z^{-1})} \left( A(z^{-1}) y_i(k-n) - \sum_{j=1}^{nu} \frac{B_{ij}(z^{-1})}{F(z^{-1})} u_j(k-nk_{ij}-n) \right) \right] = 0 \quad (\text{B.73})$$

$$\Rightarrow H_{dd}(\underline{\theta}) \triangleq 0 \quad (\text{B.74})$$

$$\begin{aligned} \frac{\partial^2 \tilde{e}_i(k, \underline{\theta})}{\partial d_m \partial f_n} &= \frac{\partial}{\partial d_m} \left[ L_i(z^{-1}) \frac{D(z^{-1})}{C_{ii}(z^{-1}) F^2(z^{-1})} \sum_{j=1}^{nu} B_{ij}(z^{-1}) u_j(k-nk_{ij}-n) \right] \\ &= \begin{cases} L_i(z^{-1}) \sum_{j=1}^{nu} \frac{B_{ij}(z^{-1})}{C_{ii}(z^{-1}) F^2(z^{-1})} u_j(k-nk_{ij}-n-m) & p = i \\ 0 & p \neq i \end{cases} \end{aligned} \quad (\text{B.75})$$

$$\Rightarrow H_{df, *i}(\underline{\theta}) \triangleq L_i(z^{-1}) \frac{1}{C_{ii}(z^{-1}) F(z^{-1})} W_{*i}(\underline{\theta}) \quad , \quad H_{df}(\underline{\theta}) \in \mathbb{R}^{N \times ny} \quad (\text{B.76})$$

$$\begin{aligned} \frac{\partial^2 \tilde{e}_i(k, \underline{\theta})}{\partial f_m \partial f_n} &= \frac{\partial}{\partial f_m} \left[ L_i(z^{-1}) \frac{D(z^{-1})}{C_{ii}(z^{-1}) F^2(z^{-1})} \sum_{j=1}^{nu} B_{ij}(z^{-1}) u_j(k-nk_{ij}-n) \right] \\ &= \begin{cases} L_i(z^{-1}) \frac{-2 D(z^{-1})}{C_{ii}(z^{-1}) F^3(z^{-1})} \sum_{j=1}^{nu} B_{ij}(z^{-1}) u_j(k-nk_{ij}-n-m) & p = i \\ 0 & p \neq i \end{cases} \end{aligned} \quad (\text{B.77})$$

$$\Rightarrow H_{ff,*i}(\underline{\theta}) \triangleq \frac{-2}{F(z^{-1})} J_{f,*i}(\underline{\theta}) \quad , \quad H_{ff}(\underline{\theta}) \in \mathbb{R}^{N \times ny} \quad (B.78)$$

The implementation in MATLAB of this minimal polynomial identification procedure is quite complex. Some simple examples can be used to verify the correctness and numerical accuracy of the algorithms.

### Example

The following true process and noise model :

$$G_t(z^{-1}) = \frac{1.5 z^{-1}}{1 - 0.9 z^{-1}} \quad , \quad H_t(z^{-1}) = \frac{1 + 0.5 z^{-1}}{1 - 0.7 z^{-1}} \quad (B.79)$$

have been used to generate identification data according to :

$$\begin{aligned} \text{OE} : \quad & \begin{cases} y_{oe}^{(i)}(k) = G_t(z^{-1}) u(k) + \xi_{oe}^{(i)}(k) = y_s(k) + \xi_{oe}^{(i)}(k) \\ \sigma_{y_s} = 3.4034 \quad , \quad \sigma_{\xi_{oe}^{(1)}} = 0 \quad , \quad \sigma_{\xi_{oe}^{(2)}} = 3.4034e-1 \end{cases} \\ \text{BJ} : \quad & \begin{cases} y_{bj}(k) = G_t(z^{-1}) u(k) + H_t(z^{-1}) \xi_{bj}(k) = y_s(k) + e_s(k) \\ \sigma_{e_s} = \sigma_{y_s} = 3.4034 \quad , \quad \sigma_{\xi_{bj}} = 1.9284 \end{cases} \end{aligned} \quad (B.80)$$

where a PRBN sequence has been used as input signal ,  $u(k) \in [-1,1]$ ,  $N = 1023$  samples, and normally distributed noise sequences  $\xi_{oe}^{(i)}(k)$  and  $\xi_{bj}(k)$  .

The identification results of these 3 examples are depicted in the Tables B.2 and B.3 for the OE and the BJ examples respectively.

**Table B.2** : Output-error estimation.

$\underline{\theta}_t$	SNR = $\infty$		SNR = 20 dB ( $\sigma_{pe} = 3.3991e-1$ )	
	$\hat{\underline{\theta}}$	%*	$\hat{\underline{\theta}}$	%*
1.5	1.5000e+0	0	1.4963e+0	0.25
-0.9	-9.0000e -1	0	-9.0067e -1	0.07

\*) Relative uncertainty :  $100 |(\hat{\underline{\theta}} - \underline{\theta}_t) / \underline{\theta}_t|$

**Table B.3** : Box-Jenkins estimation,  $\sigma_{pe} = 1.9206$ 

$\underline{\theta}_t$	$\hat{\underline{\theta}}$	%*
1.5	1.5072e+0	0.48
0.5	5.1012e -1	2.02
-0.7	-6.2799e -1	10.29
-0.9	-9.0844e -1	0.94

\*) Relative uncertainty :  $100 |(\hat{\underline{\theta}} - \underline{\theta}_t) / \underline{\theta}_t|$

These simulation examples show clearly the accuracy of the algorithms. Note however, that the exact estimation of the parameters in the noise free output-error example is a coincidence which occurred because the tolerance margins (see Eq. B.31) have been defined sufficiently small ( $\gamma_1 = \gamma_2 = 1e-4$ ). In general, numerical round-off errors will limit the accuracy of the solutions. When noise is included, the parameter estimates depend highly on the signal-to-noise ratio and how this noise affects the dynamical system. Especially for the Box-Jenkins simulation example, the estimate of the  $d_1$ -parameter is less accurate. However, the standard deviation of the corresponding prediction error,  $\sigma_{pe} = 1.9206$ , is smaller compared to the noise which has been used to generate the data,  $\sigma_{\xi_{bj}} = 1.9284$ . So prediction error minimization of the data set generated with the Box-Jenkins system converges to the global minimum, but obviously there exists a parameter estimate which can explain the data more accurate using this identification criterion.



# Parameter Covariance

---

C.1 Covariance Approximation

C.2 Confidence Computation

---

A quality measure of the parameter estimates can be obtained by estimating the covariance matrix of the parameter vector  $\underline{\theta}$  as well. The problem, however, is that in general the assumption is required that the process is contained in the model set, defined by the minimal polynomial structure in Eq. 2.17. As described in the introduction this is not a valid assumption in practice and therefore the standard expressions for the covariance matrix cannot be used. Nevertheless, the estimate  $\hat{\underline{\theta}}$  converges to a minimum point of the loss-function  $V_e(\underline{\theta})$ . When the number of data samples  $N$  tends to infinity, there holds :

$$\hat{\underline{\theta}} \rightarrow \underline{\theta}^* \triangleq \arg \min_{\underline{\theta} \in \Theta_{pe}} V_{e_{\infty}}(\underline{\theta}) \quad , \quad N \rightarrow \infty \quad (C.1)$$

which indicates the best possible approximation of the system within the selected model set. The parameter vector  $\underline{\theta}^*$  is by definition such that the prediction error  $\underline{e}(k, \underline{\theta})$  has a variance as small as possible. Taking this undermodelling explicitly into account, an expression for the parameter covariance matrix  $\text{Cov}(\hat{\underline{\theta}})$  will be derived which can be used to compute confidence intervals.

### C.1 Covariance Approximation

Under the assumption that the process is not contained in the model set, an expression of the covariance matrix can be derived as follows. Consider first the noise free case :

$$V_{\varepsilon}(\underline{\theta}) = \frac{1}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} [y_{\text{col}}(\ell) - \hat{y}_{\text{col}}(\ell, \underline{\theta})]^2 \triangleq \frac{1}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \varepsilon_{\text{col}}^2(\ell, \underline{\theta}) \quad (\text{C.2})$$

where the subscript col indicates a column vector which stores the data as described in Appendix B. A Taylor series expansion of the noise free loss-function around  $\underline{\theta}^*$  is given by :

$$\begin{aligned} V_{\varepsilon}(\underline{\theta}) &= V_{\varepsilon}(\underline{\theta}^*) + \nabla_{\underline{\theta}}^T V_{\varepsilon}(\underline{\theta}^*) (\underline{\theta} - \underline{\theta}^*) + \frac{1}{2} (\underline{\theta} - \underline{\theta}^*)^T \nabla_{\underline{\theta}\underline{\theta}}^2 V_{\varepsilon}(\underline{\theta}^*) (\underline{\theta} - \underline{\theta}^*) + O^3 \\ &= V_{\varepsilon}(\underline{\theta}^*) + \frac{1}{2} (\underline{\theta} - \underline{\theta}^*)^T \nabla_{\underline{\theta}\underline{\theta}}^2 V_{\varepsilon}(\underline{\theta}^*) (\underline{\theta} - \underline{\theta}^*) + O^3 \end{aligned} \quad (\text{C.3})$$

since there holds that for the minimum  $\underline{\theta}^*$  of the loss-function  $V_{\varepsilon}(\underline{\theta}^*) : \nabla_{\underline{\theta}} V_{\varepsilon}(\underline{\theta}^*) = 0$ , when  $N$  tends to infinity.

When including noise, the prediction error  $e_i(k, \underline{\theta})$  will still consist of a term  $\varepsilon_i(k, \underline{\theta})$  due to the undermodelling and in addition a white noise term  $\xi_i(k)$  :

$$e_i(k, \underline{\theta}) = \varepsilon_i(k, \underline{\theta}) + \xi_i(k) \quad (\text{C.4})$$

where  $\xi_i(k)$  is supposed to be a Gaussian white noise sequence with zero mean corresponding to the  $i^{\text{th}}$  output. The loss-function  $V_e(\underline{\theta})$  can then be written as :

$$\begin{aligned} V_e(\underline{\theta}) &\triangleq \frac{1}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} e_{\text{col}}^2(\ell, \underline{\theta}) = \frac{1}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} [\varepsilon_{\text{col}}(\ell, \underline{\theta}) + \xi_{\text{col}}(\ell)]^2 \\ &= V_{\varepsilon}(\underline{\theta}) + \frac{1}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \xi_{\text{col}}^2(\ell) + \frac{2}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \varepsilon_{\text{col}}(\ell, \underline{\theta}) \xi_{\text{col}}(\ell) \end{aligned} \quad (\text{C.5})$$

Due to the fact that  $\varepsilon_{\text{col}}(\ell, \underline{\theta})$  in the last term is parameter dependent, the minimum of the loss-function  $V_e(\underline{\theta})$  will not converge to  $\underline{\theta}^*$  :

$$\hat{\underline{\theta}} \triangleq \arg \min_{\underline{\theta} \in \Theta_{\text{pe}}} V_e(\underline{\theta}) \neq \underline{\theta}^* \quad (\text{C.6})$$

But if  $N$  tends to infinity, the last term in Eq. C.5 can be neglected since  $\xi_{\text{col}}(\ell)$  consists of white noise sequences which are assumed to be uncorrelated with  $\varepsilon_{\text{col}}(\ell, \underline{\theta})$ .

Therefore, the expectation satisfies :

$$E \left\{ \frac{2}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \epsilon_{\text{col}}(\ell, \underline{\theta}) \xi_{\text{col}}(\ell) \right\} = 0 \quad , \quad N \rightarrow \infty \quad (\text{C.7})$$

So, the parameter estimate will still asymptotically converge to the optimal value :

$$\text{plim}_{N \rightarrow \infty} \hat{\underline{\theta}} \rightarrow \underline{\theta}^* \quad (\text{C.8})$$

A Taylor series expansion of  $\epsilon_{\text{col}}(\ell, \underline{\theta})$  around  $\underline{\theta}^*$  gives :

$$\begin{aligned} \epsilon_{\text{col}}(\ell, \underline{\theta}) &= \epsilon_{\text{col}}(\ell, \underline{\theta}^*) + \nabla_{\underline{\theta}}^T \epsilon_{\text{col}}(\ell, \underline{\theta}^*) (\underline{\theta} - \underline{\theta}^*) + \\ &\quad \frac{1}{2} (\underline{\theta} - \underline{\theta}^*)^T \nabla_{\underline{\theta}\underline{\theta}}^2 \epsilon_{\text{col}}(\ell, \underline{\theta}^*) (\underline{\theta} - \underline{\theta}^*) + O^3 \end{aligned} \quad (\text{C.9})$$

Substitution of Eq. C.9 into Eq. C.2 then results in :

$$\begin{aligned} V_{\epsilon}(\underline{\theta}) &= \frac{1}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \epsilon_{\text{col}}^2(\ell, \underline{\theta}) \\ &= \frac{1}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \left[ \epsilon_{\text{col}}(\ell, \underline{\theta}^*) + \nabla_{\underline{\theta}}^T \epsilon_{\text{col}}(\ell, \underline{\theta}^*) (\underline{\theta} - \underline{\theta}^*) + \right. \\ &\quad \left. \frac{1}{2} (\underline{\theta} - \underline{\theta}^*)^T \nabla_{\underline{\theta}\underline{\theta}}^2 \epsilon_{\text{col}}(\ell, \underline{\theta}^*) (\underline{\theta} - \underline{\theta}^*) + O^3 \right]^2 \quad (\text{C.10}) \\ &= \frac{1}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \left[ \epsilon_{\text{col}}^2(\ell, \underline{\theta}^*) + 2 \epsilon_{\text{col}}(\ell, \underline{\theta}^*) \nabla_{\underline{\theta}}^T \epsilon_{\text{col}}(\ell, \underline{\theta}^*) (\underline{\theta} - \underline{\theta}^*) + \right. \\ &\quad \epsilon_{\text{col}}(\ell, \underline{\theta}^*) (\underline{\theta} - \underline{\theta}^*)^T \nabla_{\underline{\theta}\underline{\theta}}^2 \epsilon_{\text{col}}(\ell, \underline{\theta}^*) (\underline{\theta} - \underline{\theta}^*) + \\ &\quad \left. (\underline{\theta} - \underline{\theta}^*)^T \nabla_{\underline{\theta}} \epsilon_{\text{col}}(\ell, \underline{\theta}^*) \nabla_{\underline{\theta}}^T \epsilon_{\text{col}}(\ell, \underline{\theta}^*) (\underline{\theta} - \underline{\theta}^*) + O^3 \right] \end{aligned}$$

Comparing Eq. C.10 with Eq. C.3 results in the following expressions for the loss-function and its derivatives :

$$V_{\epsilon}(\underline{\theta}^*) = \frac{1}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \epsilon_{\text{col}}^2(\ell, \underline{\theta}^*)$$

$$\begin{aligned} \nabla_{\theta} V_{\varepsilon}(\underline{\theta}^*) &= \frac{2}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \varepsilon_{\text{col}}(\ell, \underline{\theta}^*) \nabla_{\theta}^T \varepsilon_{\text{col}}(\ell, \underline{\theta}^*) = 0 \quad , \quad N \rightarrow \infty \\ \nabla_{\theta\theta}^2 V_{\varepsilon}(\underline{\theta}^*) &= \frac{2}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \left[ \varepsilon_{\text{col}}(\ell, \underline{\theta}^*) \nabla_{\theta}^T \varepsilon_{\text{col}}(\ell, \underline{\theta}^*) + \varepsilon_{\text{col}}(\ell, \underline{\theta}^*) \nabla_{\theta\theta}^2 \varepsilon_{\text{col}}(\ell, \underline{\theta}^*) \right] \end{aligned} \quad (\text{C.10})$$

