

8. Parameterschätzung

8. Parameterschätzung

- Ziel: Schätzung der Parameter eines Signalmodells
- Am häufigsten angewandter Schätzer: **Least-Squares-Schätzer**
- Varianten bei besserer Kenntnis über die Statistik von Parametern und Fehlern:
 - **Gauß-Markov-Schätzer** (GM-Schätzer)
 - **Minimum-Varianz-Schätzer** (Sonderfall des GM-Schätzers)
- Allgemeinster Schätzer: **Bayes-Schätzer**
- Sonderfall davon: **Maximum-Likelihood-Schätzer**

8.1 Least-Squares-Schätzer

- Parameterschätzung für ein lineares Signalmodell
- Filterbank-Methode
- Anwendung zur Systemidentifikation

8.2 Gauß-Markov-Schätzer

8.3 Möglichkeiten zur Verbesserung der LS-Schätzung

8.4 Rekursiver Least-Squares-Schätzer (RLS)

- Ableitung aus dem Least-Squares-Schätzer
- Rekursiver LS-Schätzer für zeitvariante Systeme

8.5 AR-Schätzung

- Yule-Walker-Gleichung
- Prädiktionsfilter

8.6 Bayes-Schätzung

- Bayes'sche Parameterschätzung
- Maximum-Likelihood-Schätzer
- Cramér-Rao-Ungleichung

8.1 Least-Squares-Schätzer

8.1.1 Parameterschätzung f. ein lineares Signalmodell

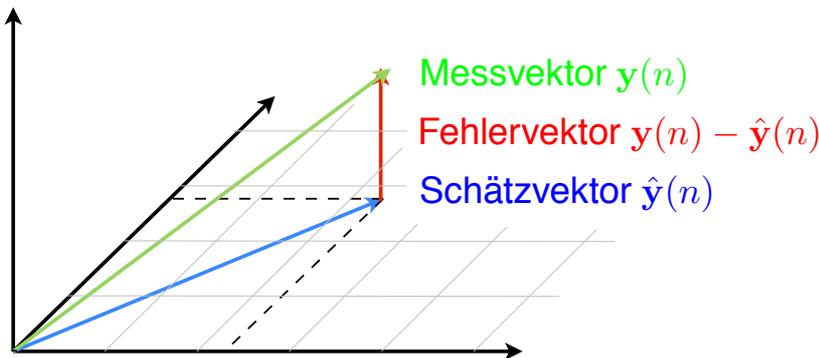
- **Wunsch:** „möglichst“ gute Approximation des Messvektors durch dessen Schätzung

→ **Minimierung des quadratischen Schätzfehlers** („Least Squares“):

$$\|\mathbf{y}(n) - \hat{\mathbf{y}}(n)\|^2 \rightarrow \min$$

- **Projektionstheorem:** optimaler Schätzvektor ist orthogonal zum Fehlervektor:

$$\langle \mathbf{y}(n) - \hat{\mathbf{y}}(n), \hat{\mathbf{y}}(n) \rangle = 0$$



5

Prof. Dr.-Ing. F. Puente León – Methoden der Signalverarbeitung

8.1.1 Parameterschätzung f. ein lineares Signalmodell

- Für den **Schätzvektor** wird als **Signalmodell** eine Linearkombination von Basisvektoren (Spaltenvektoren der Beobachtungsmatrix Φ) mit den zu schätzenden Parametern b_k als Gewichtungsfaktoren angesetzt [Abs. 7.2.2]:

$$\hat{\mathbf{y}}(n) = \sum_{k=0}^{K-1} \hat{b}_k(n) \varphi_k(n) = \Phi(n) \hat{\mathbf{b}}(n) \quad \text{mit} \quad \Phi^T(n) = \begin{bmatrix} \varphi_0^T(n) \\ \vdots \\ \varphi_{K-1}^T(n) \end{bmatrix}$$

- Wegen des Projektionstheorems ist der **Fehlervektor** auch orthogonal zu allen Basisvektoren:

$$\langle \mathbf{y}(n) - \hat{\mathbf{y}}(n), \varphi_k(n) \rangle = 0$$

$$\Leftrightarrow \varphi_k^T(n) \mathbf{y}(n) = \varphi_k^T(n) \hat{\mathbf{y}}(n)$$

$$\Leftrightarrow \Phi^T(n) \mathbf{y}(n) = \Phi^T(n) \Phi(n) \hat{\mathbf{b}}(n)$$

- **Least-Squares-Schätzer:**

$$\hat{\mathbf{b}}(n) = \underbrace{\left(\Phi^T(n) \Phi(n) \right)^{-1} \Phi^T(n)}_{\text{Pseudoinverse } \mathbf{G}(n)} \mathbf{y}(n) = \mathbf{G}(n) \mathbf{y}(n)$$

6

Prof. Dr.-Ing. F. Puente León – Methoden der Signalverarbeitung

- Der LS-Schätzer ist **erwartungstreu**, falls e mittelwertfrei ist:

$$E\{\hat{\mathbf{b}}(n)\} = E\{\underbrace{\mathbf{G}(n) \mathbf{y}(n)}_{\Phi(n) \mathbf{b} + e}\} = \mathbf{G}(n) \Phi(n) \mathbf{b} + \underbrace{\mathbf{G}(n) \cdot E\{e\}}_{=0}$$

$$\mathbf{G}(n) \Phi(n) = \left(\Phi^T(n) \Phi(n)\right)^{-1} \Phi^T(n) \Phi(n) = \mathbf{I} \quad (13)$$

Beispiel: LS-Schätzer für ein lineares Signalmodell

- Ansatz Signalmodell: $y(t) = b_0 + b_1 \cdot t$
- Störung sei additives, mittelwertfreies, weißes Rauschen $e(t)$

N Messwerte:

$$\underbrace{\begin{bmatrix} y(0T) \\ y(1T) \\ \vdots \\ y((N-1)T) \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} 1 & 0T \\ 1 & 1T \\ \vdots & \vdots \\ 1 & (N-1)T \end{bmatrix}}_{\Phi} \cdot \underbrace{\begin{bmatrix} b_0 \\ b_1 \end{bmatrix}}_{\mathbf{b}} + \underbrace{\begin{bmatrix} e(0T) \\ e(1T) \\ \vdots \\ e((N-1)T) \end{bmatrix}}_{\mathbf{e}}$$

Geschätzte Parameter: $\hat{\mathbf{b}} = (\Phi^T \Phi)^{-1} \Phi^T \cdot \mathbf{y}$

Beispiel: LS-Schätzer und Mittelwertschätzer

- Strahlungsintensität eines radioaktiven Materials ändert sich exponentiell:

$$y(t) = K_0 e^{-\alpha t}, \quad \alpha > 0$$

- Wunsch: Schätzung des Anfangswertes K_0
- Signalmodell für N Werte:

$$\begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(N-1) \end{bmatrix} = \underbrace{\begin{bmatrix} e^{-\alpha 0T} \\ e^{-\alpha 1T} \\ \vdots \\ e^{-\alpha (N-1)T} \end{bmatrix}}_{\Phi} \cdot K_0 + \begin{bmatrix} e(0) \\ e(1) \\ \vdots \\ e(N-1) \end{bmatrix}$$

- Für den **LS-Schätzwert** ergibt sich nach wenigen Schritten [KSW08]:

$$\hat{K}_0 = (\Phi^T \Phi)^{-1} \Phi^T \cdot \mathbf{y} = \dots = \frac{1 - e^{-2\alpha T}}{1 - e^{-2\alpha NT}} \sum_{n=0}^{N-1} e^{-\alpha nT} y(n)$$

- Der LS-Schätzer wird im Folgenden mit einem **Mittelwertschätzer** verglichen. Bei diesem wird K_0 aus jedem einzelnen Messwert geschätzt und anschließend der Mittelwert über die einzelnen Schätzwerte gebildet:

$$\bar{K}_0 = \frac{1}{N} \sum_{n=0}^{N-1} y(n) \cdot e^{\alpha n T}$$

$$y(n) = K_0 e^{-\alpha n T} + e(n)$$

- Beide Schätzer sind **erwartungstreu**, da der Erwartungswert ihrer Schätzfehler verschwindet:

$$\text{MW: } E\{\bar{K}_0 - K_0\} = E\left\{\frac{1}{N} \sum_{n=0}^{N-1} e^{\alpha n T} e(n)\right\} = 0$$

$$\text{LS: } E\{\hat{K}_0 - K_0\} = E\left\{\frac{1 - e^{-2\alpha T}}{1 - e^{-2\alpha NT}} \sum_{n=0}^{N-1} e^{-\alpha n T} e(n)\right\} = 0$$

- Aufgrund der Erwartungstreue kann die Varianz des Schätzfehlers als Erwartungswert des quadrierten Fehlers berechnet werden

- Konsistenzprüfung Mittelwertschätzer:

$$\sigma_{\text{MW}}^2(N) = E\left\{(\bar{K}_0 - K_0)^2\right\} = \dots = \frac{1}{N^2} \sum_{n=0}^{N-1} e^{2\alpha n T} \sigma_e^2 \xrightarrow[N \rightarrow \infty]{} \infty$$

- Konsistenzprüfung LS-Schätzer:

$$\sigma_{\text{LS}}^2(N) = E\left\{(\hat{K}_0 - K_0)^2\right\} = \dots = \frac{1 - e^{-2\alpha T}}{1 - e^{-2\alpha NT}} \sigma_e^2 \xrightarrow[N \rightarrow \infty]{} (1 - e^{-2\alpha T}) \sigma_e^2$$

- Beide Schätzer sind **nicht konsistent**

- Für den **Mittelwertschätzer** existiert ein optimales N mit minimaler Schätzfehlervarianz; bei mehr Messwerten verschlechtert sich das Ergebnis immer weiter

Grund: Betrag des Messsignals nimmt für wachsende n monoton ab, Rauschen bleibt jedoch konstant, d. h. SNR wird immer schlechter

- LS-Schätzer:** Varianz des Schätzwertes nimmt bei Hinzunahme von Messwerten monoton ab und konvergiert gegen konstanten Wert

■ Interpretation der LS-Schätzung

$$\hat{\mathbf{b}}(n) = \mathbf{G}(n) \mathbf{y}(n), \quad \mathbf{G}(n) = (\Phi^T(n) \Phi(n))^{-1} \Phi^T(n)$$

als **Filterbank**, bei der für jeden Parameter b_k die Impulsantwort \mathbf{g}_k eines zeitdiskreten Filters mit dem Messvektor gefaltet wird:

$$\hat{b}_k(n) = \mathbf{g}_k^T \mathbf{y}(n) = \sum_{m=0}^{N-1} g_k(m) y(n-m) \quad (16)$$

- In manchen Anwendungen zwar Modell der Ordnung K für eine präzise Signalbeschreibung nötig, aber nur wenige Parameter von Interesse
- Dann kann man die Schätzung auf wenige Filter \mathbf{g}_k beschränken, ohne auf die Genauigkeit des Modells zu verzichten
- Ansatz für inverse Matrix für **einzelnen** Schätzvorgang mit N Messwerten:

$$(\Phi^T \Phi)^{-1} = \begin{bmatrix} \alpha_{0,0} & \cdots & \alpha_{0,K-1} \\ \vdots & \ddots & \vdots \\ \alpha_{K-1,0} & \cdots & \alpha_{K-1,K-1} \end{bmatrix}$$

8.1.2 Filterbank-Methode

■ Damit lautet der LS-Schätzer:

$$\begin{bmatrix} \hat{b}_0 \\ \vdots \\ \hat{b}_{K-1} \end{bmatrix} = \begin{bmatrix} \alpha_{0,0} & \cdots & \alpha_{0,K-1} \\ \vdots & \ddots & \vdots \\ \alpha_{K-1,0} & \cdots & \alpha_{K-1,K-1} \end{bmatrix} \cdot \begin{bmatrix} \varphi_0^T \\ \vdots \\ \varphi_{K-1}^T \end{bmatrix} \cdot \mathbf{y} = \begin{bmatrix} \mathbf{g}_0^T \\ \vdots \\ \mathbf{g}_{K-1}^T \end{bmatrix} \cdot \mathbf{y}$$

■ Einzelner Parameter:
$$\hat{b}_k = \underbrace{\sum_{j=0}^{K-1} \alpha_{kj} \varphi_j^T \mathbf{y}}_{=\mathbf{g}_k^T} \stackrel{!}{=} \mathbf{g}_k^T \cdot \mathbf{y}$$

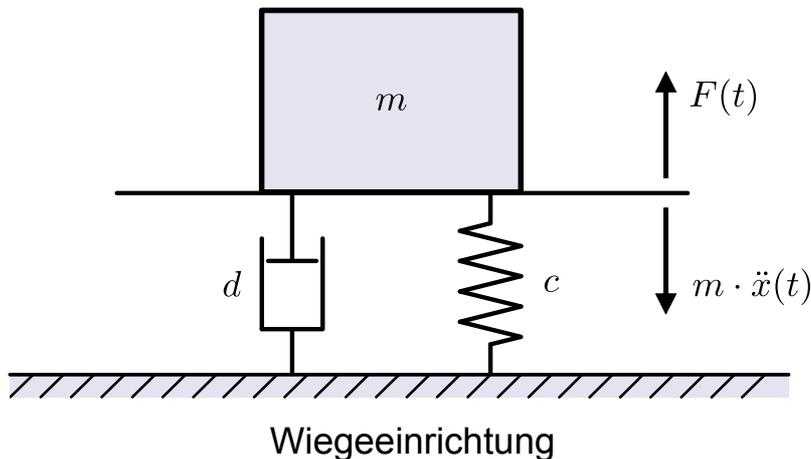
- Aus der Bedingung (13) für Erwartungstreue $\mathbf{G} \Phi = \mathbf{I}$ folgen K^2 Bestimmungsgleichungen für die α_{kj} :

$$\mathbf{g}_k^T \varphi_i = \delta_{ki}$$

$$\Rightarrow \sum_{j=0}^{K-1} \alpha_{kj} \varphi_j^T \varphi_i = \sum_{j=0}^{K-1} \alpha_{kj} \langle \varphi_j, \varphi_i \rangle = \delta_{ki} \quad (24)$$

Beispiel: Schnelle Wiegeeinrichtung

- Wunsch: Bestimmung der Masse m an einer Federwaage vor dem Abklingen der Schwingung, um mehr Messungen pro Zeit zu ermöglichen
- Strategie: Messung der Beschleunigung $\ddot{x}(t)$ (Ausgangsgröße)
- Federkonstante c und Dämpferkonstante d seien bekannt



8.1.2 Filterbank-Methode

- Aus der Kräftebilanz folgt die DGL des Masse-Feder-Dämpfer-Systems:

$$m \cdot \ddot{x}(t) + d \cdot \dot{x}(t) + c \cdot x(t) + F(t) = 0$$

- Mittels Laplace-Transformation erhält man die Übertragungsfunktion:

$$\ddot{X}(s) = -\frac{s^2}{m \cdot s^2 + d \cdot s + c} \cdot F(s) \quad \text{mit} \quad \ddot{x}(t) \circ \bullet \ddot{X}(s)$$

