

# 8. Parameterschätzung

## 8. Parameterschätzung

- Ziel: Schätzung der Parameter eines Signalmodells
- Häufig **Least-Squares-Schätzer** verwendet (Abschn. 8.1, 8.3)
- Varianten bei besserer Kenntnis über die Statistik von Parametern und Fehlern (Abschn. 8.2):
  - **Gauß-Markov-Schätzer** (GM-Schätzer)
  - **Minimum-Varianz-Schätzer** (Sonderfall des GM-Schätzers)
- Allgemeinsten Schätzer: **Bayes-Schätzer** (Abschn. 8.6)
- Sonderfall davon: **Maximum-Likelihood-Schätzer**

# 8.1 Least-Squares-Schätzer

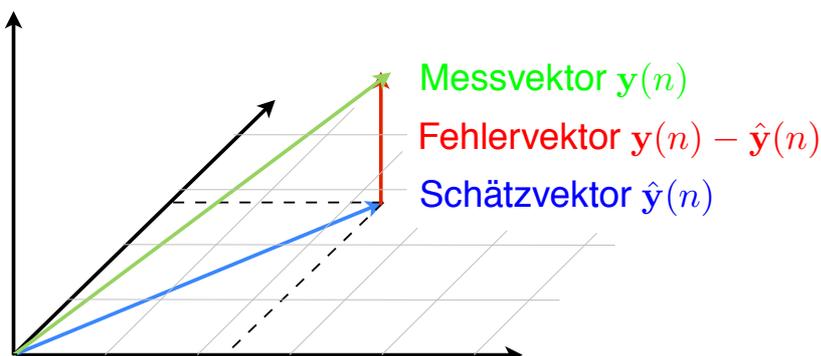
## 8.1.1 Parameterschätzung f. ein lineares Signalmodell

- **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:

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



- Für den **Schätzvektor** wird als **Signalmodell** eine Linearkombination von Basisvektoren (Spaltenvektoren der Beobachtungsmatrix  $\Phi$ ) 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:**

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

- Der LS-Schätzer ist **erwartungstreu**, falls  $\mathbf{e}$  mittelwertfrei ist:

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

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



### Beispiel: LS-Schätzer 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:

$$\bar{K}_0 = \frac{1}{N} \sum_{n=0}^{N-1} y(n) \cdot e^{\alpha nT} \quad y(n) = K_0 e^{-\alpha nT} + e(n)$$

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

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

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

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

- Konsistenzprüfung Mittelwertschätzer:

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

- Konsistenzprüfung LS-Schätzer:

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

- 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 ■

## 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  $e(n)$  sei mittelwertfreies **farbiges** Rauschen mit der bekannten **Autokovarianzmatrix**  $\mathbf{C}_{ee}$
- **Autokorrelationsmatrix**  $\mathbf{R}_{bb}$  des Parametervektors  $\mathbf{b}$  sei bekannt
- Parametervektor und Fehlervektor seien stochastisch unabhängig:

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

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

- Die Parameter  $b_k$  des Signalmodells werden als **Zufallsgrößen** interpretiert (Bayes-Theorie)
- Die Approximation einer Funktion  $y(n)$  im Funktionenraum (Folie 4) kann dann durch eine Approximation der stochastischen Variablen  $b_k$  im **Vektorraum stochastischer Variabler** ersetzt werden (MMSE):

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

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

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

- **Schätzung**  $\hat{\mathbf{b}}(n) = \mathbf{G}(n)\mathbf{y}(n)$  für jeden Parameter  $b_k$  als Faltung des Messvektors  $y(n)$  mit der Impulsantwort  $g_k(n)$  eines Schätzfilters interpretiert:

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

- 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 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:

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

- Der Gauß-Markov-Schätzer ist **nicht erwartungstreu** (vgl. Folie 6):

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

### Minimum-Varianz-Schätzer

- Sonderfall** bei unbekannter Statistik des Parametervektors  $\mathbf{b}$
- Annahme einer Gleichverteilung über unendlich großem Intervall  
→ Varianzen gehen gegen  $\infty$  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

