

# Systemidentifikation und Regelung in der Medizin

## 6. Vorlesung Modellvalidierung

Sommersemester 2020

16. Juni 2020

Thomas Schauer

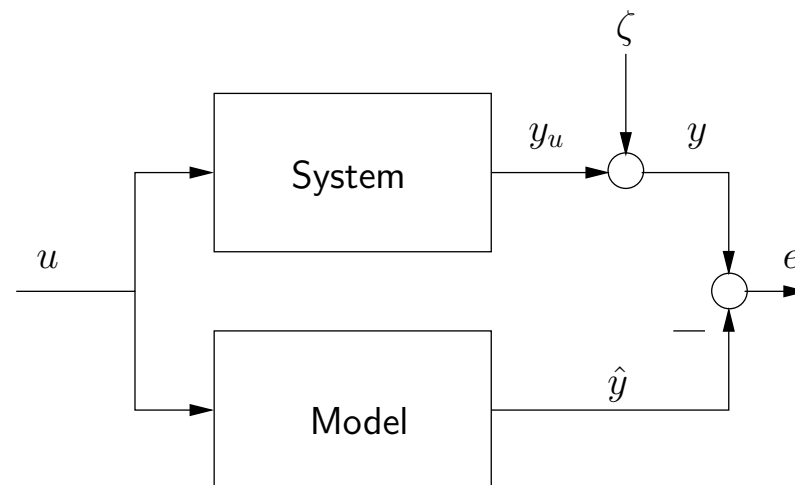
Technische Universität Berlin  
Fachgebiet Regelungssysteme

---

## 6. Modellvalidierung - Strukturbestimmung

### 6.1 Bias-Varianz-Dilemma

- Bisher Annahme der perfekten Modellstruktur
- In der Praxis muss die Struktur bei der Identifikation mit bestimmt werden.  
⇒ Frage nach der Modellkomplexität (Anzahl der Modellparameter (Freiheitsgrade))



- Analyse des Erwartungswertes des Modellfehlers  $y - \hat{y}$ :

$$\begin{aligned}\mathcal{E} \{ (y - \hat{y})^2 \} &= \mathcal{E} \{ (y_u + \zeta - \hat{y})^2 \} \\ &= \mathcal{E} \{ (y_u - \hat{y})^2 \} + \underbrace{\mathcal{E} \{ (\zeta)^2 \}}_{\text{Rauschvarianz}}\end{aligned}$$

??? Where was this?

Rauschvarianz

Bemerkung:  $\mathcal{E} \{ 2(y_u - \hat{y})\zeta \}$  verschwindet, da keine Korrelation von  $\zeta$  mit  $\hat{y}$  (Simulation hier) bzw.  $y_u$  vorliegt und  $\mathcal{E} \{ \zeta \} = 0$ .

- $\mathcal{E} \{ (\zeta)^2 \}$  kann nicht durch das Modell beeinflusst werden.
- $\mathcal{E} \{ (y_u - \hat{y})^2 \}$  wird minimal für ein perfektes Modell.
- Weiteres Zerlegen von  $\mathcal{E} \{ (y_u - \hat{y})^2 \}$  liefert mittels Verschiebungssatz von Steiner für eine Zufallsgröße  $X$  ( $\mathcal{E} \{ X^2 \} = \mathcal{E} \{ X \}^2 + \underbrace{\mathcal{E} \{ (X - \mathcal{E} \{ X \})^2 \}}_{\text{var}(X)}$ )

$$\begin{aligned}\mathcal{E} \{ (y_u - \hat{y})^2 \} &= \mathcal{E} \{ (y_u - \hat{y}) \}^2 + \mathcal{E} \{ (\hat{y} - \mathcal{E} \{ \hat{y} \})^2 \} \\ &= (y_u - \mathcal{E} \{ \hat{y} \})^2 + \mathcal{E} \{ (\hat{y} - \mathcal{E} \{ \hat{y} \})^2 \}\end{aligned}$$

$$(\text{model error})^2 = (\text{bias error})^2 + \text{variance error}$$

- Modellkomplexität hat Einfluss auf die Zusammensetzung des Modellfehlers.
- Bias- und Varianzfehler sind im Konflikt  $\Rightarrow$  nur Kompromiss möglich!