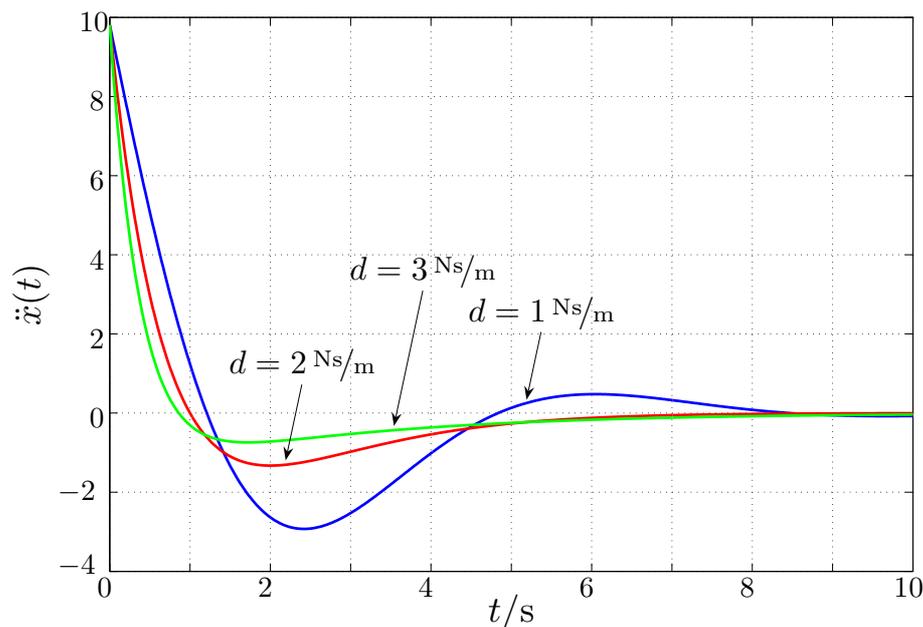
- Masse wird erst ab $t = 0$ auf die Waage gelegt \rightarrow Gewichtskraft wirkt erst ab $t = 0$:

$$F(t) = -m g \cdot \sigma(t) \quad \circ \bullet \quad F(s) = -m g \cdot \frac{1}{s}$$

- Durch inverse Laplace-Transformation erhält man mit den Abkürzungen $\alpha = d/2m$ und $\beta^2 = c/m - d^2/4m^2$ die Sprungantwort:

$$\ddot{x}(t) = \mathcal{L}^{-1}\{\ddot{X}(s)\} = \dots = g \exp(-\alpha t) \left[\cos(\beta t) - \frac{\alpha}{\beta} \sin(\beta t) \right], \quad t \geq 0$$

Verläufe der gemessenen Beschleunigung $\ddot{x}(t)$ (Sprungantworten) bei der Schätzung der Masse $m = 1 \text{ kg}$ ($c = 1 \text{ N/m}$)



$$\ddot{x}(t) = g \exp(-\alpha t) \cdot \left[\cos(\beta t) - \frac{\alpha}{\beta} \sin(\beta t) \right], \quad t \geq 0$$

- Signalmodell muss linear sein \rightarrow Linearisierung für kleine t :

$$\exp(-\alpha t) \approx 1 - \alpha t + \frac{\alpha^2}{2} t^2 - \frac{\alpha^3}{6} t^3$$

$$\cos(\beta t) \approx 1 - \frac{\beta^2}{2} t^2$$

$$\frac{\alpha}{\beta} \cdot \sin(\beta t) \approx \frac{\alpha}{\beta} \cdot \left(\beta t - \frac{1}{6} \beta^3 t^3 \right) = \alpha t - \frac{\alpha \beta^2}{6} t^3$$

- Daraus resultiert für den Einschwingvorgang das **lineare** Signalmodell:

$$\ddot{x}(t) = g \cdot \left[1 - 2\alpha t + \frac{1}{2}(3\alpha^2 - \beta^2)t^2 + \frac{2}{3}(\alpha\beta^2 - \alpha^3)t^3 \right]$$

$$= g \cdot [b_0 + b_1 \cdot t + b_2 \cdot t^2 + b_3 \cdot t^3]$$

$$= g \cdot [b_0 \cdot \varphi_0(t) + b_1 \cdot \varphi_1(t) + b_2 \cdot \varphi_2(t) + b_3 \cdot \varphi_3(t)] \quad , \quad t \geq 0$$

- Zur Bestimmung der Masse m reicht die Schätzung von b_1 aus (b.w.)

$$b_1 = -2\alpha = -2 \cdot \frac{d}{2m} \Rightarrow m = -\frac{d}{b_1}$$

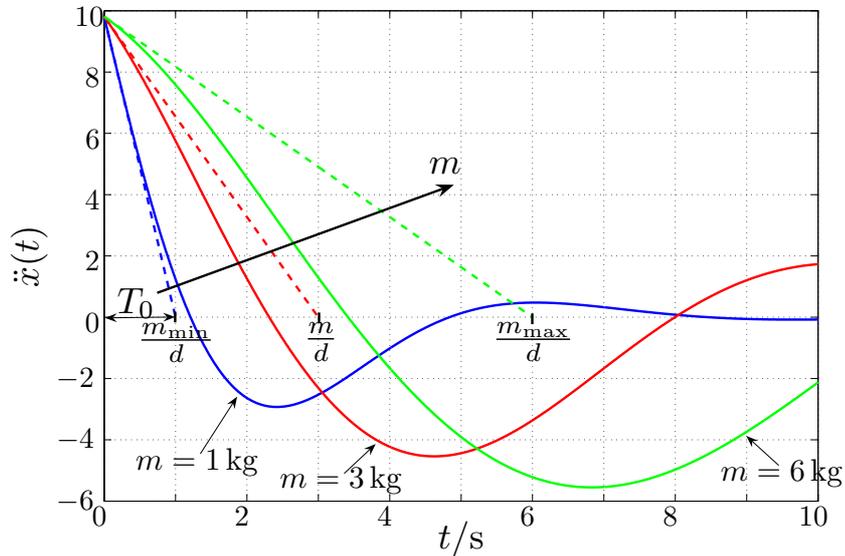
Dennoch hat das Signalmodell die Ordnung $K = 4$, um die Beschleunigung während der Anfangsphase ausreichend genau zu beschreiben

- Wird an die Sprungantwort eine Tangente in $t = 0$ angelegt, so schneidet diese die Zeitachse an der Stelle $t = m/d$

- Bis zu dieser Zeit sind die Approximationen der Exponentialfunktion und der trigonometrischen Funktionen ausreichend genau

➔ Beobachtungsintervall:

$$T_0 = \frac{m_{\min}}{d}$$



- Ansatz zur Bestimmung der Gewichtungsfaktoren α_{kj} ((24) in Matrixform)

$$\begin{bmatrix} \langle \varphi_0(t), \varphi_0(t) \rangle & \langle \varphi_1(t), \varphi_0(t) \rangle & \langle \varphi_2(t), \varphi_0(t) \rangle & \langle \varphi_3(t), \varphi_0(t) \rangle \\ \langle \varphi_0(t), \varphi_1(t) \rangle & \langle \varphi_1(t), \varphi_1(t) \rangle & \langle \varphi_2(t), \varphi_1(t) \rangle & \langle \varphi_3(t), \varphi_1(t) \rangle \\ \langle \varphi_0(t), \varphi_2(t) \rangle & \langle \varphi_1(t), \varphi_2(t) \rangle & \langle \varphi_2(t), \varphi_2(t) \rangle & \langle \varphi_3(t), \varphi_2(t) \rangle \\ \langle \varphi_0(t), \varphi_3(t) \rangle & \langle \varphi_1(t), \varphi_3(t) \rangle & \langle \varphi_2(t), \varphi_3(t) \rangle & \langle \varphi_3(t), \varphi_3(t) \rangle \end{bmatrix} \cdot \begin{bmatrix} \alpha_{k0} \\ \alpha_{k1} \\ \alpha_{k2} \\ \alpha_{k3} \end{bmatrix} = \begin{bmatrix} \delta_{k0} \\ \delta_{k1} \\ \delta_{k2} \\ \delta_{k3} \end{bmatrix}$$

- Die Innenprodukte sind dabei: $\langle \varphi_i(t), \varphi_j(t) \rangle = \int_0^{T_0} \varphi_i(t) \cdot \varphi_j(t) dt$

- Für $k = 1$ (Schätzung **nur** des Parameters b_1) ergibt sich folgender Ansatz für die Gewichtungsfaktoren:

$$\begin{bmatrix} T_0 & \frac{T_0^2}{2} & \frac{T_0^3}{3} & \frac{T_0^4}{4} \\ \frac{T_0^2}{2} & \frac{T_0^3}{3} & \frac{T_0^4}{4} & \frac{T_0^5}{5} \\ \frac{T_0^3}{3} & \frac{T_0^4}{4} & \frac{T_0^5}{5} & \frac{T_0^6}{6} \\ \frac{T_0^4}{4} & \frac{T_0^5}{5} & \frac{T_0^6}{6} & \frac{T_0^7}{7} \end{bmatrix} \cdot \begin{bmatrix} \alpha_{10} \\ \alpha_{11} \\ \alpha_{12} \\ \alpha_{13} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

- Daraus folgen die Gewichtungsfaktoren α_{1j} des Filters
- Damit erhält man für die Impulsantwort des Schätzfilters (vgl. Folie 12):

$$g_1(t) = \alpha_{10} \cdot \varphi_0(t) + \alpha_{11} \cdot \varphi_1(t) + \alpha_{12} \cdot \varphi_2(t) + \alpha_{13} \cdot \varphi_3(t)$$

$$= -\frac{120}{T_0^2} + \frac{1200}{T_0^3} \cdot t - \frac{2700}{T_0^4} \cdot t^2 + \frac{1680}{T_0^5} \cdot t^3, \quad 0 \leq t \leq T_0$$

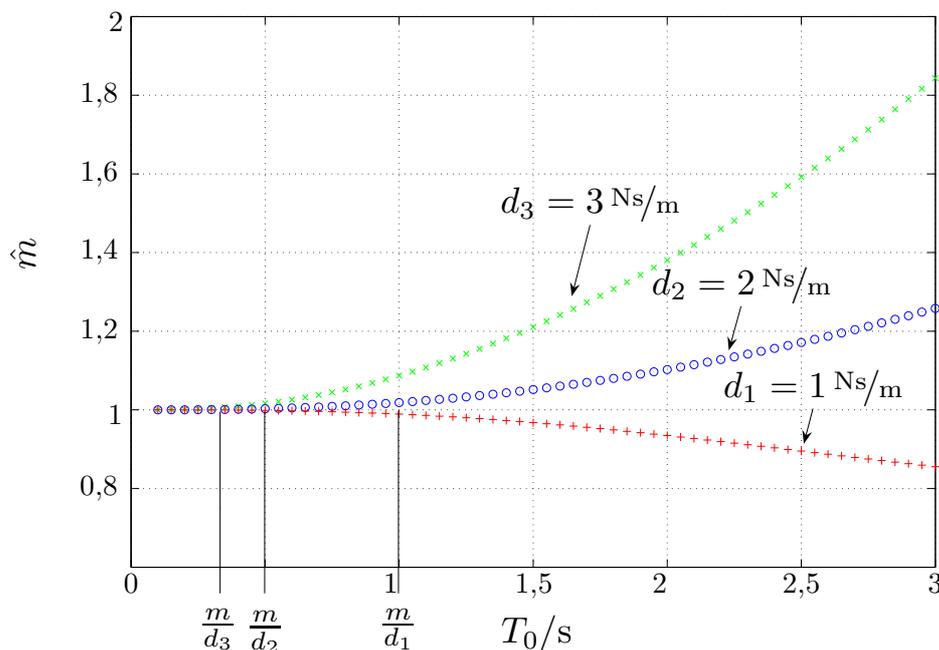
- Die Schätzung des Parameters b_1 erfolgt mit der Impulsantwort $g_1(t)$ nach (16) [Folie 11]:

$$\hat{b}_1 = \langle g_1(t), \ddot{x}(t) \rangle = \int_0^{T_0} g_1(t) \cdot \ddot{x}(t) dt$$

- Schließlich resultiert die gesuchte Masse aus Folie 17 (oben):

$$b_1 = -2\alpha = -2 \cdot \frac{d}{2m} \Rightarrow \hat{m} = -\frac{d}{\hat{b}_1}$$

- Ergebnisse der Schätzung einer Masse $m = 1$ kg für verschieden große Beobachtungsintervalle T_0 und für 3 verschiedene Dämpferkonstanten d .



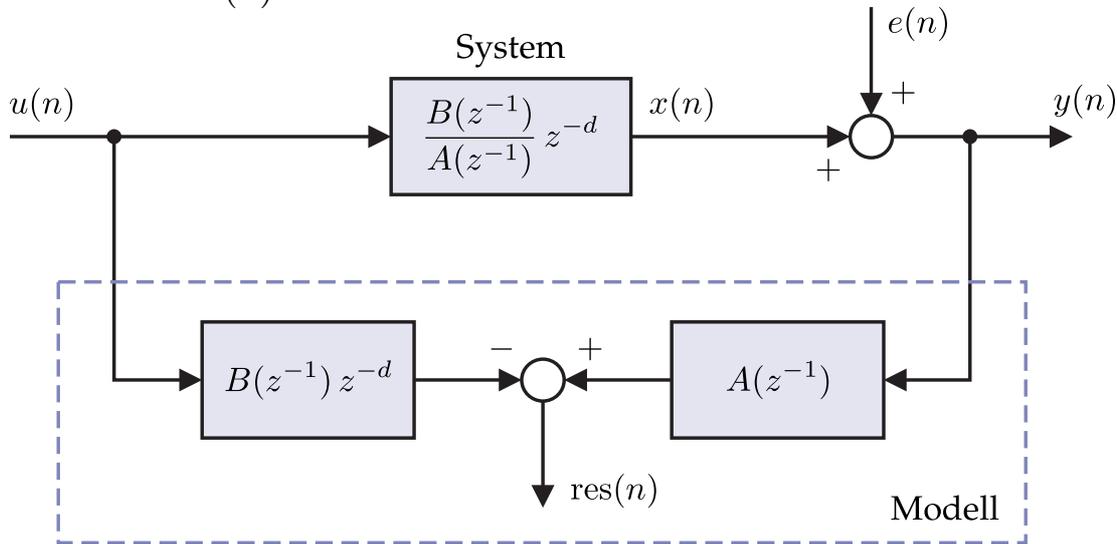
- Ergebnisse für $T_0 > m/d$ ungenau (wg. Approximationsfehlers)

8.1.3 Identifikation von Systemfunktionen

- LS-Schätzung zur Parameteridentifikation in LTI-Systemen (K, d bekannt)

- Systemmodell:
$$G(z) = \frac{B(z^{-1})}{A(z^{-1})} z^{-d} = \frac{b_0 + b_1 z^{-1} + \dots + b_K z^{-K}}{1 + a_1 z^{-1} + \dots + a_K z^{-K}} z^{-d}$$

- Ziel: Schätzung der Modellparameter b_i und a_i durch Minimierung des Residuums $\text{res}(n)$



8.1.3 Identifikation von Systemfunktionen

Beispiel: Identifikation eines Systems 2. Ordnung

- Modellansatz:
$$G(z) = \frac{X(z)}{U(z)} = \frac{b_0 + b_1 z^{-1}}{1 + a_1 z^{-1} + a_2 z^{-2}}$$