## 8.4 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:

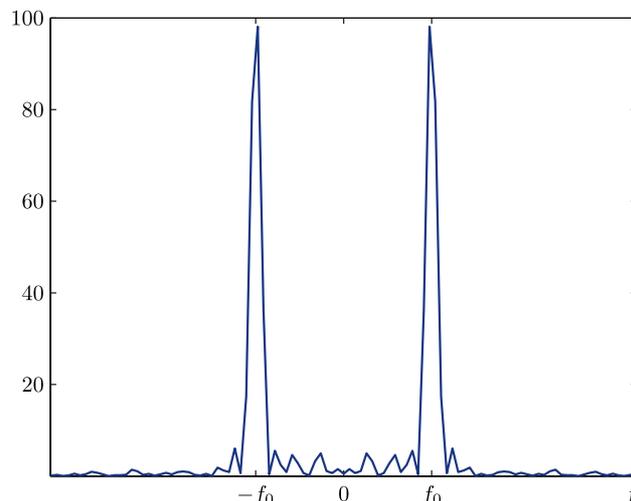
$$\lim_{n \rightarrow \infty} \mathbf{P}(n) = \mathbf{0}, \quad \lim_{n \rightarrow \infty} \mathbf{k}(n) = \mathbf{0}$$

### 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  $\varphi$
- Bestimmung der Frequenz  $f_0$  vorab aus dem Periodogramm  $\hat{S}_{yy}(f)$



- **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:

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

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

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

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

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

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

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

- 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:

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

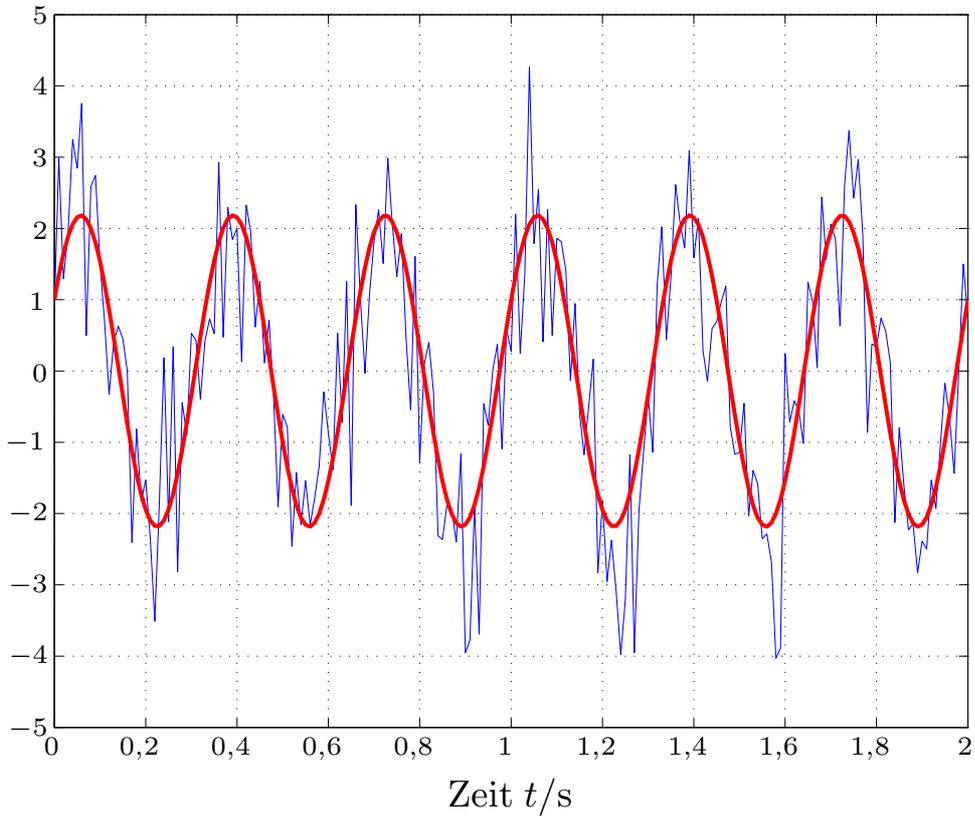
- Die letzten drei Gleichungen müssen in jedem Iterationsschritt erneut gelöst werden
- Der Schätzwert nähert sich zunehmend dem wahren Wert  $\mathbf{b}$  an
- Bei einer iterativen Schätzung (200 Messwerte;  $t_A = 10$  ms) erhält man:

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

## 8.4.1 Ableitung aus dem Least-Squares-Schätzer

Verrauschtes  
Signal

Geschätztes  
Signal

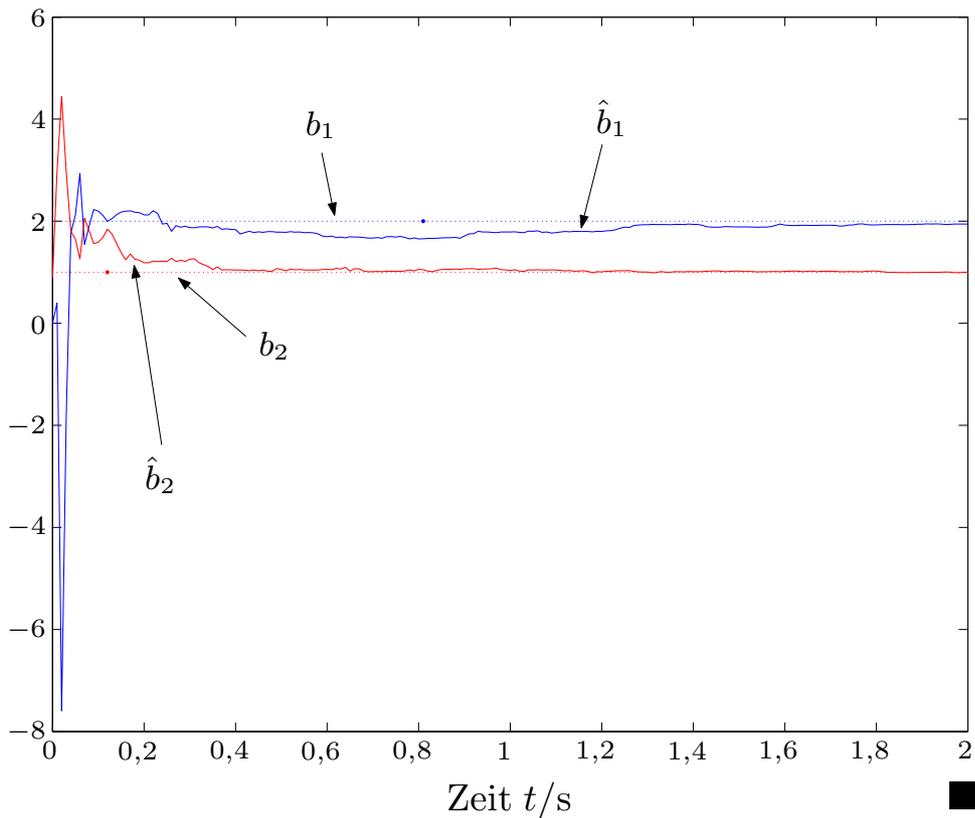


23

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

## 8.4.1 Ableitung aus dem Least-Squares-Schätzer

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



24

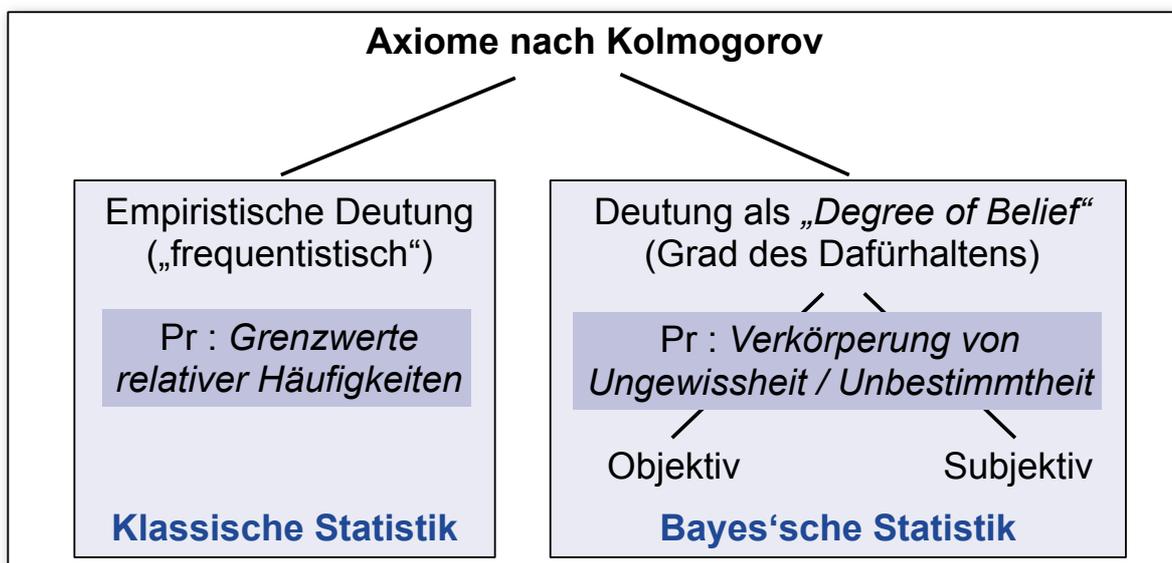
Prof. Dr.-Ing. F. Puente León – Methoden der Signalverarbeitung