The parameter estimate  $\hat{\underline{\theta}}$  can be derived by calculating :  $\nabla_{\theta} V_{\varepsilon}(\hat{\underline{\theta}}) = 0$ . Substituting Eq. C.3 and C.9 into Eq. C.5, this gradient can be written as :

$$\begin{aligned} \nabla_{\theta} V_{\varepsilon}(\underline{\theta}) &= \nabla_{\theta\theta}^2 V_{\varepsilon}(\underline{\theta}^*) (\underline{\theta} - \underline{\theta}^*) + \\ &\quad \frac{2}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \left[ \nabla_{\theta} \varepsilon_{\text{col}}(\ell, \underline{\theta}^*) + \nabla_{\theta\theta}^2 \varepsilon_{\text{col}}(\ell, \underline{\theta}^*) (\underline{\theta} - \underline{\theta}^*) \right] \xi_{\text{col}}(\ell) + O^2 \end{aligned} \quad (\text{C.12})$$

which yields :

$$\begin{aligned} &\left[ \nabla_{\theta\theta}^2 V_{\varepsilon}(\underline{\theta}^*) + \frac{2}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \nabla_{\theta\theta}^2 \varepsilon_{\text{col}}(\ell, \underline{\theta}^*) \xi_{\text{col}}(\ell) \right] \hat{\underline{\theta}} \approx \\ &\left[ \nabla_{\theta\theta}^2 V_{\varepsilon}(\underline{\theta}^*) + \frac{2}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \nabla_{\theta\theta}^2 \varepsilon_{\text{col}}(\ell, \underline{\theta}^*) \xi_{\text{col}}(\ell) \right] \underline{\theta}^* - \frac{2}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \nabla_{\theta} \varepsilon_{\text{col}}(\ell, \underline{\theta}^*) \xi_{\text{col}}(\ell) \end{aligned} \quad (\text{C.13})$$

where the higher order terms  $O^2$  are neglected, or :

$$\begin{aligned} \hat{\underline{\theta}} &= \underline{\theta}^* - \left[ \nabla_{\theta\theta}^2 V_{\varepsilon}(\underline{\theta}^*) + \frac{2}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \nabla_{\theta\theta}^2 \varepsilon_{\text{col}}(\ell, \underline{\theta}^*) \xi_{\text{col}}(\ell) \right]^{-1} \frac{2}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \nabla_{\theta} \varepsilon_{\text{col}}(\ell, \underline{\theta}^*) \xi_{\text{col}}(\ell) \\ &\approx \underline{\theta}^* - \left[ \nabla_{\theta\theta}^2 V_{\varepsilon}(\underline{\theta}^*) \right]^{-1} \frac{2}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \nabla_{\theta} \varepsilon_{\text{col}}(\ell, \underline{\theta}^*) \xi_{\text{col}}(\ell) \\ &=: \underline{\theta}^* - \underline{\eta} \end{aligned} \quad (\text{C.14})$$

where the approximation is due to the fact that the white noise sequence  $\xi_{\text{col}}(\ell)$  is assumed to be independent of the second derivative matrix  $\nabla_{\theta\theta}^2 \varepsilon_{\text{col}}(\ell, \underline{\theta}^*)$  and therefore this term tends to zero for  $N \rightarrow \infty$ . Because asymptotically the parameter estimate converges to the optimal value  $\underline{\theta}^*$  (Eq. C.8), the asymptotic covariance matrix is defined

by :

$$\text{Cov}(\underline{\theta}^*) = E \left\{ (\underline{\hat{\theta}} - \underline{\theta}^*)(\underline{\hat{\theta}} - \underline{\theta}^*)^T \right\} = E \left\{ \underline{n} \underline{n}^T \right\} \quad (\text{C.15})$$

which can be rewritten in the following form by substituting  $\underline{n}$  (Eq. C.14) :

$$\begin{aligned} \text{Cov}(\underline{\theta}^*) &= E \left\{ \left[ \nabla_{\theta\theta}^2 V_{\epsilon}(\underline{\theta}^*) \right]^{-1} \frac{4}{N_{\text{col}}^2} \sum_{\ell=1}^{N_{\text{col}}} \sum_{s=1}^{N_{\text{col}}} \left[ \nabla_{\theta} \epsilon_{\text{col}}(\ell, \underline{\theta}^*) \xi_{\text{col}}(\ell) \xi_{\text{col}}(s) \nabla_{\theta}^T \epsilon_{\text{col}}(s, \underline{\theta}^*) \right] \right. \\ &\quad \left. \left[ \nabla_{\theta\theta}^2 V_{\epsilon}(\underline{\theta}^*) \right]^{-1} \right\} \\ &= \sigma_{\xi}^2 \left[ \nabla_{\theta\theta}^2 V_{\epsilon}(\underline{\theta}^*) \right]^{-1} \frac{4}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \left[ \nabla_{\theta} \epsilon_{\text{col}}(\ell, \underline{\theta}^*) \nabla_{\theta}^T \epsilon_{\text{col}}(\ell, \underline{\theta}^*) \right] \left[ \nabla_{\theta\theta}^2 V_{\epsilon}(\underline{\theta}^*) \right]^{-1} \end{aligned} \quad (\text{C.16})$$

Since asymptotically the noisy derivatives will be independent of the white noise components affecting the data, this covariance matrix can be approximated in practice by replacing the first and second order-derivatives by those determined from the measured data resulting in :

$$\begin{aligned} \text{Cov}(\underline{\hat{\theta}}) &= \sigma_{\xi}^2 \left[ \frac{2}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \nabla_{\theta} \epsilon_{\text{col}}(\ell, \underline{\hat{\theta}}) \nabla_{\theta}^T \epsilon_{\text{col}}(\ell, \underline{\hat{\theta}}) + \epsilon_{\text{col}}(\ell, \underline{\hat{\theta}}) \nabla_{\theta\theta}^2 \epsilon_{\text{col}}(\ell, \underline{\hat{\theta}}) \right]^{-1} * \\ &\quad \left[ \frac{4}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \nabla_{\theta} \epsilon_{\text{col}}(\ell, \underline{\hat{\theta}}) \nabla_{\theta}^T \epsilon_{\text{col}}(\ell, \underline{\hat{\theta}}) \right] * \\ &\quad \left[ \frac{2}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \nabla_{\theta} \epsilon_{\text{col}}(\ell, \underline{\hat{\theta}}) \nabla_{\theta}^T \epsilon_{\text{col}}(\ell, \underline{\hat{\theta}}) + \epsilon_{\text{col}}(\ell, \underline{\hat{\theta}}) \nabla_{\theta\theta}^2 \epsilon_{\text{col}}(\ell, \underline{\hat{\theta}}) \right]^{-1} \end{aligned} \quad (\text{C.17})$$

which can be considerably simplified by substituting Eq. B.30 :

$$\text{Cov}(\underline{\hat{\theta}}) = 2 \sigma_{\xi}^2 \left[ \mathbf{R}(\underline{\hat{\theta}}) + \mathbf{S}(\underline{\hat{\theta}}) \right]^{-1} \mathbf{R}(\underline{\hat{\theta}}) \left[ \mathbf{R}(\underline{\hat{\theta}}) + \mathbf{S}(\underline{\hat{\theta}}) \right]^{-1} \quad (\text{C.18})$$

where

$$R(\hat{\underline{\theta}}) := \frac{2}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} \nabla_{\underline{\theta}} e_{\text{col}}(\ell, \hat{\underline{\theta}}) \nabla_{\underline{\theta}}^T e_{\text{col}}(\ell, \hat{\underline{\theta}}) = \frac{2}{N_{\text{col}}} \mathbf{J}^T(\hat{\underline{\theta}}) \mathbf{J}(\hat{\underline{\theta}}) \quad (\text{C.19})$$

$$S(\hat{\underline{\theta}}) := \frac{2}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} e_{\text{col}}(\ell, \hat{\underline{\theta}}) \nabla_{\underline{\theta}\underline{\theta}}^2 e_{\text{col}}(\ell, \hat{\underline{\theta}}) = \frac{2}{N_{\text{col}}} \sum_{\ell=1}^{N_{\text{col}}} e_{\text{col}}(\ell, \hat{\underline{\theta}}) \mathbf{H}_{\ell}(\hat{\underline{\theta}})$$

It has been shown in Appendix B that the Gradient  $\mathbf{J}(\hat{\underline{\theta}})$  and the Hessian  $\mathbf{H}(\hat{\underline{\theta}})$  of the prediction error, the first and second order derivatives with respect to the parameters, can be calculated analytically in an efficient way. However,  $\sigma_{\xi}^2$  has to be approximated. If the system is contained in the model set,  $\sigma_{\xi}^2$  could easily be obtained from the criterion  $V_e(\hat{\underline{\theta}}) \approx \sigma_{\xi}^2$  for  $N \rightarrow \infty$ . When the system is not contained in the model set, there holds :

$$\lim_{N \rightarrow \infty} V_e(\hat{\underline{\theta}}) = \lim_{N \rightarrow \infty} \left\{ \frac{N_{\text{col}} - n\theta}{N_{\text{col}}} \sigma_{\xi}^2 + V_{\varepsilon}(\hat{\underline{\theta}}) \right\} \approx \sigma_{\xi}^2 + V_{\varepsilon}(\hat{\underline{\theta}}) \quad (\text{C.20})$$

As assumed in Eq. C.4 the prediction error will consist of a correlated sequence  $\varepsilon_i(k, \underline{\theta})$  and a white noise sequence  $\xi_i(k)$ . To obtain an estimate  $\hat{\sigma}_{\xi}^2$  of the white noise variance, estimates of the innovations  $\xi_i(k)$  have to be formed which can be computed by applying a high order AR-model to  $e_i(k, \hat{\underline{\theta}})$  :

$$A_h(z^{-1}) e_i(k, \hat{\underline{\theta}}) = \xi_i(k) \quad (\text{C.21})$$

The reason for this approximation is twofold :

- The scaling of the data should be verified to validate the assumption :  $\hat{\sigma}_{\xi_1}^2 \approx \dots \approx \hat{\sigma}_{\xi_{n_y}}^2$  (Eq. 2.21, equal variance of the white noise sequences).
- An upper bound estimate of the true parameter covariance can be obtained, since  $\hat{\sigma}_{\xi}^2 \geq \sigma_{\xi}^2$  implies that  $\text{Cov}(\hat{\underline{\theta}}) \geq \text{Cov}(\underline{\theta}^*)$ . When constructing a white noise sequence  $\xi_i(k)$  of the prediction error  $e_i(k, \hat{\underline{\theta}})$  according to Eq. C.21, the correlation at time lag  $\tau = 0$  is completely assigned to the white noise sequence. Under the assumption of Eq. C.4, however, the auto-correlation of the prediction error at time lag  $\tau = 0$  is built of both signals,  $\varepsilon_i(k, \underline{\theta})$  and  $\xi_i(k)$ , and therefore the approximation of the white noise variance is always larger than or equal to the true variance (see Eq. C.20).

In practice, however, it should be verified that  $N_{\text{col}} \gg n\theta$  in order to guarantee that the approximation in Eq. C.20 is valid, i.e.  $(N_{\text{col}} - n\theta)/N_{\text{col}} \rightarrow 1$ .

## C.2 Confidence Computation

As described in Section 2.5, confidence ellipsoids of the parameter estimate  $\hat{\underline{\theta}}$  are described by :

$$z_{\underline{\theta}} = N_{\text{col}} \left( \hat{\underline{\theta}} - \underline{\theta}^* \right)^T \text{Cov}(\hat{\underline{\theta}}) \left( \hat{\underline{\theta}} - \underline{\theta}^* \right) \quad (\text{C.22})$$

where  $z_{\underline{\theta}}$  converges in distribution to a  $\chi^2$ -distribution with  $n\theta$  degrees of freedom as the number of data samples  $N$ , and therefore also  $N_{\text{col}}$ , tends to infinity. The  $\chi^2$ -distribution is defined as [Kre93] :

$$F(z_{\underline{\theta}}) = C_{n\theta} \int_0^{z_{\underline{\theta}}} e^{(-1/2 x)} x^{(1/2 n\theta - 1)} dx \quad (\text{C.23})$$

where

$$C_{n\theta} = \frac{1}{2^{1/2 n\theta} \Gamma(1/2 n\theta)} \quad (\text{C.24})$$

and the gamma function satisfies :

$$\begin{aligned} \Gamma(n+1) &= n! & \Gamma(n+1/2) &= \frac{(1.3.5 \dots 2n-1)}{2^n} \Gamma(1/2) \\ \Gamma(2) &= \Gamma(1) = 1 & \Gamma(1/2) &= \sqrt{\pi} \end{aligned} \quad (\text{C.25})$$

Repeated partial integration of the integral defined in Eq. C.23 results in :

$$\begin{aligned} \int_0^{z_{\underline{\theta}}} e^{-1/2 x} x^{(1/2 n\theta - 1)} dx &= \frac{e^{-1/2 x} x^{(1/2 n\theta - 1)}}{-1/2} - \frac{1/2 n\theta - 1}{-1/2} \int_0^{z_{\underline{\theta}}} e^{-1/2 x} x^{(1/2 n\theta - 2)} dx \\ &= \frac{e^{-1/2 x} x^{(1/2 n\theta - 1)}}{-1/2} - \\ &\quad \frac{1/2 n\theta - 1}{-1/2} \left[ \frac{e^{-1/2 x} x^{(1/2 n\theta - 2)}}{-1/2} - \frac{1/2 n\theta - 2}{-1/2} \int_0^{z_{\underline{\theta}}} e^{-1/2 x} x^{(1/2 n\theta - 3)} dx \right] \end{aligned}$$

$$\begin{aligned}
 &= e^{-\frac{1}{2}x} \sum_{r=0}^m (-1)^r \left[ \prod_{k=0}^r (\frac{1}{2}n\theta - k - 1) \right] \frac{x^{\frac{1}{2}n\theta - r - 1}}{(-\frac{1}{2})^{r+1}} \Bigg|_0^{z_\theta} + \\
 &(-1)^{m+1} \prod_{k=0}^m \frac{(\frac{1}{2}n\theta - k - 1)}{(-\frac{1}{2})^{m+1}} \int_0^{z_\theta} x^{-\frac{1}{2}} e^{-\frac{1}{2}x} dx
 \end{aligned} \quad (C.26)$$

where  $m$  is defined as the smallest integer of  $\frac{1}{2}n\theta - 1$ . The integral to compute the  $\chi^2$ -distribution can now be written in a more compact form :

$$\begin{aligned}
 \frac{F(z_\theta)}{C_{n\theta}} &= \sum_{r=1}^m (-1)^r \left[ \prod_{k=0}^{r-1} (\frac{1}{2}n\theta - r - 1) \right] \frac{e^{-\frac{1}{2}z_\theta} z_\theta^{\frac{1}{2}n\theta - r - 1} - 0^{\frac{1}{2}n\theta - r - 1}}{(-\frac{1}{2})^{r+1}} \Bigg|_{n\theta > 1} \\
 &+ \frac{(-1)^{n+1} \prod_{k=0}^n (\frac{1}{2}n\theta - k - 1)}{(-\frac{1}{2})^{n+1}} \int_0^{z_\theta} x^{-\frac{1}{2}} e^{-\frac{1}{2}x} dx \Bigg|_{n\theta \text{ odd}}
 \end{aligned} \quad (C.27)$$

The problem has been solved for an even number degrees of freedom. However, when  $n\theta$  is odd, the remaining integral must be computed :

$$I = \int_0^{z_\theta} e^{-\frac{1}{2}x} x^{-\frac{1}{2}} dx \quad (C.28)$$

which can be rewritten in a more convenient form by substituting  $x = 2y^2$  :

$$I = \int_0^{(\frac{1}{2}z_\theta)^{\frac{1}{2}}} \frac{e^{-y^2}}{y\sqrt{2}} 4y dy = 2\sqrt{2} \int_0^{(\frac{1}{2}z_\theta)^{\frac{1}{2}}} e^{-y^2} dy \quad (C.29)$$

Using now power series expansions :

$$e^{-y} = \sum_{k=0}^{\infty} \frac{1}{k!} y^k \quad \Rightarrow \quad e^{-y^2} = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} y^{2k} \quad (C.30)$$

this integral can be written as an infinite sum according to :



$$\begin{aligned}
 I &= 2\sqrt{2} \int_0^{(\frac{1}{2}z_\theta)^{1/2}} e^{-y^2} dy = 2\sqrt{2} \int_0^{(\frac{1}{2}z_\theta)^{1/2}} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} y^{2k} \\
 &= 2\sqrt{2} \left[ \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)k!} y^{2k+1} \right]_0^{(\frac{1}{2}z_\theta)^{1/2}} \\
 &= 2\sqrt{2} \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)k!} (\frac{1}{2}z_\theta)^{k+1/2}
 \end{aligned} \tag{C.31}$$

In practice, this integral can be approximated accurately by a finite sum of elements, because the term  $(2k+1)k!$  ensures that the elements of the sum will decrease rapidly for increasing values of  $k$ .

It is now possible to compute the confidence  $F(z_\theta)$  for a given integration range  $z_\theta$  of the  $\chi^2$ -distribution, exact for even  $n\theta$  and approximate for odd  $n\theta$ . The confidence ellipsoids for the parameter estimates are scaled, however, by  $z_\theta$  for a given  $F(z_\theta)$ . Therefore, this scaling factor  $z_\theta$  has to be computed iteratively, e.g. by bisection search within a predefined interval, to determine the required confidence  $F(z_\theta)$ .



## ***Parametric Uncertainty Descriptions***

---

D.1	Multivariable Bounded Error Models	D.3	Ellipsoidal Parameter Bounding
D.2	Exact Polytope Updating	D.4	Orthotopic Parameter Bounding

---

In this appendix the bounded error identification approach which has been described in Chapter 3 will be extended to multivariable models. In addition, several algorithms to characterize or approximate the parameter uncertainty set  $\Theta$  will be described in detail.