- Systemgleichung:
$$Y(z) = G(z) \cdot U(z) + E(z)$$

- Differenzengleichung:
$$y(n) = -a_1 y(n-1) - a_2 y(n-2) + b_0 u(n) + b_1 u(n-1) + \text{res}(n)$$

- Für $N + 2$ Messwerte ergibt sich:

$$\begin{bmatrix} y(n) \\ \vdots \\ y(n-N+1) \end{bmatrix} = \overbrace{\begin{bmatrix} -y(n-1) & -y(n-2) & u(n) & u(n-1) \\ \vdots & \vdots & \vdots & \vdots \\ -y(n-N) & -y(n-N-1) & u(n-N+1) & u(n-N) \end{bmatrix}}^{\Phi(n)} \overbrace{\begin{bmatrix} a_1 \\ a_2 \\ b_0 \\ b_1 \end{bmatrix}}^{\mathbf{b}} + \begin{bmatrix} \text{res}(n) \\ \vdots \\ \text{res}(n-N+1) \end{bmatrix}$$

- Wichtig: breitbandige Anregung $u(n)$, da sonst das identifizierte Modell nicht für alle Frequenzen gültig ist
- Verwendung von Pseudo-Rauschsignalen hoher Leistung als Eingangssignal, damit Beobachtungsmatrix gut konditioniert:

$$\hat{\mathbf{b}}(n) = (\Phi^T(n) \Phi(n))^{-1} \Phi^T(n) \mathbf{y}(n)$$

- Vorgegebene Parameter des zu identifizierenden Systems:

$$a_1 = \frac{1}{2}\sqrt{2} \approx 0,707 \quad a_2 = 0,25 \quad b_0 = 0,5 \quad b_1 = 1$$

- Geschätzte Parameter für $N = 10$:

$$\hat{a}_1 = 0,684 \quad \hat{a}_2 = 0,236 \quad \hat{b}_0 = 0,602 \quad \hat{b}_1 = 1,057$$

- Für $N = 1000$ ergeben sich nahezu dieselben Schätzwerte (Bias!)
- Bei Erhöhung der Rauschleistung von $e(n)$ vergrößert sich der Bias
- In der Praxis: Modellordnung aus theoretischen (physikalischen) Betrachtungen oder zunächst möglichst kleine Ordnung ansetzen und dann schrittweise erhöhen, bis gewünschte Genauigkeit erzielt wird ■

8.2 Gauß-Markov-Schätzer

■ Annahmen beim **LS-Schätzer**

- Modellfehler $e(n)$ sei mittelwertfreies weißes (d. h. unkorreliertes) Rauschen mit konstanter Varianz
- Statistik des Parametervektors $\mathbf{b}(n)$ unbekannt

■ Neue Annahmen beim **Gauß-Markov-Schätzer** (*Minimum Mean Squared Error, MMSE*)

- Modellfehler $\mathbf{e}(n)$ sei mittelwertfreies farbiges Rauschen
- Parametervektor und Fehlervektor seien statistisch unabhängige Zufallsvektoren:

$$E \{ \mathbf{b} \cdot \mathbf{e}^T \} = E \{ \mathbf{b} \} \cdot E \{ \mathbf{e}^T \}$$

$$E \{ \mathbf{e} \cdot \mathbf{b}^T \} = E \{ \mathbf{e} \} \cdot E \{ \mathbf{b}^T \}$$

- Die Autokorrelationsmatrix \mathbf{R}_{bb} des Parametervektors \mathbf{b} und die Autokovarianzmatrix \mathbf{C}_{ee} des Fehlervektors $\mathbf{e}(n)$ seien bekannt

- Die einzelnen Parameter b_k des Signalmodells werden bei aufeinander folgenden Schätzungen als Zufallsgrößen interpretiert (Bayes-Theorie)
- Die Approximation einer Funktion $y(n)$ im Funktionenraum (Folie 5) wird daher durch eine Approximation der stochastischen Variablen b_k im Vektorraum stochastischer Variabler ersetzt (MMSE):

$$\left\| b_k - \hat{b}_k(n) \right\|^2 \rightarrow \min, \quad E \left\{ (\mathbf{b} - \hat{\mathbf{b}}(n)) \cdot (\mathbf{b} - \hat{\mathbf{b}}(n))^T \right\} \rightarrow \min$$

- **Projektionstheorem:** Bei optimaler Schätzung ist **Schätzvektor** orthogonal zum **Fehlervektor**:

$$\left\langle b_k - \hat{b}_k(n), \hat{b}_k(n) \right\rangle = 0 \quad \forall k$$

- Für die **Schätzung** wird die zeitdiskrete Faltung (16) des Messvektors $y(n)$ mit der Filterbank $g_k(n)$ des Schätzers eingesetzt:

$$\left\langle b_k - \hat{b}_k(n), \sum_{m=0}^{N-1} g_k(m) y(n-m) \right\rangle = \sum_{m=0}^{N-1} g_k^*(m) \langle b_k - \hat{b}_k(n), y(n-m) \rangle = 0$$

- Das Innenprodukt ist bei Zufallsgrößen als Erwartungswert definiert:

$$\langle b_k - \hat{b}_k(n), y(n-m) \rangle = E\{(b_k - \hat{b}_k(n)) \cdot y(n-m)\} = 0$$

- Für $k = 0, \dots, K-1$ und $m = 0, \dots, N-1$ ergibt sich in Matrixschreibweise:

$$E\{\mathbf{b} \cdot \mathbf{y}^T(n)\} = E\{\hat{\mathbf{b}}(n) \cdot \mathbf{y}^T(n)\}$$

- Mit dem allgemeinen Ansatz $\hat{\mathbf{b}}(n) = \mathbf{G}(n) \cdot \mathbf{y}(n)$ folgt:

$$E\{\mathbf{b} \cdot \mathbf{y}^T(n)\} = \mathbf{G}(n) \cdot E\{\mathbf{y}(n) \cdot \mathbf{y}^T(n)\} = \mathbf{G}(n) \cdot \mathbf{R}_{yy} \quad (88)$$

- In die Autokorrelationsmatrix \mathbf{R}_{yy} wird $\mathbf{y}(n) = \Phi(n) \cdot \mathbf{b} + \mathbf{e}(n)$ eingesetzt:

$$\begin{aligned} \mathbf{R}_{yy} &= E\{\mathbf{y}(n) \cdot \mathbf{y}^T(n)\} = E\{\Phi(n) \cdot \mathbf{b} \cdot \mathbf{b}^T \Phi^T(n) + \underbrace{\Phi(n) \cdot \mathbf{b} \cdot \mathbf{e}^T(n)}_{=0} \\ &\quad + \underbrace{\mathbf{e}(n) \cdot \mathbf{b}^T \cdot \Phi^T(n)}_{=0} + \mathbf{e}(n) \cdot \mathbf{e}^T(n)\} \\ &= \Phi(n) \cdot \mathbf{R}_{bb} \cdot \Phi^T(n) + \mathbf{C}_{ee} \end{aligned}$$

- Entsprechend gilt mit $\mathbf{y}(n) = \Phi(n) \cdot \mathbf{b} + \mathbf{e}(n)$:

$$E\{\mathbf{b} \cdot \mathbf{y}^T(n)\} = E\{\mathbf{b} \cdot \mathbf{b}^T \cdot \Phi^T(n) + \underbrace{\mathbf{b} \cdot \mathbf{e}^T(n)}_{=0}\} = \mathbf{R}_{bb} \cdot \Phi^T(n)$$

- Damit folgt aus (88):

$$\begin{aligned} \mathbf{R}_{bb} \cdot \Phi^T(n) &= \mathbf{G}(n) \cdot \left(\Phi(n) \cdot \mathbf{R}_{bb} \cdot \Phi^T(n) + \mathbf{C}_{ee} \right) \\ \Rightarrow \mathbf{G}_{GM}(n) &= \mathbf{R}_{bb} \cdot \Phi^T(n) \cdot \underbrace{\left(\Phi(n) \cdot \mathbf{R}_{bb} \cdot \Phi^T(n) + \mathbf{C}_{ee} \right)^{-1}}_{N \times N\text{-Matrix, aufwendig zu invertieren}} \end{aligned}$$

- Weitere Umformungen [KSW08] ergeben den **Gauß-Markov-Schätzer** in einer Form, bei der nur noch eine $K \times K$ -Matrix zu invertieren ist:

$$\mathbf{G}_{GM}(n) = \underbrace{\left(\Phi^T(n) \cdot \mathbf{C}_{ee}^{-1} \cdot \Phi(n) + \mathbf{R}_{bb}^{-1} \right)^{-1}}_{\text{Schätzfehlerkovarianz } \mathbf{C}_{\tilde{b}\tilde{b}}(n)} \cdot \Phi^T(n) \cdot \mathbf{C}_{ee}^{-1}$$

- Die Autokorrelationsmatrix \mathbf{R}_{bb} des Parametervektors \mathbf{b} und die Autokovarianzmatrix \mathbf{C}_{ee} des Fehlers $\mathbf{e}(n)$ müssen bekannt sein
- Der Gauß-Markov-Schätzer ist **nicht erwartungstreu** (vgl. Folie 7):

$$\mathbf{G}_{\text{GM}}(n) \cdot \Phi(n) \neq \mathbf{I}$$

Minimum-Varianz-Schätzer

- **Sonderfall** bei unbekannter Statistik des Parametervektors \mathbf{b}
- Annahme einer Gleichverteilung über unendlich großem Intervall
→ Varianzen gehen gegen ∞ und daher gilt $\mathbf{R}_{bb}^{-1} = \mathbf{0}$:

$$\mathbf{G}_{\text{MV}}(n) = \underbrace{\left(\Phi^T(n) \cdot \mathbf{C}_{ee}^{-1} \cdot \Phi(n) \right)^{-1}}_{\text{Schätzfehlerkovarianz } \mathbf{C}_{\tilde{b}\tilde{b}}(n)} \cdot \Phi^T(n) \cdot \mathbf{C}_{ee}^{-1}$$

- MV-Schätzer ist **erwartungstreu**
- Falls Kovarianzmatrix \mathbf{C}_{ee} unbekannt, Annahme weißes Rauschens
⇒ LS-Schätzer

8.3 Verbesserung der LS-Schätzung

- **Frage:** Lässt sich mittels einer **linearen Transformation** (z. B. KLT) des Messvektors $\mathbf{y}(n)$ eine bessere LS-Schätzung der Koeffizienten erzielen?

- Lineare Transformation: $\mathbf{z}(n) = \mathbf{A} \cdot \mathbf{y}(n)$

- Einsetzen des Signalmodells: $\mathbf{z}(n) = \mathbf{A} \cdot \mathbf{y}(n) = \underbrace{\mathbf{A} \cdot \Phi(n)}_{\text{Beobachtungsmatrix}} \cdot \mathbf{b} + \underbrace{\mathbf{A} \cdot \mathbf{e}(n)}_{\text{Messfehler}}$

- Kovarianz des transformierten Messfehlers:

$$\mathbf{C}_{ee,\mathbf{A}} = E \left\{ (\mathbf{A} \cdot \mathbf{e}(n)) (\mathbf{A} \cdot \mathbf{e}(n))^T \right\} = \mathbf{A} \cdot \mathbf{C}_{ee} \cdot \mathbf{A}^T$$

- Schätzfehlerkovarianz des **MV-Schätzers**:

$$\begin{aligned} \mathbf{C}_{\tilde{b}\tilde{b},\mathbf{A}}(n) &= \left((\mathbf{A} \cdot \Phi(n))^T \cdot \mathbf{C}_{ee,\mathbf{A}}^{-1} \cdot (\mathbf{A} \cdot \Phi(n)) \right)^{-1} \\ &= \left(\Phi^T(n) \cdot \mathbf{A}^T \cdot (\mathbf{A}^T)^{-1} \cdot \mathbf{C}_{ee}^{-1} \cdot \mathbf{A}^{-1} \cdot \mathbf{A} \cdot \Phi(n) \right)^{-1} = \mathbf{C}_{\tilde{b}\tilde{b}}(n) \end{aligned}$$

- **Fazit:** Lineare Transformation vor der Schätzung kann Schätzergebnis nicht verbessern

8.3.2 Erhöhung des Stichprobenumfangs

- **Frage:** Führt eine Erhöhung des Stichprobenumfangs N zu einer Verbesserung der LS-Schätzung?

- **Kovarianzmatrix des Schätzfehlers** [Fol. 29 bei Unkorreliertheit]:

$$\mathbf{C}_{\tilde{b}\tilde{b}}(n) = \sigma_e^2 \left(\Phi^T(n) \Phi(n) \right)^{-1}$$

mit:

$$\begin{aligned} &\Phi^T(n) \Phi(n) \\ &= N \cdot \begin{bmatrix} \frac{1}{N} \sum_{m=0}^{N-1} \varphi_0(n-m)\varphi_0(n-m) & \dots & \frac{1}{N} \sum_{m=0}^{N-1} \varphi_0(n-m)\varphi_{K-1}(n-m) \\ \vdots & \ddots & \vdots \\ \frac{1}{N} \sum_{m=0}^{N-1} \varphi_{K-1}(n-m)\varphi_0(n-m) & \dots & \frac{1}{N} \sum_{m=0}^{N-1} \varphi_{K-1}(n-m)\varphi_{K-1}(n-m) \end{bmatrix} \\ &= N \cdot \mathbf{S} \end{aligned}$$

S : Streumatrix der Regressionsvektoren

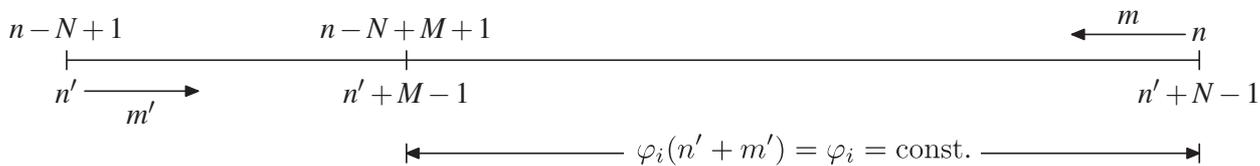
$$\Rightarrow \mathbf{C}_{\tilde{b}\tilde{b}}(n) = \frac{\sigma_e^2}{N} \cdot \mathbf{S}^{-1}$$

$$C_{\tilde{b}\tilde{b}}(n) = \frac{\sigma_e^2}{N} \cdot S^{-1}$$

- Bei **unabhängigen Messwerten** sind die $\varphi_k(n-m)$ unterschiedlich, sie streuen mehr. Die Kovarianz des Schätzfehlers wird mit wachsendem N kleiner. Der LS-Schätzer ist dann **konsistent**:

$$\lim_{N \rightarrow \infty} C_{\tilde{b}\tilde{b}}(n) = 0$$

- Bei **abhängigen Messwerten** (wenn diese ab einem $n' \geq M$ konstant sind, z. B. bei Abklingvorgängen) beinhalten diese keine weitere Information. Der LS-Schätzer ist dann nicht mehr konsistent; die Kovarianz des Schätzfehlers geht nicht gegen null.