# 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)

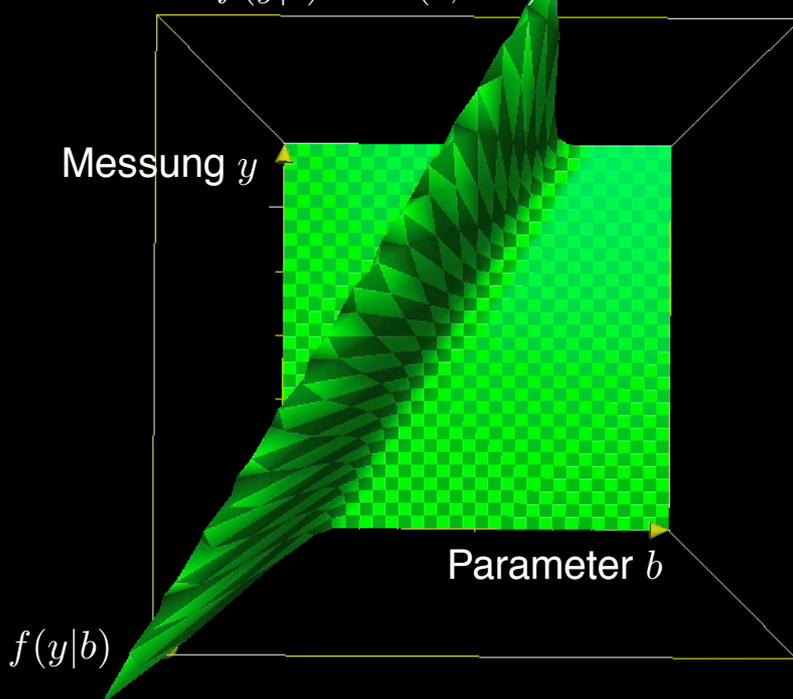
Was bedeutet Wahrscheinlichkeit? — zwei statistische „Denklager“