### ***D.1 Multivariable Bounded Error Models***

For the multivariable minimal polynomial model structure as described in Section 2.2 :

$$A(z^{-1})\underline{y}(k) = \frac{B(z^{-1})}{F(z^{-1})} \underline{u}(k) + \frac{C(z^{-1})}{D(z^{-1})} \xi(k) \quad (D.1)$$

the pseudo-linear regression form is given by :

$$\hat{\underline{y}}_{\text{col}}(\underline{\theta}) = \Phi(\underline{\theta}) \underline{\theta} \quad (D.2)$$

where  $\Phi(\underline{\theta})$  and  $\hat{y}_{col}(\ell, \underline{\theta})$  have been defined in Appendix B.3, Eq. B.17 and Eq. B.18 respectively. The unknown but bounded error identification can be extended to the multivariable case by computing lower and upper bounds of the prediction error  $\underline{e}(k, \underline{\theta})$  and of the auxiliary variables  $\underline{v}(k, \underline{\theta})$  and  $\underline{w}(k, \underline{\theta})$ , e.g.  $\underline{e}_{min}(k)$  and  $\underline{e}_{max}(k)$ . These lower and upper bounds can then be used to compute the centre values and the corresponding uncertainty for all variables :

$$\begin{aligned}
 \underline{e}_c(k) &= \frac{1}{2} \left( \underline{e}_{max}(k) + \underline{e}_{min}(k) \right) \\
 \underline{\delta}_e(k) &= \frac{1}{2} \left( \underline{e}_{max}(k) - \underline{e}_{min}(k) \right) \\
 \underline{v}_c(k) &= \frac{1}{2} \left( \underline{v}_{max}(k) + \underline{v}_{min}(k) \right) \\
 \underline{\delta}_v(k) &= \frac{1}{2} \left( \underline{v}_{max}(k) - \underline{v}_{min}(k) \right) \\
 \underline{w}_c(k) &= \frac{1}{2} \left( \underline{w}_{max}(k) + \underline{w}_{min}(k) \right) \\
 \underline{\delta}_w(k) &= \frac{1}{2} \left( \underline{w}_{max}(k) - \underline{w}_{min}(k) \right)
 \end{aligned} \quad \forall k \quad (D.3)$$

The corresponding conditions in bounded error identification (Eq. 3.14) to this multivariable regression form can now be derived in a straightforward way and are defined by :

$$\begin{aligned}
 \mathbf{H}_1 : \quad \psi_1(\ell, \underline{\theta}) \underline{\theta} &\geq c_1(\ell) \\
 \mathbf{H}_2 : \quad \psi_2(\ell, \underline{\theta}) \underline{\theta} &\leq c_2(\ell)
 \end{aligned} \quad (D.4)$$

where

$$\begin{aligned}
 \psi_1(\ell, \underline{\theta}) &= \phi(\ell, \underline{\theta}) + \Delta\phi(\ell, \underline{\theta}) & c_1(\ell) &= y_{col}(\ell) - \delta_{e,col}(\ell) \\
 \psi_2(\ell, \underline{\theta}) &= \phi(\ell, \underline{\theta}) - \Delta\phi(\ell, \underline{\theta}) & c_2(\ell) &= y_{col}(\ell) + \delta_{e,col}(\ell)
 \end{aligned} \quad (D.5)$$

and  $y_{col}(\ell)$  and  $\delta_{e,col}(\ell)$  describe the elements of the vectors :

$$\begin{aligned}
 \underline{y}_{col} &= \left[ y_1(n+1) \dots y_1(N) \ y_2(n+1) \dots y_{ny}(N) \right]^T \\
 \underline{\delta}_{e,col} &= \left[ \delta_{e,1}(n+1) \dots \delta_{e,1}(N) \ \delta_{e,2}(n+1) \dots \delta_{e,ny}(N) \right]^T
 \end{aligned} \quad (D.6)$$

The rows  $\phi(\ell, \underline{\theta})$  of the regression matrix  $\Phi(\underline{\theta})$ , which has been defined in Appendix B, Eq. B.17, contain the input/output data together with the prediction error and the auxiliary signals where the various data samples, e.g.  $w(k-1, \underline{\theta})$ , are replaced by the

corresponding centre values, e.g.  $w_c(k-1)$ . The uncertainty regression matrix  $\Delta\Phi(\underline{\theta})$  which consists of rows  $\Delta\phi(\ell, \underline{\theta})$ , can be derived in a similar way to  $\Phi(\underline{\theta})$  by introducing the following sub-matrices with respect to the several data and auxiliary signals :

$$\Delta\phi_{y_i} = \begin{bmatrix} 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{bmatrix}, \quad \Delta\phi_{u_{ij}} = \begin{bmatrix} 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{bmatrix}$$

$$\Delta\phi_{e_i}(\underline{\theta}) = \begin{bmatrix} \text{sgn}(c_{ii,1}) \delta_{e_i}(k-1) & \dots & \text{sgn}(c_{ii,nc_{ii}}) \delta_{e_i}(k-nc_{ii}) \\ \vdots & & \vdots \\ \text{sgn}(c_{ii,1}) \delta_{e_i}(k-1) & \dots & \text{sgn}(c_{ii,nc_{ii}}) \delta_{e_i}(k-nc_{ii}) \end{bmatrix} \quad (D.7)$$

$$\Delta\phi_{v_i}(\underline{\theta}) = \begin{bmatrix} \text{sgn}(d_1) \delta_{v_i}(k-1) & \dots & \text{sgn}(d_{nd}) \delta_{v_i}(k-nd) \\ \vdots & & \vdots \\ \text{sgn}(d_1) \delta_{v_i}(k-1) & \dots & \text{sgn}(d_{nd}) \delta_{v_i}(k-nd) \end{bmatrix}$$

$$\Delta\phi_{w_i}(\underline{\theta}) = \begin{bmatrix} \text{sgn}(f_1) \delta_{w_i}(k-1) & \dots & \text{sgn}(f_{nf}) \delta_{w_i}(k-nf) \\ \vdots & & \vdots \\ \text{sgn}(f_1) \delta_{w_i}(k-1) & \dots & \text{sgn}(f_{nf}) \delta_{w_i}(k-nf) \end{bmatrix}$$

which then results in :

$$\Delta\Phi(\underline{\theta}) = \begin{bmatrix} \Delta\phi_{y_1} & | & \Delta\phi_{u_{11}} & \dots & \Delta\phi_{u_{1nu}} & 0 & \dots & 0 & \dots & 0 \\ \Delta\phi_{y_2} & | & 0 & \dots & 0 & \Delta\phi_{u_{21}} & \dots & \Delta\phi_{u_{2nu}} & \ddots & \vdots \\ \vdots & | & \vdots & & 0 & \dots & 0 & \ddots & & \vdots \\ \vdots & | & \vdots & & & & & \ddots & 0 & \dots & 0 \\ \Delta\phi_{y_{ny}} & | & 0 & & \dots & & & 0 & \Delta\phi_{u_{ny1}} & \dots & \Delta\phi_{u_{nynu}} \end{bmatrix} \quad (D.8)$$

$$\begin{bmatrix} \Delta\phi_{e_1}(\underline{\theta}) & 0 & \dots & 0 & | & \Delta\phi_{v_1}(\underline{\theta}) & | & \Delta\phi_{w_1}(\underline{\theta}) \\ 0 & \Delta\phi_{e_2}(\underline{\theta}) & \ddots & \vdots & | & \Delta\phi_{v_2}(\underline{\theta}) & | & \Delta\phi_{w_2}(\underline{\theta}) \\ \vdots & 0 & \ddots & \vdots & | & \vdots & | & \vdots \\ \vdots & & \ddots & 0 & | & \vdots & | & \vdots \\ 0 & \dots & 0 & \Delta\phi_{e_{ny}}(\underline{\theta}) & | & \Delta\phi_{v_{ny}}(\underline{\theta}) & | & \Delta\phi_{w_{ny}}(\underline{\theta}) \end{bmatrix}$$

The structure of this uncertainty regression matrix  $\Delta\Phi(\underline{\theta})$  is exactly the same as the data

regression matrix  $\Phi(\underline{\theta})$  where only the data sub-matrices have been replaced by the boundary matrices, e.g.  $\phi_{e_i}(\underline{\theta})$  by  $\Delta\phi_{e_i}(\underline{\theta})$ . The construction of these regression matrices for special cases of the multivariable minimal polynomial model structure (equation-error models, output-error models, etc.) is exactly the same as for SISO models described in Section 3.1 and will therefore not be mentioned here.

## D.2 Exact Polytope Updating

The exact description of the parameter uncertainty set  $\Theta$  [Mo89] can be computed by iteratively updating the following information matrices :

- The vertex set ( $M_{vs}$ ) which contains the coordinates of the vertices in column form where the column  $p$  corresponds to the coordinates of vertex  $\underline{v}_p$ .
- The vertex-plane adjacency list ( $M_{vp}$ ) which contains in column  $p$  the indices of the hyperplanes intersecting through vertex  $\underline{v}_p$ .
- The vertex-vertex adjacency list ( $M_{vv}$ ) which contains in column  $p$  the indices of the neighbouring vertices with respect to  $\underline{v}_p$ .

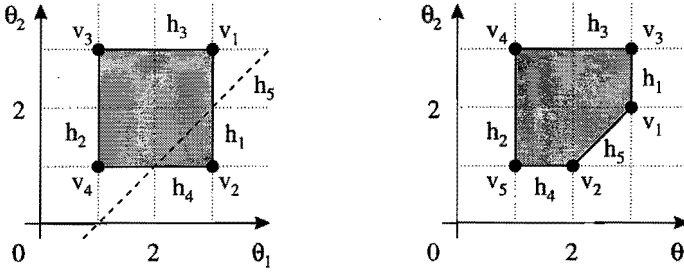
The hyperplanes which describe the boundary of the parameter uncertainty set  $\Theta$  are defined by Eq. D.4 replacing the inequality signs by equalities. A simple example will be used to illustrate the updating in more detail together with the initialization.

### D.2.1 Initial Orthotope

An initial polytope  $\Theta^0$  must be specified before the updating can start. If there is no a priori information on the bounds of the parameters,  $\Theta^0$  can be of any shape as long as its boundary does not partly define the final parameter uncertainty set  $\Theta$ . An orthotope is a suitable simple initial shape which can easily be constructed even for higher dimensional identification problems. This orthotope has  $2^{n\theta}$  vertices which are the combinations of all minimum and maximum values describing the extreme points. As example the following two-dimensional problem ( $n\theta = 2$ ) has been defined :

$$\underline{\theta}^0 = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}, \quad \underline{s}^0 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad (D.9)$$

This initial orthotope is illustrated in Fig. D.1a.



**Fig. D.1** : Exact polytope updating ; a) Initial orthotope, b) Updating step.

The corresponding hyperplanes bounding this orthotope are defined by :

$$\begin{aligned}
 h_1 &: [1 \ 0] \underline{\theta} = 3 \\
 h_2 &: [1 \ 0] \underline{\theta} = 1 \\
 h_3 &: [0 \ 1] \underline{\theta} = 3 \\
 h_4 &: [0 \ 1] \underline{\theta} = 1
 \end{aligned} \tag{D.10}$$

and the construction of the corresponding information matrices for this initial orthotope is because of its simple shape straightforward.

$$\begin{aligned}
 &\underline{v}_1 \quad \underline{v}_2 \quad \underline{v}_3 \quad \underline{v}_4 \\
 M_{vs} &= \begin{bmatrix} 3 & 3 & 1 & 1 \\ 3 & 1 & 3 & 1 \end{bmatrix} && \text{Vertex coordinates} \\
 M_{vp} &= \begin{bmatrix} 1 & 1 & 2 & 2 \\ 3 & 4 & 3 & 4 \end{bmatrix} && \text{Intersecting hyperplanes } h_p \text{ through } \underline{v}_p \\
 M_{vv} &= \begin{bmatrix} 2 & 1 & 1 & 2 \\ 3 & 4 & 4 & 3 \end{bmatrix} && \text{Neighbour vertices of } \underline{v}_p
 \end{aligned} \tag{D.11}$$

Whenever a new constraint is added from the set  $H_1$  (Eq. D.8) which is defined by  $\psi_i(\ell, \underline{\theta}) = c_i(\ell)$ ,  $i = \{1, 2\}$ , e.g. :

$$h_5 : \psi_1(1, \underline{\theta}) \underline{\theta} = c_1(1) \Rightarrow [1 \ -1] \underline{\theta} = 1 \tag{D.12}$$

the vertex set  $M_{vs}$ , the vertex-plane  $M_{vp}$  and vertex-vertex  $M_{vv}$  adjacency lists

(Eq. D.11) have to be updated.

### D.2.2 Updating Vertex Set

For every boundary from the set  $\psi_i(\ell, \underline{\theta}) = c_i(\ell)$ ,  $i = \{1, 2\}$ , which is added to the constraint set, compute :

$$\underline{\beta}^T = c_i(\ell) - \psi_i(\ell, \underline{\theta}) M_{vs} \quad (D.13)$$

where  $\ell$  indicates the selected constraint in the set  $H_i$ . The sign of  $\beta_p$  indicates the position of the new hyperplane with respect to  $\underline{v}_p$  :

- $\beta_p$  positive indicates that  $\underline{v}_p$  is inside  $\Theta$ .
- $\beta_p = 0$  indicates that  $\underline{v}_p$  is on the boundary of  $\Theta$ .
- $\beta_p$  negative indicates that  $\underline{v}_p$  is outside  $\Theta$ .

For a given  $\underline{\beta}$ , corresponding to the constraint to be added in the form of a half space  $H_i$ ,  $i = \{1, 2\}$  (Eq. D.4), the following situations can be distinguished for the updating :

- $H_i \cap \Theta^\ell = \Theta^\ell = \Theta^{\ell+1}$  if all elements of  $\underline{\beta}$  are either positive or zero for which no updating is necessary.
- $H_i \cap \Theta_k = \emptyset$  if all elements of  $\underline{\beta}$  are negative which indicates a violating constraint and suggesting that  $H_i$  is caused by an outlier.
- $H_i \cap \Theta^\ell = \Theta^{\ell+1} \subset \Theta^\ell$  if the elements of  $\underline{\beta}$  have different signs which requires updating of  $M_{vs}$ ,  $M_{vp}$  and  $M_{vv}$ .

Computing  $\underline{\beta}$  for the selected example and  $h_5$  (from Eq. D.12) :

$$\underline{\beta}^T = c_1(1) - \psi_1(1, \underline{\theta}) M_{vs} = \begin{bmatrix} \underline{v}_1 & \underline{v}_2 & \underline{v}_3 & \underline{v}_4 \\ 1 & -1 & 3 & 1 \end{bmatrix} \quad (D.14)$$

indicates that the vertex  $\underline{v}_2$  will be outside the new boundary.

To update the vertex set, all vertices inside the boundary are selected and all vertices outside the boundary will be eliminated. An edge list is constructed having  $n\theta-1$  hyperplanes in common to compute the new vertices which will then be added to the set of selected vertices of the old boundary. This edge list can be constructed by forming all combinations of adjacent vertices from  $M_{vv}$ .

The edge list  $M_{el}$  of the initial orthotope ( $n\theta = 2$ ) is build from the vertex-plane adjacency list by combining all vertices having 1 common hyperplane in each column :



$$M_{el} = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 2 & 3 & 4 & 4 \end{bmatrix} \quad (D.15)$$

When computing  $\beta_p$  for all vertices  $\underline{v}_p$  in the edge list  $M_{el}$ , a changing sign in a column indicates then an intersection. A different sign in the first and second row of  $\beta_{el}$  (Eq. D.16) indicates that one vertex is inside the boundary while the other vertex is outside the boundary. The corresponding vertices are sorted with respect to increasing  $\beta_p$  for each column :

$$\beta_{el} = \begin{bmatrix} -1 & 1 & -1 & 1 \\ 1 & 3 & 1 & 3 \end{bmatrix}, \quad M_{el}^* = \begin{bmatrix} 2 & 1 & 2 & 4 \\ 1 & 3 & 4 & 3 \end{bmatrix} \quad (D.16)$$

where  $M_{el}^*$  indicates the sorted edge list. This guarantees that for an intersection the vertex  $\underline{v}_p$  in the first row of  $M_{el}^*$  is outside the new boundary (negative sign of  $\beta$ ) while the vertex  $\underline{v}_q$  in the second row of  $M_{el}^*$  is inside the new boundary (positive sign of  $\beta$ ). Obviously, for the first and third combination in the edge list, new vertices have to be computed. The new vertices can now be computed according to :

$$\underline{v}_{\text{intersect}} = \lambda \underline{v}_p + (1 - \lambda) \underline{v}_q \quad (D.17)$$

where  $\psi_i(\ell, \underline{\theta}) \underline{v}_{\text{intersect}} = c_i(\ell)$ , and

$$\lambda = \frac{\psi_i(\ell, \underline{\theta}) \underline{v}_q - c_i(\ell)}{\psi_i(\ell, \underline{\theta}) \underline{v}_q - \psi_i(\ell, \underline{\theta}) \underline{v}_p} > 0 \quad (D.18)$$

which defines the unique solution of Eq. D.18 and the new vertex.

The updated boundary of the simulation example is shown in Fig. D.1b and the corresponding vertex set  $M_{vs}$  derived according to the described procedure becomes :

$$M_{vs} = \begin{bmatrix} \underline{v}_1 & \underline{v}_2 & \underline{v}_3 & \underline{v}_4 & \underline{v}_5 \\ 3 & 2 & 3 & 1 & 1 \\ 2 & 1 & 3 & 3 & 1 \end{bmatrix} \quad \text{Updated vertex coordinates} \quad (D.19)$$

where the first two vertices have been added due to the intersection of the new hyperplane and the last three vertices correspond to old vertices (in different order). This reordering of the vertices results from the fact that in the exact polytope updating algorithm, the updated vertex list  $M_{vs}$  (Eq. D.19) is constructed by first computing the new vertices according to Eq. D.18 and then adding the vertices which have not been modified.

Round-off errors in the computation of  $\beta$  and  $\lambda$  can cause incorrect updating of the parameter uncertainty set  $\Theta$ . Coinciding vertices might not be detected due to these round-off errors which then results in artificial adding or eliminating of vertices. The repeated occurrence of exclusion of feasible vertices and inclusion of infeasible vertices from the vertex set can lead to an inaccurate final polytope which is undesirable. This is, however, very unlikely in practice and will have very little effect on the algorithm but can be avoided by combining vertices within a predefined distance.

### D.2.3 Updating Vertex-plane Adjacency List

The updating of  $M_{vp}$  is straightforward since the intersecting hyperplanes of the old vertices do not change. This information is already known from the old list. For the new vertices (Eq. D.17) the added hyperplane is one element. The other hyperplanes intersecting the new vertex can be derived by selecting the  $n\theta-1$  common hyperplanes of the vertex combination in the edge list.

$$M_{vp} = \begin{matrix} & \underline{v}_1 & \underline{v}_2 & \underline{v}_3 & \underline{v}_4 & \underline{v}_5 \\ \begin{bmatrix} 5 & 5 & 1 & 1 & 2 \\ 1 & 4 & 3 & 3 & 4 \end{bmatrix} & \text{Updated vertex-plane adjacency list} \end{matrix} \quad (D.20)$$

In this updated list, the first element of the new vertices,  $\underline{v}_1$  and  $\underline{v}_2$ , corresponds to the index of the new hyperplane. The second element has of  $\underline{v}_1$  for example corresponds to the index of the common hyperplane of  $\underline{v}_1$  and  $\underline{v}_2$  (first column of the edge list in Eq. D.14), i.e. 1 (common index of first two columns in  $M_{vp}$  list, Eq. D.11).