8.3.3 Vorabmittelung der Messwerte

- Frage:** Führt eine Vorabmittelung der Messwerte zu einer Verbesserung der LS-Schätzung?
- Voraussetzung:** Unabhängige Messwerte
- Bei der Messwernerfassung werden N Blöcke mit jeweils L Messwerten aufgenommen:

$$y_L(n) = \left[\underbrace{y_0(n) \dots y_{L-1}(n)}_{y(n) = \frac{1}{L} \sum_l y_l(n)} \dots \underbrace{y_0(n-N+1) \dots y_{L-1}(n-N+1)}_{y(n-N+1) = \frac{1}{L} \sum_l y_l(n-N+1)} \right]^T$$

- Mittelung der L Werte in einem Block durch Multiplikation mit der $N \times LN$ -Matrix $\mathbf{B} \rightarrow$ Reduktion der Größe des Messvektors von $L \cdot N$ auf N :

$$y(n) = \mathbf{B} \cdot y_L(n) \quad \text{mit} \quad \mathbf{B} = \frac{1}{L} \cdot \begin{bmatrix} 1 \dots 1 & & 0 \\ & \ddots & \\ 0 & & \underbrace{1 \dots 1}_L \end{bmatrix}$$

- Einsetzen des linearen Signalmodells (Kap. 7)

$$y_L(n) = \Phi_L(n) \mathbf{b} + \mathbf{e}_L(n) \quad [\dots]$$

ergibt:

$$\mathbf{y}(n) = \mathbf{B} \cdot \mathbf{y}_L(n) = \mathbf{B} \cdot \Phi_L(n) \cdot \mathbf{b} + \underbrace{\mathbf{B} \cdot \mathbf{e}_L(n)}_{\text{Messfehler}}$$

■ Wegen

$$\mathbf{B} \cdot \mathbf{B}^T = \frac{1}{L} \cdot \mathbf{I}$$

erhält man eine um den Faktor L niedrigere Messfehlerkovarianzmatrix:

$$\begin{aligned} \mathbf{C}_{ee}(n) &= \mathbb{E} \{ \mathbf{B} \mathbf{e}_L(n) \mathbf{e}_L^T(n) \mathbf{B}^T \} = \mathbf{B} \cdot \mathbf{C}_{ee,L}(n) \cdot \mathbf{B}^T \\ &= \sigma_e^2 \cdot \mathbf{B} \mathbf{B}^T = \frac{\sigma_e^2}{L} \cdot \mathbf{I} \end{aligned}$$

■ Die Schätzfehlerkovarianz verringert sich um den gleichen Faktor L :

$$\mathbf{C}_{\hat{\mathbf{b}}\hat{\mathbf{b}}}(n) = \frac{\sigma_e^2}{L} \left(\Phi^T(n) \Phi(n) \right)^{-1}$$

■ **Fazit:** Bei starken Störungen wird man zur Verbesserung der Schätzung die Ordnung N des LS-Schätzers nicht zu groß wählen. Statt dessen verringert man den Einfluss der Störung durch eine Vorabmittelung.

8.4 Rekursiver Least-Squares-Schätzer (RLS)

- **Motivation:** Anstelle der Schätzung in einem Schritt soll diese **rekursiv** durch Hinzunahme neuer Messwerte verbessert werden:

$$\text{Diskrete Zeit:} \quad n - 1 \rightarrow n$$

$$\text{Stichprobenumfang:} \quad N \rightarrow N + 1$$

8.4.1 Ableitung aus dem Least-Squares-Schätzer

- Bisherige Messwerte: $\mathbf{y}(n-1) = [y(n-1) \dots y(n-N)]^T$

- Bisherige Schätzung basierend auf N Messwerten:

$$\begin{aligned} \hat{\mathbf{b}}(n-1) &= \mathbf{G}(n-1) \cdot \mathbf{y}(n-1) \\ &= \left(\Phi^T(n-1) \cdot \Phi(n-1) \right)^{-1} \cdot \Phi^T(n-1) \cdot \mathbf{y}(n-1) \end{aligned}$$

- Auf die Varianz des Messfehlers **bezogene Schätzfehlerkovarianz:**

$$\mathbf{P}(n-1) = \frac{\mathbf{C}_{\tilde{\mathbf{b}}\tilde{\mathbf{b}}}(n-1)}{\sigma_e^2} = \left(\Phi^T(n-1) \cdot \Phi(n-1) \right)^{-1}$$

8.4.1 Ableitung aus dem Least-Squares-Schätzer

- **Signalmodell** bei Hinzunahme eines neuen Messwertes $y(n)$:

$$\begin{bmatrix} y(n) \\ \mathbf{y}(n-1) \end{bmatrix} = \begin{bmatrix} \varphi^T(n) \\ \Phi(n-1) \end{bmatrix} \cdot \mathbf{b}(n) + \begin{bmatrix} e(n) \\ \mathbf{e}(n-1) \end{bmatrix}$$

oder:

$$\mathbf{y}(n) = \Phi(n) \cdot \mathbf{b}(n) + \mathbf{e}(n)$$

- Mit zunehmendem n wird **bei unabhängigen Messwerten** die bezogene Schätzfehlerkovarianz kleiner, da der LS-Schätzer dann konsistent ist:

$$\mathbf{P}(n) = \left(\Phi^T(n) \cdot \Phi(n) \right)^{-1} = \left(\varphi(n) \varphi^T(n) + \mathbf{P}^{-1}(n-1) \right)^{-1} \leq \mathbf{P}(n-1)$$

- Mit dem Matrix-Inversions-Lemma erhält man eine effiziente **Rekursionsbeziehung für die bezogene Schätzfehlerkovarianz** [KSW08]:

$$\mathbf{P}(n) = \mathbf{P}(n-1) - \mathbf{k}(n) \cdot \varphi^T(n) \cdot \mathbf{P}(n-1)$$

mit dem von $\mathbf{P}(n-1)$ abhängigen Gewichtungsvektor

$$\mathbf{k}(n) = \mathbf{P}(n-1) \cdot \varphi(n) \left[1 + \varphi^T(n) \cdot \mathbf{P}(n-1) \cdot \varphi(n) \right]^{-1}$$

- Für den **Parametervektor** ergibt sich ebenfalls eine effiziente rekursive Schätzvorschrift [KSW08]:

$$\hat{\mathbf{b}}(n) = \hat{\mathbf{b}}(n-1) + \mathbf{k}(n) \cdot \left[y(n) - \underbrace{\boldsymbol{\varphi}^T(n) \cdot \hat{\mathbf{b}}(n-1)}_{\hat{y}(n)} \right]$$

- Die Korrektur des Parametervektors berechnet sich aus der Differenz des neuen Messwertes $y(n)$ und des geschätzten Messwertes $\hat{y}(n)$, gewichtet mit dem von der Schätzfehlerkovarianz \mathbf{P} abhängigen Vektor $\mathbf{k}(n)$
- $\mathbf{k}(n)$ besagt, wie stark die neue Messung $y(n)$ den Schätzwert korrigiert
- Als Startmatrix für die rekursive Berechnung der Schätzfehlerkovarianz \mathbf{P} hat sich die folgende Wahl bewährt:

$$\mathbf{P}(0) = c \cdot \mathbf{I}, \quad 10^3 \leq c \leq 10^6$$

- Schätzung konvergiert mit wachsendem Stichprobenumfang n aufgrund der Konsistenz-Eigenschaft:

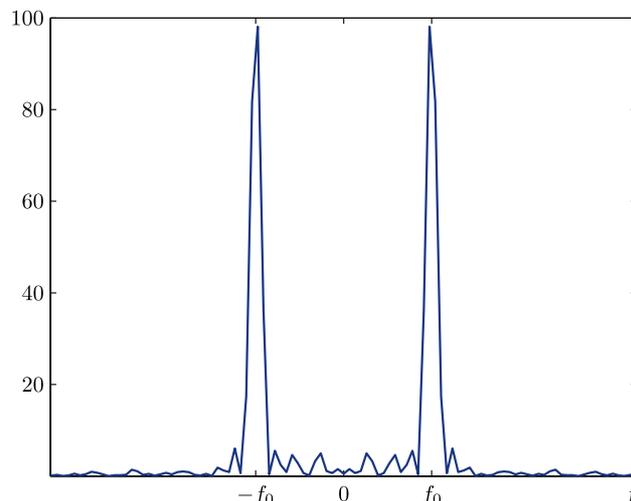
$$\lim_{n \rightarrow \infty} \mathbf{P}(n) = \mathbf{0}, \quad \lim_{n \rightarrow \infty} \mathbf{k}(n) = \mathbf{0}$$

Beispiel: Gestörte harmonische Schwingung

- Gegeben: Gestörtes harmonisches Messsignal

$$y(t) = A \sin(2\pi f_0 t + \varphi) + e(t)$$

- Ziel: Schätzung der Amplitude A und der Phase φ
- Bestimmung der Frequenz f_0 vorab aus der Autoleistungsdichte $S_{yy}(f)$



- Lineares Signalmodell durch Übergang zur äquivalenten Darstellung:

$$y(t) = b_1 \sin(2\pi f_0 t) + b_2 \cos(2\pi f_0 t) + e(t)$$

$$A = \sqrt{b_1^2 + b_2^2}, \quad \varphi = \arctan \frac{b_2}{b_1}$$

- Zeitdiskreter Regressionsvektor:

$$\varphi^T(n) = [\sin(2\pi f_0 t_A n) \cos(2\pi f_0 t_A n)]$$

- Parametervektor: $\mathbf{b}^T(n) = [b_1(n) \ b_2(n)]$

- Für die Parameter wählt man als Anfangswerte für $n = 0$:

$$y(0) = \hat{b}_1(0) \sin(0) + \hat{b}_2(0) \cos(0) \Rightarrow \hat{b}_2(0) = y(0)$$

$$\hat{b}_1(0) = 0 \quad (\text{willkürliche Wahl})$$

$$\mathbf{P}(0) = c \cdot \mathbf{I}, \quad c = 10^3, \dots, 10^6$$

$$\mathbf{k}(n) = \mathbf{P}(n-1) \cdot \varphi(n) [1 + \varphi^T(n) \cdot \mathbf{P}(n-1) \cdot \varphi(n)]^{-1}$$

- Für $n = 1$ erhält man mit dem Gewichtungsvektor

$$\mathbf{k}(1) = \mathbf{P}(0) \cdot \varphi(1) [1 + \varphi^T(1) \cdot \mathbf{P}(0) \cdot \varphi(1)]^{-1}$$

den verbesserten Schätzwert

$$\hat{\mathbf{b}}(1) = \hat{\mathbf{b}}(0) + \mathbf{k}(1) \cdot [y(1) - \varphi^T(1) \cdot \hat{\mathbf{b}}(0)]$$

sowie die zugehörige Schätzfehlerkovarianzmatrix:

$$\mathbf{P}(1) = \mathbf{P}(0) - \mathbf{k}(1) \cdot \varphi^T(1) \cdot \mathbf{P}(0)$$

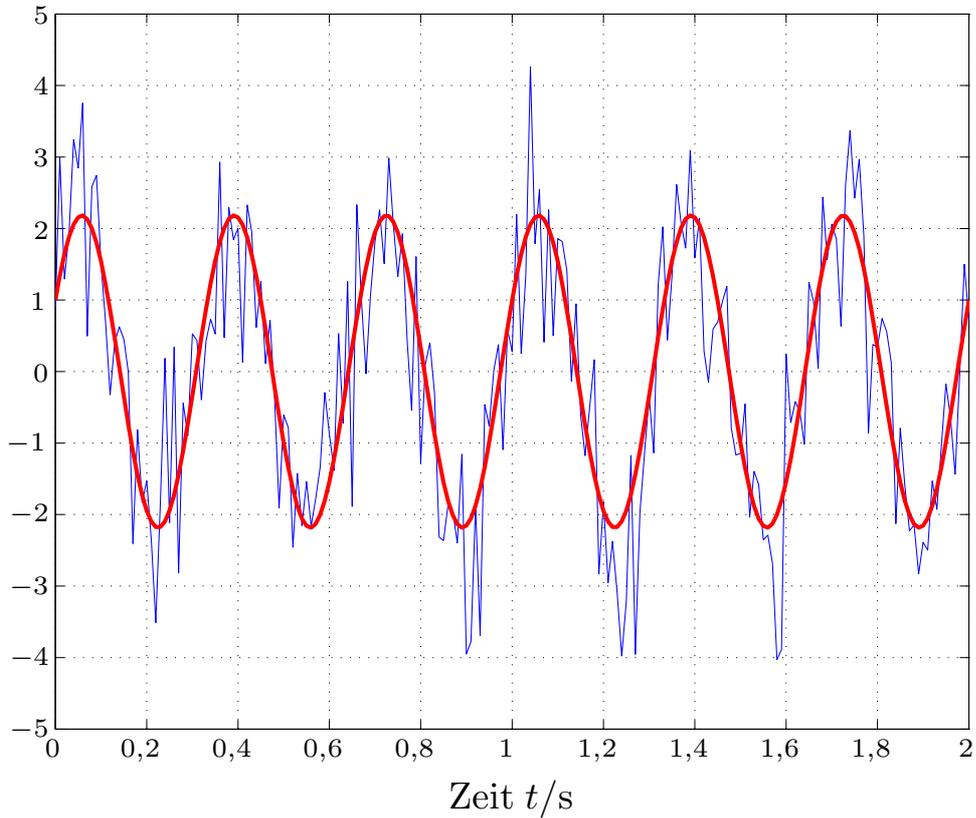
- Die letzten drei Gleichungen müssen in jedem Iterationsschritt erneut gelöst werden
- Der Schätzwert nähert sich dabei zunehmend dem wahren Wert \mathbf{b} an
- Bei einer iterativen Schätzung (200 Messwerte; $t_A = 10$ ms) erhält man:

$$\left\{ \begin{array}{l} \hat{A} = 2,18 \\ \hat{\varphi} = 0,47 \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} \hat{b}_1 = 1,94 \\ \hat{b}_2 = 0,99 \end{array} \right\} \quad \text{Wahr:} \quad \left\{ \begin{array}{l} A = 2,23 \\ \varphi = 0,46 \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} b_1 = 2 \\ b_2 = 1 \end{array} \right\}$$