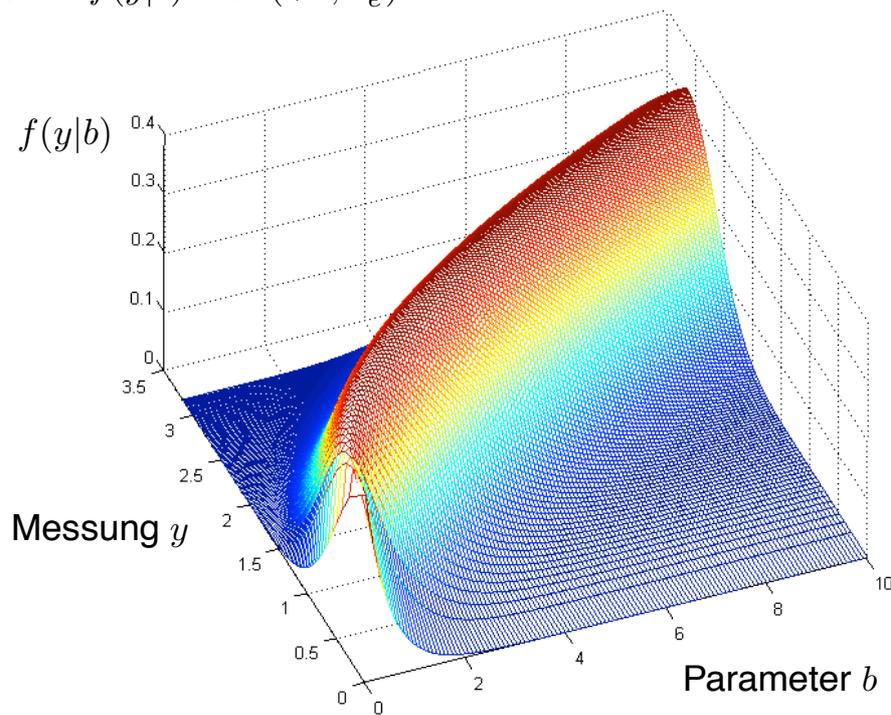
### Klassische Statistik

- Wahrscheinlichkeiten als **relative Häufigkeiten** interpretiert
- Zufallsgrößen: Ergebnis der Wiederholung eines Zufallsexperimentes
  - Messungen  $y$  werden als zufällig angesehen
  - Parametervektor  $\mathbf{b}$  wird als **deterministisch** (konstant) 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:

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

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

### 8.6.1 Bayes'sche Parameterschätzung

- Im Folgenden werden die wichtigsten Techniken zur **Bayes'schen Parameterschätzung** gezeigt
- Ansatz wie bei GM, 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  $\Delta$  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
- Außerdem kann die A-priori-Dichte des Parametervektors  $\mathbf{b}$  oft nicht genau angegeben werden

## 8.6.2 Maximum-Likelihood-Schätzer

- 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:

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

- Oft 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)**:

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

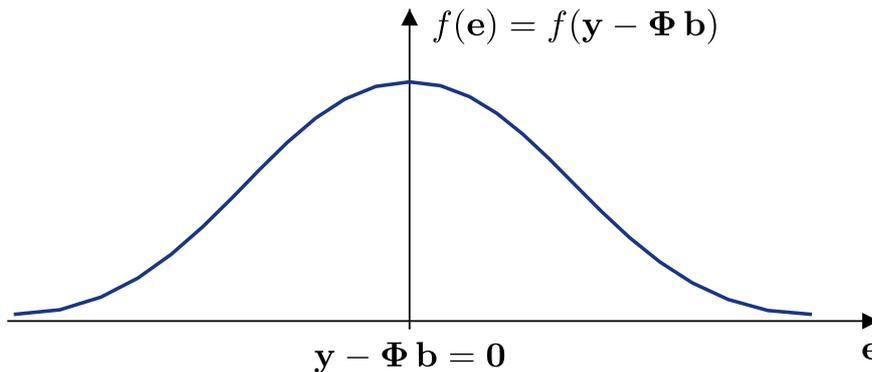
### Beispiel: MAP- und ML-Schätzer bei Normalverteilung

- 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:

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



## 8.6.2 Maximum-Likelihood-Schätzer

### 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:

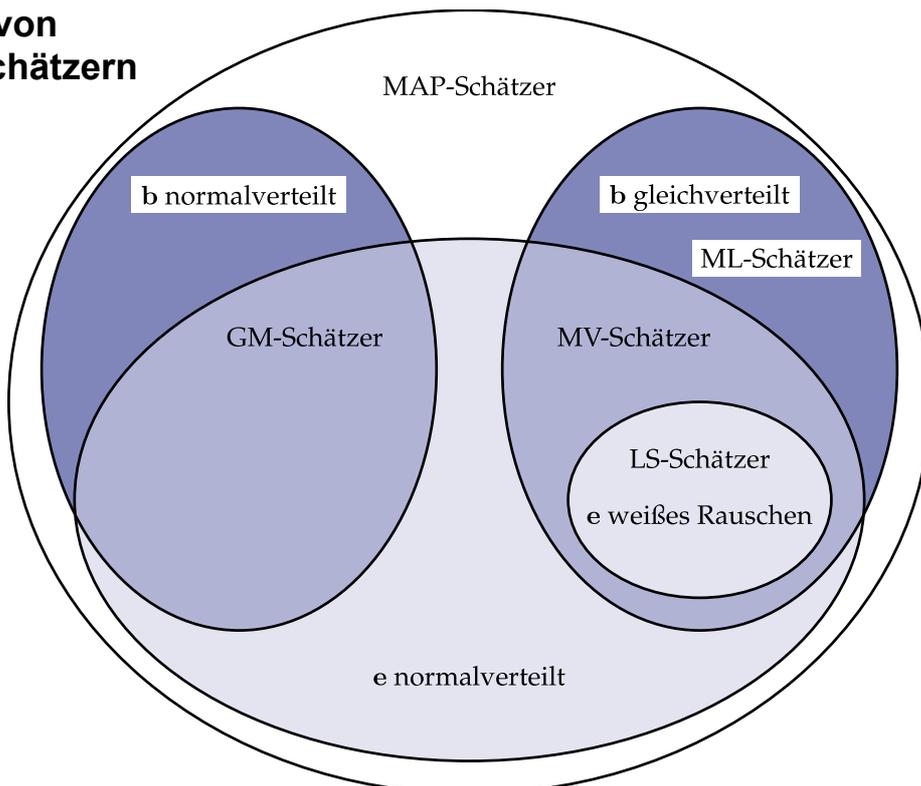
$$\frac{\partial \ln f(\mathbf{y}, \mathbf{b})}{\partial \mathbf{b}} = \frac{\partial \ln f(\mathbf{y}|\mathbf{b})}{\partial \mathbf{b}} = \Phi^T \mathbf{C}_{ee}^{-1} (\mathbf{y} - \Phi \mathbf{b}) \stackrel{!}{=} \mathbf{0}$$

- Schätzung des Parametervektors:

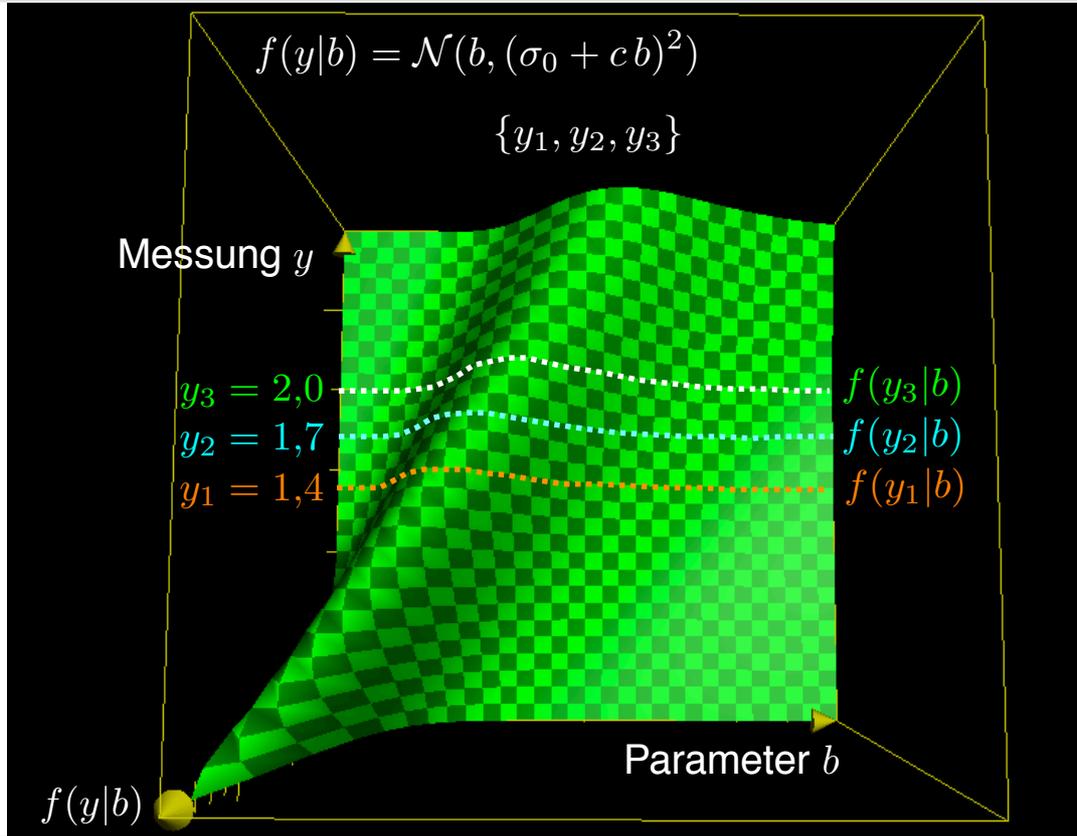
$$\hat{\mathbf{b}} = (\Phi^T \mathbf{C}_{ee}^{-1} \Phi)^{-1} \Phi^T \mathbf{C}_{ee}^{-1} \mathbf{y} \quad (\text{MV-Schätzer})$$