### D.2.4 Updating Vertex-vertex Adjacency List

This is the most time consuming step in the updating procedure since the new  $M_{vv}$  list must be constructed again for all vertices due to the reordering of these vertices. Because edges between neighbour vertices are formed by  $n\theta-1$  hyperplanes, the  $M_{vv}$  list can be constructed by searching for at least  $n\theta-1$  common hyperplanes in the  $M_{vp}$  list. Note that the number of adjacent vertices depends on the number of hyperplanes intersecting a vertex which can theoretically explode to infinity but simulation examples indicate that this number will be limited in most cases.

The procedure of exact polytope updating shows however clearly, the exploding computational complexity for high-dimensional identification problems. An overview of other exact bounding methods is given in [MN90], but all these algorithms which are

not significantly different suffer from the fundamental problem of dimensionality.

$$M_{vv} = \begin{matrix} & \underline{v}_1 & \underline{v}_2 & \underline{v}_3 & \underline{v}_4 & \underline{v}_5 \\ \begin{bmatrix} 2 & 1 & 1 & 3 & 2 \\ 3 & 5 & 4 & 5 & 4 \end{bmatrix} & \text{Updated vertex-vertex adjacency list} \end{matrix} \quad (D.21)$$

### D.3 Ellipsoidal Parameter Bounding

This approach is based on a recursive construction of ellipsoidal sets enclosing the parameter uncertainty set  $\Theta$  :

$$\mathbf{E}^{\ell-1}(\underline{\theta}_c^{\ell-1}, \mathbf{P}^{\ell-1}) = \left\{ \underline{\theta} \in \mathbb{R}^{n\theta} : \left( \underline{\theta} - \underline{\theta}_c^{\ell-1} \right)^T [\mathbf{P}^{\ell-1}]^{-1} \left( \underline{\theta} - \underline{\theta}_c^{\ell-1} \right) \leq 1 \right\} \quad (D.22)$$

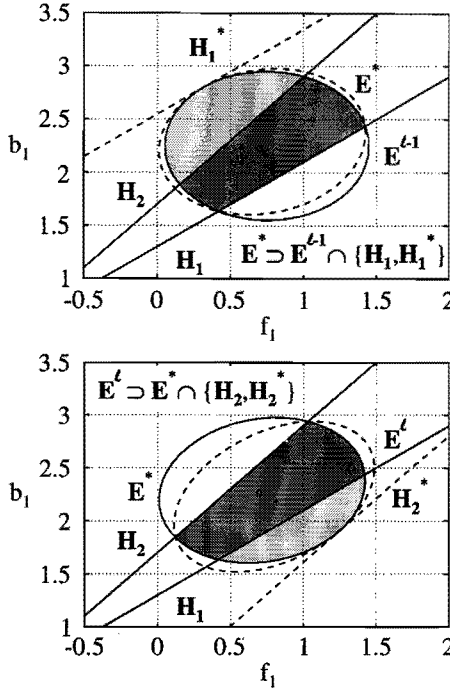
being a bounding ellipsoid from the first  $\ell-1$  constraints and a region  $\mathbf{S}^\ell$  :

$$\mathbf{S}^\ell = \left\{ \underline{\theta} \in \mathbb{R}^{n\theta} : (y_{\text{col}}(\ell) - \phi(\ell, \underline{\theta}))^2 \leq \delta_{e, \text{col}}^2(\ell) \right\} \quad (D.23)$$

being the region in the parameter space associated with the new measurements which is limited by two parallel hyperplanes (Fig. 3.4), i.e.  $\Delta\phi(\ell, \underline{\theta}) = 0$  is required in Eq. D.5.

However, when approximating models which are nonlinear in their parameters by pseudo-linear regression models to guarantee a convex parameter uncertainty set  $\Theta$ , the new parameter space  $\mathbf{S}^\ell$  will be enclosed by two non-parallel hyperplanes  $\{\mathbf{H}_1, \mathbf{H}_2\}$  (Eq. D.4).

In [CG90] a two step algorithm is proposed where parallel hyperplanes are systematically used, once or twice at each iteration step depending on the configuration of the parameter space  $\mathbf{S}^\ell$  and the ellipsoid  $\mathbf{E}^{\ell-1}$  (Fig. D.2). The algorithm is initiated by taking  $\mathbf{E}^0$  large enough to ensure that it contains the parameter uncertainty set  $\Theta$ . A hyperplane  $\mathbf{H}_1^*$  tangent to  $\mathbf{E}^{\ell-1}$  and parallel to the constraint  $\mathbf{H}_1$  is constructed to define an intermediate ellipsoid  $\mathbf{E}^*$ . Next this procedure is repeated for  $\mathbf{E}^*$  and  $\mathbf{H}_2^*$  to construct  $\mathbf{E}^\ell$ . Using this approach, it is obvious that the ellipsoids  $\mathbf{E}^*$  and  $\mathbf{E}^\ell$  describe a parameter space that is much larger than strictly necessary (compare shaded areas in Fig. D.2). Instead of using tangent hyperplanes to obtain a new ellipsoid  $\mathbf{E}^\ell$  of minimum volume, parallel hyperplanes  $\{\mathbf{H}_1^*, \mathbf{H}_2^*\}$  should be constructed which intersect  $\mathbf{E}^{\ell-1}$  and  $\{\mathbf{H}_1, \mathbf{H}_2\}$  describing the smallest parameter space possible.



**Fig. D.2 :** Two-step construction of a bounding ellipsoid from non-parallel hyperplanes using intermediate tangent-parallel hyperplanes.

Before going into detail of this last approach, the ellipsoidal bounding procedure will be summarized. If  $E^{\ell-1} \cap (H_1, H_2) = \emptyset$ , the parameter space  $\Theta$  is empty and the procedure is terminated. This occurs if :

$$\begin{aligned} c_1(\ell) &> \psi_1(\ell, \underline{\theta}) \underline{\theta}_c^{\ell-1} + \sqrt{g_1(\ell, \underline{\theta})} \quad \text{or} \\ c_2(\ell) &< \psi_2(\ell, \underline{\theta}) \underline{\theta}_c^{\ell-1} - \sqrt{g_2(\ell, \underline{\theta})} \end{aligned} \quad (\text{D.24})$$

where

$$\begin{aligned} g_1(\ell, \underline{\theta}) &= \psi_1(\ell, \underline{\theta}) P^{\ell-1} \psi_1^T(\ell, \underline{\theta}) \\ g_2(\ell, \underline{\theta}) &= \psi_2(\ell, \underline{\theta}) P^{\ell-1} \psi_2^T(\ell, \underline{\theta}) \end{aligned} \quad (\text{D.25})$$

Otherwise three different situations can occur when calculating  $E^\ell$  after testing which

of the hyperplanes  $\{H_1, H_2\}$  intersects  $E^{\ell-1}$ .

$$\begin{aligned} H_1 : \quad & |c_1(\ell) - \psi_1(\ell, \underline{\theta}) \underline{\theta}_c^{\ell-1}| \leq \sqrt{g_1(\ell, \underline{\theta})} \\ H_2 : \quad & |c_2(\ell) - \psi_2(\ell, \underline{\theta}) \underline{\theta}_c^{\ell-1}| \leq \sqrt{g_2(\ell, \underline{\theta})} \end{aligned} \quad (D.26)$$

- $\{H_1, H_2\}$  do not intersect  $E^{\ell-1}$ . Then  $E^\ell = E^{\ell-1}$ .
- Only one of the hyperplanes  $\{H_1, H_2\}$  intersects  $E^{\ell-1}$ . In this case the non intersecting hyperplane is replaced by one tangent to  $E^{\ell-1}$  and parallel to the intersecting hyperplane defining temporarily a new measurement and bound value :

$$\begin{aligned} H_1^* : \quad & \begin{cases} y^*(\ell) = \frac{1}{2} \left( c_1(\ell) + \psi_1(\ell, \underline{\theta}) \underline{\theta}_c^{\ell-1} + \sqrt{g_1(\ell, \underline{\theta})} \right) \\ \delta^*(\ell) = \frac{1}{2} \left( -c_1(\ell) + \psi_1(\ell, \underline{\theta}) \underline{\theta}_c^{\ell-1} + \sqrt{g_1(\ell, \underline{\theta})} \right) \end{cases} \\ H_2^* : \quad & \begin{cases} y^*(\ell) = \frac{1}{2} \left( c_2(\ell) + \psi_2(\ell, \underline{\theta}) \underline{\theta}_c^{\ell-1} - \sqrt{g_2(\ell, \underline{\theta})} \right) \\ \delta^*(\ell) = \frac{1}{2} \left( c_2(\ell) - \psi_2(\ell, \underline{\theta}) \underline{\theta}_c^{\ell-1} + \sqrt{g_2(\ell, \underline{\theta})} \right) \end{cases} \end{aligned} \quad \text{or} \quad (D.27)$$

- Both hyperplanes  $\{H_1, H_2\}$  intersect  $E^{\ell-1}$ . The new ellipsoid  $E^\ell$  can be obtained by a two step procedure, similar to [CG90], computing parallel and intersecting hyperplanes  $\{H_1^*, H_2^*\}$ . First an intermediate ellipsoid :  $E^* \supset E^{\ell-1} \cap \{H_2, H_2^*\}$  and then the new ellipsoid :  $E^\ell \supset E^* \cap \{H_1, H_1^*\}$  is constructed.

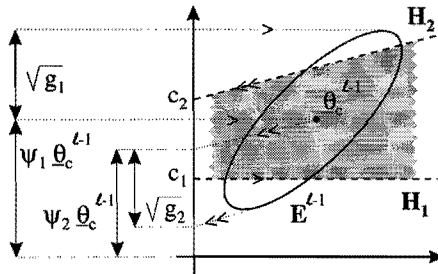


Fig. D.3 : Intersection of hyperplanes  $\{H_1, H_2\}$  with ellipsoid  $E^{\ell-1}$ .

Note that the cross-section  $\mathbf{H}^\perp$  which is considered to compute the intersecting hyperplanes  $\{\mathbf{H}_1^*, \mathbf{H}_2^*\}$  describes a special cross-section in the  $n\theta$ -dimensional parameter space : the space perpendicular to the intersection of  $\mathbf{H}_1$  and  $\mathbf{H}_2$ , i.e.  $\mathbf{H}^\perp \perp (\mathbf{H}_1 \cap \mathbf{H}_2)$ . The construction of parallel and intersecting hyperplanes  $\{\mathbf{H}_1^*, \mathbf{H}_2^*\}$  will be explained by constructing a hyperplane  $\mathbf{H}_2^*$  parallel to  $\mathbf{H}_2$  and intersecting at  $\mathbf{E}^{\ell-1} \cap \mathbf{H}_1$ . First apply a coordinate transformation using a Cholesky decomposition ( $\mathbf{P} = \mathbf{R}^T \mathbf{R}$ ), such that the ellipsoid  $\mathbf{E}^{\ell-1}$  transforms into a unit hypersphere in the origin :

$$\mathbf{V}_1 = \boldsymbol{\psi}_1 \mathbf{R}^T, \quad \mathbf{V}_{1,n} = \frac{\mathbf{V}_1}{\|\mathbf{V}_1\|_2}, \quad c_{1,n} = \frac{c_1 - \boldsymbol{\psi}_1 \boldsymbol{\theta}_c}{\|\mathbf{V}_1\|_2} \quad (\text{D.28})$$

$$\mathbf{V}_2 = \boldsymbol{\psi}_2 \mathbf{R}^T, \quad \mathbf{V}_{2,n} = \frac{\mathbf{V}_2}{\|\mathbf{V}_2\|_2}, \quad c_{2,n} = \frac{c_2 - \boldsymbol{\psi}_2 \boldsymbol{\theta}_c}{\|\mathbf{V}_2\|_2}$$

where the arguments  $\ell$  and  $\boldsymbol{\theta}$  have been omitted to simplify the notation. The resulting situation is illustrated in Fig. D.4. The next step is to calculate the intersection point  $\underline{s}_1$  that has a maximum distance to  $\mathbf{H}_2$  in the following way. Let :

$$\underline{s}_0 = c_{1,n} \mathbf{V}_{1,n}^T \quad (\text{D.29})$$

define the projection of the origin on  $\mathbf{H}_1$ . This hyperplane will intersect  $\mathbf{E}_{\text{unit}}$  in  $\underline{s}_1$  and  $\underline{s}_2$  which define the intersection points of maximum and minimum distance to  $\mathbf{H}_2$  respectively. The distance between  $\underline{s}_0$  and the intersection points  $\underline{s}_1$  and  $\underline{s}_2$  is defined by :

$$d = \sqrt{1 - \underline{s}_0^T \underline{s}_0} \quad (\text{D.30})$$

There exists a hyperplane  $\mathbf{H}_0$  parallel to  $\mathbf{H}_1$  containing the origin :  $\mathbf{V}_{1,n} \boldsymbol{\theta} = 0$ . The direction vector from  $\underline{s}_0$  to the intersection points can now be computed by deriving the projection of  $\mathbf{V}_{2,n}$  onto the hyperplane  $\mathbf{H}_0$  :

$$\mathbf{p} = \mathbf{V}_{2,n}^T - \left( \mathbf{V}_{1,n} \mathbf{V}_{2,n}^T \right) \mathbf{V}_{1,n}^T, \quad \mathbf{p}_n = \frac{\mathbf{p}}{\|\mathbf{p}\|_2} \quad (\text{D.31})$$

This normalized direction vector  $\mathbf{p}_n$  can be used to compute the intersection points :  $\underline{s}_i = \underline{s}_0 \pm d \mathbf{p}_n$ . The point which has the maximum distance to the hyperplane  $\mathbf{H}_2$  is selected as  $\underline{s}_1$ . Inverse coordinate transformation and adding the centre point  $\boldsymbol{\theta}_c$  results in the intersection point in the original coordinate system :

$$\underline{s} = R \underline{s}_1 + \underline{\theta}_c \quad (D.32)$$

Define  $c_2^*(\ell) = \psi_2(\ell, \underline{\theta}) \underline{s}$ . The intersecting hyperplane  $H_2^*$  can now be constructed as follows :

$$H_2^* = \left\{ \underline{\theta} \in \mathbb{R}^{n\theta} : \psi_2(\ell, \underline{\theta}) \underline{\theta} \leq c_2^*(\ell) \right\} \quad (D.33)$$

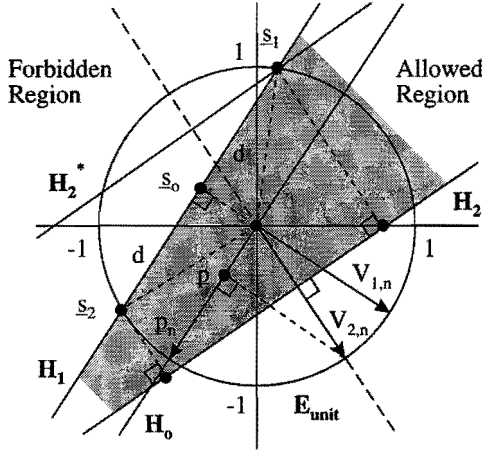
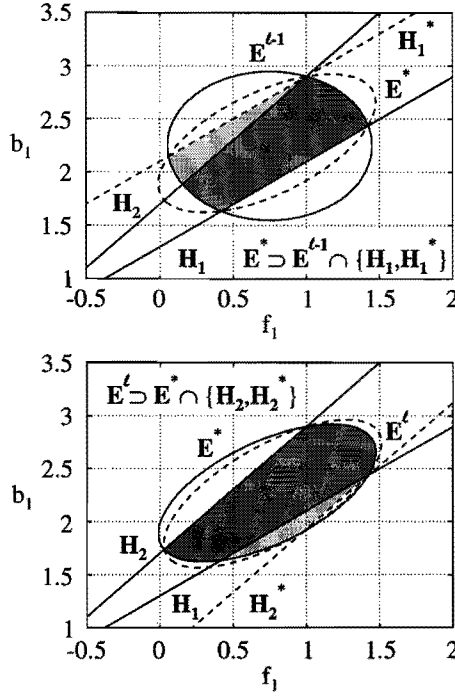


Fig. D.4 : Construction of intersection point in cross-section  $H^1$ .

Consider now  $H_2$  and the parallel hyperplane  $H_2^*$ . The ellipsoid  $E^* \supset E^{t-1} \cap \{H_2, H_2^*\}$  can be computed using for example the basic ellipsoidal bounding formulas. In the final setup, the procedure just explained is first applied to construct an intermediate ellipsoid  $E^*$  based on  $H_2$  and  $H_1^* // H_1$  and next the procedure is repeated applying it to  $E^*$ ,  $H_2^* // H_2$  and  $H_1$  to yield  $E^\ell$ . This procedure is illustrated in Fig. D.5.

The two-step construction of intersecting hyperplanes for ellipsoidal bounding can be summarized as follows :

$$H_1^* \begin{cases} y^*(\ell) = \frac{1}{2} (c_1(\ell) + \psi_1(\ell, \underline{\theta}) \underline{s}) \\ \delta^*(\ell) = \frac{1}{2} (-c_1(\ell) + \psi_1(\ell, \underline{\theta}) \underline{s}) \end{cases}, \quad H_2^* \begin{cases} y^*(\ell) = \frac{1}{2} (c_2(\ell) + \psi_2(\ell, \underline{\theta}) \underline{s}) \\ \delta^*(\ell) = \frac{1}{2} (c_2(\ell) - \psi_2(\ell, \underline{\theta}) \underline{s}) \end{cases} \quad (D.34)$$



**Fig. D.5** : Two-step construction of bounding ellipsoid from non-parallel hyperplanes using intermediate intersecting-parallel hyperplanes.