### 6.1.1 Biasfehler

- aufgrund fehlender Modellflexibilität (zu wenig Parameter (Freiheitsgrade))
- systematischer Strukturfehler zwischen System und Modell
- Verkleinerung des Fehlers durch erhöhen der Modellkomplexität

### 6.1.2 Varianzfehler

- Aufgrund verrauschter Messdaten und begrenzter Datensätze
- Der Varianzfehler geht gegen Null für  $K \rightarrow \infty$  ( $K$ -Anzahl der Messwerte).
- Der Varianzfehler ist maximal für  $K = P$  ( $P$ -Anzahl der Modellparameter). In diesem Fall kann das Modell perfekt an die Messwerte angepasst werden und lernt somit das Rauschen auswendig.

- Varianzfehler  $\sim \sigma_{\zeta}^2 \frac{P}{K} \sim \text{Rauschvarianz} \frac{\text{Anzahl der Modellparameter}}{\text{Anzahl der Messungen}}$

## 6.2 Goodness of Fit / Compare Fit

Perfekt bei Wert=0 / Perfekt bei Wert=1

Mean Squared Error:

$$\text{MSE} = \frac{\|\mathbf{y} - \hat{\mathbf{y}}\|_2^2}{K}$$

Normalized Root Mean Squared Error

$$\text{fit}_{\text{GoF},\text{NRMSE}} = \frac{\|\mathbf{y} - \hat{\mathbf{y}}\|_2}{\|\mathbf{y} - \bar{\mathbf{y}}\|_2}, \quad 0 \leq \text{fit}_{\text{GoF},\text{NRMSE}} \leq 1$$

$$\text{fit}_{\text{Compare},\text{NRMSE}} = 1 - \text{fit}_{\text{GoF},\text{NRMSE}} \cdot 100\%, \quad 0 \leq \text{fit}_{\text{GoF},\text{NRMSE}} \leq 100\%$$

Normalized Mean Squared Error

$$\text{fit}_{\text{GoF},\text{NMSE}} = \frac{\|\mathbf{y} - \hat{\mathbf{y}}\|_2^2}{\|\mathbf{y} - \bar{\mathbf{y}}\|_2^2}, \quad 0 \leq \text{fit}_{\text{GoF},\text{NMSE}} \leq 1$$