8.4.1 Ableitung aus dem Least-Squares-Schätzer

Verrauschtes
Signal

Geschätztes
Signal

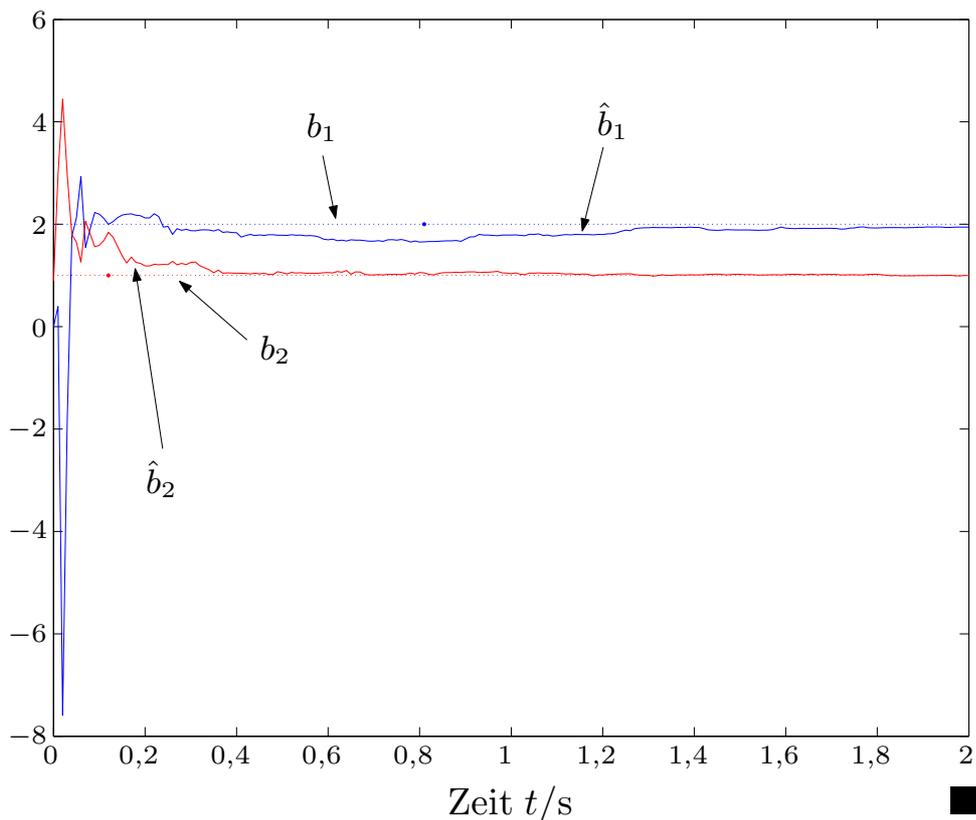


43

Prof. Dr.-Ing. F. Puente León – Methoden der Signalverarbeitung

8.4.1 Ableitung aus dem Least-Squares-Schätzer

Verlauf der
Parameter
 $\hat{b}_1(n), \hat{b}_2(n)$



44

Prof. Dr.-Ing. F. Puente León – Methoden der Signalverarbeitung

- **Bisher:** Schätzung zeitlich konstanter Parameter
- **Wunsch:** rekursive Schätzung **zeitlich veränderlicher** Parameter
- Konvergenzeigenschaft des rekursiven LS-Schätzers problematisch: neue Messwerte haben immer weniger Einfluss auf die Schätzung
- **Lösung:** Vergessen alter Messwerte (verschiedene heuristische Ansätze)

1. Ansatz: Exponentielle Gewichtung

- Gütekriterium wird um **Vergessensfaktor** λ erweitert (typisch: $\lambda = 0,95$):

$$J(n) = \sum_i (y(n-i) - \hat{y}(n-i))^2 \cdot \lambda^i$$

- Es folgen verallgemeinerte Rekursionsbeziehungen für $\mathbf{k}(n)$ und $\mathbf{P}(n)$:

$$\mathbf{k}_{\text{EG}}(n) = \mathbf{P}(n-1) \cdot \boldsymbol{\varphi}(n) [\lambda + \boldsymbol{\varphi}^T(n) \cdot \mathbf{P}(n-1) \cdot \boldsymbol{\varphi}(n)]^{-1}$$

$$\mathbf{P}(n) = \frac{1}{\lambda} \cdot [\mathbf{P}(n-1) - \mathbf{k}_{\text{EG}}(n) \cdot \boldsymbol{\varphi}^T(n) \cdot \mathbf{P}(n-1)]$$

→ Elemente von \mathbf{k} wg. $\lambda < 1$ größer; \mathbf{P} konvergiert langsamer gegen $\mathbf{0}$

2. Ansatz: Zyklisches Zurücksetzen

- Die laufende Schätzung wird zyklisch nach einer festen Anzahl von Rekursionsschritten beendet
- Der im letzten Rekursionsschritt geschätzte Parametervektor dient als Anfangswert für eine neue Rekursion; für die Matrix der Schätzfehlerkovarianz wird der Anfangswert $\mathbf{P}(0) = c \mathbf{I}$ gewählt
- Verfahren umgeht die Schwierigkeit, den Vergessensfaktor λ geeignet zu wählen

3. Ansatz: Berücksichtigung des aktuellen Schätzfehlers

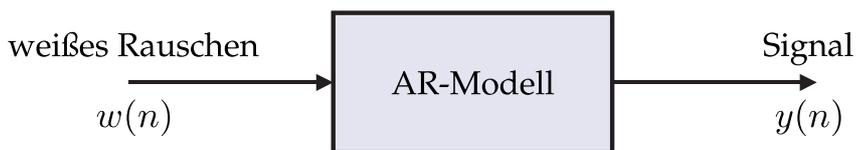
- Das „Einschlafen“ des rekursiven LS-Schätzers aufgrund der Konsistenzeigenschaft hat bei zeitvarianten Systemen einen steigenden Schätzfehler zur Folge
- Durch Addition des mit 10^3 gewichteten quadr. Schätzfehlers sinkt die Diagonale der Schätzfehlerkovarianzmatrix nicht unter diesen Wert ab:

$$\mathbf{P}(n) = \mathbf{P}(n-1) - \mathbf{k}(n) \cdot \boldsymbol{\varphi}^T(n) \cdot \mathbf{P}(n-1) + 10^3 \cdot (y(n) - \hat{y}(n))^2 \cdot \mathbf{I}$$

8.5 AR-Schätzung

8.5 AR-Schätzung

- Die AR-Schätzung basiert auf dem AR-Signalmodell aus Abschnitt 7.2.3:



Annahmen:

- $y(n)$ schwach stationär
- w, y unkorreliert

- Schätzung der AR-Modellparameter (Koeffizienten der Systemfunktion)
- Anwendungen: Spektralschätzung, Signalprädiktion, Fehlererkennung

8.5.1 Yule-Walker-Gleichung

- Übertragungsfunktion des AR-Systems besitzt nur ein Nennerpolynom:

$$G(z) = \frac{Y(z)}{W(z)} = \frac{1}{1 + \sum_{k=1}^K a(k) z^{-k}} = \frac{1}{A(z)}, \quad a(K) \neq 0$$

- Durch inverse z-Transformation erhält man die Differenzengleichung:

$$y(n) + a(1) \cdot y(n-1) + \dots + a(K) \cdot y(n-K) = w(n)$$

$$y(n) + a(1) \cdot y(n-1) + \dots + a(K) \cdot y(n-K) = w(n)$$

- Differenzengleichung in Vektorform:

$$y(n) + \mathbf{c}^T \mathbf{y}(n-1) = w(n)$$

$$\mathbf{c} = [a(1) \ a(2) \ \dots \ a(K)]^T \quad \text{AR-Koeffizienten}$$

$$\mathbf{y}(n-1) = [y(n-1) \ \dots \ y(n-K)]^T$$

- Multiplikation von rechts mit $\mathbf{y}^T(n-1)$ und Erwartungswertbildung:

$$\underbrace{E \{ y(n) \mathbf{y}^T(n-1) \}}_{= \mathbf{r}_{yy}^T(1)} + \mathbf{c}^T \cdot \underbrace{E \{ \mathbf{y}(n-1) \mathbf{y}^T(n-1) \}}_{= \mathbf{R}_{yy}} = \underbrace{E \{ w(n) \mathbf{y}^T(n-1) \}}_{= \mathbf{0}^T \text{ da unkorreliert}}$$

- Durch Invertierung der Autokorrelationsmatrix erhält man die **Yule-Walker-Gleichung**:

$$\mathbf{c} = -\mathbf{R}_{yy}^{-1} \mathbf{r}_{yy}(1)$$

Spektralschätzung

- **Idee:** Signal $y(t)$, dessen Spektrum zu schätzen ist, wird als Ausgang eines AR-Modells interpretiert
- Approximative Berechnung der Autokorrelationsmatrix \mathbf{R}_{yy} und des Korrelationsvektors $\mathbf{r}_{yy}(1)$ durch Mittelung aus N Messwerten
- Berechnung der K Koeffizienten des AR-Modells über Yule-Walker-Gleichung
- Die **Leistungsdichte des Ausgangssignals** folgt aus:

$$G(z) = \frac{Y(z)}{W(z)} = \frac{1}{A(z)} \quad \Longleftrightarrow \quad Y(z) = \frac{W(z)}{A(z)}$$

durch Multiplikation mit der konjugiert komplexen Größe zu:

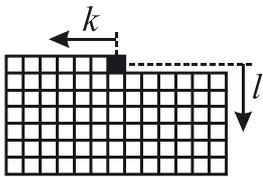
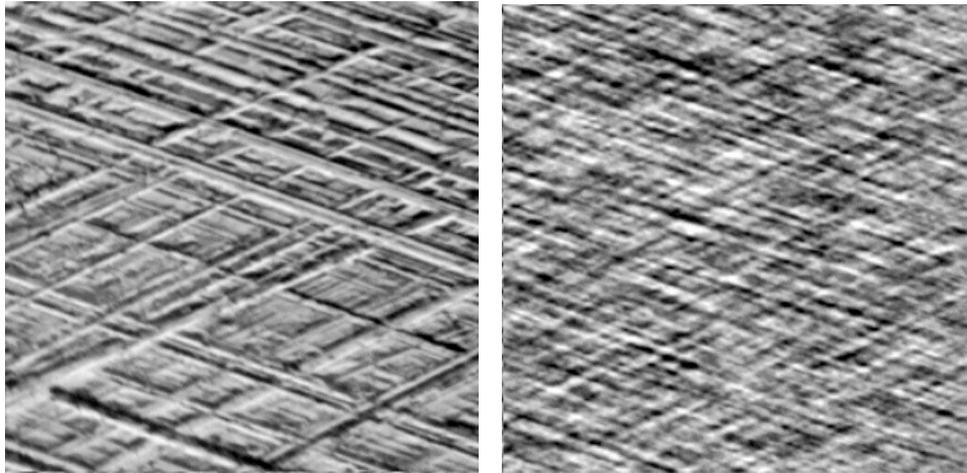
$$S_{yy}(z) = \frac{\sigma_w^2}{A(z) \cdot A(z^{-1})}$$

- Bei niedriger Ordnung K des Parametervektors \mathbf{c} ist dieser Ansatz deutlich effizienter als eine Mittelung von Periodogrammen der Länge N

8.5.1 Yule-Walker-Gleichung

Beispiel: Modellierung und parametrische Beschreibung von Texturen

- Gegeben: Intensitätsbild einer realen Textur
- Gesucht: Merkmale zur Beschreibung der Textur (→ AR-Koeffizienten)



Original

Synthese mit AR-Modell
der Ordnung 84
(85 Parameter)

\mathcal{U} : kausale Umgebung des Ursprungs

8.5.2 Prädiktionsfilter

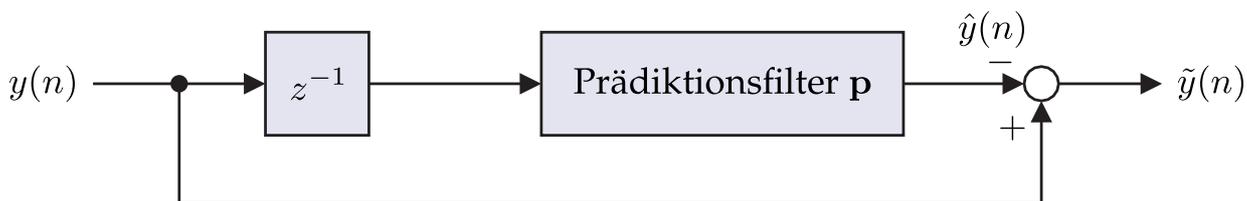
- **Wunsch:** Prädiktion des aktuellen Signalwerts $y(n)$ aus den K vergangenen Werten durch lineares Filter:

$$\hat{y}(n) = \mathbf{p}^T \cdot \mathbf{y}(n-1)$$

- Gütemaß zur Bestimmung von \mathbf{p} : $E \left\{ (y(n) - \hat{y}(n))^2 \right\} \rightarrow \min$
- Aus dem Projektionstheorem erhält man nach kurzer Rechnung [KSW08]:

$$\mathbf{p} = \mathbf{R}_{yy}^{-1} \mathbf{r}_{yy}(1) = -\mathbf{c}$$

- Für den Prädiktionsfehler folgt unmittelbar (vgl. Folie 49): $\tilde{y}(n) = w(n)$
- Das AR-Modell ist optimal angepasst, wenn der Prädiktionsfehler minimale Varianz hat



Prädiktionsfehlerfilter

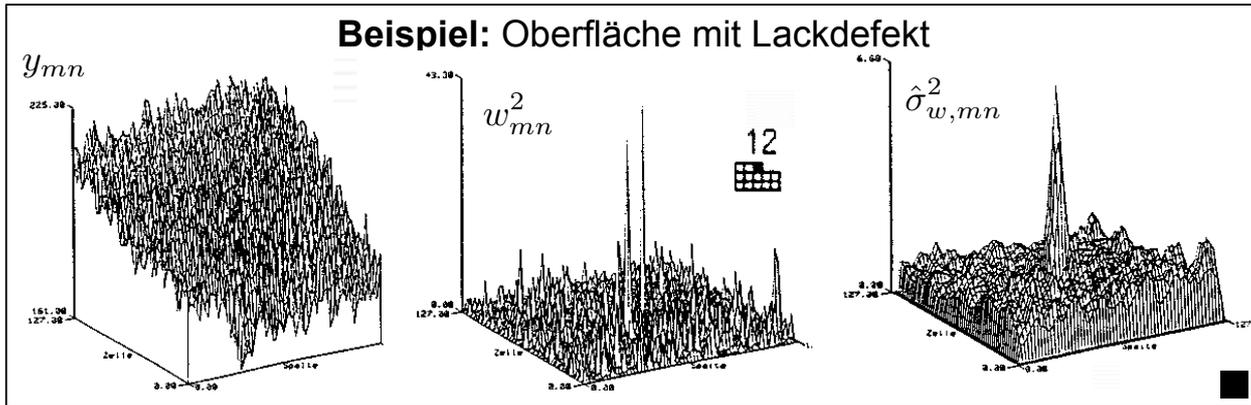
Beispiel: Prädiktionsfehlerfilter für Texturen basierend auf AR-Modell

■ Prädiktionsfehler:

$$w_{mn} = y_{mn} - \underbrace{\mathbf{p}^T \mathbf{y}_{mn}}_{\hat{y}_{mn}}$$

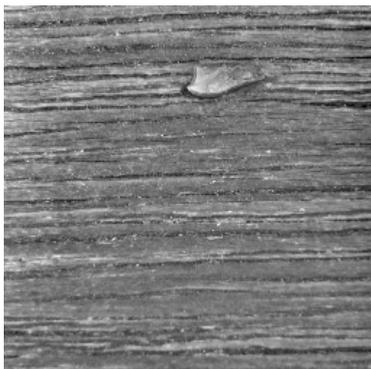
\mathbf{p}^T verkörpert Wissen \mathbf{R}_{yy} über Nutzsignal
 \hat{y}_{mn} ← Vorhersage aus einer Umgebung von (m, n)
 w_{mn} ← Prädiktionsfehler (Vorhersagefehler)