Whenever parallel hyperplanes have been constructed, basic ellipsoidal bounding algorithms can be applied to update the ellipsoid  $E^{l-1}$ . An ellipsoid of minimal volume can be derived according to [FH82] by computing the auxiliary variables :

$$\begin{aligned}
 \varepsilon &= y^*(\ell) - \psi_i(\ell, \underline{\theta}) \underline{\theta}_c^{\ell-1} \\
 a &= (n\theta - 1) g_i^2(\ell, \underline{\theta}) \\
 b &= g_i(\ell, \underline{\theta}) \left( (2n\theta - 1) \delta^{*2}(\ell) - g_i(\ell, \underline{\theta}) + \varepsilon^2 \right) \\
 c &= \delta^{*2}(\ell) \left( n\theta (\delta^{*2}(\ell) - \varepsilon^2) - g_i(\ell, \underline{\theta}) \right)
 \end{aligned} \tag{D.35}$$

and



$$\rho = \max \left\{ 0, \frac{-b + \sqrt{b^2 - 4ac}}{2a} \right\} \quad (D.36)$$

where depending on the construction of parallel hyperplanes  $\mathbf{H}_1^*$  and  $\mathbf{H}_2^*$ ,  $i$  corresponds with  $\{1,2\}$ . The parameter estimate can then be updated by :

$$\begin{aligned} \mathbf{P}^{\ell-1} &:= \mathbf{P}^{\ell-1} - \frac{\rho \mathbf{P}^{\ell-1} \boldsymbol{\Psi}_i^T(\ell, \underline{\theta}) \boldsymbol{\Psi}_i(\ell, \underline{\theta}) \mathbf{P}^{\ell-1}}{\delta^{*2}(\ell) + \rho g_i(\ell, \underline{\theta})} \\ \mathbf{P}^{\ell} &:= \mathbf{P}^{\ell-1} \left( 1 + \rho - \frac{\rho \varepsilon^2}{\delta^{*2}(\ell) + \rho g_i(\ell, \underline{\theta})} \right) \\ \underline{\theta}_c^{\ell} &:= \underline{\theta}_c^{\ell-1} + \frac{\rho \varepsilon}{\delta^{*2}(\ell)} \mathbf{P}^{\ell-1} \boldsymbol{\Psi}_i^T(\ell, \underline{\theta}) \end{aligned} \quad (D.37)$$

Instead of using the Modified Ellipsoidal Bounding (MEB) algorithm (updating according to Eq. D.37 combined with the modification of Eq. D.27), the Ellipsoidal bounding with Parallel Cuts (EPC) algorithm described in [WPL90] can be applied as well which is mathematically equivalent to MEB [PWPL89]. The ellipsoidal updating in these two algorithms is computed numerically in a different way, but simulation examples did not indicate a preference due to a higher numerical accuracy or faster updating. The minimization of the volume of the ellipsoid enclosing the parameter uncertainty set  $\Theta$  involves the analysis of a quadratic function (Eq. D.36) where as alternative the minimization of the trace of the ellipsoid involves the analysis of a cubic equation which requires somewhat more computational effort. A detailed description of this approach can be found in [FH82].

Simulations showed that the order in which the intermediate hyperplanes  $\{\mathbf{H}_1^*, \mathbf{H}_2^*\}$  are constructed, can result in a final ellipsoid  $\mathbf{E}^{\ell}$  of different volume for both approaches (tangent as well as intersecting hyperplanes). It is easy to give examples which illustrate this fact. To obtain an optimal solution with respect to the minimum volume of the ellipsoid  $\mathbf{E}^{\ell}$  enclosing  $\mathbf{S}^{\ell} \cap \mathbf{E}^{\ell-1}$ , both orders,  $\{\mathbf{H}_1^*, \mathbf{H}_2^*\}$  and  $\{\mathbf{H}_2^*, \mathbf{H}_1^*\}$ , must be applied to derive  $\mathbf{E}^{\ell}$  and select the ellipsoid of minimum volume. This increases, of course, the computational complexity.

The volume of the ellipsoid  $V_e$  is proportional to the determinant of  $\mathbf{P}^{\ell}$  :

$$V_e = \frac{\pi^{1/2} n\theta}{\Gamma(1/2 n\theta)} \sqrt{\det(P^\ell)} \quad (D.38)$$

where

$$\Gamma(1/2 n\theta) = \begin{cases} \prod_{k=1}^{1/2 (n\theta-2)} k & n\theta \text{ even} \\ \frac{\sqrt{\pi}}{2^{1/2 (n\theta-1)}} \prod_{k=1}^{1/2 (n\theta-1)} 2k-1 & n\theta \text{ odd} \end{cases} \quad (D.39)$$

The minimum and maximum parameter values of the corresponding ellipsoid  $E^\ell$  are defined by :

$$\theta_{\min,i} = \theta_{c,i} - \sqrt{P^\ell(i,i)} \quad , \quad \theta_{\max,i} = \theta_{c,i} + \sqrt{P^\ell(i,i)} \quad (D.40)$$

describing the uncertainty intervals associated with each parameter  $\theta_i$ .

#### D.4 Orthotopic Parameter Bounding

The parameter uncertainty set  $\Theta$  is described by an orthotope aligned with the coordinate axis. The final bounds  $\theta_{\min,i}$  and  $\theta_{\max,i}$  of the uncertainty interval associated with the  $i^{\text{th}}$  parameter are given by the minimum and maximum values of the criterion :

$$f(\underline{\theta}) = \theta_i \quad (D.41)$$

under  $2N_{\text{col}}$  linear constraints :

$$\Psi(\underline{\theta}) \underline{\theta} \leq \underline{c} \quad (D.42)$$

where

$$\Psi(\underline{\theta}) = \begin{bmatrix} -\Psi_1(\underline{\theta}) \\ \Psi_2(\underline{\theta}) \end{bmatrix} \quad , \quad \underline{c} = \begin{bmatrix} -\underline{c}_1 \\ \underline{c}_2 \end{bmatrix} \quad (D.43)$$

and the elements of  $\Psi_i(\underline{\theta})$  and  $\underline{c}_i$ ,  $i = \{1,2\}$ , are defined in Eq. D.5. Each bound can therefore be obtained by solving a linear programming problem. The computation of the orthotopic bounding thus requires the solution of  $2n\theta$  linear programming problems,

each with  $2N_{\text{col}}$  linear constraints which can be solved using a simplex method as proposed in [WPL90]. However, as shown in Fig. 3.6, only a small set of constraints is active in an optimal point  $\underline{\theta}$ . In addition, equality constraints are easier to treat than inequalities, which suffer from the fundamental difficulty that the active set is unknown. These two facts form a motivation to use active set methods. The idea of active set methods is to develop a prediction (working set) of the constraints active at the solution. The working set may change at each iteration and constraints in the working set are (temporarily) treated as equalities during a given iteration.

The solution to the problem can then be found in two steps. First an initial feasible solution  $\underline{\theta}^0$  together with the corresponding initial working set has to be derived. A feasible solution for the  $2N_{\text{col}}$  inequality constraints, is constructed by adding a single artificial variable, denoted by  $\xi$ . Consider the  $\ell^{\text{th}}$  constraint,  $\psi(\ell, \underline{\theta}) \underline{\theta}^0 \leq c(\ell)$ , of the set of original constraints defined in Eq. D.42. This constraint can always be satisfied by subtracting a sufficiently large positive quantity :  $\psi(\ell, \underline{\theta}) \underline{\theta}^0 - \xi \leq c(\ell)$ . If  $\underline{\theta}^0$  satisfies this constraint when  $\xi$  is zero,  $\underline{\theta}^0$  is feasible with respect to the original constraint. To find a feasible point, it is sufficient to drive the artificial variable to zero, which can be achieved by an appropriate definition of the objective function. A new linear programming problem to find an initial feasible solution has been formulated:

$$\begin{aligned} & \text{minimize} \quad \xi, \quad \Psi(\underline{\theta}) \underline{\theta}^0 - \xi \underline{1} \leq \underline{c} \\ & \underline{\theta}^0 \in \mathbb{R}^{n\theta}, \xi \in \mathbb{R}^+ \end{aligned} \quad (\text{D.44})$$

where  $\underline{1} \in \mathbb{R}^{2N_{\text{col}}}$ . The new variable  $\underline{\theta}^*$ , the normal vector of the objective function  $\underline{f}^*$  and the  $2N_{\text{col}}+1$  inequality constraints :  $\Psi^*(\underline{\theta}) \underline{\theta}^* \leq \underline{c}^*$ , are given by :

$$\underline{\theta}^* = \begin{pmatrix} \underline{\theta} \\ \xi \end{pmatrix}, \quad \underline{f}^* = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \quad \Psi^*(\underline{\theta}) = \begin{pmatrix} \Psi(\underline{\theta}) & -\underline{1} \\ 0 & -1 \end{pmatrix}, \quad \underline{c}^* = \begin{pmatrix} \underline{c} \\ r \end{pmatrix} \quad (\text{D.45})$$

If the smallest value of  $\xi$  is zero (in which case a feasible point has been found with respect to the original constraints) the added constraint -  $\xi \leq 0$  ( $r = 0$ ) is active and the feasible point is a vertex of the polytope describing the parameter uncertainty set  $\Theta$ . Due to finite precision in numerical computation, however, the minimum value of  $\xi$  can be strictly positive, so that the constraint -  $\xi \leq 0$  is not active. Therefore, this formulation with one artificial variable to find a feasible initial point, is slightly modified by choosing  $r$  as a small positive number, -  $\xi \leq r$  ( $r = 10^{-5}$ ) [GMW91]. Note that it has to be verified afterwards that this constraint is active for  $\underline{\theta}^*$ .

Starting at an arbitrary point  $\underline{\theta}^*$ , the search direction is defined by  $\underline{p}^k = -Z^k [Z^k]^T \underline{f}^*$ , where  $Z^k$  is the null space of the active set ( $Z^0 = I$ ). Selecting the minimum distance  $d_{\min}$  in the search direction before a constraint ( $\Psi(\underline{\theta}^*) [\underline{\theta}^*]^T \leq \underline{c}^*$ ) is violated, the new point  $[\underline{\theta}^*]^{k+1}$  is defined by:  $[\underline{\theta}^*]^{k+1} = [\underline{\theta}^*]^k + d_{\min} \underline{p}^k$ . As long as the active set  $\Psi_a^*(\underline{\theta})$  is smaller than the number of parameters ( $n\theta+1$ ), the constraint along the search direction is added to the active set. When the active set is equal to the number of parameters, the new constraint along the search direction  $\underline{p}^k$  is added to the working set and the constraint corresponding with the smallest Lagrange multiplier  $\underline{f}^* = \Psi_a^*(\underline{\theta}) \underline{\lambda}_a^{\min}$  is removed. The procedure is terminated and an optimal initial point is reached, if there exists a point  $\underline{\theta}_{\text{opt}}^*$  satisfying  $\Psi^*(\underline{\theta}) \underline{\theta}_{\text{opt}}^* \leq \underline{c}^*$ ,  $\underline{f}^* = \Psi_a^*(\underline{\theta}) \underline{\lambda}_a^{\text{opt}}$  and when all Lagrange multipliers satisfy  $\underline{\lambda}_a^{\text{opt}} \geq 0$ . After deriving an initial feasible point together with the  $n\theta+1$  active constraints, the artificial variable can be eliminated again together with the corresponding constraint. This gives the initialization conditions, a feasible point  $\underline{\theta}^0$  that is a vertex of the polytope and the initial working set ( $n\theta$  constraints) for the non simplex active set method.

The final solution can now be found by solving, the same way, the  $2n\theta$  linear programming problems  $f(\underline{\theta})$  for the original constraints (Eq. D.42) using the derived working set  $\Psi_a(\underline{\theta})$  and the initial feasible point  $\underline{\theta}^0$ . A more detailed explanation can be found in [GMW91].

# References

---

- H. Akçay and P. Khargonekar** (1993). *The least-squares algorithm, parametric system identification and bounded noise*. Automatica, Vol. 29, No. 6, pp. 1535-1540.
- L.V.R. Arruda and G. Favier** (1991). *A review and a comparison of robust estimation methods*. 9<sup>th</sup> IFAC/IFORS Symp. on Identification and System Parameter Estimation, Budapest, Hungary, pp. 1027-1032.
- L.V.R. Arruda, G. Favier and W. do Amaral** (1991). *Reformulation of optimal bounding ellipsoid algorithms as robust identification algorithms with dead zone*. Proc. 1<sup>st</sup> ECC '91, Grenoble, France, pp. 1194-1199.
- K.J. Åström** (1980). *Maximum likelihood and prediction error methods* Automatica, Vol. 16, pp. 551-574.
- A.C.P.M. Backx** (1987). *Identification of an industrial process : A Markov parameter approach*. Ph.D. Thesis, Eindhoven University of Technology, The Netherlands.
- A.C.P.M. Backx and A.A.H. Damen** (1989a). *Identification of industrial MIMO processes for fixed controllers - Part 1 : General theory and practice*. Journal A, Vol. 30, No. 1, pp. 3-12.
- A.C.P.M. Backx and A.A.H. Damen** (1989b). *Identification of industrial MIMO processes for fixed controllers - Part 2 : Case studies*. Journal A, Vol. 30, No. 2, pp. 33-43.
- A.C.P.M. Backx and A.A.H. Damen** (1992). *Identification for the control of MIMO*

- industrial processes*. IEEE Trans. on Aut. Control, Vol. 37, No. 7, pp. 980-986.
- G.J. Balas and J.C. Doyle** (1989). *Identification for robust control of flexible structures* Proc. ACC '89, pp. 2566-2571.
- B. Barmish, P. Khargonekar, Z. Shi and R. Tempo** (1989). *Robustness margin need not to be a continuous function of the problem data*. Systems and Control Letters, Vol. 15, pp. 91-98.
- A.C. Bartlett, A. Tesi and A. Vicino** (1993). *Frequency response of uncertain systems with interval plants*. IEEE Trans. on Aut. Control, Vol. 38, No. 6, pp. 929-933.
- S. Beghelli and R.P. Guidorzi** (1983). *Transformations between input-output multistructural models : Properties and applications*. Int. J. of Control, Vol. 37, pp. 1385-1400.
- S. Beghelli, R.P. Guidorzi and U. Soverini** (1987). *Structural identification of multivariable systems by the eigenvector method*. Int. J. of Control, Vol. 46, No. 2, pp. 671-678.
- S. Beghelli, R.P. Guidorzi and U. Soverini** (1990). *The Frisch scheme in dynamic system identification*. Automatica, Vol. 26, No. 1, pp. 171-176.
- G.A. Bekey and S.F. Masri** (1983). *Random search techniques for optimization of nonlinear systems with many parameters*. Math. and Comp. in Simulation 25, pp. 210-213.
- G. Belforte and B. Bona** (1985). *An improved parameter identification algorithm for signals with unknown-but-bounded errors*. IFAC Symp. on Identification and System Parameter Estimation, York, UK, pp. 1507-1511.
- G. Belforte, B. Bona and V. Cerone** (1988a). *Parameter estimation with set membership uncertainty : nonlinear families of models*. IFAC Symp. on Identification and System Parameter Estimation, Beijing, PRC, pp. 399-404.
- G. Belforte, B. Bona and V. Cerone** (1988b). *A bounded error approach to the tuning of a digital voltmeter*. 12<sup>th</sup> IMACS World Congress on Scientific Computation, Vol. 2, Paris, France, pp. 498-501.
- G. Belforte, B. Bona and V. Cerone** (1990a). *Parameter estimation algorithms for a set-membership description of uncertainty*. Automatica, Vol. 26, No. 5, pp. 887-898.
- G. Belforte, B. Bona and V. Cerone** (1990b). *Identification, structure selection and validation on uncertain models with set-membership error description*. Math. and Comp. in Simulation, Vol. 32, No. 5 & 6, pp. 561-570.
- G. Belforte and T.T. Tay** (1991). *Recursive estimation for linear models with set membership measurement error*. 9<sup>th</sup> IFAC/IFORS Symp. on Identification and System Parameter Estimation, Budapest, Hungary, pp. 872-877.
- G. Belforte and T.T. Tay** (1993). *Two new estimation algorithms for linear models with unknown but bounded measurement noise*. IEEE Trans. on Aut. Control, Vol. 38, No. 8, pp. 1273-1279.