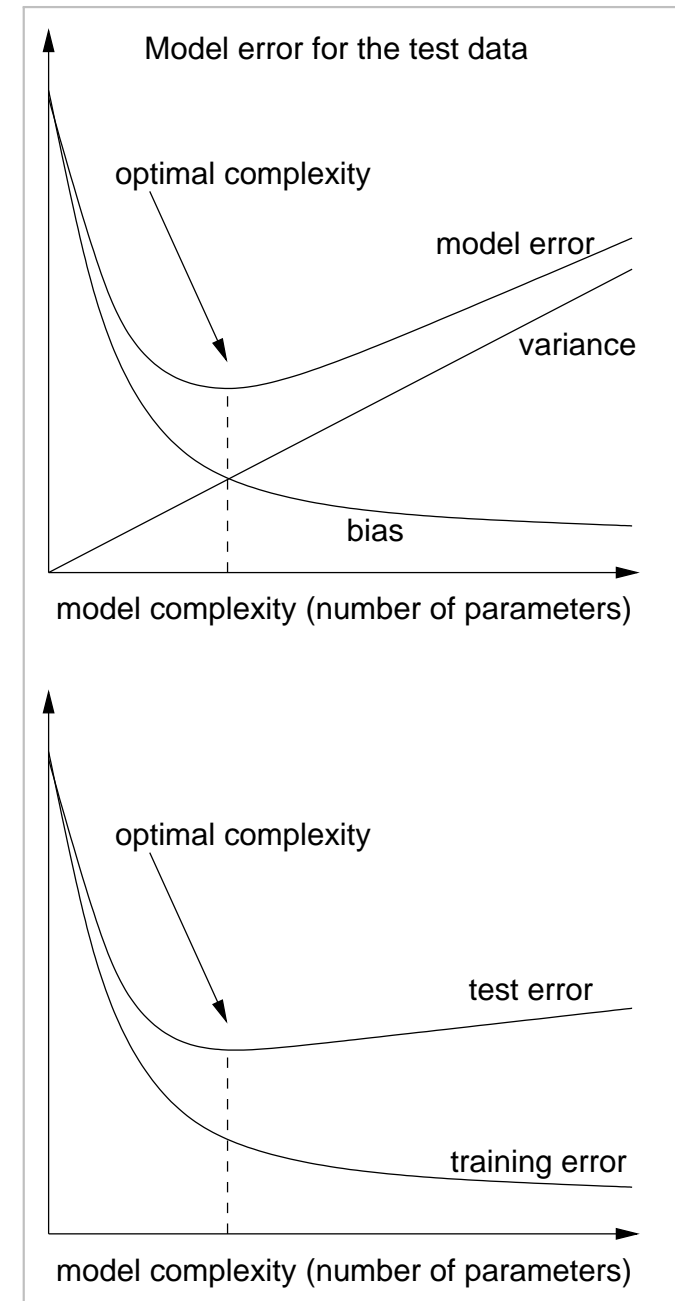
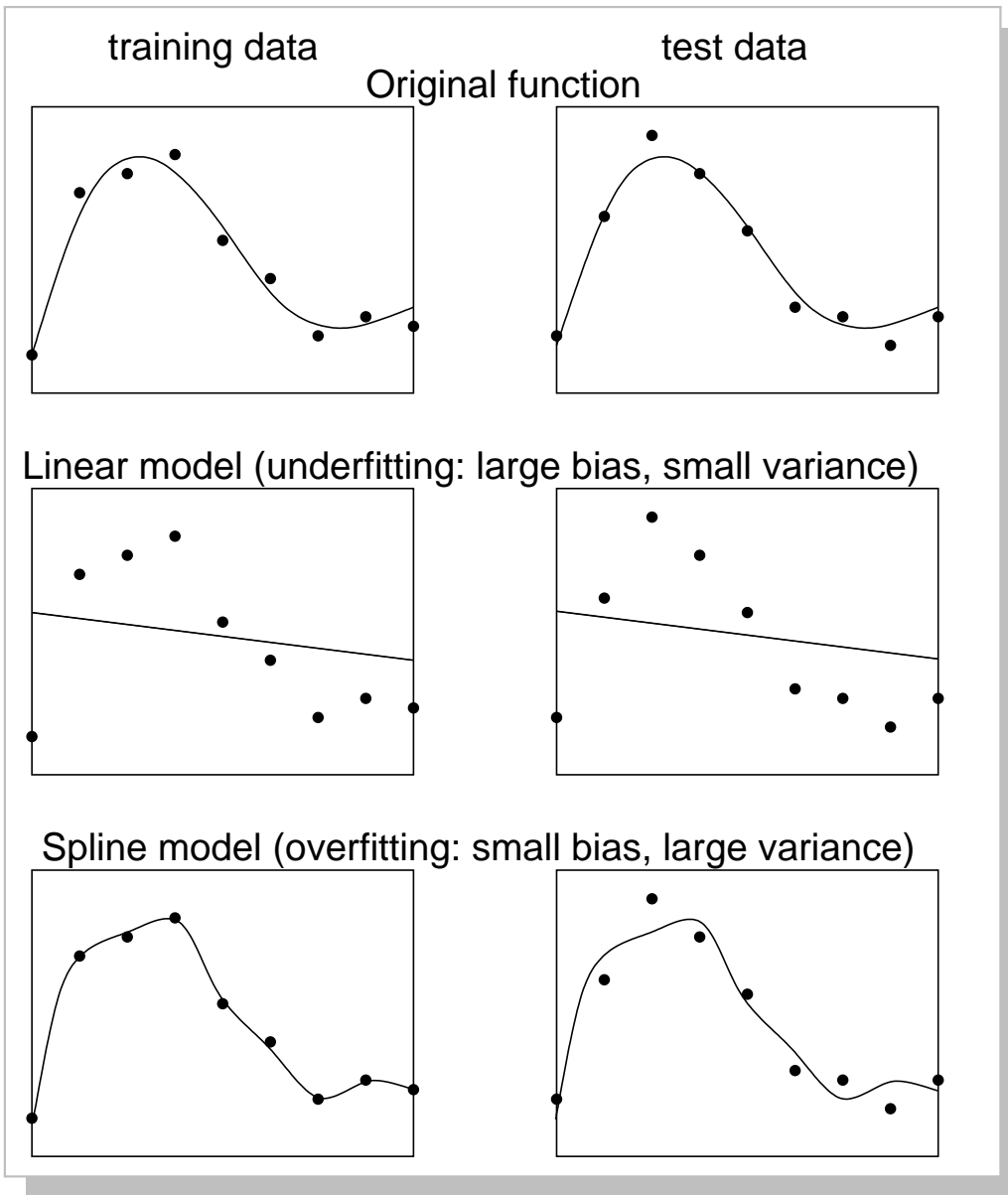
- $K$  - Anzahl der Messungen (Samples),  $\bar{\mathbf{y}}$  - Mittelwert der Ausgangsbeobachtungen  $\mathbf{y}$ ,  $\hat{\mathbf{y}}$  Vektor der Prädiktionen
- $\text{fit}_{\text{GoF},\text{NRMSE}} = 0 \Rightarrow$  perfekter Fit,
- $\text{fit}_{\text{Compare},\text{NRMSE}} = 100\% \Rightarrow$  perfekter Fit