- Bei normalverteilten Größen entspricht
  - der MAP-Schätzer dem Gauß-Markov-Schätzer
  - der ML-Schätzer dem Minimum-Varianz-Schätzer

### Klassen von Bayes-Schätzern

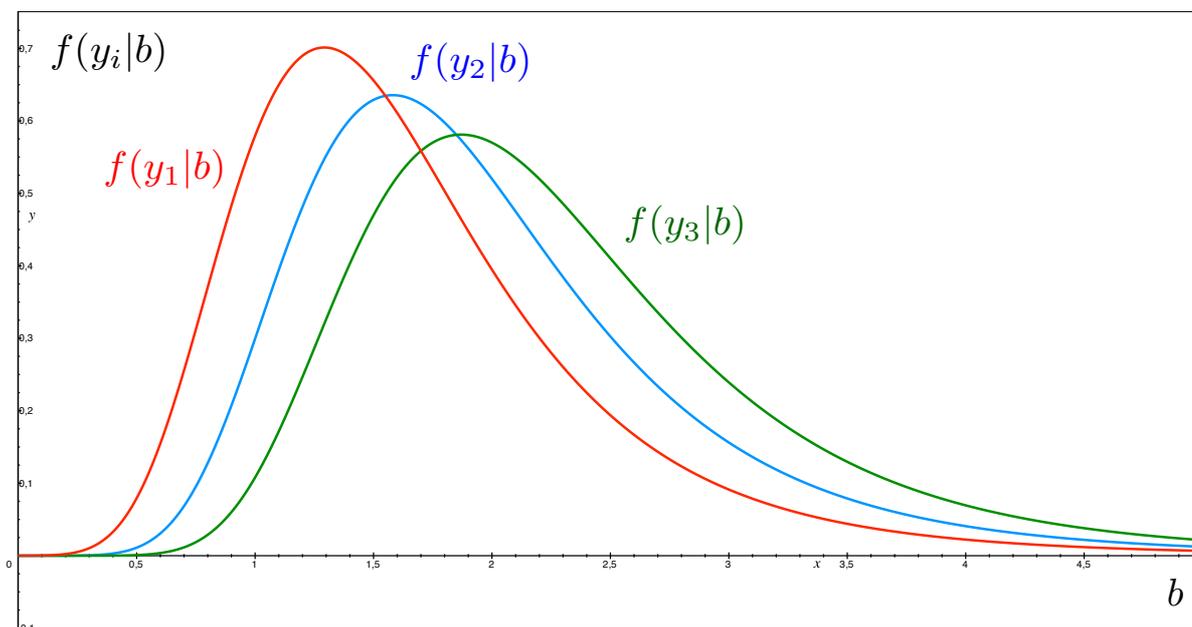


Bsp.