- 
- G. Belforte, R. Tempo and A. Vicino** (1988). *Robust parameter estimation for linear models with set-membership uncertainty*. IFAC Symp. on Identification and System Parameter Estimation, Beijing, China, pp. 809-813.
- S.A. Billings, S. Chen and M.J. Korenberg** (1989). *Identification of MIMO non-linear systems using a forward-regression orthogonal estimator*. Int. J. of Control, Vol. 49, No. 6, pp. 2175-2189.
- R.G. Bland, D. Goldfarb and M.J. Todd** (1981). *The ellipsoid method : A survey*. Operations Research, Vol. 29, No. 6, pp. 1039-1091.
- S.P. Boyd** (1986). *A note on parametric and nonparametric uncertainties in control systems*. Proceedings ACC '86, pp. 1847-1849.
- S.P. Boyd and C.H. Barrat** (1991). *Linear controller design ; Limits of performance*. Prentice Hall, Englewood Cliffs, N.J.
- S.P. Boyd and J.C. Doyle** (1987). *Comparison of peak and RMS gains for discrete-time systems*. Systems & Control Letters 9, pp. 1-6.
- V. Broman and M.J. Shensea** (1990). *A compact algorithm for the intersection and approximation of N-dimensional polytopes*. Math. and Comp. in Simulation, Vol. 32, No. 5 & 6, pp. 469-480.
- C. Canudas de Wit and J. Carrillo** (1990). *A modified EW-RLS algorithm for systems with bounded disturbances*. Automatica, Vol. 26, No. 3, pp. 599-606.
- V. Cerone** (1991a). *Feasible parameter set for bilinear systems from records with bounded output errors*. Proc. ACC '91, pp. 37-42.
- V. Cerone** (1991b). *Parameter bounding in ARMAX models from records with bounded errors in variables*. 9<sup>th</sup> IFAC/IFORS Symp. on Identification and System Parameter Estimation, Budapest, Hungary, pp. 1419-1424.
- V. Cerone** (1991c). *Parameter bounds for models with bounded errors in all variables*. 9<sup>th</sup> IFAC/IFORS Symp. on Identification and System Parameter Estimation, Budapest, Hungary, pp. 1518-1523.
- V. Cerone** (1993). *Feasible parameter set for linear models with bounded errors in all variables*. Automatica, Vol. 29, No. 6, pp. 1551-1555.
- M.F. Cheung, S. Yurkovich and K.M. Passino** (1991). *An optimal volume ellipsoid algorithm for parameter set estimation*. Proc. 30<sup>th</sup> CDC, Brighton, England, pp. 969-974.
- T. Clément and S. Gentil** (1988). *Reformulation of parameter identification with unknown-but-bounded errors*. Math. and Comp. in Simulation 30, pp. 257-270.
- T. Clément and S. Gentil** (1990). *Recursive membership estimation for output-error models*. Math. and Comp. in Simulation, Vol. 32, No. 5 & 6, pp. 505-514.
- G.O. Corrêa and K. Glover** (1984a). *Pseudo-canonical forms, identifiable parametrizations and simple parameter estimation for linear multivariable systems : Input-output models*. Automatica, Vol. 20, pp. 429-442.
- G.O. Corrêa and K. Glover** (1984b). *Pseudo-canonical forms, identifiable*

- parametrizations and simple parameter estimation for linear multivariable systems : Parameter estimation. Automatica, Vol. 20, pp. 443-452.*
- M.A. Dahley and M.H. Khammash** (1993). *Controller design for plants with structured uncertainty. Automatica, Vol. 29, No. 1, pp. 37-56.*
- A.A.H. Damen, H.M. Falkus and J.P.H.M Bouwels** (1994). *Modeling and control of a floating platform. IEEE Trans. on Aut. Control, Vol. 39, No. 5, pp. 1075-1078.*
- S. Dasgupta and Y.F. Huang** (1985). *Asymptotically convergent modified recursive least-square with data-dependent updating and forgetting factor. Proc. 24<sup>th</sup> CDC, Ft. Lauderdale, Florida, pp. 1067-1071.*
- S. Dasgupta and Y.F. Huang** (1987). *Asymptotically convergent modified recursive least-squares with data-dependent updating and forgetting factor for systems with bounded noise. IEEE Trans. on Information Theory, Vol. IT-33, No. 3, pp. 383-392.*
- B. De Moor** (1993). *Total least squares for affinely structured matrices and the noisy realization problem. Accepted for publication in IEEE J. on Signal Processing.*
- B. De Moor, M. Gevers and G. Goodwin** (1991). *Overbiased, underbiased and unbiased estimation of transfer functions. Proc. 1<sup>st</sup> ECC, Grenoble, France, pp. 1372-1377.*
- J.E. Dennis, Jr., D.M. Gay and R.E. Welsch** (1981). *An adaptive nonlinear least-squares algorithm. ACM Trans. on Mathematical Software, Vol. 7, No. 3, pp. 348-368.*
- D.K. de Vries and P.M.J. van den Hof** (1993). *Quantification of uncertainty in transfer function estimation : A mixed deterministic-probabilistic approach. Preprints of 12<sup>th</sup> IFAC World Congress, Sydney, Australia, Vol. 8, pp. 157-160.*
- J.C. Doyle** (1982). *Analysis of feedback systems with structured uncertainties. IEE Proc., Vol. 129, Pt. D., No. 6, pp. 242-250.*
- J.C. Doyle** (1985). *Structured uncertainty in control system design. Proc. 24<sup>th</sup> CDC, Ft. Lauderdale, Florida, pp. 260-265.*
- J.C. Doyle** (1987). *A review of  $\mu$  : For case studies in robust control. Proc. IFAC '87, pp. 395-402.*
- J.C. Doyle, B.A. Francis and A.R. Tannenbaum** (1992). *Feedback control theory. Macmillan, New York.*
- J.C. Doyle, K. Glover, P.P. Khargonekar and B.A. Francis** (1989). *State-space solutions to standard  $H_2$  and  $H_\infty$  control problems. IEEE Trans. on Aut. Control, Vol. 34, No. 8, pp. 831-847.*
- L. Dugard and I.D. Landau** (1980). *Recursive output error identification algorithms ; Theory and evaluation. Automatica, Vol. 16, pp. 443-462.*
- J.G. Ecker** (1980). *Geometric programming : Methods, computations and applications. SIAM, Rev. 22, pp. 338-362.*
- P. Eykhoff** (1974). *System identification : Parameter and state estimation. Wiley and*



Sons, London.

- J. Falk** (1973). *Global solutions of signomial programs*. Technical Report T-274, George Washington University, Washington D.C.
- H.M. Falkus** (1991).  *$H_\infty$  robust control design for an electromechanical servo system*. Proc. 1<sup>st</sup> ECC, Grenoble, France, pp. 2478-2483.
- H.M. Falkus** (1992). *A comparison of set estimation methods for output error models with bounded errors*. Proc. ACC '92, Chicago Illinois, pp. 2169-2173.
- H.M. Falkus and A.A.H. Damen** (1993). *Robust parameter set estimation : Applied to a watervessel process*. Preprints of 12<sup>th</sup> IFAC World Congress, Sydney, Australia, Vol. 5, pp. 59-62.
- H.M. Falkus and A.A.H. Damen** (1994). *Multivariable  $H_\infty$  control design toolbox : User manual*. EUT Report 94-E-282, Eindhoven University of Technology, The Netherlands.
- H.M. Falkus, A.A.H. Damen and A.C.P.M. Backx** (1993). *Identification of a tube glass production process : Point vs. set estimation*. Proc. 2<sup>nd</sup> ECC, Groningen, The Netherlands, pp. 2344-2349.
- H.M. Falkus, A.A.H. Damen and A.C.P.M. Backx** (1994a). *Disturbance Modelling in a Tube Glass Production Process*. To be published in proc. ACC '94, Baltimore, Maryland.
- H.M. Falkus, A.A.H. Damen and A.C.P.M. Backx** (1994b). *Minimal polynomial identification : User guide*. To be published as EUT Report, Eindhoven University of Technology, The Netherlands.
- H.M. Falkus, A.A.H. Damen and J. Bouwels** (1992). *General MIMO  $H_\infty$  control design framework*. Proc. 31<sup>st</sup> IEEE CDC, Tucson, Arizona, pp. 2181-2186.
- M.K.H. Fan, A.L. Tits and J.C. Doyle** (1991). *Robustness in the presence of mixed parametric uncertainty and unmodeled dynamics*. IEEE Trans. on Aut. Control, Vol. 36, No. 1, pp. 25-38.
- E. Fogel** (1979). *System identification via membership set constraints with energy constrained noise*. IEEE Trans. on Aut. Control, Vol. AC-24, No. 5, pp. 752-758.
- E. Fogel and Y.F. Huang** (1982). *On the value of information in system identification - Bounded noise case*. Automatica, Vol. 18, No. 2, pp. 229-238.
- B.A. Francis** (1987). *A course in  $H_\infty$ -control theory*. Lecture notes in Control and Information Sciences, Vol. 88, Springer Verlag, London.
- J.P. Garner and J.C. Spall** (1989). *Identification of state-space parameters in the presence of uncertain nuisance parameters*. Proc. ACC '89, pp. 1226-1230.
- A. Gauthier and I.D. Landau** (1978). *On the recursive identification of multi-input multi-output systems*. Automatica, Vol. 14, pp. 609-614.
- M. Gevers** (1991). *Connecting identification and robust control : A new challenge*. 9<sup>th</sup> IFAC/IFORS Symp. on Identification and System Parameter Estimation,

- Budapest, Hungary, pp. 1-10.
- M. Gevers and A.C. Tsoi** (1984). *Structural identification of linear multivariable systems using overlapping forms : a new parametrization*. Int. J. of Control, Vol. 40, No. 5, pp. 971-987.
- M. Gevers and V. Wertz** (1984). *Uniquely identifiable state-space and ARMA parametrizations for multivariable linear systems*. Automatica, Vol. 20, No. 3, pp. 333-347.
- P.E. Gill, W. Murray and M.H. Wright** (1981). *Practical optimization*. Academic Press Inc., San Diego.
- P.E. Gill, W. Murray and M.H. Wright** (1991). *Numerical linear algebra and optimization, Volume 1*. Addison-Wesley, California.
- K. Glover and J.C. Doyle** (1988). *State-space formulae for all stabilizing controllers that satisfy an  $H_\infty$ -norm bound and relations to risk sensitivity*. Systems & Control Letters 11, pp. 167-172.
- G.H. Golub and C.F. van Loan** (1989). *Matrix computations, 2<sup>nd</sup> edition*. John Hopkins University Press, Baltimore.
- G.C. Goodwin** (1985). *Some observations on robust estimation and control*. Proc. IFAC Symp. on Identification and System Parameter Estimation, York, UK, pp. 851-859.
- G.C. Goodwin, M. Gevers and B. Ninness** (1992). *Quantifying the error in estimated transfer functions with application to model order selection*. IEEE Trans. on Aut. Control, Vol. 37, No. 7, pp. 913-928.
- G.C. Goodwin, B. Ninness and M.E. Salgado** (1990). *Quantification of uncertainty in estimation*. Proc. ACC '90, pp. 2400-2405.
- G.C. Goodwin and M.E. Salgado** (1990). *A stochastic embedding approach for quantifying uncertainty in the estimation of restricted complexity models*. Int. J. of Adaptive Control and Signal Processing, Vol. 3, pp. 333-356.
- G.C. Goodwin, M.E. Salgado and D.Q. Mayne** (1989). *A Bayesian approach to estimation with restricted complexity models*. Internal Report EE 8953, Department of Electrical Engineering & Computer Science, University of Newcastle, Australia.
- G.C. Goodwin and K.S. Sin** (1984). *Adaptive filtering, prediction and control*. Prentice-Hall, Englewood-Cliffs, N.J.
- G. Gu and P.P. Khargonekar** (1992). *A class of algorithms for identification in  $H_\infty$* . Automatica, Vol. 28, No. 2, pp. 229-312.
- R.P. Guidorzi** (1975). *Canonical structures in the identification of multivariable systems*. Automatica, Vol. 11, pp. 361-374.
- R.P. Guidorzi** (1981). *Invariants and canonical forms for systems structural and parametric identification*. Automatica, Vol. 17, No. 1, pp. 117-133.
- R.P. Guidorzi** (1990). *Certain models from uncertain data: The algebraic case*. 2<sup>nd</sup>

SIAM Conf. on LASSC, San Francisco.

- R.P. Guidorzi** (1991a). *Errors-in-variables identification in presence of non-independent noise*. Recent Advances in Mathematical Theory of Systems, Control, Networks and Signal Processing I, MTNS '91, Kobe, Japan, pp. 563-568.
- R.P. Guidorzi** (1991b). *Certain models from uncertain data: The dynamic case*. ICIAM '91, Washington.
- R.P. Guidorzi and S. Beghelli** (1982). *Input-output multistructural models in multivariable systems identification*. IFAC Symp. on Identification and System Parameter Estimation, Washington D.C., USA, pp. 539-543.
- R.P. Guidorzi, M.P. Losito and T. Muratori** (1982). *The range error test in the structural identification of linear multivariable systems*. IEEE Trans. on Aut. Control, Vol. AC-27, No. 5. pp. 1044-1054.
- R.D. Gupta and F.W. Fairman** (1974). *Luenberger's canonical form revisited*. IEEE Trans. on Aut. Control, Vol. AC-19, pp. 440-441.
- R. Haber and H. Unbehauen** (1990). *Structure identification of nonlinear dynamic systems - A survey on input/output approaches*. Automatica, Vol. 26, No. 4, pp. 651-677
- A.K. Hajdasiński** (1980). *Estimation of the order of multivariable dynamical systems*. Journal A, Vol. 21, No. 2, pp. 21-32.
- R.G. Hakvoort** (1991). *Identification of an upper bound for the  $l^1$ -norm of the model uncertainty*. Selected Topics in Identification, Modelling and Control.
- R.G. Hakvoort** (1993). *Worst-case system identification in  $H_\infty$  : Error bounds and optimal models*. Preprints of 12<sup>th</sup> IFAC World Congress, Sydney, Australia, Vol. 8, pp. 161-164.
- E.J. Hannan and M. Deistler** (1988). *The statistical theory of linear systems*. John Wiley & Sons, New York.
- A.J. Helmicki, C.A. Jacobson and C.N. Nett** (1989).  *$H_\infty$  identification of stable LSI systems : A scheme with direct application to controller design*. Proc. ACC '89, pp. 1428-1434.
- A.J. Helmicki, C.A. Jacobson and C.N. Nett** (1990). *Identification in  $H_\infty$  : A robustly convergent, nonlinear algorithm*. Proc. ACC '90, pp. 386-391.
- A.J. Helmicki, C.A. Jacobson and C.N. Nett** (1991a). *Control oriented system identification : A worst-case / deterministic approach in  $H_\infty$* . IEEE Trans. on Aut. Control, Vol. 36, No. 10, pp. 1163-1176.
- A.J. Helmicki, C.A. Jacobson and C.N. Nett** (1991b). *Fundamentals of control oriented system identification and their application for identification in  $H_\infty$* . Proc. ACC '91, pp. 89-99.
- A.J. Helmicki, C.A. Jacobson and C.N. Nett** (1991c). *Identification in  $H_\infty$  : The continuous-time case*. Proc. ACC '91, pp. 1893-1898.
- A.J. Helmicki, C.A. Jacobson and C.N. Nett** (1991d). *Identification in  $H_\infty$  : Linear*

- algorithms. Proc. ACC '91, pp. 2418-2423.
- O. Hernandez-Lerma and J.B. Lasserre** (1990). *Error bounds for rolling horizon policies in discrete-time markov control processes*. IEEE Trans. on Aut. Control, Vol. 35, No. 10, pp. 1118-1124.
- B.L. Ho and R.E. Kalman** (1965). *Effective construction of linear state-variable models from input/output functions*. Regelungstechnik, Vol. 12, pp. 545-548.
- G.H. Hostetter, C.J. Savant, Jr. and R.T. Stefani** (1982). *Design of feedback control systems*. Holt-Saunders, Japan.
- A. İftar and Ü. Özgüner** (1991). *Modeling of uncertain dynamics for robust controller design in state space*. Automatica, Vol. 27, No. 1, pp. 141-146.
- P. Janssen** (1988). *On model parametrization and model structure selection for identification of MIMO systems*. PhD. Thesis, Department of Electrical Engineering, Eindhoven University of Technology, The Netherlands.
- P. Janssen, P. Stoica, T. Söderström and P. Eykhoff** (1988). *Model structure selection for multivariable systems by cross-validation methods*. Int. J. of Control, Vol. 47, No. 6, pp. 1737-1758.
- L. Jaulin and E. Walter** (1993). *Set inversion via interval analysis for nonlinear bounded-error estimation*. Automatica, Vol. 29, No. 4, pp. 1053-1064.
- B.Z. Kacwicz** (1991). *Solving linear problems with uncertain information*. 9<sup>th</sup> IFAC/IFORS Symp. on Identification and System Parameter Estimation, Budapest, Hungary, pp. 868-871.
- T. Kailath** (1980). *Linear Systems*. Prentice Hall, Englewood Cliffs, N.J.
- L.H. Keel, S.P. Bhattacharyya and J.W. Howze** (1988). *Robust control with structured perturbations*. IEEE Trans. on Aut. Control, Vol. 33, No. 1, pp. 68-78.
- K.J. Keesman** (1990). *Membership-set estimation using random scanning and principal component analysis*. Math. and Comp. in Simulation, Vol. 32, No. 5 & 6, pp. 535-544.
- K.J. Keesman** (1991). *Robust identification and prediction for output-error models*. 9<sup>th</sup> IFAC/IFORS Symp. on Identification and System Parameter Estimation, Budapest, Hungary, pp. 878-881.
- K.J. Keesman** (1992). *Determination of a minimum-volume orthotopic enclosure of a finite vector set*. Internal Report MRS-92-1, Dept. of Agriculture Engineering and Physics, Wageningen Agriculture University, The Netherlands.
- K.J. Keesman** (1993). *Set-membership identification and prediction of environmental systems*. First Workshop on Forecasting Environmental Change IIASA, Laxenburg, Austria, pp. 1-10.
- K.J. Keesman and G. van Straten** (1989). *Identification and prediction propagation of uncertainty in models with bounded noise*. Int. J. of Control, Vol. 49, No. 6, pp. 2259-2269.