■ An lokalen Abweichungen von y_{mn} bezüglich \hat{y}_{mn} hat w_{mn} eine hohe Energiedichte → lokale Varianz von w_{mn} hat Maxima bei Defekten

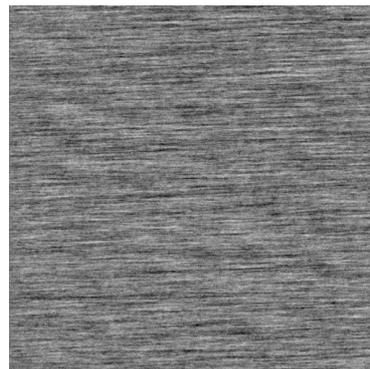


Beispiel: Fehlerdetektion in einer Holztextur

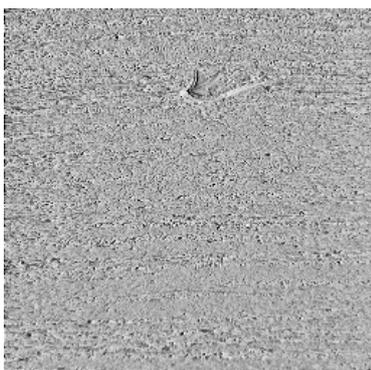
Original-
aufnahme
 y_{mn}



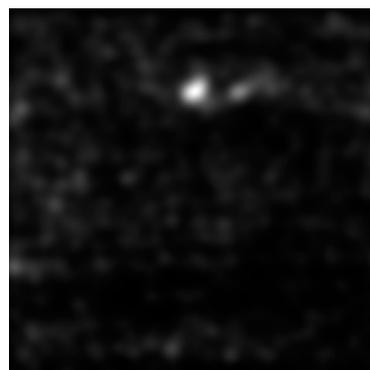
Prädiktion
 \hat{y}_{mn}



Prädiktions-
fehler
 w_{mn}



Schätzung
der lokalen
Varianz
von w_{mn}
($\text{TP}\{w_{mn}^2\}$)

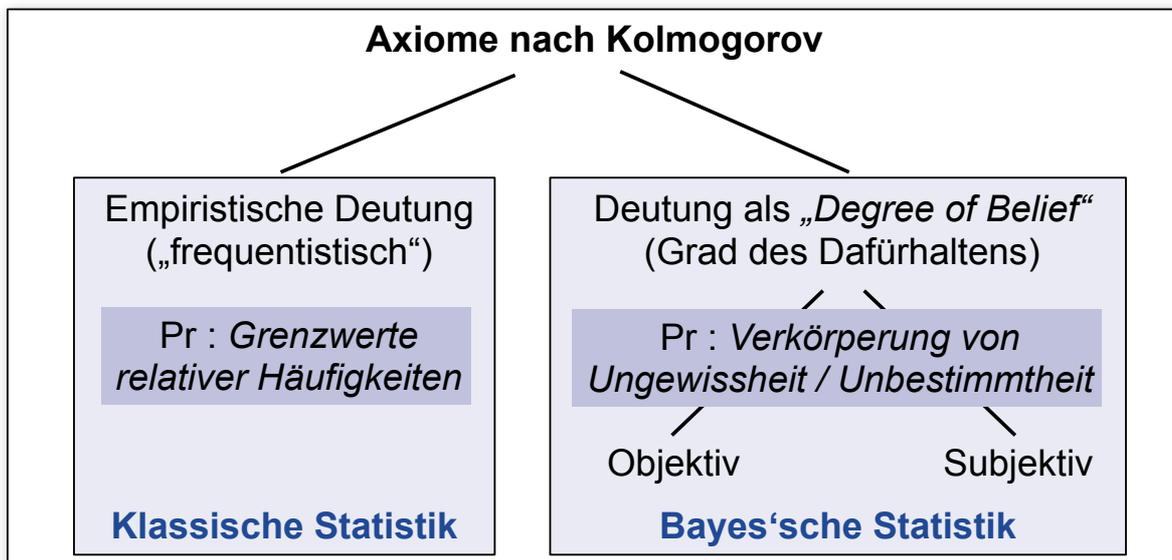


8.6 Bayes-Schätzung

8.6 Bayes-Schätzung

- **Bisher:** Annahme einer vom Parametervektor \mathbf{b} unabhängigen mittelwertfreien Störung e (Beschreibung mittels der Kovarianzmatrix \mathbf{C}_{ee})
- **Jetzt:** Allgemeine Wahrscheinlichkeitsverteilungen (WV)

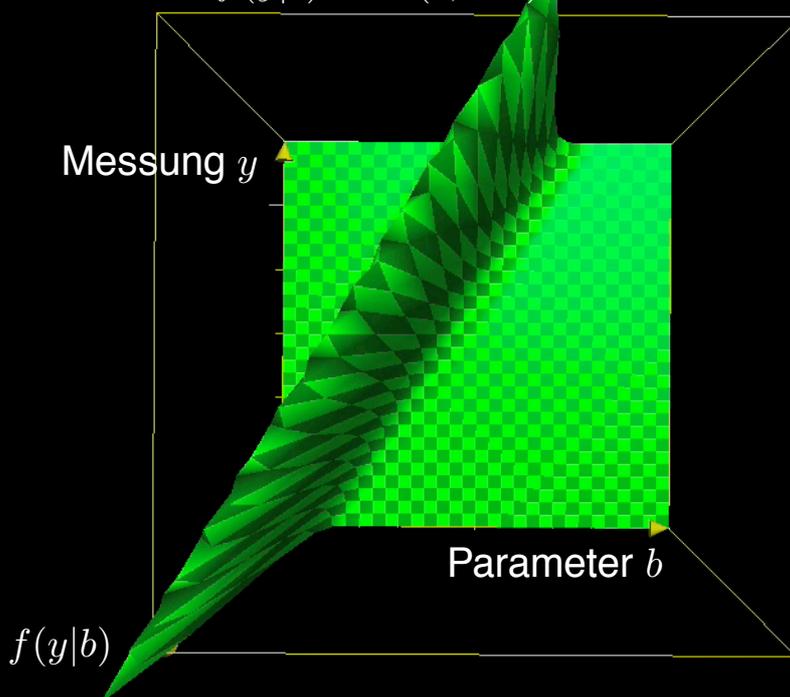
Was bedeutet Wahrscheinlichkeit? — zwei statistische „Denklager“



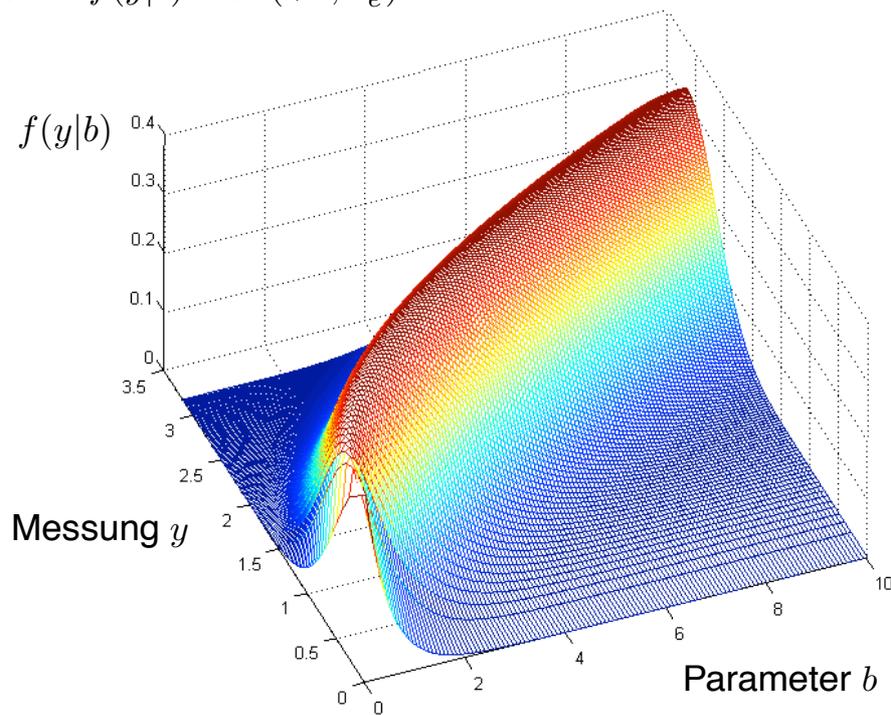
Klassische Statistik

- Wahrscheinlichkeiten als **relative Häufigkeiten** interpretiert
- Zufallsgrößen: Ergebnis der Wiederholung eines Zufallsexperimentes
 - Messungen y werden als zufällig angesehen
 - Parametervektor \mathbf{b} wird als **deterministische** Konstante erachtet
- Einsatz der Wahrscheinlichkeitsrechnung auf „Messverfahren“ $\mathbf{b} \mapsto y$ beschränkt
- **Messverfahren** mathematisch mittels der bedingten WV $f(y|\mathbf{b})$ (Likelihood-Funktion) beschrieben
- **Konsequenzen:**
 - **Unsymmetrische** Behandlung von Parametern \mathbf{b} und Messungen y
 - Wahrscheinlichkeitsaussagen können nicht auf \mathbf{b} bezogen werden
 - Keine spezifischen Methoden für Rückschluss von y auf \mathbf{b} vorhanden
 - Wichtigstes Schätzverfahren: **Maximum-Likelihood-Verfahren**
 - Likelihood-Funktion macht keine Wahrscheinlichkeitsaussage über \mathbf{b}

Likelihood-Funktion $f(y|b) = \mathcal{N}(b, c b^2)$



Beispiel: $f(y|b) = \mathcal{N}(\sqrt{b}, \sigma_e^2)$



Bayes'sche Statistik

- Wahrscheinlichkeiten als „*Degree of Belief*“ interpretiert
- WV verkörpern den Grad der Ungewissheit bzw. den Wissensstand bzgl. der betrachteten Größen
- Klassische Interpretation (relative Häufigkeiten) in der Bayes'schen Statistik zulässig (spezielle Ausprägung von Wissen)
- Vorhandenes Wissen muss in eine WV überführt werden
- Möglichst objektive Auffassung der „*Degree of Belief*“-Interpretation
- Sowohl Parameter \mathbf{b} als auch Messungen \mathbf{y} sind Zufallsgrößen
→ **symmetrische** Betrachtung mittels Verbund-WDF möglich:

$$f(\mathbf{b}, \mathbf{y}) = f(\mathbf{b}|\mathbf{y}) \cdot f(\mathbf{y}) = f(\mathbf{y}|\mathbf{b}) \cdot f(\mathbf{b})$$

Satz von Bayes

$f(\mathbf{b})$: A-priori-Dichte

$f(\mathbf{y}|\mathbf{b})$: Likelihood-Funktion

$f(\mathbf{b}|\mathbf{y})$: A-posteriori-Dichte

- Erst durch die Interessenslage entsteht eine Unsymmetrie
→ Hilfsgrößen (hier die Messungen \mathbf{y}) werden herausintegriert
- Für eine konkrete Messung \mathbf{y}_0 erhält man durch Integration:

$$\int_{\mathcal{Y}} f(\mathbf{b}, \mathbf{y}) d\mathbf{y} = \int_{\mathcal{Y}} f(\mathbf{b}|\mathbf{y}) f(\mathbf{y}) d\mathbf{y} \xrightarrow{\mathbf{y}=\mathbf{y}_0} \int_{\mathcal{Y}} f(\mathbf{b}|\mathbf{y}) \underbrace{\delta(\mathbf{y} - \mathbf{y}_0)}_{f(\mathbf{y}_0)} d\mathbf{y} = f(\mathbf{b}|\mathbf{y}_0)$$

- Das Ergebnis ist die **A-posteriori-WDF**

$$f(\mathbf{b}|\mathbf{y}) = \frac{f(\mathbf{y}|\mathbf{b}) \cdot f(\mathbf{b})}{f(\mathbf{y})},$$

auf welche die Bayes'sche Statistik ihre Aussagen stützt

- Im Gegensatz zur Likelihood-Funktion ist die A-posteriori-WDF eine **echte WDF** bzgl. des Parametervektors \mathbf{b}
- In der Bayes'schen Statistik ergibt sich eine **schlanke Theorie** durch die einheitliche Behandlung der beteiligten Größen

	Klassische Statistik	Bayes'sche Statistik
Wahrscheinlichkeitsdeutung	relative Häufigkeit	<i>Degree of Belief</i>
Parametervektor	\mathbf{b} : deterministische Konstante	\mathbf{b} : Zufallsgröße
Messungen	\mathbf{y} : Zufallsgröße	\mathbf{y} : Zufallsgröße
Beschreibung von	$\mathbf{b} \mapsto \mathbf{y}$	\mathbf{b} und $\mathbf{b} \mapsto \mathbf{y}$
Qualität der Aussagen	<ul style="list-style-type: none"> • pre-experimentell • lokal (abhängig vom unbekanntem wahren Wert des Parametervektors \mathbf{b}) 	<ul style="list-style-type: none"> • pre- oder post-experimentell • global (gemittelt über \mathbf{b} → unabhängig vom „wahren Wert“ des Parametervektors \mathbf{b})
Theorie	kompliziert wegen asymmetrischer Behandlung von \mathbf{b} und \mathbf{y}	übersichtlich, intuitiv einsichtig

- Im Folgenden werden die wichtigsten Techniken zur **Bayes'schen Parameterschätzung** gezeigt
- Ansatz wie bisher, jedoch um nichtnegative, konvexe **Kostenfunktion** C (z. B. quadratische Funktion) zur Gewichtung des Schätzfehlers ergänzt:

$$C(\mathbf{b} - \hat{\mathbf{b}}) \geq 0$$

- Verschiedene Kostenfunktionen C führen zu unterschiedlichen Schätzern
- Dabei ist immer der **Erwartungswert** der Kostenfunktion zu minimieren:

$$E_{\mathbf{b}, \mathbf{y}} \{C(\mathbf{b} - \hat{\mathbf{b}})\} = \int_{\mathcal{Y}} \int_{\mathcal{B}} C(\mathbf{b} - \hat{\mathbf{b}}) f(\mathbf{b}, \mathbf{y}) d\mathbf{b} d\mathbf{y} \rightarrow \min$$

$$E_{\mathbf{b}, \mathbf{y}} \{C(\mathbf{b} - \hat{\mathbf{b}})\} = \int_{\mathcal{Y}} \left(\int_{\mathcal{B}} C(\mathbf{b} - \hat{\mathbf{b}}) \cdot f(\mathbf{b}|\mathbf{y}) d\mathbf{b} \right) \cdot f(\mathbf{y}) d\mathbf{y} \rightarrow \min$$

$$\Leftrightarrow Q = \int_{\mathcal{B}} C(\mathbf{b} - \hat{\mathbf{b}}) \cdot f(\mathbf{b}|\mathbf{y}) d\mathbf{b} = E_{\mathbf{b}|\mathbf{y}} \{C(\mathbf{b} - \hat{\mathbf{b}})\} \rightarrow \min$$

Ansatz 1: Minimaler quadratischer Schätzfehler

- Quadratische Kostenfunktion:

$$C(\mathbf{b} - \hat{\mathbf{b}}) = (\hat{\mathbf{b}} - \mathbf{b})^T (\hat{\mathbf{b}} - \mathbf{b})$$