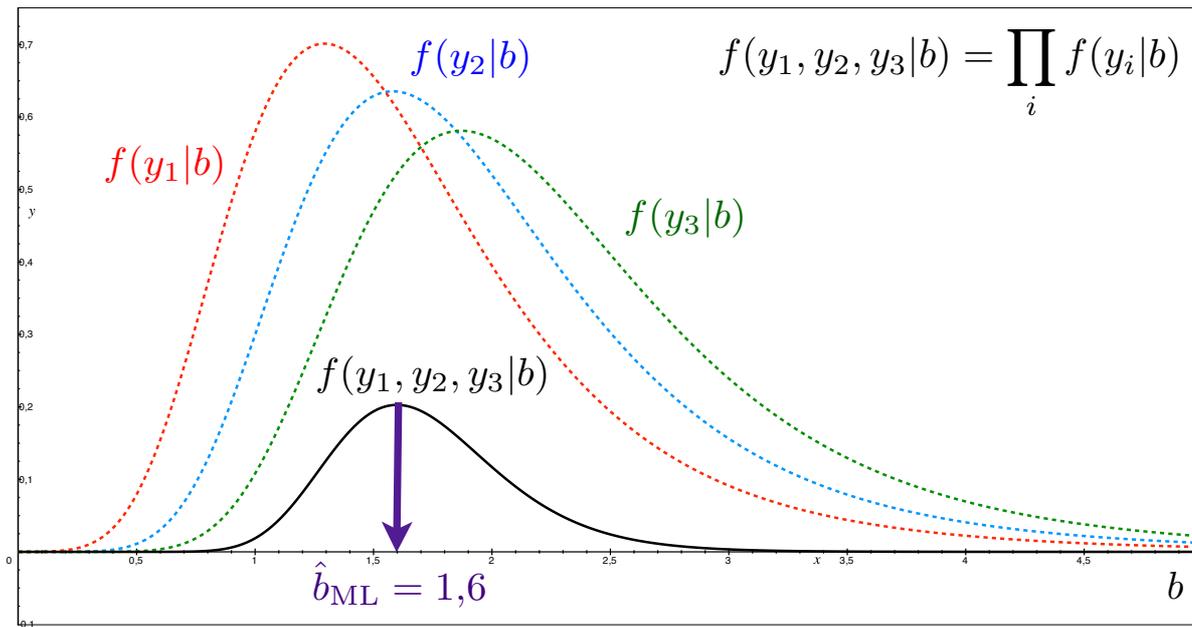


## Fusion der einzelnen Likelihood-Funktionen

- Annahme: statistisch unabhängige Messungen



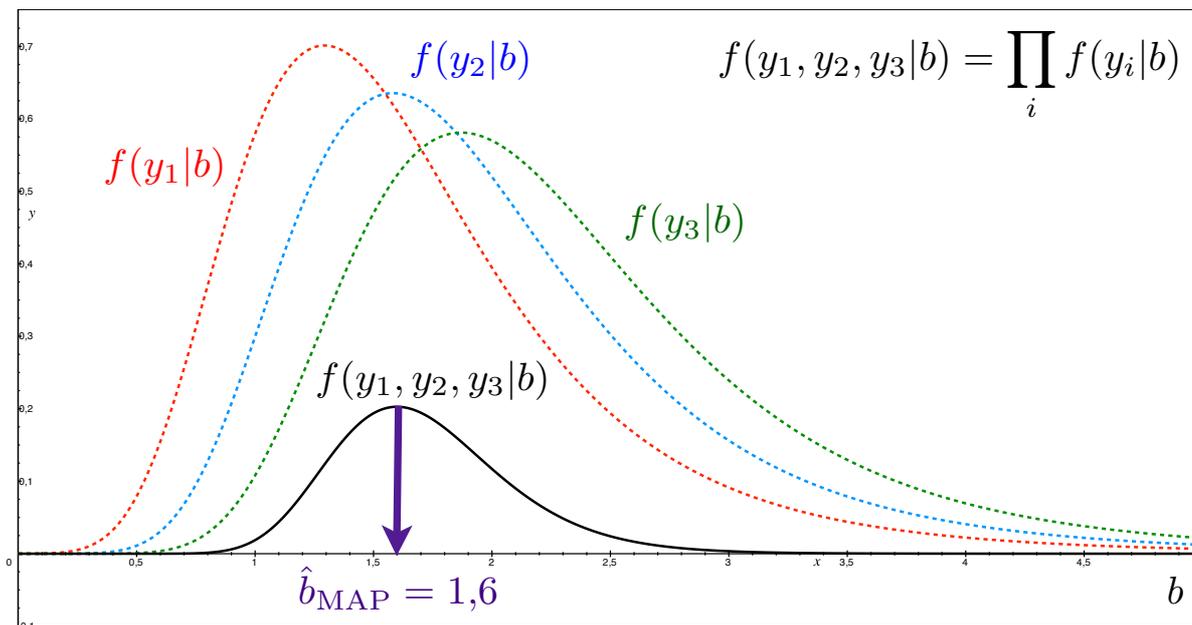
## ML-Schätzer



# 8.6.2 Maximum-Likelihood-Schätzer

## MAP-Schätzer bei vagem A-priori-Wissen

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



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

■ Annahme:  $f(b) = \text{const.}$