- K.J. Keesman and G. van Straten** (1990). *Set membership approach to identification and prediction of lake Eutrophication*. Water Resources Research, Vol. 26, No. 11, pp. 2643-2652.
- M. Keulers** (1993). *Identification and control of a fed-batch process ; Application to culture of Saccharomyces cerevisiae*. Ph.D. Thesis, Department of Electrical Engineering, Eindhoven University of Technology, The Netherlands.
- M. Keulers, L. Ariaans and M.L.F. Giuseppin** (1994). *The dynamic step response of a fed-batch bakers' yeast process*. Progress in Biotechnology, Vol. 9, Proc. 6<sup>th</sup> ECB, Florence, Italy, 1993, pp. 879-882.
- M. Khammash and J.B. Pearson** (1991). *Performance robustness of discrete-time systems with structured uncertainty*. IEEE Trans. on Aut. Control, Vol. 36, No. 4, pp. 398-412.
- H. König and D. Pallaschke** (1981). *On Khachian's algorithm and minimal ellipsoids*. Numerische Mathematik 36, pp. 211-223.
- R.L. Kosut** (1988). *Adaptive control via parameter set estimation* Int. J. of Adaptive control and Signal Processing, Vol. 2, pp. 371-399.
- R.L. Kosut, M.K. Lau and S.P. Boyd** (1990). *Identification of systems with parametric and nonparametric uncertainty*. Proc. ACC '90, pp. 2412-2417.
- R.L. Kosut, M.K. Lau and S.P. Boyd** (1991). *System identification for robust control design*. Proc. 1<sup>st</sup> ECC, Grenoble, France, pp. 1384-1389.
- R.L. Kosut, M.K. Lau and S.P. Boyd** (1992). *Set-membership identification of systems with parametric and nonparametric uncertainty*. IEEE Trans. on Aut. Control, Vol. 37, No.7, pp. 929-941.
- B. Kouvaritakis and M.S. Trimboli** (1990). *Bounded-error data and frequency response design*. Math. and Comp. in Simulation, Vol. 32, No. 5 & 6, pp. 597-608.
- J.M. Krause and P.P. Khargonekar** (1990). *Parameter identification in the presence of non-parametric dynamic uncertainty*. Automatica, Vol. 26, No. 1, pp. 113-123.
- E. Kreyszig** (1993). *Advance engineering mathematics, 7<sup>th</sup> edition*. John Wiley & Sons, New York.
- H. Kwakernaak** (1993). *Robust control and  $H_\infty$ -optimization - Tutorial paper*. Automatica, Vol. 29, No. 2, pp. 255-273.
- H. Lahanier, E. Walter and R. Gomeni** (1987). *OMNE: A new robust membership-set estimator for the parameters of nonlinear models*. J. of Pharmaccokinetics and Biopharmaceutics, Vol. 15, No. 2, pp. 203-219.
- R.O. Lemaire, L. Valavani, M. Athans and G. Stein** (1991). *A frequency-domain estimator for use in adaptive control systems*. Automatica, Vol. 27, pp.23-38.
- P.F. Lambrechts, J. Terlouw, S. Bennani and M Steinbuch** (1992). *Parametric uncertainty modelling using LFTs*. Selected Topics in Identification, Modelling and Control, Vol. 5, pp. 1-10.

- M.K. Lau, S.P. Boyd, R.L. Kosut and G.F. Franklin (1991a). *A robust control design for FIR plants with parameter set uncertainty*. Proc. ACC '91, pp. 83-88.
- M.K. Lau, S. Boyd, R.L. Kosut and G.F. Franklin (1991b). *Robust control design for ellipsoidal plant set* Proc. 30<sup>th</sup> CDC, Brighton, England, pp. 291-296.
- M.K. Lau, R.L. Kosut and S. Boyd (1990). *Parameter set estimation of systems with uncertain nonparametric dynamics and disturbances*. Proc. 29<sup>th</sup> CDC, Honolulu, Hawaii, pp. 3162-3167.
- D.L. Laughlin, K.G. Jordan and M. Morari (1986). *Internal model control and process uncertainty : Mapping uncertainty regions for SISO controller design*. Int. J. of Control, Vol. 44, No. 6, pp. 1675-1698.
- W.R.H.M. Liebregts (1991). *A new identification technique for  $H_\infty$ -robust control design : Identification of the watervessel process*. M.Sc. Thesis, Department of Electrical Engineering, Eindhoven University of Technology, The Netherlands.
- L. Ljung (1978). *Convergence analysis of parametric identification methods*. IEEE Trans. on Aut. Control, Vol. AC-23, No. 5, pp. 770-783.
- L. Ljung (1985). *Asymptotic variance expressions for identified black-box transfer function models*. IEEE Trans. on Aut. Control, Vol. AC-30, pp. 834-844.
- L. Ljung (1987). *System identification : Theory for the user*. Prentice-Hall, Engelwood Cliffs, N.J.
- L. Ljung and P.E. Caines (1979). *Asymptotic normality of prediction error estimators for approximate system models*. Stochastics, Vol. 3, pp. 29-46.
- L. Ljung and K. Glover (1981). *Frequency domain versus time domain methods in system identification*. Automatica, Vol. 17, No. 1, pp. 71-86.
- L. Ljung, B. Wahlberg and H. Hjalmarsson (1991). *Model quality : The roles of prior knowledge and data information*. Proc. 30<sup>th</sup> CDC, Brighton, England, pp. 273-277.
- R. Lozano-Leal and R. Ortega (1987). *Reformulation of the parameter identification problem for systems with bounded disturbances*. Automatica, Vol. 23, No. 2, pp. 247-251.
- D.G. Luenberger (1973). *Introduction to linear and nonlinear programming*. Addison-Wesley, California.
- J. Lunze (1988). *Robust multivariable feedback control*. Prentice-Hall, New York.
- J.M. Maciejowsky (1989). *Multivariable feedback design*. Addison-Wesley, California.
- D.Q. Mayne and F. Firoozan (1982). *Linear identification of ARMA processes*. Automatica, Vol. 18, No. 4, pp. 461-466.
- W. Mendenhall and T. Sincich (1988). *Statistics for the engineering and computer sciences, 2<sup>nd</sup> edition*. Dellen Publishing Company, San Francisco, California.
- Y.A. Merkurjev and V.A. Popov (1991). *Identification of linear objects with bounded disturbances in both input and output channels*. 9<sup>th</sup> IFAC/IFORS Symp. on Identification and System Parameter Estimation, Budapest, Hungary,

- pp. 1733-1737.
- M. Milanese** (1989). *Estimation and prediction in the presence of unknown but bounded uncertainty : A survey*. Robustness in Identification and Control, Plenum Press, New York.
- M. Milanese and G. Belforte** (1982). *Estimation theory and uncertainty intervals evaluation in presence of unknown but bounded errors : Linear families of models and estimators*. IEEE Trans. on Aut. Control, Vol. AC-27, No. 2, pp. 408-414.
- M. Milanese and R. Tempo** (1985). *Optimal algorithms theory for robust estimation and prediction* IEEE Trans. on Aut. Control, Vol. AC-30, No.8, pp. 730-738.
- M. Milanese, R. Tempo and A. Vicino (Eds.)** (1988). *Robustness in identification and control*. Applied Information Technology, Plenum Press, New York.
- M. Milinase and A. Vicino** (1989). *Robust estimation and exact uncertainty intervals evaluation for nonlinear models*. System Modelling and Simulation, Amsterdam, pp. 91-96.
- M. Milanese and A. Vicino** (1991a). *Optimal estimation theory for dynamic systems with set membership uncertainty : An overview*. Automatica, Vol. 27, No. 6, pp. 997-1009.
- M. Milanese and A. Vicino** (1991b). *Estimation theory for dynamic systems with unknown but bounded uncertainty : An overview*. 9<sup>th</sup> IFAC/IFORS Symp. on Identification and System Parameter Estimation, Budapest, Hungary, pp. 859-867.
- S.H. Mo** (1989). *Computation of bounded-parameter models of dynamical systems from bounded-noise records*. Ph.D. Thesis, School of Electronic & Electrical Engineering, University of Birmingham.
- S.H. Mo and J.P. Norton** (1990). *Fast and robust algorithm to compute exact polytope bounds*. Math. and Comp. in Simulation, Vol. 32, No. 5 & 6, pp. 481-494.
- M. Morari and E. Zafiriou** (1989). *Robust process control*. Prentice-Hall, Englewood Cliffs, N.J.
- T.S. Motzkin, H. Raiffa, G.L. Thompson and R.M. Thrall** (1953). *The double description method*. Anals of Math. Study, 28, pp. 51-73.
- J.P. Norton** (1975). *Optimal smoothing in the identification of linear time-varying systems*. Proc. IEE, Vol. 122, No. 6, pp. 663-668.
- J.P. Norton** (1986). *An Introduction to Identification*. Academic Press
- J.P. Norton** (1987a). *Identification and application of bounded-parameter models*. Automatica, Vol. 23, No. 4, pp. 497-507.
- J.P. Norton** (1987b). *Identification of parameter bounds for ARMAX models from records with bounded noise*. Int. J. of Control, Vol. 45, No. 2, pp. 273-390.
- J.P. Norton** (1989). *Recursive computation of inner bounds for the parameters of linear models*. Int. J. of Control, Vol. 50, No. 6, pp. 2423-2430.

- J.P. Norton and S.H. Mo** (1990). *Parameter bounding for time-varying systems*. Math. and Comp. in Simulation, Vol. 32, No. 5 & 6, pp. 527-534.
- J.P. Norton and S.M. Veres** (1991). *Identification of nonlinear state-space models by deterministic search*. 9<sup>th</sup> IFAC/IFORS Symp. on Identification and System Parameter Estimation, Budapest, Hungary, pp. 363-368.
- A. Packard and J. Doyle** (1993). *The complex structured singular value*. Automatica, Vol. 29, No. 1, pp. 71-109.
- P.J. Parker and R.R. Bitmead** (1987). *Adaptive frequency response identification*. Proc. 26<sup>th</sup> CDC, pp. 348-353.
- H. Piet-Lahanier and E. Walter** (1988). *Practical implementation of an exact and recursive algorithm for characterizing likelihood sets*. Proc. 12<sup>th</sup> IMACS World Congress, Vol. 2, Paris, pp. 481-483.
- H. Piet-Lahanier and E. Walter** (1989). *Further results on recursive polyhedral description of parameter uncertainty in the bounded-error context*. Proc. 28<sup>th</sup> CDC, Tampa, Florida, pp. 1964-1966.
- H. Piet-Lahanier and E. Walter** (1990a). *Exact recursive characterization of feasible parameter sets in the linear case*. Math. and Comp. in Simulation, Vol. 32, No. 5 & 6, pp. 495-504.
- H. Piet-Lahanier and E. Walter** (1990b). *Characterization of non-connected parameter uncertainty regions*. Math. and Comp. in Simulation, Vol. 32, No. 5 & 6, pp. 553-560.
- L. Pronzato and E. Walter** (1984). *A general-purpose global optimizer : Implementation and applications*. Math. and Comp. in Simulation 26, pp. 412-422.
- L. Pronzato and E. Walter** (1990). *Experiment design for bounded-error models*. Math. and Comp. in Simulation, Vol. 32, No. 5 & 6, pp. 571-584.
- L. Pronzato and E. Walter** (1991a). *Sequential experimental design for parameter bounding*. Proc. of the 1<sup>st</sup> ECC, Grenoble, France, pp. 1181-1186.
- L. Pronzato and E. Walter** (1991b). *Robustness to outliers of bounded-error estimators ; Consequences on experiment design*. 9<sup>th</sup> IFAC/IFORS Symp. on Identification and System Parameter Estimation, Budapest, Hungary, pp. 821-826.
- L. Pronzato, E. Walter and H. Piet-Lahanier** (1989). *Mathematical equivalence of two ellipsoidal algorithms for bounded-error estimation*. Proc. 28<sup>th</sup> CDC, Tampa, Florida, pp. 1952-1955.
- A.K. Rao and Y.F. Huang** (1990). *Recent developments in optimal bounding ellipsoidal parameter estimation*. Math. and Comp. in Simulation, Vol. 32, No. 5 & 6, pp. 515-526.
- H.F. Raynaud and S. Besset** (1991).  *$H_\infty$  - Ellipsoidal controller - An approach to robustness with respect to parameter uncertainties*. Proc. 1<sup>st</sup> ECC, Grenoble,



- France, pp. 2601-2604.
- D.E. Rivera, J.F. Pollard, L.E. Stermann and C.E. Garcia** (1990). *An industrial perspective on control-relevant identification*. Proc. ACC '90, pp. 2406-2411.
- L.E. Scales** (1985). *Introduction to Non-Linear Optimization*. MacMillan Publishers Ltd, Hong Kong.
- R.J.P. Schrama** (1992). *Approximate identification and control design*. PhD. Thesis, Delft University of Technology, The Netherlands.
- F.C. Schweppe** (1968). *Recursive state estimations : Unknown but bounded errors and system inputs*. IEEE Trans. on Aut. Control, Vol. AC-13, pp. 22-28.
- F.C. Schweppe** (1973). *Uncertain dynamic systems*. Prentice Hall, Englewood Cliffs, N.J.
- F.G. Shinskey** (1984). *Distillation Control, 2<sup>nd</sup> Edition*. MacGraw-Hill, New York.
- N.K. Sinha and Y.H. Kwong** (1979). *Recursive estimation of the parameters of linear multivariable systems*. Automatica, Vol. 15, pp. 471-475.
- S. Skogestad** (1993). *Dynamics and control of distillation columns - A critical survey*. Proc. IFAC Dynamics and Control of Chemical Reactors, Distillation Columns and Batch Processes, Vol. 2, Dycord '92, pp. 11-35.
- M.K. Smit** (1983). *A novel approach to the solution of indirect measurement problems with minimal error propagation*. Measurement 1, pp. 181-190.
- M.K. Smit and J.W. Verhoof** (1990). *A bounded-error approach to accuracy analysis in ellipsometry*. Math. and Comp. in Simulation, Vol. 32, No. 5 & 6, pp. 545-552.
- R.S. Smith and M. Dahleh (Eds.)** (1994). *The modeling of uncertainty in control systems. Proceedings of the 1992 Santa Barbara workshop*. Lecture Notes in Control and Information Sciences 192, Springer Verlag.
- R.S. Smith and J.C. Doyle** (1989). *Model invalidation : A connection between robust control and identification*. Proc. ACC '89, pp. 1435-1440.
- R.S. Smith and J.C. Doyle** (1990). *Towards a methodology for robust parameter identification*. Proc. ACC '90, pp. 2394-2399.
- R.S. Smith, J.C. Doyle, M. Morari and A. Skjellum** (1987). *A case study using  $\mu$  : Laboratory process control problem*. Proc. IFAC '87, pp. 403-415.
- T. Söderström** (1977). *On model structure testing in system identification*. Int. J. of Control, Vol. 26, No. 1, pp. 1-18.
- T. Söderström and P. Stoica** (1989). *System identification*. Prentice-Hall, Englewood Cliffs, New Jersey.
- K. Steiglitz and L.E. McBride** (1965). *A technique for the identification of linear systems*. IEEE Trans. on Aut. Control, Vol. AC-10, pp. 461-464.
- M. Steinbuch, S.G. Smit, G. Schootstra and O.H. Bosgra** (1991).  *$\mu$ -Synthesis of a flexible mechanical servo system*. Proc. ACC '91, pp. 593-598.
- M. Steinbuch, J.C. Terlouw and O.H. Bosgra** (1991). *Robustness analysis for real*

- and complex perturbations applied to an electro-mechanical system. Proc. ACC '91, pp. 556-561.
- P. Stoica, P. Eykhoff, P. Janssen and T. Söderström** (1986). *Model-structure selection by cross-validation*. Int. J. of Control, Vol. 43, No. 6, pp. 1841-1878.
- R. Tempo and A. Vicino** (1990). *Optimal algorithms for system identification: a review of some recent results*. Math. and Comp. in Simulation, Vol. 32, No. 5 & 6, pp. 584-596.
- J.C. Terlouw** (1990). *Robustness analysis of control systems with parametric uncertainty*. Matlab Technical Note, Nr. 099/90, Philips Research Laboratories Eindhoven, The Netherlands.
- J.C. Terlouw and S.G. Smit** (1990). *Robust stability analysis of a flexible mechanism assuming real or complex parametric uncertainty*. Selected Topics in Identification, Modelling and Control, Vol. 1, pp. 51-55.
- Y. Tomita, A.A.H. Damen and P.M.J. van den Hof** (1992). *Equation error versus output error methods*. Ergonomics, Vol. 35, No. 5/6, pp. 551-564.
- H. Tulleken** (1992). *Grey-box modelling and identification topics*. Ph.D. Thesis, Delft University of Technology, The Netherlands.
- H.J.A.F. Tulleken** (1993). *Grey-box Modelling and Identification Using Physical Knowledge and Bayesian Techniques*. Automatica, Vol. 29, No. 2, pp. 285-308.
- T.J.J. van den Boom** (1992). *System identification for coprime factor plant description with bounded error*. Proc. ACC '92, Chicago, Illinois, pp. 1248-1252.
- T.J.J. van den Boom** (1993a). *MIMO system identification for  $H_\infty$  robust control ; A frequency domain approach with minimum error bounds*. Ph.D. Thesis, Department of Electrical Engineering, Eindhoven University of Technology, The Netherlands.
- T.J.J. van den Boom** (1993b). *MIMO-systems identification by minimum error bounds*. Preprints of 12<sup>th</sup> IFAC World Congress, Sydney, Australia, Vol. 8, pp.153-156.
- T. van den Boom, M. Klompstra and A. Damen** (1991). *System identification for  $H_\infty$ -robust control design*. Preprints 9<sup>th</sup> IFAC/IFORS Symp. on Identification and System Parameter Estimation, pp. 1431-1436.
- P.M.J. van den Hof** (1989). *On residual-based parametrization and identification of multivariable systems*. Ph.D. Thesis, Eindhoven University of Technology, The Netherlands.
- P.M.J. van den Hof** (1994). *Model sets and parametrizations for identification of multivariable equation error models*. Automatica, Vol. 30, No. 3, pp. 433-446.
- S. van Huffel and J. Vandewalle** (1991). *The total least squares problem ; Computational aspects and analysis*. Frontiers in Applied Mathematics, SIAM.
- A.J.M. van Overbeek and L. Ljung** (1982). *On-line structure selection for multivariable state-space models* Automatica, Vol. 18, No. 5, pp. 529-543.
- P. van Overschee and B. De Moor** (1994). *N4SID : Subspace algorithms for the*