- Goodness of Fit (Funktion `goodnessOfFit` in MATLAB) und Compare Fit (Funktion `compare` in MATLAB) verbessern sich mit größerer Modellkomplexität  $\Rightarrow$  daher nicht geeignet für das Finden der optimalen Modellkomplexität

### **6.3 Kreuzvalidierung (Cross Validation) zur Bestimmung der Modellkomplexität**

Einführung von Trainings- und Testdaten:

- Trainingsdaten: Benutzung für die Parameterschätzung
  - Testdaten: Benutzung für die Modellvalidierung
  - Die beiden Datensätze beinhalten verschiedene Rauschrealisierungen. Die Eingangsgrößenverläufe sollten identisch sein.
-





## 6.4 Information Criteria

- Information criteria = IC(Gütefunktional, Modellkomplexität)
  - ◇ Varianz-Anteil des Modellfehlers wird berücksichtigt.
  - ◇ Das “beste Modell” hat den **kleinsten IC Wert**.
  - ◇ IC **erhöht** sich mit **steigender** Anzahl von Modellparametern  **$P$**  und **verkleinert** sich mit **steigender** Anzahl von Messwerten  **$K$** .
  - ◇ IC wird angewendet, wenn **keine Testdaten** verfügbar sind.
  - ◇ Akaike’s information criterion (AIC):

$$\text{AIC} = K \ln(J_{\text{LS}}(\hat{\Theta})) + 2I$$

- ◇ **Corrected** Akaike’s information criterion ( $\text{AIC}_c$ ) (für  $K/I \leq 40$ )

$$\text{AIC}_c = \text{AIC} + (2I(I + 1))/(K - I - 1)$$

- ◇ Final Prediction Error (FPE):

$$\text{FPE} = J_{\text{LS}}(\hat{\Theta})(1 + I/K)/(1 - I/K)$$

Was ist LS?