- Einsetzen in Gütekriterium Q (inneres Integral):

$$Q = \int_{\mathcal{B}} (\hat{\mathbf{b}} - \mathbf{b})^T (\hat{\mathbf{b}} - \mathbf{b}) \cdot f(\mathbf{b}|\mathbf{y}) d\mathbf{b} = \int_{\mathcal{B}} \sum_i (\hat{b}_i(\mathbf{y}) - b_i)^2 f(\mathbf{b}|\mathbf{y}) d\mathbf{b} \rightarrow \min$$

- Notwendige Bedingung für Minimum:

$$\frac{\partial Q}{\partial \hat{b}_i(\mathbf{y})} \stackrel{!}{=} 0 \Leftrightarrow \int_{\mathcal{B}} 2(\hat{b}_i(\mathbf{y}) - b_i) f(\mathbf{b}|\mathbf{y}) d\mathbf{b} = 0 \Leftrightarrow \hat{b}_i(\mathbf{y}) = \int_{\mathcal{B}} b_i f(\mathbf{b}|\mathbf{y}) d\mathbf{b}$$

- **Schätzer mit minimalem mittleren quadratischen Fehler:**

$$\hat{\mathbf{b}}(\mathbf{y}) = \int_{\mathcal{B}} \mathbf{b} f(\mathbf{b}|\mathbf{y}) d\mathbf{b} = E_{\mathbf{b}|\mathbf{y}} \{\mathbf{b}\}$$

A-posteriori-Erwartungswert von \mathbf{b}

Ansatz 2: Konstante Gewichtung großer Fehler

- Durch die folgende Kostenfunktion werden Fehler mit Beträgen kleiner als Δ ignoriert:

$$C(\mathbf{b} - \hat{\mathbf{b}}) = \begin{cases} 0 & \text{für } \|\mathbf{b} - \hat{\mathbf{b}}\| < \Delta, \quad \Delta > 0 \\ 1 & \text{sonst} \end{cases}$$

- Einsetzen in Gütekriterium Q (inneres Integral):

$$Q = \int_{\mathcal{B}} C(\mathbf{b} - \hat{\mathbf{b}}) \cdot f(\mathbf{b}|\mathbf{y}) \, d\mathbf{b} = 1 - \int_{\{\mathbf{b} \mid \|\mathbf{b} - \hat{\mathbf{b}}\| < \Delta\}} f(\mathbf{b}|\mathbf{y}) \, d\mathbf{b} \rightarrow \min$$

kleine Fehler

- Für $\Delta \rightarrow 0$ folgt:

$$\hat{\mathbf{b}}_{\text{MAP}}(\mathbf{y}) = \arg \max_{\mathbf{b}} \{f(\mathbf{b}|\mathbf{y})\}$$

MAP-Schätzer

- Wegen der Monotonie der Logarithmusfunktion kann man äquivalent die logarithmierte A-posteriori-WDF maximieren (bringt oft analytische Vorteile)

Diskussion

- In vielen praktisch relevanten Fällen liefern A-posteriori-Erwartungswert und MAP-Schätzung das gleiche Ergebnis
- Bayes'sche Schätzung: allgemeinste und umfassendste Parameterschätzmethode
- Durch Spezialisierung lassen sich andere grundlegende Schätzmethoden als Sonderfälle ableiten
- Oftmals wird der Einsatz der Bayes'schen Parameterschätzung dadurch erschwert, dass die Bestimmung der A-posteriori-Dichte die numerische Berechnung mehrdimensionaler Integrale erfordert – etwa im Kontext von Klassifikationsaufgaben, siehe z. B. [Duda, Hart & Stork: Pattern Classification, 2001]
- Außerdem kann die A-priori-Dichte des Parametervektors \mathbf{b} oft nicht genau angegeben werden

- Logarithmus ist monoton; daher ist das Maximum von $f(\mathbf{b}|\mathbf{y})$ mit dem Maximum von $\ln f(\mathbf{b}|\mathbf{y})$ identisch
- Aus dem Satz von Bayes folgt durch Logarithmierung und Differentiation:

$$\frac{\partial \ln f(\mathbf{b}, \mathbf{y})}{\partial \mathbf{b}} = \frac{\partial \ln f(\mathbf{b}|\mathbf{y})}{\partial \mathbf{b}} + \underbrace{\frac{\partial \ln f(\mathbf{y})}{\partial \mathbf{b}}}_{=0} = \frac{\partial \ln f(\mathbf{y}|\mathbf{b})}{\partial \mathbf{b}} + \frac{\partial \ln f(\mathbf{b})}{\partial \mathbf{b}} \stackrel{!}{=} \mathbf{0}$$

- Meistens ist $f(\mathbf{b})$ unbekannt; daher Annahme einer Gleichverteilung über unendlich breites Intervall (vgl. Laplace'sches Prinzip des unzureichenden Grundes):

$$\Rightarrow \frac{\partial \ln f(\mathbf{b})}{\partial \mathbf{b}} = \mathbf{0}$$

- Dann resultiert als Sonderfall des MAP-Schätzers der in der Klassischen Statistik verwendete **Maximum-Likelihood-Schätzer (ML)**:

$$\frac{\partial \ln f(\mathbf{y}|\mathbf{b})}{\partial \mathbf{b}} = \mathbf{0} \quad \Leftrightarrow \quad \hat{\mathbf{b}}_{\text{ML}}(\mathbf{y}) = \arg \max_{\mathbf{b}} \{f(\mathbf{y}|\mathbf{b})\}$$

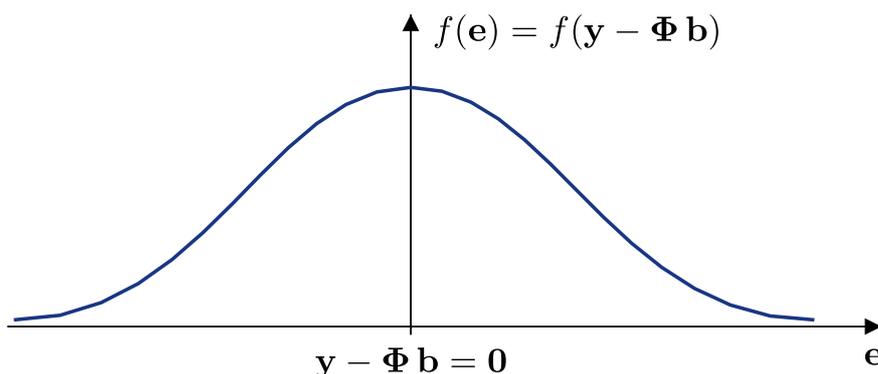
Beispiel: MAP- und ML-Schätzer bei Normalverteilung

- Lineares Signalmodell: $\mathbf{y} = \Phi \mathbf{b} + \mathbf{e}$, \mathbf{C}_{ee} und \mathbf{R}_{bb} bekannt
- Likelihood-Funktion $f(\mathbf{y}|\mathbf{b})$ stimmt dann bis auf die Verschiebung um $\Phi \mathbf{b}$ mit der Dichte $f(\mathbf{e})$ des Fehlers überein:

$$\mathbf{e} = \mathbf{y} - \Phi \mathbf{b}$$

- Likelihood-Funktion sei normalverteilt:

$$f(\mathbf{y}|\mathbf{b}) = \frac{1}{(2\pi)^{N/2} |\mathbf{C}_{ee}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y} - \Phi \mathbf{b})^T \mathbf{C}_{ee}^{-1}(\mathbf{y} - \Phi \mathbf{b})\right)$$



MAP-Schätzer:

- Annahme: Parametervektor \mathbf{b} ebenfalls normalverteilt

$$f(\mathbf{b}) = \frac{1}{(2\pi)^{N/2} |\mathbf{R}_{bb}|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{b}^T \mathbf{R}_{bb}^{-1} \mathbf{b}\right) \quad \text{A-priori-WDF}$$

- Logarithmierte Verbund-WDF:

$$\begin{aligned} \ln f(\mathbf{y}, \mathbf{b}) &= \ln f(\mathbf{y}|\mathbf{b}) + \ln f(\mathbf{b}) \\ &= \text{const} - \frac{1}{2} (\mathbf{y} - \Phi \mathbf{b})^T \mathbf{C}_{ee}^{-1} (\mathbf{y} - \Phi \mathbf{b}) - \frac{1}{2} \mathbf{b}^T \mathbf{R}_{bb}^{-1} \mathbf{b} \end{aligned}$$

- Maximumsuche:

$$\frac{\partial \ln f(\mathbf{y}, \mathbf{b})}{\partial \mathbf{b}} = \Phi^T \mathbf{C}_{ee}^{-1} (\mathbf{y} - \Phi \mathbf{b}) - \mathbf{R}_{bb}^{-1} \mathbf{b} \stackrel{!}{=} \mathbf{0}$$

- Schätzung des Parametervektors (Auflösen nach \mathbf{b}):

$$\hat{\mathbf{b}} = \left(\Phi^T \mathbf{C}_{ee}^{-1} \Phi + \mathbf{R}_{bb}^{-1} \right)^{-1} \Phi^T \mathbf{C}_{ee}^{-1} \mathbf{y} \quad \text{(GM-Schätzer)}$$

ML-Schätzer:

- Annahme: Parametervektor \mathbf{b} gleichverteilt über unendlichem Intervall (unbekannte Statistik)
- Partielle Ableitung der logarithmierten Verbund-WDF:

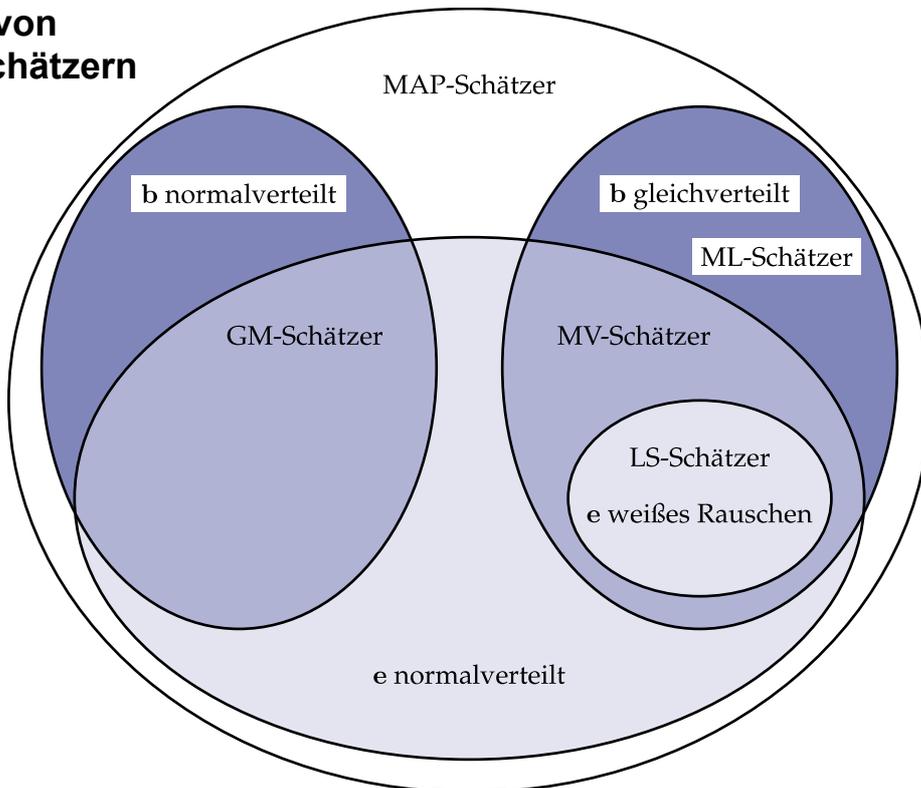
$$\frac{\partial \ln f(\mathbf{y}, \mathbf{b})}{\partial \mathbf{b}} = \frac{\partial \ln f(\mathbf{y}|\mathbf{b})}{\partial \mathbf{b}} = \Phi^T \mathbf{C}_{ee}^{-1} (\mathbf{y} - \Phi \mathbf{b}) \stackrel{!}{=} \mathbf{0}$$

- Schätzung des Parametervektors:

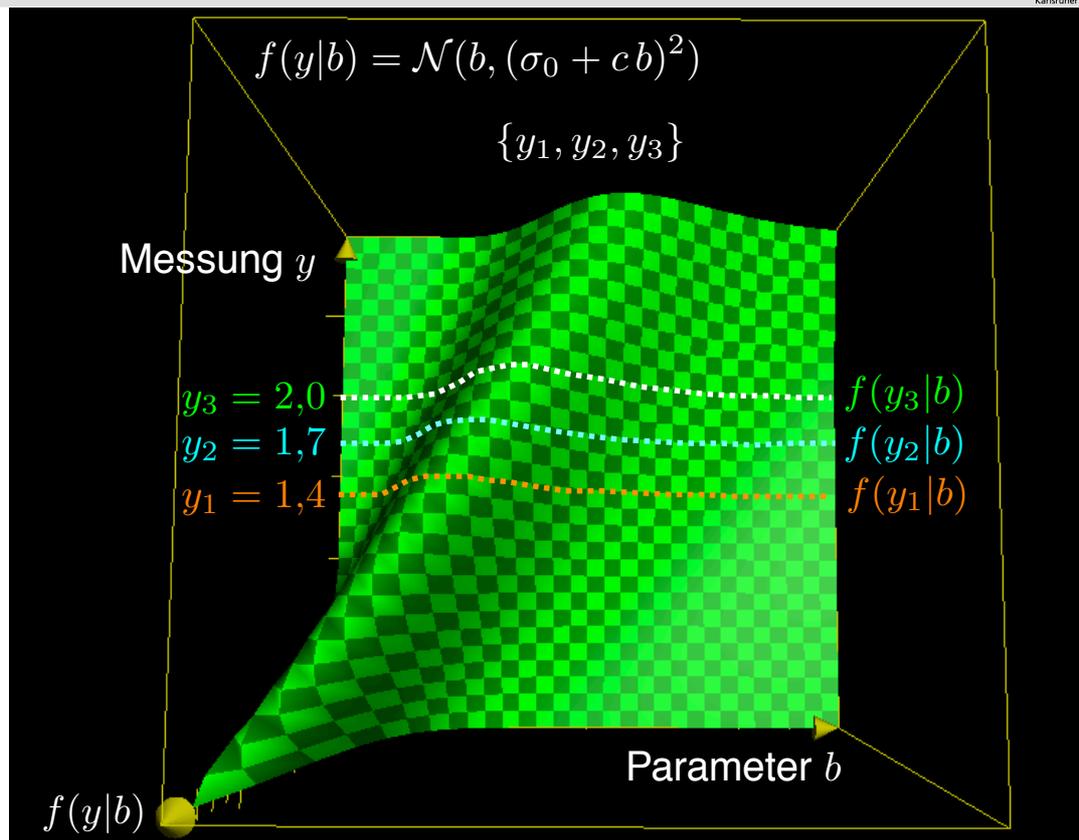
$$\hat{\mathbf{b}} = \left(\Phi^T \mathbf{C}_{ee}^{-1} \Phi \right)^{-1} \Phi^T \mathbf{C}_{ee}^{-1} \mathbf{y} \quad \text{(MV-Schätzer)}$$