- identification of combined deterministic-stochastic systems*. Automatica, Vol. 30, No. 1, pp. 75-93.
- S.M. Veres and J.P. Norton** (1989). *Structure identification of parameter-bounding models by use of noise-structure bounds*. Int. J. of Control, Vol. 50, No. 2, pp. 639-649.
- S.M. Veres and J.P. Norton** (1991a). *Parameter-bounding algorithms for linear errors in variables models*. 9<sup>th</sup> IFAC/IFORS Symp. on Identification and System Parameter Estimation, Budapest, Hungary, pp. 1038-1043.
- S.M. Veres and J.P. Norton** (1991b). *Structure selection for bounded parameter models: Consistency conditions and selection criterion*. IEEE Trans. on Aut. Control, Vol. 36, No. 4, pp. 474-481.
- S.M. Veres and J.P. Norton** (1993). *Predictive self-tuning control by parameter bounding and worst-case design*. Automatica, Vol. 29, No. 4, pp. 911-928.
- M. Verhaegen** (1993). *Application of a subspace model identification technique to identify LTI systems operating in closed-loop*. Automatica, Vol. 29, No. 4, pp. 1027-1040.
- M. Verhaegen** (1994). *Identification of the deterministic part of MIMO state-space models given in innovations form from input-output data*. Automatica, Vol. 30, No. 1, pp. 61-74.
- M. Verhaegen and P. Dewilde** (1992a). *Subspace model identification. Part 1 : The output-error state-space model identification class of algorithms*. Int. J. of Control, Vol. 56, No. 5, pp. 1187-1210.
- M. Verhaegen and P. Dewilde** (1992b). *Subspace model identification. Part 2 : Analysis of the elementary output-error state-space model identification algorithm*. Int. J. of Control, Vol. 56, No. 5, pp. 1211-1241.
- A. Vicino and M. Milanese** (1989). *Optimal inner bounds of feasible parameter set in linear estimation with bounded noise*. Proc. 28<sup>th</sup> CDC, Tampa, Florida, pp. 2576-2580.
- M. Vidyasagar** (1986). *Optimal rejection of persistent bounded disturbances*. IEEE Trans. on Aut. Control, Vol. AC-31, No. 6, pp. 527-534.
- B. Wahlberg and L. Ljung** (1991). *On estimation of transfer function error bounds*. Proc. 1<sup>st</sup> ECC, Grenoble, France, pp. 1378-1383.
- B. Wahlberg and L. Ljung** (1992). *Hard frequency-domain model error bounds from least-squares like identification techniques*. IEEE Trans. on Aut. Control, Vol. 37, No. 7, pp. 900-912.
- E. Walter and H. Piet-Lahanier** (1986a). *Estimation of non-uniquely identifiable parameters via exhaustive modeling and membership theory*. Math. and Comp. in Simulation 28, pp. 479-490.
- E. Walter and H. Piet-Lahanier** (1986b). *Robust nonlinear parameter estimation in the bounded noise case*. Proc. 25<sup>th</sup> CDC, Athene, Greece, pp. 1037-1042.

- E. Walter and H. Piet-Lahanier (1988). *Estimation of the parameter uncertainty resulting from bounded-error data*. Mathematical Biosciences 92, pp. 55-74.
- E. Walter and H. Piet-Lahanier (1989). *Exact recursive polyhedral description of the feasible parameter set for bounded-error models*. IEEE Trans. on Aut. Control, Vol. 34, No. 8, pp. 911-915.
- E. Walter and H. Piet-Lahanier (1990). *Estimation of parameter bounds from bounded-error data : a survey*. Math. and Comp. in Simulation, Vol. 32, No. 5 & 6, pp. 449-468.
- E. Walter and H. Piet-Lahanier (1991). *Recursive robust minimax estimation for models linear in their parameters*. 9<sup>th</sup> IFAC/IFORS Symp. on Identification and System Parameter Estimation, Budapest, Hungary, pp. 763-768.
- A. Weinmann (1991). *Uncertain models and robust control*. Springer-Verlag, London.
- H.S. Witsenhausen (1968). *Sets of possible states of linear states given perturbed observations*. IEEE Trans. of Aut. Control, Vol. AC-13, pp. 556-558.
- L. Xie, C.E. de Souza and M. Fu (1991).  *$H_\infty$  estimation for discrete-time uncertain systems*. Proc. 1<sup>st</sup> ECC, Grenoble, France, pp. 432-437.
- R.C. Younce and C.E. Rohrs (1992). *Identification with nonparametric uncertainty*. IEEE Trans. on Aut. Control, Vol. 37, No. 6.
- P.M. Young and J.C. Doyle (1990). *Computation of  $\mu$  with real and complex uncertainties*. Proc. 28<sup>th</sup> CDC, Honolulu, Hawaii, pp. 1230-1235.
- Y. Zhu and T. Backx (1993). *Identification of multivariable industrial processes for simulation, diagnosis and control*. Advances in Industrial Control, Springer-Verlag.
- Y.C. Zhu (1989). *Estimation of transfer function : Asymptotic theory and a bound of model uncertainty*. Int. J. of Control, Vol. 49, pp. 2241-2258.
- Y.C. Zhu (1990). *Identification and control of MIMO industrial processes : An integration approach*. Ph.D. Thesis, Department of Electrical Engineering, Eindhoven University of Technology, The Netherlands.
- Y.C. Zhu, A.C.P.M. Backx and P. Eykhoff (1991). *Multivariable process identification for robust control*. EUT Report 91-E-249, Eindhoven University of Technology, The Netherlands.
- Y.C. Zhu and M.H. Klompstra (1991). *Spectral analysis in linear process identification - a new look at an old problem*. Int. J. of Control, Vol. 53, No. 6, pp. 1449-1465.

# ***Samenvatting***

---

Identificatieprocedures in het tijddomein zijn bestudeerd voor lineaire tijd-invariante multivariabele systemen. Hierbij wordt een model afgeleid met onzekerheidsgrenzen op de modelparameters. Deze parametrische onzekerheid kan zowel stochastisch als deterministisch zijn, afhankelijk van de veronderstellingen die gemaakt zijn m.b.t. de ruis die het systeem verstoort (statistische eigenschappen of een absolute begrenzing).

De klassieke predictiefout-identificatiemethode wordt kort uiteengezet, waarbij eerst de SISO modelstructuren en de (pseudo-) canonieke uitbreidingen naar multivariabele systemen beschreven worden. Deze multivariabele modelstructuren zijn echter niet geschikt voor procesidentificatie in de praktijk. In het bijzonder in een industriële omgeving vanwege de structuuronzekerheid. Daarom is een minimum polynoom-structuur geïntroduceerd welke grote overeenkomsten vertoont met de SISO situatie. Het modelschattingsprobleem wordt opgelost m.b.v. niet-lineaire kleinste-kwadraten optimalisatie. Hierbij wordt de som van de gekwadrateerde predictiefouten geminimaliseerd, gebruik makend van analytische eerste en tweede afgeleiden die zeer efficiënt berekend kunnen worden. In de klassieke aanpak is de veronderstelling noodzakelijk dat het proces in de gedefinieerde modelverzameling zit om stochastische onzekerheidsgrenzen te kunnen bepalen. In de praktijk kan echter slechts een lage orde procesbenadering geschat worden. Voor de situatie dat het proces niet in de modelverzameling zit, wordt een benadering van de covariantie matrix afgeleid, die gebruikt

kan worden om stochastische onzekerheidsgrenzen op de parameters te berekenen. Deterministische onzekerheden op de modelparameters worden verkregen m.b.v. begrensde foutidentificatie of setschattingen. De veronderstelling van begrensde ruis (predictiefout), die het proces verstoort, wordt vertaald naar begrenzingen in de parameter ruimte die de toegestane parameterverzameling definiëren. Om te garanderen dat een convexe en aaneengesloten parameterverzameling (polytope) in de parameter ruimte gedefinieerd wordt, moeten de begrenzingen lineair in de modelparameters zijn. Voor modellen die niet-lineair in de parameters zijn, wordt een lineaire benadering van de begrenzingen gegeven. De parameterverzameling die begrensd wordt door een polytope kan niet exact beschreven worden voor complexe identificatieproblemen. Daarom wordt de voorkeur gegeven aan eenvoudigere, maar benaderende beschrijvingen van de parameterbegrenzingen, zoals ellipsoides of rechthoeken. De minst conservatieve benadering van de parameterverzameling wordt verkregen met een ellips georiënteerde rechthoek die de voordelen van beide methodieken combineert. Desalniettemin is de toepassing van deze identificatiemethode in de praktijk beperkt, als gevolg van pieken in de data die niet voldoen aan de gedefinieerde foutgrenzen. Bovendien is deze methode zeer gevoelig voor overparametrisatie. Uitschieters in de data kunnen veroorzaakt worden door fouten tijdens de data acquisitie, de data voorbewerking, maar ook door te optimistische foutgrenzen of niet gemodelleerde dynamica. Om te voorkomen dat de parameterverzameling leeg wordt, is een robuuste uitbreiding van deze methode voorgesteld. Hierbij wordt een referentiemodel gebruikt om de datapieken in de begrenzingen te detecteren. De gevoeligheid voor overparametrisatie ontstaat door het conservatisme van de setschattingsmethodiek zelf en door de conservatieve beschrijving van de parameterverzameling. Indien de parameteronzekerheden zeer groot zijn, geeft de daarbij behorende centrale schatting geen nauwkeurige beschrijving van het proces. Dit conservatisme kan beperkt worden, door de parameters met de grootste onzekerheid te fixeren. Hierdoor zal de onzekerheid in de overige parameters kleiner worden en de centrale schatting verbeteren. Deze moeilijkheden beperken de toepasbaarheid van de identificatie met begrensde ruis in de praktijk, met name in een industriële omgeving. Voor de compatibiliteit met robuust regelaarontwerp, wordt de parametrische onzekerheid in de polynoommodellen, vertaald in een onzekerheidsmodel met een LFT representatie. De parameterfouten worden weergegeven in een diagonaal matrix en het nominale model samen met de verbindingsmatrices (tussen de parameterfouten en het nominale model) worden gecombineerd in een samengesteld toestandsruimtemodel. De voorgestelde identificatiemethodieken zijn getest op verschillende praktijk voorbeelden, die variëren van SISO laboratoriumprocessen tot multivariabele industriële productieprocessen. In deze voorbeelden is de aandacht vooral gericht op de toepassing van de twee identificatiemethodieken in de praktijk om daarmee de voor- en nadelen aan te tonen. Het vergelijken van beide methodieken geeft aan dat de predictiefout aanpak de voorkeur verdient voor de identificatie van industriële processen.

# Curriculum Vitae

---

Falkus, Heinz Manfred

21 augustus 1966

Geboren te Oldenburg.

mei 1984

Eindexamen Atheneum aan het Eijkhagencollege te Landgraaf.

september 1984 - augustus 1990

Studie Elektrotechniek aan de Technische Universiteit Eindhoven. Afgestudeerd bij de vakgroep Meten en Regelen.

juli 1988 - december 1988

Werkzaam bij IBM Canada HQ. Ltd, Information Systems onder leiding van E.E. Embacher.

september 1989 - juni 1990

Werkzaam bij Philips, Centrum voor Fabricage Technologie (CFT) : *Toepassing van  $H_{\infty}$  regeltechniek op een elektro-mechanisch servo-systeem*, onder leiding van prof.dr.ir. P. Eykhoff en dr.ir. K. Machielsen.

juli 1990 - augustus 1990

Student assistent bij de vakgroep Meten en Regelen : *Implementatie van  $H_\infty$  regeltechniek algoritmen t.b.v. laboratorium processen.*

september 1990 - september 1994

Promotiewerk bij de vakgroep Meten en Regelen, Technische Universiteit Eindhoven, onder leiding van prof.dr.ir. A.C.P.M. Backx en dr.ir. A.A.H. Damen : *Identificatie t.b.v. robuust regelen.*



STELLINGEN

behorende bij het proefschrift :

***PARAMETRIC UNCERTAINTY  
IN SYSTEM IDENTIFICATION***

van

**Heinz Falkus**

Eindhoven, 22 september 1994

1. Ondanks veel onderzoek op het gebied van robuuste regeltechniek en identificatie, zullen modelverzamelingen met een uniforme weging van de elementen te ruim blijven, om tegelijkertijd stabiliteit en hoge prestaties te garanderen voor de gehele modelverzameling. De oplossing kan gevonden worden in een combinatie met stochastiek, die een regelaarontwerp mogelijk maakt op basis van een modelverzameling met statistisch gewogen elementen.
2. Predictiefout-identificatie onder de aanname dat het proces in de modelverzameling zit, conflicteert met de aangenomen voorkennis voor robuust regelaarontwerp. Dit reduceert redelijk recente en vaak geciteerde boeken over systeemidentificatie tot theoretische beschouwingen.
  - L. Ljung (1987). *System Identification : Theory for the user*. Prentice Hall, Englewood Cliffs, N.J.
  - T. Söderström and P. Stoica (1989). *System Identification*. Prentice Hall, Englewood Cliffs, N.J.
3. In de literatuur wordt identificatie met gebruikmaking van set-schattingen, gepresenteerd als een alternatief voor predictiefout-identificatie. De praktijk leert echter, dat deze technieken als wederzijds ondersteunend gebruikt moeten worden.
  - M. Milanese (1989). *Estimation and prediction in the presence of unknown but bounded uncertainty : A survey*. Robustness in Identification and Control, Plenum Press, N.Y.
  - E. Walter and H. Piet-Lahanier (1990). *Estimation of parameter bounds from bounded-error data : A survey*. Mathematics and Computers in Simulation, Vol. 32, No. 5 & 6, pp. 449-468
  - Dit proefschrift.
4. In de praktijk kan geen enkele garantie gegeven worden, dat een voor regeldoelinden gespecificeerde modelonzekerheid ten opzichte van een nominaal systeemgedrag, niet overschreden wordt.
5. De nauwkeurigheid van een geïdentificeerd model is fundamenteel begrensd door de kwaliteit en de kwantiteit van de data.

6. Nieuwe technieken zullen in de industrie alleen worden toegepast, indien economische belangen de drempel van de introductie van deze technologieën overwinnen.
7. In het algemeen zijn vertragingen (looptijden) rampzalig voor de karakteristieken van het teruggekoppelde proces. Anderzijds kunnen deze vertragingen juist compensatie van verstoringen met een voorwaartse regeling mogelijk maken.
  - H.M. Falkus, A.A.H. Damen and A.C.P.M. Backx (1994). *Disturbance modelling in a tube glass production process*. To be published in proc. ACC '94, June 29 - July 1, Baltimore, Maryland.
8. Met het presenteren van de toestandsruimte oplossing, is voor wiskundigen het  $H_\infty$  regelaarontwerp opgelost. Voor regeltechnici daarentegen begint dan pas het probleem om een regelaar te ontwerpen die aan de specificaties voldoet.
  - K. Glover and J. Doyle (1988). *State-space formulae for all stabilizing controllers that satisfy an  $H_\infty$ -norm bound and relations to risk sensitivity*. Systems & Control Letters, Vol 11, pp. 167-172
  - H.M. Falkus, A.A.H. Damen and J. Bouwels (1992). *General MIMO  $H_\infty$  control design framework*. Proc. 31<sup>st</sup> IEEE CDC, December 16-18, Tucson, Arizona, pp. 2181-2186.
9. De spontaniteit en interesse bij de jeugd tijdens demonstraties van laboratorium opstellingen, is omgekeerd evenredig met hun leeftijd.
10. De voortdurende kortingen enerzijds op technisch onderwijs en technische onderzoeksinstellingen en de aanhoudende pogingen anderzijds om het volgen van een technische opleiding te stimuleren, geven aan dat politici geen consistent beleid voeren.
11. Een verenigd Europa is net als een sprookje : het is mooi maar zal nooit werkelijkheid worden. Nationale belangen zullen, vooral wanneer het slechter gaat met de economie, een volledig geïntegreerde samenwerking verhinderen.

12. De impliciete verplichting tot het gebruik van moderne tekstverwerkers, tekenprogramma's en geavanceerde afdrukapparatuur leidt tot een onevenwichtige tijdsverdeling bij het afronden van een promotie door de beperkingen van deze middelen.
13. Van de racketsporten badminton, squash, tafeltennis en tennis, is squash het meest geschikt om frustraties, ontstaan tijdens de afronding van een promotie, af te reageren.
14. Het afronden van een promotie met een proefschrift, is in zichzelf geen reden tot stress. Een ½ woord per minuut is voldoende om in 9 maanden met een werkweek van 48 uur een proefschrift van 60000 woorden te schrijven.
15. De door de leescommissie van dit proefschrift gesuggereerde correctie van het woord "mandril" in "mandrill" is onjuist. Een "mandrill" is een Afrikaanse aap. Een "mandril" (voorkeursspelting "mandrel") daarentegen is een nauwkeurig gepositioneerde buis.
  - The compact edition of the Oxford English dictionary (1971), Vol. 1, Oxford University Press.
  - Webster's new twentieth century dictionary : Unabridged (1978), second edition, World Publishing Co., Inc.