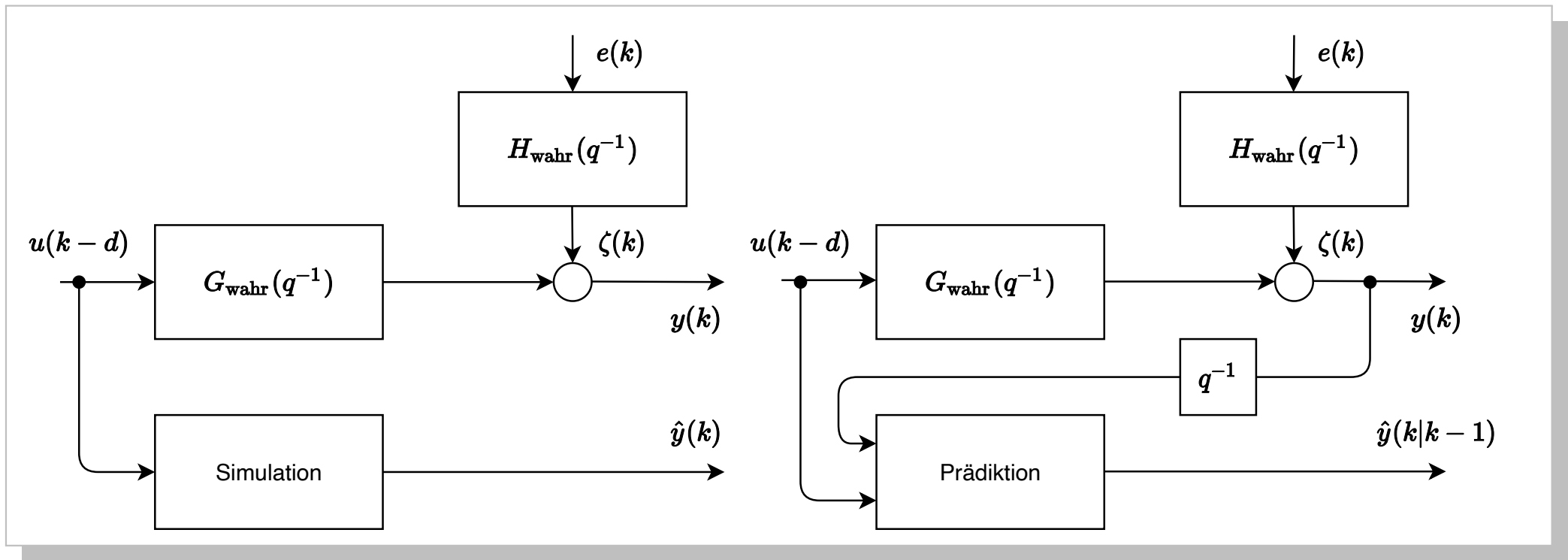
## 6.5 Pol-/Nullstellen-Kürzung

- Pol-/Nullstellen-Kürzungen in der deterministischen Transferfunktion  $G$  bedeuten, dass ein Modell kleinerer Ordnung verwendet werden sollte. Die zusätzlichen Pole wurden nur benutzt, um das Rauschen zu beschreiben → anstelle von ARX sollte man ARMAX,OE oder BJ benutzen.

## 6.6 Betrachtung des Simulationsfehlers

Unterschied zwischen Simulation und Prädiktion:

- *Simulation*: Berechnung der Modellantwort für eine gegebene Eingangsgrößensequenz, während die Ausgangsgrößen des Systems unbekannt sind.
- *Prädiktion*: Berechnung der Antwort des Modells  $l$  Schritte in die Zukunft für eine gegebene Stellgrößentrajektorie basierend auf den gemessenem Systemausgang bis zum Zeitpunkt  $k - 1$  ( $l = 1$ : 1-Schritt-Prädiktion).



Beispiel für ARX-Modell:

- Simulation:

$$\hat{y}(k) = -a_1 \hat{y}(k-1) - \dots - a_{n_A} \hat{y}(k-n_A) + b_0 u(k-d) + \dots + b_{n_B} u(k-d-n_B)$$

- Ein-Schritt-voraus-Prädiktion:

$$\begin{aligned}\hat{y}(k|k-1) &= -a_1 y(k-1) - \dots - a_{n_A} y(k-n_A) \\ &\quad + b_0 u(k-d) + \dots + b_{n_B} u(k-d-n_B)\end{aligned}$$

Simulationsfehler:

$$\text{MSE} = \frac{1}{K} \sum_{k=1}^K (y(k) - \underbrace{G(q^{-1}, \hat{\Theta})u(k-d)}_{\hat{y}(k)})^2$$

- Der Simulationsfehler wird sich nicht wesentlich verbessern, wenn das Modell die Realisierung des Rauschens (der Störungen) lernt.
- Anhand der Simulation kann man Stabilität und Verstärkung schnell beurteilen und das Modellverhalten besser einschätzen als mit der Prädiktion.