- Bei einer Statistik zweiter Ordnung (normalverteilten Größen) entspricht
 - der MAP-Schätzer dem Gauß-Markov-Schätzer
 - der ML-Schätzer dem Minimum-Varianz-Schätzer

Klassen von Bayes-Schätzern

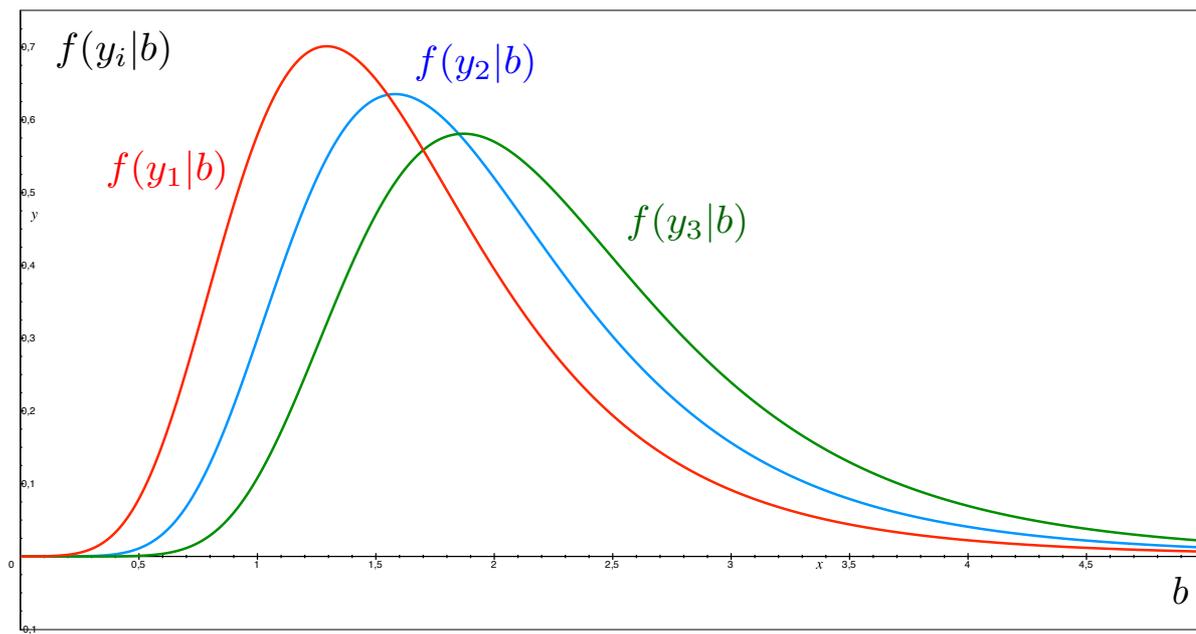


Bsp.

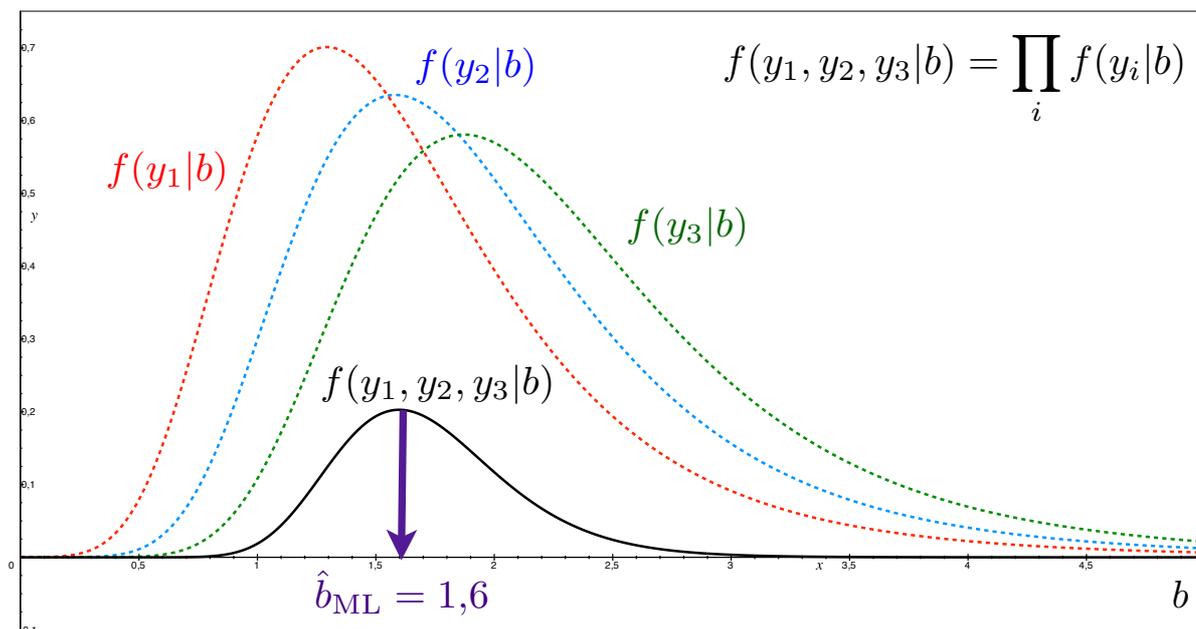


Fusion der einzelnen Likelihood-Funktionen

- Annahme: statistisch unabhängige Messungen

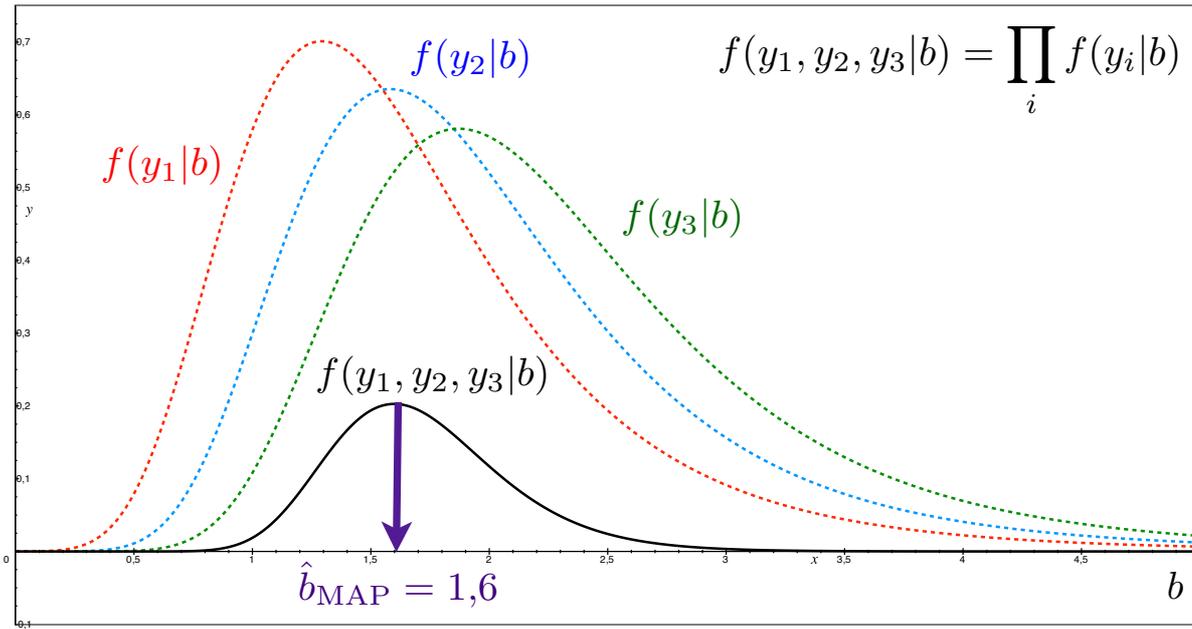


ML-Schätzer



MAP-Schätzer bei vagem A-priori-Wissen

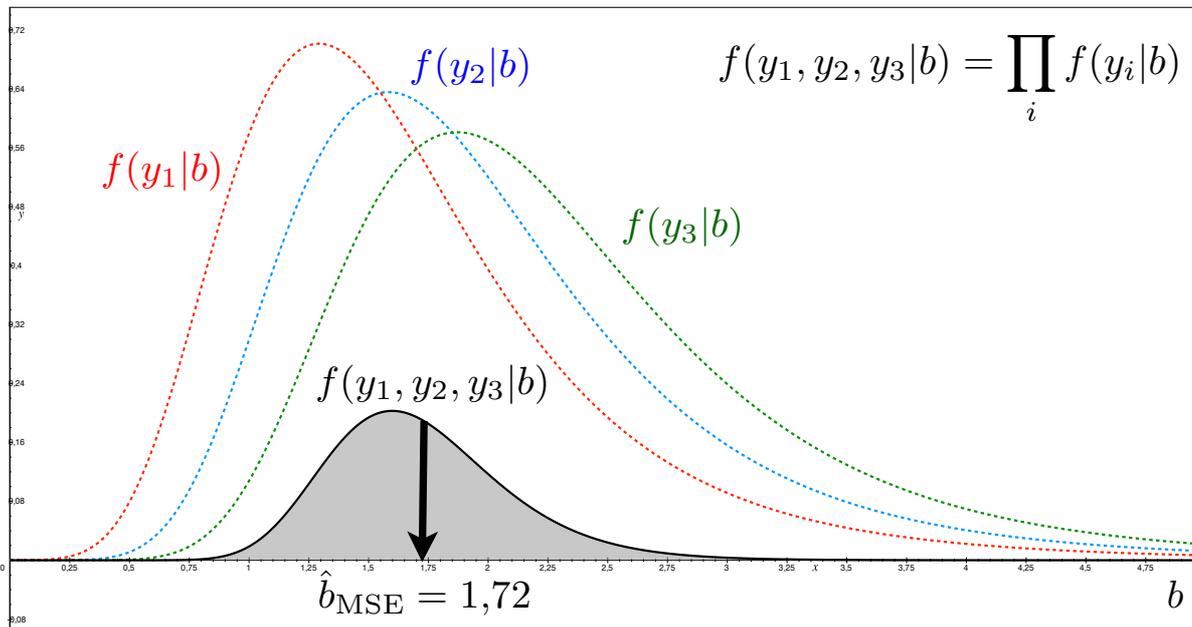
■ Annahme: $f(b) = \text{const.} \Rightarrow \hat{b}_{\text{MAP}} = \hat{b}_{\text{ML}}$



8.6.2 Maximum-Likelihood-Schätzer

A-posteriori-Erwartungswert bei vagem A-priori-Wissen

■ Annahme: $f(b) = \text{const.}$



■ Fisher'sche Informationsmatrix:

$$\mathbf{J} = \left(-\mathbb{E}_{\mathbf{y}|\mathbf{b}} \left\{ \frac{\partial^2}{\partial b_i \partial b_j} \ln f(\mathbf{y}|\mathbf{b}) \right\} \right)_{i,j}$$

Diese $K \times K$ -Matrix liefert ein Maß für die Information, die der (beobachtbare) Zufallsvektor \mathbf{y} über den Parametervektor \mathbf{b} trägt

Cramér-Rao-Schranke (CRB, *Cramér-Rao bound*, ohne Beweis):

$$\mathbf{J}^{-1} \leq \mathbf{C}_{\tilde{b}\tilde{b}}$$

- Die Schätzfehlerkovarianzmatrix jedes erwartungstreuen Schätzers ist größer oder gleich der inversen Fisher'schen Informationsmatrix – letztere ist also eine **untere Schranke** für die **Varianz eines Schätzers**
- Bei Gleichheit heißt der Schätzer **effizient**

Beispiel: Maximum-Likelihood-Schätzer

- Mobilfunk: Häufig keine direkte Sichtverbindung zw. Sender u. Empfänger
→ Rayleigh-Verteilung der Amplitude A des Empfangssignals

$$f(A) = \begin{cases} \frac{2A}{P} \cdot \exp\left(-\frac{A^2}{P}\right) & A \geq 0 \\ 0 & \text{sonst} \end{cases}$$

- Dabei ist P die mittlere Empfangsleistung: $\mathbb{E}\{A^2\} = P$
- **Ziel:** ML-Schätzer für die **mittlere Leistung** P aus N Messungen $\{A_i\}$ der Amplitude A
- Log-Likelihood-Funktion:

$$\begin{aligned} \ln f(\mathbf{A}|P) &= \sum_{i=1}^N \ln f(A_i|P) = \sum_{i=1}^N \left(\ln 2 + \ln A_i - \ln P - \frac{A_i^2}{P} \right) \\ &= N \ln 2 - N \ln P + \sum_{i=1}^N \left(\ln A_i - \frac{A_i^2}{P} \right) \end{aligned}$$

8.6.4 Cramér-Rao-Ungleichung

- Herleitung **ML-Schätzer** für mittlere Empfangsleistung P :

$$\frac{\partial \ln f(\mathbf{A}|P)}{\partial P} = -\frac{N}{P} + \sum_{i=1}^N \frac{A_i^2}{P^2} \stackrel{!}{=} 0 \quad \Rightarrow \quad \hat{P} = \frac{1}{N} \sum_{i=1}^N A_i^2$$

- Der ML-Schätzer ist erwartungstreu:

$$\mathbb{E}\{\hat{P}\} = \mathbb{E}\left\{\frac{1}{N} \sum_{i=1}^N A_i^2\right\} = \frac{1}{N} \sum_{i=1}^N \underbrace{\mathbb{E}\{A_i^2\}}_{=P} = P$$

- Berechnung der **Schätzfehlervarianz** zur Überprüfung von Konsistenz und Effizienz:

$$\sigma_{\hat{P}}^2 = \mathbb{E}\left\{(P - \hat{P})^2\right\} = \mathbb{E}\left\{\left(P - \frac{1}{N} \sum_{i=1}^N A_i^2\right)^2\right\} = \dots = \frac{P^2}{N}$$

8.6.4 Cramér-Rao-Ungleichung

- Schätzer ist **konsistent** wegen:

$$\lim_{N \rightarrow \infty} \sigma_{\hat{P}}^2 = \lim_{N \rightarrow \infty} \frac{P^2}{N} = 0$$

- Überprüfung der **Effizienz** mittels der Cramér-Rao-Ungleichung [Fol. 77]:

$$\sigma_{\hat{P}}^2 \geq -\left(\mathbb{E}\left\{\frac{\partial^2 \ln f(\mathbf{A}|P)}{\partial P^2}\right\}\right)^{-1} \quad \sigma_{\hat{P}}^2 = \frac{P^2}{N}$$

mit

$$\frac{\partial^2 \ln f(\mathbf{A}|P)}{\partial P^2} = \frac{N}{P^2} - 2 \sum_{i=1}^N \frac{A_i^2}{P^3}$$

Erwartungswertbildung:

$$\mathbb{E}\left\{\frac{\partial^2 \ln f(\mathbf{A}|P)}{\partial P^2}\right\} = \frac{N}{P^2} - 2 \sum_{i=1}^N \frac{\mathbb{E}\{A_i^2\}}{P^3} = \frac{N}{P^2} - \frac{2N}{P^2} = -\frac{N}{P^2}$$

- Die Schätzfehlervarianz erreicht die Cramér-Rao-Schranke.
Der Schätzer ist somit **effizient**.