## 6.7 Statistische Analyse des Prädiktionsfehlers

- Analyse des Restprädiktionsfehlers (Residual analysis)
  - ◇ Bei einer konsistenten Modellschätzung, wandelt sich der Prädiktionsfehler  $\epsilon(k) = \hat{y}(k|k-1) - y(k)$  asymptotisch zu einem weißem Rauschsignal.
  - ◇ Eine konsistente Schätzung der Eingangstransferfunktion  $G(q, \hat{\Theta})$  lässt darauf schließen, dass der Prädiktionsfehler nicht mit alten Eingangsgrößen korreliert ist.

Basierend auf diesen Feststellungen kann man zwei Modellannahmen im statistischen Sinne formulieren (Nullhypothesen):

- (a)  $\epsilon(k, \hat{\Theta})$  ist eine Realisierung eines mittelwertfreien weißen Rauschprozesses;
- (b)  $\epsilon(k, \hat{\Theta})$  ist eine Realisierung eines stochastischen Prozesses, der  $\mathcal{E}\epsilon(k)u(s) = 0, t > s$  erfüllt.

In der Praxis arbeitet man mit normierten zeit-diskreten Korrelationsfunktionen, um zu überprüfen, ob die Modellannahmen erfüllt

sind:

$$\hat{R}_{\epsilon}^K(\tau) = \frac{\sum_{k=1}^{K-\tau} \epsilon(k+\tau)\epsilon(k)}{\sum_{k=1}^K \epsilon(k)^2} = \begin{cases} 1, \tau = 0 \\ 0, \tau \neq 0 \end{cases}$$

$$\hat{R}_{\epsilon u}^K(\tau) = \frac{\sum_{k=1}^{K-\tau} \epsilon(k+\tau)u(k)}{\left(\sum_{k=1}^K \epsilon^2(k) \sum_{k=1}^K u^2(k)\right)^{1/2}} = 0, \quad \forall \tau$$

◇ Gewöhnlich überprüft man diese Funktionen für Verzögerungen im Intervall  $\tau \in [-20, 20]$ . Null bedeutet, dass der Wert der Korrelationsfunktion innerhalb eines 95% Konfidenzintervall, i.e., if  $-1.96/\sqrt{K} < \hat{R}^K < 1.96/\sqrt{K}$ , liegt.

◇  $K$  sollte größer als 100 sein.

◇ Die Analyse sollte anhand der Testdaten erfolgen.

◇ Diese Analysemethode versagt, wenn der Rauschpegel sehr klein im Vergleich zum Ausgangssignalbereich ist (mehr als 60 dB Unterschied).

## 6.8 Schätzung der Totzeit

Idee 1:

- Identifikation eines Modells mit hoher Ordnung und Totzeitannahme  $d = 0$
- Die ersten  $d$  Koeffizienten von  $\mathbf{B}$  werden Null sein (annähernd) bei einer vorhandenen Totzeit  $d$

Idee 2: Schätzung der Impulsantwort mittels LS

$$y(k) = \sum_{i=1}^P b_i u(k-i), \quad P = 20 \dots 30$$

$$|\hat{b}_i| \ll |\hat{b}_{d+1}|, \quad i = 1, 2, \dots, d$$

Idee 3: Anzeige der Sprungantworten  $\Rightarrow$  Ablesen der Totzeit

Idee 4: Kreuzkorrelationsanalyse von  $y$  und  $u$

## 6.9 Wahl der anderen Polynomgrade

- Startwerte: a-priori
- Schrittweise Erhöhung der Ordnung und Anwendung der Modellvalidierungsmethoden (zeitaufwendig)

*Vorschlag mit Kreuzvalidierung:*

1. Start mit BJ-Modell (Grund auf nächster Folie) mit  $d = 0$ : große Ordnung für B, z.B. 10; Ordnung aller anderen Polynome z.B.  $n_{\max} = 3$   
 $\Rightarrow d$  aus geschätzten B bestimmen (Anzahl führender „Null“-Koeffizienten)
2.  $n_A = n_B$  schrittweise von 1 bis  $n_{\max}$  erhöhen;  $\deg(D) = \deg(C) = n_{\max}$   
 $\Rightarrow n_A$  mittels Kreuzvalidierung bestimmen
3. Ordnung von  $n_B$  schrittweise von 0 bis  $n_A$  erhöhen;  $\deg(D) = \deg(C) = n_{\max}$   
 $\Rightarrow n_B$  mittels Kreuzvalidierung bestimmen
4. Ordnung von  $n_C = n_D$  schrittweise von 0 bis  $n_{\max}$  erhöhen  
 $\Rightarrow n_C = n_D$  mittels Kreuzvalidierung bestimmen
5. Überprüfen, ob  $A \approx D \Rightarrow$  dann ARX oder ARMAX Modell ansetzen



## 6.10 Konsistenz für allgemeine Modellstrukturen

- Konsistenz: Biasfehler verschwindet für  $K \rightarrow \infty$ .
- Um eine konsistente Schätzung zu erzielen, muss  $\hat{G}$  das reale System  $G$  beschreiben können. Dies bedeutet, dass  $\hat{G}$  flexibel genug sein muss; anders ausgedrückt,  $\hat{G}$  muss eine genügend große Ordnung haben.
- *Bei allen Modellen, die unabhängig parametrisierte Transferfunktionen  $G(q)$  und  $H(q)$  haben, kann man die Parameter von  $G(q)$  konsistent schätzen, auch wenn  $H(q)$  nicht korrekt gewählt wurden ist. (nicht gültig bei einer Identifikation im geschlossenen Kreis)*

## 6.11 Datenvorverarbeitung

- Ziel: Bestimmung eines linearen Modells um einen Gleichgewichtspunkt
- Ein- und Ausgang des Modells sind Abweichungen vom Gleichgewichtspunkt  $(y^s, u^s)$
- Es sei angenommen, dass  $u = u^s + \Delta \mathbf{u}$  und  $y = y^s + \Delta \mathbf{y}$  die gemessenen Daten sind.
- Es gibt 3 Ansätze, wie mit den Offsets umgegangen werden kann:
  - (a) Entfernen der Offsets, Driften etc. durch explizite Vorverarbeitung der gemessenen Daten,
  - (b) Aufstellen einer Modellstruktur für  $u$  und  $y$  mit einem Offsetterm, welcher mit geschätzt werden kann,
  - (c) oder Erweiterung des Rausch-/Störmodells ( $H$ ).

Realisierungen dieser Ansätze:

1. (zu (a)) Falls der Gleichgewichtspunkt bekannt ist:

$$\Delta u(k) = u(k) - u^s, \quad \Delta y(k) = y(k) - y^s \quad (1)$$

Falls der Gleichgewichtspunkt unbekannt ist, kann er durch die Mittelwerte von  $u(k)$  und  $y(k)$  bestimmt werden.

2. (zu (a) und (c)) Offsets können auch durch Differenzierung der Daten eliminiert werden. Dies kann durch Vorfilterung von  $y$  und  $u$  mit

$$L(q) = 1 - q^{-1} \quad (2)$$

erfolgen oder durch Erweiterung der Transferfunktion  $H$  mit einem Integrator

$$H(q^{-1}) = \frac{1}{1 - q^{-1}} \tilde{H} \quad (3)$$

Beides erhöht die Genauigkeit des Modells für hohe Frequenzen.

3. (zu (c)) Die Grad des Nennerpolynoms der Transferfunktion  $H$  kann erweitert werden, so dass die PEM eine Polstelle bei  $q = 1$  finden kann.

4. (zu (b)) Berücksichtigung der Offsets bei der Parameterschätzung (direkte Schätzung des Offsetterme).

Beispiel ARX Modell. Siehe Mitschrift

Die Ansätze können leicht erweitert werden, um mit anderen Störungen (Driften, Schwingungen etc.) umgehen zu können.

Aus numerischen Gründen empfiehlt sich immer eine Normierung der Offset-/Drift-bereinigten Ein- und Ausgangsdaten auf Mittelwert 0 und Standardabweichung 1 (macht das Gütefunktional schöner - Vermeidung langer Täler).

---